CLASSE DI SCIENZE

Tesi di Perfezionamento in Matematica per la Finanza

Models of Dynamical Networks with Applications to Finance

Author:
Domenico Di Gangi

Supervisors:
Prof. Dr. Giacomo Bormetti
Prof. Dr. Fabrizio Lillo

2021/22

Acknowledgements

First of all, I want to thank my supervisors Fabrizio Lillo and Giacomo Bormetti. This thesis benefited enormously from their collaboration and expert advice. I am deeply thankful for their constant support and unlimited patience across the years of my PhD.

I am also grateful to the head of our PhD program, Stefano Marmi, for his support.

A special thanks goes to all my colleagues at the Quantitative Finance group in Pisa. In particular, I thank Giulia for kindly helping me in multiple occasions, Giuseppe for his advises on score driven models and joyful company over multiple meals, and Clemente for always being available for a chat and for his helpful tips on life after a PhD.

I thank my coauthors Carlo, Daniele and Davide who have contributed to some of the works featured in this thesis, as well as my collaborators in other works not included here.

I am also deeply grateful for the selfless support that my family has always provided me.

Finally, I thank my girlfriend Silvia for her constant encouragement and the comfort provided in the hardest moments of my PhD journey. Without her I would have not managed to complete it.

Contents

List of Abbreviations

DGP	**D**ata **G**enerating **P**rocess
PMF	**P**robability **M**ass **F**unction
ERGM	**E**xponential **R**andom **G**raph **M**odel
TERGM	**T**emporal **E**xponential **R**andom **G**raph **M**odel
TGRG	**T**emporally **G**eneralized **R**andom **G**raph
ME	**M**aximum **E**ntropy
MLE	**M**aximum **L**ikelihood **E**stimate
PMLE	**P**seudo **M**aximum **L**ikelihood **E**stimate
SD	**S**core **D**riven
SDERGM	**S**core **D**riven **E**xponential **R**andom **G**raph **M**odel
KIM	**K**inetik **I**sing **M**odel
SDKIM	**S**core **D**riven **K**inetik **I**sing **M**odel
DyNoKIM	**D**ynamic **N**oise **K**inetik **I**sing **M**odel
DyEnKIM	**D**ynamic **E**ndogeneity **K**inetik **I**sing **M**odel
MSE	**M**ean **S**quared **E**rror
MAD	**M**ean **A**bsolute **D**ifference
BIC	**B**ayesian **I**nformation **C**riterion
AUC	**A**rea **U**nder the **C**urve
LM	**L**agrange **M**ultiplier
EBA	**E**uropean **B**anking **A**uthority
FFIEC	**F**ederal **F**inancial **I**nstitutions **E**xamination **C**ouncil

A Silvia

List of Original Contributions

- Domenico Di Gangi, Fabrizio Lillo, and Davide Pirino (2018). "Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction". In: *Journal of Economic Dynamics and Control* 94, pp. 117–141
- Domenico Di Gangi, Giacomo Bormetti, and Fabrizio Lillo (2019). "Score-Driven Exponential Random Graphs: A New Class of Time-Varying Parameter Models for Dynamical Networks". In: *arXiv:1905.10806*
- Domenico Di Gangi, Giacomo Bormetti, and Fabrizio Lillo (2021). "Score-Driven Fitness Model for Sparse Weighted Temporal Networks". In: *preparation*
- Carlo Campajola, Domenico Di Gangi, Fabrizio Lillo, and Daniele Tantari (2020). "Modelling time-varying interactions in complex systems: the Score Driven Kinetic Ising Model". In: *arXiv:2007.15545*

Chapter 1

Introduction

Complex financial, economical and social systems are characterized by fascinating structures and non trivial behaviours that emerge from the interactions of their elementary, often simple parts (Bar-Yam, 2002; Mitchell, 2006). The description of elementary components of complex systems and their relations by means of simple abstractions has very old roots and permeates scientific investigations across various fields (Newman, 2011). Two of the most diffused descriptions of complex systems are formally very simple and, in their most basic version, both rely on sets of binary variables that describe the state of the system at hand.

The first description that we mention consists in describing systems of interacting elements as *graphs*. Describing a complex system as a graph, in its simplest form, amounts to associate a binary variable, called *link*, to the presence or absence of a relation between two entities. Every entity in this setting is referred to as a *node*. When used to describe real world systems, graphs are also called *networks* (Newman, 2010) and in this thesis we use the two names interchangeably. Nodes and links are the only components of a *binary network*, where the only information of interest is whether a relation between two nodes is present or not. When a binary network is complemented with information on the intensity of the pairwise relations, we say that a weight is associated to a link, which is usually in the form of a positive continuous variable. In this cases we speak of *weighted networks*.

A second framework uses a similar description for a complex system as that of the famous Ising model (Ising, 1925), and its dynamical version the Kinetic Ising Model (KIM) (Derrida, Gardner, and Zippelius, 1987; Crisanti and Sompolinsky, 1988). This approach consists in associating one binary variable to each element of the system, known as *spin*, then collectively describing the joint probability mass function exactly as is done in the physics literature. The latter is extremely well known in physics, where it has been studied for decades and declined in multiple versions, for example to explore the emergence of collective phenomena in physical systems. The crucial difference with the applications to complex systems is the meaning associated to each spin and their states. This approach has found various applications in the realm of complexity science as a mean of exploring and understanding spin like data-sets (Nguyen, Zecchina, and Berg, 2017).

Our main focus in this thesis is on the network description of complex systems. The main contributions regard the proposal and exploration of new statistical models to describe static and temporal networks. Nevertheless, we also consider statistical models for spin like descriptions in a dynamical context, and discuss their relations with temporal networks. We provide novel contributions in various streams of literature. We start with the problem of reconstructing systemic risk, arising from possible fire sales spillovers, from partial information on the networks structure. In doing so we take the most diffused approach in the literature of systemic risk reconstruction, and model each snapshot of a temporal network as an individual

static network. In the subsequent chapters we change approach and to take into account explicitly the longitudinal dimension of temporal networks in our modelling approach. Instead of resorting to the sequential application of models for static networks we focus on a dynamical description of a binary temporal networks. Specifically in our second original contribution, we propose an extension of a large and well known family of statistical models for networks, known as Exponential Random Graph Models (ERGM), to a version with time varying parameters, well suited for the description of temporal networks. In this contribution, and also in the following two, we leverage a recently proposed framework to define time varying parameters models often called Score Driven models. After the score driven generalization of ERGMs, we define a new model for weighted temporal networks, that is able to describe weighted networks with abundance of missing links, and to account for the influence of regressors on both the probability of observing a link and the expected weight of present links. In the last contribution included in this thesis, we extend the Kinetic Ising Model to a version with time varying parameters and discuss its application as a statistical model for temporal networks.

In summary, this work contains a total of nine Chapters, this Introduction, three Chapters of literature review, four Chapters presenting original contributions, and a final Chapter of conclusions. In the following we first give an introduction of the main topic of this thesis, i.e. static and temporal networks and statistical models used to describe them. Finally, we briefly summarize the contents presented in each of the following ones.

Although the interest in graphs and graph descriptions of complex problems is quite old (Zermelo, 1929; Erdős and Rényi, 1959; Erdős and Rényi, 1961), in the last two decades we have witnessed an explosion of interest in graphs as simple abstractions to model pairwise relations in complex systems (Barabási, 2002; Cohen and Havlin, 2010). This explosion of interest has been fuelled by the increasing availability of unstructured dyadic data[1] in the form of graphs. On pair with data availability came the realization that real world networks are very often not compatible with the hypothesis of random link formations between each possible pairs of links. This hypothesis has been formalized in the seminal work of Paul Erdős and Alfréd Rényi in Erdős and Rényi, 1959, where they derived the statistical properties of what was the benchmark statistical model for network data for some time, the *Random Graph*. The basic assumption of the random graph model is that each link is an independent binary random variable and they all have equal probability of being 1 or 0 depending on the presence or absence of link, respectively. A first example of a real world network feature that is not compatible with a Random Graph description is the shape of the distribution of the number of connections per node. While from the random graph assumption it follows a Poisson distribution of the number of connections per node, many real world networks are found to be better described by power law distributions and they are often indicated as scale free networks (Song, Havlin, and Makse, 2005; Albert and Barabási, 2002). Another peculiar feature of real world networks is the low number of connections that in average separate two nodes. This turns out to be much lower than what is consistent with a random graph description, and it is called the small world property of networks. Interestingly this was first noticed in the context of social networks by the psychologist Stanley Milgram (Milgram, 1967; Travers and Milgram, 2011), that empirically estimated the average

[1] Data that does not come in the standard tabular form, i.e. as a list of rows describing entities and their features, but describes the presence or absence of relations between all the possible pairs of entities called *nodes*.

distance between two actors to be about 6 steps [2], an empirical result that is often referred to as the *six degrees of separations* "law" also in non technical literature. Those examples of differences between random graphs and observed networks, and many others that we do not mention here, motivated multiple contributions proposing statistical models, or network formation mechanisms, better suited to describe the observed features (Holland and Leinhardt, 1981a; Albert and Barabási, 2002; Watts and Strogatz, 1998; Dorogovtsev, Mendes, and Samukhin, 2003; Park and Newman, 2004a; Caldarelli, 2007). A crucially important family of statistical models for networks is that of ERGMs. That is the family of statistical models over all possible binary networks, with a given number of nodes, whose probability mass function (PMF) belongs to the exponential family, and the sufficient statistics are referred to as network functions[3]. ERGMs have been widely applied and are probably the best known family of statistical models for binary networks[4].

The number of systems that can be described as networks is enormous and an exhaustive list is impossible here. For example, networks permeate our day to day lives, obvious examples being social networks followed by transportation, electric, internet networks. Networks are also ubiquitous in the financial sector, where standard credit relations between pairs of financial entities, e.g. banks, are the standard example of pairwise relations that can be described as financial networks. This and other kinds of financial inter-linkages are considered of paramount importance for financial stability and the mitigation of systemic risk and have attracted an enormous amount of attention in the literature (see Allen and Gale, 2000; Haldane et al., 2009; Gai and Kapadia, 2010; Haldane and May, 2011; Gai, Haldane, and Kapadia, 2011; Mistrulli, 2011; Corsi, Marmi, and Lillo, 2016; Greenwood, Landier, and Thesmar, 2015, among many contributions). Aside direct credit relations, there are several other channels through which financial distress may propagate from one institution to another and, eventually, affect a vast portion of the global economy. One example, that we focus on in this thesis, is that of fire sales spillovers due to assets' illiquidity and common portfolio holdings. In short, the risk of fire sales spillovers arises since shared investments create a significant overlap of portfolios between couples of financial institutions. Such (indirect) financial interconnectedness is an important source of contagion, since partial liquidation of assets by a single market player is expected to affect all other market participants that share with it a large fraction of their own investments (see Corsi, Marmi, and Lillo, 2016; Huang, Vodenska, Havlin, and Stanley, 2013; Caccioli, Shrestha, Moore, and Farmer, 2014; Lillo and Pirino, 2015). Fire sales move prices due to the finite liquidity of assets and to market impact. In a perfectly liquid market there will be no fire sale contagion at all (see Adrian and Shin, 2008, for a review on the role of liquidity in financial contagion). Finally, leverage amplifies such feedbacks. In fact, as described in detail by Adrian and Shin, 2010; Adrian and Shin, 2014, levered institutions continuously rebalance their positions inflating positive and, most importantly, negative assets' price variations. Financial networks are considered so important that large efforts have been conducted at the

[2] The experimental set up based on the U.S. postal service, and kind requests to random citizens to forward mails to all their acquaintances, is a remarkable example of the issue of sampling real world networks and gives an idea of how technological progress is largely the main reason for the increased availability of network data.

[3] As we discuss diffusely in the following, a network is typically associated to a matrix known as *adjacency matrix*, as follows. The nodes are labelled, with indices running from 1 to N, and a link is identified by the pair of nodes it connects (i, j). To each G, we can assign one *adjacency matrix* $\mathbf{A}$ such that $A_{ij} = 1$ if link (i, j) is present in E and $A_{ij} = 0$ otherwise.

[4] As it turns out, the seminal random graph model of Erdős and Rényi, 1959 is the simplest example of this family.

highest regulatory levels to collect large amount of data on various types of financial networks (Giannone, Lenza, Pill, and Reichlin, 2012; Mancini, Ranaldo, and Wrampelmeyer, 2016; Fache Rousová, Jukonis, Letizia, Gravanis, et al., 2020; Lenoci and Letizia, 2021; Ehrmann and Schure, 2020). Interestingly, in the financial context, the possibility to associate a weight to a financial link is probably more important than in many other fields. For example, the difference between a credit worth 1 million and one worth 100 millions should be taken into account in the description of credit networks, especially from the point of view of the propagation of financial distress in case of counter-party default. For this reason many contributions in the literature of financial networks focused on network models that are able to describe also link weights, and to reconstruct them from partial information whenever the full network structure is not available (Mistrulli, 2011; Cimini et al., 2014; Anand, Craig, and Von Peter, 2015; Anand et al., 2017).

Clearly many complex systems that are described as networks are not static. As intuitive examples we can think for example at the internet, where billions of packages are routed everyday following multiple paths, or air transportation networks where links can describe routes between two airports and their weights the traffic associated with them, over a certain time period. The behaviour of networks and the relations between their parts are often inherently dynamical and their description requires more than a static network. This is becoming more and more evident as the diffused collection of network data is allowing to explore also the longitudinal dimension of pairwise relations that change, i.e. the temporal evolution of networks. As more and more data-sets describing networks evolving in time become available, the network description needs to be extended to account also for the time dimension (Holme and Saramäki, 2012). Temporal network data can be described in more than one way, depending on the sampling frequency and the scope of the analysis. In this thesis we consider one of the most popular descriptions for temporal networks that defines them as sequences of time stamped networks evolving in discrete time. Alongside with the representation of temporal network data, also the need to appropriately model them has motivated multiple efforts in the recent literature, and a number of models have been proposed to model binary temporal networks (Hanneke, Fu, Xing, et al., 2010; Krivitsky and Handcock, 2014). Many models for binary temporal networks are related to the ERGM framework. The most well known temporal extension of ERGM is probably the one known as Temporal Exponential Random Graph Model (TERGM) (Robins and Pattison, 2001; Hanneke, Fu, Xing, et al., 2010; Cranmer and Desmarais, 2011). A TERGM can be loosely introduced as an ERGM where the network statistics that define the PMF are allowed to be functions of the current network state of a temporal network, but also of the network at K previous time steps. Where K is an integer number. More precisely, the PMF for the observation at time t in a temporal network, depends on statistics that are functions of the networks that go from time $t - K$ to time t.

The literature on models for temporal weighted networks instead is still quite scarce. This is most certainly due to the additional model complexity required to properly describe the weights. Moreover, most real world networks are found to have an abundance of zero entries that, in the weighted case, require some care in modelling the probability for a link to be present and the distributions of link weights.

In the following we give a brief overview of the contents of each of the following chapters.

Chapter 2 In this Chapter we set the notation used throughout this thesis to indicate static binary and weighted networks and review the literature on statistical models to describe them. In reviewing statistical models for static binary networks we give a detailed introduction of the ERGM family. As mentioned, ERGMs are statistical models for binary networks having an exponential PMF. An ERGM specification is defined by the set of network functions, i.e. functions that associate one scalar number to a network. We discuss inference methods for ERGMs and give a detailed description of two particular specifications. The emphasis on this model class is motivated by the relevance that they have in the literature and by the fact that in Chapter 6 we extend the whole ERGMs family to a temporal version that accommodates time varying parameters.

In the final part of the chapter we review the Maximum Entropy (ME) approach to define models for binary and weighted networks. Similarly to the ERGM family, the maximum entropy approach starts from a set of network statistics. In the context of ME the statistics are introduced as constraints on the probability distribution. In fact, given a set of network statistics and one numerical value for each of them, the so called grand canonical maximum entropy ensemble is defined by the PMF that maximizes the entropy under the constraints that the expected value of the set of network statistics is equal to the mentioned numerical values. This approach is widely used in network reconstruction from partial information, and is an important ingredient of the original contribution that we present in Chapter 3.

Chapter 3 In this Chapter we discuss a novel contribution on reconstruction of systemic risk, generated by the risk of fire sales spillovers, from partial information. Our efforts are motivated by the importance of monitoring and assessing systemic risk in financial markets. That often requires data that are unavailable or available at a very low frequency. For this reason, systemic risk assessment with partial information is potentially very useful for regulators and other stakeholders. We propose to apply the maximum entropy approach to the inference of the network of portfolio weights in order to estimate metrics of systemic risk due to fire sales spillovers. Specifically, we show how indirect vulnerability, systemicness (as defined by Greenwood, Landier, and Thesmar, 2015) and the aggregate systemic risk of US commercial banks can be estimated when only a partial information (the size of each bank and the capitalization of each asset) is available. Differently from the interbank studies (as in Mistrulli, 2011; Mastromatteo, Zarinelli, and Marsili, 2012; Anand, Craig, and Von Peter, 2015) we deal with *bipartite networks*, i.e. networks whose nodes can be divided into two sharply distinguished sets that, in our case, are commercial banks and asset classes. More specifically, we analyze the quarterly networks of US commercial banks' exposures in the period 2001-2013 using the *Federal Financial Institutions Examination Council* (FFIEC) through the *Call Report* files. We compute, for each quarter, systemicness and vulnerability of each bank and the aggregate vulnerability of the system. We compare them with the values inferred assuming the balance sheet compositions of the banks were not known.

The contribution of this chapter is divided into two main parts. First, following a practice that is largely diffused among researchers of both academic institutions and central banks (see, among others, Sheldon and Maurer, 1998; Upper and Worms, 2004; Wells, 2004; Mistrulli, 2011; Sachs, 2014), we reconstruct the matrix of portfolio holdings as such that minimizes the cross entropy (or Kullback-Leibler divergence) from a initial guess. Despite this approach has often been referred to as maximum entropy, or matrix balancing, in order to avoid confusions with different methods discussed in the following, we refer to it as Cross-Entropy method. We show that

this approach does a very good job in our case, providing unbiased estimates of the systemic risk metrics defined by Greenwood, Landier, and Thesmar, 2015. Besides, we show that the reconstructed matrix corresponds to that implied by the Capital Asset Pricing Model, hence it possesses a clear economic meaning.

Second, we compare Cross-Entropy with a different approach to entropy maximization, which allows to define a probability mass function for graphs (ensemble) by maximizing entropy under suitable constraints where some average quantities are set equal to the ones observed in data. Despite the economic intuition of this approach is less sharp than the previous one, the method is widespread in the literature and allows performing scenario generation. We propose a new ensemble, termed MECAPM, which (i) satisfies a set of economically motivated constraints, (ii) behaves in average as the cross entropy method proposed before, and (iii) allows for scenario generation, potentially useful for supervisory authorities to test if a specific institution has increased its systemicness with respect to the past.

Chapter 4 In this Chapter we give an overview of the literature on models for temporal networks. We start with a discussion of models for sequences of binary temporal networks, with particular attention to temporal extensions of the ERGMs class. As mentioned above, probably the most famous extension of ERGMs to the temporal case are the so called Temporal Exponential Random Graphs. This approach builds on ERGMs, but allows the network statistics defining the probability at time t to depend on current and previous networks up to time $t - K$. This K-step Markov assumption is a defining feature of the TERGMs. For example, considering only lag 1 dependencies, a TERGM is defined as a sequence of ERGMs, one for each time step, where the PMF for the network at time t can depend on a set of functions of the current network realization and that at the previous time $t - 1$. The introduction of TERGMs in this Chapter is relevant for our discussion in Chapter 8, where we show how they are related with the Kinetic Ising Model, and to the extensions of KIM that we propose therein.

We then discuss examples from a second class of extensions of ERGMs, those that allow ERGMs parameters to evolve in time. We mention two main contributions. The first one being that of Mazzarisi, Barucca, Lillo, and Tantari, 2020, where the authors consider a fitness model with time varying fitness. They allow each fitness to follow an autoregressive process of order one. Moreover, they combine this random evolution of the fitness with a probabilistic description of a link copying mechanism that captures the possibility of a link to remain exactly the same between two time steps. The second work that we review is the VCERGM of Lee, Li, and Wilson, 2020 that propose what in econometrics jargon can be seen as a smoother for the time varying coefficients of ERGMs. They assume the parameters to be smoothly time varying and decompose their time varying paths in terms of a set of splines basis functions. Finally, they use all the observations from a temporal network to estimate the coefficients of the splines decomposition.

In the last part of the Chapter we review the much smaller literature on models for temporal weighted networks, mainly presenting the proposal of Giraitis, Kapetanios, Wetherilt, and Žikeš, 2016 to use a Tobit model for weighted temporal networks.

Chapter 5 In this Chapter we review score driven econometric models, also known as Dynamic Conditional Score (DCSs) (Harvey, 2013a) or Generalized Auto-regressive

Score models (GAS) (Creal, Koopman, and Lucas, 2013). We start by briefly discussing time varying parameter models in general. Then we give the formal definition of score driven models and discuss their application as misspecified filters with one explicit example. We discuss a theoretical motivation for using the score in the update rule for the parameters. Finally, we conclude discussing how to quantify uncertainty and test for temporal heterogeneity in the score driven framework.

In extreme syntheses, the framework of score driven models can be described as an econometric approach to introduce time-variation on the parameters of a static model. Indeed, a score driven model is defined starting from a probability distribution that, in its static version depends on a set of parameters. The idea is then to promote the static parameters to be time varying by means of an update rule. In this rule a central role is played by the derivative of the log-likelihood, the score, with respect to the parameter that is to be considered time varying. The score driven update rule is defined by an auto-regressive component and a driving term defined by the score and a scaling matrix. Such a choice for the update rule has been found to be optimal from an information theoretical point of view (Blasques, Koopman, and Lucas, 2015), and allows for an easy to apply Lagrange multiplier test for temporal heterogeneity of the parameters (Engle and Russell, 1998; Calvori, Creal, Koopman, and Lucas, 2017).

Interestingly many well known econometric models belong to the score driven class. Probably the most famous example is the Normal specification of the Generalized Auto-regressive Conditional Heteroscedasticity (GARCH) model of Bollerslev, 1986a. Indeed, as we discuss in this Chapter, we can obtain the GARCH update equation for the time varying volatility by starting from a Gaussian distribution and promoting its variance to be time varying following the score driven recipe. Just as the GARCH model, any score driven model can be seen both as a data generating process (DGP) and as a misspecified filter for a generic dynamical evolution of the latent parameters. One of their convenient features is that, when used as filters, they are extremely easy to estimate as the likelihood of a sequence of observations can be usually written in closed form.

Chapter 6 In this Chapter we present a novel approach to modeling time-varying networks. We focus on sets of links among nodes evolving in discrete time, and consider a temporal network as a sequence of networks. The original contribution presented in this chapter stems from a combination of ERGMs with score driven models. Our main contribution is an extension of the ERGMs where each parameter is allowed to evolve in time, according to the auto-regressive updating rule driven by the score of the ERGM distribution. We refer to this class as *Score-Driven Exponential Random Graph Models* (SD-ERGMs). This approach results in a framework for the description of time-varying networks, more than in a single model, in very much the same way as ERGM is considered a modeling framework for static networks. Our method, is able to capture time varying network dependencies and allows for a test discriminating between static or time-varying parameters.

In SD-ERGM, the PMF at each time t is that of an ERGM where the vector of parameters associated with the sufficient statistics is allowed to evolve in time according to the score driven update rule. It follows that a generic SD-ERGM can be used to generate synthetic sequences of graphs, i.e. it can be considered as DGP. Alternatively it can be interpreted as an effective filter of latent time-varying parameters, regardless of what the true DGP might be. With this spirit, in our work we run extensive numerical experiments for three SD-ERGM specifications showing

that SD-ERGM consistently outperforms ERGM in filtering the patterns of simulated time varying ERGM parameters.

Often in ERGMs the PMF depends on a normalization function of the parameters that cannot be computed in closed form. When this is the case, we propose to use the pseudo-likelihood (Strauss and Ikeda, 1990) to compute the score and estimate the static parameters. In this way, our approach can describe the dynamic dependence of the PMF from virtually all the network statistics usually considered in ERGM applications.

One of the ERGM specifications that we discuss in detail is the well known *fitness model* (Caldarelli, Capocci, De Los Rios, and Muñoz, 2002; Garlaschelli and Loffredo, 2008), also known as *beta model* or *configuration model*. That is an ERGM designed to capture the heterogeneity in the number of connections for each node. For the directed network case that we consider, connections have a direction and the fitness model is defined by counting the number of incoming and outgoing connections for each node. This results in a model with two parameters per each node, the in fitness and the out fitness, that capture the tendency of each node to form incoming and outgoing connections. In order to support the proposed modeling framework, we run extensive numerical experiments for three SD-ERGM specifications, showing that SD-ERGM is extremely effective in filtering the patterns of simulated time varying ERGM parameters. We consider a first application to data from the electronic Market of Interbank Deposit (e-MID), a market where banks can extend loans to one another for a specified term and/or collateral. Our dataset contains the list of all credit transactions in each day from June 6, 2009 to February 27, 2015. In our analysis, we investigate the interbank network of overnight loans, aggregated weekly, and following the literature we disregard the size of the exposures, i.e. the weights of the links. This results in a set of $T = 298$ weekly aggregated networks of 132 nodes. By means of a forecasting exercise we show that our score driven fitness model preforms best in both one step and multi steps ahead link prediction, with respect to two considered alternatives. Moreover, our approach allows us to predict links from partial information. In fact, knowing only the current sequences of in and out degrees we can predict the future probabilities of existence for each link. Finally, we presented a second empirical application to the co-voting relations between members of US the congress. For each one of 74 congresses formed between 1867 and 2015, we define a network of co-voting relations, where a link between two senators indicates that they voted in agreement on over 75% of the votes. For this empirical application, we considered an ERGM with two statistics: one measuring total connectivity and one measuring transitivity. We then show that the test for temporal heterogeneity, based on SD-ERGM, allows us assess what parameters are to be described as time varying, and the temporal evolution significantly differs from assuming static parameters.

Chapter 7 In this Chapter, we propose a novel time varying parameter model for sparse weighted temporal networks as a combination of the fitness model, appropriately extended to handle also the weights, and the score driven framework.

While the vast majority of the literature on models for time varying networks focuses on *binary graphs*, i.e. graphs that are defined solely by a set of nodes and a set of links between pairs of nodes, often we can associate a weight to each link. In such cases the data is better described by a weighted, or valued, network. One important well known fact is that real world valued networks are very often found to be sparse, i.e. their adjacency matrices have an abundance of zero entries. That is the case,

for example, of interbank networks (Anand et al., 2017), a class of weighted temporal networks of paramount importance (Allen and Babus, 2011), that are known to be extremely relevant to financial stability (Haldane and May, 2011), and have motivated the application and development of a number of statistical models for networks (Bargigli et al., 2015b; Mazzarisi, Barucca, Lillo, and Tantari, 2017).

Our main contribution is a model for sparse weighted dynamical networks, that also accommodates for the dependency of the network dynamics on external variables, and its application to weighted temporal network data, describing overnight exposures in the European interbank market. Our work contributes to the extremely scarce literature on dynamical models for sparse weighted networks by extending the very well known fitness model for static binary networks. We consider a zero augmented generalized linear model to handle the weights and a state of the art econometric approach to describe time varying parameters. This results in a flexible model that allows us to decouple the probability of a link to exist from its expected weight, and to explore the influence of external regressors on the network's dynamics. We then exploit such flexibility to investigate how the relevance of EONIA rates on the e-Mid interbank market changed over time.

We start from the well known *fitness model* for binary networks and extend it to handle sparse weighted temporal networks, and their dependency on external covariates. Alongside the *binary fitness* parameters, we associate to each node i two new parameters that we name *weighted fitness*. They describe the propensity of a node to have more or less heavy weights in incoming and outgoing links respectively, and use them to model the weighted adjacency matrix using a zero augmented distribution. We then extend this model to the dynamical context by allowing the fitness, both binary and weighted, and the regression parameters to change over time, following the Score Driven approach. We run extensive numerical simulations to make sure that this update rule defines an effective way to *filter* the time varying parameters, also when their temporal evolution is governed by a different DGP. Moreover, we show that the introduction of score driven fitness attenuates estimation errors in cases where some external variables that are relevant for the network dynamics are not available. As an empirical application, we run an exercise in weights forecasting on a portion of the European interbank market, e-Mid. Moreover, we exploit the flexibility of our model to explore the relation between the dynamics of e-Mid and the EONIA rate and to investigate persistence of the weights.

Chapter 8 In this Chapter, we tackle the problem of effectively treating time-varying interactions in Kinetic Ising Models by means of two score driven extensions of the standard KIM (Crisanti and Sompolinsky, 1988). We also discuss the application of KIM and its score driven extensions to model temporal networks.

It is often the case that a system adapts in response to its own behavior or to external inputs. For modeling, this frequently implies the necessity to use computationally costly statistical learning methods or demanding restrictions on sample selection. We introduce an optimal observation-driven framework to treat time-varying interactions within the paradigm of popular complex systems models. Indeed, a common issue when analyzing real-world complex systems is that the interactions between the elements often change over time: this makes it difficult to find optimal models that describe this evolution and that can be estimated from data, particularly when the driving mechanisms are not known. As mentioned, a well known approach to model interactions in complex systems that evolve in time is the KIM. The KIM is a minimalistic pairwise constant interactions model which has found applications in multiple scientific disciplines. It describes the time evolution of a set of

N binary variables, typically called "spins", which can influence each other through a time lagged interaction. We focus on its applications to time series analysis and extend it to allow the presence of time-varying parameters with score-driven dynamics. Keeping arbitrary choices of dynamics to a minimum and seeking information theoretical optimality, the Score-Driven methodology lets us significantly increase the knowledge that can be extracted from data using the simple KIM. In particular, we first identify a parameter whose value at a given time can be directly associated with the local predictability of the dynamics. Then we introduce a method to dynamically filter value of such parameter from the data, without the need of specifying parametrically its dynamics. Finally, we extend our framework to disentangle different sources (e.g. endogenous vs exogenous) of predictability in real time. Additionally, we discuss the possibility to use KIM and our score driven extensions to model temporal networks. In doing so we highlight an interesting relation between KIM and the TERGMs family. To prove the flexibility and various potential applications of our methodology, we apply our models to several complex systems including financial markets, temporal (social) networks, and neuronal populations. Our results show that the Score-Driven KIM produces insightful descriptions of the systems, allowing to predict forecasting accuracy in real time as well as to separate different components of the dynamics. This provides a significant methodological improvement for data analysis in a wide range of disciplines.

Chapter 2

Models for Static Networks

Synopsis of the Chapter In this chapter we set the notation used throughout this thesis to indicate static binary and weighted networks and introduce statistical models to describe them. We review the main statistical models for static networks available in the literature starting with the family of Exponential Random Graphs Models (ERGMs). We discuss inference methods for ERGMs and give a detailed description of two particular specifications. The emphasis on this model class is motivated by the relevance that they have in the literature and by the fact that in Chapter 6 we extend to a temporal version that accommodates time varying parameters. We then present the latent parameters modelling framework, and finally introduce the Maximum Entropy approach to define models for binary and weighted networks. This approach is widely used in network reconstruction from partial information, as we discuss in the original contributions presented in Chapter 3.

2.1 Definitions

A network, or graph[1], is a useful abstraction for a system composed by a number of single elements that have some pairwise relation among them. The simplified description of social, economic, biological, transportation systems, in terms of nodes and links attracted and still attracts an enormous amount of attention, in a number of different streams of literature (Albert and Barabási, 2002; Bullmore and Sporns, 2009; Newman, 2010; Jackson, 2010; Easley, Kleinberg, et al., 2010; Allen and Babus, 2011).

Formally, a graph G is a pair (V, E) where V is a set of N nodes and E is a set of node pairs named links, or edges. The nodes are labelled, with indices running from 1 to N, and a link is identified by the pair of nodes it connects (i, j). To each G, we can assign one *adjacency matrix* $\mathbf{A}$ such that $A_{ij} = 1$ if link (i, j) is present in E and $A_{ij} = 0$ otherwise. If the links have a direction, i.e. link (i, j) describes different information than (j, i), than we speak of a *directed* graph. Otherwise the graph is said to be *undirected* and the adjacency matrix is symmetrical, i.e. $A_{ij} = A_{ji} \quad \forall i, j = 1, \dots, N$. While the majority of the literature focuses on *binary graphs*, i.e. graphs that are defined solely by a set of nodes and a set of links between pairs of nodes, often more information is available to describe pairwise relations, and in many cases we can associate a weight to each link. When this is the case, we speak of *weighted networks*, to distinguish them from *binary networks*, where the only information available is whether a link is present or not. Weights are typically positive, discrete or continuous, numbers and can be associated, for example to the strength of the relation described by each link. In these cases the data is better described by a weighted, or valued, network, that can be associated with a positive, real valued matrix $\mathbf{Y}$ such

[1] The two names are used interchangeably in this work.

that $Y_{ij} \in \mathcal{R}^+ \quad \forall i = 1, \ldots, N, j = 1, \ldots, N$. Where Y_{ij} is the value of the link between node i and node j, and $Y_{ij} = 0$ if the link is not present.

It is well known that real world networks, both weighted and binary, are often found to be sparse. Technically a network is said to be sparse if the number of links is of the order of the number of nodes N, but more loosely a sparse network is one where majority of the possible links are not present. This fact is especially important for weighted networks as it often requires a separated description for a link's presence and it's expected weight.

