Equivalence and bifurcations of finite order stochastic processes

Cees Diks and Florian Wagener

Equivalence and bifurcations of finite order stochastic processes [*]

C. Diks and F.O.O. Wagener [†]

25th April 2005

Abstract

This article presents an equivalence notion of finite order stochastic processes. Local dependence measures are defined in terms of ratios of joint and marginal probability densities. The dependence measures are classified topologically using level sets. The corresponding bifurcation theory is illustrated with some simple examples.

1 Introduction

Bifurcation theory has been an extremely successful tool to investigate the qualitative (or structural) properties of deterministic nonlinear systems. But in many practical situations, deterministic models fit the available data only imperfectly, and stochastic models are proposed to describe the behaviour of a given system; the stochastic components can model genuinely random events, but they can also be introduced for quantities of which not enough is known to describe them otherwise.

Motivated by the success of deterministic bifurcation theory, there have been several attempts to develop bifurcation theory for stochastic processes; however, to find a natural replacement of the notion of 'topological equivalence' has been the main problem. For at the base of any bifurcation theory, there is a notion of 'form' or 'shape', formalised as an equivalence relation between systems: two systems are said to be of the same form if they are in the same equivalence class. A meaningful bifurcation theory can only be developed if there are equivalence classes with non-empty interior; note that this presupposes a topology on the space of systems. Elements in the interior of an equivalence class are 'structurally stable': if a system parameter is changed slowly, the system will remain in the equivalence class and the form of the system does not change. All other elements, associated to changes of form, are called 'bifurcating'.

In this article, we quickly review the previously proposed notions of phenomenological and dynamical bifurcation of stochastic processes. We introduce a new equivalence relation, based on the dependence structure of the process. This notion is related, and in some cases equal, to the better known copula density of dependent stochastic variables. We show that our equivalence relation has 'many' structurally stable elements and that it avoids some limitations of older notions. Finally, we illustrate it by giving several applications.

[†]Center for Nonlinear Dynamics in Economics and Finance (CeNDEF), Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, *C.G.H.Diks@uva.nl*, *F.O.O.Wagener@uva.nl*

1.1 Phenomenological bifurcations

Without always stating it explicitly, in this article we shall always assume that the stochastic processes considered are ergodic, and that they have therefore unique invariant probability distributions; we shall moreover assume that these invariant distributions are absolutely continuous with respect to a measure of the Lebesgue class, and that the corresponding probability density is a smooth differentiable function.

The natural first attempt to attain at a classification of stochastic processes is to apply the Morse classification of real valued functions to invariant probability densities p of processes [3, 15]. The corresponding equivalence relation is that of smooth coordinate transformations of domain and range of p, the stable elements being Morse functions with all critical values distinct from each other. For the purposes of this article, we shall call the equivalence relation *P-equivalence*, in analogy with the associated bifurcation notion, which has been called *phenomenological bifurcation* or *P-bifurcation* (see Arnold [2], p. 471-473).

The main problem of this approach, acknowledged in [15], is that the equivalence classes are not invariant under diffeomorphisms of the underlying space. For instance, let $\{X_t\}$ and $\{Y_t\}$ be two processes on $\mathbb{R}^m$ with invariant densities p_X and p_Y, and let φ be an invertible transformation of $\mathbb{R}^m$. If the processes are related by $Y_t = \varphi(X_t)$, then the invariant densities are related as

$$p_X(x) = p_Y(\varphi(x))\,|\det D\varphi(x)|;$$

we see that, in the language of physicists, the function value of the invariant density 'depends on the coordinates'. It is clear that equivalence classes are preserved if only volume-preserving diffeomorphisms are admitted as coordinate changes, for which $|\det D\varphi(x)| = 1$ for all x. Note that these comprise the class of Riemannian isometries proposed in [15].

The restriction of admissible transformations to volume-preserving diffeomorphisms of the domain of the invariant density p is necessary. For if all diffeomorphisms were admissible, then all processes on the real line would be equivalent, since there is always a coordinate transformation such that the invariant density of the transformed process is the uniform density on the unit interval. In the words of Zeeman [15]: *'However, this* [i.e. admitting all diffeomorphisms] *would be exactly the* wrong *thing to do, because it would render the tool* [i.e. P-equivalence] *useless by making everything equivalent (...)'* (the comments in square bracket are our interpolations).

Underlying this difficulty is the fact that the probability density $p(x)$, unlike the measure $p(x)\,\mathrm{d}x$, is not an invariant geometrical object. By consequence, P–equivalence is an inconvenient notion for practical applications: for instance, recording data on linear or logarithmic scale might point to different conclusions.

1.2 Dynamical bifurcations

A second bifurcation notion for stochastic processes has been introduced by Arnold and his co-workers (see [2] for an extensive exposition). We shall try to sketch its main ideas briefly using the simple first-order process $\{X_t\}$ on $\mathbb{R}$ satisfying

$$X_{t+1} = g(X_t) + \varepsilon_t, \tag{1}$$

where the $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables. This process can be considered as a deterministic dynamical system on an infinite dimensional phase space $\Omega \times \mathbb{R}$ as follows. The elements of Ω are the possible realisations $\omega = (\varepsilon_0, \varepsilon_1, \cdots)$ of the process $\{\varepsilon_t\}$. Introducing the projection $\pi(\omega) = \varepsilon_0$ and the shift $\sigma(\omega) = (\varepsilon_1, \varepsilon_2, \cdots)$, we have for instance that $\varepsilon_t = \pi \circ \sigma^t(\omega)$. Define now the map Φ on $\Omega \times \mathbb{R}$ by

$$\Phi(\omega, x) = (\varphi_1(\omega), \varphi_2(\omega, x)) = (\sigma(\omega), g(x) + \pi(\omega)).$$

Note that this is a deterministic system; the stochastics are 'hidden' in the fact that the initial condition $\omega \in \Omega$ is unknown. The realisations X_t of the process (1) are the values of the second component of $\Phi^t(\omega, x)$.

Note that Φ is a skew system: the shift dynamics φ_1 in the space Ω are driving the dynamics φ_2 in $\mathbb{R}$. For Φ, a *random fixed point* is defined as a map $\xi : \Omega \to \mathbb{R}$, which satisfies the invariance condition

$$\varphi_2(\omega, \xi(\omega)) = \xi(\varphi_1(\omega))$$

for all (or almost all) ω. Stability is now be defined in the usual way: a random fixed point ξ is stable if all nearby orbits converge to ξ. A random, or, following the terminology in [2], *dynamical bifurcation* or *D-bifurcation* of the process occurs for instance if a random fixed point loses stability.

At this point, a drawback of the notion of dynamical bifurcation becomes apparent: to determine stability of a random fixed point, two orbits of Φ with identical noise realisations have to be compared. This seems to make it rather difficult to apply the notion of D-bifurcation to practical problems (see however [4] and related literature). We therefore leave aside this theory, and try rather to improve on the notion of P-equivalence.

1.3 Local dependency structure

By considering stochastic analogues of concepts used in catastrophe theory, Hartelman *et al.* [12] arrived at a classification for stochastic differential equations of diffusion type on the real line that is invariant under invertible transformations. At the basis of this classification are level crossing statistics and derived quantities, which are invariant under monotonous transformations of the underlasying space. Although level crossing statistics can also be used for discrete time systems, the corresponding classification would be rather restrictive. The reason is that discrete time dynamical systems are 'essentially richer' than discretely sampled continuous time diffusions, because finite time transition densities induced by diffusions only represent a subclass of transition densities for discrete time dynamical systems.

Instead of level crossing statistics, we base our proposed equivalence relation on another function that can be associated to stochastic processes of finite order. For the purposes of this introduction, we explain the main ideas in the case of a first order process $\{X_t\}$ on the real line; general definitions will be given below. Assume therefore that the process $\{X_t\}$ is generated by

$$X_{t+1} = g(X_t) + \varepsilon_t, \tag{2}$$

where $\{\varepsilon_t\}$ is a sequence of independent, identically distributed (IID) innovations. Recall that equation (2) imply a transformation law for probability densities. Assume that X_t is distributed with marginal probability density p^t, that is $P(a \leq X_t < b) = \int_a^b p^t(x)\,dx$, and that ε_t has density φ. Then the probability density p^{t+1} of X_{t+1} is given by

$$p^{t+1}(x_{t+1}) = \int \tau(x_{t+1}|x_t)p^t(x_t)\,dx_t,$$

where $\tau(x|y) = \varphi(x - g(y))$ is the transition probability density of the process.

Consider in particular the strictly stationary process $\{X_t\}$ where X_1, and hence every X_t, is distributed according to the invariant probability density $p^1 = p$. Denote the invariant joint probability density of (X_t, X_{t+1}) by $p^{t,t+1}$; this joint density does not depend on t and it is therefore equal to $p^{1,2}$. Moreover, the measure $p^{1,2}(x_1, x_2)\,dx_1\,dx_2$ is absolutely continuous with respect to $p(x_1)p(x_2)\,dx_1\,dx_2$. By the Radon-Nikodym theorem, the following Jacobian exists:

$$f(x_1, x_2) \stackrel{\text{def}}{=} \frac{p^{1,2}(x_1, x_2)\,dx_1\,dx_2}{p(x_1)p(x_2)\,dx_1\,dx_2} = \frac{p^{1,2}(x_1, x_2)}{p(x_1)p(x_2)} = \frac{\tau(x_2|x_1)}{p(x_2)};$$

3

The function f is called the *dependency ratio* of the process. Note that f is identically 1 if X_t and X_{t+1} are independent; the difference $|f(x_1,x_2) - 1|$ can therefore be seen as a measure of the local dependency structure of the process. Moreover, f contains all essential information regarding the dependence structure of the process: for if coordinates are chosen such that $p(x) = 1$ for $x \in [0,1]$ and 0 otherwise, then $f(x_1,x_2) = p^{1,2}(x_1,x_2) = \tau(x_2|x_1)$ determines the entire process. By construction, dependency ratios are invariant — as geometrical objects — under any invertible transformation the underlying space of the process under consideration. Our notion of equivalent processes will be based on these ratios; invariance of the equivalence classes under nonlinear invertible transformations is then obtained automatically.

Several other local dependence measures have recently been described in the statistical literature (see e.g. [7], [8], and [10]). The dependence measures in this literature are various localised versions of the Pearson correlation coefficient, and as such are motivated entirely differently than our dependency ratio. In particular they do not share the invariance property with our dependency ratio.

The definition of our equivalence relation is a bit involved. We therefore postpone this definition to section 3. First, in section 2, we give the definition of dependency ratio for more general processes, and we show how this quantity is connected with other quantities like copula density, mutual information and entropy of a stochastic process. In section 3, the definition of our equivalence relation is given, together with topological properties of the associated equivalence classes. Applications are given in section 4.

2 Dependency ratios, copulas and information theory

Our attention will be restricted to stationary discrete time processes of finite order that are generated by equations like

$$X_t = g(X_{t-n}, \cdots, X_{t-1}, \varepsilon_t).$$

Here the variables X_t take values in some n–dimensional orientable manifold M, and the ε_t are identically and independently distributed random variables. In most cases M will be equal to $\mathbb{R}^n$. We wish however to bring out the dependency of the process on the variables chosen; in order to achieve this, a coordinate–independent formulation is chosen. Considering the problem on a manifold will come at little extra cost.

2.1 Basic definitions

Recall that any stochastic process $\{X_t\}$ is given by the joint probabilities

$$P^{t_1,\cdots,t_\ell}(A_1 \times \cdots \times A_\ell) = P\{X_{t_1} \in A_1, \cdots, X_{t_\ell} \in A_\ell\},$$

which are defined for all ℓ–tuples $(t_1,\cdots,t_\ell) \in \mathbb{Z}^\ell$. A stochastic process is called *strictly stationary* if its finite dimensional joint probabilities are time invariant

$$P^{t_1,\cdots,t_\ell} = P^{t_1+h,\cdots,t_\ell+h},$$

for all $(t_1,\cdots,t_\ell,h)$. Two stochastic variables X and Y are said to be equivalent in distribution, written as $X \sim Y$, if $P(X \in A) = P(Y \in A)$ for every P-measurable set A. A strictly stationary process $\{X_t\}$ is said to be of order n if, for all $k > n$, the conditional distribution of X_t given $X_{t-k},\ldots,X_{t-1}$ is a function of $(X_{t-n},\ldots,X_{t-1})$ only:

$$X_t|X_{t-k},\ldots,X_{t-1} \sim X_t|X_{t-n},\ldots,X_{t-1}. \tag{3}$$

A measure on a manifold M is of *Lebesgue class*, if at every point of the manifold it is absolutely continuous with respect to the Lebesgue measure in some (and hence any) coordinate chart. If M is orientable, there is

a *volume form* on M: a differentiable n-form that is everywhere non-zero. It is clear that any volume form is of Lebesgue class. We assume that M is orientable, and we fix a volume form on M, denoting it by $\mathrm{d}x$. It will be assumed that the processes we are dealing with have all joint probabilities $P^{t_1,\cdots,t_\ell}$ of Lebesgue class, with twice differentiable probability densities $p^{t_1,\cdots,t_\ell}$.

