STRONG LAWS OF LARGE NUMBERS FOR DEPENDENT HETEROGENEOUS PROCESSES: A SYNTHESIS OF RECENT AND NEW RESULTS

James Davidson
University of Wales, Cardiff
Cardiff Business School
Colum Drive,
Cardiff CF1 3EU
UK

Robert de Jong
Tilburg University
Department of Econometrics
PO Box 90153
5000 LE Tilburg
The Netherlands

Keywords: Strong law of large numbers, mixing, mixingales, near-epoch dependence

JEL Classification: C19

Abstract

This paper surveys recent developments in the strong law of large numbers for dependent heterogeneous processes. We prove a generalised version of a strong law for L_2-mixingales that seems to have gone unnoticed in the econometrics literature, and also a new strong law for L_p-mixingales. These results greatly relax the dependence and heterogeneity conditions relative to the results that are currently cited, at the same time as introducing explicit trade-offs between dependence and heterogeneity. The results are applied to proving strong laws for near-epoch dependent functions of mixing processes. We contrast several methods for obtaining these results, including mapping directly to the mixingale properties, and applying a truncation argument.

1 Introduction

This paper offers a unifying treatment of the strong law of large numbers for dependent heterogeneous processes. The main results are based on three strong laws for mixingale sequences. One of these is essentially a result by Masry and Györfi (1987), but in the more general version of De Jong (1994) who independently obtained the same result for nonadapted mixingales; the second is Theorem 2.3 of the present paper; and the third is given in De Jong (1995a). The novel feature of these results is that they permit a direct trade-off between the degrees of dependence and heterogeneity. They also permit much more dependence than is allowed in, for example, the strong laws of McLeish (1975), Hansen (1991,1992), and Davidson (1994, Theorem 20.21). These theorems impose the requirement that the mixingale numbers be either square-summable, for processes bounded in L_2-norm, or otherwise summable. By contrast, the new results allow local dependence which can decline at an arbitrarily slow rate as the lag increases. Related results are given in De Jong (1995b), which provides the only strong law of large numbers for triangular mixingale arrays that is available at this moment.

We characterize a strong law of large numbers as a theorem citing conditions such that $a_n^{-1} \sum_{t=1}^{n} X_t \to 0$ a.s., where a_n is a sequence of positive constants tending to infinity. The standard case of the sample mean $\overline{X}_n$ has $a_n = n$, but the more general formulation is useful in several contexts, and introduces at worst a minor complication into the proofs. On the one hand, to have proved the strong law in respect of a sequence a_n going to infinity more slowly than n places a bound on the rate of convergence of the sample mean, in the sense that $\overline{X}_n = O(a_n/n)$, almost surely. On the other hand, if the sample mean fails to converge, identifying a sequence a_n for which convergence does obtain provides information about the behaviour of the sample mean itself. For example, if the strong law holds for the case $a_n = n(\log n)^{\delta}$ for any real δ, we know that $\overline{X}_n$ is at worst slowly varying in the limit.

Section 2 of the paper discusses the mixingale results. Section 3 gives extensions to the class of dependent processes which can be characterized as near-epoch dependent functions of a mixing process. Section 4 presents a systematic comparison of the various strong laws, and shows that all the results proved in the paper have a contribution to make in certain circumstances. Section 5 concludes the paper, and the various proofs are gathered in Section 6.

2 Mixingale Strong Laws

In a probability space $(\Omega, \mathcal{F}, P)$, consider a sequence $\{X_t, \mathcal{F}_t\}_{-\infty}^{\infty}$ where the first member of the pair is a real, $\mathcal{F}$-measurable, zero-mean stochastic sequence and the second member is an increasing sequence of sub-σ-fields of $\mathcal{F}$. Following McLeish (1975), and Andrews (1988), the sequence is called a L_p-mixingale with respect to constants $\{c_t\}_{-\infty}^{\infty}$ and mixingale numbers $\{\psi_j\}_0^{\infty}$, if for all t and all $j \geq 0$,

$$\|E(X_t \mid \mathcal{F}_{t-j})\|_p \leq c_t \psi_j,$$

$$\|X_t - E(X_t \mid \mathcal{F}_{t+j})\|_p \leq c_t \psi_{j+1}.$$

We say that the mixingale is of size $-\lambda$ if $\psi_j = O(j^{-\lambda-\varepsilon})$ for $\varepsilon > 0$. See for example Davidson (1994) Chapter 16, and also Gallant and White (1988), Hall and Heyde (1980), and Pötscher and Prucha (1991), for additional details.

The size of the sequence is a useful summary measure of the decline of dependence. It is not obvious that this is so, for the mixingale strong laws given here and elsewhere characteristically impose conditions of the form $\sum_{j=0}^{\infty}\psi_j^r a_j < \infty$ for some sequence $\{a_j\}$, and some $r > 0$. Take the case with $a_j = 1$ for illustration. A mixingale size of $-1/r$ is sufficient for summability, but evidently not necessary (for example, let $\psi_j = 0$ except when j is the square of a natural number). In such cases, size conditions would be excessively stringent constraints on dependence. However, such cases can be ruled out since there is virtually no loss of generality in assuming that the mixingale numbers are monotone. In view of the monotonicity of the sequence $\{\mathcal{F}_t\}$, they must satisfy the inequalities $\psi_{j+k} \le 2\psi_j$, for $k > 0$ (see Davidson 1994, Theorems 10.27 and 10.28). The monotonized sequence $\psi'_m = \sup_{j\ge m}\psi_j$ has the property $\psi'_j \le 2\psi_j$ for every $j > 0$.

The standard applications of mixingale theory to the strong law of large numbers are reviewed in Davidson (1994), Chapter 20. There are, characteristically, separate results for the cases $p = 2$ and $1 < p < 2$. The following composite theorem gives the generalized versions (i.e., allowing divisor $a_n \ne n$) of the theorems for sample means due to McLeish (1975) (the case $p = 2$) and Hansen (1991, 1992) (the case $1 < p < 2$).

Theorem 2.1 *Let $\{X_t, \mathcal{F}_t\}_{-\infty}^{\infty}$ be a L_p-mixingale, with either (i) $p = 2$, of size $-\frac{1}{2}$, or (ii) $1 < p < 2$, of size -1; if, for a sequence of positive constants $\{a_n\}_{n=0}^{\infty}$ such that $a_n \uparrow \infty$,*[1]

$$\sum_{t=1}^{\infty}(c_t/a_t)^p < \infty,$$

then $a_n^{-1}\sum_{t=1}^{n}X_t \to 0$, a.s.

(For proof see Davidson 1994, Corollary 20.16.).

We cite three theorems which jointly represent a substantial improvement over these results. The first is basically the extension to nonadapted L_2-mixingales of Masry and Györfi's (1987) Theorem 3.1.

Theorem 2.2 *Let $\{X_t, \mathcal{F}_t\}_{-\infty}^{\infty}$ be a L_2-mixingale, and a_n a sequence of positive constants such that $a_n/\sqrt{n} \uparrow \infty$. If for some $\delta > 0$,*

$$\sum_{t=1}^{\infty}\left(1 + \sum_{j=1}^{t}\psi_j^2 \log(j+1)(\log\log(j+2))^{1+\delta}\right)c_t^2/a_t^2 < \infty \tag{2.1}$$

then $a_n^{-1}\sum_{t=1}^{n}X_t \to 0$, a.s.

Note that condition (2.1) is equivalent to

$$\sum_{t=1}^{\infty}c_t^2/a_t^2 + \sum_{j=1}^{\infty}\psi_j^2\log(j+1)(\log\log(j+2))^{1+\delta}\sum_{t=j}^{\infty}c_t^2/a_t^2 < \infty. \tag{2.2}$$

Theorems 2.2 and 2.1 coincide in the martingale difference case, in which $\psi_j = 0$ for $j > 0$ and the square summability of the coefficients c_t/a_t is seen to be a sufficient condition. Expressing the summability condition in terms of size and trend numbers leads directly to the following.

Corollary 2.1 *Let $\{X_t, \mathcal{F}_t\}_{-\infty}^{\infty}$ be a L_2-mixingale of size $-\lambda$. If $a_n/\sqrt{n} \uparrow \infty$, and $c_t/a_t = O(t^{\alpha})$ where $\alpha < \min\{-1/2, \lambda - 1\}$, then $a_n^{-1}\sum_{t=1}^{n}X_t \to 0$ a.s.*

[1] The symbol $\uparrow$ is used to denote that the sequence in question increases monotonically.

This version of the strong law loses some sharpness, but is a convenient vehicle for applying the result, as we show in Section 3.

The next strong law, given here for the first time, covers the L_p-mixingale case with $p < 2$, and contains Hansen's theorem.

