Analysis of bank leverage via dynamical systems and deep neural networks

Fabrizio Lillo* Giulia Livieri† Stefano Marmi‡
Anton Solomko§ Sandro Vaienti¶

4 August 2023

Abstract

We consider a model of a simple financial system consisting of a leveraged investor that invests in a risky asset and manages risk by using Value-at-Risk (VaR). The VaR is estimated by using past data via an adaptive expectation scheme. We show that the leverage dynamics can be described by a dynamical system of slow-fast type associated with a unimodal map on $[0, 1]$ with an additive heteroscedastic noise whose variance is related to the portfolio rebalancing frequency to target leverage. In absence of noise the model is purely deterministic and the parameter space splits in two regions: (i) a region with a globally attracting fixed point or a 2-cycle; (ii) a dynamical core region, where the map could exhibit chaotic behavior. Whenever the model is randomly perturbed, we prove the existence of a unique stationary density with bounded variation, the stochastic stability of the process and the almost certain existence and continuity of the Lyapunov exponent for the stationary measure. We then use deep neural networks to estimate map parameters from a short time series. Using this method, we estimate the model in a large dataset of US commercial banks over the period 2001–2014. We find that the parameters of a substantial fraction of banks lie in the dynamical core, and their leverage time series are consistent with a chaotic behavior. We also present evidence that the time series of the leverage of large banks tend to exhibit chaoticity more frequently than those of small banks.

Keywords: leverage cycles, risk management, systemic risk, random dynamical systems, unimodal maps, Lyapunov exponents, neural networks.

2020 Mathematics Subject Classication: primary 91G80, 34F05; secondary 37H15, 62M45.

*Dipartimento di Matematica, Universitá di Bologna and Scuola Normale Superiore, Pisa, Italy. Email address: fabrizio.lillo@unibo.it

†Scuola Normale Superiore, Pisa, Italy. Email address: giulia.livieri@sns.it

‡Scuola Normale Superiore, Pisa, Italy. Email address: stefano.marmi@sns.it

§Scuola Normale Superiore, Pisa, Italy. Email address: solomko.anton@gmail.com

¶Aix Marseille Université, Université de Toulon, CNRS, CPT, 13009 Marseille, France. Email address: vaienti@cpt.univ-mrs.fr

Contents

1 Introduction

Leverage is one of the most critical and controversial concepts in finance. On one side, borrowing is essential in many economic activities, while, on the other, it is intrinsically connected with risk. The recent literature (see, e.g., [37, 2, 45, 3, 1, 61]) has highlighted, both theoretically and empirically, an essential source of risk deriving from leverage, namely its procyclical nature. The seminal papers [2] and [3] argued that when assets are evaluated at mark-to-market, an increase in market prices of assets decreases the so-called "quasi-market leverage ratio" –, roughly the ratio of total assets to equity capital – and this leaves room to build up debt for banks that operate through leverage or Value-at-Risk (VaR) constraints[1]. Banks typically use the additional debt to expand the asset side of the balance sheet by purchasing more assets and, due to finite liquidity, this leads to a further increase in prices. This positive feedback between prices and balance sheets provides the additional source of systemic risk associated with procyclicality. As an empirical support, [2] and [3] shows that cycles of expansion (contraction) in the banks' balance sheet size go hand in hand with increases (decreases) in leverage, a behavior that has been witnessed since the 1960s but exacerbated during the 2007–2009 financial crisis. The creation of negative externalities in financial systems populated by VaR constrained financial institutions following standard mark-to-market and risk management rules has been investigated in e.g., [62, 31, 52]. These works confirm that the balance sheet dynamics of financial intermediaries are far from being passive and exogenous and can create market instability and result in what has been called by [32] *endogenous risk*.

This risk sometimes manifests itself in an abrupt and violent manner, via the so-called *fire sales*. A sudden drop in prices leads to a revaluation of the asset side of the balance sheet, and when the VaR constraint is violated, banks must deleverage by massively selling part of their portfolio. Due to the finite liquidity, this leads to a further decrease of prices affecting other banks' balance sheets. This mechanism creates an exceptionally threatening environment if many banks hold similar positions and use the same VaR model to manage their risk since they are forced to sell the same assets contemporaneously (overlapping portfolio contagion), leading to a destabilizing spiral. The theoretical and empirical literature on fire sales and their impact on the market price dynamics is vast and growing, see e.g.,[50, 22, 43, 28, 23, 29, 33].

Investors' expectations provide a further positive feedback between past and

[1]Here, VaR is the loss in market value of the bank's portfolio over one period that is exceeded with probability $1-\alpha$, where α is the associated confidence level. VaR constraints were imposed on the banks' trading book under Basel II. However, many other leveraged institutions, not subject to Basel II, also used VaR constraints in their internal risk management.

future risk. To implement the VaR constraint (and any risk management mechanism), banks must estimate the riskiness of the investments and their dependencies. The estimation of the *expected* risk is usually performed by using historical data; thus when past volatility is high, the VaR constraint becomes more binding, forcing financial institutions to trade more, increasing the future volatility endogenously. This additional feedback adds new threats to the systemic stability of financial markets, as shown, for example, by [55].

Our contribution. In order to fully understand the leverage dynamics and its role in financial systemic risk, it is paramount to model the effect of leverage targeting, risk expectations, asset prices, and of all the above-described feedbacks. In this paper, we present and analyze a stylized yet realistic model, and we show that leverage dynamics can be described by a discrete-time slow-fast dynamical system ([36] and [16]) with heteroscedastic noise. This fact makes our model amenable to the tools of dynamical systems theory. It allows us to fully characterize in a mathematically rigorous way the stability properties of the financial system depending on its parameters. We then employ a methodology based on deep neural networks to estimate the model's parameters on empirical data. Finally, we use such a methodology to estimate the model in a large dataset of US commercial banks over 2001–2014, finding that the leverage of large banks tends to exhibit chaotic behavior more frequently than small banks.

More precisely, building on [30] and [55], we consider a representative investor (bank for short), which evaluates the VaR capital requirement at discrete points in time $t \in \mathbb{N}$. This evaluation leads the bank to choose the maximum possible leverage, which is inversely proportional to the portfolio's expected risk. At each t the bank computes the expected volatility by using a straightforward yet realistic backward-looking method using past returns and uses it to set its desired leverage. The unitary time scale of the interval used to make decisions on leverage is the slow component of our model. During the unitary time interval $(t, t+1]$, the bank rebalances its portfolio to target the leverage without changing the risk expectations. The rebalancing occurs in $\mathtt{n}$ time sub-intervals within $(t, t+1]$. The time scale $1/\mathtt{n}$, with $\mathtt{n} \in \mathbb{N}$ characterizes the so-called fast component of the model. In particular, slow variables evolve as a function of averages over fast variables. Starting from this model, we make the following contributions.

First, we show that a deterministic unimodal map perturbed with an additive and heteroscedastic noise describes the dynamics of leverage in our model on the unit interval. The variance of the noise is related to the frequency of portfolio rebalancing $\mathtt{n}$ to target leverage. The parameter space of the deterministic map has two regions: (i) a region where the map has a globally attracting fixed point or a 2-cycle; (ii) the so-called *dynamical core* region, where the map can exhibit chaotic behavior[2]. Then, in order to improve the understanding of

[2]Roughly speaking, a dynamical system is defined chaotic if small changes in initial conditions produce large changes in long-term outcomes and any auto-correlation function decays

the anatomy of leverage cycles, we consider a more general class of maps and describe the leverage dynamic using a Markov chain parametrized by the *rebalance time* $\mathbb{n}$ and we study the regime of finite $\mathbb{n}$, as well as the limit for $\mathbb{n} \to \infty$. The stability of Markov chains is a relatively well studied topic (see e.g., [5] or [57] and references therein). However, some specific properties of the stochastic kernel that defines our model do not allow us to apply general results available [14, 15] (e.g., we do not know if our chain is Harris recurrent) and therefore we introduce an original approach to this type maps; see [40] for another type of heteroscedastic nonlinear autoregressive process applied to financial time series. In particular, we prove the existence of many stationary measures with bounded variation densities. The proof hinges on the spectral properties of the Markov operator associated with our chain on suitable Banach spaces and the quasi-compactness of such an operator. The stationary measure's uniqueness is achieved when the chain perturbs the unimodal map, which is either topologically transitive or admits an attracting periodic orbit; such maps correspond to a major part of the parameter space.

Second, in order to rigorously establish the existence of chaotic behavior in leverage time series of banks and to detect it in financial datasets, we define an average Lyapunov[3] exponent by integrating the logarithm of the derivative of the unimodal map with respect to the stationary measure, a definition that is suitable for the Markov chain approach. We show that the average Lyapunov exponent still allows us to discriminate periodic and chaotic behaviors: it is negative when we perturb a contracting map (and then the realizations of the process fluctuate around the fixed point), and it becomes positive by perturbing the dynamical core region. We also show that the average Lyapunov exponent depends continuously on the Markov chain parameters and relate it to the different chaotic behavior of the unperturbed unimodal map. Finally, we show the weak convergence of the unique stationary measure to the invariant measure of the unimodal map; namely we prove the weak stochastic stability of the system.

The third contribution of the paper is the estimation of the proposed model to real data to identify evidences of chaotic behavior. Our dataset consists of the time series of leverage of about $5,000$ US commercial banks in the period 2001-2014. The time series are very short (59 observations), thus standard estimation methods perform poorly. For instance, maximum likelihood estimation is not feasible for two reasons.First, the likelihood function is highly non-convex, so that standard optimization methods may perform poorly on short time series. Second, although the likelihood function for the process itself can be written explicitly, it may happen that in many cases we observe only a certain iterate of the process, e.g., we observe only one slow time scale portfolio decision event out of two. Therefore, we propose to use a Convolutional Neural Network (CNN) ([51]) to estimate the parameters of the map. Our CNN takes as input the one-dimensional time-series and gives the map's corresponding parameters as output. We train the CNN via extensive simulations of the model, considering

to zero, resulting in unpredictability of the system dynamics in the long-run ([34])

[3]We remind that for deterministic systems, the Lyapunov exponents characterize the divergence of nearby orbits, allowing us to distinguish between regular and chaotic dynamics.

different regions of the parameter space, and test its robustness and effectiveness [4]. Remarkably, we find that the parameters of a sizable fraction of banks lie in the map's dynamical core and that the large banks' leverage tends to be more chaotic than one of the small ones. As a robustness check, the identification of chaotic/periodic behavior is compared with a non-parametric approach where the map is not specified and estimators of indicators (such as the Lyapunov exponent [73]), which assume different values in the two regimes, are used to classify finite length time series. We use the Chaos Decision Tree Algorithm [71], which combines several tools into an automated processing pipeline that can detect the presence (or absence) of chaos in noisy recordings, even for difficult edge cases. When applying the Chaos Decision Tree Algorithm to our data set., the results corroborate the CNN approach's findings concerning the chaotic behavior for a significant subset of typically large banks.

Related literature. The present paper aims to combine several strands of literature.

From a methodological point of view, it is naturally related to the literature on the application of dynamical systems theory to the problem of systemic risk in financial markets [26, 27, 63, 25, 11], while, from the financial perspective, it is related to the literature on the impact of risk management on the leverage cycle [2, 3, 68]. These two strands were partially connected, for instance, in [10, 11, 30, 55] which study the impact of risk management practice on the dynamical properties of leverage cycles and how leverages cycles might be controlled. Similarly to these works, we develop a fully dynamic model of endogenous leverage cycles, and we show that the endogenous dynamics induced by leverage management and historical risk estimation can be very rich (i.e., one can observe dynamics ranging from stable fixed points to chaos), but, contrary to them, we study in a mathematically rigorous way the properties of the resulting model. The approach we proposed also differs from [39] who shows the existence of leverage cycles but uses a two-period general equilibrium model. Finally, we mention the works of [9] and [24] which show that the deleveraging of banks may amplify asset return shocks and lead to large fluctuations in realized returns which in turn can cause spillover effects between different assets. Differently from these papers, we explicitly model the existence of an additional feedback driven by the estimation of risk used in leverage decision.

Our work is also partially related to the literature on fire sales and their impact on the market price dynamics, e.g., [50, 22, 43, 28, 23, 29, 33]. The main differences are the following. First, in our model we do not consider "extreme" events leading to fire sale spillovers, but "normal" market conditions. Second, the previous works do not focus on the role of risk expectation feedbacks in financial systems. Instead, our model quantifies the impact of a possible mismatch between perception and reality of market condition due to the historical estimation of banks' perceived portfolio risk.

[4] We decide to use CNN because when employed for time-series analysis they may increase the accuracy up to 30% and train the models twice faster than other algorithms such as Recurrent Neural Networks and Long Short Term Memory ([64]).

In addition, the present paper has a point of comparison with the agent-based model of leveraged investors in [63] and [70]. The main difference is that in their model the bank is a dummy agent with infinite capital whose only role is to provide credit to funds. In contrast, in our model the banks are the key strategic agents. Again, another important difference is the mathematically rigorous approach that we adopt.

Finally, our estimation method connects with the vast literature on parameter estimation of dynamical systems via NN such as, for example, [58, 53, 65] which employ multi-layer NN and recurrent networks to identify and control nonlinear deterministic dynamical systems.

Outline of the paper. In Section 2 we present the financial model of a representative bank managing its leverage. In Section 3 we recall some facts about unimodal maps and Markov chains and then define the class of chains that we study. We also represent our model in terms of random transformations. In Section 4 we show the existence and uniqueness of an absolutely continuous stationary measure and establish its convergence to the invariant measure of the deterministic map. This allows us to define the Lyapunov exponent and prove its continuity with respect to the model parameters. We also discuss chaotic indicators naturally arising from the random maps representation of the process. The last part of the paper concerns numerical and empirical analyses. Specifically, Section 5 presents some numerical investigations of the bifurcation diagram and Lyapunov exponent of the map. Section 6 proposes an estimation method of the map based on the use of deep neural networks and Section 7 presents an empirical application to a large set of leverage time series of US banks, showing evidence of chaotic behavior. Finally, in Section 8 we draw some conclusions and outline some potential extensions of our work.

2 From the structural model to the dynamical system

The stylized model of the leverage dynamics we are going to present is a special case of the model of [55] (which in turn builds on [30]) restricted to the case of a single (representative) financial institution and of a single investment asset. Specifically, we consider a representative financial institution (a bank, hereafter) taking investment decisions at discrete times $t \in \mathbb{N}$, which defines the slow time scale. At each time t, the balance sheet of the financial institution is endowed with an amount of equity E_t and of asset A_t; the leverage is defined as $\lambda_t := A_t/E_t$. Financial institutions are confronted with a VaR type of constraints, in line with the literature on bank behaviour ([2, 21]), given by $\text{VaR} = \alpha\, \sigma_{e,t}\, A_t \leq E_t$. The scaling constant α depends on the return distribution[5] and on the VaR constraint, whereas $\sigma_{e,t}$ is the expected volatility at time t of the portfolio, which in this simple model is composed by a risky representative investment. Thus, at each time t the bank (re-)computes $\sigma_{e,t}$ and chooses the optimal leverage,

[5] For example, if returns are Gaussian and the probability of VaR is 5%, it is $\alpha = 1.64$.

which results to be inversely related to the volatility $\sigma_{e,t}$ and given by:

$$\lambda_t = \frac{1}{\alpha \sigma_{e,t}}. \tag{1}$$

Then, in the interval $[t, t+1]$ the bank trades the risky investment to keep the leverage close to the above defined target (1). The trading process occurs on the points of a grid obtained by subdividing $[t, t+1]$ in $\mathtt{n}$ sub-intervals of length $1/\mathtt{n}$ (the fast time scale). The dynamics of the investment return is made of two components:

$$r_{t+k/\mathtt{n}} = \varepsilon_{t+k/\mathtt{n}} + e_{t+(k-1)/\mathtt{n}}, \quad k = 1, 2, \ldots, \mathtt{n}, \tag{2}$$

the exogenous component $\varepsilon_{t+k/\mathtt{n}}$ coming from external shocks and the endogenous component $e_{t+(k-1)/\mathtt{n}}$. The former is a white noise term with variance σ_ϵ^2, while the latter depends on the banks' demand for the risky investment in the previous step. Having identified its optimal leverage, the bank adjusts the demand for the risky investment at time $t + k/\mathtt{n}$ by computing the difference between the desired amount of asset A_t^* to reach λ_t and the actual one A_t. Because the bank's asset is composed by the risky investment, an investment return $r_{t+k/\mathtt{n}}$ modifies A_t and the bank trades at each grid point to reach the target leverage. In order to achieve this, at each time $t+k/\mathtt{n}$ the bank's demand for the risky investment is

$$D_{t+k/\mathtt{n}} = (\lambda_t - 1) A^*_{t+(k-1)/\mathtt{n}} r_{t+k/\mathtt{n}}, \tag{3}$$

where $A^*_{t+(k-1)}$ is the target asset size in the previous step; see [30, 55] for an explicit derivation of (3). In particular, any profit and loss from investments in the chosen risky asset (which is given by $A^*_{t+(k-1)/\mathtt{n}} r_{t+k/\mathtt{n}}$) will directly result in a change in the risky asset value amplified by the current degree of leverage (being the leverage greater than one). Assuming a linear price impact function, the endogenous component of the return $r_{t+k/\mathtt{n}}$ is given by:

$$e_{t+k/\mathtt{n}} = \frac{1}{\gamma} \frac{D_{t+k/\mathtt{n}}}{C_{t+k/\mathtt{n}}},$$

where $C_{t+k/\mathtt{n}} = A^*_{t+(k-1)/\mathtt{n}}$ is a proxy of the market capitalization of the risky asset, and γ is a parameter expressing the market liquidity of the investment. Using the expression in Equation (3) for $D_{t+k/\mathtt{n}}$, the endogenous component is given by

$$e_{t+k/\mathtt{n}} = \frac{\lambda_t - 1}{\gamma} e_{t+(k-1)/\mathtt{n}} := \phi_t e_{t+(k-1)/\mathtt{n}},$$

where we set $\phi_t := \frac{\lambda_t - 1}{\gamma}$. Therefore, in the period $[t, t+1]$ the return $r_{t+k/\mathtt{n}}$ follows an AR(1) process with autoregression parameter ϕ_t and idiosyncratic variance σ_ϵ^2.

