Random dynamical systems: A Bayesian approach

Konstantinos Kaloudis
(joint work with Christos Merkatas, Spyridon J. Hatjispyros)

Department of Statistics & Actuarial-Financial Mathematics
University of the Aegean

15th Summer School in Stochastic Finance
Athens, 09-13/07/2018

Table of contents

Outline

Dynamical systems

A deterministic mathematical prescription for evolving the state of a system forward in time (discrete/continuous).
Examples: Solar system, meteorology, population evolution, chemical reactions, ...

Discrete time evolution: Difference equations

State space $\mathscr{X}$, map $g : \mathscr{X} \to \mathscr{X}$, set of times T (e.g. $T = \mathbb{N}$)

Initial condition x_0, $x_1 = g(x_0), \ldots$

At n-th time step: $x_n = g^n(x_0)$, with $g^n = \underbrace{g \circ g \circ \ldots \circ g}_{n\text{–times}}$

$\mathscr{O}_g = \{g^n(x_0)\}_{n \in \mathbb{N}_0}$: Trajectory (orbit) of the initial condition x_0 under g

Different types of noise

- Systems evolve in the presence of noise.
- Observational noise: Does not give rise to new dynamical effects. Blurring evolution effect.
- Dynamical noise: The noise affects the evolution equations and can be additive, multiplicative or both.
- State space models combine dynamical and observational noise e.g. for $i = 1, \ldots, n$

$$\begin{aligned} x_i &= f(x_{i-1}) + e_i \\ y_i &= g(x_i) + \epsilon_i \end{aligned}$$

Description

- Random Dynamical System: System subjected to the effects of random dynamical noise.
- Projection of the effects of many parameters:
  High-dimensional noise.
- Deterministic dynamics can be drastically modified.

Main Idea

The states of the system are modeled as random variables parametrized with respect to time, over a probability space $(\Omega, \mathscr{F}, \mathscr{P})$.

$$X_i(\omega) = f(X_{i-1}(\omega)) + z_i(\omega)$$

Assumption: Noise distribution

$$z_i \overset{\text{iid}}{\sim} \mathcal{N}(\cdot \mid 0, \lambda^{-1}) \Rightarrow x_i \mid x_{i-1} \overset{\text{ind}}{\sim} \mathcal{N}(x_i \mid f(x_{i-1}), \lambda^{-1})$$

Bayesian modeling: Knowledge regarding dynamics of the deterministic counterpart is utilized in the form of a prior distribution over the parameter space or (and) restrictions on the dynamics of the process modeling the states of the system .

Problems in the RDS field

1. Reconstruction: Dynamical equations, known/unknown functional form.
2. Prediction of unobserved observations of the specific noisy realization of the system, finite horizon, forward/backward in time.
3. Noise reduction: Reduce the level of incorporated noise, using different methods for dynamic and observational noise.
4. Numerical approximation of dynamical quantities: generalizations of Lyapunov exponents, dimensions etc

Motivation

- The assumption of normal errors may cause problems (outlying errors).
- In nonparametric Bayesian modeling, no parametric form is assumed for the probability distribution.
- Prior beliefs of the noise process are assigned to the probability distributions.
- The priors are distributions over a suitable space of probability measures

A Bayesian nonparametric model is a model with the prior defined on an infinite dimensional parameter space.

Dirichlet Distribution

$$\mathscr{D}(p_1,\ldots,p_k \mid \alpha_1,\ldots,\alpha_k) \propto \prod_{i=1}^{k} p_i^{\alpha_i-1} \mathbb{1}\{(p_1,\ldots,p_k) \in S_k\}$$

- S_k is the k-dimensional probability simplex:

$$\sum_{i=1}^{k} p_i = 1, \quad p_i > 0, i = 1,\ldots,k$$

- Conjugate prior of Multinomial distribution.
- Multivariate analogue of Beta distribution.

Dirichlet Process I

- Draws from a Dirichlet Process are random probability measures denoted by $\mathbb{P} \sim \mathscr{DP}(\cdot \mid c, P_0)$

Ferguson, 1973

- For each finite partition $\{A_1, \ldots, A_m\}$ we have that

$$(\mathbb{P}(A_1), \ldots, \mathbb{P}(A_m)) \sim \mathscr{D}(cP_0(A_1), \ldots, cP_0(A_m))$$

- For appropriate sets A:

$$\mathbb{E}\{\mathbb{P}(A)\} = P_0(A)$$
$$\mathbb{V}\{\mathbb{P}(A)\} \propto (c+1)^{-1}.$$

- Whenever $X_1, \ldots, X_n \mid \mathbb{P} \overset{\text{iid}}{\sim} \mathbb{P}(\cdot)$ we have that the posterior random measure $\mathbb{P} \mid X_1, \ldots, X_n$ is also Dirichlet

Dirichlet Process II

- Probability measures drawn from a DP are almost surely discrete (Sethuraman, 1994)
- Stick Breaking representation: Countable mixture of point masses at random locations
- Constructive definition of Dirichlet random measures:

$$\mathbb{P}(A) = \sum_{j=1}^{\infty} w_j \delta_{\theta_j}(A), A \in \mathscr{F}$$

$$w_1 = v_1, w_j = v_j \prod_{i<j}(1-v_i), j \geq 2$$

$$v_j \overset{\text{iid}}{\sim} \mathscr{B}(1,c), \quad \theta_j \overset{\text{iid}}{\sim} P_0$$

- Cannot used for modeling continuous noise distributions.

Dirichlet Process Mixture Models

DPM (Antoniak, 1974 and Lo, 1984): To overcome the discrete nature of $\mathbb{P}$ we model the observations x_i for $i = 1, \ldots, n$ as

$$x_i \mid \mathbb{P} \overset{\text{iid}}{\sim} \int_{\Theta} K(\cdot \mid \theta)\mathbb{P}(d\theta)$$

Hierarchically we obtain

$$\begin{aligned} x_i \mid \theta_i &\overset{\text{ind}}{\sim} K(\cdot \mid \theta_i) \\ \theta_i \mid \mathbb{P} &\overset{\text{iid}}{\sim} \mathbb{P}(\cdot) \\ \mathbb{P} &\sim \mathscr{DP}(\cdot \mid c, P_0) \end{aligned}$$

Mixture of kernels with mixing measure the a.s. discrete random probability measure $\mathbb{P}$. Then, given $w = (w_j)_{j\geq 1}, \theta = (\theta_j)_{j\geq 1}$

$$x_i \mid w, \theta \sim \sum_{j=1}^{\infty} w_j K(\cdot \mid \theta_j)$$

Geometric Stick Breaking measures

Instead of considering Sethuraman's stick breaking weights, we consider their expectation (Fuéntes García et al. 2010)

$$\phi_j = \mathbb{E}\{v_j \prod_{l=1}^{j-1}(1-v_l)\}$$

Then we obtain the geometric weights

$$\phi_j = \lambda(1-\lambda)^{j-1}, \qquad \lambda = (c+1)^{-1}.$$

Geometric weights prior

$$\mathbb{P} = \sum_{j=1}^{\infty} \phi_j \delta_{\theta_j},$$

with $\lambda \sim \mathscr{B}(\alpha_\lambda, \beta_\lambda)$ and $\theta_j \overset{\text{iid}}{\sim} P_0$

