NONPARAMETRIC ESTIMATION IN A NONLINEAR COINTEGRATION TYPE MODEL

By Hans Arnfinn Karlsen, Terje Myklebust and Dag Tjøstheim

University of Bergen

Department of Mathematics, Johannes Bruns gt.12, 5008 Bergen, Norway

6th March 2000

Abstract

We derive an asymptotic theory of nonparametric estimation for a nonlinear transfer function model $Z_t = f(X_t)+W_t$, where $\{X_t\}$ and $\{Z_t\}$ are observed nonstationary processes, and $\{W_t\}$ is a stationary process. In econometrics this can be interpreted as a nonlinear cointegration type relationship, but we believe that our result have wider interest. The class of nonstationary processes allowed for $\{X_t\}$ is a subclass of the class of null recurrent Markov chains. This subclass contains the random walk model and the unit root processes. We derive the asymptotics of a nonparametric estimate of $f(x)$ under two alternative sets of assumptions on $\{W_t\}$: i) $\{W_t\}$ is a linear process. ii) $\{W_t\}$ is a Markov chain satisfying some mixing conditions. The latter requires considerably more work but also holds larger promise for further developments. The finite sample properties of $\widehat{f}(x)$ are studied via a set of simulation experiments.

Key words and phrases. Cointegration, Nonstationary time series models, null recurrent Markov chain, nonparametric kernel estimators, transfer function model.

AMS 1991 subject classification. Primary 62M10, 62G07; secondary 60J05.

1 Introduction

Two time series $\{X_t\}$ and $\{Z_t\}$ are said to be linearly cointegrated if they are both nonstationary (persistent) of unit root type, and if there exists a linear combination $aX_t + bZ_t = W_t$ such that $\{W_t\}$ is stationary. This means that the series $\{X_t, Z_t\}$ move together in the long run. The concept of cointegration was introduced by Granger and further developed by Engle and Granger (1987), and since its introduction there have been numerous papers in econometrics exploring various aspects of it. Some main result are given in Johansen (1996).

The long term relationship between two econometric time series may not necessarily be linear, however, and the processes $\{X_t\}$ and $\{Z_t\}$ may not be linearly generated unit root processes. This has lead to a search for nonlinear cointegration type relationship such as $Z_t = f(X_t) + W_t$ for some nonlinear function f and some possibly nonlinearly generated input process $\{X_t\}$. Indeed, functional relationship of this type have been fitted to economic data (see e.g. Granger and Hallman 1991 and Aparicio and Escribano 1997), but to our knowledge the properties of the resulting nonparametric estimates have not been established. (See Xia 1998 for a consistency property in a simplified situation, however).

There are at least two difficulties (cf. Granger 1995 and others): which class of processes should be chosen as a basic class of persistent processes, and how should an estimation theory for an estimate of f be constructed? The main goal of this paper is to try to one answer these questions; i.e. we wish to establish a nonparametric estimation theory for the function f in the nonlinear transfer function model

$$Z_t = f(X_t) + W_t \tag{1.1}$$

where $\{W_t\}$ is a non-observed stationary process, and $\{X_t\}$ and $\{Z_t\}$ are observed processes which are nonstationary in a sense to be made more precise. The connection to the nonlinear cointegration problem is obvious, but we would like to point out that the estimation of the transfer function f in the general context we are considering should also be of interest in other areas. In a traditional transfer function problem some sort of mixing condition is often assumed for $\{X_t\}$ to obtain a central limit theorem for $\widehat{f}(x)$. However, mixing assumptions on $\{X_t\}$ are ruled out in the general situation we look at. A minimal condition for doing asymptotic analysis for $\widehat{f}(x)$ is that, as the number of observations on $\{X_t\}$ increases, there must be infinitely many observations in any neighborhood of x. This means $\{X_t\}$ must return to a neighborhood of x infinitely often, which in turn implies that the framework of a recurrent Markov chain is especially convenient. Since $\{X_t\}$ may be nonstationary, null recurrent processes have to be included, and it should be noted that the class of null recurrent processes contain unit root processes (cf. Myklebust et al 2000). Unlike the parametric situation, where a unit root speeds up the convergence of (global) estimates due to the large spread of the observations, in the nonparametric case, which is concerned with local estimates, the nonstationarity slows down the convergence, because the time until the process returns to the local neighborhood increases. (The expected time being infinite in the null recurrent case).

In Karlsen and Tjøstheim (1998) (hereafter referred to as KT) we developed an asymptotic theory for nonparametric estimation for a nonstationary univariate nonlinear autoregressive model in the framework of so-called β-null recurrent processes. This is a subclass of the null recurrent processes which contains the random walk. For an alternative theoretical approach in the random walk case we refer to Phillips and Park (1998).

We will rely on central parts of the theory of Karlsen and Tjøstheim (1998) for our derivations in this paper. But a host of new problems emerges in the transfer function case as will be made clear in the following. We also present several examples of the theory. In particular it will be seen that the verification of asymptotic theory via simulation on null recurrent series presents some special difficulties, which are absent in the ordinary stationary mixing situation.

2 Notation and some basic conditions

We will follow the notation of KT since our proofs and results will be closely based on that paper. Thus, we denote by $\{X_t, t \geq 0\}$ a ϕ-irreducible Markov chain on a general state space $(E, \mathcal{E})$ with transition probability P. In this paper we take $E \subset R$, and we denote the class of non-negative measurable functions with ϕ-positive support by $\mathcal{E}^+$. For a set $A \in \mathcal{E}$ we write $A \in \mathcal{E}^+$ if the indicator function $1_A \in \mathcal{E}^+$. The process $\{X_t, t \geq 0\}$ will be assumed to be Harris recurrent. This essentially implies that given a neighborhood $\mathcal{N}_x$ of x with $\phi(\mathcal{N}_x) > 0$, $\{X_t\}$ will return to $\mathcal{N}_x$ with probability one, and this is what makes asymptotics for a nonparametric estimation possible. The chain is positive recurrent if there exists an invariant probability measure such that $\{X_t \geq 0\}$ is strictly stationary, and it is null recurrent otherwise. In this paper we are primarily interested in the null recurrent situation, in which case there exists a (unique up to a constant, non-probability) invariant measure, which will be denoted by π.

We assume aperiodicity and that the transition probability for the X_t-process satisfies the so-called minorization inequality

$$P \geq s \otimes \nu \tag{2.1}$$

where s is a small function and ν a small measure with $\nu(E) = 1$. The pair (s, ν) is called an atom. This is not a strong restriction (cf. KT and Nummelin, 1984). Moreover, the minorization condition may be relaxed. Clearly, (2.1) implies that $0 \leq s(x) \leq 1$ for all $x \in E$.

If a set $C \in \mathcal{E}^+$ is such that 1_C is a small function, C itself is said to be small. Under quite wide conditions (cf. KT) a compact set will be small. From (2.1) we get the identity

$$\begin{aligned} P(x,A) &= (1-s(x))\left\{\frac{P(x,A)-s(x)\nu(A)}{1-s(x)}\right\}1(s(x)<1) + 1_A(x)1(s(x)=1) \\ &\qquad + s(x)\nu(A) \\ &= (1-s(x))Q(x,A) + s(x)\nu(A) \end{aligned}$$

so that the transition probability P can be thought of as a mixture of the transition probability Q and the small measure ν. Since ν is independent of x this means that the chain regenerates each time ν is chosen. This occurs with probability $s(x)$. The reasoning can be formalized by introducing the split chain $\{X_t, Y_t\}$ where the auxiliary chain, $\{Y_t\}$, can only take values 0 and 1. Given that $X_t = x$, $Y_{t-1} = y_{t-1}$, Y_t takes the value one with probability $s(x)$ so that $\alpha = E \times \{1\}$ is a proper atom (cf. Nummelin, 1984, p 51) for the split chain. We denote by

$$S_\alpha = \min\{t \geq 1\colon\ Y_t = 1\}$$

the corresponding recurrence time. We will also make use of the consecutive sequence of recurrence times starting at time $t = 0$;

$$\tau_k = \min\{t > \tau_{k-1}\colon\ Y_t = 1\}, \quad \tau_{-1} \stackrel{\text{def}}{=} -1 \quad \text{for } k \geq 0, \quad \tau = \tau_\alpha = \tau_0 \tag{2.2}$$

and the number of regenerations in the time interval $[0, n]$, i.e.,

$$T(n) = \max_k\{k\colon\ \tau_k \leq n\} \vee 0\ .$$

An invariant measure π_s can be defined in terms of the atom (s, ν) of (2.1). In fact (KT , Section 2.3)

$$\pi_s \stackrel{\text{def}}{=} \nu G_{s,\nu}, \quad G_{s,\nu} \stackrel{\text{def}}{=} \sum_{\ell=0}^{\infty} (P - s \otimes \nu)^\ell\ . \tag{2.3}$$

If the measure π_s is absolutely continuous with respect to Lebesgue measure, we denote by p_s the corresponding density so that $p_s(x)dx = \pi_s(dx)$. Similarly, we define the density $p_C(x) = p_s(x)/\pi_s(C)$. For a π-integrable function g on R we use the notation $\pi_s g$ for

$$\pi_s g = \pi_s(g) = \int g(x) \pi_s(dx)$$

Corresponding to $T(n)$ for a set $C \in \mathcal{E}^+$, the number of times $\{X_t\}$ is visiting C up to time n is denoted by

$$T_C(n) = \sum_{t=0}^{n} 1_C(X_t)$$

From KT (Remark 2.7) we have that $\lim_{n\to\infty} T_C(n)/T(n) = \pi_s 1_C$.

The kernel $G_{s,\nu}$ of (2.3) plays an important role in Section 4 and it easily follows from the above that for a π_s-integrable g defined on E,

$$\mathrm{E}_x \sum_{t=0}^{\tau} g(X_t) = G_{s,\nu} g(x)\ . \tag{2.4}$$

The minorization condition and the accompanying split chain permit decomposing the chain into separate and identical parts defined by the regeneration points. We have for a function g,

$$S_n(g) \stackrel{\text{def}}{=} \sum_{t=0}^{n} g(X_t) = U_0 + \sum_{k=1}^{T(n)} U_k + U_{(n)} \tag{2.5}$$

where

$$U_k = \begin{cases} \sum_{t=\tau_{k-1}+1}^{\tau_k} g(X_t), & \text{when } k \geq 0; \\ \sum_{t=\tau_{T(n)}+1}^{n} g(X_t), & \text{when } k = (n) \end{cases} \tag{2.6}$$

The sequence $\{(U_k, (\tau_k - \tau_{k-1})),\ k \geq 1\}$ consists of independent identically distributed (iid) random variables. This partition of the chain is basic for the subsequent asymptotic analysis.

We have to make a restriction on the way the process regenerates: the chain $\{X_t\}$ is β-null recurrent if there exists a small non-negative function h, an initial measure λ, a constant $\beta \in (0,1)$ and a slowly varying function L_h so that

$$\mathrm{E}_\lambda\Big[\sum_{t=0}^{n} h(X_t)\Big] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_h(n) \tag{2.7}$$

as $n \to \infty$. This condition is equivalent with (cf. KT) a restriction on the tail distribution of the recurrence time S_α in that

$$\mathrm{P}_\alpha(S_\alpha > n) = \frac{1}{\Gamma(1-\beta) n^\beta L_s(n)}\big(1 + o(1)\big) \tag{2.8}$$

where L_s is a slowly varying function depending on s. In the sequel (2.8) will be referred to as the tail condition.

A random walk process is β-null recurrent with $\beta = 1/2$.

2.1 Basic conditions

In all of the proofs we use $c_1, c_2, \ldots$ as a sequence of generic constants in our proofs, and if $\{a_n\}$ and $\{b_n\}$ are two real-valued strictly positive sequences, then we write $a_n \ll b_n$ if $a_n = o(b_n)$.

If η is a non-negative measurable function and λ is a measure, then the kernel $\eta \otimes \lambda$ is defined by

$$\eta \otimes \lambda(x, A) = \eta(x)\lambda(A), \quad (x, A) \in (E, \mathcal{E}) .$$

If H is a general kernel, the function $H\eta$, the measure λH and the number $\lambda H \eta$ are defined by

$$H\eta(x) = \int H(x, dy)\eta(y), \quad \lambda H(A) = \int \lambda(dx) H(x, A), \quad \lambda H \eta = \int \lambda(dx) H(x, dy)\eta(y) .$$

The convolution of two kernels H_1 and H_2 gives another kernel defined by

$$H_1 H_2(x, A) = \int H_1(x, dy) H_2(y, A) .$$

Due to associative laws the number $\lambda H_1 H_2 \eta$ is uniquely defined. If $A \in \mathcal{E}$ and 1_A is the corresponding indicator variable, then $H 1_A(x) = H(x, A)$. The kernel I_η is defined by $I_\eta(x, A) = \eta(x) 1_A(x)$.

We denote by $h = h_n$ the bandwidth used in the nonparametric estimation. It is assumed to satisfy $h_n \to 0$, and with no loss of generality we also assume $h_n \leq 1$. Let $K : R \to R$ be a kernel function, and for a fixed x, let $K_{x,h}(y) = h^{-1}K((y-x)/h)$, $\mathcal{N}_x(h) = \{y : K_{x,h}(y) \neq 0\}$ and $\mathcal{N}_x = \mathcal{N}_x(1)$. In our context a locally bounded function will be taken to mean bounded in a neighborhood of x, and a locally continuous function is continuous at the point x. Without loss of generality we may assume that this neighborhood equals $\mathcal{N}_x$, and that local continuity implies local boundness. This is since $\mathcal{N}_x(h) = x \oplus h\mathcal{N}_0$.

In the analysis of a kernel estimator of the transfer function f of (1.1) we will look at a slight generalization in that we allow an instantaneous transformation of W_t as well, resulting in

$$Z_t = f(X_t) + g_W(W_t)\ . \tag{2.9}$$

Here, g_W is assumed to be known, e.g. $g_W(w) \equiv w$, and f is to be estimated. We will consider this problem under two sets of conditions on $\{W_t\}$. In Section 3 $\{W_t\}$ will essentially be assumed to be a linear process, whereas a Markov assumption will be adopted in Section 4.

The following set of conditions is always assumed:

K_0: The kernel K is non-negative, $\int K(u)du < \infty$ and $\int K^2(u)du < \infty$.
P_0: The $\{X_t\}$ process is a Harris recurrent Markov chain.
F_0: The transfer function f is continuous at the point x.

Subsets of the conditions listed below are used according to need. The concepts of ϕ-mixing and α-mixing are defined in Hall & Heyde (1980, p. 277).

K_1: The support $\mathcal{N}_0$ of the kernel is contained in a compact set.
K_2: The kernel is bounded and $\mathcal{N}_x$ is a small set.
K_3: The kernel is normalized so that $\int K(u)du = 1$.
K_4: The kernel satisfies $\int uK(u)du = 0$.
P_1: The invariant measure π_s has a locally bounded density p_s.
P_2: The density p_s is locally continuous.
P_3: The density p_s possesses locally continuous partial derivatives of a given specified order.
P_4: The density is locally strictly positive, i.e., $\underline{\lim}_{y\to x}\, p_s(y) > 0$.
P_5: For all $h > 0$, the function: $y \leadsto P(y, \mathcal{N}_x(h))$ is continuous
P_6: The tail condition, (2.8), holds.
P_6': $\beta \overset{\text{def}}{=} \sup\{p \geq 0\text{: } \mathrm{E}_\alpha S_\alpha^p < \infty\} \wedge 1 > 0$.
T_1: The set C in T_C is a small set.
F_1: The function f possesses locally continuous partial derivatives of a given order.
W_1: The process $\{W_t\}$ is strictly stationary, $\mathrm{E}g_W(W_0) = 0$, $\mathrm{E}g_W^4(W_0) < \infty$, and ϕ-mixing with a finite mixing rate; $\sum_\ell \phi_\ell^{1/2} < \infty$.
W_2: The process $\{W_t\}$ is a strictly stationary linear process, $W_t = \sum_k a_k e_{t-k}$, with $\sum_k |a_k| < \infty$, $\{e_t\}$ satisfies W_1 and $g_W(w) \equiv w$.

W$_3$: The process $\{W_t\}$ is an irreducible ergodic Markov chain which satisfies (2.1).

W$_4$: In addition to W$_3$, $\{W_t\}$ is strongly α-mixing with mixing rate
$\sum_\ell \ell^{2m-1+\frac{2m}{k}}\alpha_\ell < \infty$, $\pi g_W = 0$ and $\pi g_W^{2m(k+1)}$ is finite for some $k, m \geq 1$, where π is the unique stationary measure of $\{W_t\}$.

3 Nonparametric estimation of f when $\{W_t\}$ is linear

The kernel estimate of f in (2.9) is given by

$$\widehat{f}(x) = \frac{\sum_{t=0}^n K_{x,h}(X_t)Z_t}{\sum_{t=0}^n K_{x,h}(X_t)} . \tag{3.1}$$

We get a central limit theorem by an adaption of Th. 4.4 in KT and its proof.