Finally, to close the model, we specify how the bank forms expectations $\sigma_{e,t}$

on future volatility at time t. Here we assume, as in [55], that the representative bank uses adaptive risk expectations, which implies that

$$\sigma^2_{e,t} = \omega\sigma^2_{e,t-1} + (1-\omega)\hat{\sigma}^2_{e,t},$$

where $\omega \in [0,1]$ is a parameter weighting between the expectation at $t-1$ and the estimation $\hat{\sigma}^2_{e,t}$ of volatility obtained by the return data in $[t-1,t]$. As done in practice (e.g., [44]), we assume that the representative bank estimates the sample variance of the returns in $[t-1,t]$ to compute $\hat{\sigma}^2_{e,t}$, i.e.

$$\begin{aligned}\hat{\sigma}^2_{e,t} &= \widehat{\mathrm{Var}}\left[\sum_{k=1}^{\mathtt{n}} r_{t-1+k/\mathtt{n}}\right] \\ &= \left(1 + 2\frac{\hat{\phi}_{t-1}(1-\hat{\phi}^{\mathtt{n}}_{t-1})}{1-\hat{\phi}_{t-1}} - 2\frac{(\mathtt{n}\hat{\phi}_{t-1} - \mathtt{n} - 1)\hat{\phi}^{\mathtt{n}+1}_{t-1} + \hat{\phi}_{t-1}}{\mathtt{n}(1-\hat{\phi}_{t-1})^2}\right)\frac{\mathtt{n}\hat{\sigma}^2_\epsilon}{1-\hat{\phi}^2_{t-1}}, \quad (4)\end{aligned}$$

where the last expression gives the aggregated variance of an AR(1) process as a function of the AR estimated parameters $\hat{\phi}_{t-1}$ and $\hat{\sigma}^2_\epsilon$. In particular, the estimator $\sigma^2_{e,t}$ can be seen as a stochastic term depending on λ_{t-1} and whose variance goes to zero when $\mathtt{n} \to \infty$. In the following, we will assume that these are the Maximum Likelihood (ML) estimators. We remind that when $\mathtt{n}$ is large, $\hat{\phi}_{t-1}$ is a Gaussian distributed variable with mean ϕ_{t-1} and variance $(1-\phi^2_{t-1})/\mathtt{n}$.

In conclusion, the following system of equations describes the leverage dynamics:

$$\begin{cases} \lambda_t = \left(\omega\frac{1}{\lambda^2_{t-1}} + (1-\omega)\alpha^2\hat{\sigma}^2_{e,t}\right)^{-1/2}, \\ r_s = \phi_{t-1}r_{s-1/\mathtt{n}} + \epsilon_s, \qquad s = t-1+k/\mathtt{n}, \quad k = 1,2,\ldots,\mathtt{n}, \end{cases} \quad (5)$$

In the first equation, the variable λ_t defines the slow variable and describes the slow component of the dynamics. In the second, the return evolution of the risky investment r_s describes the fast component of the dynamics. Since slow variables evolve depending on the averages of the fast variables, the model is a slow-fast deterministic-random dynamical system. We now derive an expression for the variable ϕ_t valid when $\mathtt{n}$ is large. For these values of $\mathtt{n}$, the map for λ_t reduces to

$$\lambda_t = \left(\omega\frac{1}{\lambda^2_{t-1}} + \frac{(1-\omega)\alpha^2\mathtt{n}\hat{\sigma}^2_\epsilon}{(1-\hat{\phi}_{t-1})^2}\right)^{-1/2},$$

and so, using the relation $\phi_t = \frac{\lambda_t - 1}{\gamma}$, the desired map becomes

$$\phi_t = -\frac{1}{\gamma} + \frac{1}{\gamma}\left(\frac{\omega}{(1+\gamma\phi_{t-1})^2} + \frac{(1-\omega)\alpha^2\mathtt{n}\hat{\sigma}^2_\epsilon}{(1-\hat{\phi}_{t-1})^2}\right)^{-1/2}. \quad (6)$$

At this point, a few remarks are in order. When changing $\mathtt{n}$ also σ^2_ϵ changes, since the AR(1) can be seen as the discretization of a continuous time stochastic

process (namely an Ornstein-Uhlenbeck process). A simple scaling argument shows that the quantity $\Sigma_\epsilon = \sigma_\epsilon^2 \mathtt{n}$ is instead constant and independent from the discretization step $1/\mathtt{n}$. In the limit $\mathtt{n} \to \infty$, it is $\hat{\phi}_t \to \phi_t$, thus the map in Equation (6) becomes purely deterministic[6]. The map in this case has a fixed point $\phi^* = \frac{1-\alpha\sqrt{\Sigma_\epsilon}}{1+\alpha\gamma\sqrt{\Sigma_\epsilon}}$. By replacing this condition in (6) and assuming that the risky asset is very liquid ($\gamma \gg 1$), the map becomes

$$\phi_t \simeq \left(\frac{\omega}{\phi_{t-1}^2} + \left(\frac{1-\phi^*}{\phi^*} \right)^2 \frac{(1-\omega)}{(1-\hat{\phi}_{t-1})^2} \right)^{-1/2}. \tag{7}$$

In the large $\mathtt{n}$ limit the ML estimator $\hat{\phi}_{t-1}$ is a Gaussian variable with mean ϕ_{t-1} and variance $(1-\phi_{t-1}^2)/\mathtt{n}$. Therefore, it holds that

$$\hat{\phi}_{t-1} = \phi_{t-1} + \eta_{t-1},$$

where $\eta_{t-1} \sim \mathcal{N}(0, (1-\phi_{t-1}^2)/\mathtt{n})$. If the noise η_{t-1} is small (i.e. $\mathtt{n}$ is large), we can perform a series expansion, obtaining

$$\phi_t \simeq \frac{|\phi_{t-1}(1-\phi_{t-1})|}{\sqrt{b\phi_{t-1}^2 + \omega(1-\phi_{t-1})^2}}(1+\zeta_{t-1}),$$

where $b = b(\omega, \phi^*)$ is given by

$$b = (1-\omega)\left(\frac{1-\phi^*}{\phi^*} \right)^2 \tag{8}$$

and the noise term

$$\zeta_{t-1} = \frac{-b\phi_{t-1}^2}{(1-\phi_{t-1})(b\phi_{t-1}^2 + \omega(1-\phi_{t-1})^2)}\eta_{t-1}.$$

Finally, in this approximation, the map becomes:

$$\phi_{t+1} = T(\phi_t; \theta) + \sigma(\phi_t; \theta)\epsilon_t, \tag{9}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$ and θ is a vector of parameters. In our setting $\theta = (b, \omega, \mathtt{n})$ and the deterministic map T does not depend on $\mathtt{n}$. Specifically,

$$T(\phi_t; \theta) = \frac{|\phi_t(1-\phi_t)|}{\sqrt{b\phi_t^2 + \omega(1-\phi_t)^2}} \tag{10}$$

and

$$\sigma(\phi_t; \theta) = \frac{b\phi_t^3\sqrt{1-\phi_t^2}}{\sqrt{\mathtt{n}}\left(b\phi_t^2 + \omega(1-\phi_t)^2\right)^{3/2}}. \tag{11}$$

In particular, the process ϕ_t is constructed by perturbing with a heteroscedastic additive noise $\sigma(\phi_t; \theta)\epsilon_t$ the deterministic map $T(\phi_t; \theta)$ of the unit interval

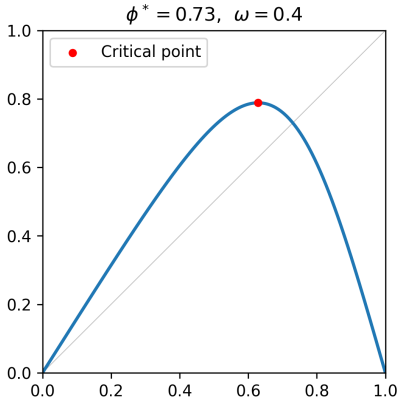


Figure 1: Plot of the deterministic component $T(\phi)$, $\phi^* = 0.73$, $\omega = 0.4$ ($b = 0.082$).

$I = [0, 1]$. Fig 1 provides a pictorial representation of the map T. Notice that the process ϕ_t is a (continuous state) Markov chain since the distribution of ϕ_t only depends on ϕ_{t-1}: this is a simple yet crucial observation that will be heavily used in the next section.

As said in the Introduction, one of the purposes of the present paper is to rigorously establish the possibility of chaotic behavior in leverage time series of banks and to detect it in financial datasets. To this end, in the next sections we develop a rigorous mathematical theory of additive unimodal maps with heteroscedastic noise as in Equation (9), with (10) being its representative. This theory will allow us to study, e.g., the existence of a stationary measure, the stochastic stability, and the Lyapunov exponent for this class of models. It is worth noting that our results remain valid for any noise ϵ_t in (9), not only Gaussian.

3 The mathematical model

We first describe the class of unimodal maps with which we will work (Subsection 3.1). Then, we define the Markov chain describing our model (Subsections 3.2), and we show that our leverage model can be described in terms of this Markov chain (Subsection 3.3). Finally, we present a different yet equivalent approach based on random transformations to define our model (Subsection 3.4).

[6] This is the deterministic skeleton, whose properties are discussed in detail in [55].

3.1 Unimodal maps

We consider unimodal maps [15] $T\colon I \to I$ of class at least C^3 with $T(0) = T(1) = 0$ and with a non-degenerate critical point[7] at c: $T'(c) = 0$. The map T is strictly increasing on $[0, c)$ and strictly decreasing on $(c, 1]$. Moreover, we suppose that T satisfies the following assumptions:

(A1) T has negative Schwarzian derivative: $S(T) := \dfrac{T'''}{T'} - \dfrac{3}{2}\left(\dfrac{T''}{T'}\right)^2 < 0.$

(A2) the maximum $\Delta := T(c) < 1$,

(A3) the critical point is *quadratic*: $T''(c) \neq 0$.

The map T arising from our financial model (10) satisfies the above assumptions, as we verified numerically.

In order to establish the possibility of chaotic behavior in leverage time series, we partition the parameter space for the map (10) by using the theory of unimodal maps (see, for instance, the review [69]):

(C1) If $\Delta \leq c$, then a globally attractive fixed point exists.

(C2) If $c < T(\Delta) < \Delta$, then there is a globally attracting fixed point or a 2-cycle in (c, Δ).

(C3) If $T(\Delta) < c < \Delta$, we can reduce the study to the so-called *dynamical core* $[T(\Delta), \Delta]$, which is mapped onto itself and absorbs all initial conditions (except 0 which is a fixed point).

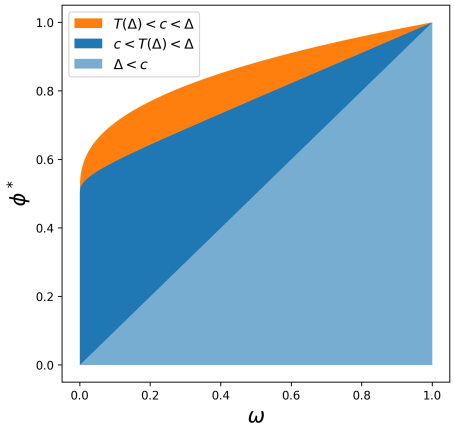


Figure 2: Partition of the parameter space for the unimodal map (10) according to the classification (C1)–(C3).

[7] The critical point for (10) is $c = \left(1 + \sqrt[3]{b/\omega}\right)^{-1}$.

Figure 2 reports a pictorial representation of the partition of the parameter space according to the classification (C1)–(C3). Following this classification, we will say that:

(1) T is **periodic** if there is a globally attracting fixed point or a globally attracting cycle.

(2) T is **chaotic** if T is topologically transitive[8] on the interval $[T(\Delta), \Delta]$ and T admits a unique invariant Borel probability measure which is absolutely continuous with respect to the Lebesgue measure [9].

Subsection A.1 of Appendix A collects additional important results on the class of unimodal maps presented in this section. We will refer to these results in the rest of the paper. Notice that there are unimodal maps that are neither periodic nor chaotic, as follows from Theorem A.1.

3.2 Coupling with a stochastic process

We define now the continuous state Markov chain $\{X_t : t \in \mathbb{N}\}$ describing our model. For the reader's convenience, Subsection A.2 of Appendix A recalls some basic, yet important, properties of such chains. The chain $\{X_t : t \in \mathbb{N}\}$ is obtained as a deterministic unimodal map $T\colon I \to I$ satisfying (A1)–(A3), coupled with a stochastic process, namely, by perturbing T with an additive noise. Starting from (9) as our main motivation, on the one hand, we consider a more general class, and on the other, we impose some mild technical restrictions that are necessary for rigorous analysis.

Since the noise varies in a neighborhood of 0, we will need to extend the state space on the negative axes. However, we will see in a moment that such an extension is irrelevant for the asymptotic behavior of the perturbed system, whose random trajectories spend all the time, but a relatively short transient, on the positive unit interval.

We fix T and parametrize the chain by the *rebalance time* $\mathtt{n}$ (which is roughly inverse to the variance of the noise), consequently indexing with $\mathtt{n}$ the chain $(X_t^{(\mathtt{n})})$, the transition probabilities $P_x^{(\mathtt{n})}$ and the stochastic kernel $p_{\mathtt{n}}(x, y)$ where it is necessary. We will be interested in the limit for $\mathtt{n} \to \infty$.

We also need to assume that the noise is compactly supported, so trajectories of the process stay bounded. Compared to the Gaussian noise in (9), this is done by truncating the distribution tails that are exponentially small for large $\mathtt{n}$, see Section 3.3 for the main example.

Denote by $\Gamma := 1 - \Delta$ the gap between $T(c)$ and 1. We now extend the domain of definition of T to the larger interval $[-\Gamma, 1]$ (which, by abuse of notation, will be still denoted by I) so that T is continuous at 0 and on $[-\Gamma, 0)$

[8] A map T on a topological space X is called *topologically transitive* if for all nonempty open sets $U, V \subset X$, there exists n such that $T^{-n}U \cap V \neq \emptyset$.

[9] Notice that, since T satisfies (C3), $T(x) > x$ for any $x < T(\Delta)$, in particular, $T'(0) > 1$.

is C^4 smooth, positive and decreasing, with $T(-\Gamma) < \Delta$.[10]

To construct the chain, we need to define transition probabilities. Let $g_{x,\mathtt{n}}(y)$ be a probability density supported on a compact interval $[-s(x), s(x)]$. We assume that

- $0 < s(x) < \Gamma/2$ for $x \in (0,1)$,
- $T(x) - s(x) > 0$ for $x \in (0, 1-\Gamma/2]$,
- $T(x) - s(x) > x$ for x small (in particular, $T'(0) > 1$).

We set for simplicity $s(x) := 0$ for $x \leq 0$ (meaning $P_x^{(\mathtt{n})} = \delta_{T(x)}$); Lemma B.4 shows that this choice does not affect the dynamics. We will also assume that both the mean and the variance of $g_{x,\mathtt{n}}$ decrease to 0 as $\mathtt{n} \to \infty$ and, for every $\varepsilon > 0$,

$$\sup_{x \in [\varepsilon, 1-\varepsilon]} |g_{x,\mathtt{n}}|_{TV} < \infty. \tag{12}$$

Fix any initial distribution $\rho_0 \in \mathrm{BV}$ and define transition probabilities

$$P_x^{(\mathtt{n})}(A) := \int_{-s(x)}^{s(x)} \mathbb{1}_A(T(x)+y) g_{x,\mathtt{n}}(y) dy, \tag{13}$$

which correspond to the stochastic kernel

$$p_{\mathtt{n}}(x,y) = g_{x,\mathtt{n}}(y - T(x)). \tag{14}$$

Informally speaking, the probability that the chain steps from x to A will be high whenever $T(x)$ falls in A. Equivalently, we can write

$$X_{t+1}^{(\mathtt{n})} = T(X_t^{(\mathtt{n})}) + Y_{t+1}, \quad \text{where} \quad Y_{t+1} \sim g_{x,\mathtt{n}}. \tag{15}$$

The values of $X_{t+1}^{(\mathtt{n})}$ are spread in a neighborhood of $T(x)$ due to the addition of the random variable Y_{t+1}.

3.3 The leverage model

Now, we will slightly modify (9) to satisfy the technical assumptions listed above. The unimodal map is

$$T(x) = \frac{|x(1-x)|}{\sqrt{bx^2 + \omega(1-x)^2}}, \tag{16}$$

[10] A similar extension was considered in [15] to allow perturbations with additive noise; in particular, it was supposed that $T(I) \subset \mathrm{int}(I)$ and that T admits an extension to some compact interval $J \supset I$, preserving all the previous properties and satisfying $T(\partial J) \subset \partial J$. Notice that with this extensions the map T is not any more of class C^4 as prescribed in Appendix A.1. However, this regularity persists on the interval $(0,1)$ and this will be enough for the subsequent considerations.

where the parameters ϕ^* and ω are such that T satisfies assumptions (A1)–(A3). It always has a negative Schwarzian derivative, as we verified numerically, and the critical point is quadratic. The condition $T(\Delta) < c < \Delta < 1$ defines a nonempty subset of parameters (see Fig. 2). Notice that $T'(0) = 1/\sqrt{\omega} > 1$.

We want $g_{x,\mathtt{n}}$ to be (truncated) normal with a variance close to (11). For this, let us denote by $\mathcal{N}_a(0,\sigma)$ the smoothed truncated normal distribution with the density $g(y) = c_{a,\sigma}\chi_a(y)e^{-\frac{y^2}{2\sigma^2}}$, where $c_{a,\sigma}$ is so that $\int g(y)dy = 1$ and χ_a is a smooth bump function supported on $[-a,a]$.[11] For instance, we may set

$$\chi_a(y) = \begin{cases} 1, & \text{if } |y| \le (1-\varepsilon)a, \\ \Psi(\frac{y \pm (1-\varepsilon)a}{\varepsilon a}), & \text{if } (1-\varepsilon)a < |y| \le a, \\ 0, & \text{if } |y| > a, \end{cases}$$

where $\Psi(t) = e^{1-\frac{1}{1-t^2}}$ is the standard C^∞ bump function on $[-1,1]$. We set

$$\sigma_{\mathtt{n}}(x) := \frac{bx^3\sqrt{1-x^2}}{\sqrt{\mathtt{n}}\big(bx^2+\omega(1-x)^2\big)^{3/2}}. \tag{17}$$

Denote $\sigma(x) := \sigma_1(x)$ and $\sigma_{\max} := \max_{x\in[0,1]}\sigma(x)$. Set

$$g_{x,\mathtt{n}}(y) := c_{x,\mathtt{n}}\chi_{s(x)}(y)e^{-\frac{y^2}{2\sigma^2_{\mathtt{n}}(x)}}, \tag{18}$$

where

$$s(x) := \frac{\sigma(x)}{\sigma_{\max}}\min\Big\{\frac{\Gamma}{2}, T(1-\frac{\Gamma}{2})\Big\} \quad \text{and} \quad c_{x,\mathtt{n}} := \left(\int_{-s(x)}^{s(x)} \chi_{s(x)}(y)e^{-\frac{y^2}{2\sigma^2_{\mathtt{n}}(x)}}dy\right)^{-1}.$$

We can then rewrite (15) as

$$X^{(\mathtt{n})}_{t+1} = T(X^{(\mathtt{n})}_t) + \sigma_{\mathtt{n}}(X_t)Z_{t+1}, \quad Z_{t+1} \sim \mathcal{N}_{b_{\mathtt{n}}}(0,1), \tag{19}$$

with $b_{\mathtt{n}} := \frac{\sqrt{\mathtt{n}}}{\sigma_{\max}}\min\{\frac{\Gamma}{2}, T(1-\frac{\Gamma}{2})\} \to \infty$ as $\mathtt{n}\to\infty$. The fact that for $\mathtt{n}$ fixed, $g_{x,\mathtt{n}}$ are rescaled copies of the same distribution will be used in Section 3.4 to explicitly describe random maps associated to the process. We get the following stochastic kernel:

$$p_{\mathtt{n}}(x,y) = c_{x,\mathtt{n}}\chi_{s(x)}(y-T(x))e^{-\frac{(y-T(x))^2}{2\sigma^2_{\mathtt{n}}(x)}}. \tag{20}$$

Notice also that the support of $p_{\mathtt{n}}(x,y)$ does not depend on $\mathtt{n}$ (see Fig. 3).

