

ET

62

05348

1988

SERIE RESEARCH MEMORANDA

CONDITIONING AND DEPENDENCE

Herman J. Bierens

Research Memorandum 1988-62



**VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM**

CONDITIONING AND DEPENDENCE *

Herman J. Bierens
Department of Econometrics
Free University, Amsterdam

December 1988

* Chapter 9 of:

PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS

To be published by Marcel Dekker Inc., New York.



TABLE OF CONTENTS:

PREFACE

PART 1: PRELIMINARY MATHEMATICS

1. BASIC PROBABILITY THEORY
 - 1.1 Measure-theoretical foundation of probability theory
 - 1.2 Independence
 - 1.3 Borel measurable functions
 - 1.4 Mathematical expectation
 - 1.5 Characteristic functions
 - 1.6 Random functions
2. CONVERGENCE
 - 2.1 Weak and strong convergence of random variables
 - 2.2 Convergence of mathematical expectations
 - 2.3 Convergence of distributions
 - 2.4 Central limit theorems
 - 2.5 Further results on convergence of distributions and mathematical expectations, and laws of large numbers
 - 2.6 Convergence of random functions
 - 2.7 Uniform strong and weak laws of large numbers
3. INTRODUCTION TO CONDITIONING
 - 3.1 Definition of conditional expectation
 - 3.2. Basic properties of conditional expectations
 - 3.3 Identification of conditional expectations

PART 2: REGRESSION ANALYSIS UNDER INDEPENDENCE

4. NONLINEAR PARAMETRIC REGRESSION ANALYSIS
 - 4.1 Nonlinear regression models and the nonlinear least squares estimator
 - 4.2 Consistency and asymptotic normality: General theory
 - 4.2.1 Consistency
 - 4.2.2 Asymptotic normality
 - 4.3 Consistency and asymptotic normality of nonlinear least squares estimators in the i.i.d. case
 - 4.3.1 Consistency

- 4.3.2 Asymptotic normality
 - 4.3.3 Consistent estimation of the asymptotic variance matrix
- 4.4 Consistency and asymptotic normality of the nonlinear least squares estimator under data heterogeneity
 - 4.4.1 Data heterogeneity
 - 4.4.2 Strong and weak consistency
 - 4.4.3 Asymptotic normality
- 4.5 Testing parameter restrictions: The Wald test
5. TESTS FOR MODEL MISSPECIFICATION
 - 5.1 White's version of Hausman's test
 - 5.2 Newey's M-test
 - 5.2.1 Introduction
 - 5.2.2 The conditional M-test
 - 5.3 A consistent conditional M-test
 - 5.4 The integrated M-test
6. THE NADARAYA-WATSON KERNEL REGRESSION FUNCTION ESTIMATOR
 - 6.1 Introduction
 - 6.2 Asymptotic normality in the continuous case
 - 6.3 Uniform consistency in the continuous case
 - 6.4 Discrete and mixed continuous-discrete regressors
 - 6.4.1 The discrete case
 - 6.4.2 The mixed continuous-discrete case
 - 6.5 The choice of the kernel
 - 6.6 The choice of the window width
 - 6.7 An empirical application to specification of household expenditure systems and equivalence scales
 - 6.7.1 Introduction
 - 6.7.2 Model and data
 - 6.7.3 The results
 - 6.7.4 Sample selection
7. SAMPLE MOMENTS INTEGRATING NORMAL KERNEL ESTIMATORS
 - 7.1 Kernel density estimators
 - 7.1.1 Integral conditions
 - 7.1.2 Sample moments integrating kernel density estimators
 - 7.1.3 The Dirac-catastrophe and the modified SMINK density estimator

- 7.1.4 How to choose the window width of the modified SMINK density estimator
- 7.2 Regression
 - 7.2.1 SMINK estimators of a regression function
 - 7.2.2 How to choose the window width parameters
- 7.3 A numerical example
- 7.4 Proofs
- 8. NONLINEAR REGRESSION WITH DISCRETE EXPLANATORY VARIABLES
 - 8.1. The earnings function
 - 8.2. The functional form of a regression model with discrete explanatory variables
 - 8.3. The choice of the linear separator
 - 8.4. Estimating and testing the regression function
 - 8.4.1 Estimation
 - 8.4.2 Model specification testing
 - 8.4.3 The selection of the polynomial order
 - 8.5 Proofs
 - 8.5.1 Proof of theorem 8.4.2
 - 8.5.2 Proof of theorem 8.4.3

PART 3: TIME SERIES

- 9. CONDITIONING AND DEPENDENCE
 - 9.1 Conditional expectations relative to a Borel field
 - 9.1.1 Definition and basic properties
 - 9.1.2 Martingales
 - 9.1.3 Martingale convergence theorems
 - 9.1.4 A martingale difference central limit theorem
 - 9.2 Measures of dependence
 - 9.2.1 Mixingales
 - 9.2.2 Uniform and strong mixing
 - 9.2.3 ν -Stability
 - 9.3 Weak laws of large numbers for dependent random variables
 - 9.4 Proper heterogeneity and uniform laws for functions of infinitely many random variables
- 10. FUNCTIONAL SPECIFICATION OF TIME SERIES MODELS
 - 10.1 Linear times series regression models
 - 10.1.1 Introduction

- 10.1.2 The Wold decomposition
- 10.1.3 Linear vector time series models
- 10.2 ARMA memory index models
 - 10.2.1 Introduction
 - 10.2.2 Finite conditioning of univariate rational-valued time series
 - 10.2.3 Infinite conditioning of univariate rational-valued time series
 - 10.2.4 The multivariate case
 - 10.2.5 The nature of the ARMA memory index parameters and the response functions
 - 10.2.5 Discussion
- 10.3 Nonlinear ARMAX models
- 11. ARMAX MODELS: ESTIMATION AND TESTING
 - 11.1 Estimation of linear ARMAX models
 - 11.1.1 Introduction
 - 11.1.2 Consistency and asymptotic normality
 - 11.2 Estimation of nonlinear ARMAX models
 - 11.3 A consistent $N(0,1)$ model specification test
 - 11.4 A consistent Hausman-type model specification test
 - 11.5 An autocorrelation test
- 12. NONPARAMETRIC TIME SERIES REGRESSION
 - 12.1 Assumptions and preliminary lemmas
 - 12.2 Consistency and asymptotic normality of time series kernel regression function estimators

PREFACE

This book deals with statistical inference of nonlinear regression models from two opposite points of view, namely the case where the functional form of the model is completely specified as a known function of regressors and unknown parameters, and the opposite case where the functional form of the model is completely unknown. First it is assumed that the response function of the regression model under review belongs to a certain well-specified parametric family of functional forms, by which estimation of the model merely amounts to estimation of the unknown parameters. For this class of models we review the asymptotic properties of the nonlinear least squares estimator for independent data as well as for time series.

In practice assumptions on the functional form are often made on the basis of computational convenience rather than on the basis of precise a priori knowledge of the empirical phenomenon under review. Therefore the linear regression model is still the most popular model specification in applied research. However, even if the specification of the functional form is based on sound theoretical considerations there is quite often a large range of functional forms that are theoretically admissible, so that there is no guarantee that the actually chosen functional form is true. Functional specification of a parametric nonlinear regression model should therefore always be verified by conducting model misspecification tests. Various model misspecification tests will therefore be discussed, in particular consistent tests which have asymptotic power 1 against all deviations from the null hypothesis that the model is correct.

The opposite case of parametric regression is nonparametric regression. Nonparametric regression analysis is concerned with estimation of a regression model without specifying in advance its functional form. Thus the only source of information about the functional form of the model is the data set itself. In this book we shall review various nonparametric regression approaches, with special emphasis on the kernel method, under various distributional assumptions.

This book is divided into three parts. In the first part we review the elements of abstract probability theory we need in part 2. Part 2 is devoted to the asymptotic theory of para-

metric and nonparametric regression analysis in the case of independent data generating processes. In part 3 we extend the analysis involved to time series.

The selection of the topics mainly reflexes my own interest in the subject. Instead of providing an encyclopedic survey of the literature, I have chosen for a setup which aims to fill the gap between intermediate statistics (including linear time series analysis) and the level necessary to get access to the recent literature on nonlinear and nonparametric regression analysis, with emphasis on my own contributions. The ultimate goal is to provide the student with the tools for his own independent research in this area, by showing what tools I and others have used and what they have been used for. Thus, this book may be viewed as an account of my own struggle with the material involved. I think this book is particularly suitable for self-tuition (at least it aims to be), and may prove useful in a graduate course in mathematical statistics and advanced econometrics.

Acknowledgements:

The first five chapters of this book have been disseminated in draft form as working papers. I am grateful to Anil Bera, Alexander Georgiev and Jan Magnus for suggesting additional references, and in particular to Lourens Broersma, Johan Smits and Ton Steerneman who suggested various improvements.

A large body of the material in chapter 6 has been published earlier in Truman F. Bewley (ed.), *Advances in Econometrics, Fifth World Congress*, Cambridge University Press. I am indebted to Cambridge University Press for granting permission to reprint it.

9. CONDITIONING AND DEPENDENCE

Time series models usually aim to represent, directly or indirectly, the conditional expectation of a time series variable relative to the entire past of the time series process involved. The reason for this is that this conditioned expectation is the best forecasting scheme; best in the sense that it yields forecasts with minimal mean square forecast error. The concept of a conditional expectation relative to a one-sided infinite sequence of 'past' variables, however, cannot be made clear on the basis of the elementary notion of conditional expectation known from intermediate mathematical-statistical textbooks, and even the more general approach in chapter 3 is not suitable. What we need here is the concept of a conditional expectation relative to a Borel field. We shall discuss this concept and its consequences (in particular martingale theory) in section 9.1. In section 9.2 we consider various measures of dependence as some, though rather weak, conditions have to be imposed on the dependence of a time series process to prove weak (uniform) laws of large numbers. These weak laws are the topics of sections 9.3 and 9.4.

Throughout we assume that the reader is familiar with the basic elements of linear time series analysis, say on the level of Harvey's (1981) textbook.