The transition probability density τ of a process of order n is given by

$$\tau(x_{n+1}|x_1,\cdots,x_n) = \frac{p^{1,\cdots,n+1}(x_1,\cdots,x_{n+1})}{p^{1,\cdots,n}(x_1,\cdots,x_n)}. \tag{4}$$

Note that from the order property, it follows that for $k \geq n$

$$\tau(x_t|x_{t-k},\cdots,x_{t-1}) = \tau(x_t|x_{t-n},\cdots,x_{t-1}).$$

Given any initial probability density of $X_1,\cdots,X_n$, we can determine all joint probabilities by first determining

$$p^{1,\cdots,n+1}(x_1,\cdots,x_{n+1}) = \tau(x_{n+1}|x_1,\cdots,x_n)p^{1,\cdots,n}(x_1,\cdots,x_n),$$
$$p^{1,\cdots,n+2}(x_1,\cdots,x_{n+2}) = \tau(x_{n+2}|x_2,\cdots,x_{n+1})\tau(x_{n+1}|x_1,\cdots,x_n)p^{1,\cdots,n}(x_1,\cdots,x_n),$$
$$\vdots$$

and then integrating out unwanted variables. In particular, the distribution of $X_2,\cdots,X_{n+1}$ is obtained by

$$p^{2,\cdots,n+1}(x_2,\cdots,x_{n+1}) = \int \tau(x_{n+1}|x_1,\cdots,x_n)p^{1,\cdots,n}(x_1,\cdots,x_n)\,\mathrm{d}x_1.$$

The central notion of dependency ratio of a stochastic process is given in terms of joint probability densities.

Definition. *The n'th order* **dependency ratio** *f of a strictly stationary stochastic process $\{X_t\}$ is given by*

$$f(x_1,\cdots,x_{n+1}) = \frac{p^{1,\cdots,n+1}(x_1,\cdots,x_{n+1})}{p^{1,\cdots,n}(x_1,\cdots,x_n)p^1(x_{n+1})} = \frac{\tau(x_{n+1}|x_1,\cdots,x_n)}{p(x_{n+1})}.$$

Note that f is the Jacobian of the measure $p^{1,\cdots,n+1}(x_1,\cdots,x_{n+1})\,\mathrm{d}x_1\cdots\mathrm{d}x_{n+1}$ with respect to the measure $p^{1,\cdots,n}(x_1,\cdots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n \cdot p^1(x_{n+1})\,\mathrm{d}x_{n+1}$. In particular, f is chart-independent.

Proposition 1. *If for two processes of order n the dependency ratios of order n are equal, then for every $m \geq n$, the dependency ratios of order m are equal as well.*

Proof.
We consider two processes $\{X_t\}$ and $\{Y_t\}$ of order n. It suffices to show that order-m equality of the dependency ratio for any $m \geq n$ implies, and is implied by, order-n equality. For $m \geq n$, we may write

$$\begin{aligned} p^{1,\cdots,m+1}(x_1,\ldots,x_{m+1}) &= p^{1,\cdots,m}(x_1,\ldots,x_m)\tau(x_{m+1}\,|\,x_1,\ldots,x_m) \\ &= p^{1,\cdots,m}(x_1,\ldots,x_m)\tau(x_{m+1}\,|\,x_{m-n+1},\ldots,x_m). \end{aligned}$$

The dependency ratio of $\{X_t\}$, for $m \geq n$, can be written as a function of $n+1$ variables only:

$$\frac{p^{1,\cdots,m+1}(x_1,\ldots,x_{m+1})}{p^{1,\cdots,m}(x_1,\ldots,x_m)p(x_{m+1})} = \frac{\tau(x_{m+1}\,|\,x_{m-n+1},\ldots,x_m)}{p(x_{m+1})} = \frac{p^{1,\cdots,n+1}(x_{m-n+1},\ldots,x_{m+1})}{p^{1,\cdots,n}(x_{m-n+1},\ldots,x_m)p(x_{m+1})},$$

which is nothing but the dependency ratio of order n in terms of the last $n+1$ variables of the vector $(x_1,\ldots,x_{m+1})$. The dependency ratio of $\{Y_t\}$ can be reduced similarly. Clearly, order-m equality for $m \geq n$ implies, and is implied by, order-n equality. ∎

2.2 Relation with copulas

For a strictly stationary real-valued time series $\{X_t\}$ with continuous joint cumulative distribution functions (CDF) $F_{t_1,\dots,t_\ell}(x_{t_1}, \cdots, x_{t_\ell})$ and marginal CDF $F(x)$, the copula associated with a delay vector $(X_{t-n+1}, \dots, X_t)$ is the quantity

$$C_n(u_1,\dots,u_n) = F_{1,\dots,n+1}(F^{-1}(u_1),\dots,F^{-1}(u_n)),$$

where $u_j \in [0,1]$ for $j = 1, \cdots, n+1$. The correspondeing copula density function is

$$c_n(u_1,\dots,u_n) = \frac{\partial^{n+1} C_n(u_1,\dots u_n)}{\partial u_1 \cdots \partial u_n} = \frac{p^{1,\dots,n+1}(F^{-1}(u_1),\dots,F^{-1}(u_n))}{p(F^{-1}(u_1))\cdots p(F^{-1}(u_n))},$$

or

$$c_n(F(x_1),\dots,F(x_n)) = \frac{p^{1,\dots,n}(x_1,\dots,x_n)}{p(x_1)\cdots p(x_n)}.$$

In the case of a real valued first order time series, our definitions imply that the dependency ratio $f(x_1, x_{n+1})$ is equal to the copula density function c_2 evaluated in the 'standard' coordinates $(u_1,u_2) = (F(x_1),F(x_2))$, which are equivariant under monotonously increasing transformations of X.

Although the $n+1$-dimensional copula $c(u_1,\dots,u_{n+1})$ characterises the dependence structure within $n+1$ consecutive values $(X_{t-n}, \dots X_t)$, it does not take into account the ordering of the observations in time. In time series analysis one is often interested in the question of how X_t depends on $(X_{t-n},\dots,X_{t-1})$. It is then more natural to take into account the ordering of the observations in time and to focus on the conditional distribution of X_t given past values of X_t. In this way, we are led to the definition of dependency ratio given above. This allows, in principle, to distinguish between time reversals in processes of order higher than one. For orders larger than one the dependency ratio can be expressed as:

$$f(x_1,\dots,x_{n+1}) = \frac{c_{n+1}(F(x_1),\dots,F(x_{n+1}))}{c_n(F(x_1),\dots,F(x_n))}.$$

2.3 Relation with information theory

Information theoretic dependence measures can often be expressed in terms of copulas. Well-known examples are mutual information and the redundancy, which both are special instances of Kullback-Leibler divergences (relative entropies).

The *Kullback-Leibler divergence* between two probability distributions with probability densities $p(x)$ and $q(x)$ is defined as

$$D(p,q) = \int p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx$$

The *mutual information* $I_{(X_1;X_2)}$ between two random variables X_1 and X_2, is a dependence measure defined as the Kullback-Leibler divergence between their joint probability density function $p^{1,2}(x_1,x_2)$ and the product of their marginals $p(x_1)p(x_2)$ (we consider the case with identical marginals, appropriate for stationary time series). Since $p^{1,2}(x_1,x_2)$ denotes the joint probability density function of (X_1,X_2), the integral over $p(x) = p(x_1,x_2)$ can be concisely expressed as an expectation, that is,

$$I_{(X_1;X_2)} = \iint p^{1,2}(x_1,x_2) \ln\left(\frac{p^{1,2}(x_1,x_2)}{p(x_1)p(x_2)}\right) dx_1\,dx_2 = E\left[\ln\left(\frac{p^{1,2}(X_1,X_2)}{p(X_1)p(X_2)}\right)\right],$$

which is non-negative and equals zero if and only if X_1 and X_2 are independent. This can be seen as follows: the expected value of the random variable $Z = p(X_1)p(X_2)/p^{1,2}(X_1, X_2)$ is

$$E[Z] = \iint p^{1,2}(x_1, x_2) \frac{p(x_1)p(x_2)}{p^{1,2}(x_1, x_2)} \, dx_1 \, dx_2 = 1.$$

By convexity of the function $\ln(z^{-1}) = -\ln(z)$, Jensen's inequality gives

$$I_{(X_1; X_2)} = E[\ln(Z^{-1})] \geq -\ln(E[Z]) = 0,$$

and equality holds if and only if $Z = 1$ almost surely.

The generalisation of the mutual information to the multivariate case is known as the *redundancy*:

$$R_{(X_1; \ldots; X_{n+1})} = E\left[\ln\left(\frac{p^{1,\ldots,n+1}(X_1, \cdots, X_{n+1})}{p(X_1) \cdot \ldots \cdot p(X_{n+1})}\right)\right],$$

which is zero if and only if $(X_1, \ldots, X_{n+1})$ are jointly independent, and positive otherwise. In analogy with the above discussion on copulas, a generalisation which is more suitable within a time series context is the *entropy*, given by

$$H_{(X_{n+1}; X_1, \ldots, X_n)} = E\left[\ln\left(\frac{p^{1,\ldots,n+1}(X_1, \cdots, X_{n+1})}{p^{1,\ldots,n}(X_1, \ldots X_n)p(X_{n+1})}\right)\right] = E\left[\ln f(x_1, \ldots, x_{n+1})\right].$$

3 Equivalence notions

In this section we introduce and motivate our notion of equivalence of stochastic processes and we give some of its fundamental properties.

3.1 Structural stability and bifurcations

We recall briefly the fundamentals of bifurcation theory. The two main ingredients of any such theory are a topological space X and an equivalence relation between elements of X. An element f of X is *structurally stable* if there is a neighbourhood $N(f)$ such that all elements g in that neighbourhood are equivalent to f; that is $g \sim f$ for all $g \in N(f)$. Intuitively speaking, a structurally stable element f can be 'perturbed' slightly without being pushed out of its equivalence class. Such an element is sometimes called 'persistent'. Clearly, the equivalence class of any structurally stable element is an open set. A structurally stable equivalence class can be thought of as defining a set of elements of the same 'shape' or 'form' (see [14]): form remains 'stable' if perturbed slightly.

All elements of X that are not structurally stable are called *bifurcating*. This notion is usually familiar from the context of parametrised families: if λ is some q-dimensional parameter, and $\lambda \mapsto f_\lambda$ a family of elements of X, then $\lambda = \lambda_0$ is a bifurcating parameter value of the family if f_{λ_0} is not structurally stable; it might be said that at bifurcating parameter values the 'form' of f_λ changes. Since the set of structurally stable elements is open, the set of bifurcating elements, and therefore also the set of bifurcating parameter values in a parametrised family, is closed.

An equivalence relation will give rise to a useful bifurcation theory on X only if there at all exist structurally stable elements. The optimal situation is attained if the set of structurally stable elements, while not consisting of a single equivalence class, is 'topologically big', since then we will be able to associate to 'most' elements a form. In a topological space, a set is 'big' if it is open and dense, or if it is at least a countable intersection of open and dense sets (a so-called 'generic' or 'second category' set, see [11]).