Theorem 2.3 *Let $\{X_t, \mathcal{F}_t\}_{-\infty}^{\infty}$ be a L_p-mixingale sequence for $1 < p \leq 2$, and a_n a sequence of positive constants such that $a_n/n^{1-1/p} \uparrow \infty$, and*

$$\sum_{j=0}^{\infty} \psi_j \left(\sum_{t=\max\{1,[j^{p/(p-1)}]\}}^{\infty} c_t^p a_t^{-p} \right)^{1/p} < \infty. \tag{2.3}$$

Then $a_n^{-1} \sum_{t=1}^{n} X_t \to 0$, a.s.

As before, there is a corollary summarizing the result in terms of size and trend numbers.

Corollary 2.2 *Let $\{X_t, \mathcal{F}_t\}_{-\infty}^{\infty}$ be a L_p-mixingale, $1 < p \leq 2$, of size $-\lambda$. If $a_n/n^{1-1/p} \uparrow \infty$, and $c_t/a_t = O(t^{\alpha})$ where $\alpha < \min\{-1/p, (1-1/p)\lambda - 1\}$, then $a_n^{-1} \sum_{t=1}^{n} X_t \to 0$ a.s.*

In the case $p = 2$, the condition in Corollary 2.2 reduces to

$$\alpha < \min\{-1/2, \lambda/2 - 1\}.$$

Comparing this with Corollary 2.1 shows how Theorem 2.2 improves on Theorem 2.3 in this case.

The last result, the generalized form of the mixingale strong law in De Jong (1995a), has already been given in Davidson (1994), Theorem 20.17, and we merely state here the corollary for size and trend numbers which may be derived from the latter theorem. This is as follows.

Corollary 2.3 *Let X_t be a L_p-bounded L_1-mixingale of size $-\lambda$, for $p \geq 1$, let c_t and a_t be regularly varying sequences with $c_t \leq \|X_t\|_p$ and $a_t \to \infty$, and $c_t/a_t = O(t^{\alpha})$ where*

$$\alpha < -\frac{(1+p)\lambda + 2(p-1)}{2p\lambda + 2(p-1)},$$

then $a_n^{-1} \sum_{t=1}^{n} X_t \to 0$ a.s.

A L_p-mixingale for $p > 1$ is also a L_1-mixingale with respect to the same constants, and so this set-up is comparable with the previous cases. Unlike those, this is a result for which mere integrability ($p = 1$) can be sufficient, although we would need $\alpha < -1$, so that there is no strong law for the sample mean in that instance. The regular variation requirement on the sequences c_t and a_t should be easy to satisfy in practice, even though it is not imposed explicitly by the previous results. Hence, this constraint on the parameters p, α, and λ is directly comparable to those of Corollaries 2.1 and 2.2.

Consider the 'standard' sample mean case, with $a_t = t$, and also assume that c_t is bounded uniformly in t. In this case, $\alpha = -1$. We see, by referring to either Corollary 2.2 or Corollary 2.3, that a.s. convergence of the sample mean holds whenever both $\lambda > 0$ and $p > 1$, and indeed, the first of these inequalities has to be strict only because of the coarsening of the result in the size-based Corollary 2.2. If we go directly to Theorem 2.3, we see that setting $c_t/a_t = O(t^{-1})$ reduces (2.3) to

$$\sum_{j=0}^{\infty} \psi_j j^{-1} < \infty, \tag{2.4}$$

and accordingly, it suffices for $\psi_j = O((\log j)^{-1-\delta})$ for $\delta > 0$, so that $\lambda = 0$ is permitted. These conditions may be compared with a L_1-convergence law for mixingales such as Davidson (1993) (see also Theorem 19.11 of Davidson 1994). Theorem 2.3 tightens the conditions of the latter result for the 'standard' case slightly, by substituting uniform boundedness of the sequence $\{\|X_t\|_p\}$, some $p > 1$, for uniform integrability of the sequence $\{X_t\}$, and also places *some* constraint on the rate of convergence of the ψ_j, beyond simply $\psi_j \to 0$ as $j \to \infty$. But this is a substantially smaller gap between dependent weak and strong laws than has been achieved previously, to our knowledge.

More generally, consider values of α exceeding -1. This could arise because of the c_t sequence trending upward over time, but if we continue to think of c_t as uniformly bounded, such a result shows convergence with $a_n = n^\beta$, with $\beta = -\alpha > 0$, and hence establishes $n^{\beta-1}$ as a bound on the rate at which $\overline{X}_n$ converges to 0, as pointed out in Section 1. Of particular interest here are the the relative strengths of Corollaries 2.2 and 2.3. Sample calculations show that Corollary 2.3 dominates Corollary 2.2 for values of p close to 1 and λ close to 0, but more generally the ranking is reversed. To illustrate, Figure 1 plots the suprema of permitted α against λ, for the case $p = 1.1$.

Combining the three corollaries, taking the maximum available upper bound on α for each pair (λ, p), we can propound a composite strong law for mixingales. Its statement would be algebraically cumbersome, but the corresponding constraint on the three parameters is graphed as a surface in three dimensions in Figure 2. Note that this diagram contains all the requisite information, since the profile of the surface in the p-direction is the same for all values of λ exceeding unity.

3 Strong Laws for NED Functions of Mixing Processes

A stochastic sequence $\{X_t\}_{-\infty}^{\infty}$, is said to be near-epoch-dependent in L_p-norm (or L_p-NED) on a (possibly vector-valued) process $\{V_t\}_{-\infty}^{\infty}$, with respect to constants d_t and numbers $\{\nu_m\}_0^\infty$, if

$$\left\|X_t - E(X_t \mid \mathcal{F}_{t-m}^{t+m})\right\|_p \le d_t \nu_m$$

where $\mathcal{F}_s^t = \sigma(V_s, ..., V_t)$ for $t \ge s$. It is said to be of size $-\lambda$ if $\nu_m = O(m^{-\lambda-\varepsilon})$ for $\varepsilon > 0$. See for example Davidson (1994) Chapter 17, or Gallant and White (1988), for additional details.

This has been proposed as a good characterization of the dependence structure of time series processes in economics which are generated by the distributed lag effect of a sequence of random shocks. For example, if the process has a MA(∞) representation,

$$X_t = \sum_{j=0}^{\infty} \theta_j V_{t-j},$$

it is L_p-NED on $\{V_t\}$ with respect to constants $d_t = \sup_{s\le t} \|V_s\|_p$ if $\sum_{j=0}^{\infty} |\theta_j| < \infty$, and in this case $\nu_m = \sum_{j=m+1}^{\infty} |\theta_j|$. The stationary invertible ARMA(p, q) process is a case in point, with $\lambda = \infty$, and considerably more dependence than this model generates is clearly permitted. On the other hand, a fractional-difference $(I(d))$ process with $d > 0$ violates the summability condition (see Hosking 1981, for example) and is accordingly characterized as having 'long memory' (or as we might say in the present context, far-epoch dependence on $\{V_t\}$).

After centering by subtraction of the mean, a L_p-NED function of an α-mixing or ϕ-mixing process is (subject to the existence of higher-order moments in the former case) a L_p-mixingale. Straightforward formulae are available for determining the mixingale size and mixingale constants, given the relevant NED characteristics and those of the underlying mixing processes. These results are summarized in, for example, Theorem 17.5 of Davidson (1994). In principle, it is therefore a simple matter to generate strong laws direct from the mixingale results. Although more primitive results are available direct from the theorems, it will suffice for most practical purposes to derive results from the various corollaries of the previous section. Apart from other considerations, it is only through these results that we can feasibly compare the theorems for particular values of the various parameters of the processes. These now include the mixing size (to be denoted by $-a$), the NED size (to be denoted by $-b$), the NED order (p), the maximum order of existing moments (q), and in addition, the rates of increase of the sequences $\| X_t - \mu_t \|_q$ and a_t.

In each of the following theorems, $\{X_t\}_{-\infty}^{\infty}$ is a L_q-bounded stochastic sequence, $q \geq 1$, with mean sequence $\{\mu_t\}_{-\infty}^{\infty}$, and is assumed to be L_p-NED, of size $-b$, with respect to constants $d_t \ll \| X_t - \mu_t \|_q$ for $1 \leq p \leq q$,[2] on $\{V_t\}_{-\infty}^{\infty}$ which is either (i) α-mixing or (ii) ϕ-mixing, of size $-a$. Each theorem has two parts, according to the mixing mode assumed. First we have a result based on Corollary 2.1.

