Stockis, Jean-Pierre (1997) *On some problems related to machine-generated noise.* Doctor of Philosophy (PhD) thesis, University of Kent.

ON SOME PROBLEMS RELATED TO MACHINE-GENERATED NOISE

A thesis submitted to
the University of Kent at Canterbury
in the subject of Statistics
for the degree
of Doctor of Philosophy.

By

Jean-Pierre Stockis

October 1997

Contents

Abstract iv

Acknowledgements v

1 Introduction 1

2 Attractors under dynamic noise 7

2.1 Attractors 7

2.2 Theorems on the stability of attractors under dynamic noise . . . 15

2.3 Comments and examples 23

2.4 Extensions towards state-dependent noise and randomly varying parameters 29

3 Chaotic sequences: introduction and some results 36

3.1 Lyapunov exponents 36

3.2 Chaotic sequences: introduction and motivations 39

3.3 Properties derived from the ergodic theorem 50

3.4 Lyapunov exponents of chaos driven systems 58

4 Statistical inference on chaos driven AR models 63

4.1 Asymptotic bias of the Yule-Walker estimator for an AR(1) 63

4.2 Asymptotic normality of the Yule-Walker estimator for an AR(1) 65

4.3 Simulations 74

4.4 Extensions of the results to AR(p) 92

4.5 Simulations 99
4.6 Some comments on noisy chaos driven AR models 139

5 Statistical inference on chaos driven linear stochastic regression models **143**

5.1 The framework 144
5.2 Consistency and asymptotic variance of the linear regression estimator 147
5.3 Asymptotic normality of the linear regression estimator 150
5.4 Simulations 154
5.5 Towards a central limit theorem for general chaotic sequences . . 173

6 Conclusion **178**

References **181**

Abstract

Computer calculations do not exactly follow classical theoretical models: it is enough to think of rounding errors or of pseudo-random number generators, typically chaotic maps, simulating *iid* noise. The thesis aims to look at their impacts on statistical inference.

We prove that the attractors of dynamical systems are stable under some kind of infinitesimal random perturbation which is a good approximation to the rounding errors. Concerning the autoregressive models, we have obtained the asymptotic bias and the limiting distribution for the Yule-Walker estimator of the autoregressive parameter under considerably weaker assumption than that of independence in the noise sequence. In the same way, we have proved consistency and asymptotic normality of the linear regression estimator for quite general chaos driven linear stochastic regression models.

In particular, these suggest robustness of the corresponding classical asymptotic results and throw some light on the use of simulations in verifying these results.

Acknowledgements

I thank Professor H. Tong for advising and guiding my study. He has introduced me into the world of non-linearity. He has provided me with many good ideas and suggestions and has given me much support. I am greatly indebted to him.

I thank Doctor B. Cheng for his advice. His useful instructions helped me to enter well into the thesis.

I thank Professor K. S. Chan (University of Iowa) and Doctor Q. Yao for many valuable suggestions. I thank various members of the Department of Statistics of the University of Kent at Canterbury (UKC) for their help and concern, especially Doctor E. E. Bassett and Doctor C. G. H. Diks.

I thank the secretaries of the Institute of Mathematics and Statistics of UKC for their commitment. I thank the University of Kent at Canterbury for allowing me free access to very good computer and library facilities.

I thank the Engineering and Physical Sciences Research Council (EPSRC) for their financial support.

I thank Mrs. Sullivan for allowing me to live in her fine house and for offering me confidence and freedom.

I thank my parents for their wonderful support throughout my years of study.

Chapter 1

Introduction

In numerous fields, observations of a phenomenon are made sequentially over time; these sequences of observations are called time series. In fact, the use of time series goes far into the past and can be said to have its origin at the invention of the writing: in each civilisation, almost simultaneously with writing appear numbers, dating and calendars. And these days, there is a multitude of time series to be found in the fields of economics, engineering, science or sociology, to give a few examples.

From these observations the aim is to construct a model which captures both important features of the data and the available underlying theory. The modelling, 'as much an art as a science' (Tong (1990)), is not going to be part as such of the thesis; important references are Box and Jenkins (1970), with the formulation of an iterative procedure consisting of the stages of identification, estimation and diagnostic checking, and Akaike (1985). We shall concentrate on discrete time series models.

Now, we can distinguish between two types of models. First, we have the so-called deterministic models, based on non-linear dynamics: the system is defined by a known (non-linear) map $f : M \to M : x_{t+1} = f(x_t)$. In this case, there is a perfect knowledge of the dynamic laws governing the system and randomness occurs just in the choice of the initial conditions. Note that, as t tends to ∞, the system may not converge or if it does, it converges to a set called attractor, notion which will be dealt with in the next chapter. Now, without entering into the details at this stage, some of these deterministic models exhibit an irregular and seemingly random behaviour. This so-called chaotic dynamics has been known since the beginning of this century (Hadamard (1898) and Poincaré (1905)) and is closely related to the notion of 'sensitive dependence on the initial condition'. Roughly speaking, this means that, starting from any typical value in a neighbourhood of any initial condition x_0, the orbit diverges from that starting from x_0. So, even the tiniest of errors in the initial condition can lead to huge errors in (long-term) prediction (see, for example, Berliner (1991)). Within the last decade, there has been a fast growing interest in this chaotic dynamics (see, for example, Ott (1993) for a broad coverage) and some important applications have been established, in particular in engineering and in physics.

Besides the deterministic models, there are the so-called stochastic models, well known to the statisticians, which incorporate randomness at each time-step. We can mention the autoregressive (AR) models (Yule (1927)) of the form: $X_t = \sum_{j=1}^{k} \alpha_j X_{t-j} + \varepsilon_t$, where $k \geq 1$ is called the order of the AR model and $\{\varepsilon_t\}$ will be in this thesis a sequence of independent and identically distributed random variables with mean 0 and positive variance $\sigma^2 < \infty$ (that is $\varepsilon_t \sim iid(0, \sigma^2)$). This class of models can be enlarged to the class of autoregressive/moving average (ARMA) models (for example, Brockwell and Davis (1989) gives a systematic study of linear time series models in both time domain and frequency domain); when linearity fails, non-linear models are used (see, for example, Tong (1990)). Of course, stochastic models also apply in the context of regression.

Now, an important fact to note is that links and interactions exist between deterministic and stochastic models (Tong (1990) and especially (1995)) and that each subject can benefit from the other.

The main aim of this thesis is to show the role of noise (and more particularly of machine-generated noise) in linking the two kinds of models and to estimate the effect of this interaction on the theory.

On the one hand, observations practically never evolve according to a deterministic (noise-free) system; noise is unavoidable in the real world (Tong (1990 and 1995), Takens (1994a)). We can mention, for example, the impossibility of absolute accuracy in the measurements leading to the existence of measurement errors. A stochastic environment is also required if the true model is unknown to us and our deteministic model is just an approximation if there exist unpredictable environmental changes or unknown external interactions. Finally, it is worthwhile for us to notice that computers commit round-off errors and so any orbit obtained from computer calculations is subject to noise.

So, the deterministic model $x_{t+1} = f(x_t)$ should be replaced by the stochastic model $X_{t+1} = g(X_t, \varepsilon_t)$ in the practical applications. Now, what does the noise-free attractor become under the effect of noise? In particular, can we rely on the computers and think that the attractors on the computer screen are a good approximation to the noise-free (true) attractors? These questions are of course of special interest in the case of chaotic attractors. Ruelle (1981) and Kifer (1988) have analysed the related problem of random perturbations of dynamical systems from a mathematical point of view. We study the effect of different kinds of noise on the noise-free attractors from a statistical point of view which is conceptually simpler and allows a more general approach.

On the other hand, deterministic attractors bring a profound contribution to stochastic models in the following way: the pseudo-random numbers used by the computers to simulate sequences $\{\varepsilon_t\}$ of *iid* random variables are generated by purely deterministic dynamical systems (see, for example, Knuth (1981)). So, in simulations, the sequence $\{\varepsilon_t\}$ is replaced by $\{E_t\}$ which is a sequence of identically distributed random variables generated by a deterministic map f such that $E_{t+1} = f(E_t)$ and each E_t has the invariant distribution associated with f as its marginal distribution. Typically, the map f is chaotic and so the sequence $\{E_t\}$ is concentrated on chaotic attractors. A chaotic sequence can be treated as a time series in its own right (see, for example, Hall and Wolff (1995b)) and a chapter of this thesis focuses on properties of chaotic sequences.

Now, Tong (1995) has commented that simulation studies in the statistical literature typically *assume* albeit implicitly that central limit properties existing in the case of a model driven by *iid* noise continue to hold even when we replace $\{\varepsilon_t\}$ by pseudo-random numbers (that is a chaotic sequence $\{E_t\}$). It is therefore practically relevant to enquire why the simulated results seem to support the conclusion of central limit properties even though the assumption of independence no longer holds in the chaos driven model. We note that dynamicists have recently also shown considerable interests in the chaos driven models for their own sake (see, for example, Takens (1994b)). In this thesis, we focus on the statistical inference aspect of chaos driven models. In the statistical literature, Lawrance (1992) and Lawrance and Spencer (1997) have studied the related but different problem of connecting chaotic maps in reversed time with multiplicative congruential random number generators.

Consider an AR(k) model:

$$X_t = \sum_{j=1}^{k} \alpha_j X_{t-j} + \varepsilon_t$$

where $\alpha_k \neq 0$ and $\varepsilon_t \sim iid(0, \sigma^2)$. Suppose that it is causal. It is then well-known that central limit theorems are available for the Yule-Walker estimator $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha}_k)'$ of $\alpha = (\alpha_1, ..., \alpha k)'$. Let Γ_k denote the covariance matrix $[\gamma(i-j)]_{i,j=1}^{k}$, $\hat{R}_k$ denote the sample autocorrelation matrix $[\hat{\rho}(i-j)]_{i,j=1}^{k}$ and $\hat{\rho}_k = (\hat{\rho}(1), ..., \hat{\rho}(k))'$. Then $\hat{\alpha} = \hat{R}_k^{-1}\hat{\rho}_k$ and $n^{1/2}(\hat{\alpha} - \alpha) \rightarrow^d \mathcal{N}(0, \sigma^2\Gamma_k^{-1})$, where $\rightarrow^d$ denotes the convergence in distribution, or equivalently $\hat{\alpha} \doteq \mathcal{N}(\alpha, \sigma^2\Gamma_k^{-1}/n)$ for large sample size n.

Now, the thesis will deal with causal chaos driven AR(k) models of the form:

$$X_t = \sum_{j=1}^{k} \alpha_j X_{t-j} + E_t$$

and we shall consider the same Yule-Walker estimators because we want to know if the asymptotic properties of an AR(k) model with *iid* noise sequence are preserved when the *iid* noise sequence is replaced by a purely deterministic chaotic system. Note that the results and simulations relative to the AR(1) case closely follow Stockis and Tong (1996, submitted to the Journal of the Royal Statistical Society, Series B).

Consider the multiple linear stochastic regression model

$$Y_j = \beta_1 X_{j1} + ... + \beta_p X_{jp} + \varepsilon_j, \quad j = 1, 2, ..., n,$$

where $X_{j1}, ..., X_{jp}$ are random variables and the ε_js form a martingale difference sequence. Let $X = (X_{il})_{1 \leq i \leq n,\ 1 \leq l \leq p}$ and $Y = (Y_1, ..., Y_n)'$; $X'X$ is assumed to be non-singular. Then Lai and Wei (1982) show that the linear regression estimate $\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_p)' = (X'X)^{-1}X'Y$ is consistent and asymptotically normal under some weak additional assumptions. Now, we shall deal with chaos driven linear stochastic regression models; are the properties of consistency and asymptotical normality still present for the linear regression estimate if the noise sequence is a chaotic (purely deterministic) sequence?

The outline of the thesis is as follows. In chapter 2, after introducing precisely the notion of attractor in the same way as in Ruelle (1981), we have a close look at the effects of different kinds of dynamic noise on the dynamical systems and more particularly on their attractors. Chapter 3 introduces the notion of chaotic sequence, a notion which will be extensively used throughout the thesis. The degree of stochasticity of chaotic sequences is examined and their practical importance is highlighted. Finally, some useful properties of chaotic sequences are derived. Chapter 4 deals with the asymptotic properties of the classical Yule-Walker estimator of the parameters for autoregressive models driven by chaotic sequences. The last section of this chapter says a few words on AR models driven by noisy chaos. In chapter 5, we focus on chaos driven linear stochastic regression models. After making four assumptions on the chaotic sequence and on the explanatory variables, we analyse some asymptotic properties of the linear regression estimator: its consistency, its asymptotic variance and its asymptotic normality. The last section of chapter 5 indicates a possible direction for future research. Finally, chapter 6 summarizes our results and discusses future directions. Note that throughout the thesis numerous simulations illustrate the theoretical results; all the simulations use as basic elements the well-known modulo map and logistic map which are introduced in section 2.1.

Chapter 2

Attractors under dynamic noise

The structure of the chapter is as follows. Section 2.1 introduces a precise definition of attractors due to Ruelle (1981) and illustrates it by means of the well-known logistic map. Section 2.2 first describes some kinds of noise; we then focus on systems submitted to small independent absolutely continuous dynamic noise and more particularly on their attractors. We prove that, given the noise level, it is possible to construct the attractors of these perturbed systems; moreover, it is shown that the attractors of noise-free systems are stable under such infinitesimal random perturbations. Section 2.3 makes some important comments on the results of section 2.2 and illustrates them. Finally, section 2.4 deals with the case of systems perturbed by state-dependent dynamic noise. While it is possible to adapt some results of section 2.2 to this framework, the stability of the noise-free attractors under this kind of noise is not guaranteed.

2.1 Attractors

This chapter deals with non-linear dynamics: generally, a non-linear dynamical system is defined either by an evolution equation of the form $dx_t/dt = F(x_t)$ (continuous-time case) or by a mapping $x_{t+1} = f(x_t)$, where $f : M \to M$ is a non-linear deterministic map. There exist many well-known examples of continuous-time non-linear dynamics: let us just mention the Navier-Stokes equation (time evolution of a hydrodynamic system) or the Rössler system (stirred chemical reaction); see Ruelle (1989) or Diks (1996) for more details.

Now, this thesis will restrict itself to the discrete-time case; the main reason for this is that we shall be interested in the statistical analysis of digitised data. Here are two examples of discrete dynamical systems; both of them will be used extensively throughout the thesis.

- The so-called modulo map $f : [0,1] \to [0,1] : x_{t+1} = 2x_t$ modulo 1, that is

  $$f : [0,1] \to [0,1] : x \to f(x) = \begin{cases} 2x, & \text{if } x \in [0,1/2] \\ 2x-1, & \text{if } x \in (1/2,1] \end{cases}$$

  This simple map is illustrated in Figure 2.1 (a). A way of viewing this map is to write the initial condition x_0 in a binary expansion: $x_0 = 0.a_0a_1a_2a_3... = \sum_{j=0}^{\infty} 2^{-j-1}a_j$, where each a_j is either 0 or 1. Then, $x_1 = f(x_0) = 0.a_1a_2a_3...$, $x_2 = f(x_1) = 0.a_2a_3a_4...$ and so on. Thus, the orbits starting from two initial conditions equal up to the 50th digit (that is a change of x_0 by 2^{-50}) will diverge more and more and eventually, at time 51, there will be an important change between them. The modulo map exhibits sensitive dependence on the initial condition and so is chaotic (see chapter 1).

- the famous logistic map (May (1976), inspirated from animal population dynamics) $f : [0,1] \to [0,1] : x_{t+1} = \theta x_t(1-x_t), \quad 0 \le \theta \le 4$.

  Figure 2.1 (b) shows this map for different values of the parameter θ. We can notice that the map becomes more concave as θ increases. Note that for $0 \le \theta \le 4$, if x_t belongs to $[0,1]$, then so does x_{t+1}; now, if the domain of definition of the logistic map is extended to the whole real line, then an initial condition outside $[0,1]$ implies that x_t will move off to $-\infty$.

Now, this chapter is centered on the behaviour of discrete-time dynamical

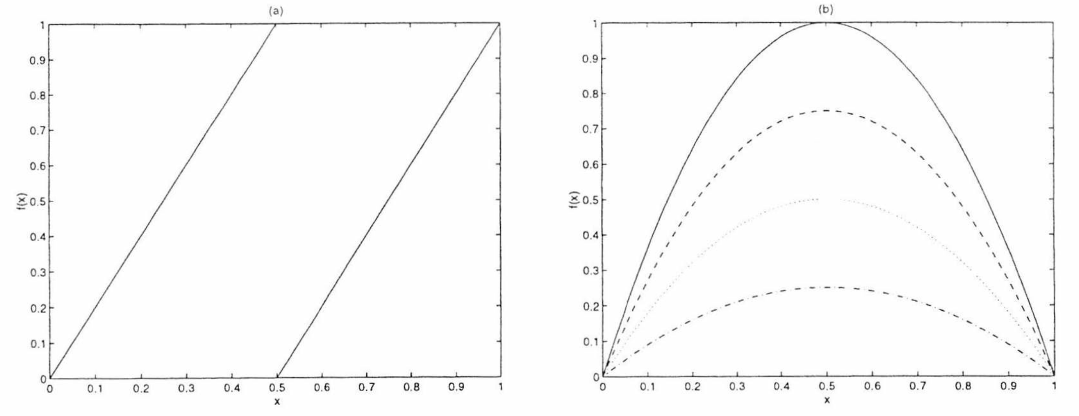


Figure 2.1: (a) The modulo map; (b) The logistic map with $\theta = 1.00$ $(- \cdot -\cdot)$, $\theta = 2.00$ $(\cdots)$, $\theta = 3.00$ $(- - -)$, $\theta = 4.00$ (solid line).

systems as $t \to \infty$. For linear dynamical systems, there are generically speaking two possibilities:

- the system does not converge, that is $|x_t| \to \infty$ as $t \to \infty$,
- the system converges to a limit point.

For non-linear dynamical systems, in addition of these two cases, there exist basically three more:

- the system converges to a limit cycle of fixed finite period,
- the system is of the form $x_t = f(\omega_1 t, \omega_2 t, ..., \omega_k t)$, where f is a periodic function of period 2π in each of its arguments and ω_1, ω_2, ..., ω_k are rationally independent frequencies; the motion then constitutes a quasi-periodic attractor,
- the system is chaotic (see chapter 1) and it converges to a non-empty closed set. This set is then called chaos or chaotic attractor. Generically, if a system converges to a closed set which is not a limit point, a limit cycle or a quasi-periodic attractor, then the system is chaotic and therefore the attracting set is called chaos. Now, there is a formal way to quantify sensitive dependence

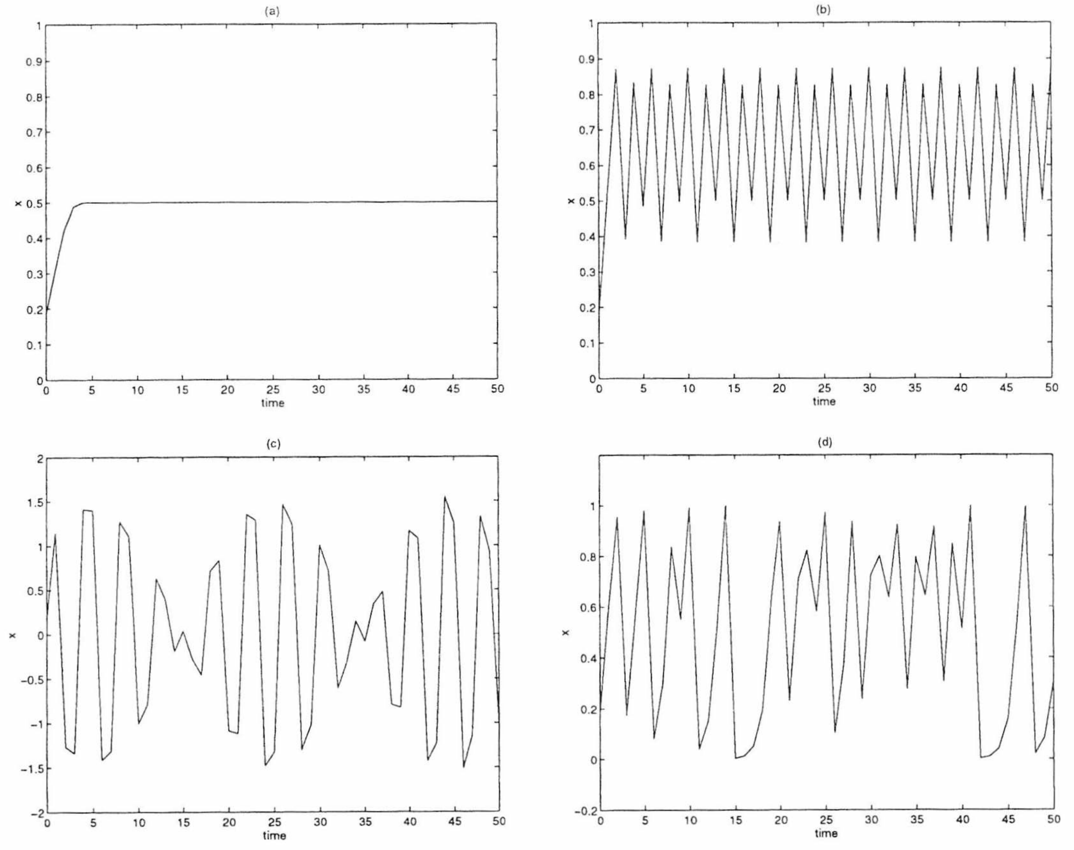


Figure 2.2: Time series, starting from randomly selected initial values between 0 and 1, generated by (a) the logistic map with $\theta = 2.00$, (b) the logistic map with $\theta = 3.50$, (c) $x_t = cos(\sqrt{2}\ t) + sin(\sqrt{3}\ t)$, (d) the logistic map with $\theta = 4.00$.

on initial conditions and so to detect chaotic dynamics, by means of the so-called Lyapunov exponents. This will be dealt with in the next chapter.

Figure 2.2 illustrates the four possible types of convergence for a time series, namely to a limit point (0.5 in this case), a limit cycle (of period 4 here), a quasi-periodic attractor and a chaos. For each of the four maps, the initial condition has been selected at random between 0 and 1.

As a verification, the logistic map with parameter $\theta = 4.00$ exhibits well sensitive dependence on the initial condition: Table 1.1 illustrates it for two typical very close initial conditions.

Now, as already hinted by Figure 2.2, the logistic map $f : [0, 1] \to [0, 1] : x \to$

t=0	0.1	0.1000001
1	0.36	0.36000032
2	0.9216	0.92160036
10	0.14783656	0.14771543
15	0.00393603	0.00328172
20	0.82001387	0.93222609
23	0.12638436	0.73905231

Table 1.1: Time series generated by the logistic map with $\theta = 4.00$ for initial values 0.1 and 0.1000001.

$\theta x(1-x)$ goes through a range of different asymptotic behaviors as θ varies from 0 to 4. When $\theta < 3$, it has an attracting fixed point, which becomes unstable at $\theta = 3$; there is then an attracting limit cycle of period 2^n with n tending to ∞ as θ tends to $3.57... = r_\infty$. At $\theta = r_\infty$, the so-called Feigenbaum Cantor set is produced. For $r_\infty < \theta \leq 4$, the attractor is chaotic, except in narrow windows throughout the chaotic range, where the attracting orbit is periodic (see, for example, Ruelle (1989) and especially Ott (1993) for more detais).

Figure 2.3 shows the attractors for the range $\theta \in [3.5, 4]$, with increase jumps of 0.01 for θ, and then for the subrange $[3.82, 3.83]$ with smaller increase jumps for θ. These figure suggest that the attractors are "often" chaotic in the range $[3.58, 4.00]$; in the same time, we should keep in mind the presence of small windows of periodic orbits throughout the range. Note the increase of the support of the chaotic attractor as θ increases: it can easily be seen that the maximum of the support is given by $f(0.5)$ and the minimum of the support by $f^2(0.5)$.

Now, depending on the initial condition, the system may converge to different asymptotic sets. For example, for a chaotic logistic map with parameter θ, besides the chaotic motion obtained with probability 1 for uniformly distributed initial conditions included in $[0, 1]$, there are two fixed points, 0 and $(\theta - 1)/\theta$, and an infinite number of (unstable) periodic orbits (see, for example, Ott (1993) stating Sarkovskii's theorem). Therefore, the supports of the chaotic motions are in fact discontinuous.

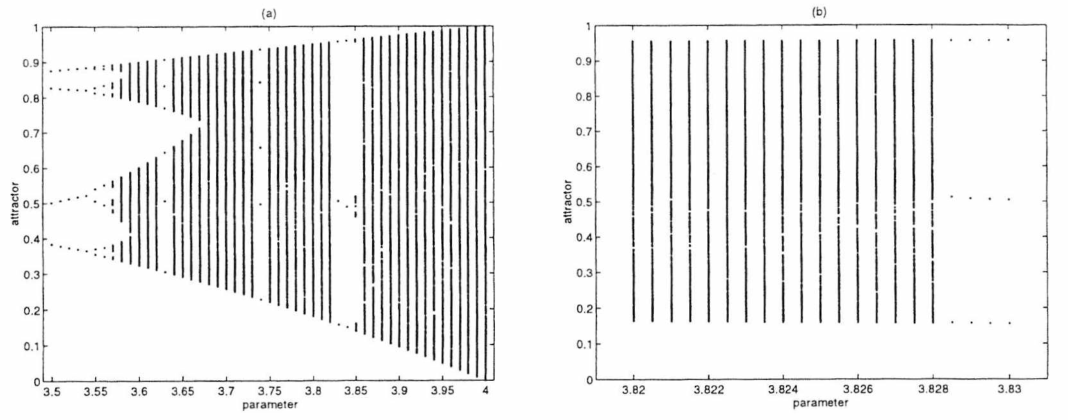


Figure 2.3: Attractors for the logistic map with, (a) $\theta = 3.5, 3.51, ..., 4.00$, (b) $\theta = 3.82, 3.8205, ..., 3.83$.
For each θ, the time series starts at random between 0 and 1. the map is iterated 1000 times without plotting anything (in view of discarding transient states); the next 1000 iterations are plotted on the graph.

So, it is quite clear that we need a precise definition of what we shall actually call an attractor. In fact, Ruelle (1981) has introduced two distinct notions, that is attracting set and attractor, which both include invariance and attractivity (but in different meanings). We shall say a few words on attracting sets but we shall mainly focus on attractors.

Definition 2.1 *Let $M \equiv (M, \mathcal{T})$ be a Hausdorff topological space. Let $f : M \to M$ be a map defining a dynamical system; the $f^t : M \to M$ are assumed to be continuous.*
Then, a set $B \subset M$ is said to be an **attracting set** *if the following conditions hold.*

1. *B is closed.*

2. *B has a neighbourhood U such that, for every neighbourhood V of B, we have $f^tU \subset V$ when t is large enough.*

3. *$f^tB = B$ for all t.*

Remark: Given the two first conditions, there are equivalent conditions to the last one. See Ruelle (1981) for the details.

Now, the notion of attracting set regroups the ideas of closedness, invariance, attractivity and open basin of attraction. (The basin of attraction of a set denotes the totality of initial states which iterate to this set.) The last idea is implicit in the definition but the basin of attraction of each attracting set has to be open (see my M.Sc. dissertation for a formal proof). This requirement of an open basin of attraction is not met by some sets, for example by the above-mentioned Feigenbaum Cantor set, which are usually seen as "attractors" (see, e.g., Ruelle (1981)). So, we need to concentrate on a mathematically weaker notion, namely the notion of attractor.

For the rest of the section, we assume that we work on a Hausdorff topological metric space M; the induced distance is denoted "*dist*". Moreover, the maps $f^t : M \to M$ are assumed to be continuous, for all $t \in Z_+$.

Definition 2.2 *A curve, i.e. a family* $(x_t)_{t=t_0,t_0+1,...,t_1-1,t_1}$ *of points of* M*, is called an* **ε-pseudoorbit** *if* $dist(fx_t, x_{t+1}) \leq \varepsilon \quad$ *for* $t = t_0, t_0+1, ..., t_1 - 1$.

Remarks: (i) Ruelle (1981) uses a more general definition including time continuous case. (ii) Note that the fact $dist(x, y) \leq \varepsilon$ does not imply that there is an ε-pseudoorbit going from x to y.

Definition 2.3 *We write* $\boldsymbol{a \succ b}$ *("a goes to b") if, for arbitrarily small* $\varepsilon > 0$*, there is an* ε*-pseudoorbit going from* a *to* b*.*

The relation $\succ$ is obviously a preorder, that is reflexive and transitive. (It is not an order since $(a \succ b$ and $b \succ a) \not\Rightarrow (a = b)$).) Since $\succ$ is a preorder, the relation $\sim$ ($a \sim b$ if $a \succ b$ and $b \succ a$) is an equivalence relation (i.e. reflexive, transitive, symmetric) and we can introduce equivalence classes $[a]$. Moreover, the preorder $\succ$ induces an order on the equivalence classes: $[a] \geq [b]$ if $a \succ b$, $[a] = [b]$ if $a \sim b$.

Now, here is the definition of an attractor.

Definition 2.4 *An* **attractor** *is a minimal equivalence class, that is a class* $[b]$ *such that* $\nexists a : [a] < [b]$ *(or, equivalently,* $\nexists a : b \succ a$ *and* $a \not\succ b$*).*

We can have a look at the basic properties of an attractor.
An attractor is:

- closed. This is an easy corollary of the fact that the relation $\succ$ is closed (i.e. if x_n and y_n tend to a and b respectively and if $x_n \succ y_n$, $\forall n$, then $a \succ b$).
- invariant. There is no orbit starting from an attractor which goes outside this attractor.
- attracting. An attractor is a limit set of (slightly) *perturbed* trajectories.
- irreducible. Each point in the attractor can be reached from any other point in the attractor by means of a slightly perturbed trajectory.
- stable under infinitesimal random perturbations. See sections 2.2 and 2.3.

Now, let us illustrate this concept by means of the logistic map. The space $[0, 1]$ is provided with the Euclidean topology; in particular, $dist(a, b) = |b - a|$.

We start with $f : [0, 1] \to [0, 1] : x \to 2x(1 - x)$ (see Figures 2.1 (b) and 2.2). There are two fixed points, $\{0\}$ and $\{1/2\}$, with respective basins of iteration $\{0\} \cup \{1\}$ and $]0, 1[$.
The class $[1/2] = \{1/2\}$ is an attractor: $\nexists b \neq 1/2 : 1/2 \succ b$. Now, all the points $a \in]0, 1[$ are such that $a \succ 1/2$; therefore, $\{1/2\}$ is the only attractor on $]0, 1[$.
Is $[0]$ an attractor? For arbitrarily small $\varepsilon > 0$, there exists $b_\varepsilon \in]0, 1[$ such that $|f(0) - b_\varepsilon| = b_\varepsilon - 0 \leq \varepsilon$. For example, we can take $b_\varepsilon = \varepsilon$. So, for arbitrarily small $\varepsilon > 0$, there exists an ε-pseudoorbit going from 0 to 1/2; $[0]$ is therefore not an attractor.
The only attractor (on $[0, 1]$) is $\{1/2\}$.

In the same way, the map $f : [0,1] \to [0,1] : x \to 3.5x(1-x)$ has one attractor, namely $[0.500884] = \{0.500884, 0.874997, 0.382820, 0.826941\}$.

Now, what is the attractor for the chaotic map $f : [0,1] \to [0,1] : x \to 4x(1-x)$?

Let C denote the chaotic motion; it is dense in $[0,1]$. Let $c \in C$, what is $[c]$? $[c]$ obviously contains all the points of the chaotic motion. Moreover, $[c]$ contains all the unstable fixed points or unstable periodic orbits arbitrarily close to the chaotic motion. For: consider a limit cycle (or a limit point) P and $p \in P$. Let ε be any number > 0; then, at some stage, the chaotic motion will come closer to this limit cycle than ε since the chaotic motion is dense in $[0,1]$. So, $c \succ p$ and $p \succ c$.

Now, the chaotic motion is dense in $[0,1]$. So, $[0,1]$ is the chaotic attractor of the logistic map with $\theta = 4.00$ on $[0,1]$.

So, by definition of an attractor, a chaotic attractor does not necessarily contain only the chaotic motion but also, if they exist, the limit points and the limit cycles arbitrarily close to the chaotic motion. As a corollary, in the particular case of the logistic map, the supports of the chaotic attractors (see Figure 2.3) are continuous (unlike the supports of the chaotic motions).

Now, after clarifying the notion of attractor, we can analyse in which way they are stable under small random perturbations.

2.2 Theorems on the stability of attractors under dynamic noise

In the last section, we have defined the notion of attractor of a (noise-free) dynamical system. Now, because of the omnipresence of noise in the real world, and in particular in the computer calculations (by way of round-off errors, truncations, ...), we would like for the attractors to be stable under small random perturbations:

the motion of the noisy system should be concentrated on the noise-free attractor. Note that a randomly perturbed chaotic system is such that no long-term prediction is possible because of its dynamical unstability (sensitive dependence on the initial conditions); here, we would like for the asymptotic (statistical) properties of such a system to be preserved under small noise fluctuations.

In other terms, is an attractor a meaningful notion in practice? By extension, how is an attractor affected by noise of given finite amplitude? Can we rely on (noisy) attractors on the computer screen, are they close to the noise-free attractors?

Now, we essentially can distinguish between two types of noise, namely the additive noise and the dynamic noise.

The additive noise consists of the measurement noise: the system remains $x_{t+1} = f(x_t)$, but now instead of observing x_{t+1} as in the noise-free case, we observe Y_{t+1}, where $Y_{t+1} = x_{t+1} + \varepsilon_{t+1}$ and $\{\varepsilon_t\}$ is assumed to be a sequence of *iid* random variables. Note that additive noise does not affect the state of the system, nor the evolution of the system. We are not going to deal with additive noise here; let us just mention that there exist methods for cleaning a time series from additive noise (see Grassberger et al. (1993) for a review).

Now, the most important errors from a practical point of view come from the dynamic noise, which affects the state of the system and therefore its evolution. A system affected by dynamic noise is of the form $X_{t+1} = F(X_t, \varepsilon_{t+1})$, where the ε_t are *iid* random variables. Examples of dynamic noise are given by external interactions, round-off errors, approximations to the true system, ...

Now, there exist many ways of introducing small random perturbations into a dynamical system.

Ruelle (1981) uses the mathematical notion of diffusions with compact support. Let M be a metric space and let $f : M \to M$ be a continuous map, generating a discrete-time dynamic system (f^t). Let $\varepsilon > \delta > 0$. Then, an affine map F from the space of probability measures with compact support in M to itself is an

(ε, δ)-diffusion associated with f if the following conditions are satisfied.

1. supp. $F\delta_x \subset f\bar{B}_x(\varepsilon)$.

2. supp. $F\delta_x \supset f\bar{B}_x(\delta)$

3. If $\phi : M \to R$ is continuous, then $x \to (F\delta_x)\phi$ is continuous

   and $(F\mu)\phi = \int \mu(dx)[(F\delta_x)\phi]$. In particular,

   supp. $F\mu =$ closure $\bigcup_{x \in supp.\mu}$ supp. $F\delta_x$.

4. If $\phi : M \to R$ is continuous, the set $\{(F\delta_y)\phi : y \in \bar{B}_x(\delta)\}$ is a closed interval.

Note that supp. denotes the support of a measure or of a set, δ_x denotes the unit mass at x, $B_x(\varepsilon)$ the open ball of radius ε centered at x and $\bar{B}_x(\varepsilon)$ its closure.

Ruelle (1981) then proves the following theorem: " Let (f^t) be a discrete-time dynamical system, F an (ε, δ)-diffusion associated with f, and ν a probability measure with compact support in the basin of attraction of an attracting set Λ. We assume that Λ has a neighborhood on which f is uniformly continuous. We denote by A the union of all attractors contained in Λ, by $\bar{A}$ the closure of A, and we write $A^* = \bigcup_{z \in A}[z]$.

Then, for every neighborhood Θ of A^*, if ε is small enough,

$$\lim_{t \to \infty} F^t\nu(M \setminus \Theta) = 0."$$

So, for a discrete time dynamical system submitted to sufficiently small such diffusions, the motion is asymptotically concentrated on attractors, if some weak conditions hold.

Now, we simplify the matter: the dynamical system (f^t) is submitted to dynamical noise in the following way: $X_{t+1} = f(X_t) + \varepsilon_{t+1}$, where $\{\varepsilon_t\}$ is a sequence of *iid* bounded absolutely continuous random variables. So, we take a simple statistical point of view. In presence of noise of given finite amplitude, we shall be able to construct theoretically the noisy attractor and to compare it with the noise-free attractor. This will allow us to see that the attractors are stable under infinitesimal random perturbations provided very general conditions are satisfied.

Let us summarise our framework. As in the previous section, M is a Hausdorff topological metric space; for simplicity of notation, the induced distance $dist(a, b)$ will be denoted $|b - a|$. The maps $f^t : M \to M$ are assumed to be continuous. Moreover, M is assumed to be the largest possible domain of definition of f. Note that, if dynamic noise is introduced, it is possible in some cases for the dynamical system to leave M; we rule out this possibility (for instance, by allowing for boundary conditions).

We work on a probability space $(M, \mathcal{F}, P)$. Now, here is our first assumption.

Assumption 1 *A dynamical system submitted to small random perturbations is such that* $X_{t+1} = f(X_t) + \varepsilon_{t+1}$, $\forall t$, *where* X_t *is the random variable describing the state of the system at time* t *and* $\{\varepsilon_t\}$ *is a sequence of iid zero mean, bounded, absolutely continuous random variables with symmetric support such that* $P(|X_{t+1} - f(X_t| \leq \varepsilon) = 1$ *and* ε *is the lower bound such that the above probability is equal to* 1.

In particular, if $P(X_t = x) = 1$ at time t, then Assumption 1 means that the probability to lie in $B_{f(x)}(\varepsilon)$ (or in $\bar{B}_{f(x)}(\varepsilon)$) at time $t + 1$ is equal to 1. Note that for $M = R$, the support of the random variables ε_t is simply $[-\varepsilon, \varepsilon]$.

On the other hand, as noticed in the previous section, it is possible for a dynamical system not to converge (that is, there is no attractor). Now, for the rest of the section, we would like that each system has at least one attractor. More particularly, we would like that for all $a \in M$, there exists an attractor $[b]$ such that $a \succ b$. To ensure this, we make the following assumption.

Assumption 2 *M is compactified. That is, if M is compact, we do not change anything; otherwise, we compactify M, i.e. we identify M (as a topological space) to a dense subspace of a compact space.*

We refer to any textbook in differential geometry for more details. In particular, if $M = R$, we can compactify R in the following way: we construct the space $\bar{R} = R \cup \{-\infty, \infty\}$ with, as base of topology, the Euclidean open sets, i.e. the sets $]-\infty, N[,]N, \infty[$ $(N \in R)$.

Now, given the noise level ε, what is the attractor of the noisy system? Let us introduce the relation (the beginning of the reasoning follows my M.Sc. dissertation (1994)).

Definition 2.5 $\boldsymbol{a} \succ_\varepsilon \boldsymbol{b}$ *means: there is an ε-pseudoorbit (see Definition 2.2) going from a to b.*

The relation $\succ_\varepsilon$ is a preorder since the relation $\succ_\varepsilon$ is reflexive ($a \succ_\varepsilon a$) and transitive (if there is an ε-pseudoorbit $(a, a_1, ..., a_n, b)$ going from a to b and if there is an ε-pseudoorbit $(b, b_1, ..., b_m, c)$ going from b to c, then $a \succ_\varepsilon c$ since $(a, a_1, a_n, b, b_1, ..., b_m, c)$ is an ε-pseudoorbit).

Now, the same way of reasoning applies for the relation $\succ_\varepsilon$ as for the relation $\succ$ (see last section). In particular, an equivalence class for the relation $\succ_\varepsilon$ will be denoted $[a]_\varepsilon$. Note that, by definition, the set $[a]_\varepsilon$ contains $[a]$ since $a \sim b \Rightarrow a \sim_\varepsilon b$.

Definition 2.6 *A minimal equivalence class $[b]_\varepsilon$ for the relation $\succ_\varepsilon$ is called an* **attractor for the relation $\succ_\varepsilon$.**

We can notice that each equivalence class for the relation $\succ_\varepsilon$ is closed; therefore, each attractor for the relation $\succ_\varepsilon$ (also called noisy attractor) is closed.

Now, for all t, let D_t be the following set of events, $D_t = \{\omega : |f(X_{t-1}(\omega)) - X_{t-1}(\omega)| < \varepsilon\}$, and let $D = \bigcap_{t=1}^{\infty} D_t$.

Lemma 2.1 *Under Assumption 1, $P(D) = 1$.*

Proof: for each t, $P(D_t) = 1$ by Assumption 1, therefore $P(M \setminus D_t) = P(D_t') = 0$.

Now, $D = \bigcap_t D_t$ implies $D' = \bigcup_t D'_t$, and, by Boole's inequality,

$$P(D') = P(\bigcup_t D'_t) \leq \sum_t P(D'_t).$$

So, $P(D') = 0$ and $P(D) = 1$. □

At this stage, we recall the notions of transience and essentiality (in the context of a Markov process); note that the sequence $\{X_t\}$ as defined in Assumption 1 is obviously a Markov process.

Definition 2.7 *A set is S* **transient** *if the probability that there exists an infinity of i's such that X_i belongs to S is equal to* 0.

Note that a set which is not transient is called recurrent.

Definition 2.8 *A set S is* **essential** *if for all sets T such that $\exists n : P(X_{t+n} \subset T | X_t \subset S) > 0$, $\exists m : P(X_{t+m} \subset S | X_t \subset T) > 0$.*

It is well-known that any recurrent set is essential. Lemma 2.2 follows directly.

Lemma 2.2 *Any inessential set is transient.*

The next lemma plays an important role in the proof of the theorems.

Lemma 2.3 *Let Assumptions 1 and 2 apply. Let $f : M \to M$ be continuous. Then, if a does not belong to an attractor for the relation $\succ_\varepsilon$, then there exists $\delta(a)$ such that the subspace $B_a(\delta)$ is transient with probability* 1.

Proof: By Lemma 2.1, we can concentrate on sample points coming from event D.

Now, $[a]_\varepsilon$ is not an attractor; therefore, by Definition 2.6 and Assumption 2, there exists b belonging to an attractor for the relation $\succ_\varepsilon$ such that $a \succ_\varepsilon b$ and $b \not\succ_\varepsilon a$. In particular, there exists an ε-pseudoorbit going from a to b, say $(a, x_1, ..., x_{m-1}, b)$.

On the one hand, since $a \notin [b]_\varepsilon$, closed set, there exists δ^* such that $B_a(\delta^*) \cap [b]_\varepsilon = \emptyset$.

On the other hand, since f is continuous on M, there exists δ_1 such that

$$|y-a| \leq \delta_1 \Rightarrow |f(y)-f(a)| \leq \varepsilon - |x_1 - f(a)|.$$

This implies

$$\begin{aligned} |f(y)-x_1| &\leq |f(y)-f(a)| + |f(a)-x_1| \\ &\leq \varepsilon - |x_1 - f(a)| + |f(a)-x_1| \\ &\leq \varepsilon. \end{aligned}$$

In the same way, there exists δ_2 such that

$$|y-x_1| \leq \delta_2 \Rightarrow |f(y)-f(x_1)| \leq \varepsilon - |x_2 - f(x_1)|$$

and so

$$|f(y)-x_2| \leq \varepsilon, ...,$$

and finally there exists δ_{m+1} such that

$$|y-b| \leq \delta_{m+1} \Rightarrow |f(y)-f(b)| \leq \varepsilon.$$

Now, let $\delta = \min(\delta^*, \delta_1)$. We get a path

$$B_a(\delta) \to B_{x_1}(\delta_2) \to ... \to B_{x_{m-1}}(\delta_m) \to B_b(\delta_{m+1}) \to B_{f(b)}(\varepsilon)$$

such that, by Assumption 1, the probability to go from a ball to the next one in one step is strictly greater than 0.

So,

$$P(X_{m+1} \subset B_{f(b)}(\varepsilon)|X_0 \subset B_a(\delta)) > 0.$$

Now, $B_{f(b)}(\varepsilon) \subset [b]_\varepsilon$; so, since $[b]_\varepsilon$ is an attractor, we have

$$\forall t > 0, \quad P(X_t \subset B_a(\delta) | X_0 \subset [b]_\varepsilon) = 0.$$

Thus, $B_a(\delta)$ is inessential; the conclusion follows from Lemma 2.2. □

Now, given the noise level ε, the following theorem provides us with the attractor(s) of the noisy system.

Theorem 2.1 *Let Assumptions 1 and 2 apply. Let f be continuous.*
Let A denote the union of all the attractors for the relation $\succ$.
Let L_ε denote the union of the attractors for the relation $\succ$, which do not belong to attractors for the relation $\succ_\varepsilon$.
Let $E_\varepsilon = A \setminus L_\varepsilon$.
Then, as t tends to ∞, the (noisy) system will tend to $A_\varepsilon = \bigcup_{z \in E_\varepsilon} [z]_\varepsilon$ with probability 1.

Proof: Let $a \in M$. If $[a]_\varepsilon$ is not an attractor, then, by Lemma 2.3, there exists δ such that $B_a(\delta)$ is transient with probability 1. So, as $t \to \infty$, the system converges to the attractor(s) for the relation $\succ_\varepsilon$ (note that we are sure of the existence of such attractor(s) because of Assumption 2).

Now, each attractor for $\succ_\varepsilon$ contains at least one "true" attractor (that is, for $\succ$). For: let us suppose it is not true: $[d]_\varepsilon$ is a minimal equivalence class for $\succ_\varepsilon$ and there is no element from an attractor for $\succ$ in $[d]_\varepsilon$. We know, however, that $d \succ b$ for an attractor $[b]$; so, there exists an ε-pseudoorbit going from d to b. Thus, $b \in [d]_\varepsilon$ and $[b]_\varepsilon = [d]_\varepsilon \supset [b]$.

By putting together the two parts of the proof, we get the conclusion. □

Note that L_ε can be non-empty: some attractors for $\succ$ may not be part of an attractor for $\succ_\varepsilon$. See next section for some examples.

Now, the next theorem shows that, as $\varepsilon \to 0$, the noisy attractor A_ε tends to the noise-free attractor A. In other words, a (noise-free) attractor is stable under infinitesimal random perturbation.

Theorem 2.2 *Let Assumptions 1 and 2 apply. Let f be continuous.*

Under the same notations as in Theorem 2.1,

$$\lim_{\varepsilon \to 0} A_\varepsilon = A.$$

Proof:

First, we prove that

$$\lim_{\varepsilon \to 0} A_\varepsilon \subset A.$$

If $[a]$ is not an attractor for the relation $\succ$, then, for sufficiently small ε, there exists an ε-pseudoorbit going from a to a point b but no ε-pseudoorbit going from b to a. So, for sufficiently small ε, $[a]_\varepsilon$ is not an attractor for the relation $\succ_\varepsilon$ and therefore $a \notin \lim_{\varepsilon \to 0} A_\varepsilon$ by Lemma 2.3 and by definition of A_ε.

Next, we prove that

$$A \subset \lim_{\varepsilon \to 0} A_\varepsilon.$$

Let $[b]$ be an attractor for $\succ$ (that is, $[b] \subset A$). Then, there exists ε^* such that $\forall \varepsilon \leq \varepsilon^*$, $[b]_\varepsilon$ is an attractor for $\succ_\varepsilon$. Otherwise, for arbitrarily small ε, there is an ε-pseudoorbit going outside the attractor $[b]$ which is in contradiction with the definition of an attractor (see Definition 2.2). So, $\forall \varepsilon \leq \varepsilon^*$, $[b]_\varepsilon \subset A_\varepsilon$. □

2.3 Comments and examples

First, before considering the two theorems of the previous section into more details, we need to have a close look at their set-up. We work on a space M; M is a Hausdorff topological metric space and, by Assumption 2, M is compactified. In particular, a compact space M ensures the existence of (noise-free or noisy) attractor(s). This compactification of M is easy to visualise if $M = R^n$; in this section, we shall focus on a simple compactification of R given by $\bar{R} = R \cup \{-\infty, +\infty\}$.

Now, the dynamical system $x_{t+1} = f(x_t)$ is submitted to small random perturbations and becomes the noisy system $X_{t+1} = f(X_t)+\varepsilon_{t+1}$, $t = 1, 2...$, the dynamic noise $\{\varepsilon_t\}$ being, by Assumption 1, a sequence of *iid* zero mean, bounded, absolutely continuous random variables with symmetric support of amplitude ε.
Note that this kind of noise is a good approximation to the round-off errors of the computers. The round-off errors are certainly bounded and it is reasonable to assume the other conditions. Of course, however, this is just an approximation to the round-off errors: a computer is a physical system and the concept of independence, to give an example, is unknown to him.

Finally, there is an important point to notice: we require for M, as it is the case in computing, to be the largest possible domain of definition of f in view not to constrain artificially the effect of the noise (but, of course, we rule out the possibility for the noisy system to leave M, for example by allowing for boundary conditions).

This has important consequences. As an example, consider the logistic map f with parameter $\theta = 4$. Now, M has to be the largest possible domain of definition of f, that is R. So, keeping Assumption 2 in mind, we consider $f : \bar{R} \rightarrow \bar{R} : x \rightarrow 4x(1 - x)$.

But the set $[0, 1]$ which was an attractor on the space $[0, 1]$ (see section 2.1) is not an attractor any more on $\bar{R}$ since $0 \succ -\infty$ and $-\infty \not\succ 0$. For: on the one hand, for arbitrarily small $\varepsilon > 0$, there is an ε-pseudoorbit going from 0 to $-\infty$, namely $(0 = f(0), -\varepsilon, f(-\varepsilon), f^2(-\varepsilon), ..., -\infty)$; on the other hand, $-\varepsilon$, so a fortiori $-\infty$, $\not\succ 0$ since $|f(-\varepsilon) - 0| = -4\varepsilon(1 + \varepsilon) > \varepsilon$ and f is an increasing function on $]-\infty, 0[$.
So, the only attractor for the logistic map with $\theta = 4.00$ on $\bar{R}$ is $[-\infty] = \{-\infty\}$. Therefore, the following comments apply to $[-\infty]$ and not to the set $[0, 1]$, which is not an attractor on R; or, in other words, since Theorems 2.1 and 2.2 say that the only attractor of the noisy system is $\{-\infty\}$, whatever the noise level is, it means that the set $[0, 1]$ is unstable under random perturbations.

Now, can we trust the (noisy) attractors which appear on the screen? In fact, Theorem 2.2 tells us that, provided f is continuous on M, the noise-free attractors are stable under infinitesimal random perturbations (that is, the amplitude of the noise level ε tends to 0). So, for sufficiently small ε, the noisy attractor(s) will be close to the true (noise-free) attractor(s). What does "sufficiently small" mean? Theorem 2.1 provides us with the answer. Note that, in most cases, the round-off errors are sufficiently small to ensure that the attractors on the computer screen are close to the true attractors.

For a given noise level ε, Theorem 2.1 shows that, provided f is continuous on M, the motion will asymptotically concentrate on A_ε as $t \to \infty$; if the noise-free attractors are known, it is possible to construct A_ε theoretically.

Before considering examples, a few comments are in order here.

1. As $t \to \infty$, the noisy system tends to the set of points $A_\varepsilon = \bigcup_{z \in A \setminus L_\varepsilon} [z]_\varepsilon$, where A denotes the union of all noise-free attractors and L_ε denotes the union of the noise-free attractors which are not part of noisy attractors. So, L_ε contains the true attractors which are destabilized by noise of amplitude ε. Of course, L_ε depends on the noise level; in particular, L_ε tends to an empty set as $\varepsilon \to 0$ by Theorem 2.2.

   Note that, by Assumption 2, A_ε cannot be empty and so, for any noise level ε, there will be noise-free (true) attractor(s) which are not destabilized, that is which are part of noisy attractor(s). In particular, if there is just one true attractor, it cannot be destabilized.

   So, to summarize this comment, we could say that, for a given noise level ε, some (but not all) of the noise-free attractors can be "lost".

2. Since A_ε consists on a union on z's belonging to true attractors, each noisy attractor contains inside itself a noise-free attractor ; that is, for any given noise level, there is no created "completely fictitious" noisy attractor.

3. By definition of the equivalent classes $[z]$ and $[z]_\varepsilon$, the support of each noisy attractor will be greater or equal to the support(s) of the true attractor(s)

corresponding to it. (Note that disjoint attractors can belong to a same noisy attractor for some noise levels.)

Of course, the increase level of the support depends both on the dynamical system under consideration and on the level of noise, a lower noise level leading to a better approximation to the true attractor.

Now let us illustrate the theorems by means of some examples.

First, suppose that the logistic map $f : \bar{R} \to \bar{R} : x \to 2x(1-x)$ is affected by small random uniformly distributed perturbations $U(-0.01, 0.01)$; this means that the noise level ε is equal to 0.01. Of course, f is continuous and the dynamic noise $U(-0.01, 0.01)$ satisfies the requirements of Assumption 1, so both theorems apply.

Now, the noise-free attractors are $[1/2] = \{1/2\}$ and $[-\infty] = \{-\infty\}$ (see section 2.1). Comment 2 tells us that the only possible noisy attractors are $[1/2]_{0.01}$ and $[-\infty]_{0.01} = \{-\infty\}$. Of course, $\{-\infty\}$ is a noisy attractor as it would be for any noise level but what about $[1/2]_{0.01}$? And how to construct it?

In fact, it can easily be seen that the minimum of $[1/2]_{0.01}$ is given by $2x(1-x) - 0.01 = x$, that is $x \simeq 0.489792$: there exists an 0.01-pseudoorbit going from 1/2 to 0.489792, namely $0.5 \to f(0.5) - 0.01 = 0.49 \to f(0.49) - 0.01 = 0.4898 \to f(0.4898) - 0.01 = 0.489792$.

On the other hand, $2x(1-x) + 0.01 = x$ gives $x \simeq 0.509808$ but we know that there is the following 0.01-pseudoorbit: $0.5 \to f(0.5) + 0.1 = 0.51$, so the maximum is 0.51.

Therefore, after noting that $1/2 \not\succ_{0.01} -\infty$, we can say, following Theorem 2.1, that the noisy attractors are $\{-\infty\}$ and the closed interval $[0.489792, 0.51]$ for small uniform perturbations $U(-0.01, 0.01)$.

Now, the analysis of the noisy attractors for the logistic map with $\theta = 2.00$ can be extended to uniform noise perturbations $U(-\varepsilon, \varepsilon)$, ε any number > 0. Figure 2.4 shows the theoretical noisy attractor corresponding to the noise-free

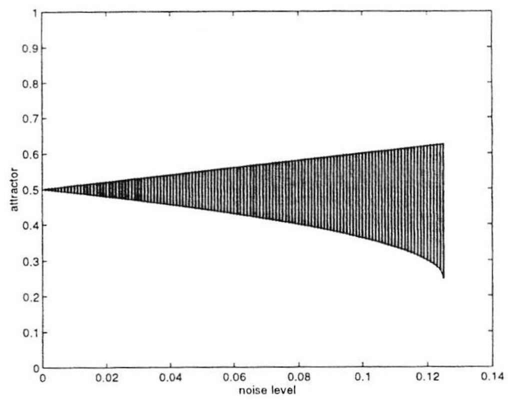


Figure 2.4: Theoretical noisy attractors corresponding to the noise-free attractor $\{1/2\}$; the dynamic noise is uniformly distributed between $-\varepsilon$ and ε. For $\varepsilon > 0.125$, the attractor is destabilized.

attracting point $\{1/2\}$ for different values of ε. If $\varepsilon > 0.125$, there is only one noisy attractor, namely $\{-\infty\}$ (the attractor $\{1/2\}$ has been destabilized, see comment 1). If $\varepsilon \leq 0.125$, in addition to $\{-\infty\}$, there is another noisy attractor, that is the closed interval $[x, 0.5+\varepsilon]$ where x is the positive root of $2x^2 - x + \varepsilon = 0$. The simulations confirmed these theoretical results. Note that, in accordance with Theorem 2.2, the interval tends to $\{1/2\}$ as $\varepsilon \to 0$.

Now, consider the logistic map $f : \bar{R} \to \bar{R} : x \to \theta x(1-x)$, with $\theta = 3.5, 3.51, ..., 4.00$. The noise-free attractors are given by $\{-\infty\}$ only in the case $\theta = 4.00$ (see the beginning of this section) and by $\{-\infty\}$ and the attractor shown by Figure 2.3 for $3.5 \leq \theta < 4.00$.

Let the logistic map be submitted to small random uniformly distributed perturbations $U(-\varepsilon, \varepsilon)$, with $\varepsilon > 0$. Theorems 2.1 and 2.2 then apply. Figure 2.5 shows the simulated noisy attractors for different values of ε. If no attractor appears for some ε and θ, it means that the only noisy attractor is $\{-\infty\}$. The simulated attractors which are in agreement with our theorems form by far the majority. Now, in a few cases, Theorem 2.1 tells us that the only noisy attractor is $\{-\infty\}$ (that is, the other attractor has been destabilized) and still the simulations provide us with an attractor between 0 and 1. Let

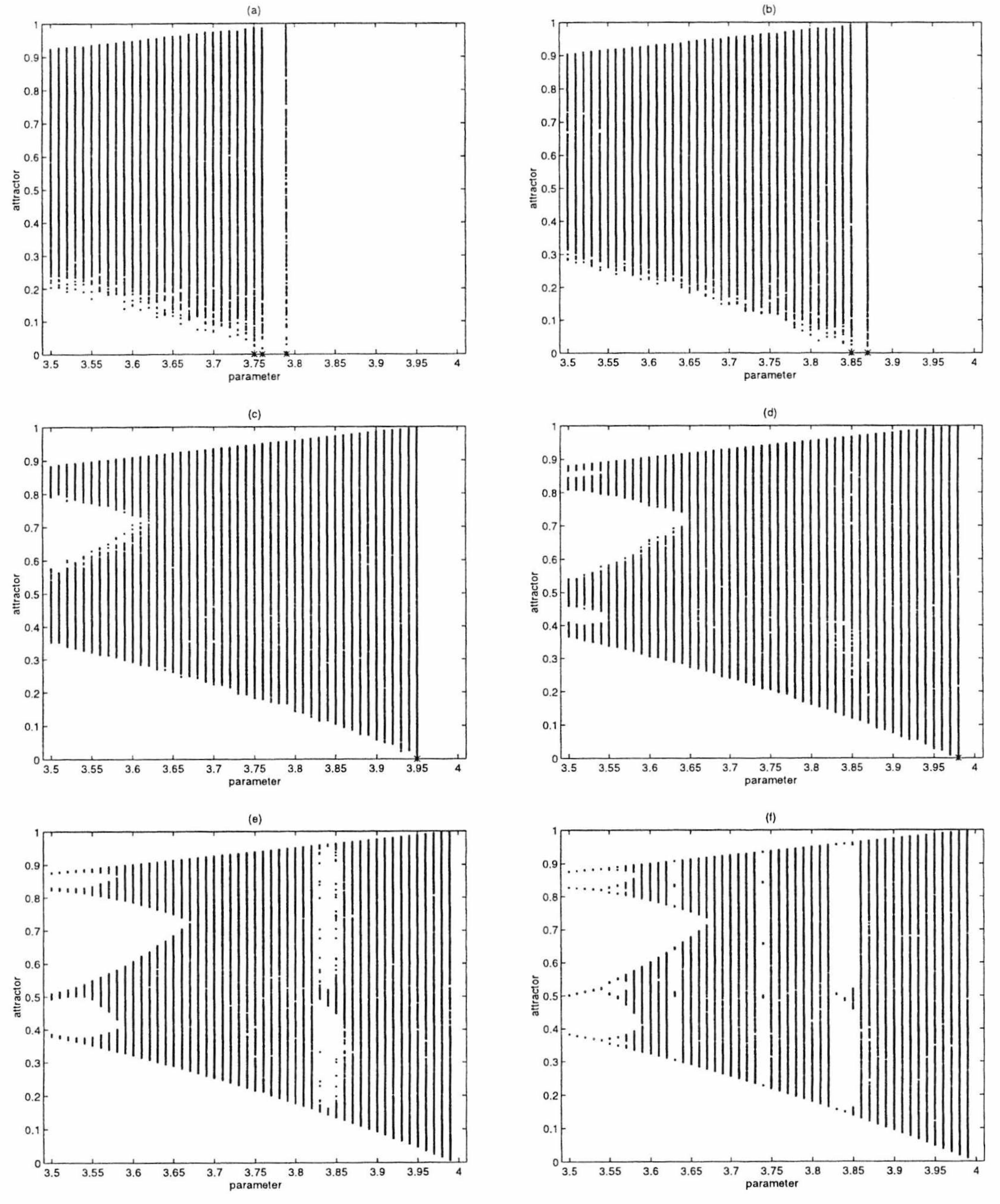


Figure 2.5: Simulated noisy attractors of $X_t = f(X_{t-1}) + \varepsilon_t$, where f is the logistic map with $\theta = 3.5, 3.51, ..., 4.00$ and the ε_t are *iid* with unform distribution $U(-\varepsilon, \varepsilon)$ with, (a) $\varepsilon = 0.05$, (b) $\varepsilon = 0.03$, (c) $\varepsilon = 0.01$, (d) $\varepsilon = 0.005$, (e) $\varepsilon = 0.001$, (f) $\varepsilon = 0.0001$.
For each θ and ε, the time series starts at random between 0 and 1. The map is iterated 1000 times without plotting anything; then, if the motion is still between 0 and 1, the next 1000 iterations are plotted on the figure. The simulated attractors which are not in agreement with our theory are indicated by an asterisk $(*)$.

us have a closer look at one example: $\varepsilon = 0.05$ and $\theta = 3.75$. There is an 0.05-pseudoorbit going from 0.5 (and so from the attractor) to $-\infty$: in fact, $0.5 \to f(0.5) + 0.05 = 0.9875 \to f(0.9875) - 0.05 = -0.0037 < 0$ and so the true attractor between 0 and 1 is destabilized.