Since in this thesis we consider both weighted and binary networks, whenever we state something that holds true both for binary and weighted networks we use the $\mathbf{Y}$ notation to indicate the adjacency matrix, and leave the $\mathbf{A}$ notation for cases where the network considered is restricted to be binary. Given a weighted network, described by $\mathbf{Y}$ we can always neglect the weights and consider only the presence or absence of a link. In terms of the matrix description, this amounts to considering the binary adjacency matrix $\mathbf{A}$, with elements $A_{ij} = \Theta\left(Y_{ij}\right)$, where the function Θ is known as *Heaviside* or *indicator* function and is defined to be zero when its argument is less or equal then zero and one otherwise,

$$\Theta\left(y\right) = \begin{cases} 0 & if \quad y \leq 0, \\ 1 & if \quad y > 0. \end{cases} \tag{2.1}$$

Given a network representation of data describing pairwise relations, we can analyze it with the tools of probability and statistics. That typically entails the definition of a network model, that are the main focus of this chapter. Our main reference for the formal definition of a network model is the the book by Kolaczyk, 2009b. A network model is defined by an *ensemble* $\mathcal{G}$ (of possible graphs) and a family $\text{Prob}_{\boldsymbol{\vartheta}}\left(G\right)$ of probability distribution indexed by some parameters

$$\left\{ \text{Prob}_{\boldsymbol{\vartheta}}\left(G\right), G \in \mathcal{G} : \boldsymbol{\vartheta} \in \Xi \right\},$$

where $\mathcal{G}$ is a countable set whose element G are graphs and $\text{Prob}_{\boldsymbol{\vartheta}}\left(G\right)$ is a probability density function on $\mathcal{G}$, that is a function

$$\text{Prob}_{\boldsymbol{\vartheta}} : \mathcal{G} \rightarrow [0, 1]$$

such that $\sum_{G \in \mathcal{G}} \text{Prob}_{\boldsymbol{\vartheta}}\left(G\right) = 1$. Such a probability is allowed to depend on a vector of real parameters $\boldsymbol{\vartheta} \in \Xi$, where Ξ is a convex subset of $\mathbb{R}^C$, with C the total number of parameters (or the dimension) of the model. For an arbitrary, vector valued, network statistic $q : \mathcal{G} \rightarrow \mathbb{R}^M$, where M is the number of elements of q, defined on the set $\mathcal{G}$, the expected value of q on the ensemble $\mathcal{G}$ is defined as

$$\mathbb{E}_{\mathcal{G}}\left[q\right] = \sum_{G \in \mathcal{G}} q\left(G\right) \text{Prob}_{\boldsymbol{\vartheta}}\left(G\right). \tag{2.2}$$

A model can be defined by explicitly giving the ensemble, the probability density function along with the space Ξ of the parameters, or by inducing $\text{Prob}_{\boldsymbol{\vartheta}}$ through the recurrent application of some generative mechanism or rule, either starting from an empty graph or by applying a randomization procedure to a reference graph.

Since there is a one to one correspondence between a graph and the appropriate version of the adjacency matrix $\mathbf{Y} \leftrightarrow G$, network models are usually defined through $\mathbf{Y}$, and network statistics are simply functions of the adjacency matrix $h\left(G\right) = h\left(\mathbf{Y}\right)$. This means that the probability distribution can be written as a function of the matrix

associated with a graph

$$\text{Prob}_{\vartheta}(G) = \text{Prob}_{\vartheta}(\mathbf{Y}), \quad \text{for} \quad \mathbf{Y} \leftrightarrow G.$$

Statistical models play a central role in the study of real world network data and have been applied in many different contexts. For example, network models are used to compare and test different possible generative mechanisms (Albert and Barabási, 2002; Watts and Strogatz, 1998). They can be exploited to estimate the relevance of a network feature in explaining observed network data (Garlaschelli, Hollander, and Roccaverde, 2016). Additionally, link probabilities and expected values, obtained from network models, are used to predict the possibility of link to be observed in the future (Martínez, Berzal, and Cubero, 2016; Wang, Xu, Wu, and Zhou, 2015), or to be present in a partially observed network (Anand et al., 2017). Last but not least, network models can be used as null models to find unlikely patterns that might reveal meaningful information about the underlying system (Serrano, Boguñá, and Vespignani, 2009; Tumminello et al., 2011).

2.2 Exponential Random Graphs

In general, for a binary network, if the distribution belongs to the exponential family, then the model is named Exponential Random Graph Model (ERGM). In this section we review ERGMs in some details and discuss two examples that are relevant for the original contributions presented in Chapters 6 and 7.

To introduce ERGMs, let us mention the first and probably most famous example of this class: the Erdös-Rényi random graph model of Erdős and Rényi, 1959. In this model, fixed the number of nodes N, each of the possible $N(N-1)/2$ links [2] is present with constant probability p, equal for all links. The probability to observe the adjacency matrix $\mathbf{A}$ is

$$P(\mathbf{A}) = \prod_{i<j} p^{A_{ij}} (1-p)^{(1-A_{ij})}. \tag{2.3}$$

In the context of exponential distributions, it is possible to consider more general structures for the probability of a link to be present, and even depart from the assumption that each link is independent from the others.

Examples of more general ERGMs have been first proposed by Holland and Leinhardt, 1981a, under the name of log-linear, or $p^{\star}$, models. Specifically, they named $p1$ the model defined by

$$\log P(\mathbf{A}) = \sum_{ij} \left[A_{ij}A_{ji}\rho_{ij} + A_{ji}\phi_{ij}\right] - \log\left(\mathcal{K}(\boldsymbol{\rho}, \boldsymbol{\phi})\right), \tag{2.4}$$

where $\boldsymbol{\rho}$ and $\boldsymbol{\phi}$ are two matrices of parameters, and $\mathcal{K}(\boldsymbol{\rho}, \boldsymbol{\phi})$ is a normalization factor[3], that ensures that the probabilities defined over all the possible adjacency matrices sum to one. This model can be estimated in parsimonious specifications, e.g. $\phi_{ij} = \phi_i + \phi_j$, known as *sender plus receiver effect*, and $\rho_{ij} = \rho$ that describes the tendency to reciprocate links. Additionally, p1 models can be enriched with dependencies on node attributes (Fienberg and Wasserman, 1981) or predetermined

[2] In this thesis, we do not consider links that start and end at the same node, so named *self-loops*. However, including them in would be trivial.

[3] Also known as partition function in the statistical physics literature.

(exogenous or endogenous) covariates X_{ij} (Wasserman and Pattison, 1996). The requirement of independence among dyads has been relaxed since Frank and Strauss, 1986 in order to take into account neighborhood effects, such as the tendency to form 2 stars, quantified by the function $q_{\text{2-stars}} = \sum_{ijk} A_{ik}A_{jk}$ or triangles $q_{\text{triangles}} = \sum_{ijk} A_{ik}A_{kj}A_{ji}$. These functions are examples of network statistics, i.e. functions of the adjacency matrix, that play a central role in ERGMs.

ERGMs can be seen as an application of the family of discrete exponential distributions (Barndorff-Nielsen, 2014) to the description of graphs. The sufficient statistics, known as network statistics, are functions of the adjacency matrix $q_s(\mathbf{A})$ for $q = 1, \dots, Q$ and the PMF is defined by

$$P(\mathbf{A}|\theta) = \frac{e^{\sum_s \theta_s q_s(\mathbf{A})}}{\mathcal{K}(\theta)}. \tag{2.5}$$

where $\theta \in \mathcal{R}^Q$ is the vector of parameters whose component θ_s is associated with the network statistic $q_s(\mathbf{A})$, and the normalization $\mathcal{K}$ is defined by

$$\mathcal{K}(\theta) = \sum_{\{\mathbf{A}\}} e^{\theta_s q_s(\mathbf{A})}.$$

The normalizing factor $\mathcal{K}(\theta)$ is often not available as a closed-form function of the parameters θ. Each matrix element is a binary random variable, and its probability depends only on the value of the network statistics appearing in (2.5).

The literature on ERGMs is extremely vast and still growing (see Schweinberger, Krivitsky, Butts, and Stewart, 2020, for a recent literature review). Without being exhaustive, in the following, we introduce two specific examples of ERGMs. They describe distinct features of the network and require different approaches to the parameter inference. The first statistic we will consider is meant to capture the heterogeneity in the number of connections that each link can have. It allows for straightforward maximum likelihood estimation. The second one describes transitivity in the formation of links, i.e. the tendency of connected nodes to have common neighbors. For this case, inference is instead complicated by the fact that the normalizing factor in (2.5) as a function of the parameters is not available in closed-form. The choice of these examples is instrumental to the main focus of this work, i.e. the time-varying parameter extension of the general ERGM in (2.5). In fact, they allow us to discuss different estimation techniques that will be crucial for our methodology: Maximum Likelihood Estimation (MLE) for node specific parameters and approximate pseudo-likelihood inference.

2.2.1 The Fitness / Beta / Configuration Model

The first example we consider is quite simple but, at the same time, largely employed in different streams of literature (Zermelo, 1929; Bradley and Terry, 1952; Holland and Leinhardt, 1981a; Caldarelli, Capocci, De Los Rios, and Muñoz, 2002; Park and Newman, 2004b; Garlaschelli and Loffredo, 2008; Chatterjee, Diaconis, Sly, et al., 2011). The range of applications for this model is so broad that researchers were often not aware of previous works using exactly the same model. For this reason it can be found under at least three different names: *beta model*, *fitness model*, and *configuration model*. They all refer to a probability distribution that can be rewritten as an ERGM where each node i has two parameters: $\overrightarrow{\theta}_i$, that captures the propensity of node i to form outgoing connections, and $\overleftarrow{\theta}_i$ those incoming. It is standard to

indicate the number of connections a node has as its *degree*. For the directed network case considered here, we have – for node i – *out-degree* $\overrightarrow{D}_i$ and *in-degree* $\overleftarrow{D}_i$ defined as

$$\overrightarrow{D}_i = \sum_j A_{ij} \,, \qquad \overleftarrow{D}_i = \sum_j A_{ji} \,.$$

With these definitions, and since it is possible to compute the normalization factor $\mathcal{K}\left(\overleftarrow{\theta}, \overrightarrow{\theta}\right)$, the log-PMF for the whole matrix in terms of the degrees, reads

$$\log P(\mathbf{A}) = \sum_{i=1}^{N} \left(\overleftarrow{\theta}_i \overleftarrow{D}_i + \overrightarrow{\theta}_i \overrightarrow{D}_i \right) - \sum_{ij} \log \left(1 + e^{\overleftarrow{\theta}_i + \overrightarrow{\theta}_j} \right), \tag{2.6}$$

and the probability for link (i, j) to be observed is

$$P\left(A_{ij} = 1 | \overleftarrow{\theta}_i, \overrightarrow{\theta}_j \right) = \frac{1}{1 + e^{-\left(\overleftarrow{\theta}_i + \overrightarrow{\theta}_j \right)}}.$$

This formulation is often used when the heterogeneity in the degrees is expected to play a prominent role in explaining the presence or absence of links. It is worth to notice that the static version of the *fitness model*, in the directed case, is not identified. If we add any constant to each $\overleftarrow{\theta}_i$ and subtract it from each $\overrightarrow{\theta}_i$, the PMF remains unchanged. To fix it, one needs to introduce an identification restriction. This is essential, to compare the parameter values estimated for different observations of the same network.

Interestingly, the MLE can be performed using a fixed point algorithm, described for example in Yan, Leng, Zhu, et al., 2016, that reaches the optimal solution in a fast way. Moreover, we point out the existence of interesting results on the asymptotic behavior of the maximum likelihood estimates for $\left(\overleftarrow{\theta}, \overrightarrow{\theta}\right)$ when the number of nodes increases. Indeed, consistency results have been proved in Chatterjee, Diaconis, Sly, et al., 2011 for the undirected case and in Yan, Leng, Zhu, et al., 2016 for the directed case (see also Graham, 2017; Yan, Jiang, Fienberg, and Leng, 2018; Jochmans, 2018, for discussions of the statistical properties of the beta model). A necessary condition for these results to hold is that the network density remains constant as N increases. An alternative, and often more realistic, possibility is that the average degree remains constant when N increases, implying that the density decreases as[4] $1/N$. Networks belonging to this density regime are named *sparse*. Notably, no consistency results are known for large N in the sparse regime.

2.2.2 Curved Exponential Random Graphs

Our second example is going to motivate the discussion of the next section on how to deal with cases when the normalization function $\mathcal{K}(\theta)$ is not available as a closed-form function of the parameters θ. It is well known that, when network statistics involve products of matrix elements[5], this is often the case. This lack of analytical tractability has been arguably the main obstacle in estimation and understanding of the properties of ERGMs. Moreover, it is nowadays well known that, when dealing with ERGMs, the use of network statistics involving products of matrix elements,

[4] For a network with N nodes, the number of possible links is of order N^2. Instead, when all nodes have a fixed average degree d, the number of present link is dN, and the density is of order $1/N$.

[5] Examples of such statistics are the count of 2 stars present in the network, or the number of triangles (Wasserman and Pattison, 1996).

such as the number of triangles, requires some care, in order to avoid statistical issues (as discussed for example in Handcock, 2003a; Handcock, 2003b). The main issue, with consequences on estimation, simulation, and interpretability of ERGMs, is known as *degeneracy*. An ERGM is degenerate if it concentrates a large portion of its probability on a small set of configurations, typically the uninteresting graphs that are completely connected or void of links. When this phenomenon occurs, estimating the model becomes very hard, and often the estimated model does not provide a meaningful description of real networks. Indeed, a great effort has been dedicated to investigating this problem, and characterizing degeneracy (see for example Schweinberger, 2011, and references therein). Interestingly, a subset of the possible ERGM specifications have been found to be able to describe properties of the whole network without being subject to the degeneracy issue. They are known as *curved exponential random graphs* and are defined as standard exponential random graphs with the additional constraint that the parameters are required to live in a certain region of the parameter space (Snijders, Pattison, Robins, and Handcock, 2006; Robins et al., 2007; Hunter and Handcock, 2006). In practice that amounts to consider a weighted sum of simple network statistics as a single network statistic, depending on a small number of parameters, instead of considering one parameter for each element of the sum. One example of this kind of statistics is the Geometrically Weighted Edgewise Shared Partners (GWESP). This function has recently been applied extensively to describe transitivity in social networks (see Hunter and Handcock, 2006). It captures the tendency of nodes to form triangles, without the degeneracy issues that emerge when the direct triangle count is used as a statistic in ERGM. To get an intuition of the formula defining GWESP, let us consider two nodes, that are connected by an edge, and count the number of nodes to which they are both connected, i.e. the number of neighbors that they share. Let us indicate with $\mathrm{ESP}_k(\mathbf{A})$ the number of edgewise shared partners, i.e. connected node pairs [6] that share exactly k neighbors in the network described by $\mathbf{A}$. Then GWESP is defined as

$$\mathrm{GWESP}(\mathbf{A}, \lambda) = e^{\lambda} \sum_{k=1}^{n-2} \left[1 - \left(1 - e^{-\lambda} \right)^k \right] \mathrm{ESP}_k(\mathbf{A}) \ .$$

In the following, we will stick to the usual approach in the literature treating the parameter λ as fixed and known, i.e. $\lambda = 0.5$.

2.2.3 Inference

To conclude this partial overview of ERGMs, we discuss parameter inference when the likelihood is not available in closed-form. The two standard approaches to ERGM inference consist in maximizing alternative functions that are known to share the same optimum as the exact likelihood. The first possibility (described, for example, in Snijders, 2002) is to maximize an objective function obtained from a sufficiently large sample drawn from the PMF with an arbitrary (but close enough to the true one) parameter. As a consequence of the non independence of the links in the general ERGM, sampling from (2.5) necessary relies on Markov Chain Monte Carlo (MCMC) approaches (see Hunter et al., 2008, for a description of a popular software that implements it). The computational burden of MCMC-based estimation can be prohibitive for large enough graphs. For this reason, a second approximate inference procedure, known as Maximum Pseudo-Likelihood Estimation (MPLE), first proposed for ERGMs in the seminal work of Strauss and Ikeda, 1990,

[6] *Edgewise* precisely means that we count partners only if shared by nodes that are connected.

is often used in empirical applications. MPLE is based on the optimization of the pseudo-likelihood function, that is in turn defined from link specific variables (one for each element of the adjacency matrix) named *change statistics*. Given an ERGM, the change statistic for the link between node j and i, associated with network statistic q_s is $\delta_{ij}^s = q_s\left(A_{ij}{}^+\right) - q_s\left(A_{ij}{}^-\right)$, where $A_{ij}{}^+$ is a matrix such that $A_{ij}{}^+ = 1$ and it is equal to $\mathbf{A}$ in all other elements. Similarly, $A_{ij}{}^-$ has $A_{ij}^- = 0$ and it is equal to $\mathbf{A}$ in all other entries. Given these definitions, the pseudo-likelihood reads

$$PL\left(\mathbf{A}\right) = \prod_{ij} \pi_{ij}^{A_{ij}} \left(1 - \pi_{ij}\right)^{\left(1 - A_{ij}\right)} \tag{2.7}$$

where $\pi_{ij} = \left(1 + e^{-\sum_s \theta_s \delta_{ij}^s}\right)^{-1}$. The maximum pseudo-likelihood estimates correspond to the parameter values θ that maximize the pseudo-likelihood. On one hand, obtaining the pesudo-likelihood estimates is extremely faster than the MLE based on MCMC, and easy to implement, since the pseudo-likelihood in (2.7) is equal to the likelihood of a logistic regression, and it can be easily maximized with standard software for logistic regressions. On the other hand, the analogy with logistic regression is typically pushed too far. It has become widespread malpractice to associate to MPLEs the confidence intervals obtained from the maximum-likelihood theory for logistic regressions, that are known to be theoretically unjustified, as already noted in Strauss and Ikeda, 1990 and Handcock, 2003b, and thoroughly discussed in Varin, Reid, and Firth, 2011. It is nowadays common knowledge that such a naive approach to MPLE inference results in a systematic underestimation of confidence intervals' width (see, for example, Van Duijn, Gile, and Handcock, 2009; Desmarais and Cranmer, 2012; Schmid and Desmarais, 2017). More principled methods to estimate uncertainties of MPLEs, based on non-parametric and parametric bootstrap, have been proposed in Desmarais and Cranmer, 2012 and Schmid and Desmarais, 2017, respectively. These contributions clearly showed that the computational convenience of MPLE for ERGMs can indeed be reconciled with a reliable estimation of statistical uncertainties.

2.3 Latent Space Models

Another class of models for binary and weighted networks, that has received much attention in the literature, is that of Latent Space Models (Hoff, Raftery, and Handcock, 2002). Although, in the binary case, they formally belong to the class of ERGMs, latent space models have been used also for weighted networks and are typically discussed separately from ERGMs. In the following, for ease of exposition, we review the main examples of the latent variables family of models, only for binary networks. Such models associate to each node a position in a K dimensional latent space. Then for each pair of nodes they compute a measure of distance, or similarity, between the respective latent variables and assign higher probabilities to links between nodes that are closer or more similar. This geometry-based assumption, allows the latent space models to provide the useful visualization and interpretation of network or data, and for thus have been used widely in the literature.

The general definition of a latent space model associates to each node i a latent variable $z_i \in R^K$ and assumes that the probability of observing a given value for each link is independent from that of the values of other links, conditionally on the

latent variables. Hence, the probability mass function takes the following form

$$P\left(\mathbf{A}|z,\mathbf{X},\theta\right)=\prod_{ij}P\left(A_{ij}|z_i,z_j,X_{ij},\theta\right),$$

where $\mathbf{X}$ observed link-specific covariates and θ denotes the corresponding set of regression parameters. There are mainly two families of latent space models, that differ on how the latent variables enter the probability mass function, the distance model and the projection model. They can be applied to both directed and undirected networks, with different interpretations of the parameters. For ease of discussion, we introduce here only their version for directed networks. In the distance model the probability of link (i,j) depends on the Euclidean distance between vectors z_i and z_j. For example, in the binary network it takes the following form

$$\log P\left(A_{ij}|z_i,z_j,X_{ij},\theta\right)=\theta_0+\theta_1 X_{ij}-\left\|z_i-z_j\right\|.$$

Hence, for a given value of the external covariates and the respective parameters, the probability of node to exist tends to be low for nodes that are further apart. While in the projection model the probability of link (i,j) depends on the angle between the two latent parameters' vectors:

$$\log P\left(A_{ij}|z_i,z_j,X_{ij},\theta\right)=\theta_0+\theta_1 X_{ij}-\frac{z_i'z_j}{\left\|z_j\right\|}.$$

Thus, in the projection version of the model the probability of observing link (i,j) increases if the two vectors point in the same direction. Both versions of the latent space model can be easily applied to weighted networks by simply relating the above definitions to expected link weights by means of appropriate link functions. Moreover, various extensions of the original latent space models of Hoff, Raftery, and Handcock, 2002, for example to model explicitly nodes' clusters, have been considered in the literature as reviewed for example in Kim, Lee, Xue, Niu, et al., 2018.

2.4 Maximum Entropy Models

The ERGM framework is intrinsically linked to the very well known *principle of maximum entropy* (Shannon, 1948) and its applications to statistical physics (Jaynes, 1957). Indeed, an ERGM with sufficient statistics $q\left(\theta\right)$ naturally arises when looking for the probability distribution, over the set of all possible binary networks, which maximizes the entropy under a linear equality constraint on the statistics $q\left(\theta\right)$ (Park and Newman, 2004b; Garlaschelli and Loffredo, 2008).

In this section we review the Maximum Entropy (ME) approach to define network models, that will be relevant for the original contribution presented in Chapter 3. Since the maximum entropy principle allows for the definition of a more general set of network models, not restricted to binary networks, here we review it as a mean to define statistical models for weighted networks, thus including binary networks as a subset. Hence in this section we denote the adjacency matrix as $\mathbf{Y}$ and the maximum entropy methodology that we review can and has been applied to both binary and weighted networks (Park and Newman, 2004b; Garlaschelli and Loffredo, 2009).

In what follows we indicate with $S\left(\mathbb{P}\right)\left[\mathbf{Y}\right]$ the Shannon's entropy function

$$S\left(\mathbb{P}\left[\mathbf{Y}\right]\right)=-\sum_{\mathbf{Y}\in\mathcal{Y}}\mathbb{P}\left[\mathbf{Y}\right]\log\left(\mathbb{P}\left[\mathbf{Y}\right]\right), \tag{2.8}$$

and with $\phi_i(\mathbf{Y})$, $i = 1, ..., I$, a set of functions of the adjacency matrix, that describe the information that we want to use to define the maximum entropy model. To different sets of functions correspond different maximum entropy models. In its most general formulation, the ME principle postulates to obtain the probability mass function $\mathbb{P}$ as that which maximizes (2.8), subject to the normalization constraint

$$\sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}[\mathbf{Y}] = 1,$$

and, to further additional constraints on the values of $\phi_i(\mathbf{Y})$, $i = 1, ..., I$.

There are two ways of imposing the constraints. In the first one, termed *microcanonical ensemble*, constraints are imposed *exactly*, i.e. only the graphs fulfilling all the constraints have non zero probability. In the second one, termed *canonical ensemble*, all the graphs have non zero probability and the constraints are satisfied on average over the distribution. There are advantages and disadvantages in both approaches. The microcanonical ensemble can be more realistic in some applications. On the contrary, in the canonical ensemble also graphs where these values are very different from the real data might have non zero probability. Applications of the maximum entropy principle typically consider the canonical ensemble (see Bargigli and Gallegati, 2011; Squartini, Fagiolo, and Garlaschelli, 2011; Fagiolo, Squartini, and Garlaschelli, 2013; Mastrandrea, Squartini, Fagiolo, and Garlaschelli, 2014; Saracco, Di Clemente, Gabrielli, and Squartini, 2015; Almog, Squartini, and Garlaschelli, 2017, for example) for at least two reasons. First, solving the problem in the microcanonical ensemble is typically extremely hard or it requires extensive numerical simulations, randomizing the network by allowing moves that preserve all the constraints. On the contrary, canonical ensemble can often be obtained much more directly, as testified by their widespread use in Statistical Mechanics (Huang, 2008). Moreover when other constraints are added to the optimization problem, microcanonical ME becomes intractable, thus limiting their practical use when regulators want to add additional knowledge on the system. Second, the flexibility of canonical ME allows exploring the relative role of information set and constraints in network reconstruction.

In the canonical ensemble, the probability distribution is defined such that those functions are required to have fixed expected values $\phi_i^\star$. Examples of $\phi_i(\mathbf{Y})$ are $\phi_{nk} = Y_{nk}$ and $\phi_n = \sum_k Y_{nk}$ or $\phi_k = \sum_n Y_{nk}$. The ME approach consists in defining the probability mass function which solves the constrained optimization problem

$$\begin{aligned} \max_{\mathbb{P}(\mathbf{Y})} \quad & S(\mathbb{P}[\mathbf{Y}]) \\ \text{s.t.} \quad & \sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}(\mathbf{Y}) = 1, \\ & \mathbb{E}_{\mathcal{Y}}[\phi_i] = \phi_i^\star,\ i = 1, ..., I. \end{aligned} \tag{2.9}$$

If each matrix element can take only a finite set of values, i.e. the space $\mathcal{Y}$ is finite, the optimization problem is easily solved, at least formally, using Lagrange multipliers. The Lagrangian associated to the problem is written as

$$\mathcal{L} = S(\mathbb{P}[\mathbf{Y}]) + \alpha\left(1 - \sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}(\mathbf{Y})\right) + \sum_{i=1}^{I} \vartheta_i \left(\phi_i^\star - \sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}(\mathbf{Y})\,\phi_i(\mathbf{Y})\right),$$

where α and ϑ_i are Lagrange multipliers. Taking the first derivative w.r.t. $\mathbb{P}(x) \equiv \mathbb{P}(Y = x)$, where x indicates an element of $\mathcal{Y}$, we get[7]

$$\frac{\partial \mathcal{L}}{\partial \mathbb{P}(x)} = -\log(\mathbb{P}(x)) - 1 - \alpha - \sum_{i=1}^{I} \vartheta_i \phi_i = 0,$$

whose solution is

$$\mathbb{P}_{\boldsymbol{\vartheta}}(\mathbf{Y}) = \frac{e^{-\sum_{i=1}^{I} \vartheta_i \phi_i(\mathbf{Y})}}{K_{\boldsymbol{\vartheta}}}, \tag{2.10}$$

where ϑ indicates the set of all Lagrange multipliers, and $K_{\boldsymbol{\vartheta}}$ is a normalizing factor given by

$$K_{\boldsymbol{\vartheta}} = \sum_{\mathbf{Y} \in \mathcal{Y}} e^{-\sum_{i=1}^{I} \vartheta_i \phi_i(\mathbf{Y})}. \tag{2.11}$$

Interestingly, in the case of binary networks, the ME PMF obtained above corresponds to the ERGM specification having as sufficient statistics the functions $\phi_i(\mathbf{Y})$, $i = 1, ..., I$ used in defining the ME model. Given a set of constraints $\phi_i^\star$ the corresponding Lagrange multipliers $\boldsymbol{\vartheta}^\star$ are obtained as the (unique[8]) solution to

$$\sum_{\mathbf{Y} \in \mathcal{Y}} \mathbb{P}_{\boldsymbol{\vartheta}^\star}(\mathbf{Y}) \phi_i(\mathbf{Y}) = \phi_i^\star, \; i = 1, ..., I.$$

The latter parameters are determined, either analytically or numerically, from the information codified in the values $\phi_i^\star$. In the following, when not strictly required, we will omit the dependency of $\mathbb{P}$ on $\boldsymbol{\vartheta}$, to ease the notation.

Let us consider models that allow Y_{nk} to assume any positive integer values, in which case the Lagrange multipliers derivation of the probability mass function is only heuristic. Nevertheless, it is well known (see for example Campbell, 1970; Barndorff-Nielsen, 2014) that, even in the infinite case, the maximum entropy probability mass function is the one in (2.10) . In fact, it can be easily shown that every other probability mass function, satisfying the constraints in (2.9), has lower entropy than the one obtained with Lagrange multipliers.

2.4.1 Maximum Entropy Model Fixing Strengths

An interesting application of the maximum entropy principle to undirected weighted networks, considered in Park and Newman, 2004b and Garlaschelli, Hollander, and Roccaverde, 2016, is the one where the constrained network functions are the strengths of all nodes, $\phi_n = \sum_k Y_{nk}$, in a setting where links' weights can take only positive integer values.

Assigning one Lagrange multiplier λ_n to each ϕ_n, we can go on with the computation by explicitly carrying out the sum in the definition of the partition function,

[7] We stress that we are considering the derivative w.r.t. the probability of each of the (finite) possible realizations of $\mathbf{Y} \in \mathcal{Y}$.

[8] The uniqueness of the solution is well-known in maximum entropy literature.

as follows

$$\begin{aligned} K_{\vartheta} &= \sum_{\mathbf{Y}\in\mathcal{Y}} \prod_{n=1}^{N}\prod_{k=1}^{N} e^{-Y_{nk}\,(\lambda_n+\lambda_k)} \\ &= a \sum_{Y_{1,1}=0}^{\infty} \cdots \sum_{Y_{1,N}=0}^{\infty} \cdots \sum_{Y_{n,1}=0}^{\infty} \cdots \sum_{Y_{n,N}=0}^{\infty} \prod_{n=1}^{N}\prod_{k=1}^{N} e^{-Y_{n,N}\,(\lambda_n+\lambda_k)} \\ &= \prod_{n=1}^{N}\prod_{k=1}^{N} \sum_{Y_{n,k}=0}^{\infty} e^{-Y_{n,k}\,(\lambda_n+\lambda_k)} \\ &= \prod_{n=1}^{N}\prod_{k=1}^{N} \frac{1}{1-e^{-(\lambda_n+\lambda_k)}} = \prod_{n=1}^{N}\prod_{k=1}^{N} \frac{1}{1-\varphi_n\,\varphi_k}, \end{aligned}$$

where $\varphi_n = e^{-\lambda_n}$, hence

$$\begin{aligned} \mathbb{P}\left(\mathbf{Y}\right) &= \frac{\prod_{n=1}^{N}\prod_{k=1}^{K} e^{-Y_{nk}\,(\lambda_n+\lambda_k)}}{\prod_{n=1}^{N}\prod_{k=1}^{K} \frac{1}{1-\varphi_n\,\varphi_k}} \\ &= \prod_{n=1}^{N}\prod_{k=1}^{N} \left(\varphi_n\,\varphi_k\right)^{Y_{nk}}\,\left(1-\varphi_n\,\varphi_k\right), \end{aligned} \tag{2.12}$$

The value of the Lagrange multipliers are determined by imposing that the expected values of $\sum_k Y_{nk}$ on the ensemble $\mathcal{Y}$ are equal to $\phi_n^{\star}$. Note that K_{ϑ} is such that

$$\mathbb{E}_{\mathcal{Y}}\left[\phi_n\right] = -\frac{\partial \log\left(K_{\vartheta}\right)}{\partial \lambda_n}.$$

Hence we can compute $\mathbb{E}_{\mathcal{Y}}\left[\phi_n\right]$ as a function of the Lagrange multipliers, that is

$$\begin{aligned} \mathbb{E}_{\mathcal{Y}}\left[\phi_n\right] &= \frac{\partial}{\partial \lambda_n} \sum_{n=1}^{N}\sum_{k=1}^{K} \log\left(1-e^{-(\lambda_n+\lambda_k)}\right) \\ &= \sum_{k=1}^{K} \frac{e^{-(\lambda_n+\lambda_k)}}{1-e^{-(\lambda_n+\lambda_k)}}. \end{aligned} \tag{2.13}$$

Therefore the Lagrange multipliers are determined by numerically solving the non-linear system of equations

$$\sum_{k=1}^{K} \frac{\varphi_n\,\varphi_k}{1-\varphi_n\,\varphi_k} = \phi_n^{\star},\ n = 1, ..., n. \tag{2.14}$$

It is important to notice that the expected matrix that follows from such a maximum entropy description

$$\mathbb{E}_{\mathcal{Y}}\left[Y_{nk}\right] = \frac{\varphi_n\,\varphi_k}{1-\varphi_n\,\varphi_k} \tag{2.15}$$

always describes a fully connected network[9]. This is common in maximum entropy models that use only the strengths as constraints. In Chapter 3 we discuss maximum entropy ensembles that use both strengths and degrees as constraints. Moreover in Chapter 7 we propose an original model for temporal weighted networks that separately models the probability of a link to be observed from its weight.

[9] Unless, of course, a node k is such that $\phi_k = 0$, in which case it could be omitted to begin with.

Chapter 3

Reconstruction of Systemic Risk due to Fire Sales Spillover

Synopsis of the Chapter. Monitoring and assessing systemic risk in financial markets is of great importance but it often requires data that are unavailable or available at a very low frequency. For this reason, systemic risk assessment with partial information is potentially very useful for regulators and other stakeholders. In this Chapter we consider systemic risk due to fire sales spillovers and portfolio rebalancing by using the risk metrics defined by Greenwood, Landier, and Thesmar, 2015. By using a method based on the constrained minimization of the Cross Entropy, we show that it is possible to assess aggregated and single bank's systemicness and vulnerability, using only the information on the size of each bank and the capitalization of each investment asset. We also compare our approach with an alternative widespread application of the Maximum Entropy principle allowing to derive graph probability distributions and generating scenarios and we use it to propose a statistical test for a change in banks' vulnerability to systemic events.

Almost all the contents of this chapter previously appeared in Di Gangi, Lillo, and Pirino, 2018.