Outline

Motivation

Financial data - complexity:

- Market instability and chaos: Unpredictability
- Noise and uncertainty: nonlinear stochastic models
- "Noisy chaos" - RDS: interaction between different types of traders
- Memory effects, volatility clustering, and non-normality
- Proper map - proper noise process

MG-(G)ARCH model

Mackey-Glass GARCH(1,1) (Kyrtsou C., Terraza M., 2003):

$$x_t = \underbrace{\alpha \frac{x_{t-\tau}}{1 + x_{t-\tau}^c} - \delta x_{t-1}}_{\text{Mackey-Glass}} + \epsilon_t$$

$$\epsilon | I_t \sim \mathcal{N}(0, h_t), \quad h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}$$

c fixed, $\tau = 1$

- Delay-difference equation
- Many applications: Physiology (Mackey, Glass 1977), dynamical diseases (Glass 2015)

MG-(G)ARCH model

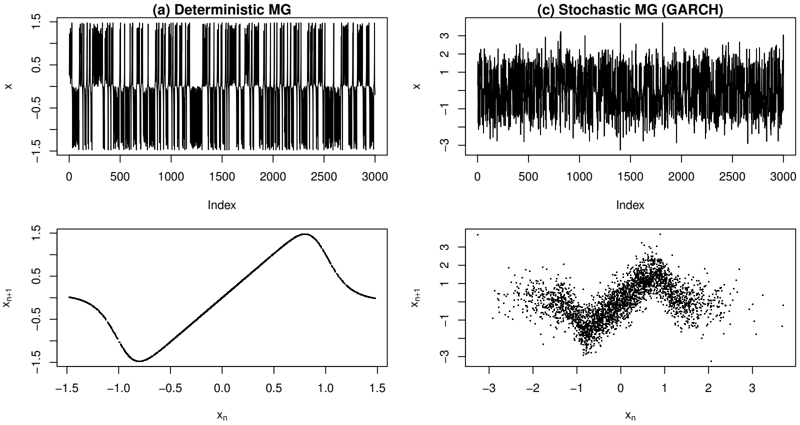

Figure 1: Time series of deterministic & stochastic Mackey-Glass map, for $c = 10, \tau = 1, \alpha = 2.1, \delta = 0.05, x_0 = 1.2$.

Data

The data used are the daily index series (CAC40) of the French Stock Exchange, during the period 09/07/1987-05/28/1999, giving 3,060 observations.

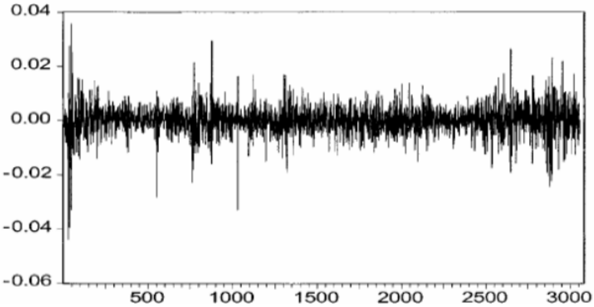

Figure 2: Paris Stock Exchange returns series

Results

- Test findings: Short memory (not iid)
- Lyapunov exponents, correlation dimension: Noisy chaotic or pure stochastic
- Statistically significant map parameters (capture nonlinearity effects)
- Volatility-clustering effects: significant α_1 coefficient
- MG-GARCH outperforms both chaotic and GARCH models (prediction MSE)

Comments

- RDS: efficient alternative modeling of financial time series
- Applications: case of Korea, 1997 crisis, structural shifts in causal relations (Karanasos, Kyrtsou 2005)
- Rejection of the nonlinearity hypothesis would not necessarily mean "linear structure in the mean equation". In the presence of high-dimensional dynamics, such tests cannot easily isolate and analyze underlying complexity (Kyrtsou 2005)
- Highly non-linear models can exhibit heteroskedasticity, when no heteroskedastic structure is assumed by construction (Kyrtsou 2008)

Outline

Aims

1. Fully reconstruct dynamical equations and predict future values: geometric stick breaking (GSB) mixture as prior for the noise process.
2. Compare modeling via GSB - DP mixtures.
3. Obtain information for the long term behavior of the underlying process: (quasi-invariant measure of the system).

The setting

We have observed the time series $x^{(n)} = (x_1, \ldots, x_n)$ generated by the stochastic process $\{X_i\}_{i \geq 0}$ as follows:

$$X_i = T_\vartheta(X_{i-1}, z_i) = g(\vartheta, X_{i-1}) + z_i,$$

where $\vartheta = (\theta_1, \ldots, \theta_d) \in \Theta$ parameter space, $g : \Theta \times X \to X$ a nonlinear and (for simplicity) continuous in X_{i-1} map.

The time series $x^{(n)}$ is fully determined by the parameters of the deterministic part, the initial condition and the realization of the stochastic part (the noise component).

Nonparametric noise component

$z_i \overset{iid}{\sim} f(\cdot)$: 0-mean nonparametric density with Gaussian Kernels

$$f(\cdot \mid \mathbb{G}) = \int_{\mathbb{R}^+} \mathcal{N}(\cdot \mid 0, \tau^{-1})\mathbb{G}(d\tau) = \sum_{i=1}^{\infty} \pi_j \mathcal{N}(\cdot \mid 0, \tau_i^{-1})$$

- $\mathbb{G} = \sum_{j \geq 1} \pi_j \delta_{\tau_j}$ is the mixing measure
- $\pi = (\pi_i)_{i \geq 1}$ and $\tau = (\tau_i)_{i \geq 1}$ are infinite sequences of weights and locations respectively.
- Transition kernels

$$f(x_i \mid x_{i-1}, \theta, \pi, \tau) = \sum_{i=1}^{\infty} \pi_j \mathcal{N}(x_i \mid g(\vartheta, x_{i-1}), \tau_j^{-1})$$

Slice sets

- Infinite mixtures: Need for finite-dimensional Gibbs samplers
- Introduce clustering variables d_i and proper slice sets A_i such that $d_i|A_i$ attains a discrete uniform distribution over -a.s. finite set of indices- A_i.
- Then, given the sets A_i, the observations are coming from an a.s. finite mixture of normal kernels

$$f\,(x_i \mid A_i) = \frac{1}{|A_i|}\sum_{j\in A_i} \mathcal{N}\left(x_i \mid g(\vartheta, x_{i-1}), \tau_j^{-1}\right)$$

- How to construct the slice sets?