Theorem 3.1 *Let $p = 2$, and let $a_n/\sqrt{n} \uparrow \infty$ as $n \to \infty$. If*

$$\frac{\| X_t - \mu_t \|_q}{a_t} = O(t^{\xi}), \tag{3.1}$$

where either (i) V_t is α-mixing, X_t is L_q-bounded for $q > 2$, and

$$\xi < \min\{-1/2, \min\{b, a(1/2 - 1/q)\} - 1\},$$

or (ii) V_t is ϕ-mixing, X_t is L_q-bounded for $q \geq 2$, and

$$\xi < \min\{-1/2, \min\{b, a(1 - 1/q)\} - 1\},$$

then $a_n^{-1} \sum_{t=1}^{n} (X_t - \mu_t) \to 0$ a.s.

Second, here is the result from Corollary 2.2:

Theorem 3.2 *Let $1 < p \leq 2$, and let $a_n/n^{1-1/p} \to \infty$ as $n \to \infty$. If*

$$\frac{\| X_t - \mu_t \|_q}{a_t} = O(t^{\xi}), \tag{3.2}$$

where either (i) V_t is α-mixing, $q > p$, and

$$\xi < \min\{-1/p, (1 - 1/p) \min\{b, a(1/p - 1/q)\} - 1\},$$

or (ii) V_t is ϕ-mixing, $q \geq p$, and

$$\xi < \min\{-1/p, (1 - 1/p) \min\{b, a(1 - 1/q)\} - 1\};$$

then $a_n^{-1} \sum_{t=1}^{n} (X_t - \mu_t) \to 0$, a.s.

Third, here is the result from Corollary 2.3.

---

[2] The symbol $x \ll y$ is used to denote that there exists a finite constant $C > 0$ such that $x \leq Cy$.

Theorem 3.3 *Let $1 \leq p \leq 2$, and let $a_n \to \infty$ as $n \to \infty$. If*

$$\frac{\| X_t - \mu_t \|_q}{a_t} = O(t^\xi), \tag{3.3}$$

where $q \geq p$ and $q > 1$, and

$$\xi < \min \left\{ -\frac{(1+q)b + 2(q-1)}{2qb + 2(q-1)}, -\frac{(1+q)a + 2q}{2q(a+1)} \right\},$$

and V_t is either (i) α-mixing or (ii) ϕ-mixing, then $a_n^{-1} \sum_{t=1}^n (X_t - \mu_t) \to 0$, a.s.

However, it is possible that better results for certain cases can be obtained using a truncation argument. With truncated random variables, one can set $p = 2$ to take advantage of the more powerful results for L_2-mixingales, and also set $q = \infty$ to get the best mapping from mixing properties to NED properties. One must find a way of showing that the remainder from the truncation also obeys a strong law, or can otherwise be ignored. Care is needed in specifying the truncations, because discontinuous transformations may not preserve the NED property. The specification of a suitable continuous truncation is discussed in Davidson (1994, Section 17.3).

There are various ways in which the truncation can be employed. Using the method introduced in De Jong (1995a), which proves the strong law directly for the remainder terms, leads to the following result.

Theorem 3.4 *Let a sequence $\{X_t\}_{-\infty}^{\infty}$ with means $\{\mu_t\}_{-\infty}^{\infty}$ be L_p-NED, $1 < p \leq 2$, of size $-b$, with respect to constants $d_t \ll \|X_t - \mu_t\|_p$, on a sequence $\{V_t\}_{-\infty}^{\infty}$ which is either α-mixing or ϕ-mixing of size $-a$. If $a_n/\sqrt{n} \uparrow \infty$ as $n \to \infty$, and*

$$\frac{\|X_t - \mu_t\|_p^{4-p-2/p}}{a_t} = O(t^\xi), \tag{3.4}$$

where (i) in the α-mixing case,

$$\xi < 1 - 2/p + 2(1 - 1/p)\min\{-1/2, \min\{bp/2, a/2\} - 1\},$$

and (ii) in the ϕ-mixing case,

$$\xi < 1 - 2/p + 2(1 - 1/p)\min\{-1/2, \min\{bp/2, a\} - 1\},$$

Then, $a_n^{-1} \sum_{t=1}^n (X_t - \mu_t) \to 0$, a.s.

The alternative, traditional, way of applying the truncation method is to combine it with the 'equivalent sequences' argument, applying the first Borel-Cantelli lemma to the probabilities that the remainders differ from 0. The original 'three series theorem' of Kolmogorov is a result in this mould, and see also Davidson (1994), Theorem 20.21, for an application of the method to NED processes. When we apply this approach, we obtain the following strong law.

Theorem 3.5 *Let a sequence $\{X_t\}_{-\infty}^{\infty}$ with means $\{\mu_t\}_{-\infty}^{\infty}$ be L_p-NED, $1 \leq p \leq 2$, of size $-b$, with respect to constants $d_t \ll \|(X_t - \mu_t)\|_p$, on a sequence $\{V_t\}_{-\infty}^{\infty}$ which is either α-mixing or ϕ-mixing of size $-a$. If $a_n/\sqrt{n} \uparrow \infty$ as $n \to \infty$, and*

$$\frac{\|(X_t - \mu_t)\|_p^{2-p/2}}{a_t} = O(t^\xi), \tag{3.5}$$

where (i) in the α-mixing case,

$$\xi < 1/2 - 1/p + \min\{-1/2, \min\{bp/2, a/2\} - 1\},$$

and (ii) in the ϕ-mixing case,

$$\xi < 1/2 - 1/p + \min\{-1/2, \min\{bp/2, a\} - 1\}.$$

Then, $a_n^{-1} \sum_{t=1}^{n}(X_t - \mu_t) \to 0$, a.s.

4 Comparison of the NED strong laws

We would clearly like to know if one of our five strong laws is dominated by the others, or indeed whether one dominates all the rest. More generally, we would like to know in which circumstances one or other may be useful. A systematic comparison of the conditions is not a trivial exercise, since it entails exploring a six-dimensional parameter space; if we write $a_t = t^\beta$ and $\|X_t - \mu_t\|_p = t^\gamma$, the dimensions are $(p, q, a, b, \beta, \gamma)$, with the added complication that a can represent either the rate of α-mixing or the rate of ϕ-mixing, with different conditions for each. To simplify the investigation, we considered three 'polar' dependence cases: an α-mixing process; a ϕ-mixing process; and a L_p-NED function of an independent process. In other words, we set either $a = \infty$, or $b = \infty$ while varying the other dependence parameter. We also looked at two bases for comparison. Firstly, we assumed a uniformly L_q-bounded process ($\gamma = 0$) and compared the conditions on the basis of the infimal value of β. As pointed out above, this establishes the rate of convergence of $\overline{X}_n$ as $O(n^{\beta-1})$. Secondly, we compare the supremal values of γ under which $\overline{X}_n$ converges. These cases still exhibit three dimensions of variation, the dependence measure plus p and q, although interest obviously focuses on the case $p = q$.

We will not attempt to report the results of our investigations in detail. This would require very extensive tabulations, which could easily be duplicated by the interested reader using a spreadsheet or other calculating device. It will suffice for us to report the two basic findings; first, that Theorem 3.4 dominates all the other theorems in the large majority of cases examined; but second, each of the others dominates it in certain situations, so that none of the theorems appears to be entirely dispensable.

To demonstrate the basic superiority of the truncation approach, compare Theorem 3.4 with Theorem 3.1. Putting $p = 2$ in condition (3.4), we obtain the condition

$$\frac{\|X_t - \mu_t\|_2}{a_t} = O(t^\xi), \tag{4.1}$$

where $\xi < \min\{-1/2, \min\{b, a/2\} - 1\}$ in the α-mixing case, and $\xi < \min\{-1/2, \min\{b, a\} - 1\}$ in the ϕ-mixing case. It is immediately clear how these conditions dominate (3.1). While the truncation argument allows us to apply a result for L_2-bounded random variables to the L_p-bounded case for $p < 2$, the potential improvement is more fundamental than this fact alone would suggest, because the truncated variable possesses all its moments.

However, in the uniformly L_q-bounded case where q strictly exceeds p, there are situations in which Theorem 3.4 is dominated by Theorem 3.3 in the sense of supplying a faster rate of convergence of $\overline{X}_n$. The latter theorem permits $p = 1$, and has more value than might be guessed from the mixingale strong law from which it derives (Corollary 2.3) because of the improved mapping from NED to mixingale dependence which is available for the L_1-NED case. In the event that we can establish the existence of higher-order moments, but cannot establish the L_p-NED property except for $p = 1$, Theorem 3.3 is worth having.

On the other hand, in the L_p-NED case with p in the region of $\sqrt{2}$ (which is where the exponent of $\|X_t - \mu_t\|_p$ in (3.4) is at its maximum), and with the L_p-NED size large enough, Theorem 3.2 dominates Theorem 3.4 in the sense that a higher rate of trend in the L_p-norm is compatible with convergence of $\overline{X}_n$.