Finally, (12) holds, because $|g_{x,\mathtt{n}}|_{TV} = 2c_{x,\mathtt{n}}$ and the latter is proportional to $1/\sigma_{\mathtt{n}}(x)$, which is bounded on $[\varepsilon, 1-\varepsilon]$.

In particular, the evolution of our system/model is given by the Markov chain $\{X_t : t \in \mathbb{N}\}$, which produces a perturbation affecting the deterministic map

[11] The smoothness of the truncation function is only used in the proof of Theorem 4.12.

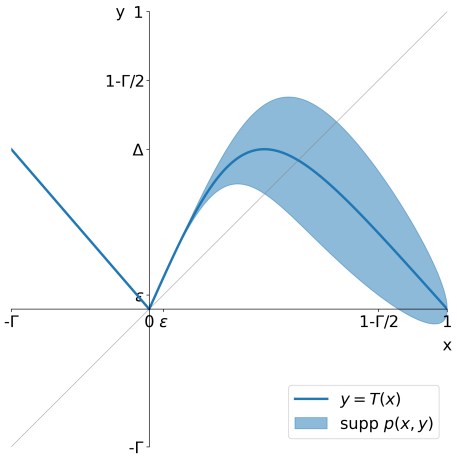


Figure 3: Support of the kernel (20).

T at each step. In the next subsection, we proceed differently by replacing the deterministic orbit $T^n(x), x \in I$, with a composition of (possibly different) maps T_k close to T and chosen in an i.i.d. way, which is called *random transformations*. Importantly, we will provide an explicit construction of the maps T_k together with their distribution, which will be crucial in proving the weak stochastic stability of our system.

3.4 Random transformations

Our model was defined as a Markov chain. We now present a slightly different, yet equivalent, point of view. Namely, we will pick up a family of maps $T_\eta \colon I \to I$, $\eta \in [0,1]$, in such a way that

$$\mathrm{Leb}\{\eta : T_\eta(x) \in A\} = P_x(A) \tag{21}$$

for all $A \subset I$. We can then define a stochastic process

$$\bar{x}_{t+1} = T_{\eta_{t+1}}(\bar{x}_t), \tag{22}$$

where η_t are independent and uniformly distributed in $[0,1]$. We can write $\bar{x}_t = T_{\eta_t} \circ \cdots \circ T_{\eta_1} \bar{x}_0$, where $(\eta_t)_{t\in\mathbb{N}}$ is an i.i.d. stochastic process, i.e. the process (22) follows the orbits under the concatenation of randomly chosen maps from the family. One can show that the two processes are equivalent, see, for instance, [47]. Conversely, starting with a family of maps T_η, one can use (21) to define transition probabilities P_x and thus a Markov chain.

Rewriting (21) as $P_x(A) = \int_0^1 \mathbb{1}_A(T_\eta(x))d\eta$ and plugging into (31) we get the *disintegration formula* for the Markov operator:

$$\mathcal{L} = \int L_\eta d\eta, \tag{23}$$

where L_η are the transfer operators associated with T_η. In particular, a measure μ is stationary for the Markov chain if and only if it satisfies $\mu = \int L_\eta \mu d\eta$, i.e., for all $A \subset I$,

$$\mu(A) = \int \mu(T_\eta^{-1} A) d\eta. \tag{24}$$

Equation (24) is usually taken as the definition of a stationary measure for the family of random maps. Every such measure corresponds to a product measure that is invariant for the skew-product with the Bernoulli shift in the base and the maps T_η in the fibers; we refer to [4, Section 2] for details.

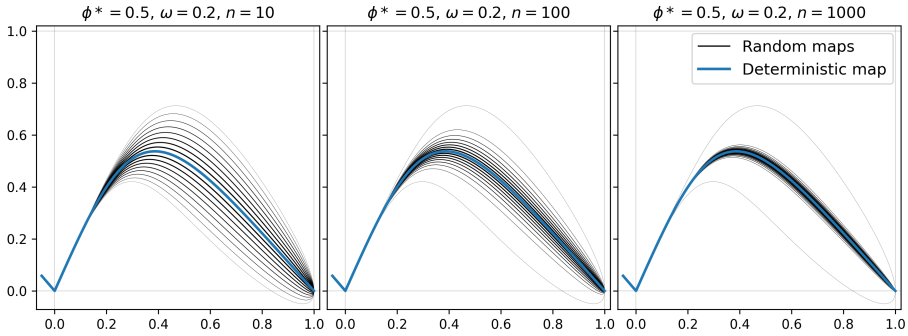


Figure 4: Random maps (25) for $\phi^* = 0.5$, $\omega = 0.2$, $\mathtt{n} = 10, 100, 1000$, $\eta = k/16$.

As follows from (19), the random maps for the main example have the form

$$T_\eta(x) = T(x) + q_{\mathtt{n}}(\eta)\sigma_{\mathtt{n}}(x), \tag{25}$$

where $q_{\mathtt{n}}$ is the quantile function of the truncated normal distribution $\mathcal{N}_{b_{\mathtt{n}}}(0,1)$. Indeed, since $q_{\mathtt{n}}$ maps the uniform measure on $[0,1]$ to the truncated standard Gaussian measure on $[-b_{\mathtt{n}}, b_{\mathtt{n}}]$, we have

$$\begin{aligned} P_x(A) =& \mathbb{P}\{T(x) + \sigma_{\mathtt{n}}(x) Z_t \in A\} \\ =& \mathrm{Leb}\{\eta : T(x) + \sigma_{\mathtt{n}}(x) q_{\mathtt{n}}(\eta) \in A\} \\ =& \mathrm{Leb}\{\eta : T_\eta(x) \in A\}. \end{aligned}$$

We can equivalently rewrite (25) as

$$T_\eta(x) = T(x) + \tilde{q}_{\mathtt{n}}(\eta)\sigma(x), \tag{26}$$

where $\tilde{q}_{\mathtt{n}}$ is the quantile function of $\mathcal{N}_{b_1}(0, \frac{1}{\mathtt{n}})$. Notice that for different $\mathtt{n}$, the set $\{T_\eta\}_{\eta\in[0,1]}$ consists of the same maps, however the ones close to $T = T_{\frac{1}{2}}$ are given bigger weights for large $\mathtt{n}$. More precisely, for every $\delta > 0$, we have

$$\sup_{\eta\in[\delta,1-\delta]} \sup_{x\in I} |T_\eta(x) - T(x)| \le \sigma_{\max} \sup_{\eta\in[\delta,1-\delta]} |\tilde{q}_{\mathtt{n}}(\eta)| \xrightarrow[\mathtt{n}\to\infty]{} 0. \tag{27}$$

4 Theoretical study of the mathematical model

We first establish the existence of a unique stationary measure for the Markov chain defined in Subsection 3.2 (Subsection 4.1). Then, we prove its weak stochastic stability (Subsection 4.2). Finally, we prove the existence and some regularity properties of the Lyapunov exponent for the slow component of our dynamical system (Subsection 4.3). In order to not weigh down the reading of this section, we decide to confine all the proofs in Appendix B.

4.1 Existence and uniqueness of the stationary measure

We establish the existence of a unique stationary measure for the chain. We here only mention the techniques used in the proofs. The existence of stationary measures will be accomplished in the following steps: we first prove the Lasota-Yorke inequality; as a consequence, we will get a finite number of ergodic absolutely continuous stationary measures whose supports are mutually disjoint up to sets of zero Lebesgue measure. The uniqueness will be proved by showing that all the previous components share a measurable set of positive Lebesgue measure.

Before stating the existence Theorem 4.1, we need to introduce the following notation. For any $\varepsilon > 0$, we define the interval $I_{\varepsilon,\Gamma} := [\varepsilon, 1 - \Gamma/2]$.

Theorem 4.1. *The Markov chain defined in Section 3.2 admits finitely many ergodic stationary measures with densities in BV, the space of bounded variation function on the unit interval equipped with the complete norm $\|f\|_{BV} = |f|_{TV} + \|f\|_1$; see Appendix A, Subsection A.1. Moreover, there is $\varepsilon > 0$ such that $\operatorname{supp}\mu \subset I_{\varepsilon,\Gamma}$ for any such measure μ.*

Proof. See Appendix B, Subsection B.1. □

Remark 4.2. It is worth noticing that the prior result is entirely independent of the structure of the unimodal map T. In this respect, we could consider maps admitting attracting periodic points or Cantor sets of measure zero but still producing smooth stationary measures when perturbed with our additive noise.

We now state the main result of this section.

Theorem 4.3. *If T is either periodic or chaotic, then the Markov chain defined in Section 3.2 admits a unique stationary measure μ with BV density. Moreover, $\operatorname{supp}\mu$ contains a neighbourhood of the periodic cycle if T is periodic, or the interval $[T(\Delta), \Delta]$ if T is chaotic.*

Proof. See Appendix B, Subsection B.2. □

We conclude this subsection with the following consideration. From a financial point of view, should the stationary measure not be unique, it would imply that, depending on the initial conditions, different banks could experience completely different dynamics corresponding to different stationary measures.

When this occurs in physical systems, one speaks of phase transitions and coexistence of different mutually singular states. In our case, this would imply, for instance, that policy measures could not be universal.

4.2 Stochastic stability

Once we consider random perturbations of a deterministic dynamics, an important question is to investigate the stochastic stability of the system, which means to determine if a sequence of stationary measure will converge, in a sense to be precised, to the invariant measure of the unperturbed map. In our case, the sequence of probability measures is given by $\mu_{\mathbb{n}} := h_{\mathbb{n}} dx$. These measures belong to the set of Borel probability measures on the unit interval, which is a compact metric space with the weak-* topology[12]. There will be therefore at least one subsequence $(\mu_{\mathbb{n}_k})_{k\geq 1}$ converging to a probability measure μ_∞ on I. Our objective is to prove that: (i) μ_∞ is invariant, (ii) it is the same for any convergent subsequence, if more than one, and (iii) it coincides with μ. Whenever that happens, we will say that our random system is *weakly stochastic stable*. This result could be strengthened by showing that $\|h_{\mathbb{n}} - h\|_1 \to 0$, which is called the *strong stochastic stability*. However, we are not able to obtain this result. Instead, we now give a *sufficient condition* to get the weak stochastic stability:

(A_q) *There exist* $q > 1$ *and* $C_q > 0$ *such that for all* $\mathbb{n} \geq 1$ *we have* $\|h_{\mathbb{n}}\|_q \leq C_q$.

We will see in the next subsection that with the preceding assumption, we can prove the convergence of the Lyapunov exponent (Theorem 4.10) and then verify it numerically, which is an indirect indication of the validity of (A_q).

Theorem 4.4. *Under Assumption* (A_q) *and when the map* T *is chaotic, the Markov chain defined in Section 3.2 is weakly stochastic stable, i.e., the stationary probabilities converge to the unique* T*-invariant probability in the weak-* topology as* $\mathbb{n} \to \infty$.

Proof. See Appendix B, Subsection B.3 □

It follows from Proposition B.8 that Assumption (A_q) cannot be satisfied in the periodic case, since the limiting T-invariant measure is singular and supported on the periodic orbit, so Theorem 4.4 only covers the chaotic case. We will now give proof in the periodic case under the following assumption:

(A_s) *For all* $\mathbb{n}$ *sufficiently large and all* $x \in \operatorname{supp} \mu_{\mathbb{n}}$ *we have* $|T'(x)| \leq \tau < 1$.

Proposition 4.5. *If* T *is periodic and satisfies* (A_s)*, then the Markov chain defined in Section 3.3 is weakly stochastic stable.*

Proof. See Appendix B, Subsection B.4. □

[12]The *weak-* topology* is given by the family of seminorms $\|\rho\|_\varphi = \int \varphi d\rho$, $\rho \in \mathcal{M}$, $\varphi \in C^0$.

Remark 4.6. It follows from the proof that Proposition 4.5 remains valid for the general class of Markov chains defined in Section 3.2 whenever (27) holds, which is in turn the case when T and $g_{x,\mathbb{n}}$ are sufficiently smooth.

Remark 4.7. We conjecture that, if T is periodic with the attracting periodic orbit $\mathcal{O}$, then $h_{\mathbb{n}} \to 0$ uniformly on compact sets $K \subset I \setminus \mathcal{O}$ as $\mathbb{n} \to \infty$. This property, which we checked numerically, straightens the previous result. In particular,

(A_c) *If T is periodic and the critical point c does not belong to the attracting periodic orbit, then $h_{\mathbb{n}} \to 0$ uniformly in a neighbourhood of c as $\mathbb{n} \to \infty$.*

4.3 Lyapunov exponent

4.3.1 Average Lyapunov exponent

We are interested in the existence of the *Lyapunov exponent* for the slow component, which in our case is defined $\mathbb{P}_\mu$-almost surely as the limit

$$\Lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \log |T'(X_t)| \tag{28}$$

along the chain $(X_t)_{t\geq 0}$. We provide two motivations for such a choice. First, we want to reproduce the Lyapunov exponent of the unperturbed map T in the limit of zero noise, which we will get in Theorem 4.10; Second, we want an indicator that kept the memory of the underlying *slow dynamics* played by the map T. We will, in particular, show that such an exponent is negative for periodic T, even in the presence of mixing stationary measures.

We now return to (28); if the chain admits a unique stationary probability μ, then, by the ergodic theorem for Markov chains, the above limit equals

$$\int \log |T'| d\mu, \tag{29}$$

assuming $\log |T'| \in L^1(\mu)$.

Remark 4.8. The Lyapunov exponent (29) was called the *average Lyapunov exponent* in [38, 59]. It was associated with the phenomenon of *noise-induced order*, which happens when the perturbed systems admit a unique stationary measure depending on some parameter, say θ, and the Lyapunov exponent depends continuously on θ and exhibits a transition from positive to negative values, see also [54] for experimental evidence of this fact. We will partially prove this phenomenon below by combining Corollary 4.11 and Theorem 4.12, and show it numerically in Section 5.2.

A unimodal map T is said to have a *critical point of order l* if there is a constant D such that $D^{-1}|x-c|^{l-1} \leq |T'(x)| \leq D|x-c|^{l-1}$. In this case it was proved in [60] that the invariant density for T is in L^q, with $q < \frac{l}{l-1}$. We will assume that T has a critical point of order 2. It is easy to check that (16) satisfies this assumption. Consequently, $\log |T'|$ is in L^p for any $p \geq 1$.

Theorem 4.9. *If T is periodic or chaotic, the limit* (28) *exists almost surely.*

Proof. The integral (29) is finite, because $\log|T'|$ is in L^1 and the unique stationary measure μ has bounded density, as we proved in Section 4.1. □

Once we know that the Lyapunov exponent exists almost surely, it is natural to ask how it depends on the model parameters, for instance, the length $\mathtt{n}$ of the fast component series. We have the following result:

Theorem 4.10. *Suppose one of the following is satisfied: (a) T verifies (A_q); (b) T is periodic and verifies (A_s) and (A_c). Then the Lyapunov exponent* (28) *converges to the Lyapunov exponent of the deterministic map T as $\mathtt{n} \to \infty$.*

Proof. See Appendix B, Subsection B.5. □

Corollary 4.11. *Under the assumptions of Theorem 4.10 and for $\mathtt{n}$ large enough, Λ is positive if T is chaotic, and negative if T is periodic.*

In some cases, the negativity of the Lyapunov exponent can be shown relatively easily, see Example B.9. We also provide some numerical examples in Section 5.2.

4.3.2 Continuity of the Lyapunov exponent

Denote by $\Theta := \{\theta = (\phi^*, \omega, \mathtt{n}) \in (0,1)^2 \times (0,\infty) \mid \max T_\theta < 1\}$ the (extended) parameter space. In order to prove the continuity of the Lyapunov exponent, we will assume that $T_\theta(x) \in C^3(\Theta \times [0,1])$ and $p_\theta(x,y) \in C^2(\Theta \times (0,1)^2)$. It is straightforward that our main example defined in Section 3.3 satisfies this assumption. Let $\tilde{\Theta} \subset \Theta$ be the set of parameters θ for which there is a unique stationary measure μ_θ with a density $h_\theta \in BV$; we proved in Section 4.1 that this is the case if T_θ is periodic or chaotic, but our numerical investigations confirm that in fact $\mathrm{Leb}(\Theta \setminus \tilde{\Theta}) = 0$.

Theorem 4.12. *The mapping $\tilde{\Theta} \ni \theta \mapsto \Lambda_\theta \in \mathbb{R}$ is continuous.*

Proof. See Appendix B, Subsection B.6. □

We conclude this section with the following two remarks.

Remark 4.13. In (28) we use the derivative of the deterministic map only to define the Lyapunov exponent. However, in Subsection 3.4 we define our process using the perspective of the random transformation. This leads naturally to define the Lyapunov exponent of the cocycle given by the derivative computed along the random orbit, named *Random Lyapunov Exponent (RLE)*. We analyze it in Appendix B, Subsection B.7. Also, We discuss the so-called *entropy formula* allowing us to equate the random Lyapunov exponent with the *random entropy*, which is the random generalization of the Kolmogorov-Sinai entropy. Finally, we describe the interesting situation of a stationary state with zero random entropy but which mixes exponentially fast.

Remark 4.14. For the stochastic stability and the Lyapunov exponent, we mainly discuss the relation between finite $\mathtt{n}$ (thus noisy system) and infinite $\mathtt{n}$ (purely deterministic system). Thus our results indicate in which sense what we learn for a noisy system is informative about the deterministic backbone. In our empirical analysis we will do not study or use directly neither the stationary measure nor the Lyapunov exponent (mainly because we have very short time series), however the "continuity" we observe from finite to infinite $\mathtt{n}$ suggests that the properties we observe empirically for finite $\mathtt{n}$ are informative of the underlying deterministic dynamics.

5 Numerical results

In this section, we describe and discuss some numerical experiments in support of our rigorous theoretical investigations. Specifically, we present the bifurcation diagram associated with the unimodal map (16) and compute the corresponding Lyapunov exponents for both the deterministic and the stochastic version of the map, see Subsections 5.1 and 5.2, respectively.[13]

5.1 Bifurcation diagram

In this subsection, we analyse the dynamics of the unimodal map (16). The bifurcation diagram of a dynamical system shows how a typical orbit's asymptotic distribution varies as a parameter's function. For our map, either the memory parameter ω or the parameter ϕ^* can be employed as the bifurcation parameter. Fig. 5 shows the bifurcation diagram as a function of ϕ^* (the bifurcation diagram as a function of ω looks similar). The choice of the parameters for this plot corresponds to a vertical segment in the parameter space (see Fig. 2) with $\omega = 0.5$ and ϕ^* varying in a neighbourhood of the dynamical core area. As explained in Appendix A.1, the invariant set of the unimodal map (16) could be an attracting periodic orbit, a Cantor set of measure zero, or a finite union of intervals with a dense orbit, depending on the parameters ω and ϕ^*. Specifically, there is an attracting fixed point or a 2-cycle outside the dynamical core region, whereas in the dynamical core, the situation is more complex as small parameter variations can change the dynamics from chaotic to periodic and back, as we see in Fig. 5. To identify the signature of a chaotic behavior more precisely, in the following subsection, we compute the Lyapunov exponent as a function of ϕ^*.

5.2 Lyapunov exponent

The Lyapunov exponent for the deterministic map (16) is positive if and only if T admits an absolutely continuous invariant measure, see Theorem A.3 in

[13]The code to reproduce all the figures and tables of this section are available from the corresponding author upon reasonable request.