This apparent disagreement between simulations and theory disappears if we let the number of iterations increase by a sufficient amount. We have to remember that Theorem 2.1 deals with the asymptotic (i.e. as $t \to \infty$) motion.

Finally, two remarks are worth to notice; as $\varepsilon \to 0$, the noisy attractor tends to the true attractor, as announced by Theorem 2.2. Note also for true periodic attractors and high noise levels the important increase of the support for the noisy attractors corresponding to them.

2.4 Extensions towards state-dependent noise and randomly varying parameters

In this section, we concentrate on unidimensional maps; the main reason for this is the complexity of the notations for higher dimensional maps, which would have made our aims less clear.

Now, in the previous sections, the dynamical system $x_t = f(x_{t-1})$ was subject to *iid* dynamic noise $\{\varepsilon_t\}$ with ε_t independent from the past states $X_{t-1}, X_{t-2}, ...$ (see Assumption 1). Let us relax this assumption in the following way.

Assumption 3 *A dynamical system submitted to small random perturbations is such that* $X_t = f(X_{t-1}) + g(X_{t-1}, \delta_t)$*, where* $\{\delta_t\}$ *is a sequence of zero mean iid absolutely continuous random variables with support* $[-\delta, \delta]$*,* δ *positive constant.*

This assumption is satisfied by dynamical systems subject to state dependent noise. Boundary conditions typically imply this kind of noise. Also, the computer makes precisely the same round-off error when acting on the same number and therefore state-dependent noise occurs.

Now, there is another important case in practice, namely a parameter of a dynamical system varies slightly in a random manner with time, for which Assumption 3 is often satisfied. Let $x_t = f(a, x_{t-1})$ a dynamical system, where a is a known real parameter influencing f and so the system. Now, replace a by $a + \delta_t$ where δ_t is as described in the above assumption. The parameter is now varying with time or more precisely randomly varying with time and our system becomes $x_t = f(a + \delta_t, x_{t-1})$.

Now, for a lot of unidimensional maps f, our noisy system can be decomposed like $x_t = f(a, x_{t-1}) + g(a, x_{t-1}, \delta_t)$ and therefore satisfies Assumption 3. As an example, we can take the logistic map with parameter θ (see section 2.1): $x_t = (\theta + \delta_t)x_{t-1}(1 - x_{t-1}) = \theta x_{t-1}(1 - x_{t-1}) + \delta_t x_{t-1}(1 - x_{t-1}) = f(x_{t-1}) + g(x_{t-1}, \delta_t)$. We shall come back to the logistic map later in this section. Note that for some maps, it is not possible to express the noisy system in the terms of Assumption 3; an example of such maps is given by $x_t = cos(ax_{t-1})$ with the noisy system corresponding to it $x_t = cos((a + \delta_t)x_{t-1})$.

Now, the framework remains the same as in section 2.2: M is a Hausdorff topological metric space with $dist(a, b) = |b - a|$. The maps $f^t : M \rightarrow M$ are assumed to be continuous and M is the largest possible domain of definition of f. As in section 2.2, we require for M to be compactified (Assumption 2).

We need to impose some conditions on the function g (that is, on the noise term) in view to get analogous theorems to those of section 2.2. We regroup these conditions on $g(x, y)$ in Assumption 4. Note that they are reasonable to make.

Assumption 4 • $\forall x \in M, g(x, 0) = 0.$

- $\forall x \in M$, *g is continuous, strictly monotonic and bounded in y.*

- $\forall y$, *g is continuous in x.*

Now, by analogy with sections 2.1 and 2.2, we can introduce the following notions.

Definition 2.9 *A sequence $(x_0, x_1, ..., x_n)$ is called an* **δ-pseudoorbit** *if there exist $\delta_i \in [-\delta, \delta]$, $i = 1, 2...n$ such that $x_i = f(x_{i-1}) + g(x_{i-1}, \delta_i)$, $i = 1, ..., n$.*

Definition 2.10 $\boldsymbol{a \succ_\delta b}$ *means: there is an δ-pseudoorbit going from a to b.*

The relation $\succ_\delta$ is a preorder. Therefore, equivalence classes for the relation $\succ_\delta$, denoted $[a]_\delta$, can be introduced and

Definition 2.11 *A minimal equivalence class $[b]_\delta$ for the relation $\succ_\delta$ is called an* **attractor for the relation $\succ_\delta$**.

Now, is it possible to link δ-pseudoorbits and noise-free attractors (Definition 2.2)? In other words, if for arbitrarily small $\delta > 0$, there is an δ-pseudoorbit going from a to b, do we come back to the noise-free case $a \succ b$? In fact, the answer is: this depends from the particular case under consideration. So, since we do not necessarily come back to our true attractors, we need to go further.

Definition 2.12 $\boldsymbol{a \succ_{\delta \to 0} b}$ *means: for arbitrarily small $\delta > 0$, there is an δ-pseudoorbit going from a to b.*

This relation is a preorder. So, we can introduce equivalence classes $[a]_{\delta \to 0}$ and attractors for this relation.

Definition 2.13 *An* **attractor for the relation $\succ_{\delta \to 0}$** *is a minimal equivalence class for this relation.*

We should keep in mind that these attractors do not necessarily coincide with the true (noise-free) attractors, especially in the cases where the parameter varies randomly over time.

Then, in the same way as in section 2.2, we seek theorems describing the asymptotic motion of systems submitted to random perturbations in the sense of Assumption 3. In particular, the following lemma plays an important role in proving these theorems.

Lemma 2.4 *Let Assumptions 2,3 and 4 apply. Let $f : M \to M$ be continuous. Then, if a does not belong to an attractor for the relation $\succ_\delta$, then there exists $\gamma(a)$ such that the subspace $B_a(\gamma)$ is transient with probability* 1.

The proof of this lemma follows the same scheme as Lemma 2.3 's proof and some technical points, although not very hard, lenghten the proof quite a lot. So, by sake of concision, we shall leave out the proof, which would not bring anything more to the discussion.

Now, given the noise level δ, the following theorem gives us the attractor(s) of the noisy system.

Theorem 2.3 *Let Assumptions 2,3 and 4 apply. Let f be continuous.*
Let $A_{\delta\to 0}$ denote the union of all attractors for the relation $\succ_{\delta\to 0}$.
Let L_δ denote the union of the attractors for the relation $\succ_{\delta\to 0}$, which do not belong to attractors for the relation $\succ_\delta$.
Let $E_\delta = A_{\delta\to 0} \setminus L_\delta$.
Then, as $t \to \infty$, the (noisy) system tends to $A_\delta = \bigcup_{z\in E_\delta}[z]_\delta$ with probability 1.

Proof: The proof is similar to Theorem 2.1 ' s proof.

The next theorem deals with infinitesimal random perturbations (that is, $\delta \to 0$) and provides us with the attractor(s) for a system so perturbed.

Theorem 2.4 *Let Assumptions 2,3 and 4 apply. Let f be continuous.*

Under the same notations as in Theorem 2.3,

$$\lim_{\delta\to 0} A_\delta = A_{\delta\to 0}.$$

Proof: The proof is similar to Theorem 2.2 's proof.

If, for the system under consideration, $A_{\delta\to 0}$ is the true (noise-free) attractor, then we can say that the system is stable under such infinitesimal random perturbations. At this stage, we would like to insist on one point: in some cases, it is straightforward to see if the system is stable or not but, in other cases, the calculation of $A_{\delta\to 0}$ involves highly complex hidden dynamical structures and even sometimes a definite answer cannot be given. So, saying that a system is stable or not can be an awkward task.

We can now look at an example. Consider the logistic map $f : \bar{R} \to \bar{R} : x \to \theta x(1-x)$ (see section 2.1) and replace θ by $\theta + \delta_t$, where $\{\delta_t\}$ is a sequence of independent random variables with uniform distribution $U(-\delta, \delta)$. Then, $x_t = (\theta+\delta_t)x_{t-1}(1-x_{t-1}) = \theta x_{t-1}(1-x_{t-1}) + \delta_t x_{t-1}(1-x_{t-1}) = f(x_{t-1}) + g(x_{t-1}, \delta_t)$ and both Assumptions 3 and 4 are satisfied.

Theorem 2.3 allows us to construct the noisy attractor(s) for a given noise level δ. Of course, for all θ, $\{-\infty\}$ is an attractor; if it is not the only one, then the second attractor should contain all the noise-free attractors of the range $[\theta-\delta, \theta+\delta]$. So, the second attractor, if it exists, is at least as big as the biggest true attractor in the range (in fact, in our case, it is even bigger).

Figure 2.6 provides us with simulations of the noisy attractor situated between 0 and 1, if it exists, for different values of δ. All the simulations are in accordance with the theory. Comparing these graphs with Figure 2.3, we can see the dramatic changes for noise-free periodic attractors. This important increase in the support of the noisy attractors corresponding to them is easily explained: since the chaotic attractors are dense in $[r_\infty, 4.00]$, for a given noise level δ, there will be chaotic attractors in the range $[\theta+\delta, \theta+\delta]$ and so the noisy attractor is, by Theorem 2.3, at least as big as these chaotic attractors.

Theorem 2.4 states that the asymptotic motion is concentrated on $A_{\delta\to 0}$ for a system submitted to infinitesimal random perturbations in the sense of Assumption 3. Now, for the logistic map, the calculation of $A_{\delta\to 0}$ needs the knowledge of very fine structures of the dynamics; Basing ourselves on the results and conjectures in Ott (1993), it seems well that the chaotic attractors are stable under such types of perturbations; the result is uncertain for periodic attractors (the simulations seem to suggest that they also are stable).

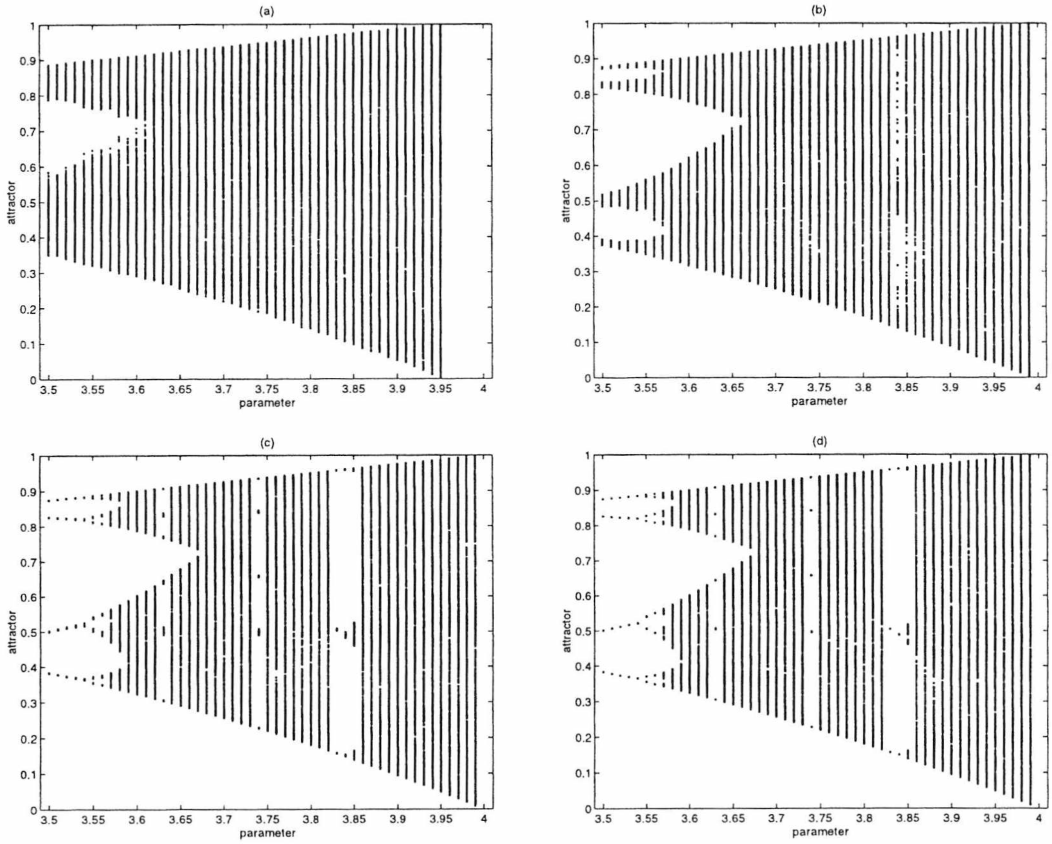


Figure 2.6: Simulated noisy attractors of $X_t = (\theta + \delta_t)X_{t-1}(1 - X_{t-1})$, where $\theta = 3.5, 3.51, ..., 4.00$ and the δ_t are *iid* with uniform distribution $U(-\delta, \delta)$. The different noise levels are: (a) $\delta = 0.05$, (b) $\delta = 0.03$, (c) $\delta = 0.01$, and (d) $\delta = 0.001$.
For each θ and δ, the time series starts at random between 0 and 1. The map is iterated 1000 times without plotting anything; then, if the motion is still between 0 and 1, the next 1000 iterations are plotted on the graph.

To conclude this section, we can say that, for systems satisfying Assumptions 3 and 4, it is possible to construct the noisy attractors for a given noise level δ; it is also possible, at least theoretically, to construct the attractors for systems submitted to infinitesimal random perturbations. If the system is stable under these perturbations (that is, $A_{\delta\to 0}$ coincides with the noise-free attractors), which is not necessarily the case, then the comments of section 2.3 obviously apply. This means that, for a given noise level δ, some (but not all) of the noise-free attractors can be "lost"; there is no created "completely fictitious" noisy attractor and the support of each noisy attractor is greater or equal to the support(s) of the true attractor(s) corresponding to it.

Chapter 3

Chaotic sequences: introduction and some results

Here is the plan of chapter 3. Section 3.1 introduces the Lyapunov exponent which allows us to give a precise definition of a chaotic system. Section 3.2 explains first what we shall actually call a chaotic sequence, a notion which plays a central part in the thesis. The degree of stochasticity of chaotic sequences is then examined. Finally, section 3.2 stresses the importance of this notion in various topics. Section 3.3 states and proves a theorem for chaotic sequences, which is both a corollary of and an extension to the ergodic theorem. This theorem has important consequences, in particular on the sample moments and on the empirical distribution function; they are highlighted at the end of the section. Section 3.4 computes the Lyapunov exponents of some chaos driven models analysed later in the thesis and links these Lyapunov exponents to the fractal dimension of attractors.

3.1 Lyapunov exponents

In the previous chapters, the so-called chaotic dynamics has been mentioned on a few occasions. In an informal way, chaotic dynamics has been associated with 'sensitive dependence on initial conditions': starting from any typical value in a neighbourhood of any initial condition x_0 the orbit diverges from that starting from x_0. So, small errors in the initial value grow with time; in fact, they grow very rapidly (exponentially) with time, which makes long-term prediction impossible in practice (that is, when noise comes into account). Note, however, that statistical properties, unlike dynamical properties, can be preserved in a noisy environment: chapter 2 has shown that the noise-free attractors, so in particular the chaotic attractors, are stable under some kinds of random perturbations.

Now, coming back to a noise-free dynamical system, section 2.1 names the chaotic attractors (also called chaos) in the list of the different kinds of attractors and gives examples picked from the logistic map. At first sight, it could seem surprising to be able to get chaotic attractors, which are bounded, as chaotic dynamics is about exponential divergence of the orbits. In fact, this can easily be explained: besides the stretching process which leads to exponential divergence of nearby trajectories, the system has also a folding process which keeps the orbits bounded.

There is a well-known formal way of measuring the degree of dependence on initial conditions for a dynamical system, namely the Lyapunov exponents. We shall have a look at them now; we refer, for example, to Ruelle (1989) or Ott (1993) for more details.

First, consider the unidimensional differentiable map f generating the orbit $(x_0, ..., x_n, f(x_n) = x_{n+1}, ...)$. Let x_0 and x'_0 denote two nearby initial points, x_n and x'_n, $n = 0, 1, 2...$ denote the orbits starting respectively from x_0 and x'_0 and let f' denote the differential of f.

Then,

$$x'_n - x_n = f^n(x'_0) - f^n(x_0) \simeq \frac{df^n}{dx}(x_0)(x_0 - x'_0)$$

.

Now, the chain rule of differentiation , that is

$$\frac{df^n}{dx}(x_0) = \frac{df}{dx}(x_{n-1})\frac{df}{dx}(x_{n-2})...\frac{df}{dx}(x_0),$$

allows us to define the average rate of change of $\frac{df^n}{dx}(x_0)$ as

$$\lambda(x_0) = \lim_{n\to\infty} \frac{1}{n}\, ln|\frac{df^n}{dx}(x_0)|,$$

the Lyapunov exponent at x_0. Here, we have assumed that the limit exists.

We notice that $x_n' - x_n \simeq e^{n\lambda(x_0)}(x_0' - x_0)$; so a positive $\lambda(x_0)$ leads to a magnification of the initial difference. Now, this $\lambda(x_0)$ is a local Lyapunov exponent in that it depends on x_0. In fact, if the system is ergodic, which is the case of all chaotic attractors, (see next section for a definition of ergodicity), it can be proved (see Ruelle (1989) and the references therein) that $\lambda(x_0) = \lambda$ for ρ-almost all initial conditions x_0 where ρ denotes an ergodic distribution of the system. We call λ the (global) Lyapunov exponent of the system, which is chaotic if and only if $\lambda > 0$. Note that $\lambda = \int ln|f'(y)|\rho(dy)$.

Let us generalise the notion of Lyapunov exponents to N-dimensional maps M. A N-dimensional dynamical system has N (not necessarily distinct) Lyapunov exponents. They can be defined in the following way. Let the map M be differentiable and the system generated by M be ergodic; then, for typical initial condition x_0 (in the sense of previous paragraph), the (global) Lyapunov exponent corresponding to the orientation u_0 (the vector u_0 is assumed to have unit modulus) is equal to

$$h(u_0) = \lim_{n\to\infty} \frac{1}{n}\, ln|DM^n(x_0)u_0|,$$

where $DM^n(x_0)$ is the Jacobian matrix of M^n evaluated at x_0. Now, let h_{in} denote the modulus of the ith eigenvalue of $DM^n(x_0)$, ordered so that $h_{1n} \geq h_{2n} \geq ... \geq h_{Nn}$. The ith Lyapunov exponent λ_i is then

$$\lambda_i = \lim_{n\to\infty} \frac{1}{n}\, ln\ h_{in}, \quad i = 1, ..., N.$$

Clearly, if $\lambda_i > 0$, small differences in the initial conditions are stretched in the direction of the ith eigenvector of $DM^n(x_0)$.

Definition 3.1 *A bounded dynamical system is said to be* **chaotic** *if it possesses at least one positive Lyapunov exponent; hence a bounded system is chaotic if and only if* $\lambda_1 > 0$.

Finally, note that the Lyapunov exponents are not only useful in detecting chaotic systems but also in quantifying the sensitive dependence on initial conditions.

Now, we can compute the Lyapunov exponent of the logistic map for the same values of the parameter θ as in Figure 2.3, which pictures the attractors of the dynamical systems generated by $f_\theta : [0,1] \to [0,1] : x \to \theta x(1-x), \quad 3.5 \leq \theta \leq 4.0$. What we have been actually calculating is the quantity

$$\lambda_\theta = \frac{1}{n} \sum_{i=0}^{n-1} ln \left| \frac{df_\theta}{dx}(f_\theta^i(x_0)) \right|,$$

with $n = 5000$, where x_0 is a typical initial condition for the ergodic distribution ρ_θ. This is justified by the facts that the logistic map defined on $[0,1]$ is ergodic for all θ between 0 and 4 and that the ergodic theorem applies in our case (see the next two sections for more details on these notions).

The results are shown in Figure 3.1. As expected, the Lyapunov exponent is non-positive for systems exhibiting limit cycles and $\lambda > 0$ for systems which converge to what we had called 'chaos' in section 2.1 (this terminology is now fully justified). Note that the Lyapunov exponent is roughly increasing as θ increases and that the maximum is reached at $\theta = 4$; λ is then equal to $ln\ 2$.

3.2 Chaotic sequences: introduction and motivations

In this section, we focus on unidimensional maps; the main reason for this is the simplicity of the notations in this case. Now, in chapter 2 and more precisely

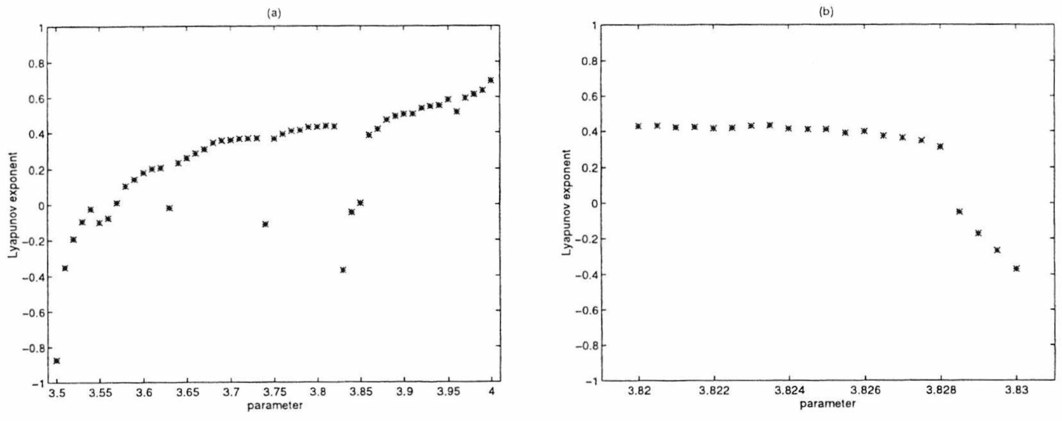


Figure 3.1: Lyapunov exponent for the logistic map with (a) $\theta = 3.5, 3.51, ..., 4.00$, (b) $\theta = 3.82, 3.8205, ..., 3.83$.

in section 2.1 we dealt with the support of attractors for dynamical systems. Knowing that the asymptotic motion is concentrated on the attractor, the next question is: does there exist a probability measure which is invariant under the effect of the dynamical system? In fact, it has been proved (see, for example, Ruelle (1989)) that if a compact set A is invariant under the time evolution f^t, where for each t the map $f^t : A \to A$ is continuous, then there is a probability measure π invariant under f^t and with support contained in A; moreover, one may choose π to be ergodic (that is, indecomposable: there does not exist any invariant measures π_1 and π_2, $\pi_1 \neq \pi_2$, such that $\pi = \alpha_1\pi_1 + \alpha_2\pi_2$, $\alpha_1, \alpha_2 \in (0, 1)$).

So, we know that, for many dynamical systems, there exists a (not necessarily unique) invariant distribution associated to them. Note, however, that there is no general way for constructing these invariant distributions. For some maps, it has been possible to derive an invariant distribution: for example, it can easily be seen that an invariant (ergodic) density for the logistic map $f : [0, 1] \to [0, 1] : x \to 4x(1 - x)$ is the Beta$(1/2; 1/2)$ probability distribution function (that is, $g(x) = \pi^{-1}x^{-1/2}(1 - x)^{-1/2}$ on $]0, 1[$). And an ergodic probability distribution of the modulo map is given by the uniform distribution on $(0, 1)$. These invariant distributions are shown in Figure 3.2.

Now, to find invariant distributions explicitly is a very hard (even impossible) task in many cases. In particular, exact calculations for many logistic maps leading to chaos are suspected to be impossible (see Hall and Wolff (1995a)). Simulations, however, allow us to get a good idea of the invariant distributions. Moreover, section 3.3 will ensure us, under very weak additional conditions, that, as the sampling size of the simulations tends to ∞, the simulated probability distribution functions tend to the invariant distributions with probability 1.

Four logistic maps will be more particularly examined throughout this thesis, namely the logistic maps with parameter $\theta = 4.00, 3.98, 3.825$ and 3.58. These values have been selected because they appear in Hall and Wolff (1995b) and they lead to quite different behaviors as we shall see in the thesis. Their attractors can be seen in Figure 2.3; note that all of them appear to be chaotic (see section 3.1 and Figure 3.1). Figure 3.2 shows their invariant distributions; except for $\theta = 4.00$, the invariant distributions have been obtained through simulations: the first 1000 iterations being discarded as a warming-up, the next 5000 iterations come into account to construct the invariant distribution.

At this stage, it is time to introduce the notion of chaotic sequence. Without loss of generality, we can assume that E_t has zero mean.

Definition 3.2 *A* **chaotic sequence** $\{E_t\}$ *is a sequence of identically distributed random variables generated by a real chaotic map f such that $E_{t+1} = f(E_t)$ and each E_t has an ergodic distribution associated with f as its marginal distribution.*

Now, an important question we could ask ourselves is: how random are the chaotic sequences? What is their degree of stochasticity?

First, we shall precise our framework and make a list of notions, list in order of 'increasing randomness'; then, we shall fix where the chaotic sequences are on the list. Note that many notions defined here will appear again later in the thesis.

For the rest of this section, we suppose that $\{X_k,\ k \in Z\}$ is a sequence of strictly stationary random variables; in particular, this implies that the dynamical

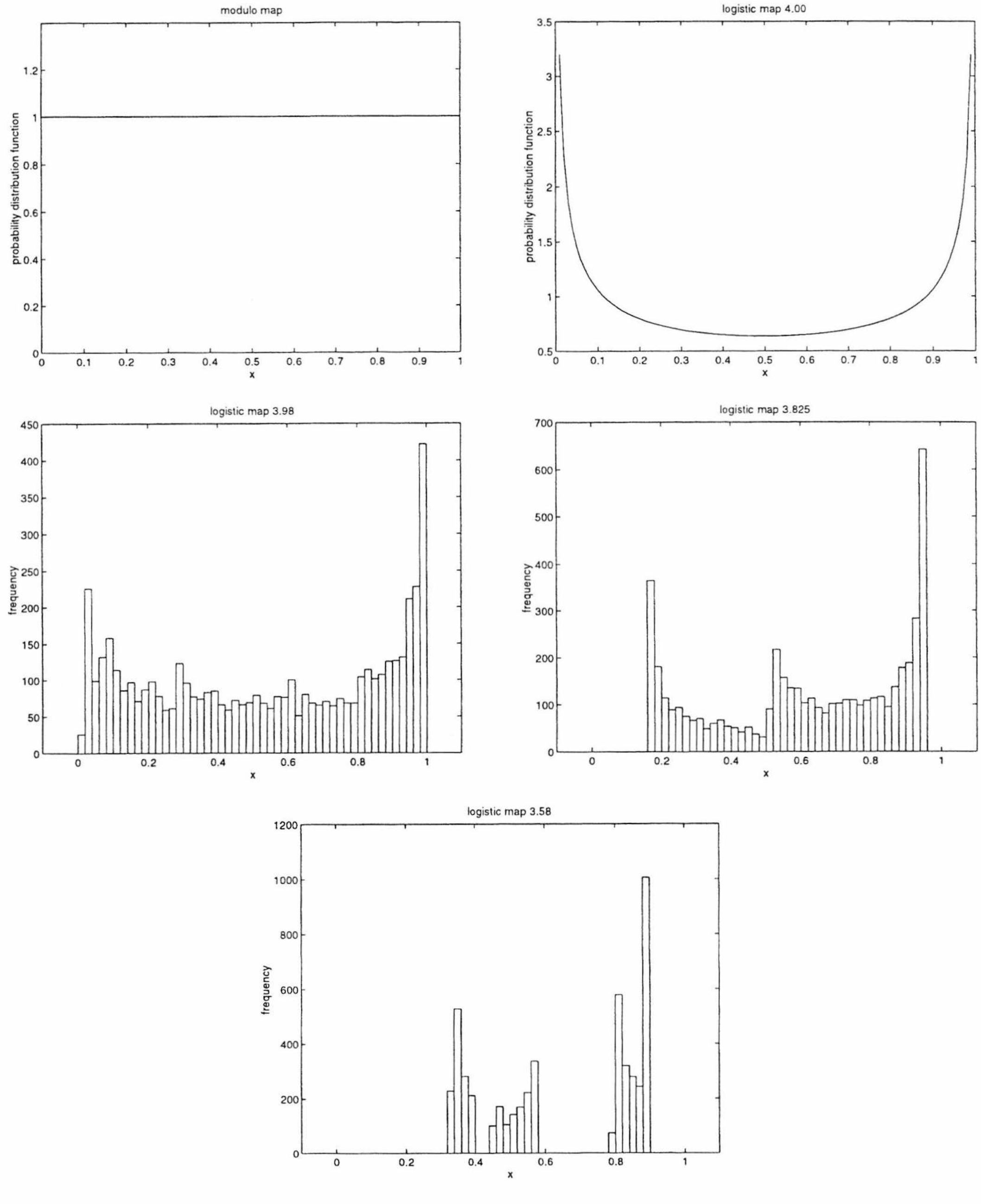


Figure 3.2: Ergodic probability distribution functions
The ergodic distributions of the modulo map and of the logistic map with $\theta = 4.00$ are obtained theoretically. The last three distributions have been obtained through simulations: the first 1000 iterations are discarded, the next 5000 iterations are used to build the histogram.

system under consideration has an invariant measure. We assume that our probability space is $(R^Z, \mathcal{B}^Z, P)$, where $\mathcal{B}$ denotes the standard Borel σ-field; we shall assume that for each $\omega \in R^Z$ and each $k \in Z$, $X_k(\omega) = \omega_k$. For $-\infty \leq J \leq L \leq \infty$ define $\mathcal{F}_J^L = \sigma(X_k, J \leq k \leq L)$, where $\sigma(X_k, J \leq k \leq L)$ denotes the σ-field generated by this family of random variables. Let T denote the shift operator; that is, for each $\omega \in R^Z$, $T\omega$ is defined by $(T\omega)_k = \omega_{k+1}$ and for any event $A \in \mathcal{B}^Z$, $TA = \{\omega : T^{-1}\omega \in A\}$. Now, the following definitions come mainly from Bradley (1986), Sinai (1989) and Billingsley (1965).

Definition 3.3 *A dynamical system is said to be* **ergodic** *(with respect to T) if each invariant set (that is, a set such that $T^{-1}A = A$, or equivalently $TA = A$) is trivial in the sense of having measure either 0 or 1.*

Definition 3.4 *The sequence $\{X_k\}$ is* **mixing in the ergodic-theoretic sense** *if,*

$$\forall A, B \in \mathcal{B}^Z, \quad \lim_{n \to \infty} P(A \cap T^{-n}B) = P(A)P(B).$$

Definition 3.5 *The sequence $\{X_k\}$ is said to be a* **K-system** *if its future tail σ-field $\bigcap_{n=1}^{\infty} \mathcal{F}_n^{\infty}$ is trivial (that is, contains only events of probability 0 or 1).*

Definition 3.6 *The system is said to be a* **Bernoulli system** *if it can be represented as a symbolic dynamics consisting on a shift T on a finite number of symbols.*

Definition 3.7 *The process $\{X_k\}$ is said to be* **strongly mixing** *if $\alpha(n)$ tends to 0 as $n \to \infty$, where*

$$\alpha(n) = \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^{\infty}) = sup|P(A \cap B) - P(A)P(B)|, \quad A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^{\infty}.$$

Definition 3.8 *The process $\{X_k\}$ is said to be* **absolutely regular** *if $\beta(n)$ tends to 0 as $n \to \infty$, where*

$$\beta(n) = \beta(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_{n}^{\infty}) = sup\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}|P(A_i \cap B_j) - P(A_i)P(B_j)|,$$

the sup being taken over all pairs of partitions $\{A_1, ..., A_I\}$ and $\{B_1, ..., B_J\}$ such that $A_i \in \mathcal{F}_{-\infty}^{0}$ for all i and $B_j \in \mathcal{F}_{n}^{\infty}$ for all j.

Definition 3.9 *The sequence $\{X_k\}$ is* **ϕ-mixing** *if $\phi(n)$ tends to 0 as $n \to \infty$, where*

$$\phi(n) = \phi(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_{n}^{\infty}) = sup|P(B|A) - P(B)|, \quad A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{n}^{\infty}.$$

All these definitions are modified in a straightforward way in the case of a singly-infinite strictly stationary sequence $\{X_k : k = 1, 2, 3...\}$. We recall that the list is in order of 'increasing randomness', that is a process which is ϕ-mixing is absolutely regular, a process which is absolutely regular is strongly mixing and so on... . Note that there exist other notions for measuring the degree of dependence of a process (see, for example, Bradley (1986) for a review) and that an *iid* sequence is, of course, ϕ-mixing, absolutely regular,..., ergodic.

Now, the following proposition tells us where the chaotic sequences are in the hierarchy.

Proposition 3.1 *A deterministic strictly stationary sequence $\{X_k\}$ is chaotic if and only if it is a K-system.*

Proof: see Billingsley (1965) and especially Sinai (1989) and Ott (1993). □

Note that a chaotic dynamical system (in the sense of Definition 3.1) does not necessarily lead to a K-system since we need to have a strictly stationary sequence for the above proposition to apply. But the chaotic sequences (see Definition 3.2) are strictly stationary and so are K-systems, therefore mixing in the ergodic-theoretic sense and ergodic.

Some chaotic sequences are not only K-systems but also Bernoulli systems. Such an example is given by the sequence $\{E_t\}$, where the chaotic map f is the modulo map going from $[0,1]$ to itself (see section 2.1), that is $E_{t+1} = 2E_t \pmod 1$, and the associated invariant measure is the uniform distribution on $[0,1]$. Now,

$$E_t - \frac{1}{2} = \sum_{j=-\infty}^{0} 2^{j-1} Z_{t-j},$$

where for each t, $P(Z_t = 0.5) = P(Z_t = -0.5) = 0.5$ (see section 2.1 for more details). So, the modulo map is a Bernoulli system (consisting on a shift on two symbols).

Now, is it possible that some chaotic sequences are strongly mixing? The answer is no, since $\alpha(n) = \alpha(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_{n}^{\infty}) \geq \alpha(\mathcal{F}_{0}^{0}, \mathcal{F}_{n}^{n}) = \alpha(\sigma(X_0), \sigma(X_n))$ and X_n is a (Borel-measurable) function of X_0, so $\alpha(n) \geq \alpha(\sigma(X_0, \sigma(X_n)) \geq \alpha(\sigma(X_0), \sigma(X_0))$ $= c$, strictly positive constant for all n. Thus, α_n does not converge to 0 as $n \to \infty$.

The following proposition is quite interesting.

Proposition 3.2 *If $\{E_t\}$ is a chaotic sequence and $g : R \to R$ is a Borel-measurable function, then the sequence $\{g(E_t)\}$ is a K-system.*

Proof: the proof follows directly from Proposition 3.1 and from the fact that $\sigma(g(E_t)) \subset \sigma(E_t)$ for all t, so the future tail σ-field of $\{g(E_t)\}$ is also trivial. □

Note that the sequence $\{g(E_t)\}$ is not necessarily a chaotic sequence (in the sense of Definition 3.2) since it is possible that $g(E_{t+1})$ is not equal to a (chaotic) function of $g(E_t)$. We shall expand this in section 3.4.

A chaotic sequence $\{E_t\}$, that is a strictly stationary process where $E_{t+1} =$

$f(E_t)$, f being a chaotic map, (see Definition 3.2), obviously has extremely strong structural dependence since any data value may be represented as a deterministic function of any of the previous values. Nevertheless, many chaotic sequences exhibit short-range statistical dependence (that is, $\sum_{i\geq 2} |corr(E_1, E_i)| < \infty$); some of them, like the logistic map with $\theta = 4.00$, are even such that the sequence $\{E_t\}$ is uncorrelated for all lags (see, e.g., Hall and Wolff (1995b)).

Figure 3.3 shows the autocorrelation functions of sequences generated by the modulo map and the logistic map with $\theta = 4.00, 3.98, 3.825$ and 3.58. For the modulo map, the autocorrelation function is positive at all lags and decays in an exponential way. For the three logistic maps with $\theta = 3.98, 3.825$ and 3.58, the (simulated; note that section 3.3 will show that the simulated values are very close to the true (unknown) ones) autocorrelation functions respectively appear to decay quite quickly, quite slowly and not at all. We notice that the function is negative at its first lag for these three maps.

Considering the last paragraph, it seems that there is a link between the parameter value θ and the statistical dependence, $\sum_{i\geq 2} |corr(E_1, E_i)|$, namely that an increase in θ could imply a decrease in the statistical dependence. We had a closer look at this: Figure 3.4 computes via simulations $\sum_{i\geq 2} |corr(E_1, E_i)|$ and plots it versus the parameter θ. Note the obvious facts that the expression $\sum_{i\geq 2} |corr(E_1, E_i)|$ is not defined if E_t is a fixed point and that $\sum_{i\geq 2} |corr(E_1, E_i)| = \infty$ if E_t is a limit cycle. Now, we can go further and detect a similarity between Figure 3.1 and Figure 3.4 in the sense that a higher Lyapunov exponent seems to mean less statistical dependence, which can be seen as intuitively logical.

The motivations for introducing chaotic sequences are of different kinds. One can treat a chaotic sequence as a time series in its own right and compute the moments and the asymptotic properties for sequences generated by a particular

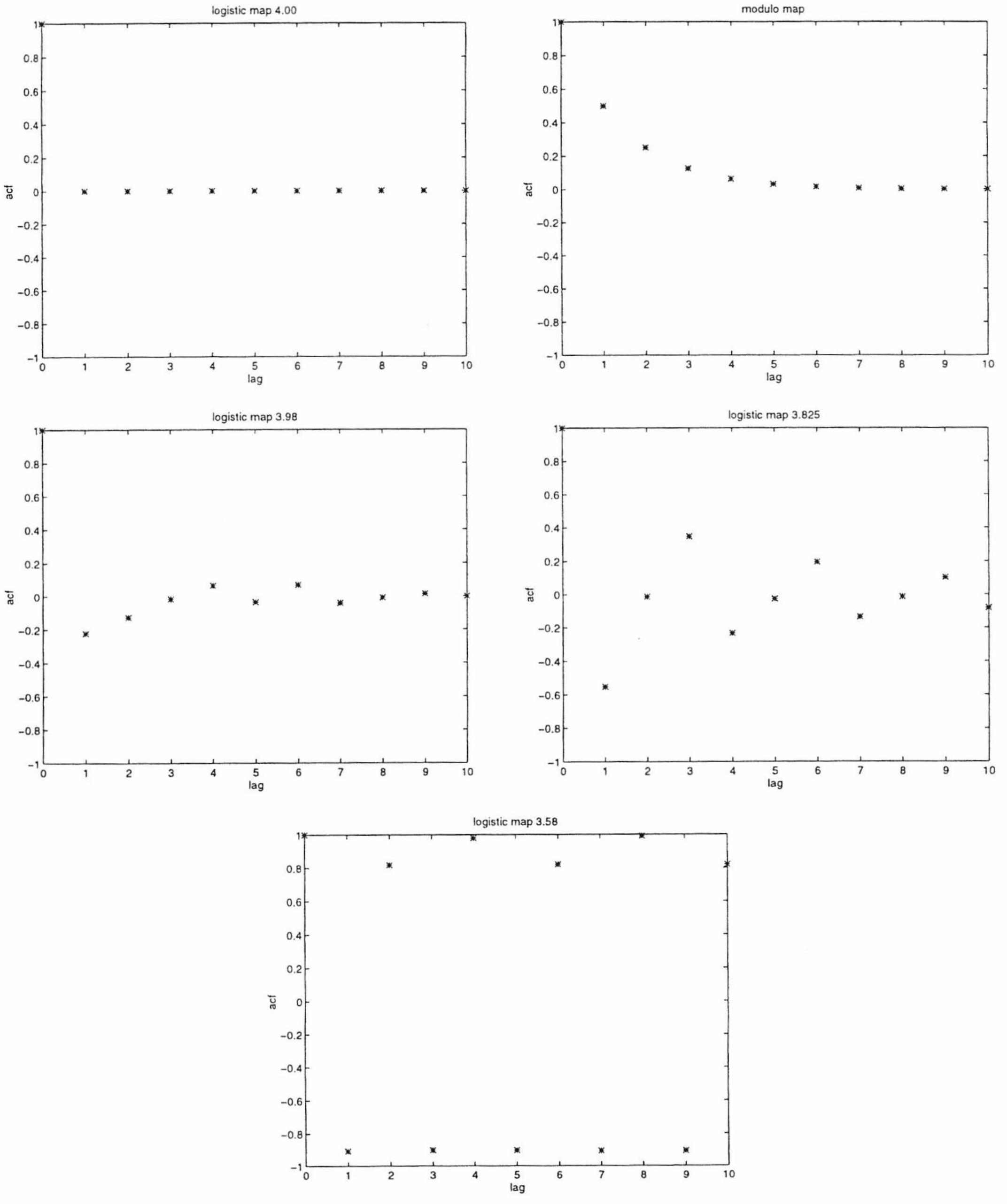


Figure 3.3: Autocorrelation functions.
The theoretical values, when available, are displayed. There are no theoretical values available for the last three logistic maps since the invariant distributions are not known analitically. Consequently we have used simulations to get the autocorrelation functions for these maps.

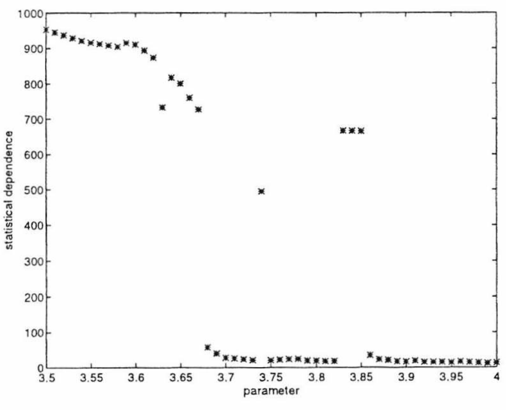


Figure 3.4: The value $\sum_{i\geq 2}^{n} |corr(E_1, E_i)|$ (statistical dependence) is plotted versus the parameter θ of the logistic map. The sample size n is taken to be 1000.

map, like Hall and Wolff (1995b) analysing in detail properties of sequences generated by the logistic map with $\theta = 4.00$. The beginning of this section and the next sections of this chapter will provide us with important properties of $\{E_t\}$, which are valid for sequences generated by any real chaotic map f.

On the other hand, consider stochastic models $X_t = g(X_{t-1}, Y_t, \varepsilon_t)$ but now instead of taking $\{\varepsilon_t\}$ a sequence of *iid* random variables, we take $\{E_t\}$ a chaotic sequence and we look at $X_t = g(X_{t-1}, Y_t, E_t)$. Dynamicists have recently shown considerable interests in these chaos driven models (see, in particular, Takens (1994b)). Clearly, attractors of the two-dimensional deterministic system $R^2 \to R^2$ defined by

$$X_{t+1} = \alpha X_t + E_{t+1}, \quad E_{t+1} = f(E_t)$$

are of dimension no larger than 2. Their primary interests include the investigation of the extent to which such a low-dimensional deterministic dynamical system can mimic a stochastic dynamical system, which may be loosely speaking likened to an infinite dimensional deterministic dynamical system. In this thesis, we are not going to spend much time about these interests; note, however, that section 3.4 will touch the point by calculating the Lyapunov exponents of chaos driven systems.

The next two chapters will focus on the statistical inference on chaos driven AR models and chaos driven linear stochastic regression models. As already announced in chapter 1, this is particularly interesting since the computers, which are physical systems, cannot generate sequences $\{\varepsilon_t\}$ of *iid* random variables but instead simulate them by means of pseudo-random number generators which are typically chaotic maps.

Now, many papers in the statistical literature use computers to confirm their results by means of simulations. Tong (1995) has commented that simulation studies typically assume albeit implicitly that central limit properties existing in the case of a model driven by *iid* noise continue to hold even when we replace $\{\varepsilon_t\}$ by a chaotic sequence $\{E_t\}$. The next two chapters are going to examine why this is the case for two well-known types of models.

We can say here a few words on pseudo-random number generators. In fact, if we have a generator of pseudo-random variables uniformly distributed between 0 and 1, then we can simulate any sequence of independent random variables. We can give a rough idea of the reasoning in the following way. Let U be uniformly distributed on $[0, 1]$ and X a random variable obtained by $X = F^{-1}(U)$, where F is any cumulative distribution function. The distribution of X follows then a law whose cumulative distribution function is $F(x)$; for: $P(X < x) = P(F(X) < F(x)) = P(U < u) = u$, since U is uniformly distributed on $[0, 1]$, and therefore $F(x) = u$.

Now, the pseudo-random number generator of $\varepsilon_t \sim iid\ U(0, 1)$ has, among other things, to be as close as possible to independent output (there are numerous tests, empirically or theoretically based, to look at this; see, for example, Knuth (1981) and the references therein) but also it should allow for repeatability and it should be fast (see, e.g., Ripley (1990) for more details). The most popular random number generators use the linear congruential method: that is $E_{n+1} = (aE_n + c)$ mod m, with $m > 0$, $0 \leq a < m$, $0 \leq c < m$ and $0 \leq X_0 < m$ (see, e.g., Knuth (1981) for more details). In this thesis, we make use of the NAG FORTRAN library; the routine G05CAF uses the multiplicative congruential method $X_{n+1} =$

$13^{13}X_n \bmod 2^{59}$.

Now, as pointed out by a referee of Stockis and Tong (1996), some of these pseudo-random number generators operate on congruential sequences $\{E_t\}$ which are not strictly chaotic since they have periods and discrete support. Of course, these periods are very long (for example, the G05CAF routine has a period equal to 2^{57}) and we can say that a chaotic sequence provides with a very good approximation to pseudo-random number generators. Moreover, we note that many results of this thesis still hold if, keeping the other assumptions, we only require f to be a real deterministic map and we assume that E_t admits an invariant distribution which is also ergodic.

3.3 Properties derived from the ergodic theorem

It has been shown in the last section that a chaotic sequence $\{E_t\}$ was ergodic. Now, the principal consequence of ergodicity is the so-called ergodic theorem:
" Suppose a probability space is fixed. Let $\{X_t\}$ be an ergodic sequence, then, if $g : R \to R$ is an integrable function, we get

$$\frac{1}{n}\sum_{t=1}^{n} g(X_t) \to^{a.s.} \mathrm{E}[g(X_t)],$$

where $\to^{a.s.}$ denotes the almost sure convergence." (see, for example, Billingsley (1965) or Ruelle (1989) for more details). In the case of a chaotic sequence, the following theorem is easily derived from the ergodic theorem; throughout the section, we assume the existence of a probability space.

Theorem 3.1 *Let $\{E_t\}$ be a chaotic sequence. Then, if the function $h : R^{c+1} \to R : (E_t, E_{t+1}, ..., E_{t+c}) \to h(E_t, E_{t+1}, ..., E_{t+c})$, c finite non-negative number, is integrable, we get*

$$\frac{1}{n}\sum_{t=1}^{n} h(E_t, E_{t+1}, ..., E_{t+c}) \to^{a.s.} \mathrm{E}[h(E_t, E_{t+1}, ..., E_{t+c})].$$

Proof: We have $h(E_t, E_{t+1}, ..., E_{t+c}) = g^*(E_t)$, since $\{E_t\}$ is a deterministic sequence and therefore $E_{t+1} = f(E_t), ..., E_{t+c} = f^c(E_t)$.

Now, g^* is an integrable function and $\{E_t\}$ is an ergodic sequence; so, we may use the ergodic theorem.

We get:

$$\frac{1}{n}\sum_{t=1}^{n} h(E_t, E_{t+1}, ..., E_{t+c}) = \frac{1}{n}\sum_{t} g^*(E_t) \to^{a.s.} \mathrm{E}[g^*(E_t)] = \mathrm{E}[h(E_t, E_{t+1}, ..., E_{t+c})].$$

□

A few remarks are in order here. First, we can notice that the condition for h to be integrable is quite a weak one, especially if we remember that chaotic sequences are concentrated on chaotic attractors, which are compact sets (see section 2.1 for more details).

Now, an important comment to note is the role played in the theorem by the fact that $\{E_t\}$ is a deterministic sequence. If $\{E_t\}$ would not have been deterministic, then we would have been back to the ergodic theorem, that is $c = 0$. So, the theorem is an extension to the ergodic theorem and brings very strong results. In particular, the sample joint moments converge almost surely to the true values of the joint moments, for moments of any order.

Table 3.1 and Table 3.2 respectively show the means and the variances of simulations of the first ten lags of some (normalised, that is the variance of the sequences is made equal to 1) autocovariance functions. Table 3.1 indicates that the means converge to the theoretical values (the same as those pictured in Figure 3.3) and Table 3.2 reveals that the variances of the sample autocovariances tend to 0. Now, this is an indication of the fact that $\hat{\gamma}_{n,E}(k) \to^p \gamma_E(k)$ for $k \geq 1$, where $\to^p$ denotes the convergence in probability, for all the considered chaotic sequences (see Proposition 6.2.4 in Brockwell and Davis (1989)). Theorem 3.1 goes even further and tells us that the sample autocovariances converge almost surely to the theoretical autocovariances for all the considered chaotic sequences.

	Simulated mean n = 2000	Simulated mean n = 5000	Asymptotic mean
Modulo map			
Lag 1	0.497741	0.499386	0.500000
Lag 2	0.248095	0.248750	0.250000
Lag 3	0.122717	0.123948	0.125000
Lag 4	0.061427	0.061987	0.062500
Lag 5	0.030003	0.031003	0.031250
Lag 6	0.013649	0.015028	0.015625
Lag 7	0.006010	0.007675	0.007813
Lag 8	0.002913	0.002770	0.003906
Lag 9	0.000688	0.001404	0.001953
Lag 10	- 0.001060	0.000555	0.000977
Logistic map $\theta = 4.00$			
Lag 1	- 0.000941	- 0.000325	0.000000
Lag 2	0.000624	- 0.000092	0.000000
Lag 3	- 0.000727	- 0.000126	0.000000
Lag 4	- 0.000125	0.000310	0.000000
Lag 5	- 0.000972	- 0.000653	0.000000
Lag 6	- 0.001020	- 0.000160	0.000000
Lag 7	- 0.000301	- 0.000590	0.000000
Lag 8	- 0.001083	- 0.000302	0.000000
Lag 9	- 0.001272	- 0.000554	0.000000
Lag 10	- 0.000205	- 0.000275	0.000000

Table 3.1: Means of the sample estimators $\hat{\rho}_{E,n}(i)$ for $i = 1, 2, ..., 10$.
The simulated expectations are obtained using 2000 replications of $\hat{\rho}_{E,n}(i)$; the asymptotic means are the autocorrelations shown in Figure 3.3.

Table 3.1: continued

Logistic map $\theta = 3.98$ Lag 1	- 0.221728	- 0.221348	- 0.221000
Lag 2	- 0.129503	- 0.130129	- 0.130000
Lag 3	- 0.017341	- 0.018023	- 0.018000
Lag 4	0.074693	0.075358	0.075000
Lag 5	- 0.045181	- 0.044766	- 0.045000
Lag 6	0.065456	0.065408	0.066000
Lag 7	- 0.033570	- 0.032867	- 0.032000
Lag 8	- 0.015399	- 0.015458	- 0.016000
Lag 9	0.018175	0.018930	0.019000
Lag 10	0.001239	0.000314	0.000000
Logistic map $\theta = 3.825$ Lag 1	- 0.551449	- 0.551449	- 0.552000
Lag 2	- 0.014987	- 0.014882	- 0.016000
Lag 3	0.351581	0.351232	0.352000
Lag 4	- 0.233459	- 0.233741	- 0.234000
Lag 5	- 0.029524	- 0.029378	- 0.029000
Lag 6	0.200452	0.200379	0.200000
Lag 7	- 0.143747	- 0.144598	- 0.144000
Lag 8	- 0.007277	- 0.007664	- 0.008000
Lag 9	0.097111	0.096846	0.097000
Lag 10	- 0.081481	- 0.081677	- 0.082000
Logistic map $\theta = 3.58$ Lag 1	- 0.907325	- 0.907574	- 0.907616
Lag 2	0.818079	0.818574	0.818578
Lag 3	- 0.899827	- 0.900650	- 0.901047
Lag 4	0.978341	0.979511	0.980343
Lag 5	- 0.899740	- 0.901073	- 0.901785
Lag 6	0.819181	0.820676	0.821246
Lag 7	- 0.902655	- 0.904582	- 0.905610
Lag 8	0.987892	0.990266	0.991786
Lag 9	- 0.901291	- 0.903716	- 0.905074
Lag 10	0.815620	0.818105	0.819257

	Simulated variance n = 2000	Simulated variance n = 5000
Modulo map		
Lag 1	0.000375	0.000152
Lag 2	0.000636	0.000262
Lag 3	0.000816	0.000313
Lag 4	0.000800	0.000319
Lag 5	0.000815	0.000332
Lag 6	0.000792	0.000313
Lag 7	0.000808	0.000320
Lag 8	0.000836	0.000328
Lag 9	0.000792	0.000308
Lag 10	0.000810	0.000329
Logistic map $\theta = 4.00$		
Lag 1	0.000512	0.000207
Lag 2	0.000501	0.000212
Lag 3	0.000485	0.000198
Lag 4	0.000974	0.000695
Lag 5	0.000509	0.000198
Lag 6	0.000500	0.000201
Lag 7	0.000494	0.000199
Lag 8	0.000498	0.000198
Lag 9	0.000484	0.000200
Lag 10	0.000498	0.000201

Table 3.2: Variances of the sample estimators $\hat{\rho}_{E,n}(i)$ for $i = 1, 2, ..., 10$. The simulated variances are obtained using 2000 replications of $\hat{\rho}_{E,n}(i)$; the variances are expected to tend to 0 as $n \to \infty$.

Table 3.2 continued

Logistic map $\theta = 3.98$ Lag 1	0.000132	0.000053
Lag 2	0.000339	0.000136
Lag 3	0.000563	0.000226
Lag 4	0.000603	0.000245
Lag 5	0.000475	0.000193
Lag 6	0.000629	0.000263
Lag 7	0.000554	0.000213
Lag 8	0.000542	0.000203
Lag 9	0.000576	0.000236
Lag 10	0.000585	0.000223
Logistic map $\theta = 3.825$ Lag 1	0.000064	0.000026
Lag 2	0.000510	0.000201
Lag 3	0.001139	0.000428
Lag 4	0.000579	0.000233
Lag 5	0.000541	0.000225
Lag 6	0.000928	0.000363
Lag 7	0.000704	0.000275
Lag 8	0.000646	0.000253
Lag 9	0.000811	0.000320
Lag 10	0.000766	0.000307
Logistic map $\theta = 3.58$ Lag 1	0.000000	0.000000
Lag 2	0.000001	0.000000
Lag 3	0.000000	0.000000
Lag 4	0.000000	0.000000
Lag 5	0.000000	0.000000
Lag 6	0.000001	0.000000
Lag 7	0.000000	0.000000
Lag 8	0.000000	0.000000
Lag 9	0.000000	0.000000
Lag 10	0.000001	0.000000

As a last comment, we would like to stress again the importance of this theorem, which will be often used from now on, by pointing out the weakness of the required conditions compared to existing theorems (see, for example, Hall and Heyde (1980) which deals with the almost sure convergence of sample autocovariances). Finally, we can notice that Theorem 3.1 does not only apply to chaotic sequences but also to any deterministic ergodic sequence.

The next application of Theorem 3.1 will justify the use of simulations and histograms as a way to getting a good idea of the ergodic distribution. In particular, the last three graphs of Figure 3.2 are shown to be good approximations to the unknown invariant distributions.

The question we are going to answer is the following one: given a sample $E_1, ..., E_n$ obtained from a chaotic sequence, can we be sure that the sample distribution function

$$F_n(x) = \frac{1}{n}\sum_{j=1}^{n} I_{[E_j<x]}$$

(I is the indicator function, that is $I_{[E_j<x]} = 1$ if $E_j < x$, 0 otherwise) is close in some way to the distribution function $F(x)$ of E_1?

Glivenko-Cantelli's theorem states: " If $\{X_n, n \geq 1\}$ are *iid* random variables with distribution function F and F_n is the empirical distribution function based on $X_1, ..., X_n$, then

$$\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to^{a.s.} 0."$$

(see, for example, Chow and Teicher (1988)). Now, Györfi, Härdle, Sarda and Vieu (1989) relaxed the assumption of *iid* random variables and replaced it by the condition that $\{X_n\}$ is strongly mixing (see Definition 3.7) with a sufficiently fast mixing rate. We know from section 3.2 that no chaotic sequence $\{E_t\}$ is strongly mixing, so it is worthwhile for us to look at a possible Glivenko-Cantelli's theorem for chaotic sequences. In fact, such a theorem exists: it is stated and proved below. The proof mainly follows the proof in the *iid* case and makes use of Theorem 3.1 (more precisely of the simple ergodic theorem).

Theorem 3.2 *If* $\{E_n : n \geq 1\}$ *is a chaotic sequence (in the sense of Definition 3.2) with distribution function* F *and if* F_n *is the empirical distribution function based on* $E_1, ..., E_n$*, then*

$$\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \rightarrow^{a.s.} 0.$$

Proof: For every $x \in]-\infty, \infty[$,

$$Y_j = I_{[E_j<x]} \quad and \quad Z_j = I_{[E_j\leq x]}, \quad j \geq 1$$

constitute integrable (remember that we work on a probability space) functions of the random variable E_j.

By Theorem 3.1, we get

$$F_n(x) = \frac{1}{n}\sum_{j=1}^{n} Y_j \rightarrow^{a.s.} \mathrm{E}(Y_j) = F(x)$$

and

$$F_n(x+) = \frac{1}{n}\sum_{j=1}^{n} Z_j \rightarrow^{a.s.} \mathrm{E}(Z_j) = F(x+).$$

Now, we follow the theorem's proof for *iid* random variables (see Chow and Teicher (1988) for more details). Thus, if $C = \{c_j, j \geq 1\}$ = set of rational numbers enlarged by any irrational discontinuity points of F and

$A_1 = \{\omega : F_n^\omega(c_j\pm), j \geq 1\}$,

$A_2 = \{\omega : F_n^\omega(x)$ is a distribution function,$n \geq 1\}$,

it follows that $P(A_1) = 1$. Moreover, $P(A_2) = 1$ and so $P(A_1 \cap A_2) = 1$. Since for $\omega \in A_1 \cap A_2$, $\{F_n^\omega(x) : n \geq 1\}$ is a sequence of distribution functions with $F_n^\omega(c_j\pm) \rightarrow F(c_j\pm)$, $j \geq 1$, we can be sure that the random variable E_n^ω associated to F_n^ω converges in distribution to E_1 for $\omega \in A_1 \cap A_2$ whence $F_n^\omega(x)$ converges uniformly to $F(x)$ for $\omega \in A_1 \cap A_2$ (there is a lemma in Chow and Teicher (1988) telling: " If X_n converges in distribution to X and the associated distribution function $F_n(x\pm)$ converges to $F(x\pm)$ at all discontinuity points x of F, then F_n converges uniformly to F in $]-\infty, \infty[)$. $\square$

Remark that Theorem 3.2 applies in the same way to any ergodic (not necessarily deterministic) sequence since we simply use the ergodic theorem in the proof.