The equivalent sequences argument proves generally to be the inferior method of exploiting the truncation approach, and Theorem 3.5 is almost always dominated by Theorem 3.4, even though it usually dominates Theorems 3.2 and 3.3 in turn. It is however available for cases with $p = 1$, and it dominates Theorem 3.4 (in the sense of faster convergence of $\overline{X}_n$) in certain cases with $p > 1$; although these are presumably of fairly minor importance, since they feature the condition $\|(X_t - \mu_t)\|_p = O(t^{-\delta})$ for $\delta > 0$.

Figures 3-8 show three-dimensional surfaces bounding permitted parameter combinations for the strong law, for α-mixing and ϕ-mixing processes (with $b = \infty$), and L_p-NED functions of an independent process (with $a = \infty$). Here, we restrict consideration to cases with $q = p$. Figures 3-5 plot β for the case $\gamma = 0$, and so indicate the rate of convergence of $\overline{X}_n$ with uniformly bounded moments. All of these plots represent the bounds set by (3.4), since Theorem 3.4 dominates the others in the cases represented.

On the other hand, Figures 6-8 show γ against p and either a or b assuming $\beta = 1$, so that these are bounding conditions for convergence of $\overline{X}_n$, with possibly trending moments. Figures 6 and 7 are based only on (3.4), whereas Figure 8 represents the maximum of the conditions specified by (3.4) and (3.2). since in this instance Theorem 3.2 sometimes dominates. As in Figure 2, note that these plots are complete in the sense that the surface has the same profile with respect to p for all size values exceeding those shown. The shapes of the α-mixing and ϕ-mixing plots are actually identical, except that the size parameter is increased by a factor of 2 in the latter case.

5 Conclusion

This paper reports and assesses a number of new strong laws for dependent heterogeneous processes. We believe that our results may not be too far from being definitive, at least with respect to the chosen characterization of dependence, by near-epoch dependence on a mixing process, including mixing processes. Our results are codified and ranked in terms of dependence, heterogeneity and moment parameters, in such as way as to simplify practical applications as far as possible.

We considerably narrow the gap between the available strong and weak (more precisely, L_p-convergence) laws for mixingales, and we view this as an important aspect of the work. Given the well-known fact that almost-sure convergence need not imply L_p-convergence, the comparatively severe restrictions sufficient to establish 'strong' laws to date have appeared as a slightly artificial feature of the theory. Hopefully our work may contribute to a better understanding of the relationship between the alternative modes of stochastic convergence.

6 Proofs

In the proofs, we use the notation $E_s X_t$ to stand for $E(X_t|\mathcal{F}_s)$, for any s and t.

The proofs of Theorems 2.2 and 2.3 are based on the well-known Convergence Lemma:

Lemma 6.1 *Let $\{Y_t\}_{t=1}^{\infty}$ be a stochastic sequence, let $S_n = \sum_{t=1}^{n} Y_t$, and let there exist a sequence* *of positive constants* $\{C_t\}_{t=1}^{\infty}$*, and* $p > 0$*, such that*

$$E \max_{m<k\leq n} |S_k - S_m|^p \ll \sum_{t=m+1}^{n} C_t^p$$

for each $n > m$ *and* $m \geq 1$*. If* $\sum_{t=1}^{\infty} C_t^p < \infty$*, then* $S_n \to S$ *a.s.*

This is a consequence of the Cauchy criterion for convergence and the Markov inequality, see for example Davidson (1994), Corollary 20.2.

Proof of Theorem 2.2 Defining $Y_{tj} = E_{t+j}X_t - E_{t+j-1}X_t$, note that the sequences $\{Y_{tj}, t = 1, ..., n\}$ are serially uncorrelated, and also, letting $\xi_{tj} = \|E_{t+j}X_t\|_2$ and $\zeta_{tj} = \|X_t - E_{t+j}X_t\|_2$, note that

$$E(Y_{tj}^2) = \xi_{tj}^2 - \xi_{t,j-1}^2 = \zeta_{t,j-1}^2 - \zeta_{tj}^2.$$

Write

$$\frac{1}{a_n}\sum_{t=1}^{n} X_t = T_{n1} + T_{n2} + T_{n3}$$

where

$$T_{n1} = \frac{1}{a_n}\sum_{t=1}^{n} E_{-2}X_t,$$

$$T_{n2} = \frac{1}{a_n}\sum_{t=1}^{n}\sum_{j=-t-1}^{t+1} Y_{tj}$$

and

$$T_{n3} = \frac{1}{a_n}\sum_{t=1}^{n}\left(X_t - E_{2t+1}X_t\right).$$

By the Kronecker lemma, $T_{n2} \to 0$ if T_{n2}' converges, where

$$T_{n2}' = \sum_{t=1}^{n} a_t^{-1} \sum_{j=-t-1}^{t+1} Y_{tj}. \tag{6.1}$$

Also, according to Lemma 6.1, T_{n2}' converges a.s. if there exists a square-summable sequence of constants, $\{C_t\}_{t=1}^{\infty}$, such that

$$E \max_{m<k\leq n} (T_{k2}' - T_{m2}')^2 \ll \sum_{t=m+1}^{n} C_t^2 \tag{6.2}$$

for $n > m \geq 0$. Inequality (6.2) is established much as in Masry and Györfi (1987), Theorem 3.1. After rearranging the order of summation in 6.1, we have

$$T_{k2}' - T_{m2}' = \sum_{j=-n-1}^{n+1} \sum_{t=\max\{m+1,|j|-1\}}^{n} a_t^{-1} Y_{tj}. \tag{6.3}$$

Let $m_0 = 1$, $m_n = \left(n\log(n+1)(\log\log(n+2))^{1+\delta}\right)^{-1}$, $\delta > 0$, for $n \geq 1$, and $m_{-n} = m_n$, such that the sequence $\{m_n\}_{n=-\infty}^{\infty}$ is summable. Applying a variant of McLeish's maximal inequality to $(T_{k2}' - T_{m2}')^2$ using the form (6.3), and then applying the properties of Y_{tj}, yields

$$E \max_{1\leq k\leq n} (T_{k2}' - T_{m2}')^2$$

$$\leq \left(\sum_{j=-n-1}^{n+1} m_j\right) \sum_{j=-n-1}^{n+1} m_j^{-1} E\left(\sum_{t=\max\{m+1,|j|-1\}}^{n} a_t^{-1} Y_{tj}\right)^2$$

$$\ll \sum_{j=-n-1}^{n+1} m_j^{-1} \sum_{t=\max\{m+1,|j|-1\}}^{n} a_t^{-2} E(Y_{tj}^2)$$

$$\ll \sum_{t=m+1}^{n} C_t^2$$

where, after re-ordering the summation according to (6.1) we have

$$C_t^2 = a_t^{-2}\left(m_0^{-1}(\xi_{t0}^2 - \xi_{t,-1}^2) + \sum_{j=-t-1}^{-1} m_j^{-1}(\xi_{tj}^2 - \xi_{t,j-1}^2)\right.$$

$$\left. + \sum_{j=1}^{t+1} m_j^{-1}(\zeta_{t,j-1}^2 - \zeta_{tj}^2)\right). \tag{6.4}$$

According to the L_2-mixingale property, $\xi_{tj} \leq c_t\psi_{|j|}$ for $j \leq 0$, and $\zeta_{tj} \leq c_t\psi_{j+1}$ for $j > 0$. Hence, applying Abel's partial summation formula to (6.4) gives

$$C_t^2 \leq c_t^2 a_t^{-2}\left(\psi_0^2 + \psi_1^2 + 2\sum_{j=1}^{t+1}(m_j^{-1} - m_{j-1}^{-1})\psi_j^2 - 2m_{t+1}^{-1}\psi_{t+2}^2\right).$$

Condition (2.1) is therefore sufficient for $\sum_{t=1}^{\infty} C_t^2 < \infty$, which completes the proof of $T_{n2} \to 0$ a.s.