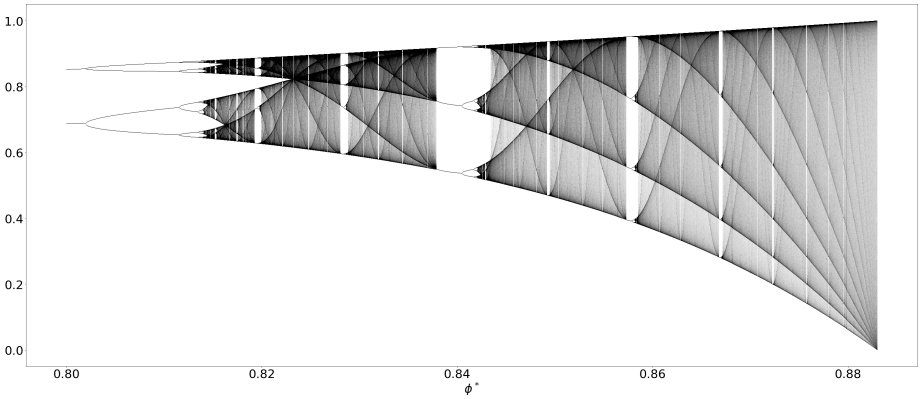


Figure 5: Bifurcation diagram for T in the dynamical core region ($\omega = 0.5$).

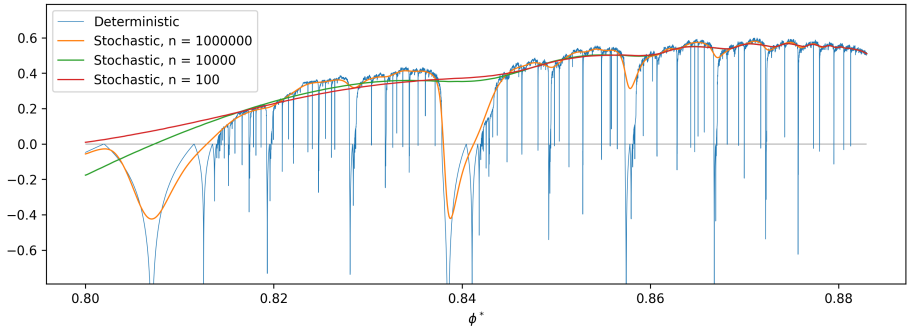


Figure 6: Lyapunov exponents for deterministic and stochastic maps ($\omega = 0.5$).

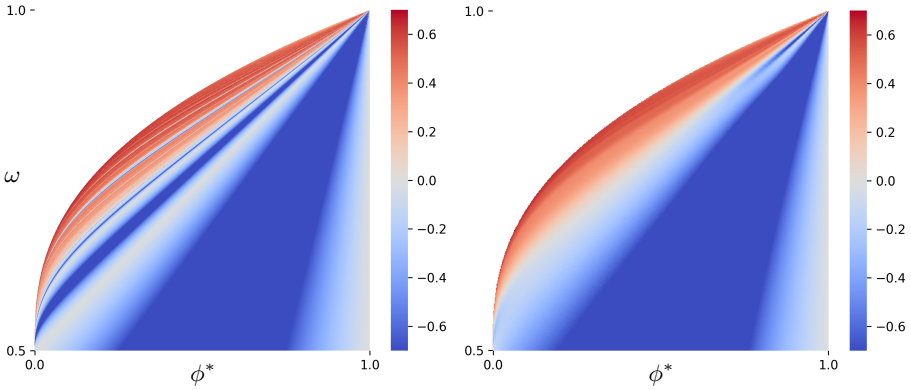


Figure 7: Contour plot of the average Lyapunov exponents for (a) deterministic and (b) stochastic maps ($\mathtt{n} = 1$).

Appendix A.1. Fig. 6 shows the estimated Lyapunov exponent in the same slice of the parameter space as in Fig. 5. Notice that the exponent becomes a smooth function of ϕ^* when we add even a very small noise, in agreement with the results of Section 4.3.2. Fig. 7 shows the contour plot of the Lyapunov exponent as a function of ϕ^* and ω both for the deterministic map (16) and for the stochastic process described in Section 3.3. The right panel shows a clear noise-induced regularization phenomenon. In fact, for the stochastic version of the map, the intricate fine structure in the parameter dependence of the Lyapunov exponent disappears and is replaced by a smooth dependence.

To provide a numerical exemplification of the stochastic stability and of Theorem 4.10, we computed average (28) and random (44) Lyapunov exponents, as well as the Lyapunov exponent for the deterministic map (16), for different values of ϕ^*, ω and $\mathtt{n}$. The results are presented in Table 1. Within each row, the two sub rows are ALE and RLE, respectively: the two agree very well, in most cases, up to the precision of the numerical computation. In both cases, we sampled 128 independent realizations of the process, each of length of $10,000$.

ϕ^*	ω	DC	Per.	Lyapunov exponent				
				$\mathtt{n}$				Det.
				1	10^3	10^6	10^9	
0.845	0.557	yes	no	0.287 0.286	0.287 0.286	0.349 0.349	0.341 0.341	0.340
0.795	0.390	yes	no	0.345 0.345	0.346 0.346	0.389 0.389	0.398 0.399	0.400
0.904	0.627	yes	no	0.557 0.558	0.557 0.557	0.560 0.560	0.550 0.550	0.552
0.821	0.439	yes	yes	0.378 0.378	0.375 0.375	−0.051 −0.052	−0.159 −0.159	−0.158
0.944	0.826	yes	yes	0.296 0.297	0.297 0.296	0.049 0.052	−0.123 −0.123	−0.122
0.766	0.323	yes	yes	0.320 0.320	0.324 0.325	0.286 0.286	−0.076 −0.076	−0.046
0.258	0.837	no	yes	−0.243 −0.286	−0.248 −0.248	−0.248 −0.248	−0.248 −0.248	−0.248
0.908	0.804	no	yes	−0.284 −0.286	−0.285 −0.287	−0.365 −0.366	−0.362 −0.362	−0.362
0.541	0.227	no	yes	−0.619 −0.578	−0.441 −0.446	−0.380 −0.380	−0.380 −0.380	−0.380

Table 1: Average and random Lyapunov exponents for different values of $\mathtt{n}$ compared to the Lyapunov exponent for the deterministic map. DC stands for dynamical core, Per. for periodic, and Det. for the Lyapunov exponent of the deterministic map.

6 Estimating the map parameters via deep neural networks

We now consider the problem of estimating the map's parameters from (short) time series. This fact is motivated by the fact that in Section 7 we empirically investigate a dataset of US commercial bank's leverage. We will consider the time series of the bank's leverage as realizations of the process described in Section 3.3 and will estimate for each bank the model parameters. Each time series is very short, composed of only 59 points.

Given the random nature of the map, one could use maximum likelihood estimation to estimate the parameters. However, this approach is not feasible for two reasons. First, the likelihood function is highly non-convex, so standard optimization methods may perform poorly. Second, although the likelihood function for the process itself can be written explicitly, in many cases, the observed time series are systematically undersampled, preventing an explicit calculation of the likelihood function. For example, we may observe only one slow time scale, corresponding to portfolio rebalancing, out of two, or even out of three (i.e. the bank's risk assessment and portfolio composition may be updated more frequently than our quarterly observations, for instance at a monthly frequency). If we observe, for instance, only the second iterate of the process,

$$\phi_{t+2} = T(T(\phi_t;\theta);\theta) + \sigma(\phi_t;\theta)\epsilon_t) + \sigma(\phi_{t+1};\theta)\epsilon_{t+1}, \quad t \in \mathbb{Z},$$

the transition probabilities $p(\phi_{t+2}|\phi_t;\theta)$, $t \in \mathbb{Z}$, are no longer Gaussian (as it would be the case if we observe the first iterate). Hence, there is no effective formula for the likelihood function.

For this reason, to estimate the parameters of the map, we propose to use a convolutional neural network (CNN) consisting of a sequence of convolutional layers followed by a sequence of dense, or fully connected, layers (see [41, Chapter 9] for more details). In order to deal with the possibility that the observed time series are realizations of certain iterates of the process, we separately optimize two CNN architectures to be used sequentially. First, we optimize a CNN (henceforth denoted by $\mathsf{CNN1}$) for estimating the number of iterates between observations: it takes as input time series of length 59 (as our empirical data) and outputs the corresponding value k of the map's iterate that generated the time-series. Second, for each value of k, we optimize a CNN (henceforth denoted by $\mathsf{CNN2}(k)$), having the same inputs, to output the corresponding parameters (ϕ^*, ω) that generated the time-series. Once the parameters (ϕ^*, ω) are estimated, the noise variance (and therefore $\mathtt{n}$) can be estimated by standard methods. To train $\mathsf{CNN1}$ and $\mathsf{CNN2}(k)$ we used a training set of one million samples simulated from the model in Section 3.3 with values of the parameters $\theta = (\phi^*, \omega, \mathtt{n})$ which uniformly span the parameter space. For both steps, when simulating the series, the system's initial state was taken randomly from a uniform distribution on $[0, 1]$. This fact is especially important because of the relatively short length of the series. Therefore, based only on simulations, the NN approach, contrary to the maximum likelihood one, can also work for

partial observations. The selected[14] CNN models show a good performance. We tested our methods on a testing set of 100,000 out of sample time series. Fig. 8 shows the accuracy of CNN1 to estimate the iterates on test data. We choose $k = 1, 2, 3$ because of our empirical application of Section 7. The MSE of CNN2(k) on the test set is about 0.001 for each k. Since both ϕ^* and ω are uniformly distributed in $[0, 1]$, the MSE is quite small and the NN effective.

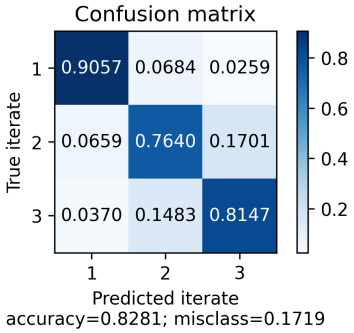


Figure 8: Accuracy of the CNN1 model used to estimate the iterates.

7 Chaos in real bank leverage time series

In this section, we perform an empirical analysis of a large set of banks' leverage time series. We first describe the data set. Next, through CNN1 and CNN2(k), we estimate the iterate k as well as the parameters (ϕ^*, ω) of the model and discuss the results, investigating the relationship between the estimated parameters and the bank's size. Finally, to perform an independent analysis supporting our conclusions, we apply the Chaos Decision Tree Algorithm [71] to these time series and compare the resulting classification with the one obtained with CNN estimates.

7.1 Data set

We use the data set of US *Commercial Banks and Savings and Loans Associations* provided by the *Federal Financial Institutions Examination Council* (FFIEC). For the sake of completeness, we provide a description of it, referring to [35] and references therein for further details. A Commercial Bank is defined[15] officially by the FFIEC as: *"[. . .] a financial institution that is owned by stockholders, operates for a profit, and engages in various lending activities"*. Commercial banks quarterly fill the *Consolidated Report of Condition and Income* (generally referred to as *Call Report*) as required by the FFIEC. A Savings

[14] Appendix C reports details on the selected CNN models, and on the procedure we employed to choose such models.

[15] See `http://www.ffiec.gov/nicSearch/FAQ/Glossary.html`.

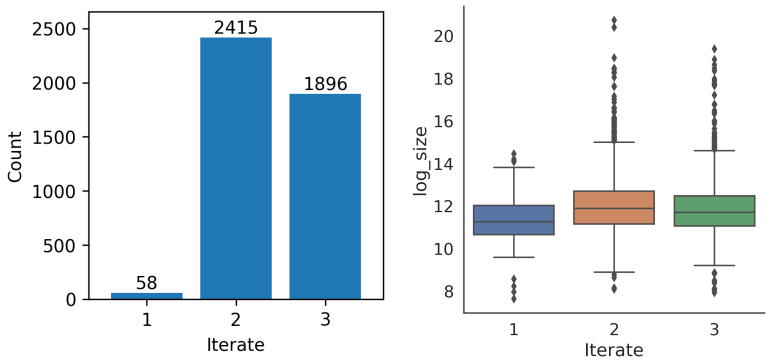
and Loan Associations, instead, is a financial institution that accepts deposits primarily from individuals and channels its funds primarily into residential mortgage loans. Starting from the first quarter of 2012 they are required to file the same reports of Commercial Banks; thus they have been included in the data set since then. The data provided by the *Call reports* have been publicly available since 1986, although the required details have increased over time. To have a good compromise between the fine structure of data and reasonably populated statistics, we follow [35] and consider the period time going from March 2001 to December 2014, for a total of 59 quarters. Also, we consider only the financial institutions present in the data set in all the quarters for a total of $5,031$ banks. The financial leverage λ_t of each institution at time t is defined as the ratio between the sum of its assets and its equity at time t. In particular, the latter is given by $E_t = A_t - L_t$ where L_t represents the liabilities and A_t the assets of the bank, thus $\lambda_t = A_t/E_t$.

7.2 Estimation via neural networks

In order to estimate the parameters of the map on the just-described data set, we need to fix the value of the liquidity parameter γ; remind that we consider the linear transformation $\phi_t = (\lambda_t - 1)/\gamma$. In this work, we assume that the liquidity parameter γ of the risk investment is the same for all the banks in our data set. Admittedly, this is a simplifying assumption, coherent with the so-called assumption of statistical equivalence for risky investments (see also [55]), which allows for an analytical tractability of the model; a complete exploration of a relaxation of this hypothesis is beyond the scopes of the present paper and is, therefore, left for future work. In order to fix its value, we exclude 662 time series (out of $5,031$) that contain outliers, which we define to be values that are two standard deviations away from the mean. We then set γ to the maximum over the remaining $4,369$ series, obtaining $\gamma = 15.969$.

Since the time series that we analyze contain quarterly data and portfolio decisions may be made more frequently, it is natural to assume that the observed time series are realizations of certain iterates of the process; we assume $k \in \{1, 2, 3\}$. Fig. 9a displays the output of $\mathsf{CNN1}$. It turns out that only a small percentage (about 1%) of the banks in our data set rebalance their portfolios at a quarterly frequency. Most banks seem to rebalance either every six weeks ($k = 2$, about 55%) or every month ($k = 3$, about 43%). One may ask if the portfolio re-balancing frequency is related to the size of the bank (defined as the average across the 59 quarters of the sum of the dollar-amount of all the types of assets detained by it), for example, because larger banks manage more actively their portfolio. Fig. 9b shows the box plots of the logarithm of the size of the banks for $k = 1, 2, 3$. We observe that there is not a statistically significant difference among them.

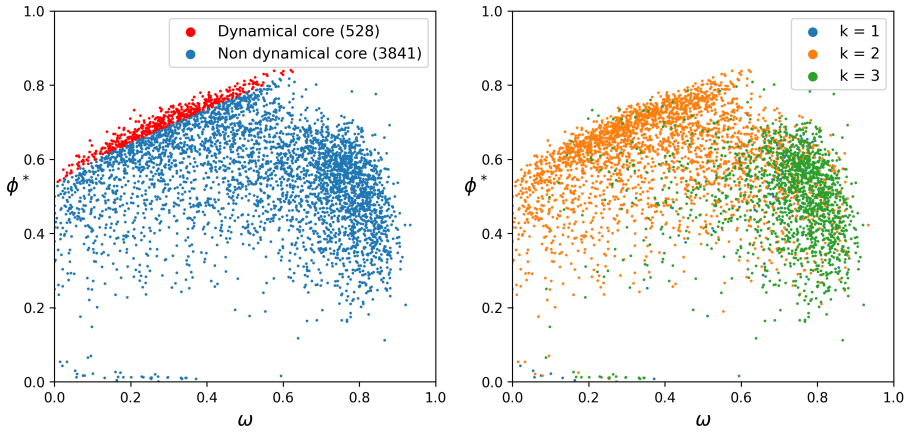
Once the number of iterates k has been identified, we proceed to the study of the chaotic behaviour of the time series. We divide the banks in the three groups identified by k and employ $\mathsf{CNN2}(k)$ in order to estimate the parameters (ϕ^*, ω). In Fig. 10a we plot the estimates of ϕ^* against those of ω; pairs belonging to



(a) Number of banks by iterate. *(b)* Log bank size by iterate.

Figure 9: Results on the iterate of the process.

the dynamical core region are displayed in red, whereas those falling outside the dynamical core region are displayed in blue. Interestingly, the percentage of banks for which the estimates (ϕ^*, ω) are in the dynamical core region is about 12%. Moreover, Fig. 10b indicates that k is very often equal to two for these banks.



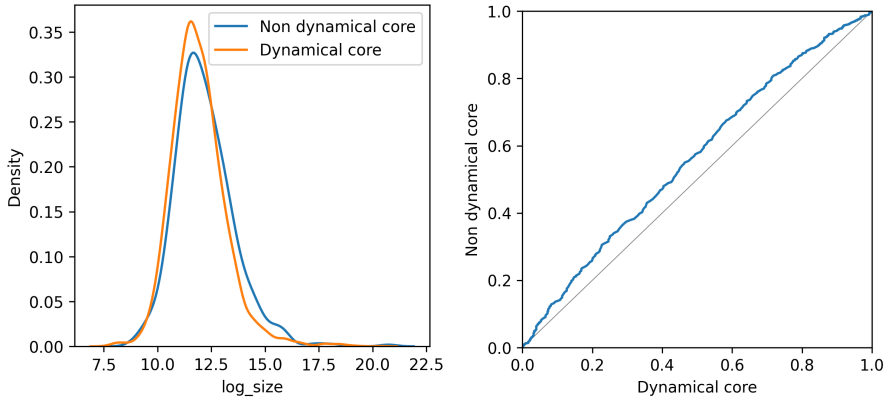
(a) Parameters distribution by DC. *(b)* Parameters distribution by iterate.

Figure 10: Estimated parameters for iterates in $\{1, 2, 3\}$.

We now ask if bank size is related with the fact that the estimated pair (ϕ^*, ω) is or not in the dynamical core region. Fig. 11a shows the probability density functions of the logarithm of the banks size, by considering separately banks inside and outside the dynamical core, and Fig. 11b displays the corresponding probability-probability plot. To test that the difference between the distribution

of bank sizes in and outside the dynamical core region is statistically significant, we perform the Kolmogorov-Smirnov test of the null hypothesis that the two samples have the same distribution. The statistics of the test is 9.31×10^{-2} corresponding to a p-value of 5.9×10^{-4}. This latter value shows that the two subsamples have different distributions.

Summarizing, we have found that the parameters of a sizable fraction of banks lie in the dynamical core region and that the dynamics of the leverage of the larger banks tend to be more frequent in the dynamical core than that of the smaller banks.



(a) Probability density functions. *(b)* Probability-probability plot.

Figure 11: Results on the parameters of the process and bank size.

7.3 Classification via the Chaos Decision Tree Algorithm

Finally, we perform an independent analysis on the bank's leverage time series by making use of the recently proposed Chaos Decision Tree Algorithm (CDTA) [71], described in detail in Appendix F. This is a non-parametric method which classifies an input time series as chaotic, periodic, or stochastic[16]. We perform this analysis for two reasons. First, we know that chaotic behavior can be present only for series generated by our map with parameters in the dynamical core. Thus we test whether the series classified as chaotic by CDTA have estimated parameters in the dynamical core. The second reason is to count how many banks in the dynamical core are identified as chaotic or periodic by CDTA. The Appendix also contains the results of the application of CDTA to data simulated by our map for different time series length, level of noise $\mathbb{n}$, and number of iterates k.