To finish this section, we illustrate Theorem 3.2 by means of Figure 3.5, which shows the empirical probability distribution functions constructed from samples of different sizes taken from the logistic map with $\theta = 4.00$ and compares them with the true probability distribution functions. For large sample sizes, the empirical and the theoretical probability densities are similar.

3.4 Lyapunov exponents of chaos driven systems

We assume throughout this section that we work on a probability space. Before considering chaos driven systems, we would like to come back to Proposition 3.2, such that if $\{E_t\}$ is a chaotic sequence, then the sequence $\{g(E_t)\}$ is a K-system (if g is a measurable function). This does not mean that $\{g(E_t)\}$ is a chaotic sequence (in the sense of Definition 3.2).

If g is invertible (that is g^{-1} exists), then $g(E_{t+1}) = gfg^{-1}g(E_t)$ and gfg^{-1} generates a chaotic sequence (in the sense of Definition 3.2) by Propositions 3.2 and 3.1; in particular, $\{g(E_t)\}$ has a positive Lyapunov exponent.

Now, if g is not invertible, then we do not deal with a dynamical system any more and the notion of Lyapunov exponent (as defined in section 3.1) has no meaning in this case.

In the introduction and in section 3.2, we have stressed the importance of chaos driven models. Before looking at the statistical inference on such models in the next two chapters, it is interesting to compute the Lyapunov exponents of some chaos driven models in view to seeing how close such models are to their

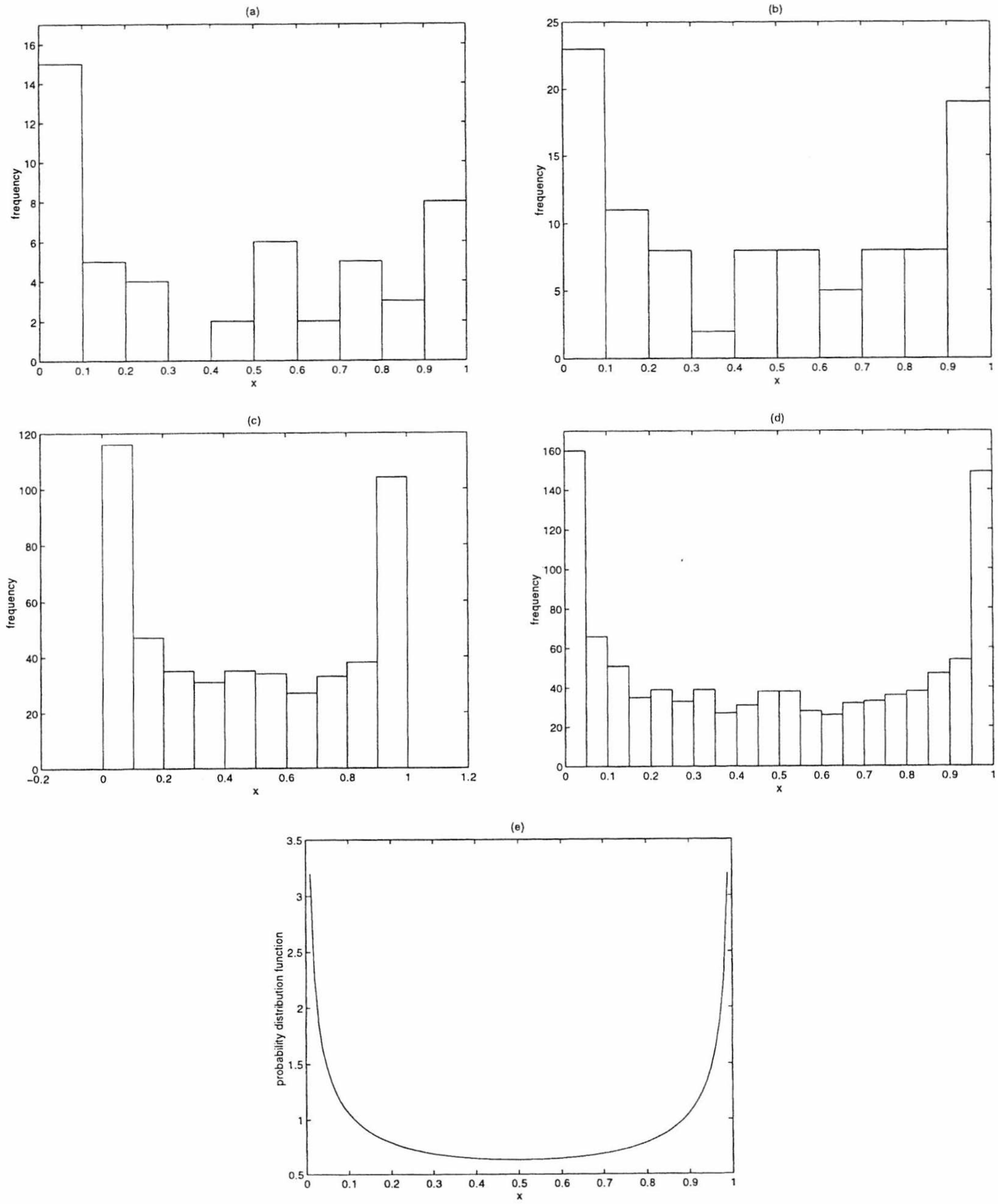


Figure 3.5: Empirical probability distribution functions of the logistic map with $\theta = 4.00$.
The sample size is (a) 50, (b) 100, (c) 500, (d) 1000. The graph (e) shows the theoretical p.d.f. .

counterparts with stochastically generated noise.

Consider the following chaos driven non linear model:

$$X_t = g(X_{t-1}) + E_t,$$

where $\{E_t\}$ is a chaotic sequence. The (positive) Lyapunov exponent for $\{E_t\}$ is denoted λ_f; moreover, we assume that the Lyapunov exponent exists for a dynamical system generated by g, we denote it λ_g.

We have here a deterministic dynamical system on R^2:

$$M : R^2 \to R^2 : \begin{pmatrix} X_{t-1} \\ E_{t-1} \end{pmatrix} \to \begin{pmatrix} X_t \\ E_t \end{pmatrix} : X_t = g(X_{t-1}) + f(E_{t-1}), \quad E_t = f(E_{t-1}).$$

Now, as in section 3.1, we compute $DM^n(x_0)$, the Jacobian matrix of M^n evaluated at x_0.

$$DM^n(x_0) = \begin{pmatrix} \frac{\partial g}{\partial x}(x_{n-1}) & \frac{\partial f}{\partial e}(e_{n-1}) \\ 0 & \frac{\partial f}{\partial e}(e_{n-1}) \end{pmatrix} \cdots \begin{pmatrix} \frac{\partial g}{\partial x}(x_0) & \frac{\partial f}{\partial e}(e_0) \\ 0 & \frac{\partial f}{\partial e}(e_0) \end{pmatrix}$$

So,

$$DM^n(x_0) = \begin{pmatrix} \frac{\partial g}{\partial x}(x_{n-1}) \ldots \frac{\partial g}{\partial x}(x_0) & * \\ 0 & \frac{\partial f}{\partial e}(e_{n-1}) \ldots \frac{\partial f}{\partial e}(e_0) \end{pmatrix}$$

and, for large n, the eigenvalues of $DM^n(x_0)$ are very similar to $e^{n\lambda_f}$ and $e^{n\lambda_g}$. So, in the same way as in section 3.1, the Lyapunov exponents of M are λ_f and λ_g.

We see that at least one of the Lyapunov exponents, namely λ_f, is positive and so the map M is chaotic. Now, $\lambda_f + \lambda_g > 0$ would mean that on average areas are stretched by M; on the other hand, $\lambda_f + \lambda_g < 0$ would imply that on average areas are contracted by M.

Consider an n-dimensional chaotic attractor with Lyapunov exponents $\lambda_1 \geq$

$\lambda_2 \geq ...\lambda_n$. We can introduce the following notion (see, for example, Ott (1993) for more details).

Definition 3.10 *Let K be the largest integer such that $\sum_{j=1}^{K} \lambda_j \geq 0$.*

The **Lyapunov dimension** *of the chaotic attractor is then defined as*

$$D_L = K + \frac{1}{|\lambda_{k+1}|} \sum_{j=1}^{K} \lambda_j.$$

The interest of this notion comes from the fact that it has been conjectured by Kaplan and Yorke (1979) that the Lyapunov dimension is the same as the so-called information dimension (fractal dimension; see, for example, Ott (1993) for the definition and more details) of the attractor for 'typical' attractors'. At present, the conjecture seems to be true but no formal proof has been given yet. Now, if the conjecture would effectively be true, then it would mean that the fractal dimension of an attractor could be given in terms of its Lyapunov exponents.

Coming back to our chaos driven non-linear model, let us deal with a particular case which will be analysed in the next chapter, namely the causal AR(1) driven by chaos $X_t = \alpha X_{t-1} + E_t$, $|\alpha| < 1$. In this case, the function $g : R \to R$ is simply $g(x) = \alpha x$ and $\lambda_g = ln|\alpha| < 0$ ($\lambda_g < 0$ is not a surprise since g has an attracting point $\{0\}$). The two Lyapunov exponents are $\lambda_1 = \lambda_f$ and $\lambda_2 = ln|\alpha|$.

Note that, for a fixed chaotic map f, as $\alpha \to 1$, the stretching process of the map M (that is the action of f) dominates the contracting process (attraction to $\{0\}$): $\lambda_1 + \lambda_2 > 0$. On the other hand, if $\alpha \to 0$, the contracting process becomes predominant: $\lambda_1 + \lambda_2 < 0$.

For our map,

$$D_L = \begin{cases} 2, & \text{if } \lambda_f + ln|\alpha| \geq 0 \\ 1 + \frac{\lambda_f}{|ln|\alpha||} = 1 - \frac{\lambda_f}{ln|\alpha|}, & \text{if } \lambda_f + ln|\alpha| < 0 \end{cases}$$

Now, for chaotic sequences $\{E_t\}$ generated by the logistic map with $\theta = 4.00$

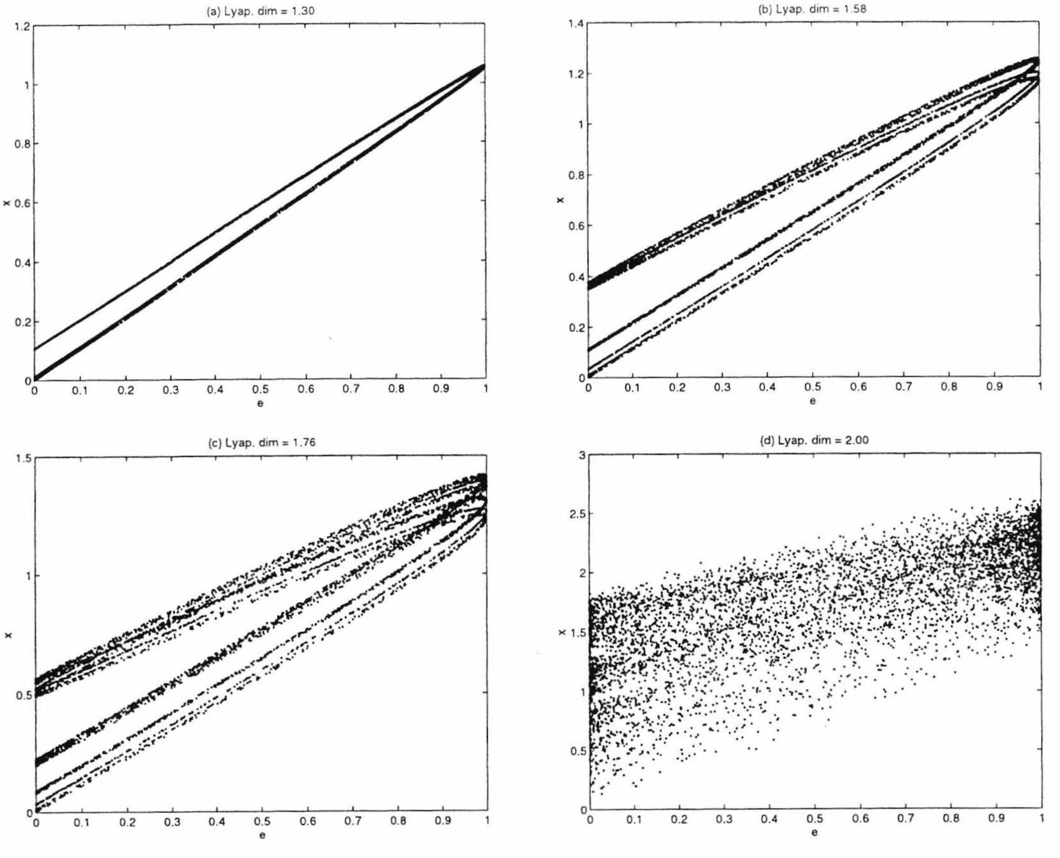


Figure 3.6: Attractors of the 2-dimensional map

$$X_t = \alpha X_{t-1} + E_t, \quad E_t = 4E_{t-1}(1 - E_{t-1})$$

with (a) $\alpha = 0.1$, (b) $\alpha = 0.3$, (c) $\alpha = 0.4$, (d) $\alpha = 0.7$.

(Lyapunov exponent $\lambda_f = ln\ 2$) and for different values for α, we had a look at the resulting attractor and we computed the corresponding Lyapunov dimensions. This is illustrated in Figure 3.6.

Chapter 4

Statistical inference on chaos driven AR models

Sections 4.1, 4.2 and 4.3 closely follow Stockis and Tong (1996); they all deal with the statistical properties of the Yule-Walker estimators in chaos driven AR(1) models. Section 4.1 concentrates on the asymptotic bias, section 4.2 on the asymptotic normality and section 4.3 presents simulations concerning the first two sections. Section 4.4 gives the asymptotic results for the parameter estimators in the case of chaos driven AR(p) models, $p \geq 1$. Section 4.5 illustrates the results of section 4.4 by means of examples and simulations. Finally, section 4.6 introduces AR models driven by noisy chaos and gives results on the asymptotic normality of the autoregressive parameter estimators.

4.1 Asymptotic bias of the Yule-Walker estimator for an AR(1)

We consider a causal AR(1) model driven by chaos: $X_{t+1} = \alpha X_t + E_{t+1}$, that is $|\alpha| < 1$ (in particular, the AR(1) model is stationary) and $\{E_t\}$ is a chaotic sequence, thus ergodic (see section 3.2). Moreover, E_t is assumed to have zero mean and finite variance.

We shall be interested in the well-known Yule-Walker estimators of α, namely $\hat{\alpha}_n = \sum_{t=1}^{n-1} X_t X_{t+1} / \sum_{t=1}^{n} X_t^2$ and $\hat{\alpha_n}' = \sum_{t=1}^{n-1} X_t' X_{t+1}' / \sum_{t=1}^{n} X_t'^2$ of α, where $X_t' = X_t - \bar{X}_n$ and $\bar{X}_n$ denotes the sample mean.

For an AR(1) model with *iid* noise $\{\varepsilon_t\}$, these estimators are asymptotically unbiased and $\hat{\alpha}_n \to^p \alpha$ and $\hat{\alpha}_n' \to^p \alpha$. We consider the same estimators in the present set-up because we want to know if this asymptotic property is preserved when the *iid* noise sequence is replaced by a chaotic sequence. We note that our results still hold if, keeping the other assumptions, we only require f to be a real deterministic map and we assume that E_t admits an invariant distribution which is also ergodic.

Let $\rho_E(i)(\gamma_E(i))$ denote the autocorrelation (autocovariance) at lag i of the sequence $\{E_t\}$, $\xi = \sum_{i=1}^{\infty} \alpha^{i-1} \rho_E(i)$, $\delta = 1 + 2\alpha\xi$ and $c = (1-\alpha^2)/\delta$, where it is assumed that $\delta \neq 0$.

Theorem 4.1

$$\hat{\alpha}_n \to^p \alpha + c\xi.$$

Proof :

$$\hat{\alpha}_n = \frac{\frac{1}{n}\sum_{t=1}^{n-1} X_t X_{t+1}}{\frac{1}{n}\sum_{t=1}^{n} X_t^2} = \alpha + \frac{\frac{1}{n}\sum_{t=1}^{n-1} E_{t+1} X_t}{\frac{1}{n}\sum_{t=1}^{n} X_t^2} + O_p(\frac{1}{n}).$$

First, denote $\frac{1}{n}\sum_{t=1}^{n} E_{t+u} E_{t+v}$ by $\hat{\gamma}_{n,E}(u-v)$ and $\hat{\gamma}_{n,E}(j)/\hat{\gamma}_{n,E}(0)$ by $\hat{\rho}_{n,E}(j)$. Now,

$$\begin{aligned}
& \frac{1}{n}[\sum_{t=1}^{n-1} E_{t+1} X_t] \\
= & \frac{1}{n}[\sum_{t=1}^{n-1} E_{t+1} E_t + \alpha \sum_{t=2}^{n-1} E_{t+1} E_{t-1} + \alpha^2 \sum_{t=3}^{n-1} E_{t+1} E_{t-2} + \ldots] \\
= & \hat{\gamma}_{n,E}(1) + \alpha \hat{\gamma}_{n,E}(2) + \alpha^2 \hat{\gamma}_{n,E}(3) + \ldots + O_p(\frac{1}{n}).
\end{aligned}$$

Similar calculations give

$$\frac{1}{n}\sum_{t=1}^{n} X_t^2 = \frac{1}{1-\alpha^2}[\hat{\gamma}_{n,E}(0) + 2\alpha\hat{\gamma}_{n,E}(1) + 2\alpha^2\hat{\gamma}_{n,E}(2) + ...] + O_p(\frac{1}{n}).$$

By Theorem 3.1, we may say that $\hat{\gamma}_{n,E}(i) \rightarrow^{a.s.} \gamma_E(i)$, $i = 0, 1, ...$ Hence $\hat{\rho}_{n,E}(i) \rightarrow^p \rho_E(i)$, $i = 1, 2, ...$ since for all $i \geq 1$, $\hat{\gamma}_{n,E}(i) \rightarrow^d \gamma_E(i)$ and $\hat{\gamma}_{n,E}(0) \rightarrow^d \gamma_E(0)$, so $\hat{\rho}_{n,E}(i) \rightarrow^d \rho_E(i)$, which implies that $\hat{\rho}_{n,E}(i) \rightarrow^d \rho_E(i)$, by using Propositions 6.3.8 and 6.3.5 of Brockwell and Davis (1989).

We conclude by noting that $h : x \rightarrow (1-\alpha^2)x/(1+2\alpha x)$ is a continuous mapping, so Proposition 6.3.4 of Brockwell and Davis (1989) applies. □

Corollary 4.1

$$\hat{\alpha}'_n \rightarrow^p \alpha + c\xi.$$

Proof : It is enough to show that $\hat{\alpha}_n - \hat{\alpha}'_n \rightarrow^p 0$. This is easily verified since $\bar{X}_n \rightarrow^{a.s.} 0$ by the ergodic theorem (see section 3.3) and so $(\bar{X}_n)^2 \rightarrow^p 0$ ($h : x \rightarrow x^2$ is a continuous mapping). □

For all the chaotic sequences (with finite variance), we have been able to quantify the asymptotic bias of the Yule-Walker estimators. These estimators are not necessarily asymptotically unbiased: in fact, the autocorrelation function for $\{E_t\}$ plays a vital role for the asymptotic bias of the estimators. Moreover, if there is bias, the value of α comes into the bias term. Examples and simulations will be given in section 4.3. Now, we shall turn our attention to the asymptotic normality.

4.2 Asymptotic normality of the Yule-Walker estimator for an AR(1)

As in section 4.1, we consider an AR(1) model driven by a zero mean chaotic sequence $\{E_t\}$ with finite variance: $X_{t+1} = \alpha X_t + E_{t+1}$, $|\alpha| < 1$.

For a causal AR(1) model with *iid* noise $\{\varepsilon_t\}$, it is well known that the Yule-Walker estimators $\hat{\alpha}_n$ and $\hat{\alpha}'_n$ have central limit properties. In fact,

$$n^{\frac{1}{n}}(\alpha_n^{(\prime)} - \alpha) \to^d \mathcal{N}(0, 1 - \alpha^2).$$

Is this asymptotical normality preserved for chaos-driven models? Unlike the asymptotic bias, we shall not answer this question for all chaotic sequences but we shall give sufficient conditions on the chaotic sequence $\{E_t\}$ under which the asymptotic normality is preserved. As a by-product, we shall get the theoretical asymptotic variance, which will be, as it may be expected, often different from the *iid* case.

To study the asymptotic distributions of the estimators $\hat{\alpha}_n$ and $\hat{\alpha}'_n$, one way is to appeal to the central limit properties of the sample autocorrelation function which hold under some conditions. One commonly used condition is a linear representation of X_t in terms of *iid* random variables Z_t's with finite variance:
" If $\{X_t\}$ is the stationary process

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim iid(0, \sigma^2)$$

where $\mathrm{E}(X_t) = \mu$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $\sum_{j=-\infty}^{\infty} |j|\psi_j^2 < \infty$, then
for each $h \in \{1, 2, ...\}$, $\hat{\rho}(h)$ is asymptotically $\mathcal{N}(\rho(h), n^{-1}W)$, where W is the covariance matrix whose (i,j)-element is given by Bartlett's formula,

$$\begin{aligned} W &= (w_{ij})_{i=1,\dots,j=1,\dots} \\ &= \Big(\sum_{k=-\infty}^{\infty} \{\rho_E(k+i)\rho_E(k+j) + \rho_E(k-i)\rho_E(k+j) + 2\rho_E(i)\rho_E(j)\rho_E^2(k) \\ &\quad -2\rho_E(i)\rho_E(k)\rho_E(k+j) - 2\rho_E(j)\rho_E(k)\rho_E(k+i)\}\Big)." \end{aligned}$$

(see, e.g., Theorem 7.2.2 of Brockwell and Davis (1989)).

This is clearly not directly applicable to our case because $X_t = \sum_{j=0}^{\infty} \alpha^j E_{t-j}$, where the E_j's are not mutually independent in view of the fact that $E_{t+1} = f(E_t)$. However, the technique is still useful indirectly as we shall now show.

Note that it is possible to relax the assumptions on the Z_t's and basically require them to be uncorrelated and that $\mathrm{E}(Z_t|Z_{t-1}, Z_{t-2}, ...) = 0$ almost surely for all t (see, e.g., Hall and Heyde (1980)). However, the last condition is not satisfied by our chaotic sequences $\{E_t\}$ since $\mathrm{E}(E_t|E_{t-1}, E_{t-2}, ...) = f(E_{t-1})$.

Now, some deterministic dynamical systems are Bernoulli systems. This means that they can be represented as a symbolic dynamics consisting of a full shift on a finite number of symbols (following Definition 3.6). So, the chaotic sequences $\{E_t\}$ generated by such systems admit the linear representation $E_t - \mathrm{E}(E_t) = \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i}$, where $\{Z_t\}$ is a set of *iid* random variables with zero mean and finite variance (see, for example, Billingsley (1965) or Ott (1993) for more details). An example of a Bernoulli system is given by the modulo map $f : x \to 2x$ (mod 1) from $[0, 1]$ to itself (see sections 2.1 and 3.2). The following theorem gives the central limit properties for the Yule-Walker estimators of AR(1) models driven by such sequences $\{E_t\}$.

Theorem 4.2 *If E_t can be written as*

$$E_t - \mu = \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i},$$

where $\mu = \mathrm{E}(E_t)$, $\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$, $\sum_{i=-\infty}^{\infty} |i|\psi_i^2 < \infty$, and $\{Z_t\}$ is a set of iid random variables with $\mathrm{E}(Z_t) = 0$ and $\mathrm{E}(Z_t^2) = \sigma^2 < \infty$, then

$$n^{\frac{1}{2}}(\hat{\alpha}_n - \alpha - c\xi) \to^d \mathcal{N}(0, \delta^{-2}c^2 a^T W a),$$

where $a^T = (1, \alpha, \alpha^2, ...)$ and

$$\begin{aligned} W &= (w_{ij})_{i=1,...,j=1,...} \\ &= \Big(\sum_{k=-\infty}^{\infty} \{\rho_E(k+i)\rho_E(k+j) + \rho_E(k-i)\rho_E(k+j) + 2\rho_E(i)\rho_E(j)\rho_E^2(k) \end{aligned}$$

$$-2\rho_E(i)\rho_E(k)\rho_E(k+j) - 2\rho_E(j)\rho_E(k)\rho_E(k+i)\}).$$

Proof :

$$\hat{\alpha}_n - \alpha \simeq \frac{\frac{1}{n}\sum E_t X_{t-1}}{\frac{1}{n}\sum X_t^2}.$$

Here and later the symbol '$\simeq$' means 'equal up to $O_p(\frac{1}{n})$'. Now, let

$$Z_n = \frac{\frac{1}{n}\sum E_t X_{t-1}}{\frac{1}{n}\sum E_t^2}, \text{ and } Y_n = \frac{\frac{1}{n}\sum X_t^2}{\frac{1}{n}\sum E_t^2}.$$

Then

$$Z_n \simeq R_n = \hat{\rho}_{E,n-1}(1) + \alpha\hat{\rho}_{E,n-2}(2) + \alpha^2\hat{\rho}_{E,n-3}(3) + \ldots$$

and R_n is asymptotically $\mathcal{N}(\xi, \frac{1}{n}a^T W a)$ since for each j, the asymptotic joint distribution of $(\hat{\rho}_E(1), ..., \hat{\rho}_E(j))$ is normal (by Theorem 7.2.2 of Brockwell and Davis (1989), stated earlier in this section) The result then follows easily by using the characteristic functions or equivalently Proposition 6.3.9 of Brockwell and Davis (1989). Now,

$$Y_n \simeq \frac{1 + 2\alpha R_n}{1 - \alpha^2}$$

and therefore $\hat{\alpha}_n - \alpha \simeq g(R_n)$ with g, differentiable at ξ, defined as

$$g(\zeta) = \frac{(1-\alpha^2)\zeta}{1 + 2\alpha\zeta}.$$

Thus, by Proposition 6.4.1 of Brockwell and Davis (1989),

$$n^{\frac{1}{2}}(\hat{\alpha}_n - \alpha - c\xi) \rightarrow^d \mathcal{N}(0, \delta^{-2}c^2 a^T W a),$$

which concludes the proof. □

Corollary 4.2 *Under the conditions of Theorem 4.2,*

$$n^{\frac{1}{2}}(\hat{\alpha}_n' - \alpha - c\xi) \to^d \mathcal{N}(0, \delta^{-2}c^2 a^T W a).$$

Proof : It is enough to show that $\sum_{i=-\infty}^{\infty} |cov(X_t, X_{t+i})| < \infty$. Now,

$$\begin{aligned} & \sum_{i=-\infty}^{\infty} |cov(X_t, X_{t+i})| \\ = & \sum_{i=-\infty}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \alpha^j \alpha^k |cov(E_{t-j}, E_{t+i-k})| \\ = & \sum_{j=0}^{\infty} \alpha^j \sum_{k=0}^{\infty} \alpha^k \sum_{i=-\infty}^{\infty} |cov(E_{t-j}, E_{t+i-k})|. \end{aligned}$$

Further, $\sum_{i=-\infty}^{\infty} |cov(E_{t-j}, E_{t+i-k})|$ = a finite constant, independent of j and k, due to the linear representation of E_t and $\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$. □

Now, there exist similar propositions to Theorem 4.2. For example, we can trade off some of the assumptions on the sequence $\{\psi_t\}$ with a finite fourth moment assumption on $\{Z_t\}$.

Theorem 4.3 *If E_t can be written as $E_t - \mu = \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i}$, where $\mu = \mathrm{E}(E_t)$, $\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$ and $\{Z_t\}$ is a set of iid random variables with $\mathrm{E}(Z_t) = 0$ and $\mathrm{E}(Z_t^4) < \infty$, then*

$$n^{\frac{1}{2}}(\hat{\alpha}_n - \alpha - c\xi) \to^d \mathcal{N}(0, \delta^{-2}c^2 a^T W a),$$

where a and W are as in Theorem 4.2.

Proof : The proof is similar to the proof of Theorem 4.2 but this time Theorem 7.2.1 in Brockwell and Davis (1989) is used. □

Corollary 4.3 *Under the conditions of Theorem 4.3,*

$$n^{\frac{1}{2}}(\hat{\alpha}_n' - \alpha - c\xi) \to^d \mathcal{N}(0, \delta^{-2}c^2a^TWa).$$

Proof: The proof is identical to that for Corollary 4.2. □

An example which meets the conditions of Theorems 4.2 and 4.3 is given by the above-mentioned map: $E_{t+1} = 2E_t \pmod 1$, which we shall analyse in the next section but for now we enlarge the class of chaotic maps such that the asymptotic normality for $\hat{\alpha}_n$ still holds. To this end, we need to introduce the notion of U-statistics.

Definition 4.1 *Let $Y_1, Y_2, ...$ be a strictly stationary stochastic process taking values in R^d with distribution F. Let $h : \underbrace{R^d \times ... \times R^d}_{m \text{ times}} \to R$ be measurable and symmetric in its m arguments; h is called the kernel for*

$$\theta = \int ... \int h(y_1, ..., y_m) \prod_{i=1}^{n} dF(y_i).$$

Then a **U-statistic** *U_n is given by*

$$U_n = \begin{pmatrix} n \\ m \end{pmatrix}^{-1} \sum_{1 \leq t_1 < t_2 < ... < t_m \leq n} h(Y_{t_1}, ..., Y_{t_m}) \quad (n \geq m).$$

We shall focus here on kernels of the type $h : R \to R$ and so $\theta = \int h(y_1)dF(y_1) = \mathrm{E}_F(h(y_1))$.

We shall appeal to Denker and Keller (1986), which provides asymptotic results including central limit theorems for U-statistics of sufficiently well-behaved functionals of an absolutely regular stationary process (see Definition 3.8 for the definition of an absolutely regular process). Here is Theorem 1 of Denker and Keller (1986) (for the case in which the kernel h goes from R to R): " The sequence

U_n of U-statistics converges to θ in probability, and the sequence $n^{1/2}(U_n - \theta)$ converges in distribution to the normal distribution with expectation 0 and variance $\sigma^2 = \mathrm{E}_F[h(X_1)^2] - \theta^2 + 2\sum_{r\geq 2} \mathrm{E}[(h(X_1) - \theta)(h(X_r) - \theta)]$, provided the following conditions hold.

1. h satisfies the Lipschitz condition: there are $L > 0$, $r \geq 0$ and $\rho > 0$ such that

   $|h(x) - h(y)| \leq |x - y|^\rho (1 + |x|^r + |y|^r)$, $x, y \in R$.

2. $\{Z_n : n \geq 1\}$ is an absolutely regular stationary sequence with $\beta(n)^{\frac{\eta}{2+\eta}} = O(n^{-2-\varepsilon})$ for some $\varepsilon, \eta > 0$.

3. $\{X_n : n \geq 1\}$ is a Lipschitz functional of $\{Z_n : n \geq 1\}$: we assume that there exists a function $g(u_1, u_2, ...)$ satisfying $X_j = g(Z_j, Z_{j+1}, ...)$ for $j \geq 1$ and g is Lipschitz-continuous in the sense that there are some $\alpha < 1$ and $c < \infty$ such that

   $|g(z_1, z_2, ...) - g(z'_1, z'_2, ...)| \leq c\alpha^n$ if $z_1 = z'_1, ..., z_n = z'_n$. ".

Now, by Denker and Keller (1986), some dynamical systems $\{E_t\}$, in particular the so-called interval transformations, can be regarded as functionals of absolutely regular processes $\{Z_t\}$ and many of the more complicated systems behave very much like an interval transformation . The next theorem follows from Theorem 1 of Denker and Keller (1986).

Theorem 4.4 *Let $X_{t+1} = \alpha X_t + E_{t+1}$ and $\hat{\alpha}_n$ be defined as before. Suppose that the following conditions hold.*

1. *$\{E_n : n \geq 1\}$ is a Lipschitz functional of $\{Z_n : n \geq 1\}$. Specifically there exists a function $g(u_1, u_2, ...)$ satisfying $E_j = g(Z_j, Z_{j+1}, ...)$ for $j \geq 1$ and there are some $\lambda < 1$ and $c < \infty$ such that*

$$|g(z_1, z_2, ...) - g(z'_1, z'_2, ...)| \leq c\lambda^n \text{ if } z_1 = z'_1, ..., z_n = z'_n.$$

2. *$\{Z_n, n \geq 1\}$ is an absolutely regular stationary sequence with $\beta(n)^{\frac{\eta}{2+\eta}} = O(n^{-2-\epsilon})$ for some $\epsilon, \eta > 0$.*

3. *$h(E_t) = E_t[\sum_{i=1}^{\infty} \alpha^{i-1} f^i(E_t)]$ satisfies the Lipschitz condition. Specifically, there are $L > 0$, $r \geq 0$ and $\rho > 0$ such that*

$$|x[\sum_{i=1}^{\infty} \alpha^{i-1} f^i(x)] - y[\sum_{i=1}^{\infty} \alpha^{i-1} f^i(y)]| \leq L|x-y|^{\rho}(1+|x|^r+|y|^r)$$

for all x and y belonging to the support of E_t.

Then, $\hat{\alpha}_n$ is asymptotically $\mathcal{N}(\alpha + c\theta\gamma_E^{-1}(0), c^2\sigma^2/(n\gamma_E^2(0)\delta^2))$, where $\theta = \mathrm{E}_F[h(E_1)]$ with respect to the invariant distribution F for $\{E_t\}$, $\sigma^2 = \mathrm{E}_F[h(E_1)]^2 - \theta^2 + 2\sum_{r\geq 2} \mathrm{E}_F[(h(E_1)-\theta)(h(E_r)-\theta)]$ and c and δ are as defined in Theorem 4.1.

Remarks: (i) Condition 3 is met if f is continuously differentiable. (ii) There are maps which satisfy the conditions of Theorems 4.2 and 4.3 but not those of Theorem 4.4. One such map is given by $E_t = 2E_{t-1}$ (mod 1).

Proof :

$$\hat{\alpha}_n - \alpha \simeq \frac{\frac{1}{n}\sum E_t X_{t-1}}{\frac{1}{n}\sum X_t^2}.$$

Let $Z_n = \frac{1}{n}\sum E_t X_{t-1}$ and $Y_n = \frac{1}{n}\sum X_t^2$. Now,

$$Z_n \simeq \frac{\sum_{i=1}^{n-1} E_i E_{i+1}}{n} + \alpha\frac{\sum_{i=1}^{n-2} E_i E_{i+2}}{n} + \alpha^2\frac{\sum_{i=1}^{n-3} E_i E_{i+3}}{n} + \ldots$$

and

$$E_{i+1} = f(E_i), E_{i+2} = f^2(E_i),$$

and so on. Therefore,

$$Z_n \simeq \frac{1}{n}\sum_{i=1}^{n-1} E_i[f(E_i) + \alpha f^2(E_i) + \ldots].$$

Also, $Z_n \simeq \frac{1}{n}\sum_{i=1}^{n-1} E_i[\sum_{j=1}^{\infty} \alpha^{j-1} f^j(E_i)]$. Now, Z_n is, up to order $1/n$, equal to a U-statistic whose associated kernel is $h : x \to x[\sum_{i=1}^{\infty} \alpha^{i-1} f^i(x)]$. Thus, by Theorem 1 of Denker and Keller (1986), we get that Z_n is asymptotically $\mathcal{N}(\theta, \sigma^2/n)$, where $\theta = \mathrm{E}_F[h(E_1)]$ and

$$\sigma^2 = \mathrm{E}_F[h(E_1)]^2 - \theta^2 + 2\sum_{r\geq 2} \mathrm{E}_F[(h(E_1) - \theta)(h(E_r) - \theta)].$$

So, $\{\gamma_E(0)\}^{-1} Z_n \simeq \mathcal{N}(\theta/\gamma_E(0), \sigma^2/(n\gamma_E^2(0))$. Now, we can use the fact that $\{\gamma_E(0)\}^{-1} Y_n \simeq (1 + 2\alpha\{\gamma_E(0)\}^{-1} Z_n)/(1-\alpha^2)$ in a similar way to the proof of Proposition 2.2 to get the result. □

Note that the well-known logistic map $E \to 4E(1-E)$, $E \in [0,1]$ is an example of dynamical systems which satisfy the conditions required in the above theorem.

Lemma 4.1 *Let $X_{t+1} = \alpha X_t + E_{t+1}$ be defined as before. Suppose that the following conditions hold.*

1. *$\{E_n, n \geq 1\}$ is a Lipschitz functional of $\{Z_n, n \geq 1\}$ in the sense of Theorem 4.4.*
2. *$\{Z_n, n \geq 1\}$ is an absolutely regular stationary sequence with $\beta(n)^{\frac{\eta}{2+\eta}} = O(n^{-2-\epsilon})$ for some $\epsilon, \eta > 0$.*

Then $\bar{X}_n$ is asymptotically $\mathcal{N}(0, v^2/\{(1-\alpha)^2 n\})$,
where $v^2 = var(E_1) + 2\sum_{r\geq 2} cov(E_1, E_r)$.

Proof :

$$\begin{aligned}
\bar{X}_n &= \frac{E_n + ... + E_2 + E_1 + \alpha E_{n-1} + ... + \alpha E_1 + ...}{n} \\
&\simeq \frac{(1 + \alpha + ... + \alpha^{n-1})}{n} E_1 + ... + \frac{(1 + \alpha + ... + \alpha^{n-1})}{n} E_n \\
&\simeq \frac{1}{n}\sum_{i=1}^{n}(\sum_{j=0}^{\infty} \alpha^j) E_i = \frac{1}{n}\frac{1}{1-\alpha}\sum_{i=1}^{n} E_i.
\end{aligned}$$

So the function $h : x \to x$ can be seen as the suitable kernel. We can apply Theorem 1 of Denker and Keller (1986) and deduce that $\bar{X}_n$ is asymptotically $\mathcal{N}(0, v^2/\{(1-\alpha)^2 n\})$, where $v^2 = var(E_1) + 2\sum_{r\geq 2} cov(E_1, E_r)$. □

Corollary 4.4 *Under the same conditions as in Theorem 4.4, $\hat{\alpha}'_n$ is asymptotically $\mathcal{N}(\alpha + c\theta/\gamma_E(0), c^2\sigma^2/(n\gamma_E^2(0)\delta^2))$.*

Proof: It is enough to prove that $\lim_{n\to\infty} n\mathrm{Var}(\bar{X}_n) < \infty$ which follows from the above lemma. □

We shall illustrate the above asymptotic results with numerical examples in the next section.

4.3 Simulations

The purpose of our simulations is to illustrate the theoretical results of sections 4.1 and 4.2 by means of sequences $\{E_t\}$ generated by some maps introduced earlier in this thesis. As bench-marks, we include two pseudo-random number generators from the NAG Fortran library, namely $G05DAF$ (for uniform distribution) and $G05DDF$ (for Gaussian distribution), both of which are based on chaotic maps of the same form as G05CAF (see section 3.2 for more details) . Other maps include (i) $f(E_t) = 2E_t$ (mod 1), the so-called modulo map, which fulfils the conditions of Theorems 4.2 and 4.3 because

$$E_t - \frac{1}{2} = \sum_{j=-\infty}^{0} 2^{j-1} Z_{t-j},$$

where, for each t, $Z_t = -0.5$ or 0.5 with equal probability, (ii) the logistic map $E_{t+1} = \theta E_t(1-E_t)$, $0 \leq E_t \leq 1$, $\theta = 4.00$, which fulfils the conditions of Theorem 4.4 (Denker and Keller (1986)) and (iii) three other logistic maps, namely $\theta = 3.98$, $\theta = 3.825$ and $\theta = 3.58$, for the first of which the autocovariance appears to decay quite quickly, for the second of which quite slowly and for the last of which not at all (see Figure 3.3). These three maps satisfy the requirements of Theorem 4.1 (like all the other maps included in this section) but do not fulfil the conditions of either Theorem 4.2, 4.3 or 4.4.

For the simulations, we standardize every chaotic sequence $\{E_t\}$ to zero mean and unit variance. A minor remark is in order here: for the map $x \mapsto 2x(mod 1)$, we replaced 2 by 1.99999 in the simulation in order to avoid degeneracy due to finite precision arithmetic.

Figure 4.1 compares, on the one hand, the means of observed $\hat{\alpha}'_n$ and those based on our theory. The results appear to be in accord with Theorem 4.1 and Corollary 4.1. It also provides with the asymptotic (theoretical) means for all the possible values of α. We can notice the role of both the noise autocorrelation function and the true value of α in the determination of the bias. In particular, we get a positive bias for the models driven by the modulo map since the modulo map has a positive autocorrelation function and we get a negative bias for the models driven by the last three logistic maps since the first lag of their autocorrelation functions is (clearly) negative and predominant in the computation of the bias.

Figure 4.2 gives the sample variances of the observed $\hat{\alpha}'_n$s and the asymptotic variances. For the first four maps, their asymptotic (theoretical) counterparts are available because these maps meet the requirements of either Theorem 4.2, Theorem 4.3 or Theorem 4.4. The observed results appear to be in accord with the theoretical ones. In particular, we can note that, for some values of α, the asymptotic variances can be smaller in the case of chaotic noise than in the *iid* noise case. For the sake of curiosity, we have also included the results based on Theorem 4.4 for the three other logistic maps (with $\theta \neq 4$) although strictly speaking these maps do not meet the requirements of the said proposition. There is apparently good agreement with the sample variances only for the case of $\theta = 3.98$.

Finally, Table 4.1 and Figures 4.3, 4.4, 4.5, 4.6 and 4.7 (all located at the end of this section) focus on the asymptotic normality of the estimator of α.

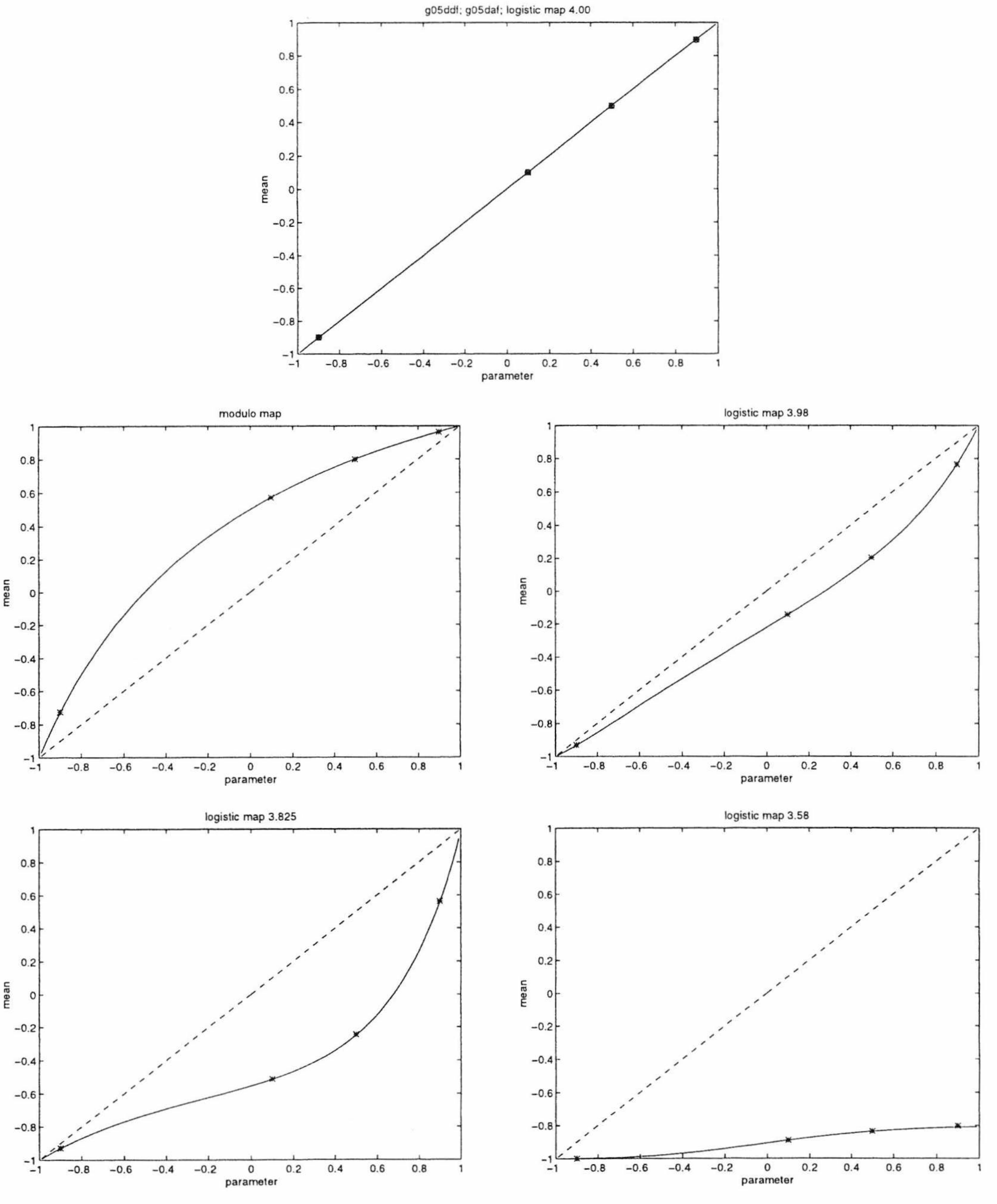


Figure 4.1: Means of the estimators of α.
The simulated expectations are obtained using 2000 replications of $\hat{\alpha}'_n$; n, the estimators sample size, is taken equal to 2000. Simulated means (denoted by an asterisk on the graphs) are obtained for four values of α (-0.9, 0.1, 0.5 and 0.9). The theoretical (asymptotic) means are based on Corollary 4.1 (solid line on the graphs). When different from the theoretical values, the (unbiased) means of the *iid* case are displayed. (dashed line on the graphs).
Note that in the first graph, '+', 'x' and 'o' respectively denote the simulated expectations for g05ddf, g05daf and the logistic map θ=4.00.

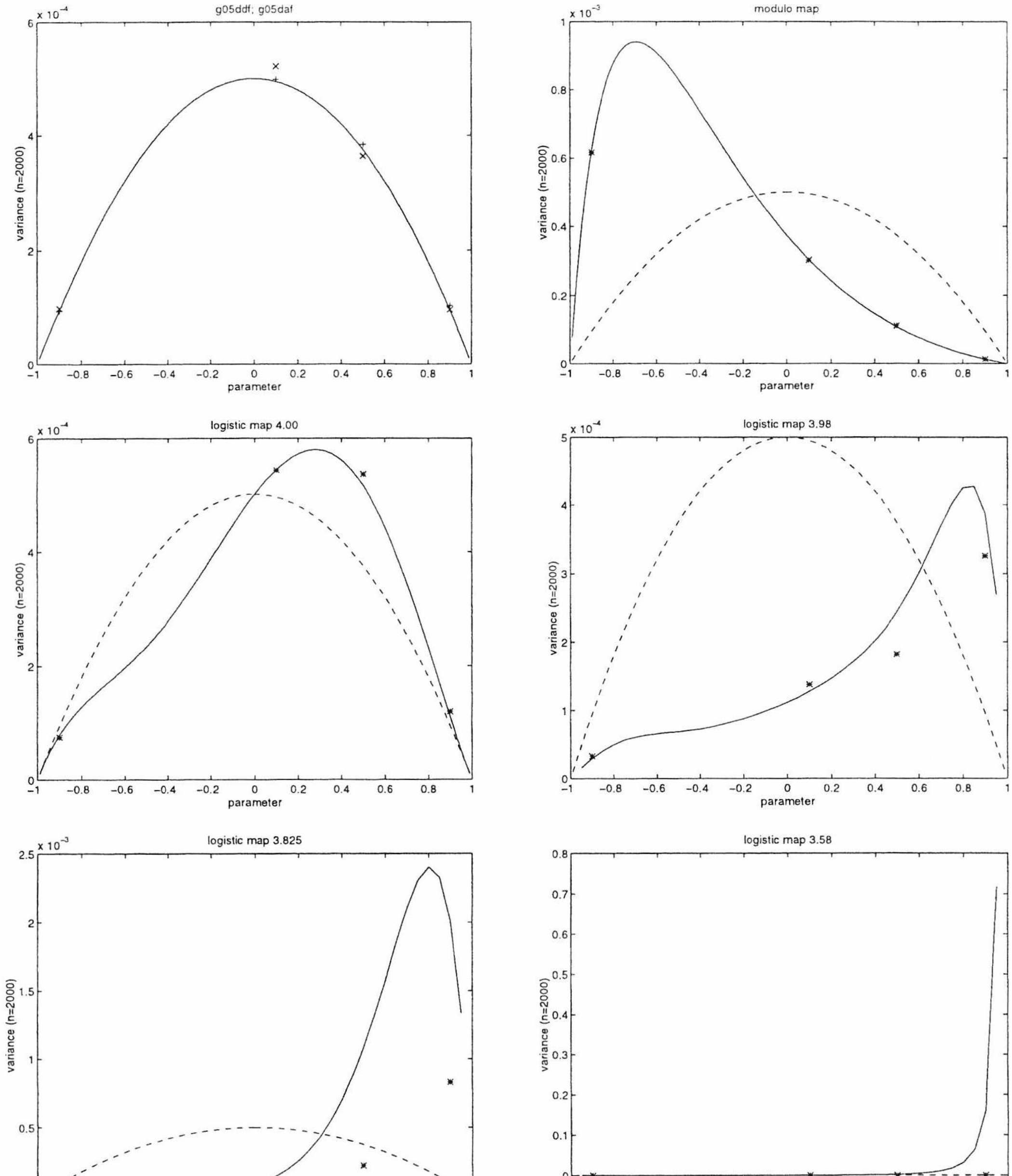


Figure 4.2: Variances of the estimators of α.
The simulated variances are obtained using 2000 replications of $\hat{\alpha}'_n$; n, the estimators sample size, is taken equal to 2000. Simulated variances (denoted by an asterisk on the graphs) are obtained for four values of α (-0.9, 0.1, 0.5 and 0.9). The theoretical (asymptotic) variances are based on the results of our paper (solid line). (The asymptotic variances of the last three logistic maps are computed using Corollary 4.4 although these maps do not satisfy its conditions. They are included for the sake of curiosity.) When different from the theoretical values, the variances in the *iid* case are displayed (dashed line).
Note that in the first graph, '+' and 'x' respectively denote the simulated variances for g05ddf and g05daf.

The Lin-Mudholkar test statistic, which is calculated in Figure 4.3, is a formal test for normality of a sample $\{\hat{e}_1, ..., \hat{e}_N\}$ against asymmetric alternatives. Specifically, we calculate

$$y_i = \{\frac{1}{N}[\sum_{j \neq i}^{N} \hat{e}_j^2 - \frac{1}{N-1}(\sum_{j \neq i}^{N} \hat{e}_j)^2]\}^{1/3}, \quad i = 1, 2, ..., N$$

and then

$$R = \frac{\sum_{i=1}^{N}(\hat{e}_i - \bar{e})(\hat{y}_i - \bar{y})}{[\sum_{i=1} N(\hat{e}_i - \bar{e})^2 \sum_{i=1} N(\hat{y}_i - \bar{y}_i)^2]^{1/2}}.$$

The Lin-Mudholkar test statistic $\frac{1}{2}(N/3)^{1/2} ln[(1+R)/(1-R)]$ is asymptotically Gaussian with zero mean and unit variance, under the null hypothesis of Gaussian $\hat{e}_t$ (see, e.g., Tong (1990) for more details).

In conformity with our theoretical results, approximate normality is apparent with the cases corresponding to the two bench-mark pseudo-random number generators and the maps $E_{t+1} = 2E_t$ (mod 1) and $E_{t+1} = 4.00E_t(1 - E_t)$ across a range of α-values but with slower rates when $|\alpha|$ gets close to 1. We also note some dissimilarity between results for negative and positive α's of the same absolute value and this is especially pronounced when $|\alpha|$ is close to 1. Interestingly, apparent normality is also discernible for the cases of $E_{t+1} = 3.98E_t(1 - E_t)$ and $E_{t+1} = 3.825E_t(1 - E_t)$ although these maps are not covered by our theorems. Perhaps simply a sufficiently fast decaying autocorrelation function will suffice as far as asymptotic normality of $\hat{\alpha}_n$ is concerned (see comments in section 5.5).

	Sample Skewness	Sample Kurtosis
g05ddf		
n = 2000		
α = 0.5	- 0.102929	3.156206
α = 0.1	- 0.035588	3.013313
α = - 0.9	0.205289	3.080298
α = 0.9	- 0.213887	2.946009
n = 5000		
α = 0.5	- 0.031205	3.097154
α = 0.1	- 0.036171	2.970640
α = - 0.9	0.137397	2.932504
α = 0.9	- 0.127471	2.798081
g05daf		
n = 2000		
α = 0.5	- 0.044629	3.009571
α = 0.1	- 0.076833	2.892512
α = - 0.9	0.325429	3.099805
α = 0.9	- 0.312897	3.100372
n = 5000		
α = 0.5	- 0.065595	2.873174
α = 0.1	- 0.037893	2.961994
α = - 0.9	0.167542	3.163506
α = 0.9	- 0.175853	3.128923
modulo map		
n = 2000		
α = 0.5	- 0.073572	2.901026
α = 0.1	- 0.142594	3.077332
α = - 0.9	0.193375	2.767712
α = 0.9	- 0.299238	3.023678
n = 5000		
α = 0.5	- 0.026970	2.992609
α = 0.1	- 0.030186	2.996988
α = - 0.9	0.157215	2.941212
α = 0.9	- 0.145436	2.780902

Table 4.1: Skewness $(E(X-\mu)^3/(E(X-\mu)^2)^{3/2} = 0.0$ if normality) and kurtosis $(E(X-\mu)^4/(E(X-\mu)^2)^2 = 3.0$ if normality) were calculated using 2000 replications of $\hat{\alpha}'_n$.

Table 4.1 continued

logistic map $\theta = 4.00$ n = 2000 $\alpha = 0.5$	- 0.081083	2.803469
$\alpha = 0.1$	- 0.016817	3.057636
$\alpha =$ - 0.9	0.154694	2.959089
$\alpha = 0.9$	- 0.334866	3.424140
n = 5000 $\alpha = 0.5$	- 0.107170	2.756208
$\alpha = 0.1$	- 0.036229	3.048160
$\alpha =$ - 0.9	0.120772	2.901569
$\alpha = 0.9$	- 0.119712	3.025298
logistic map $\theta = 3.98$ n = 2000 $\alpha = 0.5$	0.034626	2.980752
$\alpha = 0.1$	- 0.106701	2.921137
$\alpha =$ - 0.9	0.140162	2.966413
$\alpha = 0.9$	- 0.241007	3.097172
n = 5000 $\alpha = 0.5$	0.030967	2.985265
$\alpha = 0.1$	0.021639	2.918680
$\alpha =$ - 0.9	0.091189	3.052831
$\alpha = 0.9$	- 0.172674	3.092035

Table 4.1: continued

logistic map $\theta = 3.825$ n = 2000 $\alpha = 0.5$	0.157950	2.921348
$\alpha = 0.1$	- 0.099438	3.208453
$\alpha =$ - 0.9	0.315055	3.150483
$\alpha = 0.9$	- 0.120731	2.912444
n = 5000 $\alpha = 0.5$	0.086102	3.155930
$\alpha = 0.1$	- 0.075569	3.024323
$\alpha =$ - 0.9	0.105426	2.906152
$\alpha = 0.9$	- 0.078113	2.938930
logistic map $\theta = 3.58$ n = 2000 $\alpha = 0.5$	0.848187	3.615705
$\alpha = 0.1$	0.181564	2.980237
$\alpha =$ - 0.9	0.264270	2.243885
$\alpha = 0.9$	1.445372	4.436787
n = 5000 $\alpha = 0.5$	0.563053	3.412199
$\alpha = 0.1$	- 0.000096	3.204436
$\alpha =$ - 0.9	0.247537	2.312345
$\alpha = 0.9$	1.458308	4.499411

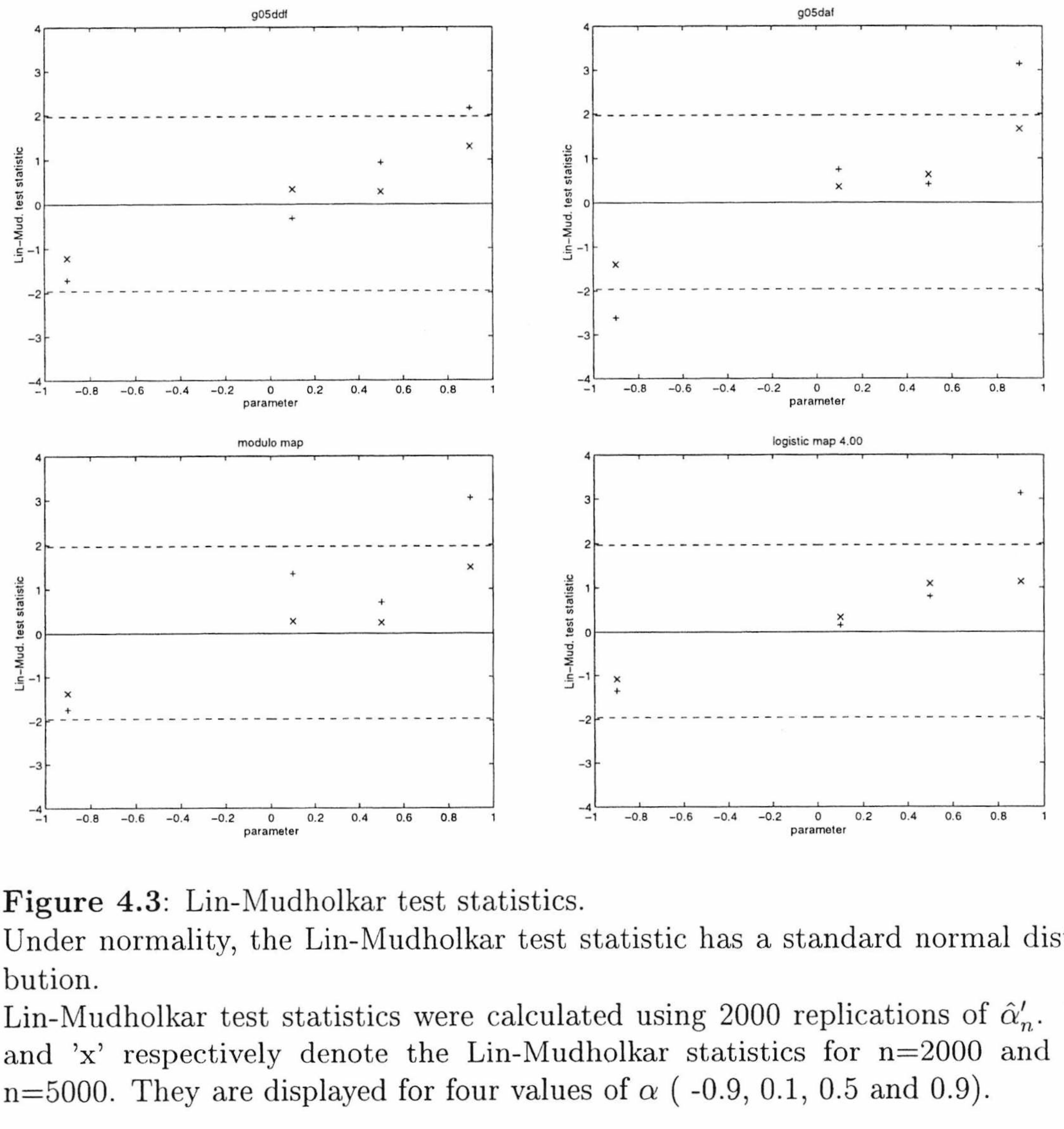
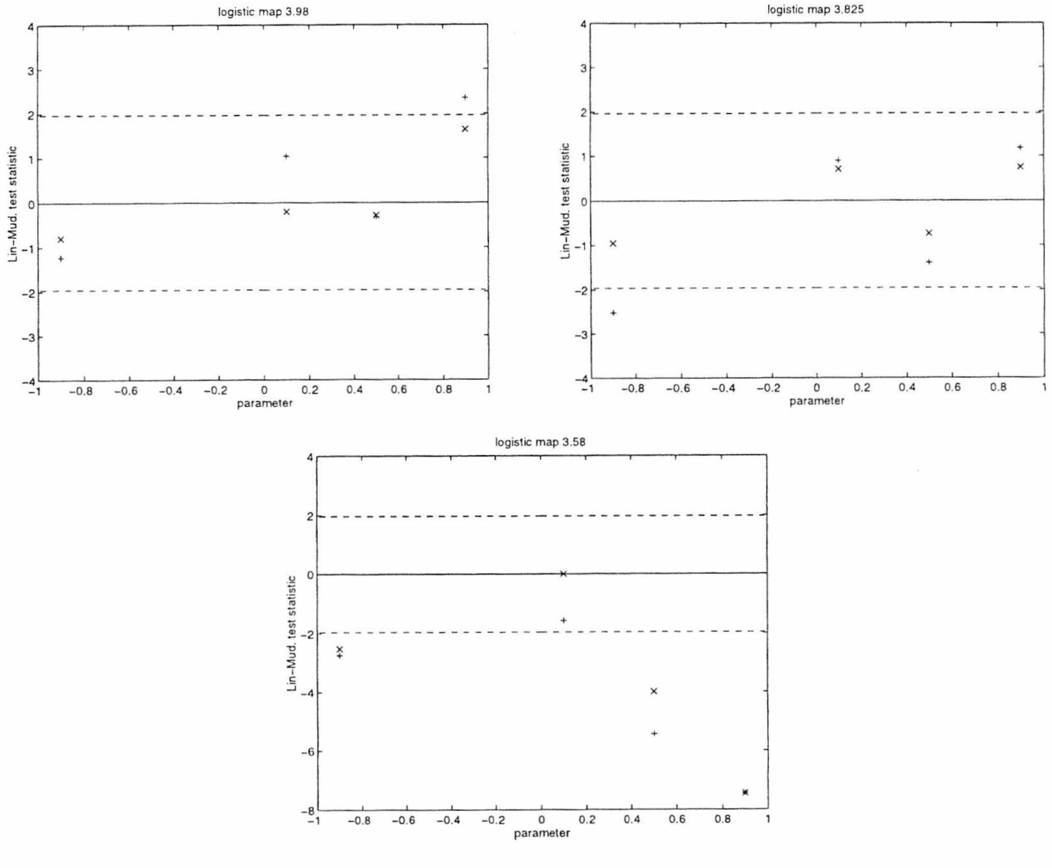


Figure 4.3: Lin-Mudholkar test statistics.
Under normality, the Lin-Mudholkar test statistic has a standard normal distribution.
Lin-Mudholkar test statistics were calculated using 2000 replications of $\hat{\alpha}'_n$. '+' and 'x' respectively denote the Lin-Mudholkar statistics for n=2000 and for n=5000. They are displayed for four values of α (-0.9, 0.1, 0.5 and 0.9).