3.2 Strong equivalence

A natural requirement to impose on an equivalence relation of stochastic processes on a manifold M is that processes which only differ by a diffeomorphism of M, that is, which are the 'same' up to a coordinate change, fall in the same equivalence class. Let for instance $\{X_t\}$, $\{Y_t\}$ denote two first order processes on M. We will certainly want to call two processes equivalent if there is a diffeomorphism $\varphi : M \to M$ such that the variables (Y_t, Y_{t-1}) and $(\varphi(X_t), \varphi(X_{t-1}))$ are identically distributed. If this is the case, we call $\{X_t\}$ and $\{Y_t\}$ *strongly equivalent*. Let $f_X(x_1, x_2)$ and $f_Y(x_1, x_2)$ denote the dependency ratios of $\{X_t\}$ and $\{Y_t\}$ respectively; if the processes are strongly equivalent, it follows from the invariance of the dependency ratios under diffeomorphisms that

$$f_X(x_1, x_2) = f_Y(\varphi(x_1), \varphi(x_2)) \quad \text{for all} \ (x_1, x_2) \in M \times M. \tag{5}$$

If we would take strong equivalence as our equivalence relation, in general we would obtain an uncountable infinity of equivalence classes, and no class would be a neighbourhood to any of its points, that is, no process would be structurally stable. To see this in a simple example, assume that f_X and f_Y are two dependency ratios defined on the square $(-1, 1) \times (-1, 1) \subset \mathbb{R}^2$, and that they are given as

$$f_X(x_1, x_2) = \frac{2 - \mu}{3} + x_1^2 + \mu x_2^2, \qquad f_Y(x_1, x_2) = \frac{2 - \nu}{3} + x_1^2 + \nu x_2^2.$$

Taking the invariant density in both cases to be $p(x) = \frac{1}{2} I_{[-1,1]}(x)$, where $I_A(x)$ denotes the indicator function, we have specified two stochastic processes. The point $(0, 0)$ is the only non-degenerate critical point for both f_X and f_Y; therefore, if f_X and f_Y are strongly equivalent, we should have that $\Phi(x_1, x_2) = (\varphi(x_1), \varphi(x_2))$ satisfies $\Phi(0, 0) = (0, 0)$. But there is no real-valued smooth diffeomorphism φ such that (5) holds simultaneously with $\varphi(0) = 0$, for the values of f_X and f_Y at $(0, 0)$ are different if $\mu \neq \nu$. We see that every value of μ defines a different equivalence class.

3.3 Topology of dependency ratios

In section 2 we have defined the dependency ratio f of an n'th order process $\{X_t\}$ by

$$f(x_1, \cdots, x_{n+1}) = \frac{p^{1, \cdots, n+1}(x_1, \cdots, x_{n+1})}{p^{1, \cdots, n}(x_1, \cdots, x_n) p(x_{n+1})};$$

this quantity is invariant under coordinate transformations, and it is therefore a characteristic of the stochastic process.

As the space of dependency ratios is infinite dimensional, this characteristic itself is too fine-grained to be useful to classify such processes. In the previous section we have seen that one way of extracting 'coarser' information from a dependency ratio f is to determine the expected value of some monotone transformation of f. But defining equivalence of two ratios by equality of such expected values would not lead to structurally stable elements, for a very small perturbation of the process might already change the expected value.

Using however topological information of the dependency ratios turns out to give characteristics that are sufficiently coarse. To give a very simple example: we clearly want to call two functions defined on the same domain to be of different shape if they have a different number of nondegenerate critical points. The number of such points is a numerical characteristic of the 'shape' of a function, and it in fact defines an equivalence class. Moreover, if we choose a suitable topology on the set of functions, we find that the equivalence classes are open sets, and that its members are structurally stable. We can specify different

equivalence relations by taking into account more detailed information; for instance, we can classify the critical points according to their topological type.

In any case, we need a function topology on the set of dependency ratios; we choose the C^2-topology, which is the 'coarsest' topology for which the number of nondegenerate critical points defines open equivalence classes. Recall that in the C^2 topology an ε-neighbourhood $N_\varepsilon(f)$ of a function $f : M \to \mathbb{R}$ defined on a compact manifold M consists of all functions g such that, with respect to a fixed Riemannian metric and the induced norm $|\cdot|_x$ on the tangent spaces $T_x M$ we have

$$|f(x) - g(x)|_x, |Df(x) - Dg(x)|_x, |D^2 f(x) - D^2 g(x)|_x < \varepsilon.$$

If M is a noncompact manifold, the constants $\varepsilon > 0$ are replaced by positive functions $\varepsilon(x) > 0$ on M in the above definition; in this context the topology obtained is called the 'strong' C^2-topology (see e.g. [6]).

As argued above, specifying an equivalence notion on the space of dependency ratios of stochastic processes induces a notion of stochastic bifurcation. In the following we shall sometimes use the words 'process' and 'dependency ratio' indiscriminately; in particular, a 'structurally stable process' will denote a stochastic process whose dependency ratio is a structurally stable element under the equivalence relation under consideration. A first rough formulation of our definition would be the following: we propose to call two dependency ratios of stochastic processes equivalent, if every non-degenerate critical point of a certain type of the first dependency ratio can be mapped to a critical point of the same type of the second dependency ratio by a transformation of $M \times M$ induced by a diffeomorphism of M. In the next section, we shall make this more precise.

3.4   Ratio equivalence on compact manifolds

Let M be an m-dimensional orientable compact manifold and M_{n+1} the $(n+1)$-fold Cartesian product $M \times \cdots \times M$; denote by $\pi_\ell : M_{n+1} \to M$ the projection on the ℓ'th component

$$\pi_\ell(x_1, \cdots, x_{n+1}) = x_\ell.$$

An n'th order dependency ratio is a real valued function defined on M_{n+1}.

Recall the following definitions (see e.g. [5], subsections 10.2 and 10.4, p. 79 and p. 86 respectively). If $f : U \to \mathbb{R}$ is a twice continuously differentiable function defined on an open set $U \subset \mathbb{R}^n$, a point $x \in U$ is a *critical point* of f if the derivative of f vanishes at x: $Df(x) = 0$. The value $f(x)$ of f at a critical point x is called the *critical value* of f at x. A critical point x is *non-degenerate* if the Hessian matrix $Hf(x)$ corresponding to the second derivative $D^2 f(x)$ of f at x is invertible. The number of negative eigenvalues of this matrix is called the *(Morse) index* of the critical point. Clearly, the notions of critical point, critical value, index and non-degeneracy carry over to functions defined on manifolds.

Our definition is based on the critical points of a given dependency ratio; this makes it necessary to restrict attention to twice differentiable ratios only, and to consider the C^2 function topology introduced above on the space of these ratios. For in any C^0-neighbourhood of a given function, there are other functions to be found with a different number of critical points; and the same is true for any C^1-neighbourhood of a function that has itself at least one critical point.

Definition.   *If M is a manifold, a twice differentiable dependency ratio $f : M_{n+1} \to (0, \infty)$ is called **regular** if all its critical points are non-degenerate, if no two critical values are equal and if no two critical points have the same image under any projection π_ℓ, for $\ell \in \{1, \cdots, n+1\}$.*

Since the manifolds M and M_{n+1} are compact, a regular dependency ratio has only finitely many critical points $\{\xi_1, \cdots, \xi_k\}$; we assume that these are ordered such that the corresponding critical values $v_i = f(\xi_i)$ are in ascending order, that is, $v_i < v_j$ if $i < j$. We associate to the critical point ξ_i its index t_i (see subsection 3.2). Note that $0 \leq t_i \leq m(n+1)$. In this way we obtain the *index sequence* $t(f) = (t_1, \cdots, t_k)$ of a regular dependency ratio f.

Definition. *Two order n processes defined on a compact manifold M, with dependency ratios $f, g : M_{n+1} \to (0, \infty)$, are called* **ratio equivalent**, *if either both f and g are non-regular, or if f and g are both regular and*

   1. *their index sequences are equal;*

   2. *there is a diffeomorphism $\varphi : M \to M$, homotopic to the identity mapping on M, such that the map $\Phi : M_{n+1} \to M_{n+1}$ defined as*

$$\Phi(x_1, \cdots, x_{n+1}) = \Big(\varphi(x_1), \cdots, \varphi(x_{n+1}) \Big) \tag{6}$$

   maps the i'th critical point of f to the i'th critical point of g.

 It follows from the first point that the number of critical points of f and g is equal as well.
        The following two propositions tell us that this definition has good properties: we can characterise all structurally stable processes, and these form an open and dense set in the space of all stochastic processes.

Proposition 2. *On a compact manifold M, a dependency ratio is structurally stable under ratio equivalence if and only if it is regular.*

Proposition 3. *On a compact manifold M, the set of regular dependency ratios is everywhere dense.*

The proofs of these propositions can be found in appendix A.

3.5 Ratio equivalence for non-compact manifolds

Though the results for compact manifolds are already useful in themselves, in practice most stochastic processes are defined on the non-compact manifold $\mathbb{R}^m$. The direct generalisation of the notion of ratio equivalence is given in the following definition.

Definition. *Two processes on a manifold M are* **weakly ratio equivalent**, *if they are ratio equivalent on each compact set that is the closure of a bounded submanifold of M_{n+1}.*

As the following example shows, this notion is unfortunately too weak for our purposes.

Example. Consider two first order processes $\{X_t\}$ and $\{Y_t\}$ on the interval $(-1, 1)$ with invariant densities $p(x) = \frac{1}{2}I_{[-1,1]}(x)$ and dependency ratios

$$f_X(x_1, x_2) = 1 - \frac{1}{2}x_1 x_2 + \frac{1}{4}x_1^3, \qquad \text{and} \qquad f_Y(x_1, x_2) = 1 + \frac{1}{2}x_1 x_2 - \frac{1}{4}x_1^3.$$

Both ratios have a unique critical point of index 1 at the origin, and hence they are ratio equivalent on compact sets. But if we consider the values of f_X and f_Y along the curve $\gamma(t) = (t, t)$ as $t \downarrow -1$, we note that $f_X \circ \gamma(t)$ approaches the infimum of f_X on $(-1, 1) \times (-1, 1)$, while $f_Y \circ \gamma(t)$ approaches the supremum of f_Y on the same space. That means in particular that if X_t is close to -1, the probability is very low that X_{t+1} is close to -1 as well, whereas if Y_t is close to -1, the probability is rather high that Y_{t+1} is close to -1 as well. Weak ratio equivalence is not sufficiently fine to distinguish between these processes. ∎

For our second generalisation, we restrict ourselves to processes defined on a subclass of manifolds which we call manifolds of 'constant type'; these are defined as follows.

Let M be an orientable m-dimensional manifold. If there exists a family $\{M_t\}$ of bounded open submanifolds-with-boundary of M depending on a real parameter $t > T$, for which $M_t \subset M_{t'}$ if $t < t'$, $\bigcup_t M_t = M$ and for which ∂M_t is a smooth boundary for every $t > T$, then we call $\{M_t\}$ an *exhaustion* of M. Moreover, we say that an exhaustion is of *constant type* if there is a constant $T > 0$ such that ∂M_t and $\partial M_{t'}$ are diffeomorphic for all $t, t' > T$. A manifold M that admits an exhaustion of constant type will itself be called to be of *constant type*. For instance, the plane is a manifold of constant type with exhaustion $\mathbb{R}^2_t = \{x \in \mathbb{R}^2 : \|x\|^2 < t\}$, whereas the plane with all points with integer coordinates removed is not a manifold of constant type.

A convenient way to define an exhaustion of M is to take a differentiable real valued function J on M such that for a fixed point $x_0 \in M$ and a metric d on M we have that $J(x) \to \infty$ as $d(x_0, x) \to \infty$, and set $M_t = \{x \in M : J(x) < t\}$. If all values of J larger than T are regular, then it follows from Morse theory that ∂M_t is diffeomorphic to $\partial M_{t'}$ for $t, t' > T$, and consequently that $\{M_t\}$ is an exhaustion of M of constant type.

Consider the set

$$(M_t)_{n+1} = M_t \times M_t \times \cdots \times M_t \subset M_{n+1}.$$

This set can be decomposed into 2^{n+1} component manifolds $\{C_t^j\}_{j=1}^N$; each component is a Cartesian products of k factors ∂M_t and $n + 1 - k$ factors M_t, for $k = 0, \cdots, n+1$. We order these components such that a product of more factors M_t precedes a component with less, and for components with equal numbers of factors M_t, the component with more factors M_t in the first ℓ positions precedes a component with less, with ℓ taking the values $1, 2, \cdots, n+1$ consecutively. That is,

$$\partial M_t \times M_t \times M_t \qquad \text{precedes} \qquad M_t \times \partial M_t \times \partial M_t,$$

and

$$\partial M_t \times M_t \times \partial M_t \qquad \text{precedes} \qquad \partial M_t \times \partial M_t \times M_t.$$

In the important special case that M is one-dimensional, the component $(\partial M_t)_{n+1}$ consists of a finite number of points. By definition, we consider these point as non-degenerate critical points, associating the index 0 to them by default.

In the following three definitions, M is a manifold of constant type with exhaustion $\{M_t\}$ and corresponding ordered decomposition $\{C_t^j\}$ of $(M_t)_{n+1}$; by $|\cdot|_x$ we denote a norm induced from a fixed Riemannian metric as in subsection 3.3. Moreover, the restriction of f to C_t^j is denoted by f_t^j with the index j always ranging from 1 to 2^{n+1}.