Theorem 3.1 *Assume that the processes X and W are independent in* (2.9) *and that* $\{X_t,\ t \geq 0\}$ *is started with an arbitrary initial measure* λ. *Moreover, assume that* K$_1$-K$_4$, P$_1$, P$_2$, P$_4$-P$_6$ *and* W$_2$ *hold. Let* $\sigma_W^2 = \mathrm{E}g_W^2(W_0)$, $\epsilon > 0$ *and* $\psi_x(y) = f(y) - f(x)$. *If* $h_n^{-1} \ll n^{\beta-\epsilon}$ *with* β *defined in* (2.7), *then*

$$\Big\{h_n \sum_{t=0}^n K_{x,h_n}(X_t)\Big\}^{1/2}\Big\{\widehat{f}(x) - f(x) - \frac{\pi_s I_{K_{x,h_n}}\psi_x}{\pi_s K_{x,h_n}}\Big\} \xrightarrow[n]{d} \mathcal{N}(0, \sigma_W^2 \int K^2(u)du) . \tag{3.2}$$

If P$_3$ *holds with order 2 and* f *possesses continuous derivatives of second order, then the bias term* $\pi_s I_{K_{x,h_n}}\psi_x / \pi_s K_{x,h_n}$ *is negligible when* $h_n^{-1} \gg n^{\beta/5+\epsilon}$.

Proof: Denote the left hand side of (3.2) by Λ_{n,h_n}. Since $\widehat{f}(x)$ is a ratio, we get some additional bias terms.

Write,

$$\begin{aligned} Z_t =& \ \{Z_t - f(X_t)\} + (f(X_t) - f(x)) + f(x) \\ Z_t K_{x,h}(X_t) =& \ g_h(X_t, W_t) + \psi_x(X_t)K_{x,h}(X_t) + f(x)K_{x,h}(X_t) \end{aligned} \tag{3.3}$$

where $g_h(z,u) = g_W(u) \cdot K_{x,h}(z)$. In the S_n-notation of (2.5) this gives

$$\widehat{f}(x) - f(x) = S_n^{-1}(K_{x,h})\Big\{S_n(g_h) + S_n(\psi_x \cdot K_{x,h})\Big\} .$$

The last term on the right hand side represents the bias. It contains a stochastic quantity, and we want to replace it by a deterministic bias term. Let

$$a_h \stackrel{\text{def}}{=} \frac{\pi_s I_{K_{x,h}}\psi_x}{\pi_s I_{K_{x,h}} 1} .$$

Then

$$\begin{aligned} \widehat{f}(x) - f(x) - a_h =& \ S_n^{-1}(K_{x,h})\Big\{S_n(g_h) + S_n(\psi_x \cdot K_{x,h}) - a_h S_n(K_{x,h})\Big\} \\ =& \ S_n^{-1}(K_{x,h})\Big\{S_n(g_h) + S_n(b_h)\Big\}, \end{aligned}$$

where

$$b_h = I_{K_{x,h}}(\psi_x - a_h) .$$

We note that $\mu_{b_h} = \pi_s b_h = 0$. We can write

$$\Lambda_{n,h} = \Delta_{n,h} + \Delta'_{n,h}$$

where

$$\begin{aligned} \Delta_{n,h} &= S_n^{-1/2}(K_{x,h}) h^{1/2} S_n(g_h) \\ \Delta'_{n,h} &= \left\{\widehat{p}_C(x)\right\}^{-1/2} T_C^{-1/2}(n) h^{1/2} S_n(b_h) \\ \widehat{p}_C(x) &= T_C^{-1}(n) S_n(K_{x,h_n}), \end{aligned}$$

and where C is a purely auxiliary set chosen such that T_1 holds. In the proof of Th. 4.4. in KT we replace $P\xi$ with f, then using K_1, K_2, P_2 and P_6,

$$\Delta'_{n,h_n} = o_P(1) \tag{3.4}$$

and by the second part of Th. 4.3 (cf. also the proof of Th 4.4.) in KT ,

$$\widehat{p}_C(x) = p_C(x) + o_P(1) \tag{3.5}$$

using K_1, K_2, P_1, P_4, P_6.

It remains to verify that Δ_{n,h_n} satisfies the prescribed CLT. We assume first that W_1 holds and then generalize to W_2 towards the end of the proof.

Let $c_W(k) = \mathrm{Cov}(g_W(W_0), g_W(W_k))$, $\sigma^2 = c_W(0) \int K^2(u)du$ and

$$\xi_{h,t} = \frac{h^{1/2} K_{x,h}(X_t)}{\left[\sum_{j=0}^n K_{x,h}(X_j)\right]^{1/2}} = \frac{h^{1/2} K_{x,h}(X_t)}{S_n^{1/2}(K_{x,h})} .$$

Then

$$\Delta_{n,h} = \frac{h^{1/2} S_n(g_h)}{S_n^{1/2}(K_{x,h})} = \sum_{t=0}^n \xi_{h,t} g_W(W_t) . \tag{3.6}$$

Let $\mathcal{F}^X$ be the σ-algebra generated by $\{X_t\}$. It is sufficient to prove that the conditional distribution of Δ_{n,h_n} given $\mathcal{F}^X$ converges in distribution to $\mathcal{N}(0,\sigma^2)$ where σ^2 is independent of x. Indeed, let z be fixed. Then if

$$P(\Delta_{n,h_n} \le z \mid \mathcal{F}^X) \underset{n}{\longrightarrow} \Phi(z/\sigma) \quad \text{a.s. } [\mathrm{P}_\lambda] \tag{3.7}$$

it follows from

$$\mathrm{P}_\lambda(\Delta_{n,h_n} \le z) = \mathrm{E}_\lambda\Big[P(\Delta_{n,h_n} \le z \mid \mathcal{F}^X)\Big]$$

and by the dominated convergence theorem that

$$\lim_{n\to\infty} \mathrm{P}_\lambda(\Delta_{n,h_n} \le z) = \mathrm{E}_\lambda\Big[\lim_{n\to\infty} P(\Delta_{n,h_n} \le z \mid \mathcal{F}^X)\Big] = \Phi(z/\sigma) . \tag{3.8}$$

In order to prove (3.7) we first show that

$$\sigma^2_{n,h_n} \stackrel{\text{def}}{=} \mathrm{E}(\Delta^2_{n,h_n} \mid \mathcal{F}^X) \xrightarrow[n]{\text{plim}} \sigma^2 \tag{3.9}$$

and then verify that conditional on $\mathcal{F}^X$, $\Delta_{n,h_n}/\sigma_{n,h_n}$ has an asymptotic standard normal distribution.

By (3.6),

$$\begin{aligned}\sigma^2_{n,h} &= c_W(0)\sum_{t=0}^{n}\xi^2_{h,t} + \sum_{0<|k|\le n} c_W(k) \sum_{t=-k\vee 0}^{n\wedge n-k} \xi_{h,t}\xi_{h,t+k} \\ &= c_W(0)\eta_{n,h,0} + \sum_{0<|k|\le n} c_W(k)\eta_{n,h,k}, \quad \text{say} \end{aligned} \tag{3.10}$$

and by K_2

$$|\eta_{n,h,k}| \le \sum_{t=0}^{n} \xi^2_{h,t} \le \frac{h\sum_{t=0}^{n} K^2_{x,h}(X_t)}{\sum_{t=0}^{n} K^2_{x,h}(X_t)} \le \sup_u \{K(u)\}\ .$$

Hence by the dominated convergence theorem it is enough to prove that

$$\eta_{n,h_n,k} \overset{\text{plim}}{\underset{n}{\longrightarrow}} \begin{cases} 0, & \text{for all } k \ne 0; \\ \int K^2(u)du, & \text{when } k=0 \end{cases}$$

since W_1 implies that

$$\sum_k |c_W(k)| < \infty\ .$$

Let C be small. We have

$$\begin{aligned}\eta_{n,h_n,k} &= \{\widehat{p}_C(x)\}^{-1}\Big\{T_C^{-1}(n)\sum_{t=-k\vee 0}^{n\wedge n-k} h_n K_{x,h_n}(X_t)K_{x,h_n}(X_{t+k})\Big\} \\ &\le \{\widehat{p}_C(x)\}^{-1}\Big\{T_C^{-1}(n)\sum_{t=0}^{n} h_n K_{x,h_n}(X_t)K_{x,h_n}(X_{t+k})\Big\} \\ &\le c_1\{\widehat{p}_C(x)\}^{-1}\Big\{T_C^{-1}(n)\sum_{t=0}^{n} K_{x,h_n}(X_t)1_{\mathcal{N}_x(h_n)}(X_{t+k})\Big\}\end{aligned}$$

since by K_1-K_3,

$$K_{x,h}(u) \le h^{-1}c_1 1_{\mathcal{N}_x(h)}(u)\ .$$

By P_1 and P_3 we have that $P^k(x,\{x\}) = 0$ (cf. the proof of Lemma 4.2 in KT) for all $k \ge 1$. Hence

$$\lim_{h\downarrow 0} P^k(x,\mathcal{N}_x(h)) = 0\ . \tag{3.11}$$

Let h_0 be fixed and arbitrary and consider n so large that $h_n \le h_0$. Then

$$\eta_{n,h_n,k} \le c_1\{\widehat{p}_C(x)\}^{-1}\theta_{n,h_n,h_0,k} \tag{3.12}$$

where

$$\theta_{n,h_n,h_0,k} = T_C^{-1}(n)\sum_{t=0}^{n} K_{x,h_n}(X_t)1_{\mathcal{N}_x(h_0)}(X_{t+k})\ .$$

Then by extending the proof of Th. 4.2 in KT using K_1, K_2, P_1, P_5, P_6' (which is implied by P_6)

$$\overline{\lim}\,|\theta_{n,h,h_0,k}| \le c_2 P^k(x,\mathcal{N}_x(h_0)) \quad \text{a.s.} \tag{3.13}$$

where c_2 is a constant independent of h_0. By (3.11), (3.12) and (3.13) it follows that

$$\eta_{n,h_n,k} = o(1) \quad \text{a.s.} \tag{3.14}$$

when $k > 0$. Exactly the same arguments can be used to show that (3.14) holds for $k < 0$. When $k = 0$ it follows by KT (Th. 4.1) using K_1-K_3, P_1, P_2, P_6' and Bochners theorem,

$$\eta_{n,h,0} = \int K^2(u)du + o(1) \quad \text{a.s.} \quad .$$

Thus we have proved (3.9).

To prove (3.7) (and hence (3.8)) it remains to verify that a fourth order Liapunov condition holds. By W_1

$$\begin{aligned}
\sum_{t=0}^{n} \mathrm{E}(\xi_{h_n,t}^4\, g_W^4(W_t) \mid \mathcal{F}^X) \leq\ & c_1 \sum_{t=0}^{n} \xi_{h_n,t}^4 \\
=\ & c_1 \frac{h_n^2 \sum_{t=0}^n K_{x,h_n}^4(X_t)}{\left[\sum_{t=0}^n K_{x,h_n}(X_t)\right]^2} \\
=\ & c_0 \left\{ \frac{1}{h_n \sum_{t=0}^n K_{x,h_n}(X_t)} \right\} \left\{ \frac{\sum_{t=0}^n h_n^3 K_{x,h_n}^4(X_t)}{\sum_{t=0}^n K_{x,h_n}(X_t)} \right\} \\
=\ & o(1) \quad \text{a.s.}
\end{aligned}$$

since

$$\frac{\sum_{t=0}^n h_n^3 K_{x,h_n}^4(X_t)}{\sum_{t=0}^n K_{x,h_n}(X_t)} \leq \{\sup_u |K(u)|^3\}$$

and

$$\Big\{T_C(n)h_n\Big\}\widehat{p}_C(x) = h_n \sum_{t=0}^{n} K_{h_n,t} \uparrow \infty \quad \text{a.s, because } \lim_{n\to\infty} T_C(n)h_n = \infty \text{ a.s .} \tag{3.15}$$

Thus the array $\{Z_{n,t}\}_{t=0}^n$ satisfies the conditions in a dependent central limit theorem (Bergstrøm, 1982, Th. 1., p 161) for given $\mathcal{F}^X$ where $Z_{n,t} \stackrel{\text{def}}{=} \xi_{h_n,t} g_W(W_t)/\sigma_{n,h_n}$.

It is a standard procedure to generalize from condition W_1 to condition W_2. The linear representation is truncated

$$W_t = W_{t,m} + W_t^{(m)}, \quad W_{t,m} = \sum_{|j|<m} a_j e_{t-j}, \quad m \geq 1,$$

so that

$$\Delta_{n,h} = \Delta_{n,h,m} + \Delta_{n,h}^{(m)} \ .$$

It is enough to prove that for all z,

$$\begin{aligned}
&\text{i): } \forall m \geq 1: \quad P(\Delta_{n,h_n,m} \leq z \mid \mathcal{F}^X) \underset{n}{\longrightarrow} \Phi(z/\sigma_m) \quad \text{a.s. } [\mathrm{P}_\lambda] \\
&\text{ii): } \quad \sigma_m^2 \underset{m}{\longrightarrow} \sigma^2 \\
&\text{iii): } \quad \overline{\lim_n}\, \mathrm{E}\Big[\Big(\Delta_{n,h_n}^{(m)}\Big)^2 \mid \mathcal{F}^X\Big] \underset{m}{\longrightarrow} 0 \quad \text{a.s. } [\mathrm{P}_\lambda]
\end{aligned}$$

(cf. Billingsley, 1968, Th. 4.2, p 25).

For all fixed m the process $W_{(m)} = \{W_{k,m},\ k \geq 0\}$ satisfies W_1. Hence part i) holds. Let $c_{W_{(m)}}$ denote the covariance function that belongs to $\{W_{k,m},\ k \geq 0\}$. Then since c_W and the sequence $a = \{a_k\}$ by assumption W_2 are summable, it follows that $\lim_m \sum_k |c_{W_{(m)}}(k) - c_W(k)| = 0$ which implies that part ii) holds by (3.10). Moreover, we have that

$$\mathrm{E}\Big[|\Delta_{n,h_n,m} - \Delta_{n,h_n}|^2 \mid \mathcal{F}^X\Big] = \mathrm{E}\Big[\Big(\Delta_{n,h_n}^{(m)}\Big)^2 \mid \mathcal{F}^X\Big] = \sum_{|k|<n} c_{W^{(m)}}(k)\eta_{n,k,h_n}\ .$$

By the above reasoning,

$$\sum_{|k|<n} c_{W^{(m)}}(k)\eta_{n,k,h_n} = c_{W^{(m)}}(0)\int K^2(u)du \quad \text{a.s.}$$

Since $\lim_{m\to\infty} c_{W^{(m)}}(0) = 0$, we have verified part iii).

The bias term is neglible if h_n goes to zero sufficiently fast. This is shown as in KT (proof of Th. 4.4) . □

The tail condition (2.8) is difficult to verify since it almost requires exact knowledge of the first order asymptotics of the rate of $P^n(x, A)$. We believe that P_6' is considerable weaker, and it may be easier to verify than P_6, since it is fulfilled if $\mathrm{E}_\nu \tau^p < \infty$ for some $p > 0$. It turns out that it is possible to incorporate this modification of the tail condition in the proof of Theorem 3.1: (Note that we use the notation $a_n \ll b_n$ to denote $a_n = o(b_n)$ as $n \to \infty$ for 2 sequences $\{a_n\}$ and $\{b_n\}$)

Corollary 3.1 *Assume that the conditions of Theorem* 3.1 *hold and that the tail condition,* P_6, *is replaced by* P_6'. *We also assume* F_1 *of order 1 and for some* $\epsilon, \delta > 0$,

$$n^\delta \ll h_n^{-1} \ll n^{\beta/2-\epsilon}\ . \tag{3.16}$$

Then the conclusion in Theorem 3.1 *still holds.*

Proof of Corollary 3.1: We have to prove that (3.4), (3.5) and (3.15) still hold when we replace P_6 by P_6'. By KT (Th. 4.1) (3.5) is true using K_1-K_3, P_1,P_2, P_6' and the right hand side of (3.16). By KT (Lemma 3.1, second part) ,

$$n^{\beta-\epsilon} \ll T(n)$$

which entails that (3.15) holds.