Figure 4.3: continued



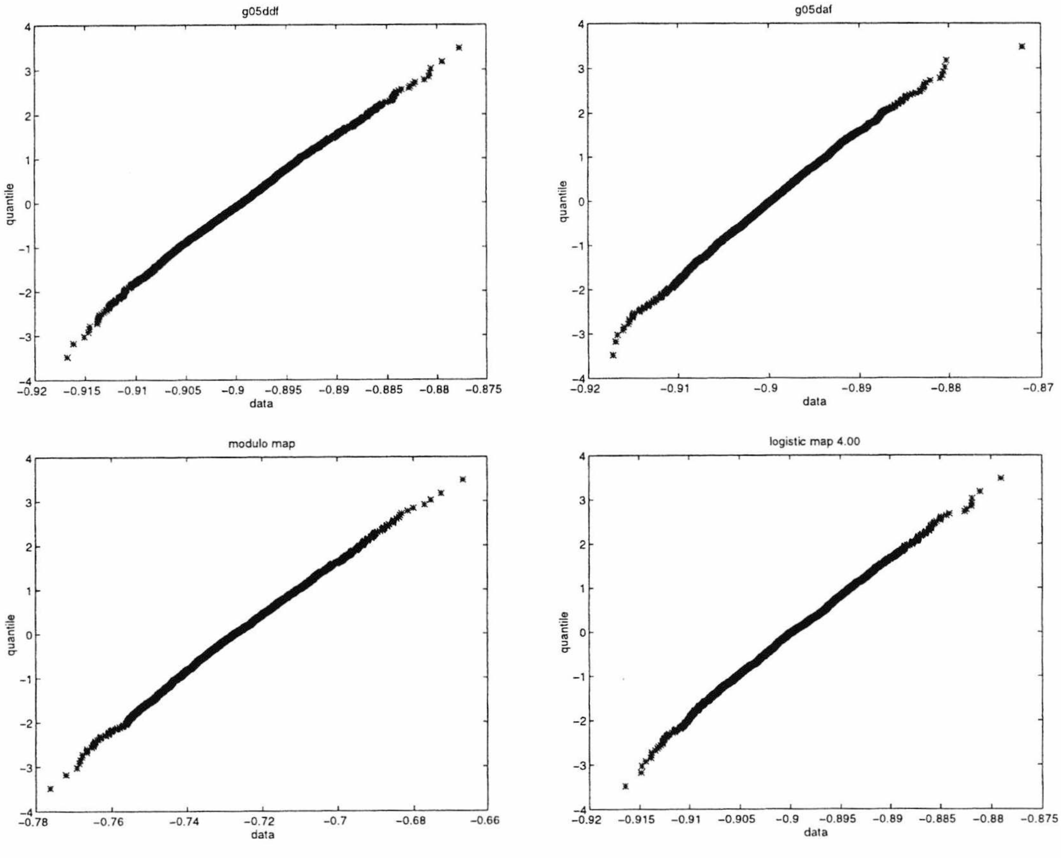
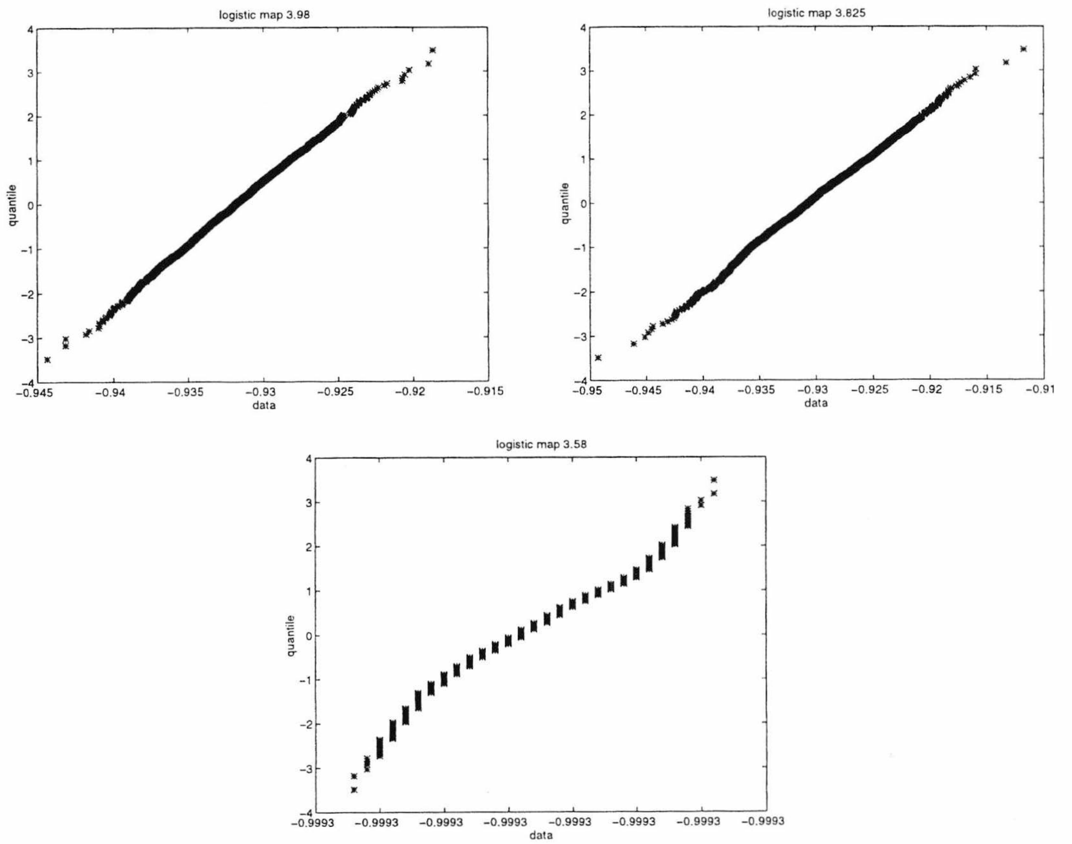


Figure 4.4: Normal probability plots for the case $\alpha = -0.9$.
The normal probability plots were obtained by using 2000 replications of $\hat{\alpha}'_n$; n was taken to be equal to 5000.

Figure 4.4: continued



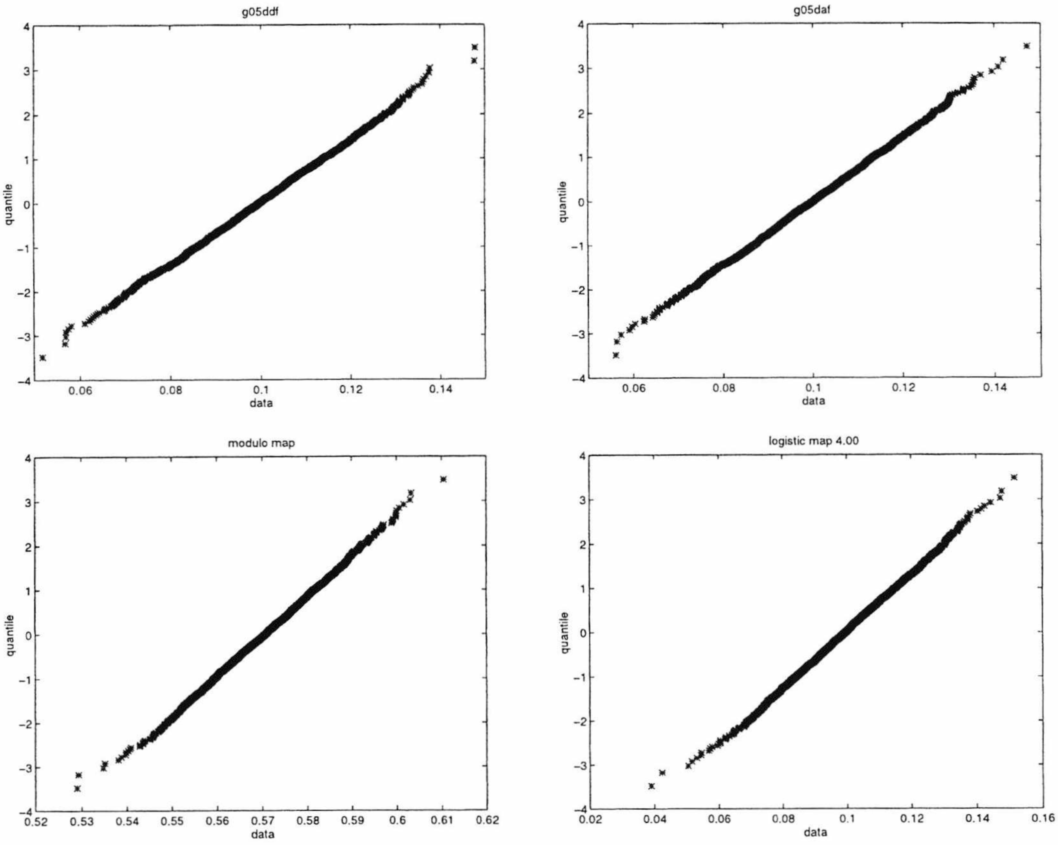
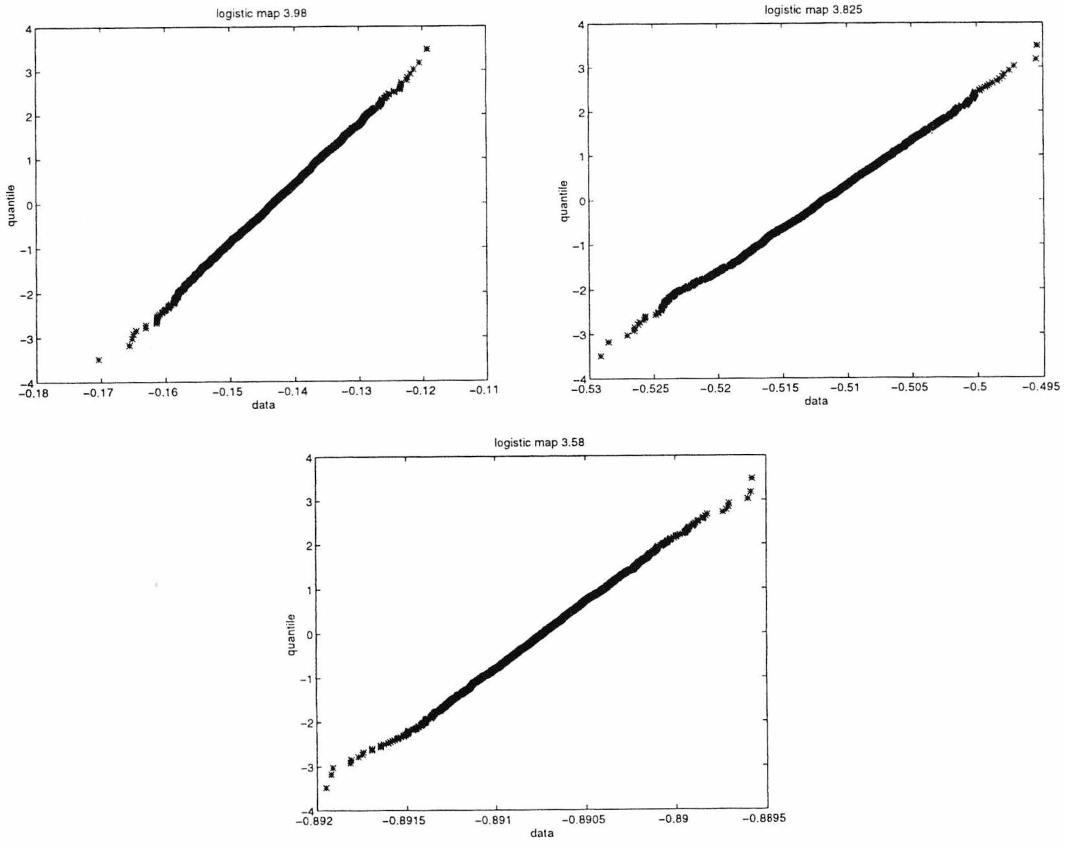


Figure 4.5: Normal probability plots for the case $\alpha = 0.1$.
The normal probability plots were obtained by using 2000 replications of $\hat{\alpha}'_n$; n was taken to be equal to 5000.

Figure 4.5: continued



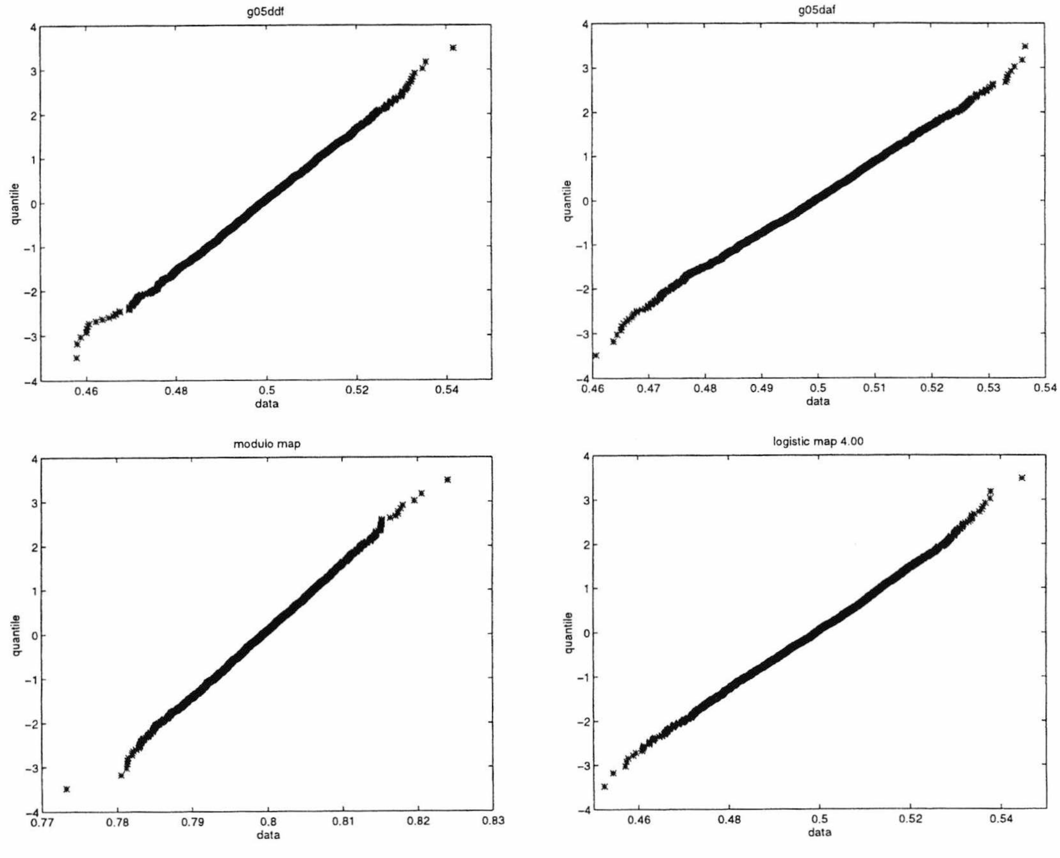
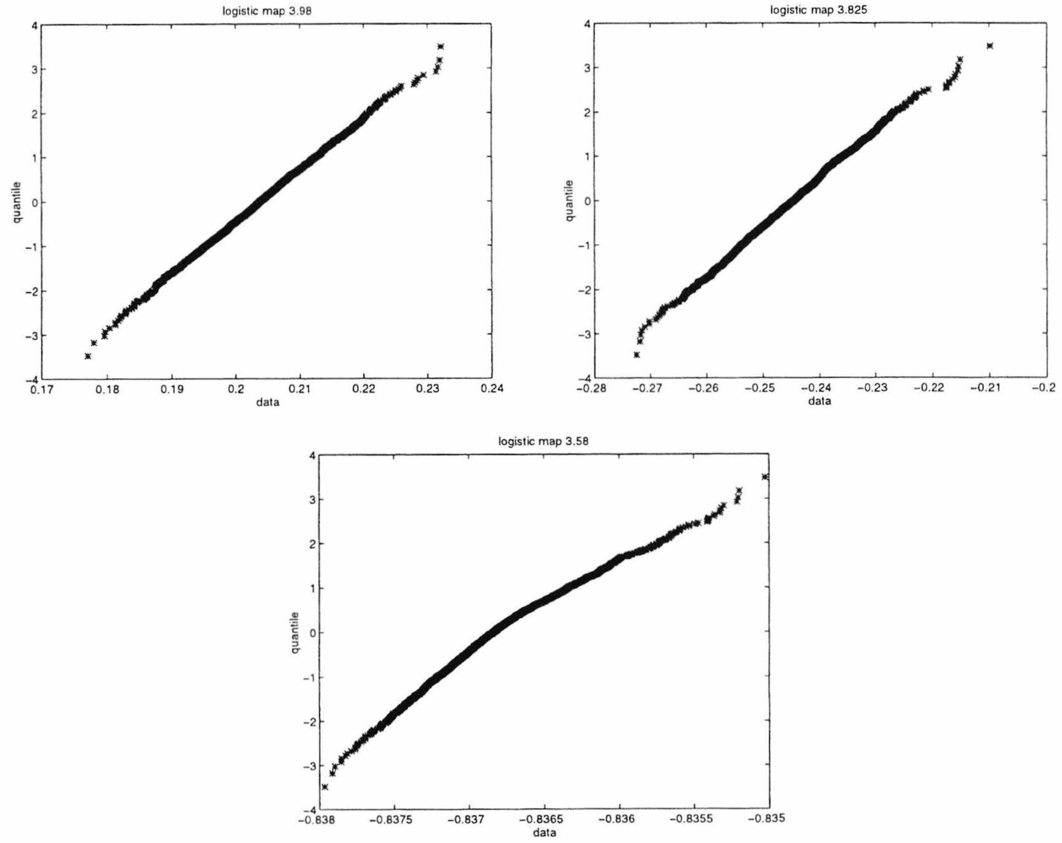


Figure 4.6: Normal probability plots for the case $\alpha = 0.5$.
The normal probability plots were obtained by using 2000 replications of $\hat{\alpha}'_n$; n was taken to be equal to 5000.

Figure 4.6: continued



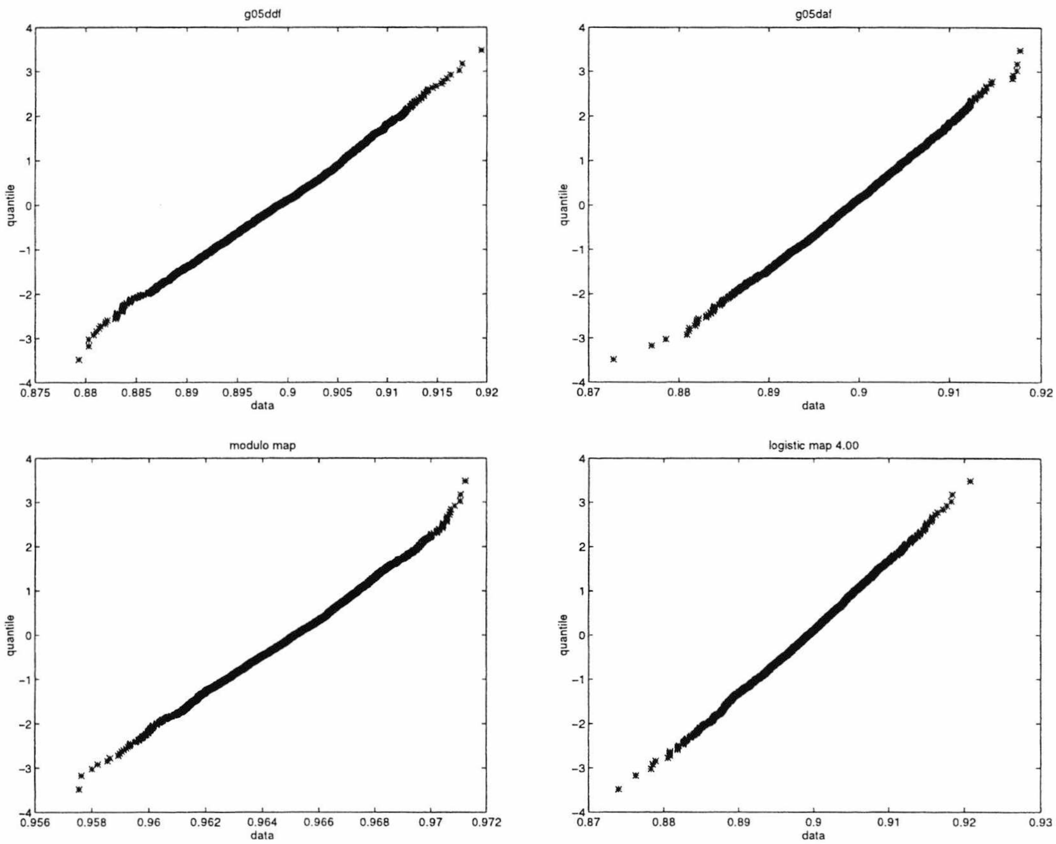
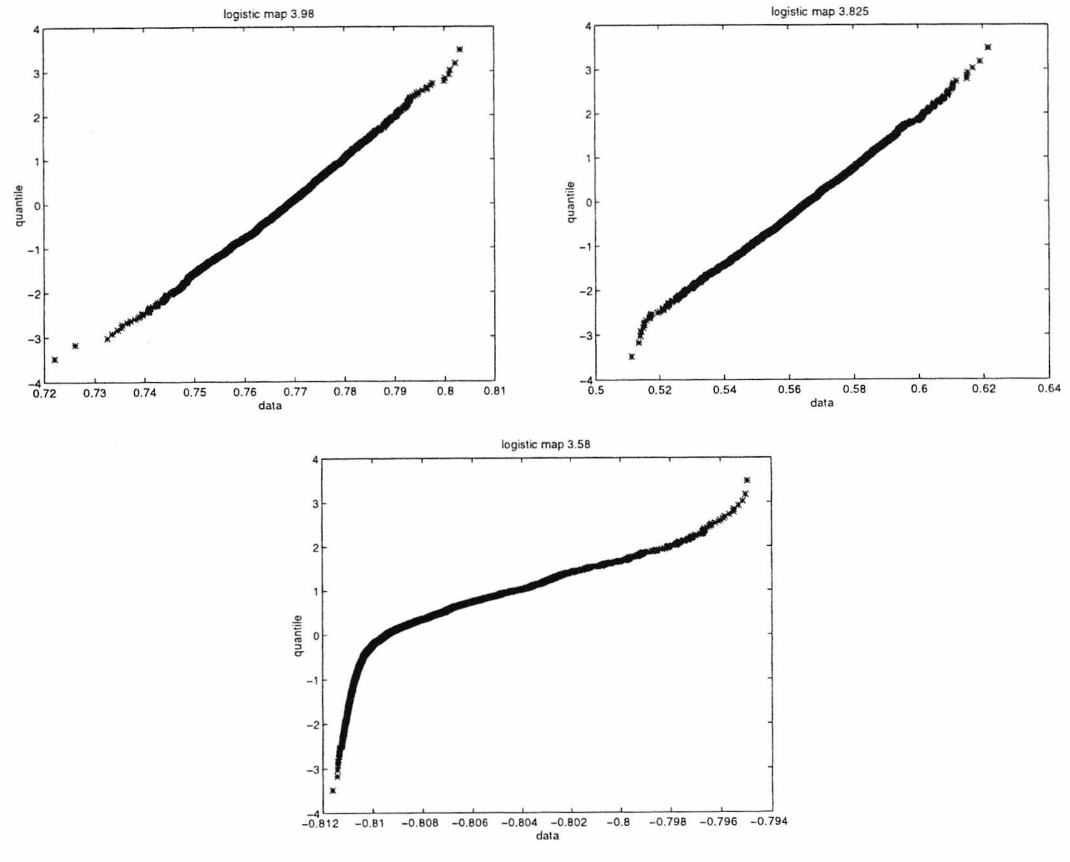


Figure 4.7: Normal probability plots for the case $\alpha = 0.9$.
The normal probability plots were obtained by using 2000 replications of $\hat{\alpha}'_n$; n was taken to be equal to 5000.

Figure 4.7: continued



4.4 Extensions of the results to AR(p)

We consider now a causal AR(p), $p \geq 1$, model driven by chaos:

$$X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + E_t, \quad \alpha_p \neq 0.$$

We recall that an AR(p) model is said to be a causal function of $\{E_t\}$ if there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $X_t = \sum_{j=0}^{\infty} \psi_j E_{t-j}$, $t = 0, 1, ...$ See, for example, Brockwell and Davis (1989) for more details. In particular, an AR(p) is causal if and only if $\phi(z) = 1 - \alpha_1 z - ... - \alpha_p z^p$, the characteristic equation of the model, has all its roots outside the unit circle (that is, $|\lambda_1| > 1, ..., |\lambda_p| > 1$). The fact that an AR(p) is causal implies thus that it is stationary. As for the AR(1) case, $\{E_t\}$ is assumed to be a chaotic sequence (in the sense of Definition 3.2) with zero mean and finite variance.

For a causal AR(p) model with *iid* noise $\{\varepsilon_t\}$, $\varepsilon_t \sim iid(0, \sigma^2)$, it is well known that

$$n^{\frac{1}{2}}(\hat{\alpha}_n - \alpha) \rightarrow^d \mathcal{N}(0, \sigma^2 \Gamma_{p,X}^{-1}),$$

where $\rightarrow^d$ denotes the convergence in distribution, $\Gamma_{p,X} = [\gamma_X(i-j)]_{i,j=1}^p$, and the Yule-Walker estimators $\hat{\alpha}_n = (\hat{\alpha}_{1,n}, \hat{\alpha}_{2,n}, ..., \hat{\alpha}_{p,n}) = \hat{R}_{p,n,X}^{-1} \hat{\rho}_{p,n,X}$, where $\hat{R}_{p,n,X} = [\hat{\rho}_{n,X}(i-j)]_{i,j=1}^p$, $\hat{\rho}_{p,n,X} = (\hat{\rho}_{n,X}(1), ..., \hat{\rho}_{n,X}(p))'$ and

$$\hat{\rho}_{n,X}(j) = \frac{\hat{\gamma}_{n,X}(j)}{\hat{\gamma}_{n,X}(0)} = \frac{\frac{1}{n}\sum_{t=1}^{n-j} X_t X_{t+j}}{\frac{1}{n}\sum_{t=1}^{n} X_t^2}.$$

The rest of the section will deal with the following question: when the sequence $\{\varepsilon_t\}$ is replaced by a chaotic sequence $\{E_t\}$, are the asymptotic properties (i.e. unbiasedness, normality) of the Yule-Walker estimators preserved?

Two remarks are in order here. Firstly, similar to the AR(1) case, there are also estimators $\hat{\alpha}'_n = \hat{R}'^{-1}_{p,n,X} \hat{\rho}'_{p,n,X}$. Note that all the properties which will be derived for $\hat{\alpha}_n$ apply to $\hat{\alpha}'_n$. The proofs, which are similar to those of sections 4.1 and 4.2, are omitted for the sake of brevity; so, we are not going to speak about $\hat{\alpha}'_n$

in this section. Secondly, our results still hold if, keeping the other assumptions, we only require f to be a real deterministic map and we assume that E_t admits an invariant distribution which is also ergodic.

By causality, there exists a sequence $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $X_t = \sum_{j=0}^{\infty} \psi_j E_{t-j}$, $t = 0, 1, \ldots$. Now, the coefficients $\{\psi_j\}$ are determined by the relation

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{1}{\phi(z)}.$$

A possible way of obtaining these coefficients is as follows: $(1-\alpha_1 z - \ldots - \alpha_p z^p)(\psi_0 + \psi_1 z + \psi_2 z^2 + \ldots) = 1$.

Equating the coefficients of z^j, we get $\psi_0 = 1$ and

$$\psi_k = \begin{cases} \sum_{i=1}^{k} \alpha_i \psi_{k-i}, & \text{if } 0 < k \leq p \\ \sum_{i=1}^{p} \alpha_i \psi_{k-i}, & \text{if } p \leq k \end{cases}$$

Therefore, by recursion, $\psi_0 = 1, \psi_1 = \alpha_1, \psi_2 = \alpha_2 + \alpha_1^2, \ldots$ (see, for example, Brockwell and Davis (1989) for more details and other ways of getting the ψ_j's). So, in the following, we should keep in mind when facing the ψ_j's that they can easily be calculated in practice.

We study the bias first.

Theorem 4.5

$$\hat{\alpha}_n \to^p R_{p,X}^{-1} \rho_{p,X},$$

where $R_{p,X} = [\rho_X(i-j)]_{i,j=1}^p$, $\rho_{p,X} = (\rho_X(1), \ldots, \rho_X(p))'$.

Proof:

$$\begin{aligned} \hat{\gamma}_{n,X}(0) &= \frac{1}{n}\sum_{t=1}^{n} X_t^2 = \frac{1}{n}\sum(\psi_0 E_t + \psi_1 E_{t-1} + \ldots)(\psi_0 E_t + \psi_1 E_{t-1} + \ldots) \\ &\simeq (\psi_0^2 + \psi_1^2 + \ldots)\hat{\gamma}_{n,E}(0) + 2(\psi_0\psi_1 + \psi_1\psi_2 + \ldots)\hat{\gamma}_{n,E}(1) \end{aligned}$$

$$+2(\psi_0\psi_2 + \psi_1\psi_3 + ...)\hat{\gamma}_{n,E}(2) + ...$$

In the same way,

$$\begin{aligned}
\hat{\gamma}_{n,X}(k) &= \frac{1}{n}\sum_{t=1}^{n-k} X_t X_{t+k} \\
&= \frac{1}{n}\sum(\psi_0 E_t + \psi_1 E_{t-1} + ...)(\psi_0 E_{t+k} + \psi_1 E_{t+k-1} + ...),\ k = 0, 1, ..., p.
\end{aligned}$$

Thus,

$$\begin{aligned}
\hat{\gamma}_{n,X}(k) &\simeq \sum_{j=0}^{\infty}(\psi_j\psi_{j+k})\hat{\gamma}_{n,E}(0) \\
&\quad + \sum_{i=1}^{\infty}\sum_{j=0}^{\infty}(\psi_j\psi_{j+|i-k|}) + \psi_j\psi_{j+i+k})\hat{\gamma}_{n,E}(i), \quad k = 0, 1, ..., p.
\end{aligned}$$

Now, $\hat{\gamma}_{n,E}(i) \rightarrow^{a.s.} \gamma_E(i)$, $i = 0, 1, 2, ...$ by Theorem 3.1 and $\hat{\gamma}_{n,X}(k) = f_k(\hat{\gamma}_{n,E}(0), \hat{\gamma}_{n,E}(1), ...)$, with f_k continuous function, $k = 0, 1, ..., p$. So, by Proposition 6.3.4 of Brockwell and Davis (1989), we get $\hat{\gamma}_{n,X}(k) \rightarrow^{p} f_k(\gamma_E(0), \gamma_E(1), ...)$, $k = 0, 1, ..., p$ and thus, by Proposition 6.3.8 of Brockwell and Davis (1989),

$$\hat{\rho}_{n,X}(k) \rightarrow^{p} \frac{f_k(\gamma_E(0), \gamma_E(1), ...)}{f_0(\gamma_E(0), \gamma_E(1), ...)}, \quad k = 1, 2, ..., p.$$

Now, for $k = 0, 1, ..., p$,

$$\begin{aligned}
\gamma_X(k) &= cov(X_t, X_{t+k}) \\
&= cov(\sum_{j=0}^{\infty}\psi_j E_{t-j}, \sum_{i=0}^{\infty}\psi_i E_{t+k-i}) \\
&= \sum_{j=0}^{\infty}(\psi_j\psi_{j+k})\gamma_E(0) + \sum_{i=1}^{\infty}\sum_{j=0}^{\infty}(\psi_j\psi_{j+|i-k|} + \psi_j\psi_{j+i+k})\gamma_E(i) \\
&= f_k(\gamma_E(0), \gamma_E(1), ...)
\end{aligned}$$

and so,

$$\hat{\rho}_{n,X}(k) \rightarrow^{p} \rho_X(k).$$

Now, for $i = 1, 2, ..., p$, $\hat{\alpha}_{i,n}$ is a continuous function of $\hat{\rho}_{n,X}(1), ..., \hat{\rho}_{n,X}(p)$ and thus, by Propositions 6.3.4 and 6.3.7 of Brockwell and Davis (1989), we get $\hat{\alpha}_n \to^p R_{p,X}^{-1}\rho_{p,X}$. □

Now, $R_{p,X}^{-1}\rho_{p,X}$ is not necessarily equal to α. In fact, except in the case where the sequence $\{E_t\}$ is uncorrelated, we get an asymptotic bias. This can easily be seen from the form of $\rho_X(k) = \frac{f_k(\gamma_E(0),\gamma_E(1),...)}{f_0(\gamma_E(0),\gamma_E(1),...)}$. In particular, in the case of an AR(1) model, we have

$$\gamma_X(0) = \frac{1}{1-\alpha^2}\gamma_E(0) + \frac{2}{1-\alpha^2}\sum_{i=1}^{\infty}\alpha^i\gamma_E(i),$$

and $\gamma_X(j) = \alpha\gamma_X(j-1) + \sum_{i=j}^{\infty}\alpha^{i-j}\gamma_E(i)$, $j = 1, 2, ...$.

So,

$$\rho_X(0) = 1, \rho_X(j) = \alpha\rho_X(j-1) + \frac{\sum_{i=j}^{\infty}\alpha^{i-j}\gamma_E(i)}{\gamma_X(0)}, j \geq 1,$$

which can be very different from the *iid* (or, equivalently, from the uncorrelated case) where $\rho_X(j) = \alpha^j$, $j \geq 0$.

For all the chaotic sequences (with finite variance), we have been able to quantify the asymptotic bias of the Yule-Walker estimators. In the next section, examples and simulations will be given in the case of an AR(2) model.

We now focus on the possible asymptotic normality of the Yule-Walker estimators. Remember that it has been possible to obtain the asymptotic normality of $\hat{\alpha}$ in the case of AR(1) basically for two types of chaotic sequences, namely the Bernoulli systems and chaotic sequences satisfying a version of Denker and Keller (1986)'s theorem. We are going to show that for AR(p), $p \geq 1$, we can get the asymptotic normality of $\hat{\alpha}$ for the Bernoulli chaotic sequences and for chaotic sequences satisfying a version (slightly stronger than in the AR(1) case) of Denker and Keller's theorem.

The next theorem corresponds to Theorem 4.2 in the AR(1) case. Before stating it, we need to introduce the following notation. By the proof of Theorem 4.5, we know that, for $k = 1, 2, ..., p$, $\rho_X(k)$ is a (continuous) function of $\rho_E(1), \rho_E(2), ...$; thus, each component of the column vector $R_{p,X}^{-1}\rho_{p,X}$ is a function of $\rho_E(1), \rho_E(2), ...$: we denote $(R_{p,X}^{-1}\rho_{p,X})_k = g_k(\rho_E(1), \rho_E(2), ...)$, $k = 1, 2, ..., p$.

Theorem 4.6 *If E_t can be written as*

$$E_t - \mu = \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i},$$

where $\mu = \mathrm{E}(E_t)$, $\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$, $\sum_{i=-\infty}^{\infty} |i|\psi_i^2 < \infty$, and $\{Z_t\}$ is a set of iid random variables with $\mathrm{E}(Z_t) = 0$ and $\mathrm{E}(Z_t^2) = \sigma^2 < \infty$, then

$$n^{\frac{1}{2}}(\hat{\alpha}_n - R_{p,X}^{-1}\rho_{p,X}) \to^d \mathcal{N}(0, DWD'),$$

where $a^T = (1, \alpha, \alpha^2, ...)$ and

$$\begin{aligned} W &= (w_{ij})_{i=1,...,j=1,...} \\ &= (\sum_{k=-\infty}^{\infty} \{\rho_E(k+i)\rho_E(k+j) + \rho_E(k-i)\rho_E(k+j) + 2\rho_E(i)\rho_E(j)\rho_E^2(k) \\ &\quad -2\rho_E(i)\rho_E(k)\rho_E(k+j) - 2\rho_E(j)\rho_E(k)\rho_E(k+i)\}). \end{aligned}$$

and D is the matrix $[(\frac{\partial g_i}{\partial x_j})(\mu^*)]_{i=1,2,...,p;\ j=1,2,...}$, where $\mu^* = (\rho_E(1), \rho_E(2), ...)'$.

Proof: For each j, the asymptotic joint distribution of $(\hat{\rho}_E(1), ..., \hat{\rho}_E(j))$ is normal (see, e.g., Theorem 7.2.2 of Brockwell and Davis (1989)).

So,

$(\hat{\rho}_E(1), \hat{\rho}_E(2), ...)'$ is asymptotically $\mathcal{N}((\rho_E(1), \rho_E(2), ...)', n^{-1}W)$

Now,

$$\hat{\alpha}_{k,n} \simeq g_k(\hat{\rho}_{n,E}(1), \hat{\rho}_{n,E}(2), ...), \quad k = 1, 2, ..., p$$

and each g_k is continuously differentiable in a neighborhood of μ^*; on the other hand, the matrix DWD' has all of its diagonal elements non-zero.

So, we can get the conclusion by applying Proposition 6.4.3 of Brockwell and Davis (1989). □

The following theorem is the analog of Theorem 4.3 which addresses the AR(1) case.

Theorem 4.7 *If E_t can be written as $E_t - \mu = \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i}$, where $\mu = \mathrm{E}(E_t)$, $\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$ and $\{Z_t\}$ is a set of iid random variables with $\mathrm{E}(Z_t) = 0$ and $\mathrm{E}(Z_t^4) < \infty$, then*

$$n^{\frac{1}{2}}(\hat{\alpha}_n - R_{p,X}^{-1}\rho_{p,X}) \to^d \mathcal{N}(0, DWD'),$$

where W, D, g_k, $k = 1, 2, ..., p$, and μ^* are as in Theorem 4.6.

Proof: The proof is similar to the proof of Theorem 4.6 but this time Theorem 7.2.1 of Brockwell and Davis (1989) is used. □

Note that the chaotic sequences $\{E_t\}$ generated by the modulo map satisfy the requirements of Theorems 4.6 and 4.7.

Now, we shall state and prove the theorem which corresponds to Theorem 4.4 in the AR(1) case. The condition 3 in Theorem 4.4 has to be slightly strenghtened if we wish to ensure the asymptotic normality of the Yule-Walker estimators for any order p of the causal AR(p) model. We can, however, notice that, if f is continuously differentiable, then the (new) condition 3 of Theorem 4.8 will be met as was the condition 3 of Theorem 4.4.

Before Theorem 4.8, we introduce the following notation:
$R_{p,X}^{-1}\rho_{p,X} = \Gamma_{p,X}^{-1}\gamma_{p,X} = (f_1(\gamma_E(0), \gamma_E(1), ...), ..., f_p(\gamma_E(0), \gamma_E(1), ...))'$

Theorem 4.8 *Suppose that the following conditions hold:*

1. *$\{E_n : n \geq 1\}$ is a Lipschitz functional of $\{Z_n : n \geq 1\}$. Specifically there exists a function $g(u_1, u_2, ...)$ satisfying $E_j = g(Z_j, Z_{j+1}, ...)$ for $j \geq 1$ and*

there are some $\lambda < 1$ *and* $c < \infty$ *such that*

$$|g(z_1, z_2, ...) - g(z_1', z_2', ...)| \leq c\lambda^n \text{ if } z_1 = z_1', ..., z_n = z_n'.$$

2. $\{Z_n, n \geq 1\}$ *is an absolutely regular stationary sequence with* $\beta(n)^{\frac{\eta}{2+\eta}} = O(n^{-2-\epsilon})$ *for some* $\epsilon, \eta > 0$.

3. $h_k(E_t) = E_t f^k(E_t)$, $k = 0, 1, 2, ...,$ *and every linear combination of* $h_k(E_t)$*'s satisfies the following Lipschitz condition:*

   let h *be any of the above functions; then there are* $L > 0$, $r \geq 0$ *and* $\rho > 0$ *such that*

   $|h(x) - h(y)| \leq |x - y|^\rho (1 + |x|^r + |y|^r)$,

   for all x *and* y *belonging to the support of* E_t.

Then,

$$n^{\frac{1}{2}} (\hat{\alpha}_n - R_{p,X}^{-1} \rho_{p,X}) \rightarrow^d \mathcal{N}(0, G\Sigma G'),$$

where

$$\begin{aligned}
\Sigma &= (\sigma_{ij})_{i=1,...;j=1,...} \\
&= (\sum_{s=-\infty}^{\infty} (\mathrm{E}(E_t E_{t+i} E_{t+s} E_{t+s+j}) - \gamma_E(i)\gamma_E(j)))
\end{aligned}$$

and G *is the matrix* $[(\frac{\partial f_i}{\partial x_j})(\nu)]_{i=1,2,...,p;j=1,2,...}$ *where* $\nu = (\gamma_E(0), \gamma_E(1), ...)'$.

Proof: For each k, the kernel $h_k : x \rightarrow x f^k(x)$ satisfies condition (i) of Theorem 1 in Denker and Keller (1986). Now, $\hat{\gamma}_{n,E}(k)$ is, up to order $1/n$, equal to a U-statistic whose associated kernel is h_k. Therefore, by Theorem 1 in Denker and Keller (1986), $\hat{\gamma}_E(k)$ is asymptotically normal; in fact, $\hat{\gamma}_E(k)$ is asymptotically $\mathcal{N}(\gamma_E(k), \frac{1}{n}(\mathrm{E}[h_k(E_1)]^2 - \theta_k^2 + 2\sum_{r\geq 2} \mathrm{E}[(h_k(E_1) - \theta_k)(h_k(E_r) - \theta_k)]))$, $k = 0, 1, 2, ...,$ where $\theta_k = \mathrm{E}[h_k(E_1)]$.

Moreover, it can easily be seen that

$$cov(\hat{\gamma}_{n,E}(i), \hat{\gamma}_{n,E}(j)) \simeq \frac{1}{n} \sum_{s=-\infty}^{\infty} (\mathrm{E}(E_t E_{t+i} E_{t+s} E_{t+s+j}) - \gamma_E(i)\gamma_E(j)).$$

Now, every linear combination of $\hat{\gamma}_E(k)$'s is asymptotically normal. For: let $K > 0$, $d_0, d_1, ..., d_K \geq 0$, then

$$\begin{aligned}
\sum_{i=0}^{K} d_i \hat{\gamma}_E(i) &= d_0 \frac{1}{n} \sum_{t=1}^{n} E_t^2 + d_1 \frac{1}{n} \sum_{t=1}^{n-1} E_t E_{t+1} + ... + d_K \frac{1}{n} \sum_{t=1}^{n-K} E_t E_{t+K} \\
&\simeq \frac{1}{n} \sum_{t=1}^{n} E_t (d_0 E_t + d_1 f(E_t) + ... + d_K f^K(E_t)) \\
&\simeq \frac{1}{n} \sum_{t=1}^{n} E_t f^*(E_t),
\end{aligned}$$

and so $\sum_{i=0}^{K} d_i \hat{\gamma}_E(i)$ is asymptotically normal. Theorem 1 in Denker and Keller (1986) applies because of condition 3 of our theorem. So, $(\hat{\gamma}_E(0), \hat{\gamma}_E(1), ...)'$ is asymptotically $\mathcal{N}((\gamma_E(0), \gamma_E(1), ...)', n^{-1}\Sigma)$.

Now, $\hat{\alpha}_{k,n} \simeq f_k(\hat{\gamma}_{n,E}(0), \hat{\gamma}_{n,E}(1), ...)$, $k = 1, 2, ..., p$.

Each f_k is continuously differentiable in a neighbourhood of ν and the matrix $G\Sigma G'$ has all of its diagonal elements non-zero. So, we can get the conclusion by applying Proposition 6.4.3 in Brockwell and Davis (1989). □

An example of a map which satisfies the conditions of Theorem 4.8 is given by the logistic map with $\theta = 4.00$, that is $E_t \to 4E_t(1 - E_t)$. Now, we shall illustrate the above asymptotic results in the next section by means of the chaos driven AR(2) case.

4.5 Simulations

In this section, we concentrate on chaos driven AR(2) models (i.e. $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + E_t$) with a view to illustrate in a simple way the results of section 4.4. First, we recall the Yule-Walker estimators of an AR(2) model and some of its properties:

$$\hat{\alpha}_{1,n} = \frac{\hat{\rho}_{n,X}(1)(1-\hat{\rho}_{n,X}(2))}{1-(\hat{\rho}_{n,X}(1))^2}$$

and

$$\hat{\alpha}_{2,n} = \frac{\hat{\rho}_{n,X}(2)-(\hat{\rho}_{n,X}(1))^2}{1-(\hat{\rho}_{n,X}(1))^2}.$$

Provided the roots of $1 - \alpha_1 z - \alpha_2 z^2 = 0$ are both outside the unit circle, we can write $X_t = \sum_{j=0}^{\infty} \psi_j E_{t-j}$, $t = 0, 1...$ and the ψ_j's can easily be obtained (see section 4.4).

The calculations of the asymptotic means and variances are based on the following formula (see section 4.4 for more details):

$$\rho_X(k) = \frac{\sum_{j=0}^{\infty}(\psi_j\psi_{j+k})\gamma_E(0) + \sum_{i=1}^{\infty}\sum_{j=0}^{\infty}(\psi_j\psi_{j+|i-k|} + \psi_j\psi_{j+i+k})\gamma_E(i)}{\sum_{j=0}^{\infty}\psi_j^2\gamma_E(0) + 2\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}\psi_j\psi_{i+j}\gamma_E(i)}, \quad k = 1, 2.$$

Now, as in section 4.3, we consider chaotic sequences $\{E_t\}$ generated by the following maps: the two pseudo-random number generators G05DAF and G05DDF, the modulo map and the logistic maps with $\theta = 4.00$, $\theta = 3.98$, $\theta = 3.825$, $\theta = 3.58$. Their autocorrelation functions can be seen in Figure 3.3. All the maps satisfy the set-up of Theorem 4.5 (concerning the asymptotic bias); on the other hand, the three last logistic maps do not fulfil the conditions of either Theorem 4.6, 4.7 or 4.8, which means that we do not expect a priori that their Yule-Walker estimators are asymptotically normal and that we do not know their asymptotic variances.

For the simulations, we standardize every chaotic sequence $\{E_t\}$ to zero mean and unit variance. Note that for the modulo map, we replaced 2 by 1.99999 in the simulation in order to avoid degeneracy due to finite precision arithmetic.

Numerous simulations have been made; for the sake of brevity, only a typical part of them has been reproduced in this thesis. Note that all the figures and tables related to this section are located at the end of this section.

The simulations concern the asymptotic bias (Figures 4.8 to 4.23), the asymptotic variance (Figures 4.24 to 4.32) and the asymptotic normality (Table 4.2 and Figures 4.33 to 4.36). On the one hand, simulated results are compared with our theoretical results. When the conditions of our theorems are satisfied, the agreement is good. On the other hand, our theoretical results are illustrated for different values of the parameters. Note that some of the figures display non-smooth theoretical curves (this is particularly clear for Figure 4.32). In fact, the non-smooth character is due to discretisation. Now, we can make some general remarks.

Concerning the asymptotic bias of $\hat{\alpha}_1$ (or, equivalently, of $\hat{\alpha}_2$), we notice the role of the noise autocorrelation function and, if the noise is correlated, of the true values of α_1 and α_2. So, in the case of a non-zero asymptotic bias , the value of α_2 affects the asymptotic bias of $\hat{\alpha}_1$ (as does obviously α_1).

When looking at the figures on the asymptotic variance, we can see that the asymptotic variances of $\hat{\alpha}_1$ and $\hat{\alpha}_2$, which are both equal to $(1-\alpha_2^2)/n$ in the *iid* case, depend not only of α_2 but also of α_1 for all the other considered sequences; note, however, that the asymptotic variances of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are identical in the case of the modulo map. For the sake of curiosity, we have also included some results based on Theorem 4.8 for the three other logistic maps (with $\theta \neq 4.00$) although strictly speaking these maps do not meet the requirements of the said theorem. There is apparently good agreement with the sample variances for the cases of $\theta = 3.98$ (as in the AR(1) case) and of $\theta = 3.825$.

In conformity with our theoretical results, approximate normality is apparent with the cases corresponding to the two bench-mark pseudo-random number generators, the modulo map and the logistic map with $\theta = 4.00$. As in the AR(1) case, the Yule-Walker estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ for the cases of the logistic maps $\theta = 3.98$ and $\theta = 3.825$ also appear to be asymptotically normal although these maps are not covered by our theorems. Perhaps, a sufficiently fast decaying autocorrelation function is enough to get asymptotic normality of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ (see comments in section 5.5).

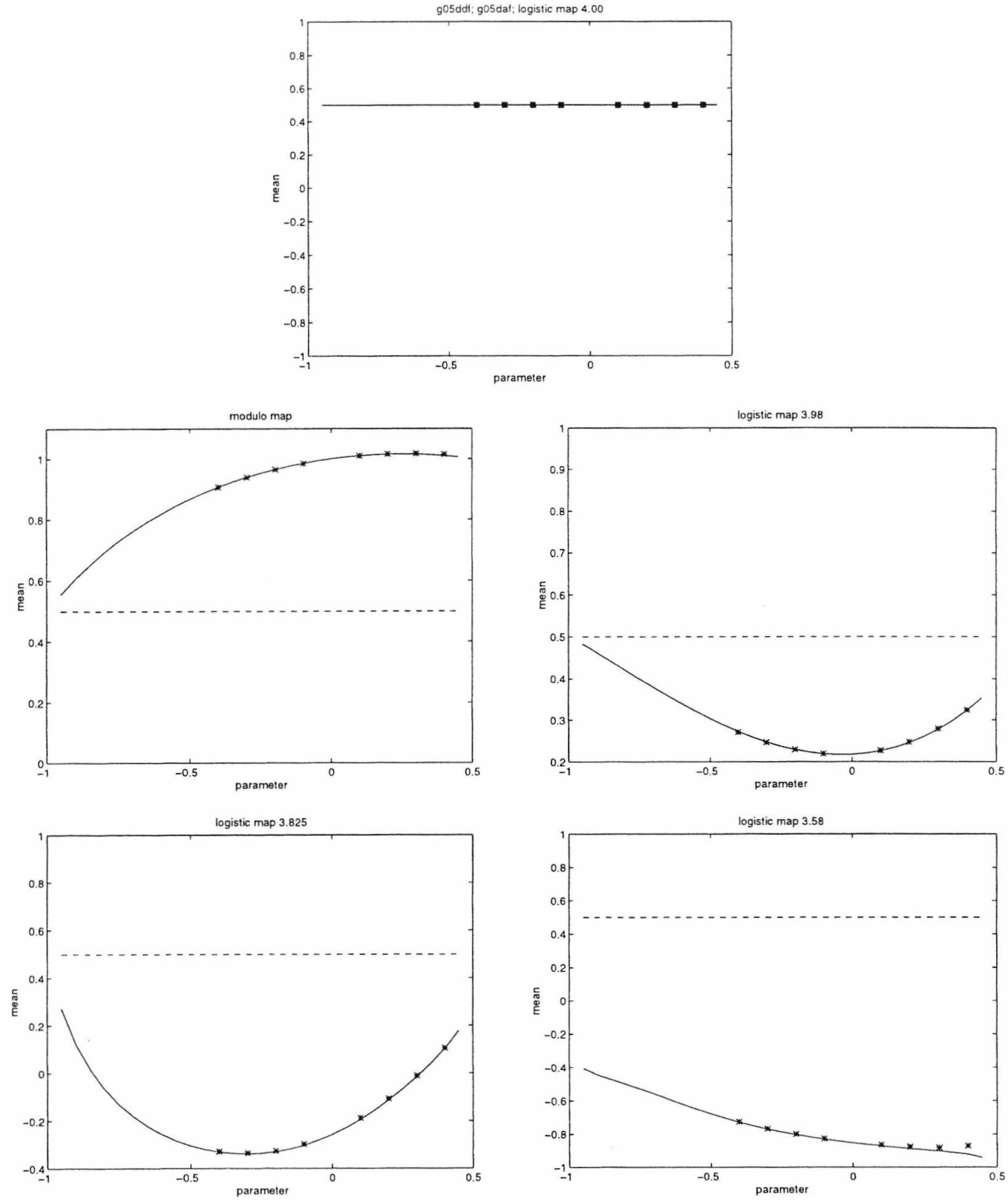


Figure 4.8: Means of the estimators of α_1 for the case $\alpha_1 = 0.5$, while α_2 varies through the range of values for which the AR(2) model is causal.
The simulated expectations are obtained using 2000 replications of $\hat{\alpha}_{1,n}$; n, the estimators sample size, is taken equal to 2000. Simulated means (denoted by an asterisk on the graphs) are obtained for eight values of α_2 (-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3 and 0.4) The theoretical (asymptotic) means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the (unbiased) means of the *iid* case are displayed. (dashed line on the graphs). Note that in the first graph, '+', 'x' and 'o' respectively denote the simulated expectations for g05ddf, g05daf and the logistic map θ=4.00.

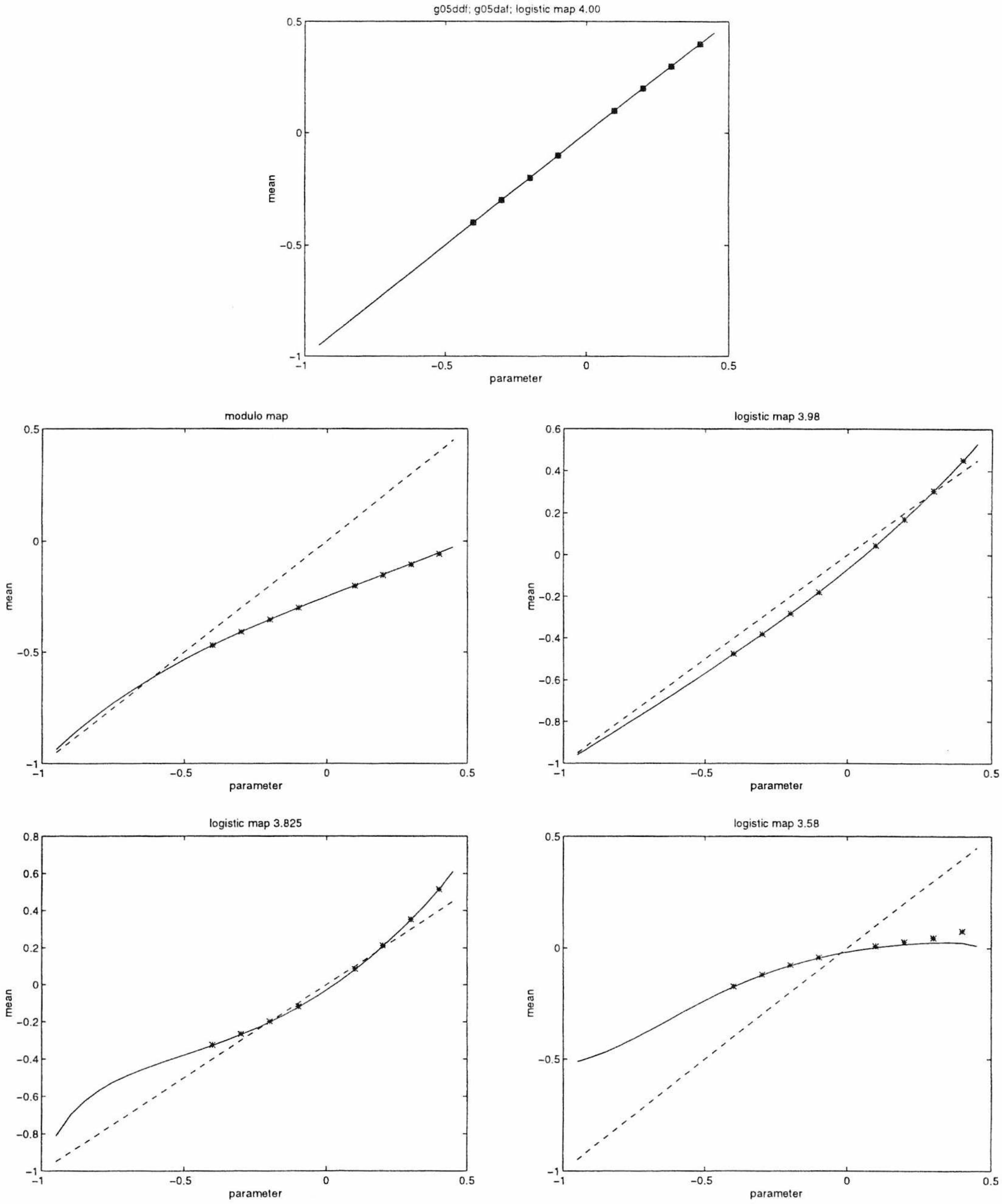


Figure 4.9: Means of the estimators of α_2 for the case $\alpha_1 = 0.5$, while α_2 varies through the range of values for which the AR(2) model is causal.
The simulated expectations are obtained using 2000 replications of $\hat{\alpha}_{2,n}$; n, the estimators sample size, is taken equal to 2000. Simulated means (denoted by an asterisk on the graphs) are obtained for eight values of α_2 (-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3 and 0.4) The theoretical (asymptotic) means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the (unbiased) means of the *iid* case are displayed. (dashed line on the graphs). Note that in the first graph, '+', 'x' and 'o' respectively denote the simulated expectations for g05ddf, g05daf and the logistic map θ=4.00.