3.1 Introduction

After the recent troubled years for the global economy, in which two severe crises (the 2007 crisis of financial markets and the 2010 sovereign debt crisis) have put the whole economic system in dramatic distress, vulnerability of banks to systemic events is now the main focus of a growing number of investigations of the academic community. Simultaneously, many research efforts are devoted to understand the role of banks or, broadly speaking, of financial institutions in the creation and in the spreading of systemic risk. Given the prominent importance of the topic and its multifaceted nature, the literature on evaluation and anticipation of systemic events is huge (see Demirgüç-Kunt and Detragiache, 1998; Kaminsky and Reinhart, 1999; Harrington, 2009; Scheffer et al., 2009; Barrell, Davis, Karim, and Liadze, 2010; Duttagupta and Cashin, 2011; Kritzman, Li, Page, and Rigobon, 2011; Allen, Bali, and Tang, 2012; Arnold, Borio, Ellis, and Moshirian, 2012; Bisias, Flood, Lo, and Valavanis, 2012; Scheffer et al., 2012; Merton et al., 2013; Oet, Bianco, Gramlich, and Ong, 2013, among many contributions).

Several are the channels through which financial distress may propagate from one institution to another and, eventually, affect a vast portion of the global economy. Fire sales spillovers due to assets' illiquidity and common portfolio holdings are definitely one of the main drivers of systemic risk. Shared investments create a significant overlap of portfolios between couples of financial institutions. Such (indirect) financial interconnectedness is an important source of contagion, since partial liquidation of assets by a single market player is expected to affect all other market participants that share with it a large fraction of their own investments (see Corsi, Marmi, and Lillo, 2016; Huang, Vodenska, Havlin, and Stanley, 2013; Caccioli, Shrestha, Moore, and Farmer, 2014; Lillo and Pirino, 2015). Fire sales move prices due to the finite liquidity of assets and to market impact. In a perfectly liquid market there will be no fire sale contagion at all (see Adrian and Shin, 2008, for a review on the role of liquidity in financial contagion). Finally, leverage amplifies such feedbacks. In fact, as described in detail by Adrian and Shin, 2010; Adrian and Shin, 2014, levered institutions continuously rebalance their positions inflating positive and, most importantly, negative assets' price variations.

Assessing and monitoring systemic risk due to fire sales spillover is therefore of paramount importance for regulators, policy makers, and other participants to the financial markets. Greenwood, Landier, and Thesmar, 2015 introduced recently a stylized model of fire sales, where illiquidity, target leverage, and portfolio overlap are the constituent bricks. They used the model to propose two systemic risk metrics: systemicness and vulnerability of a bank. Given a market shock, the first is the total percentage loss induced on the system by the distress of the bank, whereas the second is the total percentage loss experienced by the bank when the whole system is in distress. In order to compute these quantities, a full knowledge of the portfolio composition of all banks is needed, because the systemicness and vulnerability of a bank depends on the portfolio and leverage of the other banks.

Greenwood, Landier, and Thesmar, 2015 applied their method to the European Banking Authority (EBA) data that resulted from the July 2011 European stress tests. These data provide detailed balance sheets for the 90 largest banks in the European Union. Duarte and Eisenbach, 2013 exploited a publicly available dataset of balance sheets of US bank holding companies to apply the framework of Greenwood, Landier, and Thesmar, 2015. They derive a measure of aggregate vulnerability that *[...] reaches a peak in the fall of 2008 but shows a notable increase starting in 2005, ahead of many other systemic risk indicators*.

In general, however, the detailed information set required to compute such systemic risk indicators might not be available. For example European stress test data are sporadic. Moreover the sampling frequency of balance sheet data is rarely higher than quarterly. Thus an important question is whether it is possible to estimate systemic risk due to fire sales spillovers in absence of data on portfolio composition of financial intermediaries.

Two possible approaches have been proposed in the literature. The first one (see, among others, Adrian and Brunnermeier, 2011; Acharya, Engle, and Richardson, 2012; Banulescu and Dumitrescu, 2015; Corsi, Lillo, Pirino, and Trapin, 2018) is purely econometric and it is typically based on publicly available data on price of assets and market equity value of publicly quoted financial institutions. Generically the method consists in estimating conditional variables, such as conditional Value-at-Risk or conditional Expected Shortfall. The econometric approach circumvents the unavailability of data on portfolio holdings, but pays this advantage with the introduction of a strong stationarity assumption: estimates based on the past information are assumed to be always good predictors of the future behavior of the system. Nevertheless, due to the nature of a global financial crisis, it is in the very moment of the onset of a period of distress that the stationarity assumption may fail to work properly. Moreover it is often restricted to publicly quoted institutions for which equity value are available at daily frequency.

A second possible approach[1], followed in the present chapter, consists in inferring the matrix of portfolio holdings using only a reduced, but easily available, information set, and/or deriving a probability distribution for the portfolio weights according to some criterion. This is typically achieved summoning the *maximum entropy principle* which postulates that (Anand et al., 2013) *[...] subject to known constraints [...] the probability distribution that best represents our current knowledge and that is least biased is the one with maximal entropy*. The approach of Maximum Entropy, can be applied in at least two different ways that we distinguish clearly in the following, and is not new in systemic risk studies (Mistrulli, 2011; Anand et al., 2013; Musmeci et al., 2013; Squartini, Lelyveld, and Garlaschelli, 2013; Bargigli et al., 2015a). It is widely used for inferring the structure of the interbank network when only data of total interbank lending and borrowing for each bank (plus possibly other information) are available (for a comparison of different methods, see Anand et al., 2017; Gandy and Veraart, 2016).

The seminal contribution by Mistrulli, 2011, comparing the empirical Italian interbank network with that reconstructed via a Maximum Entropy optimization procedure, has shown that the latter is fully connected while the former is very sparse (see also Mastromatteo, Zarinelli, and Marsili, 2012) and,as a consequence of this misestimation, the reconstructed network underestimates the risk contagion[2]. Recently a comparison of network reconstructions techniques has been carried out also for bipartite networks (Ramadiah, Caccioli, and Fricke, 2020).

[1] There are, of course, many different approach to assess systemic risk in financial networks. For example, Amini, Cont, and Minca, 2013 propose a rigorous asymptotic theory that allows to predict the spread of distress in interbank networks.

[2] A complementary method is proposed by Anand, Craig, and Von Peter, 2015. Here the authors reconstruct the network of bilateral exposures for the German banking system via the matrix that, preserving some constraints, has the minimum density. Nevertheless, if cross entropy method underestimates systemic risk by overestimating the network density, Anand, Craig, and Von Peter, 2015 show that, for a similar reason, minimum density returns positively biased estimates. Hence, the two approaches can be used jointly together to create a corridor in which the true systemic risk should lay.

A part from network reconstruction, the use of entropic methods is widespread in economic sciences. For example, it is widely used in econometrics for the estimation of probability densities, as it is witnessed by a vast stream of contributions in this direction (see, among others, the contributions by Zellner and Highfield, 1988; Ryu, 1993; Wu, 2003; Kouskoulas, Pierce, and Ulaby, 2004; Park and Bera, 2009; Usta and Kantar, 2011; Chen, 2015). An interesting point of coomparison for our work is the use of entropy in the theory of portfolio choice[3]. When investors are uncertain about the probability structure of reality and, being averse to ambiguity, use the relative entropy as a way of penalizing (see, among the most notable contributions in this field, the works by Hansen and Sargent, 2001; Maccheroni, Marinacci, and Rustichini, 2006; Bera and Park, 2008; Zhou, Liu, and Qiu, 2008; Gilboa and Marinacci, 2016; Zhou, Cai, and Tong, 2013). In this respect it is important to clarify that we implicitly adopt the maximum entropy principle from the point of view of a regulator (or a social planner) who, irrespectively of the decisional process which is behind the creation of the network, is solely interested in having an unbiased estimate of systemic risk. Hence, our perspective could be thought as orthogonal to that adopted by the ambiguity aversion literature, in the sense that our declination of the maximum entropy principle is purely inferential and it is not meant to mimic, in any way, the banks' decisional processes that have created the network and, accordingly, the prevailing level of systemic risk.

In this chapter we propose to apply maximum entropy approach to the inference of the network of portfolio weights in order to estimate metrics of systemic risk due to fire sales spillovers. Specifically, we show how indirect vulnerability, systemicness (as defined by Greenwood, Landier, and Thesmar, 2015) and the aggregate systemic risk of US commercial banks can be estimated when only a partial information (the size of each bank and the capitalization of each asset) is available. Differently from the interbank studies (as in Mistrulli, 2011; Mastromatteo, Zarinelli, and Marsili, 2012; Anand, Craig, and Von Peter, 2015) we deal with bipartite networks, namely graphs[4] whose nodes can be divided into two sharply distinguished sets that, in our case, are commercial banks and asset classes. More specifically, we analyze the quarterly networks of US commercial banks' exposures in the period 2001-2013 using the *Federal Financial Institutions Examination Council* (FFIEC) through the *Call Report* files[5]. We compute, for each quarter, systemicness and vulnerability of each bank and the aggregate vulnerability of the system. We compare them with the values inferred assuming the balance sheet compositions of the banks were not known. In this sense our contribution is similar to Mistrulli, 2011, but applied to systemic risk due to fire sale spillover rather than to cascades in the interbank network. Differently from the interbank case, we find that newly introduced maximum entropy methods are very accurate in assessing systemic risk due to fire sales spillover when partial information is available.

The contribution of this chapter is divided into two main parts. First, following a practice that is largely diffused among researchers of both academic institutions and central banks (see, among others, Sheldon and Maurer, 1998; Upper and Worms, 2004; Wells, 2004; Mistrulli, 2011; Sachs, 2014), we reconstruct the matrix of portfolio holdings as such that minimizes the cross entropy (or Kullback-Leibler divergence) from a initial guess. Despite this approach has often been referred to as maximum

[3] We thank an anonymous referee for this suggestion.

[4] Throughout all the manuscript we use the terms "network" and "graph" interchangeably.

[5] Hence our dataset is quite similar to that adopted by Duarte and Eisenbach, 2013, but it profits from a larger sample of banks, since commercial banks are fairly more numerous than bank holding companies. All data are available at: https://cdr.ffiec.gov/public/

entropy, or matrix balancing, in order to avoid confusions with different methods discussed in the following, we refer to it as Cross-Entropy method. We show that this approach does a very good job in our case, providing unbiased estimates of the systemic risk metrics defined by Greenwood, Landier, and Thesmar, 2015. Besides, we show that the reconstructed matrix corresponds to that implied by the Capital Asset Pricing Model, hence it possesses a clear economic meaning.

Second, we compare Cross-Entropy with a different approach to entropy maximization, which allows to define a probability mass function for graphs (ensemble) by maximizing entropy under suitable constraints where some average quantities are set equal to the ones observed in data. Despite the economic intuition of this approach is less sharp than the previous one, the method is widespread in the literature and allows performing scenario generation. We propose a new ensemble, termed MECAPM, which (i) satisfies a set of economically motivated constraints, (ii) behaves in average as the cross entropy method proposed before, and (iii) allows for scenario generation, potentially useful for supervisory authorities to test if a specific institution has increased its systemicness with respect to the past.

We structure our chapter as follows. Section 3.2 introduces some nomenclature and briefly describes the risk metrics of Greenwood, Landier, and Thesmar, 2015. The dataset of US commercial banks provided by the FFIEC is discussed in Section 3.3. In Section 3.4 we present the cross entropy method and show its performances for the estimation of systemic risk. In Section 3.5 we compare the cross entropy method with the maximum entropy alternative which derives a probability distribution of graphs. This is useful, among other things, to introduce a statistical test for surveillance activities by central banks and other regulatory institutions. Finally Section 3.6 summarizes the main contributions of the chapter. Appendices provide additional information on the construction of the dataset of bank portfolio holdings and all the analytical computations omitted in the main text.

3.2 Systemic risk metrics: Vulnerability and Systemicness

In this chapter we use some metrics of systemic risk due to fire sales, which have been recently introduced by Greenwood, Landier, and Thesmar, 2015. They consider a system composed by N banks and K asset classes. Portfolio holdings are described by the $N \times K$ matrix $\mathbf{Y}$, whose element Y_{nk} is the dollar-amount of k-type assets detained by bank n. The corresponding matrix of portfolio weights is thus

$$W_{nk}(\mathbf{Y}) = \frac{Y_{nk}}{\sum_{k'=1}^{K} Y_{nk'}}.$$

In what follows, we introduce a discretization of the elements of Y's, in such a way that the matrix $\mathbf{Y}$ belongs to the space $\mathbb{N}^{N\times K}$ of $N \times K$ integer valued matrices. In the empirical application we will use the resolution of the dataset which is 10^3\$.

The total asset size, or bank dimension, D_n of the n-th bank and the total capitalization[6] C_k of the k-th asset class are easily computed as, respectively, the total row and column sums of the matrix Y_{nk}, in formula

$$D_n(\mathbf{Y}) = \sum_{k=1}^{K} Y_{nk}, \qquad C_k(\mathbf{Y}) = \sum_{n=1}^{N} Y_{nk}, \tag{3.1}$$

[6]More precisely, the quantity C_k is the total amount of asset's k capitalization due to the banking sector. To simplify the notation we will call it capitalization.

where we have explicitly expressed the dependence of D_n and C_k from $\mathbf{Y}$.

The rectangular matrix $\mathbf{Y}$ can be naturally associated to a bipartite network, i.e. a graph whose vertices can be divided into two disjoint sets such that every edge connects a vertex in one set to one in the other set, the two sets being the banks and the asset classes. In the network jargon, $D_n(\mathbf{Y})$ and $C_k(\mathbf{Y})$ are called the strength sequences.

A relevant information concerning the balance sheet of each bank n is the total equity E_n, from which one can compute the leverage as $T_n = \frac{D_n - E_n}{E_n}$ (as in Greenwood, Landier, and Thesmar, 2015). Finally, each asset class is characterized by an illiquidity parameter ℓ_k, with $k = 1, ..., K$, defined as the return per dollar of net purchase of asset k[7].

This setting is used in Greenwood, Landier, and Thesmar, 2015 to define three metrics of systemic risk, capturing the effect of fire sales in response to a shock on the price of the assets. This is described by the K dimensional vector $-\boldsymbol{\varepsilon} = (-\varepsilon_1, ..., -\varepsilon_K)$, whose components are the assets' shocks. They define:

- **Aggregate vulnerability** AV as *[...] the percentage of aggregate bank equity that would be wiped out by bank deleveraging if there was a shock [...] to asset returns.*
- **Bank systemicness** S_n as *the contribution of bank n to aggregate vulnerability.*
- **Bank's indirect vulnerability** IV_n as *[...] the impact of the shock on its equity through the deleveraging of other banks.*

By assuming that banks follow the practice of leverage targeting and that, in response to a negative asset shock, they sell assets proportionally to their pre-shock portfolio holdings, Greenwood, Landier, and Thesmar, 2015 show that S_n can be decomposed as

$$S_n = \Gamma_n \frac{D_n}{E} T_n r_n, \tag{3.2}$$

where E is the total equity, $E = \sum_{n=1}^{N} E_n$, r_n is the n-th element of the vector $\mathbf{r} = \mathbf{W}\boldsymbol{\varepsilon}$, i.e. the portfolio return of bank n due to the shock $\boldsymbol{\varepsilon}$, and

$$\Gamma_n = \sum_{k=1}^{K} \left(\sum_{m=1}^{N} D_m W_{m,k} \right) \ell_k W_{nk}.$$

The aggregate vulnerability is computed simply as

$$\mathrm{AV} = \sum_{n=1}^{N} S_n. \tag{3.3}$$

Finally, the indirect vulnerability of a bank is

$$\mathrm{IV}_n = (1 + T_n) \sum_{k=1}^{K} \ell_k W_{nk} \sum_{n'=1}^{N} W_{n',k} D_{n'} B_{n'} r_{n'}. \tag{3.4}$$

In what follows we often assume, as in Duarte and Eisenbach, 2013, that $\epsilon_k = 1\%$ for all $k = 1, ..., K$, which in turns implies that $r_n = 1\%$ in equations (3.2) and (3.4).

[7] The assumption of linear price impact comes directly from the framework of Greenwood, Landier, and Thesmar, 2015. Although a square-root law fits the data better, the linear assumption has been widely adopted in the literature (see Gatheral, Schied, and Slynko, 2012; Cont and Wagalath, 2014; Lillo and Pirino, 2015, among others) and has been empirically validated at daily frequency by Obizhaeva, 2008.

Note however that if all the assets are shocked by the same amount, our results do not depend on it, since the systemic risk measures will have only a different prefactor. In Section 3.4.1 we consider other shock scenarios to test the robustness of our methods. Finally, we set the liquidity parameter at $\ell_k = 10^{-10}$ for all asset classes except for cash, for which we put $\ell_k = 0$ (as in Greenwood, Landier, and Thesmar, 2015; Duarte and Eisenbach, 2013). As a final comment it should be noted that (Greenwood, Landier, and Thesmar, 2015) add two more constraints to the problem. First, when direct losses of a bank exceed its equity, the bank liquidates all the assets. Second, leverage is capped to the value 30. In our empirical investigation we have followed (Duarte and Eisenbach, 2013) who do not add these constraints. We have however compared the aggregate vulnerability of the US banking system (see next Section for the data used) under the two model specifications. We have found that the difference is less than 1% with the exception of few quarters around the end of 2009 when it reaches 10%.

It is important to stress that the Greenwood, Landier, and Thesmar, 2015 method to estimate systemic risk metrics is essentially static. As it is standard in stress testing, a given scenario of price changes at a given time is considered and then, given the balance sheet and portfolio composition of banks at that time, the consequences of deleveraging and fire sales are computed. Thus no past information (even when available) on balance sheets or prices is ever used in the methodology. This is of course a limitation since the decision on how to deleverage in a certain quarter depends in reality also on past market price behavior as well as on deleveraging in the last quarters. Such an extension, although interesting, is beyond the scope of Greenwood, Landier, and Thesmar, 2015 model as well as of the vast majority of stress test methods. It would require to choose scenarios containing the price changes at more than one quarter as well as being able to disentangle the price changes due to fundamental reasons from those due to past deleveraging. The definition of a dynamic stress test is clearly beyond the scope of our work and we will stick to the standard static stress test approach. As in Duarte and Eisenbach, 2013, in the empirical application below we will consider a stress test for each available quarter, discarding all the information coming from past quarters. Thus even if apparently we are treating a time series of portfolios of length T, as a matter of fact we are repeating T times the (static) stress test.

In the next section we present the dataset that we use in our analysis to measure systemic risk, as captured by the metrics of Greenwood, Landier, and Thesmar, 2015, in the US banking sector. Such a dataset allows us to have quarterly estimates of systemicness, aggregate, and indirect vulnerability and to compare these estimates with those inferred from the Cross-Entropy approach and the Maximum Entropy principle. Since we have to deal with both real and reconstructed (or sampled form a statistical ensemble) networks, from now on we follow the convention to add a superscript $x^\star$ to any variable x whenever it is referred to a real (observed) network, while the variable x is represented without the superscript $\star$ every time it is referred to a reconstructed network (e.g. one sampled from a statistical ensemble as described in Section 3.5).

3.3 Data

All regulated financial institutions in the United States are required to file periodic financial information with their incumbent regulators. The *Federal Financial Institutions Examination Council*, is the regulatory institution responsible to collect and

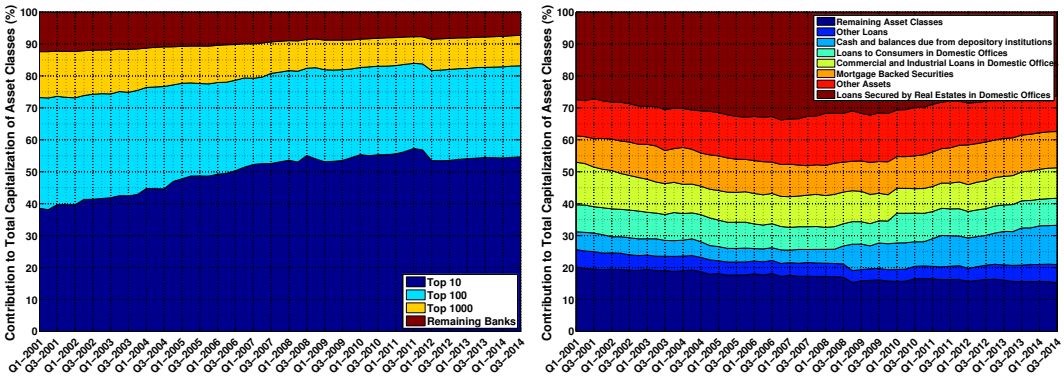

FIGURE 3.1: We report in the left panel the percentage of total assets detained by top 10, top 100, top 1000 and the remaining banks in shaded areas of different colors. A vast portion of total assets is controlled by the top 10 banks. In the right panel we report, for each quarters, the contribution of the top seven asset classes (in terms of capitalization) to the total capitalization. A large percentage to total asset capitalization is due to Loan Secured by Real Estates in Domestic Offices.

maintain the data used in our analysis. The financial institutions subject of our investigation are *Commercial Banks and Savings and Loans Associations*. The FFIEC defines officially a commercial bank as[8]: *"[...] a financial institution that is owned by stockholders, operates for a profit, and engages in various lending activities"*. FFIEC requires commercial banks to file the quarterly *Consolidated Report of Condition and Income*, generally referred to as *Call Report*. Each bank is required to fill a form with detailed information on its financial status, in particular on its balance sheet. The specific reporting requirements depend upon the size of the bank and whether or not it has any foreign office. The form FFIEC031 is used for banks with both domestic (U.S.) and foreign (non-U.S.) offices while form FFIEC041 is designed for banks with domestic (U.S.) offices only. A *Saving and Loan Association* is a financial institution that accepts deposits primarily from individuals and channels its funds primarily into residential mortgage loans. From the first quarter of 2012, all *Savings and Loan Associations* are required to file the same reports, thus they are included in the dataset since then.

The data provided by the *Call reports* are publicly available[9] since 1986, although the form changed considerably throughout the years, showing an increasing level of details requested. To have a good compromise between the fine structure of data and a reasonably populated statistics we considered the time period going from March 2001 to September 2013, for a total of 55 quarters. The number of financial institutions present in the data is pretty stable during quarters, starting from approximately $9,000$ entities in the first quarter and ending in roughly $6,500$ in the last one. The asset categories have been created as coherent sums of codes. We describe the procedure adopted to form asset classes in Appendix 3.7.1 along with some data statistics. In particular, we aggregate data in a set of 20 asset classes following the rationale of Duarte and Eisenbach, 2013, that is each of the 20 asset classes is composed in such

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

[9] See https://www.chicagofed.org/banking/financial-institution-reports/commercial-bank-data.

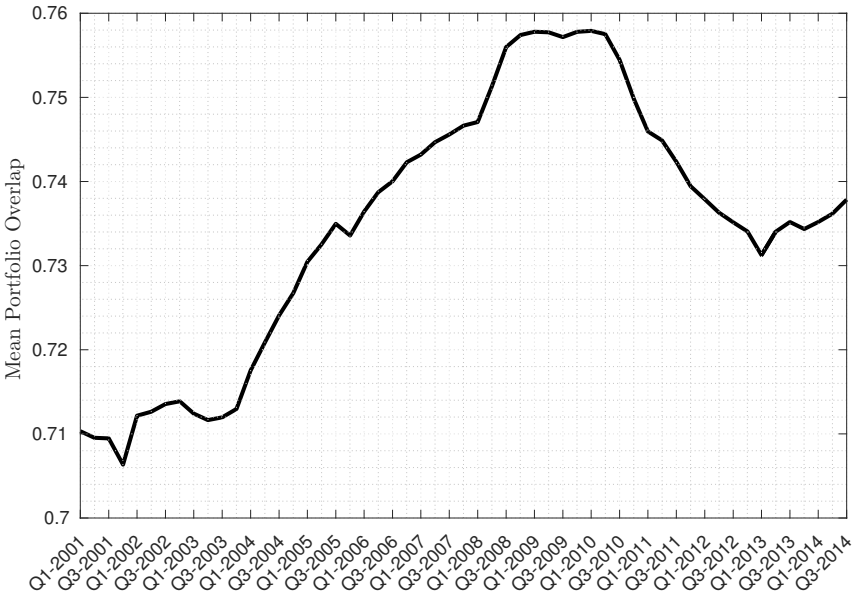

FIGURE 3.2: Mean cosine (or L_2) similarity of pairs of portfolios of US commercial banks.

a way that, in case of a fire sale of assets belonging to a specific class, the price impact would be restricted mainly to the assets in the same class. In other words, it is reasonable to assume that the co-illiquidity (or cross impact) of two different asset classes is negligible. The twenty macro asset classes used to build the network are described in Table 3.2 of Appendix 3.7.1, which also documents in detail how they have been formed. In the left panel of Figure 3.1 we show how the total asset value is concentrated on the top tiered banks. The right panel of Figure 3.1 shows the relative importance of the top seven assets classes (in terms of total capitalization), revealing that a large portion of the total capitalization is due to *Loan secured by real estates in domestic offices*.

To test the role of portfolio similarity in systemic risk, we report in Figure 3.2 the time series of the mean similarity between all the pairs of banks' portfolios. Similarity is measured with the cosine (or L_2) norm i.e. the cosine of the angle formed by the two vectors $\{Y^*_{nk}\}_{k=1,..,K}$ and $\{Y^*_{m,k}\}_{k=1,..,K}$ representing the portfolios of bank n and m, respectively. The plot shows clearly that similarity between portfolios has increased significantly before the 2008 crisis, making the systemic risk higher. A pretty similar pattern is shown by the aggregate vulnerability in Figure 3.6, we will turn later on this point.

In conclusion, for each quarter we are able to construct a matrix $\mathbf{Y}^\star$ of bank holding whose element $Y_{nk}{}^\star$ is the total dollars invested by the n-th bank in the k-th asset class. It is important to note that the matrix $\mathbf{Y}^\star$ has around 50% of zero entries. Thus the network is relatively dense, but far from being fully connected. Simply put, the portfolio of the typical bank in the dataset does not contain investments in all the 20 assets classes.

3.4 The Cross-Entropy Approach for Systemic Risk Assessment

Cross-Entropy is a method[10], largely adopted by scholars and researchers of central banks, used to reconstruct a target matrix (as the interbank matrix) from partial knowledge of its properties. The idea is to select an *a priori* guess for the matrix and then to find its closest matrix subject to some constraints. In the simplest case, such constraints are non negativity conditions of matrix elements and the total row and column sums. Finally, as a measure of distance to be minimized between the guess and the target matrix one uses the Kullback-Leibler divergence (also called relative entropy).

For the specific case of the system of bank holdings for US commercial banks we assume to have at our disposal, for each quarter, only the information on the total asset size $D_n^\star$ for the n-th bank and the total capitalization $C_k^\star$ for the k-th asset class. The Cross-Entropy approach derives the target matrix $\mathbf{Y}$ as that which solves the optimization problem

$$\begin{aligned} \min_{\mathbf{Y}} \quad & \sum_{n=1}^{N}\sum_{k=1}^{K} Y_{nk} \log\left(\frac{Y_{nk}}{\widetilde{Y}_{nk}}\right) \\ \text{s.t.} \quad & \sum_{n=1}^{N} Y_{nk} = D_n^\star,\ n = 1,...,N, \\ & \sum_{k=1}^{K} Y_{nk} = C_k^\star,\ k = 1,...,K, \\ & Y_{nk} \geq 0, \end{aligned} \tag{3.5}$$

where $\widetilde{Y}_{nk}$ are the entries of a given guess matrix. Note that the cases analyzed in the interbank lending literature (Mistrulli, 2011) typically have an additional constraint that diagonal elements vanish, required to avoid a single institution to be simultaneously a borrower and lender to itself (see, for example, the Appendix B in Mistrulli, 2011). The matrix of portfolio holdings analyzed here does not require any of such kind of restrictions.

We suggest to use the capital asset pricing model (CAPM) to form an economically motivated initial guess. In a standard CAPM, investors choose their portfolio in such a way that each weight on a stock is the fraction of that stock's market value relative to the total market value of all stocks (Sharpe, 1964; Lintner, 1965; Mossin, 1966). Since $D_n^\star$ is the total asset size of the n-th bank and since the total market value of all stocks is given by $L^\star = \sum_{k=1}^{K} C_k^\star$, the portfolio weights expected by CAPM are given by[11]

$$Y_{n,k}^{\text{CAPM}} = \frac{C_k^\star}{L^*} D_n^\star. \tag{3.6}$$

Notice that this choice of the initial guess is the same used in Mistrulli, 2011 for the interbank market, even if in that case the CAPM interpretation is less direct.

[10] The method is also known as matrix balancing, or maximum entropy matrix reconstruction. We refer to it as Cross-Entropy in order to clearly distinguish it from a method deriving probability distributions over all possible matrices (graphs). The latter is introduced in Section 3.5 as a competing approach for systemic risk reconstruction, and we refer to it as Canonical Maximum Entropy Ensemble method.

[11] More precisely, we are deriving what can be addressed as a "banking-CAPM" since, as mentioned before, $C_k^\star$ represents the total amount of asset's k capitalization due to the banking sector and not its total market capitalization.

Given that in (3.5) the condition on the diagonal elements is absent and since the Kullback-Leibler divergence is always positive, the optimal solution of the Cross-Entropy problem in (3.5) when $\widetilde{Y}_{nk} = Y_{n,k}^{\text{CAPM}}$ is nothing but the $Y_{n,k}^{\text{CAPM}}$ itself. To distinguish the estimator from the Capital Asset Pricing Model of (Sharpe, 1964) we will call the former Cross Entropy CAPM (CECAPM) estimator. Note that thanks to the bipartite nature of the network under study we do not have to resort to numerical routines to solve problem (3.5). If other constraints are added to the problem (e.g. that some banks cannot invest in some asset classes) one could numerically solve the problem (3.5) with the additional constraints.

3.4.1 Assessing aggregate vulnerability

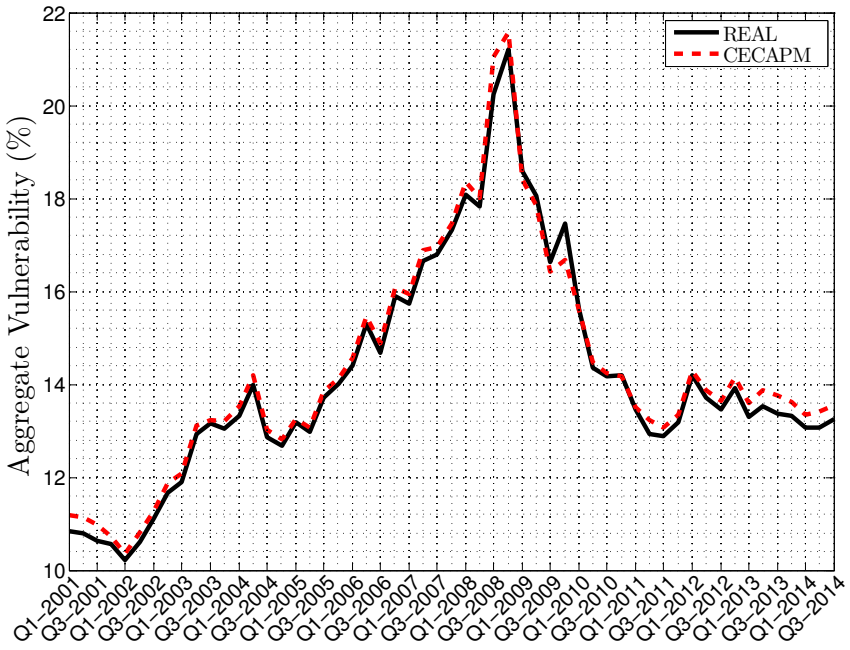

FIGURE 3.3: This figure reports as a black continuous line the aggregated vulnerability, as defined by equation (3.3), computed on the matrix $Y_{nk}{}^{\star}$ of portfolio holdings as provided by the FFIEC dataset of US commercial bank holdings, described in Section 3.3. The dashed red line refers to the aggregate vulnerability reconstructed with CECAPM.

We now empirically test the validity of the cross entropy approach in estimating the aggregate vulnerability on our data.

Figure 3.3 compares the true value of the aggregated vulnerability, obtained by using the real matrix of portfolio compositions, with the one obtained with the cross entropy method. It is clear that CECAPM provides estimates of AVs in excellent agreement with the real one, in spite of the fact that the true portfolio matrix is quite different from that of CECAPM, because in the former roughly half of the matrix elements are zero while the latter models have adjacency matrices with all non vanishing elements.