The treatment of the terms T_{n1} and T_{n3} is identical, and we take the case T_{n1} as representative. By the Jensen inequality for averages,

$$T_{n1}^2 \leq a_n^{-2} n \sum_{t=1}^{n} (E_{-2}X_t)^2. \tag{6.5}$$

Hence, $T_{n1} \to 0$ a.s. if this also holds for the majorant of (6.5). Since $a_t^{-2}t \uparrow \infty$ by assumption, it follows by the Kronecker lemma that

$$\sum_{t=1}^{\infty} a_t^{-2} t (E_{-2}X_t)^2 < \infty, \text{ a.s.} \tag{6.6}$$

is also sufficient for the convergence, and (6.6) follows in turn from

$$E\left(\sum_{t=1}^{\infty} a_t^{-2} t (E_{-2}X_t)^2\right) < \infty. \tag{6.7}$$

But

$$E\left(\sum_{t=1}^{\infty} a_t^{-2} t (E_{-2}X_t)^2\right) = \sum_{t=1}^{\infty} a_t^{-2} t E(E_{-2}X_t)^2$$

$$\leq \sum_{t=1}^{\infty} a_t^{-2} \sum_{j=1}^{t} E(E_{-j-2}X_t)^2 \leq \sum_{t=1}^{\infty} c_t^2 a_t^{-2} \sum_{j=1}^{t} \psi_{j+2}^2 < \infty, \tag{6.8}$$

where the equality in (6.8) holds by monotone convergence, and the finiteness is by (2.1). ■

Proof of Corollary 2.1 On the assumptions,

$$\sum_{t=j}^{\infty} c_t^2 a_t^{-2} = O(j^{1+2\alpha}). \tag{6.9}$$

In order for (2.2) to hold, which is equivalent to (2.1) as noted, it is sufficient if $c_t^2/a_t^2 = O(t^{-1-\varepsilon})$ for $\varepsilon > 0$, and also,

$$\sum_{j=1}^{\infty} j^{-2\lambda}(\log(j+1))(\log\log(j+2))^{1+\delta} j^{1+2\alpha} < \infty. \tag{6.10}$$

The former condition holds if $2\alpha < -1$, and the latter if $-2\lambda+1+2\alpha < -1$, and these inequalities are jointly equivalent to the stated condition. ■

Proof of Theorem 2.3 Define $q(j) = [j^{p/(p-1)}]$, and $b_t = \inf\{j \in \{0,1,2,\ldots\} : q(j) \geq t\}$. Let

$$W_t = E_{t+b_t}X_t - E_{t-b_t}X_t. \tag{6.11}$$

Note that for all $j \geq 0$,

$$\begin{aligned}\|E_{t-j}W_t\|_p &\leq \|E_{t-j}(E_{t-b_t}X_t) - E_{t-j}(E_{t+b_t}X_t)\|_p 1_{\{j\leq b_t\}} \\ &\leq \|E_{t-b_t}(E_{t-j}X_t)\|_p 1_{\{j\leq b_t\}} + \|E_{t-j}X_t\|_p 1_{\{j\leq b_t\}} \\ &\leq 2c_t\psi_j 1_{\{j\leq b_t\}},\end{aligned} \tag{6.12}$$

and

$$\begin{aligned}&\|W_t - E_{t+j}W_t\|_p \\ &= \|E_{t-b_t}X_t - E_{t+b_t}X_t - E_{t+j}(E_{t-b_t}X_t) + E_{t+j}(E_{t+b_t}X_t)\|_p \\ &= \|E_{t+j}(E_{t+b_t}X_t) - E_{t+b_t}X_t\|_p \\ &= \|E_{t+b_t}(X_t - E_{t+j}X_t)\|_p \\ &\leq c_t\psi_j 1_{\{j\leq b_t\}}.\end{aligned} \tag{6.13}$$

Moreover, letting $Y_{tj} = E_{t+j}W_t - E_{t+j-1}W_t$ we may write, formally, $W_t = \sum_{j=-\infty}^{\infty} Y_{tj}$, where the sequences $\{Y_{tj}, \mathcal{F}_{t+j}\}_{j=-\infty}^{\infty}$ for $t = 1,2,...,n$ are martingale differences, $Y_{tj} = 0$ a.s. for $\mid j \mid > b_t$, and also

$$\| Y_{tj} \|_p \leq 2c_t\psi_{|j|}1_{\{|j|\leq b_t\}}. \tag{6.14}$$

Thus, consider the decomposition

$$\begin{aligned}&\frac{1}{a_n}\sum_{t=1}^{n} X_t \\ &= \frac{1}{a_n}\sum_{t=1}^{n} E_{t-b_t}X_t + \frac{1}{a_n}\sum_{t=1}^{n} W_t + \frac{1}{a_n}\sum_{t=1}^{n}(X_t - E_{t+b_t}X_t) \\ &= T_{n1} + T_{n2} + T_{n3}.\end{aligned} \tag{6.15}$$

Define $T'_{n2} = \sum_{t=1}^{n} a_t^{-1}W_t$, so that by the Kronecker lemma, $T_{n2} \to 0$ a.s. if T'_{n2} converges a.s. Let $\{m_j\}_{j=0}^{\infty}$ denote a summable sequence to be chosen, and let $m_{-j} = m_j$. Applying in turn the

Jensen, Doob, and Burkholder inequalities (cf. Davidson 1994, Lemma 16.8, and also Hansen 1991, 1992), and (6.14), we obtain for any $p > 1$ and $n > m \geq 0$,

$$\begin{aligned} & E \max_{m<k\leq n} |T'_{k2} - T'_{m2}|^p \\ \leq & \left(\frac{p}{1-p}\right)^p \left(\sum_{j=-\infty}^{\infty} m_j\right)^{p-1} \sum_{j=-\infty}^{\infty} m_j^{1-p} E \left|\sum_{t=m+1}^{n} Y_{tj}/a_t\right|^p \\ & \ll \sum_{j=0}^{\infty} m_j^{1-p} \psi_j^p \sum_{t=m+1}^{n} c_t^p a_t^{-p} 1_{\{j\leq b_t\}} \\ & = \sum_{t=m+1}^{n} C_t^p \end{aligned} \tag{6.16}$$

where

$$C_t = \left(c_t^p a_t^{-p} \sum_{j=0}^{b_t} m_j^{1-p} \psi_j^p\right)^{1/p}. \tag{6.17}$$

Now, choose $m_j = \psi_j \left(\sum_{t=\max\{1,q(j)\}}^{\infty} c_t^p a_t^{-p}\right)^{1/p}$, which is summable under condition (2.3). We also find in this case that

$$\sum_{t=1}^{\infty} C_t^p = \sum_{j=0}^{\infty} m_j^{1-p} \psi_j^p \sum_{t=1}^{\infty} c_t^p a_t^{-p} 1_{\{j\leq b_t\}}$$

$$= \sum_{j=0}^{\infty} m_j^{1-p} \psi_j^p \sum_{t=\max\{1,q(j)\}}^{\infty} c_t^p a_t^{-p} = \sum_{j=0}^{\infty} m_j < \infty.$$

Applying Lemma 6.1 now completes the proof of $T_{n2} \to 0$ a.s.

The treatment of the terms T_{n1} and T_{n3} is identical, and we take the case T_{n1} as representative. $T_{n1} \to 0$ a.s. if and only if

$$\lim_{m\to\infty} P\left(\sup_{n\geq m} |T_{n1}| > \varepsilon\right) = 0$$

for every $\varepsilon > 0$. For $n \geq 1$, and any positive integer k, define

$$T'_{n1}(k) = \frac{1}{a_n} \sum_{t=1}^{\min\{k,n\}} E_{t-b_t} X_t,$$

and also $T''_{n1}(k) = T_{n1} - T'_{n1}(k)$, so that $T''_{n1}(k) = 0$ for $k \geq n$. Then,

$$\begin{aligned} & \lim_{m\to\infty} P\left(\sup_{n\geq m} |T_{n1}| > \varepsilon\right) \\ \leq & \lim_{m\to\infty} P\left(\sup_{n\geq m} \left|T'_{n1}(k)\right| > \frac{\varepsilon}{2}\right) + \lim_{m\to\infty} P\left(\sup_{n\geq m} \left|T''_{n1}(k)\right| > \frac{\varepsilon}{2}\right). \end{aligned} \tag{6.18}$$