9.1 Conditional expectations relative to a Borel field

9.1.1 Definition and basic properties

In section 3.1 we have defined the conditional expectation of a random variable Y relative to a random vector $X \in \mathbb{R}^k$ as a Borel measurable real function g on $\mathbb{R}^k$ such that

$$E[Y - g(X)]\psi(X) = 0$$

for all bounded Borel measurable real functions ψ on $\mathbb{R}^k$. Approximating ψ by simple functions (cf. definition 1.3.2 and theorem 1.3.5) it follows that $g(X) = E(Y|X)$ a.s. if and only if

$$E(Y - g(X))I(X \in B) = 0 \quad (9.1.1)$$

for all Borel sets B in $\mathbb{R}^k$, where $I(\cdot)$ is the indicator function. Now let $(\Omega, \mathcal{F}, P)$ be the probability space on which (Y, X)

is defined and let F_X be the Borel field generated by X , i.e. F_X is the collection of all sets of the type

$$\{\omega \in \Omega: x(\omega) \in B\}, B \in \mathcal{B}^k.$$

Cf. theorem 1.1.1. Then condition (9.1.1) is equivalent to

$$\int_A [y(\omega) - g(x(\omega))] P(d\omega) = 0$$

for all $A \in F_X$. This justifies the notation

$$E(Y|X) = E(Y|F_X).$$

Note that F_X is a Borel field contained in F :

$$F_X \subset F.$$

Actually, along this line we can define conditional expectations relative to an arbitrary sub-Borel field G of F without reference to the random variable or vector which generates G :

Definition 9.1.1. Let Y be a random variable defined on the probability space (Ω, F, P) satisfying $E|Y| < \infty$, and let G be a Borel field contained in F . Then the conditional expectation of Y relative to G is a random variable Z defined on (Ω, G, P) such that for every $A \in G$,

$$\int_A [y(\omega) - z(\omega)] P(d\omega) = 0$$

The existence of $Z = E(Y|G)$ is guaranteed by the condition $E|Y| < \infty$, i.e. if EY exists, so does $E(Y|G)$. Cf. Chung (1974, theorem 9.1.1). Moreover, the uniqueness of Z follows from the Radon-Nikodym theorem [cf. Royden (1968), p.238], i.e. if both Z_1 and Z_2 satisfy the conditions in definition 9.1.1 then $P(Z_1 = Z_2) = 1$. We then say that Z_1 and Z_2 belong to the same equivalence class. Thus, $E(Y|G)$ is almost surely unique.

Next, we define the Borel field generated by a one-sided infinite sequence (X_t) , $t=1,2,\dots$, of random vectors in R^k . First, let F_n be the Borel field generated by $X_1, \dots, X_n$, i.e. F_n is the collection of sets

$$(\omega \in \Omega: x_1(\omega) \in B_1, \dots, x_n(\omega) \in B_n),$$

where $B_1, \dots, B_n$ are arbitrary Borel sets in R^k . Now (F_n) is an increasing sequence of Borel fields:

$$F_n \subset F_{n+1} \quad (9.1.2)$$

If $\bigcup_{n \geq 1} F_n$ is a Borel field itself, it is just the Borel field generated by (X_t) , $t=1,2,\dots$. However, it is in general not a Borel field, for

$$\Lambda_t \in \bigcup_{n \geq 1} F_n, \quad t=1,2,\dots$$

does not imply

$$\bigcup_{t \geq 1} \Lambda_t \in \bigcup_{n \geq 1} F_n.$$

Therefore we define the Borel field F_∞ generated by (X_t) , $t=1,2,\dots$, as the minimum Borel field containing $\bigcup_{n \geq 1} F_n$.

Definition 9.1.2. Let (X_t) , $t=1,2,\dots$, be a one-sided infinite sequence of random vectors defined on a common probability space (Ω, F, P) , and let F_n be the Borel field generated by $X_1, \dots, X_n$. Then the Borel field generated by (X_t) , $t=1,2,\dots$ is the minimum Borel field containing $\bigcup_{n \geq 1} F_n$, which is denoted by $F_\infty = \bigvee_{n \geq 1} F_n$.

Consequently, for every Y defined on (Ω, F, P) satisfying $E|Y| < \infty$ we have

$$E[Y|X_1, X_2, X_3, \dots] = E[Y|F_\infty].$$

The properties of conditional expectations relative to a Borel field are quite similar to those in section 3.2. In particular, theorem 3.2.1 now reads as:

Theorem 9.1.1. Let Y and Z be random variables defined on a probability space (Ω, F, P) satisfying $E|Y| < \infty$, $E|Z| < \infty$. Let G and H be Borel fields satisfying $G \subset H \subset F$. We have:

$$(I) \quad E[E(Y|H)|G] = E(Y|G) = E[E(Y|G)|H];$$

(II) (a) $E[E(Y|G)] = E Y$; (b) $E(Y|F) = Y$ a.s.;

(III) $U = Y - E(Y|G) \Rightarrow E(U|G) = 0$ a.s.;

(IV) $E(Y + Z|G) = E(Y|G) + E(Z|G)$ a.s.;

(V) $Y \leq Z \Rightarrow E(Y|G) \leq E(Z|G)$ a.s.;

(VI) $|E(Y|G)| \leq E(|Y| | G)$ a.s.;

(VII) If X is defined on (Ω, G, P) and $E|X| < \infty$ then $E(X \cdot Y|G) = X \cdot E(Y|G)$;

(VIII) If every random variable defined on (Ω, G, P) is independent of Y , i.e.,

$$\Lambda_1 \in F_Y, \Lambda_2 \in G \Rightarrow P(\Lambda_1 \cap \Lambda_2) = P(\Lambda_1) \cdot P(\Lambda_2),$$

where F_Y is the Borel Field generated by Y , then $E(Y|G) = E Y$.

Proof: Exercise 2.

Note that part (VIII) of theorem 3.2.1 does not apply anymore. The Borel field generated by $\Gamma(X)$ is contained in the Borel field generated by X , so that the first result of theorem 3.2.1 (VIII) is just part (I) of theorem 9.1.1. Moreover, if Γ is a one-to-one mapping the aforementioned Borel fields are the same, and so are the corresponding conditional expectations.

9.1.2 Martingales

A fundamental concept in the theory of stochastic processes is the martingale concept. This concept is particularly important in regression analysis, as the errors of a properly specified time series regression model are, by construction, martingale differences and martingale differences obey similar central limit theorems as independent random variables.

Consider a sequence (X_j) of independent random variables defined on a common probability space, satisfying $E|X_j| < \infty$, $E X_j = 0$, and let

$$Y_n = \sum_{j=1}^n X_j.$$

Moreover, let F_n be the Borel field generated by $(X_1, \dots, X_n)$. Then

$$\begin{aligned} E[Y_n | F_{n-1}] &= E[X_n + Y_{n-1} | F_{n-1}] = E[X_n | F_{n-1}] + E[Y_{n-1} | F_{n-1}] \\ &= Y_{n-1} \text{ a.s.}, \end{aligned} \quad (9.1.3)$$

for $E[X_n | F_{n-1}] = 0$ by independence and the condition $E X_n = 0$ [cf. theorem 9.1.1 (VIII)], whereas by theorem 9.1.1 (IIb) and the fact that

$$Y_n \text{ is defined on } (\Omega, F_n, P), \quad (9.1.4)$$

we have $E[Y_{n-1} | F_{n-1}] = Y_{n-1}$ a.s. Furthermore, observe that

$$E|Y_n| < \infty \quad (9.1.5)$$

and

$$F_n \supset F_{n-1} \quad (9.1.6)$$

The properties (9.1.3) through (9.1.6) are just the defining conditions for a martingale:

Definition 9.1.3 Let (Y) be a sequence of random variables satisfying (a) $E|Y_n| < \infty$, (b) Y_n is defined on (Ω, F_n, P) , (c) $F_n \subset F_{n+1}$, and (d) $Y_n = E[Y_{n+1} | F_n]$ a.s. Then Y_n is called a martingale.

Next, let

$$U_n = Y_n - Y_{n-1}.$$

Then by (d),

$$E[U_n | F_{n-1}] = 0 \text{ a.s.},$$

whereas conditions (a), (b) and (c) in definition 9.1.3 hold with Y_n replaced by U_n . Such a sequence (U_n) is called a martingale difference sequence. Conversely, let (U_n) be such that

$$E|U_n| < \infty, U_n \text{ is defined on } (\Omega, F_n, P), \text{ with } F_n \subset F_{n+1}, \text{ and } E[U_n | F_{n-1}] = 0 \text{ a.s.} \quad (9.1.7)$$

Can we construct a martingale (Y_n) such that

$$U_n = Y_n - Y_{n-1}?$$

In general the answer is no. The only possible candidate for Y_n is

$$Y_n = \sum_{j \leq n} U_j,$$

but in general $E Y_n$ is not defined. For example, if the U_n 's are independent random drawings from the standard normal distribution then $E|Y_n| = \infty$. However, for every one-sided infinite sequence (U_n) , $n=1,2,\dots$, satisfying (9.1.7) we can define a martingale (Y_n) such that $U_n = Y_n - Y_{n-1}$ for $n \geq 1$, namely

$$Y_n = \sum_{j=1}^n U_j \text{ for } n \geq 1, \quad (9.1.8)$$

$$Y_n = 0 \text{ for } n < 1.$$

Therefore we restrict the martingale difference concept to one-sided infinite sequences:

Definition 9.1.4. Let (U_n) , $n=1,2,\dots$, be a sequence of random variables such that for $n \geq 1$, (a) $E|U_n| < \infty$, (b) U_n is defined on (Ω, F_n, P) , (c) $F_n \subset F_{n+1}$, $F_0 = (\Omega, \emptyset)$, (d) $E[U_n | F_{n-1}] = 0$ a.s. Then (U_n) , $n=1,2,\dots$, is called a martingale difference sequence.