Non-sequential Slice sets

- Assign $A_i = \{j \in \mathbb{N} : 0 < u_i < \pi_j\}$ to each observation x_i
- A_i depends on the weights π through the auxiliary random variable u_i such that $f_\pi(d_i = j | u_i) \propto \pi_j \, \mathcal{U}(u_i | 0, \pi_j)$

DP mixture based augmented random density

$$f_{w,\lambda}(x_i, u_i, d_i = j) = w_j \, \mathcal{U}(u_i | 0, w_j) \times \mathcal{N}(x_i | 0, \lambda_j^{-1})$$

with stick-breaking weights $\pi = w$, $w_1 = z_1$ and for $j > 1$

$$w_j = z_j \prod_{s<j} (1 - z_s),$$

with the $z_i \overset{\text{iid}}{\sim} \mathcal{B}(1, c)$.

Sequential Slice sets

- Introduce $N_i \stackrel{iid}{\sim} f_N$ a.s. finite discrete random variables of mass f_N (Fuéntes-García et al (2010)).
- Given N_i we have $f(d_i = j | N_i) = N_i^{-1} \mathbb{1}\{j \le N_i\}, \quad 1 \le i \le n$

GSB mixture based augmented random density

$$f_\lambda(x_i, N_i = l, d_i = j) = f_N(l\,|\,p)\, l^{-1} \times \mathbb{1}\{j \le l\}\, \mathcal{N}(x_i | 0, \lambda_j^{-1})$$

with geometric weights

$$\pi_j = \sum_{l=j}^{\infty} l^{-1} f_N(l\,|\,p) = p\,(1-p)^{j-1} \mathbb{1}\{j \ge 1\}$$

if $f_N(l\,|\,p) = \mathcal{NB}(l\,|\,2, p) = l p^2 (1-p)^{l-1} \mathbb{1}\{l \ge 1\}$

Reconstruction models

DP mixture based model

For $i = 1, \ldots, n$ and $j \geq 1$:

$$(x_i | x_{i-1}, d_i = j, \theta, \lambda) \overset{\text{ind}}{\sim} \mathcal{N}(g(\theta, x_{i-1}), \lambda_j^{-1})$$

$$(u_i | d_i = j, w) \overset{\text{ind}}{\sim} \mathcal{U}(0, w_j)$$

$$\mathrm{P}(d_i = j | w) = w_j$$

$$w_j = z_j \prod_{s<j} (1 - z_s),\ z_j \overset{\text{iid}}{\sim} \mathcal{B}e(1, c)$$

$$\lambda_j \overset{\text{iid}}{\sim} P_0$$

GSB mixture based model

For $i = 1, \ldots, n$ and $j \geq 1$:

$$(x_i | x_{i-1}, d_i = j, \theta, \lambda) \overset{\text{ind}}{\sim} \mathcal{N}(g(\theta, x_{i-1}), \lambda_j^{-1})$$

$$(d_i | N_i = l) \overset{\text{ind}}{\sim} \mathcal{U}\{1, \ldots, l\}$$

$$\pi_j = \mathcal{NB}(j | 1, p),\ N_i \overset{\text{iid}}{\sim} \mathcal{NB}(2, p)$$

$$\lambda_j \overset{\text{iid}}{\sim} P_0$$

Sampling algorithm

- Non-standard full conditionals (FCs): Proper augmentation schemes, following Damien et al (1999), Hatjispyros et al (2009)
- Fully stochastic version of DPR (rDPR), randomizing concentration parameter $c \sim \mathcal{G}(\alpha, \beta)$
- "Synchronized" prior specifications for c and $p = (1+c)^{-1}$ for comparison purposes:

$$f(p) = \mathcal{T}\mathcal{G}(p \,|\, \alpha, \beta) = \frac{\beta^{\alpha} e^{\beta}}{\Gamma(\alpha)} \, p^{-(\alpha+1)} e^{-\beta/p} (1-p)^{\alpha-1},$$

with $p \in (0,1)$

A random cubic map

We will generate observations from a cubic random map with a deterministic part given by

$$\tilde{g}(\vartheta, x) = 0.05 + \vartheta x - 0.99x^3$$

- $x_i = \tilde{g}(\vartheta, x_{i-1}) + z_i$, for parameter value $\vartheta^* = 2.55$ and initial condition $x_0 = 1$
- Coexistence of a repelling strange set and an attracting strange set
- We model the deterministic part of the map with a polynomial in x of degree $m = 5$.

Noise Processes

1. The equally weighted normal 4-mixture

$$f_1 = \sum_{r=0}^{3} \frac{1}{4} \mathcal{N}\left(0, (5r+1)\sigma^2\right), \ \sigma = 10^{-2} \tag{1}$$

2. The normal 2-mixtures, which exhibit progressively heavier tails for $1 \leq l \leq 4$

$$f_{2,l} = \frac{5+l}{10} \mathcal{N}(0, \sigma^2) + \frac{5-l}{10} \mathcal{N}\left(0, (200\sigma)^2\right), \ \sigma = 10^{-3} \tag{2}$$

Measure of tail fatness of the density $z \sim f$:

$$TF_f = \mathbb{E}|z| / \sqrt{\mathbb{E}|z|^2}$$

The closer TF_f is to 1, the thinner the tails are. It can be verified numerically that

$$TF_{f_1} > TF_{f_{2,1}} > \cdots > TF_{f_{2,4}}$$

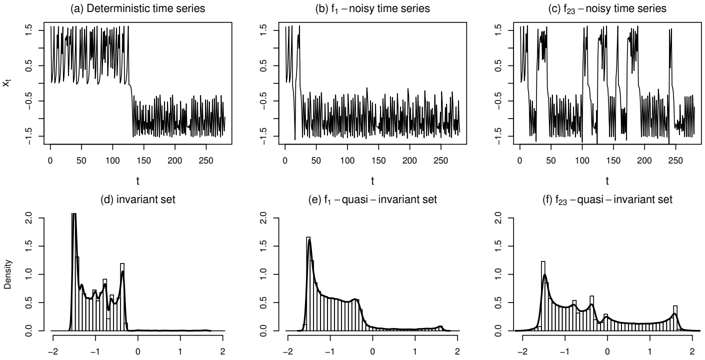

Figure 3: In figures 3(a)-(c) we display the deterministic orbit and f_1 and $f_{2,3}$ data-realizations with initial condition $x_0 = 1$. In figures 3(d)-(f) we display the deterministic and the f_1 and $f_{2,3}$ quasi-invariant set approximations respectively.

rDPR -GSBR Synnchronized Prior Specifications

$$c \sim \mathcal{G}(\alpha, \beta), \quad p \sim \mathcal{T}\mathcal{G}(\alpha, \beta), \quad \{\lambda_j \sim \mathcal{G}(a, b) : j \geq 1\}$$
$$\theta \sim \mathcal{U}((-M, M)^{k+1}), \quad x_0 \sim \mathcal{U}(-M_0, M_0),$$

where k is the degree of the modeling polynomial.

- Noninformative reconstruction and prediction:

$$\mathcal{P}\mathcal{S}_{\mathrm{NRP}} : \alpha = \beta \geq 10^{-1},\ a = b \geq 10^{-4},\ M \gg 1,\ M_0 \gg 1$$

- Informative reconstruction and prediction:

$$\mathcal{P}\mathcal{S}_{\mathrm{IRP}} : \alpha > \beta \geq 10^{-1},\ a > b \geq 10^{-4},\ M \gg 1,\ M_0 \gg 1.$$

How to choose prior set up?