Definition. *A dependency ratio f on M_{n+1} is* **well-behaved at infinity** *if there are constants $c_t, T > 0$ such that for every $t > T$*

1. *for every j there is a compact set $K_t^j \subset C_t^j$ such that $|Df_t^j(x)|_x > c_t$ if $x \in C_t^j \backslash K_t^j$, unless C_t^j is 0-dimensional, and*

2. *f_t^j is weakly ratio equivalent to $f_{t'}^j$ for all $t, t' > T$.*

Definition. *A dependency ratio f on M_{n+1} is* **well-behaved** *if f is well-behaved at infinity and f_t^j is regular on C_t^j for every j and every $t > T$.*

Definition.   *If M is a manifold of constant type, two well-behaved dependency ratios f and g are called* **ratio equivalent**, *if there is a value of t such that f_t^j and g_t^j are weakly ratio equivalent for every j.*

Note that if f and g are weakly ratio equivalent on each component C_t^j for a single value $t > T$, they are in fact equivalent for all such values, since $f_t^j \sim f_{t'}^j$ for all $t, t' > T$.

Example.   With this definition, we can distinguish between the two ratios f_X and f_Y introduced at the end of the previous subsection. Set $a_t = t/(t+1)$, and consider the exhaustion $I_t = (-a_t, a_t)$ of $(-1, 1)$. Note that $\partial(I_t \times I_t)$ can be decomposed into

$$C_t^1 = (-a_t, a_t) \times \{-a_t, a_t\}, \quad C_t^2 = \{-a_t, a_t\} \times (-a_t, a_t), \quad C_t^3 = \{-a_t, a_t\} \times \{-a_t, a_t\}.$$

Restricted to C_t^1 and C_t^2, neither f_X nor f_Y have any critical points. The set C_t^3 consists of four isolated critical points, which are critical by definition. The maximum of f_X restricted to C_t^3 is assumed in the points $(a_t, -a_t)$, whereas f_Y takes its minimum there. Since the only diffeomorphism of C_t^3 homotopic to the identity is the identity itself, corresponding critical points of f_X and f_Y cannot be mapped onto each other. ∎

The following propositions describe the topological properties of ratio equivalence. The results are weaker than in the compact case, as was to be expected; we obtain that well-behaved processes are stable elements of ratio equivalence. However, restricted to the space of processes that are well-behaved at infinity, the well-behaved processes form again an open and dense set.

The next proposition is a corollary to propositions 2 and 3.

Proposition 4.   *On a manifold M of constant type, a well-behaved dependency ratio is stable with respect to the strong topology under ratio equivalence.*

The proof of this proposition is given in appendix A.

4   Examples

4.1   Stochastic dynamics on the circle

As an illustration of a stochastic process on a compact manifold, we consider the stochastic dynamical system on the unit circle $M = S^1$ defined by

$$X_{t+1} = X_t + a\sin(X_t) + 0.25\sin^2(X_t) + 0.25 + \varepsilon_{t+1} \quad \mathrm{mod}\ 2\pi, \tag{7}$$

with $\{\varepsilon_t\}$ IID and $N(0, \sigma^2)$ distributed. The state variable is taken modulo 2π; we represent states by points on the interval $[-\pi, \pi)$. For the above system we fix σ at the value 0.7 and consider qualitative changes in the stochastic dynamics as a varies. The term $0.25(\sin^2(x_{t-1}) + 1)$ is added to break the $x \mapsto -x$ symmetry of the dynamics. In the symmetric case some particular additional properties arise which will be discussed in the next subsection.

Figure 1 shows a contour plot for the dependency ratio f for values of a decreasing from -0.85 to -0.95. For $a = -0.85$, the contour plot shows two extrema, a maximum and a minimum, together with two saddle points. These are the minimal number of critical points of each type that can be attained for a non-degenerate function f on the torus $M_2 = S^1 \times S^1$. As the bifurcation parameter a decreases, the system shows a stochastic bifurcation at which f develops a new local extremum, together with a new saddle point.
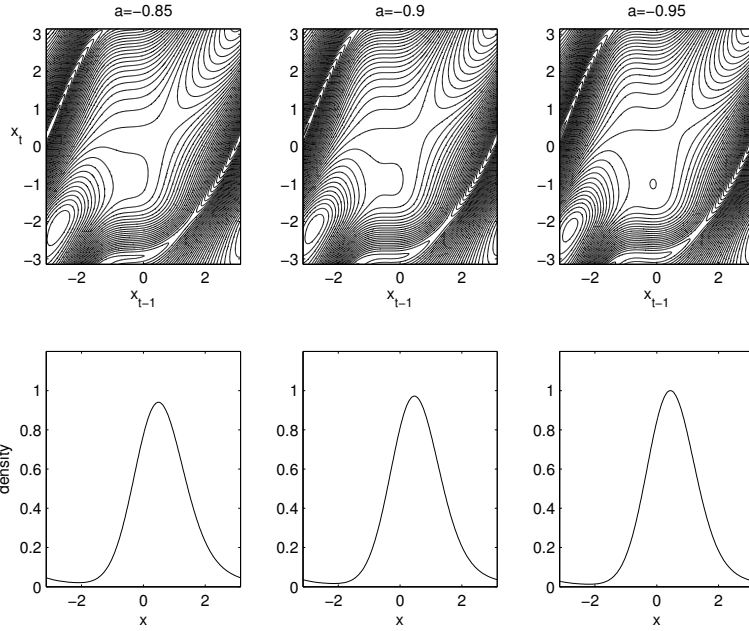
Figure 1: Level sets for the map $X_t = X_{t-1} + a\sin(X_{t-1}) + 0.25\sin^2(x_{t-1}) + 0.25 + \varepsilon_t$ for decreasing values of a (top panels). The lower panels show the invariant probability density of X_t.

4.2 Antisymmetric dynamics

In applications dynamical systems are often symmetric. Though we leave the theoretical development of our equivalence relation for symmetric systems for later work, we want to make some remarks about this situation. We restrict to processes that are the sum of an antisymmetric deterministic part and a unimodal symmetric noise term (e.g. Gaussian). Surprisingly, it turns out that for these systems the 'ratio bifurcation' coincides with a P-bifurcation.

Consider the process

$$X_{t+1} = g(X_t) + \varepsilon_t,$$

where $g(x)$ is a smooth odd function, that is, for which $g(-s) = -g(s)$. The ε_t are independent and identically distributed according to a symmetric unimodal distribution. It can be readily checked that the deterministic dynamics is equivariant under the transformation $x \mapsto -x$. Because this transformation affects both X_{t-1} and X_t, the effect on the joint variable (X_{t-1}, X_t) is a point reflection in the origin.

For the conditional density of X_t given $X_{t-1} = x$, one may write

$$p_{X_{t+1}|X_t}(x_2|x_1) = \tau(x_2|x_1) = \frac{p^{1,2}(x_1, x_2)}{p(x_1)} = \frac{1}{\sigma} h\left(\frac{x_2 - g(x_1)}{\sigma}\right)$$

where $h(\cdot)$ is the probability density function of ε_t; we have that $h(s) = h(-s)$ and that h has a unique local maximum at $s = 0$. The map $g(\cdot)$ as well as the probability density function $h(\cdot)$ are assumed to be twice continuously differentiable.

Moreover we assume that the process has an invariant density $p(x)$, which is unique and twice continuously differentiable. These conditions on $p(x)$ can easily be met by imposing some additional requirements on g and h. For instance, a sufficient condition is that for each x_0, the support of $\frac{1}{\sigma} h\left(\frac{x_1 - g(x_0)}{\sigma}\right)$

13

is M (strong mixing). If the invariant density is unique, it is necessarily an even function, since otherwise its mirror image $p(-x)$ would be a different invariant density.

Our aim is to examine the properties of the dependency ratio

$$f(x_1, x_2) = \frac{p^{1,2}(x_1, x_2)}{p(x_1)p(x_2)} = \frac{1}{\sigma p(x_2)} h\left(\frac{x_2 - g(x_1)}{\sigma}\right)$$

near the origin. We see that

$$f(-x_1, -x_2) = \frac{1}{\sigma p(-x_2)} h\left(\frac{-x_2 - g(-x_1)}{\sigma}\right) = \frac{1}{\sigma p(x_2)} h\left(-\frac{x_2 - g(x_1)}{\sigma}\right) = f(x_1, x_2).$$

It follows that the partial derivatives $\frac{\partial f}{\partial x_1}$ and $\frac{\partial f}{\partial x_2}$ vanish at the point $(x_1, x_2) = (0,0)$, so that the origin is always critical. The index of this critical point is determined by the Hessian matrix $Hf(0,0)$. If g is odd and h is even and unimodal, we have

$$p'(0) = 0, \qquad g''(0) = 0, \qquad h'(0) = 0, \qquad \text{and} \qquad h''(0) < 0.$$

After some algebra one finds that the Hessian evaluated at the origin reduces to

$$Hf(0,0) = \frac{h''(0)}{\sigma^3 p(0)} \begin{pmatrix} g'(0)^2 & -g'(0) \\ -g'(0) & 1 - \sigma^2 \frac{h(0)p''(0)}{h''(0)p(0)} \end{pmatrix}.$$

Since $h''(0)$ is negative, the Hessian matrix has a negative determinant if and only if the second derivative of the invariant density $p(x)$ satisfies

$$p''(0) < 0,$$

in which case the origin is a saddle-point. If however $p''(0) > 0$, the determinant is positive, while the trace of the matrix is negative, and the dependency ratio $f(x_1, x_2)$ has a local maximum at the origin.

Apparently, the critical point at the origin changes from a saddle point to a local maximum if $p''(0)$ becomes positive. Because this is exactly the condition for which the local maximum of $p(x)$ at the origin changes to a local minimum with a pair of maxima bifurcating off, it follows that for antisymmetric maps with symmetric unimodal noise, the 'ratio bifurcation' coincides with a phenomenological bifurcation.

4.2.1 On the circle

We illustrate this by figure 2 which shows the level sets of the dependency ratio, as well as the invariant density, for the map

$$X_{t+1} = a\left(0.5 \sin(X_t) + 0.25 \sin(2X_t)\right) + \varepsilon_{t+1},$$

where again M is the unit circle S^1 and where $\varepsilon_t \sim N(0, \sigma_{t+1}^2)$ with $\sigma_{t+1}^2 = 0.6 - 0.12 \cos(X_t)$. The noise variance is made state dependent to avoid the critical point at $(\pi, 0)$ to bifurcate simultaneously with the bifurcation at the origin. As noted above, the local minimum in $p(x)$ at $x = 0$ occurs exactly when $f(x_1, x_2)$ develops a local maximum at the origin. This is related to the fact that the denominator of the dependency ration contains a product of marginals which simultaneously develop a local minimum.
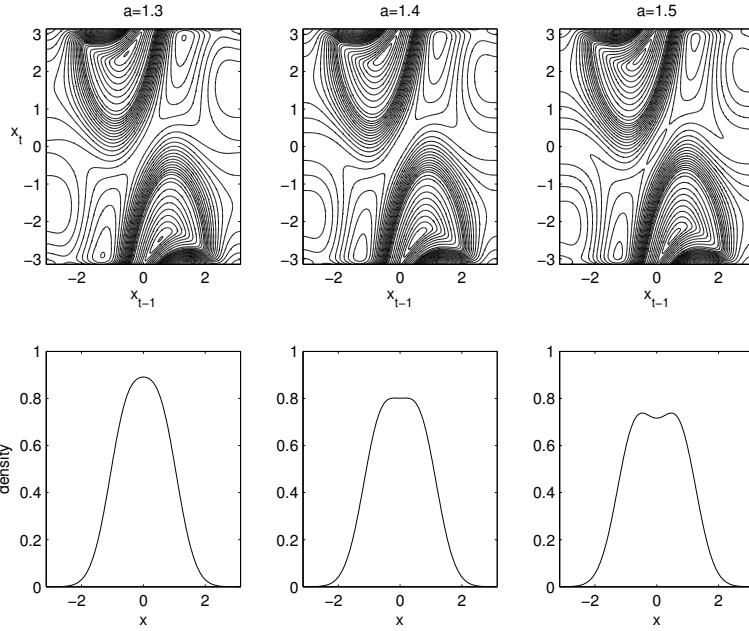
Figure 2: Level sets for the map $X_{t+1} = a\left(0.5\sin(X_t) + 0.25\sin(2X_t)\right) + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim N(0, \sigma_{t+1}^2)$ and $\sigma_{t+1}^2 = 0.6 - 0.12\cos(X_t)$ for increasing values of a (upper panels) and the corresponding marginal density functions (lower panels). The levels are not uniformly spaced.