Note that for $n=1$ condition (d) follows from the definition of F_0 , which is called a trivial Borel field, and the condition $E U_1 = 0$, for

$$E(U_1 | F_0) = E U_1 \quad (9.1.9)$$

9.1.3 Martingale convergence theorems

Given a random variable X defined on $(\Omega, \mathcal{F}, P)$, with $E|X| < \infty$, a martingale (Y_n) can be constructed as follows: let $(\mathcal{F}_n)$ be any increasing sequence of Borel fields contained in $\mathcal{F}$ and let

$$Y_n = E(X|\mathcal{F}_n) \quad (9.1.10)$$

Then by theorem 9.1.1 (VI),

$$|Y_n| = |E(X|\mathcal{F}_n)| \leq E(|X| | \mathcal{F}_n),$$

hence by theorem 9.1.1(IIa),

$$E|Y_n| \leq E|X| < \infty$$

Moreover, by definition 9.1.1, Y_n is defined on $(\Omega, \mathcal{F}_n, P)$. Finally,

$$E[Y_{n+1} | \mathcal{F}_n] = E[E(X|\mathcal{F}_{n+1}) | \mathcal{F}_n] = E(X|\mathcal{F}_n) = Y_n$$

by theorem 9.1.1(I). Thus (Y_n) is a martingale. This construction is important because it enables us to prove the existence of

$$\lim_{n \rightarrow \infty} E(X|\mathcal{F}_n),$$

on the basis of the following dominated convergence theorem for martingales.

Theorem 9.1.2 Let (Y_n) be a martingale satisfying

$$\sup_n E|Y_n| < \infty.$$

Then Y_n converges a.s. to an a.s. finite limit Y_∞ .

Proof: Chung (1974, theorem 9.4.4).

Thus the martingale (Y_n) defined in (9.1.10) satisfies:

$$\lim_{n \rightarrow \infty} E(X|\mathcal{F}_n) = Y_\infty \text{ a.s., with } |Y_\infty| < \infty \text{ a.s.}$$

Moreover, we can identify Y_∞ as a conditional expectation of X :

Theorem 9.1.3 Let X be a random variable defined on $(\Omega, \mathcal{F}, P)$, satisfying $E|X| < \infty$, and let $\mathcal{F}_n$ be an increasing sequence of Borel fields contained in $\mathcal{F}$. Then $\lim_{n \rightarrow \infty} E[X | \mathcal{F}_n] = E[X | \mathcal{F}_\infty]$, where $\mathcal{F}_\infty = \bigvee_{n \geq 1} \mathcal{F}_n$.

Proof: Chung (1974, theorem 9.4.8).

A direct consequence of theorem 9.1.3 and definition 9.1.2 is:

Theorem 9.1.4 Let X be defined as in theorem 9.1.3 and let (Z_t) , $t=1,2,\dots$, be a sequence of random variables or vectors defined on $(\Omega, \mathcal{F}, P)$. Then

$$\lim_{n \rightarrow \infty} E[X | Z_1, Z_2, \dots, Z_n] = E[X | Z_1, Z_2, \dots] \text{ a.s.}$$

Another application of theorem 9.1.3 we need later on is the following:

Theorem 9.1.5 Let the conditions of theorem 9.1.4 hold. Let $Z_t^{(\ell)}$ be the random vector consisting of the components of Z_t truncated to ℓ decimal digits. Then

$$\lim_{\ell \rightarrow \infty} E[X | Z_1^{(\ell)}, Z_2^{(\ell)}, \dots] = E[X | Z_1, Z_2, \dots] \text{ a.s.}$$

Proof: Let G_ℓ be the Borel field generated by $(Z_t^{(\ell)})$, $t=1,2,\dots$ and let $G_\infty = \bigvee_{\ell \geq 1} G_\ell$. Since for each ℓ , $G_\ell \subset \mathcal{F}_\infty$, we have $G_\infty \subset \mathcal{F}_\infty$. Moreover, since for each ℓ , $G_\ell \subset G_{\ell+1}$, it follows from theorem 9.1.3 that

$$\lim_{\ell \rightarrow \infty} E[X | Z_1^{(\ell)}, Z_2^{(\ell)}, \dots] = E[X | G_\infty] \text{ a.s.}$$

The theorem can now be proved by showing

$$\mathcal{F}_\infty \subset G_\infty, \tag{9.1.11}$$

as then $\mathcal{F}_\infty = G_\infty$. To prove (9.1.11), assume first that the Z_t ($= z_t(\omega)$) are scalar random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$. Next, prove along the line of the proof of theorem 1.3.2 that for arbitrary $\xi \in \mathbb{R}$,

$$\{\omega \in \Omega : z_t(\omega) \leq \xi\} = \{\omega \in \Omega : \limsup_{\ell \rightarrow \infty} z_t^{(\ell)}(\omega) \leq \xi\} \in G_\infty$$

and conclude from this [cf. theorem 1.1.1] that for every Borel set B_t in $\mathbb{R}$,

$$\{\omega \in \Omega : z_t(\omega) \in B_t\} \in G_\infty.$$

See further exercise 4.

Q.E.D.

The importance of theorem 9.1.5 is that the sequence $(Z_t^{(\ell)})$ is rational-valued, hence countably-valued. Cf. Royden (1968, proposition 6, p.21). This property will enable us to identify $E[X|Z_1, Z_2, \dots]$ in terms of unconditional expectations, similarly to theorem 3.3.4. See Bierens (1987, 1988) and chapter 11.

9.1.4 A martingale difference central limit theorem

The martingale difference central limit theorem we consider here is an extension of theorem 2.4.3 to double arrays $(X_{n,j})$ of martingale differences, based on the following theorem of McLeish (1974).

Theorem 9.1.6 Let $(X_{n,j})$ be a martingale difference array, i.e.

$$E(X_{n,j} | X_{n,j-1}, \dots, X_{n,1}) = 0 \text{ for } j \geq 2, \quad E X_{n,1} = 0,$$

satisfying

$$(a) \sup_{n \geq 1} E[\max_{j \leq k_n} |X_{n,j}|]^2 < \infty$$

$$(b) \max_{j \leq k_n} |X_{n,j}| \rightarrow 0 \text{ in prob.}$$

$$(c) \sum_{j=1}^{k_n} X_{n,j}^2 \rightarrow 1 \text{ in prob.,}$$

where $k_n \rightarrow \infty$ as $n \rightarrow \infty$. Then

$$\sum_{j=1}^{k_n} X_{n,j} \rightarrow N(0,1) \text{ in distribution.}$$

Proof: McLeish (1974, theorem 2.3)

Using this theorem it is not too hard to prove the following generalisation of theorem 2.4.3.

Theorem 9.1.7. Let $(X_{n,j})$ be a double array of martingale differences. If

$$\text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n X_{n,j}^2 = \lim_{n \rightarrow \infty} (1/n) \sum_{j=1}^n E X_{n,j}^2 = \sigma^2 \in (0, \infty) \quad (9.1.12)$$

and

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n E |X_{n,j} / \sqrt{n}|^{2+\delta} = 0 \text{ for some } \delta > 0 \quad (9.1.13)$$

then $(1/\sqrt{n}) \sum_{j=1}^n X_{n,j} \rightarrow N(0, \sigma^2)$ in distr.

Proof: Let

$$Y_{n,j} = X_{n,j} / \left\{ (1/n) \sum_{j=1}^n E X_{n,j}^2 \right\}^{1/2}$$

Verify that $(Y_{n,j})$ satisfies the conditions of theorem 3.4.6
Cf. exercise 5. Q.E.D.

Remark: Note that condition (9.1.13) is implied by condition

$$\sup_n (1/n) \sum_{j=1}^n E |X_{n,j}|^{2+\delta} < \infty \text{ for some } \delta > 0. \quad (9.1.14)$$

Exercises:

1. Prove that (9.1.1) for arbitrary Borel sets B in R^k implies

$$E(Y|X) = g(X).$$

2. Prove theorem 9.1.1

3. Prove (9.1.9)

4. Prove (9.1.11)

5. Prove theorem 9.1.7

9.2 Measures of dependence

9.2.1 Mixingales

To prove laws of large numbers for sequences of dependent random variables we need some conditions on the extent of dependence of the random variables involved. For example, let (X_t) be a sequence of dependent random variables with $E X_t^2 < \infty$ for each t . Then

$$\begin{aligned} \text{var}((1/n)\sum_{t=1}^n X_t) &= (1/n^2)\sum_{t=1}^n \text{var}(X_t) \\ &+ 2(1/n^2)\sum_{t=1}^{n-1}\sum_{m=1}^{n-t} \text{cov}(X_{t+m}, X_t) \end{aligned} \quad (9.2.1)$$

If

$$\sup_{n \geq 1} (1/n)\sum_{t=1}^n \text{var}(X_t) = M < \infty \quad (9.2.2)$$

then the first term in (9.2.1) is $O(1/n)$. So for showing that (9.2.1) converges to zero, and hence by Chebishev's inequality,

$$\text{plim}_{n \rightarrow \infty} (1/n)\sum_{t=1}^n (X_t - E X_t) = 0,$$

we need a condition that ensures that the second term in (9.2.1) is $o(1)$. So let us have a closer look at the covariances in (9.2.1).

First we introduce some notation, which will be maintained throughout this section. For $-\infty \leq n \leq m \leq \infty$,

F_n^m is the Borel field generated by $X_n, X_{n+1}, X_{n+2}, \dots, X_m$.

Now by theorem 9.1.1(II,VII) and Cauchy-Schwarz-inequality,

$$\begin{aligned} |\text{cov}(X_{t+m}, X_t)| &= |E(X_{t+m} - E X_{t+m})(X_t - E X_t)| \\ &= |E(E(X_{t+m} | F_{-\infty}^t) - E X_{t+m})(X_t - E X_t)| \\ &\leq [E(E(X_{t+m} | F_{-\infty}^t) - E X_{t+m})^2]^{1/2} [\text{var}(X_t)]^{1/2}. \end{aligned} \quad (9.2.3)$$

Denoting

$$\eta(m) = \sup_{t \geq 1} [E(E(X_{t+m} | F_{-\infty}^t) - E X_{t+m})^2]^{1/2} \quad (9.2.4)$$

it follows now from (9.2.1) through (9.2.3) and Liapounov's inequality

$$\begin{aligned}
 & \text{var}[(1/n)\sum_{t=1}^n X_t] \\
 & \leq (1/n)M + 2(1/n)\sum_{t=1}^n [\text{var}(X_t)]^{1/2}(1/n)\sum_{m=1}^n \eta(m) \\
 & \leq (1/n)M + 2\{(1/n)\sum_{t=1}^n \text{var}(X_t)\}^{1/2}(1/n)\sum_{m=1}^n \eta(m) \\
 & \leq (1/n)M + 2M^{1/2}(1/n)\sum_{m=1}^n \eta(m) \rightarrow 0 \quad (9.2.5)
 \end{aligned}$$

if

$$\lim_{m \rightarrow \infty} \eta(m) = 0. \quad (9.2.6)$$

Condition (9.2.6), together with (9.2.2), is therefore a sufficient condition for a weak law of large numbers. Sequences satisfying condition (9.2.6) are called *mixingales* [Cf. McLeish (1975)].