An important implication of Figure 3.3 is that, at least for the dataset under analysis, it is not necessary to know the matrix $\mathbf{Y}^{\star}$ to assess the systemic risk as measured

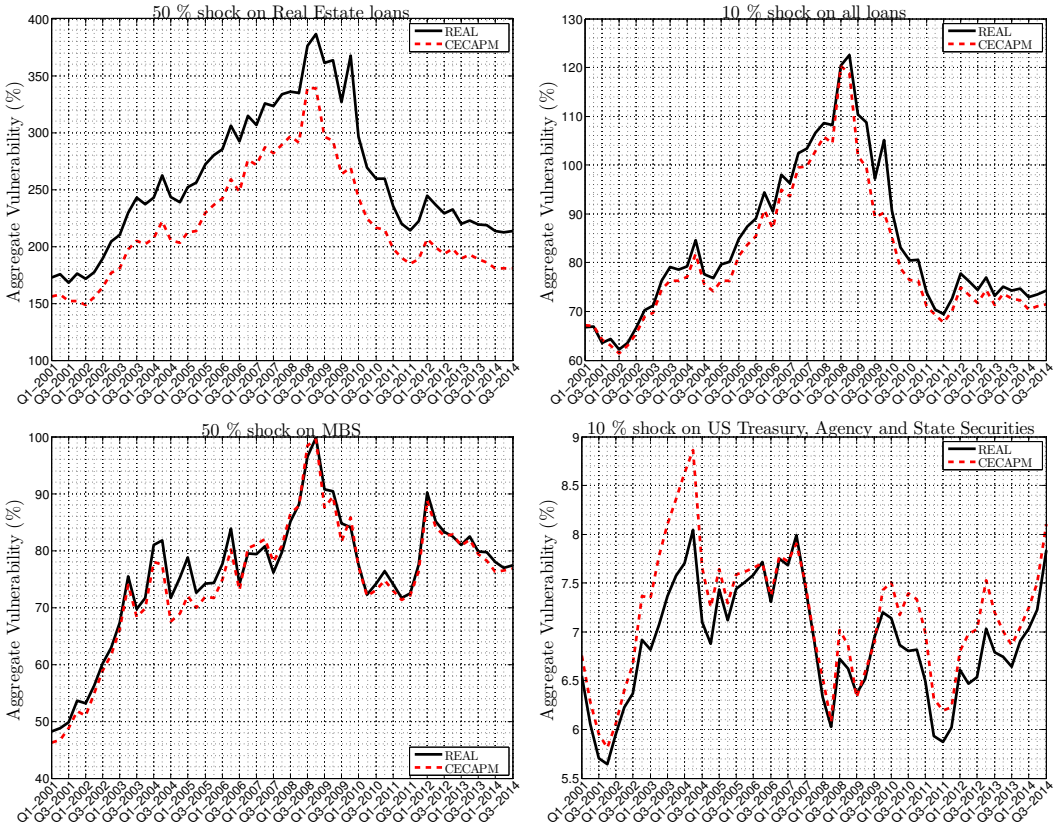

FIGURE 3.4: Aggregate vulnerability under different shock scenarios. Each panel reports the AV obtained from the full knowledge of portfolios composition and those obtained using the CECAPM reconstruction.

by the aggregate vulnerability. The knowledge of banks' size and assets' capitalization is enough to infer the matrix $Y^{\text{CAPM}}_{n,k}$, which very well reproduces the aggregate behavior (in terms of systemicness) of the system. This is different from the result of Mistrulli, 2011 for the interbank network, since he finds that the Cross-Entropy approach significantly underestimates systemic risk, while in our case the bias is negligible.

Robustness to different shock scenarios

The estimation and reconstruction of AV has been performed by assuming a uniform shock of 1% across all asset classes. However our results are robust also to other shock scenarios. To show this, we have repeated the above analysis by considering other cases, namely: (i) a 50 % shock on Real Estate loans (2 asset classes); (ii) a 10 % shock on all loans (8 asset classes); (iii) a 50% shock on Mortgage Backed Securities (1 asset class); and (iv) a 10% shock on U.S. treasury securities, U.S agency securities, Securities issued by state and local governments (3 asset classes). The resulting aggregate vulnerability with real data and estimated with CECAPM is shown in Figure 3.4. In all cases the CECAPM estimation tracks quite closely the AV obtained from the full knowledge of portfolios composition. We therefore conclude that our result is not due to the uniform shock assumption, but is more generically applicable.

European Banking Authority Data

We now show that our results hold also for different banking systems. To this end we investigate the public dataset made available[12] by the European Banking Authority (EBA), after the 2011 stress tests conducted on the largest 90 European banks at the time. The data consists of the exposures of each bank toward a set of 42 asset classes, as well as their book leverages. We defined the asset classes following precisely Greenwood, Landier, and Thesmar, 2015 [13]. Moreover the initial shock is (following Greenwood, Landier, and Thesmar, 2015) "a 50% write-off of all GIIPS debt". As for the U.S. commercial banks, we compared the AV obtained from the full network data with that obtained using CECAPM. The relative percentage bias of AV estimated using partial information is 3.7%. As a robustness check, we repeated the exercise for two different shocks: a 10% write off of either all EU debt or all sovereign debt, including non E.U. countries. The percentage bias of the CECAPM estimation is 3.6% in the former case and 5.1% in the latter. Clearly also for this dataset, CECAPM gives a faithful estimation of AV, since the bias is around $3\% - 5\%$, showing the robustness of the method.

3.4.2 Assessing systemic risk for individual banks

We now test the performance of CECAPM in assessing systemic risk for individual banks. In order to assess the performance of an estimator that produces estimates $\widehat{S}_n$ and $\widehat{\mathrm{IV}}_n$ of, respectively, systemicness and indirect vulnerability of the n-th bank in a given quarter we compute the relative error as

$$s_n = \frac{\widehat{S}_n - S_n^\star}{S_n^\star}, \qquad v_n = \frac{\widehat{\mathrm{IV}}_n - \mathrm{IV}_n^\star}{\mathrm{IV}_n^\star}. \tag{3.7}$$

In each quarter we have from $N = 6,500$ to $N = 9,000$ values of relative errors for each metric.

To visualize the result we plot the distribution of s_n and v_n across all banks present at a given year.

Figure 3.5 shows the results for 4 different years, and in Appendix 3.7.3 we show that the results obtained are stable across all the period considered. We observe that the distributions of percentage bias in the reconstruction of S and IV are similar for all the years considered. Their support is roughly comprised between -50% and $+60\%$ and they are peaked around -25%, indicating that Cross Entropy tends to underestimate single banks measures of systemic risk. In summary, the estimates of systemicness and indirect vulnerably for each single bank as provided by the CECAPM-implied matrix are satisfactorily accurate.

3.5 Comparison with the Max-Entropy ensembles

The Cross-Entropy approach described in the previous section assumes that the unknown matrix elements are those with minimal distance (as proxied by the cross-entropy function) from an a priori matrix. In this approach, the available economic

[12] http://www.eba.europa.eu/risk-analysis-and-data/eu-wide-stress-testing/2011/results

[13] The asset classes are: "sovereign debt of each of the 27 EU countries plus 10 others, commercial real estate, mortgages, corporate loans, small and medium enterprise loans, and retail revolving credit lines."

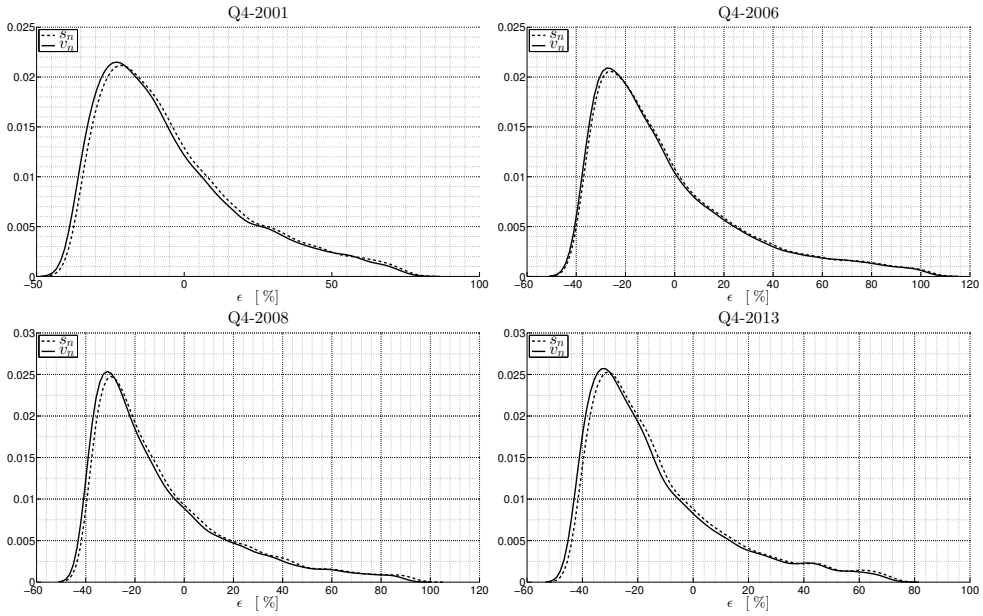

FIGURE 3.5: Distributions of the relative error of bank systemicness (dashed lines) and indirect vulnerability (solid lines) with respect to real data as estimated by the CECAPM for 4 different quarters.

information, which, in the specific case described above, consists of the quantities $D_n^\star$ and $C_k^\star$, is used to construct the guess following an economic intuition.

A different rationale constitutes the foundation of the Maximum-Entropy (ME) ensembles approach. This reconstruction method assumes, as standard in contexts of partial information, that the undisclosed quantities (in our case, the banks' portfolio holdings Y_{nk}) are random variables generated from an unknown statistical distribution. The ME approach amounts to take, among all the possible probability distributions, the one which maximizes the informational content of the economic constraints imposed during the maximization. This property follows directly from the definition of information as stated in the seminal paper by Shannon, 1948.

We define[14] a network statistical model as a set $\mathcal{Y}$ of graphs, called *ensemble*, and a probability mass function $\mathbb{P}_{\boldsymbol{\vartheta}}$ indexed by a vector of model parameters $\boldsymbol{\vartheta}$. In formula it is expressed as the triplet

$$\left\{ \mathbb{P}_{\boldsymbol{\vartheta}}, \mathcal{Y}, \boldsymbol{\vartheta} \in \Xi \right\},$$

where Ξ is a convex subset of $\mathbb{R}^P$, with P the total number of parameters of the model. The set $\mathcal{Y}$ is a countable set whose elements are graphs. In what follows, we will not distinguish between the graph and the associated matrix $\mathbf{Y}$, i.e. the probability mass function is defined in the space of integer valued matrices. Moreover the probability mass function $\mathbb{P}_{\boldsymbol{\vartheta}} : \mathcal{Y} \to [0,1]$ is such that $\sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}_{\boldsymbol{\vartheta}}(\mathbf{Y}) = 1$ and is allowed to depend on a vector of real parameters $\boldsymbol{\vartheta} \in \Xi$. A model can be defined by explicitly giving the ensemble, the probability mass function along with the space Ξ of the parameters, or by deriving $\mathbb{P}_{\boldsymbol{\vartheta}}[\mathbf{Y}]$ through the recurrent application of some generative mechanism or rule, either starting from an empty graph or by applying a randomization procedure to a reference graph.

[14] In describing our ME approach, we largely follow the theoretical framework of Kolaczyk, 2009a.

In its most general formulation, the ME principle postulates to obtain the probability mass function $\mathbb{P}$ as that which maximizes the Shannon's entropy

$$S = -\sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}\left[\mathbf{Y}\right] \log\left(\mathbb{P}\left[\mathbf{Y}\right]\right)$$

subject to the normalization constraint

$$\sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}\left[\mathbf{Y}\right] = 1,$$

and, possibly, to further additional constraints.

There are two ways of imposing the constraints. In the first one, termed *microcanonical ensemble*, constraints are imposed *exactly*, i.e. only the graphs fulfilling all the constraints have non zero probability. In the second one, termed *canonical ensemble*, all the graphs have non zero probability and the constraints are satisfied on average over the distribution. There are advantages and disadvantages in both approaches. The microcanonical ensemble is economically more grounded, for example in the system under investigation here it implies that a given network realization has non zero probability only if each bank (asset class) has the same asset size (capitalization) as in real data. On the contrary, in the canonical ensemble also graphs where these values are very different from the real data might have non zero probability. Despite this undesirable property, we believe it is worth performing a comparison of the Cross-Entropy approach with the canonical ME for the following reasons:

1. Solving the problem in the microcanonical ensemble is typically extremely hard or it requires extensive numerical simulations, randomizing the network by allowing moves that preserve all the constraints. On the contrary, canonical ensemble can often be obtained much more directly, as testified by their widespread use in Statistical Mechanics (Huang, 2008). Moreover when other constraints are added to the optimization problem, microcanonical ME (as well as Cross-Entropy) becomes intractable, thus limiting their practical use when regulators want to add additional knowledge on the system.

2. The flexibility of canonical ME allows exploring the relative role of information set and constraints in network reconstruction. For example we will show below that, using the same information sets (the strength sequences) but *different* constraints can lead to very different performance in the estimation of systemic risk, indicating its main determinants.

3. The excellent performance of CECAPM for systemic risk assessment calls for the construction of a network probability distribution which performs on average as CECAPM, but allows for scenario generation. The canonical ME ensemble we will introduce below (MECAPM) does exactly this job.

4. Last but not least, the application of canonical ME network ensembles in Economics and Finance is quite widespread, see, for example, Bargigli and Gallegati, 2011; Squartini, Fagiolo, and Garlaschelli, 2011; Fagiolo, Squartini, and Garlaschelli, 2013; Mastrandrea, Squartini, Fagiolo, and Garlaschelli, 2014; Saracco, Di Clemente, Gabrielli, and Squartini, 2015; Almog, Squartini, and Garlaschelli, 2017

3.5.1 Maximum Entropy ensembles

We will consider three ME ensembles in this chapter[15]. First, we propose a new max entropy ensemble which is based on the role of CAPM in the problem at hand. The probability mass function $\mathbb{P}$ is the solution of the optimization problem

$$\begin{aligned} \max_{\mathbb{P}} \quad & -\sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}[\mathbf{Y}] \log(\mathbb{P}[\mathbf{Y}]) \\ \text{s.t.} \quad & \sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}[\mathbf{Y}] = 1 \\ & \mathbb{E}_{\mathcal{Y}}[Y_{nk}] = Y_{n,k}^{\text{CAPM}},\ n = 1,...,N,\ k = 1,...,K. \end{aligned} \tag{3.8}$$

We call this model Maximum Entropy Capital Asset Pricing Model (shortened in MECAPM henceforth). In Appendix 3.7.2 we prove that the MECAPM has the unique solution

$$\mathbb{P}[\mathbf{Y}] = \prod_{n=1}^{N}\prod_{k=1}^{K} \left(\frac{Y_{n,k}^{\text{CAPM}}}{1+Y_{n,k}^{\text{CAPM}}}\right)^{Y_{nk}} \frac{1}{1+Y_{n,k}^{\text{CAPM}}}, \tag{3.9}$$

hence each single matrix entry Y_{nk} is geometrically distributed with mean $Y_{n,k}^{\text{CAPM}}$. To understand the rationale behind this ensemble, we notice an interesting relation between the CECAPM and MECAPM estimation of AV under a uniform shock of asset returns. As shown in Appendix 3.7.4,

$$\mathbb{E}[S_n(\mathbf{Y})] = S_n\left(\mathbf{Y}^{\text{CAPM}}\right)\left(1+\frac{D_n^*}{L^*}+\frac{\sum_{k=2}^{K} C_k}{\sum_{k=2}^{K} C_k^{\star 2}}\right), \tag{3.10}$$

where $\mathbb{E}[S_n(\mathbf{Y})]$ is the expected systemicness of bank n under the MECAPM ensemble and $S_n\left(\mathbf{Y}^{\text{CAPM}}\right)$ is the one according to the CECAPM. We notice that the former is larger than the latter, but the correction is small if $D_n^* \ll L^*$, since the last term in parenthesis is generally small. This result can also be used to compute systemicness and AV in the MECAPM ensemble *without* sampling but using the expression above. A similar result holds for indirect vulnerability (see Appendix 3.7.4 for details).

Since the other specifications of maximum entropy are quite popular in the literature of network reconstruction, for comparison purposes we take into considerations two other ensembles, mainly inspired by the paper by Mastrandrea, Squartini, Fagiolo, and Garlaschelli, 2014 and Saracco, Di Clemente, Gabrielli, and Squartini, 2015. Each of them is characterized by different constraints imposed on the maximization of the Shannon's entropy.

In the first ensemble, termed Bipartite Weighted Configuration Model (BIPWCM), the constrained maximization is

$$\begin{aligned} \max_{\mathbb{P}} \quad & -\sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}[\mathbf{Y}] \log(\mathbb{P}[\mathbf{Y}]) \\ \text{s.t.} \quad & \sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}[\mathbf{Y}] = 1 \\ & \mathbb{E}_{\mathcal{Y}}[D_n] = D_n^\star,\ n = 1,...,N, \\ & \mathbb{E}_{\mathcal{Y}}[C_k] = C_k^\star,\ k = 1,...,K. \end{aligned}$$

[15] All numerical routines, accompanied with an instruction manual, can be downloaded from http://mathfinance.sns.it/network_reconstruction/

Appendix 3.7.2 reports the derivation and calibration of the ensemble. Note that BIPWCM imposes weaker constraints with respect to MECAPM, while exploiting the same information set, namely the strength sequences.

Finally, we consider another (richer) statistical ensemble whose probability mass function, derived in Appendix 3.7.2, corresponds in our bipartite framework to the enhanced configuration model of Mastrandrea, Squartini, Fagiolo, and Garlaschelli, 2014. This newly defined ensemble, that we address as Bipartite Enhanced Configuration Model (BIPECM), is obtained via Maximum Entropy imposing both the mean value of strengths (as in BIPWCM) and the mean value of degrees, that is the number of edges incident in each vertex. In other words, we reconstruct the matrix by assuming the knowledge of the number of assets in which each bank invests as well as the number of banks investing in each asset. Despite the fact that this information is typically not known, we consider this ensemble to show that even with an information set significantly larger than the one used in MECAPM it is very difficult to outperform it. Mathematically, the BIPECM is obtained by solving the optimization problem

$$\begin{aligned}
\max_{\mathbb{P}} \quad & -\sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}\left[\mathbf{Y}\right] \log\left(\mathbb{P}\left[\mathbf{Y}\right]\right) \\
\text{s.t.} \quad & \sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}\left[\mathbf{Y}\right] = 1 \\
& \mathbb{E}_{\mathcal{Y}}\left[D_n\right] = D_n^{\star}, \\
& \mathbb{E}_{\mathcal{Y}}\left[D_n^{\text{row}}\right] = D_n^{\text{row}^{\star}},\ n = 1, ..., N, \\
& \mathbb{E}_{\mathcal{Y}}\left[C_k\right] = C_k^{\star}, \\
& \mathbb{E}_{\mathcal{Y}}\left[D_k^{\text{col}}\right] = D_k^{\text{col}^{\star}},\ k = 1, ..., K,
\end{aligned} \tag{3.11}$$

where D_n^{row} and D_k^{col} are, respectively, the row and the column degree sequences (see Appendix 3.7.2 for more details). The peculiarity of BIPECM is the addition of the information on the degree sequences that is absent in both BIPWCM and MECAPM[16]. Note that the three ensembles can be used not only for statistical inferences, but to produce estimates of any function defined on the network, which is the topic of the next section.

3.5.2 Results

Figure 3.6 compares the true value of the aggregated vulnerability, obtained by using the real matrix of portfolio compositions, with those obtained with entropic methods. It is clear that all the methods track qualitatively well the temporal pattern of AV in the investigated period, but it is worth noticing that CECAPM has a very tiny bias, providing estimates of AVs in excellent agreement with the real one. As expected from the above argument, AV under MECAPM is always slightly larger than under CECAPM. Among the max entropy methods, MECAPM outperforms BIPWCM and BIPECM. Since the information set required to derive the BIPECM is larger than that used for the MECAPM, this means that it is not the amount of

[16]One could consider another maximum entropy ensemble where the constraints are the same as in MECAPM plus the degree sequences. This is an enhanced MECAPM because additional information on the number of asset classes in each portfolio (and the number of banks investing in each asset class) is used. The optimization can be performed but the application on the US banks data shows no appreciable improvement in the systemic risk assessment with respect to the CECAPM (data available on request). For this reason and for the sake of simplicity in this chapter we will not present results on this ensemble.

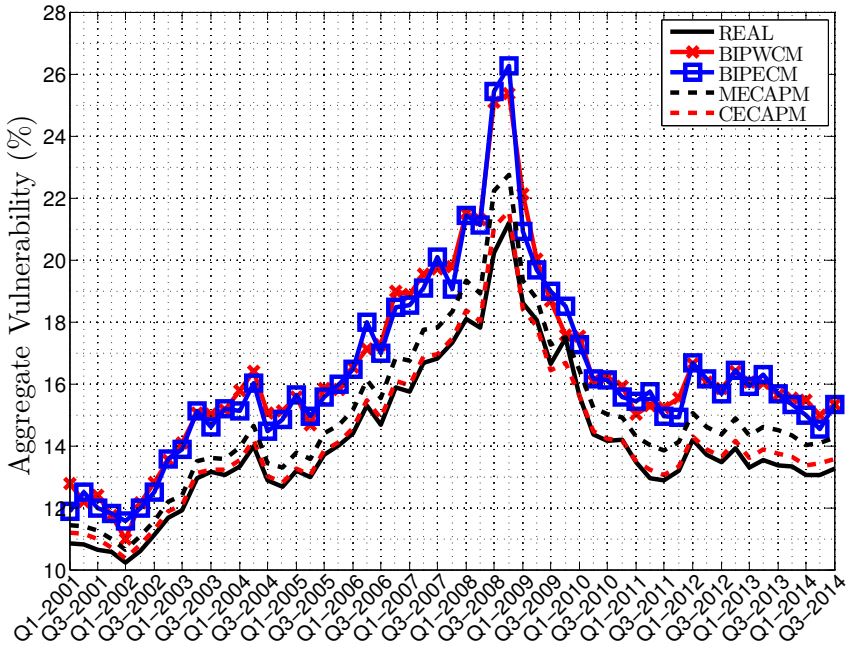

FIGURE 3.6: This figure reports as a black continuous line the aggregated vulnerability, as defined by equation (3.3), computed on the matrix $Y_{nk}{}^{\star}$ of portfolio holdings as provided by the FFIEC dataset of US commercial bank holdings, described in Section 3.3. The other lines refer to the aggregate vulnerability reconstructed with the four entropic methods.

information that matters, rather the way in which information is conveyed in the reconstruction algorithm. Finally, notice that the true portfolio matrix is quite different from the matrices of CECAPM, MECAPM, and BIPWCM because in the former half of the matrix elements are zero while the latter models have adjacency matrices with all non vanishing elements.

A similar comparative result holds by considering different shock scenario, as those studied in Section 3.4 (see Fig. 3.9 of the Appendix 3.7.3) as well as for the European Banking Authority Data (see Table 3.1. Among the max entropy methods, MECAPM significantly outperforms BIPWCM and BIPECM in estimating the AV obtained with the full knowledge of the portfolio composition of banks.

Finally we considered the assessment of systemic risk for individual banks. Figure 3.10 of the Appendix 3.7.3 shows that for each quarter BIPWCM strongly underestimates individual bank systemicness and indirect vulnerability. The median relative error ranges roughly between -60% and -70% and the interquartile range is very far from zero. The estimator based on BIPECM (using the additional information on degrees) gives slightly better results, even if a strong underestimation is still present. The median relative error ranges roughly between -50% and -40% and again the interquartile range is far from zero. On the contrary the estimator based on MECAPM (or CECAPM) performs much better. The median relative error never goes below -20% and almost always the interquartile range is centered around zero.[17]

[17] If instead we focus on the banks with higher systemicness or indirect vulnerability, the performances of the estimator based on MECAPM worsen. In particular, for the quartile of banks with largest systemicness, the median percentage bias of the MECAPM estimator of systemicness is always

	50% GIIPS $AV = 496.3\%$		10% E.U. Gov. $AV = 270.8\%$	10% All Gov. $AV = 357\%$
	$\widehat{AV}$	% Bias	% Bias	% Bias
CECAPM	480.4 %	3.2%	3.6 %	5.1 %
BIPWCM	361.9%	27.1 %	28.6 %	22.4 %
BIPECM	392.9 %	20.8%	20.6 %	12.5 %
MECAPM	436.7 %	12%	12.4 %	4.1 %

TABLE 3.1: Comparison, for the EBA data, between real and estimated values of the Aggregate Vulnerability (AV). The three columns correspond to different shocks. Under the name of each shock we report the corresponding real AV, computed from the complete knowledge of banks' portfolios. In the first column we report the estimated AV and the percentage bias for the 4 different ensembles, resulting from a 50% value loss of GIIPS sovereign debt, as considered in Greenwood, Landier, and Thesmar, 2015. In the second and third columns we report the percentage biases of estimated AV for two alternative scenarios: a 10% loss of value for either all the E.U. sovereign debt or the sovereign debt of all countries.

In summary, the estimates of systemicness and indirect vulnerably for each single bank as provided by the CECAPM-implied matrix are almost identical to those obtained as the corresponding expected values on the MECAPM ensemble. Besides, they are satisfactorily accurate and surely more reliable than those provided by standard maximum entropy ensembles. Once more, the important message is that it is possible to achieve pretty accurate estimates of systemic risk metrics, at the aggregate or individual institution level, due to fire sales spillover *without* a full knowledge of the portfolio holdings of financial institutions.

3.5.3 Monitoring and testing changes in systemicness

As another application of the ensembles of graphs obtained with the Maximum Entropy method, we consider here the problem of assessing whether the systemicness of a given bank (or of the whole system) has changed in a statistically significant way. In order to answer this question, it is necessary to have a null hypothesis and we propose to use network ensembles to this end. Since the MECAPM shows superior performances in estimating risk metrics, in this section we use it and we propose a possible application for statistical validation. Our objective here is not to study all the banks and all the quarters, but only to show how the testing method can be implemented.

In particular, imagine a regulator who monitors a given bank, measuring its systemicness and searching for evidences of a significant increase. Having a given quarter as reference, the regulator can extract the distribution of bank's systemicness and, in the subsequent quarters, identify when the systemicness is outside a given confidence interval around the reference period. As a special case, we select four banks among the top fifty in the first quarter and that exist for the entire time

between -20% and -30%. Similarly, the median of the percentage bias in the estimation of indirect vulnerability via MECAPM is always between -20% and -35%. Nevertheless, the ranking among the three estimation methods remains unchanged.

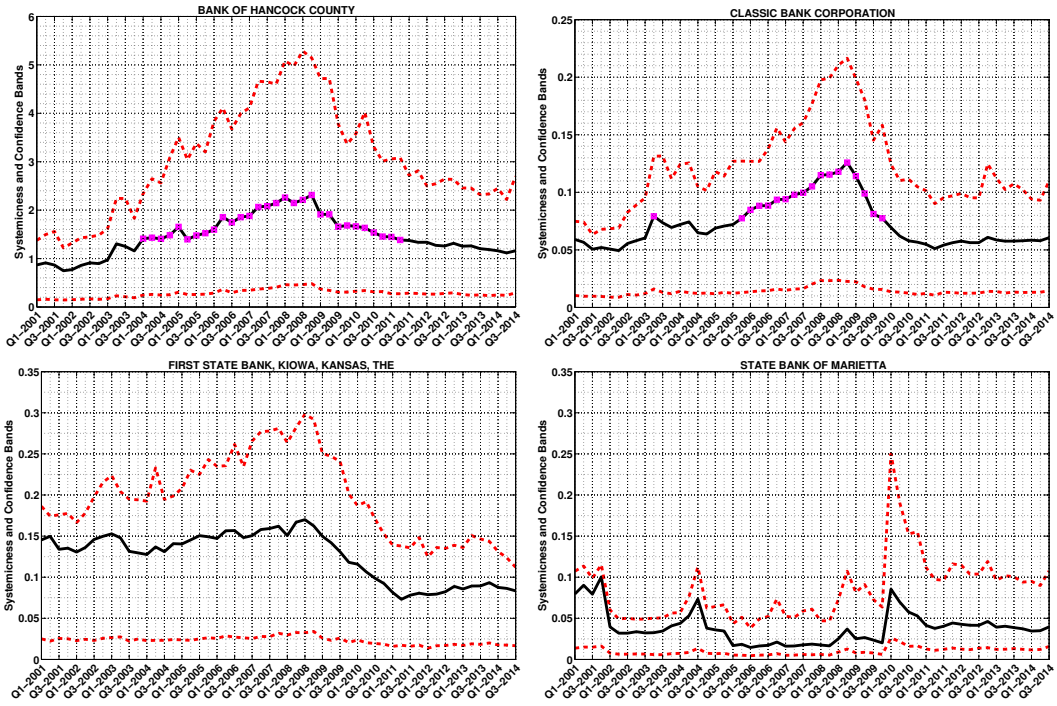

FIGURE 3.7: We report, for four selected banks, the true systemicness (thick dotted lines) and the 5%-95% confidence bands according to the MECAPM ensemble. A magenta square is added in every quarter in which the systemicness of the bank is above the 95% confidence level of the first quarter of 2001.

period (i.e. they do not exit the dataset). For each quarter we compute the true bank systemicness and the 5%-95% confidence bands according to the MECAPM ensemble (see Figure 3.7). We then added a magenta square in each quarter when the true systemicness is above the 95% confidence band of the first quarter, used as reference. Hence, a magenta square is indicating a quarter when the systemicness of the bank is statistically larger (according to the MECAPM) than at beginning of 2001. We show two banks for which a statistically significant change in systemicness is observed (top row) and two for which no change is observed (bottom row). Notably, for the former case we find that the systemicness of the banks analyzed increased significantly much before the onset of the 2007-2008 financial crisis. This phenomenon persisted along the entire period of the crisis and vanished not before the end of 2009. This suggests that network statistical models could be of valuable help in the surveillance activity of central banks and other supervisory authorities as monitoring tools and in constructing early warning indicators.

3.6 Conclusions

In this chapter we focused on the problem of estimating metrics of systemic risk due to fire sale spillover in presence of limited information on the composition of portfolios of financial institutions. A full knowledge of the portfolio holdings of each institution in the economy is generally required to have a precise estimate of any risk metrics that, as those proposed by Greenwood, Landier, and Thesmar, 2015, is based on the mechanism of portfolio rebalancing through fire sales. Nevertheless,

such a huge and detailed information may not be available, especially at frequency higher than quarterly, making the estimation of systemic risk quite difficult. In this chapter we circumvent the problem by providing accurate estimates of systemic risk metrics that are based on a partial knowledge of the system, more precisely only on the sizes of balance sheets and the capitalization of assets (or asset classes), which are much easier to trace. In this respect, we have shown that the method of Cross-Entropy minimization does a very good job in estimating aggregate vulnerability and individual bank systemicness without requiring any knowledge of the underlying matrix of bank portfolio holdings.

Furthermore, we have compared the results with a Max Entropy ensembles. Specifically we have introduced a new ensemble (MECAPM), which reproduces, on average, the CECAPM and performs quite well in estimating systemicness and indirect vulnerability of single institutions, outperforming standard Max Entropy competitors. Moreover the estimation of systemic risk metrics could provide valuable information to any policy maker, but variations in systemicness and indirect vulnerability are difficult to interpret in absence of a statistical validation. For this reason, as a final contribution, we have proposed to use the Max Entropy ensemble to assess the statistical significance of systemic risk metrics. On a selection of banks of our dataset we documented that their systemicness significantly increased, with respect to the level observed at the beginning of the 2001, much before the onset of the 2007-2008 financial crisis. Even if deeper investigations are required in this direction, we believe that this approach could be easily implemented as an early warning indicator of systemic risk.

Finally, we would like to comment again on the scope of the Greenwood, Landier, and Thesmar, 2015 model as well of our work. As discussed in the main text, the considered methodology belongs to the classic static stress test approach. Only the portfolios and balance sheets at the time of the tested shock are used and no intertemporal dynamics is ever considered. This is a serious limitation, since financial distress and deleveraging might occur on longer periods and the bank's decision at a given quarter can depend, not only on the present price changes and portfolio composition, but also on past market state and banks' behavior. We believe that extending the Greenwood, Landier, and Thesmar, 2015 approach to a dynamic stress test setting is a very interesting avenue for research both for academicians and for regulators.

3.7 Appendix

3.7.1 Data Description and Dataset Creation

This appendix provides some descriptive features of the data along with the method adopted to build the 20 asset classes of the bank-asset network analyzed in the chapter. The left panel (first row) of Figure 3.8 reports, on a log-log scale, the kernel density of the bank sizes (i.e. the total amount of assets detained by the bank) pooled across all quarters. It is evident that bank sizes are quite heterogeneous. The right panel (first row) of Figure 3.8 reports the density of the bank leverages T_n pooled across all quarters. In this case we observe a much less heterogeneous distribution, with most banks showing a leverage around 10. Finally, the second row of Figure 3.8 reports the relation between size and leverage. The plot is achieved by sorting all records of bank size from the smallest to the largest and then applying a moving-window procedure. As expected from the density plots, there is no relation between leverage and bank size, having most bank a leverage of 10 and a highly heterogeneous size.

Concerning the formation of the asset classes used in the main text, we provide in what follows details on how they have been created. As mentioned in the main text, the focus of the chapter is on commercial banks, whose precise definition is given by the FFIEC as *[...] every national bank, state member bank, insured state nonmember bank, and savings association is required to file a consolidated Call Report normally as of the close of business on the last calendar day of each calendar quarter, i.e., the report date. The specific reporting requirements depend upon the size of the bank and whether it has any "foreign" offices [...]*. This is the set of institutions that is referred as Commercial Banks throughout all the chapter.

Forms FFIEC031 and FFIEC041 are dedicated to, respectively, banks with only domestic offices and banks with domestic and foreign offices. However, in both forms, it is adopted the same coding system. More specifically there are only two types of codes, RCON and RCFD, which are followed by a four digits alphanumerical code. The alphanumerical code identifies the budget item, for example 2170 refers to total assets of the bank. The prefix RCON is used for financial items relative to domestic offices, while RCFD encompasses both domestic and foreign offices. Hence RCON2170 is the code for the total assets of the bank detained in U.S. offices, while RCFD2170 is relative to the sum of total assets detained in U.S. plus offices abroad. Of course, for banks that fill the FFIEC031 the two codes RCON and RCFD report the same value if they have the same alphanumerical code.