- Informative structure in data $\rightarrow$ predictability
- Forecastable component analysis Ω (ForeCA): Large Ω values characterize more predictable time series.
- Data sets $\{X^n_{f_{2,l}} : 1 \leq l \leq 4\}$ have the more informative structure: $\Omega(X^n_{f_{2l}}) > \Omega(X^n_{f_1}), \quad n > 80, 1 \leq l \leq 4$

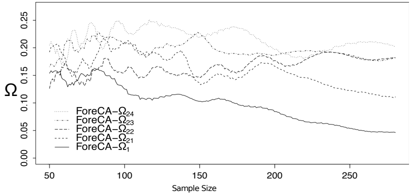

Figure 4: Here we display the Ω curves relating to the data sets $X^n_{f_1}$ and $\{X^n_{f_{2,l}} : 1 \leq l \leq 4\}$ for n between 50 and 280.

Informative reconstruction-prediction under f_1 noise

- rDPR,GSBR Gibbs samplers for $T = 20$, data set $X_{f_1}^{200}$, 500,000 iterations with 10,000 burn-in period
- $\mathscr{P}\mathscr{S}_{\mathrm{IRP}}$: $\alpha = 3$, $\beta = 0.3$, $a = 1$, $b = 10^{-3}$ and $M = M_0 = 10$
- Bayesian estimators: Sample mean (SM)-Sample mode (MAP)

Model	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	x_0
Param.	1.98	0.37	0.03	0.58	0.00	0.04	$x_{\mathrm{M}}: 3.87$
rDPR	0.81	0.29	0.01	0.09	0.04	0.14	$x_{\mathrm{M}}: 0.80$
GSBR	0.19	0.27	0.05	0.04	0.02	0.18	$x_{\mathrm{R}}: 0.60$
Estim.	x_{201}	x_{202}	x_{203}	x_{204}	x_{205}	GSBR-Av	Par-Av
SM	6.43	7.35	29.70	5.48	13.68	12.53	53.49
MAP	3.84	11.48	19.16	2.15	149.06	37.14	53.25

Table 1: (θ, x_0) reconstruction PAREs ($T = 0$) under the informative prior configuration.

Results I

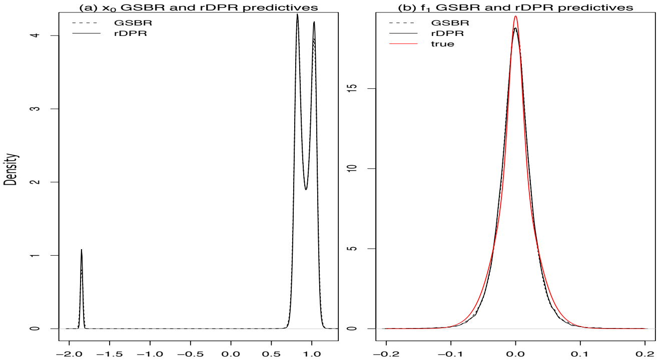

Figure 5: In (a) we give superimposed the KDE's based on the posterior marginal predictive samples of the initial condition variable x_0. In (b) we superimpose the GSBR and the rDPR noise density estimations together with the true dynamical error density.

Results II

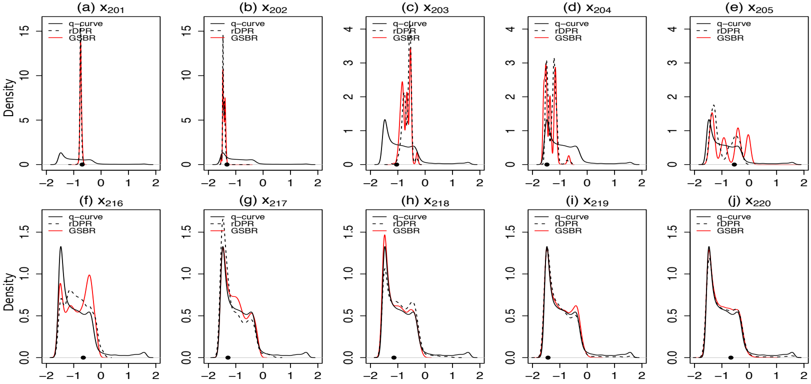

Figure 6: In (a)-(j) we display superimposed the first five and the last five KDE's of the out-of-sample posterior marginal predictive based on data set $X_{f_1}^{200}$ under $\mathscr{P}\mathscr{S}_{\text{IRP}}$. Together we superimpose the KDE of the f_1 quasi invariant density (solid black line). Bullet points represent the corresponding true future value.

Noninformative reconstruction-prediction under $f_{2,l}$ noise I

- rDPR,GSBR Gibbs samplers for $T = 20$, data sets $\{X^{200}_{f_{2,l}}\}$, 500,000 iterations with 10,000 burn-in period
- $\mathscr{P}\mathscr{S}_{\mathrm{NRP}}$: $\alpha = \beta = 0.3, a = b = 10^{-3}, M = M_0 = 10$

Noise	Model	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	x_0
$f_{2,1}$	Param.	19.95	1.54	4.83	4.39	2.52	1.01	7.27
	GSBR	0.51	0.01	0.06	0.02	0.02	0.00	x_R : 0.03
$f_{2,2}$	Param.	2.89	0.94	4.07	2.37	2.07	0.76	7.49
	GSBR	0.54	0.05	0.06	0.12	0.03	0.03	x_R : 0.03
$f_{2,3}$	Param.	29.97	0.40	4.97	1.25	1.88	0.41	7.55
	GSBR	0.20	0.04	0.04	0.13	0.02	0.04	x_R : 0.03
$f_{2,4}$	Param.	15.57	1.07	1.33	3.71	0.43	1.03	6.40
	GSBR	0.10	0.01	0.05	0.03	0.01	0.00	x_R : 0.03

Table 2: Simultaneous reconstruction-prediction under the noninformative prior specification. The (θ, x_0) PARE's are based on the data sets $\{X^{200}_{f_{2,l}} : 1 \leq l \leq 4\}$ for $T = 20$.

Noninformative reconstruction-prediction under $f_{2,l}$ noise II

Noise	Estim.	x_{201}	x_{202}	x_{203}	x_{204}	x_{205}	GSBR-Av	Par-Av
$f_{2,1}$	SM	12.50	0.86	12.57	44.04	82.11	30.42	58.72
	MAP	12.86	2.10	77.13	25.89	39.99	31.59	69.62
$f_{2,2}$	SM	0.52	0.70	8.07	167.16	15.17	38.32	65.08
	MAP	0.29	1.72	0.50	103.00	20.96	25.29	65.57
$f_{2,3}$	SM	0.72	7.99	0.01	9.74	49.94	13.68	233.53
	MAP	0.14	0.47	2.34	0.39	1.38	0.93	234.80
$f_{2,4}$	SM	0.24	1.01	2.95	3.79	40.25	9.65	60.69
	MAP	0.07	0.86	4.78	0.13	21.00	5.37	109.23

Table 3: Simultaneous reconstruction-prediction under the noninformative prior specification. The out-of-sample PARE's are based on data sets $\{X^{200}_{f_{2,l}} : 1 \leq l \leq 4\}$ for $T = 20$. The GSBR-Av and Par-Av columns are the PARE means of the first five out-of-sample estimations using the GSBR and the parametric Gibbs samplers respectively.