It remains to verify (3.4). It is enough to show that

$$T_C^{-1/2}(n)h_n^{1/2}S_n(b_{h_n}) \quad \xrightarrow[n]{\text{a.s.}} \quad 0$$

holds without the tail condition. Recall that

$$\psi_x(y) = f(y) - f(x), \quad a_h = \frac{\pi_s I_{K_{x,h}}\psi_x}{\pi_s I_{K_{x,h}}1}, \quad b_h(y) = K_{x,h}(y)(\psi_x(y) - a_h) \tag{3.17}$$

so that $\pi_s b_h = 0$. Smoothness of order 1 for the function f implies by (3.17)

$$|a_h| \le c_0 h, \quad |b_h| \le c_1 1_{\mathcal{N}_x(h)} \ .$$

In accordance with (2.5) and (2.6), let

$$S_n(b_h) = U_{0,h_n} + V_{T(n),h_n} + U_{(n),h_n}, \quad V_{T(n),h_n} = \sum_{k=1}^{T(n)} U_{k,h_n}$$

which implies that

$$|U_{k,h_n}| \le U^0_{k,h_n} \le U^0_{k,1}, \quad k \in \{0, 1, \ldots, \infty, \quad (n)\} \tag{3.18}$$

where the U^0_{k,h_n}'s are generated by $g_h = c_1 1_{\mathcal{N}_x(h)}$ instead of b_h. We find that

$$|U^0_{(n),1}| \le U^0_{T(n)+1,1} = T^{1/2}(n) F^{1/2}_{T(n)}$$

where $F_n \stackrel{\text{def}}{=} U^2_{n+1,1}/n$. Since $\mathrm{E}_\nu U^2_{k,1} < \infty$ then by the strong law of the large numbers $F_n = o(1)$ a.s. . Likewise we have that $U_{0,h_n} = o(T^{1/2}(n))$ a.s. . Let δ be defined by (3.16). Then

$$\begin{aligned} \left|T(n)^{-1/2} h_n^{1/2} V_{T(n),h_n}\right| &= \left|\{T^\delta(n) h_n\}^{1/2} T^{-p}(n) V_{T(n),h_n}\right|, \quad p = 1/2 + \delta/2 \\ &\le \{n^\delta h_n\} \left|T^{-p}(n) V_{T(n),h_n}\right| \end{aligned} \tag{3.19}$$

where, by assumption, $n^\delta h_n = o(1)$ a.s. . Rephrasing notation, we

$$T^{-p}(n) V_{T(n),h_n} = n_1^{-p} V_{n_1,h_{n_2}}, \quad (T(n), n) = (n_1, n_2) \ .$$

Let $a = \beta - \epsilon > 0$. Then $T(n) \gg n^a$ a.s. . Hence by (3.19), (3.4) follows if

$$\lim_{\substack{|n_1 n_2| \to \infty \\ n_1 \ge n_2^a}} n_1^{-p} V_{n_1,h_{n_2}} = 0 \quad \text{a.s.} \ . \tag{3.20}$$

By the Borel Cantelli lemma (3.20) is true if for some $m \ge 1$

$$\sum_{n_1=1}^{\infty} \sum_{n_2=1}^{n_1^{1/a}} \mathrm{E} \left| n_1^{-p} \sum_{k=1}^{n_1} U_{k,h_{n_2}} \right|^{2m} < \infty \ .$$

Let $m > \delta^{-1}(1 + a^{-1})$. Then by (3.18)

$$\mathrm{E} \left| n_1^{-p} \sum_{k=1}^{n_1} U_{k,h_{n_2}} \right|^{2m} = n_1^{-m\delta} \mathrm{E} \left| n_1^{-1/2} \sum_{k=1}^{n_1} U_{k,h_{n_2}} \right|^{2m} \le c_2 n_1^{-m\delta} \mathrm{E}_\nu \left| U^0_{0,1} \right|^{2m} \le c_3 n_1^{-m\delta}$$

where c_3 is finite since $\mathcal{N}_x$ is small (cf. KT , proof of Lemma 4.1).

□

Remark 3.1 *It is possible to allow for a stochastic bandwidth, i.e., h_n is replaced by $q_{T_C(n)}$ where the function q is independent of β (cf.* KT *, Th. 3.4).*

□

4 Nonparametric estimation of f when $\{W_t\}$ is Markov process

In this section the linear process assumption W_2 is replaced by the Markov assumption W_3. This also makes it possible to relax the ϕ-mixing of W_1 to α-mixing as in W_4. In the preceding section we did not directly use a central limit theorem for the X-process. But now we have to rely on a null recurrent central limit theorem, more precisely Theorem 4.1 based upon Corollary 3.2, of KT . However, because we cannot in general restrict g_W to be a small function (consider e.g., $g_W(w) \equiv w$), Lemma 4.1 and Lemma 4.2 of KT cannot be used and this complicates matters considerably.

Throughout Sections 4.3 to 4.6 we make the assumption that $\{X_t\}$ and $\{W_t\}$ are independent. However, we believe that it is possible to avoid this assumption, and that this type of generalization is easier to achieve in the framework of Markov theory.

Before embarking on a proof of the asymptotic properties of $\widehat{f}$ we need a series of lemmas on the regeneration structure of the compound Markov process $\{(X_t, W_t)\}$. We think that these results are of independent interest, and that they are potentially useful in other situations.

We denote by g a general measurable functions defined on R^2, g_X, g_W are measurable functions defined on R, and $g = g_X \otimes g_W$ means that $g(x,w) = g_X(x) g_W(w)$.

4.1 A nonparametric CLT for null recurrent processes

Assume that $\{X_t\}$ is a Markov chain which satisfies the minorization condition (2.1) and the tail condition (2.8). Let

$$V_{g_h} = \sum_{j=0}^{\tau} g_h(X_j), \quad \mu_{g_h} = \mathrm{E}_\nu V_{g_h}, \quad \sigma^2_{g_h} = \mathrm{E}_\nu V^2_{g_h} - \mu^2_{g_h}$$

where g_h is a real-valued function defined on E for all $h > 0$, and τ is defined in (2.2). Consider the following conditions where $-\infty < \mu, \mu' < \infty$, $0 < \sigma, \sigma' < \infty$, $v \in [0,1]$, $m \geq 2$, $\epsilon > 0$, $0 < d_m, d'_m < \infty$, β is defined in (2.8), λ is an initial measure and $h \downarrow 0$.

$$\mu_{g_h} = \mu + o(1), \quad \mu_{g_h} = \mu' + o(1) \ . \tag{4.1}$$

$$h\sigma^2_{g_h} = \sigma^2 + o(1) \ . \tag{4.2}$$

$$h\sigma^2_{|g_h|} = \sigma'^2 + o(1) \ . \tag{4.3}$$

$$\mathrm{E}_\nu |V_{g_h} - \mu_{g_h}|^{2m} \leq d_m h^{-2m+v} \ . \tag{4.4}$$

$$\mathrm{E}_\nu |V_{|g_h|} - \mu_{|g_h|}|^{2m} \leq d'_m h^{-2m+v} \ . \tag{4.5}$$

$$h_n^{-1} \ll n^{\beta\delta_m - \epsilon}, \quad \delta_m = \frac{m-1}{m-v} \ . \tag{4.6}$$

$$\exists g_0\text{: } h|g_h| \leq g_0 \quad \text{and } \mathrm{P}_\lambda(V_{g_0} < \infty). \tag{4.7}$$

The following theorem is essentially a translation of a CLT-result in KT . It will be used to prove the main CLT-results of the present paper.

Theorem 4.1 *Let C be a small set. Assume that (4.1) - (4.6) hold for an $m \geq 2$ and a $v \in [0,1]$. Then for any initial measure λ for X_0 so that (4.7) holds,*

$$h_n^{1/2} S_n^{-1/2}(1_C) S_n(g_{h_n}) \xrightarrow[n]{d} \mathcal{N}\left(0, \sigma^2 \pi_s^{-1} 1_C\right) .$$

Proof of Theorem 4.1: The proof is essentially based upon KT (Th. 3.2) . Since g_h is a function in one variable the conditions in that theorem simplifies. Moreover, (4.1) - (4.3) are conditions implied by the corresponding conditions in Th. 3.2. In the conditions (4.1) - (4.7) the quantity v is allowed to vary everywhere in the interval $[0,1]$ and that represent a minor extension of Th. 3.2. It is permitted by a trivial extension of the proof of Lemma 3.3 of KT .

□

Before we can put Theorem 4.1 to use we need to analyze the regeneration structure of $\{X_t, W_t\}$ closer. This is done in Sections 4.2 - 4.7. We then state our main result in Section 4.8.

4.2 Decomposition of $S_n(g)$

We assume that the compound chain, $\{(X_t, W_t)\}$ satisfies (2.1), so that it can be extended by the split chain method, with $\{(X_t, W_t, Y_t)\}$ being a split chain. Note that if $\{X_t\}$ and $\{W_t\}$ separately satisfy the minorization inequality condition (2.1), it is not obvious that the compound chain $\{(X_t, W_t)\}$ does. If $\{X_t\}$ and $\{W_t\}$ are independent, then it is trivial to verify (2.1) as is shown in the beginning of Section 4.3. Let

$$\eta_k = \inf\{t > \eta_{k-1} : Y_t = 1\}, \quad k \geq 0, \quad \eta_{-1} = -1 .$$

Then, the sequence $\{\eta_k\}$ represents the regeneration times for the compound process. The basic decomposition, (2.5), gives

$$S_n(g) = V_0 + \sum_{k=1}^{T(n)} V_k + V_{(n)}, \quad T(n) = \sup\{k: \quad \eta_k \leq n\} \vee 0 \tag{4.8}$$

where

$$V_k = \begin{cases} \sum_{t=\eta_{k-1}+1}^{\eta_k} g(X_t, W_t), & \text{for } k \geq 0 \\ \sum_{t=\eta_{T(n)}+1}^{n} g(X_t, W_t), & \text{for } k = (n). \end{cases}$$

According to the general theory (4.8) represents essentially a decomposition into independent variables where $\{(V_k, (\eta_k - \eta_{k-1})),\ k \geq 1\}$ are iid and where V_k has expectation μ_g, variance σ_g^2, i.e.

$$\mu_g = \mathrm{E}_\nu(V_0), \quad \sigma_g^2 = \mathrm{var}_\nu(V_0) = \mathrm{E}_\nu(V_0^2) - \mu_g^2$$

where ν refers to the compound chain $\{(X_t, W_t)\}$.

Our first problem is to find conditions which ensure that μ_g and σ_g^2 are finite. Again, by reference to general theory (cf. Appendix A, (A.10)) we have that, with s referring to the compound chain,

$$\mu_g = \pi_s g, \quad \sigma_g^2 = \pi_s(g^2) + 2\pi_s I_g H\, G_{s,\nu} g - \pi_s^2(g) \tag{4.9}$$

where

$$H = P - s \otimes \nu, \quad G_{s,\nu} = \sum_{j=0}^{\infty} H^j\ . \tag{4.10}$$

The conditions ensuring $\mu_g = 0$ and $\sigma_g^2 < \infty$ are not evident from (4.10) if we want to avoid the strong restriction that g_W is a small function. If $g_W(w) \equiv w$, requiring g_W to be small is equivalent to ϕ-mixing which is not satisfied for an autoregressive process, say. The problem is linked to the term $G_{s,\nu}$. In fact, we also need to show existence of higher moments and to verify conditions connected to the bandwidth as exemplified in (4.1) -(4.7).

In Sections 4.3 - 4.6, we make the assumption that $\{X_t\}$ and $\{W_t\}$ are independent.

4.3 β null recurrence for the compound process

Let P denote the transition probability for the Markov process $\{(X_t, W_t)\}$. We label quantities associated with the X-process by one and with the W-process by two. The transition probability P satisfies (2.1) when P_1 and P_2 do, since

$$P = P_1 \otimes P_2 \geq (s_1 \otimes s_2) \otimes (\nu_1 \otimes \nu_2) = s \otimes \nu\ .$$

Lemma 4.1 *Assume that $\{X_t\}$ and $\{W_t\}$ are independent and that the tail condition (2.8) holds for $\{X_t\}$. Then the compound process $\{(X_t, W_t)\}$ is β-null recurrent, i.e. the tail condition holds for the compound process.*

Proof: Let C_1 and C_2 be small sets and $\nu = \nu_1 \otimes \nu_2$.

$$\begin{aligned} \mathrm{E}_\nu \sum_{t=0}^{n} 1_{C_1}(X_t) 1_{C_2}(W_t) &= \sum_{t=0}^{n} \left[\nu_1 P_1^t 1_{C_1}\right]\left[\nu_2 P_2^t 1_{C_2}\right] \\ &= [\pi_2 1_{C_2}] \sum_{t=0}^{n} \left[\nu_1 P_1^t 1_{C_1}\right] + \sum_{t=0}^{n} \left[\nu_1 P_1^t 1_{C_1}\right] b_t \end{aligned}$$

where $b_t = \nu_2 P_2^t 1_{C_2} - \pi_2 1_{C_2}$ and where π_2 is the stationary measure for $\{W_t\}$. Since $\{W_t\}$ is ergodic $b_t = o(1)$. Since $\{X_t\}$ is β-null (cf. KT (Lemma 2.2 and (2.31))), we have that

$$\sum_{t=0}^{n} \Big[\nu_1 P_1^t 1_{C_1}\Big] = [\pi_{s_1} 1_{C_1}]\psi_1(n)(1+a_n), \quad \psi_1(n) = n^\beta L_{s_1}(n), \quad a_n = o(1) .$$

Let $\psi_M = \sup_{t\le M}\psi_1(t)$, $A = \sup_t |a_t|$, $B = \sup_t |b_t|$, $B^{(M)} = \sup_{t>M}|b_t|$. Then, for all $M > 0$,

$$\Big|\sum_{t=0}^{n} \Big[\nu_1 P_1^t 1_{C_1}\Big] b_t\Big| \le B\Big\{[\pi_{s_1} 1_{C_1}]\psi_M(1+A)\Big\} + B^{(M)}\Big\{[\pi_{s_1} 1_{C_1}]\psi_1(n)(1+|a_n|)\Big\}$$

so that

$$\begin{aligned} &\overline{\lim_n} \frac{\Big|\sum_{t=0}^{n} \Big[\nu_1 P_1^t 1_{C_1}\Big] b_t\Big|}{\psi_1(n)} \\ &\le \overline{\lim_M}\,\overline{\lim_n}\, B\Big\{\{\pi_{s_1} 1_{C_1}\}\frac{\psi_M}{\psi_1(n)}(1+A)\Big\} + \overline{\lim_M}\,\overline{\lim_n}\, B^{(M)}\Big\{\{\pi_{s_1} 1_{C_1}\}(1+|a_n|)\Big\} = 0 . \qquad (4.11)\end{aligned}$$

Hence, $\sum_{t=0}^{n} \Big[\nu_1 P_1^t 1_{C_1}\Big] b_t = d_n \psi_1(n)$, $\quad d_n = o(1)$, and therefore

$$\mathrm{E}_\nu\Big\{\sum_{t=0}^{n} 1_{C_1}(X_t) 1_{C_2}(W_t)\Big\} = [\pi_s 1_C]\psi(n)(1+d_n), \quad \psi(n) \overset{\text{def}}{=} \psi_1(n)\pi_2(s_2) .$$

□

4.4 Decomposition structure

In the following $H = P - s \otimes \nu$ and $H_j = P_j - s_j \otimes \nu_j$ for $j = 1, 2$.

We extend both chains with the split chain method so that we have $\{(X_t, Y_t^1)\}$ and $\{(W_t, Y_t^2)\}$. Due to independence $\{X_t, W_t, Y_t\}$ is the split chain for the compound process (X, W) where $Y_t = Y_t^1 Y_t^2$. Thus

$$\eta_k = \inf\{t > \eta_{k-1} \colon Y_t^1 = Y_t^2 = 1\}, \quad k \ge 0, \quad \eta_{-1} = -1 .$$

We look at the decomposition structure in a different way where we try to benefit from the marginal decomposition of the X-process, i.e., the regenerations defined by $\{\tau_k^1\}$,

$$\tau_k^1 = \inf\{t > \tau_{k-1}^1 \colon \quad Y_t^1 = 1\}, \quad k \ge 0, \quad \tau_{-1}^1 = -1 .$$

Let

$$U_j = U_{g,j} = \sum_{t=\tau_{j-1}^1+1}^{\tau_j^1} g(X_t, W_t), \quad s \ge 0 . \qquad (4.12)$$

We note that in general the U_j's are neither unconditionally nor conditionally independent. The simultaneous regeneration times are also marginal regeneration times, i.e.,

$$\{\eta_k,\ k \geq 0\} \subseteq \{\tau^1_j,\ j \geq 0\}$$

Let $\mathcal{T}_{-1} = -1$ and

$$\mathcal{T}_k = \inf\{k > \mathcal{T}^1_{k-1}\colon \quad Y^2_{\tau^1_k} = 1\}, \quad k \geq 0, \quad \mathcal{T} = \mathcal{T}_0 .$$

Then

$$\eta_k = \tau^1_{\mathcal{T}_k}, \quad k \geq 0$$

which gives

$$V_0 = V_g = \sum_{t=0}^{\tau^1_{\mathcal{T}}} g(X_t, W_t) = \sum_{j=0}^{\mathcal{T}} \sum_{t=\tau^1_{j-1}+1}^{\tau^1_j} g(X_t, W_t) = \sum_{j=0}^{\mathcal{T}} U_j \tag{4.13}$$

and in general

$$V_k = \sum_{j=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} U_j, \quad k \geq 0 .$$

The following lemma contains necessary information about the process $\{W_{\tau^1_k}, k \geq 0\}$

Lemma 4.2 *The process $\{W_{\tau^1_k}, k \geq 0\}$ is a Markov process with transition probability $\underset{\sim}{\mathrm{P}} = P_2 \Phi_{\nu_1}$ where $\Phi_{\nu_1} = \sum_{\ell=0}^{\infty} \{\nu_1 H_1^\ell s_1\} P_2^\ell$. Moreover,*

$$\underset{\sim}{\mathrm{P}} \geq \underset{\sim}{s} \otimes \underset{\sim}{\nu} \tag{4.14}$$

with

$$(\underset{\sim}{s}, \underset{\sim}{\nu}) = \Big(s_2,\ \nu_2 \Phi_{\nu_1}\Big) .$$

Let $\lambda = \lambda_1 \otimes \lambda_2$ be the initial measure for $\{(X_t, W_t)\}$. Let $\underset{\sim}{\mathrm{W}}^\star = \{(\underset{\sim}{\mathrm{W}}_t, \underset{\sim}{\mathrm{Y}}_t),\ t \geq 0\}$ be the split chain generated by $\underset{\sim}{\mathrm{P}}$ and $(\underset{\sim}{s}, \underset{\sim}{\nu})$ and let $W'_\tau = \{(W_{\tau^1_k}, Y^2_{\tau^1_k}),\ k \geq 0\}$. Then

$$W'_\tau \overset{\mathrm{d}}{=} \underset{\sim}{\mathrm{W}}^\star \tag{4.15}$$

when the initial measure for $\underset{\sim}{\mathrm{W}}_0$ is $\widetilde{\lambda} = \lambda_2 \Phi_{\lambda_1}$. In particular, let $\underset{\sim}{\mathcal{T}}$ denote the first regeneration time for $\underset{\sim}{\mathrm{W}}^\star$. Then the occupation time formula is given by

$$\mathrm{E}_\lambda \sum_{k=0}^{\mathcal{T}} 1_A(W_{\tau^1_k}) = \mathrm{E}_{\widetilde{\lambda}} \sum_{k=0}^{\underset{\sim}{\mathcal{T}}} 1_A(\underset{\sim}{\mathrm{W}}_k) = \begin{cases} \widetilde{\lambda} \underset{\sim}{\mathrm{G}}_{\underset{\sim}{s}, \underset{\sim}{\nu}} 1_A, & \text{in general;} \\ \pi_2 \underset{\sim}{\mathrm{G}}_{\underset{\sim}{s}, \underset{\sim}{\nu}} 1_A, & \text{if } \lambda_2 = \pi_2; \\ \pi_{s_2} 1_A, & \text{if } \lambda = \nu. \end{cases} \tag{4.16}$$

where $\underset{\sim}{\mathrm{G}}_{\underset{\sim}{s}, \underset{\sim}{\nu}} = \sum_{\ell=0}^{\infty} (\underset{\sim}{\mathrm{P}} - \underset{\sim}{s} \otimes \underset{\sim}{\nu})^\ell$.