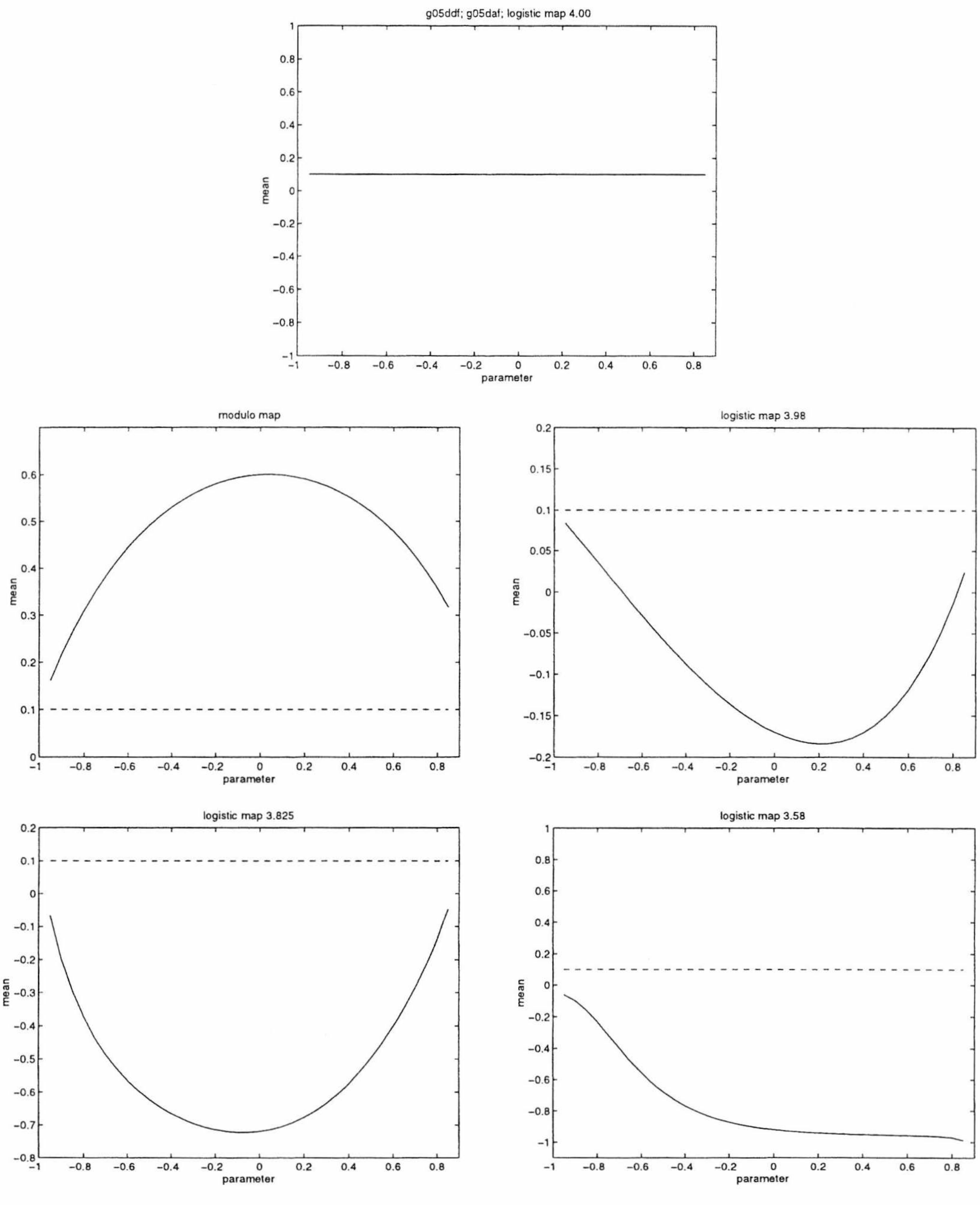


Figure 4.10: Theoretical means of the estimators of α_1 for the case $\alpha_1 = 0.1$, while α_2 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

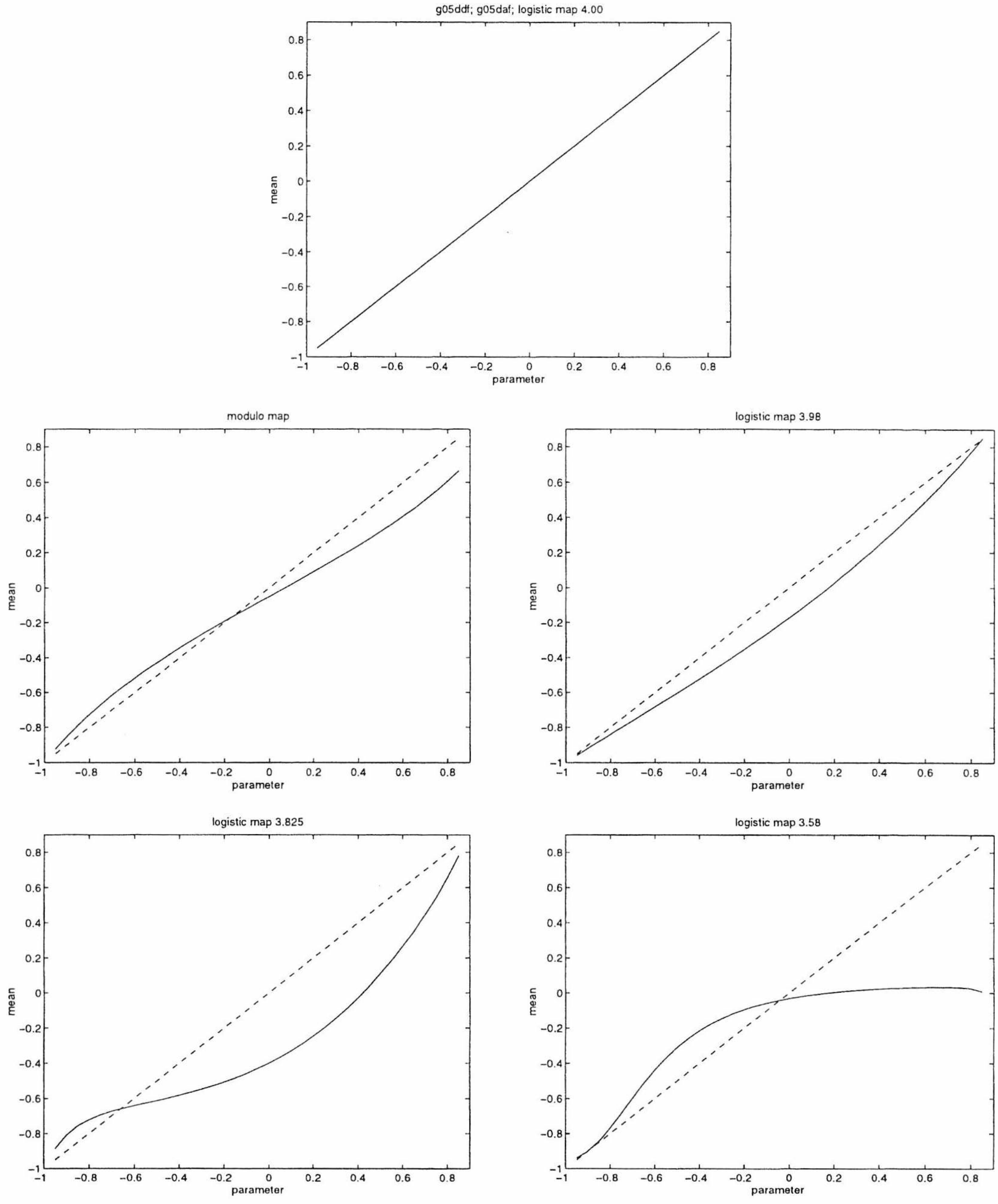


Figure 4.11: Theoretical means of the estimators of α_2 for the case $\alpha_1 = 0.1$, while α_2 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

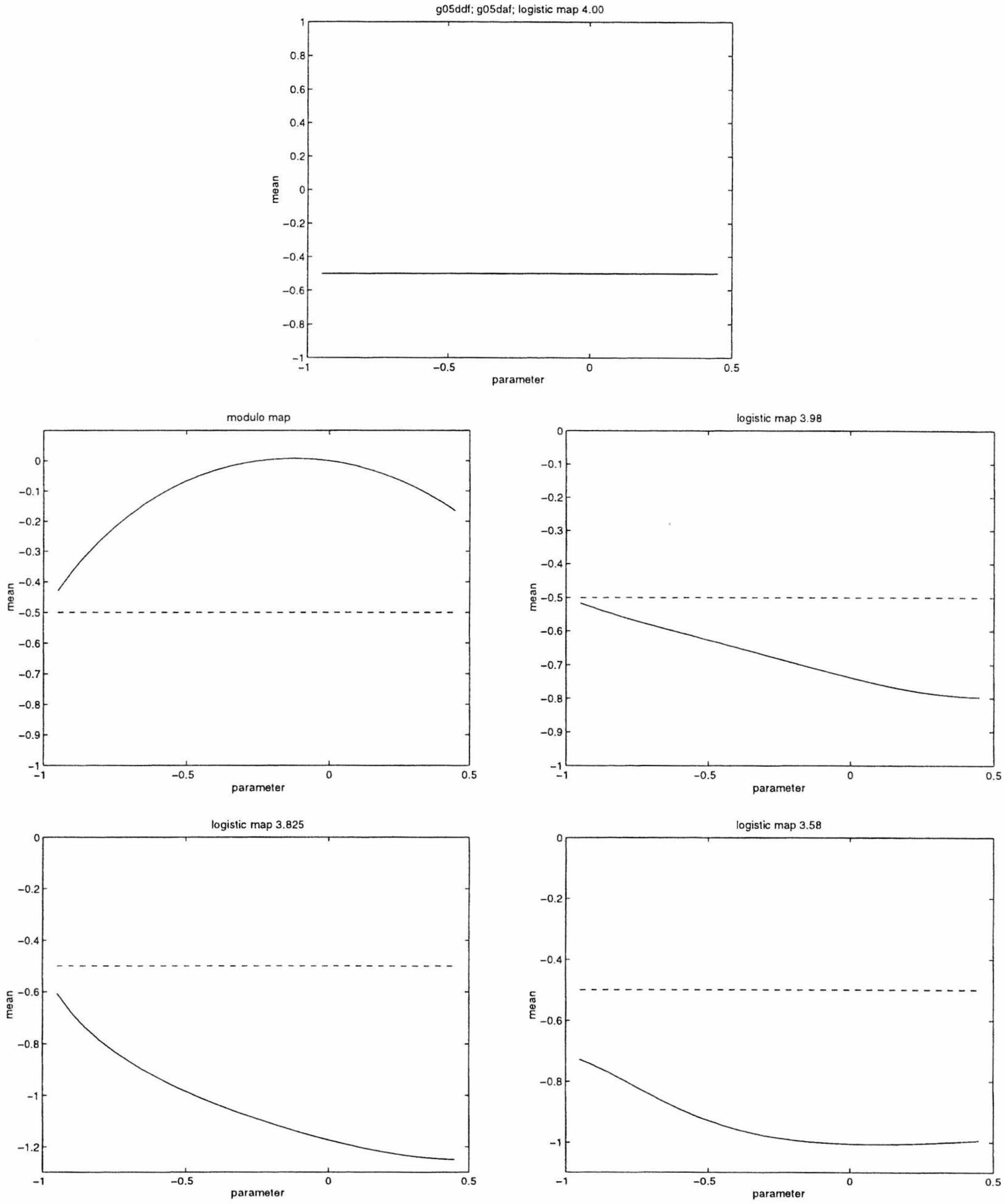


Figure 4.12: Theoretical means of the estimators of α_1 for the case $\alpha_1 = -0.5$, while α_2 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

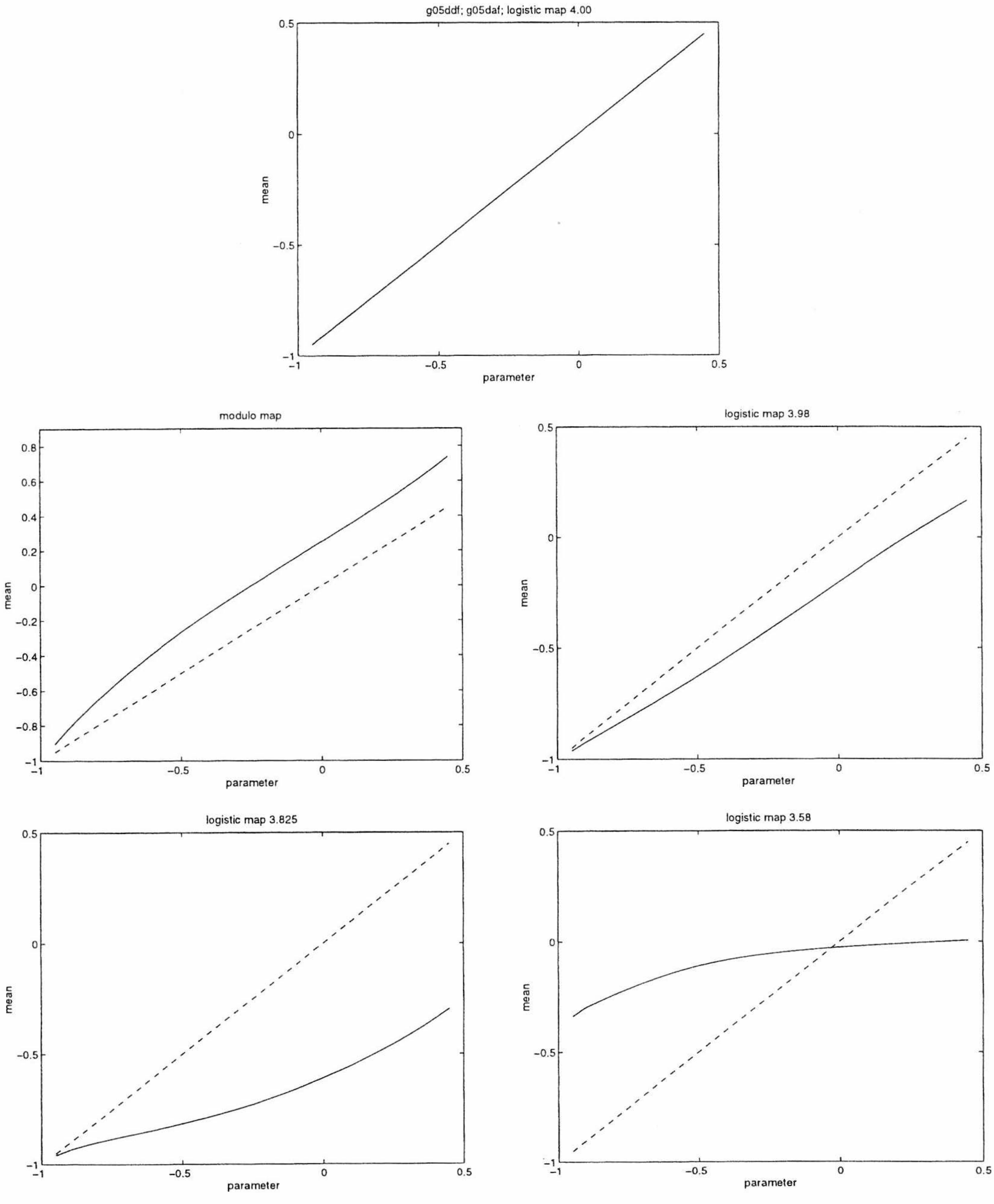


Figure 4.13: Theoretical means of the estimators of α_2 for the case $\alpha_1 = -0.5$, while α_2 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

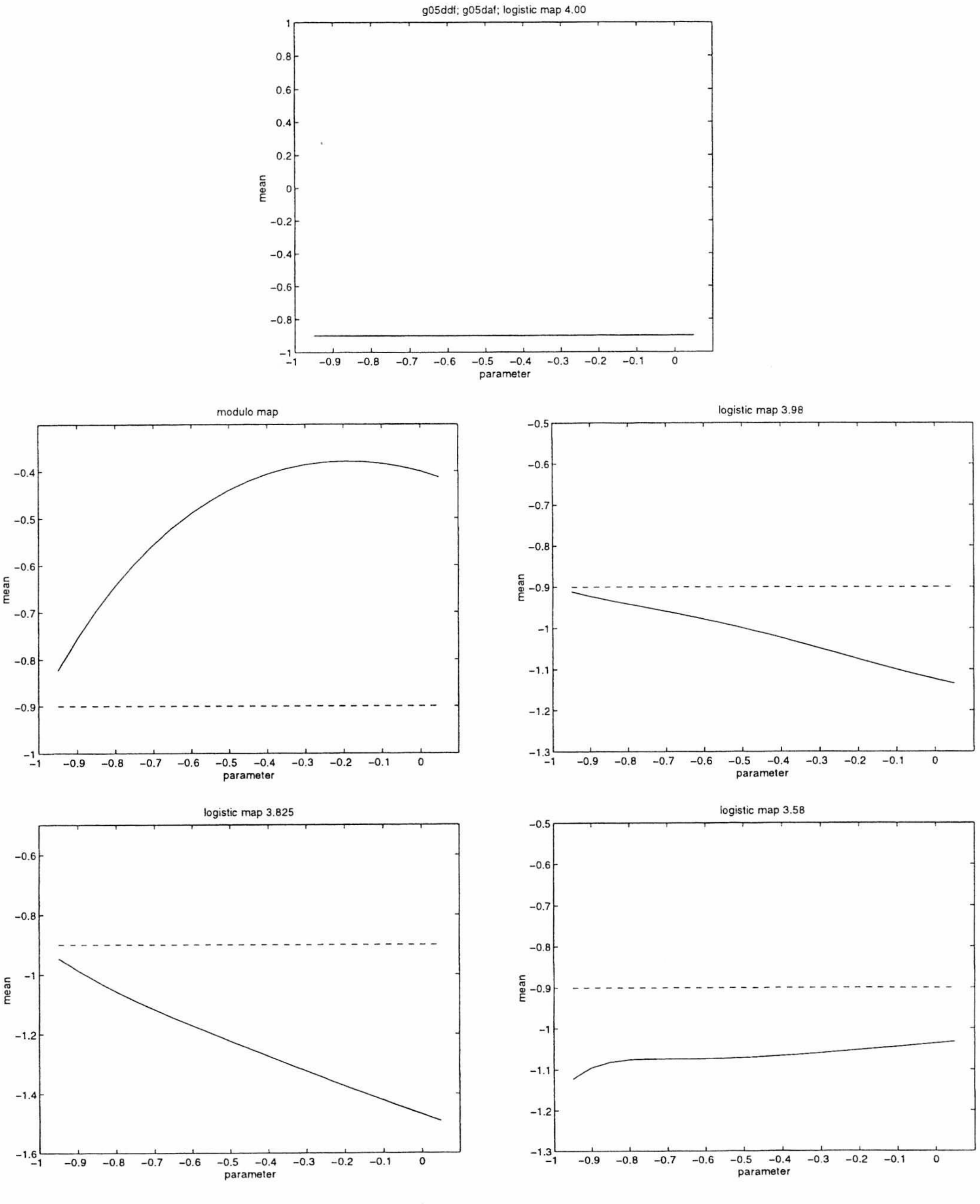


Figure 4.14: Theoretical means of the estimators of α_1 for the case $\alpha_1 = -0.9$, while α_2 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

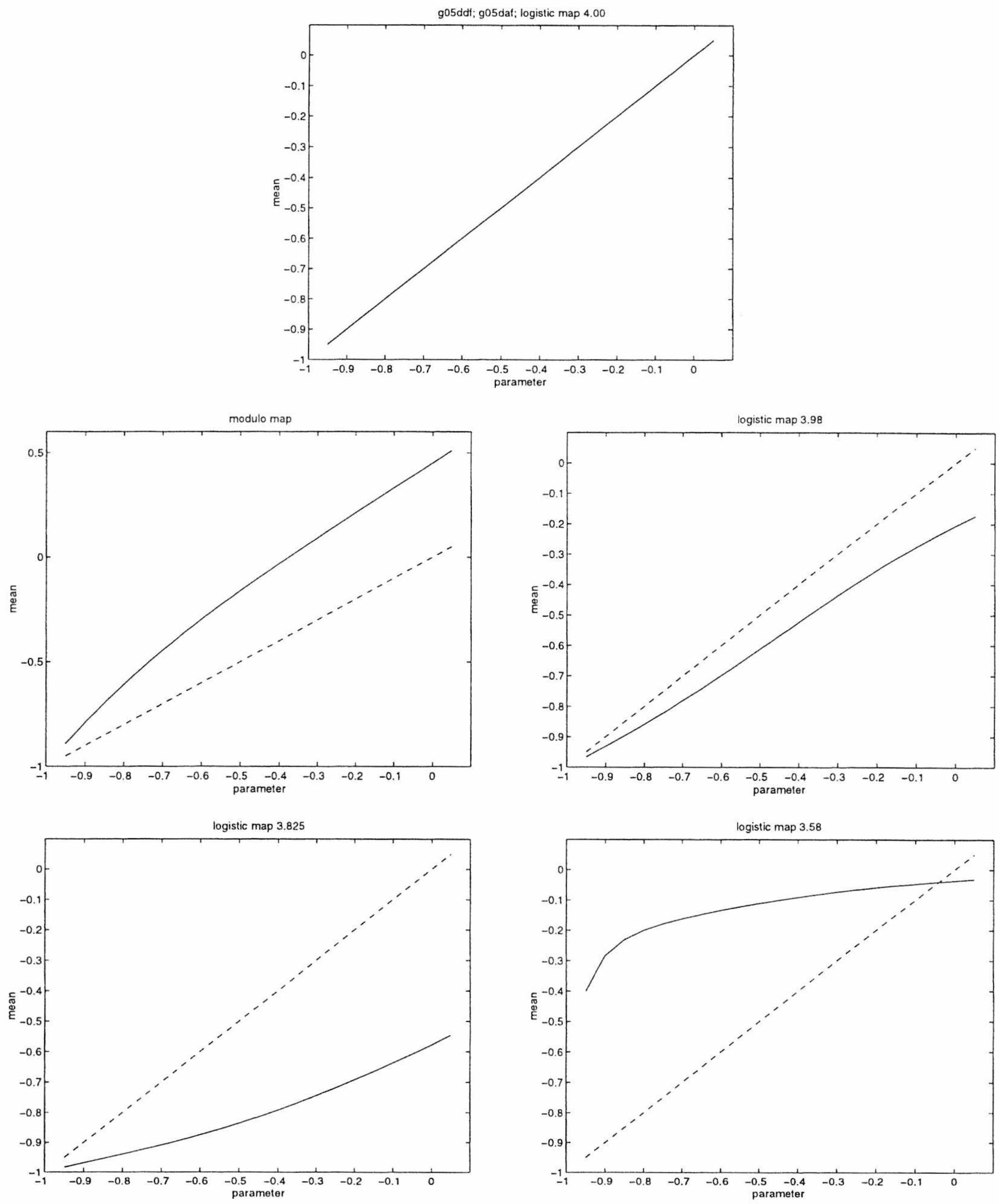


Figure 4.15: Theoretical means of the estimators of α_2 for the case $\alpha_1 = -0.9$, while α_2 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

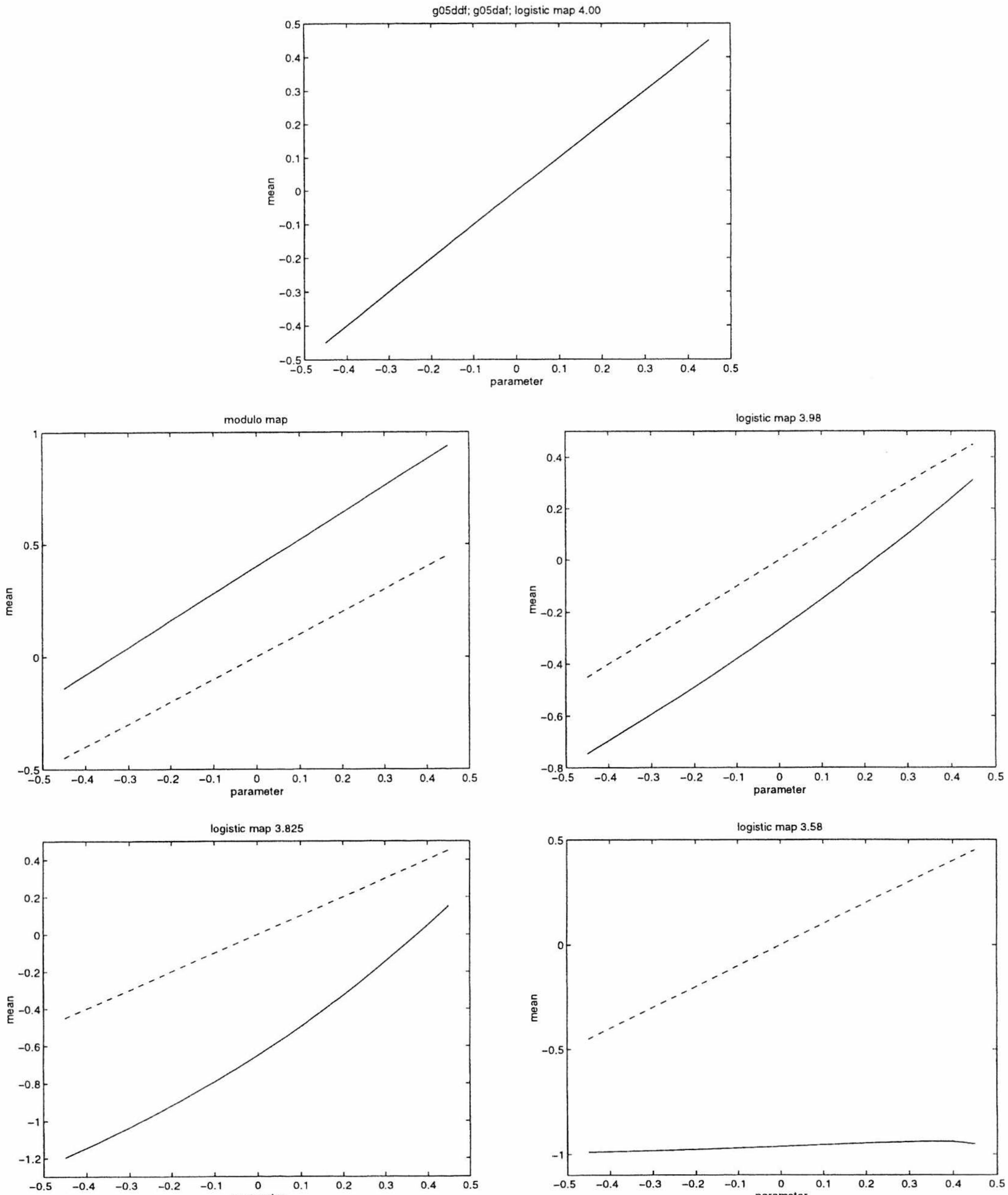


Figure 4.16: Theoretical means of the estimators of α_1 for the case $\alpha_2 = 0.5$, while α_1 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

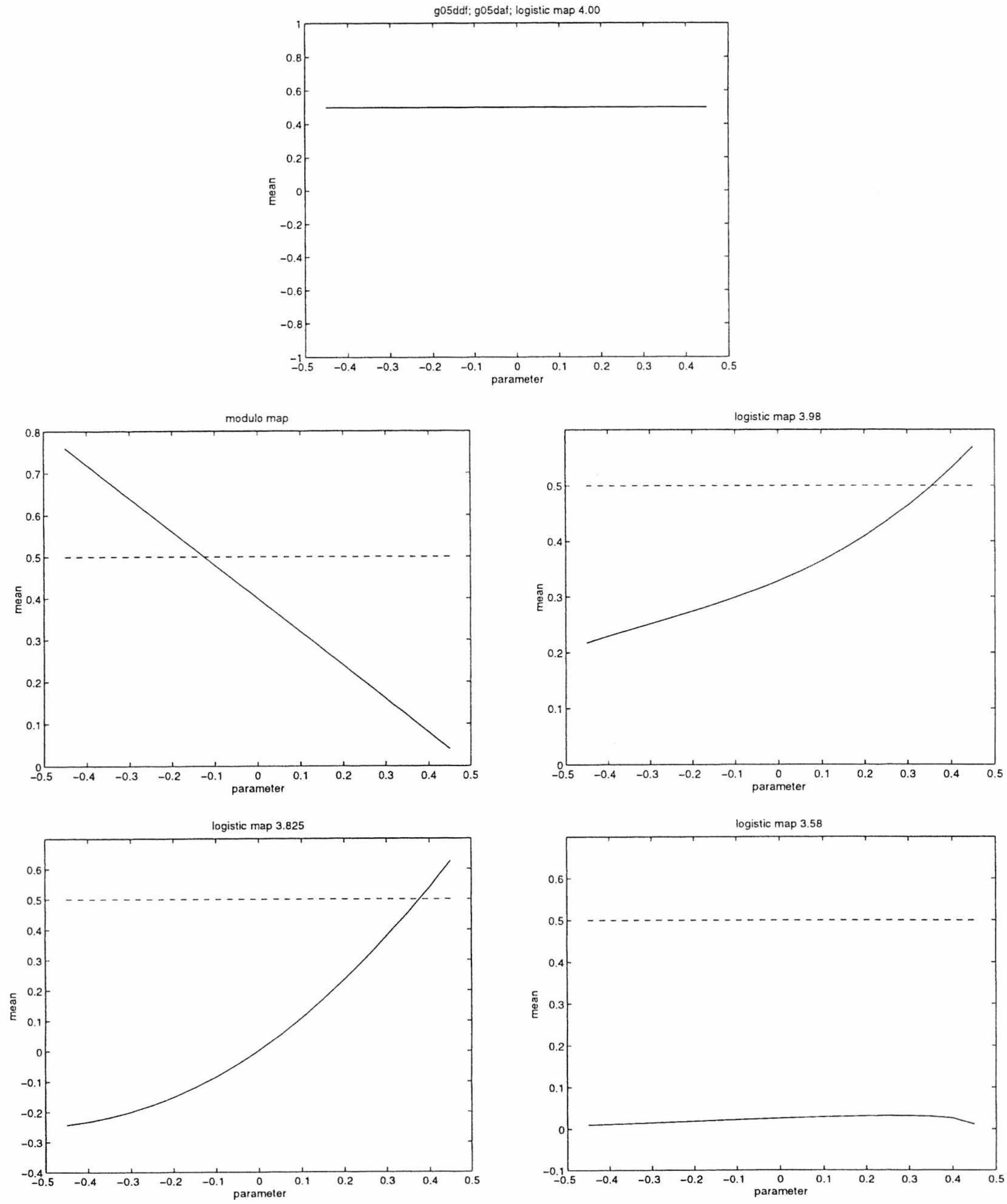


Figure 4.17: Theoretical means of the estimators of α_2 for the case $\alpha_2 = 0.5$, while α_1 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

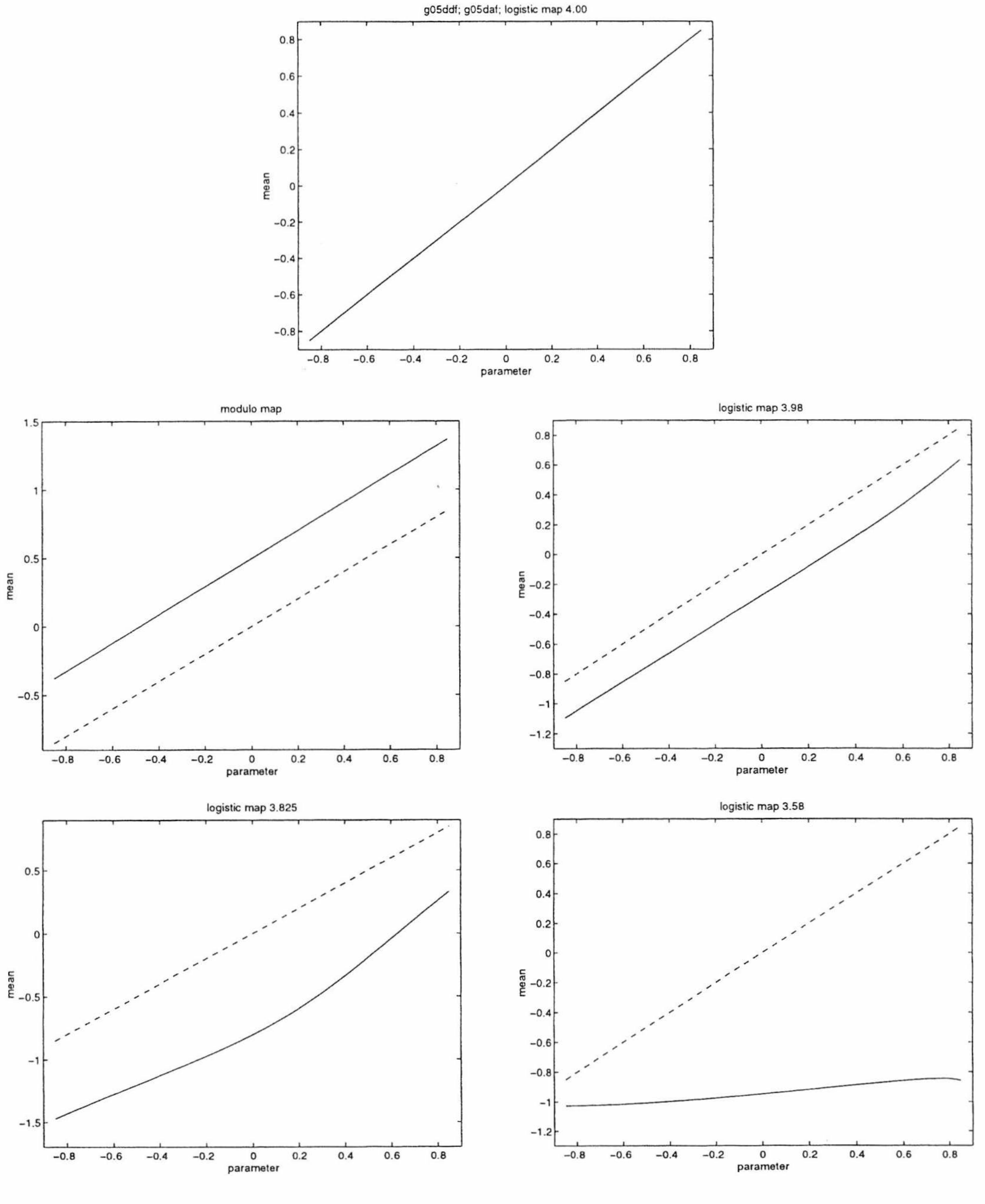


Figure 4.18: Theoretical means of the estimators of α_1 for the case $\alpha_2 = 0.1$, while α_1 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

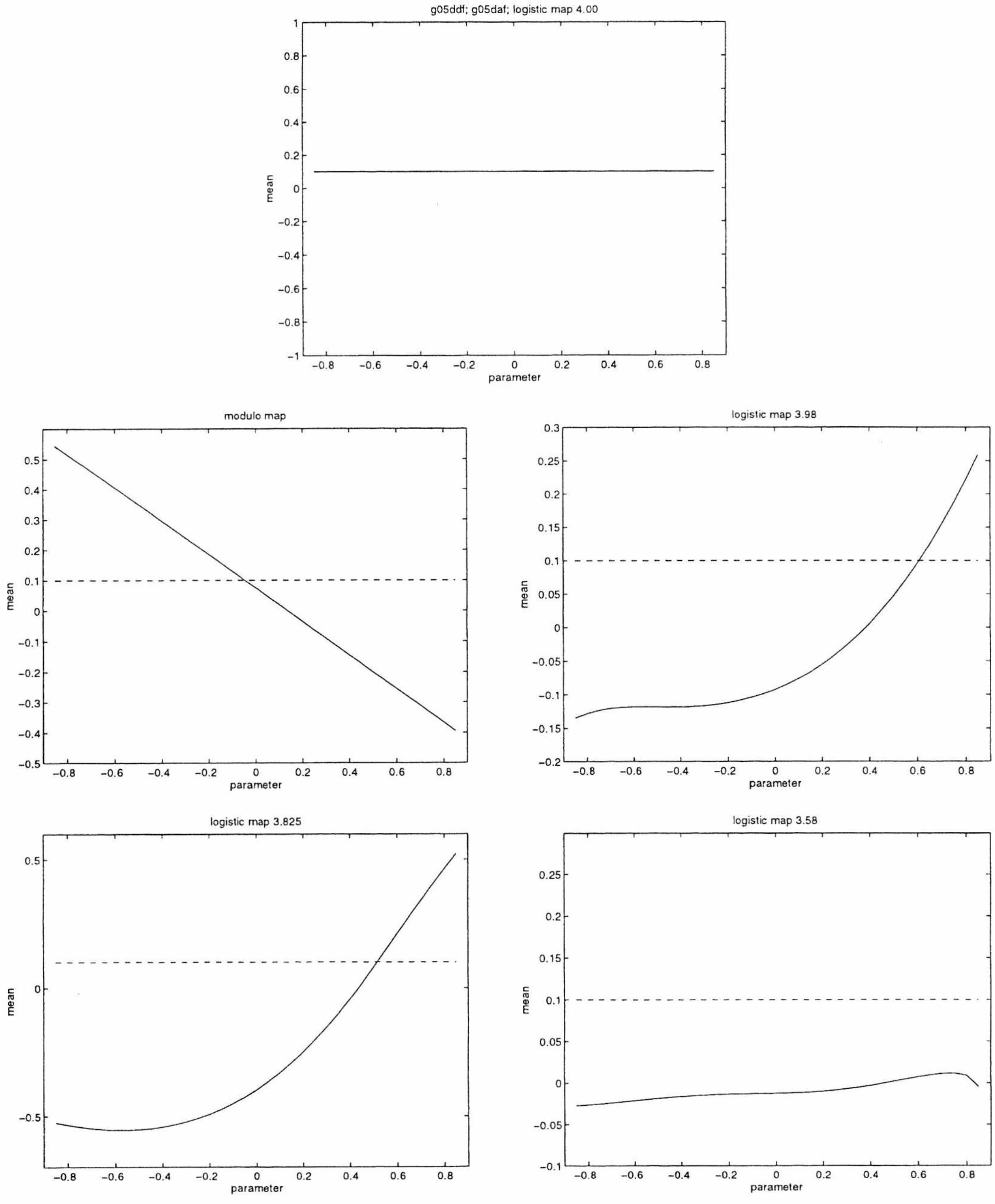


Figure 4.19: Theoretical means of the estimators of α_2 for the case $\alpha_2 = 0.1$, while α_1 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

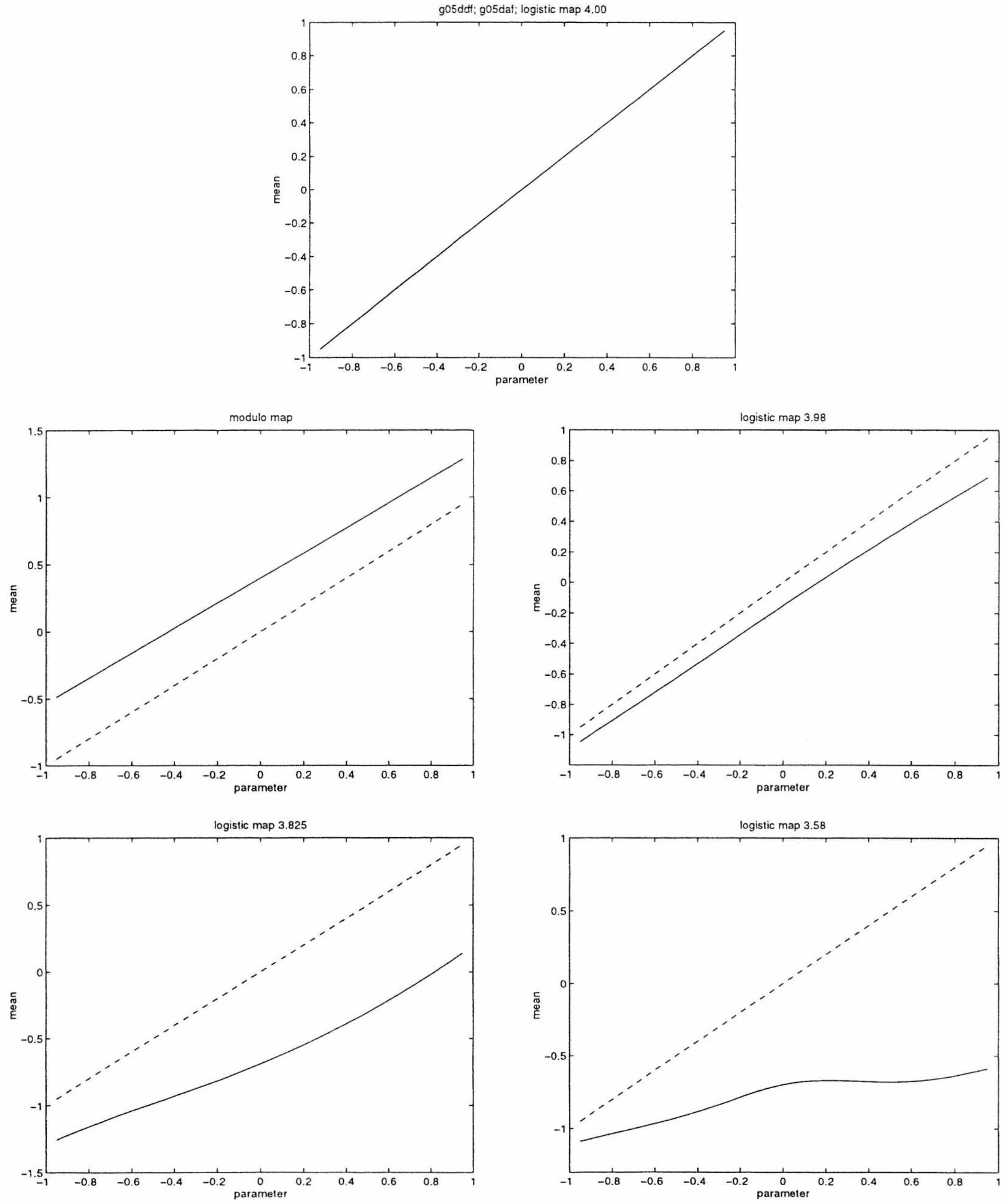


Figure 4.20: Theoretical means of the estimators of α_1 for the case $\alpha_2 = -0.5$, while α_1 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

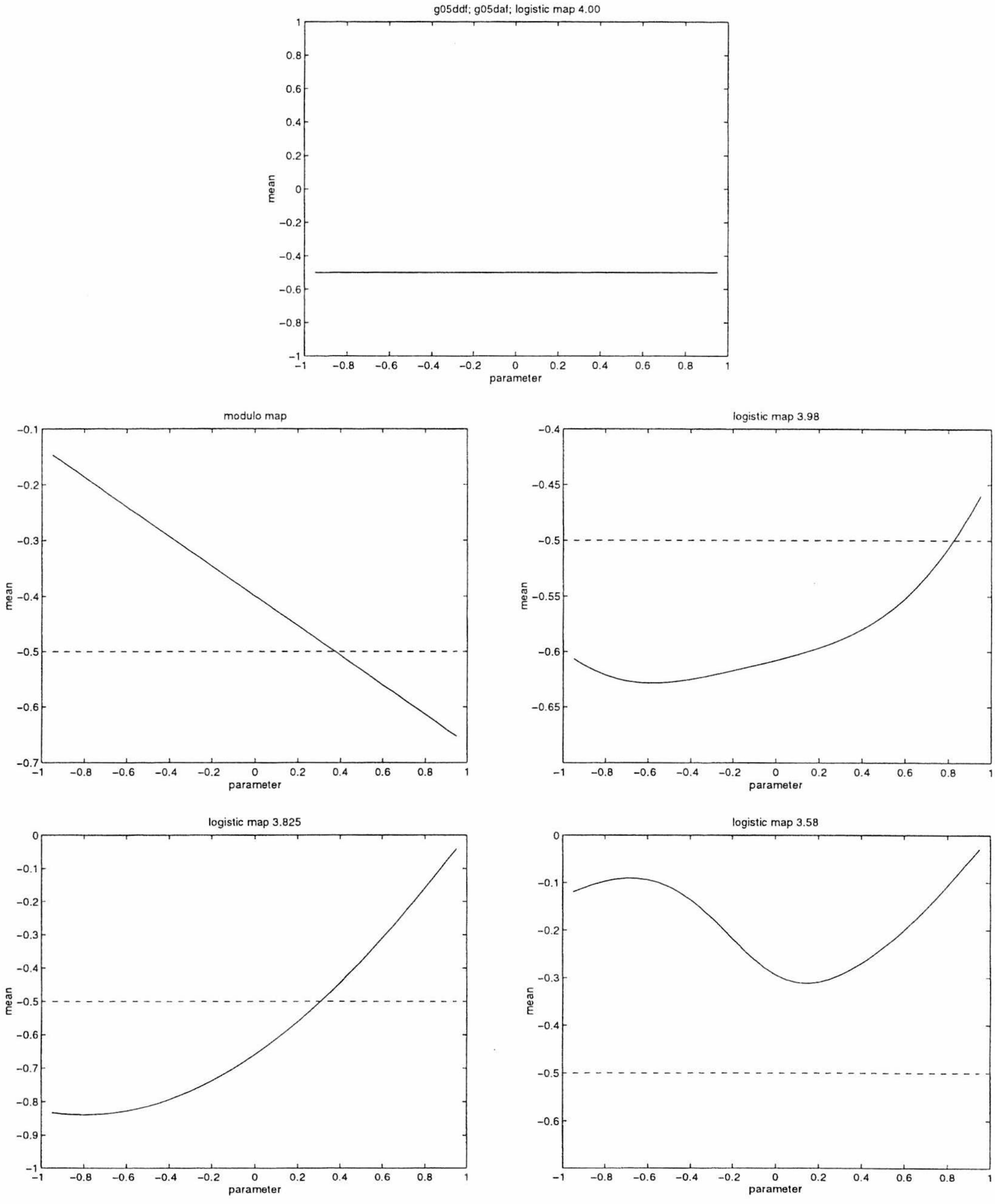


Figure 4.21: Theoretical means of the estimators of α_2 for the case $\alpha_2 = -0.5$, while α_1 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

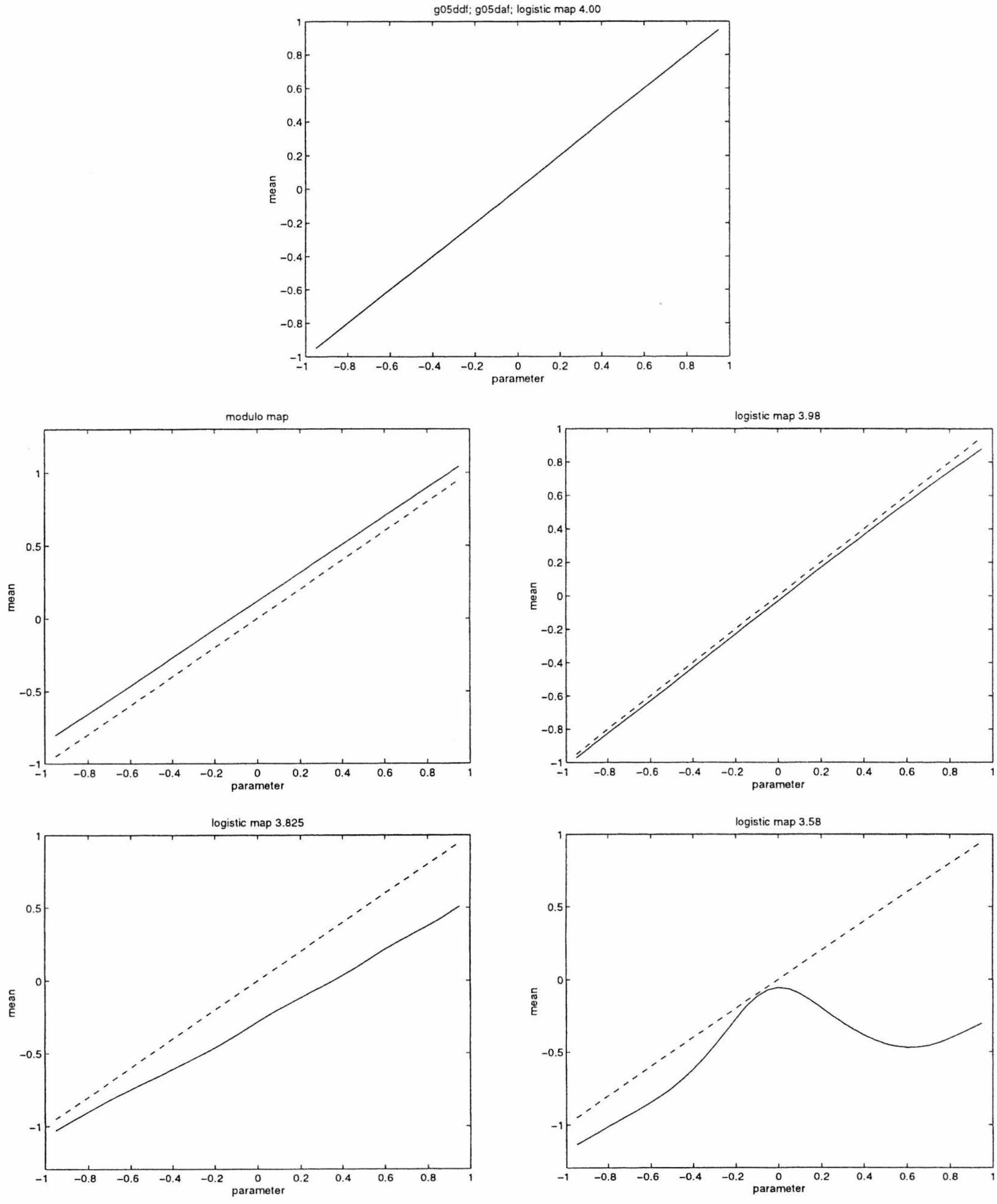


Figure 4.22: Theoretical means of the estimators of α_1 for the case $\alpha_2 = -0.9$, while α_1 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

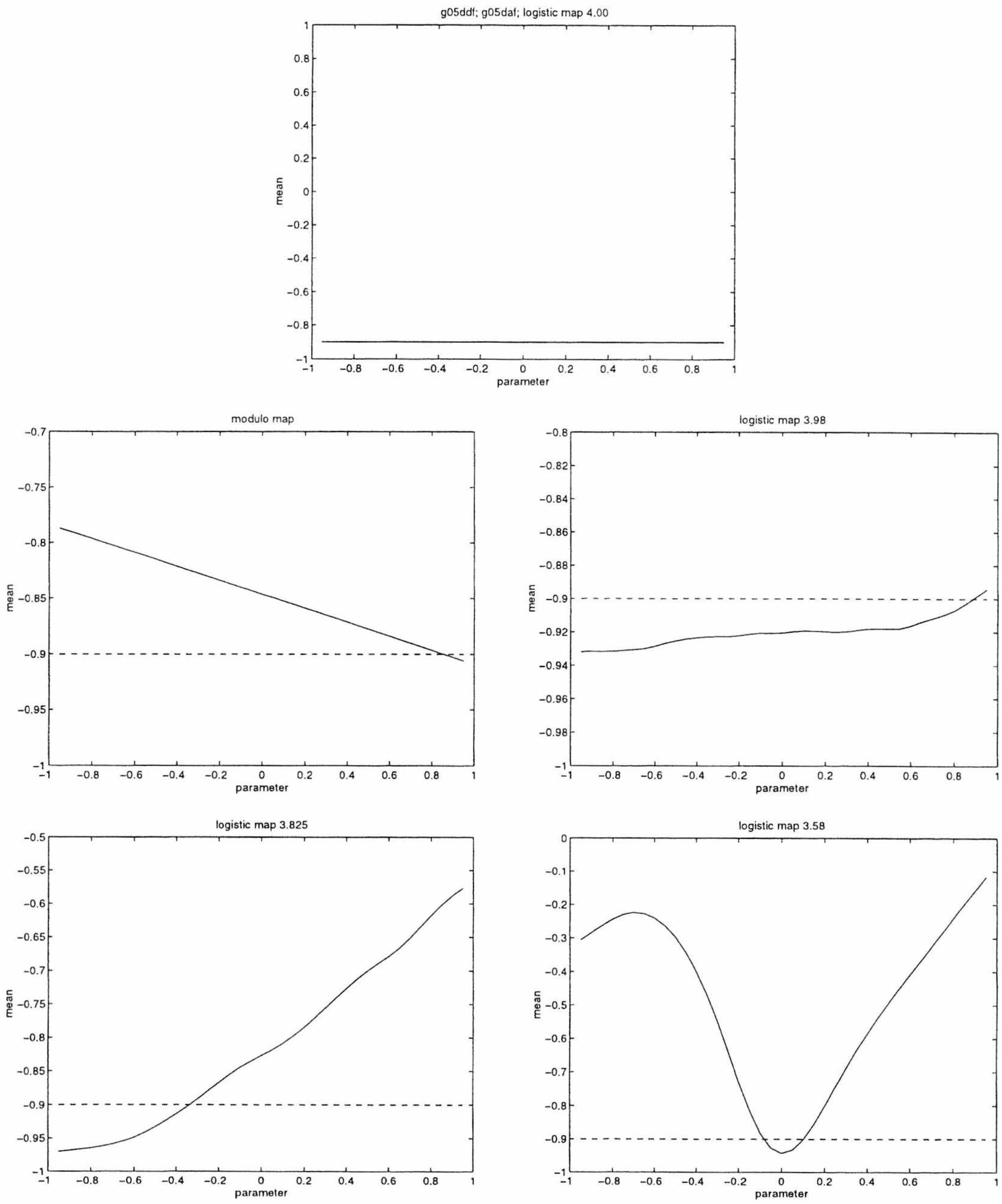


Figure 4.23: Theoretical means of the estimators of α_2 for the case $\alpha_2 = -0.9$, while α_1 varies through the range of values for which the AR(2) model is causal. The asymptotic means are based on Theorem 4.5 (solid line on the graphs). When different from the theoretical values, the unbiased means of the *iid* case are displayed (dashed line on the graphs).

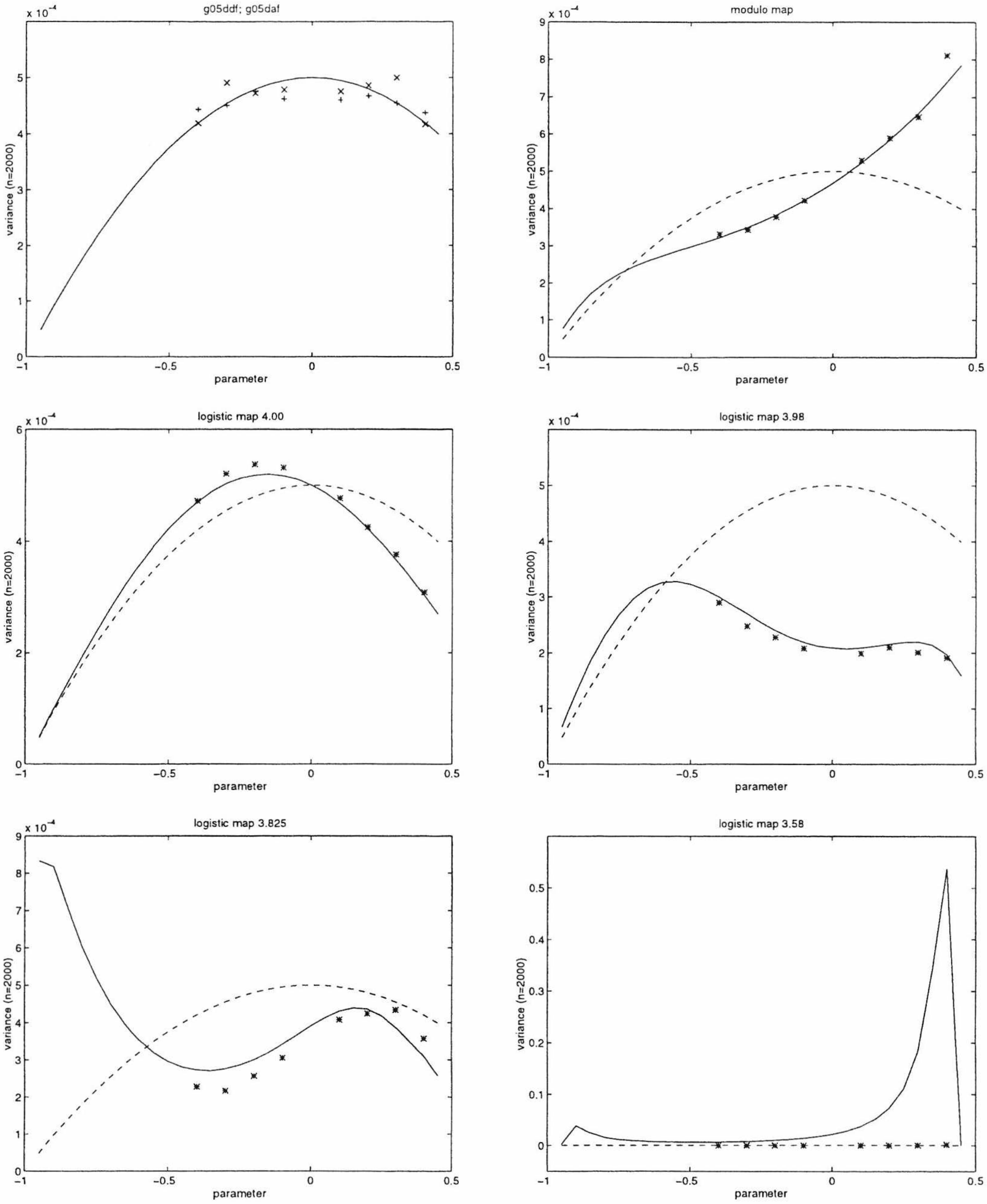


Figure 4.24: Variances of the estimators of α_1 for the case $\alpha_1 = 0.5$, while α_2 varies through the range of values for which the AR(2) model is causal.
The simulated variances are obtained using 2000 replications of $\hat{\alpha}'_n$; n, the estimators sample size, is taken equal to 2000. Simulated variances (denoted by an asterisk on the graphs) are obtained for eight values of α_2 (-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3 and 0.4). The theoretical (asymptotic) variances are based on the results of our paper (solid line). (The asymptotic variances of the last three logistic maps are computed using Theorem 4.8 although these maps do not satisfy its conditions. They are included for the sake of curiosity.) When different from the theoretical values, the variances in the *iid* case are displayed (dashed line). Note that in the first graph, '+' and 'x' respectively denote the simulated variances for g05ddf and g05daf.