Results

- Parameter-initial condition-noise density estimation
- Non-efficiency of simple MCMC samplers
- Quasi invariant density approximation: Prediction barrier
- GSBR: Same level performance with rDPR - Smaller execution times

Data set $X_{f_1}^{200}$

Prior spec.	Algorithm	$T = 0$	$T = 20$
$\mathscr{P}\mathscr{S}_{\text{IRP}}$	rDPR	5.44	11.76
$\mathscr{P}\mathscr{S}_{\text{IRP}}$	GSBR	2.24	8.65

Table 4: Mean execution times in seconds per 10^3 iterations.

Outline

Introduction

We have in our disposal the time series $\boldsymbol{x} = (x_1, \ldots, x_n)$ generated by the stochastic process $\{x_i\}_{i \geq 0}$ as follows:

$$x_i = T_\vartheta(x_{1:d}, e_i) = g(\vartheta, x_{1:d}) + e_i \tag{3}$$

where $\vartheta = (\theta_1, \ldots, \theta_k) \in \Theta$ parameter space, $g : \Theta \times X^d \to X$ a nonlinear and continuous in $x_{1:d} = (x_{i-1}, \ldots, x_{i-d})$ map. We assume the noise variables e_i independent of the states $x_{i-1}, \ldots, x_{i-d}$ and independent to each other.

- Orbits contaminated with dynamical noise are pseudoorbits of the underlying dynamics $g(\cdot)$, $|x_i - g(\vartheta, x_{1:d})| < \alpha, i = 1, \ldots, n$
- Invariant measure $\to$ deformation into a quasi-invariant measure

Useful measures

Original orbit ($\boldsymbol{x}$) - Noise-reduced orbit ($\boldsymbol{y}$)

1. Noise level $\eta = \dfrac{\sigma_{\text{noise}}}{\sigma_{\text{signal}}}$
2. Measure of overall deviation of an orbit $\boldsymbol{x}$ from determinism is
$$E_{\text{dyn}}(\boldsymbol{x}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - g(\vartheta, x_{1:d}))^2}$$
3. Measure of overall distance between orbits (average correction)
$$E_0(\boldsymbol{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$$
4. Percentage of dynamical noise reduction
$$\text{NR}(\boldsymbol{y}, \boldsymbol{x}) = 100\left|\frac{E_{\text{dyn}}(\boldsymbol{y}) - E_{\text{dyn}}(\boldsymbol{x})}{E_{\text{dyn}}(\boldsymbol{x})}\right| \%$$

Effects of dynamical noise

Hénon map:

$$x_i = g(x_{i-1}, x_{i-2}) + e_i = 1.38 - x_{i-1}^2 + 0.27 x_{i-2} + e_i, \tag{4}$$

where $e_i \overset{\text{i.i.d.}}{\sim} 0.6\,\mathcal{N}(0, \sigma^2) + 0.4\,\mathcal{N}(0, 100\sigma^2), \quad \sigma^2 = 0.21 \times 10^{-4}$

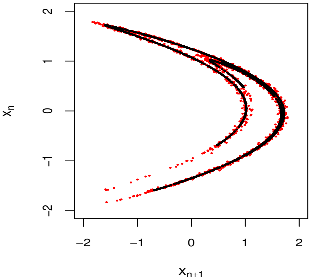

Figure 7: Noisy and deterministic Hénon trajectories, 3% dynamical noise level.

- Hyperbolic case - shadowing
- Nonhyperbolic case - dynamical noise
- Homoclinic tangencies (HT) and noise amplification
- Noise-induced prolongations

Aim

Estimate a denoised orbit $\boldsymbol{y} = \{y_i\}_{i=1}^n$, fulfilling approximately the same estimated dynamics and evolving in a neighborhood of the observed noisy orbit, under lower level of incorporated dynamical noise.

- High $\mathrm{NR}(\boldsymbol{y}, \boldsymbol{x})$ & $E_{\mathrm{dyn}}(\boldsymbol{y}) < E_{\mathrm{dyn}}(\boldsymbol{x})$ (high noise reduction)
- $E_0(\boldsymbol{y}) \simeq \eta$ (small average correction)
- Assumption: Zero-mean, symmetric noise process

Generic probability model

$$
\begin{aligned}
x_i &= g(\theta, x_{i:d}) + e_i, \quad e_i \overset{\text{i.i.d.}}{\sim} f(\cdot) \\
f(\cdot) &= \sum_{k=1}^{\infty} w_k \mathcal{N}(\cdot|0, \lambda_k^{-1}),\ 1 \le i \le n \\
y_i &= g(\theta, y_{i:d}) + \zeta_i, \quad \zeta_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\cdot|0, \delta) \\
y_{1:d} &= x_{1:d},\ \mathrm{P}-\text{a.s.} \quad \text{and} \quad |x_i - y_i| < \gamma_i, \quad \gamma_i \overset{\text{i.i.d.}}{\sim} h(\cdot)
\end{aligned}
\tag{5}
$$

- Reasonable prior over δ - Posterior concentrated near zero
- Proximity restriction: y_i's γ_i-close to x_i
- Minimization of the overall deviation of y_i-orbit from determinism

Full posterior distribution

$$
\begin{aligned}
\pi(p,\lambda^{\infty},d^{n},N^{n},\theta,x_{1:d},\tau,y^{(n)}\,|\,x^{(n)},\rho,\mathscr{R},\mathscr{P},\mathscr{M}) &\propto \pi(p,\lambda^{\infty},\tau,\theta,x_{1:d}) \\
&\times \prod_{\substack{i=1 \\ d_i : d_i \le N_i}}^{n} p^2(1-p)^{N_i-1}\lambda_{d_i}^{1/2}\exp\left\{-\frac{\lambda_{d_i}}{2}(x_i-g(\theta,x_{i:d}))^2\right\} \\
&\times \quad \mathbb{1}\{y_{1:d}=x_{1:d}\}\;\tau^{n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[\tau\underbrace{(y_i-g(\theta,y_{i:d}))^2}_{indeterminism}+\rho\underbrace{(y_i-x_i)^2}_{proximity}\right]\right\}.
\end{aligned}
$$