Proof of Lemma 4.2: See Appendix B. □

The following result indicates that the rate of convergence of the transition probability towards the stationary measure is at least as good for the $\underset{\sim}{W}$-process as for the W-process.

Lemma 4.3 *Suppose that W is geometric ergodic. Then this is also true for $\underset{\sim}{W}$. If W is strongly mixing with mixing rate defined by $\alpha = \{\alpha_j\}$, then $\underset{\sim}{W}$ is strongly mixing with mixing rate $\underset{\sim}{\alpha}$ which is equal to or faster than α. In particular for $k \geq 0$*

$$\sum_{\ell=1}^{\infty} \ell^k \alpha_\ell < \infty \quad \Longrightarrow \quad \mathrm{E}_{\pi_2} \underset{\sim}{\mathcal{T}}^{k+1} < \infty \,. \tag{4.17}$$

Proof of Lemma 4.3: See Appendix B. □

4.5 Upper bounds for $\mathrm{E}|V_g|^m$

In this subsection we assume that (cf. (3.3))

$$g(x,w) = (g_X \otimes g_W)(x,w) = g_X(x) g_W(w) \,. \tag{4.18}$$

In the proof of the CLT we need the moments of V_0 of (4.13). The following theorem is our main result in this direction.

Theorem 4.2 *Let $m \geq 1$ and U_k is defined by (4.12) and (4.18), then*

$$\mathrm{E}_\nu \Big\{ \sum_{k=0}^{\mathcal{T}} |U_k|^m \Big\} \leq \Big\{ \pi_{s_2} |g_W|^m \Big\} \mathrm{E}_{\nu_1} \Big\{ U^m_{|g_X|} \Big\} \,. \tag{4.19}$$

We use the notation

$$\delta_j = \tau^1_j - \tau^1_{j-1}, \quad j \geq 0, \quad \mathcal{H}_j = \mathcal{F}^X_{\tau^1_j} \vee \mathcal{F}^{Y^1}_{\tau^1_j} \vee \mathcal{F}^W \vee \mathcal{F}^{Y^2} \,.$$

Then U_j is measurable $\mathcal{H}_j$ and $\{\mathcal{T}_0 \geq j\} \in \mathcal{H}_{j-1}$. By (4.13), $V_0 = \sum_{j=0}^{\infty} U_j 1(\mathcal{T} \geq j)$ and for $m \geq 1$,

$$\begin{aligned} \mathrm{E}_\lambda \Big[U^m_j 1(\mathcal{T} \geq j) \Big] &= \mathrm{E}_\lambda \Big[U^m_j 1(\mathcal{T} \geq j) \mid \mathcal{H}_{j-1} \Big] \\ &= \mathrm{E}_\lambda \Bigg[1(\mathcal{T} \geq j) \mathrm{E}_\lambda \Big[U^m_j \mid \mathcal{H}_{j-1} \Big] \Bigg] \,. \end{aligned}$$

The following technical result, which is the first step in the proof of Theorem 4.2, use the independence of X and W together with the regeneration property of X.

Lemma 4.4 [Decoupling]
Let $\lambda = \lambda_1 \otimes \lambda_2$. Let $j \geq 0$ be fixed and let $\{X'_t\}$ be an independent copy of $\{X_t\}$ so that $\{X'_t\}$ is independent of both $\{X_t\}$ and $\{W_t\}$. Let ξ_W be a real-valued function defined on $R \times \{0,1\}$ and for fixed j let

$$a_\ell = \xi_W(W_{\tau^1_{j-1}+\ell+1}, Y^2_{\tau^1_{j-1}}), \quad \ell \geq 0, \quad Y^2_{\tau^1_{-1}} = y$$

and $U_{\xi,j}$ be an extension of (4.12) *given by*

$$U_{\xi,j} = \sum_{t=\tau^1_{j-1}+1}^{\tau^1_j} g_X(X_t)\xi_W(W_t, Y^2_{\tau^1_{j-1}}), \quad s \geq 0\,. \tag{4.20}$$

Then for $m \geq 1$,

$$\mathrm{E}_\lambda\Big[U^m_{\xi,j} \mid \mathcal{H}_{j-1}\Big] = \begin{cases} \mathrm{E}_{\lambda_1}\Big[U_{\underline{a}}\Big]^m, & \text{for } j = 0 \\ \mathrm{E}_{\nu_1}\Big[U_{\underline{a}}\Big]^m, & \text{for } j \geq 1 \end{cases}$$

where $U_{\underline{a}} = \sum_{\ell=0}^{\tau_1} g_X(X'_\ell)a_\ell$.

Proof of Lemma 4.4:
Let $j \geq 1$. By (4.20),

$$U_{\xi,j} = \sum_{t=\tau^1_{j-1}+1}^{\tau^1_j} g_X(X_t)\xi_W(W_t, Y^2_{\tau^1_{j-1}}) = \sum_{\ell=1}^{\tau^1_j-\tau^1_{j-1}} g_X(X_{\tau^1_{j-1}+\ell})\xi_W(W_{\tau^1_{j-1}+\ell}, Y^2_{\tau^1_{j-1}})$$

so that

$$\begin{aligned}
\mathrm{E}_\lambda\Big[U^m_{\xi,j} \mid \mathcal{H}_{j-1}\Big] &= \mathrm{E}_\lambda\Big[\Big\{\sum_{\ell=1}^{\delta_j} g_X(X_{\tau^1_{j-1}+\ell})\xi_W(W_{\tau^1_{j-1}+\ell}, Y^2_{\tau^1_{j-1}})\Big\}^m \mid \mathcal{H}_{j-1}\Big] \\
&= \mathrm{E}_\alpha\Big[\sum_{\ell=1}^{S^1_\alpha} g_X(X'_\ell)a_{\ell-1}\Big]^m \\
&= \mathrm{E}_{\nu_1}\Big[\sum_{\ell=0}^{\tau^1_0} g_X(X''_\ell)a_\ell\Big]^m \\
&= \mathrm{E}_{\nu_1} U^m_{\underline{a}},
\end{aligned}$$

where we have used that

$$\mathcal{L}_{\lambda_1}\Big\{X_{\tau^1_{j-1}+\ell},\ 1 \leq \ell \leq \delta_j\Big\} = \mathcal{L}_\alpha\Big\{X'_\ell,\ 1 \leq \ell \leq S_\alpha\Big\} = \mathcal{L}_{\nu_1}\Big\{X''_\ell,\ 0 \leq \ell \leq \tau^1_0\Big\}$$

where $\mathcal{L}_\lambda$ denote the simultaneous distribution with initial measure λ and X'' has the same distribution as X'.

If $j = 0$, then

$$\begin{aligned} \mathrm{E}_\lambda\Big[U^m_{\xi,j} \mid \mathcal{H}_{j-1}\Big] &= \mathrm{E}_\lambda\Big[\Big\{\sum_{t=0}^{\tau_0^1} g_X(X_t)\xi_W(W_t,y)\Big\}^m \mid \mathcal{F}^W \vee \mathcal{F}^{Y^2}\Big] \\ &= \mathrm{E}_{\lambda_1} U^m_{\underline{a}} \,. \end{aligned}$$

□

Using the previous lemma and a general moment formula given in Corollary A.1 in Appendix A we obtain a useful exact formula.

Lemma 4.5 *Let U_k be defined by (4.12). Let $m \geq 1$. Then*

$$\mathrm{E}_\nu\Big\{\sum_{k=0}^{\mathcal{T}} U_k^m\Big\} = \sum_{r=1}^{m}\sum_{\ell\in\Delta_{m,r}}\binom{m}{\ell}\sum_{j^{(2)}\in\mathcal{N}_+^{r-1}}\Big\{\pi_{s_1} f^X_{j^{(2)},\ell}\Big\}\Big\{\pi_{s_2} f^W_{j^{(2)},\ell}\Big\} \tag{4.21}$$

where $\mathcal{N}\times\mathcal{N}_+^{r-1} = \{(j_1,\ldots,j_r)\in\mathcal{N}^r,\ j_1\geq 0,\ j_2\geq 1\ \cdots\ j_r\geq 1\}$, $\Delta_{m,r} = \{\ell\in\mathcal{N}_+^r\colon \sum_{j=1}^r \ell_j = m\}$, $j^{(2)} = (j_2,\ldots,j_r)$ *and*

$$f^X_{j^{(2)},\ell} = I_{g_X^{\ell_1}} H_1^{j_2} I_{g_X^{\ell_2}} \cdots H_1^{j_r} I_{g_X^{\ell_r}} 1, \quad f^W_{j^{(2)},\ell} = I_{g_W^{\ell_1}} P_2^{j_2} I_{g_W^{\ell_2}} \cdots P_2^{j_r} I_{g_W^{\ell_r}} 1 \,.$$

More generally we have for arbitrary $\lambda = \lambda_1\otimes\lambda_2$ *with* $f^X_{j,\ell} = H_1^{j_1} f^X_{j^{(2)},\ell}$ *and* $f^W_{j,\ell} = P_2^{j_1} f^W_{j^{(2)},\ell}$,

$$\begin{aligned} &\mathrm{E}_\lambda\Big\{\sum_{k=0}^{\mathcal{T}} U_k^m\Big\} \\ &\quad= \sum_{r=1}^{m}\sum_{\ell\in\Delta_{m,r}}\binom{m}{\ell}\sum_{j\in\mathcal{N}\times\mathcal{N}_+^{r-1}}\Big\{(\lambda-\nu)(f^X_{j,\ell}\otimes f^W_{j,\ell}) + \Big\{\nu_1 f^X_{j,\ell}\Big\}\Big\{\tilde{\lambda}\underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} P_2 f^W_{j,\ell}\Big\}\Big\}. \end{aligned} \tag{4.22}$$

Remark 4.1 *If $\lambda = \nu$, then*

$$\tilde{\lambda}\underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} P_2 f^W_{j,\ell} = \underset{\sim}{\nu}\underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} P_2 P_2^{j_1} f^W_{j^{(2)},\ell} = \pi_{s_2} P_2^{j_1+1} f^W_{j^{(2)},\ell} = \pi_{s_2} f^W_{j^{(2)},\ell}$$

and

$$\sum_{j_1=0}^{\infty} \nu_1 f^X_{j,\ell} = \nu_1\Big\{\sum_{j_1=0}^{\infty} H^{j_1}\Big\} f^X_{j^{(2)},\ell} = \pi_{s_1} f^X_{j^{(2)},\ell} \,.$$

Thus (4.22) reduces to (4.21) when $\lambda = \nu$.

<u>Proof of Lemma 4.5:</u>

We rewrite the first term on the left hand side of (4.21) using that $\{\mathcal{T} \geq k\} = \{\mathcal{T} \geq k-1\} \cap \{Y^2_{\tau^1_{k-1}} = 0\}$, so that

$$\mathrm{E}_\nu(\sum_{k=0}^{\mathcal{T}} U_k^m) = \mathrm{E}_\nu(U_0^m) + \sum_{k=1}^{\infty} \mathrm{E}_\nu\Big[1(\mathcal{T} \geq k-1)\mathrm{E}_\nu(U_k^m 1(Y^2_{\tau^1_{k-1}} = 0) \mid \mathcal{H}_{k-1})\Big] .$$

Let $U_{\xi,k}$ be defined by (4.20) where $\xi_W(w,y) = g_W(w)(1-y)$. By Lemma 4.4 and Corollary A.1 we find that for $k \geq 1$

$$\begin{aligned}
&\mathrm{E}_\nu(U_k^m 1(Y^2_{\tau^1_{k-1}} = 0) \mid \mathcal{H}_{k-1}) \\
&\quad = \mathrm{E}_\nu(U_{\xi,k}^m \mid \mathcal{H}_{k-1}) \\
&\quad = \sum_{r=1}^{m} \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} \sum_{j \in \mathcal{N} \times \mathcal{N}_+^{r-1}} \{\nu_1 f^X_{j,\ell}\}\Big\{\prod_{i=1}^{r} g_W^{\ell_i}(W_{\tau_{k-1}+t_i+1})\Big\} 1(Y^2_{\tau^1_{k-1}} = 0)
\end{aligned} \tag{4.23}$$

where $t_i = j_1 + \cdots + j_i$. Let

$$\mathcal{G}_k = \mathcal{F}^X_{\tau^1_k} \vee \mathcal{F}^{Y^1}_{\tau^1_k} \vee \mathcal{F}^W_{\tau_k} \vee \mathcal{F}^{Y^2}_{\tau^1_{k-1}} .$$

Then by conditioning with respect to $\mathcal{G}_{k-1}$ we find that

$$\begin{aligned}
&\mathrm{E}_\nu\Big\{1(\mathcal{T} \geq k-1)\Big\{\prod_{i=1}^{r} g_W^{\ell_i}(W_{\tau_{k-1}+t_k+1})\Big\} 1(Y^2_{\tau^1_{k-1}} = 0)\Big\} \\
&\qquad\qquad = \mathrm{E}_\nu\Big\{1(\mathcal{T} \geq k-1) H_2 f^W_{j,\ell}(W_{\tau_{k-1}})\Big\} .
\end{aligned} \tag{4.24}$$

Hence

$$\begin{aligned}
\mathrm{E}_\nu\Big\{\sum_{k=1}^{\mathcal{T}} U_k^m\Big\} &= \sum_{k=1}^{\infty} \mathrm{E}_\nu\Big[1(\mathcal{T} \geq k-1)\mathrm{E}_\nu(U_k^m 1(Y^2_{\tau^1_{k-1}} = 0) \mid \mathcal{H}_{k-1})\Big] \\
&= \sum_{r=1}^{m} \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} \sum_{j \in \mathcal{N} \times \mathcal{N}_+^{r-1}} \{\nu_1 f^X_{j,\ell}\}\Big\{\mathrm{E}_\nu\Big\{\sum_{k=1}^{\infty} 1(\mathcal{T} \geq k-1) H_2 f^W_{j,\ell}(W_{\tau_{k-1}})\Big\} \\
&= \sum_{r=1}^{m} \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} \sum_{j \in \mathcal{N} \times \mathcal{N}_+^{r-1}} \{\nu_1 f^X_{j,\ell}\}\Big\{\pi_{s_2} H_2 f^W_{j,\ell}\Big\} .
\end{aligned} \tag{4.25}$$

Similarly, we find that

$$\mathrm{E}_\nu\Big\{U_0^m\Big\} = \sum_{r=1}^{m} \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} \sum_{j \in \mathcal{N} \times \mathcal{N}_+^{r-1}} \{\nu_1 f^X_{j,\ell}\}\Big\{\nu_2 f^W_{j,\ell}\Big\} \tag{4.26}$$

and combining (4.25), (4.26) and inserting $\pi_{s_2} H_2 = \pi_{s_2} - \nu_2$ we get

$$\mathrm{E}_\nu\Big\{\sum_{k=0}^{\mathcal{T}} U_k^m\Big\} = \sum_{r=1}^{m} \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} \sum_{j \in \mathcal{N} \times \mathcal{N}_+^{r-1}} \{\nu_1 f^X_{j,\ell}\}\Big\{\pi_{s_2} f^W_{j,\ell}\Big\} .$$

Now

$$\pi_{s_2} f^W_{j,\ell} = \pi_{s_2} f^W_{j(2),\ell}\,.$$

$$\begin{aligned}\sum_{j_1=0}^{\infty} \nu_1 f^X_{j_1,j(2),\ell} &= \sum_{j_1=0}^{\infty} \nu_1 H_1^{j_1} I_{g_X^{\ell_1}} H_1^{j_2} I_{g_X^{\ell_2}} \cdots H_1^{j_r} I_{g_X^{\ell_r}} 1 \\ &= \pi_{s_1} I_{g_X^{\ell_1}} H_1^{j_2} I_{g_X^{\ell_2}} \cdots H_1^{j_r} I_{g_X^{\ell_r}} 1 \\ &= \pi_{s_1} f^X_{j(2),\ell}\end{aligned}$$

and (4.21) is proved.