4.2.2 On the real line

When we derived the coincidence of a P-bifurcation and a copula bifurcation at the origin in the antisymmetric case, apart from the global requirement of symmetry of the invariant density, only local arguments were used. Therefore, provided that we confine ourselves to cases with symmetric invariant densities, the result that a P-bifurcation coincides with a (local) ratio bifurcation, directly extends to stochastic dynamics on the real line.

As an example we consider the stochastic process on $\mathbb{R}$ defined by

$$X_{t+1} = \tanh(aX_t) + \varepsilon_{t+1}. \tag{8}$$

Figure 3 shows the level sets of the dependency ratio and the corresponding invariant probability density function for this map with $N(0, \sigma^2)$ distributed noise, taking $\sigma = 0.7$.

Note that the bifurcation parameter value differs from that of the analogous deterministic system ($\sigma = 0$): for the tanh map the stochastic analogue of the usual pitchfork bifurcation at $a = 1$ is shifted to a larger value of a. Apparently the value of the bifurcation depends on the noise level. A natural question, therefore, is whether for increasing noise levels the bifurcation parameter merely shifts, or whether the bifurcation can disappear altogether.

Intuitively, if the map is bounded and has a small range relative to the noise level, the dynamics is mainly governed by the noise and the deterministic part has little influence on the dynamics. In fact a simple argument shows that if the noise is fixed at a sufficiently large level, and if the family of odd maps $\{g_a\}$ is uniformly bounded, then there is no phenomenological bifurcation at $x = 0$, and therefore also no ratio bifurcation at $(x_1, x_2) = (0, 0)$, for symmetric processes of the form
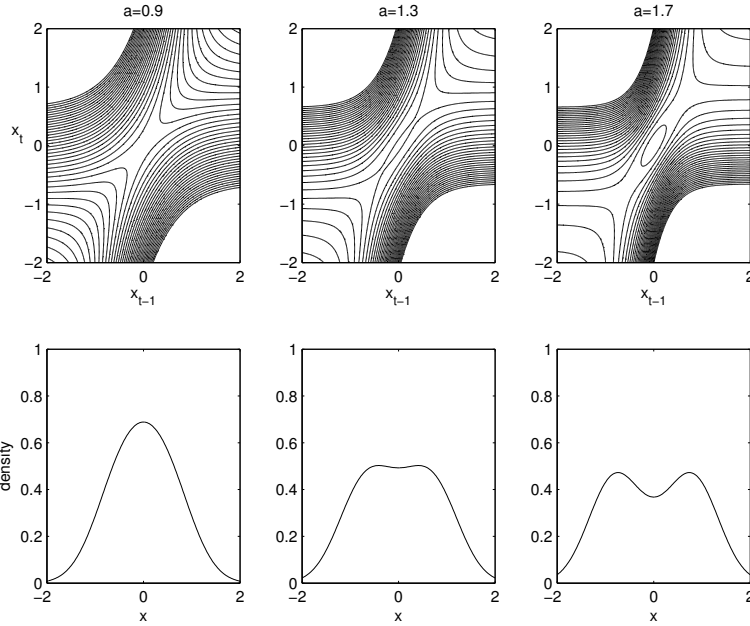
$$X_{t+1} = g_a(X_t) + \varepsilon_{t+1}. \tag{9}$$

15

Figure 3: Level sets for the map $X_{t+1} = \tanh(aX_t) + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim N(0, 0.5^2)$ for increasing values of a (top panels) and corresponding marginal density function (lower panels).

The argument runs as follows. By stationarity the invariant density p should satisfy

$$p(x) = \int \frac{1}{\sigma} h\left(\frac{x - g_a(s)}{\sigma}\right) p(s)\, ds.$$

A necessary condition for $p(x)$ to have a local minimum at $x = 0$ is that $p''(0) > 0$, where

$$p''(0) = \int \frac{1}{\sigma} h''\left(\frac{-g_a(s)}{\sigma}\right) p(s)\, ds.$$

Since $h(s)$ is a unimodal probability density function, its second derivative $h''(s)$ is negative in a neighbourhood of $s = 0$. It follows that, for g_a uniformly bounded, for large σ the integral on the right hand side of the last equation may remain negative as a varies.

4.3 Estimated dependency ratios from time series

In order to see whether dependency ratios can be used for classification of processes of which only a time series is available, a common situation in empirical applications, we estimate dependency ratios from simulated time series. We generate relatively short series $\{X_t\}$ from the stochastic models considered earlier in this section; we estimate from these series bivariate invariant densities and use them to reconstruct the dependency ratios. It is well known [1, 13] that fixed bandwidth nonparametric kernel density estimates become rather poor in regions with only few observations. One way to avoid this would be to use a data driven adaptive bandwidth which depends on the density locally, becoming larger as fewer observations are present locally. Instead of using an adaptive bandwidth we suggest, for real valued time series, to transform the data using the probability integral transform, that is, we construct

$$U_t = \widehat{F}_X(X_t) = \frac{\text{rank of } X_t \text{ among } \{X_s\}_{s=1}^N}{N}.$$

16

This amounts to transforming the invariant distribution to a uniform distribution on the unit interval, which tends to stabilise the estimation of the dependency ratio as the marginals no longer need to be estimated. In the case of first order ratios, the estimated empirical dependency ratio is equal to the empirical copula density

$$\widehat{f}(u_1, u_2) = \frac{1}{N-1} \sum_{t=1}^{N-1} K_b(u_1 - U_i, u_2 - U_{i+1}).$$

Here $K_b(u_1, u_2)$ is a bivariate probability kernel, which we take to be the commonly used Gaussian kernel:

$$K_b(u_1, u_2) = \frac{1}{\sqrt{2\pi}b} e^{-(u_1^2 + u_2^2)/(2b^2)}.$$

To avoid 'probability mass' from disappearing out of the unit square by this smoothing procedure, we impose periodic boundary conditions if $M = S^1$ and reflecting boundary conditions if $M = \mathbb{R}$.
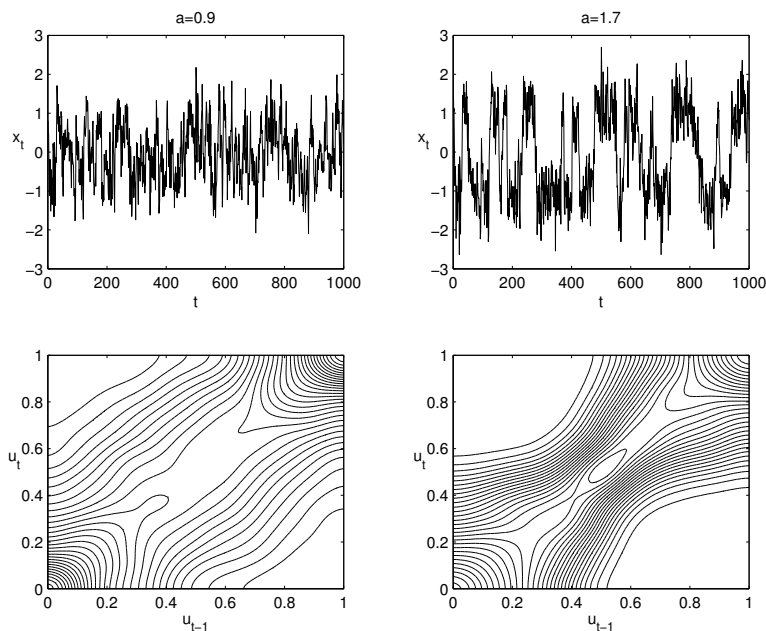

Figure 4: First 1000 values (top panels) of 4000 consecutive X_t-values generated by the map $X_{t+1} = \tanh(aX_t) + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim N(0, 0.5^2)$. for $a = 0.9$ (left) and $a = 1.7$ (right). The lower panels show the corresponding empirical level sets estimated with a Gaussian kernel (bandwidth $b = 0.07$).

Figure 4 shows level sets of the empirical dependency ratio obtained from time series of length 4000 from the symmetric hyperbolic tangent map given in equation (8) for different parameter values. The dependency ratio is estimated by smoothing the empirical copula with a bivariate normal probability density function (bandwidth $b = 0.07$). The empirical dependency ratio clearly reflects the fine structure of the theoretical dependency ratio.

Figure 5 shows an attempt at performing a similar reconstruction for the asymmetric sine map given by equation (7). In this case the topology of the reconstructed level sets does not correspond with that obtained earlier; this is due to estimation error. Probably longer time series (along with smaller bandwidths for the smoothers) are required for this case. We consider the optimal estimation and the related issue of data requirements for estimating dependency ratios as an important area for future research.
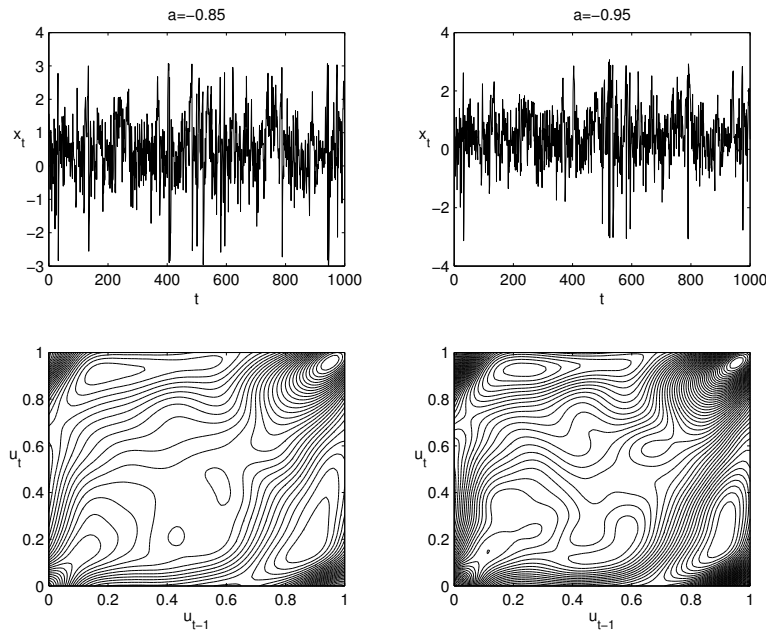
17

Figure 5: First 1000 values (top panels) of 4000 consecutive X_t-values generated by the map $X_{t+1} = X_t + a\sin(X_t) + 0.25\sin^2(X_t) + 0.25 + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim N(0, 0.7^2)$ for $a = -0.85$ (left) and $a = -0.95$ (right). The lower panels show the level sets of the corresponding empirical dependency ratio.

References

[1] I.S. Abramson, *On bandwidth variation in kernel estimates - a square root law*, Annals of Statistics **10** (1982), 1217–1223.

[2] L. Arnold, *Random dynamical systems*, Springer, Heidelberg, 1998.

[3] L. Cobb, *Stochastic catastrophe models and multimodal distributions*, Behavioral Science **23** (1978), 360–374.

[4] M.A.H. Dempster, I.V. Evstigneev, and K.R. Schenk-Hoppé, *Exponential growth of fixed-mix strategies in stationary asset markets*, Finance and Stochastics **7** (2003), 263–276.

[5] B.A. Dubrovin, A.T. Fomenko, and S.P. Novikov, *Modern Geometry − Methods and Applications. Part II: The Geometry and Topology of Manifolds*, Graduate Texts in Mathematics, vol. 104, Springer, New York, 1985.

[6] M.W. Hirsch, *Differential Topology*, Graduate Texts in Mathematics, vol. 33, Springer, New York, 1976.

[7] P. W. Holland and Y. J. Wang, *Dependence function for bivariate densities*, Communications in Statistics **A 16** (1987), 863–876.

[8] M. C. Jones, *The local dependence function*, Biometrika **83** (1996), 899–904.

[9] A. Lasota and M.C. Mackey, *Chaos, fractals, and noise: Stochastic aspects of dynamics*, Springer, Heidelberg, 1994, 2nd edition.

[10] S. Nadarajah, K. Mitov, and S. Kotz, *Local dependence functions for extreme value distributions*, Journal of Applied Statistics **30** (2003), 1081–1100.

[11] John C. Oxtoby, *Measure and category. A survey of the analogies between topological and measure spaces. 2nd ed.*, Graduate Texts in Mathematics, vol. 2, Springer, 1980.

[12] A. Ploeger, H. L. J. van der Maas, and P. Hartelman, *Catastrophe Analysis of Switches in the Perception of Apparent Motion*, Psychonomic Bulletin & Review **9** (2002), 26–42.