Full conditional distributions of the denoised variables y_i

$$\pi(y_j|\cdots) \propto e^{-C(y_j|\cdots)/2}$$

Letting $h_\theta(y_j, y_{j:d}) := (y_j - g(\theta, y_{j:d}))^2$, the function $C(y_j|\cdots)$, for $j = 1, \ldots, d$ is given by

$$C(y_j|\cdots) = \tau \sum_{k=0}^{d} h_\theta(y_{j+k}, y_{j+k:d}) \times \mathbb{1}\left\{y_0 = x_0, \ldots, y_{-d+j} = x_{-d+j}\right\} + \rho(y_j - x_j)^2,$$

for $j = d+1, \ldots, n-d$ is given by

$$C(y_j|\cdots) = \tau \sum_{k=0}^{d} h_\theta(y_{j+k}, y_{j+k:d}) + \rho(y_j - x_j)^2,$$

and, for $j = n-d+1, \ldots, n$, by

$$C(y_j|\cdots) = \tau \sum_{k=0}^{j-n} h_\theta(y_{j+k}, y_{j+k:d}) + \rho(y_j - x_j)^2.$$

DNRR model: Algorithm (I)

We first specify initial values for the variables $x_{1:n}$, θ, τ, and we iterate for $t = 1, \ldots, K$ the following sampling scheme:

1. For $i = 1, \ldots, n$, generate the state space range variable $N_i^{(t)} \sim \pi(N_i | \cdots)$, of the allocation variable $d_i^{(t)}$.
2. For $i = 1, \ldots, n$, generate the infinite mixture allocation variable $d_i^{(t)} \sim \pi(d_i | \cdots)$.
3. For $i = 1, \ldots, N^*$, with $N^* = \max_{1 \le k \le n} N_k$, sample $\lambda_i^{(t)} \sim \pi(\lambda_i | \cdots)$.
4. Generate the initial condition vector $(x_{1:n})^{(t)} \sim \pi(x_{1:n} | \cdots)$
5. Generate $\theta^{(t)} \sim \pi(\theta | \cdots)$.
6. Sample the geometric probability $p^{(t)} \sim \pi(p | \cdots)$.
7. Having updated $p^{(t)}$ and $\lambda^{(t)}$ up to N^*, sample from the noise process $\hat{f}_{x^n}$

$$z_{n+1}^{(t)} \sim \sum_{j=1}^{N^*} p^{(t)} (1 - p^{(t)})^{j-1} \mathcal{N}\left(z_{n+1} \,|\, 0,\, 1/\lambda_j^{(t)}\right).$$

DNRR model: Algorithm (II)

8. Initialize the vector of initial conditions $(y_{1:n})^{(t)}$ of the noise reduced trajectory to the previously sampled initial condition $(x_{1:n})^{(t)}$ of the x^n, and iterate for $j = 1, \ldots, n$ the following Metropolis-within-Gibbs sampling scheme:
   1. Generate proposal
   $$y_j^* \sim y_j^{(t-1)} + \nu \mathcal{N}(0,1). \tag{6}$$
   2. Calculate the acceptance probability $\alpha(y_j^{(t-1)}, y_j^*)$ given by
   $$\min\left\{1, \exp\left\{-\frac{1}{2}\left[C(y_j^*|\cdots) - C\left(y_j^{(t-1)}|\cdots\right)\right]\right\}\right\}.$$
   3. Accept $y_j^{(t)} = y_j^*$ with probability $\alpha(y_j^{(t-1)}, y_j^*)$.
9. Generate $\tau^{(t)} \sim \pi(\tau|\cdots)$.

Hénon map

We consider a time series realization x^n of size $n = 1,000$, coming from the random Hénon map:

$$x_i = 1.38 - x_{i-1}^2 + 0.27 x_{i-2} + e_i,$$

with $e_i \overset{\text{i.i.d.}}{\sim} f_{2,1}$, variance $\sigma^2 = 0.21 \times 10^{-4}$ and initial condition $x_0 = x_{-1} = 0.5$ for noise level at approximately 3%.
We model the deterministic part g, with the complete quadratic polynomial in the two variables, namely

$$g(\theta, x_{i-1}, x_{i-2}) = \theta_0 + \theta_1 x_{i-1} + \theta_2 x_{i-2} + \theta_3 x_{i-1} x_{i-2} + \theta_4 x_{i-1}^2 + \theta_5 x_{i-2}^2$$

The rms in this illustration is $\eta = 0.03$

Prior specifications

The prior distributions assigned were:

- $p \sim \mathscr{B}(0.5, 0.5)$
- $\lambda_j \sim \mathscr{G}(10^{-3}, 10^{-3})$
- $\pi(\theta) \propto 1$ and $\pi(x_{1:d}) \propto 1$
- $\tau \sim \mathscr{G}(10^4, 10^{-2})$, $\delta = \tau^{-1}$
- Proposal variance calibration

We ran the Gibbs sampler for $250,000$ iterations with a burn-in period of $50,000$ iterations.

Results I

After the implementation of the proposed DNRR model we:

- Reconstructed the dynamic equations that generated the data
- Estimated the density of the dynamical noise
- Estimated a noise-reduced orbit close to the original one

The average noise reduction achieved by the DNRR sampler is larger than two orders of magnitude, with $R_{\text{dyn}}(y^n, x^n; \hat{g}_{x^n}) = 0.902$, $E_{\text{dyn}}(y^n; \hat{g}_{x^n}) = 0.00286$ and $E_0(x^n, y^n) = 0.0428$.

Validation

In order to further validate the results of the DNRR model, we perform reconstruction on the estimated noise-reduced orbit $\boldsymbol{y} = \{y_1, \ldots, y_n\}$, using the GSBR model.

Results II

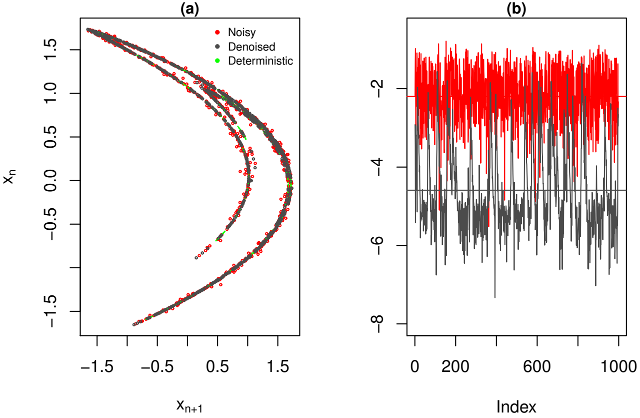

Figure 8: In figure (a), we present superimposed delay plots of the noisy, the noise reduced and the deterministic trajectories of the Heńon map, of length $n = 1,000$. The associated $\log_{10}$ −determinism plot is given in figure (b).

Noise density estimation

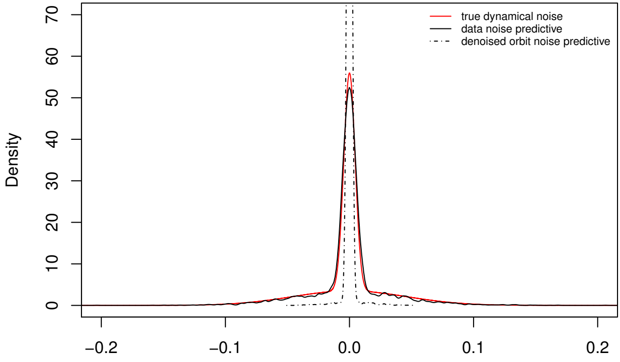

Figure 9: The true noise density $f = f_{2,1}$, for $\sigma^2 = 0.21 \times 10^{-4}$, is the red continuous curve. Along, we superimpose the x^n-estimated noise density $\hat{f}_{x^n}$ as a black continuous curve, and the y^n-estimated 'weaker' interactive noise density $\hat{f}_{y^n}$ as a black dashed curve.