Definition 9.2.1. Let (X_t) be a sequence of random variables defined on a common probability space, satisfying $E X_t^2 < \infty$ for each $t \geq 1$. Then (X_t) is called a *mixingale* if there exists nonnegative constants c_t , ψ_m , with $\psi_m \rightarrow 0$ as $m \rightarrow \infty$, such that for all $t \geq 1$, $m \geq 0$,

$$[E(E(X_{t+m} | F_{-\infty}^t) - E X_{t+m})^2]^{1/2} \leq c_{t+m} \psi_m$$

Thus, condition (9.2.6) now holds if $\sup_{t \geq 1} c_t < \infty$, as then $\eta(m) = O(\psi_m)$.

Examples of mixingales:

1. *Independent sequences.* If the X_t 's are independent then by theorem 9.1.1 (VIII),

$$E(X_{t+m} | F_{-\infty}^t) = E X_{t+m} \text{ for } m \geq 1,$$

hence $\psi_m = 0$ for $m \geq 1$. For $m = 0$ we have

$$E(X_t | F_{-\infty}^t) = X_t$$

[cf. theorem 9.1.1(IIb)], hence we may choose

$$c_t = [\text{var}(X_t)]^{1/2}, \psi_0 = 1.$$

2. *Finite dependent sequences.* The sequence (X_t) is finite dependent if for some nonnegative integer ℓ , $X_{t+\ell}$, $X_{t+\ell+1}$, ... are independent of X_t , X_{t-1} , X_{t-2} , ... Then for $m \geq \ell$,

$$E(X_{t+m} | F_{-\infty}^t) = E X_{t+m},$$

hence $\psi_m = 0$ for $m \geq \ell$. We may now choose

$$\psi_m = 1 \text{ for } m < \ell$$

$$c_t = \max_{0 \leq m \leq \ell} [E(E(X_{t+m} | F_{-\infty}^t) - E X_{t+m})^2]^{1/2}$$

3. *AR(1) processes.* Let

$$X_t = \rho X_{t-1} + U_t$$

where $|\rho| < 1$ and (U_t) is an independent process with

$$E U_t = 0, E U_t^2 = \sigma^2 < \infty.$$

Then

$$X_t = \sum_{j=0}^{\infty} \rho^j U_{t-j}.$$

Since X_t , X_{t-1} , X_{t-2} , ... can be constructed from U_t , U_{t-1} , U_{t-2} , ... and vice versa, the Borel field $F_{-\infty}^t$ is also the Borel field generated by the latter sequence. We now have

$$\begin{aligned} E(X_{t+m} | F_{-\infty}^t) &= E[\sum_{j=0}^{\infty} \rho^j U_{t+m-j} | U_t, U_{t-1}, \dots] \\ &= E[\sum_{j=0}^{m-1} \rho^j U_{t+m-j} + \sum_{j=m}^{\infty} \rho^j U_{t+m-j} | U_t, U_{t-1}, \dots] \\ &= \sum_{j=m}^{\infty} \rho^j U_{t+m-j} \end{aligned}$$

and $E X_{t+m} = 0$. Thus

$$[E(E(X_{t+m} | F_{-\infty}^t) - E X_{t+m})^2]^{1/2} = (\sum_{j=m}^{\infty} \rho^{2j} \sigma^2)^{1/2} = |\rho|^m (\sum_{j=0}^{\infty} \rho^{2j} \sigma^2)^{1/2}.$$

So we may choose

$$c_t = \sigma, \psi_m = |\rho|^m / \sqrt{1-\rho^2}$$

4. ARMA(p,q) processes. Let

$$X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + \dots + \rho_p X_{t-p} + U_t - \tau_1 U_{t-1} - \dots - \tau_q U_{t-q}, \quad (9.2.7)$$

where (U_t) is an independent sequence with

$$E U_t = 0, E U_t^2 = \sigma^2 < \infty,$$

and the lag polynomials

$$\rho(L) = 1 - \rho_1 L - \rho_2 L^2 - \dots - \rho_p L^p \quad (9.2.8)$$

$$\tau(L) = 1 - \tau_1 L - \tau_2 L^2 - \dots - \tau_q L^q \quad (9.2.9)$$

have roots all outside the complex unit circle. Then X_t has a moving average representation:

$$X_t = \sum_{j=0}^{\infty} \beta_j U_{t-j} \quad (9.2.10)$$

where the β_j 's are exponentially decreasing:

$$|\beta_j| \leq c_* \rho^j \text{ for some } c_* > 0, \rho \in (0,1), \quad (9.2.11)$$

and U_t has a moving average representation:

$$U_t = \sum_{j=0}^{\infty} \gamma_j X_{t-j}. \quad (9.2.12)$$

Cf. Harvey (1981). Therefore, again,

$$F_{-\infty}^t \text{ is also the Borel field generated by } U_t, U_{t-1}, U_{t-2}, \dots,$$

and so

$$E(X_{t+m} | F_{-\infty}^t) = E(\sum_{j=0}^{\infty} \beta_j U_{t+m-j} | U_t, U_{t-1}, \dots) = \sum_{j=m}^{\infty} \beta_j U_{t+m-j},$$

$$E X_{t+m} = 0.$$

Consequently,

$$\begin{aligned} \sup_{t \geq 1} [E(E(X_{t+m} | F_{-\infty}^t) - E X_{t+m})^2]^{1/2} \\ = (\sum_{j=-m}^{\infty} \beta_j^2 \sigma^2)^{1/2} \rightarrow 0 \text{ as } m \rightarrow \infty \end{aligned}$$

So we may take

$$c_t = \sigma, \quad \psi_m = (\sum_{j=-m}^{\infty} \beta_j^2)^{1/2} \leq c_* |\rho|^m / \sqrt{1-\rho^2}.$$

9.2.2 Uniform and strong mixing

A disadvantage of the mixingale concept is that it does not necessarily carry over to functions of mixingales. However, it is desirable to work with conditions that are invariant under Borel measurable transformations. We consider here two of these invariant conditions, namely the uniform (or φ -) mixing condition and the strong (or α -) mixing condition. Cf. Iosifescu and Theodorescu (1969) and Rosenblatt (1956).

Definition 9.2.2. Let (X_t) be a sequence of random variables or vectors defined on a common probability space. Denote for $m \geq 0$

$$\begin{aligned} \varphi(m) &= \sup_t \sup_{A \in F_{-\infty}^t, B \in F_{t+m}^{\infty}, P(A) > 0} |P(B|A) - P(B)| \\ \alpha(m) &= \sup_t \sup_{A \in F_{-\infty}^t, B \in F_{t+m}^{\infty}} |P(A \cap B) - P(A)P(B)| \end{aligned}$$

If

$$\lim_{m \rightarrow \infty} \varphi(m) = 0$$

then (X_t) is called a uniform (or φ -) mixing process, and if

$$\lim_{m \rightarrow \infty} \alpha(m) = 0$$

then (X_t) is called a strong (or α -) mixing process, with mixing coefficients φ and α , respectively.

The quantities $\varphi(m)$ and $\alpha(m)$ measure how much dependence exists between events separated by at least m time periods. Clearly,

$$c_t = (E|X_t|^r)^{1/r} < \infty \text{ for some } r > 2 \quad (9.2.14)$$

then the mixingale coefficient ψ_m equals

$$\psi_m = 2\varphi(m)^{1-1/r}, \quad \psi_m = 2(1 + \sqrt{2})\alpha(m)^{1/2-1/r}, \quad (9.2.15)$$

respectively.

9.2.3 ν -Stability

A disadvantage of the mixing conditions is that they are hard to verify. The traditional time series models, in particular ARMA(p,q) and AR(p) models, however, can be represented by moving averages of mixing variables, provided the errors are assumed to be independent. For example, a standard ARMA(p,q) model may be considered as an AR(p) model with q-dependent hence φ - and α -mixing, errors, and if the AR lag polynomial (9.2.8) has roots all outside the unit circle the process can be represented by an infinite moving average of these q-dependent errors. A natural extension of these kind of processes is therefore to consider processes of the type

$$X_t = f_t(U_t, U_{t-1}, U_{t-2}, \dots), \quad (9.2.16)$$

where (U_t) is a mixing process and f_t is a mapping from the space of one-sided infinite sequences in the range of the U_t 's such that the righthand side of (9.2.16) is a properly defined random variable or vector.

In order to limit the memory of this process (X_t) we need a condition which ensures that the impact of remote U_{t-m} 's on X_t vanishes:

Definition 9.2.3. Let (U_t) be a sequence of random vectors in R^k and let $X_t \in R^k$ be defined by (9.2.16). For some $r \geq 1$ and all $t \geq 1$, let $E|X_t|^r < \infty$, and denote

$$\nu(m) = \sup_{t \geq 1} (E|E(X_t | G_{t-m}^t) - X_t|^r)^{1/r},$$

where

the φ -mixing concept implies the α -mixing concept, for $P(B|A) = P(A \cap B)/P(A)$, hence $\alpha(m) \leq \varphi(m)$. Thus the strong mixing concept is weaker a condition than the uniform mixing concept.

Examples of φ -mixing processes are independent processes ($\varphi(m) = 0$ for $m \geq 1$), finite dependent processes ($\varphi(m) = 0$ for m larger than some finite integer) and stationary Markov processes with finite state space [cf. Billingsley (1968, pp. 167-168)]. An example of an α -mixing process that is not φ -mixing is a Gaussian AR(1) process. [Cf. Ibragimov and Linnik (1971, pp. 312-313)].