Applying CDTA, we find that 64% of the banks are classified as stochastic, $\sim 12\%$ as periodic and $\sim 23\%$ as chaotic. The consistency between the classifi-

[16]Notice that the definitions of *chaotic* and *periodic* for the CDTA differ from the ones given in Subsection 3.1, see Appendix D.

cation made by CDTA and the partition 'dynamical core' and 'not dynamical core' of the parameters space (ϕ^*, ω) found by the NN model can be assessed by looking at Table 2. We find that a large fraction of series outside the dynamical core are classified as stochastic by CDTA, while a third of banks in the dynamical core is classified as chaotic. This fraction is significantly smaller for banks outside the dynamical core. Thus, despite the agreement is not perfect, we find a reasonable consistency between the conclusions of the two methods and, more importantly, find significant (and independent) support to the conclusion that a sizable fraction of bank time series are described by a chaotic dynamics.

	Periodic	Chaotic	Stochastic
Non dynamical core	382 (9.98%)	648 (16.93%)	2798 (73.09%)
Dynamical core	107 (20.34%)	176 (33.46%)	243 (46.20%)

Table 2: Number of banks by classes.

Finally, the findings reported in the previous subsection suggest a positive relation between the size of a bank and the probability that the dynamics of the (corresponding) leverage time series is chaotic. To verify this observation by using the CDTA classification, we first rank the banks in quintiles according to their size and within each quintiles we compute the percentage of banks that are detected to be stochastic, periodic and chaotic. Table 3 collects the results. In a nutshell, banks having a larger size have, on average, a larger percentage of leverage time series detected as chaotic. A χ^2-test applied to contiguous quintiles rejects the hypothesis of independence of the CDTA classification from the quintile, indicating that the difference in frequencies across quintiles are statistically significant. Thus also the CDTA analysis confirms that larger banks are more likely characterized by chaotic time series of leverage.

Statistics	q_1	q_2	q_3	q_4	q_5
Chaotic (%)	18.2	20.2	21.7	23.6	29.1
Periodic (%)	12.9	11.6	12.9	13	11.8
Stochastic (%)	68.6	67.3	65	63	58.9

Table 3: Fraction of banks classified as chaotic, periodic, or stochastic by CDTA conditionally to the decile of the bank size.

8 Conclusions

Most risk management practices (for example, Value-at-Risk) assume that prices are not affected by the actions of other financial institutions managing the risk of their portfolio. In other words, these practices assume that risk is exogenous. In reality, in the presence of limited liquidity, coordinated and homogeneous risk management can create market instability and result in what is known

as endogenous risk. This has the potential to amplify market instabilities and create crashes through the well-known feedback between leverage, risk, and asset prices. Additional and less considered feedback between past and future risks is present because financial institutions use historical data to estimate both the riskiness of their investments and their correlations. This creates new threats to the systemic stability of financial markets. Studying how these two feedbacks affect the leverage dynamics is paramount for understanding systemic risk.

In this paper, we consider a stylized model where both feedbacks are present. We showed that the dynamics of the bank's leverage are described by a unimodal map on $[0, 1]$ perturbed with additive and heteroscedastic noise. The perturbed system can be described in two equivalent ways as a stationary Markov chain or in terms of random transformations. In both cases, a fundamental object is the *stationary measure* of the process, which allow us to properly define and state all the statistical properties of the system. We are able to construct such a measure and prove its uniqueness. Moreover, under a few assumptions, we show the stochastic stability of the perturbed system, namely the weak convergence of the stationary measure to the invariant measure for the unimodal map in the zero noise limit. We also define an average Lyapunov exponent, still in terms of the stationary measure, as a sensitive indicator of slow motion and prove its continuity with respect to the parameters defining the system. We show that, depending on the parameters, the average Lyapunov exponent can be either negative or positive, leading to two qualitatively different (periodic- and chaotic-like) leverage dynamics.

We then estimate the map's parameters via a method based on deep neural networks, whose efficiency was tested in a large testing set. Assuming the proposed unimodal map with heteroscedastic noise as data generating process for the banks leverage, we estimated the parameters on quarterly data of about 5,000 US Commercial Banks via the proposes CNN architecture. By investigating the period from March 2001 to December 2014, for a total of 59 quarters, we found that the parameters of a sizable fraction of banks lie in the dynamical core region of the parameter space and that the large banks' leverage tends to be more chaotic than one of small ones. The latter finding was also corroborated by using a non-parametric approach.

We believe that the proposed methodologies may offer revealing perspectives for future works. For instance, it would be interesting to extend the employed mathematical techniques to study a model in which more than one asset and one bank are present in the system.

9 Acknowledgements

This research was supported by the research project 'Dynamics and Information Research Institute - Quantum Information, Quantum Technologies' within the agreement between UniCredit Bank and Scuola Normale Superiore. F.L. and S.M. acknowledge partial support by the European Program scheme 'INFRAIA-01-2018-2019: Research and Innovation action', grant agreement #871042 'So-

BigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics'. S.V. is grateful to the Centro di Ricerca Matematica Ennio de Giorgi, Scuola Normale Superiore, Laboratoire International Associé LIA LYSM, the INdAM (Italy) and the UMI-CNRS 3483, Laboratoire Fibonacci (Pisa), where this work has been initiated and completed under a CNRS delegation, for the support. S.V. thanks C. Gonzalez-Tokman for useful discussion about Subsection 3.4.

Appendix A: Additional mathematical results of Section 3

A.1: Unimodal maps

In this section, we remind some known results on the class of unimodal maps introduced by [15].

First, we remind the following theorem of Blokh and Lyubich (we quote the statement given in [69, Theorem 6]) on the structure of the invariant sets (attractors) of a unimodal map.

Theorem A.1 ([18])**.** Let $T : I \to I$ be an S-unimodal map with nonflat critical point c. Then T has a unique metric attractor A, such that the ω-limit set $\omega(x) = A$ for Lebesgue almost all $x \in I$. The attractor A is of one of the following types:

1. an attracting periodic orbit;
2. a Cantor set of measure zero;
3. a finite union of intervals with a dense orbit.

In the first two cases, $A = \omega(c)$.

Second, we remind that associated to T there is the *transfer operator* (also called the *Perron-Frobenius operator*) $L\colon L^1 \to L^1$ which is the positive linear operator defined by the duality relation[17]

$$\int_I Lf\, g = \int_I f\, g \circ T, \quad f \in L^1, g \in L^\infty.$$

In order to get useful information from this operator, we need to restrict the functional space where it acts; we choose here the Banach space BV of bounded variation functions on the unit interval equipped with the complete norm

$$\|f\|_{BV} = |f|_{TV} + \|f\|_1,$$

where $|f|_{TV}$ is the total variation of the function $f \in L^1$. A **chaotic** map T will admit a unique absolutely continuous invariant measure $\nu = \nu \circ T^{-1}$ supported

[17]Without mention of the contrary, all the L^p spaces in the paper will be intended with respect to the Lebesgue measure. The latter will be denoted as dx or Leb.

on the interval $[T(\Delta), \Delta]$ with a density $h \in BV$ ([15, Section 5, Corollary 1]). Moreover, ν is mixing with exponential decay of correlations on BV observable, namely there exists $0 < v < 1$ and a constant $C > 0$ such that

$$\left| \int L^n f \, g dx - \int f dx \int g dx \right| \leq C v^n \|f\|_{BV} \|g\|_\infty, \tag{30}$$

see [15, Section 5, Corollary 3] and [72, Proposition 5.15].

Also, we quote a sort of analog of the theorem of Blokh and Lyubich given above, for what concerns invariant measures for the map T. We give here the statement of Theorem 9 in [69], where the fact that the map T is S-unimodal is equivalent to say that it has a negative Schwarzian derivative. In particular, the (weak) accumulation points of the empirical measures constructed by iterating Lebesgue almost all initial points are characterized by Theorem A.2. These limiting invariant measure will be equivalent to the Lebesgue measure on the attractor if and only if the corresponding Lyapunov exponent is positive; this is the content of Theorem A.3 below.

Theorem A.2 (see [56, Chapter V.1])**.** Let T be an S-unimodal map with nonflat critical point. If T has a periodic attractor, or a Cantor attractor, then T admits a unique SRB measure[18] supported on the attractor.
If T admits an absolutely continuous invariant probability measure μ, then:

1. μ is a SRB measure;
2. the attractor A of T is an interval attractor;
3. $\operatorname{supp} \mu = A$, in particular, μ is equivalent to the Lebesgue measure on A.

Finally, we remind the following theorem by G. Keller, useful to compute the Lyapunov exponent Λ for the map T.

Theorem A.3 ([46])**.** Let $T \colon I \to I$ be an S-unimodal map with nonflat critical point. Then T admits an absolutely continuous invariant probability measure if and only if $\lim_{n\to\infty} \frac{1}{n} \log |DT^n(x)| = \Lambda > 0$, for almost all $x \in I$.

Of course if T has a periodic attractor, the Lyapunov exponent will be negative. The situation is different whenever T has a Cantor attractor. For an S-unimodal map with nondegenerate critical point that also has a Cantor attractor, the Lyapunov exponent will be 0, while there are families of unimodal maps with critical point of sufficiently high order, which have Cantor attractors with sensitive dependence on initial conditions, see [69, Section 5].

[18] We remind that an invariant measure μ is called a Sinai-Ruelle-Bowen (SRB) measure if

$$\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k(x)}$$

for Leb-a.e. $x \in [0, 1]$, where δ_x is the Dirac mass at x.

A.2: Continuous state Markov chains

In this section, we remind some basic properties of the continuous state Markov chain. A Markov chain $\{X_t : t \in \mathbb{N}\}$ on the interval I is given by the *transition probabilities*

$$P_x(A) = \mathbb{P}\{X_{t+1} \in A \mid X_t = x\}$$

(the probability that a chain at $x \in I$ will be in a set[19] $A \subset I$ after one step) and an initial distribution $\rho_0(A) = \mathbb{P}\{X_0 \in A\}$. In the particular case where all P_x, $x \in I$, and ρ_0 are absolutely continuous (with respect to Leb) and are given by densities $p(x, \cdot)$ and h_0, respectively, we have

$$P_x(A) = \int_A p(x, y)dy, \quad \rho_0(A) = \int_A h_0(y)dy.$$

The map $p\colon I \times I \to \mathbb{R}_+$ (known as the *stochastic kernel*) plays the role that the transition matrix does in the theory of Markov chains with a finite state space. For P_x to be a probability, it should satisfy $\int p(x, y)dy = 1$ for every $x \in I$.

Denote with $\mathcal{M}$ the space of (real-valued) Radon measures on I. There is an associated operator $\mathcal{L}\colon \mathcal{M} \to \mathcal{M}$ (called the *Markov operator* corresponding to P) acting by

$$\mathcal{L}\rho = \int P_x d\rho(x), \quad \rho \in \mathcal{M}, \tag{31}$$

that is, $\mathcal{L}\rho(A) = \int P_x(A)d\rho(x)$ for every $A \subset I$, or, equivalently,

$$\int \varphi d\mathcal{L}\rho = \iint \varphi(y)dP_x(y)d\rho(x)$$

for all $\varphi \in C^0$, where C^0 denotes the Banach space of continuous functions on I with the sup norm. We note that $\mathcal{L}\colon L^1 \to L^1$ is an isometry, where L^1 is intended, from now on, with respect to the Lebesgue measure. If the chain is given by the kernel p, formula (31) restricted to L^1 becomes

$$(\mathcal{L}h)(y) = \int p(x, y)h(x)dx, \quad h \in L^1. \tag{32}$$

If ρ_t denotes the distribution of the random variable X_t, then the distribution of X_{t+1} is $\rho_{t+1} = \mathcal{L}\rho_t$. In other words, by fixing the distribution ρ_0 for X_0, the entire sequence of future distributions can be obtained by iterating with $\mathcal{L}$.

A measure $\mu \in \mathcal{M}_+$ is said to be *stationary* if

$$\mathcal{L}\mu = \mu.$$

Every stationary measure μ gives rise to a shift-invariant measure $\mathbb{P}_\mu$ on the sequences space $\Omega = \{(x_t)_{t\in\mathbb{N}} : x_t \in I\}$ of realizations of the process, such that $\mathbb{P}_\mu(x_t \in A) = \mu(A)$ for all $t \in \mathbb{N}$ (see, e.g., [49]). We say that μ is *ergodic* if $\mathbb{P}_\mu$

[19] All sets considered will be assumed to be measurable. For brevity's sake, we omit the word 'measurable' everywhere in this text.

is ergodic. In the next section, we will show that, under some mild conditions, our model admits a unique (and thus ergodic) absolutely continuous stationary probability μ with a density $h \in BV$. Then, by the Ergodic Theorem (see, e.g., [17, Remark C4.1]), for every $f \in L^1(\mu)$,

$$\frac{1}{n}\sum_{t=1}^{n} f(X_t) \xrightarrow[n\to\infty]{} \int f d\mu, \quad \mathbb{P}_\mu\text{-almost surely.}$$

In particular, realizations of the process are distributed in the state space according to the measure μ.

Appendix B: Proofs of the results in Section 4

B.1: Proof of Theorem 4.1

In order to prove the theorem, we need some auxiliary results. We start by showing that there are *finitely many* ergodic stationary densities of bounded variation. The following lemma will be useful to that end.

Lemma B.1. *For any $\rho \in BV$, if $C := \operatorname{ess\,sup}_{x\in\operatorname{supp}\rho} |p(x,\cdot)|_{TV} < \infty$, then*

$$|\mathcal{L}\rho|_{TV} \le C\|\rho\|_1 \quad \textit{and} \quad \|\mathcal{L}\rho\|_{BV} \le (C+1)\|\rho\|_1.$$

Proof. For the first inequality, we have

$$\begin{aligned}
|\mathcal{L}\rho|_{TV} &= \sup \sum_i \left| \int p(x,y_{i+1})\rho(x)dx - \int p(x,y_i)\rho(x)dx \right| \\
&\le \sup \sum_i \int |p(x,y_{i+1})\rho(x) - p(x,y_i)\rho(x)|\, dx \\
&\le \int |p(x,\cdot)|_{TV}\rho(x)dx \le C\|\rho\|_1.
\end{aligned}$$

The second inequality follows from the first one, since the Markov operator is an isometry, i.e. $\|\mathcal{L}\rho\|_1 = \|\rho\|_1$ for all $\rho \in L^1$. □

Then, we show in Proposition B.2 below that we get a finite number of ergodic absolutely continuous stationary measures whose supports are mutually disjoint up to sets of zero Lebesgue measure.

We say that a stochastic kernel $p(x,y)$ has *uniformly bounded variations* if $|p(x,\cdot)|_{TV} \in L^\infty$, i.e. there is $C > 0$ such that $|p(x,\cdot)|_{TV} \le C$ for almost every $x \in I$.

Proposition B.2. *If the kernel p has uniformly bounded variations, then the operator $\mathcal{L}$ is quasi-compact, and there exist finitely many ergodic stationary measures with densities in BV and, moreover, their supports are mutually disjoint up to sets of zero Lebesgue measure.*

Proof. By Lemma B.1, for every n,

$$\|\mathcal{L}^n\rho\|_{BV} = |\mathcal{L}^n\rho|_{TV} + \|\mathcal{L}^n\rho\|_1 \leq C\|\mathcal{L}^{n-1}\rho\|_1 + \|\rho\|_1 = (C+1)\|\rho\|_1.$$

In particular,

$$\|\mathcal{L}\rho\|_{BV} \leq (C+1)\|\rho\|_1 \leq \eta\|\rho\|_{BV} + (C+1)\|\rho\|_1 \tag{33}$$

for any $\eta < 1$. This is the *Lasota-Yorke inequality*. The latter, plus the fact that BV is compactly embedded in L^1, implies that the peripheral spectrum of $\mathcal{L}$ is discrete and therefore the chain will admit *finitely many* (at least one) absolutely continuous ergodic stationary measures, with supports that are mutually disjoint up to sets of zero Lebesgue measure. Moreover, the essential spectral radius is strictly smaller than the spectral radius (*spectral gap*). These properties, which are consequences of the Ionescu-Tulcea-Marinescu theorem, are summarized by saying that the operator $\mathcal{L}$ acting on BV is *quasi-compact*, see, e.g., [13, 20, 19] for an exhaustive presentation of these results and [12, Section 2.3] for a specific application to random systems. □

Remark B.3. Let us mention that whenever the operator $\mathcal{L}$ is quasi-compact and the largest eigenvalue, which is 1 in our case, is simple and therefore there is only one stationary measure with density in BV, then the norm of $\|\mathcal{L}^k f\|_{BV}$ goes exponentially fast to zero when $k \to \infty$, for $f \in BV$ and $\int f dx = 0$ (exponential decay of correlations). This fact will be extensively used in Section 4.3.

At this point, the following remark is in order. Since the variance (17) vanishes at 0 and 1, the kernel (20) is in fact unbounded. However, we can still apply Proposition B.2 under a suitable restriction of the domain of $\mathcal{L}$. We first state a general result which allows us to confine the stationary measures.

Lemma B.4. *Under the assumptions of Section 3.2, there is $\varepsilon > 0$ such that any stationary measure μ has* $\operatorname{supp}\mu \subset \{0\} \cup I_{\varepsilon,\Gamma}$. *If μ is continuous,* $\operatorname{supp}\mu \subset I_{\varepsilon,\Gamma}$.

Proof. First, notice that any stationary measure is supported on the interval $K_\Gamma := [-\Gamma/2, 1-\Gamma/2]$. Indeed, by invariance, $\mu(K_\Gamma^c) = \int P_x(K_\Gamma^c)d\mu(x) = 0$, because $P_x(K_\Gamma^c) = 0$ for all x.

Fix $\varepsilon > 0$ such that $T(x) - s(x) > x$ for $x \in (0, \varepsilon)$. By choosing a smaller ε if needed, we may also assume that $T(x) - s(x) > \varepsilon$ for $x \in [\varepsilon, 1-\Gamma/2]$. Then $\inf\operatorname{supp} P_x > \min\{x, \varepsilon\}$ for every $x \in K_\Gamma \setminus \{0\}$. This means that for any realization (x_t) of the process, either all $x_t = 0$ (clearly, 0 is a fixed point, since $P_0 = \delta_0$) or *all but finitely many* $x_t > \varepsilon$.

On the other hand, if $\mu([-\Gamma, \varepsilon) \setminus \{0\}) > 0$, then by the Poincaré recurrence theorem, applied to the shift on $(\Omega, \mathbb{P}_\mu)$, $\mathbb{P}_\mu$-almost surely there would exist a realization (x_t) with *infinitely many* $0 \neq x_t < \varepsilon$, which is not possible, as we showed above. This finishes the proof. □

Proof of Theorem 4.1. By Lemma B.4, the density of any absolutely continuous stationary measure belongs to the subspace $Y := \{h \in L^1 \mid \operatorname{supp} h \subset I_{\varepsilon,\Gamma}\}$.