The proof of (4.22) is similar. Instead of (4.25) we get

$$\begin{aligned}\mathrm{E}_\lambda\Big\{\sum_{k=1}^{\mathcal{T}} U_k^m\Big\} &= \sum_{r=1}^{m}\sum_{\ell\in\Delta_{m,r}} \binom{m}{\ell} \sum_{j\in\mathcal{N}\times\mathcal{N}_+^{r-1}} \{\nu_1 f^X_{j,\ell}\}\Big\{\mathrm{E}_\lambda \sum_{k=0}^{\mathcal{T}} H_2 f^W_{j,\ell}(W_{\tau_k^1})\Big\} \\ &= \sum_{r=1}^{m}\sum_{\ell\in\Delta_{m,r}} \binom{m}{\ell} \sum_{j\in\mathcal{N}\times\mathcal{N}_+^{r-1}} \{\nu_1 f^X_{j,\ell}\}\Big\{\tilde{\lambda} \underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} H_2 f^W_{j,\ell}\Big\} \qquad (4.27)\end{aligned}$$

and (4.26) is changed to

$$\mathrm{E}_\lambda\Big\{U_0^m\Big\} = \sum_{r=1}^{m}\sum_{\ell\in\Delta_{m,r}} \binom{m}{\ell} \sum_{j\in\mathcal{N}\times\mathcal{N}_+^{r-1}} \{\lambda_1 f^X_{j,\ell}\}\Big\{\lambda_2 f^W_{j,\ell}\Big\}\,. \qquad (4.28)$$

Combining (4.27), (4.28) and inserting that $\tilde{\lambda} \underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} H_2 = \tilde{\lambda} \underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} P_2 - \nu_2$ we obtain (4.22).

□

Remark 4.2 *If $m = 1$, then by (4.21) we find that*

$$\mu_g = \mathrm{E}_\nu\Big\{\sum_{j=0}^{\mathcal{T}} U_j\Big\} = \{\pi_{s_1} g_X\}\{\pi_{s_2} g_W\}$$

which is consistent with (4.9).

Remark 4.3 *If $m = 2$. Then*

$$\mathrm{E}_\nu\Big\{\sum_{j=0}^{\mathcal{T}} U_j^2\Big\} = \{\pi_{s_1} g_X^2\}\{\pi_{s_2} g_W^2\} + 2\sum_{\ell=1}^{\infty}\Big\{\pi_{s_1} I_{g_X} H_1^\ell I_{g_X} 1\Big\}\Big\{\pi_{s_2} I_{g_W} P^\ell I_{g_W} 1\Big\}\,.$$

Remark 4.4 *By* (4.10), (4.13) *and* (4.22) *we find that for general* $\lambda = \lambda_1 \otimes \lambda_2$, $g = g_X \otimes g_W$,

$$\mathrm{E}_\lambda V_0 = \mathrm{E}_\lambda V_g = \sum_{j=0}^{\infty}(\lambda - \nu)(P_1^j g_X \otimes P_2^j g_W) + \sum_{j=0}^{\infty}\Big\{\nu_1 H_1^j g_X\Big\}\Big\{\tilde{\lambda}\underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} P_2^{j+1} g_W\Big\}\,.$$

If g_X *is small,* $\lambda_2 = \pi_2$ *and* $\sup_j \pi_2 \underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} P_2^{j+1}|g_W| < \infty$ *then* $\mathrm{E}_\lambda V_g$ *is finite. By Lemma* A.1 *taking* $p = 1+\delta$, $f = P_2^{j+1}|g_W|$ *and* $\lambda = \pi_2$ *we have that*

$$\pi_2\Big(\underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} P_2^{j+1}|g_W|\Big)^{\frac{1+\delta}{1+\eta\delta}} \le c_2\Big\{\mathrm{E}_{\pi_2}^{\frac{1}{1+\eta\delta}} \underset{\sim}{\mathcal{T}}^{1+2\delta}\Big\}\Big\{\pi_2^{\frac{\eta\delta}{1+\eta\delta}}|g_W|^{\frac{1+\delta}{\eta\delta}}\Big\}$$

with $\eta \in (0,1)$ *and* $\delta > 0$ *arbitrary.*

Proof of Theorem 4.2: Assume that $g_X \ge 0$. By Cauchy-Schwartz, recalling that $\sum_{j=1}^m \ell_j = m$,

$$\begin{aligned} |\pi_2 f^W_{j^{(2)},\ell}| &= |\pi_2 I_{g_W^{\ell_1}} P_2^{j_2} I_{g_W^{\ell_2}} \cdots P_2^{j_r} I_{g_W^{\ell_r}} 1| \\ &\le \mathrm{E}_{\pi_2}|g_W|^{\ell_1}(W_0)\cdots|g_W|^{\ell_r}(W_{j_r}) \\ &\le \prod_{r=1}^m \mathrm{E}_{\pi_2}^{(\frac{\ell_r}{m})}|g_W|^{(\frac{m}{\ell_r})\,\ell_r}(W_0) \\ &= \mathrm{E}_{\pi_2}|g_W|^m(W_0) \\ &= \pi_2|g_W|^m\,. \end{aligned} \tag{4.29}$$

Inserting (4.29) into (4.21) we obtain

$$\begin{aligned} \mathrm{E}_\nu\Big\{\sum_{k=0}^{\mathcal{T}} U_k^m\Big\} \le\ & \Big\{\pi_{s_2}|g_W|^m\Big\}\sum_{r=1}^m \sum_{\ell\in\Delta_{m,r}} \binom{m}{\ell} \sum_{j^{(2)}\in\mathcal{N}_+^{r-1}} \Big\{\pi_{s_1} I_{g_X^{\ell_1}} H_1^{j_2} I_{g_X^{\ell_2}} \cdots H_1^{j_r} I_{g_X^{\ell_r}} 1\Big\} \\ =\ & \Big\{\pi_{s_2}|g_W|^m\Big\}\mathrm{E}_\nu\Big\{U_{g_X}^m\Big\} \end{aligned}$$

from which (4.19) follows trivially.

□

4.6 Moment bounds of V_{g_h} expressed in terms of bandwidth

The following results describe how higher order moments of V_0 behave as a function of the bandwidth. This is directly related to (4.4) and (4.5).

Theorem 4.3 *Let $g_X = K_{x,h}$ and assume that the conditions* K_1, K_2, P_1 *and* (4.18) *hold. Then for all $k, m \geq 1$,*

$$\mathrm{E}_\nu |V_0|^{2m} = \mathrm{E}_\nu |V_{g_h}|^{2m} \leq d_{m,k} h^{-2m+1/(k+1)} \tag{4.30}$$

where

$$d_{m,k} \stackrel{\text{def}}{=} \left\{\pi_{s_2}^{\frac{1}{k+1}} g_W^{2m(k+1)}\right\}\left\{\mathrm{E}_\nu^{\frac{k}{k+1}} \left(\mathcal{T}+1\right)^{2m\left(\frac{k+1}{k}\right)}\right\}\left\{d'^{\frac{1}{k+1}}_{2m}\right\}$$

and the sequence of constants $\{d'_m\}$ is only dependent upon $\mathcal{N}_x$ and $\sup_u K(u)$.

First we need the following lemma which gives a more direct upper bound of $\mathrm{E}|V_0|^p$.

Lemma 4.6 *For all $p > 0$ and $\delta \in (0, \infty)$*

$$\mathrm{E}_\nu |V_0|^p \leq \mathrm{E}_\nu^{\frac{1}{(1+\delta)}} \left|\sum_{j=0}^{\mathcal{T}} |U_j|^{p(1+\delta)}\right| \mathrm{E}_\nu^{\frac{\delta}{(1+\delta)}} |\mathcal{T}+1|^{p(1+\delta^{-1})} .$$

Proof of Lemma 4.6: Let $r = 1+\delta$ and $q = 1+\delta^{-1}$

$$\begin{aligned}
\mathrm{E}_\nu |V_0|^p &= \mathrm{E}_\nu \Big|\sum_{j=0}^{\mathcal{T}} U_j\Big|^p \\
&\leq \mathrm{E}_\nu \left|\max_{0\leq j\leq \mathcal{T}} |U_j|(\mathcal{T}+1)\right|^p \\
&= \mathrm{E}_\nu \max_{0\leq j\leq \mathcal{T}} |U_j|^p \, |\mathcal{T}+1|^p \\
&\leq \mathrm{E}_\nu^{\frac{1}{r}} \left|\max_{0\leq j\leq \mathcal{T}} |U_j|^{pr}\right| \mathrm{E}_\nu^{\frac{1}{q}} |\mathcal{T}+1|^{pq} \\
&\leq \mathrm{E}_\nu^{\frac{1}{r}} \left|\sum_{j=0}^{\mathcal{T}} |U_j|^{pr}\right| \mathrm{E}_\nu^{\frac{1}{q}} |\mathcal{T}+1|^{pq} .
\end{aligned}$$

□

Proof of Theorem 4.3:
By Lemma 4.6 with $p = 2m$ and $\delta = k$,

$$\mathrm{E}_\nu |V_0|^{2m} \leq \mathrm{E}_\nu^{\frac{1}{k+1}} \left|\sum_{j=0}^{\mathcal{T}} |U_j|^{2m(k+1)}\right| \mathrm{E}_\nu^{\frac{k}{k+1}} |\mathcal{T}+1|^{2m\left(\frac{k+1}{k}\right)} . \tag{4.31}$$

From Theorem 4.2 we have

$$\mathrm{E}_\nu \left|\sum_{j=0}^{\mathcal{T}} U_j^{2m(k+1)}\right| \leq \left\{\pi_{s_2} |g_W|^{2m(k+1)}\right\} \mathrm{E}_\nu \left|U_{g_X}^{2m(k+1)}\right| \tag{4.32}$$

and by KT (Lemma 4.1 with $\xi_0 \equiv 1$) ,

$$\mathrm{E}_\nu \left|U_{g_X}\right|^{2m(k+1)} \le d'_{2m} h^{-2m(k+1)+1} \tag{4.33}$$

where it is also shown that the sequence of constants $\{d'_m\}$ is only dependent upon $\mathcal{N}_x$ and $\sup_u K(u)$.

Inserting (4.32) and (4.33) into (4.31) we get (4.30).

□

4.7 Asymptotic variance

Exact information about the first order properties of the asymptotic variance is important (cf. (4.2) and (4.3)). This is the content of the next result. Our method of proof use a truncation technique based on the notion of a generalized autocovariance function. We believe that the latter concept has some independent interest.

Theorem 4.4 *Assume that* $g_{X,h} = K_{x,h}$ *and that the conditions* K_1, K_2, P_1,P_5, *and* W_4 *with* $m = 1$ *and* $k \ge 1$ *hold. Then, if* $\mu_{g_W} = 0$, *as* $h \downarrow 0$

$$h\sigma^2_{g_{X,h}\otimes g_W} = p_{s_1}(x)\Big\{\int K^2(u)du\Big\}\Big\{\pi_{s_2} g_W^2\Big\} + o(1) \ .$$

We also have that

$$h\sigma^2_{|g_{X,h}\otimes g_W|} = p_{s_1}(x)\Big\{\int K^2(u)du\Big\}\Big\{\pi_{s_2} g_W^2\Big\} + o(1) \ . \tag{4.34}$$

In the proof of Theorem 4.4 we use an extension of the cross covariance function defined for ergodic processes. Let $\mathcal{Z}$ denote set of integers and g, f real valued measurable functions defined on E. Then

$$\gamma_{g,f}(\ell) \stackrel{\text{def}}{=} \pi_s I_g P^{|\ell|} I_f 1, \quad \ell \in \mathcal{Z} \tag{4.35}$$

and $\gamma_g = \gamma_{g,g}$. Note that $\gamma_{g_W} = \{\pi_2 s_2\}^{-1} c_W$ where c_W is the covariance function for the stationary process $\{g_W(W_t),\ t \ge 0\}$ (cf. Section 3).

Lemma 4.7 *Assume that* $g(x,w) = g_X(x) g_W(w)$, g_X *is small,* $\mu_{g_W} = 0$, *and the W-process satisfies* W_4 *with* $m = 1$ *and* $k \ge 1$. *Then*

$$\sigma_g^2 = \sum_{\ell=-\infty}^{\infty} \gamma_{g_X}(\ell)\gamma_{g_W}(\ell) \ . \tag{4.36}$$

We also have that

$$\sigma^2_{|g|} = \sum_{\ell=-\infty}^{\infty} \gamma_{(|g|,|g|-s\mu_{|g|})}(\ell) - \pi_s|g|\pi_s|g\cdot s| \ . \tag{4.37}$$

Proof of Lemma 4.7:

Our proof is based upon Theorem A.2. By (4.31) and (4.32) with $m = 1$ and the smallness of g_X,

$$\pi_s(g^2) + 2\pi_s I_g P\, G_{s,\nu} g = \mathrm{E}_\nu V_0^2 \le c_0\, \pi_2^{\frac{1}{k+1}} |g_W|^{2(k+1)}\, \mathrm{E}_\nu^{\frac{k}{k+1}} \mathcal{T}^{\frac{2k+2}{k}} < \infty\,.$$

To prove that the last term is finite note first that

$$\mathrm{E}_\nu \mathcal{T}^{\frac{2k+2}{k}} = \mathrm{E}_{\underset{\sim}{\nu}} \underset{\sim}{\mathcal{T}}^{\frac{2k+2}{k}}\,. \tag{4.38}$$

By (2.1), $\pi_2 \ge \{\pi_2 s_2\}\nu_2$. Hence $\pi_2 \ge c_1 \nu_2$ and the right hand side of (4.38) is finite if

$$\mathrm{E}_{\pi_2} \underset{\sim}{\mathcal{T}}^{\frac{2k+2}{k}} < \infty\,.$$

By Lemma 4.3 this is true if

$$\sum_{\ell=1}^{\infty} \ell^{(k+2)/k} \underset{\sim}{\alpha}_\ell < \infty \tag{4.39}$$

But (4.39) follows from W_4 and hence (A.14) is fulfilled. Assume that $g_{W,m}$ is defined by (A.13). Let

$$g_m(x,w) = g_X(x) g_{W,m}(w)$$

and $\xi_m = g_{W,m} - s_2\mu_m$ with $\mu_m = \pi_{s_2} g_{W,m}$. Then, using that for general functions g and g'

$$\gamma_{g_X \otimes g_W, g'_X \otimes g'_W} = \gamma_{g_X, g'_X} \cdot \gamma_{g_W, g'_W}\,,$$

we obtain

$$\begin{aligned}
&|\gamma_{g_m, g_m - g_X \otimes s_2 \mu_m}(\ell)| \\
&\quad = \left|\gamma_{g_X \otimes g_{W,m}, g_X \otimes \xi_m)}(\ell)\right| \\
&\quad = |\gamma_{g_X}(\ell)|\, \left|\gamma_{g_{W,m}, \xi_m}(\ell)\right| \\
&\quad = |\gamma_{g_X}(\ell)|\, \{\pi_2(s_2)\}^{-1} \left|\pi_2 I_{g_{W,m}} P_2^\ell \xi_m\right| \\
&\quad \le \{\pi_2(s_2)\}^{-1} \gamma_X(0) \left|\mathrm{Cov}_{\pi_2}\Big(g_{W,m}(W_0),\ \xi_m(W_\ell)\Big) + \mathrm{E}_{\pi_2}\Big[g_{W,m}(W_0)\Big] \mathrm{E}_{\pi_2}\Big[\xi_m\Big]\right| \\
&\quad \le \{\pi_2(s_2)\}^{-1} \gamma_X(0)\, 8\, \|g_{W,m}\|_{1/2(k+1)}\, \|\xi_m\|_{1/2(k+1)}\, \alpha_\ell^{k/(k+1)} \\
&\quad \le \{\pi_2(s_2)\}^{-1} \gamma_X(0)\, 16\, \|g_W\|^2_{1/2(k+1)}\, \alpha_\ell^{k/(k+1)}
\end{aligned} \tag{4.40}$$

where we have used a strong mixing inequality (cf. Hall & Heyde, 1980, Corollary A.2, p. 278),

$$\left|\mathrm{Cov}_{\pi_2}\Big(g_{W,m}(W_0),\ \xi_m(W_\ell)\Big)\right| \le 8 \|g_W\|_{1/2(k+1)}\, \|\xi_m\|_{1/2(k+1)}\, \alpha_\ell^{k/(k+1)}$$

and that $\mathrm{E}_{\pi_2}\Big[\xi_m\Big] = 0$. By W_4

$$\sum_{\ell=1}^{\infty} \ell^{(k+2)/k} \alpha_\ell < \infty\,.$$

Since for $p > 0$ $\sum_{\ell=1}^{\infty} \ell^p \alpha_\ell < \infty$ implies that $\sum_{\ell=1}^{\infty} \alpha_\ell^{1/p} < \infty$, it follows that

$$\sum_{\ell=1}^{\infty} \alpha_\ell^{k/(k+2)} < \infty \, .$$

Together with

$$\sum_{\ell=1}^{\infty} \alpha_\ell^{k/(k+1)} \le \sum_{\ell=1}^{\infty} \alpha_\ell^{k/(k+2)} < \infty$$

this shows that (A.16) holds and (4.36) follows from Theorem A.2.