Changing the proximity restriction

Table 5: Relative dynamical noise reductions, average indeterminisms and average distances, for two different values of ρ.

ρ	$E_{\text{dyn}}(x^n, \hat{g}_{x^n})$	$E_{\text{dyn}}(y^n, \hat{g}_{x^n})$	R_{dyn}	E_0
10^2	0.02932	0.00286	0.9023	0.0428
5×10^5	0.02932	0.00710	0.7577	0.0223

Table 6: PAREs for the estimated coefficients of the deterministic part of the perturbed Hénon map in (4), based on the noisy and the corresponding noise reduced trajectories, for two different values of ρ.

Time series	ρ	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	$\bar{\theta}$
x^n	10^2	0.089	0.096	0.046	0.044	0.011	0.070	0.059
y^n		0.063	0.043	0.022	0.028	0.020	0.038	0.036
x^n	5×10^5	0.079	0.071	0.041	0.031	0.002	0.059	0.047
y^n		0.177	0.155	0.015	0.023	0.005	0.157	0.089

The average distance E_0 as a function of ρ

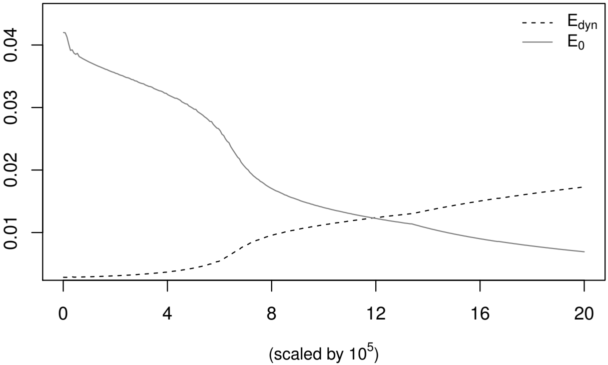

Figure 10: The average distance $E_0(y^n, x^n)$ and the average dynamic error $E_{\text{dyn}}(y^n, \hat{g}_{x^n})$ as functions of the parameter ρ.

Effect of (primary) homoclinic tangencies

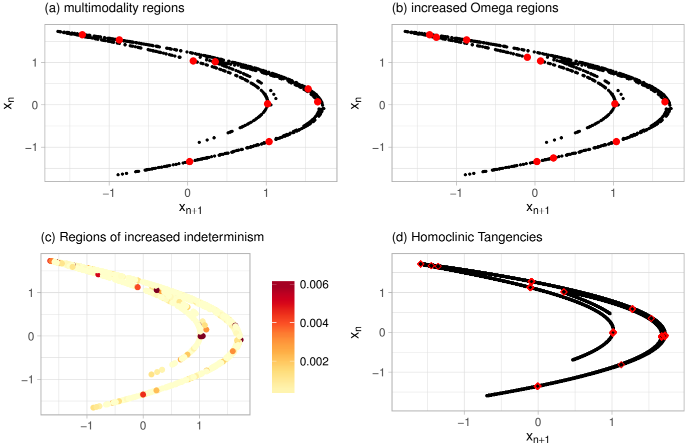

Figure 11: In Figure (a) we present a delay plot of the points in the set M_{HT} of the point estimators of the Y_i-posterior marginals, passing Hartigan's test for unimodality. In Figure (b) we depict the delay plot of the points in the set ˙$_{\mathrm{HT}}$ that are above the 99th percentile of the histogram of ˙. Regions of high E_{dyn} are depicted in Figure (c), and in Figure (d) we present the primary homoclinic tangencies of the corresponding deterministic attractor.

Fixed noise levels imply fixed relative noise reduction

Simulations using the $f_{2,l}$ noise formulation:

$$f_{2,l} = \frac{5+l}{10}\mathcal{N}(0,\sigma^2) + \frac{5-l}{10}\mathcal{N}\left(0,100\sigma^2\right)$$

We used dynamically noisy corrupted orbits of length $1,000$ with the same control parameters, initial conditions and prior specifications as above, using the DNRR model.

The variances of the noise processes, and each realization has been chosen, such that, η is fixed at about 3%.

We ran the chains for $250,000$ iterations with a burn-in period of length $50,000$. PAREs and measures of noise reduction efficiency are presented in Table 7.

Fixed noise levels imply fixed relative noise reduction

Table 7: Measures of reconstruction and noise reduction efficiency for the $f_{2,l}$ noise processes. The variances of the noise processes, and each realization has been chosen, such that, η is fixed at about 3%, where $E_{\text{dyn}} = E_{\text{dyn}}(y^n, \hat{g}_{x^n})$.

Noise	$\sigma^2 \times 10^4$	E_0	E_{dyn}	R_{dyn}	$\bar{\theta}_{x^n}$	$\bar{\theta}_{y^n}$
$f_{2,1}$	0.21	0.0428	0.00286	0.902	0.059	0.036
$f_{2,2}$	0.29	0.0514	0.00371	0.871	0.115	0.062
$f_{2,3}$	0.40	0.0490	0.00392	0.871	0.072	0.098
$f_{2,4}$	0.77	0.0627	0.00323	0.892	0.054	0.059

Random cubic map

Here, we consider the cubic map

$$x_i = g(\vartheta, x_{i-1}) = 0.05 + \vartheta x_{i-1} - 0.99 x_{i-1}^3$$

For $\vartheta \in \Theta_{\mathrm{bi}} = [1.27, 2.54]$ the map is bistable.

We let $\vartheta = \vartheta^* = 2.55$ and we consider the dynamically perturbed map $x_i = g(\vartheta^*, x_{i-1}) + e_i$ with $e_i \overset{\text{i.i.d.}}{\sim} f_{2,1}$, $\sigma^2 = 0.55 \times 10^{-4}$, and $\rho = 10^2$ (neutral proximity restriction)

- Noise-induced jumps
- Shadowing
- Modeling polynomial: $g(\theta, x_{i-1}) = \sum_{k=0}^{5} \theta_j x_{i-1}^k$

Results I

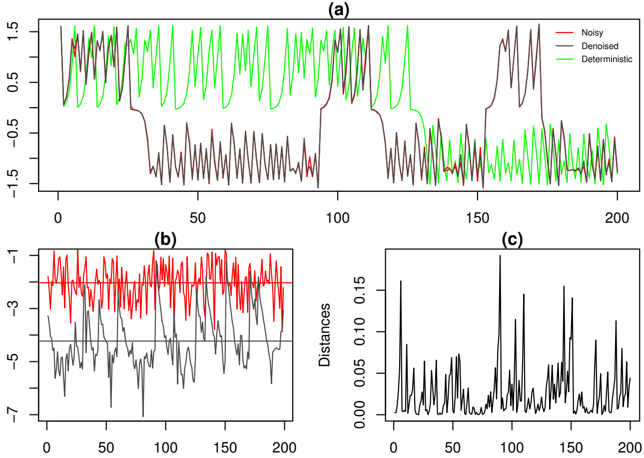

Figure 12: In (a), we give superimposed, the deterministic, the noisy (x^n) and the estimated (y^n) trajectories. In (b) we present the $\log_{10}$ indeterminism plot. The trace of the individual distances between the original and the noise reduced trajectory is given in (c).