An important feature of the mixing concepts is that they are invariant under arbitrary Borel measurable transformations:

If (X_t) is an φ -[α -] mixing process, then so is $(\psi_t(X_t))$,
(9.2.13)

where (ψ_t) is an arbitrary sequence of conformable Borel measurable functions. Moreover, a φ - or α -mixing process is a mixingale, provided a certain moment condition holds. The following theorem points out the relation between the mixing and the mixingale concepts. x

Theorem 9.2.1. Let (X_t) be a sequence of random vectors satisfying

$$(E|X_t|^r)^{1/r} < \infty \text{ for some } r \geq 1, r \leq \infty.$$

Then for $1 \leq p \leq r$,

$$\begin{aligned} [E|E[X_{t+m}|F_{-\infty}^t] - E X_{t+m}|^p]^{1/p} &\leq 2 \varphi(m)^{1-1/r} (E|X_{t+m}|^r)^{1/r}, \\ [E|E[X_{t+m}|F_{-\infty}^t] - E X_{t+m}|^p]^{1/p} \\ &\leq 2(2^{1/p}+1)\alpha(m)^{1/p-1/r} (E|X_{t+m}|^r)^{1/r}. \end{aligned}$$

Proof: Serfling (1968) for the φ -mixing case and McLeish (1975, Lemma 2.1) for the α -mixing case.

Thus, if

G_{t-m}^t is the Borel field generated by $U_t, U_{t-1}, \dots, U_{t-m}$.

If

$$\lim_{m \rightarrow \infty} \nu(m) = 0,$$

then (X_t) is a ν -stable process in L^r with respect to the base (U_t) .

Cf. McLeish (1975) and Bierens (1983). This concept is reminiscent of the stochastic stability concept of Bierens (1981), which is somewhat weaker a condition.

The actual contents of this definition is not in the first place the condition that $\nu(m) \rightarrow 0$, for this is always satisfied if the processes (X_t) and (U_t) are strictly stationary, as in that case we may delete " $\sup_{t \geq 1}$ " and take for t some arbitrary fixed index, say $t = 1$:

$$\nu(m) = (E|E(X_1 | G_{1-m}^1) - X_1|^r)^{1/r}$$

From theorem 9.1.3 it then follows

$$\lim_{m \rightarrow \infty} E(X_1 | G_{1-m}^1) = E(X_1 | G_{-\infty}^1) \text{ a.s.}$$

where

$$G_{-\infty}^1 = \bigvee_{m \geq 0} G_{1-m}^1$$

is the Borel Field generated by $U_1, U_0, U_{-1}, \dots$. Since X_1 is a function of this one-sided infinite sequence, the Borel field generated by X_1 is contained in $G_{-\infty}^1$, by which

$$E(X_1 | G_{-\infty}^1) = X_1 \text{ a.s.}$$

Cf. theorem 9.1.1 (IIb). Thus,

$$\lim_{m \rightarrow \infty} E(X_1 | G_{1-m}^1) = X_1 \text{ a.s.}$$

and consequently by dominated convergence [theorem 2.2.2],

$$\nu(m) \rightarrow 0 \text{ as } m \rightarrow \infty.$$

Summarizing, we have shown:

Theorem 9.2.2. Let (U_t) be a strictly stationary stochastic process in R^k and let for each t

$$X_t = f(U_t, U_{t-1}, U_{t-2}, \dots)$$

be a well-defined random vector in R^k , where f does not depend on t . If $E|X_t|^r < \infty$ for some $r \geq 1$ then (X_t) is ν -stable in L^r with respect to the base (U_t) .

Therefore, the actual contents of definition 9.2.3 is that this result goes through if the X_t 's and/or U_t 's are heterogeneously distributed (that means that the distributions of U_t and/or X_t depend on t). In other words, the ν -stability condition actually imposes restrictions on the extent of heterogeneity of (X_t) and (U_t) .

In checking the ν -stability condition the following result is useful.

Theorem 9.2.3. Let X_t be of the form (9.2.16). If there exist Borel measurable mapping $f_{t,m}$ such that for some $r \geq 1$,

$$E|X_t|^r < \infty, \quad t \geq 1,$$

$$\begin{aligned} \nu^*(m) &= \sup_{t \geq 1} (E|f_{t,m}(U_t, U_{t-1}, \dots, U_{t-m}) - X_t|^r)^{1/r} \\ &\rightarrow 0 \text{ as } m \rightarrow \infty, \end{aligned}$$

then (X_t) is ν -stable in L^r with respect to the base (U_t) , where

$$\nu(m) \leq 2\nu^*(m).$$

Proof: Observe that

$$E(X_t | \mathcal{G}_{t-m}^t) = E(X_t | U_t, U_{t-1}, \dots, U_{t-m})$$

is a Borel measurable function of the conditioning variables. By Jensen's inequality for conditional expectations (theorem 3.2.6) it follows

$$|E(X_t | U_t, U_{t-1}, \dots, U_{t-m}) - f_{t,m}(U_t, \dots, U_{t-m})|^r \\ \leq E[|X_t - f_{t,m}(U_t, \dots, U_{t-m})|^r | U_t, \dots, U_{t-m}]$$

hence

$$E|E(X_t | G_{t-m}^t) - f_{t,m}(U_t, U_{t-1}, \dots, U_{t-m})|^r \\ \leq E|X_t - f_{t,m}(U_t, U_{t-1}, \dots, U_{t-m})|^r$$

Using Minkowski's inequality (cf. section 1.4) it now follows:

$$(E|E(X_t | G_{t-m}^t) - X_t|^r)^{1/r} \leq 2(E|X_t - f_{t,m}(U_t, \dots, U_{t-m})|^r)^{1/r}$$

Q.E.D.

It should be noted that the ν -stability concept is only meaningful if we make further assumptions on the dependence of the base (U_t) , for every sequence (X_t) is ν -stable with respect to itself. In the next section we shall impose mixing conditions on (U_t) , in order to prove various laws of large numbers.

The ν -stability concept, like the mixingale concept, is not generally invariant under Borel measurable transformations. However, an invariance property holds under bounded uniformly continuous transformation, and this is just what we need to generalize the weak law of large numbers in sections 2.5 and 2.7 to ν -stable processes in L^2 with respect to a mixing base. Thus, let (X_t) obey the conditions in definition 9.2.3 and let ψ be a bounded uniformly continuous function on the domain of the X_t 's. For notational convenience, let for $m \geq 0$

$$X_t^{(m)} = E(X_t | G_{t-m}^t)$$

and let for $a \geq 0$

$$\zeta(a) = \sup_{|x_1 - x_2| \leq a} |\psi(x_1) - \psi(x_2)| \quad (9.2.17)$$

Then for arbitrary $q > 0$

$$\begin{aligned}
& E|\psi(X_t^{(m)}) - \psi(X_t)|^q \leq \zeta(a)^q \\
& + 2\sup_x |\psi(x)|^q P(|X_t^{(m)} - X_t| > a) \\
& \leq \zeta(a)^q + 2\sup_x |\psi(x)|^q E|X_t^{(m)} - X_t|^r / a^r \\
& \leq \zeta(a)^q + 2\sup_x |\psi(x)|^q \nu(m)^r / a^r,
\end{aligned}$$

where the second inequality follows from Chebishev's inequality. Then

$$\sup_{t \geq 1} (E|\psi(X_t^{(m)}) - \psi(X_t)|^q)^{1/q} \leq \nu^*(m),$$

where

$$\nu^*(m) = 2 \cdot \inf_{a \geq 0} \{\zeta(a)^q + 2\sup_x |\psi(X_1)|^q \nu(m)^r / a^r\}^{1/q} \quad (9.2.18)$$

Using the fact that $\zeta(a) \rightarrow 0$ for $a \downarrow 0$, it is easy to show $\nu^*(m) \rightarrow 0$ for $m \rightarrow \infty$. It follows now from theorem 9.2.3 that $\psi(X_t)$ is $2\nu^*(m)$ -stable in L^q with respect to the base (U_t) :

Theorem 9.2.4. Let (X_t) be ν -stable in L^r with respect to the base (U_t) . Let ψ be a bounded uniformly continuous function on the domain of the X_t 's. Let $q > 0$ be arbitrary. Then $\psi(X_t)$ is ν^* -stable in L^q with respect to the base (U_t) , where ν^* is defined by (9.2.17) and (9.2.18).

9.3 Weak Laws of large numbers for dependent random variables

In section 9.2 we already have derived a weak law of large numbers for mixingales. We shall now generalize this result to processes that are ν -stable in L^1 with respect to a φ - or α -mixing base. Let (X_t) be a sequence of bounded random variables, i.e., for some $M < \infty$ and all t ,

$$P(|X_t| \leq M) = 1, \quad (9.3.1)$$

with the structure (9.2.16). Define the Borel fields

$$F_{-\infty}^t, F_t^\infty, G_{-\infty}^t \text{ and } G_t^\infty$$

as in section 9.2. Moreover, let the base (U_t) be φ -mixing or

α -mixing and denote

$$X_t^{(m)} = E(X_t | G_{t-m}^t) = g_{t,m}(U_t, U_{t-1}, \dots, U_{t-m}), \quad (9.3.2)$$

say. Note that $X_t^{(m)}$ is bounded too by M . We shall now derive an upper bound of $|\text{cov}(X_{t+l}, X_t)|$ in terms of ν , φ and/or α . First observe that by (9.3.1) and the definition of ν -stability,

$$\begin{aligned} & |\text{cov}(X_{t+l}, X_t) - \text{cov}(X_{t+l}^{(m)}, X_t)| \\ & \leq 2 \cdot E|X_{t+l} - X_{t+l}^{(m)}| |X_t - E X_t| \\ & \leq 4M \cdot \nu(m), \quad \text{uniformly in } l. \end{aligned} \quad (9.3.3)$$

Next observe that for fixed m , and α -mixing (U_t)

$$\begin{aligned} & X_t^{(m)} - E X_t^{(m)} \\ & = g_{t,m}(U_t, U_{t-1}, \dots, U_{t-m}) - E X_t^{(m)} \text{ is } \alpha^*\text{-mixing} \end{aligned} \quad (9.3.4)$$

with

$$\begin{aligned} \alpha^*(m^*) &= 1 \text{ if } m^* < m, \\ \alpha^*(m^*) &= \alpha(m^* - m) \text{ if } m^* \geq m, \end{aligned} \quad (9.3.5)$$

Similarly, if (U_t) is φ -mixing the process (9.3.3) is φ^* -mixing, with

$$\begin{aligned} \varphi^*(m^*) &= 1 \text{ if } m^* < m, \\ \varphi^*(m^*) &= \varphi(m^* - m) \text{ if } m^* \geq m \end{aligned} \quad (9.3.6)$$