The last part of the lemma follows in the same way.

□

Remark 4.5 *The formulae* (4.36) *and* (4.37) *can be viewed as generalizations of the formulae* $Var(n^{-1/2}\sum_{j=0}^{n} X_j) = \sum_{\ell=-\infty}^{\infty} Cov(X_t, X_{t-\ell}) + o(1)$ *in case* $\{X_t\}$ *is a stationary process with an absolutely summable covariance function.*

Proof of Theorem 4.4:

By Lemma 4.7

$$h\sigma^2_{g_{X,h}\otimes g_W} = \sum_\ell \gamma_{g_{X,h}}(\ell)\gamma_{g_W}(\ell)$$

and for $\ell > 0$,

$$h\gamma_{g_{X,h}}(\ell) = h\pi_{s_1} I_{K_{x,h}} P_1^\ell K_{x,h} = o(1)$$

by P_5 (cf. KT proof of Lemma 4.2). Since

$$|h\gamma_{g_{X,h}}(\ell)| \le \int p_s(x+hu)K(u)P^\ell(x+hu, \mathcal{N}_x(h))du \le c_0,$$

we can apply the dominated convergence theorem, i.e.,

$$\begin{aligned}
\lim_{h\downarrow 0} h\sigma^2_{g_{X,h}\otimes g_W} &= \sum_\ell \Big\{\lim_{h\downarrow 0} h\gamma_{g_{X,h}}(\ell)\Big\}\gamma_{g_W}(\ell) \\
&= \Big\{\lim_{h\downarrow 0} h\gamma_{g_{X,h}}(0)\Big\}\gamma_{g_W}(0) \\
&= \Big\{\lim_{h\downarrow 0} h\pi_{s_1} K^2_{x,h}\Big\}\Big\{\pi_{s_2} g_W^2\Big\} \\
&= p_{s_1}(x)\Big\{\int K^2(u)du\Big\}\Big\{\pi_{s_2} g_W^2\Big\} \, .
\end{aligned}$$

The proof of (4.34) is similar since

$$\gamma_{(|g|,|g|-s\mu_{|g|})}(\ell) = \gamma_{g_X}(\ell)\gamma_{|g_W|}(\ell) - \gamma_{g_X,s_1}(\ell)\gamma_{|g_W|,s_2}(\ell)\{\pi_{s_1} K_{x,h}\}\pi_{s_2}|g_W|$$

and

$$\pi_s|g|\pi_s|g\cdot s| = \{\pi_{s_1} K_{x,h}\}\{\pi_{s_1} K_{x,h}\cdot s_1\}\ \{\pi_{s_2}|g_W|\}\{\pi_{s_2}|g_W\cdot s_2|\} \, .$$

□

4.8 Main results

Theorem 4.5 *Assume that* $\{X_t\}$ *and* $\{W_t\}$ *are independent and that* K_1, K_2, K_4, P_1, P_2, P_4, P_5, P_6 *and* W_4 *hold. Let* $k \geq 1$ *and* $m \geq 2$. *Assume that for some* $\epsilon > 0$,

$$h_n^{-1} \ll n^{\beta\delta_m - \epsilon}, \quad \delta_m = \frac{m-1}{m-1/(k+1)} .$$

Then for all $\lambda = \lambda_1 \otimes \pi_2$

$$\Delta_{n,h_n} = S_n^{-1/2}(K_{x,h_n}) h_n^{1/2} S_n(K_{x,h_n} \otimes g_W) \xrightarrow[n]{d} \mathcal{N}\Big(0, \Big\{\int K^2(u)du\Big\}\Big\{\pi_2 g_W^2\Big\}\Big) .$$

Proof of Theorem 4.5: Recall that for $C = C_1 \times C_2$, $C_i \in \mathcal{E}_i$, $i = 1, 2$,

$$T_C(n) = \sum_{t=0}^{n} 1_{C_1}(X_t) 1_{C_2}(W_t)$$

represents the number of visits to C of $\{(X_t, W_t)\}$ up to time n. We choose C_1 and C_2 so that both sets are small. Then by KT (second part of Th. 4.3) using K_1, K_2, P_1, P_2, P_4, P_6

$$\widehat{p}_{C_1}(x) \stackrel{\text{def}}{=} \frac{S_n(K_{x,h_n})}{T_{C_1 \times E_2}(n)} = p_{C_1}(x) + o_P(1)$$

with $E = E_1 \times E_2$, and where $p_{C_1}(x) = p_{s_1}(x)/\pi_{s_1} 1_{C_1}$. By KT (Corollary 2.1)

$$\frac{T_C(n)}{T_{C_1 \times E_2}(n)} = \frac{\pi_s 1_C}{\pi_s 1_{C_1 \times E_2}} + o(1) \quad \text{a.s.} \ = \pi_2(C_2) + o(1) \quad \text{a.s.} \ .$$

We can write

$$\Delta_{n,h_n} = \{\widehat{p}_{C_1}(x)\}^{-1/2} \Big\{\frac{T_C(n)}{T_{C_1 \times E_2}(n)}\Big\}^{1/2} \Big\{T_C^{-1/2}(n) h_n^{1/2} S_n(g_{h_n})\Big\} = A_{n,h_n}^{1/2} \Delta_{n,h_n}^0, \quad \text{say.}$$

We have that

$$A_{n,h_n} = \{p_{C_1}^{-1}(x) \pi_2(C_2)\} + o_P(1) .$$

Hence, it is enough to prove that

$$\Delta_{n,h_n}^0 \xrightarrow[n]{d} \mathcal{N}\Big(0, p_{C_1}(x) \Big\{\int K^2(u)du\Big\}\Big\{\frac{\pi_2 g_W^2}{\pi_2 1_{C_2}}\Big\}\Big) \tag{4.41}$$

where $g_h(z, w) = K_{x,h}(z) g_W(w)$.

By P_1 and P_2 and Bochner theorem, (4.1) is satisfied. From Theorem 4.3 and Theorem 4.4 the conditions (4.2) - (4.5) are fulfilled.

It remains to verify (4.7). Let $g_0 = c_0 1_{\mathcal{N}_x} |g_W|$ where c_0 is an appropriate constant. Then $|h g_h| \leq g_0$. We have to prove that

$$\mathrm{P}_\lambda(V_{g_0} < \infty) = 1$$

But this is fulfilled if $\mathrm{E}_\lambda V_{g_0} < \infty$. By Remark 4.4 this is true if

$$\mathrm{E}_{\pi_2} \left| \underset{\sim}{\mathcal{T}}^{1+2\delta} \right| \pi_2 |g_W|^{\frac{1+\delta}{\eta\delta}} < \infty \tag{4.42}$$

for some $\delta > 0$ and $\eta \in (0,1)$. By W_4 and Lemma 4.3

$$\pi_2 |g_W|^{2m(k+1)} < \infty, \quad \mathrm{E}_{\pi_2} \underset{\sim}{\mathcal{T}}^{2m+\frac{2m}{k}} < \infty .$$

Thus (4.42) holds.

Hence

$$\Delta^0_{n,h_n} \xrightarrow[n]{d} \mathcal{N}\Big(0, \sigma_C^2\Big)$$

and

$$\sigma_C^2 = \{\pi_{s_1} 1_{C_1}\}^{-1} \{\pi_{s_2} 1_{C_2}\}^{-1} p_{s_1}(x) \Big\{ \int K^2(u) du \Big\} \Big\{ \pi_{s_2} g_W^2 \Big\} .$$

It follows that (4.41) holds.

□

5 Examples and simulations

The purpose of this section is to illustrate the small sample properties of the estimator $\widehat{f}(x)$ defined by (3.1). As can be seen from Myklebust et al (1999) it is not simple to find examples of null recurrent and β-null recurrent processes. The trivial example is the random walk given by

$$X_t = X_{t-1} + e_t \tag{5.1}$$

where $\{e_t\}$ consists of independent identical distributed variables with strictly positive density function with respect to the Lebesgue measure almost everywhere on R. This process is β-null recurrent with $\beta = 1/2$. Another class of null recurrent processes is the process given by

$$X_t = 1(|X_{t-1}| \le M) g(X_{t-1}) + 1(|X_{t-1}| > M) X_{t-1} + e_t$$

for some finite $M > 0$ and some measurable function g finite on $|x| \le M$. This process behaves as a random walk for large X_t's. Other examples of null recurrent processes are the first order threshold model studied by Meyn and Tweedie (1994) and the exponential autoregressive process looked at by Cline and Pu (1997).

We will look at some of these examples in the context of (1.1) and (3.1). We will examine the finite sample properties of our estimators by means of simulations. A difficult and largely unresolved problem is that of choosing a proper bandwidth. Theorem 3.1 and Theorem 4.4 of KT only give the allowable rate as n tends to infinity. It should be noted that these rates are different from the stationary case, n effectively being replaced by n^β. It practice we have found it useful to use cross-validation and to let the bandwidth h depend on x. In fact we have typically let h_n be proportional to $n^{-1/5}\{\widehat{p}_C(x)\}^{1/5}$ where $\widehat{p}_C(x)$ could be thought of as the locally estimated density.

We will look at two aspects of the estimation problem. In Figures 1-4 we concentrate on the accuracy of the estimation of the function $f(x)$ for the model (1.1). Only single realizations are shown, but they are representative for the quality of the estimates. In Figure 5 we turn to the finite sample approximation of the asymptotic normal distribution derived in Theorems 3.1 and 4.5.

In the sequence of plots given in Figures 1-3 we look at the estimate of $f(x)$ defined in (3.1). In all cases X_t is the random walk process given in (5.1), with $\{e_t\}$ consisting of iid standard normal variables. In Figure 2 $\{W_t\}$ has been chosen to be the stationary Gaussian AR process

$$W_t = 0.8W_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ consists of iid standard normal variables independent of $\{e_t\}$, whereas in Figures 1 and 3 $\{W_t\}$ is simply identical to $\{\varepsilon_t\}$. The function $f(x)$ has taken to be $f(x) = x$, $f(x) = \sqrt{|x|}$ and $f(x) = x^2$ in Figures 1-3, respectively. Both X_t and Z_t of (1.1) have been plotted in the a)-parts of the figures. (The traces are virtually coinciding in Figure 1a due to the scale). In all of the cases $f(x)$ is reasonably well estimated for the region for which we have data – and in all cases with the moderate sample size of 500 (Note that Figure 3c gives a magnified picture in the region $-10 \leq x \leq 10$). The traces of Figures 2a and 3a do not look particularly well cointegrated perhaps, but this is due to the nonlinear nature of the cointegrating relationship, where $\{Z_t - f(X_t)\}$ is stationary, whilst both $\{X_t\}$ and $\{Z_t\}$ are nonstationary. In Figures 2c and 2d we have also included plots of the trace and of the estimated $\widehat{M}_{\hat{W}}(x) = \widehat{E}(\widehat{W}_t \mid \widehat{W}_{t-1} = x)$ for the estimated residual process

$$\widehat{W}_t = Z_t - \widehat{f}(X_t)\ .$$

It is seen from Figure 2 that the stationary AR behavior of $\{W_t\}$ is well recovered, and thus gives a very informal verification of the cointegrating property. In a more stringent approach the properties of $\widehat{M}_{\hat{W}}(x)$ would have to be evaluated along the lines of Theorem 4.4 of KT . The plots of $\widehat{M}_{\hat{W}}(x)$ for the other two cases were similar.

Our last single realization example is for the same case as for Figure 1, except that W_t is now taken to be a random walk $W_t = W_{t-1} + \varepsilon_t$, which would imply that $\{X_t\}$ is not cointegrated with $\{Z_t\}$. It is not difficult to show theoretically that we get inconsistent estimates of $f(x)$ in this situation, and this is clearly revealed in Figure 4b; it is also seen that the two traces drift apart in Figure 4a.

Estimates similar to those in Figures 1-4 have been presented in the cointegration literature. Our contribution, which we believe to be new, is that we have singled out classes of processes and assumptions for which an asymptotic theory of these estimates can be constructed, such that it should be possible to work out confidence intervals and bands (and possibly stringent tests of nonlinear cointegration in the sense discussed in this paper). Our last experiment indicates the finite sample approximation to the asymptotic distribution of Theorem 3.1.

The quality of the sample approximation has to be judged using a multitude of realizations. A problem not encountered in the stationary case is that the simulated realizations may cover very different x-regions. (This is one reason for using only one realization in Figures 1-4.) Hence, for a fixed $x = x'$, close to the starting value $X_0 = 0$, say, of each realization, some realizations may have many observations in the neighborhood of x', whereas other realizations may have none observations in the vicinity of x' for the sample size we are considering. This kind of behavior does not occur in the stationary case, where the expected time until the process reaches x' is always finite and in practice small, when $|x - x'|$ is small. This means that in the verification of Theorem 3.1 we can either keep x fixed and wait until we have sufficiently many observations close to x; the other realizations being discarded, or we can choose a central realization-dependent value of x; e.g. the modal value of the sample, for studying the normalized ratio (3.2) of Theorem 3.1

We have chosen to adopt both procedures, although clearly we introduce some extraneous stochastics into the problem in the latter case. The approximation to normality as a function of sample size, for the quantity

$$\left[\frac{h_n \sum K_{x,h_n}}{\int K^2(u)du}\right]^{1/2} [\widehat{f}(x) - x]$$

derived from the simple cointegrated system

$$X_t = X_{t-1} + e_t, \qquad Z_t = X_t + \varepsilon_t.$$

at the point $x = 7.5$ is shown in Figure 5a. 1000 realizations have been used, and a particular realization is admitted into the evaluation as respectively 100, 200, 300, 500 and 800 observations are accumulated in the interval (5,10). For Figure 7b, on the other hand, a fixed point x has not been used; rather x has been taken to be the modal value and is thus varying from one realization to another. In this case the length of the time series has been 500, 1000 and 3000, respectively.

The quality of the approximation seems to be quite good with 500 observations in the "relevant" region surrounding x.

2x

A Appendix

In this section we assume that $\{X_t\}$ is an aperiodic ϕ-irreducible Markov chain with state space $(E, \mathcal{E})$ where $\mathcal{E}$ is countably generated. We also assume that the transition probability P satisfies a minorization inequality, i.e., $P \geq s \otimes \nu$ and that $\{X_t\}$ is Harris recurrent.

A.1 Higher order moments

Theorem A.1 *Let $\underline{g} = \{g_j\}$ be a sequence of real-valued measurable functions defined on E. Let*

$$U_{\underline{g}} = U = \sum_{j=0}^{\tau} g_j(X_j)\ .$$

Then

$$\mathrm{E}_x U^m = \sum_{r=1}^{m} \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} \psi_{r,\ell}(x) \qquad (A.1)$$

where $\Delta_{m,r} = \{\ell \in \mathcal{N}^r \colon \sum_{j=1}^r \ell_j = m\}$, $\binom{m}{\ell} = \binom{m}{\ell_1! \cdots \ell_r!}$,

$$\psi_{r,\ell} = \sum_{j \in \mathcal{N} \times \mathcal{N}_+^{r-1}} H^{j_1} I_{g_{j_1}^{\ell_1}} H^{j_2} I_{g_{j_1+j_2}^{\ell_2}} \cdots H^{j_r} I_{g_{j_1+\cdots+j_r}^{\ell_r}} 1 \qquad (A.2)$$

with $H = P - s \otimes \nu$.

Remark A.1 *If* $g_j \equiv g$ *then*

$$\begin{aligned}\psi_{r,\ell} &= \sum_{j \in \mathcal{N}^r} H^{j_1} I_{g^{\ell_1}} H^{j_2+1} I_{g^{\ell_2}} \cdots H^{j_r+1} I_{g^{\ell_r}} 1 \\ &= G_{s,\nu}(I_g H) \cdots G_{s,\nu}(I_g H) G_{s,\nu} g \,.\end{aligned}$$

Hence this result partly generalize Theorem A.1. in KT .