[13] G.R. Terrell and D.W. Scott, *Variable kernel density estimation*, Annals of Statistics **20** (1992), 1236–1265.

[14] René Thom, *Structural stability and morphogenesis. An outline of a general theory of models*, W. A. Benjamin, Reading, Massachusetts, 1975.

[15] E.C. Zeeman, *Stability of dynamical systems*, Nonlinearity **1** (1988), 115–155.

A Proofs of the topological properties

In this appendix, the topological properties given in section 3 are proved.

A.1 Proofs of the propositions

We repeat the statements of propositions 2 and 3 for the convenience of the reader.

Proposition 2. *On a compact manifold M, a dependency ratio is structurally stable under ratio equivalence if and only if it is regular.*

Proposition 3. *On a compact manifold M, the set of regular dependency ratios is everywhere dense.*

Proof.
These propositions are direct corollaries from the following two lemmas.

Lemma 1. *If M is compact and if $f : M_{n+1} \to \mathbb{R}$ is a regular dependency ratio, then there is a constant $\varepsilon > 0$ such that every $g \in N_\varepsilon(f)$ is regular and equivalent to f.*

Lemma 2. *If M is compact, the set of regular dependency ratios is dense in the C^2-topology.*

From lemma 1 we infer that regular ratios are structurally stable. If however f is a structurally stable ratio, there is a neighbourhood $U = N_\varepsilon(f)$ such that every $g \in U$ is equivalent to f. But as the regular ratios are dense, according to lemma 2, there is a regular ratio in U which then equivalent to f. By definition of equivalence, the ratio f itself has to be regular. The propositions follow. ∎

For non-compact manifolds of constant type, the following result is essentially a corollary of the results for compact manifolds.

Proposition 4. *On a manifold M of constant type, a well-behaved dependency ratio is stable with respect to the strong topology under ratio equivalence.*

Proof.

Let f be a well-behaved dependency ratio, and let $\{M_t\}$ be an exhaustion of M. Let moreover $T > 0$ and $c_t > 0$ be such that for every $t, t' > T$ we have that ∂M_t and $\partial M_{t'}$ are diffeomorphic, for every component $\{C_t^j\}$ of $(M_t)_{n+1}$ the restriction f_t^j of f to C_t^j is weakly ratio equivalent to $f_{t'}^j$ and there is a compact set K_t^j such that $|Df_t^j(x)|_x > c_t$ if $x \in C_t^j \backslash K_t^j$.

For every $t > T$ and every j, there is then a constant $\varepsilon_t^j > 0$ such that for every $g \in N_{\varepsilon_t^j}(f)$ in the C^2-topology on $(M_t)_{n+1}$, the restriction g_t^j of g to C_t^j is weakly ratio equivalent to f_t^j. Let $\varepsilon_t = \min_j \varepsilon_t^j$, and $\varepsilon(x) = \max\{\varepsilon_t \,|\, x \in (M_t)_{n+1}\}$. It follows that $N_\varepsilon(f)$ is an open neighbourhood of f in the strong C^2-topology, such that every $g \in N_\varepsilon(f)$ is ratio equivalent to f. $\blacksquare$

A.2 Proofs of the lemmas

It remains to demonstrate the lemmas. For this, we first need the following technical result. In the statement of the lemma, a ball of radius r around 0 is denoted by B_r, that is, $B_r = \{x \in \mathbb{R}^k \,|\, \|x\| < r\}$; also we have

$$\|f - g\|_{C^2(U)} = \max_{0 \leq j \leq 2} \max_{x \in U} |D^j f(x) - D^j g(x)|.$$

Lemma 3. *Let $U \subset \mathbb{R}^k$ be a bounded open set, and let $f : U \to \mathbb{R}$ be a C^2 function with $Df(0) = 0$ and $Hf(0)$ non-degenerate. Then there exist constants $\delta_0, \eta_0 > 0$ such that $B_{\delta_0} \subset U$ and that for every $0 < \delta \leq \delta_0$ and $0 < \eta \leq \eta_0$ there is an $\varepsilon > 0$, such that every function g satisfying $\|f - g\|_{C^2(U)} < \varepsilon$ has a unique non-degenerate critical point $\bar{y} \in B_\delta$ with $|g(\bar{y}) - f(0)| < \eta$, with g having the same index at $\bar{y}$ as f at 0.*

Proof.

For a matrix A, let $\|A\| = \max_{\|x\|=1} \|Ax\|$ denote the matrix norm of A. Since $Hf(0)$ is non-degenerate, there is a constant $c > 0$ such that $\|Hf(0)^{-1}\| = c$. Moreover, by continuity there is then a $\delta_1 > 0$, such that

$$\|Hf^{-1}(x)\| < 2c \qquad \text{for all} \quad x \in B_{\delta_1}.$$

Introduce $\psi = g - f$ and $h_t = f + t(g - f) = f + t\psi$. Then $h_0 = f$ and $h_1 = g$. We shall solve the equation

$$Dh_t(x) = 0 \tag{10}$$

for $t \in [0, 1]$, using the implicit function theorem. Note that

$$Hh_t(x) = Hf(x) + tH\psi(x) = Hf(x)\big(I + tHf(x)^{-1}H\psi(x)\big),$$

and consequently that

$$\|Hh_t(x)^{-1}\| \leq \frac{\|Hf(x)^{-1}\|}{1 - t\|Hf(x)^{-1}\|\|H\psi(x)\|}.$$

Taking $|t| \leq 1$, $x \in B_{\delta_1}$ and $\|\psi\|_{C^2} < (4c)^{-1}$, we obtain

$$\|Hh_t(x)^{-1}\| \leq 4c.$$

In particular, we can apply the implicit function theorem to solve $x = x(t)$ from (10), first around $t = 0$, and then around every value of t for which $x(t) \in B_{\delta_1}$. Note that since Hh_t is non-degenerate everywhere, the index of the critical point cannot change.

20

Furthermore

$$Dh_t(x) = Df(0) + Df(x) - Df(0) + tD\psi(x) = \int_0^1 Hf(sx)x\,ds + tD\psi(x)$$

$$= Hf(0)x + \int_0^1 \big(Hf(sx) - Hf(0)\big)x\,ds + tD\psi(x).$$

By the continuity of Hf, there exists $0 < \delta_2 < \delta_1$ such that $\|Hf(x) - Hf(0)\| < 1/(2c)$ for all $x \in B_{\delta_2}$. Recalling the estimate $\|Ax\| \ge \|A^{-1}\|^{-1}\|x\|$, it follows for $0 < \delta < \delta_2$ that if $|t| \le 1$, $\|x\| = \delta$ and $\|\psi\|_{C^2} < \delta/(2c)$, then

$$\|Dh_t(x)\| \ge \|Hf(0)^{-1}\|^{-1}\|x\| - \int_0^1 \|Hf(sx) - Hf(0)\|\|x\|\,ds - t\|D\psi(x)\|$$

$$> \frac{\delta}{c} - \frac{\delta}{2c} - \frac{\delta}{2c} > 0.$$

We have obtained the *a priori* statement that $x(t) \notin \partial B_\delta$ for all values $t \in [0,1]$ for which $x(t)$ is defined. Therefore $x(t)$ can be continued to $t = 1$, yielding $\bar{y} = x(1)$.

To show that $x(t)$ is the unique solution of (10) in B_δ, take $y \in B_\delta$ such that $y \ne x(t)$, and compute (putting $y_s = x(t) + s(y - x(t))$):

$$Dh_t(y) = Dh_t(y) - Dh_t(x(t)) = \int_0^1 Hh_t(y_s)(y - x(t))\,ds$$

$$= Hf(0)(y - x(t)) + \int_0^1 \big(Hf(y_s) - Hf(0)\big)(y - x(t))\,ds + t\int_0^1 H\psi(y_s)(y - x(t))\,ds.$$

It follows, as in the previous paragraph, that $\|Dh_t(y)\| > 0$ for all t. But then y cannot be a critical point.

Finally, if $v(t) = h_t(x(t))$ denotes the critical value of h_t, we see by differentiating that

$$\frac{dv}{dt} = \frac{\partial h_t}{\partial t}(x(t)) + Dh_t(x(t))\frac{dx}{dt} = \psi(x(t));$$

consequently

$$g(\bar{y}) - f(0) = v(1) - v(0) = \int_0^1 \psi(x(t))\,dt,$$

and $|g(\bar{y}) - f(0)| \le \max_{B_\delta} |\psi(x)|$.

Putting $\delta_0 = \delta_2$, $\eta_0 = 1$ and $\varepsilon = \min\{\frac{1}{4c}, \frac{\delta}{2c}, \eta\}$ yields the statement of the lemma. ∎

Lemma 1. *If M is compact and if $f : M_{n+1} \to \mathbb{R}$ is a regular dependency ratio, then there is a constant $\varepsilon > 0$ such that every $g \in N_\varepsilon(f)$ is regular and equivalent to f.*

Proof.
Let $\mathbf{x} = (x_1, \cdots, x_{n+1})$ and $\mathbf{y} = (y_1, \cdots, y_{n+1})$ denote points in M_{n+1}. Note that then $\pi_\ell(\mathbf{x}) = x_\ell$ etc. For a metric d on M, define

$$d_{n+1}(\mathbf{x}, \mathbf{y}) = \max_{1 \le \ell \le n+1} d\big(\pi_\ell(\mathbf{x}), \pi_\ell(\mathbf{y})\big).$$

Then d_{n+1} is a metric on M_{n+1}.

Let $\mathbf{x}_1, \cdots, \mathbf{x}_k$ be the critical points of f, ordered such that $v_i = f(\mathbf{x}_i) < f(\mathbf{x}_j) = v_j$ if $i < j$. Put $v_0 = 0$; then $v_0 < v_1$. Introduce

$$\zeta = \min_{0 \le i < j \le k} |v_i - v_j|, \qquad \sigma = \min_{0 \le \ell \le n} \min_{1 \le i < j \le k} d(\pi_\ell(\mathbf{x}_i), \pi_\ell(\mathbf{x}_j));$$

then ζ is the smallest absolute difference of two critical values, and σ is the smallest distance of two projections of critical points on M.

For every i, choose a neighbourhood W_i of $\mathbf{x}_i$ and a coordinate chart $x_i : W_i \to \mathbb{R}^{m(n+1)}$, such that $x_i(\mathbf{x}_i) = 0$, and set $f_i = f \circ x_i^{-1}$. By assumption $Df_i(0)$ vanishes and $Hf_i(0)$ is nondegenerate. For every i, take $0 < \delta_i < \sigma$ such that $B_{\delta_i} \subset x_i(W_i)$ and such that 0 is the only critical point of f_i in B_{δ_i}.

By the lemma, we can find $\varepsilon_i > 0$, such that every function g_i defined on $x_i(W_i)$ with $\|f_i - g_i\|_{C^2} < \varepsilon_i$ has a unique nondegenerate critical point y_i in B_{δ_i}, with $|f_i(0) - g_i(y_i)| < \zeta/2$ and with y_i having the same index t_i as 0.

Introduce the open sets $U_i = x_i^{-1}(B_{\delta_i}) \subset M_{n+1}$, and let $C = M_{n+1} \setminus \bigcup_i U_i$; note that C is compact, and that $df \ne 0$ on C. Therefore, there is $\zeta_0 > 0$, such that if $g \in N_{\zeta_0}(f)$, then $dg \ne 0$ on C as well. Moreover, for every i there is ζ_i such that $g \in N_{\zeta_i}(f)$, then in the chart x_i we have that $\|f_i - g_i\|_{C^2} < \varepsilon_i$. Set $\varepsilon = \min_{0 \le i \le k} \zeta_i$.

Finally, we have to provide a diffeomorphism $\varphi : M \to M$, homotopic to the identity, such that

$$\Phi(\mathbf{x}) = \big(\varphi \circ \pi_1(\mathbf{x})), \cdots, \varphi \circ \pi_{n+1}(\mathbf{x})\big)$$

maps $\mathbf{y}_i = x_i^{-1}(y_i)$ to $\mathbf{x}_i$.

Note that by the choice of δ_i, no two projections of the sets U_i on M intersect:

$$\pi_{\ell_1}(U_{i_1}) \cap \pi_{\ell_2}(U_{i_2}) = \emptyset, \qquad \text{for all} \quad 1 \le i_1 < i_2 \le k, \quad 1 \le \ell_1 < \ell_2 \le n+1.$$

Fix i and ℓ, and consider on $\pi_\ell(U_i)$ a differentiable curve $\gamma(t)$, defined for $0 \le t \le 1$, such that $\gamma(0) = \pi_\ell(\mathbf{x}_i)$ and $\gamma(1) = \pi_\ell(\mathbf{y}_i)$. Construct a vector field $X_{i\ell}$ on M such that $\dot\gamma(t) = X_{i\ell}(\gamma(t))$ for $0 \le t \le 1$ and $X_{i\ell} = 0$ on $M \setminus \pi_\ell(U_i)$.