Table 3.2 documents the detailed composition of each asset class. For each asset class (first column) we report the composition in terms of FFEIC items in the third column and a short name given to the asset class in the second one. Such abbreviation is needed since some asset class, e.g. "loans to consumers in foreign offices", are assembled subtracting from the FFIEC codes some previously defined asset classes. There is a one-to-one correspondence between asset classes and variable names, a part for the case of "loans secured by real estates in domestic offices", which is computed as the sum of five variables, from "construction loans" to "non farm, non residential". The composition of the FFEIC formula reported in the third column may vary during time, hence we report in bold the period of validity of the formula adopted. In this respect, note that the date **12/99** refers to the last available quarter, that is the third quarter of 2014. In reporting the FFEIC formula, we adopt the convention that the prefix is omitted whenever RCON is used solely for banks with only domestic offices and RCFD solely for those that have at least on office abroad. On

the contrary, when the prefix is specified, it means that only the code with that particular prefix is being been used. For example the code RCON3532 is used only in its domestic version, hence we do not use RCFD3532 for banks with offices abroad.

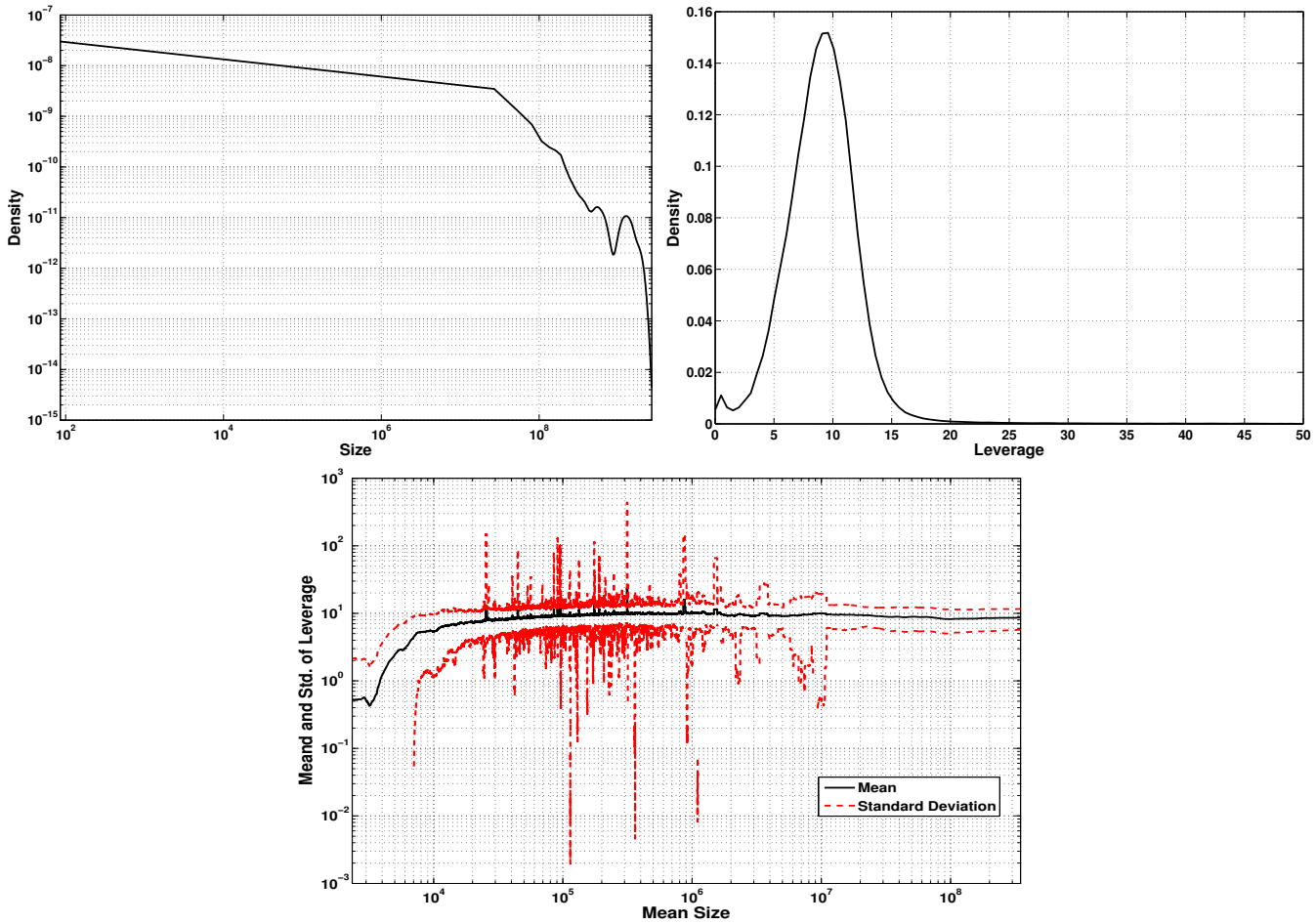

FIGURE 3.8: This Figure reports some descriptive features of the data analyzed. Top left panel plots, on a log-log scale, the kernel density of bank sizes (defined as total assets in unit of 10^3\$) while top right is the kernel density of the bank leverages. Both densities are computed using all records pooled across the entire time span. For the sake of visualization, we put a cut-off of 50 on the maximum leverage allowed, although leverages of more than 150 are (rarely) observed. The bottom panel shows that there is no relation between leverage and size. The procedure adopted to draw the plot is the following: all records of bank size are sorted from the smallest to the largest one and a rolling window of 1000 records is moved, with an incremental shift of 10 records, from the first to the last. In each window we compute the mean leverage (black continuous line) and the standard deviation of leverage (red dotted line) of banks that fall in the window. Mean and standard deviation are plotted as a function of the mean size in the window, which is reported in the horizontal axis.

TABLE 3.2: Composition of Asset Classes

Asset Class	Variable Name	FFIEC Formula
Total assets	tot_ass	**03/01-12/99**: 2170+2123+3123
Equity	equity	**03/01-03/09**: 3210+3000 **03/09-12/99**: G105
Cash and balances due from depository institutions	cahab	**03/01-12/99**: 0081+0071
U.S. treasury securities	ust_sec	**03/01-12/99**: 0211+1287+RCON3531
U.S agency securities	agency_sec	**03/11-12/99**: 1289+1294+1293+1298+RCON3532
Securities issued by state and local governments	state_sec	**03/01-12/99**: 8496+8499+RCON3533
Mortgage backed securities	mbs	**03/01-03/09**: 1698+1702+1703+1707+1709+1713+1714+1717+1718+1732+ 1733+1736+RCON3534+RCON3535+RCON3536. **06/09-12/10**: G300+G303+G304+G307+G308+G311+G312+G315+G316+ G319+G320+G323+G324+G327+G328+G331+RCONG379+RCONG380+ RCONG381+RCONG382 **03/11-12/99**: G300+G303+G304+G307+G308+G311+G312+G315+G316+ G319+G320+G323+K142+K146+K145+ K149+K150+K154+K153+K157+ RCONG379+RCONG380+RCONG381+RCONK197+RCONK198
Asset backed securities	abs	**03/01-12/05**: B838+B841+B842+B845+B846+B849+B850+B853+B854+ B857+B858+B861 03/06-03/09 C026+C027 **06/09-12/99**: C026+C027+G336+G340+G344+G339+G343+G347
Other domestic debt securities	dom_debt_oth_sec	**03/01-12/99**: 1737+1741
Foreign debt securities	for_debt_sec	**03/01-12/99**: 1742+1746
Residual securities	res_sec	**03/01-12/99**: A511
Futures, forwards sold and securities purchased under the agreement to resell (asset)	ffrepo_ass	**03/01-12/01**: 1350 **03/02-12/99**: RCONB987+B989
Loans secured by real estates in domestic offices	Construction loans	**03/01-12/07**: RCON1415 **03/08-12/99**: RCONF158+RCONF159
	Secured by farmland	**03/01-12/99**: RCON1420
	1-4 Family real estate	**03/01-12/99**: RCON5367+RCON5368+RCON1797
	Multifamily property loans	**03/01-12/99**: RCON1460
	Non farm, non residential	**03/01-12/07**: RCON1480 **03/08-12/99**: RCONF160+RCONF161
Loans secured by real estate in foreign offices	ln_re_for	**03/01-12/99**: (if present) RCFD1410 - ln_const - ln_farm - ln_rre - ln_multi - ln_nfnr, **3/01-12/99**: (otherwise) zero
Commercial and industrial loans in domestic offices	ln_ci_dom	**03/01-12/99**: RCON1766
Commercial and industrial loans in foreign offices	ln_ci_for	**03/01-12/99**: (if present) RFCD1763+RFCD1764 - RCON1766, **03/01-12/99**: (otherwise) zero
Loans to consumers in domestic offices	ln_cons_dom	**03/01-12/10**: RCON2011+RCONB538+RCONB539 **03/11-12/99**: +RCONB538+RCONB539+RCONK137+RCONK207
Loans to consumers in foreign offices	ln_cons_for	**03/01-12/10**: (if present) RCFD2011+ RCFDB538+ RCFDB539 - ln_cons_dom, (otherwise) zero **03/11-12/99** (if present) RCFDB538+RCFDB539+ RCFDK137+RCFDK207-ln_cons_dom, (otherwise) zero
Loans to depository institutions and acceptances of other banks	ln_dep_inst_banks	**03/01-12/99**: (if present) RCFDB532+RCFDB533+RCFDB534+ RCFDB536+RCFDB537, (otherwise) RCON1288
other loans	oth_loans	**03/01-12/99**: 2122+2123-ln_const-ln_farm-ln_rre-ln_multi-ln_nfnr- ln_re_for-ln_ci_dom-ln_ci_for-ln_cons_dom-ln_cons_for-ln_dep_inst_banks
Equity securities that do not have readily determinable fair value	equ_sec_nondet	**03/01-12/99**: 1752
other assets	oth_ass	**03/01-12/99**: tot_ass - all preceding assets

3.7.2 Max Entropy Ensembles

In this appendix we provide the details of the derivation of the probability mass functions for the MECAPM, BIPWCM and BIPECM ensembles.

Maximum Entropy Capital Asset Pricing Model

Considering the maximum entropy optimization problem presented in Section 2.4 with constraint functions $\phi_{nk} = Y_{nk}$, and λ_{nk} as Lagrange multipliers, we obtain the normalizing factor

$$\begin{aligned} K_{\vartheta} &= \sum_{\mathbf{Y}\in\mathcal{Y}} e^{-\sum_{n=1}^{N}\sum_{k=1}^{K}\lambda_{nk}Y_{nk}} \\ &= \sum_{\mathbf{Y}\in\mathcal{Y}}\prod_{n=1}^{N}\prod_{k=1}^{K} e^{-\lambda_{nk}Y_{nk}} \\ &= \prod_{n=1}^{N}\prod_{k=1}^{K}\sum_{Y_{nk}=0}^{\infty} e^{-\lambda_{nk}Y} \\ &= \prod_{n=1}^{N}\prod_{k=1}^{K}\frac{1}{1-e^{-\lambda_{nk}}}. \end{aligned} \tag{3.12}$$

Hence

$$\mathbb{P}\left(\mathbf{Y}\right) = \prod_{n=1}^{N}\prod_{k=1}^{K}\frac{e^{-\lambda_{nk}Y_{nk}}}{1-e^{-\lambda_{nk}}}.$$

Note that the *partition function* K_{ϑ} in (3.12) is such that

$$\frac{\partial\log\left(K_{\vartheta}\right)}{\partial\lambda_{nk}} = -\mathbb{E}_{\mathcal{Y}}\left[Y_{nk}\right]. \tag{3.13}$$

Hence, imposing the CAPM structure as required in (3.8), the Lagrange multipliers are determined by

$$-\frac{\partial\log K_{\vartheta}}{\partial\lambda_{nk}} = \frac{D_n^{\star}C_k^{\star}}{L^{\star}},$$

which gives the probability mass function for the MECAPM

$$\mathbb{P} = \prod_{n=1}^{N}\prod_{k=1}^{K}\left(\frac{Y_{nk}^{\mathrm{CAPM}}}{1+Y_{nk}^{\mathrm{CAPM}}}\right)^{Y_{nk}}\left(\frac{1}{1+Y_{nk}^{\mathrm{CAPM}}}\right), \tag{3.14}$$

where

$$Y_{nk}^{\mathrm{CAPM}} = \frac{D_n^{\star}C_k^{\star}}{L^{\star}}.$$

Bipartite Weighted Configuration Model.

When we want to use as economic information the total asset size of each bank and the total capitalization of each asset, without imposing the MECAPM structure, we need one Lagrange multiplier λ_n for each $\phi_n = \sum_k Y_{nk}$, and one Lagrange multiplier η_k for each $\phi_k = \sum_n Y_{nk}$. We can go on with computation by explicitly writing the

expression of $D_n(\mathbf{Y})$ and $C_k(\mathbf{Y})$ in terms of the elements of the matrix $\mathbf{Y}$, obtaining

$$\begin{aligned} K_{\vartheta} &= \sum_{\mathbf{Y}\in\mathcal{Y}} \prod_{n=1}^{N}\prod_{k=1}^{K} e^{-Y_{nk}(\lambda_n+\eta_k)} \\ &= \sum_{Y_{1,1}=0}^{\infty}\cdots\sum_{Y_{1,K}=0}^{\infty}\cdots\sum_{Y_{n,1}=0}^{\infty}\cdots\sum_{Y_{nk}=0}^{\infty}\prod_{n=1}^{N}\prod_{k=1}^{K} e^{-Y_{nk}(\lambda_n+\eta_k)} \\ &= \prod_{n=1}^{N}\prod_{k=1}^{K}\sum_{Y_{nk}=0}^{\infty} e^{-Y_{nk}(\lambda_n+\eta_k)} \\ &= \prod_{n=1}^{N}\prod_{k=1}^{K}\frac{1}{1-e^{-(\lambda_n+\eta_k)}} = \prod_{n=1}^{N}\prod_{k=1}^{K}\frac{1}{1-\varphi_n\xi_k}, \end{aligned}$$

where $\varphi_n = e^{-\lambda_n}$ and $\xi_k = e^{-\eta_k}$, whence

$$\begin{aligned} \mathbb{P}(\mathbf{Y}) &= \frac{\prod_{n=1}^{N}\prod_{k=1}^{K} e^{-Y_{nk}(\lambda_n+\eta_k)}}{\prod_{n=1}^{N}\prod_{k=1}^{K}\frac{1}{1-\varphi_n\xi_k}} \\ &= \prod_{n=1}^{N}\prod_{k=1}^{K}(\varphi_n\xi_k)^{Y_{nk}}(1-\varphi_n\xi_k), \end{aligned} \tag{3.15}$$

The value of the Lagrange multipliers are determined by imposing that the expected value of $D_n(\mathbf{Y})$ and $C_k(\mathbf{Y})$ on the ensemble $\mathcal{Y}$ are equal to, respectively, $D_n^{\star}$ and $C_k^{\star}$. As for the MECAPM case, note that K_{ϑ} is such that

$$\mathbb{E}_{\mathcal{Y}}[D_n] = -\frac{\partial\log(K_{\vartheta})}{\partial\lambda_n},$$

and similarly

$$\mathbb{E}_{\mathcal{Y}}[C_k] = -\frac{\partial\log(K_{\vartheta})}{\partial\eta_k}.$$

Hence we can compute $\mathbb{E}_{\mathcal{Y}}[D_n]$ and $\mathbb{E}_{\mathcal{Y}}[C_k]$ explicitly as a function of the Lagrange multipliers, that is

$$\begin{aligned} \mathbb{E}_{\mathcal{Y}}[D_n] &= \frac{\partial}{\partial\lambda_n}\sum_{n=1}^{N}\sum_{k=1}^{K}\log\left(1-e^{-(\lambda_n+\eta_k)}\right) \\ &= \sum_{k=1}^{K}\frac{e^{-(\lambda_n+\eta_k)}}{1-e^{-(\lambda_n+\eta_k)}}, \\ \mathbb{E}_{\mathcal{Y}}[C_k] &= \frac{\partial}{\partial\eta_k}\sum_{n=1}^{N}\sum_{k=1}^{K}\log\left(1-e^{-(\lambda_n+\eta_k)}\right) \\ &= \sum_{n=1}^{N}\frac{e^{-(\lambda_n+\eta_k)}}{1-e^{-(\lambda_n+\eta_k)}}. \end{aligned}$$

Therefore the Lagrange multipliers are determined by numerically solving the non-linear system of equations

$$\begin{cases} \sum_{k=1}^{K}\frac{\varphi_n\xi_k}{1-\varphi_n\xi_k} = D_n^{\star},\ n=1,...,n, \\ \sum_{n=1}^{N}\frac{\varphi_n\xi_k}{1-\varphi_n\xi_k} = C_k^{\star},\ k=1,...,K. \end{cases} \tag{3.16}$$

Bipartite Enhanced Configuration Model.

The only difference with the Weighted model described in Appendix 3.7.2 is the addition of the constraints on the number of degrees for each node. Before proceeding, we have thus to add some additional definitions.

The binary projection of Y_{nk} is defined as the matrix $Y_{nk} = \mathbb{1}_{Y_{nk}>0}$. Accordingly, the number D_n^{row} of assets in which the n-th bank invests and the number D_k^{col} of banks that own the k-th asset class are computed as

$$D_n^{\text{row}}(\mathbf{Y}) = \sum_{k=1}^{K} Y_{nk}, \qquad D_k^{\text{col}}(\mathbf{Y}) = \sum_{n=1}^{N} Y_{nk}, \tag{3.17}$$

where the capital letter D stands for *degree*, as it is common practice in network theory[18].

The maximization problem for the BIPECM case is hence stated as

$$\begin{aligned}
\max_{\mathbb{P}} \quad & S(\mathbb{P}[\mathbf{Y}]) \\
\text{s.t.} \quad & \sum_{\mathbf{Y}\in\mathcal{Y}} \mathbb{P}[\mathbf{Y}] = 1 \\
& \mathbb{E}_{\mathcal{Y}}[D_n] = D_n^{\star},\ n = 1,...,N, \\
& \mathbb{E}_{\mathcal{Y}}[D_n^{\text{row}}] = D_n^{\text{row}^{\star}},\ n = 1,...,N, \\
& \mathbb{E}_{\mathcal{Y}}[C_k] = C_k^{\star},\ k = 1,...,K, \\
& \mathbb{E}_{\mathcal{Y}}\left[D_k^{\text{col}}\right] = D_k^{\text{col}^{\star}},\ k = 1,...,K.
\end{aligned}$$

With an obvious extension of the number of Lagrange multipliers, and since $D_n^{\text{row}}(\mathbf{Y}) = \sum_{k=1}^{K} \mathbb{1}_{Y_{nk}>0}$ (similarly, $D_k^{\text{col}}(\mathbf{Y}) = \sum_{n=1}^{N} \mathbb{1}_{Y_{nk}>0}$), we obtain

$$\begin{aligned}
K_{\vartheta} &= \sum_{\mathbf{Y}\in\mathcal{Y}} \prod_{n=1}^{N}\prod_{k=1}^{K} e^{-\left(Y_{nk}(\lambda_n+\eta_k)+\mathbb{1}_{Y_{nk}>0}(\rho_n+\delta_k)\right)} \\
&= \sum_{Y_{1,1}=0}^{\infty}\cdots\sum_{Y_{1,K}=0}^{\infty}\cdots\sum_{Y_{n,1}=0}^{\infty}\cdots\sum_{Y_{nk}=0}^{\infty}\prod_{n=1}^{N}\prod_{k=1}^{K} e^{-\left(Y_{nk}(\lambda_n+\eta_k)+\mathbb{1}_{Y_{nk}>0}(\rho_n+\delta_k)\right)} \\
&= \prod_{n=1}^{N}\prod_{k=1}^{K}\sum_{Y_{nk}=0}^{\infty} e^{-\left(Y_{nk}(\lambda_n+\eta_k)+\mathbb{1}_{Y_{nk}>0}(\rho_n+\delta_k)\right)} \\
&= \prod_{n=1}^{N}\prod_{k=1}^{K}\left(1+e^{-(\rho_n+\delta_k)}\sum_{Y_{nk}=1}^{\infty} e^{-(Y_{nk}(\lambda_n+\eta_k))}\right) \\
&= \prod_{n=1}^{N}\prod_{k=1}^{K}\left(1+\frac{e^{-(\rho_n+\delta_k)}}{1-e^{-(\lambda_n+\eta_k)}}-e^{-(\rho_n+\delta_k)}\right).
\end{aligned}$$

Defining $\varphi_n = e^{-\lambda_n}$, $\xi_k = e^{-\eta_k}$, $\psi_n = e^{-\rho_n}$ and $\gamma_k = e^{-\delta_k}$ we get

$$K_{\vartheta} = \prod_{n=1}^{N}\prod_{k=1}^{K}\frac{1-\varphi_n\xi_k(1-\psi_n\gamma_k)}{1-\varphi_n\xi_k}. \tag{3.18}$$

[18] We have chosen not to use a different notation for bank and asset degrees since, although they have an immediate economic interpretation, they are not frequently used in economic analysis. In fact, while D_n and C_k are typically publicly available, the degree sequences D_n^{row} and D_k^{col} are instead difficult to trace.

Finally, we get that the probability mass function for the BIPECM is

$$\mathbb{P}_{\vartheta}(\mathbf{Y}) = \prod_{n=1}^{N}\prod_{k=1}^{K} \frac{(1-\varphi_n\xi_k)\ (\varphi_n\xi_k)^{Y_{nk}}\ (\psi_n\gamma_k)^{\mathbb{1}_{Y_{nk}>0}}}{1-\varphi_n\xi_k\ (1-\psi_n\gamma_k)}. \tag{3.19}$$

The determination of the Lagrange multipliers follows a procedure identical to that described for BIPWCM in Appendix 3.7.2, that is the expected values of $D_n(\mathbf{Y})$, $C_k(\mathbf{Y})$, $D_n^{\text{row}}(\mathbf{Y})$ and $D_k^{\text{col}}(\mathbf{Y})$ are computed as the partial derivative of $-\log(K_{\vartheta})$ with respect to the corresponding Lagrange multiplier and then equated to the observed value. This procedure produce the non-linear system of equations

$$\begin{cases} \sum_{k=1}^{K} \frac{\phi_n\xi_k\psi_n\gamma_k}{(1-\phi_n\xi_k)(1-\phi_n\xi_k(1-\psi_n\gamma_k))} &= D_n^{\star},\ n=1,...,n, \\ \sum_{k=1}^{K} \frac{\phi_n\xi_k\psi_n\gamma_k}{1-\phi_n\xi_k(1-\psi_n\gamma_k)} &= D_n^{\text{row}^{\star}},\ n=1,...,n, \\ \sum_{n=1}^{N} \frac{\phi_n\xi_k\psi_n\gamma_k}{(1-\phi_n\xi_k)(1-\phi_n\xi_k(1-\psi_n\gamma_k))} &= C_k^{\star},\ k=1,...,K, \\ \sum_{n=1}^{N} \frac{\phi_n\xi_k\psi_n\gamma_k}{1-\phi_n\xi_k(1-\psi_n\gamma_k)} &= D_k^{\text{col}^{\star}},\ k=1,...,K. \end{cases} \tag{3.20}$$

3.7.3 Comparison of Reconstruction Methods under Different Shocks

In this appendix we report additional comparisons between cross entropy and maximum entropy ensemble methods for systemic risk assessment. In Figure 3.9 we show the performances of the four methods presented in the chapter in reconstructing Aggregate Vulnerability, for four different shock scenarios. In all cases the CECAPM and MECAPM estimation outperform the other two and track quite closely the AV obtained from the full knowledge of portfolios composition. We therefore conclude that our result is not due to the uniform shock assumption, but is more generically applicable.

We then compare the performances of the different methods in assessing individual bank's quantities, it Indirect Vulnerability and Systemicness. In each quarter we have from $N = 6,500$ to $N = 9,000$ values of relative errors for each metric. To visualize the result we plot the median of the relative error and as a measure of dispersion we use the interquartile range, i.e. the difference between the upper and lower quartiles[19]. Clearly a median well centered around zero is an indication that the estimator is unbiased.

Figure 3.5 shows the results for bank systemicness (top panel) and indirect vulnerability (bottom panel). The three different colors refer to the three different ensembles (see the figure caption for more details). We do not report the results for the CECAPM because it is indistinguishable from the MECAPM one and this fact can be understood by the argument in Eq. (3.10).

We observe that for each quarter BIPWCM strongly underestimates individual bank systemicness and indirect vulnerability. The median relative error ranges roughly between -60% and -70% and the interquartile range is very far from zero. The estimator based on BIPECM (using the additional information on degrees) gives slightly better results, even if a strong underestimation is still present. The median relative error ranges roughly between -50% and -40% and again the interquartile range is

[19] We choose these metrics because they are more robust and less sensitive to outliers.

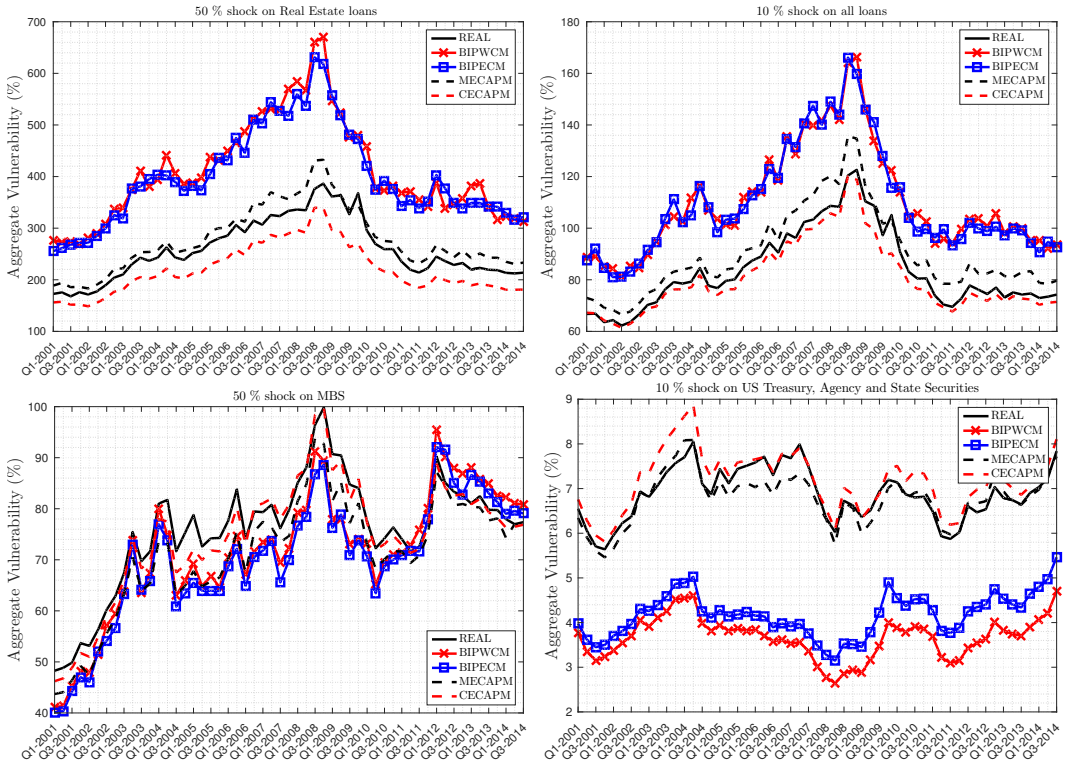

FIGURE 3.9: Aggregate vulnerability under different shock scenarios. Each panel reports the AV obtained from the full knowledge of portfolios composition and those obtained using the four reconstruction methods considered.

far from zero. On the contrary the estimator based on MECAPM (or CECAPM) performs much better. The median relative error never goes below -20% and almost always the interquartile range is centered around zero.[20]

In summary, the estimates of systemicness and indirect vulnerably for each single bank as provided by the CECAPM-implied matrix are almost identical to those obtained as the corresponding expected values on the MECAPM ensemble. Besides, they are satisfactorily accurate and surely more reliable than those provided by standard maximum entropy ensembles.

3.7.4 Relation between CECAPM and MECAPM

This appendix is dedicated to derive explicit formulas for the expected values of bank systemicness and indirect vulnerability under the MECAPM ensemble. Moreover we show that these expected values are well approximated by the corresponding values returned by the CECAPM-implied matrix. First note that, according to equation (3.14), the element Y_{nk} is geometrically distributed with parameter $p =$

[20] If instead we focus on the banks with higher systemicness or indirect vulnerability, the performances of the estimator based on MECAPM worsen. In particular, for the quartile of banks with largest systemicness, the median percentage bias of the MECAPM estimator of systemicness is always between -20% and -30%. Similarly, the median of the percentage bias in the estimation of indirect vulnerability via MECAPM is always between -20% and -35%. Nevertheless, the ranking among the three estimation methods remains unchanged.

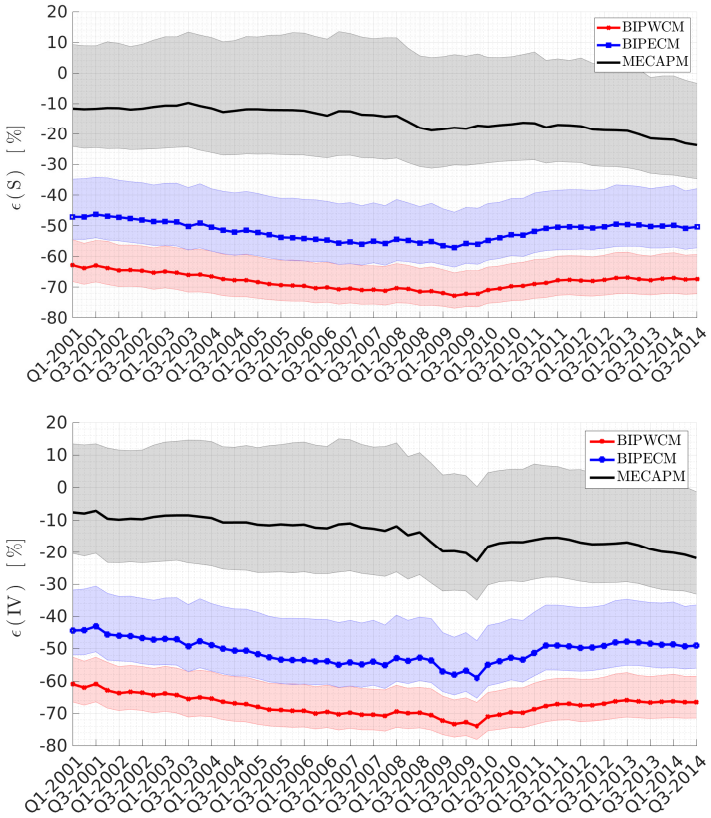

FIGURE 3.10: Time series of the relative error (see Eq. 3.7) of bank systemicness (top panel) and indirect vulnerability (bottom panel) with respect to real data as estimated by the three ensembles BIPWCM (red and squares), BIPECM (blue and circles), and MECAPM (grey and dashed line). The thick lines indicate the median and the colored areas the interquartile range.

$\frac{1}{1+Y_{nk}^{\mathrm{CAPM}}}$ and thus

$$\mathbb{E}\left[{Y_{nk}}^2\right] - \mathbb{E}\left[Y_{nk}\right]^2 = \frac{1-p}{p^2} = Y_{nk}^{\mathrm{CAPM}}\left(Y_{nk}^{\mathrm{CAPM}}+1\right).$$

Hence the expected value of the systemicness of bank n is given by[21]

$$\begin{aligned}
\mathbb{E}\left[S_n\left(\mathbf{Y}\right)\right] &= \mathbb{E}\left[\Gamma_n\right]\frac{D_n^\star}{E^\star}T_n^\star \\
&= \ell\,\frac{T_n^\star}{E^\star}\sum_{k=2}^{K}\left(\mathbb{E}\left[{Y_{nk}}^2\right] + \sum_{m\neq n}\mathbb{E}\left[Y_{nk}\right]\mathbb{E}\left[Y_{nk}\right]\right) \\
&= \ell\,\frac{D_n^\star\,T_n^\star}{E^\star\,L^\star}\sum_{k=2}^{K}C_k^\star\left(Y_{nk}^{\mathrm{CAPM}}+C_k^\star+1\right).
\end{aligned}$$

[21] Remember that in our setting all the illiquidity parameters are set to a common value ℓ except for the cash, which we assume to be the first class $k=1$, for which it is set to zero, whence the sum from $k=2$ to $k=K$.

Note that since

$$S_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right) = \ell\,\frac{D_n^\star\,T_n^\star}{E^\star\,L^\star}\,\sum_{k=2}^{K} C_k^{\star 2},$$

it is

$$\mathbb{E}\left[S_n\left(\mathbf{Y}\right)\right] = S_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right) + \ell\,\frac{D_n^\star\,T_n^\star}{E^\star\,L^\star}\,\sum_{k=2}^{K} C_k^\star\,\left(Y_{nk}^{\mathrm{CAPM}}+1\right),$$

whence

$$\frac{\mathbb{E}\left[S_n\left(\mathbf{Y}\right)\right] - S_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right)}{S_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right)} = \frac{\sum_{k=2}^{K} C_k^\star\,\left(Y_{nk}^{\mathrm{CAPM}}+1\right)}{\sum_{k=2}^{K} C_k^{\star 2}}.$$

This expression can be rewritten as

$$\mathbb{E}\left[S_n\left(\mathbf{Y}\right)\right] = S_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right)\left(1+\frac{D_n^*}{L^*}+\frac{\sum_{k=2}^{K} C_k}{\sum_{k=2}^{K} C_k^{\star 2}}\right),$$

In order to evaluate the relative error above consider the simplified case in which all the capitalization are almost equal $C_k^\star \approx C^\star = \frac{L^\star}{K}$, hence

$$\begin{aligned}\frac{\mathbb{E}\left[S_n\left(\mathbf{Y}\right)\right] - S_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right)}{S_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right)} &\approx \frac{\left(D_n^\star + K\right)}{L^\star}\\ &\approx \frac{D_n^\star}{L^\star} \ll 1,\end{aligned}$$

where the last inequality follows from the fact that $L^\star = \sum_{n=1}^{N} D_n^\star$.