Thus by theorem 9.2.1 with $p = 1$ and $r = \infty$ it follows,

$$\begin{aligned} E|E(X_{t+l}^{(m)} | G_{-\infty}^t) - E X_{t+l}| &\leq 2M\varphi^*(l), \\ E|E(X_{t+l}^{(m)} | G_{-\infty}^t) - E X_{t+l}| &\leq 6M\alpha^*(l). \end{aligned} \quad (9.3.7)$$

Hence, similarly to (9.2.3), we have

$$\begin{aligned}
|\text{cov}(X_{t+\ell}^{(m)}, X_t)| &\leq 2M^2\varphi(\ell-m), \\
|\text{cov}(X_{t+\ell}^{(m)}, X_t)| &\leq 6M^2\alpha(\ell-m),
\end{aligned}
\tag{9.3.8}$$

where $\varphi(m) = \alpha(m) = 1$ if $m \leq 0$. Combining (9.3.3) and (9.3.8) now yields:

Lemma 9.3.1. Let (X_t) be ν -stable in L^1 with respect to a base (U_t) and let φ and α be the mixing coefficients corresponding to (U_t) . Moreover, let for some $M < \infty$ and all t , $P(|X_t| \leq M) = 1$. Then for $\ell \geq 1$, $m \geq 0$

$$|\text{cov}(X_{t+\ell}, X_t)| \leq M^2\psi(\ell-m) + 4M \cdot \nu(m),$$

where

$$\psi(\ell-m) = \min[2\varphi(\ell-m), 6\alpha(\ell-m)] \text{ if } \ell-m > 0,$$

$$\psi(\ell-m) = 1 \text{ if } \ell-m \leq 0.$$

Use this result in (9.2.1) now yields:

$$|\text{var}((1/n)\sum_{t=1}^n X_t)| \leq M^2/n + 8\nu(m)M + 2n^{-1}M^2[m\psi(0) + \sum_{\ell=0}^{\infty}\psi(\ell)].$$

Letting $m \rightarrow \infty$ with n at rate $o(n)$, we now see that the following lemma holds

Lemma 9.3.2. Let the bounded stochastic process (X_t) be ν -stable in L^1 with respect to the base (U_t) . Assume either

- (a) (U_t) is φ -mixing with $\sum_{\ell=0}^{\infty}\varphi(\ell) < \infty$, or
- (b) (U_t) is α -mixing with $\sum_{\ell=0}^{\infty}\alpha(\ell) < \infty$.

Then $\lim_{n \rightarrow \infty} \text{var}[(1/n)\sum_{t=1}^n X_t] = 0$.

We are now ready to generalize theorem 2.5.2 to ν -stable processes in L^1 with respect to a mixing base.

Theorem 9.3.1. Let (X_t) be an R^k -valued v -stable process in L^1 with respect to a φ - or α -mixing base. Let ψ be a continuous function on R^k and let F_t be the distribution function of X_t . If

$$(1/n)\sum_{t=1}^n F_t \rightarrow G \text{ properly, pointwise,}$$

$$\sup_{n \geq 1} (1/n)\sum_{t=1}^n E|\psi(X_t)|^{1+\delta} < \infty \text{ for some } \delta > 0$$

and either

$$\sum_{\ell=0}^{\infty} \varphi(\ell) < \infty \text{ or } \sum_{\ell=0}^{\infty} \alpha(\ell) < \infty,$$

$$\text{then } \lim_{n \rightarrow \infty} (1/n)\sum_{t=1}^n \psi(X_t) = \int \psi(x) dG(x)$$

Proof: We prove the theorem for the case $X_t \in R$. The proof for the case $X_t \in R^k$ is almost the same. We show that (2.5.7) in the proof of theorem 2.5.2 goes through. Thus, define similarly to (2.5.1)

$$\begin{aligned} \psi_a(x) &= \psi(x) & \text{if } |\psi(x)| \leq a, \\ \psi_a(x) &= a & \text{if } \psi(x) > a, \\ \psi_a(x) &= -a & \text{if } \psi(x) < -a, \end{aligned} \tag{9.3.9}$$

and let for $b > 0$,

$$\begin{aligned} \psi_{ab}(x) &= \psi_a(x) & \text{if } |x| \leq b, \\ \psi_{ab}(x) &= \psi_a(b \cdot x/|x|) & \text{if } |x| > b, \end{aligned} \tag{9.3.10}$$

Then $\psi_{ab}(x)$ is bounded and uniformly continuous, for the set

$$\{x \in R^k: |x| \leq b\}$$

is a closed and bounded subset of R^k and hence compact, whereas continuous functions on a compact set are uniformly continuous on that set. The proof of the uniform continuity outside this set is left as an exercise (Exercise 1). It follows from theorem 9.2.4 that $\psi_{ab}(X_t)$ is v -stable in L^1 with respect to the mixing base involved. Thus it follows from lemma 9.3.1 that

$$\lim_{n \rightarrow \infty} \text{var}[(1/n)\sum_{t=1}^n \psi_{ab}(X_t)] = 0,$$

hence by Chebishev's inequality,

$$\text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n (\psi_{ab}(X_t) - E \psi_{ab}(X_t)) = 0.$$

Since

$$\begin{aligned} \lim_{n \rightarrow \infty} E & \left| (1/n) \sum_{t=1}^n (\psi_a(X_t) - E \psi_a(X_t)) \right. \\ & \left. - (1/n) \sum_{t=1}^n (\psi_{ab}(X_t) - E \psi_{ab}(X_t)) \right| \\ & \leq 2a \cdot \lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n E I(|X_t| > b) \\ & \leq 2a \cdot \lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n [1 - F_t(b) - F_t(-b)] \\ & = 2a[1 - G(b) - G(-b)] \rightarrow 0 \text{ as } b \rightarrow \infty, \end{aligned}$$

we now have

$$\text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n (\psi_a(X_t) - E \psi_a(X_t)) = 0.$$

Thus (2.5.7) in the proof of theorem 2.5.2 goes through. Since the rest of the proof of theorem 2.5.2 does not employ the independence condition, this result is sufficient for theorem 9.3.1 to hold. Q.E.D.

Referring in the proof of theorem 2.7.2 to theorem 9.3.1 rather than to theorem 2.5.2, it immediately follows:

Theorem 9.3.2. Let (X_t) be an R^k -valued ν -stable process in L^1 with respect to a φ - or α -mixing base. Let F_t be the distribution function of X_t and let $f(x, \theta)$ be a continuous function on $R^k \times \Theta$, where Θ is a compact Borel set in R^m . If

$$\begin{aligned} (1/n) \sum_{t=1}^n F_t & \rightarrow G, \text{ properly, pointwise,} \\ \sup_{n \geq 1} (1/n) \sum_{t=1}^n E \sup_{\theta \in \Theta} |f(X_t, \theta)|^{1+\delta} & < \infty \end{aligned}$$

for some $\delta > 0$, and either

$$\sum_{\ell=0}^{\infty} \varphi(\ell) < \infty \quad \text{or} \quad \sum_{\ell=0}^{\infty} \alpha(\ell) < \infty,$$

then

$$\text{plim}_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left| (1/n) \sum_{t=1}^n f(X_t, \theta) - \int f(x, \theta) dG(x) \right| = 0,$$

where the limit function involved is continuous on Θ

Next we want to allow the function ψ in theorem 9.3.1 to be Borel measurable rather than continuous. There is no problem if the process (X_t) is strictly stationary and of the type $X_t = f(U_t, U_{t-1}, \dots)$ with (U_t) a strictly stationary mixing base, as we then may apply theorem 9.2.2.

Theorem 9.3.3. Let $X_t \in R^k$ be such that

$$X_t = f(U_t, U_{t-1}, U_{t-2}, \dots)$$

where (U_t) is a strictly stationary φ - or α -mixing process and the time-invariant mapping f is such that (X_t) is a well-defined strictly stationary stochastic process. Let ψ be a Borel measurable real function on R^k . If $E|\psi(X_t)| < \infty$ and either

$$\sum_{\ell=0}^{\infty} \varphi(\ell) < \infty \text{ or } \sum_{\ell=0}^{\infty} \alpha(\ell) < \infty,$$

then $\text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n \psi(X_t) = E \psi(X_1)$.

Proof: Define ψ_a as in (9.3.9). Then for fixed $a > 0$, ψ_a is a bounded Borel measurable real function, and by theorem 9.2.2, $\psi_a(X_t)$ is ν -stable in L^1 with respect to (U_t) . From lemma 9.3.2 and Chebishev's inequality it follows that

$$\text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n \psi_a(X_t) = E \psi_a(X_1).$$

Now by (9.3.9) and theorem 1.4.1

$$\begin{aligned} & E \left| (1/n) \sum_{t=1}^n \psi(X_t) - (1/n) \sum_{t=1}^n \psi_a(X_t) \right| \\ & \leq 2 \cdot E |\psi(X_1)| I(|\psi(X_1)| > a) \rightarrow 0 \text{ as } a \rightarrow \infty, \end{aligned}$$

for $P(|\psi(X_1)| > a) \rightarrow 0$ as $a \rightarrow \infty$, as otherwise $E|\psi(X_1)| = \infty$. Similarly

$$|E \psi(X_1) - E \psi_a(X_1)| \rightarrow 0 \text{ as } a \rightarrow \infty.$$

The theorem under review now follows from the argument in the proof of theorem 2.7.2. Q.E.D.

Again, referring to theorem 9.3.3 instead of theorem 2.7.2 (with $\delta = 0$), theorem 2.7.5 carries over.

Theorem 9.3.4. Let (X_t) be as in theorem 9.3.3. Let $f(x, \theta)$ be a Borel measurable real function on $R^k \times \Theta$, where Θ is a compact Borel set in R^m , such that for each $x \in R^k$, $f(x, \theta)$ is continuous on Θ . If in addition

$$E \sup_{\theta \in \Theta} |f(X_1, \theta)| < \infty$$

then

$$\text{plim}_{n \rightarrow \infty} \sup_{\theta \in \Theta} |(1/n) \sum_{t=1}^n f(X_t, \theta) - E f(X_1, \theta)| = 0,$$

where $E f(X_1, \theta)$ is continuous on Θ .