From the first part of the proof of Lemma B.4 it also follows that Y is $\mathcal{L}$-invariant. Moreover, the kernel (14) has uniformly bounded variations when restricted to $I_{\varepsilon,\Gamma} \times I_{\varepsilon,\Gamma}$. Indeed, $|p_{\mathfrak{n}}(x,\cdot)|_{TV} = |g_{x,\mathfrak{n}}|_{TV}$ and the latter is bounded on $I_{\varepsilon,\Gamma}$ by (12). We can therefore apply Proposition B.2. □

B.2: Proof of Theorem 4.3

As for the proof Theorem 4.1, we need some preliminary lemmas. We begin with the following simple lemma that links the topological dynamics of T with the structure of any stationary measure.

Lemma B.5. *For any stationary measure μ and any open set U, if $\mu(U) = 0$, then also $\mu(T^{-k}U) = 0$ for all $k > 0$.*

Proof. It is enough to show that $\mu(T^{-1}U) = 0$, the result then follows by induction. By invariance we have $0 = \mu(U) = \int P_x(U) d\mu(x) \geq \int_{T^{-1}U} P_x(U) d\mu(x)$. But for every $x \in T^{-1}U$, $T(x) \in U \cap \operatorname{supp} P_x$, and hence $P_x(U) > 0$. Therefore the latter integral can only be zero if $\mu(T^{-1}U) = 0$. □

With the help of the following lemma we will show that the support of any stationary measure contains the support of the T-invariant measure (atomic in the periodic case). Recall that $x \in \operatorname{supp}\mu$ iff $\mu(U) > 0$ for any open $U \ni x$.

Lemma B.6. *Let $A \subset I$ be such that (1) T is topologically transitive on A and (2) $\bigcup_{k=0}^{\infty} T^{-k}U = I$ for any open set $U \supset A$. Then $A \subset \operatorname{supp}\mu$ for any stationary measure μ.*

Proof. Given an open set U with $U \cap A \neq \emptyset$, by transitivity $A \subset \bigcup_{k=0}^{\infty} T^{-k}U$, and therefore $\bigcup_{k=0}^{\infty} T^{-k}U = I$. By Lemma B.5, $\mu(U) > 0$. □

Proof of Theorem 4.3. If T is periodic, a globally attracting cycle $\mathcal{O}$ satisfies the assumptions of Lemma B.6, therefore $\mathcal{O} \subset \operatorname{supp}\mu$ for any stationary measure μ. Let us show that $\operatorname{supp}\mu$ contains an open neighbourhood of $\mathcal{O}$. Recall that $p(x,y) > 0$ if and only if $y \in (T(x) - s(x), T(x) + s(x))$, and T and s are continuous. Given $x_0 \in \mathcal{O}$, let $x_1 \in \mathcal{O}$ be such that $x_0 = T(x_1)$. Since $(x_1, x_0) \in \{(x,y) \mid p(x,y) > 0\}$ and the latter set is open, we can find open sets $U \ni x_1$ and $V \ni x_0$ such that $p(x,y) > 0$ for all $x \in U$, $y \in V$. Also $\mu(U) > 0$, because $x_1 \in \operatorname{supp}\mu$. Denoting h the density of μ, by invariance we get $h(y) \geq \int_U p(x,y) d\mu(x) > 0$ for all $y \in V$, i.e. $V \subset \operatorname{supp}\mu$.

If T is chaotic, then T is topologically transitive when restricted to $I_\Delta = [T(\Delta), \Delta]$. It also follows from (C3) that $\bigcup_{k=0}^{\infty} T^{-k} I_\Delta = I$, so again we can apply Lemma B.6.

In both cases, we conclude that $\operatorname{Leb}(\operatorname{supp}\mu_1 \cap \operatorname{supp}\mu_2) > 0$ for any stationary measures μ_1, μ_2, and therefore by Proposition B.2 they must coincide. Hence the stationary measure is unique. □

B.3: Proof of Theorem 4.4

The proof of Theorem 4.4 is based on Proposition B.8, where the following lemma is employed.

Lemma B.7. *For every $x \in I$, $P_x^{(\mathtt{n})}$ converges to δ_{Tx} in the weak-* topology as $\mathtt{n} \to \infty$, i.e. $\int \varphi dP_x^{(\mathtt{n})} \to \varphi(Tx)$ for all $\varphi \in C^0(I)$.*

Proof. For arbitrary $\varepsilon > 0$ we can split

$$\int \varphi dP_x^{(\mathtt{n})} = \int_{B_\varepsilon(Tx)} \varphi dP_x^{(\mathtt{n})} + \int_{B_\varepsilon(Tx)^c} \varphi dP_x^{(\mathtt{n})}.$$

By Chebyshev's inequality, $P_x^{(\mathtt{n})}(B_\varepsilon(Tx)^c) \le \mathrm{Var}^2 P_x^{(\mathtt{n})}/\varepsilon^2 \to 0$ as $\mathtt{n} \to \infty$, while φ is bounded, so the second integral can be made arbitrarily small for $\mathtt{n}$ large. Consequently, $P_x^{(\mathtt{n})}(B_\varepsilon(Tx)) \to 1$, and since φ is continuous, the first integral can be made arbitrarily close to $\varphi(Tx)$. □

The proof of (i) below follows a suggestion in [6, Theorem D].

Proposition B.8. *Let μ be a weak-* limit measure of a sequence $\mu_{\mathtt{n}_k} = h_{\mathtt{n}_k} dx$. If $h_{\mathtt{n}_k}$ satisfy (A_q), then*

(i) μ is absolutely continuous with density in L^q;

(ii) μ is invariant under T.

Proof. (i) Let $\varphi \in C^0(I)$. By Hölder's inequality, with $p = \frac{q}{q-1}$,

$$\left| \int \varphi d\mu \right| = \left| \lim_{k\to\infty} \int \varphi h_{\mathtt{n}_k} dx \right| \le \lim_{k\to\infty} \|h_{\mathtt{n}_k}\|_q \|\varphi\|_p \le C_q \|\varphi\|_p.$$

Therefore the map $L^p \ni \varphi \mapsto \int \varphi d\mu \in \mathbb{R}$ is continuous, since C^0 is dense in L^p, and therefore such a functional will be in L^q, namely $\mu = h dx$, $h \in L^q$, $\|h\|_q \le C_q$.

(ii) It suffices to prove that any test function $\varphi \in C^0(I)$ satisfies

$$\int \varphi h_{\mathtt{n}_k} dy - \int \varphi \circ T h_{\mathtt{n}_k} dy \to 0.$$

Since $h_{\mathtt{n}_k} = \mathcal{L} h_{\mathtt{n}_k} = \int p_{\mathtt{n}_k}(x, \cdot) h_{\mathtt{n}_k}(x) dx$, by changing the order of integration in the first integral and subtracting the second, we get

$$\int h_{\mathtt{n}_k}(x) \left[\int p_{\mathtt{n}_k}(x,y)\varphi(y) dy - \varphi(T(x)) \right] dx. \tag{34}$$

Since the function $\psi_{\mathtt{n}_k}(x) := \int p_{\mathtt{n}_k}(x,y)\varphi(y)dy - \varphi(T(x))$ is uniformly bounded,

$$(34) \le \|h_{\mathtt{n}_k}\|_q \|\psi_{\mathtt{n}_k}\|_p \le C_q \|\psi_{\mathtt{n}_k}\|_p.$$

By Lemma B.7, $\psi_{\mathtt{n}_k}(x) \to 0$ for every $x \in I$, and therefore by dominated convergence $\|\psi_{\mathtt{n}_k}\|_p \to 0$. □

The weak stochastic stability now follows easily from the above results:

Proof of Theorem 4.4. Since T admits a unique invariant measure, μ must be the same for all convergent subsequences in Proposition B.8, and therefore the entire sequence $\mu_{\mathtt{n}}$ converges to μ. □

B.4: Proof of Proposition 4.5

Proof. Let us first consider the case when T has a globally attracting fixed point x_0. We need to show that for any test function $\varphi \in C^0(I)$ we have $\int \varphi(x) h_{\mathtt{n}}(x) dx \to \varphi(x_0)$ as $\mathtt{n} \to \infty$. Since $h_{\mathtt{n}}$ is a fixed point of the random transfer operator (23) and this operator is the dual of the random Koopman operator $\varphi \mapsto \int \varphi \circ T_\eta d\eta$ (see, e.g., [4, Section 2] for details), the previous weak limit leads to prove that the following quantity

$$\int_I \int_{[0,1]^k} \big(\varphi(T_{\eta_k} \circ \cdots \circ T_{\eta_1}(x)) - \varphi(x_0)\big) h_{\mathtt{n}}(x)\, d\bar{\eta}\, dx \tag{35}$$

goes to 0 as $\mathtt{n} \to \infty$, where k is an arbitrary fixed number and $\bar{\eta} = (\eta_1, \ldots, \eta_k)$.

Given $\varepsilon > 0$, let $\zeta > 0$ be such that $|\varphi(x) - \varphi(x_0)| < \varepsilon$ when $|x - x_0| < \frac{2\zeta}{1-\tau}$. Fix k such that, for all $\mathtt{n}$,

$$\sup_{x \in \operatorname{supp} \mu_{\mathtt{n}}} |T^k(x) - x_0| < \zeta. \tag{36}$$

Next, fix $\delta > 0$ such that

$$2\|\varphi\|_\infty (1 - (1 - 2\delta)^k) < \varepsilon. \tag{37}$$

Finally, by (27), for all $\mathtt{n}$ sufficiently large, we have

$$\sup_{\eta \in [\delta, 1-\delta]} \sup_{x \in I} |T_\eta(x) - T(x)| < \zeta. \tag{38}$$

We now split the integral (35) in the $\bar{\eta}$ variable over the region $E := [\delta, 1-\delta]^k$ and its complement. On E^c the absolute value of (35) is bounded by (37). Notice that the integral over x takes place on the support of $\mu_{\mathtt{n}}$, where (A_s) holds. Moreover, since the map $[0,1] \ni \eta \mapsto T_\eta \in C^2(I)$ is continuous, each T_η maps $\operatorname{supp} \mu_{\mathtt{n}}$ to itself; see [7]. Therefore, for $\bar{\eta} \in E$ and $x \in \operatorname{supp} \mu_{\mathtt{n}}$, by (38) and (A_s), we have

$$|T_{\eta_2} \circ T_{\eta_1}(x) - T^2(x)| \le |T_{\eta_2}(T_{\eta_1}(x)) - T(T_{\eta_1}(x))| + |T(T_{\eta_1}(x)) - T^2(x)| < \zeta + \tau\zeta.$$

By induction we easily get $|T_{\eta_k} \circ \cdots \circ T_{\eta_1}(x) - T^k(x)| < \frac{\zeta}{1-\tau}$ and therefore, in view of (36), $|T_{\eta_k} \circ \cdots \circ T_{\eta_1}(x) - x_0| < \frac{2\zeta}{1-\tau}$ for all $\bar{\eta} \in E$ and $x \in \operatorname{supp} \mu_{\mathtt{n}}$. Then, by the choice of ζ, $|\varphi(T_{\eta_k} \circ \cdots \circ T_{\eta_1}(x)) - \varphi(x_0)| < \varepsilon$ and the absolute value of (35) over E is therefore bounded by ε.

It is straightforward to modify the above proof for the case when T has a globally attracting periodic orbit of length $m > 1$. One needs to replace T

with T^m, the latter will have m attracting fixed points. The corresponding random maps of the form $T_{\bar\eta} = T_{\eta_1} \circ \cdots \circ T_{\eta_m}$ will be parametrized by $\bar\eta = (\eta_1, \cdots, \eta_m) \in [0,1]^m$ endowed with the Lebesgue measure. We leave the details to the reader. □

B.5: Proof of Theorem 4.10

Proof. (a) Denote with $\mu_{\mathbb{n}} = h_{\mathbb{n}} dx$ the unique stationary measure associated to $\mathbb{n}$ and with $\mu = h dx$ the unique invariant measure for T. We need to show that

$$\int \log|T'| d\mu_{\mathbb{n}} \xrightarrow[\mathbb{n}\to\infty]{} \int \log|T'| d\mu. \tag{39}$$

Let $q > 1$ be such that $h_{\mathbb{n}}, h \in L^q$ and set $p := \frac{q}{q-1}$. Since $\log|T'| \in L^p$, for any $\varepsilon > 0$ there is $\varphi_\varepsilon \in C^0$ such that $\|\log|T'| - \varphi_\varepsilon\|_p < \varepsilon$. Write

$$\int \log|T'| d\mu_{\mathbb{n}} = \int (\log|T'| - \varphi_\varepsilon) h_{\mathbb{n}} dx + \int \varphi_\varepsilon d\mu_{\mathbb{n}}$$

and

$$\int \log|T'| d\mu = \int (\log|T'| - \varphi_\varepsilon) h dx + \int \varphi_\varepsilon d\mu.$$

Since $\log|T'| - \varphi_\varepsilon \in L^p$ and $h_{\mathbb{n}}, h \in L^q$, we have

$$\int |\log|T'| - \varphi_\varepsilon| h_{\mathbb{n}} dx \le \|\log|T'| - \varphi_\varepsilon\|_p \|h_{\mathbb{n}}\|_q \le \varepsilon C_q,$$

and the same inequality holds for the integral with respect to μ. Finally, from Theorem 4.4 we know that $\mu_{\mathbb{n}} \xrightarrow{w^*} \mu$, hence $\int \varphi_\varepsilon d\mu_{\mathbb{n}} \to \int \varphi_\varepsilon d\mu$ as $\mathbb{n} \to \infty$.

(b) We know from Proposition 4.5 that $\int \varphi d\mu_{\mathbb{n}} \to \frac{1}{|\mathcal{O}|} \sum_{x\in\mathcal{O}} \varphi(x)$ for all $\varphi \in C^0(I)$ as $\mathbb{n} \to \infty$, since the T-invariant measure μ is atomic and supported on the attracting periodic orbit $\mathcal{O}$. Let us first consider the case when the critical point c belongs to $\mathcal{O}$; the right-hand side of (39) is then $-\infty$. Denoting $f_m(x) := \max\{\log|T'(x)|, -m\} \in C^0(I)$, for every m we have

$$\int \log|T'| d\mu_{\mathbb{n}} \le \int f_m d\mu_{\mathbb{n}} \xrightarrow[\mathbb{n}\to\infty]{} \frac{1}{|\mathcal{O}|} \sum_{x\in\mathcal{O}} f_m(x) \le -\frac{m}{|\mathcal{O}|} + C,$$

because $f_m \le C := \sup \log|T'|$ and $f_m(c) = -m$. Therefore $\int \log|T'| d\mu_{\mathbb{n}} \to -\infty$ as $\mathbb{n} \to \infty$.

If $c \notin \mathcal{O}$, we can fix a neighbourhood $U \ni c$ given by (A_c) and split

$$\int \log|T'| d\mu_{\mathbb{n}} = \int_U \log|T'| d\mu_{\mathbb{n}} + \int_{U^c} \log|T'| d\mu_{\mathbb{n}}.$$

The first term is bounded by $\|\log|T'|\|_1 \sup_U h_{\mathbb{n}}$ and vanishes as $\mathbb{n} \to \infty$ by (A_c), while the second one converges to $\int \log|T'| d\mu$ by Proposition 4.5 (approximate $\log|T'| \mathbb{1}_{U^c}$ with a suitable continuous function). □

The following example illustrates Corollary 4.11.

Example B.9. Denote $\{x \in I \mid |T'(x)| \leq 1\} = [m, M]$ and $\bar{\Delta} := \sup_{x \in I} T(x) + s(x)$. If $T(x) - s(x) > \min\{x, m\}$ for all $x \in (0, \bar{\Delta}]$, then, arguing as in the proof of Lemma B.4, one can show that any continuous stationary measure μ has $\operatorname{supp} \mu \subset [m, \bar{\Delta}]$. So if, moreover, $\bar{\Delta} \leq M$, then $\Lambda < 0$. Following the classification given in Appendix A.1, let us consider the case $T(c) < c$, where the map T exhibits a globally attracting fixed point. In this case, the conditions above will be satisfied if $s(x)$ is small enough, in other words, the stationary measure will be supported in a neighbourhood of the fixed point, where $|T'(x)| \leq 1$, leading to the negative Lyapunov exponent. For other cases, we provide some numerical examples in Section 5.2.

B.6: Proof of Theorem 4.12

Proof. Fix an exhaustion of Θ by nested compact sets Θ_ι and set $\tilde{\Theta}_\iota := \Theta_\iota \cap \tilde{\Theta}$. It is enough to prove that the mapping $\tilde{\Theta}_\iota \ni \theta \mapsto \Lambda_\theta \in \mathbb{R}$ is continuous on each $\tilde{\Theta}_\iota$, and from now on we fix one of them. As we showed in Lemma B.4, for each $\theta \in \Theta$, $\operatorname{supp} \mu_\theta \subset I_{\varepsilon_\theta} = [\varepsilon_\theta, 1 - \varepsilon_\theta]$, and since ε_θ can be shown to depend continuously on θ, we can find a single $\varepsilon > 0$ that works for all $\theta \in \Theta_\iota$.

Given $\theta, \theta' \in \tilde{\Theta}_\iota$ we can write

$$|\Lambda_\theta - \Lambda_{\theta'}| = \Big| \int \log |T'_\theta| h_\theta dx - \int \log |T'_{\theta'}| h_{\theta'} dx \Big| \leq$$

$$\int \big|\log |T'_\theta|\big| |h_\theta - h_{\theta'}| dx + \int \big|\log |T'_\theta| - \log |T'_{\theta'}|\big| h_{\theta'} dx. \tag{40}$$

To bound the second term in (40), first notice that, by Lemma B.1,

$$\|h_{\theta'}\|_\infty \leq \|h_{\theta'}\|_{BV} = \|\mathcal{L}_{\theta'} h_{\theta'}\|_{BV} \leq C \|h_{\theta'}\|_1 = C,$$

where

$$C = 1 + \sup_{\theta \in \Theta_\iota} \sup_{x \in I_\varepsilon} |p_\theta(x, \cdot)|_{TV} \leq 1 + \sup_{\theta \in \Theta_\iota} \sup_{x, y \in I_\varepsilon} \big| \frac{\partial p_\theta}{\partial y}(x, y) \big| < \infty$$

is finite, because $\frac{\partial p_\theta}{\partial y}(x, y)$ is continuous and $\Theta_\iota \times I_\varepsilon^2$ is compact. The second term is thus bounded by $C \| \log |T'_\theta| - \log |T'_{\theta'}| \|_1$ and, by Lemma B.11 below, goes to 0 as $\theta' \to \theta$.