Proof of Theorem A.1:

Let

$$s_k = j_1 + \cdots + j_k, \quad k = 1, \ldots, r$$

and

$$\mathcal{B}_s = \mathcal{F}_s^X \vee \mathcal{F}_{s-1}^Y, \quad A_s = (1 - Y_s) \,.$$

By KT (Lemma A.3) we can write

$$U^m = \sum_{r=1}^m \sum_{\ell \in \Delta_{m,r}} \binom{m}{\ell} J_{r,\ell}, \quad J_{r,\ell} = \sum_{j \in \mathcal{N} \times \mathcal{N}_+^{r-1}} Z_{j,\ell} \qquad (A.3)$$

with

$$Z_{j,\ell} \stackrel{\text{def}}{=} g_{s_1}^{\ell_1}(X_{s_1}) \cdots g_{s_r}^{\ell_r}(X_{s_r}) \prod_{k=0}^{s_r - 1} A_k \,. \qquad (A.4)$$

It is sufficient to prove that

$$\mathrm{E}_x J_{r,\ell} = \psi_{r,\ell}, \quad \ell \in \Delta_{m,r}, \quad r = 1, \ldots, m \,.$$

We will prove this by induction on r. When $r = 1$, then

$$J_{1,\ell} = \sum_{j_1=0}^\infty g_{j_1}^{\ell_1}(X_{j_1}) \prod_{k=0}^{j_1 - 1} A_k = \sum_{j_1=0}^\infty g_{j_1}^{\ell_1}(X_{j_1}) 1(\tau \ge j)$$

and

$$\mathrm{E}_x J_{1,\ell} = \sum_{j_1=0}^\infty \mathrm{E}_x \Big[g_{j_1}^{\ell_1}(X_{j_1}) 1(\tau \ge j) \Big] = \sum_{j_1=0}^\infty H^{j_1} I_{g_{j_1}} 1 = \psi_{1,\ell} \,.$$

Assume that these formulae are correct for $r - 1$. Let ℓ be fixed, $Z_j = Z_{j,\ell}$ and $j_{(r)} = (j_1, \ldots, j_r)$. Then

$$Z_j = Z_{j_{(r-1)}} A_{s_{r-1}} D_{s_{r-1}, j_r}$$

where

$$D_{s_{r-1},j_r} \stackrel{\text{def}}{=} \sum_{j_r=1}^{\infty} g^{\ell_r}_{s_{r-1}+j_r}(X_{s_{r-1}+j_r})\Big\{\prod_{k=s_{r-1}+1}^{s_{r-1}+j_r-1} A_k\Big\}$$

so that

$$J_{r,\ell} = \sum_{j\in\mathcal{N}\times\mathcal{N}_+^{r-1}} Z_{j(r-1)} A_{s_{r-1}} D_{s_{r-1},j_r} \ .$$

Taking conditional expectation with respect to $\mathcal{B}_{s_{r-1}+1}$ gives

$$\mathrm{E}\Big[Z_{j(r-1)} A_{s_{r-1}} D_{s_{r-1},j_r} \mid \mathcal{B}_{s_{r-1}+1}\Big] = Z_{j(r-1)} A_{s_{r-1}} \mathrm{E}\Big[D_{s_{r-1},j_r} \mid \mathcal{B}_{s_{r-1}+1}\Big]$$

and

$$\begin{aligned}\mathrm{E}\Big[D_{s_{r-1},j_r} \mid \mathcal{B}_{s_{r-1}+1}\Big] &= \mathrm{E}\Big[g^{\ell_r}_{s_{r-1}+j_r}(X_{s_{r-1}+j_r})\Big\{\prod_{k=s_{r-1}+1}^{s_{r-1}+j_r-1} A_k\Big\} \mid \mathcal{B}_{s_{r-1}+1}\Big] \\ &= \mathrm{E}_{X_{s_{r-1}+1}}\Big[g^{\ell_r}_{s_{r-1}+j_r}(X_{j_r-1})\Big\{\prod_{k=1}^{j_r-1} A_k\Big\}\Big] \qquad (A.5)\end{aligned}$$

so that

$$\sum_{j_r=1}^{\infty} \mathrm{E}\Big[D_{s_{r-1},j_r} \mid \mathcal{B}_{s_{r-1}+1}\Big] = \phi^{l_r,0}_{s_{r-1}}(X_{s_{r-1}+1}) \qquad (A.6)$$

where

$$\phi^{l_r,0}_{s_{r-1}} = \sum_{j_r=1}^{\infty} H^{j_r-1} g^{\ell_r}_{s_{r-1}+j_r}$$

and

$$\begin{aligned}\phi^{l_r}_{s_{r-1}}(X_{s_{r-1}}) &\stackrel{\text{def}}{=} \mathrm{E}\Big[A_{s_{r-1}} \phi^{l_r,0}_{s_{r-1}}(X_{s_{r-1}+1}) \mid \mathcal{B}_{s_{r-1}}\Big] \\ &= \mathrm{E}_{X_{s_{r-1}}}\Big[\phi^{l_r,0}_{s_{r-1}}(X_1) Y_0\Big] \\ &= H\phi^{l_r,0}_{s_{r-1}}(X_{s_{r-1}}) \qquad .\end{aligned}$$

The product (A.4) is reduced from r to $r-1$ since

$$\begin{aligned}g^{l_{r-1}}_{s_{r-1}}(X_{s_{r-1}}) \cdot H\phi^{l_r}_{s_{r-1}}(X_{s_{r-1}}) &= [I_{g^{l_{r-1}}_{s_{r-1}}} H\phi^{l_r}_{s_{r-1}}](X_{s_{r-1}}) \\ &= g^{l_{r-1},0}_{s_{r-1}}(X_{s_{r-1}}), \quad \text{say} \ . \qquad (A.7)\end{aligned}$$

Hence, by (A.3), (A.5), (A.6), (A.7) we obtain

$$\begin{aligned}\mathrm{E}_x J_{r,\ell} &= \sum_{j(r-1)\in\mathcal{N}^{r-2}} Z_{j(r-1),\ell} \sum_{j_r=1}^{\infty} \mathrm{E}\Big[D_{s_{r-1},j_r} \mid \mathcal{B}_{s_{r-1}+1}\Big] \\ &= \mathrm{E}_x\Big\{\sum_{j(r-1)\in\mathcal{N}^{r-2}} Z_{j(r-1),\ell}\, \phi^{l_r,0}_{s_{r-1}}(X_{s_{r-1}})\Big\} \\ &= \mathrm{E}_x\Big\{\sum_{j(r-1)\in\mathcal{N}^{r-2}} Z'_{j(r-1),\ell}\Big\}\end{aligned}$$

where

$$Z'_{j(r-1),\ell} = g_{s_1}^{\ell_1}(X_{s_1}) \cdots g_{s_{r-2}}^{\ell_{r-2}}(X_{s_{r-2}}) g_{s_{r-1}}^{\ell_{r-1},0}(X_{s_{r-1}}) \prod_{k=0}^{s_{r-1}-1} A_k \ .$$

By the induction hypotheses we are through since

$$I_{g_{s_{r-1}}^{l_{r-1},0}} 1(x) = g_{s_{r-1}}^{l_{r-1},0}(x) = I_{g_{s_{r-1}}^{l_{r-1}}} \Big[\sum_{j_r=1}^{\infty} H^{j_r} g_{s_{r-1}+j_r}^{\ell_r} \Big](x) \ .$$

□

Corollary A.1 *Let* $U_{\underline{a}} = \sum_{j=0}^{\tau} a_j g(X_j)$. *Then* $\mathrm{E}_x\Big[U_{\underline{a}}^m\Big]$ *is given by* (A.1) *with*

$$\psi_{r,\ell} = \sum_{j \in \mathcal{N} \times \mathcal{N}_+^{r-1}} d_{j,\ell}(x) \Big\{ a_{j_1}^{\ell_1} a_{j_1+j_2}^{\ell_2} \cdots a_{j_1+\cdots+j_r}^{\ell_r} \Big\} \qquad (A.8)$$

where

$$d_{j,\ell} = H^{j_1} I_{g^{\ell_1}} \cdots H^{j_r} I_{g^{\ell_r}} 1 \ .$$

In particular for $m = 1, 2$ *we have that*

$$\mathrm{E}_x\Big[U_{\underline{a}}\Big] = \sum_{j=0}^{\infty} d_j(x) a_j, \quad d_j = H^j I_g 1$$

$$\mathrm{E}_x\Big[U_{\underline{a}}^2\Big] = \sum_{j=0}^{\infty} d_{j,0}(x) a_j^2 + 2 \sum_{j=0}^{\infty} \sum_{\ell=1}^{\infty} d_{j,\ell}(x) \{a_j a_{\ell+j}\}, \quad d_{j,\ell} = H^j I_g H^\ell I_g 1 \ . \qquad (A.9)$$

Proof of Corollary A.1:
We obtain (A.8) from (A.2) with $g_j = a_j g$. With $m = 2$ we have

$$\begin{aligned}
\mathrm{E}_x\Big[U_{\underline{a}}^2\Big] &= \sum_{\ell \in \Delta_{2,1}} \binom{2}{\ell} \psi_{r,\ell}(x) + \sum_{\ell \in \Delta_{2,2}} \binom{2}{\ell} \psi_{r,\ell}(x) \\
&= \psi_{1,2}(x) + 2\psi_{2,(1,1)}(x) \\
&= \Big[\sum_{j_1=0}^{\infty} H^{j_1} I_{g_{j_1}^2} \Big] 1(x) + 2 \Big[\sum_{j_1=0}^{\infty} H^{j_1} I_{g_{j_1}} \Big] \Big[\sum_{j_2=1}^{\infty} H^{j_2} I_{g_{j_1+j_2}} \Big] 1(x) \\
&= \sum_{j=0}^{\infty} H^j g_j^2(x) + 2 \sum_{j=0}^{\infty} \sum_{s=1}^{\infty} H^j I_{g_j} H^s g_{j+s}(x) \ .
\end{aligned}$$

Hence

$$\mathrm{E}_\nu\Big[U_{\underline{a}}^2\Big] = \sum_{j=0}^{\infty} a_j^2 \{\nu H^j g^2\} + 2 \sum_{j=0}^{\infty} \sum_{s=1}^{\infty} a_j a_{j+s} \{\nu H^j I_g H^s g\} \ .$$

□

Remark A.2 *In particular if* $a_j \equiv 1$ *then* (A.9) *gives the formulae*

$$\mathrm{E}_\nu\Big[U_{\underline{a}}\Big] = \pi_s g, \quad \mathrm{E}_\nu\Big[U_{\underline{a}}^2\Big] = \pi_s g^2 + 2\pi_s I_g H G_{s,\nu} g \ . \qquad (A.10)$$

A.2 Moment inequality

Lemma A.1 *Assume that (2.1) holds. Let $p > 1$ and $\eta \in (0,1)$ and f a real–valued measurable function defined on E. Then for any probability measure λ*

$$\lambda\Big[G_{s,\nu} f\Big]^t \leq c_2\, \mathrm{E}_\lambda^{t/p}\Big[\tau^{1+2(p-1)}\Big] \sup_{j\geq 0} \lambda^{t/q} P^j |f|^q,$$

$$t = p/(1+\eta(p-1)), \quad q = p/\eta(p-1),$$

and where c_2 is a universal constant only dependent upon p and η.

Proof of Lemma A.1:

Let $q' = p/(p-1)$, $r = q'/(1-\eta)$, $w = 1/q'$, $v = 1/q'$, $u = 2/q'$. Then $u = v + w$, $p^{-1} + q^{-1} + r^{-1} = 1$, $1/t = 1/p + 1/q$, $pu = 2(p-1)$ and $qv = \eta^{-1}$.

From the right hand side of (2.4) and by the Cauchy Schwartz inequality, we get

$$\begin{aligned}
\Big[G_{s,\nu} f\Big]^t(x) = & \Big[\sum_j \mathrm{E}_x\Big\{1(\tau \geq j) f(X_j)\Big\}\Big]^t \\
\leq & \Big[\sum_j \mathrm{P}_x^{1/p}(\tau \geq j) \mathrm{E}_x^{1/q}\Big\{|f|^q(X_j)\Big\}\Big]^t \\
\leq & \Big[\sum_j \mathrm{P}_x^{1/p}(\tau \geq j) \Big\{P^j |f|^q(x)\Big\}^{1/q}\Big]^t \\
= & \Big[\sum_j \Big(j^u \mathrm{P}_x^{1/p}(\tau \geq j)\Big)\Big(j^{-v}\Big\{P^j|f|^q(x)\Big\}^{1/q}\Big)\Big(j^{-w}\Big)\Big]^t \\
\leq & \Big(\sum_j j^{up} \mathrm{P}_x(\tau \geq j)\Big)^{t/p} \Big(\sum_j j^{-vq} P^j |f|^q(x)\Big)^{t/q} \Big(\sum_j j^{-wr}\Big)^{t/r} \\
= & c_1 V^{t/p} Z^{t/q}, \text{ say.}
\end{aligned}$$

We apply the Cauchy Schwartz inequality with with $p_1 = p/t$ and $q_1 = q/t$. This gives

$$\begin{aligned}
\lambda\Big[G_{s,\nu}\Big]^t f \leq & c_1 \lambda V^{t/p} Z^{t/q} \\
\leq & c_1 \Big[\lambda^{t/p} V\Big]\Big]\lambda^{t/q} Z\Big] \\
= & c_1 \Big[\sum_{j=0}^\infty j^{pu} \mathrm{P}_\lambda(\tau \geq j)\Big]^{t/p} \Big[\sum_{j=0}^\infty j^{-qv} \lambda P^j |f|^q\Big]^{t/q} \\
= & c_1 \Big[\sum_{j=0}^\infty j^{2(p-1)} \mathrm{P}_\lambda(\tau \geq j)\Big]^{t/p} \Big[\sum_{j=0}^\infty j^{-\eta^{-1}} \lambda P^j |f|^q\Big]^{t/q} \\
\leq & c_2 \Big[\sum_{j=0}^\infty j^{2(p-1)} \mathrm{P}_\lambda(\tau \geq j)\Big]^{t/p} \sup_{j\geq 0} \lambda^{t/q} P^j |f|^q \\
\leq & c_2 \mathrm{E}_\lambda^{t/p} \tau^{1+2(p-1)} \sup_{j\geq 0} \lambda^{t/q}(P^j |f|^q)
\end{aligned}$$

and

$$c_2 = \Big[\sum_{j=0}^\infty j^{-wr}\Big]^{t/r} \Big[\sum_{j=0}^\infty j^{-\eta^{-1}}\Big]^{t/q}.$$

□

A.3 Asymptotic formulae

Let k be an integer. Then (cf. (4.35))

$$\gamma_{g,f}(k) = \pi_s I_g P^{|k|} I_f 1$$

and $\gamma_g(k) = \gamma_{g,g}(k)$.

Lemma A.2 *Assume that g is a small function in one variable such that $\mu_g = \pi_s g = 0$. Then*

$$\sigma_g^2 = \sum_{\ell=-\infty}^{\infty} \gamma_g(\ell)$$

and when $\mu_g = \pi_s g \neq 0$, then

$$\sigma_g^2 = \pi_s(g)\pi_s(g - g \cdot s) + \sum_{\ell=-\infty}^{\infty} \gamma_{(g,g-s\mu_g)}(\ell) \ . \tag{A.11}$$

Proof of Lemma A.2:

We have by (A.10) that

$$\begin{aligned} \sigma_g^2 = &\ \pi_s g^2 - \pi_s^2 g + 2\pi_s I_g P G_{s,\nu} g - 2\mu_g \pi_s(sg) \\ = &\ \pi_s g^2 + \mu_g^2 + 2\pi_s I_g P G_{s,\nu}(g - s\mu_g) - 2\mu_g \pi_s(sg) \ . \end{aligned}$$

Let $G^{(n)} = \sum_{j=0}^{n-1} P^j$. Then

$$G_{s,\nu} = G^{(n)} + P^n G_{s,\nu} - G^{(n)} \otimes \pi_s \ . \tag{A.12}$$

Let $\lambda = \pi_s I_g$ and $f = G_{s,\nu}(g - s\mu_g)$. Then $|\lambda| = \pi_s I_{|g|}$ (the total variation of λ) and by (A.12),

$$\pi_s I_g P G_{s,\nu}(g - s\mu_g) = \sum_{j=1}^{n} \pi_s I_g P^j I_{(g-s\mu_g)} 1 + \lambda P^n f = \sum_{j=1}^{n} \gamma_{(g,g-s\mu_g)}(j) + \lambda P^n f \ .$$

Since g is small, the function f is bounded. By Nummelin(1984, Theorem 6.10, p 112),

$$|\lambda| P^n |f| = o(1) \ .$$

Hence

$$2\pi_s I_g P G_{s,\nu}(g - s\mu_g) = \sum_{\ell=-\infty}^{\infty} \gamma_{(g,g-s\mu_g)}(k) - \pi_s g^2 + \mu_g \pi_s(gs) \ .$$

Thus (A.11) holds.

□

Lemma A.3 *Let $g_0, f_0 \geq 0$ be real-valued functions such that $\pi_s I_{g_0} PG_{s,\nu} f_0 < \infty$. Moreover, let $\{g_m\}$ and $\{f_n\}$ be sequences of real-valued functions such that $g_m \underset{m}{\longrightarrow} g$ a.s. $[\pi_s]$, $f_n \underset{n}{\longrightarrow} f$ a.s. $[\pi_s]$ and $\sup_m |g_m| \leq g_0$, $\sup_n |f_n| \leq f_0$. Then*

$$\lim_{m,n} \pi_s I_{g_m} PG_{s,\nu} f_n = \pi_s I_g PG_{s,\nu} f \ .$$

Proof of Lemma A.3: Let $\xi_n = PG_{s,\nu} f_n$ and $\xi_0 = PG_{s,\nu} f_0$ so that $|\xi_n| \leq \xi_0$. We have to prove that

$$\xi_n \underset{n}{\longrightarrow} \xi = PG_{s,\nu} f \qquad \text{a.s. } [\pi_s]$$

Let D be the set of points where f_n fails to converge towards f. Then $\pi_s\{G 1_D > 0\} = 0$ with $G = \sum_{\ell=0}^{\infty} P^\ell$ since π_s is a maximal irreducibel measure. Hence $\pi_s\{PG_{s,\nu} 1_D > 0\} = 0$. The rest of the proof follows directly from the standard dominated convergence theorem since

$$\pi_s I_{g_m} PG_{s,\nu} f_n = \pi_s I_{g_m} \xi_n = \pi_s (g_m \cdot h_n),$$

where $|(g_m \cdot \xi_n)| \leq (g_0 \cdot \xi_0)$ and

$$(g_m \cdot \xi_n) \underset{m,n}{\longrightarrow} (g \cdot h) \qquad \text{a.s. } [\pi_s] \ .$$

□

A candidate for an approximating sequence $\{g_m,\}$ is given by

$$g_m = g 1_{C_m} \tag{A.13}$$

where $\{C_m\}$ be an increasing sequence of small sets so that $C_m \uparrow E$ and g is bounded on C_m for all m.