For generalizing theorem 9.3.1 to Borel measurable functions ψ we need the following lemma.

Lemma 9.3.3. Let ψ be a Borel measurable real function on R^k and let X be a random vector in R^k such that for some $r \geq 1$, $E|\psi(X)|^r < \infty$. For every $\varepsilon > 0$ there exists a function ψ_ε on R^k that is bounded, continuous and zero outside a compact set (hence uniformly continuous), such that

$$E|\psi(X) - \psi_\varepsilon(X)|^r < \varepsilon$$

Proof: Dunford and Schwartz (1957, p. 298).

A direct corollary of this lemma is:

Lemma 9.3.4. Let (X_t) be a sequence of random vectors in R^k and let (F_t) be the sequence of corresponding distribution functions. Assume

$$(1/n) \sum_{t=1}^n F_t \rightarrow G \text{ properly, setwise.}$$

Let ψ be a Borel measurable real function on R^k such that

$$\sup_{n \geq 1} (1/n) \sum_{t=1}^n E |\psi(X_t)|^{1+\delta} < \infty \text{ for some } \delta > 0.$$

For every $\varepsilon > 0$ there exists a uniformly continuous bounded function ψ_ε on R^k such that

$$\limsup_{n \rightarrow \infty} E \left| (1/n) \sum_{t=1}^n \psi(X_t) - (1/n) \sum_{t=1}^n \psi_\varepsilon(X_t) \right| < \varepsilon$$

Proof: Exercise 1 (Hint: combine lemma 9.3.3 with theorem 2.5.5).

Combining lemma 9.3.4 with theorem 9.3.2 yields:

Theorem 9.3.5. Let the conditions of theorem 9.3.1 be satisfied, except that now ψ is Borel measurable and

$$(1/n) \sum_{t=1}^n F_t \rightarrow G \text{ properly, setwise.}$$

Then the conclusion of theorem 9.3.1 carries over.

Proof: Exercise 2.

Finally, referring in the proof of theorem 2.7.4 to theorem 9.3.5 rather than to theorem 2.5.6, yields:

Theorem 9.3.6. Let the conditions of theorem 9.3.2 be satisfied, except that now

$$(1/n) \sum_{t=1}^n F_t \rightarrow G \text{ properly, setwise,}$$

$f(x, \theta)$ is Borel measurable on $R^k \times \Theta$ and for each $x \in R^k$ a continuous function on Θ . Then the conclusions of theorem 9.3.2 carry over.

Remark: It seems possible to generalize the results in this section further to strong laws, using theorem 3.1 of McLeish (1975). This, however, will require further conditions on the rate of convergence to zero of v , φ and α .

Exercises:

1. Prove that the function (9.3.10) is uniformly continuous.
2. Prove lemma 9.3.4
3. Prove theorem 9.3.5

9.4 *Proper heterogeneity and uniform laws for functions of infinitely many random variables*

For some time series models like ARMA and ARMAX models least squares parameter estimation is conducted by minimizing a sum of functions of a one-sided infinite sequence of random variables. See chapter 11. In order to prove consistency of these parameter estimators under data heterogeneity we need an extension of the proper convergence concept to distributions of one-sided infinite sequences of random variables (cf. definitions 2.3.1 and 2.5.1), as well as some generalisations of the uniform laws in section 9.3.

Definition 9.4.1. Let (X_t) be a sequence of random variables in R^k , and let $F_{t,m}$ be the distribution function of $(X_t, \dots, X_{t-m})$. The process (X_t) is said to be pointwise (setwise) properly heterogenous if there exists a one-sided infinite sequence (X_t^*) , $t \leq 0$, of R^k -valued random variables such that for $m = 0, 1, 2, \dots$,

$$(1/n) \sum_{t=1}^n F_{t,m} \rightarrow H_m \text{ properly pointwise (setwise),}$$

where H_m is the distribution function of $(X_0^*, X_{-1}^*, \dots, X_{-m}^*)$. The sequence (X_t^*) , $t \leq 0$ will be called the mean process.

The following theorem now specializes theorem 9.3.2. This theorem is the basis for the consistency results of linear ARMAX models, in chapter 11.

Theorem 9.4.1. Let (X_t) be an stochastic process in R^k which is ν -stable with in L^1 with respect to an α or ϕ -mixing base, where either

$$\sum_{j=0}^{\infty} \phi(j) < \infty \text{ or } \sum_{j=0}^{\infty} \alpha(j) < \infty,$$

and pointwise properly heterogenous with mean process (X_t^*) . Let $(\gamma_{j,i}(\theta))$, $j \geq 0$, $i=1, \dots, p$, be sequences of continuous map-

pings from a compact subset Θ of a Euclidean space into R^k , such that for $i=1,2,\dots,p$,

$$\sum_{j=0}^{\infty} \sup_{\theta \in \Theta} |\gamma_{j,i}(\theta)| < \infty. \quad (9.4.1)$$

Let ψ be a differentiable real function on R^p such that for $c \rightarrow \infty$,

$$\sup_{|\xi| \leq c} |(\partial/\partial \xi) \psi(\xi)| = O(c^\mu), \quad (9.4.2)$$

where $\mu > 0$ is such that for some $\delta > 0$,

$$\sup_t E |X_t|^{1+\mu+\delta} < \infty. \quad (9.4.3)$$

Denote

$$\Gamma_j(\theta) = (\gamma_{j,1}(\theta), \dots, \gamma_{j,p}(\theta))'.$$

Then

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \sup_{\theta \in \Theta} |(1/n) \sum_{t=1}^n \psi(\sum_{j=0}^{t-1} \Gamma_j(\theta)' X_{t-j}) \\ - E \psi(\sum_{j=0}^{\infty} \Gamma_j(\theta)' X_{-j}^*)| = 0. \end{aligned}$$

Moreover, the limit function $E \psi(\sum_{j=0}^{\infty} \Gamma_j(\theta)' X_{-j}^*)$ is a continuous real function on Θ .

Proof: Denote

$$\rho_j = \max_{i=1,\dots,p} \sup_{\theta \in \Theta} |\gamma_{j,i}(\theta)|. \quad (9.4.4)$$

Then condition (9.4.1) implies

$$\sum_{j=0}^{\infty} \rho_j < \infty. \quad (9.4.5)$$

Moreover, denote for non-negative integers s ,

$$\xi_t^{(s)}(\theta) = \sum_{j=0}^s \Gamma_j(\theta)' X_{t-j}. \quad (9.4.6)$$

We shall now prove theorem 9.4.1 in four steps, each stated in a lemma.

Lemma 9.4.1. There exists a constant K such that for every $n \geq 1$ and every $s \geq 0$,

$$\begin{aligned} E \sup_{\theta \in \Theta} |(1/n) \sum_{t=1}^n \psi(\xi_t^{(\infty)}(\theta)) - (1/n) \sum_{t=1}^n \psi(\xi_t^{(s)}(\theta))| \\ \leq K \cdot \sum_{j=s+1}^{\infty} \rho_j. \end{aligned}$$

Lemma 9.4.2. For every fixed $s \geq 0$,

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \sup_{\theta \in \Theta} |(1/n) \sum_{t=1}^n \psi(\sum_{j=0}^s \Gamma_j(\theta)' X_{t-j}) \\ - E \psi(\sum_{j=0}^s \Gamma_j(\theta)' X_{-j}^*)| = 0. \end{aligned}$$

Lemma 9.4.3. There exists a constant K such that for every $s \geq 0$,

$$\begin{aligned} \sup_{\theta \in \Theta} |E \psi(\sum_{j=0}^s \Gamma_j(\theta)' X_{-j}^*) - E \psi(\sum_{j=0}^{\infty} \Gamma_j(\theta)' X_{-j}^*)| \\ \leq K \cdot \sum_{j=s+1}^{\infty} \rho_j. \end{aligned}$$

Lemma 9.4.4. The function $E \psi(\sum_{j=0}^{\infty} \Gamma_j(\theta)' X_{-j}^*)$ is continuous on Θ .

Realizing that (9.4.5) implies

$$\lim_{m \rightarrow \infty} \sum_{j=m+1}^{\infty} \rho_j = 0,$$

the theorem under review now easily follows from these four lemmas.

Proof of lemma 9.4.1: Observe from (9.4.4) and (9.4.6) that for $\theta \in \Theta$,

$$\begin{aligned} |\xi_t^{(\infty)}(\theta) - \xi_t^{(s)}(\theta)| &\leq \sum_{j=s+1}^{\infty} |\Gamma_j(\theta)' X_{t-j}| \\ &\leq \sum_{j=s+1}^{\infty} \max_{i=1,2,\dots,p} |\gamma_{j,i}(\theta)' X_{t-j}| \leq \sum_{j=s+1}^{\infty} \rho_j |X_{t-j}|. \end{aligned}$$