We now estimate the first term in (40). Since $\log |T'_\theta| \in L^1$, it is enough to bound $\|h_\theta - h_{\theta'}\|_\infty$ which is again dominated by $\|h_\theta - h_{\theta'}\|_{BV}$. By invariance,

$$\|h_\theta - h_{\theta'}\|_{BV} = \|\mathcal{L}_\theta^k h_\theta - \mathcal{L}_{\theta'}^k h_{\theta'}\|_{BV} \leq \|\mathcal{L}_\theta^k (h_\theta - h_{\theta'})\|_{BV} + \|(\mathcal{L}_\theta^k - \mathcal{L}_{\theta'}^k) h_{\theta'}\|_{BV}.$$

As we said in Remark B.3, the Markov operator $\mathcal{L}_\theta$ enjoys the exponential bound

$$\|\mathcal{L}_\theta^k f\|_{BV} \leq C_\theta \zeta_\theta^k \|f\|_{BV}$$

for all $k > 0$ and $f \in BV$ supported on I_ε with $\int f dx = 0$, where the constants $C_\theta > 0$, $0 < \zeta_\theta < 1$ depend on the parameter θ. Since $h_\theta - h_{\theta'}$ has zero mean, we therefore have

$$\|\mathcal{L}_\theta^k(h_\theta - h_{\theta'})\|_{BV} \le C_\theta \zeta_\theta^k \|h_\theta - h_{\theta'}\|_{BV}.$$

Then

$$(1 - C_\theta \zeta_\theta^k)\|h_\theta - h_{\theta'}\|_{BV} \le \|(\mathcal{L}_\theta^k - \mathcal{L}_{\theta'}^k)h_{\theta'}\|_{BV},$$

and for k sufficiently large, $C_\theta \zeta_\theta^k < 1$. By a standard trick, expanding a telescopic sum $\mathcal{L}_\theta^k - \mathcal{L}_{\theta'}^k = \sum_{j=1}^k \mathcal{L}_\theta^{k-j}(\mathcal{L}_\theta - \mathcal{L}_{\theta'})\mathcal{L}_{\theta'}^{j-1}$, we get

$$\begin{aligned}\|(\mathcal{L}_\theta^k - \mathcal{L}_{\theta'}^k)h_{\theta'}\|_{BV} &\le \sum_{j=1}^k \|\mathcal{L}_\theta^{k-j}(\mathcal{L}_\theta - \mathcal{L}_{\theta'})h_{\theta'}\|_{BV} \\ &\le \sum_{j=1}^k C_\theta \zeta_\theta^{k-j} \|(\mathcal{L}_\theta - \mathcal{L}_{\theta'})h_{\theta'}\|_{BV} \le C_\theta \frac{1}{1-\zeta_\theta} \|(\mathcal{L}_\theta - \mathcal{L}_{\theta'})h_{\theta'}\|_{BV}.\end{aligned}$$

Combining the above inequalities we come to the following estimate:

$$\int \big|\log |T_\theta'|\big| |h_\theta - h_{\theta'}| dx \le M_\theta \|(\mathcal{L}_\theta - \mathcal{L}_{\theta'})h_{\theta'}\|_{BV},$$

where $M_\theta = \frac{C_\theta \| \log |T_\theta'| \|_1}{(1-\zeta_\theta)(1-C_\theta \zeta_\theta^k)}$. It therefore remains to bound $\|(\mathcal{L}_\theta - \mathcal{L}_{\theta'})h_{\theta'}\|_{BV}$. Since both $h_{\theta'}$ and $(\mathcal{L}_\theta - \mathcal{L}_{\theta'})h_{\theta'}$ are supported on I_ε and $\|h_{\theta'}\|_\infty \le C$, we have

$$\|(\mathcal{L}_\theta - \mathcal{L}_{\theta'})h_{\theta'}\|_1 = \int_{I_\varepsilon} \int_{I_\varepsilon} |p_\theta(x,y) - p_{\theta'}(x,y)| h_{\theta'}(x) dx dy \le M_1 \|\theta - \theta'\|,$$

where $M_1 = C \sup_{\theta \in \Theta_\iota} \sup_{x,y \in I_\varepsilon} \|\nabla_\theta p_\theta(x,y)\|$ is finite because $\nabla_\theta p_\theta(x,y)$ is continuous and $\Theta_\iota \times I_\varepsilon^2$ is compact. Similarly, arguing as in the proof of Lemma B.1, we get

$$|(\mathcal{L}_\theta - \mathcal{L}_{\theta'})h_{\theta'}|_{TV} \le \int_{I_\varepsilon} |p_\theta(x,\cdot) - p_{\theta'}(x,\cdot)|_{TV} h_{\theta'}(x) dx \le M_2 \|\theta - \theta'\|,$$

with $M_2 = C \sup_{\theta \in \Theta_\iota} \sup_{x,y \in I_\varepsilon} \|\nabla_\theta \frac{\partial p_\theta}{\partial y}(x,y)\| < \infty$. Therefore, the first term in (40) is bounded by $M_\theta(M_1 + M_2)\|\theta - \theta'\|$. This finishes the proof. □

Remark B.10. Clearly, the above proof works if we replace $\log |T'|$ with any continuous function. Therefore, the mapping $\tilde{\Theta} \ni \theta \mapsto \mu_\theta \in \mathcal{M}$ is continuous with respect to the weak-* topology on $\mathcal{M}$, i.e. $\tilde{\Theta} \ni \theta \mapsto \int \varphi \mu_\theta \in \mathbb{R}$ is continuous for any $\varphi \in C^0(I)$. Theorem 4.12 is more delicate, however, because $\log |T_\theta'|$ is neither continuous nor bounded and also depends on θ.

Lemma B.11. *$\| \log |T_\theta'| - \log |T_{\theta'}'| \|_1 \to 0$ as $\theta' \to \theta$.*

Proof. First notice that the critical point c_θ of the map T_θ depends continuously on the parameter $\theta \in \Theta$. Indeed, since T'_θ is continuous on $\Theta \times I$, the set $\{(\theta, c_\theta)\} = (T'_\theta)^{-1}(\{0\})$ is closed, and then the map $\Theta \ni \theta \mapsto c_\theta \in I$ is continuous by the closed graph theorem.

The functions $\log|T'_\theta|$ and $\log|T'_{\theta'}|$ have logarithmic singularities at c_θ and $c_{\theta'}$ respectively. We will show that these singularities cancel out as $c_{\theta'}$ approaches c_θ. As in the proof of Theorem 4.12, we may assume that $\theta, \theta' \in \Theta_\iota$, where Θ_ι is compact.

Let $\alpha := \sup_{\theta\in\Theta_\iota} |T'''_\theta|$. As a direct consequence of the mean value theorem,

$$(|T''_\theta(c_\theta)| - 2\alpha\delta)|x - c_\theta| \le |T'_\theta(x)| \le (|T''_\theta(c_\theta)| + 2\alpha\delta)|x - c_\theta|$$

for all $|x - c_\theta| \le 2\delta$, and the same inequality holds if we replace θ with θ'. Set $D^\pm_\delta := |T''_\theta(c_\theta)| \pm 3\alpha\delta$; both D^+_δ and D^-_δ are positive, since T_θ has quadratic critical point ($T''_\theta(c_\theta) < 0$). If θ' is sufficiently close to θ, then $|c_\theta - c_{\theta'}| < \delta/2$ and $|T''_\theta(c_\theta) - T''_\theta(c_{\theta'})| < \delta$, and for all $|x - c_\theta| \le \delta$ we then simultaneously have

$$D^-_\delta|x - c_\theta| \le |T'_\theta(x)| \le D^+_\delta|x - c_\theta|,$$

$$D^-_\delta|x - c_{\theta'}| \le |T'_{\theta'}(x)| \le D^+_\delta|x - c_{\theta'}|,$$

and hence

$$\big|\log|T'_\theta| - \log|T'_{\theta'}|\big| \le \log\frac{D^+_\delta}{D^-_\delta} + \Big|\log\frac{|x - c_\theta|}{|x - c_{\theta'}|}\Big|.$$

Given $\varepsilon > 0$ and $\theta \in \Theta_\iota$, we first fix $\delta > 0$ such that $\log\frac{D^+_\delta}{D^-_\delta} < \frac{\varepsilon}{3}$ and then let $\theta' \to \theta$. The integral of the second term is elementary and vanishes as $c_{\theta'} \to c_\theta$, so $\int_{B_\delta(c_\theta)}\big|\log\frac{|x-c_\theta|}{|x-c_{\theta'}|}\big|dx < \frac{\varepsilon}{3}$, provided θ' and θ are sufficiently close, and

$$\int_{B_\delta(c_\theta)}\big|\log|T'_\theta| - \log|T'_{\theta'}|\big|dx < \frac{2\varepsilon}{3}. \tag{41}$$

Let us show that $\log|T'_\theta| - \log|T'_{\theta'}| \to 0$ uniformly on $B_\delta(c_\theta)^c$. Denote $\beta := \inf_{x\in B_\delta(c_\theta)^c}|T'_\theta(x)| > 0$. We have $\|T'_\theta - T'_{\theta'}\|_\infty \le M\|\theta - \theta'\|$, where $M = \sup_{\theta\in\Theta_\iota, x\in I}\|\nabla_\theta T'_\theta(x)\|$. Therefore $\inf_{x\in B_\delta(c_\theta)^c}|T'_{\theta'}(x)| > \frac{\beta}{2}$ if $\|\theta - \theta'\| < \frac{\beta}{2M}$. Consequently, $\big|\log|T'_\theta| - \log|T'_{\theta'}|\big| < \frac{2M}{\beta}\|\theta - \theta'\|$, and for $\|\theta - \theta'\| < \frac{\varepsilon\beta}{6M}$ we have

$$\int_{B_\delta(c_\theta)^c}\big|\log|T'_\theta| - \log|T'_{\theta'}|\big|dx < \frac{\varepsilon}{3}. \tag{42}$$

Combining (42) with (41) we finally get $\|\log|T'_\theta| - \log|T'_{\theta'}|\|_1 < \varepsilon$. □

B.7: Random Lyapunov exponent and random entropy

In (28) we used the derivative of the deterministic map only. Alternatively, if we define the process using the random transformations (22), we are led to compute the Lyapunov exponent of the cocycle given by the derivative computed along

the random orbit, namely we define the *random Lyapunov exponent* (RLE) $\bar{\Lambda}$ as

$$\bar{\Lambda} := \lim_{n\to\infty} \frac{1}{n} \log |D(T_{\eta_n} \circ \cdots \circ T_{\eta_1})(x)|, \tag{43}$$

for almost every sequence $(\eta_k) \in [0,1]^{\mathbb{N}}$ with respect to the measure $\mathrm{Leb}^{\otimes\mathbb{N}}$ (see Section 3.4), and almost every $x \in I$ with respect to the stationary measure $\mu_{\mathtt{n}}$. By using the notation introduced in Section 3.4: $\bar{x}_k(x) := T_{\eta_k} \circ \cdots \circ T_{\eta_1}(x)$, with $\bar{x}_0(x) = x$, formula (43) is equal to

$$\bar{\Lambda} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log |T'_{\eta_k}(\bar{x}_{k-1})|, \tag{44}$$

again for $\mathrm{Leb}^{\otimes\mathbb{N}}$-a.e. $(\eta_k) \in [0,1]^{\mathbb{N}}$ and $\mu_{\mathtt{n}}$-a.e. $x \in I$. By using the ergodic theorem for random transformations, see [47] or [8, Section 3.1], we have that

$$\bar{\Lambda} = \int \log |T'_\eta(x)| d\mu_{\mathtt{n}}(x) d\eta.$$

Notice that if we compare this random exponent with Λ, we see that the difference between the two is bounded by

$$|\Lambda - \bar{\Lambda}| \le \int \big|\log |T'_\eta(x)| - \log |T'(x)|\big| d\mu_{\mathtt{n}}(x) d\eta.$$

By using expression (26) for T_η, we can bound the error term in a more explicit way as

$$|\Lambda - \bar{\Lambda}| \le \int \left| \log\Big|1 + \frac{\tilde{q}_{\mathtt{n}}(\eta)\sigma'(x)}{T'(x)}\Big| \right| h_{\mathtt{n}}(x) dx d\eta.$$

Since σ' is bounded over $I_{\varepsilon,\Gamma}$, $\log |T'| \in L^p, p \ge 1$, and finally $\|h_{\mathtt{n}}\|_q \le C_q$ for all $\mathtt{n}$, we have

$$|\Lambda - \bar{\Lambda}| \le C_q \int \left\| \log\Big|1 + \frac{\tilde{q}_{\mathtt{n}}(\eta)\sigma'}{T'}\Big| \right\|_p d\eta.$$

We expect that this error converges to zero for large $\mathtt{n}$, since the quantile function converge to zero almost everywhere. Table 1 in Section 5.2 shows that, in fact, the error remains very small even for small values of $\mathtt{n}$.

It is known that whenever the random Lyapunov exponent (43) is positive, then it equals the *random entropy*, which is the random generalization of the Kolmogorov-Sinai entropy: this equality is called the *entropy formula*. Roughly speaking, the random entropy computes the limit of the Shannon entropy of the join partition generated by the successive application of the backward images of the random maps on an initial generating partition, aka the entropy rate. We defer to [47, Theorem 1.3] for the precise definition and [8, Theorem 3.2] for the connection with the RLE. What is important for us is that the random entropy coincides with the much easier object which is the RLE when the latter is positive.

In Corollary 4.11 we proved that the Lyapunov exponent (28) may be negative, and there is strong numerical evidence that the RLE (44) is also negative for certain parameters; see Section 5.2. By assuming that the RLE is negative or zero, we then get that the random entropy is also zero, by using another important result connecting entropy and Lyapunov exponents, namely the Margulis-Ruelle inequality (see for instance [66]), which states that the metric entropy is bounded by the maximum between zero and the sum of the positive Lyapunov exponents. The random version of this inequality, which we use, has been proved by Kifer in [47, Theorem 1.4].

We should now stress the interesting fact that although the entropy is zero, the Markov chain *mixes exponentially fast*, as we pointed out in Remark B.3. This means that for any observables $f \in L^1$, $g \in BV$ the correlations $\int (\mathcal{L}_{\mathbb{m}}^k f)(x) g(x) dx$ converge to $\int f(x) d\mu_{\mathbb{m}}(x) \int g(x) dx$ exponentially fast when $k \to \infty$. This result can be stated in a more suggestive way by relying on random transformations. By using the notations introduced in Section 3.4, we can in fact rewrite the previous correlation in terms of composition of randomly chosen maps and say that there exist $0 < v < 1$ and $C > 0$, depending only on the system, such that, for all $k \geq 0$,

$$\left| \iint f(x) g(T_{\eta_k} \circ \cdots \circ T_{\eta_1}(x)) d\overline{\eta} dx - \int f d\mu_{\mathbb{m}} \int g(x) dx \right| \leq C v^k \|f\|_1 \|g\|_{BV}.$$

This exponential decay of correlations is a consequence of the spectral gap prescribed by the quasi-compactness of the Markov operator proved in Proposition B.2, and of the uniqueness of the absolutely continuous stationary measure, see [12] for details; of course these properties hold even when the Lyapunov exponent is negative or zero.

Appendix C: Details on the CNN's architectures

The architectures of CNN1 and CNN2(k) used in Section 6 are schematized in Fig. 12. In particular the only difference between CNN1 and CNN2(k) is on the dimension of the final layer (three and two, respectively) and on the loss (categorical cross-entropy and mean square error, respectively). In general, a convolutional layer is composed of n_f filters and each filter is associated with one kernel that is applied to a small moving window of the time-series; for instance, in our first convolutional layer (indicated by Conv1D in Fig. 12) $n_f = 128$ and all the windows are of width 2 (this number is not displayed in the standard Python output). The outputs of one convolutional layer are connected to the next layer. The weights of these connections constitute the NN parameters to be optimized. After seven convolutions, the output is passed to a sequence of dense layers, which concludes the NN. We use the rectified linear unit (ReLU) function as activation function.

Now, we describe the experimental setup. The implementation is carried out in Python. To generate training and testing data we simulate one million samples. CNN1 is optimized with the stochastic gradient descent method by

```
Model CNN1: "convolutional_categorical_model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
reshape (Reshape)            (None, 59, 1)             0
_________________________________________________________________
conv1d_1 (Conv1D)            (None, 58, 128)           384
_________________________________________________________________
conv1d_2 (Conv1D)            (None, 29, 64)            16448
_________________________________________________________________
conv1d_3 (Conv1D)            (None, 15, 64)            8256
_________________________________________________________________
conv1d_4 (Conv1D)            (None, 8, 64)             8256
_________________________________________________________________
conv1d_5 (Conv1D)            (None, 4, 64)             8256
_________________________________________________________________
conv1d_6 (Conv1D)            (None, 2, 64)             8256
_________________________________________________________________
conv1d_7 (Conv1D)            (None, 1, 64)             8256
_________________________________________________________________
flatten (Flatten)            (None, 64)                0
_________________________________________________________________
dense_1 (Dense)              (None, 128)               8320
_________________________________________________________________
dense_2 (Dense)              (None, 64)                8256
_________________________________________________________________
dense_3 (Dense)              (None, 3)                 195
=================================================================
Trainable params: 74,883

Model CNN2: "convolutional_model"
...
_________________________________________________________________
dense_3 (Dense)              (None, 2)                 130
=================================================================
Trainable params: 74,818
```

Figure 12: Architectures of the CNN1 model (Model CNN1) used to estimate the iterate k and the CNN2(k) model (Model CNN2) used to estimate the parameters (ϕ^*, ω) for each k. The two models differ only in the output layer. *Meaning of the table's header:* Layers(type) indicates the type of the employed layer (precisely, Conv1D indicates the usage of a 1D convolution layer, Dense indicates the usage of a dense layer). OutputShape : indicates the output shape of the layers, with None indicating that we do not perform batching. Finally, Param # indicates the number of layer's parameters.

using the Adam algorithm [48], the categorical cross-entropy as loss function, and the accuracy as metric, with a L^2 regularization of weights equal to 10^{-7}. Instead, to optimize CNN2 we use the Mean Squablack Error (MSE) both as loss function and as metric, again with a L^2 regularization of weights equal to 10^{-7}. The batch size is 32 in both cases. The seven convolutional and three dense layers have a total of $74,818$ trainable parameters.

We point out that different CNN models with different ranges for layers, dense units, and CNN filters were tested, but they gave us a lower accuracy and a bigger MSE with respect the one selected. For the sake of space, we do not report all the experimented architectures; the results and the Python code are available from the authors upon request. Finally, the L^2 regularization (hyper-)parameter has been chosen as part of the training process by using the

validation set as that value for which the validation MSE stops decreasing. For instance, Fig. 13 shows the training MSE and the Validation MSE as a function of the Regularization factor for CNN1.

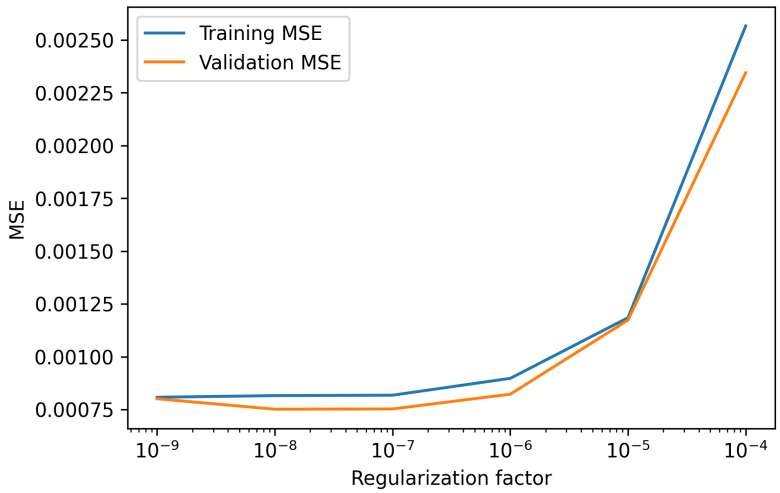


Figure 13: Training MSE and the Validation MSE as a function of the Regularization factor for CNN1

Appendix D: The Chaos Decision Tree Algorithm

The Chaos Decision Tree Algorithm [71] is a non-parametric chaos-detection tool which has been developed with the goal of being especially robust to measurement noise. It provides an automated processing pipeline which has been showed to be able to detect the presence (or absence) of chaos in noisy recordings, even for difficult edge cases. We use it in our work to identify periodic/chaotic time series without any reference to our model. It is meant to provide an independent check on the existence of chaotic behaviour in leverage time series and to support the evidence that it may depend on the bank's size.