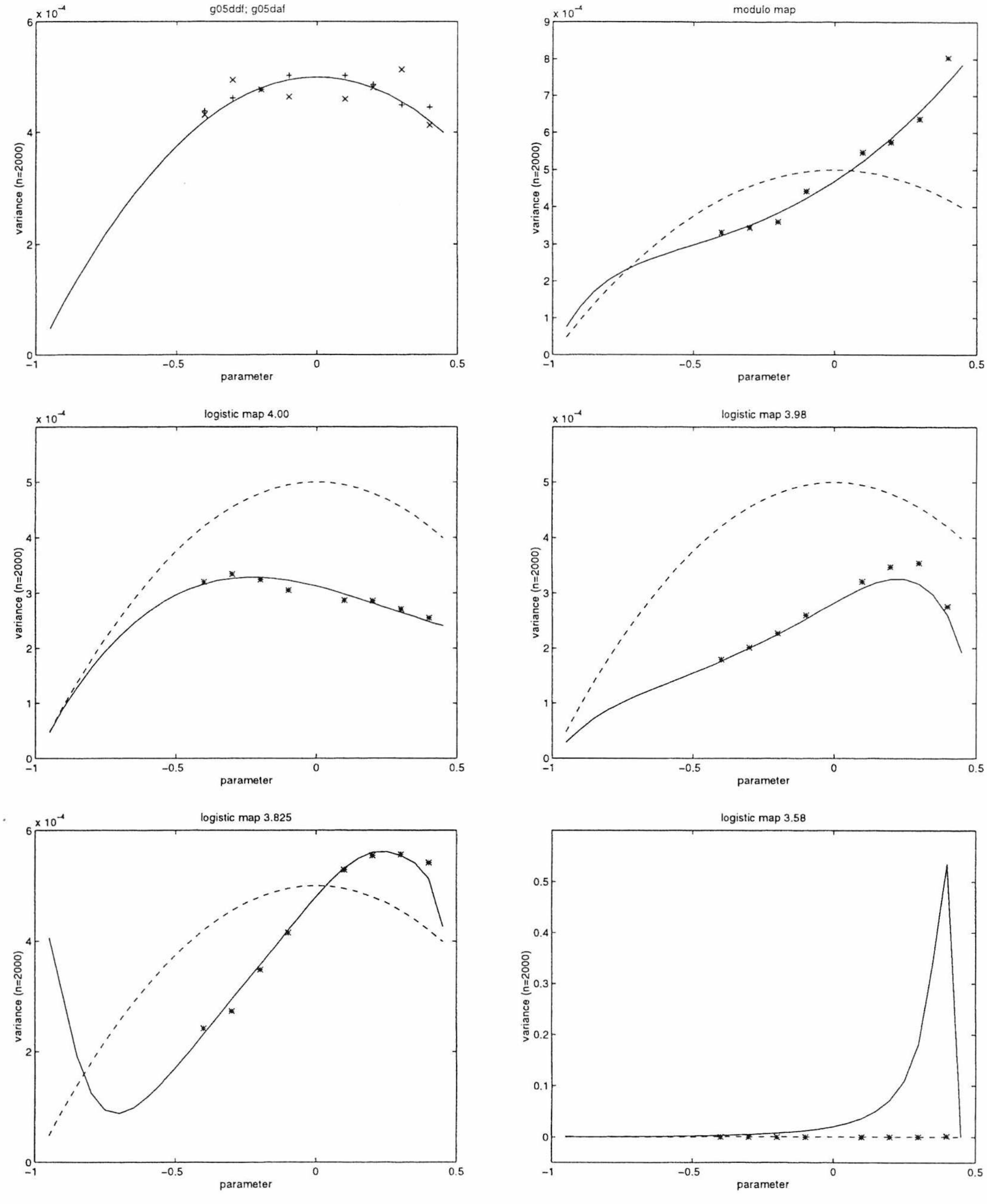


Figure 4.25: Variances of the estimators of α_2 for the case $\alpha_1 = 0.5$, while α_2 varies through the range of values for which the AR(2) model is causal.
The simulated variances are obtained using 2000 replications of $\hat{\alpha}'_n$; n, the estimators sample size, is taken equal to 2000. Simulated variances (denoted by an asterisk on the graphs) are obtained for eight values of α_2 (-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3 and 0.4). The theoretical (asymptotic) variances are based on the results of our paper (solid line). (The asymptotic variances of the last three logistic maps are computed using Theorem 4.8 although these maps do not satisfy its conditions. They are included for the sake of curiosity.) When different from the theoretical values, the variances in the *iid* case are displayed (dashed line). Note that in the first graph, '+' and 'x' respectively denote the simulated variances for g05ddf and g05daf.

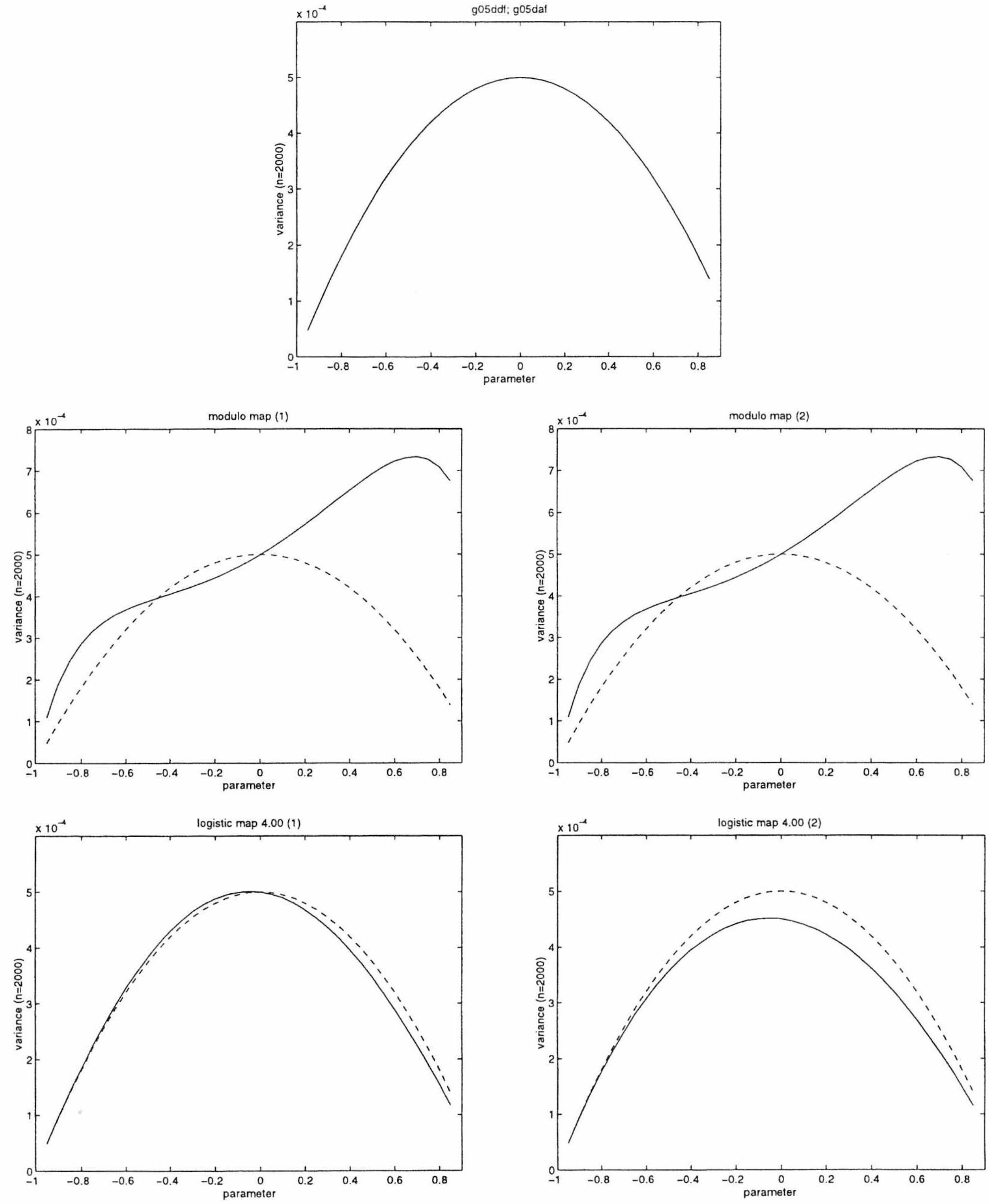


Figure 4.26: Theoretical variances of the estimators of α_1 and α_2 for the case $\alpha_1 = 0.1$, while α_2 varies through the range of values for which the AR(2) model is causal.
On the graphs labelled (1) and (2) the solid lines show the theoretical variances of the estimators of α_1 and α_2 respectively. There is only one graph in the *iid* case since it is then well-known that the theoretical variances of the estimators of α_1 and those of the estimators of α_2 have the same values. When different from the theoretical values, the variances of the *iid* case are displayed (dashed line on the graphs).

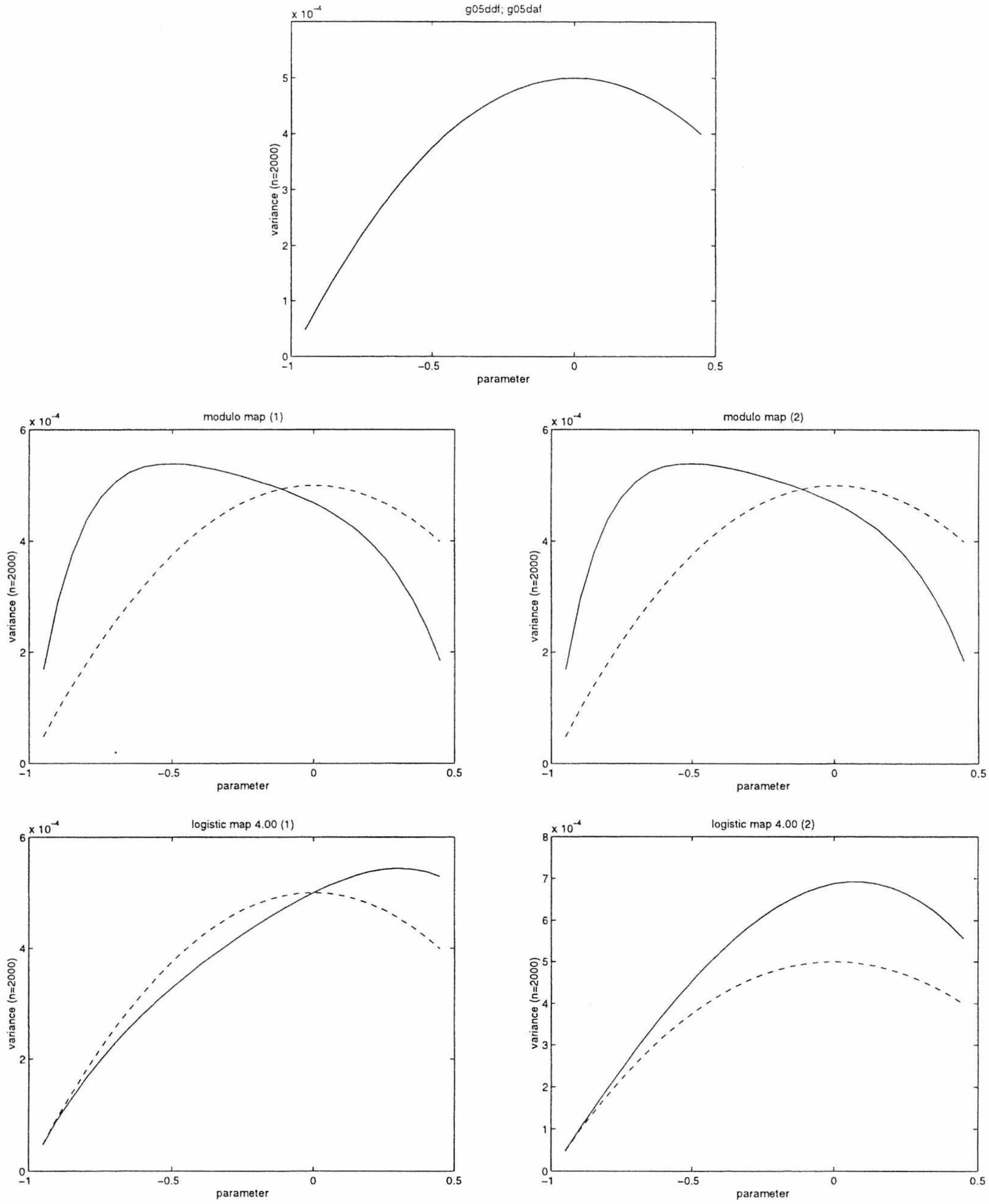


Figure 4.27: Theoretical variances of the estimators of α_1 and α_2 for the case $\alpha_1 = -0.5$, while α_2 varies through the range of values for which the AR(2) model is causal.
On the graphs labelled (1) and (2) the solid lines show the theoretical variances of the estimators of α_1 and α_2 respectively. There is only one graph in the *iid* case since it is then well-known that the theoretical variances of the estimators of α_1 and those of the estimators of α_2 have the same values. When different from the theoretical values, the variances of the *iid* case are displayed (dashed line on the graphs).

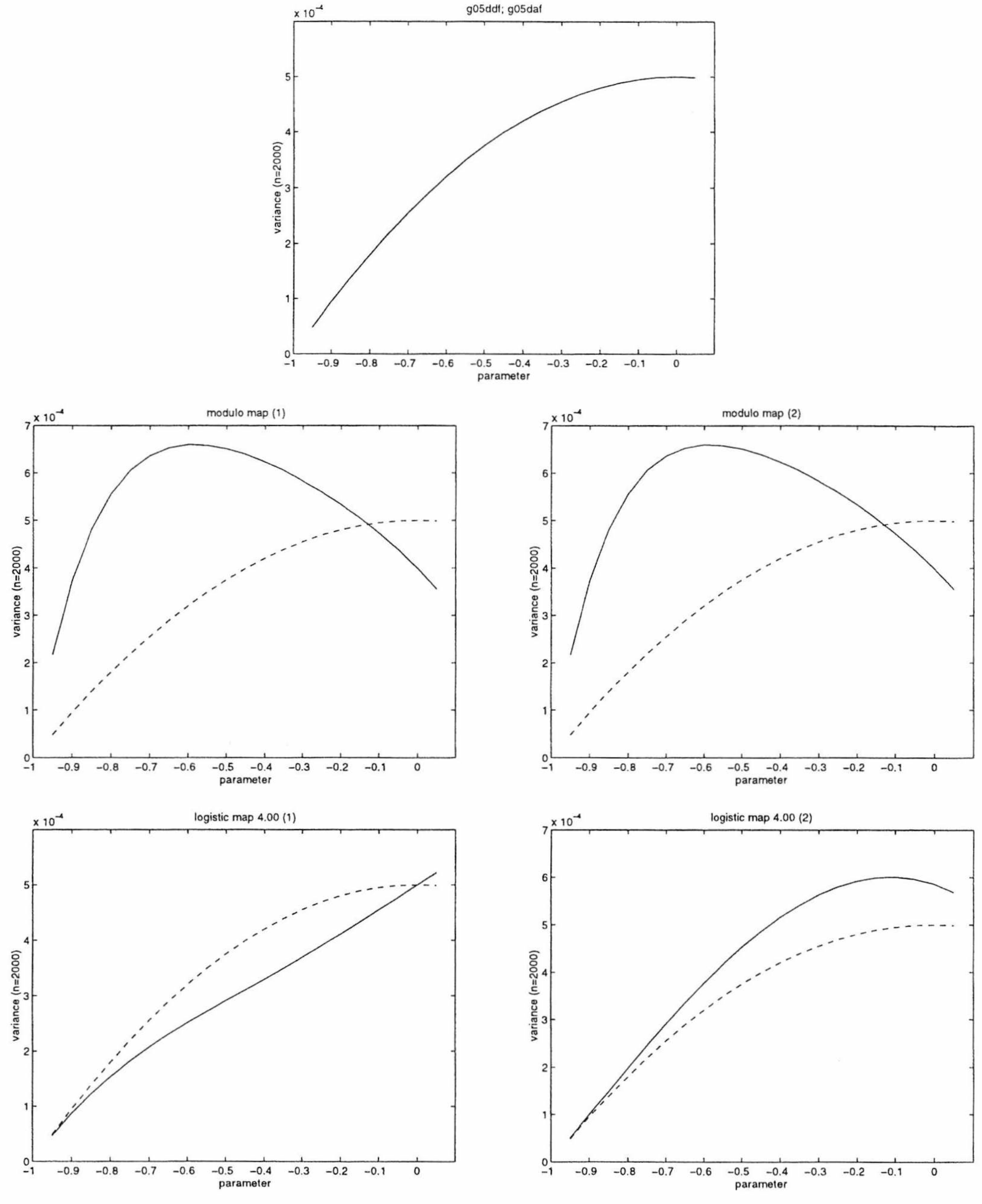


Figure 4.28: Theoretical variances of the estimators of α_1 and α_2 for the case $\alpha_1 = -0.9$, while α_2 varies through the range of values for which the AR(2) model is causal.
On the graphs labelled (1) and (2) the solid lines show the theoretical variances of the estimators of α_1 and α_2 respectively. There is only one graph in the *iid* case since it is then well-known that the theoretical variances of the estimators of α_1 and those of the estimators of α_2 have the same values. When different from the theoretical values, the variances of the *iid* case are displayed (dashed line on the graphs).

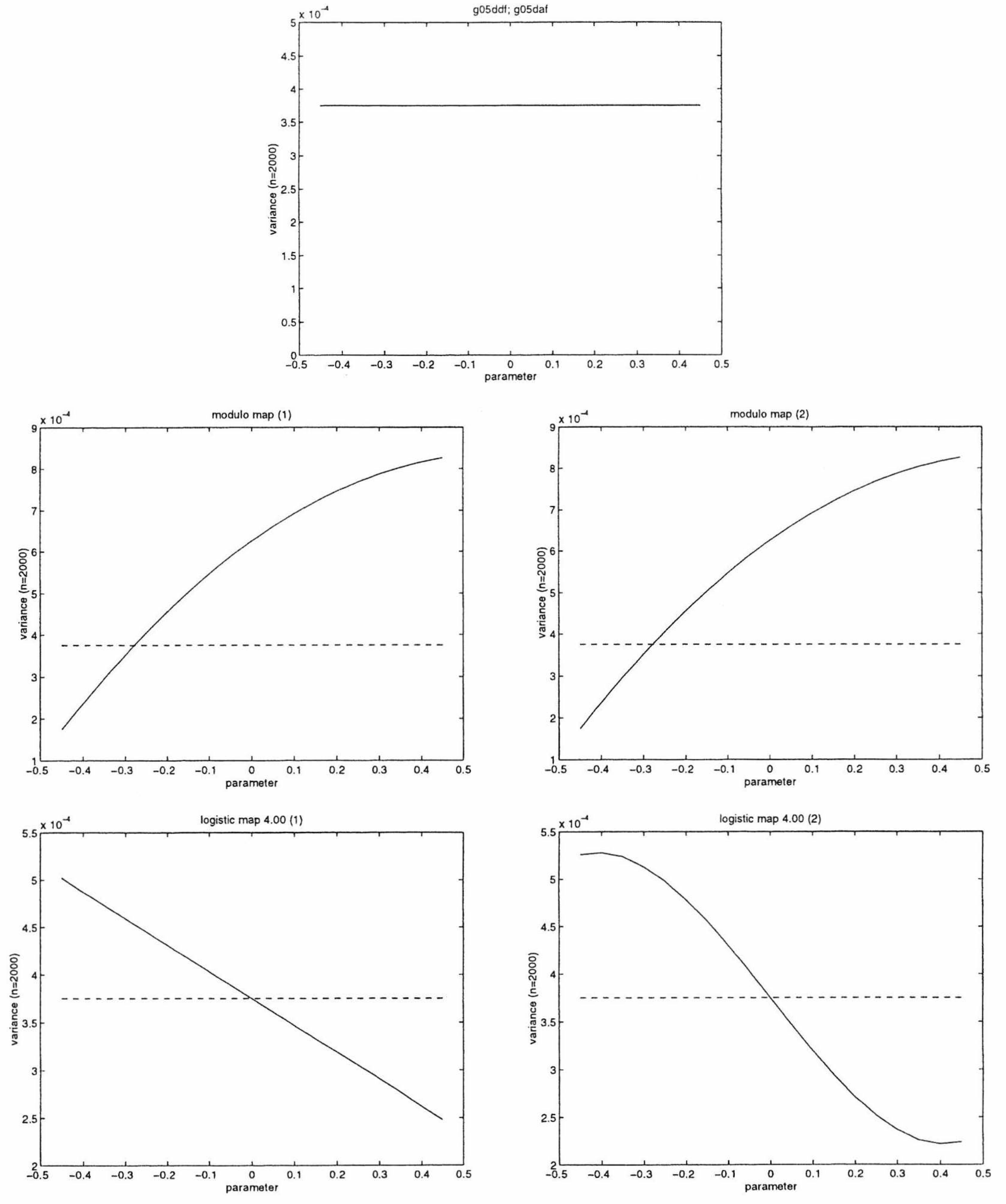


Figure 4.29: Theoretical variances of the estimators of α_1 and α_2 for the case $\alpha_2 = 0.5$, while α_1 varies through the range of values for which the AR(2) model is causal.
On the graphs labelled (1) and (2) the solid lines show the theoretical variances of the estimators of α_1 and α_2 respectively. There is only one graph in the *iid* case since it is then well-known that the theoretical variances of the estimators of α_1 and those of the estimators of α_2 have the same values. When different from the theoretical values, the variances of the *iid* case are displayed (dashed line on the graphs).

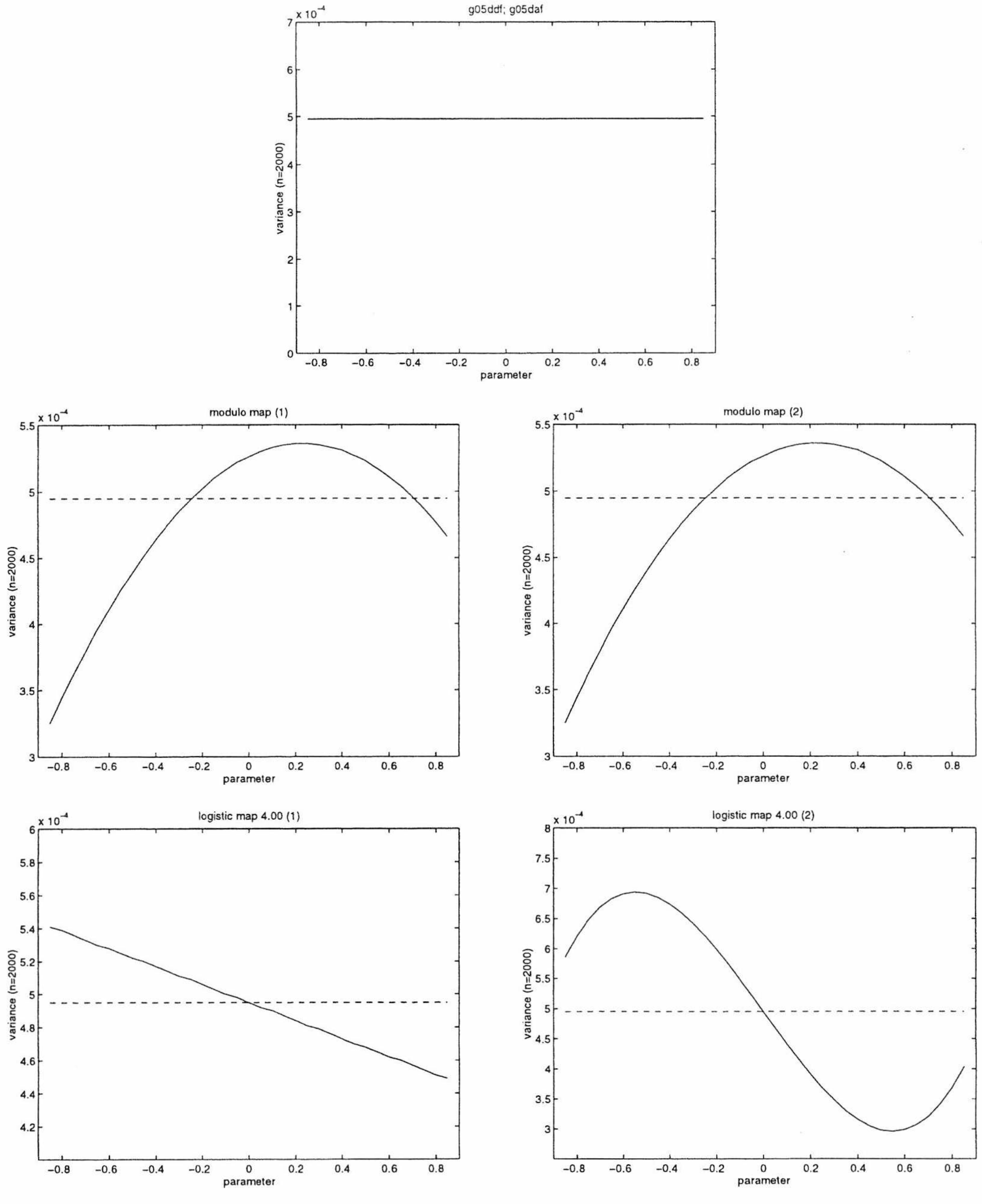


Figure 4.30: Theoretical variances of the estimators of α_1 and α_2 for the case $\alpha_2 = 0.1$, while α_1 varies through the range of values for which the AR(2) model is causal.
On the graphs labelled (1) and (2) the solid lines show the theoretical variances of the estimators of α_1 and α_2 respectively. There is only one graph in the *iid* case since it is then well-known that the theoretical variances of the estimators of α_1 and those of the estimators of α_2 have the same values. When different from the theoretical values, the variances of the *iid* case are displayed (dashed line on the graphs).

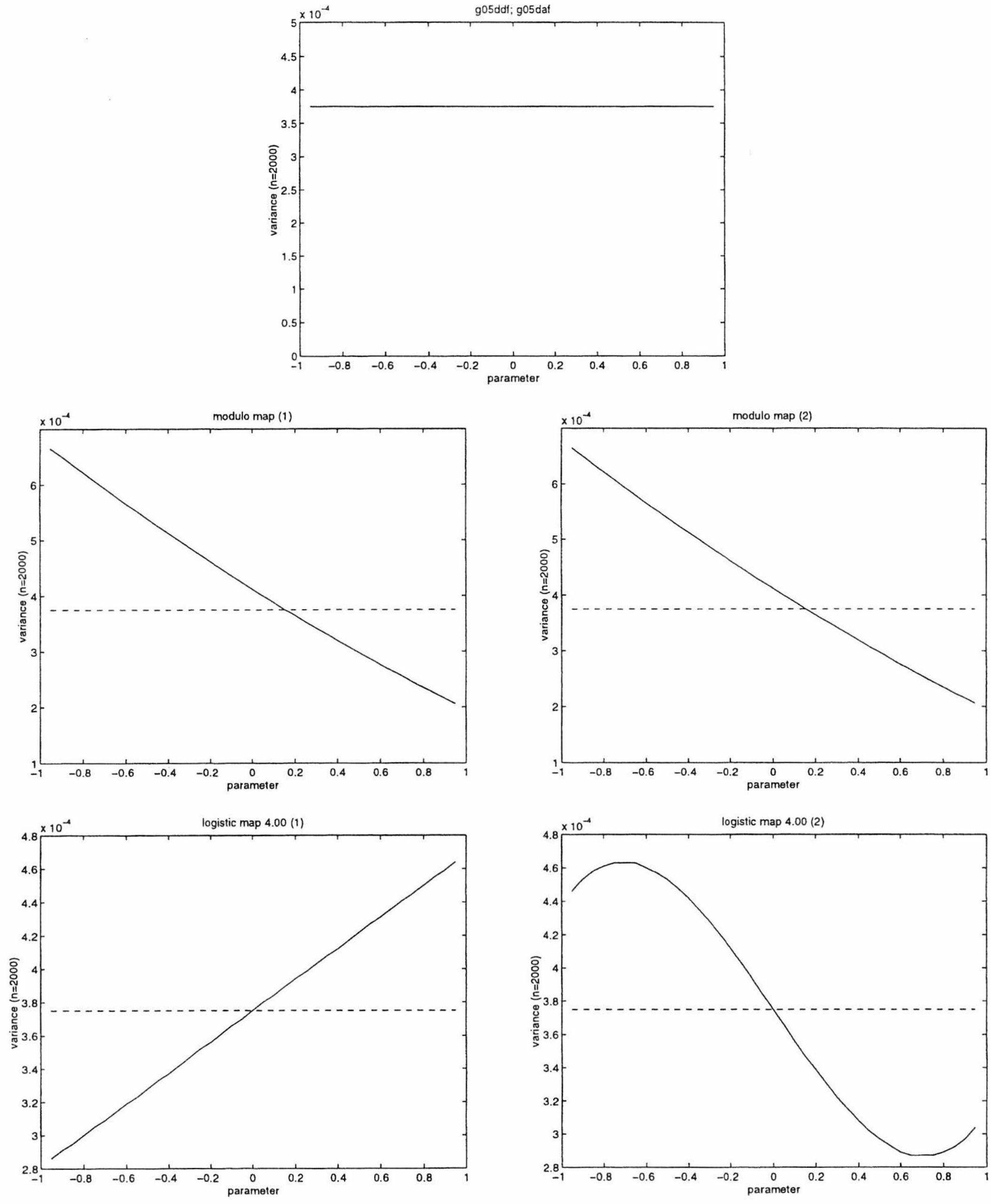


Figure 4.31: Theoretical variances of the estimators of α_1 and α_2 for the case $\alpha_2 = -0.5$, while α_1 varies through the range of values for which the AR(2) model is causal.
On the graphs labelled (1) and (2) the solid lines show the theoretical variances of the estimators of α_1 and α_2 respectively. There is only one graph in the *iid* case since it is then well-known that the theoretical variances of the estimators of α_1 and those of the estimators of α_2 have the same values. When different from the theoretical values, the variances of the *iid* case are displayed (dashed line on the graphs).

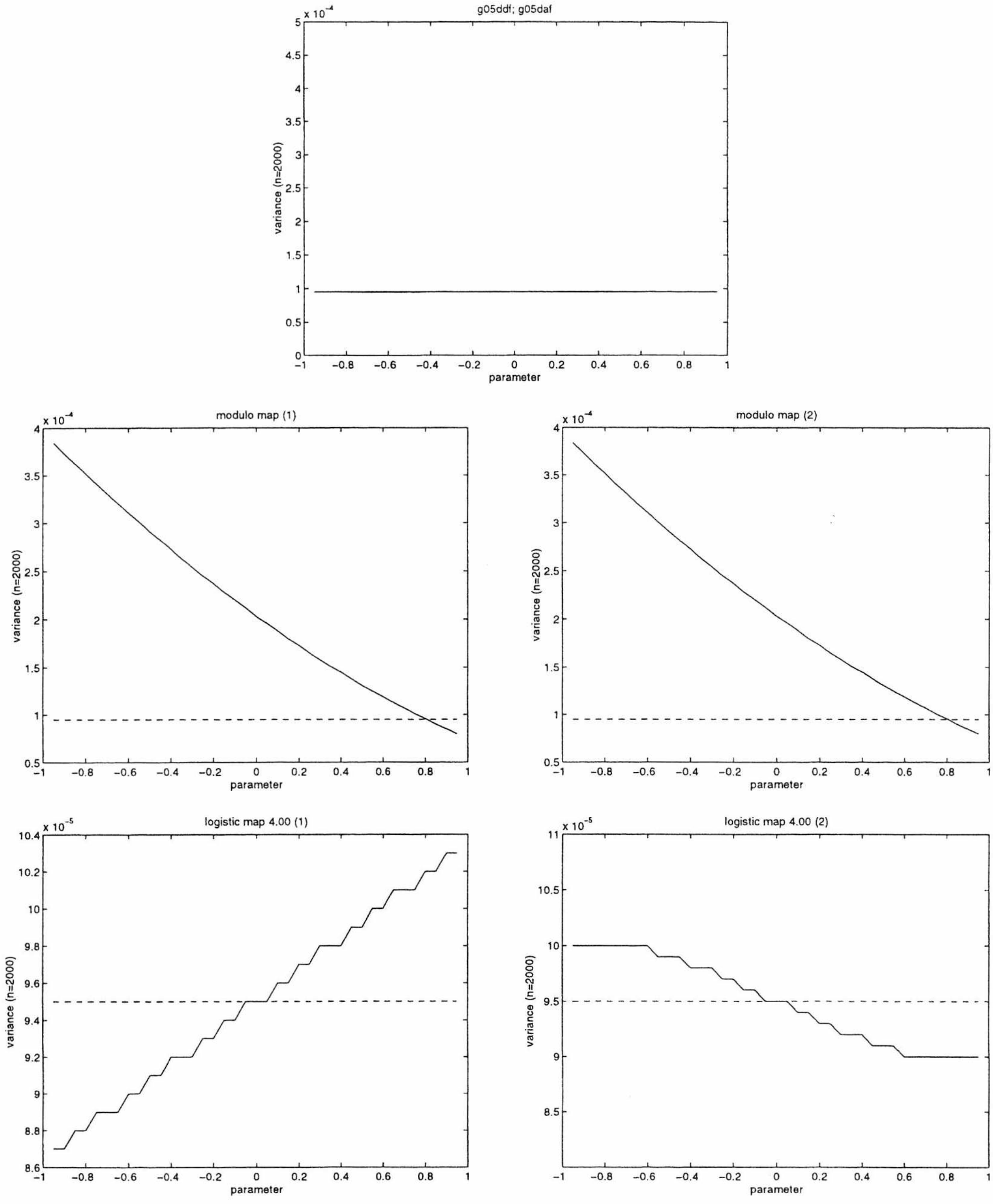


Figure 4.32: Theoretical variances of the estimators of α_1 and α_2 for the case $\alpha_2 = -0.9$, while α_1 varies through the range of values for which the AR(2) model is causal.
On the graphs labelled (1) and (2) the solid lines show the theoretical variances of the estimators of α_1 and α_2 respectively. There is only one graph in the *iid* case since it is then well-known that the theoretical variances of the estimators of α_1 and those of the estimators of α_2 have the same values. When different from the theoretical values, the variances of the *iid* case are displayed (dashed line on the graphs).

	Simulated skewness	Simulated kurtosis	Lin-Mudholkar test statistic
g05ddf			
$\hat{\alpha}_1$			
α_2 = - 0.4	0.056143	3.025619	- 0.499687
α_2 = - 0.3	0.047395	2.933155	- 0.432614
α_2 = - 0.2	- 0.016495	3.113604	0.147530
α_2 = - 0.1	- 0.092498	3.185653	0.834433
α_2 = 0.1	- 0.098066	3.215379	0.880149
α_2 = 0.2	- 0.040295	3.146723	0.360567
α_2 = 0.3	0.019822	2.939489	- 0.182022
α_2 = 0.4	0.000841	2.999467	- 0.007143
$\hat{\alpha}_2$			
α_2 = - 0.4	0.064693	2.827570	- 0.603556
α_2 = - 0.3	0.032865	2.960944	- 0.299779
α_2 = - 0.2	- 0.043812	3.009757	0.405766
α_2 = - 0.1	0.016805	3.028521	- 0.151449
α_2 = 0.1	0.008856	2.939526	- 0.082105
α_2 = 0.2	- 0.013689	3.166596	0.120540
α_2 = 0.3	- 0.087408	2.979284	0.828541
α_2 = 0.4	- 0.074197	2.962823	0.702954

Table 4.2: Sample skewness, sample kurtosis and Lin-Mudholkar test statistics are computed for the case $\alpha_1 = 0.5$, while α_2 takes different values.
The simulations are obtained using 2000 replications of $\hat{\alpha}_{i,n}$, $i = 1, 2$ and $n = 2000$.

Table 4.2 continued

g05daf $\hat{\alpha}_1$ $\alpha_2 =$ - 0.4	- 0.044829	2.990984	0.417147
$\alpha_2 =$ - 0.3	- 0.081705	2.781583	0.815780
$\alpha_2 =$ - 0.2	0.028445	3.025214	- 0.254902
$\alpha_2 =$ - 0.1	0.044251	2.969529	- 0.400858
$\alpha_2 = 0.1$	0.104280	3.007476	- 0.917276
$\alpha_2 = 0.2$	0.057774	2.952897	- 0.522681
$\alpha_2 = 0.3$	0.043040	2.769650	- 0.411203
$\alpha_2 = 0.4$	0.012764	2.961361	- 0.117167
$\hat{\alpha}_2$ $\alpha_2 =$ - 0.4	0.164917	2.948780	- 1.442408
$\alpha_2 =$ - 0.3	0.072142	3.053276	- 0.634789
$\alpha_2 =$ - 0.2	- 0.063335	2.917163	0.604258
$\alpha_2 =$ - 0.1	- 0.013733	3.087221	0.123077
$\alpha_2 = 0.1$	- 0.001236	3.214640	0.010602
$\alpha_2 = 0.2$	- 0.114359	2.842211	1.136700
$\alpha_2 = 0.3$	- 0.009161	2.929267	0.085588
$\alpha_2 = 0.4$	- 0.032681	3.015771	0.300081
modulo map $\hat{\alpha}_1$ $\alpha_2 =$ - 0.4	0.085024	3.016447	- 0.750995
$\alpha_2 =$ - 0.3	- 0.115458	2.953179	1.114223
$\alpha_2 =$ - 0.2	- 0.057722	3.090354	0.526690
$\alpha_2 =$ - 0.1	- 0.073422	2.968918	0.694670
$\alpha_2 = 0.1$	- 0.046603	2.977041	0.435785
$\alpha_2 = 0.2$	- 0.019582	3.191064	0.172211
$\alpha_2 = 0.3$	- 0.081664	3.009592	0.766757
$\alpha_2 = 0.4$	0.028477	2.962145	- 0.259843
$\hat{\alpha}_2$ $\alpha_2 =$ - 0.4	0.065479	2.902191	- 0.599386
$\alpha_2 =$ - 0.3	0.105722	2.890995	- 0.956616
$\alpha_2 =$ - 0.2	0.054996	2.940836	- 0.500338
$\alpha_2 =$ - 0.1	0.068601	3.073131	- 0.601052
$\alpha_2 = 0.1$	0.034717	3.081857	- 0.306973
$\alpha_2 = 0.2$	- 0.014552	3.186395	0.127495
$\alpha_2 = 0.3$	0.100850	2.922849	- 0.907290
$\alpha_2 = 0.4$	0.008658	2.957621	- 0.079855

Table 4.2 continued

logistic map $\theta = 4.00$ $\hat{\alpha}_1$ $\alpha_2 =$ - 0.4	- 0.005504	2.988565	0.050264
$\alpha_2 =$ - 0.3	- 0.149464	3.066566	1.417812
$\alpha_2 =$ - 0.2	- 0.152481	3.068790	1.447195
$\alpha_2 =$ - 0.1	- 0.126512	3.236125	1.142452
$\alpha_2 = 0.1$	- 0.128964	3.259592	1.159435
$\alpha_2 = 0.2$	0.024498	2.795138	- 0.234049
$\alpha_2 = 0.3$	- 0.107751	3.159073	0.983346
$\alpha_2 = 0.4$	0.046864	2.812451	- 0.441824
$\hat{\alpha}_2$ $\alpha_2 =$ - 0.4	0.075327	3.037970	- 0.664513
$\alpha_2 =$ - 0.3	0.064757	3.125559	- 0.560912
$\alpha_2 =$ - 0.2	0.072896	3.192504	- 0.620957
$\alpha_2 =$ - 0.1	0.077779	2.925641	- 0.704277
$\alpha_2 = 0.1$	0.053784	3.002637	- 0.481503
$\alpha_2 = 0.2$	- 0.040609	2.835461	0.393042
$\alpha_2 = 0.3$	0.045085	2.877693	- 0.418012
$\alpha_2 = 0.4$	- 0.116932	2.920388	1.138363
logistic map $\theta = 3.98$ $\hat{\alpha}_1$ $\alpha_2 =$ - 0.4	- 0.180173	3.120499	1.706449
$\alpha_2 =$ - 0.3	- 0.097416	2.935346	0.937792
$\alpha_2 =$ - 0.2	0.018411	3.041646	- 0.165246
$\alpha_2 =$ - 0.1	- 0.096803	3.015249	0.912523
$\alpha_2 = 0.1$	- 0.038719	3.008213	0.358266
$\alpha_2 = 0.2$	- 0.154427	3.045733	1.475744
$\alpha_2 = 0.3$	- 0.090338	3.055195	0.840785
$\alpha_2 = 0.4$	- 0.049485	2.849755	0.479127
$\hat{\alpha}_2$ $\alpha_2 =$ - 0.4	0.080220	3.015578	- 0.710049
$\alpha_2 =$ - 0.3	- 0.015999	3.201163	0.140222
$\alpha_2 =$ - 0.2	0.046973	2.942648	- 0.428120
$\alpha_2 =$ - 0.1	- 0.045464	2.930127	0.430171
$\alpha_2 = 0.1$	- 0.063257	2.920767	0.603740
$\alpha_2 = 0.2$	- 0.017877	2.957893	0.165695
$\alpha_2 = 0.3$	- 0.169963	3.141579	1.595645
$\alpha_2 = 0.4$	- 0.003875	3.019914	0.034863

Table 4.2 continued

logistic map $\theta = 3.825$ $\hat{\alpha}_1$ $\alpha_2 =$ - 0.4	- 0.070576	2.915737	0.676194
$\alpha_2 =$ - 0.3	- 0.016943	2.920389	0.158847
$\alpha_2 =$ - 0.2	- 0.018533	3.016826	0.169748
$\alpha_2 =$ - 0.1	- 0.055313	2.972819	0.518968
$\alpha_2 = 0.1$	- 0.131882	2.998510	1.265168
$\alpha_2 = 0.2$	- 0.075711	3.044402	0.702931
$\alpha_2 = 0.3$	- 0.058555	2.965702	0.551221
$\alpha_2 = 0.4$	- 0.122760	2.975999	1.180164
$\hat{\alpha}_2$ $\alpha_2 =$ - 0.4	- 0.020308	2.912147	0.191357
$\alpha_2 =$ - 0.3	0.004598	3.059198	- 0.041725
$\alpha_2 =$ - 0.2	- 0.017775	2.969566	0.164574
$\alpha_2 =$ - 0.1	0.072046	2.982660	- 0.644216
$\alpha_2 = 0.1$	0.003126	3.027819	- 0.027977
$\alpha_2 = 0.2$	- 0.070173	3.166561	0.631052
$\alpha_2 = 0.3$	- 0.075758	2.863455	0.737177
$\alpha_2 = 0.4$	- 0.192895	2.765555	2.026053
logistic map $\theta = 3.58$ $\hat{\alpha}_1$ $\alpha_2 =$ - 0.4	1.201583	4.395433	- 6.481278
$\alpha_2 =$ - 0.3	1.107133	4.288325	- 6.152710
$\alpha_2 =$ - 0.2	0.946324	4.107422	- 5.547814
$\alpha_2 =$ - 0.1	0.930832	4.028278	- 5.532063
$\alpha_2 = 0.1$	1.119891	4.220266	- 6.260561
$\alpha_2 = 0.2$	1.267612	4.335562	- 6.804667
$\alpha_2 = 0.3$	1.402347	4.525006	- 7.200866
$\alpha_2 = 0.4$	1.432739	4.532688	- 7.313896
$\hat{\alpha}_2$ $\alpha_2 =$ - 0.4	0.045339	3.182190	- 0.390921
$\alpha_2 =$ - 0.3	0.116266	3.065765	- 1.004290
$\alpha_2 =$ - 0.2	0.246374	3.208201	- 1.981487
$\alpha_2 =$ - 0.1	0.400075	3.296770	- 3.025399
$\alpha_2 = 0.1$	0.976540	3.969486	- 5.796585
$\alpha_2 = 0.2$	1.227970	4.273355	- 6.687259
$\alpha_2 = 0.3$	1.394606	4.507407	- 7.184177
$\alpha_2 = 0.4$	1.430539	4.523115	- 7.312901

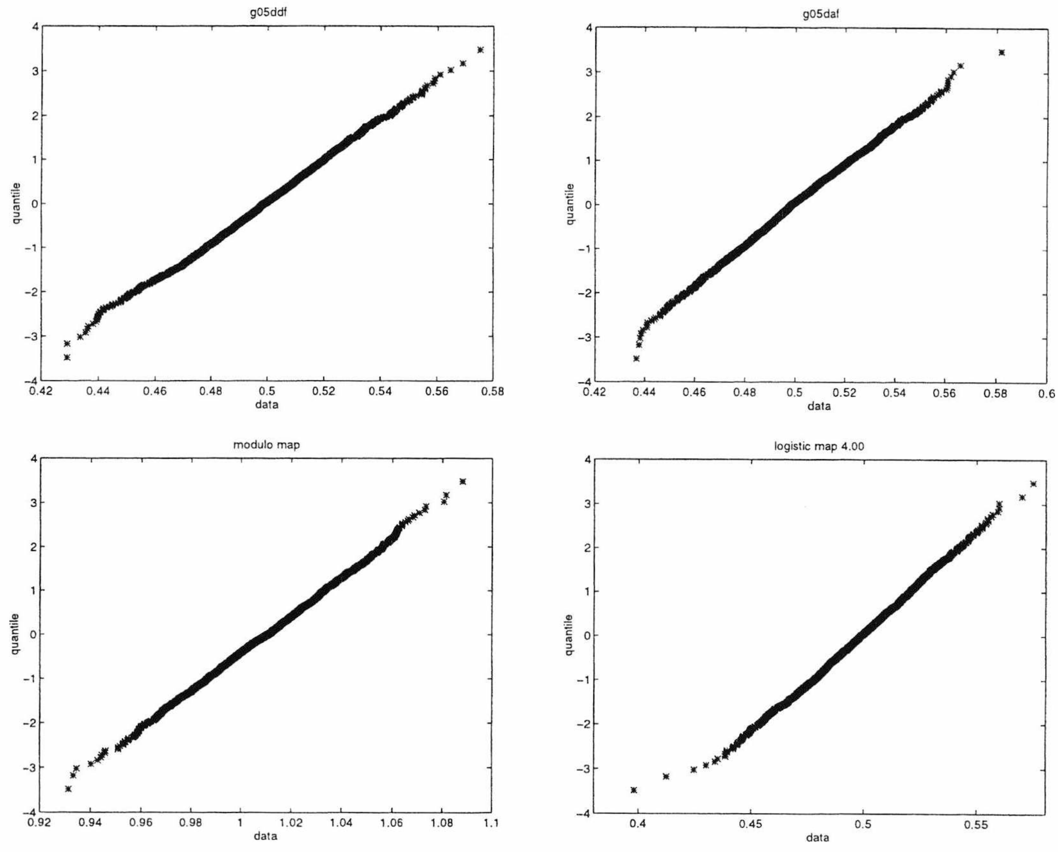
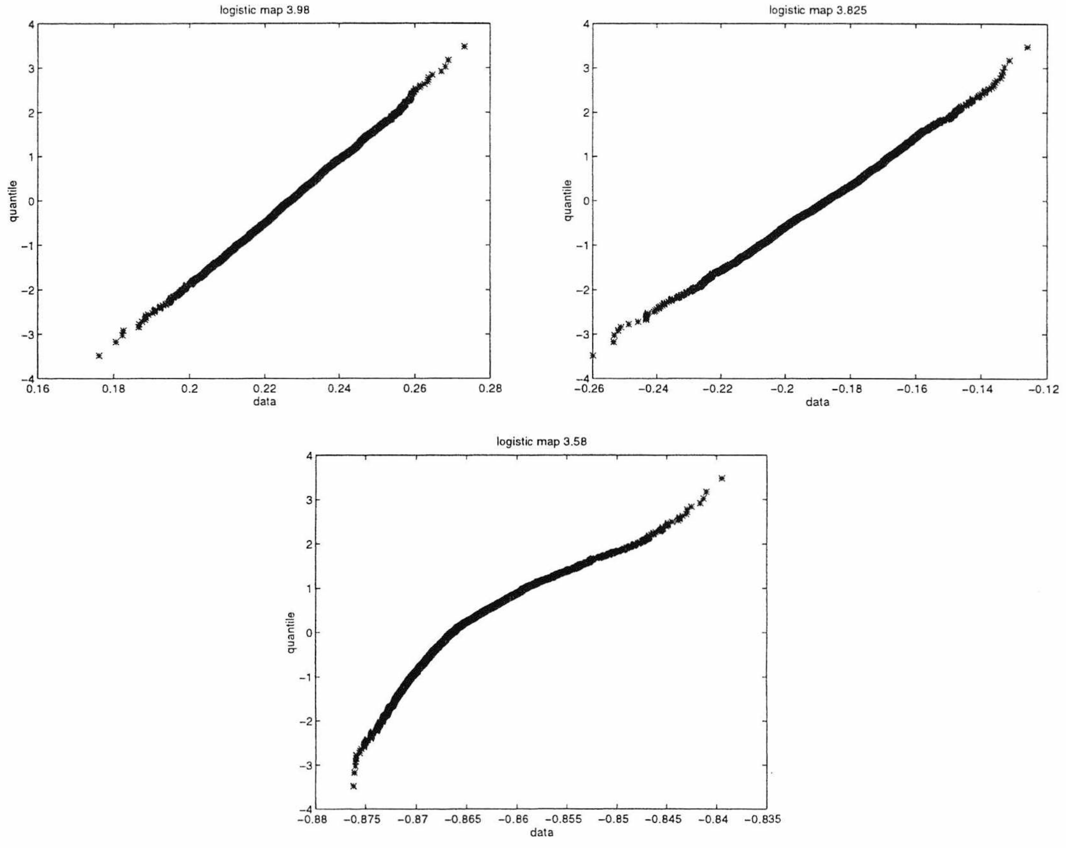


Figure 4.33: Normal probability plots for the case $\alpha_1 = 0.5$, $\alpha_2 = 0.1$. The normal probability plots were obtained by using 2000 replications of $\hat{\alpha}_{1,n}$; n was taken to be equal to 2000.

Figure 4.33: continued



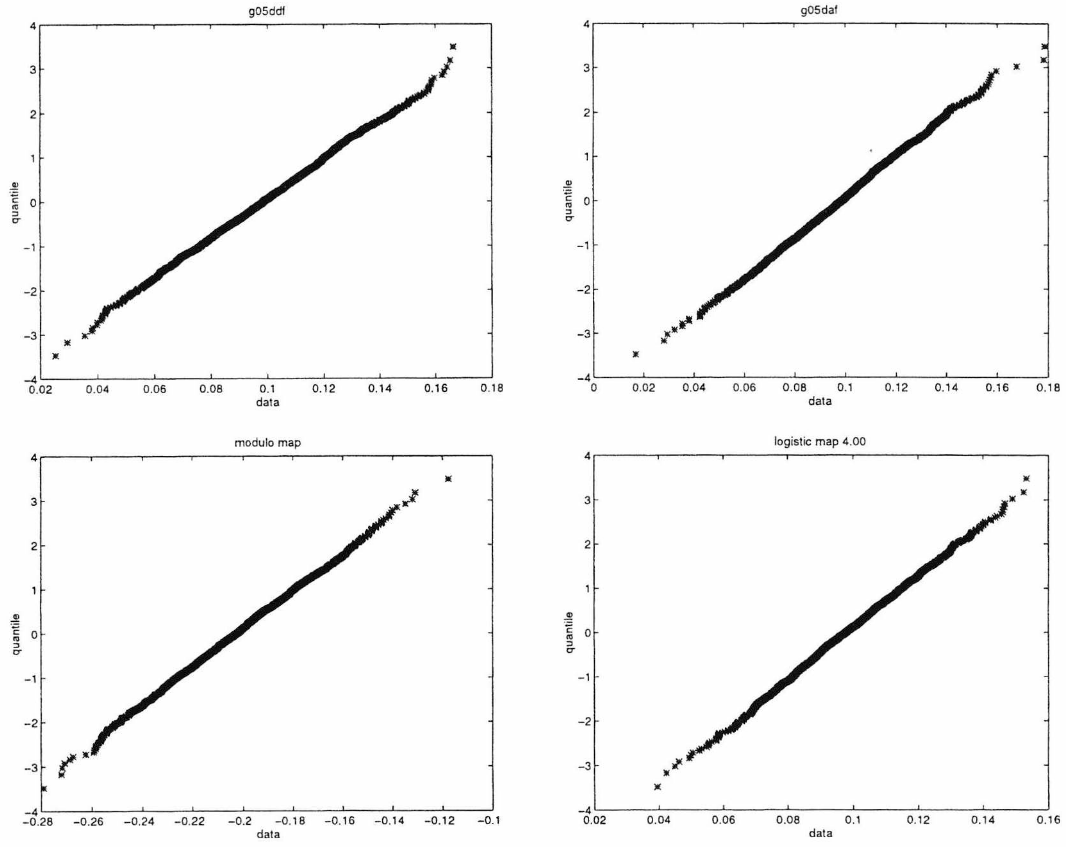
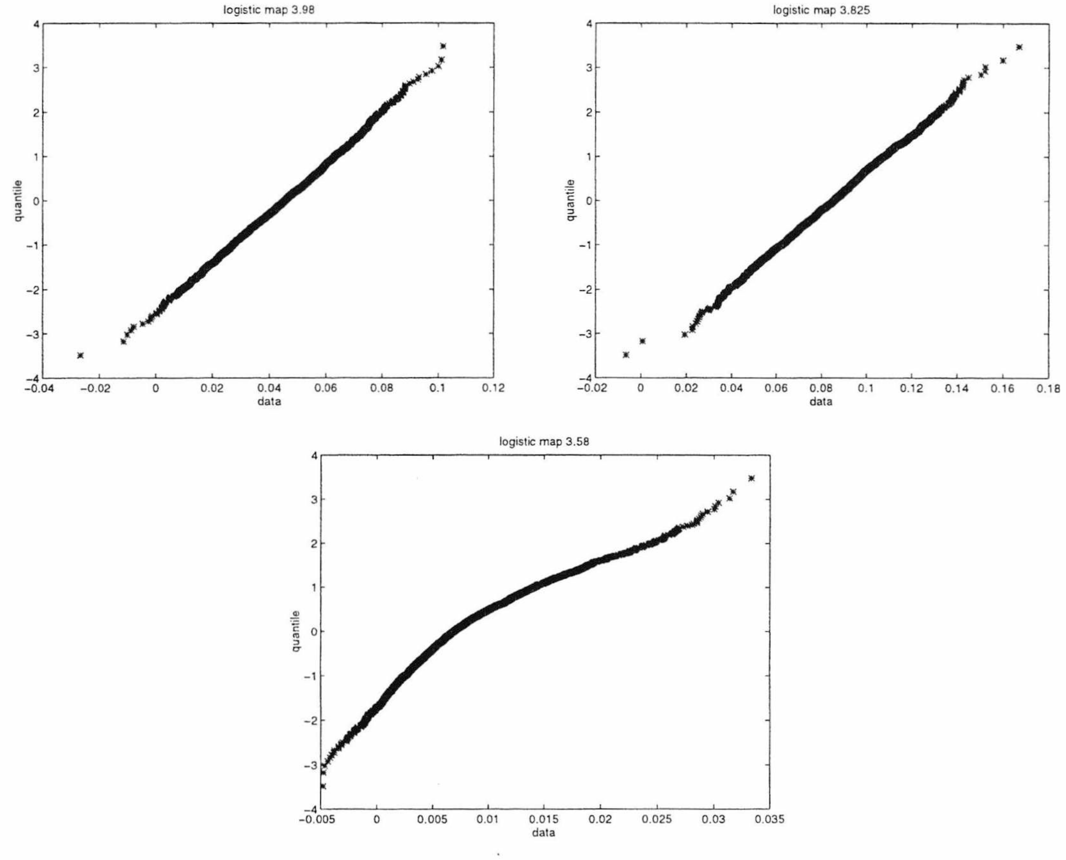


Figure 4.34: Normal probability plots for the case $\alpha_1 = 0.5$, $\alpha_2 = 0.1$.
The normal probability plots were obtained by using 2000 replications of $\hat{\alpha}_{2,n}$; n was taken to be equal to 2000.

Figure 4.34: continued



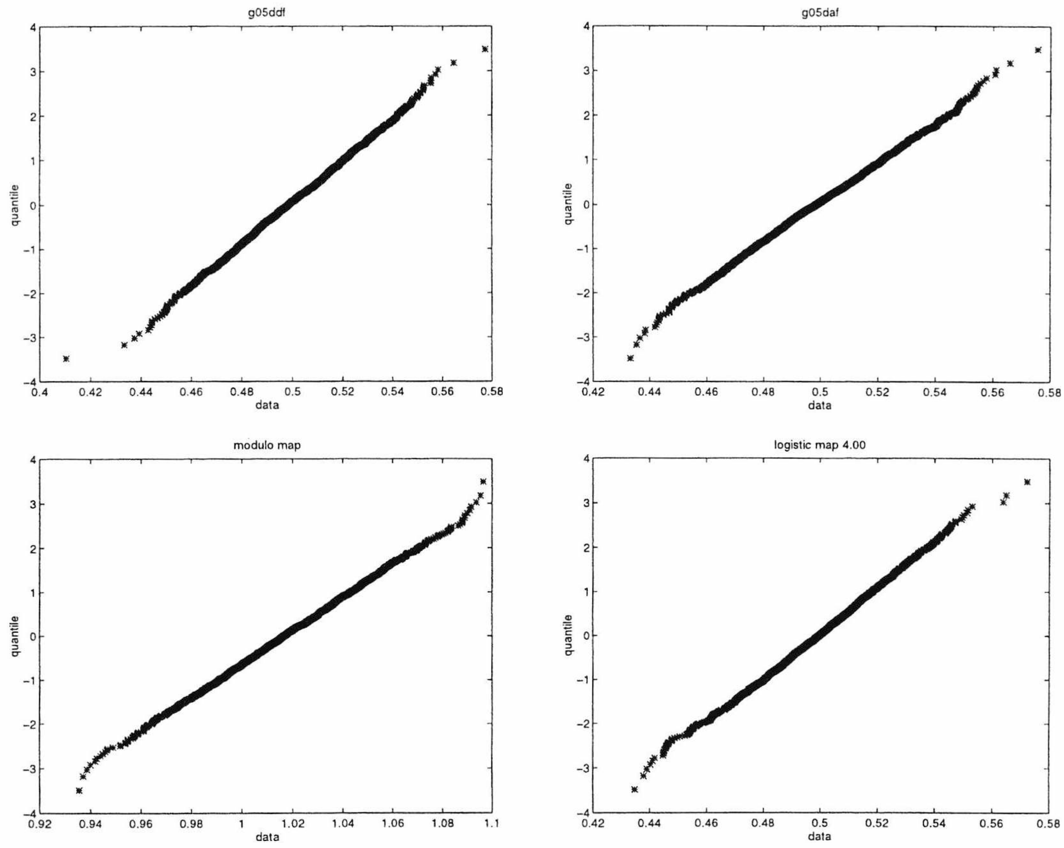
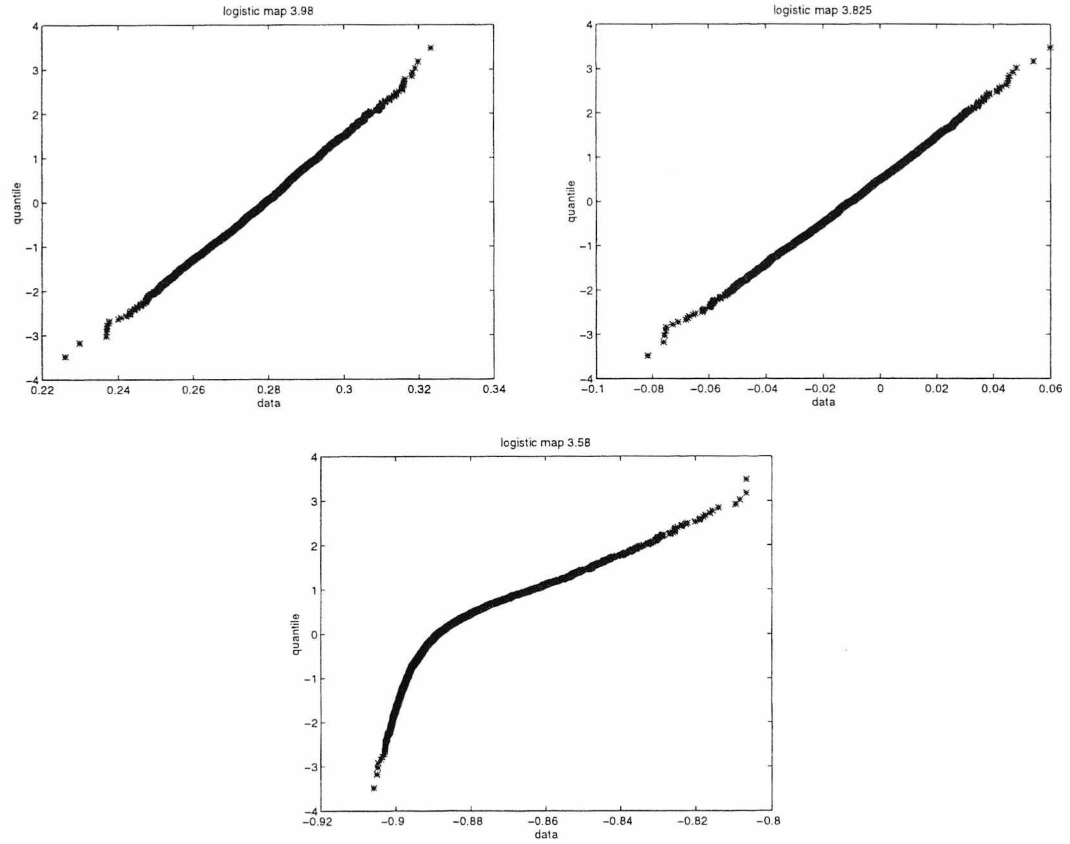


Figure 4.35: Normal probability plots for the case $\alpha_1 = 0.5$, $\alpha_2 = 0.3$.
The normal probability plots were obtained by using 2000 replications of $\hat{\alpha}_{1,n}$; n was taken to be equal to 2000.