Since $a_n \to \infty$ as $n \to \infty$, the first term of the majorant of (6.18) is equal to 0 for each fixed k, and any $\varepsilon > 0$. It therefore suffices to show that the second term can be made arbitrarily close to 0 by taking k large enough. The Markov inequality yields

$$P\left(\sup_{n>m} \left|T''_{n1}\right| > \frac{\varepsilon}{2}\right) \leq \frac{2}{\varepsilon} E \sup_{n>m} \left|T''_{n1}\right|, \tag{6.19}$$

and we exhibit a sequence of finite constants δ_k, such that $E \sup_{n>m} |T''_{n1}| \leq \delta_k$, and $\delta_k \to 0$ as $k \to \infty$. Note that for $p > 1$, and any $n > k$, we have

$$
\begin{aligned}
\left|T''_{n1}(k)\right| &\leq a_n^{-1} \left| \sum_{t=k+1}^{n} E_{t-b_t} X_t \right| \\
&= a_n^{-1} \left| \sum_{t=k+1}^{n} \left(\sum_{j=1}^{b_t} b_t^{-1} \right) E_{t-b_t} X_t \right| \\
&\leq \sum_{j=1}^{b_n} a_n^{-1} \left| \sum_{t=k+1}^{n} 1_{\{j \leq b_t\}} b_t^{-1} E_{t-b_t} X_t \right| \\
&\leq \sum_{j=1}^{b_n} \left(\frac{(n-k)^{p-1}}{a_n^p} \sum_{t=k+1}^{n} 1_{\{j \leq b_t\}} b_t^{-p} \left| E_{t-b_t} X_t \right|^p \right)^{1/p} \\
&\leq \sum_{j=1}^{b_n} \left(a_n^{-p} n^{p-1} \sum_{t=k+1}^{n} 1_{\{j \leq b_t\}} b_t^{-p} \left| E_{t-b_t} X_t \right|^p \right)^{1/p} \\
&\leq \sum_{j=1}^{b_n} \left(\sum_{t=k+1}^{n} 1_{\{j \leq b_t\}} t^{p-1} b_t^{-p} a_t^{-p} \left| E_{t-b_t} X_t \right|^p \right)^{1/p}
\end{aligned}
\tag{6.20}
$$

where the third inequality uses Jensen's inequality for averages, and the last one, the assumption that the sequence $a_n^{-p} n^{p-1}$ is non-increasing in n. Taking expectations, we therefore obtain for $k < n$,

$$
\begin{aligned}
E \left|T''_{n1}(k)\right| &\leq \sum_{j=1}^{b_n} \left(\sum_{t=k+1}^{n} 1_{\{j \leq b_t\}} t^{p-1} b_t^{-p} a_t^{-p} E \left| E_{t-b_t} X_t \right|^p \right)^{1/p} \\
&\leq \sum_{j=1}^{\infty} \psi_j \left(\sum_{t=\max\{q(j), k+1\}}^{\infty} a_t^{-p} c_t^p \right)^{1/p} = \delta_k,
\end{aligned}
\tag{6.21}
$$

where the first inequality uses Jensen's inequality applied to the concave function $(.)^{1/p}$, and the second uses the facts that $b_t^p \geq t^{p-1}$, and $E \left| E_{t-j} X_t \right|^p \geq E \left| E_{t-b_t} X_t \right|^p$ for $t \geq q(j)$. The inequality $E \left|T''_{n1}(k)\right| \leq \delta_k$ also holds trivially for $k \geq n$. Note that $\delta_k < \infty$, and $\lim_{k \to \infty} \delta_k = 0$, according to (2.3). Moreover, δ_k does not depend on n and therefore bounds $E \sup_{n>m} \left|T''_{n1}(k)\right|$ for any m, as required. This completes the proof. ■

Proof of Corollary 2.2 $(c_t/a_t)^p$ is summable if $\alpha < -1/p$. In this case, the inner sum in (2.3) is of $O(j^{p(1+\alpha p))/(p-1)})$, and (2.3) holds if $\lambda - (1 + \alpha p)/(p-1) > 1$. This is equivalent to the second of the stated conditions. ■

Proof of Corollary 2.3 This is just a restatement of Corollary 20.18 of Davidson (1994). ■

Proof of Theorems 3.1 and 3.2 These results follow directly from Corollaries 2.1 and 2.2 respectively, on noting that under the conditions stated, the sequence $\{X_t - \mu_t, \mathcal{F}_t\}$, where $\mathcal{F}_t = \sigma(V_{t-j}, j \geq 0)$, is (with $p = 2$ and $1 < p \leq 2$, respectively) a L_p-mixingale of size $-\min\{b, a(1/p - 1/q)\}$ in the α-mixing case, or of size $-\min\{b, a(1 - 1/q)\}$ in the ϕ-mixing case; in every case this is with respect to constants $c_t \ll \| X_t - \mu_t \|_q$. See Theorem 17.5 of Davidson (1994), for example, for a proof of this assertion. ■

Proof of Theorem 3.3 This is just a restatement of Theorem 20.19[3] of Davidson (1994). ■

Proof of Theorem 3.4 Assume $\mu_t = 0$ without loss of generality. Consider the continuous transformation

$$h_B(x) = 1_{\{|x|\leq B\}}x + 1_{\{x>B\}}B - 1_{\{x<-B\}}B \tag{6.22}$$

for some $B > 0$, such that $\mid h_B(x) \mid\leq B$, and $|h_B(x) - h_B(y)| \leq |x-y|$ for all x and y. For a sequence of positive constants B_t to be defined, we have the decomposition

$$\frac{1}{a_n}\sum_{t=1}^{n} X_t$$

$$= \frac{1}{a_n}\sum_{t=1}^{n} h_{B_t}(X_t) - Eh_{B_t}(X_t) + \frac{1}{a_n}\sum_{t=1}^{n} X_t - h_{B_t}(X_t) - E(X_t - h_{B_t}(X_t))$$

$$= T_{n1} + T_{n2}. \tag{6.23}$$

We show that for suitable choice of B_t, each of these terms converges to zero almost surely under condition (3.4).

Under the assumptions of the theorem, it follows by Theorem 17.13 and Example 17.14 of Davidson (1994) that $h_{B_t}(X_t)$ is L_2-NED with respect to the constant sequence $\{B_t^{1-p/2}\left\|X_t\right\|_p\}_{-\infty}^{\infty}$ on $\{V_t\}_{-\infty}^{\infty}$, of size $-bp/2$. Noting that $h_{B_t}(X_t-\mu_t)$ possesses moments of all orders, it follows by Theorem 17.5 of Davidson (1994) that the sequence $\{h_{B_t}(X_t) - E(h_{B_t}(X_t)), \mathcal{F}_t\}_{-\infty}^{\infty}$ is a L_2-mixingale with respect to constants $\{B_t^{1-p/2}\left\|X_t\right\|_p\}_{-\infty}^{\infty}$, of size $-\lambda = -\min\{bp/2, a/2\}$ in the α-mixing case, and size $-\lambda = -\min\{bp/2, a\}$ in the ϕ-mixing case. Applying Corollary 2.1, it further follows that

$$T_{n1} \to 0, \text{ a.s.} \tag{6.24}$$

if

$$a_t^{-1}B_t^{1-p/2}\left\|X_t\right\|_p = O(t^{\min\{-1/2,\lambda-1\}-\varepsilon}) \tag{6.25}$$

for $\varepsilon > 0$.

Next consider term T_{n2}. Invoking monotone convergence and the Kronecker lemma, note that $a_n^{-1}\sum_{t=1}^{n} Y_t \to 0$ a.s. if

$$\sum_{t=1}^{\infty} E|Y_t\mid/a_t < \infty. \tag{6.26}$$

In particular we may consider the case $Y_t = X_t - h_{B_t}(X_t)$, since it is also clear that $a_n^{-1}\sum_{t=1}^{n} E(Y_t) \to 0$ under condition (6.26). Note that

$$E\left|X_t - h_{B_t}(X_t)\right|/a_t \leq B_t^{1-p}\left\|X_t\right\|_p^p/a_t, \tag{6.27}$$

so that for $T_{n2} \to 0$ a.s., the majorant side of (6.27) must be summable. Accordingly, choose $B_t = (t/a_t)^{1/(p-1)}\left\|X_t\right\|_p^{p/(p-1)}(\log(t+1))^{(1+\delta)/(p-1)}$ for $\delta > 0$. Substituting this formula into expression (6.25), and simplifying, yields condition (3.4). ■

Proof of Theorem 3.5 Again, assume $\mu_t = 0$. The strategy is to show there is a sequence equivalent to X_t which satisfies the conditions of Corollary 2.1. Applying the continuous transformation in (6.22), let the bounding sequence be defined as $B_t = \left\|X_t\right\|_p t^{1/p}(\log(t+1))^{(1+\delta)/p}$,

[3] In the first printing of Davidson (1994), Theorem 20.19 is stated incorrectly. The conditions on q should be $q \geq p$ and $q > 1$ in both the α-mixing and ϕ-mixing cases. This error is corrected in subsequent printings.

for $\delta > 0$. Note that

$$\sum_{t=1}^{\infty} E\left|X_t\right|^p / B_t^p < \infty, \qquad p \geq 1, \tag{6.28}$$

and hence, by Theorems 20.4 and 20.6 of Davidson (1994),

$$\sum_{t=1}^{\infty} P(|X_t| > B_t) < \infty, \tag{6.29}$$

and

$$\sum_{t=1}^{\infty} |E(h_{B_t}(X_t)/B_t)| < \infty. \tag{6.30}$$

Inequality (6.29) implies that the sequences X_t and $h_{B_t}(X_t)$ are equivalent, and hence, applying the first Borel-Cantelli lemma, it suffices to prove the strong law for the truncated sequence. Exactly the same argument as in the proof of Theorem 3.4 leads us to the conclusion that $a_n^{-1} \sum_{t=1}^n h_{B_t}(X_t) - E(h_{B_t}(X_t)) \to 0$ a.s. if

$$a_t^{-1} B_t^{1-p/2} \left\|X_t\right\|_p = O(t^{\min\{-1/2, \lambda-1\}-\varepsilon}) \tag{6.31}$$

for $\varepsilon > 0$, and given the definition of B_t, this condition is equivalent to (3.5). Also, note that (3.5) implies

$$\frac{B_t}{a_t} = \frac{\left\|X_t\right\|_p t^{1/p} (\log(t+1))^{(1+\delta)/p}}{a_t} \to 0, \tag{6.32}$$

and hence

$$a_n^{-1} \sum_{t=1}^{n} h_{B_t}(X_t) \to 0 \quad a.s. \tag{6.33}$$

in view of (6.30) and the Kronecker lemma. ■

References

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

[2] Davidson, J. (1993), An L_1-convergence theorem for heterogeneous mixingale arrays with trending moments, *Statistics and Probability Letters* 16, 301-4.