Moreover, by the mean value theorem there exists a mean value

$$\lambda_{t,s}(\theta) \in [0,1]$$

such that

$$\begin{aligned}
& |\psi(\xi_t^{(\infty)}(\theta)) - \psi(\xi_t^{(s)}(\theta))| \\
&= |(\xi_t^{(\infty)}(\theta) - \xi_t^{(s)}(\theta))' (\partial/\partial \xi) \psi(\lambda_{t,s}(\theta) \xi_t^{(\infty)}(\theta) \\
&\quad + (1 - \lambda_{t,s}(\theta)) \xi_t^{(s)}(\theta))| \\
&\leq |\xi_t^{(\infty)}(\theta) - \xi_t^{(s)}(\theta)| \sup_{|\xi| \leq \max(|\xi_t^{(\infty)}(\theta)|, |\xi_t^{(s)}(\theta)|)} |(\partial/\partial \xi) \psi(\xi)| \\
&\leq \sum_{j=s+1}^{\infty} \rho_j |X_{t-j}| \sup_{|\xi| \leq \sum_{j=0}^{\infty} \rho_j |X_{t-j}|} |(\partial/\partial \xi) \psi(\xi)|, \quad (9.4.7)
\end{aligned}$$

where the last inequality follows from (9.4.4). According to condition (9.4.2) there exists a constant C such that

$$\sup_{|\xi| \leq a} |(\partial/\partial \xi) \psi(\xi)| = Ca^\mu, \quad (9.4.8)$$

hence by (9.4.7) and Holder's and Liapounov's inequalities

$$\begin{aligned}
& E \sup_{\theta \in \Theta} |\psi(\xi_t^{(\infty)}(\theta)) - \psi(\xi_t^{(s)}(\theta))| \\
&\leq C \cdot E\{(\sum_{j=s+1}^{\infty} \rho_j |X_{t-j}|)(\sum_{j=0}^{\infty} \rho_j |X_{t-j}|)^\mu\} \\
&\leq C\{E(\sum_{j=s+1}^{\infty} \rho_j |X_{t-j}|)^{1+\mu}\}^{1/(1+\mu)} \{E(\sum_{j=0}^{\infty} \rho_j |X_{t-j}|)^{1+\mu}\}^{\mu/(1+\mu)} \\
&\leq C\{(\sum_{j=s+1}^{\infty} \rho_j)^\mu (\sum_{j=s+1}^{\infty} \rho_j E|X_{t-j}|^{1+\mu})\}^{1/(1+\mu)} \\
&\quad \times \{(\sum_{j=0}^{\infty} \rho_j)^\mu (\sum_{j=0}^{\infty} \rho_j E|X_{t-j}|^{1+\mu})\}^{\mu/(1+\mu)} \\
&\leq C(\sum_{j=0}^{\infty} \rho_j)^\mu \sup_t E|X_t|^{1+\mu} (\sum_{j=s+1}^{\infty} \rho_j) = K \cdot \sum_{j=s+1}^{\infty} \rho_j, \quad (9.4.9)
\end{aligned}$$

say. This proves the lemma.

Q.E.D.

Proof of lemma 9.4.2: Lemma 9.4.2 follows straightforwardly from theorem 9.3.2.

Proof of lemma 9.4.3: From lemma 9.4.1 and the conditions of theorem 9.4.1 it follows

$$\lim_{n \rightarrow \infty} \sum_{t=1}^n E|X_{t-s}|^{1+\mu} = E|X_{-s}^*|^{1+\mu},$$

hence

$$\sup_{j \geq 0} E|X_{-j}^*|^{1+\mu} < \infty.$$

The lemma now follows similarly to lemma 9.4.1.

Q.E.D.

Proof of lemma 9.4.4: Let $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$. Similarly to (9.4.7) (with (9.4.8)) we have

$$\begin{aligned} & |\psi(\sum_{j=0}^{\infty} \Gamma_j(\theta_1)' X_{-j}^*) - \psi(\sum_{j=0}^{\infty} \Gamma_j(\theta_2)' X_{-j}^*)| \\ & \leq C |\sum_{j=0}^{\infty} (\Gamma_j(\theta_1) - \Gamma_j(\theta_2))' X_{-j}^*| (\sum_{j=0}^{\infty} \rho_j |X_{-j}^*|)^{\mu} \\ & \leq C (\sum_{j=0}^{\infty} (\sum_{i=0}^p |\gamma_{j,i}(\theta_1) - \gamma_{j,i}(\theta_2)|) |X_{-j}^*|) (\sum_{j=0}^{\infty} \rho_j |X_{-j}^*|)^{\mu}. \end{aligned} \quad (9.4.10)$$

Thus similarly to (9.4.9) it follows from (9.4.10),

$$\begin{aligned} & |E \psi(\sum_{j=0}^{\infty} \Gamma_j(\theta_1)' X_{-j}^*) - E \psi(\sum_{j=0}^{\infty} \Gamma_j(\theta_2)' X_{-j}^*)| \\ & \leq C. (\sum_{j=0}^{\infty} \rho_j)^{\mu} \sup_{j \geq 0} E|X_{-j}^*|^{1+\mu} \sum_{i=0}^p |\gamma_{j,i}(\theta_1) - \gamma_{j,i}(\theta_2)|. \end{aligned}$$

Since the $\gamma_{j,i}(\theta)$'s are continuous on Θ , this result proves the lemma. Q.E.D.

The next theorems are easy extensions of theorem 9.4.1. They will enable us to prove consistency of least squares parameter estimators of nonlinear ARMAX models. Cf. chapter 11.

Theorem 9.4.2. Let Θ , ψ , μ , (X_t) and (X_t^*) be as in theorem 9.4.1. Let the functions $\gamma_{i,j}(\theta, x)$, $j=0,1,2,\dots$, $i=1,2,\dots,p$, be continuous real functions on $\Theta \times \mathbb{R}^k$, such that

$$\max_{i=1,\dots,p} \sup_{\theta \in \Theta} |\gamma_{j,i}(\theta, x)| \leq \rho_j \bar{b}(x), \quad (9.4.11)$$

where

$$\sum_{j=0}^{\infty} \rho_j < \infty \quad (9.4.12)$$

and $\bar{b}(x)$ is a nonnegative continuous real function on $\mathbb{R}^k$ such that for some $\delta > 0$,

$$\sup_t E \bar{b}(X_t)^{1+\mu+\delta} < \infty. \quad (9.4.13)$$

Finally, let

$$\Gamma_j(\theta, x) = (\gamma_{j,1}(\theta, x), \dots, \gamma_{j,p}(\theta, x))'. \quad (9.4.14)$$

Then

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \sup_{\theta \in \Theta} & \left| (1/n) \sum_{t=1}^n \psi(\sum_{j=0}^{t-1} \Gamma_j(\theta, X_{t-j})) \right. \\ & \left. - E \psi(\sum_{j=0}^{\infty} \Gamma_j(\theta, X_{-j}^*)) \right| = 0. \end{aligned}$$

Moreover, the limit function $E \psi(\sum_{j=0}^{\infty} \Gamma_j(\theta, X_{-j}^*))$ is continuous on Θ .

Proof: Replacing $|X_{t-j}|$ and $|X_{-j}^*|$ in the proofs of lemmas 9.4.1 through 9.4.4 by $\bar{b}(X_{t-j})$ and $\bar{b}(X_{-j}^*)$, respectively, the theorem easily follows. Q.E.D.

Theorem 9.4.3. Let the conditions of theorem 9.4.2 be satisfied, except that (X_t) is now setwise properly heterogeneous and the functions $\gamma_{i,j}(\theta, x)$, $j=0,1,2,\dots$, $i=1,2,\dots,p$, are for each $x \in R^k$ continuous real functions on Θ and for each $\theta \in \Theta$ Borel measurable real functions on R^k . Then the conclusions of theorem 9.4.2 carry over.

Proof: Note that the function $\bar{b}(\cdot)$ may now be merely Borel measurable. The proof of this theorem is similar to the proof of theorem 9.4.2, referring to theorem 9.3.6 instead of theorem 9.3.2. Q.E.D.

References

- Bierens, H.J. (1981), *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*, New York: Springer Verlag.
- Bierens, H.J. (1983), "Uniform Consistency of Kernel Estimators of a Regression Function under Generalized Conditions", *Journal of the American Statistical Association* 78, 699-707.
- Bierens, H.J. (1987), "ARMAX Model Specification Testing, With an Application to Unemployment in the Netherlands", *Journal of Econometrics* 35, 161-190.
- Bierens, H.J. (1988), "ARMA Memory Index Modeling of Economic Time Series (with discussion)", *Econometric Theory* 4, 35-59.
- Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.
- Dunford, N. and J.T. Schwartz (1957), *Linear Operators, Part I: General Theory*, New York: Wiley Interscience.
- Harvey, A.C. (1981), *Time Series Models*, Oxford: Philip Allen.
- Ibragimov, I.A. and Y.V. Linnik (1971), *Independent and Stationary Sequences of Random Variables*, Groningen: Wolters-Noordhoff.
- Iosifescu, M. and R. Theodorescu (1969), *Random Processes and Learning*, New York: Springer Verlag.
- McLeish, D.L. (1974), "Dependent Central Limit Theorems and Invariance Principles", *Annals of Probability* 2, 620-628.
- McLeish, D.L. (1975), "A Maximal Inequality and Dependent Strong Laws", *Annals of Probability* 3, 829-839.
- Rosenblatt, M. (1956), "A Central Limit Theorem and a Strong Mixing Condition", *Proceedings of the National Academy of Science, USA* 42, 43-47.
- Serfling, R.J. (1968), "Contributions to Central Limit Theory for Dependent Variables", *Annals of Mathematical Statistics* 39, 1158-1175.

1988-1	H. Visser	Austrian thinking on international economics
1988-2	A.H.Q.M. Merckies T. van der Meer	Theoretical foundations for the 3-C model
1988-3	H.J. Bierens J. Hartog	Nonlinear regression with discrete explanatory variables, with an application to the earnings function
1988-4	N.M.van Dijk	On Jackson's product form with 'jump-over' blocking
1988-5	N.M.van Dijk M.Rumsewicz	Networks of queues with service anticipating routing
1988-6	H.Linneman C.P.van Beers	Commodity Composition of Trade in Manufactures and South-South Trade Potential
1988-7	N.M.van Dijk	A LCFS finite buffer model with batch input and non-exponential services
1988-8	J.C.W.van Ommeren	Simple approximations for the batch-arrival $M^X/G/1$ queue
1988-9	H.C.Tijms	Algorithms and approximations for batch-arrival queues
1988-10	J.P.de Groot H.Clemens	Export Agriculture and Labour Market in Nicaragua
1988-11	H.Verbruggen J.Wuijts	Patterns of South-South trade in manufactures
1988-12	H.C.Tijms J.C.W.van Ommeren	Asymptotic analysis for buffer behaviour in communication systems
1988-13	N.M.van Dijk E.Smeitink	A non-exponential queueing system with batch servicing
1988-14	J.Rouwendal	Existence and uniqueness of stochastic price equilibria in heterogeneous markets
1988-15	H.Verbruggen	GSTP, the structure of protection and South-South trade in manufactures
1988-16	Mevr.H.Weijland Mevr.R Herweijer J.de Groot	Female participation in agriculture in the Dominican Republic
1988-17	N.M.van Dijk	Product Forms for Random Access Schemes
1988-18	A.H.Q.M.Merkies I.J.Steyn	Adaptive Forecasting with Hyperfilters
1988-19	J.Rouwendal	Specification and Estimation of a Logit Model for Housing Choice in the Netherlands
1988-20	J.C.W.van Ommeren	An elementary proof of a basic result for the