Concerning the indirect vulnerability we have

$$\begin{aligned}\mathbb{E}\left[\mathrm{IV}_n\left(\mathbf{Y}\right)\right] &= \ell\,\frac{\left(1+T_n^\star\right)}{D_n^\star}\,\mathbb{E}\left[\sum_{k=2}^{K} Y_{nk}\,\sum_{m=1}^{N} Y_{m,k}\,B_m^\star\right]\\ &= \ell\,\frac{\left(1+T_n^\star\right)}{D_n^\star}\,\mathbb{E}\left[\sum_{k=2}^{K}\left({Y_{nk}}^2\,T_n^\star + \sum_{m\neq n} Y_{nk}\,Y_{m,k}\,B_m^\star\right)\right]\\ &= \ell\,\frac{\left(1+T_n^\star\right)}{D_n^\star}\,\sum_{k=2}^{K}\left[Y_{nk}^{\mathrm{CAPM}}\,\left(2\,Y_{nk}^{\mathrm{CAPM}}+1\right)T_n^\star + \sum_{m\neq n} Y_{nk}^{\mathrm{CAPM}}\,Y_{m,k}^{\mathrm{CAPM}}\,B_m^\star\right]\\ &= \ell\,\frac{\left(1+T_n^\star\right)}{D_n^\star}\,\sum_{k=2}^{K}\left[\left(2\,Y_{nk}^{\mathrm{CAPM}^2}+Y_{nk}^{\mathrm{CAPM}}\right)T_n^\star + \sum_{m\neq n} Y_{nk}^{\mathrm{CAPM}}\,Y_{m,k}^{\mathrm{CAPM}}\,B_m^\star\right]\\ &= \ell\,\frac{\left(1+T_n^\star\right)}{D_n^\star}\,\sum_{k=2}^{K}\left[\left(Y_{nk}^{\mathrm{CAPM}^2}+Y_{nk}^{\mathrm{CAPM}}\right)T_n^\star + \sum_{m=1}^{N} Y_{nk}^{\mathrm{CAPM}}\,Y_{m,k}^{\mathrm{CAPM}}\,B_m^\star\right]\\ &= \ell\,\frac{\left(1+T_n^\star\right)}{D_n^\star}\,\sum_{k=2}^{K}\left[\left(\frac{D_n^{\star 2}\,C_k^{\star 2}}{L^{\star 2}}+\frac{D_n^\star\,C_k^\star}{L^\star}\right)T_n^\star + \frac{D_n^\star\,C_k^{\star 2}}{L^{\star 2}}\,\sum_{m=1}^{N} D_m^\star\,B_m^\star\right].\end{aligned}$$

Now consider that $B_m^\star \approx \bar{B} = 10$, whence

$$\begin{aligned}\mathbb{E}\left[\mathrm{IV}_n\left(\mathbf{Y}\right)\right] &\approx \ell\,\frac{\bar{B}\,\left(1+\bar{B}\right)}{D_n^\star}\,\sum_{k=2}^{K}\left[\frac{D_n^{\star 2}\,C_k^{\star 2}}{L^{\star 2}}+\frac{D_n^\star\,C_k^\star}{L^\star}+\frac{D_n^\star\,C_k^{\star 2}}{L^\star}\right]\\ &= \ell\,\frac{\bar{B}\,\left(1+\bar{B}\right)}{L^\star}\,\sum_{k=2}^{K} C_k^\star\left[\frac{D_n^\star\,C_k^\star}{L^\star}+1+C_k^\star\right].\end{aligned}$$

Now consider that

$$\begin{aligned} \mathrm{IV}_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right) &= \ell\,\frac{(1+T_n^\star)}{D_n^\star}\sum_{k=2}^{K} Y_{nk}^{\mathrm{CAPM}}\sum_{m=1}^{N} Y_{m,k}^{\mathrm{CAPM}}\,B_m^\star \\ &\approx \ell\,\frac{\bar{B}\,\left(1+\bar{B}\right)}{L^\star}\sum_{k=2}^{K} C_k^{\star^2}, \end{aligned}$$

whence

$$\mathbb{E}\left[\mathrm{IV}_n\left(\mathbf{Y}\right)\right] \approx \mathrm{IV}_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right) + \ell\,\frac{\bar{B}\,\left(1+\bar{B}\right)}{L^\star}\sum_{k=2}^{K} C_k^\star\left[\frac{D_n^\star\,C_k^\star}{L^\star}+1\right].$$

Therefore

$$\frac{\mathbb{E}\left[\mathrm{IV}_n\left(\mathbf{Y}\right)\right]-\mathrm{IV}_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right)}{\mathrm{IV}_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right)} = \frac{\sum_{k=2}^{K} C_k^\star\left[\frac{D_n^\star\,C_k^\star}{L^\star}+1\right]}{\sum_{k=2}^{K} C_k^{\star^2}}.$$

Assume, again, for simplicity that $C_k \approx \frac{L}{K}$, thus

$$\frac{\mathbb{E}\left[\mathrm{IV}_n\left(\mathbf{Y}\right)\right]-\mathrm{IV}_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right)}{\mathrm{IV}_n\left(\mathbf{Y}^{\mathrm{CAPM}}\right)} \approx \frac{D_n^\star+K}{L^\star} \approx \frac{D_n^\star}{L^\star} \ll 1$$

Chapter 4

Models for Temporal Networks

Synopsis of the Chapter In this Chapter we give an overview of the literature on models for temporal networks. We start with a discussion of models for sequences of binary snapshots networks, with particular attention to temporal extensions of the ERGM models class. In the last part of the Chapter we review the literature on models for temporal weighted networks.

4.1 Definitions

Often systems that are fruitfully described as networks evolve in time. When pairwise interactions change over time, one usually speaks of temporal networks (Soramäki, Wetherilt, and Zimmerman, 2010; Holme and Saramäki, 2012; Craig and Von Peter, 2014). In such time-varying setting, nodes can appear and disappear at any point in time and links can last for a finite interval of time, a quantity usually referred to as duration, or be instantaneous. To represent the most general temporal network, without loss of information, each link should have a start time and an end time, and various formalism have been proposed in the literature to describe dynamical pairwise interaction (Casteigts, Flocchini, Quattrociocchi, and Santoro, 2012; Rossetti, 2015; Latapy, Viard, and Magnien, 2018). Probably the most widespread way of representing temporal networks is one that simplifies the temporal description by considering sequences of networks in discrete time. That amounts to consider a temporal partition of the sequence of links into a sequence of snapshot networks. In this thesis, we focus on the latter description of temporal networks. Hence for the rest of this work we will consider a temporal network as a sequence of networks, in discrete time, each one associated with an adjacency matrix and observed at T different points in time $\left\{ Y_{ij}^{(t)} \right\}_{t=1}^{T}$.

In general, an econometric model for temporal, possibly weighted, networks can be described in terms of a probability distribution $P\left(Y_{ij}^{(t)} | \mathbf{Y}^{(t-1)}, \dots \mathbf{Y}^{(1)}, \mathbf{X}_1^{(t)}, \dots, \mathbf{X}_K^{(t)}\right)$ that describes the probability of each link $Y_{ij}^{(t)}$ to be observed, and the distribution of its weight, as potentially depending on previous realizations of the network and a set of, matrix valued, external variables $\mathbf{X}_1 \dots \mathbf{X}_K$. The external variables can be different for each link, and X_{ij} would be the value of the external variable associated with link ij.

One of the main topics of this thesis is the proposal of new innovative econometric models to handle binary and weighted temporal networks. Nevertheless, we are by no means the first to approach this subject. In the following we review the main models for temporal networks available in the literature. In doing so, we start with models for binary temporal networks, focusing on the temporal extensions of ERGMs, reviewed in Section 2.2, as the original contributions presented in Chapters

6, 7 and 8 are all related to ERGMs and their temporal extensions. We then conclude this Chapter with an overview of the, smaller, literature on models for weighted temporal networks.

4.2 Models for Binary Temporal Networks

The literature on models for temporal binary networks is quite large but many contributions are related to the ERGM framework, that we reviewed in Section 2.2. Hence, we start this literature review from models that extend, in various ways, the ERGM family for the description of temporal networks. These extensions are meant to describe only binary temporal networks as link weights are not modelled in ERGMs.

4.2.1 TERGM

The most well known temporal extension of ERGM is probably the one known as Temporal Exponential Random Graph Model (TERGM). This approach was pioneered by Robins and Pattison, 2001 and subsequently discussed in detail in Hanneke, Fu, Xing, et al., 2010 and Cranmer and Desmarais, 2011. This approach builds on the ERGM, but allows the network statistics defining the probability at time t to depend on current and previous networks up to time $t - K$. This K-step Markov assumption is a defining feature of the TERGMs. Considering only lag 1 dependencies, a TERGM is defined by the following probability mass function

$$P(\mathbf{A}^{(t)}|\mathbf{A}^{(t-1)},\theta) = \frac{e^{\sum \theta_l q_l(\mathbf{A}^{(t)},\mathbf{A}^{(t-1)})}}{K(\theta)}, \tag{4.1}$$

where the functions $q_l(\mathbf{A}^{(t)},\mathbf{A}^{(t-1)})$, called network statistics[1], are defined to investigate the determinants of the network's dynamics, and $K(\theta)$ is a normalization coefficient, known as partition function. Examples of network's statistics, among those explored in the original work of Hanneke, Fu, Xing, et al., 2010, are $q_{stab}(\mathbf{A}^{(t)},\mathbf{A}^{(t-1)}) = \sum_{ij} A_{ij}^{(t)} A_{ij}^{(t-1)} + (1 - A_{ij}^{(t)})(1 - A_{ij}^{(t-1)})$, that captures links' stability, and $q_{dens}(\mathbf{A}^{(t)}) = \sum_{ij} A_{ij}^{(t)}$, related to current network's density.

In a related but different approach Krivitsky and Handcock, 2014 separately model the tie duration and tie formation. They define a class of models called Separable Temporal Exponential Random Graphs (STERGM) that explicitly models the transition of a network between time $t - 1$ and time t by defining two intermediate networks. They refer to them as the *formation network* $\mathbf{A}^{+}$, *consisting of the initial network* $\mathbf{A}^{(t-1)}$ *with ties formed during the time step added and the dissolution network* $\mathbf{A}^{-}$, *consisting of the initial network* $\mathbf{A}^{(t-1)}$ *with ties dissolved during the time step removed.*[2] Then they model the link formation $\mathbf{A}^{+}$ and link dissolution $\mathbf{A}-$ each with a TERGM.

4.2.2 ERGM with Time Varying Parameters

A more recent approach to define temporal extensions of ERGM consists in allowing for its parameters to be time-varying. A notable example of this approach is

[1] They can be defined as functions of the current and lagged adjacency matrices, $\mathbf{Y}^{(t)}$ and $\mathbf{Y}^{(t-1)}$, and can depend on both or only one of them, as in the examples presented in the following.

[2] Definition from Krivitsky and Handcock, 2014, and adapted to our notation.

the Varying-Coefficient-ERGM (VCERGM) proposed in Lee, Li, and Wilson, 2020. There, the authors combine the varying-coefficient models' formalism (see Fan and Zhang, 2008, for a review) with ERGM to take into account the possibility of the ERGM parameters to be smoothly time-varying. In particular, they specify a collection of basis functions $B_1(t), \ldots, B_q(t)$ for $t = 1, \ldots, T$, and then, assuming

$$P\left(\mathbf{A}^{(t)}|\theta^{(t)}\right) = \frac{e^{\sum_s \theta_s^{(t)} q_s\left(\mathbf{A}^{(t)}\right)}}{\mathcal{K}\left(\theta^{(t)}\right)} \tag{4.2}$$

they approximate $\theta^{(t)}$ by a linear combination of these functions

$$\theta_i^{(t)} = \sum_j \phi_{ij} B_j(t)$$

where ϕ_{ij} denote the basis coefficients to be estimated. They choose as basis the set of B-splines functions and, to infer parameter time-variation at time t need to use all the available observations, including those from future times $t' > t$. Thus, in time series jargon, this approach is a smoother and not a filter. Also, due to its smoother nature, it cannot be used to generate sequences of time-evolving networks.

In a related, but different, approach Mazzarisi, Barucca, Lillo, and Tantari, 2020 focus only on the ERGM specification known as fitness model, that we reviewed in 2.2.1, and consider the possibility of a random evolution of the node specific parameters. They call this model Temporally Generalized Random Graph (TGRG) and formally define it by assuming that, for each time t, the binary adjacency matrix has the following probability distribution

$$P\left(\mathbf{A}^{(t)}|\overleftarrow{\theta}^{(t)}, \overrightarrow{\theta}^{(t)}\right) = \prod_{ij} \frac{e^{A_{ij}^{(t)}\left(\overleftarrow{\theta}_i^{(t)} + \overrightarrow{\theta}_j^{(t)}\right)}}{1 + e^{\left(\overleftarrow{\theta}_i^{(t)} + \overrightarrow{\theta}_j^{(t)}\right)}}$$

and the time varying fitness each follow an auto-regressive process of order one. For example, the in-fitness of node i follows

$$\overleftarrow{\theta}_i^{(t)} = c_{0,i} + c_{1,i}\overleftarrow{\theta}_i^{(t-1)} + \epsilon_i^{(t)}$$

where $\epsilon_i^{(t)} \sim N\left(0, \sigma_i^2\right) \quad \forall i = 1, \ldots, N$, and a similar evolution is assumed for the out-fitness parameters $\overrightarrow{\theta}$. Moreover, they complement such a time varying fitness model with a link copying mechanism, inspired by the discrete auto regressive processes (DAR), to explicitly model the phenomenon of link persistence in temporal networks. The resulting model, that they call DAR-TGRG is then a mixture of a the link copying mechanism (DAR) with a time varying fitness model that allows to quantify how likely is a link to be copied from the previous realization, instead of being drawn according to the TGRG at time t. The model can also be used to filter the time-varying parameters in a very specific ERGM and, following the language of Cox et al., 1981, it belongs to a class of econometric models known as *parameter driven*, that we mention in Chapter 5.

4.2.3 Other Models for Binary Temporal Networks

Various instances of latent space models for networks, described in Section 2.3, have been extended to model binary temporal networks (see, for example, Sarkar and

Moore, 2005; Sewell and Chen, 2015). The underling idea of this stream of literature is to allow the latent positions z, defined in Section 2.3, to evolve in time, as reviewed in (Kim, Lee, Xue, Niu, et al., 2018). For example, Sewell and Chen, 2015 define a *Dynamic latent distance model with popularity and activity effects*, where the adjacency matrix depends on the value of the latent variables at time t parameters that evolve following a random walk, and a set of constant parameters to capture the nodes' activity effects r_i for $i = 1, \dots, N$.

$$P\left(\mathbf{A}^{(t)}|z^{(t)}, \beta_{in}, \beta_{out}, r\right) = \prod_{ij} \frac{e^{A_{ij}^{(t)}\eta_{ij}^{(t)}}}{1 + e^{A_{ij}^{(t)}\eta_{ij}^{(t)}}}, \tag{4.3}$$

where

$$\eta_{ij}^{(t)} = \beta_{in}\left(1 - \frac{|z_i^{(t)} - z_j^{(t)}|}{r_i}\right) + \beta_{out}\left(1 - \frac{|z_i^{(t)} - z_j^{(t)}|}{r_j}\right), \tag{4.4}$$

and the $z_i^{(t)}$ parameters each follow an independent random walk.

Finally, it is important to mention that frameworks alternative to latent space models and temporal extension of ERGM for modeling temporal networks have also been considered in the social science literature. Notable examples are the *Stochastic Actor Oriented model* (SAOM) of Snijders, 1996 and the *Relational Event Model* (REM) of Butts, 2008.

4.3 Models for Temporal Weighted Networks

The literature on models for temporal weighted networks is scarcer than that for binary case. This is most certainly due to the additional model complexity required to properly describe the weights. Moreover, as mentioned in Chapter 2 most real world networks are found to have an abundance of zero entries that, in the weighted case, require some care in modelling the probability for a link to be present, $P\left(A_{ij} = 1\right) = P\left(Y_{ij} > 0\right)$, and its expected weight $\mathbb{E}\left[Y_{ij}\right]$.

For example, Giraitis, Kapetanios, Wetherilt, and Žikeš, 2016 use Tobit models for temporal weighted networks with an abundance of zero entries. They associate to the weighted adjacency matrix $\mathbf{Y}$ a latent matrix $\mathbf{Y}^\star$ by

$$Y_{ij} = \begin{cases} Y_{ij}{}^\star, & if \quad Y_{ij}{}^\star > 0 \\ \\ 0, & if \quad Y_{ij}{}^\star \leq 0, \end{cases}$$

and model $\mathbf{Y}^\star$ in a standard way as a continuous variable. In practice, they select a set of network statistics $q_i\left(\mathbf{Y}\right)$ and, for each link, estimate the dependency $Y_{ij}^{(t)\star}$ on $q\left(\mathbf{Y}^{(t-1)}\right)$. With their choice of network statistics, they estimate a separate Tobit model for each pair of links. Finally, they use a local-likelihood method to estimate the time varying coefficients of the regression. The censored regression used in this paper has the downside issue of requiring a joint modeling of the presence of a link and of its weight. While this choice allows for straightforward estimates and builds on the well known Tobit regression, it models jointly the effect that a covariate has on the probability of observing a link and the one that it has on its expected weight.

Latent space models have also been considered for sparse weighted networks, with the aim of inferring from the weights dynamics the time varying positions of nodes in a latent space. For example Sewell and Chen, 2016, consider an extended version of their binary model, presented in the previous section, by applying a different link functions to (4.4) that allows them to handle count and positive continuous weights. In the case of positive continuous weights, they combine their approach with a Tobit model, in order to be able to model sparse networks.

Models that allow each one of the matrix elements $Y_{ij}^{(t)}$ to depend on each of the ${Y_{ij}}^{(t-1)}$ have also been considered in the literature. For example, Billio, Casarin, and Iacopini, 2018 estimate a tensor regression (very similar to a VAR on $vec(\mathbf{Y})$), with rank restrictions on the (huge) matrix of model's parameters. Differently from the previously mentioned papers, they do not take the sparse nature of networks explicitly into account.

Chapter 5

Score Driven Econometric Models

Synopsis of the Chapter In this chapter we review score driven econometric models. We start by briefly discussing time varying parameter models in general. Then we give the formal definition of score driven models and discuss their application as misspecified filters with one explicit example. We discuss a theoretical motivation for using the score in the update rule for the parameters. Finally, we conclude discussing how to quantify uncertainty and test for temporal heterogeneity in the score driven framework.

5.1 Introduction

Let us set the stage to better explain score-driven models by briefly reviewing the theory of time-varying parameters models in discrete time. There is a rich literature on the topic, which has been summarized in the review by Tucci, 1995 and more recently by Koopman, Lucas, and Scharth, 2016. In general, a time-varying parameters model can be written as

$$s^{(t)} \sim P(s^{(t)}|f^{(t)}, \mathcal{S}^{(t-1)}, \Phi_1) \tag{5.1a}$$

$$f^{(t)} = \psi(f^{(t-1)}, f^{(t-2)}, ..., \mathcal{S}^{(t-1)}, \epsilon^{(t)}, \Phi_2) \tag{5.1b}$$

where $s^{(t)}$ is a vector of observations sampled from the probability distribution function p, $\mathcal{S}^{(t-1)}$ is the set of all observations up to time $t-1$ and $f^{(t)}$ are the parameters which are assumed to be time varying. The dynamics of these parameters can either depend on past observations, on past values of the same parameters, on some external noise $\epsilon^{(t)}$ and on two sets of static parameters Φ_1 and Φ_2.

If the function ψ only contains past values of the time-varying parameters, a noise term and the static parameters, then the model is called a *parameter-driven* model, whereas if the function ψ can be written as a deterministic function only of past observations and past parameters, it is called an *observation-driven* model Cox et al., 1981.

Examples for parameter-driven models can be found in the financial econometrics literature looking at stochastic volatility models Tauchen and Pitts, 1983; Shephard, 2005, which aim at describing the time-varying nature of the volatility (*i.e.* the variance) of price variations, as well as other examples Bauwens and Veredas, 2004; Hafner and Manner, 2012.

Within the observation-driven models, the most celebrated example is the Generalized Auto-Regressive Conditional Heteroscedasticity (GARCH) model Bollerslev, 1986b, where a time series of financial log-returns is modelled using a time-varying volatility parameter depending deterministically on squared observations up to that time and past values of volatility.

One of the main topics of this thesis is the proposal of new observation driven models for temporal networks. The main advantage of adopting an observation-driven model rather than a parameter-driven one lies in its estimation: having time-varying parameters that only depend on observations through a set of static parameters results in a strong reduction of complexity in writing the likelihood of the model, whereas the calculations for most non-trivial parameter-driven models are typically extremely convoluted and computationally intensive.

In this thesis, we focus on the *Score Driven* approach for the definition of observation driven models that has received much attention in the recent econometric literature. Several are the reasons of the flexibility of a score-driven approach and of its success in time-series modeling, in the rest of this Chapter we precisely define them, discuss in details a well known example, and review features and methods from the score driven framework that motivated their widespread diffusion.

5.2 Definition

Score-driven models, also known as Generalized auto-regressive Score (GAS) or Dynamic Conditional Score (DCS) models, are a specific class of observation-driven models. Originally introduced by Creal, Koopman, and Lucas, 2013 and Harvey, 2013b, they postulate that time-varying parameters depend on observations through the score of the conditional likelihood, i.e. the gradient of its logarithm.

Specifically, let us consider a sequence of observations $\left\{y^{(t)}\right\}_{t=1}^{T}$, where each $y^{(t)} \in \mathbb{R}^M$, and a conditional probability density $P\left(y^{(t)}|f^{(t)}\right)$, that depends on a vector of time-varying parameters $f^{(t)} \in \mathbb{R}^K$. Defining the score as $\nabla^{(t)} = \frac{\partial \log P\left(y^{(t)}|f^{(t)}\right)}{\partial f^{(t)\prime}}$, a score-driven model assumes that the time evolution of $f^{(t)}$ is ruled by the recursive relation

$$f^{(t+1)} = w + \boldsymbol{b} f^{(t)} + \boldsymbol{a}\boldsymbol{\mathcal{I}}^{(t)}\nabla^{(t)}, \tag{5.2}$$

where w, $\boldsymbol{a}$ and $\boldsymbol{b}$ are static parameters, w being a K dimensional vector and $\boldsymbol{a}$ and $\boldsymbol{b}$ $K \times K$ matrices. $\boldsymbol{\mathcal{I}}^{(t)}$ is a $K \times K$ scaling matrix, that is often chosen to be the inverse of the square root of the Fisher information matrix associated with $P\left(y^{(t)}|f^{(t)}\right)$, i.e. $\boldsymbol{\mathcal{I}}^{(t)} = \mathbb{E}\left[\frac{\partial \log P\left(y^{(t)}|f^{(t)}\right)}{\partial f^{(t)\prime}} \frac{\partial \log P\left(y^{(t)}|f^{(t)}\right)}{\partial f^{(t)\prime}}'\right]^{-\frac{1}{2}}$. However, this is not the only possible specification and different choices for the scaling are discussed in Creal, Koopman, and Lucas, 2013.

A second look at eq. (5.2) reveals to the reader the similarity of the score-driven recursion with the iterative step from a Newton algorithm, whose objective function is precisely the log-likelihood function. Indeed, at each step the score pushes the parameter vector along the log-likelihood steepest direction. After scaling with the matrix $\boldsymbol{\mathcal{I}}$, the intensity of the push is modulated by the parameter $\boldsymbol{a}$, and its direction adjusted by the auto-regressive component.

A score driven model can be regarded both as *data generating process* (DGP) or as a filter. In both cases, the most important feature of (5.2) is the role of the score as a driver of the dynamics of $f^{(t)}$. The structure of the conditional observation density determines the score, from which the dependence of $f^{(t+1)}$ on the vector of observations $y^{(t)}$ follows. When the model is viewed as a DGP, the update results in a stochastic dynamics exactly thanks to the random occurrence of $y^{(t)}$. When the score-driven recursion is regarded as a filter, the update rule in (5.2) is used to obtain

a sequence of filtered $\left\{\hat{f}^{(t)}\right\}_{t=1}^{T}$. In this setting, the static parameters are estimated maximizing the log-likelihood of the whole sequence of observations (for a detailed discussion, see Harvey, 2013a; Blasques, Koopman, and Lucas, 2014),

$$\left(\hat{w}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{a}}\right) = \arg\max_{(w,\boldsymbol{b},\boldsymbol{a})} \sum_{t=1}^{T} \log P\left(y^{(t)} | f^{(t)}\left(w, \boldsymbol{b}, \boldsymbol{a}, \left\{y^{(t')}\right\}_{t'=1}^{t-1}\right)\right). \tag{5.3}$$

In the univariate case, Blasques, Koopman, and Lucas, 2014 work out the required regularity conditions ensuring the consistency and asymptotic normality for the maximum likelihood estimators of the parameter values.

5.3 An Example of Score-Driven Models as Misspecified Filters

Here we present a simple example of how score-driven models can be used to filter an unknown dynamics of a parameter, without assuming a specific model for its time evolution. We will do it by considering the classical case of a discrete time random walk model with time-varying diffusion coefficient. This type of models is very popular in finance where the (logarithm of the) price follows a random walk and the diffusion rate, termed volatility, represents the risk of the asset. As we will show, under minimal assumptions, such a filter turns out to coincide with the popular GARCH model for volatility.

All this is relatively well known. Indeed, the interpretation of GARCH processes as predictive filters is well described in this statement by Nelson, 1992: "Note that our use of the term 'estimate' corresponds to its use in the filtering literature rather than the statistics literature; that is, an ARCH model with (given) fixed parameters produces 'estimates' of the true underlying conditional covariance matrix at each point in time in the same sense that a Kalman filter produces 'estimates' of unobserved state variables in a linear system".

Let us call $s^{(t)}$ the increment of the log-price, $p(t+1) - p^{(t)}$ and consider a stochastic volatility model

$$s^{(t)} = \sigma^{(t)} \epsilon^{(t)} \qquad \epsilon^{(t)} \sim \mathcal{N}(0,1).$$

i.e. the conditional probability density function of $s^{(t)}$ is

$$p(s^{(t)} | \sigma^{(t)}) = \frac{1}{\sqrt{2\pi\sigma^{2(t)}}} e^{-\frac{s^2 W^{(t)}}{2\sigma^{2(t)}}}$$

By choosing as time-varying parameter $f^{(t)} = \sigma^{2(t)}$, the score of the likelihood is

$$\frac{\partial \log p(s^{(t)} | f^{(t)})}{\partial f^{(t)}} = -\frac{1}{2\sigma^{2(t)}} + \frac{s^{2(t)}}{2\sigma^{4(t)}}$$

hence the equation for the evolution of volatility is

$$\sigma^2(t+1) = w + b\sigma^{2(t)} + \frac{a\mathcal{I}^{-1/2(t)}}{2}\left[\frac{s^{2(t)} - \sigma^{2(t)}}{\sigma^{4(t)}}\right]$$

Thus if $s^{2(t)} \gg \sigma^{2(t)}$ ($s^{2(t)} \ll \sigma^{2(t)}$), the new $\sigma^2(t+1)$ will be larger (smaller) than $\sigma^{2(t)}$. This is exactly the mechanism which dynamically adjusts the filtered estimation of volatility taking into account the most recent observation(s).

By choosing $\boldsymbol{\mathcal{I}}^{(t)}$ as the the Fisher information matrix and using $\mathbb{E}[s^{2(t)}|\sigma^{2(t)}] = \sigma^{2(t)}$, it is

$$\mathcal{I}^{(t)} \equiv -\mathbb{E}\left[\frac{\partial^2 \log p(s^{(t)}|\sigma^{(t)})}{\partial^2 \sigma^{2(t)}}\middle|\sigma^{2(t)}\right] = -\mathbb{E}\left[\frac{1}{2\sigma^{4(t)}} - \frac{s^{2(t)}}{\sigma^{6(t)}}\middle|\sigma^{2(t)}\right] = \frac{1}{2\sigma^{4(t)}}$$

thus

$$\sigma^2(t+1) = w + b\sigma^{2(t)} + a(s^{2(t)} - \sigma^{2(t)}) = w + \alpha s^{2(t)} + \beta\sigma^{2(t)} \tag{5.4}$$

with $\alpha = a$ and $\beta = b - a$, which coincides with the GARCH model. This model has been originally proposed as a data generating process for describing realistic dynamics of volatility, while here it is derived as a result of Score-Driven modeling. The GARCH model of Eq. 5.4 is typically seen as a data generating process for the volatility, and thus the price, of financial assets. This model is routinely estimated from real data and used widely in the financial industry for risk management, portfolio allocation, systemic risk, etc.. Fig. 5.1 shows a typical simulated price pattern from a GARCH(1,1) process, displaying fat tails and clustered volatility, as observed in empirical data. It turns out that other well known models in econometrics can be expressed as score-driven models. Famous examples are the Exponential GARCH model of Nelson, 1991, the Autoregressive Conditional Duration (ACD) model of Engle and Russell, 1998, and the Multiplicative Error Model (MEM) of Engle, 2002. The introduction of this framework in its full generality opened the way to applications in various contexts.

However, the main point we want to make here concerns the use of GARCH, and more generally of score-driven models, as *filters* of a differently specified dynamics. To show this in practice, we simulate 1000 price observations from the model

$$s^{(t)} = \sigma^{(t)}\epsilon^{(t)} \quad \epsilon^{(t)} \sim \mathcal{N}(0,1) \quad \sigma^{(t)} = 2 + \frac{1}{2}\sin\left(\frac{\pi t}{100}\right) \tag{5.5}$$

The top panel of Fig. 5.2 shows the simulated price dynamics. This is clearly *not* a GARCH model and the sinuisodal shape can be modified with other deterministic or stochastic processes. Assuming the data generating process of Eq. 5.5 is unknown, one can nevertheless fit the GARCH(1,1) model and obtain, beside the static parameters w, a, and b, the filtered values of $\sigma^{(t)}$. The outcome of this procedure is shown in the bottom panel of Fig. 5.2 where the red line is the simulated $\sigma^{(t)}$, while the black circles represent the filtered values of $\sigma^{(t)}$.

The example shows how score-driven models can be used to filter the time-varying parameters with unknown dynamics from data. As mentioned in the main text, Score-driven models have been shown to be an optimal choice among observation-driven models when minimising the Kullback-Leibler divergence to an unknown generating probability distribution Blasques, Koopman, and Lucas, 2015.

5.4 Information Theoretical Optimality

There are motivations, originating in information theory, for the optimality of the score-driven updating rule. In Blasques, Koopman, and Lucas, 2015, the authors

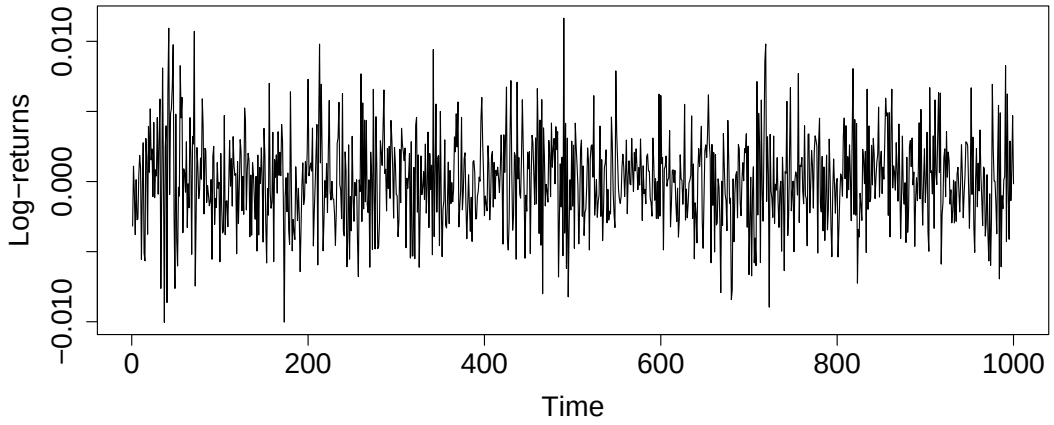

FIGURE 5.1: Artificially generated time series of length 1000 from a GARCH(1,1) model with $w = 10^{-6}$, $\alpha = 0.1$, $\beta = 0.8$. Figure from Campajola, Di Gangi, Lillo, and Tantari, 2020.