The algorithm classifies a time-series as either stochastic or periodic, or chaotic. We now briefly explain how the algorithm works. The first step is to test if data are stochastic. This is done via surrogate-based approach by comparing the permutation entropy of the original time-series to the permutation entropy of random surrogates of that time-series by using a combination of Amplitude Adjusted Fourier Transform surrogates and Cyclic Phase Permutation surrogates. If the permutation entropy of the original time-series falls within either surrogate distribution, the time-series is classified as stochastic. If the permutation entropy falls outside the surrogate distribution, then the algorithm proceeds to de-noise the inputted signal by using the Schreiber's noise-blackuction algorithm [67]. Notice that the calculation of the permutation

entropy relies on two parameters: the permutation order and the time-lag. The time-lag has been set to 1 as suggested in [71]. The choice of the order of the permutation is made in order to maximize the detection of chaotic series in our model.

To test the CDTA algorithm on our model, we first consider the deterministic map and generate 100 chaotic and 100 periodic time-series of length 59 from the dynamical core. We then apply CDTA to these series for different values of the permutation order ($\in \{3, \dots, 8\}$). While the periodicity accuracy is maximized for a value equal to 8 (92% of the periodic series are correctly detected as such and the remaining 8% are labeled chaotic), the chaos detection accuracy is maximized for a choice of the permutation order equal to 5 (65% of the series are correctly detected as chaotic and the remaining periodic). Because of the purpose of this paper, we fix the permutation order to 5, but we have checked that the conclusions of our data analysis (in particular Table 4) would have been the same with a different choice.

At this point, the algorithm checks for signal oversampling and, if the data are over-sampled, the algorithm iteratively down-samples the data until they are no longer over-sampled. Finally, CDTA performs the 0-1 chaos test [42] on the input data. Ref. [71] points out that the 0-1 chaos test has been modified from the original one to be less sensitive to noise. Then, it suppresses the correlations arising from quasi-periodicity, and normalizes the standard deviation of the test signal. The value for the parameter that suppresses signal correlations is chosen based on ROC analyses. The modified 0-1 test provides a single statistic, K, which approaches 1 for chaotic systems and approaches 0 for periodic systems. The algorithm sets up a cutoff for K based on the length of the time-series. If K is greater than the cutoff, the data are classified as chaotic, and if they are smaller than or equal to the cutoff, they are classified as periodic.

Simulations In this section we present some numerical investigations showing how CDTA performs when simulating (noisy) time-series from the map described in Section 3.3. Specifically, we run the following two numerical experiments:

(i) First, we simulate time-series from the dynamical core area (i.e. time-series for which the pairs (ϕ^*, ω) satisfy condition (C3): $T(\Delta) < c < \Delta < 1$). More precisely, we simulate 500 samples of different lengths and level of noise, which is captured by the variable $\mathtt{n}$.

(ii) Second, we simulate time-series from outside the dynamical core area (the map T thus satisfies (C1) or (C2)). The remaining simulation setting coincides with that in (i).

The procedures explained in (i)–(ii) are also repeated when T is replaced by its k-th iterate T^k, $k = 2, 3$; see discussion in Section 6. Table 4 collects the results. We observe that when both $\mathtt{n}$ and the time series length are large, CDTA classifies almost all the time series in the dynamical core as chaotic, while those outside it are never classified as such. This is quite independent

Iterate	Series length	$\mathbb{m}$	Dynamical core			Not dynamical core		
			S (%)	P (%)	C (%)	S (%)	P (%)	C (%)
1	59	5	33.5	5.14	**61.4**	**57.4**	4.81	37.7
		20	29.3	4.28	**66.5**	**69.9**	3.28	26.8
		100	24.3	6.99	**69.6**	**88.1**	2.94	8.97
	295	5	2.2	1.7	**96.1**	22.5	6.24	**71.2**
		20	0.1	1.9	**98**	43.8	10.4	**45.8**
		100	0	2.3	**97.7**	**73.8**	8.35	17.9
	590	5	0	0.7	**99.3**	13.1	6.08	**80.9**
		20	0	0.4	**99.6**	33.6	8.5	**57.9**
		100	0	0.4	**99.6**	**66.4**	8.04	25.6
	1180	5	0	0.1	**99.9**	10.9	3.44	**85.6**
		20	0	0	**100**	27.7	5.57	**66.8**
		100	0	0	**100**	**60.2**	5.29	34.5
2	59	5	**75.7**	2.17	22.2	**83.8**	1.33	14.9
		20	**80.1**	1.65	18.2	**92.8**	0.26	6.91
		100	**86.6**	1.43	11.9	**96.6**	0.56	2.81
	295	5	39.4	0	**60.6**	40.5	3.26	**56.2**
		20	38.6	0.6	**60.8**	**70**	3.7	26.3
		100	21	1.2	**77.8**	**83.9**	2.96	13.1
	590	5	27.6	0	**72.4**	25.6	3.23	**71.2**
		20	10.6	0	**89.4**	**52.3**	4.24	43.4
		100	4.8	0.6	**94.6**	**74.7**	2.88	22.4
	1180	5	11	0	**89**	13.7	2.22	**84.1**
		20	0.2	0	**99.8**	39.4	2.82	**57.7**
		100	0.2	0	**99.8**	**64.1**	2.24	33.7
3	59	5	**85.3**	0.47	14.2	**83.9**	1.33	14.9
		20	**84.5**	1.17	14.3	**92.8**	0.26	6.91
		100	**89.5**	1.43	8.79	**96.6**	0.56	2.81
	295	5	36	0	**64**	40.5	3.26	**56.2**
		20	32.6	0.4	**67**	**70**	3.7	26.3
		100	23	1.6	**75.4**	**83.9**	2.96	13.1
	590	5	20.6	0	**79.4**	25.6	3.23	**71.2**
		20	9.6	0	**90.4**	**52.3**	4.24	43.4
		100	9.4	0.8	**89.8**	**74.7**	2.88	22.4
	1180	5	6	0	**94**	13.7	2.22	**84.1**
		20	2	0	**98**	39.4	2.82	**57.7**
		100	2.8	0.8	**96.4**	**64.1**	2.24	33.7

Table 4: Percentage of time-series detected by the Chaos Decision Tree Algorithm as stochastic (S), periodic (P) and chaotic (C) in the dynamical core and its complement as a function of the time-series length and $\mathbb{m}$ for different iterates of the map.

on k. By decreasing either $\mathbb{n}$ or the time series length, the classification is less precise and this effect is stronger for larger values of k. In the regime of length comparable with our empirical data ($N = 59$), roughly a third (for $k = 1$) or up to 85% (for $k = 3$) of the time series in the dynamical core are classified as stochastic, showing the limits of the CDTA when the time series are short and/or the noise is large.

References

[1] Viral V. Acharya and Stephen G. Ryan. "Banks' Financial Reporting and Financial System Stability". In: *Journal of Accounting Research* 54.2 (2016), pp. 277–340.

[2] Tobias Adrian and Hyun Song Shin. "Liquidity and leverage". In: *Journal of Financial Intermediation* 19.3 (2010), pp. 418–437.

[3] Tobias Adrian and Hyun Song Shin. "Procyclical Leverage and Value-at-Risk". In: *The Review of Financial Studies* 27.2 (2013), pp. 373–403.

[4] Romain Aimino, Matthew Nicol, and Sandro Vaienti. "Annealed and quenched limit theorems for random expanding dynamical systems". In: *Probability Theory and Related Fields* 162.1-2 (2015), pp. 233–274.

[5] Borovkov Aleksandr. *Ergodicity and Stability of Stochastic Processes*. J. Wiley, 1998.

[6] Jose F. Alves and Vıtor Araújo. "Random perturbations of nonuniformly expanding maps". In: *Astérisque* 286 (2003), pp. 25–62.

[7] Vıtor Araújo. "Attractors and time averages for random maps". In: *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*. Vol. 17. 3. Elsevier. 2000, pp. 307–369.

[8] Vıtor Araújo and Ali Tahzibi. "Stochastic stability at the boundary of expanding maps". In: *Nonlinearity* 18.3 (2005), p. 939.

[9] Kerstin Awiszus, Agostino Capponi, and Stefan Weber. "Market efficient portfolios in a systemic economy". In: *Operations Research* 70.2 (2022), pp. 715–728.

[10] Christoph Aymanns and J Doyne Farmer. "The dynamics of the leverage cycle". In: *Journal of Economic Dynamics and Control* 50 (2015), pp. 155–179.

[11] Christoph Aymanns et al. "Taming the Basel leverage cycle". In: *Journal of Financial Stability* 27 (2016), pp. 263–277.

[12] Wael Bahsoun, Huyi Hu, and Sandro Vaienti. "Pseudo-Orbits, Stationary Measures and Metastability". In: *Dynamical Systems* 29.3 (2014), pp. 322–336.

[13] Viviane Baladi. *Positive Transfer Operators and Decay of Correlations*. Vol. 14. World Scientific, 2000.

[14] Viviane Baladi, Michael Benedicks, and Véronique Maume-Deschamps. "Almost sure rates of mixing for iid unimodal maps". In: *Annales Scientifiques de l'Ecole Normale Supérieure*. Vol. 35. 1. Elsevier. 2002, pp. 77–126.

[15] Viviane Baladi and Marcelo Viana. "Strong stochastic stability and rate of mixing for unimodal maps". In: *Annales Scientifiques de l'Ecole Normale Supérieure*. Vol. 29. 4. 1996, pp. 483–517.

[16] Nils Berglund and Barbara Gentz. *Noise-induced phenomena in slow-fast dynamical systems: a sample-paths approach*. Springer Science & Business Media, 2006.

[17] Rabindra N. Bhattacharya and Mukul Majumdar. *Random Dynamical Systems: Theory and applications*. Cambridge, 2007.

[18] Alexander Blokh and Mikhail Lyubich. "Measurable dynamics of S-unimodal maps of the interval". In: *Annales scientifiques de l'Ecole normale supérieure*. Vol. 24. 5. 1991, pp. 545–573.

[19] Abraham Boyarsky and Pawel Gora. *Laws of Chaos: Invariant Measures and Dynamical Systems in One Dimension*. Birkhäuser, 1997.

[20] Anne Broise. "Transformations dilatantes de l'intervalle et théorèmes limites". In: *Asterisque-societe mathematique de France* 238 (1996), pp. 1–109.

[21] Markus K Brunnermeier and Lasse Heje Pedersen. "Market liquidity and funding liquidity". In: *The review of financial studies* 22.6 (2009), pp. 2201–2238.

[22] Fabio Caccioli, Jean-Philippe Bouchaud, and D Farmer. "Impact-adjusted valuation and the criticality of leverage". In: *Risk* 25.12 (2012), pp. 74–77.

[23] Agostino Capponi, Paul Glasserman, and Marko Weber. "Swing pricing for mutual funds: Breaking the feedback loop between fire sales and fund redemptions". In: *Management Science* 66.8 (2020), pp. 3581–3602.

[24] Agostino Capponi and Martin Larsson. "Price contagion through balance sheet linkages". In: *The Review of Asset Pricing Studies* 5.2 (2015), pp. 227–253.

[25] Giuseppe Castellacci and Youngna Choi. "Modeling contagion in the Eurozone crisis via dynamical systems". In: *Journal of Banking & Finance* 50 (2015), pp. 400–410.

[26] Youngna Choi and Raphael Douady. "Financial crisis dynamics: attempt to define a market instability indicator". In: *Quantitative Finance* 12.9 (2012), pp. 1351–1365.

[27] Youngna Choi, Raphael Douady, et al. "Financial crisis and contagion: A dynamical systems approach". In: *Handbook on systemic risk* (2013), pp. 453–479.

[28] Rama Cont and Eric Schaanning. "Fire sales, indirect contagion and systemic stress testing". In: (2017).

[29] Rama Cont and Lakshithe Wagalath. "Fire sales forensics: measuring endogenous risk". In: *Mathematical finance* 26.4 (2016), pp. 835–866.

[30] Fulvio Corsi, Stefano Marmi, and Fabrizio Lillo. "When micro prudence increases macro risk: The destabilizing effects of financial innovation, leverage, and diversification". In: *Operations Research* 64.5 (2016), pp. 1073–1088.

[31] Jón Daníelsson, Hyun Song Shin, and Jean-Pierre Zigrand. "Procyclical Leverage and Endogenous Risk". In: *Available at SSRN, abstract=1360866* (2012).

[32] Jón Daníelsson, Hyun Shin, and Jean–Pierre Zigrand. "The impact of risk regulation on price dynamics". In: *Journal of Banking & Finance* 28.5 (2004), pp. 1069–1087.

[33] Nils Detering et al. "Suffocating fire sales". In: *SIAM Journal on Financial Mathematics* 13.1 (2022), pp. 70–108.

[34] Robert L Devaney. *An introduction to chaotic dynamical systems*. CRC press, 2018.

[35] Domenico Di Gangi, Fabrizio Lillo, and Davide Pirino. "Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction". In: *Journal of Economic Dynamics and Control* 94 (2018), pp. 117–141.

[36] Dmitrii I. Dolgopyat. "Averaging and invariant measures". In: *Moscow Mathematical Journal* 5.3 (2005), pp. 537–576.

[37] Ana Fostel and John Geanakoplos. "Leverage cycles and the anxious economy". In: *American Economic Review* 98.4 (2008).

[38] Stefano Galatolo, Maurizio Monge, and Isaia Nisoli. "Existence of Noise Induced Order, a Computer Aided Proof". In: *Nonlinearity* 33.9 (2020), p. 4237.

[39] John Geanakoplos. "The leverage cycle". In: *NBER macroeconomics annual* 24.1 (2010), pp. 1–66.

[40] Quentin Giai Gianetto, Jean-Marc Le Caillec, and Erwan Marrec. "Estimating the predictability of economic and financial time series". In: *arXiv:1212.2758v1* (2012).

[41] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[42] Georg A. Gottwald and Ian Melbourne. "On the implementation of the 0–1 test for chaos". In: *SIAM Journal on Applied Dynamical Systems* 8.1 (2009), pp. 129–145.

[43] R. Greenwood, A. Landier, and D. Thesmar. "Vulnerable banks". In: *Journal of Financial Economics* 115.3 (2015), pp. 471–485.

[44] Cars Hommes et al. "Learning in Cobweb Experiments". In: *Macroeconomic Dynamics* 11.1 (2007), pp. 8–33.

[45] Harry Huizinga and Luc Laeven. "Bank valuation and accounting discretion during a financial crisis". In: *Journal of Financial Economics* 106.3 (2012), pp. 614–634.

[46] Gerhard Keller. "Exponents, attractors and Hopf decompositions for interval maps". In: *Ergodic Theory and Dynamical Systems* 10.4 (1990), pp. 717–744.

[47] Yuri Kifier. *Ergodic Theory of Random Trasformations*. Springer, 1986.

[48] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014.

[49] Leonid Koralov and Yakov Sinai. *Theory of Probability and Random Processes*. Springer, 2012.

[50] Albert S Kyle and Wei Xiong. "Contagion as a wealth effect". In: *The Journal of Finance* 56.4 (2001), pp. 1401–1440.

[51] Yann LeCun et al. "Object Recognition with Gradient-Based Learning". In: Cambridge, 2007.

[52] Bárbara Llacay and Gilbert Peffer. "Impact of value-at-risk models on market stability". In: *Journal of Economic Dynamics and Control* 82 (2017), pp. 223–256.

[53] Andrzej Materka. "Application of artificial neural networks to parameter estimation of dynamical systems". In: *Conference Proceedings. 10th Anniversary. IMTC/94. Advanced Technologies in I & M. 1994 IEEE Instrumentation and Measurement Technolgy Conference (Cat. No. 94CH3424-9)*. IEEE. 1994, pp. 123–126.

[54] Kenji Matsumoto and Ichiro Tsuda. "Noise-induced Order". In: *Journal of Statistical Physics* 31.1 (1983), pp. 87–106.

[55] Piero Mazzarisi, Fabrizio Lillo, and Stefano Marmi. "When panic makes you blind: A chaotic route to systemic risk". In: *Journal of Economic Dynamics and Control* 100 (2019), pp. 176–199.

[56] Welington de Melo and Sebastian van Strien. *One-Dimensional Dynamics*. Vol. 25. Springer Science & Business Media, 2012.

[57] Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.

[58] Kumpati S. Narendra and Kannan Parthasarathy. "Identification and control of dynamical systems using neural networks". In: *IEEE Transactions on Neural Networks* 1.1 (1990), pp. 4–27.

[59] Isaia Nisoli. "How Does Noise Induce Order?" In: *Journal of Statistical Physics* 190.22 (2023).

[60] Tomasz Nowicki and Sebastian van Strien. "Invariant measures exist under a summability condition for unimodal maps". In: *Inventiones Mathematicae* 105.1 (1991), pp. 123–136.

[61] Galo Nuño and Carlos Thomas. "Bank Leverage Cycles". In: *American Economic Journal: Macroeconomics* 9.2 (2016), pp. 37–72.

[62] Avinash Persaud. "Sending the herd off the cliff edge: the disturbing interaction between herding and market-sensitive risk management practices". In: *Journal of Risk Finance* 2.1 (2000), pp. 59–65.

[63] Sebastian Poledna et al. "Leverage-induced systemic risk under Basel II and other credit risk policies". In: *Journal of Banking & Finance* 42 (2014), pp. 199–212.

[64] Rial A Rajagukguk, Raden AA Ramadhan, and Hyun-Jin Lee. "A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power". In: *Energies* 13.24 (2020), p. 6623.

[65] Jitendra R. Raol and Himesh Madhuranath. "Neural network architectures for parameter estimation of dynamical systems". In: *IEE Proceedings-Control Theory and Applications* 143.4 (1996), pp. 387–394.

[66] David Ruelle. "An inequality for the entropy of differentiable maps". In: *Bol. Soc. Bras. Mat.* 9 (1978), pp. 83–87.

[67] Thomas Schreiber. "Extremely simple nonlinear noise-reduction method". In: *Physical Review E* 47.4 (1993), p. 2401.

[68] Paolo Tasca and Stefano Battiston. "Market procyclicality and systemic risk". In: *Quantitative Finance* 16.8 (2016), pp. 1219–1235.

[69] Hans Thunberg. "Periodicity versus Chaos in One-Dimensional Dynamics". In: *SIAM Review* 43.1 (2001), pp. 3–30.

[70] Stefan Thurner. "Systemic financial risk: agent based models to understand the leverage cycle on national scales and its consequences". In: *IfP/fgS Working Paper* 14 (2011).

[71] Daniel Toker, Friedrich T. Sommer, and Mark D'Esposito. "A simple method for detecting chaos in nature". In: *Communications Biology* 3.1 (2020), pp. 1–13.

[72] Marcelo Viana. *Stochastic Dynamics of Deterministic Systems*. Brazillian Math. Colloquium, IMPA, 1997.

[73] Xubin Zeng, Roger A. Pielke, and Richard Eykholt. "Extracting Lyapunov exponents from short time series of low precision". In: *Modern Physics Letters B* 6.02 (1992), pp. 55–75.