Figure 4.35: continued



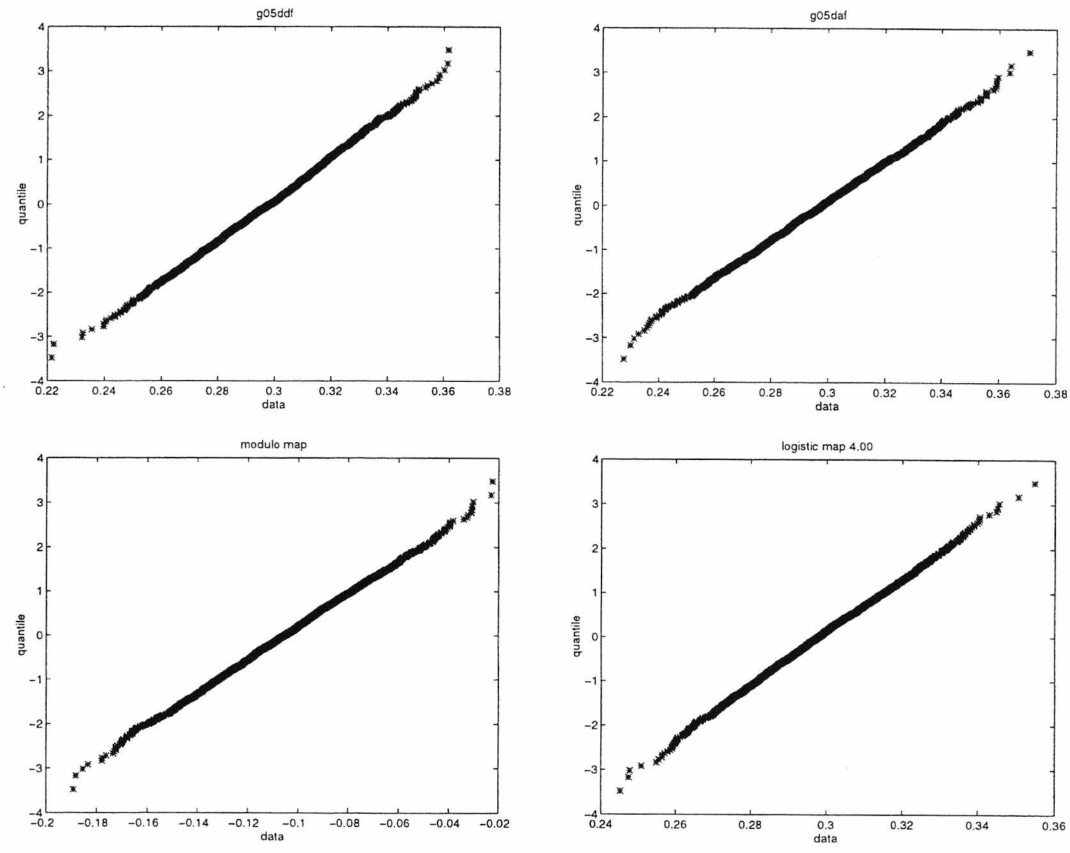
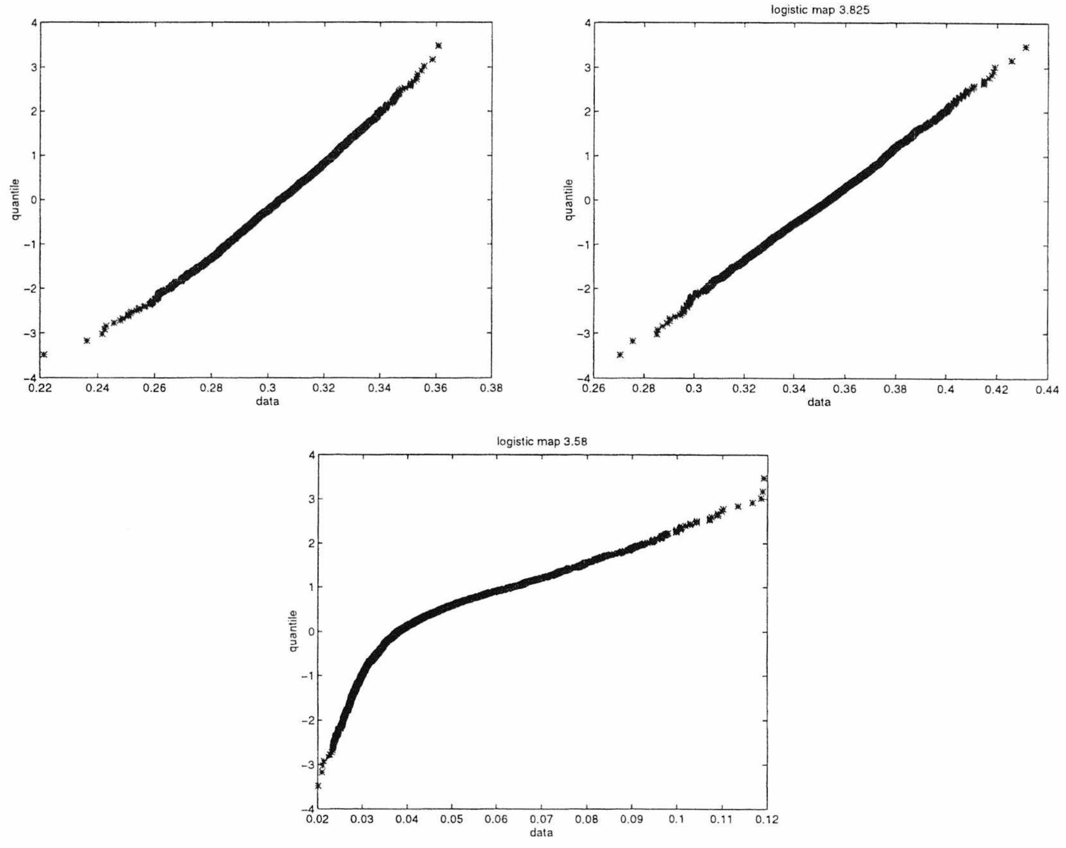


Figure 4.36: Normal probability plots for the case $\alpha_1 = 0.5$, $\alpha_2 = 0.3$. The normal probability plots were obtained by using 2000 replications of $\hat{\alpha}_{2,n}$; n was taken to be equal to 2000.

Figure 4.36: continued



4.6 Some comments on noisy chaos driven AR models

In the previous sections of this chapter, we have obtained interesting asymptotic results for the Yule-Walker estimators $\hat{\alpha}$ of chaos driven AR models $X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + E_t$, where $\{E_t\}$ is a chaotic sequence. In particular, we have evaluated their asymptotic bias and we obtained asymptotic normality of these estimators for some chaotic sequences.

Now, one of the main interests for studying chaos driven AR models is that pseudo-random number generators typically are chaotic maps. An important point to note here is that the deterministic chaotic sequences are a good approximation to the sequences generated by the computers but that they are not exactly the orbits obtained from the computers since any computer-produced orbit is subject to noise (as an example, let us mention the unavoidable rounding errors). We recall that chapter 2 of this thesis focused on the support of attractors obtained from noisy dynamical systems. Now, we shall say a few words on AR models driven by noisy chaotic maps.

Here, we shall concentrate on AR(1) models driven by chaotic sequences submitted to dynamic noise (see section 2.2) in the following way:

$$X_t^* = \alpha X_{t-1}^* + E_t^*,$$

where $E_t^* = f(E_{t-1}^*) + \varepsilon_t$ and ε_t are $iid(0, \sigma^2)$ random variables which are absolutely continuous and bounded. Note that this is a good approximation to the round-off errors of the computers. Now, stochastic randomness has returned to the model albeit at a deeper level.

This has some immediate consequences because the stochastic dynamic noise destabilizes some chaotic maps. For example, the logistic map $f : [0,1] \rightarrow [0,1] : x \rightarrow 4x(1-x)$ leads to the chaotic attractor $[0,1]$ but the same map on R has no (bounded) attractor any more (see section 2.3 for more details).

On the other hand, we can obtain the asymptotic normality of the Yule-Walker estimators for many noisy chaotic maps (provided the noise is sufficiently small) as we shall now show.

In fact, there is a theorem by Collomb stating: " Let $Z_t = G(Z_{t-1}, ..., Z_{t-q}) + \varepsilon_t$, where $\{\varepsilon_t\}$ is a sequence of *iid* random variables and G is a function going from R^{pq} to R^p; then, if G is bounded and the probability law of ε_1 is absolutely continuous w.r.t. Lebesgue measure, then $\{Z_t\}$ is ϕ-mixing (see Definition 3.9) with geometrically decreasing mixing coefficients." (see, for example, Györfi, Härdle, Sarda and Vieu (1989) for more details).

Now, provided a (bounded) noisy attractor exists (see sections 2.2 and 2.3 for more details on noisy attractors), we have

$$\begin{pmatrix} X_t^* \\ E_t^* \end{pmatrix} = \begin{pmatrix} \alpha X_{t-1}^* + f(E_{t-1}^* \\ f(E_{t-1}^*) \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \varepsilon_t \end{pmatrix}$$

and $(X_t^*, E_t^*)'$ is ϕ-mixing with geometrically decreasing coefficients (in particular, E_t^* is an ergodic distribution for the noisy chaotic map).

Then, Denker and Keller (1983) provide us with central limit theorems for U-statistics (see Definition 4.1) of ϕ-mixing sequences: " Let $h : R^m \to R$ be a non-degenerate kernel. Then the U-statistics corresponding to h is asymptotically normal provided $\{X_t\}$ is ϕ-mixing in both directions of time with mixing coefficients $\phi(n)$ satisfying $\sum \phi(n) < \infty$, $\sigma^2 \neq 0$ and $\sup_{1 \leq t_1 < t_2 < ... < t_m} \mathrm{E}(h(X_{t_1}, ..., h(X_{t_m}))^2 < \infty$.".

So, without entering into the details, we get asymptotic normality for $\hat{\alpha}$ provided there exists a (bounded) noisy attractor and $\mathrm{E}[E_t^{*^4}]$ is finite. Of course, this tells us nothing about the convergence rate to normality, which can be very slow for small dynamic noise. Now, for sufficiently small noise level, all the noisy logistic maps, except the one with $\theta = 4.00$, have bounded noisy attractors (see section 2.3 and in particular Figure 2.5). So, this and the fact that $\mathrm{E}[E_t^{*^4}]$ is finite for these noisy maps ensure asymptotic normality for $\hat{\alpha}$. An illustration of this is given in Table 4.3 and Figure 4.37 for the logistic map with $\theta = 3.58$ and a

	Sample skewness	Sample kurtosis	Lin-Mudholkar test statistic
$\alpha =$ - 0.9	0.191715	3.259277	- 1.550774
$\alpha = 0.1$	0.027516	3.319813	- 0.231448
$\alpha = 0.5$	- 0.041465	3.018742	0.382466
$\alpha = 0.9$	- 0.105891	3.060376	0.989992

Table 4.3: Sample skewness, sample kurtosis and Lin-Mudholkar test statistic of the estimators of α were obtained by using 2000 replications of $\hat{\alpha}_n$; n was taken to be equal to 2000.
The model is $X_t^* = \alpha X_{t-1}^* + E_t^*$, where $E_t^* = 3.58E_{t-1}^*(1 - E_{t-1}^*) + \varepsilon_t$ and $\varepsilon_t \sim U(-0.05, 0.05)$.

dynamic noise uniformly distributed $U(-\varepsilon, \varepsilon)$ with $\varepsilon = 0.05$. When we considered, in the earlier sections of this chapter, AR models driven by this logistic map ($\theta = 3.58$), the simulations told us that $\hat{\alpha}$ was far from being asymptotically normal but, as can be seen from Table 4.3 and Figure 4.37, for AR models driven by the logistic map $\theta = 3.58$ subject to noise, the distributions of $\hat{\alpha}_n$ with n chosen equal to 2000 are close to normality.

Now, this is just the first level of the reasoning: as argued in chapters 3 and 4, computers are physical systems and so they simulate *iid* noise by pseudo-random number generators. That is, our model becomes $X_t^{**} = \alpha X_{t-1}^{**} + E_t^{**}$, where $E_t^{**} = f(E_{t-1}^{**}) + G_t$ and $\{G_t\}$ is a chaotic sequence.

We can go further and say that noise is unavoidable in the real world, so stochasticity comes back again but at an even deeper level; now, *iid* noise is simulated by chaotic maps and so on ... . Deterministic and stochastic models are inextricably intertwined.

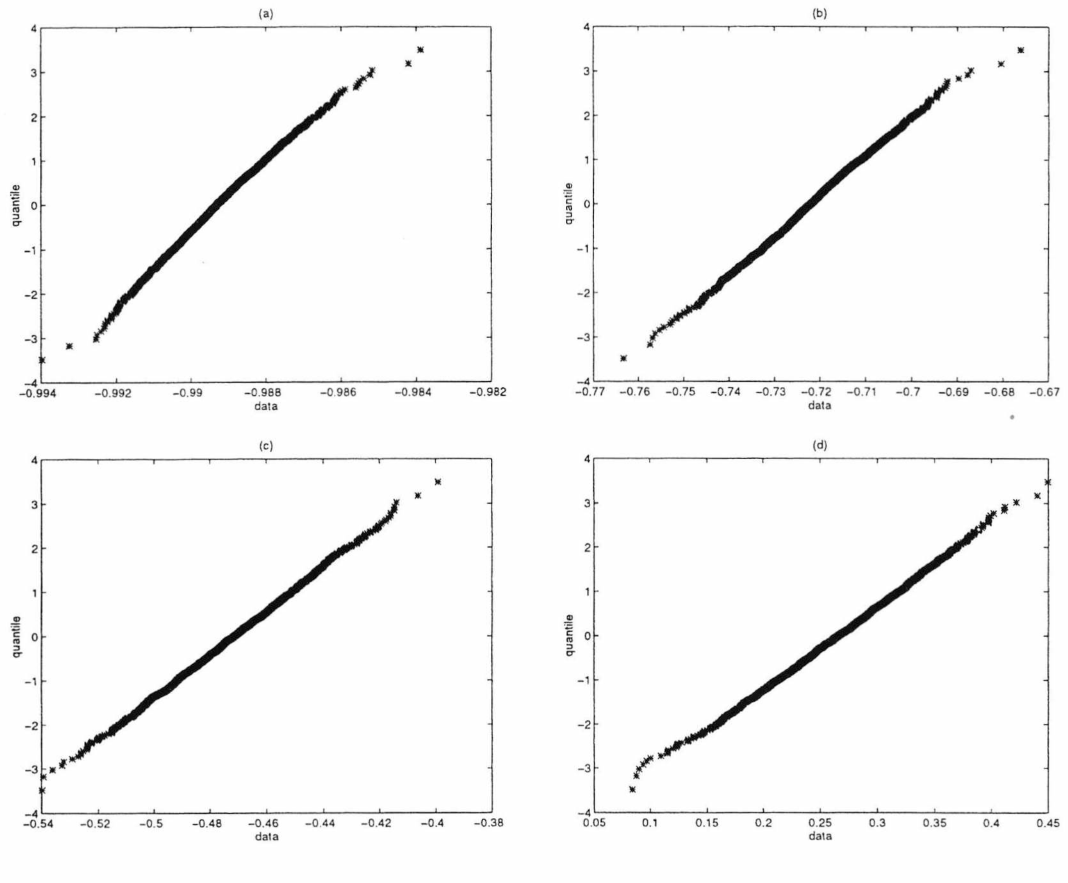


Figure 4.37: Normal probability plots for the cases: (a) $\alpha = -0.9$, (b) $\alpha = 0.1$, (c) $\alpha = 0.5$, (d) $\alpha = 0.9$. The normal probability plots were obtained by using 2000 replications of $\hat{\alpha}_n$; n was taken to be equal to 2000.
The model is $X_t^* = \alpha X_{t-1}^* + E_t^*$, where $E_t^* = 3.58E_{t-1}^*(1 - E_{t-1}^*) + \varepsilon_t$ and $\varepsilon_t \sim U(-0.05, 0.05)$.

Chapter 5

Statistical inference on chaos driven linear stochastic regression models

The structure of chapter 5 is as follows. Section 5.1 introduces linear stochastic regression models and recalls some earlier results on the subject. Finally, section 5.1 states with comments the four asssumptions which we shall make throughout the next three sections. Section 5.2 shows that the linear regression estimator is consistent under these four assumptions; we then say a few words on stochastic regression with state-dependent noise and conclude the section by giving the asymptotic variance of the estimator. Section 5.3 presents three theorems which allow us to get asymptotic normality of $\hat{\beta}$. Section 5.4 illustrates the results of sections 5.2 and 5.3 by means of examples and numerical simulations. Finally, section 5.5 suggests an explanation for the simulation results in chapters 4 and 5 by indicating a possible way of getting a central limit theorem for general chaotic sequences.

5.1 The framework

The classical linear regression model is of the form:

$$y_i = \sum_{j=1}^{p} \alpha_j x_{ij} + \varepsilon_i, \quad i = 1, 2, ..., n,$$

where the x_{ij}s are non-random constants and $\{\varepsilon_t\}$ is either a sequence of uncorrelated zero mean random variables with variance σ^2 (weak distributional assumptions) or a sequence of *iid* normally distributed $\mathcal{N}(0, \sigma^2)$ (strong distributional assumptions).

In matrix notation, the model becomes $Y = X\alpha + \varepsilon$, where $Y = (y_1, ..., y_n)'$, $X = (x_{ij})_{i=1,...,n;\ j=1,...,p}$, $\alpha = (\alpha_1, ..., \alpha_p)'$ and $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$. It is usually assumed that X is of full rank (that is $rank(X) = p$). Then, the matrix $X'X$ is invertible and the least squares estimator $\hat{\alpha}$ of α, also simply called the linear regression estimator, is given by $\hat{\alpha} = (X'X)^{-1}X'Y$. The properties of $\hat{\alpha}$ are well-known: $\mathrm{E}(\hat{\alpha}) = \alpha$, $var(\hat{\alpha}) = \sigma^2(X'X)^{-1}$, $\hat{\alpha}$ is the best linear unbiased estimator of α and, in the case of strong distributional assumptions, $\hat{\alpha}$ is normally distributed.

Consider now the (multiple) linear stochastic regression model:

$$Y_i = \sum_{j=1}^{p} \beta_j X_{ij} + \varepsilon_i, \quad i = 1, 2, ..., n$$

where the ε_is are unobservable random errors and $X_{i1}, ..., X_{ip}$ are random variables, $i = 1, ..., n$. As previously , we can write the model in the matrix form $Y = X\beta + \varepsilon$; it is assumed that $rank(X) = p$. Note that AR(p) models can be written in this form (except for the first p elements of the sequence $\{X_t\}$).

Now, Lai and Wei (1982) show that, if $\{\varepsilon_t\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_t\}$ (that is, ε_t is $\mathcal{F}_t$-measurable and $\mathrm{E}[\varepsilon_t|\mathcal{F}_{t-1}] = 0$ for every t), then the linear regression estimator $\hat{\beta} = (X'X)^{-1}X'Y$ is consistent and asymptotically normal provided some additional conditions are satisfied. Specifically, Theorem 1 of Lai and Wei (1982) states the following: " Suppose that in the regression model, $\{\varepsilon_t\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_t\}$ such that $\sup_t \mathrm{E}(|\varepsilon_t|^\alpha|\mathcal{F}_{t-1}) < \infty$ a.s. for some $\alpha > 2$. Moreover, assume that the design levels $X_{i1}, ..., X_{ip}$ at stage i are $\mathcal{F}_{i-1}$-measurable random variables such that $\lambda_{min}(n) \to \infty$ a.s. and $log\lambda_{max}(n) = o(\lambda_{min}(n))$ a.s., where $\lambda_{min}(n)$ and $\lambda_{max}(n)$ respectively denote the minimum eigenvalue of $(X'X)^{-1}$ and the maximum eigenvalue of $(X'X)^{-1}$. Then $\hat{\beta} \to^{a.s.} \beta$.". Theorem 3 of Lai and Wei (1982) states : " Suppose that in the regression model, $\{\varepsilon_t\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields $\{\mathcal{F}_t\}$ such that $\sup_t \mathrm{E}(|\varepsilon_t|^\alpha|\mathcal{F}_{t-1}) < \infty$ a.s. for some $\alpha > 2$ and $\lim_{n\to\infty} \mathrm{E}(\varepsilon_n^2|\mathcal{F}_{n-1}) = \sigma^2 < \infty$ a.s. . Moreover, assume for each i that the design vector $X_i = (X_{i1}, ..., X_{ip})'$ at stage i is $\mathcal{F}_{i-1}$-measurable and that there exists a non-random positive definite symmetric matrix B_n for which $B_n^{-1}(\sum_1^n X_i X_i')^{1/2} \to^p I_p$, identity matrix of rank p and $\max_{1\leq i \leq n} ||B_n^{-1}x_i|| \to^p 0$. Then $\hat{\beta}$ is asymptotically normal.".

In this chapter, we shall concentrate on chaos driven linear stochastic regression models, that is

$$Y_i = \sum_{j=1}^{p} \beta_j X_{ij} + E_i, \quad i = 1, 2, ..., n.$$

In matrix form, we have $Y = X\beta + E$, where $Y = (Y_1, ..., Y_n)'$, $X = (X_{ij})_{i=1,...,n;\ j=1,...,p}$, $\beta = (\beta_1, ..., \beta_p)'$, $E = (E_1, ..., E_n)'$. Now, $\{E_t\}$ is a chaotic sequence (in the sense of Definition 3.2) with finite variance. In particular, we have $E_t = f(E_{t-1})$, where f is a deterministic (chaotic) map and thus the theorems in Lai and Wei (1982) do not apply in our case since $\{E_t\}$ is not a martingale difference sequence ($\mathrm{E}[E_t|E_{t-1}, E_{t-2}, ...] = f(E_{t-1}) \neq 0$).

At this stage, we shall make four assumptions on our chaos driven linear stochastic regression model. First, we deal with three of them. The first two are straightforward; the third one is reasonable to make. After stating this third assumption, its importance for the rest of the chapter will be sketched.

Assumption 1 *The chaotic sequence $\{E_t\}$ has zero mean and finite variance.*

Assumption 2 *The p explanatory random variables* $X_1, ..., X_p$ *have zero mean and finite variances* $\sigma^2_{X_1}, ..., \sigma^2_{X_p}$.

Of course, the zero mean assumptions are made without loss of generality.

Assumption 3 *The p explanatory random variables* $X_1, ..., X_p$ *are jointly ergodic.*

The notion of joint ergodicity is a direct extension to ergodicity (see Definition 3.3). It means that the p-uple $(X_1, ..., X_p)$ is ergodic. Obviously, this implies that each X_i, $i = 1, ..., p$ is ergodic. Note that the fact that X_i is ergodic for $i = 1, 2, ..., p$ does not imply that $(X_1, ..., X_p)$ is ergodic. See, for example, Pinsker (1964) for a nice discussion on this.

Now, $\hat{\beta} - \beta = (X'X)^{-1}X'Y - \beta = (X'X)^{-1}X'(X\beta + \varepsilon) - \beta = (X'X)^{-1}X'E$. Equivalently, $\hat{\beta} - \beta = n(X'X)^{-1}(\frac{1}{n}X'E)$. Assumption 3 combined with the ergodic theorem (see section 3.3) and Assumption 2 ensure that $n(X'X)^{-1} \rightarrow^{a.s.} C$, (finite) constant matrix. In the same way, $n^{1/2}(\hat{\beta} - \beta)$ can be rewritten as $n(X'X)^{-1}(n^{-1/2}X'E)$ and once again because of our assumptions, in particular because of Assumption 3, $n(X'X)^{-1} \rightarrow^{a.s.} C$. So, we can see the role of Assumption 3 in the determination of consistency and in the construction of central limit theorems.

Note that the chaos driven AR(p) models analysed in chapter 4 are a particular case of chaos driven linear stochastic regression models (simply let $Y_i = X_i, X_{i1} = X_{i-1}, ..., X_{ip} = X_{i-p}$) and satisfy Assumptions 1, 2 and 3. Now, the chaos driven AR models will not satisfy the next assumption, which we shall state now.

Assumption 4 $\forall j = 1, ..., p$, $\forall i = 1, ..., n$, $\forall t = 1, ..., n$, X_{ij} *is independent of* E_t.

If Assumption 4 is satisfied, we shall say that 'X and E are independent'. Clearly, in the case of a chaos driven AR model, Assumption 4 is not satisfied since E_t and $X_t, X_{t+1}, \ldots$ are dependent. However, this assumption is often met in regression analysis. In fact, Assumption 4 (combined to the three other assumptions) will allow us to get consistency of $\hat{\beta}$, as we shall now show.

5.2 Consistency and asymptotic variance of the linear regression estimator

First, we can note that $\hat{\beta}$ is an unbiased estimator of β. This is obvious since $\mathrm{E}(\hat{\beta}) = \beta + \mathrm{E}((X'X)^{-1}X'E) = \beta + \mathrm{E}_X\mathrm{E}_E((X'X)^{-1}X'E|X) = \beta + \mathrm{E}_X(0)$ by Assumptions 4 and 1 and so $\mathrm{E}(\hat{\beta}) = \beta$.

Now, we can go further: the next theorem will show that the linear regression estimator $\hat{\beta}$ is a consistent estimator of β (that is, $\hat{\beta} \to^p \beta$) for the models satisfying Assumptions 1 to 4. Roughly speaking, this means that $\hat{\beta}$ is asymptotically unbiased (we know that this is right) and that $var(\hat{\beta}) \to 0$ as $n \to \infty$. It is interesting to notice that we shall get consistency for all the chaotic sequences (with finite variance), in particular even for the chaotic sequences with long-range dependence. Note that the fact that $\{E_t\}$ is not only a deterministic ergodic sequence but also a chaotic sequence plays a role in the proof.

Theorem 5.1 *Under Assumptions 1, 2, 3 and 4, $\hat{\beta}$ is a consistent estimator of β, that is $\hat{\beta} = (X'X)^{-1}X'Y$ converges in probability to β (in short, $\hat{\beta} \to^p \beta$).*

Proof: We have

$$\hat{\beta} = (X'X)^{-1}X'Y = \beta + n(X'X)^{-1}\frac{1}{n}X'E$$

On the one hand, $\frac{1}{n}X'X \to^p C$, finite constant matrix, by joint ergodicity (Assumption 3), by the existence of finite variances for the explanatory random variables (Assumption 2) and by application of the ergodic theorem (see section 3.3).

Now, let k be any of the numbers $1, \ldots, p$. We are going to show that $\frac{1}{n}(X'E)_k$

$\rightarrow^p 0$, where $(X'E)_k$ denotes the kth component of the column vector $X'E$.

As a first step, we note that the process $\{\xi_t\} = \{(X_{tk}, E_t)\}$ is ergodic since $\{E_t\}$ is mixing in the ergodic-theoretic sense (see Definition 3.4), $\{X_{tk}\}$ is ergodic (Assumption 3) and X and E are independent (Assumption 4) (see, for example, Pinsker (1964) for a proof). Note that $\{E_t\}$ being ergodic would not have been sufficient to ensure $\{\xi_t\}$ being ergodic.

Then, let $\psi_t = X_{tk}E_t\ \forall t$; the process $\{\psi_t\}$ is ergodic as a measurable function of the ergodic process $\{\xi_t\}$. Now, $\frac{1}{n}(X'E)_k = \frac{1}{n}\sum_t X_{tk}E_t \rightarrow^p \mathrm{E}[X_{tk}E_t] = 0$, by ergodicity of $\{\psi_t\}$, Assumptions 1, 2 and 4 and an application of the ergodic theorem.

Now, $\frac{1}{n}X'E \rightarrow^p (0, ..., 0)'$ by using Propositions 6.3.7 and 6.3.5 in Brockwell and Davis (1989). We then get the conclusion by using Proposition 6.3.8 in Brockwell and Davis (1989). □

Before considering the asymptotic variance of $\hat{\beta}$, we shall say a few words on more general models than the ones analysed in this chapter, namely the state-dependent noise chaos driven linear stochastic regression models of the matrix form $Y = X\beta + g(X_1, ..., X_p)E$, where Y, X, β and E are as previously and $g : R^p \rightarrow R$ is a measurable function. Then, under the same assumptions as previously and provided weak additional conditions are satisfied, we get consistency for $\hat{\beta} = (X'X)^{-1}X'Y$. The next proposition states the result; its proof is very similar to the proof of Theorem 5.1 and so it will be sketched.

Proposition 5.1 *Under Assumptions 1 to 4 and if* $g(X_1, ..., X_p)X_k$ *has zero mean and finite variance for* $k = 1, 2, ..., p$, *then* $\hat{\beta} \rightarrow^p \beta$.

Proof: (sketch)

We have

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + g(X_1, ..., X_p)E)\end{aligned}$$

$$= n(X'X)^{-1}\frac{1}{n}g(X_1,...,X_p)X'E.$$

We shall show here that $\frac{1}{n}\sum_t g(X_{t1},...,X_{tp})X_{tk}E_t \to^p 0$, $k = 1, 2, ..., p$; all the other points of the proof are identical to the previous proof.

Now, $\{\xi_t\} = \{(g(X_{t1},...,X_{tp})X_{tk}, E_t)\}$ is ergodic since $\{E_t\}$ is mixing in the ergodic-theoretic sense, $\{g(X_{t1},...,X_{tp})X_{tk}\}$ is ergodic as a measurable function of ergodic $(X_1,...,X_p)$ (Assumption 3), and $g(X)X$ and E are independent.

Then, $\{\psi_t = g(X_{t1},...,X_{tp})X_{tk}E_t\}$ is ergodic as a measurable function of the ergodic process $\{\xi_t\}$.

Now,

$$\frac{1}{n}\sum_t g(X_{t1},...,X_{tp})X_{tk}E_t \to^p \mathrm{E}[g(X_{t1},...,X_{tp})X_{tk}E_t] = 0$$

by ergodicity of $\{\psi_t\}$, Assumptions 1 and 4, the fact that $g(X_{t1},...,X_{tp})X_{tk}$ has zero mean and finite variance and by application of the ergodic theorem. □

Note that this kind of model has asymptotic properties analogous to those for the models analysed in this chapter not only for the consistency but also for the variance and the normality. We are not going to discuss them further in the thesis. Our goal was just to illustrate by means of an example the many possible simple extensions based on our results.

Now, as $\hat{\beta}$ is an unbiased estimator of β, $var(\hat{\beta}) = \mathrm{E}[(\hat{\beta} - \beta)^2]$. Since $\hat{\beta} - \beta = (X'X)^{-1}X'E$, we have $(\hat{\beta} - \beta)^2 = (X'X)^{-1}X'EE'X(X'X)^{-1}$ and so

$$var(\hat{\beta}) = \mathrm{E}[(X'X)^{-1}X'EE'X(X'X)^{-1}].$$

In particular, $var(\hat{\beta})$ does not depend on β.

Then,

$$\begin{aligned} var(\hat{\beta}) &= \mathrm{E}_X \mathrm{E}_E[(X'X)^{-1}X'EE'X(X'X)^{-1}|X] \\ &= \mathrm{E}_X[(X'X)^{-1}X'\Gamma_E X(X'X)^{-1}] \end{aligned}$$

by Assumptions 4 and 1, Γ_E denoting the variance-covariance matrix of $\{E_t\}$, that is $\Gamma_E = \{\gamma_E(i-j)\}_{i=1,\ldots,n;j=1,\ldots n}$.

In particular, in the case of one explanatory random variable (p=1), we get

$$var(\hat{\beta}) = \mathrm{E}_X[\frac{1}{\sum X_i^2}(X_1 \ldots X_n)\Gamma_E(X_1 \ldots X_n)'\frac{1}{\sum X_i^2}]$$

which leads immediately to the following theorem.

Theorem 5.2 *In the case of one explanatory random variable, under Assumptions 1 to 4, and if* $\sum_{i=1}^{\infty} \gamma_E(i)\gamma_X(i) < \infty$*, then*

$$\lim_{n\to\infty} n\ var(\hat{\beta}) = \frac{1}{(\gamma_X(0))^2}(\gamma_X(0)\gamma_E(0) + 2\sum_{i=1}^{\infty}\gamma_E(i)\gamma_X(i)).$$

Obvious extensions are possible in the case of p explanatory random variables. So, provided a weak condition is satisfied, we get the asymptotic variance of $\hat{\beta}$ in a simple form. In particular, $var(\hat{\beta})$ converges to 0 in a rate $O(1/n)$, which suggests that under suitable conditions we could get central limit theorems for $\hat{\beta}$. The asymptotic normality will be the subject of the next section. Examples and numerical simulations of asymptotic variances will be given in section 5.4.

5.3 Asymptotic normality of the linear regression estimator

In the last section, it has been proved that, under Assumptions 1 to 4, $\hat{\beta}$ was unbiased, consistent and, provided a weak additional condition is satisfied, with an asymptotic variance easy to calculate. Now, we shall focus on the (possible) asymptotic normality of $\hat{\beta}$. We recall from section 5.1 that $n^{1/2}(\hat{\beta}-\beta) = n(X'X)^{-1}(n^{-1/2}X'E)$ and, because of the joint ergodicity of $X_1, ..., X_p$ (Assumption 3) and the fact that they have finite variances (Assumption 2), we have $n(X'X)^{-1} \to^{a.s.} C$, (finite) constant matrix. So, a central limit theorem exists for $\hat{\beta}$ if and only if there is a central limit theorem for $n^{-1/2}X'E$.

No general answer can be given concerning the asymptotic normality of $X'E$. We have to impose conditions on $\{X_t\}$ and/or $\{E_t\}$. Here, we shall concentrate on the case of one explanatory random variable; the main reason for this is the large number of central limit theorems available for sequences of real random variables. Now, in case of p explanatory random variables, the univariate asymptotical normality of $\hat{\beta}_k$ can be deduced exactly in the same way as in the case p=1; the joint asymptotic normality of $\hat{\beta}$ is more complicated to get: one way to proceed is to ensure that every linear combination of the $\hat{\beta}_k$s is (univariate) asymptotically normal.

We shall state here three theorems which allow us to get asymptotic normality of $n^{-1/2}\sum_{t=1}^{n} X_t E_t$ and thus of $\hat{\beta}$. All of them are easy corollaries of well-known central limit theorems. We could have used many more theorems but our goal in this section is to indicate important ways for getting asymptotic normality of β and not to make a review of all the existing central limit theorems.

The first theorem requires for the sequence $\{X_t\}$ to be a martingale difference (that is, X_t is $\mathcal{F}_t$-measurable and $\mathrm{E}[X_t|\mathcal{F}_{t-1}] = 0$ for every t). Note that no condition is imposed on the chaotic sequence $\{E_t\}$, except that it has to satisfy Assumptions 1 and 4.

Theorem 5.3 *Under Assumptions 1 to 4 and if $\{X_t\}$ is a martingale difference related to the σ-algebra $\mathcal{F}_t$, then*

$$n^{1/2}(\hat{\beta}-\beta) \to^{d} \mathcal{N}(0, \frac{1}{(\gamma_X(0))^2}(\gamma_X(0)\gamma_E(0) + 2\sum_{i=1}^{\infty}\gamma_E(i)\gamma_X(i))).$$

Proof: If $\{X_t\}$ is a zero mean (Assumption 2) martingale difference related to the σ-algebra $\mathcal{F}_t$, then X_tE_t is a martingale difference related to $\mathcal{F}_t \vee \sigma(E_1)$ since X_tE_t is $\mathcal{F}_t \vee \sigma(E_1)$ - measurable ($\{E_t\}$ is a chaotic sequence, so $E_t = f^{t-1}(E_1)$) and $\mathrm{E}[X_tE_t|\mathcal{F}_{t-1} \vee \sigma(E_1)] = E_t\mathrm{E}[X_t|\mathcal{F}_{t-1} \vee \sigma(E_1)] = 0$

So, $n^{-1/2}\sum_t X_tE_t$ is asymptotically normal with finite variance since E_tX_t are stationary ergodic martingale differences (see, for example, Hall and Heyde (1980), p51). □

In particular, if $X_t \sim iid(0,\sigma^2)$ and $\{X_t\}$ and $\{E_t\}$ are independent, then $n^{1/2}(\hat{\beta} - \beta)$ is asymptotically normal (with mean 0 and variance $\gamma_E(0)/\sigma^2$). The next section will illustrate this.

The next theorem basically requires the sequence $\{X_t\}$ to be strongly mixing with a sufficiently fast mixing rate; again no additional condition is imposed on $\{E_t\}$.

Theorem 5.4 *Let the four assumptions be satisfied. Let $0 < \delta \leq \infty$ be fixed. Suppose that the stationary sequence $\{X_t\}$ is strongly mixing (see Definition 3.7) and $\mathrm{E}|X_1|^{2+\delta} < \infty$, in case $0 < \delta < \infty$, or $|X_1| \leq constant < \infty$ if $\delta = \infty$, while $\sum_{n=1}^{\infty}[\alpha(n)]^{\delta/2+\delta} < \infty$.*

Then, we get

$$n^{1/2}(\hat{\beta} - \beta) \rightarrow^d \mathcal{N}(0, \frac{1}{(\gamma_X(0))^2}(\gamma_X(0)\gamma_E(0) + 2\sum_{i=1}^{\infty}\gamma_E(i)\gamma_X(i))).$$

Proof: We shall show that the sequence $\{X_tE_t\}$ meets the requirements of Theorem 5.2 in Hall and Heyde (1980).

Now, $\{X_tE_t\}$ is stationary, ergodic (by the same argument as in the proof of Theorem 5.1), with $\mathrm{E}X_0 = 0$, $\mathrm{E}X_0^2 < \infty$ (Assumptions 1 and 2). Moreover, we know, by Corollary 5.1 in Hall and Heyde (1980), that $\{X_t\}$ satisfies the conditions of Theorem 5.2.

Take $\mathcal{M}_0$ as a σ-field such that X_0 is $\mathcal{M}_0$-measurable and Theorem 5.2 applies, then X_0E_0 is $\mathcal{M}_0 \vee \sigma(E_0)$ - measurable and such that

$$\begin{aligned} \mathrm{E}(X_kE_k\mathrm{E}(X_NE_N|\mathcal{M}_0 \vee \sigma(E_0)) &= \mathrm{E}(X_kE_kE_N\mathrm{E}(X_N|\mathcal{M}_0)) \text{ since } E_N = f^N(E_0) \\ &= \mathrm{E}(E_kE_N)\mathrm{E}(X_k\mathrm{E}(X_N|\mathcal{M}_0)) \text{ by Assumption 4} \end{aligned}$$

and so

$$\begin{aligned} \sum_{k=1}^{\infty} |\mathrm{E}(X_kE_k\mathrm{E}(X_NE_N|\mathcal{M}_0 \vee \sigma(E_0)))| &\leq \sum_{k=1}^{\infty} |\mathrm{E}(E_kE_N)||X_k\mathrm{E}(X_N|\mathcal{M}_0)| \\ &\leq \gamma_E(0)\sum_{k=1}^{\infty}|X_k\mathrm{E}(X_N|\mathcal{M}_0)| \end{aligned}$$

which is finite and tends to 0 as $N \to \infty$. Theorem 5.2 thus applies. □

Examples and simulations will be given in the next section.

The last theorem requires for $\{E_t\}$ and for $\{X_t\}$ to be well-behaved functionals of absolutely regular processes in the sense of Denker and Keller (1986). In particular, both $\{E_t\}$ and $\{X_t\}$ can be purely deterministic chaotic sequences.

Theorem 5.5 *Let the four assumptions be satisfied.*

Suppose that the following conditions hold.

1. $\{E_n : n \geq 1\}$ *is a Lipschitz functional of* $\{Z_n : n \geq 1\}$ *in the same sense as in Theorem 4.4.*

2. $\{Z_n : n \geq 1\}$ *is an absolutely regular stationary sequence with* $\beta(n)^{\eta/2+\eta} = O(n^{-2-\varepsilon})$ *for some* $\varepsilon, \eta > 0$.

3. $\{X_n : n \geq 1\}$ *is a Lipschitz functional of* $\{Z'_n : n \geq 1\}$ *in the same sense as in Theorem 4.4.*

4. $\{Z'_n : n \geq 1\}$ *is an absolutely regular stationary sequence with* $\beta(n)^{\eta/2+\eta} = O(n^{-2-\varepsilon})$ *for some* $\varepsilon, \eta > 0$.

5. $h(X_t, E_t) = X_tE_t$ *satisfies the Lipschitz condition.*

Specifically, there are $L > 0$, $r \geq 0$ and $\rho > 0$ such that

$$|h(y)-h(z)| = |h(y_1, y_2)-h(z_1, z_2)| = |y_1y_2-z_1z_2| \leq L|y-z|^{\rho}(1+|y|^r+|z|^r)$$

for all y_1, z_1 belonging to the support of X_t and all y_2, z_2 belonging to the support of E_t.

Then

$$n^{1/2}(\hat{\beta}-\beta) \to^d \mathcal{N}(0, \frac{1}{(\gamma_X(0))^2}(\gamma_X(0)\gamma_E(0) + 2\sum_{i=1}^{\infty}\gamma_X(i)\gamma_E(i))).$$

Proof: This is an immediate corollary of Theorem 1 of Denker and Keller (1986). □

Now, the next section will illustrate the results obtained in sections 5.2 and 5.3.

5.4 Simulations

We illustrate in this section, by means of examples and numerical simulations, the theoretical results obtained in the last two sections. We take as $\{E_t\}$ the same chaotic sequences as in chapter 4, namely sequences generated by the pseudo-random number generator G05DDF (normal distribution), the modulo map, the logistic maps with $\theta = 4.00$, $\theta = 3.98$, $\theta = 3.825$, $\theta = 3.58$. Note that $\{E_t\}$ generated by the pseudo-random number generator G05DAF (uniform distribution) brought in all cases under consideration very similar results to G05DDF, so we omit to present here these results for the sake of concision. As previously, we standardize every chaotic sequence $\{E_t\}$ to zero mean and unit variance; for the modulo map, we replaced the multiplier 2 by 1.99999 in order to avoid degeneracy.

Basically, we present here theoretical and simulated results on $\hat{\beta}$ relative to four different examples of chaos driven linear stochastic regression models $Y_t = \beta X_t + E_t$, β fixed real number. All the cases under consideration satisfy the four assumptions introduced in section 5.1

Before having a closer look at them, we make two remarks. First, all the simulations concerning the bias confirmed the obvious theoretical fact (see section 5.2) that $\hat{\beta}$ is an unbiased estimator of β. So, we omit them here. Second, we show just part of the simulated results concerning the asymptotic normality; complementary results corroborate the following commments. They are not included for the sake of brevity.

The first (simple) example is such that $X_t \sim iid\ \mathcal{N}(0,1)$ and the $\{E_t\}$ are as described above. Theorem 5.2 tells us that the asymptotic variances of $\hat{\beta}_n$ will be the same for all the chaotic sequences (since all the chaotic sequences are standardized to unit variance), namely $var(\hat{\beta}_n) = 1/n$. Simulations (Figure 5.1) confirm this. Note that the simulations also confirm that $var(\hat{\beta}_n)$ does not depend on β but this is hardly surprising (see the general form of $var(\hat{\beta})$ in section 5.2).

$\{X_t\}$ meets the requirements of Theorems 5.3 and 5.4, so we get asymptotic normality of $\hat{\beta}$ for all the chaotic sequences $\{E_t\}$, as shown in Table 5.1 and Figure 5.2.

The second and third examples are respectively $Y_t = \beta X_{t-2} + E_t$ and $Y_t = \beta(X_{t-2}^2 - 4/3) + E_t$, where $X_t = 0.5X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid\ \mathcal{N}(0,1)$ and the $\{E_t\}$ are the above-mentioned chaotic sequences. They have been inspired by a similar model in Yao and Tong (1994a) (in their article, in place of $\{E_t\}$ they worked with η_t, independent and identically distributed random variables). For both examples, Theorem 5.2 applies to all chaotic sequences $\{E_t\}$ (even to the long-range dependent sequence $\{E_t\}$ generated by the logistic map $\theta = 3.58$); Figure 5.3 (example 2) and Figure 5.5 (example 3) give the asymptotic variances and simulations, which confirm the theoretical values. Note that in both examples the lowest asymptotic variance is obtained for the logistic map with $\theta = 3.58$. This is easily explained by the facts that the autocovariance function of X_t is positive at all lags (so in particular at lag 1) and decays exponentially quickly, while $\gamma_E(1) = \rho_E(1)$ is strongly negative for $\theta = 3.58$ (see Figure 3.3).

Concerning asymptotic normality, Theorem 5.4 requires for $\{X_t\}$ to be strongly mixing with sufficiently fast mixing rate. Now, X_t is a causal AR(1) model driven by *iid* absolutely continuous noise and therefore, $\{X_t\}$ is absolutely regular with exponentially decreasing rate (see Bradley (1986) and the references therein). This implies in particular that $\{X_t\}$ and $\{X_t^2\}$ are strongly mixing with exponentially decreasing rate. Theorem 5.4 thus applies. Simulations (Tables 5.2 and 5.3, Figures 5.4 and 5.6) confirm the asymptotic normality of $\hat{\beta}$ for both examples and all the chaotic sequences $\{E_t\}$.

The last example concerns exclusively chaotic sequences: $\{E_t\}$ is taken to be generated by the logistic map $\theta = 4.00$ and the $\{X_t\}$ are generated, independently of E_t, by the modulo map and the logistic maps $\theta = 4.00$, $\theta = 3.98$, $\theta = 3.825$ and $\theta = 3.58$. Theorem 5.2 obviously applies for all the $\{X_t\}$ since $\gamma_E(i) = 0$, $i = 1, \ldots$. Simulations (Figure 5.7) confirm that $var(\hat{\beta}) = 1/n$, for all $\{X_t\}$.

Now, if $\{X_t\}$ is generated by the logistic map $\theta = 4.00$, $\{X_t\}$ and $\{E_t\}$ meet the requirements of Theorem 5.5 and thus $\hat{\beta}$ is asymptotically normal. The simulations (Table 5.4 and Figure 5.8) corroborate this; on the other hand, simulations suggest that $\hat{\beta}$ is also asymptotically normal for the other $\{X_t\}$ although the $\{X_t\}$ generated by the modulo map and the logistic maps $\theta = 3.98$, $\theta = 3.825$ and $\theta = 3.58$ do not satisfy the conditions of Theorem 5.5 or, a fortiori, Theorems 5.3 and 5.4. The next section indicates a possible reason for this.

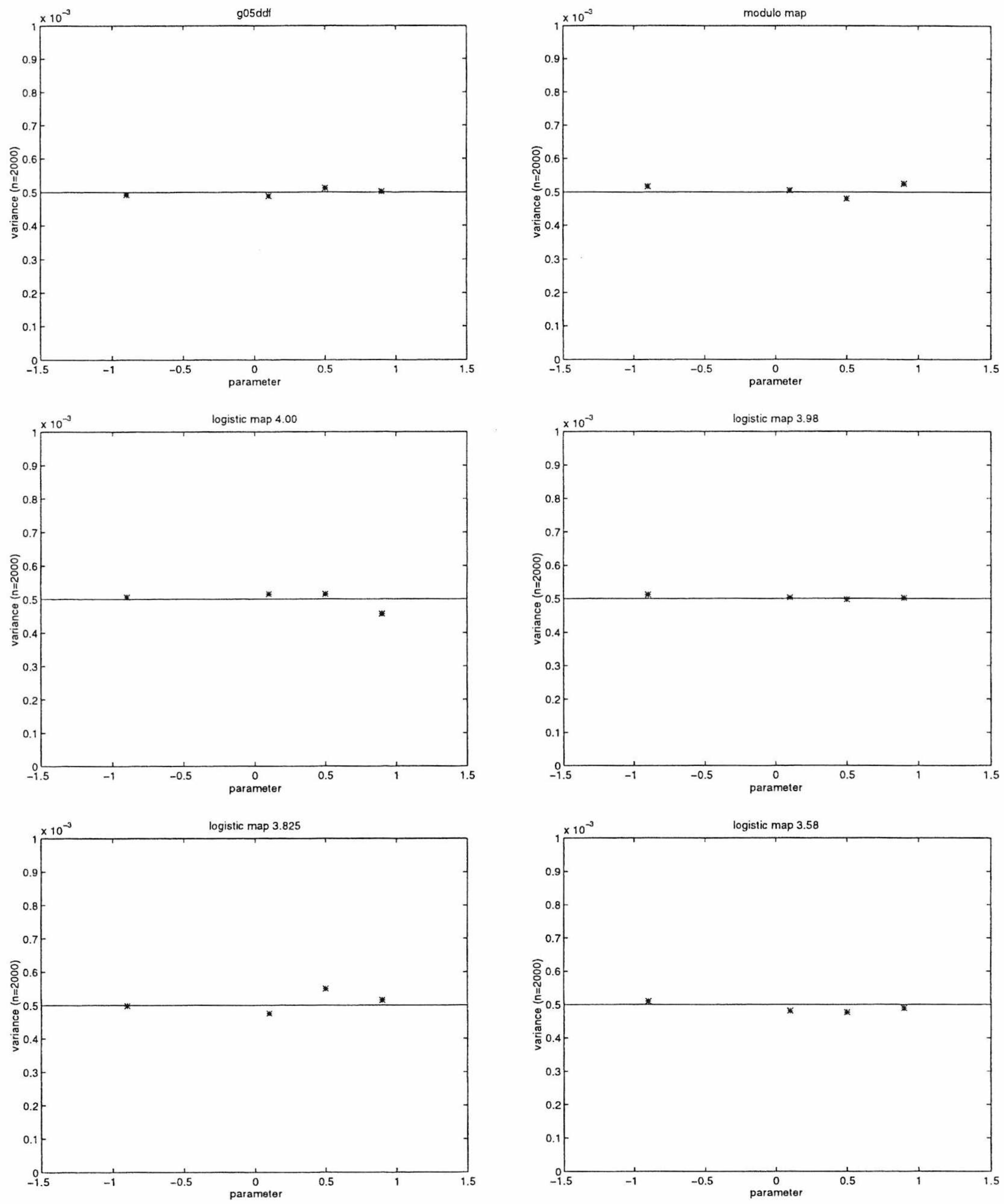


Figure 5.1: Variances of the estimators of β for example 1.
The simulated variances are obtained using 2000 replications of $\hat{\beta}_n$; n, the estimators sample size, is taken equal to 2000.
Simulated variances (denoted by an asterisk on the graphs) are obtained for four values of β (-0.9, 0.1, 0.5 and 0.9). The theoretical (asymptotic) variance is based on our paper (solid line).

	Sample skewness	Sample kurtosis	Lin-Mudholkar test statistic
g05ddf n = 2000 $\beta =$ - 0.9	0.022040	3.065404	- 0.197915
$\beta = 0.1$	- 0.042518	2.968338	0.397913
$\beta = 0.5$	0.035246	3.062191	- 0.313074
$\beta = 0.9$	- 0.007979	2.931642	0.074305
n = 5000 $\beta =$ - 0.9	0.008035	2.814271	- 0.076753
$\beta = 0.1$	- 0.101872	2.974190	0.972322
$\beta = 0.5$	0.070979	3.023089	- 0.628973
$\beta = 0.9$	- 0.028116	3.052059	0.255759
modulo map n = 2000 $\beta =$ - 0.9	- 0.058157	3.046030	0.536814
$\beta = 0.1$	- 0.023977	2.888702	0.226852
$\beta = 0.5$	- 0.004351	2.976477	0.040107
$\beta = 0.9$	- 0.009391	2.949795	0.087262
n = 5000 $\beta =$ - 0.9	0.013738	2.958020	- 0.126288
$\beta = 0.1$	- 0.032930	2.873775	0.314535
$\beta = 0.5$	- 0.075187	2.908727	0.722830
$\beta = 0.9$	- 0.046788	3.022108	0.431829
logistic map $\theta = 4.00$ n = 2000 $\beta =$ - 0.9	- 0.060127	3.181207	0.537654
$\beta = 0.1$	- 0.047553	3.098304	0.431066
$\beta = 0.5$	- 0.038284	2.867528	0.367288
$\beta = 0.9$	0.051977	3.138807	- 0.450099
n = 5000 $\beta =$ - 0.9	- 0.020754	2.892723	0.196475
$\beta = 0.1$	- 0.034389	2.868189	0.328977
$\beta = 0.5$	0.026214	2.916889	- 0.242363
$\beta = 0.9$	0.005884	3.168387	- 0.051288

Table 5.1: Asymptotic normality for example 1.
Sample skewness, sample kurtosis and Lin-Mudholkar test statistic were calculated using 2000 replications of $\hat{\beta}_n$

Table 5.1: continued

logistic map $\theta = 3.98$ n = 2000 $\beta =$ - 0.9	0.012593	2.947486	- 0.116111
$\beta = 0.1$	0.143107	2.996492	- 1.245828
$\beta = 0.5$	0.034713	3.115379	- 0.304613
$\beta = 0.9$	0.048316	2.915902	- 0.443895
n = 5000 $\beta =$ - 0.9	0.093684	3.162456	- 0.798489
$\beta = 0.1$	- 0.059229	3.061085	0.544222
$\beta = 0.5$	0.013736	3.123518	- 0.121468
$\beta = 0.9$	0.047675	2.981877	- 0.429504
logistic map $\theta = 3.825$ n = 2000 $\beta =$ - 0.9	0.024587	2.943978	- 0.225472
$\beta = 0.1$	- 0.049649	3.061751	0.454371
$\beta = 0.5$	- 0.030124	3.224413	0.264067
$\beta = 0.9$	0.004856	2.937390	- 0.044770
n = 5000 $\beta =$ - 0.9	0.120924	2.878059	- 1.092088
$\beta = 0.1$	- 0.001516	2.992549	0.014189
$\beta = 0.5$	0.020398	3.001345	- 0.185057
$\beta = 0.9$	0.098038	2.907311	- 0.885603
logistic map $\theta = 3.58$ n = 2000 $\beta =$ - 0.9	- 0.105007	3.126289	0.966190
$\beta = 0.1$	0.058001	2.961774	- 0.524104
$\beta = 0.5$	- 0.091358	2.926848	0.879781
$\beta = 0.9$	0.000254	3.266967	- 0.002079
n = 5000 $\beta =$ - 0.9	- 0.051014	3.080481	0.465252
$\beta = 0.1$	- 0.086969	3.112080	0.797515
$\beta = 0.5$	0.132109	3.292118	- 1.080722
$\beta = 0.9$	- 0.038028	2.947302	0.356980

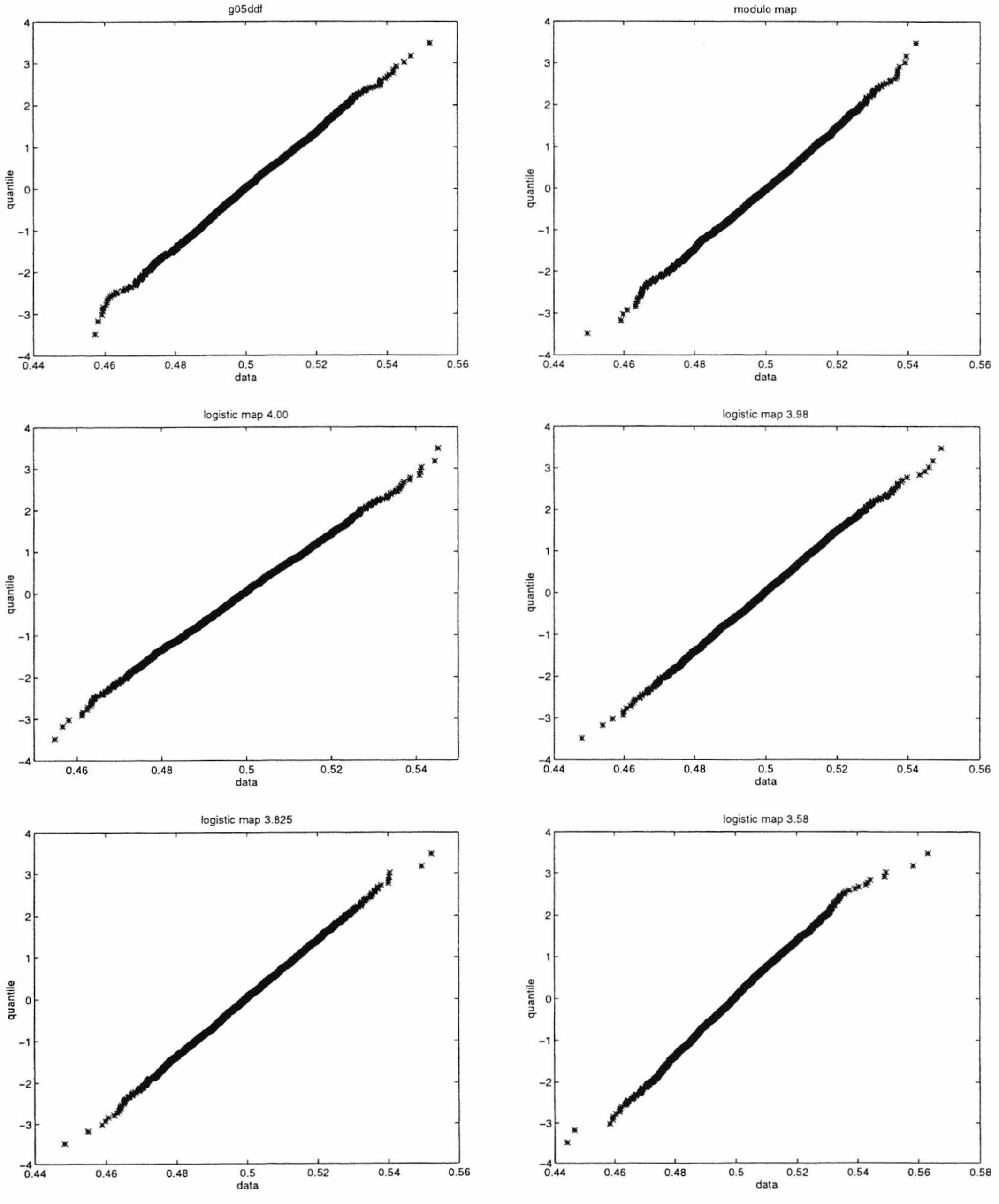


Figure 5.2: Normal probability plots for example 1.
The normal probability plots were obtained by using 2000 replications of $\hat{\beta}_n$; n was taken to be equal to 5000. We reproduce the plots for the value β =0.5. The plots for the other values of β are not significantly different from the ones here.