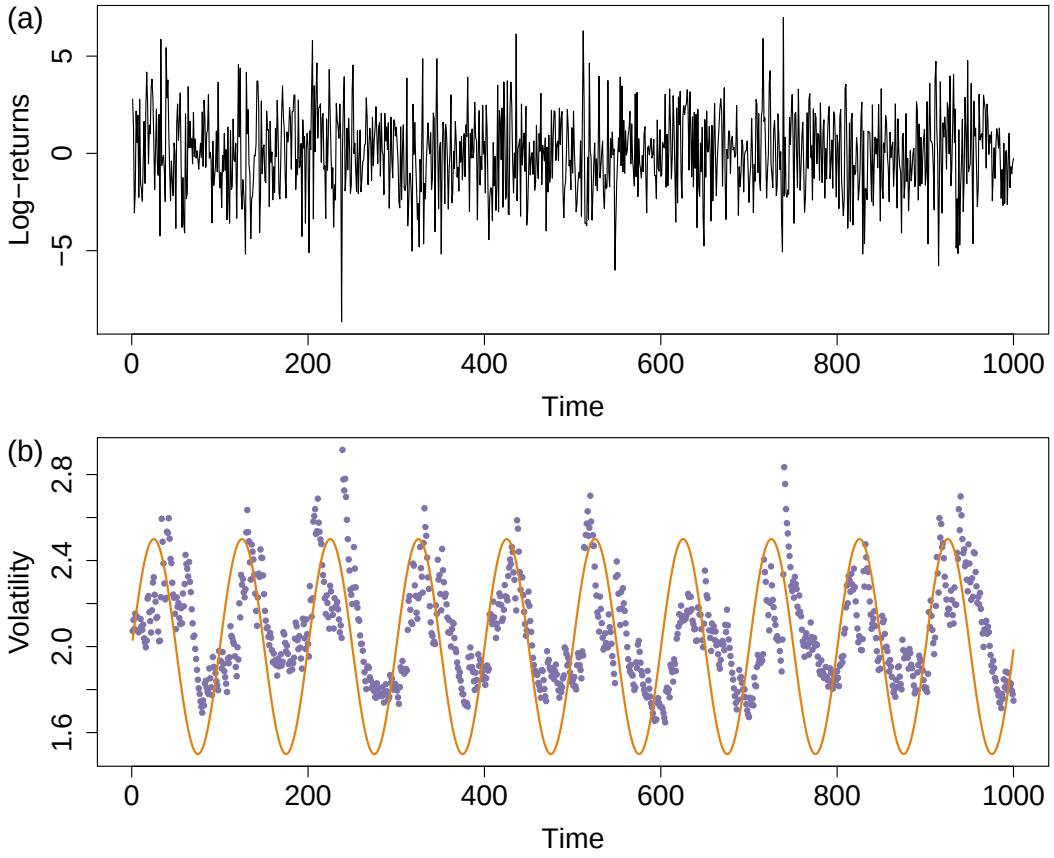

FIGURE 5.2: (a) Artificially generated time series of returns according to the model of Eq. 5.5. (b) Simulated (orange line) and filtered (purple dots) values of $\sigma^{(t)}$. The latter are obtained by fitting a GARCH(1,1) model on the data in the top panel. Figures from Campajola, Di Gangi, Lillo, and Tantari, 2020.

consider a true and unobserved DGP $y^{(t)} \sim P\left(y^{(t)}|f^{(t)}\right)$. They assume a given and in general mispecified conditional observation density $\tilde{P}^{(t)} = \tilde{P}\left(\,.\,|\tilde{f}^{(t)}\right)$, and consider the Kullback-Leibler (K-L) divergence

$$\mathcal{D}_{\mathcal{KL}}\left(P^{(t)}, \tilde{P}^{(t+1)}\right) = \int_A P\left(y|f^{(t)}\right) \log \frac{P\left(y|f^{(t)}\right)}{\tilde{P}\left(y|\tilde{f}^{(t+1)}\right)}\, dy,$$

where $A \subseteq \mathbb{R}$. Building on the minimum discrimination information principle (Kullback, 1997), they argue that, when the new observation y_t becomes available, $\tilde{f}^{(t+1)}$ should ideally be such that the updated density $\tilde{P}^{(t+1)}$ is as close as possible to the true density $P^{(t)}$. Given that the real DGP is not known, an optimal update that minimizes $\mathcal{D}_{\mathcal{KL}}$ cannot be defined in practice. For this reason, Blasques, Koopman, and Lucas, 2015 focus on the improvements of $\mathcal{D}_{\mathcal{KL}}$ that an updating step produces irrespectively of the true DGP. One way of quantifying the improvement for a parameter update from $\tilde{f}^{(t)}$ to $\tilde{f}^{(t+1)}$ is to consider the realized variation of $\mathcal{D}_{\mathcal{KL}}$

$$\Delta_{t|t} \equiv \mathcal{D}_{\mathcal{KL}}\left(P^{(t)}, \tilde{P}^{(t+1)}\right) - \mathcal{D}_{\mathcal{KL}}\left(P^{(t)}, \tilde{P}^{(t)}\right) = \int_A P\left(y|f^{(t)}\right) \log \frac{\tilde{P}\left(y|\tilde{f}^{(t)}\right)}{\tilde{P}\left(y|\tilde{f}^{(t+1)}\right)}\, dy\,. \tag{5.6}$$

Based on this definition, a parameter update is realized K-L optimal when $\Delta_{t|t} < 0$ for every $\left(y^{(t)}, \tilde{f}^{(t)}, f^{(t)}\right)$. The authors prove that, under reasonable assumptions, the updating rule (5.2) based on the score of $\tilde{P}^{(t+1)}$ is locally realized K-L optimal. For more details, and alternative definitions of optimality, we direct the reader to the original work and the more recent Blasques, Lucas, and Vlodrop, 2020.

5.5 Confidence Bands

Any filtering tool should provide an estimation of the uncertainty and confidence bands for the estimates. Blasques, Koopman, Łasak, and Lucas, 2016 discussed methods to quantify the uncertainty associated with the score driven filters, when the DGP is itself a score driven model. Specifically, they proposed a simulation based method to define in sample confidence bands around the filtered time varying parameters. Their procedure starts from the maximum likelihood estimate of the static parameters, given observations $\left\{y^{(t')}\right\}_{t'=1}^{t-1}$, as defined in (5.3). Given the MLE estimate, the method prescribes to repeatedly sample new parameters $(w, \boldsymbol{b}_i, \boldsymbol{a})_i$ from a multivariate normal, centered around the MLE estimates, and variance-covariance matrix estimated with the Huber-White estimator (Huber et al., 1967; White, 1980). Then one uses each sample to filter a different sequence of time varying parameters, from the same time series of observations, thus obtaining a sample of filtered paths $\hat{f}_i^{(t)} = \hat{f}_i^{(t)}\left(w_i, \boldsymbol{b}_i, \boldsymbol{a}_i, \left\{y^{(t')}\right\}_{t'=1}^{t-1}\right)$ for $i = 1, \ldots, K$, where K is the number of samples. Finally, for each time t one uses the obtained distribution $\hat{f}_i^{(t)}$ to calculate the appropriate percentiles defining the confidence bands. While this construction is intuitive and easy to implement in practice, it is meant to capture only the uncertainty due to the estimation of the static parameters, often referred to as *parameter uncertainty*. Hence the confidence bands reliably quantify uncertainty only when the DGP is score driven. In other words, these bands do not take into account what is known

as *filtering uncertainty*. This is the uncertainty due to the fact that in general we do not know the true DGP and the score-driven filter may be regarded only as an approximate filter. Recently Buccheri, Bormetti, Corsi, and Lillo, 2018 investigated the approximation error made by applying a score driven filter to a time varying parameter model following a different DGP. They found that, for a class of DGPs where the parameters follow an auto-regressive process, the approximation becomes exact in the limit of small variance of the latent parameters. Moreover they proposed a method to define confidence bands, inspired by Hamilton, 1986, that accounts for both filtering and parameter uncertainty in Score Driven filters. While we refer to their paper for the details of the derivation, here we briefly describe the key steps of the procedure. The total conditional variance of the latent parameters is decomposed as the sum of two terms. One term captures the parameter uncertainty similarly to the approach of Blasques, Koopman, Łasak, and Lucas, 2016. The other term captures the filtering uncertainty and can be written in terms of the static parameters $(w, \boldsymbol{b}, \boldsymbol{a})$ and the scaling matrix $\boldsymbol{\mathcal{I}}^{(t)}$ from (5.2) as $P^{(t)} = \boldsymbol{b}^{-1}\boldsymbol{a}\boldsymbol{\mathcal{I}}^{(t)}$. In practice, the procedure consists in sampling $(w, \boldsymbol{b}, \boldsymbol{a})_i$ and obtaining a distribution of filtered paths, as in Blasques, Koopman, Łasak, and Lucas, 2016. Then for each time step t the variance of the latent parameters is obtained as $\frac{1}{K}\sum_i \left(\hat{f}_i^{(t)} - \hat{f}^{(t)}\right)^2 + \frac{1}{K}\sum_i \boldsymbol{b}_i^{-1}\boldsymbol{a}_i S_i^{(t)}$, where $\hat{f}^{(t)}$ is the path filtered using the maximum likelihood estimates from (5.3).

5.6 Test for Temporal Heterogeneity

As a final aspect, score-driven models allow for a test discriminating whether the observations are better described by a model with time-varying parameters or static ones. In fact, following Engle, 1982, Calvori, Creal, Koopman, and Lucas, 2017 discuss, and extensively evaluate, the performances of a test for parameter temporal variation tailored for score-driven models. For a detailed description of the test, we refer the reader to Calvori, Creal, Koopman, and Lucas, 2017. Here, we shortly review the main idea. The method consists in a Lagrange Multiplier (LM) test for the parameter a that multiplies the score in the one dimensional version of the recursion (5.2). The null hypothesis H_0 is that the parameter $f^{(t)}$ is actually static, i.e. $\beta = \alpha = 0$, and it corresponds to w. As explained in Davidson, MacKinnon, et al., 2004, the LM statistic for the hypothesis H_0, versus the alternative $a = b \neq 0$ can be conveniently obtained from an auxiliary regression. To allow for a coefficient b different from a, one can use the same arguments as in Lee, 1991. As discussed in Calvori, Creal, Koopman, and Lucas, 2017, the LM statistic can be written as the explained sum of squares from the regression

$$\mathbf{1} = c_w \nabla_w^{(t)} + c_a \boldsymbol{\mathcal{I}}^{(t-1)} \nabla_w^{(t-1)} \nabla_w^{(t)\prime} + \text{residual},$$

where c_w and c_a are regression coefficients that can be estimated with any statistical software. It is worth noticing that, under the null, the score of the conditional density with respect to $f^{(t)}$ is equal to the score with respect to w. From standard asymptotic theory, it follows that the LM statistic is distributed as a χ^2 with one degree of freedom.

Chapter 6

Score Driven Exponential Random Graphs

Synopsis of the Chapter Motivated by the increasing abundance of data describing real-world networks that exhibit dynamical features, we propose an extension of the ERGMs that accommodates the time variation of its parameters. Inspired by the fast growing literature on Dynamic Conditional Score-driven models each parameter evolves according to an updating rule driven by the score of the ERGM distribution. We demonstrate the flexibility of the score-driven ERGMs (SD-ERGMs), both as data generating processes and as filters, and we show the advantages of the dynamic version with respect to the static one. We discuss two applications to time-varying networks from financial and political systems. First, we consider the prediction of future links in the Italian interbank credit network. Second, we show that the SD-ERGM allows to discriminate between static or time-varying parameters when used to model the dynamics of the US congress co-voting network.

Almost all the contents of this chapter previously appeared in Di Gangi, Bormetti, and Lillo, 2019.

6.1 Introduction

In this Chapter, we present an original approach to time-varying networks that is based on two main ingredients: (i) a parametric probabilistic model, according to which one can sample a network realization, i.e. an adjacency matrix; (ii) a simple mechanism to introduce time-variation on the model parameters and, consequently, to induce a dynamics on the network sequence. Concerning the former point, a natural choice is the class of statistical models for networks, known as ERGMs. As far as point (ii) is concerned, a flexible candidate is suggested by the fast growing literature on the Dynamic Conditional Score-driven models. The goal of this paper is to present a new class of models for temporal networks and to provide evidence that the novel approach is versatile and effective in capturing time-varying features. To the best of our knowledge, this is the first time the two frameworks are combined to provide a dynamic description of networks.

Our main contribution is an extension of the ERGM framework that allows model parameters to change over time in a score-driven fashion. The result of our efforts is a class of models for time-varying networks where the information encoded in $\mathcal{F}_{t-1}$ is exploited to filter the time-varying parameters $\theta^{(t)}$ at time t. We refer to this class as *Score-Driven Exponential Random Graph Models*. At this point, it is worth to comment that a generic SD-ERGM can be also used to generate synthetic sequences

of graphs, i.e. it can be considered as a DGP. However, we are more inclined to interpret it as an effective filter of latent time-varying parameters, regardless of what the true DGP might be.

The rest of the Chapter is organized as follows. In Section 6.2 we introduce the new class of models and validate it with extensive numerical experiments for three specific instances of the SD-ERGM. Section 6.3 presents the results from an application to two real temporal networks: The e-Mid interbank network for liquidity supply and demand and the U.S. Congress co-voting political network. Section 6.4 draws the relevant conclusions.

6.2 Score-Driven Exponential Random Graphs

In this section, we introduce the general SD-ERGM framework, discuss in detail the applicability of the score-driven approach to three different ERGMs, and validate their performances with extensive numerical simulations.

We propose to apply the score-driven methodology, presented in Chapter 5 to the ERGMs, presented in Chapter 2. This combination will allow any of the parameters θ_s in (2.5) to have a stochastic evolution driven by the score of the static ERGM model, computed at different points in time. This approach results in a framework for the description of time-varying networks, more than in a single model, in very much the same way as ERGM is considered a modeling framework for static networks. We refer to such class of models as Score-Driven Exponential Random Graphs Models.

Conceptually, applying the score-driven approach is fairly straightforward. Given the observations $\left\{Y_{ij}^{(t)}\right\}_{t=1}^{T}$, we can apply the update rule in (5.2) to all or some elements of θ, each of which is associated with a network statistic in (2.5). In order to do this, we need to compute the derivative of the log-likelihood at every time step, i.e. for each adjacency matrix $\mathbf{Y}^{(t)}$. For the general ERGM, the elements of the score take the form

$$\nabla_s^{(t)}(\theta) = \frac{\partial \log P\left(\mathbf{Y}^{(t)}|\theta\right)}{\partial \theta_s} = h_s\left(\mathbf{Y}^{(t)}\right) - \frac{\partial \log \mathcal{K}(\theta)}{\partial \theta_s}.$$

It follows that the vector of time-varying parameters evolves according to

$$\theta^{(t+1)} = w + \boldsymbol{b}\theta^{(t)} + \boldsymbol{a}\boldsymbol{\mathcal{I}}^{(\boldsymbol{t})}\nabla^{(t)}\left(\theta^{(t)}\right). \tag{6.1}$$

Hence, conditionally on the value of the parameters $\theta^{(t)}$ at time t and the observed adjacency matrix $\mathbf{Y}^{(t)}$, the parameters at time $t+1$ are deterministic. When used as a DGP, the SD-ERGM describes a stochastic dynamics because, at each time t, the adjacency matrix is not known in advance. It is randomly sampled from $P\left(\mathbf{Y}^{(t)}|\theta^{(t)}\right)$ and then used to compute the score that, as a consequence, becomes itself stochastic. When the sequence of networks $\left\{\mathbf{Y}^{(t)}\right\}_{t=1}^{T}$ is observed, the static parameters $(w, \boldsymbol{b}, \boldsymbol{a})$, that best fit the data, can be estimated maximizing the log-likelihood of the whole time series. Taking into account that each network $\mathbf{Y}^{(t)}$ is independent from all the others *conditionally* on the value of $\theta^{(t)}$, the log-likelihood can be written as

$$\log P\left(\left\{\mathbf{Y}^{(t)}\right\}_{t=1}^{T}|w,b,a\right) = \sum_{t=1}^{T}\log P\left(\mathbf{Y}^{(t)}|\theta^{(t)}\left(w,b,a,\left\{\mathbf{Y}^{(t')}\right\}_{t'=1}^{t-1}\right)\right). \tag{6.2}$$

It is evident that the computation of the normalizing factor, and its derivative with respect to the parameters, is essential for the SD-ERGM. Not only it enters the definition of the update, but it is also required for the optimization of (6.2).

Our main motivation for the introduction of SD-ERGM is to describe the time evolution of a sequence of networks by means of the evolution of the parameters of an ERGM. We assume to know, from the context or from previous studies of static networks in terms of ERGM, which statistics are more appropriate in the description of a given network. Hence, we do not discuss the choice of statistics in the context of dynamical networks, but refer the reader to Goodreau, 2007 and Hunter, Goodreau, and Handcock, 2008 for examples of feature selection and Goodness Of Fit (GOF) evaluation, as well as to Shore and Lubin, 2015 for a recent proposal to quantify GOF specifically in network models.

In the rest of this section, we discuss in details the SD extension of ERGMs with given statistics. The first example allows for the exact computation of the likelihood, but the number of parameters can become large for large network. In the second example, we discuss how an SD-ERGM can be defined when the log-likelihood is not known in closed form. Using extensive numerical simulations, we show that SD-ERGMs are very efficient at recovering the paths of time-varying parameters when the DGP is known and the score-driven model is employed as a misspecified filter. Moreover, we show a first application of the LM test in assessing the time-variation of ERGM parameters.

6.2.1 Score Driven Beta Model

Our first specific example is the score-driven version of the *beta model*, introduced in Sec. 2.2.1. We start with this model not only because of its wide applications and relevance in various streams of literature, but also because the likelihood of the ERGM, and its score, can be computed exactly. Moreover, the number of local statistics, the degrees, and parameters can become very large for large networks. Since we need to describe the dynamics of a large amount of parameters, this last feature poses a challenge to any time-varying parameter version of the beta model. At the end of this Section we will show how the SD framework allows for a parsimonious description of such a high dimensional dynamics.

As anticipated, the SD-beta model is defined by the application of (6.1) to each one of the $\overrightarrow{\theta}$ and $\overleftarrow{\theta}$ parameters. Among the possible choices, we use as scaling the diagonal matrix $\mathcal{I}_{ij}^{(t)} = \delta_{ij} I_{ij}^{(t)^{-1/2}}$, where $\boldsymbol{I}^{(t)} = \mathbb{E}[\nabla^{(t)} \nabla^{(t)'}]$, i.e. we scale each element of the score by the square root of its variance. It is very common, in score-driven models with numerous time-varying parameters, to restrict the matrices $\boldsymbol{a}$ and $\boldsymbol{b}$ of (5.2) to be diagonal. In this work, we consider a version of the score update having only three static parameters (w_s, b_s, a_s) for each dynamical parameter θ_s. The resulting update rule for the beta model is

$$\begin{aligned}
\overleftarrow{\theta}_s^{(t+1)} &= w_s^{\text{in}} + b_s^{\text{in}} \overleftarrow{\theta}_s^{(t)} + a_s^{\text{in}} \left(\frac{\sum_i \left(Y_{is}^{(t)} - p_{is}^{(t)} \right)}{\sqrt{\sum_i p_{is}^{(t)} \left(1 - p_{is}^{(t)} \right)}} \right) \\
\overrightarrow{\theta}_s^{(t+1)} &= w_s^{\text{out}} + b_s^{\text{out}} \overrightarrow{\theta}_s^{(t)} + a_s^{\text{out}} \left(\frac{\sum_i \left(Y_{si}^{(t)} - p_{si}^{(t)} \right)}{\sqrt{\sum_i p_{si}^{(t)} \left(1 - p_{si}^{(t)} \right)}} \right),
\end{aligned} \tag{6.3}$$

where the superscripts *in* and *out* indicate the first and second half of the parameter vectors, respectively. In order to simplify the inference procedure, we consider a two-step approach. First, we fix the node specific parameters w_i in order to target the unconditional means of $\overleftarrow{\theta}$ and $\overrightarrow{\theta}$ resulting from an ERGM with static parameters. Conditionally on the target values, we estimate the remaining parameters a^{in}, a^{out}, b^{in}, and b^{out}. We verified that the bias introduced by the two-step procedure is negligible and results remain similar when the joint estimation is performed.

SD-ERGMs as filters: Numerical Simulations

As mentioned in the Introduction, SD-ERGMs (as other observation driven model, e.g. GARCH) can be seen either as DGP and estimated on real time series or as predictive filters (Nelson, 1996), since time-varying parameters follow one-step-ahead predictable processes. In this Section we show the power of the ERGMs in this second setting. Specifically, we simulate generic non-stationary temporal evolution for the parameters $\theta^{(t)}$ of temporal networks. We then use the SD-ERGM to filter the paths of the parameters and evaluate its performances. It is important to note that the simulated dynamics of the parameters is different from the score-driven one used in the estimation.

In practice, at each time t, we sample the adjacency matrix from the PMF of an ERGM with parameters[1] $\bar{\theta}^{(t)}$, evolving according to known temporal patterns, that define different DGPs. We then use the realizations of the sampled adjacency matrices to filter the patterns. We consider a sequence of $T = 250$ time steps for a network of 10 nodes, each with parameters $\overleftarrow{\bar{\theta}}_i^{(t)}$ and $\overrightarrow{\bar{\theta}}_i^{(t)}$ evolving with predetermined patterns. We test four different DGPs. The first one is a naive case with constant parameters $\bar{\theta}^{(t)} = \bar{\theta}_0$. The elements of $\bar{\theta}_0$ are chosen in order to ensure heterogeneity in the expected degrees of the nodes under the static beta model. For the remaining three DGPs, half of the parameters is static and half is time-varying evolving with either a deterministic sinusoidal function, a deterministic step function and a stochastic AR(1) dynamics. More details on the definition of such DGPs are given in 2.2.1.

In the following, we benchmark the performance of the SD-ERGMs with that of a sequence of cross sectional estimates of static ERGMs, i.e. one ERGM estimated for each t. We quantify the performance of the two approaches computing the Root Mean Square Error $\frac{1}{T}\sqrt{\sum_t \left(\bar{\theta}_s^{(t)} - \hat{\theta}_s^{(t)}\right)^2}$, that describes the distance between the known simulated path and the filtered. We then average the RMSE across all the time-varying parameters and 100 simulations, and report the results in Table 6.1. These results confirm that the SD beta model outperforms the standard beta model in recovering the true time-varying pattern. Notably, this holds true even when the DGP is inherently non stationary, as in the case of the DGP where each parameter has a step like evolution. Indeed the results of this Section and of Section 6.2.2 confirm that, while the SD update rule (5.2) defines a stationary DGP (see Creal, Koopman, and Lucas, 2013), using SD models as filters, we can effectively recover non stationary parameters' dynamics.

Our last numerical simulations for the SD beta model explore its applicability and performances for networks of increasing size. We explore this setting for two reasons. The first one is that many real systems are described by networks with a

[1] In the following, the notation with a bar refers to the true parameters used in the DGP.

TABLE 6.1: RMSEs (on a percentage base) of the filtered paths averaged over all time-varying parameters and all Monte Carlo replicas of the numerical experiment. Left column: results from the cross sectional estimates of the beta model; right column: score-driven beta model results. Each row correspond to one of the four DGPs.

DGPs	Average RMSE	
	beta model	SD-beta model
Const	1.75	0.20
Sin	2.76	0.34
Steps	2.46	0.28
AR(1)	1.82	0.24

large number of nodes. The second reason is that we want to compare the performances of our approach with those of the standard beta model in regimes where the latter is known to perform better, under suitable conditions. Indeed, as mentioned in Section 2.2.1, asymptotic results on the single observation estimates Chatterjee, Diaconis, Sly, et al., 2011 guarantee that, if the network density remains constant as N grows larger, the accuracy of the cross sectional estimates increases. We want to check numerically that, within the regime of dense networks, the accuracy of the static and SD versions of the beta model reaches the same level. In order to check whether the SD approach provides any advantage for large networks, we perform numerical experiments similar to the previous ones, but in a different and more realistic regime of sparse networks, i.e. keeping constant the average degree. Moreover, to ease the computational burden for the estimates, we consider a restricted version of the SD-Beta model, as detailed in Section 6.5.1 of the Appendix, having only one set of parameters $\left(b^{\text{in}}, b^{\text{out}}, a^{\text{in}}, a^{\text{out}}\right)$ for the whole network, instead of one set per each node.

In this analysis, we consider only one dynamical DGP and many different values of N. Among the DGPs used above, we focus on the one with smooth and periodic time variation. Most importantly, we set a maximum degree attainable for a node and we let it depend on N in two distinct ways, each one corresponding to a different density regime: one generating *sparse* networks and the other *dense* ones. It is important to notice that the asymptotic results of (Chatterjee, Diaconis, Sly, et al., 2011) are expected to hold only in the dense case. The average densities, for different values of N, in the two regimes are shown in the left panel of Figure 6.1. Then, for both regimes and each value of N, we compute the average RMSE across all time-varying parameters and all Monte Carlo replicas. In the right panel of Figure 6.1, the average RMSEs for different values of N clearly indicate that, also for large networks, the SD version of the beta model attains better results compared with the cross sectional estimates. As expected, in the dense network regimes, both approaches reach the same accuracy as long as N becomes larger. However, in the more realistic sparse regime, the performance of the SD-ERGM remains much superior for both small and large network dimensions.

6.2.2 Pseudo Likelihood SD-ERGM

As mentioned earlier, the dependence of the normalizing function on the θ parameters is often unknown. This fact prevents us from computing the score function and directly applying the update rule (5.2) to a large class of ERGMs. To circumvent

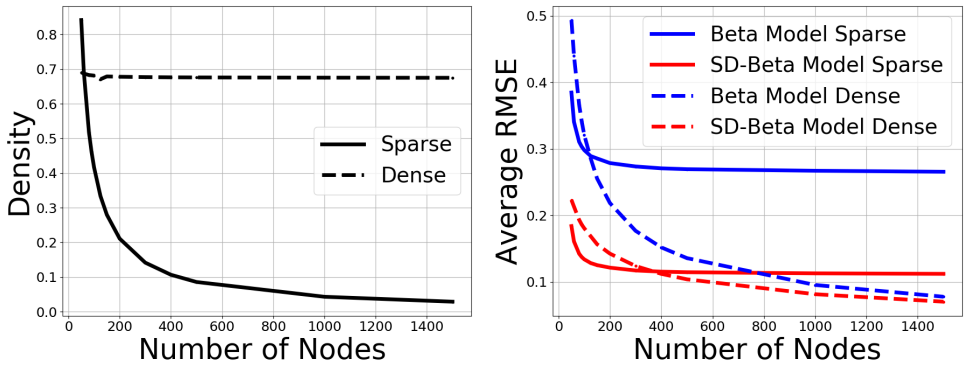

FIGURE 6.1: Left panel: average density as a function of the number of nodes N in the dense (dashed line) and sparse (solid line) regimes. Right panel: average RMSE of the filtered parameters with respect to the simulated DGP in both the dense (dashed lines) and sparse (solid lines) regimes. The average RMSE from the ERGM is plotted in blue, while the one from the SD-ERGM in red.

this obstacle, instead of the unattainable score of the exact likelihood, we propose to use the score of the pseudo-likelihood, discussed in Sec. 2.2.3, that we refer to as pseudo-score

$$\nabla^{(t)}(\theta) = \frac{\partial \log PL\left(\mathbf{Y}^{(t)}|\theta\right)}{\partial \theta_s^{(t)\prime}} = \sum_{ij} \delta_{ij}^s \left(Y_{ij}^{(t)} - \frac{1}{1+e^{-\sum_l \theta_l \delta_{ij}^l}} \right), \tag{6.4}$$

in place of the exact score in the definition the SD-ERGM update (5.2). Additionally, we use the pseudo-likelihood for each observation $\mathbf{Y}^{(t)}$ in (6.2) for the inference of the static parameters.

Our approach, based on the score of the pseudo-likelihood, requires as input the change statistics for each function $h_s\left(\mathbf{Y}^{(t)}\right)$ [2]. In the following, we show that the update based on the score of the pseudo-likelihood is effective in filtering the path of time-varying parameters. Remarkably, this is true even when the probability distribution in the DGP is the exact one, i.e. when we sample from the exact likelihood and then use the SD-ERGM based on the pseudo-likelihood to filter.

At this point we point out that the realized optimality, proven in Blasques, Koopman, and Lucas, 2015 and reviewed in Section 5.4, defines a class of updates; it does not represent a single update with a unique functional form. For instance, $\Delta_{t|t}$ defined in (5.6), is clearly specific of the chosen $\tilde{P}$. A different choice of $\tilde{P}$, e.g. one inspired by the pseudo-likelihood specification, translates into an alternative optimal choice for the update. In general, there can be an infinite number of realized Kullback-Leibler optimal updates.

SD-ERGM for Transitivity and Network Density

In this section we discuss numerical simulations for an ERGM whose normalization is not known in closed form, that we apply also to real data in Section 6.3.2. We show the concrete applicability of the SD-ERGM approach based on the pseudo-score and

[2] For practical applications, it is very convenient that, for a large number of network functions, an efficient implementation to compute change statistics is made available in the R package *ergm* Hunter et al., 2008.

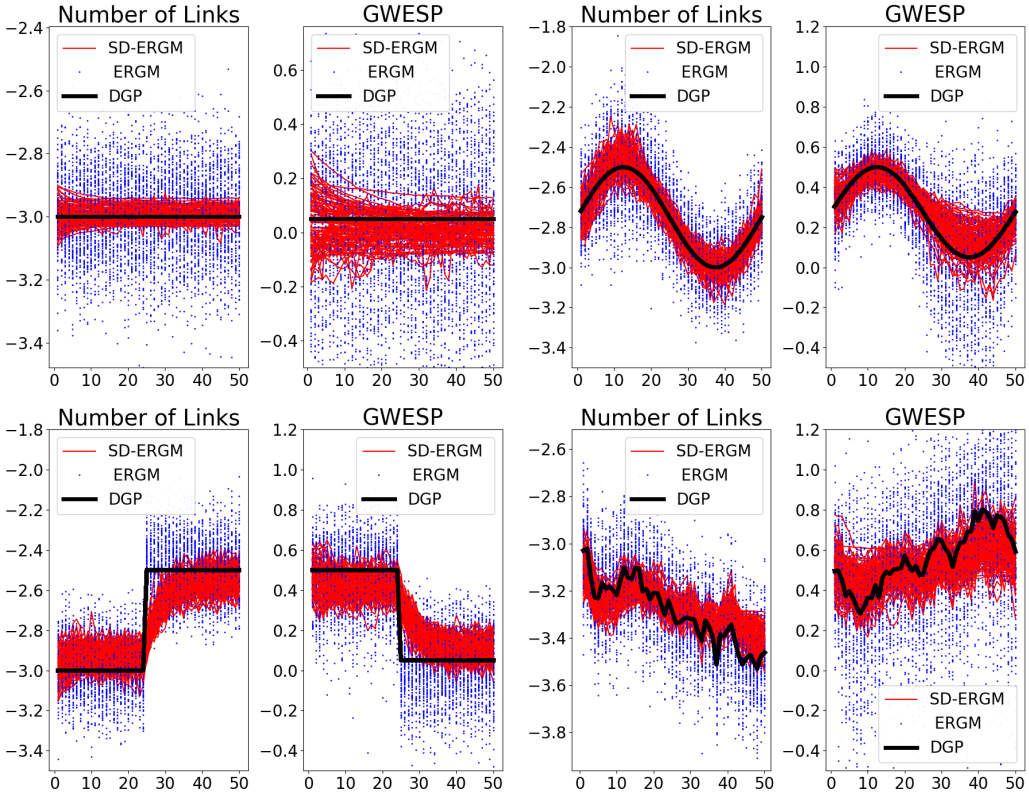

FIGURE 6.2: Filtered paths of the parameters of the ERGM in (6.5) with time-varying parameters. The path from the true DGP is in black. The blue dots are the cross sectional ERGM estimates, and the red lines the SD-ERGM filtered paths.

its performance as a filter in comparison with the cross sectional MCMC estimates of the standard ERGM. The models we consider have two statistics. The first one is the total number of links present in the network. The second statistics is the GWESP, introduced in Section 2.2.2. The ERGM is thus defined by

$$\sum_s \theta_s q_s \left(\mathbf{Y}^{(t)}\right) = \theta_1 \sum_{ij} Y_{ij}^{(t)} + \theta_2 \text{GWESP}\left(\mathbf{Y}^{(t)}\right). \tag{6.5}$$

To test the efficiency of the SD-ERGM, we simulate a known temporal evolution for the parameters and, at each time step, we sample the exact PMF from the resulting ERGMs. Finally, we use the observed change statistics for each time step to estimate two alternative models: a sequence of cross sectional ERGMs and the SD-ERGM. In what follows, we indicate the values from the DGP of parameter s at time t as $\bar{\theta}_s^{(t)}$.

We investigate four DGPs similar to those analyzed in Section 6.2.1. We sample and estimate the models 50 times for each DGP. Figure 6.2 compares the cross sectional estimates and the score-driven filtered paths. Table 6.2 reports the RMSE of the GWESP time-varying parameters, averaged over the different realizations for the whole sequence $t = 1, 2, \ldots, T$. It is evident that the SD-ERGM outperforms the cross sectional ERGM estimates for all the investigated time-varying patterns. Moreover, when the constant DGP is considered, i.e. $\bar{\theta}_1^{(t)} = \bar{\theta}_1$ and $\bar{\theta}_2^{(t)} = \bar{\theta}_2$, the average RMSE of the SD-ERGM is larger, but comparable, than the correctly specified ERGM

TABLE 6.2: First four columns: RMSEs for the filtered paths of the time-varying parameters, averaged over 50 repetitions, for the evolutions of Figure 6.2. The last three columns describe the accuracy of the test for dynamics in the parameters, considering the DGPs in Figure 6.2, as well as alternative DGPs where only one parameter is time varying. We report the percentage of times that the LM test correctly identifies the parameter as time-varying (or static in the case of the first DGP). The chosen threshold for the p-values is 0.05.