Let $X = \sum_{i,\ell} X_{i,\ell}$. The time-1 map $\varphi = e^X$ has the required properties. ∎

Lemma 2. *If M is compact, the set of regular dependency ratios is dense in the C^2-topology.*

Proof.
Recall that the joint densities of an n-th order stochastic process propagate via the Perron-Frobenius operator (see e.g. [9]), giving the equation

$$p^{t-n+1,\cdots,t}(x_{t-n+1}, \cdots, x_t) = \int_M \tau(x_t | x_{t-n}, \cdots, x_{t-1}) p^{t-n,\cdots,t-1}(x_{t-n}, \cdots, x_{t-1}) \, dx_{t-n}.$$

If the process has a unique invariant density $p(x_1, \cdots, x_n)$, the process with the transition probability density

$$\tilde\tau(x_t | x_{t-n}, \cdots, x_{t-1}) = \tau(x_t | x_{t-n}, \cdots, x_{t-1}) + \frac{q(x_{t-n}, x_{t-n+1}, \cdots, x_t)}{p(x_{t-n}, \cdots, x_{t-1})}$$

has the same invariant density p, if q is small enough, such that $\tilde p$ is indeed a probability density, and if $\int_M q \, dx_{t-(n+1)+j} = 0$ for every $1 \le j \le n+1$.

For every point $\xi \in M$, we can find a chart $x = (x^1, \cdots, x^m)$ on M, such that $x(\xi) = 0$. Take $\delta > 0$ such that $U = x^{-1}(B_\delta)$ and $V = x^{-1}(B_{2\delta})$ are in the domain of x. Let $\varphi, \psi : M \to \mathbb{R}$ be smooth functions such that $\varphi = 1$ on U, $\varphi = 0$ on $M \setminus V$, and $\psi = 0$ on $U \cup M \setminus V$ and $\int_M \psi \, dx > 0$.

For $1 \leq j \leq m$, let $\ell_j : M \to \mathbb{R}$ be defined by setting

$$\ell_j(x) = \begin{cases} x^j \varphi + \beta_j \psi & \text{on} \quad V, \\ 0 & \text{on} \quad M \backslash V. \end{cases}$$

Moreover, set $\ell_0(x) = \varphi + \beta_0 \psi$. The constants β_j are chosen such that $\int_M \ell_j \, \mathrm{d}x = 0$ for all j.

For $\mathbf{x} \in M_{n+1}$, let $x = (x_1, \cdots, x_{n+1})$ be a chart such that $x(\mathbf{x}) = 0$. For $1 \leq i \leq n+1$, $1 \leq j \leq m$, and $p : M_{n+1} \to \mathbb{R}$ a function that is everywhere positive (this will be the invariant probability density $p^{n+1}(x_{n+1})$ later on), define

$$L^{\mathbf{x}}_{kj}(x) = \frac{\ell_{kj}(x_k)}{\ell_{k0}(x_k)} \frac{\prod_{i=1}^{n+1} \ell_{i0}(x_i)}{p(x)} = \frac{\ell_{10}(x_1) \cdot \ldots \cdot \ell_{kj}(x_k) \cdot \ldots \cdot \ell_{n+1,0}(x_{n+1})}{p(x)}.$$

Writing $x_k = (x_k^1, \cdots, x_k^m)$, we find that

$$\frac{\partial L^{\mathbf{x}}_{kj}}{\partial x_{k'}^{j'}}(0) = \begin{cases} \dfrac{1}{p(0)} & \text{if} \quad k' = k, \ j' = j, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that for $\delta > 0$ sufficiently small and $x(\mathbf{y}) \in B_\delta \times \cdots \times B_\delta$, the differentials of the functions $L^{\mathbf{x}}_{kj} \circ x$ are linearly independent vectors in $T_{\mathbf{y}}^* M_{n+1}$.

Choose for every $\mathbf{x} \in M_{n+1}$ such a value for δ, and set $U_{\mathbf{x}} = x^{-1}(B_\delta \times \cdots \times B_\delta)$. Since M is compact, it is covered by a finite number of the $U_{\mathbf{x}}$, say $U_{\mathbf{x}_1}, \cdots, U_{\mathbf{x}_K}$. Set

$$q^k_{ij} = p L^{\mathbf{x}_k}_{ij}.$$

Then q^k_{ij}/p is a finite collection of functions on M_{n+1} such that their differentials span $T_{\mathbf{x}}^* M_{n+1}$ at every point $\mathbf{x} \in M_{n+1}$. Moreover

$$\int_M q^k_{ij} \pi_\ell^* \, \mathrm{d}x_\ell = 0$$

for all ℓ.

Recall the remark made at the beginning of the proof; let the stochastic process defined by the transition probability $\tau(x_{n+1} | x_1, \cdots, x_n)$ have invariant probability densities $p^{1, \cdots, k}(x_1, \cdots, x_k)$ and dependency ratio

$$f(x_1, \cdots, x_{n+1}) = \frac{p^{1, \cdots, n+1}(x_1, \cdots, x_{n+1})}{p^{1, \cdots, n}(x_1, \cdots, x_n) p^{n+1}(x_{n+1})} = \frac{\tau(x_{n+1} | x_1, \cdots, x_n)}{p^{n+1}(x_{n+1})}.$$

Let moreover $a = (a^k_{ij})$ be such that

$$\tau(x_{n+1} | x_1, \cdots, x_n) + \sum_{ijk} a^k_{ij} \frac{q^k_{ij}(x_1, \cdots, x_{n+1})}{p^{1, \cdots, n}(x_1, \cdots, x_n)}$$

defines a parameterised joint probability density: this is always the case if the $|a^k_{ij}|$ are sufficiently small, since the transition probability density is assumed to be positive everywhere on the compact manifold M_{n+1}. Then the dependency ratio of the new process is given by

$$g(a, x) = f(x) + \sum_{ijk} a^k_{ij} \frac{q^k_{ij}(x)}{p^{n+1}(x_{n+1})},$$

where $a = (a_{ij}^k) \in A \subset \mathbb{R}^{Km(n+1)}$, where A is an open neighbourhood of 0.

Recall the definition of transversality (see e.g. [5], definition 10.3.1, p. 83): if X and Y are smooth manifolds, W a smooth submanifold of Y, the map $f : X \to Y$ smooth, and $x \in X$, then f intersects W transversally at x, if either $f(x) \notin W$ or $f(x) \in W$ and $T_{f(x)}Y = T_{f(x)}W + df(x)(T_xX)$. More generally, we say that f intersects W transversally at $A \subset X$, if f intersects W transversally at x for every $x \in A$.

We have the theorem that if A, X and Y are smooth manifolds, W a smooth submanifold of Y and $f : A \times X \to Y$ a smooth map which intersects W transversally, then the set of points $a \in A$ for which $f_a = f(a,.) : X \to Y$ intersects W transversally is everywhere dense in A (see [5], theorem 10.3.3, p. 85).

The derivative df of a function $f : M \to \mathbb{R}$ on a manifold M defines a section s of the cotangent bundle T^*M; in a sufficiently small neighbourhood U of a point in M, the restriction T_U^*M of the bundle to U is isomorphic to $U \times \mathbb{R}^m$, and the section takes the form $s(x) = (x, Df(x))$. The zero section M_0 of T^*M, which is isomorphic to M, is locally of the form $U \times \{0\}$.

The section s is transversal to M_0 at a point $x \in M_0$, if either $s(x) \notin M_0$, or if

$$T_{(x,0)}T^*M = ds(x)T_xM + T_{(x,0)}M_0 = (I, Hf(x))\mathbb{R}^m + \mathbb{R}^m \times \{0\}.$$

Note that this is equivalent to saying that s is transversal to M_0 everywhere if and only if the function f has only nondegenerate critical points. Such a function is called a *regular function* or a *Morse function*.

Consider now the function $g : A \times M \to \mathbb{R}$ and the associated map $s : A \times M \to T^*M$ given by $s(a,x) = (x, d_xg(a,x))$. Note that s is transversal to M_0, since in local coordinates

$$ds(a,x)T_{(a,x)}A \times M + T_{(x,0)}M_0 = \begin{pmatrix} 0 & I \\ d\frac{q_{ij}^k}{p^{n+1}} & H_xg(a,x) \end{pmatrix} \mathbb{R}^{Km(n+1)} \times \mathbb{R}^m + \mathbb{R}^m \times \{0\},$$

and since by construction the $d(q_{ij}^k/p^{n+1})$ span $\mathbb{R}^m$ everywhere on M. By the theorem mentioned above, the set of $a \in A$ for which $g_a = g(a,.)$ is a regular function which is everywhere dense in A.

For every $\varepsilon > 0$, we can choose a so small that $g = g_a$ is a regular function and $g \in N_\varepsilon(f)$, where $N_\varepsilon(f)$ is a neighbourhood in the C^∞ topology. It remains to show that by a second arbitrarily small perturbation, we can achieve regularity of the dependency ratio.

Note that since g is a regular function, its critical points are isolated. Denote them by $\mathbf{x}_1, \cdots, \mathbf{x}_N$. Assume that the points $\mathbf{x}_1$ up to $\mathbf{x}_{k-1}$ have different critical values, and that they are such that $\pi_\ell(\mathbf{x}_i) \neq \pi_\ell(\mathbf{x}_j)$ if $1 \leq i < j \leq k-1$.

We choose a neighbourhood $U \subset M_{n+1}$ of $\mathbf{x}_k$ such that U is contained in the domain of a chart x for which $x(\mathbf{x}_k) = 0$, and such that $\mathbf{x}_k$ is the only critical point of g in U. Let $a \in \mathbb{R}^{m(n+1)}$ be such that $\langle a, Hg(0)^{-1}a \rangle \neq 0$, where $\langle x, y \rangle$ denotes the inner product of the vectors x and y; the inverse of $Hg(0)$ exists since g is nondegenerate in 0; and the set of vectors a that do not satisfy the condition form a union of a smooth manifold of codimension 1 with the point $\{0\}$.

Consider the function

$$h_t(x) = h(t,x) = g(x) - t\sum_{ij} a_{ij}L_{ij}^{\mathbf{x}_k}.$$

The critical points of h_t are determined by the equation

$$0 = D_xh_t(x).$$

This equation can be solved using the implicit function theorem around $x = 0$ and $t = 0$ since $Hg(0)$ is invertible. For the solution $x = x(t)$, we find

$$\frac{dx}{dt}(0) = \frac{1}{p}Hg(0)^{-1}a. \tag{11}$$

24

Note that by the assumption on a, this derivative is nonzero. We restrict the possible choice of a further by requiring that

$$\pi_{\ell*}\frac{\mathrm{d}x}{\mathrm{d}t}(0) = \pi_{\ell*}\frac{1}{p}Hg(0)^{-1}a \neq 0.$$

Moreover, if $v(t) = h_t(x(t))$, then

$$\frac{\mathrm{d}v}{\mathrm{d}t}(t) = -\sum_{ij}a_{ij}L_{ij}^{\mathbf{x}_k} + D_xh_t(x) = -\sum_{ij}a_{ij}L_{ij}^{\mathbf{x}_k},$$

and

$$\frac{\mathrm{d}^2v}{\mathrm{d}t^2}(0) = -\frac{1}{p}\langle a, Hg(0)^{-1}a\rangle \neq 0. \tag{12}$$

Because of our choices, there are only finitely many values of t for which $v(t)$ is equal to one of the critical values $g(\mathbf{x}_1), \cdots, g(\mathbf{x}_{k-1})$, or for which the projections $\pi_\ell(\mathbf{x}_k)$ and $\pi_\ell(\mathbf{x}_i)$ coincide for some $1 \leq i < k$ and $1 \leq \ell \leq n+1$. From equations (11) and (12) it follows that the set of values of t avoiding these special values is everywhere dense in a neighbourhood of $t = 0$. This finishes the proof of the lemma. $\blacksquare$

List of other working papers:

2005

1. Shaun Bond and Soosung Hwang, Smoothing, Nonsynchronous Appraisal and Cross-Sectional Aggreagation in Real Estate Price Indices, WP05-17
2. Mark Salmon, Gordon Gemmill and Soosung Hwang, Performance Measurement with Loss Aversion, WP05-16
3. Philippe Curty and Matteo Marsili, Phase coexistence in a forecasting game, WP05-15
4. Matthew Hurd, Mark Salmon and Christoph Schleicher, Using Copulas to Construct Bivariate Foreign Exchange Distributions with an Application to the Sterling Exchange Rate Index (Revised), WP05-14
5. Lucio Sarno, Daniel Thornton and Giorgio Valente, The Empirical Failure of the Expectations Hypothesis of the Term Structure of Bond Yields, WP05-13
6. Lucio Sarno, Ashoka Mody and Mark Taylor, A Cross-Country Financial Accelorator: Evidence from North America and Europe, WP05-12
7. Lucio Sarno, Towards a Solution to the Puzzles in Exchange Rate Economics: Where Do We Stand?, WP05-11
8. James Hodder and Jens Carsten Jackwerth, Incentive Contracts and Hedge Fund Management, WP05-10
9. James Hodder and Jens Carsten Jackwerth, Employee Stock Options: Much More Valuable Than You Thought, WP05-09
10. Gordon Gemmill, Soosung Hwang and Mark Salmon, Performance Measurement with Loss Aversion, WP05-08
11. George Constantinides, Jens Carsten Jackwerth and Stylianos Perrakis, Mispricing of S&P 500 Index Options, WP05-07
12. Elisa Luciano and Wim Schoutens, A Multivariate Jump-Driven Financial Asset Model, WP05-06
13. Cees Diks and Florian Wagener, Equivalence and bifurcations of finite order stochastic processes, WP05-05
14. Devraj Basu and Alexander Stremme, CAY Revisited: Can Optimal Scaling Resurrect the (C)CAPM?, WP05-04
15. Ginwestra Bianconi and Matteo Marsili, Emergence of large cliques in random scale-free networks, WP05-03
16. Simone Alfarano, Thomas Lux and Friedrich Wagner, Time-Variation of Higher Moments in a Financial Market with Heterogeneous Agents: An Analytical Approach, WP05-02
17. Abhay Abhayankar, Devraj Basu and Alexander Stremme, Portfolio Efficiency and Discount Factor Bounds with Conditioning Information: A Unified Approach, WP05-01

2004

1. Xiaohong Chen, Yanqin Fan and Andrew Patton, Simple Tests for Models of Dependence Between Multiple Financial Time Series, with Applications to U.S. Equity Returns and Exchange Rates, WP04-19
2. Valentina Corradi and Walter Distaso, Testing for One-Factor Models versus Stochastic Volatility Models, WP04-18
3. Valentina Corradi and Walter Distaso, Estimating and Testing Sochastic Volatility Models using Realized Measures, WP04-17
4. Valentina Corradi and Norman Swanson, Predictive Density Accuracy Tests, WP04-16
5. Roel Oomen, Properties of Bias Corrected Realized Variance Under Alternative Sampling Schemes, WP04-15
6. Roel Oomen, Properties of Realized Variance for a Pure Jump Process: Calendar Time Sampling versus Business Time Sampling, WP04-14

7. Richard Clarida, Lucio Sarno, Mark Taylor and Giorgio Valente, The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates, WP04-13
8. Lucio Sarno, Daniel Thornton and Giorgio Valente, Federal Funds Rate Prediction, WP04-12
9. Lucio Sarno and Giorgio Valente, Modeling and Forecasting Stock Returns: Exploiting the Futures Market, Regime Shifts and International Spillovers, WP04-11
10. Lucio Sarno and Giorgio Valente, Empirical Exchange Rate Models and Currency Risk: Some Evidence from Density Forecasts, WP04-10
11. Ilias Tsiakas, Periodic Stochastic Volatility and Fat Tails, WP04-09
12. Ilias Tsiakas, Is Seasonal Heteroscedasticity Real? An International Perspective, WP04-08
13. Damin Challet, Andrea De Martino, Matteo Marsili and Isaac Castillo, Minority games with finite score memory, WP04-07
14. Basel Awartani, Valentina Corradi and Walter Distaso, Testing and Modelling Market Microstructure Effects with an Application to the Dow Jones Industrial Average, WP04-06
15. Andrew Patton and Allan Timmermann, Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity, WP04-05
16. Andrew Patton, Modelling Asymmetric Exchange Rate Dependence, WP04-04
17. Alessio Sancetta, Decoupling and Convergence to Independence with Applications to Functional Limit Theorems, WP04-03
18. Alessio Sancetta, Copula Based Monte Carlo Integration in Financial Problems, WP04-02
19. Abhay Abhayankar, Lucio Sarno and Giorgio Valente, Exchange Rates and Fundamentals: Evidence on the Economic Value of Predictability, WP04-01

2002

1. Paolo Zaffaroni, Gaussian inference on Certain Long-Range Dependent Volatility Models, WP02-12
2. Paolo Zaffaroni, Aggregation and Memory of Models of Changing Volatility, WP02-11
3. Jerry Coakley, Ana-Maria Fuertes and Andrew Wood, Reinterpreting the Real Exchange Rate - Yield Diffential Nexus, WP02-10
4. Gordon Gemmill and Dylan Thomas , Noise Training, Costly Arbitrage and Asset Prices: evidence from closed-end funds, WP02-09
5. Gordon Gemmill, Testing Merton's Model for Credit Spreads on Zero-Coupon Bonds, WP02-08
6. George Christodoulakis and Steve Satchell, On th Evolution of Global Style Factors in the MSCI Universe of Assets, WP02-07
7. George Christodoulakis, Sharp Style Analysis in the MSCI Sector Portfolios: A Monte Caro Integration Approach, WP02-06
8. George Christodoulakis, Generating Composite Volatility Forecasts with Random Factor Betas, WP02-05
9. Claudia Riveiro and Nick Webber, Valuing Path Dependent Options in the Variance-Gamma Model by Monte Carlo with a Gamma Bridge, WP02-04
10. Christian Pedersen and Soosung Hwang, On Empirical Risk Measurement with Asymmetric Returns Data, WP02-03
11. Roy Batchelor and Ismail Orgakcioglu, Event-related GARCH: the impact of stock dividends in Turkey, WP02-02
12. George Albanis and Roy Batchelor, Combining Heterogeneous Classifiers for Stock Selection, WP02-01

2001

1. Soosung Hwang and Steve Satchell , GARCH Model with Cross-sectional Volatility; GARCHX Models, WP01-16
2. Soosung Hwang and Steve Satchell, Tracking Error: Ex-Ante versus Ex-Post Measures, WP01-15
3. Soosung Hwang and Steve Satchell, The Asset Allocation Decision in a Loss Aversion World, WP01-14
4. Soosung Hwang and Mark Salmon, An Analysis of Performance Measures Using Copulae, WP01-13
5. Soosung Hwang and Mark Salmon, A New Measure of Herding and Empirical Evidence, WP01-12

6. Richard Lewin and Steve Satchell, The Derivation of New Model of Equity Duration, WP01-11
7. Massimiliano Marcellino and Mark Salmon, Robust Decision Theory and the Lucas Critique, WP01-10
8. Jerry Coakley, Ana-Maria Fuertes and Maria-Teresa Perez, Numerical Issues in Threshold Autoregressive Modelling of Time Series, WP01-09
9. Jerry Coakley, Ana-Maria Fuertes and Ron Smith, Small Sample Properties of Panel Time-series Estimators with I(1) Errors, WP01-08
10. Jerry Coakley and Ana-Maria Fuertes, The Felsdtein-Horioka Puzzle is Not as Bad as You Think, WP01-07
11. Jerry Coakley and Ana-Maria Fuertes, Rethinking the Forward Premium Puzzle in a Non-linear Framework, WP01-06
12. George Christodoulakis, Co-Volatility and Correlation Clustering: A Multivariate Correlated ARCH Framework, WP01-05
13. Frank Critchley, Paul Marriott and Mark Salmon, On Preferred Point Geometry in Statistics, WP01-04
14. Eric Bouyé and Nicolas Gaussel and Mark Salmon, Investigating Dynamic Dependence Using Copulae, WP01-03
15. Eric Bouyé, Multivariate Extremes at Work for Portfolio Risk Measurement, WP01-02
16. Erick Bouyé, Vado Durrleman, Ashkan Nikeghbali, Gael Riboulet and Thierry Roncalli, Copulas: an Open Field for Risk Management, WP01-01

2000

1. Soosung Hwang and Steve Satchell , Valuing Information Using Utility Functions, WP00-06
2. Soosung Hwang, Properties of Cross-sectional Volatility, WP00-05
3. Soosung Hwang and Steve Satchell, Calculating the Miss-specification in Beta from Using a Proxy for the Market Portfolio, WP00-04
4. Laun Middleton and Stephen Satchell, Deriving the APT when the Number of Factors is Unknown, WP00-03
5. George A. Christodoulakis and Steve Satchell, Evolving Systems of Financial Returns: Auto-Regressive Conditional Beta, WP00-02
6. Christian S. Pedersen and Stephen Satchell, Evaluating the Performance of Nearest Neighbour Algorithms when Forecasting US Industry Returns, WP00-01

1999

1. Yin-Wong Cheung, Menzie Chinn and Ian Marsh, How do UK-Based Foreign Exchange Dealers Think Their Market Operates?, WP99-21
2. Soosung Hwang, John Knight and Stephen Satchell, Forecasting Volatility using LINEX Loss Functions, WP99-20
3. Soosung Hwang and Steve Satchell, Improved Testing for the Efficiency of Asset Pricing Theories in Linear Factor Models, WP99-19
4. Soosung Hwang and Stephen Satchell, The Disappearance of Style in the US Equity Market, WP99-18
5. Soosung Hwang and Stephen Satchell, Modelling Emerging Market Risk Premia Using Higher Moments, WP99-17
6. Soosung Hwang and Stephen Satchell, Market Risk and the Concept of Fundamental Volatility: Measuring Volatility Across Asset and Derivative Markets and Testing for the Impact of Derivatives Markets on Financial Markets, WP99-16
7. Soosung Hwang, The Effects of Systematic Sampling and Temporal Aggregation on Discrete Time Long Memory Processes and their Finite Sample Properties, WP99-15
8. Ronald MacDonald and Ian Marsh, Currency Spillovers and Tri-Polarity: a Simultaneous Model of the US Dollar, German Mark and Japanese Yen, WP99-14
9. Robert Hillman, Forecasting Inflation with a Non-linear Output Gap Model, WP99-13
10. Robert Hillman and Mark Salmon , From Market Micro-structure to Macro Fundamentals: is there Predictability in the Dollar-Deutsche Mark Exchange Rate?, WP99-12
11. Renzo Avesani, Giampiero Gallo and Mark Salmon, On the Evolution of Credibility and Flexible Exchange Rate Target Zones, WP99-11
12. Paul Marriott and Mark Salmon, An Introduction to Differential Geometry in Econometrics, WP99-10

13. Mark Dixon, Anthony Ledford and Paul Marriott, Finite Sample Inference for Extreme Value Distributions, WP99-09
14. Ian Marsh and David Power, A Panel-Based Investigation into the Relationship Between Stock Prices and Dividends, WP99-08
15. Ian Marsh, An Analysis of the Performance of European Foreign Exchange Forecasters, WP99-07
16. Frank Critchley, Paul Marriott and Mark Salmon, An Elementary Account of Amari's Expected Geometry, WP99-06
17. Demos Tambakis and Anne-Sophie Van Royen, Bootstrap Predictability of Daily Exchange Rates in ARMA Models, WP99-05
18. Christopher Neely and Paul Weller, Technical Analysis and Central Bank Intervention, WP99-04
19. Christopher Neely and Paul Weller, Predictability in International Asset Returns: A Re-examination, WP99-03
20. Christopher Neely and Paul Weller, Intraday Technical Trading in the Foreign Exchange Market, WP99-02
21. Anthony Hall, Soosung Hwang and Stephen Satchell, Using Bayesian Variable Selection Methods to Choose Style Factors in Global Stock Return Models, WP99-01

1998

1. Soosung Hwang and Stephen Satchell, Implied Volatility Forecasting: A Compaison of Different Procedures Including Fractionally Integrated Models with Applications to UK Equity Options, WP98-05
2. Roy Batchelor and David Peel, Rationality Testing under Asymmetric Loss, WP98-04
3. Roy Batchelor, Forecasting T-Bill Yields: Accuracy versus Profitability, WP98-03
4. Adam Kurpiel and Thierry Roncalli , Option Hedging with Stochastic Volatility, WP98-02
5. Adam Kurpiel and Thierry Roncalli, Hopscotch Methods for Two State Financial Models, WP98-01