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

[4] De Jong, R. M. (1994), A strong law for L_2-mixingale sequences, Discussion Paper, University of Tilburg.

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

[6] De Jong, R.M. (1995b), A strong law of large numbers for triangular mixingale arrays , forthcoming in *Statistics and Probability Letters.*

[7] Gallant, A. R. and H. White (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* New York: Basil Blackwell.

[8] Hall, P. and C. C. Heyde (1980), *Martingale limit theory and its applications.* New York: Academic Press.

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

[10] Hansen, B. E. (1992), Strong laws for dependent heterogeneous processes. Errata. *Econometric Theory* 8, 421–422.

[11] Hosking, J. R. M. (1981), Fractional Differencing, *Biometrika* 68,1, 165-76

[12] Masry, E. and L. Györfi (1987), Strong Consistency and Rates for Recursive Probability Density Estimators of Stationary Processes, *Journal of Multivariate Analysis* 22, 79-93

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

[14] Pötscher, B. M. and I. R. Prucha (1991), Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part 1: consistency and approximation concepts. *Econometric Reviews* 10, 125–216.

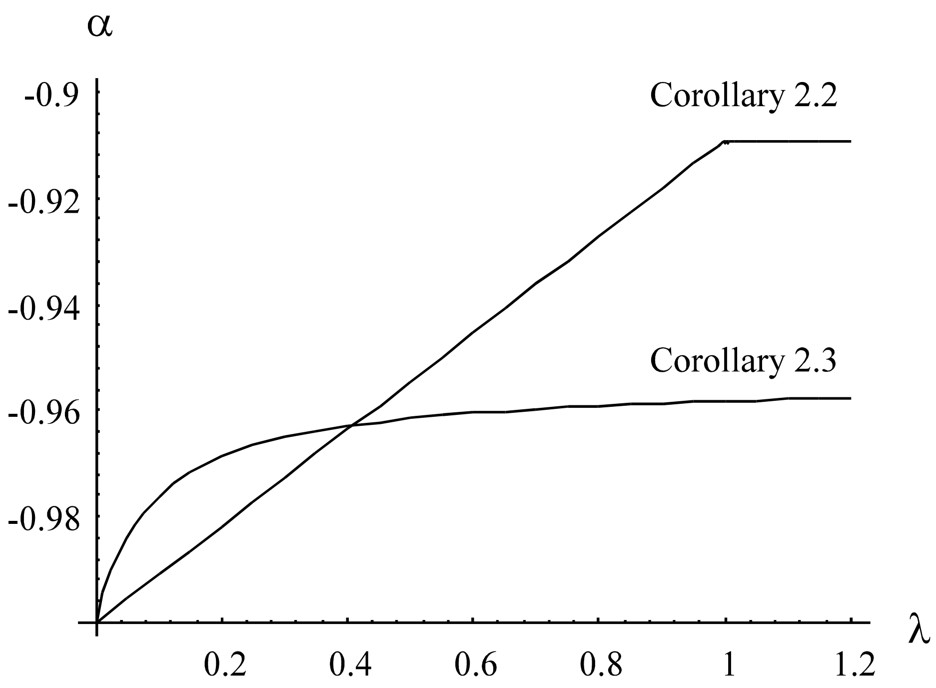

Figure 1

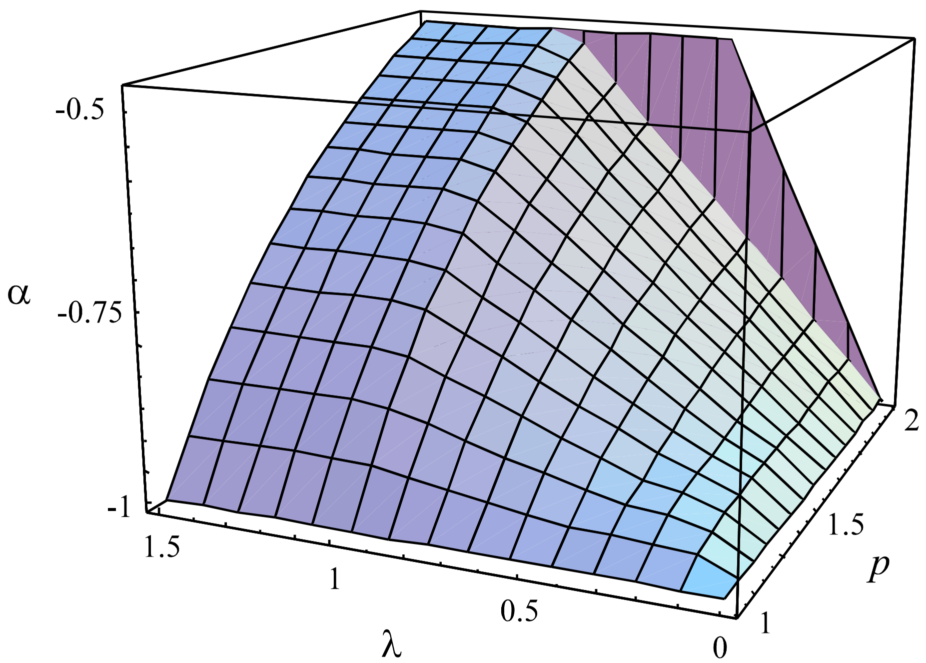

Figure 2

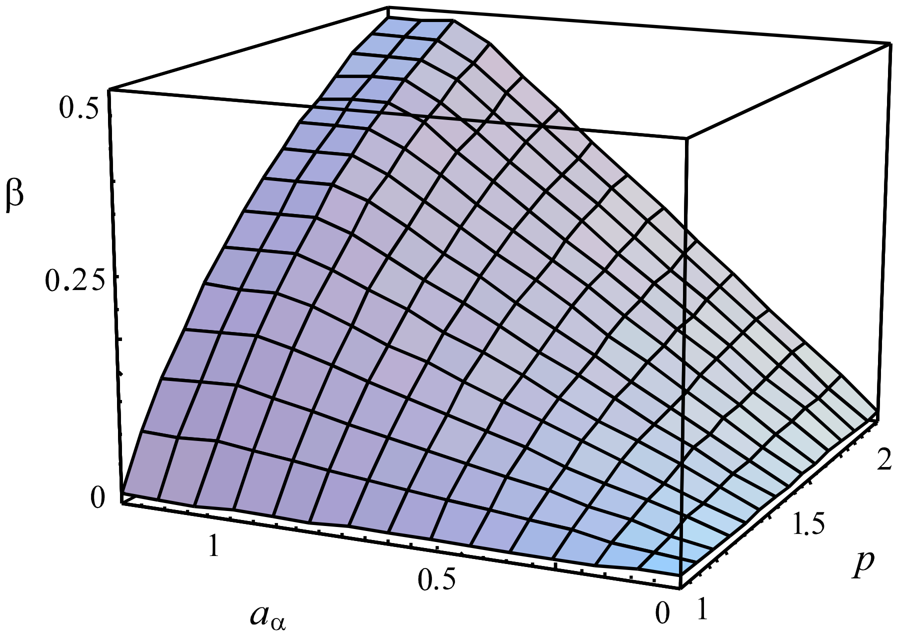

Figure 3

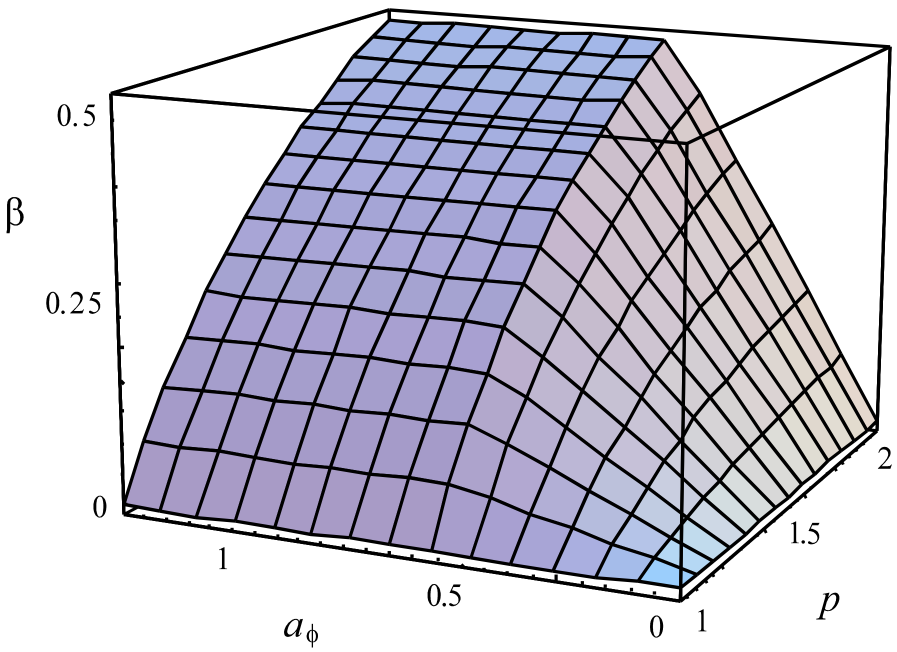

Figure 4

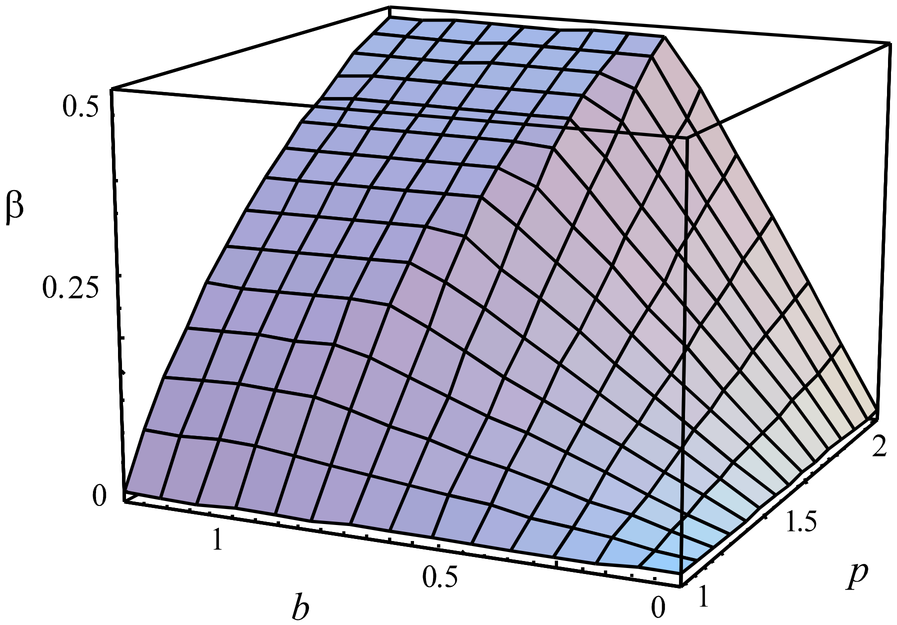

Figure 5

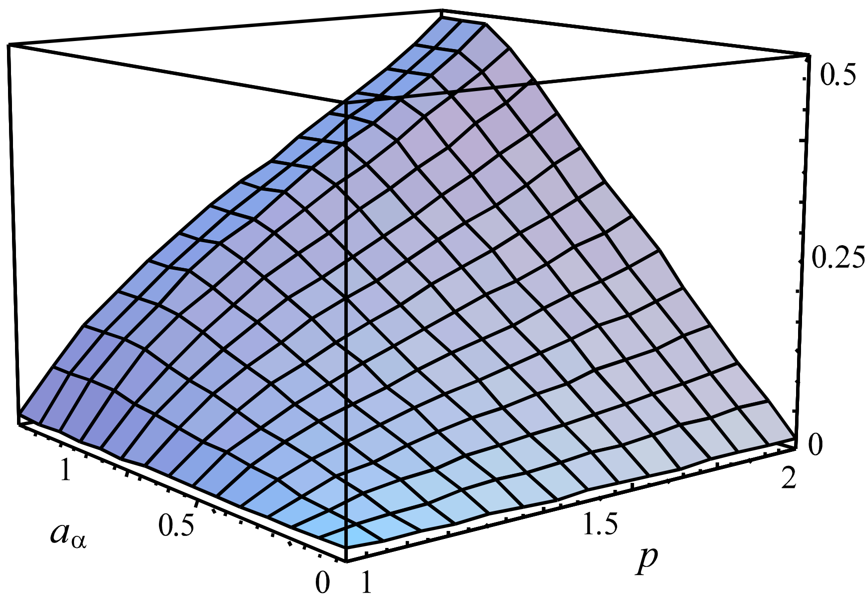

Figure 6

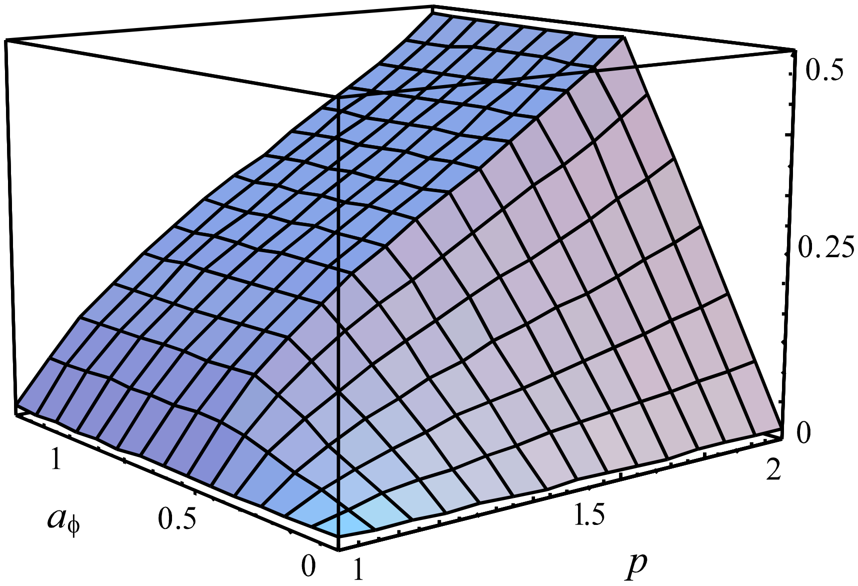

Figure 7

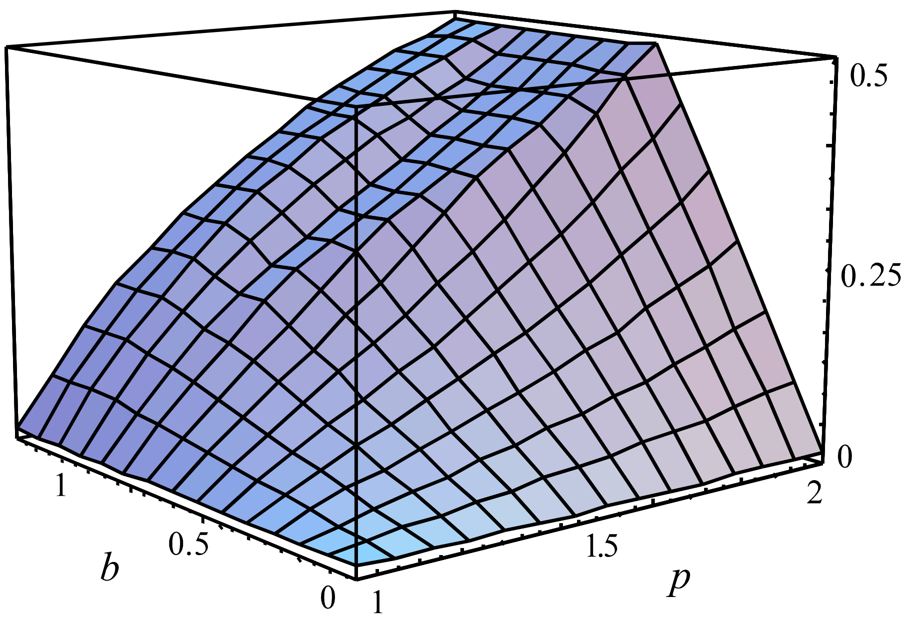

Figure 8