Results II

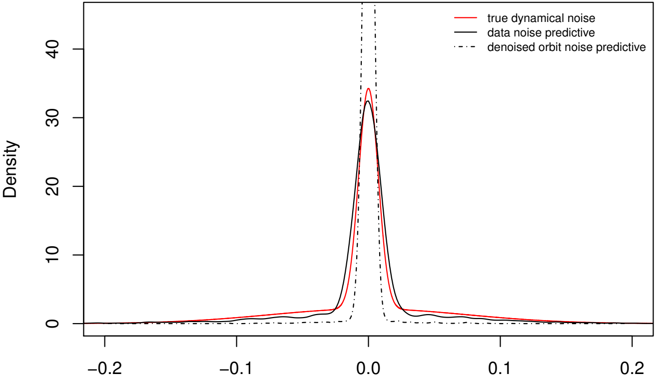

Figure 13: Kernel density estimations based on the predictive samples of $\hat{f}_{x^n}$ (continuous black curve), the predictive samples of $\hat{f}_{y^n}$ (dashed black curve) along with the true dynamical noise density (continuous red curve).

Results III

Table 8: Measures of reconstruction and noise reduction efficiency for the cubic map, for various σ^2's for the $f_{2,1}$ noise processes, where $E_{\text{dyn}} = E_{\text{dyn}}(y^n, \hat{g}_{x^n})$.

$\sigma^2 \times 10^4$	η %	E_0	E_{dyn}	R_{dyn}	$\bar{\theta}_{x^n}$	$\bar{\theta}_{y^n}$
0.33	3.5	0.0395	0.00749	0.812	0.281	0.425
0.55	4.5	0.0413	0.00695	0.842	0.605	0.804
0.59	5.5	0.0631	0.00952	0.826	0.438	0.262
0.67	6.5	0.0453	0.00847	0.848	0.872	0.958
1.00	7.5	0.0630	0.00819	0.867	0.856	0.987

Outline

Summary I

- Flexibility in modeling: RDS & BNP
- When dynamical noise departs from normality, simple MCMC methods are inefficient.
- Bayesian nonparametric framework allows us to drop the normality assumption, assign a nonparametric prior (DP/GSB) over the noise process.
- In terms of reconstruction and prediction, the proposed rDPR-GSBR models give satisfying results, with the GSBR model being faster and easier to implement.
- Even for short time series, the quasi-invariant density can be estimated and appears as natural prediction barrier.

Summary II

- We can use a similar framework with the DNRR model, in order to perform noise reduction.
- The denoised orbits obtained by the DNRR come from the same system as the data under “weaker” noise process and lower indeterminacy, while staying close to the original orbits.
- The proposed methods are robust under different noise tail fatness, formulated by the $f_{2,l}$-type noise processes. In fact, infinite mixtures of zero mean Gaussians, can mimic the effect of any heavy tailed symmetric noise processes, of finite or infinite kurtosis to an arbitrary level of accuracy.

Future Work I

Natural direction of future research include:

1. Perform reconstruction-prediction-noise reduction in cases where missing data are available.
2. Perform noise reduction in cases of asymmetric noise (α-stable processes) and state space models (Sequential Monte Carlo methods).
3. Drop the assumption of known functional form (Neural Networks/Gaussian Processes) and use the proposed models on real data.

Future Work II

4. Perform prediction on complex financial data using GSBR model - comparison.
5. Perform noise reduction on financial data using DNRR - interpretation (e.g. seasonal effects).
6. Extend MG-GARCH using Bayesian nonparametric modelling.

References I

McGoff K, Mukherjee S, Pillai NS (2012)
Statistical inference for dynamical systems: a review
Statistics Surveys

Antoniak C.E. (1974)
Mixture of Dirichlet processes with applications to Bayesian nonparametric problems
Annals of Statistics

Damien P, Wakefield J, Walker SG (1999)
Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables
Journal of the Royal Statistical Society (B)

Davies M (1997)
Nonlinear nose reduction through Monte Carlo sampling
Chaos

Ferguson, T. S. (1973)
A Bayesian analysis of some nonparametric problems
Annals of Statistics

Fuentes-García et al (2010)
A new Bayesian nonparametric mixture model
Communication in Statistics: Simulation and Computation

Jaeger L., Kantz, H. (1997)
Effective deterministic models for chaotic dynamics perturbed by noise
Physical Review E

References II

Hatjispyros SJ, Nicoleris T, Walker SG (2007)
Parameter estimation for random dynamical systems using slice sampling
Physica A

Hatjispyros SJ, Nicoleris T, Walker SG (2009)
A Bayesian nonparametric study of a dynamic nonlinear model
Computational Statistics and Data Analysis

Kaloudis K., Hatjispyros S.J. (2018)
A Bayesian nonparametric approach to dynamical noise reduction
Chaos

Merkatas C., Kaloudis K., Hatjispyros S.J. (2017)
A Bayesian nonparametric approach to reconstruction and prediction of random dynamical systems
Chaos

Kyrtsou C., Terraza M. (2003)
Is it Possible to Study Chaotic and ARCH Behaviour Jointly? Application of a Noisy Mackey-Glass Equation with Heteroskedastic Errors to the Paris Stock Exchange Returns Series
Computational Economics

Kyrtsou C., Karanasos M. (2008)
Analyzing the link between stock volatility and volume by a Mackey-Glass GARCH-type model: the case of Korea
Global Finance conference (Dublin, 2005)

Kyrtsou C. (2005)
Evidence for neglected nonlinearity in noisy chaotic models
International Journal of Bifurcation and Chaos

References III

Kyrtsou C. (2008)
Re-examining the sources of heteroskedasticity: The paradigm of noisy chaotic models
Physica A

Mackey, M. C. and Glass, L. (1977)
Oscillation and chaos in physiological control systems
Science

Glass, L. (2015)
Dynamical disease: Challenges for nonlinear dynamics and medicine
Chaos

Meyer R., Christensen N. (2000)
Bayesian reconstruction of chaotic dynamical systems
Physical Review E

Sethuraman, J. (1994)
A constructive definition of Dirichlet priors
Statistica Sinica

Walker SG (2007)
Sampling the Dirichlet Process with Slices
Communication in Statistics: Simulation and Computation

Thank you!

- **Acknowledgments**: "YPATIA" Fellowship Program for PhD Students, University of the Aegean