Theorem A.2 *Let $L^p(\pi_s)$ be the space of L^p π_s-integrable functions and assume that*

$$g \in L^1(\pi_s) \cap L^2(\pi_s), \quad \pi_s g^2 + 2\pi_s I_{|g|} PG_{s,\nu} |g| < \infty \ . \tag{A.14}$$

Let $\{g_m\}$ be given by (A.13) and $\mu_m = \mu_{g_m}$. Then for each ℓ

$$\gamma_{g_m, g_m - s\mu_m}(\ell) \underset{m}{\longrightarrow} \gamma_{g, g - s\mu_g}(\ell) \tag{A.15}$$

and

$$\mathrm{E}_\nu U_g^2 = \pi_s(g)\pi_s(g - g \cdot s) + \lim_m \Big\{ \sum_{\ell=-\infty}^{\infty} \gamma_{g_m, g_m - s\mu_m}(\ell) \Big\} \ .$$

Suppose that for all $m \geq 1$ and $\ell > 0$,

$$|\gamma_{g_m, g_m - s\mu_m}(\ell)| \leq a_\ell, \quad \sum_{\ell=1}^{\infty} a_\ell < \infty \ . \tag{A.16}$$

Then with absolutely convergence,

$$\mathrm{E}_\nu U_g^2 = \sum_{\ell=1}^{\infty} \gamma_g(\ell) \tag{A.17}$$

when $\pi_s g = 0$ and in general

$$\mathrm{E}_\nu U_g^2 = \pi_s(g)\pi_s(g - g\cdot s) + \sum_{\ell=1}^{\infty} \gamma_{g,g-s\mu}(\ell)\ .$$

Proof of Theorem A.2:

Let $f_m = g_m - s\pi_s g_m$, $g_0 = |g|$ and $f_0 = |g| + s$. By Lemma A.2 and Lemma A.3

$$\begin{aligned}
\mathrm{E}_\nu U_g^2 &= \pi_s g^2 + \pi_s^2 g - 2\pi_s g\pi_s g\cdot s + 2\pi_s I_g P G_{s,\nu}(g - s\mu_g) \\
&= \pi_s g^2 + \pi_s^2 g - 2\pi_s g\pi_s g\cdot s + \lim_m \big\{2\pi_s I_{g_m} P G_{s,\nu}(g_m - s\mu_m)\big\} \\
&= \pi_s g^2 + \pi_s^2 g - 2\pi_s g\pi_s g\cdot s + \lim_m \Big\{2\sum_{\ell=1}^{\infty} \gamma_{g_m,g_m-s\mu_m}(\ell)\Big\} \\
&= \pi_s g\pi_s(g - g\cdot s) + \lim_m \Big\{\sum_{\ell=1}^{\infty} \gamma_{g_m,g_m-s\mu_m}(\ell)\Big\}\ .
\end{aligned}$$

It is obvious that (A.15) holds and if (A.16) holds then

$$\sum_{\ell=1}^{\infty} \gamma_{g_m,g_m-s\mu_m}(\ell) \underset{m}{\longrightarrow} \sum_{\ell=1}^{\infty} \gamma_{g,g-s\mu}(\ell) \tag{A.18}$$

by the dominated convergence theorem. From (A.14) and (A.18) we can conclude that (A.17) is true.

□

B Appendix

Proof of Lemma 4.2:

Let $W_\tau = \{W_{\tau_k^1},\ k \geq 0\}$. Then, we start by showing that

$$W_\tau \overset{\mathrm{d}}{=} \underset{\sim}{\mathrm{W}} \quad \text{when } \underset{\sim}{\lambda} = \widetilde{\lambda} \tag{B.1}$$

where

$$\widetilde{\lambda} = \lambda_2\Phi_{\lambda_1}, \quad \Phi_{\lambda_1} = \sum_{\ell=0}^{\infty}\{\lambda_1 H_1^\ell s_1\} P_2^\ell, \quad \underset{\sim}{\mathrm{P}} = P_2\Phi_{\nu_1}\ . \tag{B.2}$$

In order to prove (B.1) it is enough to show that for all $r \geq 0$ and for all $A_i \in \mathcal{E}_2^+$,

$$\mathrm{P}_\lambda\Big(W_{\tau_0^1} \in A_0, \ldots, W_{\tau_0^r} \in A_r\Big) = \mathrm{P}_{\widetilde{\lambda}}\Big(\underset{\sim}{\mathrm{W}}_0 \in A_0, \ldots, \underset{\sim}{\mathrm{W}}_r \in A_r\Big)\ . \tag{B.3}$$

Let $k_0 = j_0$ and $k_\ell = j_0 + j_1 + \cdots + j_\ell$ for $\ell = 0, \ldots, r$. We have

$$\begin{aligned}
&\mathrm{P}_\lambda\Big(W_{\tau_0^1} \in A_0 \cdots W_{\tau_0^r} \in A_r\Big) \\
&\quad = \sum_{j_0=0}^{\infty} \sum_{j_1=1}^{\infty} \cdots \sum_{j_r=1}^{\infty} \mathrm{P}_{\lambda_2}\Big(W_{k_0} \in A_0 \cdots W_{k_r} \in A_r\Big) \mathrm{P}_{\lambda_1}\Big(\tau_0^1 = j_0, \ldots, \tau_r^1 = j_r\Big) \\
&\quad = \sum_{j_0=0}^{\infty} \sum_{j_1=1}^{\infty} \cdots \sum_{j_r=1}^{\infty} \Big\{\lambda_2 P_2^{j_0} I_{A_0} \cdots P_2^{j_r} I_{A_r} 1\Big\}\Big\{\{\lambda_1 H_1^{j_0} s_1\} b_{j_1} \cdots b_{j_r}\Big\} \\
&\quad = \widetilde{\lambda} I_{A_0} \underset{\sim}{\mathrm{P}} I_{A_1} \cdots \underset{\sim}{\mathrm{P}} I_{A_r} 1 \qquad (B.4)
\end{aligned}$$

where

$$b_\ell = \nu_1 H_1^{\ell-1} s_1, \quad \ell \geq 0 \,.$$

Hence (B.3) holds.

From

$$\underset{\sim}{\mathrm{P}} \geq (s_2 \otimes \nu_2)\Phi_{\nu_1} = s_2 \otimes \nu_2 \Phi_{\nu_1}$$

we obtain the minorization inequality (4.14). Let $\underset{\sim}{\mathrm{H}} = \underset{\sim}{\mathrm{P}} - \underset{\sim}{s} \otimes \underset{\sim}{\nu}$. Then

$$\underset{\sim}{\mathrm{H}} = P_2 \Phi_{\nu_1} - s_2 \otimes \nu_2 \Phi_{\nu_1} = (P_2 - s_2 \otimes \nu_2)\Phi_{\nu_1} = H_2 \Phi_{\nu_1}, \quad \underset{\sim}{\mathrm{Q}} = Q_2 \Phi_{\nu_1}$$

with $Q_2(x, A) = (1 - s_2(w))^{-1} H_2(w, A) 1(s_2(w) < 1) + 1(s_2(w) = 1)$. The next task is to prove (4.15), i.e.

$$W'_\tau = (W_\tau, Y_\tau^2) \overset{\mathrm{d}}{=} \underset{\sim}{\mathrm{W}}^\star \quad \text{when } \underset{\sim}{\lambda} = \widetilde{\lambda}$$

where $\underset{\sim}{\mathrm{W}}^\star$ denotes the split chain generated by $\underset{\sim}{\mathrm{P}}$ and $(\underset{\sim}{s}, \underset{\sim}{\nu})$. Let $\underset{\sim}{\mathrm{P}}^\star$ be the transition probability function for this split chain and let $\underset{\sim}{\mathrm{P}}'$ be the transition probability for W'_τ. We have to prove that

$$\underset{\sim}{\mathrm{P}}' = \underset{\sim}{\mathrm{P}}^\star \,.$$

First we recall the structure of a split chain. Suppose that P is a transition probability which satisfies $P \geq s \otimes \nu$. Then the corresponding the split chain has transition probability $\underline{P}$ which satisfies,

$$\begin{aligned}
\underline{P}^n(x_0 \times y_0, dx \times y) &= y_0 \nu P^{n-1}(dx)\Big[y s(x) + (1-y)(1 - s(x))\Big] \\
&\quad + (1 - y_0) Q P^{n-1}(x_0, dx)\Big[y s(x) + (1-y)(1 - s(x))\Big], \quad n \geq 1 \qquad (B.5)
\end{aligned}$$

In our case this gives for $n = 1$;

$$\begin{aligned}
\underset{\sim}{\mathrm{P}}^\star(w_0 \times y_0, dw \times y) &= y_0 \underset{\sim}{\nu}(dw)\Big[y \underset{\sim}{s}(w) + (1-y)(1 - \underset{\sim}{s}(w))\Big] \\
&\quad + (1 - y_0) \underset{\sim}{\mathrm{Q}}(w_0, dw)\Big[y \underset{\sim}{s}(w) + (1-y)(1 - \underset{\sim}{s}(w))\Big] \qquad (B.6)
\end{aligned}$$

by definition of the split chain. We look closer at $\underset{\sim}{\mathrm{P}}'$ which by (B.2) satisfies

$$\underset{\sim}{\mathrm{P}}' = \sum_{\ell=1}^{\infty} b_\ell \underline{P}_2^\ell$$

We replace $\underline{P}_2^\ell$ by the right hand side of the following expression

$$\underline{P}_2^\ell(w_0 \times y_0, dw \times y) = y_0\nu_2 P_2^{\ell-1}(dw)\Big[ys(w) + (1-y)(1-s_2(w))\Big]$$
$$+(1-y_0)Q_2P_2^{\ell-1}(w_0, dw)\Big[ys(w)+\Big](1-y)(1-s(w)) \quad (B.7)$$

and

$$\sum_{\ell=1}^{\infty} b_\ell \nu_2 P_2^{\ell-1} = \nu_2\Phi_{\nu_1} = \underset{\sim}{\nu}, \quad \sum_{\ell=1}^{\infty} b_\ell Q_2 P_2^{\ell-1} = Q_2\Phi_{\nu_1} = \underset{\sim}{\mathrm{Q}} \ . \qquad (B.8)$$

Then we obtain (4.15) from (B.6) -(B.8). The first equality in (4.16) follows from (4.15) and the second one is the occupation formulae given by (2.4). When $\lambda = \lambda_1 \times \pi_2$ we get

$$\widetilde{\lambda} = \pi_2\Phi_{\lambda_1} = \sum_{\ell=0}^{\infty}\{\lambda_1 H_1^\ell s_1\}\pi_2 P^\ell = \pi_2\{\lambda_1 G_{s_1,\nu_1} s_1\} = \pi_2$$

If $\lambda = \nu = \nu_1 \times v_2$, then $\widetilde{\lambda} = \underset{\sim}{\nu}$ and $\underset{\sim}{\nu}\underset{\sim}{\mathrm{G}}_{\underset{\sim}{s},\underset{\sim}{\nu}} = \pi_2$ since

$$\underset{\sim}{\pi}_{\underset{\sim}{\mathrm{s}}} = \pi_{s_2} \ .$$

□

Proof of Lemma 4.3: The waiting times, $\{\delta_j,\ j \geq 0\}$ are given by $\delta_j = \tau_j^1 - \tau_{j-1}^1$. Let

$$b_{n,k} = P(\delta_1 + \ \cdots\ \delta_n = k), \quad k \geq n$$

and $b_{1,k} = b_k$. Then

$$b_{n,k} = \begin{cases} \nu_1 H_1^{k-1} s_1, & \text{for } n = 1 \text{ and } k \geq 1 \\ b_k^{\star n}, & \text{for } n \geq 1 \text{ and } k \geq 1 \end{cases}$$

where "$\star n$ " denotes n-times convolution. The n-steps transition probability $\underset{\sim}{\mathrm{P}}^n$ is given by

$$\underset{\sim}{\mathrm{P}}^n = \Big\{\sum_{j=0}^{\infty} b_{n,n+j} P_2^{n+j}\Big\}, \quad n \geq 1 \ . \qquad (B.9)$$

Since W is geometric ergodic, Nummelin(1984, Th.6.14) thereby there exists a non-negative function M so that $\pi_2(M) < \infty$ and a constant $\rho \in (0,1)$ so that

$$\|P_2^n(x,\cdot) - \pi_2\| \leq M(x)\rho^n, \quad x \in E, \quad n \geq 0 \ .$$

Thus by (B.9)

$$\begin{aligned} \|\underset{\sim}{\mathrm{P}}^n(x,\cdot) - \pi\| &\leq \sum_{j=0}^{\infty} b_{n,n+j}\|P_2^n(x,\cdot) - \pi\| \\ &\leq \sum_{j=0}^{\infty} b_{n,n+j} M(x)\rho^n \\ &\leq M(x)\rho^n \sum_{j=0}^{\infty} b_{n,n+j}\rho^j \\ &\leq M(x)\rho^n \ . \end{aligned} \qquad (B.10)$$

Hence by (B.10) the $\underset{\sim}{W}$-process is geometric ergodic.

For the ergodic W-process, we have that

$$\alpha_\ell = \sup_{A,B\in E} \theta_\ell(A,B), \quad \theta_\ell(A,B) = \pi I_A P_2^\ell I_B 1 - \pi 1_A \pi 1_B \ .$$

Here,

$$\underset{\sim\ell}{\theta}(A,B) = \pi I_A \underset{\sim}{\mathrm{P}}^\ell I_B 1 - \pi 1_A \pi 1_B = \sum_{j=\ell}^{\infty} b_{\ell,j}\Big\{I_A P_2^j I_B - \pi 1_A \pi 1_B\Big\} = \sum_{j=\ell}^{\infty} b_{\ell,j}\theta_j(A,B) \ .$$

That is

$$\underset{\sim\ell}{\alpha} \le \sum_{j=\ell}^{\infty} b_{\ell,j}\Big\{\sup_{A,B\in E} \theta_j(A,B)\Big\} = \sum_{j=\ell}^{\infty} b_{\ell,j}\alpha_j \le \alpha_\ell \ . \qquad (B.11)$$

By Bolthausen (1982) in general

$$\sum_{\ell=1}^{\infty} \ell^k \alpha_\ell < \infty \quad \Longrightarrow \quad \mathrm{E}_{\pi_2}\tau_0^{k+1} < \infty \ .$$

By (B.11) it follows that

$$\sum_{\ell=1}^{\infty} \ell^k \alpha_\ell < \infty \quad \Longrightarrow \quad \sum_{\ell=1}^{\infty} \ell^k \underset{\sim\ell}{\alpha} < \infty \ .$$

Hence (4.17) is true. □

References

Aparicio, F. M. and Escribano, A. (1997) Searching for linear and nonlinear cointegration; a new approach. Working paper 97-65, Universidad Carlos III de Madrid, Statistics and Economics Series.

Bergstrøm, H. (1981) *Weak Convergence.* Academic Press, New York.

Billingsley, P. (1968) *Convergence of Probability.* John Wiley & Sons, New York.

Bolthausen, E. (1982) The Berry-Esseén Theorem for strongly mixing Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **60** 283 -289.

Cline, D. B. H. and Pu, H. H. (1997) Stability of nonlinear AR(1) time series with delay. *Texas A&M University, College Station TX 77843.*

Engle, R. F. and Granger, C. W. J. (1987) Co-integration and error correction: representation, estimation and testing. *Econometrica* **55** 251-276.

Granger, C. W. J. and Hallman, J. (1991) Long memory series with attractors. Oxford Bulletin of Economics and Statistics **53** 11-26.

Granger, C. W. J. (1995) Modelling nonlinear relationships between extended - memory variables. *Econometrica* **63** 265-279.

Hall, P. and Heyde, C. C. (1980) *Martingale Limit Theory* Academic Press, New York.

Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, Oxford.

Karlsen, H. A. and Tjøstheim, D. (1998) Nonparamteric estimation in null recurrent time series nonlinear generalization. *Sonderforschungsbereich* **373**, Humboldt-Universität zu Berlin. Under revision for Annals of Statistics.

Meyn, S. P. and Tweedie, R. L. (1994) *Markov Chains and Stochastic Stability.* Springer Verlag, London.

Myklebust, T., Karlsen, H., Tjøstheim, D. (2000) A Markov chain characterization of unit root processes. and the problem of nonlinear cointegration. In preparation.

Nummelin, E. (1984) *General Irreducible Markov Chains and Non-negative Operators.* Cambridge University Press, Cambridge.

Phillips, P. C. B. and Park, J. (1998) Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper No. 1181, Yale University.

Xia, Y. (1998) Doctoral thesis University of Hong Kong.