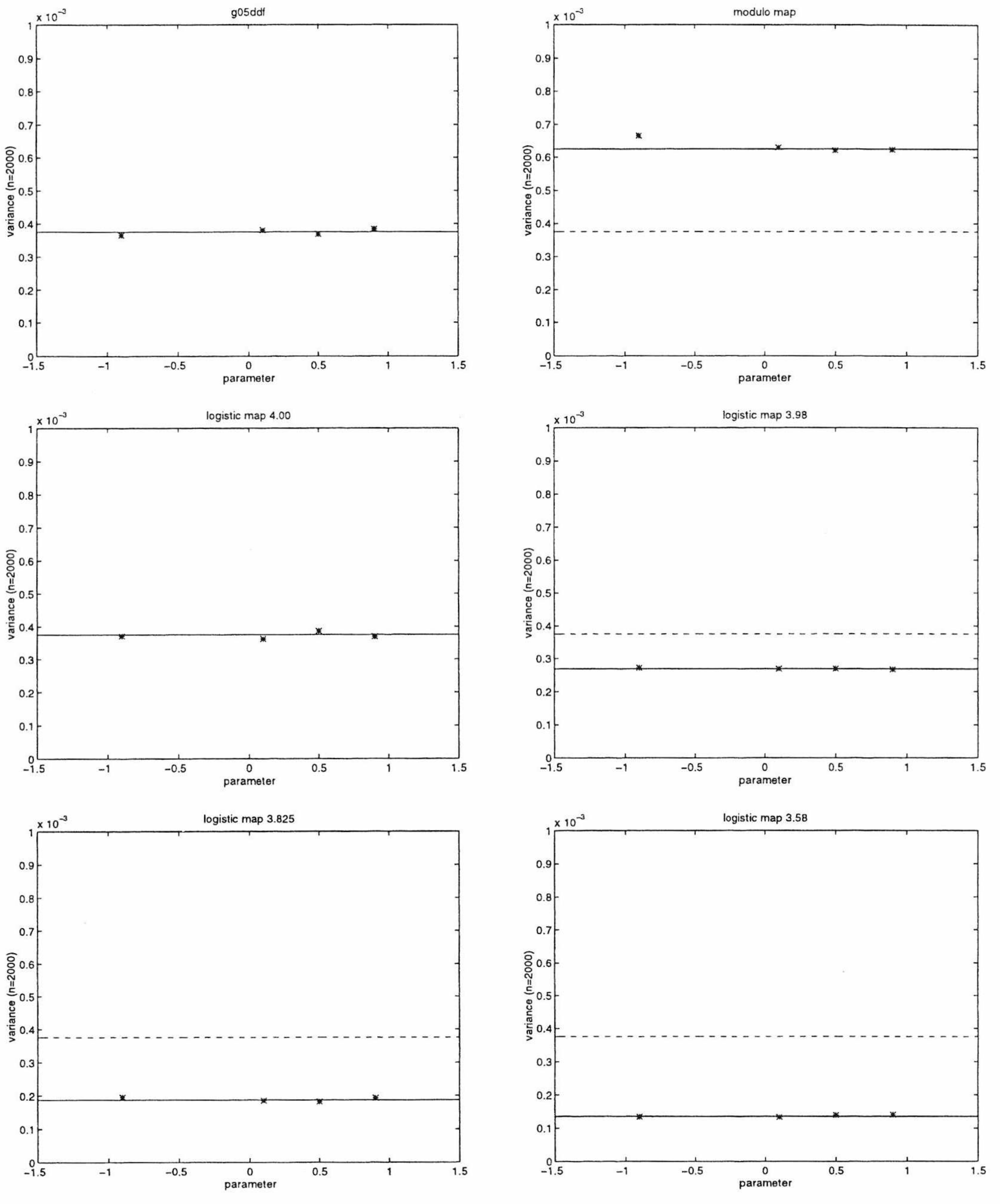


Figure 5.3: Variances of the estimators of β for example 2.
The simulated variances are obtained using 2000 replications of $\hat{\beta}_n$; n, the estimators sample size, is taken equal to 2000. Simulated variances (denoted by an asterisk on the graphs) are obtained for four values of β (-0.9, 0.1, 0.5 and 0.9). The theoretical (asymptotic) variance is based on our paper (solid line). When different from the theoretical value, the variance of the *iid* case is displayed (dashed line on the graphs).

	Sample skewness	Sample kurtosis	Lin-Mudholkar test statistic
g05ddf n = 2000 $\beta =$ - 0.9	0.058911	2.949253	- 0.534025
$\beta = 0.1$	- 0.112042	3.108897	1.037340
$\beta = 0.5$	- 0.004271	3.010583	0.039125
$\beta = 0.9$	- 0.029514	2.804668	0.287091
n = 5000 $\beta =$ - 0.9	- 0.075556	2.969972	0.714864
$\beta = 0.1$	0.088653	3.080090	- 0.770456
$\beta = 0.5$	0.032157	2.923131	- 0.295914
$\beta = 0.9$	0.017921	3.017702	- 0.162576
modulo map n = 2000 $\beta =$ - 0.9	0.028176	2.927859	- 0.258923
$\beta = 0.1$	- 0.050053	3.172654	0.446716
$\beta = 0.5$	0.028631	2.996737	- 0.259330
$\beta = 0.9$	- 0.028922	2.869930	0.276386
n = 5000 $\beta =$ - 0.9	0.090252	2.900915	- 0.819010
$\beta = 0.1$	- 0.065741	2.917982	0.628627
$\beta = 0.5$	0.008976	3.046495	- 0.080650
$\beta = 0.9$	0.008946	3.046467	- 0.080633
logistic map $\theta = 4.00$ n = 2000 $\beta =$ - 0.9	0.049587	2.974481	- 0.447866
$\beta = 0.1$	0.040944	2.969714	- 0.371021
$\beta = 0.5$	- 0.030980	3.014691	0.284563
$\beta = 0.9$	0.043746	2.910120	- 0.401801
n = 5000 $\beta =$ - 0.9	0.051413	2.872158	- 0.476549
$\beta = 0.1$	- 0.022345	2.935574	0.208649
$\beta = 0.5$	- 0.047335	2.831869	0.459553
$\beta = 0.9$	0.064447	2.983945	- 0.578053

Table 5.2: Asymptotic normality for example 2.
Sample skewness, sample kurtosis and Lin-Mudholkar test statistic were calculated using 2000 replications of $\hat{\beta}_n$

Table 5.2: continued

logistic map $\theta = 3.98$ n = 2000 $\beta =$ - 0.9	- 0.48422	3.262085	0.423641
$\beta = 0.1$	- 0.063960	3.155572	0.575278
$\beta = 0.5$	0.023472	3.078451	- 0.210030
$\beta = 0.9$	0.006850	2.815946	- 0.065586
n = 5000 $\beta =$ - 0.9	- 0.024245	3.004036	0.222940
$\beta = 0.1$	- 0.043841	2.919484	0.415906
$\beta = 0.5$	0.014915	3.016785	- 0.135478
$\beta = 0.9$	0.009675	3.074029	- 0.085941
logistic map $\theta = 3.825$ n = 2000 $\beta =$ - 0.9	0.031782	3.016561	- 0.286269
$\beta = 0.1$	- 0.001464	2.907003	0.013604
$\beta = 0.5$	0.127194	2.908697	- 1.137833
$\beta = 0.9$	- 0.075823	3.051904	0.702037
n = 5000 $\beta =$ - 0.9	0.072705	2.956622	- 0.654292
$\beta = 0.1$	- 0.004777	3.176480	0.042034
$\beta = 0.5$	0.018868	2.980779	- 0.171731
$\beta = 0.9$	0.040121	3.007966	- 0.360992
logistic map $\theta = 3.58$ n = 2000 $\beta =$ - 0.9	- 0.037055	2.892379	0.352641
$\beta = 0.1$	- 0.022058	2.903171	0.207893
$\beta = 0.5$	0.059152	3.078970	- 0.519473
$\beta = 0.9$	0.105597	2.852842	- 0.965040
n = 5000 $\beta =$ - 0.9	- 0.028649	3.035158	0.262295
$\beta = 0.1$	0.061373	2.955145	- 0.554767
$\beta = 0.5$	- 0.153308	3.107112	1.442278
$\beta = 0.9$	0.029376	2.906723	- 0.272188

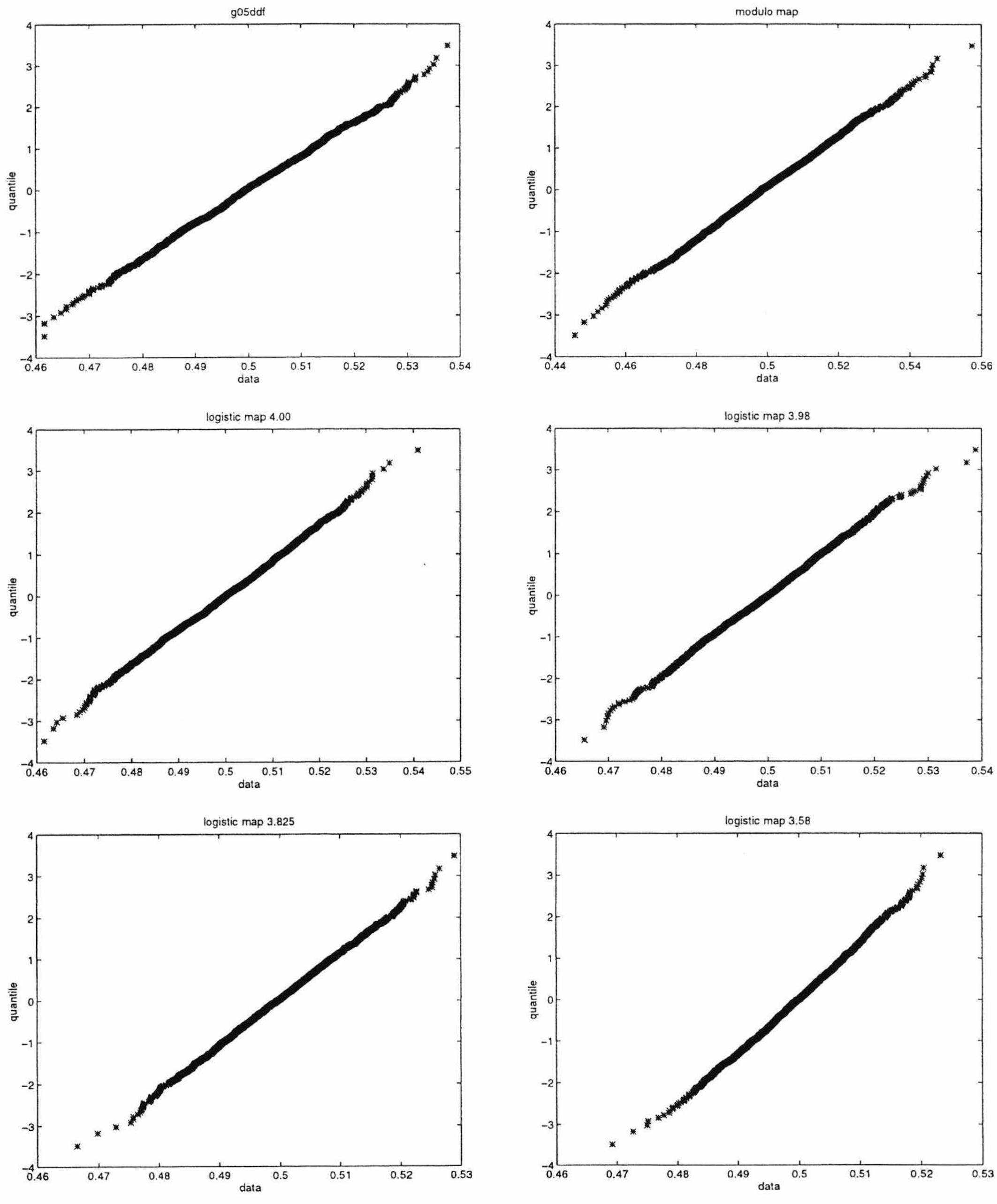


Figure 5.4: Normal probability plots for example 2.
The normal probability plots were obtained by using 2000 replications of $\hat{\beta}_n$; n was taken to be equal to 5000. We reproduce the plots for the value β =0.5. The plots for the other values of β are not significantly different from the ones here.

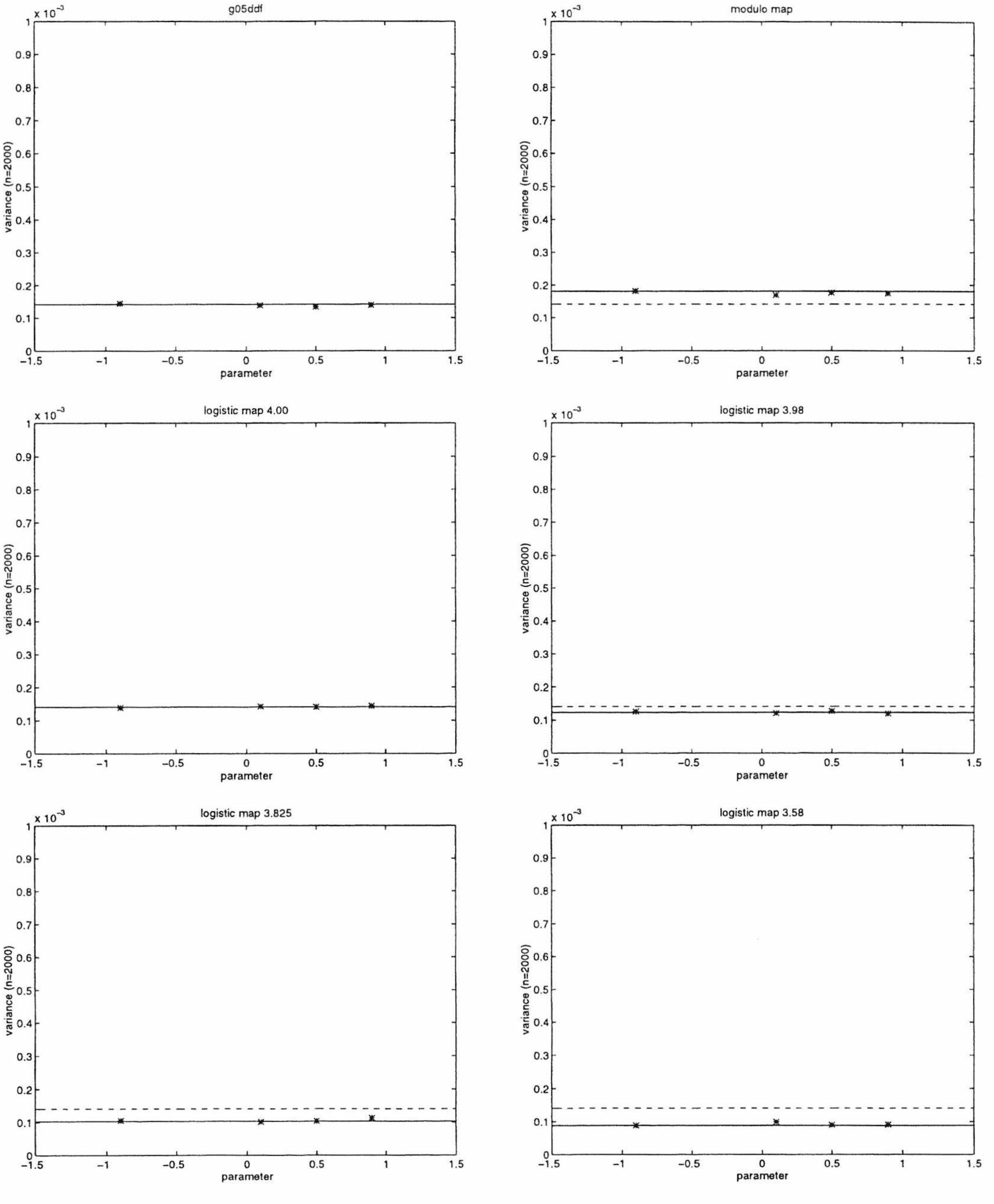


Figure 5.5: Variances of the estimators of β for example 3.
The simulated variances are obtained using 2000 replications of $\hat{\beta}_n$; n, the estimators sample size, is taken equal to 2000.
Simulated variances (denoted by an asterisk on the graphs) are obtained for four values of β (-0.9, 0.1, 0.5 and 0.9). The theoretical (asymptotic) variance is based on our paper (solid line). When different from the theoretical value, the variance of the *iid* case is displayed (dashed line on the graphs).

	Sample skewness	Sample kurtosis	Lin-Mudholkar test statistic
g05ddf n = 2000 β = - 0.9	- 0.043167	2.940596	0.406848
$\beta = 0.1$	- 0.047389	3.048025	0.434569
$\beta = 0.5$	0.031714	3.008956	- 0.285180
$\beta = 0.9$	- 0.012763	3.226253	0.110752
n = 5000 β = - 0.9	- 0.091213	3.002028	0.860572
$\beta = 0.1$	0.025228	3.040130	- 0.226248
$\beta = 0.5$	- 0.009994	2.918578	0.093641
$\beta = 0.9$	0.063891	2.931593	- 0.580189
modulo map n = 2000 β = - 0.9	- 0.048941	3.174955	0.435398
$\beta = 0.1$	0.048774	2.920711	- 0.447384
$\beta = 0.5$	- 0.027007	3.042101	0.246740
$\beta = 0.9$	- 0.039739	2.926235	0.375028
n = 5000 β = - 0.9	- 0.100851	3.054374	0.942476
$\beta = 0.1$	0.005110	3.063058	- 0.046004
$\beta = 0.5$	0.018668	3.107051	- 0.165370
$\beta = 0.9$	0.045393	2.996605	- 0.409007
logistic map $\theta = 4.00$ n = 2000 β = - 0.9	0.074366	3.004418	- 0.661013
$\beta = 0.1$	0.042516	3.028228	- 0.380105
$\beta = 0.5$	- 0.019117	3.008019	0.175145
$\beta = 0.9$	- 0.004901	3.051311	0.044189
n = 5000 β = - 0.9	- 0.057201	3.145481	0.514260
$\beta = 0.1$	- 0.124447	3.071069	1.168209
$\beta = 0.5$	- 0.024543	3.087424	0.221605
$\beta = 0.9$	- 0.093894	3.058190	0.874372

Table 5.3: Asymptotic normality for example 3.
Sample skewness, sample kurtosis and Lin-Mudholkar test statistic were calculated using 2000 replications of $\hat{\beta}_n$

Table 5.3: continued

logistic map $\theta = 3.98$ n = 2000 $\beta =$ - 0.9	- 0.042204	2.981463	0.393277
$\beta = 0.1$	0.058367	2.880236	- 0.538539
$\beta = 0.5$	- 0.168714	2.965149	1.656494
$\beta = 0.9$	0.030751	2.982472	- 0.279745
n = 5000 $\beta =$ - 0.9	- 0.076668	3.062892	0.708560
$\beta = 0.1$	0.019419	3.131635	- 0.171000
$\beta = 0.5$	- 0.101357	3.003555	0.960283
$\beta = 0.9$	0.091537	3.095664	- 0.791954
logistic map $\theta = 3.825$ n = 2000 $\beta =$ - 0.9	- 0.059207	2.991525	0.553546
$\beta = 0.1$	- 0.043354	2.999630	0.402653
$\beta = 0.5$	- 0.072246	3.020246	0.674211
$\beta = 0.9$	- 0.118830	3.169564	1.086786
n = 5000 $\beta =$ - 0.9	0.054531	3.019285	- 0.486623
$\beta = 0.1$	- 0.057383	3.077809	0.526235
$\beta = 0.5$	0.037175	2.737798	- 0.359020
$\beta = 0.9$	- 0.005072	3.176264	0.044606
logistic map $\theta = 3.58$ n = 2000 $\beta =$ - 0.9	0.087362	3.109046	- 0.753990
$\beta = 0.1$	- 0.039842	2.887086	0.380489
$\beta = 0.5$	- 0.031480	2.959448	0.293830
$\beta = 0.9$	- 0.062334	3.236238	0.550208
n = 5000 $\beta =$ - 0.9	0.001519	2.987700	- 0.013745
$\beta = 0.1$	- 0.017323	2.911859	0.162914
$\beta = 0.5$	- 0.037747	2.949071	0.354008
$\beta = 0.9$	0.011144	2.931654	- 0.103300

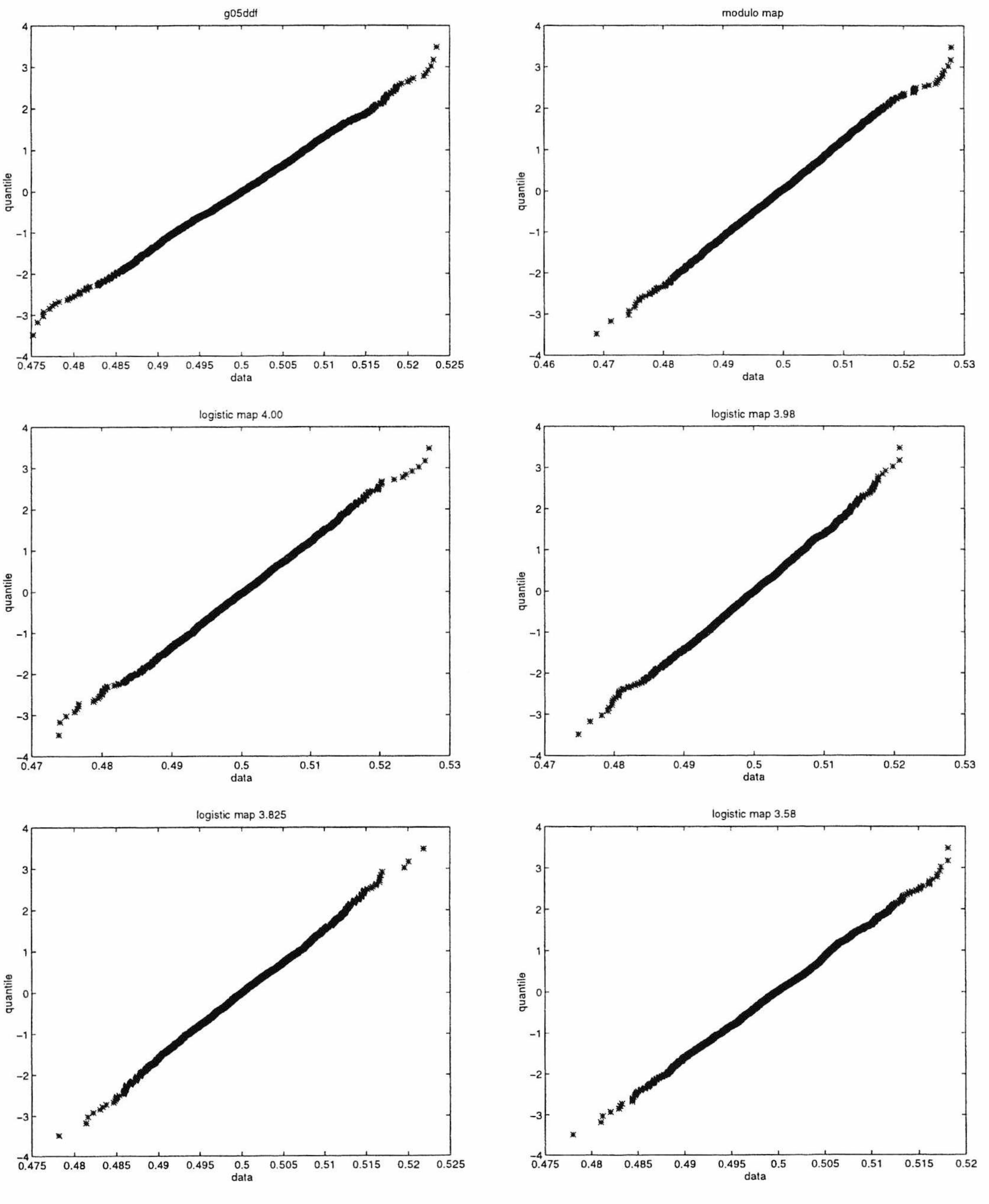


Figure 5.6 Normal probability plots for example 3.
The normal probability plots were obtained by using 2000 replications of $\hat{\beta}_n$; n was taken to be equal to 5000. We reproduce the plots for the value β =0.5. The plots for the other values of β are not significantly different from the ones here.

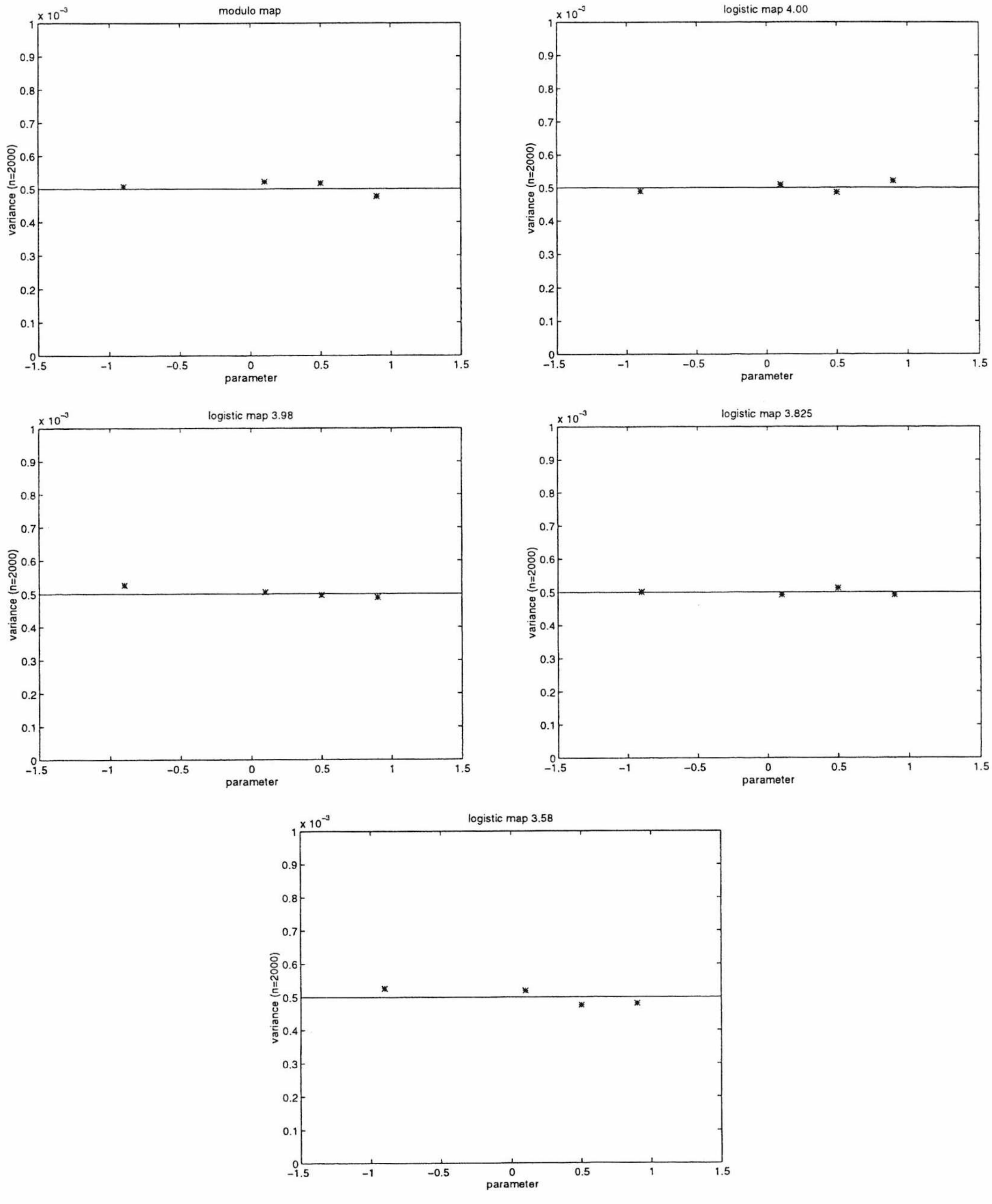


Figure 5.7: Variances of the estimators of β for example 4.
The simulated variances are obtained using 2000 replications of $\hat{\beta}_n$; n, the estimators sample size, is taken equal to 2000.
Simulated variances (denoted by an asterisk on the graphs) are obtained for four values of β (-0.9, 0.1, 0.5 and 0.9). The theoretical (asymptotic) variance is based on our paper (solid line).

	Sample skewness	Sample kurtosis	Lin-Mudholkar test statistic
modulo map n = 2000 β = - 0.9	0.052629	2.935793	- 0.479439
$\beta = 0.1$	0.073487	3.081560	- 0.642097
$\beta = 0.5$	0.066237	3.016511	- 0.589392
$\beta = 0.9$	- 0.105216	2.957195	1.009930
n = 5000 β = - 0.9	- 0.005980	3.021834	0.054218
$\beta = 0.1$	0.015619	3.126208	- 0.137381
$\beta = 0.5$	- 0.029980	3.072433	0.272365
$\beta = 0.9$	- 0.095176	3.040944	0.891138
logistic map $\theta = 4.00$ n = 2000 β = - 0.9	- 0.016154	3.061507	0.146867
$\beta = 0.1$	0.052299	2.936906	- 0.476442
$\beta = 0.5$	- 0.022526	2.916575	0.212134
$\beta = 0.9$	0.028213	3.031125	- 0.253730
n = 5000 β = - 0.9	0.006230	2.957358	- 0.057118
$\beta = 0.1$	0.050465	2.925851	- 0.461161
$\beta = 0.5$	- 0.052442	3.011357	0.486137
$\beta = 0.9$	- 0.023180	2.920517	0.218344
logistic map $\theta = 3.98$ n = 2000 β = - 0.9	- 0.016376	3.128877	0.146202
$\beta = 0.1$	0.055360	2.890137	- 0.509674
$\beta = 0.5$	0.090501	2.986479	- 0.804223
$\beta = 0.9$	0.067328	2.896332	- 0.616014
n = 5000 β = - 0.9	0.019800	2.923412	- 0.182340
$\beta = 0.1$	0.130499	3.084124	- 1.117711
$\beta = 0.5$	- 0.014515	2.887800	0.136886
$\beta = 0.9$	0.123112	3.111031	- 1.050643

Table 5.4: Asymptotic normality for example 4.
Sample skewness, sample kurtosis and Lin-Mudholkar test statistic were calculated using 2000 replications of $\hat{\beta}_n$

Table 5.4: continued

logistic map θ = 3.825 n = 2000 β = - 0.9	- 0.031183	2.910835	0.294799
β = 0.1	0.006320	3.076893	- 0.055962
β = 0.5	0.051307	3.044352	- 0.455348
β = 0.9	0.012728	2.961693	- 0.117078
n = 5000 β = - 0.9	0.003875	2.931882	- 0.035626
β = 0.1	0.038739	2.907221	- 0.357371
β = 0.5	- 0.046381	2.922019	0.439429
β = 0.9	- 0.013851	2.872886	0.131478
logistic map θ = 3.58 n = 2000 β = - 0.9	- 0.004454	2.956451	0.041193
β = 0.1	- 0.002275	3.136049	0.020371
β = 0.5	0.067461	2.998111	- 0.602345
β = 0.9	0.063379	3.048641	- 0.560099
n = 5000 β = - 0.9	0.009459	2.950808	- 0.087037
β = 0.1	0.058007	2.936228	- 0.527561
β = 0.5	0.047564	3.084741	- 0.418523
β = 0.9	0.019707	2.996737	- 0.179183

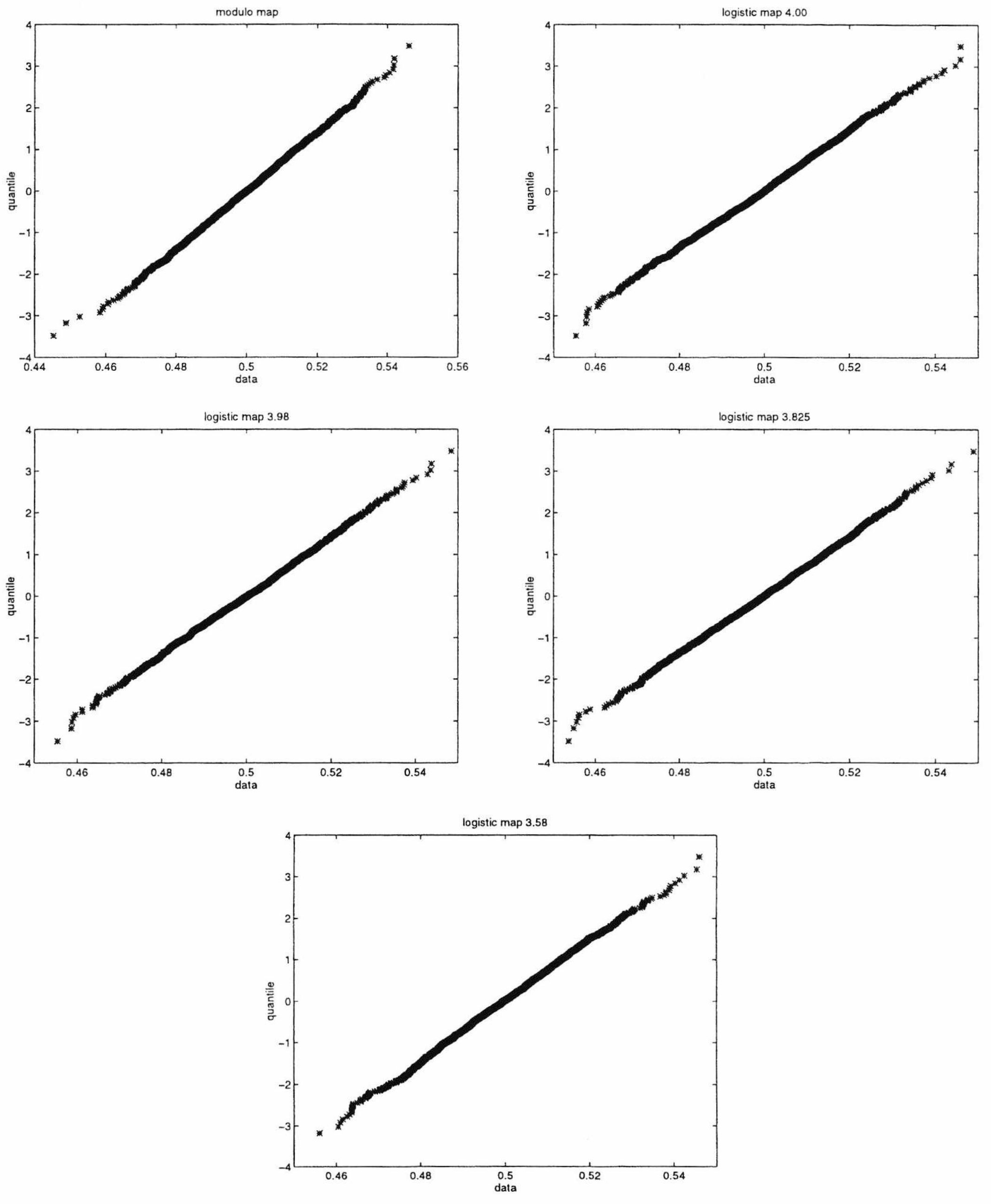


Figure 5.8 Normal probability plots for example 4.
The normal probability plots were obtained by using 2000 replications of $\hat{\beta}_n$; n was taken to be equal to 5000. We reproduce the plots for the value β =0.5. The plots for the other values of β are not significantly different from the ones here.

5.5 Towards a central limit theorem for general chaotic sequences

In the chapters 4 and 5, we have got central limit properties for a lot of parameter estimators of chaos driven models. Some of them were predicted by our theory but many others were not a priori expected. Let us mention for example the autoregressive models driven by logistic maps with $\theta = 3.98$ and $\theta = 3.825$, or the stochastic regression models only involving chaotic sequences.

My conviction is that there exists a central limit theorem for chaotic sequences $\{E_t\}$ (and, by extension, for $\{h(E_t)\}$) provided some mild conditions are satisfied, like the absolute summability of the autocorrelations ($\sum_{i=1}^{\infty} |\rho_E(i)| < \infty$). In particular, if this theorem exists, it would be the underlying reason explaining our simulation results. In the following of this section, I shall motivate my conviction by describing informally a very promising possible way for getting this central limit theorem; before doing that, some simulations are shown, which reinforce the idea.

Table 5.5 and Figure 5.9 deal with the asymptotic normality of $n^{-1/2} \sum_{i=1}^{n} E_i$, where $\{E_t\}$ are zero mean chaotic sequences generated by the modulo map and the four logistic maps under consideration throughout the thesis. We can discern apparent normality for all of them; in particular, the sequences generated by the logistic maps with $\theta = 3.98$, $\theta = 3.825$ and $\theta = 3.58$ seem to admit a central limit theorem, although there is no theoretical justification to this. We now indicate a possible way for proving formally this central limit property.

	Sample skewness	Sample kurtosis	Lin-Mudholkar test statistic
modulo map n = 2000	- 0.028703	3.250566	0.249478
n = 5000	- 0.040171	3.166374	0.357585
logistic map $\theta = 4.00$ n = 2000	- 0.113510	2.892336	1.112318
n = 5000	- 0.079357	2.944676	0.757063
logistic map $\theta = 3.98$ n = 2000	0.063462	3.063883	- 0.558807
n = 5000	0.030343	2.986421	- 0.275722
logistic map $\theta = 3.825$ n = 2000	0.148756	2.920087	- 1.317247
n = 5000	0.071949	3.155663	- 0.617811
logistic map $\theta = 3.58$ n = 2000	0.212373	2.942397	- 1.831800
n = 5000	0.185259	3.201798	- 1.520086

Table 5.5: Asymptotic normality.
Sample skewness, sample kurtosis and Lin-Mudholkar test statistic were calculated using 2000 replications.

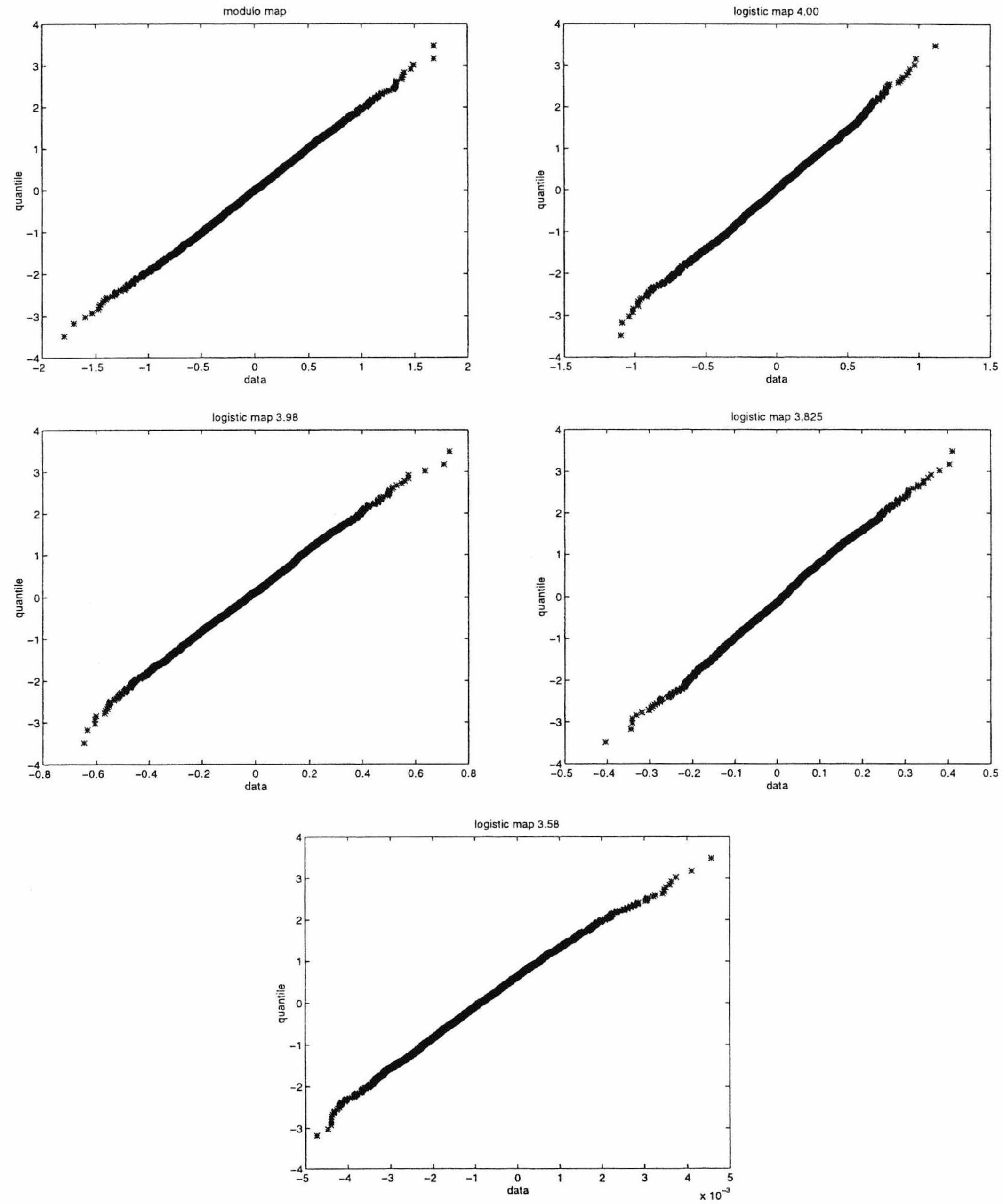


Figure 5.9: Normal probability plots.
The normal probability plots were obtained by using 2000 replications; n was taken to be equal to 5000.

We start from the following observation: the chaotic sequences $\{E_t\}$ are not strongly mixing but they are not far from it in the sense that they are K-mixing systems (see section 3.2). In fact, the main reason for not being strongly mixing is that $E_{t+1} = f(E_t)$, which means that the future can be (perfectly) predicted if the past is (perfectly) known.

Now, consider the following sequence $\{E_t + \varepsilon_t\}$, where $\{\varepsilon_t\}$ is a sequence of *iid* random variables with mean 0 and finite variance σ^2 ($\{\varepsilon_t\}$ can be seen as additive noise). The sequence $\{E_t + \varepsilon_t\}$ is not purely deterministic any more and the fact to know $E_0 + \varepsilon_0$ does not help at all to locate E_n, or a fortiori $E_n + \varepsilon_n$, for large n because of the sensitive dependence on the initial conditions of the chaotic map f. So, it seems reasonable to conjecture that $\{E_t + \varepsilon_t\}$ is strongly mixing. Note, however, that proving formally the strong mixing property for such a sequence is a difficult task.

If the conjecture ($\{E_t + \varepsilon_t\}$ is strongly mixing) is true, we can then make use of one of the many central limit theorems existing for strongly mixing sequences. For example, there is a theorem by Denker (1986) which states: " For a strongly mixing sequence, the central limit theorem holds if and only if the squares of the normalised partial sums are uniformly integrable.". Basically, for our sequence, a sufficient condition for getting the central limit theorem is that $\mathrm{E}|\sum_{i=1}^{n} E_i|^{2+\delta} = o(n^{1+\delta})$ for some $\delta > 0$.

Now, the last step is to show that

$$(n^{-1/2}\sum_{i=1}^{n}(E_i + \varepsilon_i) \to^d \mathcal{N}(0, \sigma^2_{E+\varepsilon})) \Rightarrow (n^{-1/2}\sum_{i=1}^{n} E_i \to^d \mathcal{N}(0, \sigma^2_E)).$$

Let $E_n^* = n^{-1/2}\sum_{i=1}^{n} E_i$ and $\varepsilon_n^* = n^{-1/2}\sum_{i=1}^{n} \varepsilon_i$.

Looking at the characteristic functions, we get

$$\mathrm{E}[e^{it(E_n^* + \varepsilon_n^*)}] \to e^{(-t^2/2)\sigma^2_{E+\varepsilon}}$$

and by independence of $\{E_t\}$ and $\{\varepsilon_t\}$

$$\mathrm{E}[e^{itE_n^*}]\mathrm{E}[e^{it\varepsilon_n^*}] \to e^{(-t^2/2)\sigma^2_{E+\varepsilon}}.$$

Now, since $\varepsilon_t \sim iid(0, \sigma^2)$, we have

$$\mathrm{E}[e^{it\varepsilon_n^*}] \to e^{(-t^2/2)\sigma^2_\varepsilon}.$$

So,

$$\mathrm{E}[e^{itE_n^*}] \to \frac{e^{(-t^2/2)\sigma^2_{E+\varepsilon}}}{e^{(-t^2/2)\sigma^2_\varepsilon}} = e^{(-t^2/2)\sigma^2_E}$$

since $\sigma^2_{E+\varepsilon} = \sigma^2_E + \sigma^2_\varepsilon$ by independence of $\{E_t\}$ and $\{\varepsilon_t\}$.

Since, $e^{(-t^2/2)\sigma^2_E}$ is continuous at $t = 0$, we get

$$n^{-1/2}(E_1 + ... + E_n) \to^d \mathcal{N}(0, \sigma^2_E).$$

Moreover, it can easily be seen that $\sigma^2_E = \sum_{j=-\infty}^{j=\infty} \gamma_E(j)$ (see, for example, Theorem 7.1.1 of Brockwell and Davis (1989)).

So, if the conjecture is true, we get a central limit theorem for chaotic sequences $\{E_t\}$ provided $\mathrm{E}|\sum_{i=1}^n E_i|^{2+\delta} = o(n^{1+\delta})$ for some $\delta > 0$. Obviously, a similar reasoning applies for sequences $\{h(E_t)\}$.

Now, on-going research suggests that the conjecture could be true but there is no definite answer yet.

Chapter 6

Conclusion

We have analysed the effect of dynamic noise on the attractors of dynamical systems. This is of particular practical interest since any orbit obtained from computer calculations is subject to rounding errors. We have shown that the attractors (defined as in Ruelle (1981)) are stable under infinitesimal random perturbations which are bounded, independent and identically distributed with absolutely continuous distribution. Moreover, it has been possible, given the noise level, to construct the noisy attractors corresponding to systems subject to more general forms of dynamic noise.

We have introduced chaotic sequences. Besides their academic interests, these sequences are important from a practical point of view because their analyses cast light on computer simulations. We have obtained interesting properties for chaotic sequences; in particular, a strong law of large numbers can easily be proved for their autocovariances.

We have analysed AR(k) models driven by chaotic sequences. We were especially interested to enquire why the simulated results seem to support the conclusion of central limit properties even though the assumption of independence no longer holds for pseudo-random number generators. We have evaluated the asymptotic bias of the classical estimator of α for such chaos driven AR(k) models. Moreover, it has been possible to obtain asymptotic normality of this estimator for some chaotic random sequences. This gives the asymptotic variance as a by-product.

We have considered chaos driven linear stochastic regression models. Under the assumption that the explanatory random vector and the chaotic sequence are independent, we have obtained consistency of the classical linear regression estimator. It has also been possible to give its asymptotic variance. Moreover, provided some reasonable additional conditions on the explanatory random variable are satisfied, we have also proved the asymptotic normality of the regression estimator.

Now, here are some directions of generalization.

The simulation results in chapters 4 and 5 suggest that asymptotic normality of the parameter estimator might still be obtained even if we relax some of our conditions. A possible explanation and theoretical justification could come from a central limit theorem for chaotic sequences under weak conditions. Section 5.5 sketches a promising way for proving such a theorem. There is, however, no definite answer yet; there could exist an easier or more general proof or it is even possible that a central limit theorem does not exist for general chaotic sequences.

Another direction of generalization is to extend to chaos driven non linear time series and to nonparametric regression techniques and nonparametric estimation when the *iid* sample is replaced by a chaotic sample. Without entering into the details here, we would like to mention that the first results are promising: in particular, for the kernel density estimation in the presence of a chaotic sample, it can be seen, provided some weak conditions are satisfied, that the kernel estimator is consistent and that the ideal window width is of order $n^{-1/5}$ as in the *iid* case.

Finally, we point out the inextricable interactions between the noisy dynamical systems dealt with in chapter 2 and the chaos driven models analysed in chapters 4 and 5. On the one hand, chapter 2 analyses dynamical systems subject to some kind of *iid* noise. However, computers simulate *iid* noise by means of pseudo-random number generators which are typically sensitive to initial values. So, a better model approximation of dynamical systems under computer noise is a (complex) chaos driven model. On the other hand, the chaos driven models of chapters 4 and 5 are themselves only approximations of computer simulations since no computer generated numbers can be free from rounding errors. In this case, stochastic randomness returns to the model albeit at a deeper level. Section 4.6 illustrates the situation for an AR(1) model and provides us with asymptotic properties of $\hat{\alpha}$ for quite general AR(1) models driven by noisy chaos. Now, it is clear that section 4.6 is just a first step and further investigations are necessary.

References

Akaike, H. (1985). Prediction and entropy. In *A celebration of statistics: the ISI centenary volume.* New York: Springer. A. C. Atkinson and S. E. Fienberg editors.

Anderson, T. W. (1971). *The statistical analysis of time series.* New York: Wiley.

Barnsley, M. (1988). *Fractals everywhere.* New York: Academic Press.

Bartlett, M. S. (1955). *An introduction to stochastic processes (with special reference to methods and applications).* Cambridge: Cambridge University Press.

Basawa, I. V. and B. L. S. Prakasa Rao (1980). *Statistical inference for stochastic processes.* New York: Academic Press.

Beran, J. (1994). *Statistics for long-memory processes.* New York: Chapman and Hall.

Berge, C. (1971). *Principles of combinatorics.* New York: Academic Press.

Berge, P., Y. Pomeau, and C. Vidal (1984). *L' ordre dans le chaos: vers une approche déterministe de la turbulence.* Paris: Hermann.

Berliner, L. M. (1991). Likelihood and Bayesian prediction of chaotic systems. *Journal of the American Statistical Association 86*, 938–952.

Billingsley, P. (1965). *Ergodic theory and information.* Wiley series in probability and mathematical statistics. New York: Wiley.

Blank, M. (1994). Pathologies generated by round-off in dynamical systems. *Physica D 78*, 93–114.

Bosq, D. and D. Guégan (1994). Nonparametric estimation of the chaotic function and the invariant measure of a dynamical system. Technical Report, INSEE-ENSAE.

Box, G. E. P. and G. M. Jenkins (1970). *Time series analysis, forecasting and control.* San Francisco: Holden-Day.

Bradley, R. C. (1983). Absolute regularity and functions of markov chains. *Stochastic Processes and their Applications 14*, 67–77.

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in probability and statistics: a survey of recent results*, pp. 165–192. Basel: Birkhäuser. E. Eberlein, M. S. Taqqu editors.

Brockwell, P. J. and R. A. Davis (1989). *Time series: theory and methods* (2 ed.). New York: Springer-Verlag.

Chan, K. S. and H. Tong (1994). A note on noisy chaos. *Journal of the Royal Statistical Society, Series B 56*, 301–311.

Chan, K. S. and H. Tong (1997). *Chaos: from a statistical perspective.* Heidelberg: Springer-Verlag. To appear.

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

Cheng, B. and H. Tong (1992). On consistent non-parametric order determination and chaos. *Journal of the Royal Statistical Society, Series B 54*, 427–449.

Cheng, B. and H. Tong (1994). Orthogonal projection, embedding dimension and sample size in chaotic time series from a statistical perspective. *Philosophical Transactions of the Royal Society A348*, 325–341.

Chow, Y. S. and H. Teicher (1988). *Probability theory: independence, interchangeability, martingales* (2 ed.). Springer Texts in Statistics. New York:

Springer-Verlag.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Cutler, C. D. (1993). A review of the theory and estimation of fractal dimension. In H. Tong (Ed.), *Dimension estimation and models*, Nonlinear Time Series and Chaos, pp. 1–107. London: World Scientific.

De Wilde, M. (1990). Géométrie différentielle. Lecture Notes, University of Liège.

Denker, M. (1986). Uniform integrability and the central limit theorem for strongly mixing processes. In *Dependence in probability and statistics: a survey of recent results*, pp. 269–274. Basel: Birkhäuser. E. Eberlein, M. S. Taqqu editors.

Denker, M. and G. Keller (1983). On U-statistics and von Mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 64*, 505–522.

Denker, M. and G. Keller (1986). Rigorous statistical procedures for data from dynamical systems. *Journal of Statistical Physics 44*, 67–93.

Diks, C. (1996). *On nonlinear time series analysis, spatio-temporal complexity and noise.* Ph. D. thesis, University of Leiden.

Diks, C., J. C. van Houwelingen, F. Takens, and J. DeGoede (1995). Reversibility as a criterion for discriminating time series. *Physics Letters A 201*, 221–228.

Doob, J. L. (1953). *Stochastic processes.* New York: Wiley.

Doukhan, P. (1994). *Mixing: properties and examples.* New York: Springer-Verlag.

Draper, N. R. and H. Smith (1966). *Applied regression analysis.* New York: Wiley.

Fan, J. and I. Gijbels (1996). *Local polynomial modelling and its applications.* London: Chapman and Hall.

Fan, J., Q. Yao, and H. Tong (1993). Estimating measures of sensitivity of initial values to nonlinear stochastic systems with chaos. Technical Report, University of Kent.

Gelfand, I. M. and A. M. Yaglom (1957). Calculation of the amount of information about a random function contained in another such function. *Upsekhi Matem. Nauk. 12*, 3–52.

Giraitis, L. and D. Surgailis (1986). Multivariate Appell polynomials and the central limit theorem. In *Dependence in probability and statistics: a survey of recent results*, pp. 21–71. Basel: Birkhäuser. E. Eberlein, M. S. Taqqu editors.

Grassberger, P., R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber (1993). On noise reduction methods for chaotic data. *Chaos 3*, 127–140.

Györfi, L., W. Härdle, P. Sarda, and P. Vieu (1989). *Nonparametric curve estimation from time series.* Lecture Notes in Statistics. New York: Springer-Verlag.

Hadamard, J. (1898). Les surfaces à courbures opposées et leurs lignes géodésiques. *Journal des Mathématiques Pures et Appliquées 4*, 27.

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

Hall, P. and R. C. L. Wolff (1995a). Properties of invariant distributions and Lyapunov exponents for chaotic logistic maps. *Journal of the Royal Statistical society, Series B 57*, 439–452.

Hall, P. and R. C. L. Wolff (1995b). On the strength of dependence of a time series generated by a chaotic map. *Journal of Time Series Analysis 16*, 571–584.

Halmos, P. R. (1956). *Lectures on ergodic theory.* New York: Chelsea.

Hannan, E. J. (1970). *Multiple time series.* New York: Wiley.

Hannan, E. J. (1976). The asymptotic distribution of serial covariances. *The Annals of Statistics 4*, 396–399.

Hofbauer, F. and G. Keller (1982). Ergodic properties of invariant measures for piecewise monotonic transformations. *Mathematische Zeitschrift 180*, 119–140.

Holton, D. and R. M. May (1993). Distinguishing chaos from noise. In *The nature of chaos*, pp. 149–182. Oxford: Clarendon Press. Edited by T. Mullin.

Ibragimov, I. A. and Y. V. Linnik (1971). *Independent and stationary sequences of random variables.* Groningen: Wolters-Noordhoff.

Kaplan, J. L. and J. A. Yorke (1979). Chaotic behavior of multidimensional difference equations. In *Functional differential equations and approximations of fixed points*, pp. 204–227. Berlin: Springer. H.-O. Peitgen and H.-O. Walter editors.

Keller, G. and R. Sporer (1996). Remarks on the linear regression approach to dimension estimation. In *Stochastic and spatial structures of dynamical systems*, pp. 17–27. Amsterdam: North-Holland. Edited by S. J. van Strien and S. M. Verduyn Lunel.

Khinchin, A. I. (1957). *Mathematical foundations of information theory.* New York: Dover.

Kifer, Y. (1988). *Random perturbations of dynamical systems.* Basel: Birkhäuser.

Knuth, D. E. (1981). *The art of computer programming. Vol. 2: seminumerical algorithms* (2 ed.). New York: Addison-Wesley.

Kolmogorov, A. N. (1956). On the Shannon theory of information transmission in the case of continuous signals. *IRE Transactions on Information Theory*, 102–108.

Koroljuk, V. S. and I. V. Borovskikh (1994). *Theory of U-statistics.* Dordrecht: Kluwer Academic.

Krengel, U. (1985). *Ergodic theorems.* Berlin: W. De Gruyter.

Lai, T. L. (1980). Asymptotic properties of nonlinear least squares estimates in stochastic regression models. Technical Report, University of Stanford.

Lai, T. L. (1983). Stochastic regression models and consistency of the least squares identification scheme. In *Lecture Notes in Statistics. Mathematical Learning models- Theory and Algorithms*, pp. 115–125. Edited by U. Herkenrath, D. Kalin and W. Vogel.

Lai, T. L. and C. Z. Wei (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics 10*, 154–166.

Lai, T. L. and C. Z. Wei (1983). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *Journal of Multivariate Analysis 13*, 1–23.

Lawrance, A. J. (1992). Uniformly distributed first-order autoregressive time series models and multiplicative congruential random number generators. *Journal of Applied Probability 29*, 896–903.

Lawrance, A. J. and N. M. Spencer (1997). Curved chaotic map time series models and their stochastic reversals. (Under revision for *Scandinavian Journal of Statistics*).

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature 261*, 459–467.

Nadaraya, E. A. (1989). *Nonparametric estimation of probability densities and regression curves.* Dordrecht: Kluwer Academic.

Ott, E. (1993). *Chaos in dynamical systems.* Cambridge: Cambridge University Press.

Parry, W. (1981). *Topics in ergodic theory.* Cambridge: Cambridge University Press.

Peitgen, H.-O., H. Jürgens, and D. Saupe (1992). *Chaos and fractals, new frontiers of science.* Berlin: Springer-Verlag.

Peligrad, M. (1986). Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey). In *Dependence in probability and statistics: a survey of recent results*, pp. 193–223. Basel: Birkhäuser. E. Eberlein, M. S. Taqqu editors.

Petersen, K. (1983). *Ergodic theory.* Cambridge: Cambridge University Press.

Pinsker, M. S. (1964). *Information and information stability of random variables.* San Francisco: Holden-Day.

Poincaré, H. (1905). Science and hypothesis. Reprinted in student edition by Dover, New York, in 1952.

Ripley, B. D. (1990). Thoughts on pseudorandom number generators. *Journal of Computational and Applied Mathematics 31*, 153–163.

Robinson, P. M. (1983). Non-parametric estimation for time series models. *Journal of Time Series Analysis 4*, 185–208.

Romano, J. P. and L. A. Thombs (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association 91*, 590–600.

Rosenblatt, M. (1971). *Markov processes: structure and asymptotic behavior.* Berlin: Springer-Verlag.

Rosenblatt, M. (1980). Linear processes and bispectra. *Journal of Applied Probability 17*, 265–270.

Roubens, M. (1990). Calcul des probabilités. Lecture notes, University of Liège.

Ruelle, D. (1981). Small random perturbations of dynamical systems and the definition of attractors. *Communications in Mathematical Physics 82*, 137–151.

Ruelle, D. (1989). *Chaotic evolution and strange attractors.* Cambridge: Cambridge University Press.

Schmets, J. (1989). Théorie de la mesure. Lecture Notes, University of Liège.

Schwartz, L. (1973). *Radon measures on arbitrary topological spaces and cylindrical measures.* Oxford: Oxford University Press.

Shiryayev, A. N. (1984). *Probability.* New York: Springer-Verlag.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* New York: Chapman and Hall.

Sinai, Y. G. (1962). On limit theorems for stationary processes. *Theory of Probability and its Applications 7*, 205–210.

Sinai, Y. G. (1989). *Dynamical systems, volume 2 (Ergodic theory with applications to dynamical systems and statistical mechanics).* New York: Springer-Verlag.

Smith, R. L. (1992). Estimating dimension in noisy chaotic time series. *Journal of the Royal Statistical Society, Series B 54*, 329–352.

Stockis, J.-P. (1994). Attractors under dynamic noise. M. Sc. Dissertation, University of Kent.

Stockis, J.-P. and H. Tong (1996). On the statistical inference of a machine-generated AR(1) model. Technical Report, University of Kent.

Sun, T. C. and H. C. Ho (1986). On central and non-central limit theorems for non-linear functions of a stationary gaussian process. In *Dependence in probability and statistics: a survey of recent results*, pp. 5–19. Basel: Birkhäuser. E. Eberlein, M. S. Taqqu editors.

Takens, F. (1994a). Analysis of non-linear time series with noise. Technical report, University of Groningen.

Takens, F. (1994b). Analysis of non-linear time series, a survey. Technical Report, University of Groningen.

Takens, F. (1996). The effect of small noise on systems with chaotic dynamics. In *Stochastic and spatial structures of dynamical systems*, pp. 3–15. Amsterdam: North-Holland. Edited by S. J. van Strien and S. M. Verduyn Lunel.

Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford: Clarendon Press.

Tong, H. (1992). Lecture notes on time series analysis. University of Kent at Canterbury.

Tong, H. (1995). A personal overview of non-linear time series analysis from a chaos perspective. with discussion and comments. *Scandinavian Journal of Statistics 22*, 399–445.

Tong, H. and R. L. Smith (1991). An informal guide to chaos. Unpublished handout given at the Chaos Day meeting of the Royal Statistical Society on 16 October 1991.

Volkonskii, V. A. and Y. A. Rozanov (1959). Some limit theorems for random functions, part 1. *Theory of Probability and its Applications 4*, 178–197.

Volkonskii, V. A. and Y. A. Rozanov (1961). Some limit theorems for random functions, part 2. *Theory of Probability and its Applications 6*, 186–198.

Yao, Q. and H. Tong (1994a). On subset selection in non-parametric stochastic regression. *Statistica Sinica 4*, 51–70.

Yao, Q. and H. Tong (1994b). On prediction and chaos in stochastic systems. *Philosophical Transactions of the Royal Society A348*, 357–369.

Yao, Q. and H. Tong (1994c). Quantifying the influence of initial values on non-linear prediction. *Journal of the Royal Statistical Society, Series B 56*, 701–725.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers,. *Philosophical Transactions of the Royal Society A226*, 267–298.