DGP	Average RMSE				LM Test	
	ERGM		SD-ERGM		% Correct Results	
	$\theta_1^{(t)}$	$\theta_2^{(t)}$	$\theta_1^{(t)}$	$\theta_2^{(t)}$	$\left(\theta_1^{(t)}, \theta_2\right)$	$\left(\theta_1, \theta_2^{(t)}\right)$
Const	0.02	0.1	0.0006	0.004		
Sin	0.02	0.04	0.003	0.005	94%	93%
Steps	0.02	0.03	0.01	0.001	92%	96%
AR(1)	0.02	0.2	0.007	0.01	93%	90%

that uses all the longitudinal observations to estimate the parameters. The latter result confirms that, even for the static case, the SD-ERGM is a reliable and consistent choice.

It is worth noticing that, for sampling and cross sectional inference, we employed the R package *ergm* that uses state of the art MCMC techniques for both tasks (see Hunter et al., 2008, for a description of the software). Hence, we compared the SD-ERGM based on the approximate pseudo-likelihood – both in the definition of the time-varying parameter update and inference of the static parameters – with a sequence of exact cross sectional estimates, that are in general known to be better performing than the pseudo-likelihood alternative, as mentioned in Section 2.2. Even if the cross sectional estimates are based on the exact likelihood, while the SD approach is based on an approximation, the SD-ERGM remains the best performing solution. In our opinion, this provides further evidence of the advantages of SD-ERGM as a filtering tool. Finally, the last column of Table 6.2 reports the percentage number of times the LM test of Calvori, Creal, Koopman, and Lucas, 2017 applied to the SD-ERGM correctly classifies the parameters as time-varying (or static for the constant DGP). The test performs correctly in all the cases considered.

Comparison of Pseudo and Exact Likelihood SD-ERGM

To further investigate the proposed SD-ERGM, and its version based on the pseudo-likelihood, in this section we focus on the ERGM having the total number of links and the total number of mutual links as network statistics:

$$\sum_s \theta_s q_s \left(\mathbf{Y}^{(t)}\right) = \theta_L \sum_{ij} Y_{ij}^{(t)} + \theta_M \sum_{ij} Y_{ij}^{(t)} Y_{ji}^{(t)}. \tag{6.6}$$

The static version of this model is known as *reciprocity* $p^\star$ model (Snijders, 2002). This model is relevant for our discussion because it allows us to compare the SD time-varying extension based on the pseudo-likelihood with the one based on the exact likelihood. Indeed it is simple enough that the normalizing function is known in closed form, but it has enough structure such that its pseudo-likelihood differs

TABLE 6.3: RMSEs of ML-SD-ERGM and PML-SD-ERGM, relative to that of the cross sectional ERGM. The averages are obtained over 50 repetitions, for the $AR(1)$ DGP, described in the text. For each value of T and N we report the RMSE of the SD-ERGMs divided by the RMSE of the cross sectional ERGM.

	PML-SD-ERGM			ML-SD-ERGM		
T \ N	50	100	500	50	100	500
100	0.016	0.011	0.006	0.015	0.011	0.006
300	0.015	0.011	0.006	0.014	0.011	0.007
600	0.014	0.012	0.007	0.014	0.011	0.006

from its exact likelihood. In fact, the model results in dyads, i.e. pairs of mutual links (A_{ij}, A_{ji}), being independent, while the pseudo-likelihood amounts to assuming independent links. Moreover, since its partition function is available in closed form, such a model can be sampled efficiently, without the need to resort to MCMC methods. This allows us to run extensive numerical simulations, in reasonable time, to investigate the properties of the confidence bands proposed by Buccheri, Bormetti, Corsi, and Lillo, 2018 in the context of SD-ERGM models.

In this section we will refer to the pseudo-likelihood based SD-ERGM as PML-SD-ERGM, and to the exact likelihood case as ML-SD-ERGM. We compare the capacity of the two models, used as filters, to recover a misspecified dynamics using the same approach as in the previous sections, i.e. we simulate a known DGP for $\theta_L^{(t)}$ and $\theta_M^{(t)}$. We focus on a DGP where θ_L and θ_M follow two independent $AR(1)$ processes, as the one discussed in 6.2.1. Each $AR(1)$ has $\Phi_1 = 0.98$ and $\epsilon \sim N(0, \sigma)$ with $\sigma = 0.005$. The Φ_0 parameters are chosen such that, on average, the network density is equal to 0.3 and the fraction of reciprocated pairs is 0.075. We choose this value because it is between the maximum and minimum fraction of reciprocated links possible for a network of density 0.3, which is 0 and $0.3(N^2 - N)/2$. When comparing results for different network sizes, we keep the density fixed for all network sizes N, thus exploring a dense regime[3]. In our numerical experiment, we first sample repeatedly sequences of synthetically generated observations from different specifications of the DGP. We then estimate the PML and ML versions of the SD-ERGM on those observations, and filter the time varying parameters. Finally we quantify their accuracy, with the average RMSE, across 50 samples, with respect to the simulated DGP. In Table 6.3 we report the RMSE, for both PML-SD-ERGM and ML-SD-ERGM, divided by the RMSE of the cross sectional standard ERGM, for various combinations of network size N and number of observations T. It clearly emerges that both versions of SD-ERGM strongly outperform the cross sectional ERGM. Moreover, the performances of PML-SD-ERGM is similar to the one of the exact ML-SD-ERGM.

In the final part of this section, we investigate the possibility of using the method of Buccheri, Bormetti, Corsi, and Lillo, 2018, that we describe in Section 5.5, to define confidence bands for the parameters filtered with SD-ERGM. In Buccheri, Bormetti, Corsi, and Lillo, 2018 the authors characterize the approximation error when the SD approach is used to filter a set of latent parameters, whose true DGP is an autoregressive process. While we refer to Buccheri, Bormetti, Corsi, and Lillo, 2018 for the details, we point out that their procedure rests upon the assumption that the SD

[3]We found the conclusions of this section to hold true also in a sparse network density regime.

TABLE 6.4: Coverages of the 95% confidence bands averaged over 50 repetitions, for the $AR(1)$ DGP, described in the text, and $N = 100$.

T	ML-SD-ERGM	PML-SD-ERGM
300	99.1 %	99.9 %
3000	94.5%	95.7 %

filter approximates the true underlying DGP. The authors prove that this approximation becomes exact in the limit of small variance for the latent parameters. Hence, the confidence bands obtained with their method are theoretically guaranteed to be reliable only in this limit. In practice, it is appropriate to assess whether for a given value of the variance of the DGP the application of the confidence bands is justified. This can be done by numerical experiments to assess their coverage with a simulated DGP. For example, for the model and the DGP considered in this section, we check the coverage of the confidence bands obtained, and report the results in Table 6.4, for $N = 100$. We find that that the coverage of the confidence bands, for both ML-SD-ERGM and PML-SD-ERGM, approaches the nominal value in the limit of large T, while for short time series, their coverages is higher than the nominal value. Hence in small samples they should be interpreted as having a confidence of *at least* their nominal values.

6.3 Applications to Real Data

After the analysis of synthetic data, this section presents two applications to real dynamical networks. Our goal is to show the value of SD-ERGM as a methodology to model temporal networks, irrespective of the specific system that a researcher wants to investigate. The two real networks that we consider have been the object of multiple studies in different streams of literature. They have been investigated, in the context of ERGMs, using different network statistics. We first consider a network of credit relations among Italian banks. The second real world application focuses on a network of interest for the social and political science community, namely the network of U.S senators cosponsoring legislative bills.

6.3.1 Link Prediction in Interbank Networks

Our first empirical application is to data from the electronic Market of Interbank Deposit (e-MID), a market where banks can extend loans to one another for a specified term and/or collateral. Interbank markets are an important point of encounter for banks' supply and demand of extra liquidity. In particular, e-MID has been investigated in many papers (see, for example Iori et al., 2008; Finger, Fricke, and Lux, 2013; Mazzarisi, Barucca, Lillo, and Tantari, 2020; Barucca and Lillo, 2018, and references therein). Our dataset contains the list of all credit transactions in each day from June 6, 2009 to February 27, 2015. In our analysis, we investigate the interbank network of overnight loans, aggregated weekly. We follow the literature and disregard the size of the exposures, i.e. the weights of the links. We thus consider a link from bank j to bank i present at week t if bank j lent money overnight to bank i, at least once during that week, irrespective of the amount lent. This results in a set of $T = 298$ weekly aggregated networks. For a detailed description of the dataset, we refer the reader to Barucca and Lillo, 2018.

In recent years, the amount of lending in e-MID has significantly declined. In particular, as discussed in Barucca and Lillo, 2018, it abruptly declined at the beginning of 2012, as a consequence of important unconventional measures (Long Term Refinancing Operations) by the European Central Bank, that guaranteed an alternative source of liquidity to European banks. The evident non-stationary nature of the evolution of the interbank network is of extreme interest for our purposes. In fact, as mentioned in Sections 6.2.1 and 6.2.2, one of the key strengths of SD-ERGM, used as a filter, is precisely the ability of recovering such non-stationary dynamics.

In the following, we use the SD beta model for links forecasting. Specifically, we consider the version with a restricted number of static parameters discussed at the end of Sec. 6.2.1. We divide the data set in two samples. We consider rolling windows of 100 observations and estimate the parameters a^{out}, b^{out}, a^{in} and b^{in} on each one of those rolling windows. For each window, we then test the forecasting performances, up to 8 steps ahead (i.e. roughly two months). The forecast works as follows. Assuming that at time t, the last date of the rolling window, we have filtered the value for the parameters $\overleftarrow{\theta}^{(t)}$ and $\overrightarrow{\theta}^{(t)}$, we plug the estimated static parameters and the matrix $\mathbf{Y}^{(t)}$ in the SD update and compute the time-varying parameters $\overleftarrow{\theta}^{(t+1)}$ and $\overrightarrow{\theta}^{(t+1)}$. From the latter, we readily obtain the forecast of the adjacency matrix

$$\mathbb{E}\left[\mathbf{Y}^{(t+1)} \middle| \overleftarrow{\theta}^{(t+1)}, \overrightarrow{\theta}^{(t+1)}\right],$$

where $t+1$ is the first date of the test sample. The K-step-ahead forecast for the SD-ERGM model is obtained simulating the SD dynamics up to $t+K$ 100 times[4], thus obtaining $\overrightarrow{\theta}_n^{(t+K)}$ and $\overleftarrow{\theta}_n^{(t+K)}$ for $n = 1, \ldots, 100$, and then taking the average of the expected adjacency matrices $\frac{1}{100}\sum_n \mathbb{E}\left[\mathbf{Y}^{(t+K)} \middle| \overleftarrow{\theta}_n^{(t+K)}, \overrightarrow{\theta}_n^{(t+K)}\right]$. Given the forecast values, we compute the rate of false positives and false negatives. Then, we drop the first element from the train set and add to it the first element of the test sample. We repeat the forecasting exercise estimating the SD-ERGM parameters on the new train set and testing the performance on the new test sample. We name this procedure rolling estimate and iterate it until the test sample contains the last 8 elements of the time-series.

Given a forecast for the adjacency matrix, we evaluate the accuracy of the binary classifier by computing the Receiving Operating Characteristic (ROC) curve. All results are collected and presented in Figure 6.3. The left panel reports the ROC curve for one-step-ahead link forecasting obtained according to the SD-ERGM rolling estimate. The panel also shows other three curves based on the static beta model. Specifically, the green curve results from a naive prediction, where the presence of a link tomorrow is forecasted assuming that the $t+1$ ERGM parameter values are equal to those estimated at time t. Once the sequence of cross sectional estimates of the static ERGM is completed, we take the estimated values $\overleftharpoon{\theta}^{(t)}$ and $\overrightharpoon{\theta}^{(t)}$ as observed and model their evolution with an auto-regressive model of order one, AR(1). That amounts to assuming $\overleftharpoon{\theta}^{(t+1)} = c_0 + c_1 \overleftharpoon{\theta}^{(t)} + \epsilon^{(t)}$, where c_0 and c_1 are the static parameters of the AR(1), and $\epsilon^{(t)}$ is a sequence of i.i.d. normal random variables with zero mean and variance σ^2. A similar equation holds for the out-degree parameters. Using the observations from the training sample, we estimate the parameters c_0, c_1, and σ^2 and use them for a standard AR(1) forecasting exercise on the test sample.

[4] It is worth to stress that the results become stable after 20 simulations.

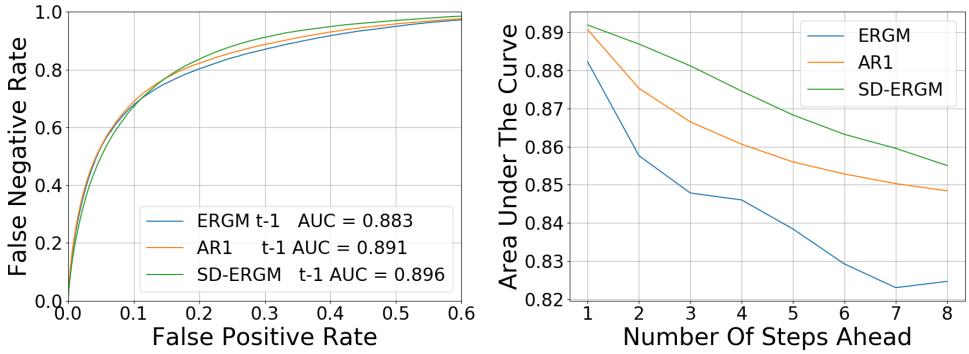

FIGURE 6.3: Left panel: ROC curves for one-step-ahead link forecasting. The green and orange ROC curves describe the one-step-ahead forecasting with SD and cross sectional AR(1) beta model, respectively. The blue curve corresponds to the forecast based on previous-time-step ERGM. Right panel: AUC for the multi-step-ahead forecast.

The results correspond to the orange curve. It is important to stress that, while the SD-ERGM forecast requires one static and one time-varying estimation on the train set, in the latter procedure we have to estimate the static parameters for each date in the train sample.

The left plot of Fig. 6.3 shows that the naive one-step-ahead forecast, in spite of its simplicity, provides a quite reasonable result. The best performance corresponds however to the forecast based on the SD-ERGM. The AR(1) static ERGM improves on the naive forecast and it is slightly worst than the SD-ERGM. However, as commented before, it is more computationally intensive. More importantly, the right panel of Fig. 6.3 presents the result from a multi-step-head forecasting analysis. It emerges clearly that the performance of the naive forecast (blue curve), tested up to $K = 8$, rapidly deteriorates, while the SD-ERGM multi-step forecast remains the best performing. [5].

To conclude this section, we point out that all three methods considered here for link forecasting do not require the full knowledge of the adjacency matrices in order to forecast the presence of links. This is due to the fact that they are all based on the directed *beta model*. The latter, as we discussed in Section 2.2.1, can be defined as the ERGM with the sequences of in and out degrees as sufficient statistics. In particular, these statistics are sufficient to estimate our score driven version of the beta model and to compute the score in the update rule (6.3), needed for the link prediction. As discussed in Chapter 3, the problem of reconstructing networks from contemporaneous partial information[6] has received much attention in the literature, and is particularly relevant from the perspective of systemic risk reconstruction (Mistrulli, 2011; Anand et al., 2017). It is thus very interesting that our methodology, which has been shown to outperform the alternatives discussed here, can be effectively used to forecast directed binary networks from the sole knowledge of the degree sequences.[7]

[5] In all the results on link forecasting – one- or multi-step-ahead – we excluded the links that are always zero, i.e. they never appear in the train and test samples. The reason is that those are extremely easy to predict and keeping them would give an unrealistically optimistic picture on the predictability of links in the data set. Importantly, the ranking of the methods remain unaltered when we keep all links for performance evaluation.

[6] Reconstructing a network at time t from partial information describing the network at time t.

[7] And with very simple modifications it can be extended to undirected networks.

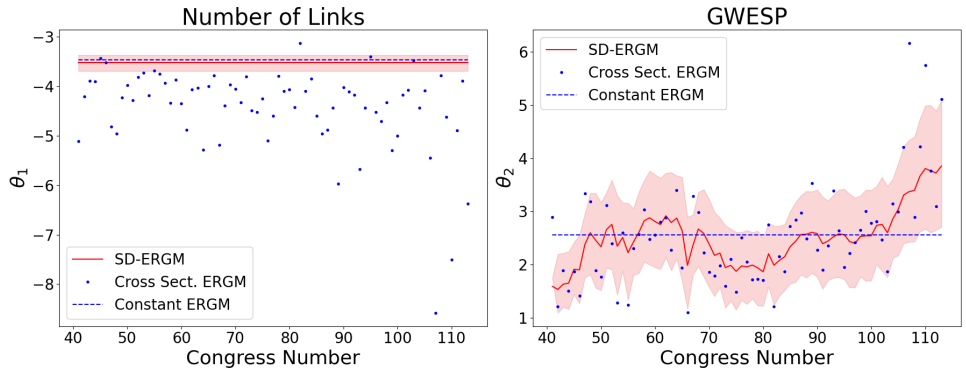

FIGURE 6.4: Estimates for the time-varying parameters associated with the number of links and the GWESP statistics. Blue dots correspond to the cross sectional ERGM estimates, while the red lines are the estimates from the SD-ERGM, with the corresponding 95% confidence intervals denoted by the red shaded regions.

Finally, let us mention that being able to run link prediction based on the sole knowledge of the degree sequence can be extremely convenient when applying the method to very large networks. Indeed, from a practical point of view, the memory required to store and process the time series of degree sequences, of length $2N$ for each time step, can be much lower than that needed to store the time series of full adjacency matrices, each having $N(N-1)$ elements.

6.3.2 Temporal Heterogeneity in U.S. Congress Co-Voting Political Network

Networks describing the U.S. congress' bills have been the object of multiple studies (see, for example Fowler, 2006; Faust and Skvoretz, 2002; Zhang et al., 2008; Cranmer and Desmarais, 2011; Moody and Mucha, 2013; Wilson, Stevens, and Woodall, 2016; Lee, Li, and Wilson, 2020). It is thus an appropriate real system for our second application of the SD-ERGM framework. In particular, we want to show that the update rule based on the pseudo-score defined in (6.4) can be concretely applied to a real network, and that it draws a different picture when compared to the sequence of cross sectional ERGM estimates. In order to build the network, we use the freely available data of voting records in the US Senate (see Lewis et al., 2019) covering the period from 1867 to 2015, for a total of 74 Congresses. We define the network of co-voting following Roy, Atchadé, and Michailidis, 2017 and Lee, Li, and Wilson, 2020, where a link between two senators indicates that they voted in agreement on over 75% of the votes, among those held in a given senate when they were both present. This procedure results in a sequence of 74 networks, one for each different Congress starting from the 40th. For this empirical application, we consider the SD-ERGM with the two network statistics discussed in Section 6.2.2. As defined in (6.5), parameter $\theta_1^{(t)}$ is associated with the number of edges, while $\theta_2^{(t)}$ with the GWESP statistic. The fact that the number of nodes is not constant over time is not a problem for our application, since we do not consider statistics associated to single nodes. That case – as for instance considering the degrees of the beta model – would require the number of time-varying parameters to be different at each time step.

As we did for the numerical simulations and the previous empirical application, we compare our framework with a sequence of standard ERGMs. The goal of this

empirical exercise is not to draw conclusions about the specific network at hand. Our main aim is to show that the two approaches return a qualitatively different picture. The choice between the alternative models, and combinations of statistics – possibly based on model selection techniques – is beyond the scope of our exercise.

Using the test for temporal heterogeneity based on SD-ERGM, only the parameter θ_2 turns out to be time-varying. In fact, testing the null hypothesis that each parameter is static, we obtain a p-value of 0.1 for the link density and 10^{-4} for GWESP. In order to check whether the sequence of cross sectional estimates is consistent with the hypothesis that the parameters remain constant, we estimate the values θ_1^c, θ_2^c from an ERGM using all observations. This amounts to compute $\theta^c = \arg\max_{\theta} \sum_{t=1}^{74} \log P\left(\mathbf{Y}^{(t)}, \theta\right)$. Then, for each sequence of cross sectional estimates $\theta_1^{(t)}$ and $\theta_2^{(t)}$, we test the hypothesis of them being normally distributed around the constant values with unknown variance. The p-values resulting from the t-tests are 1.4×10^{-6} and 0.03 for parameters θ_1 and θ_2, respectively. This simple test confirms that the two approaches imply quantitatively different behaviors for the parameters. This clearly emerges from Figure 6.4 that reports the estimates from the SD-ERGM (thick red lines), with their respective 95% confidence intervals (shaded red bands), as well as the cross sectional ERGM estimates – one per date (blue dots) or using the entire sample (dashed blue line).

In order to compute the confidence bands as in Buccheri, Bormetti, Corsi, and Lillo, 2018, we checked numerically whether the data is compatible with a DGP having a small variance. In practice, we first estimate the SD-ERGM. Then we quantify the variance of the latent parameters by estimating an AR(1) on the filtered time series[8]. Finally we repeatedly simulate such an $AR(1)$ DGP, similarly to what done at the end of section 6.2.2, and check the coverage of the confidence bands. We find that, for the current application, the coverage of the confidence bands is 99.9%, hence greater than the nominal value. These simulation-based results support the reliability of the approximate SD filter and provide a conservative estimate of the confidence bands. Thus allowing us, for example, to deduce that the data is not compatible with a model where one of the two parameters is zero.

6.4 Conclusions

In this chapter, we proposed a framework for the description of temporal networks that extends the well known Exponential Random Graph Models. In the new approach, the parameters of the ERGM have a stochastic dynamics driven by the score of the conditional likelihood. If the latter is not available in closed-form, we showed how to adapt the score-driven updating rule to a generic ERGM by resorting to the conditional pseudo-likelihood. In this way, our approach can describe the dynamic dependence of the PMF from virtually all the network statistics usually considered in ERGM applications. We investigated two specific ERGM instances by means of an extensive Monte Carlo analysis of the SD-ERGM reliability as a filter for time-varying parameters. The chosen examples allowed us to highlight the applicability of our method to models with a large number of parameters and to models for which

[8] We extensively tested via simulation that, for the model at hand and T and N taken from the data, estimating the variance of the latent parameters in such a way results, on average, in a small underestimation. In checking the coverage of the confidence bands we considered a DGP with variance increased, with respect to the one estimated on the filtered time series, to compensate for this bias.

the normalization of the PMF is not available in closed form. The numerical simulations proved the clear superior performance of the SD-ERGM over a sequence of standard cross sectional ERGM estimates. This is not only true in the sparse network regime, but also in the dense case when the number of nodes is far from the asymptotic limit. Finally, we run two empirical exercises on real networks data. The first application to e-MID interbank network showed that the SD-ERGM provides a quantifiable advantage in a link forecasting exercise over different time horizons. The second example on the U.S. Congress co-voting political network enlightened that the ERGM and the SD-ERGM could provide a significantly different picture in describing the parameter dynamics.

Our work opens a number of possibilities for future research. First, the applicability of the test for parameter instability in the context of SD-ERGM with multiple network statistics could be investigated much further. Second, the SD-ERGM could be applied on multiple instances of real world dynamical networks. An interesting application would be the study of networks describing the dynamical correlation of neural activity in different parts of the brain (see, for example, Karahanoğlu and Van De Ville, 2017, for a review of the topic and list of references). In this context, the application of the static ERGM have already proven to be extremely successful (as, for example, in Simpson, Hayasaka, and Laurienti, 2011). Moreover, as mentioned at the end of Section 6.3.1, the score driven beta model can be used to forecast binary networks from the sole knowledge of the nodes' in and out degrees. We believe that this is extremely relevant for possible extensions of the literature on network reconstruction from partial information (Anand et al., 2017), to the context of temporal networks. The last future development that we plan to explore is the extension of the score-driven framework to the description of weighted dynamical network. Regretfully, this setting has not received enough attention in the literature (one isolate example is Giraitis, Kapetanios, Wetherilt, and Žikeš, 2016), but it is of extreme relevance, particularly from the financial stability perspective and its implications for systemic risk.

6.5 Appendix

6.5.1 Details of Numerical Simulations

In the main text we referred to a set of DGPs used for numerical simulations. Although the different numerical experiments that we presented differ for the meaning and number of parameters, in every experiment each of the parameters can be constant or evolve according to one of the following dynamical DGPs:

- abrupt change of half the parameters at $t = T/2$, i.e. for odd s we have $\overline{\theta}_s^{(t)} = \overline{\theta}_{1_s}$ for $t \leq T/2$ and $\overline{\theta}_s^{(t)} = \overline{\theta}_{2_s}$ for $t > T/2$, while for even s it is $\overline{\theta}_s^{(t)} = \overline{\theta}_{0_s}$ for $t = 1 \dots T$;
- smooth periodic variation for half the parameters, i.e. for odd s we have $\overline{\theta}_s^{(t)} = \overline{\theta}_{0_s} + (\overline{\theta}_{2_s} - \overline{\theta}_{1_s}) \sin(4\pi t/T + \phi_s)$ for $t = 1 \dots T$, where the ϕ_s are randomly chosen for each node, while for even s it is $\overline{\theta}_s^{(t)} = \overline{\theta}_{0_s}$ for $t = 1 \dots T$;
- autoregressive of order 1 (AR(1)), i.e. for odd s we have $\overline{\theta}_s^{(t)} = \Phi_{0_s} + \Phi_1 \overline{\theta}_s^{(t-1)} + \epsilon^{(t)}$ for $t = 1 \dots T$, where $\Phi_1 = 0.99$, Φ_{0_s} is chosen such that the unconditional mean is equal to θ_{0_s}, $\epsilon \sim N(0, \sigma)$ and $\sigma = 0.1$. As in the previous cases, for even s we keep $\overline{\theta}_s^{(t)} = \overline{\theta}_{0_s}$ for $t = 1 \dots T$.

The dynamics considered are such that element s of vector θ remains bounded between θ_{1_s} and θ_{2_s}. The values of θ_1 and θ_2 are fixed in order to allow fluctuations in the in and out degrees of the nodes, as follows. The vector $\bar{\theta}_0$ is obtained by first generating two degree sequences (in and out) such that the degrees linearly interpolate between a minimum degree $D_m = 3$ and a maximum of $D_M = 8$. Then, we need to ensure that the degree sequence is graphicable, i.e. such that it exists one matrix of zeros and ones from which it can be obtained. We iteratively match links that make up the out-degree sequence with those that make up the in-degree sequence, starting with the largest in- and out-degrees. In practice, we start with an empty matrix, select the largest out degree and set to one the matrix element between this node and the node with largest in degree. If at some point we cannot entirely allocate a given out-degree, we disregard the leftover links outgoing from that node and move to the next one. This procedure amounts to populating the adjacency matrix, until no more links can be allocated. The degree sequence associated to this adiacency matrix is guaranteed to be graphicable. The numerical values of $\bar{\theta}_0$ follow from the estimation of the static beta model. Finally, in order to gain additional heterogeneity in the amplitude of the fluctuations, we define N values evenly spaced between 0.4 and 1, i.e. c_s for $s = 1 \dots N$. We use them to define

$$\bar{\theta}_{1_s} = \bar{\theta}_{0_s} + c_s \left(\bar{\theta}_{0_{s+1}} - \bar{\theta}_{0_s} \right)$$

$$\bar{\theta}_{2_s} = \bar{\theta}_{0_s} - c_s \left(\bar{\theta}_{0_{s+1}} - \bar{\theta}_{0_s} \right).$$

Figure 6.5 shows the temporal evolution for one randomly chosen parameter of the beta model for all the DGPs, together with the paths filtered from the observations using the SD beta model and the sequence of cross sectional static estimates. The score-driven filtering and cross sectional estimation are repeated over 100 simulated network sequences. As discussed in the main text, it clearly emerges that the paths filtered with the SD beta model are on average much more accurate than those recovered from a standard beta model.

SD-Beta Model for large N

In one of the numerical simulations that we presented in the main text we consider networks of increasing size. Here we present some additional details on how the DGPs are defined for networks of increasing size. Practically, we have to fix the vectors $\bar{\theta}_{0_s}$, $\bar{\theta}_{1_s}$, and $\bar{\theta}_{2_s}$ in a similar way, with the only difference being the numerical values for D_m and D_M. Specifically, in the sparse case we keep for each N $D_m = 10$ and $D_M = 40$. In the dense case, we set $D_M = 0.8N$, i.e. the maximum degree and the average degree both increase.

One peculiarity of the beta model is that the number of parameters, i.e. the length of the vectors $\overleftarrow{\theta}$ and $\overrightarrow{\theta}$, increases with the number of nodes. This is not the case for many ERGMs, as for example the one that we will discuss in the following section. Consistently, when we use the score-driven extension described so far, the length of the vectors w, a and b increases too.

Recall that the numerical values of θ_0, θ_1 and θ_2 are chosen in order to fix the values of average degrees over time and the amplitude of their fluctuations, as described in Appendix 6.5.1. For each value of N, we choose them in order to guarantee heterogeneity in the degrees across nodes and significant fluctuation in time. Most importantly, we set a maximum degree attainable for a node and we let it depend on N in two distinct ways, each one corresponding to a different density regime: one

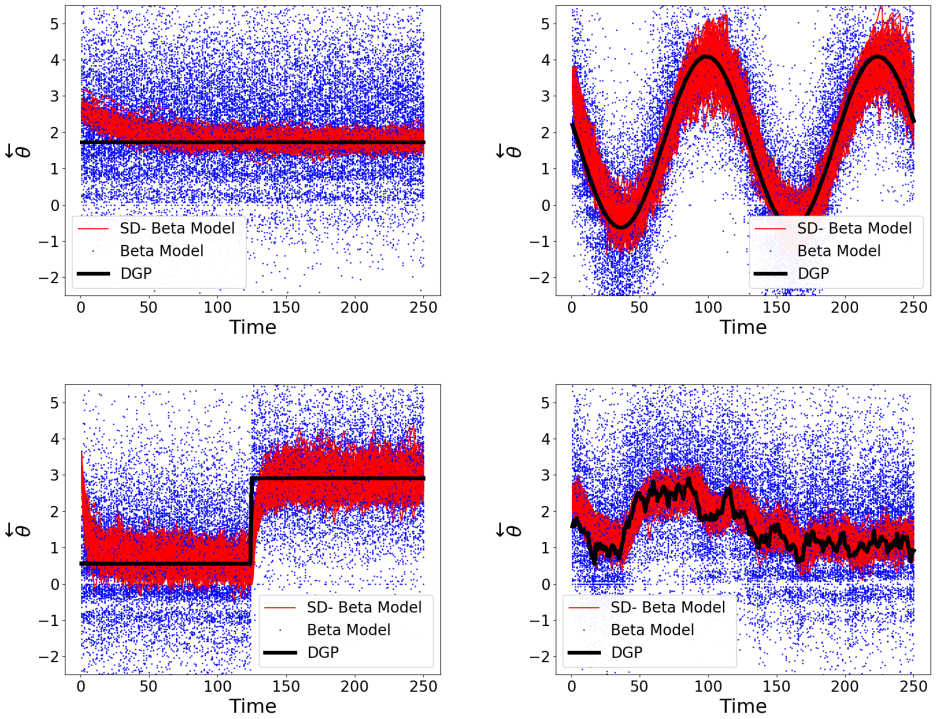

FIGURE 6.5: Temporal evolution of one of the parameters, randomly selected, for the considered DGPs. The black line is the true path of the parameter of the DGP, the red ones are those filtered using the SD-beta model, and the blue dots correspond to the cross sectional estimates of the beta model.

generating *sparse* networks and the other *dense* ones. It is important to notice that the asymptotic results of (Chatterjee, Diaconis, Sly, et al., 2011) are expected to hold only in the dense case.

In the first numerical experiment testing the SD-beta model as misspecified filter, we estimated a total of 60 parameters, 6 static parameters for each one of the ten nodes, 3 for the time-varying in-degree and 3 for the time-varying out-degree. Here we present in detail the further parameter restriction mentioned in the main text, that proved to be useful when the number of nodes increases. Specifically, we assume that the parameters a^{out} and b^{out} are common to all out-degree time-varying parameters $\overrightarrow{\theta}^{(t)}$. Similarly, all in-degree time-varying parameters $\overleftarrow{\theta}^{(t)}$ share the same a^{in} and b^{in}. The coefficients w_s^{in} and w_s^{out} remain node specific. The resulting update rule is

$$\begin{aligned}
\overleftarrow{\theta}_s^{(t+1)} &= w_s^{\text{in}} + b^{\text{in}} \overleftarrow{\theta}_s^{(t)} + a^{\text{in}} \left(\frac{\sum_i Y_{is}^{(t)} - p_{is}^{(t)}}{\sqrt{\sum_i p_{is}^{(t)} \left(1 - p_{is}^{(t)}\right)}} \right) \\
\overrightarrow{\theta}_s^{(t+1)} &= w_s^{\text{out}} + b^{\text{out}} \overrightarrow{\theta}_s^{(t)} + a^{\text{out}} \left(\frac{\sum_i Y_{si}^{(t)} - p_{si}^{(t)}}{\sqrt{\sum_i p_{si}^{(t)} \left(1 - p_{si}^{(t)}\right)}} \right)
\end{aligned} . \tag{6.7}$$

Chapter 7

Score Driven Model for Sparse Weighted Temporal Networks

Synopsis of The Chapter In this Chapter, we propose a novel time varying parameter model for sparse weighted temporal networks as a combination of the fitness model, appropriately extended to handle also the weights, and the score driven framework. While the vast majority of the literature on models for time varying networks focuses on *binary graphs*, i.e. graphs that are defined solely by a set of nodes and a set of links between pairs of nodes, often we can associate a weight to each link. In such cases the data is better described by a weighted, or valued, network. One important well known fact is that real world valued networks are very often found to be sparse, i.e. their adjacency matrices have an abundance of zero entries. Our main contribution is a model for sparse weighted dynamical networks, that also accommodates for the dependency of the network dynamics on external variables, and its application to weighted temporal network data, describing overnight exposures in the European interbank market. Our work contributes to the extremely scarce literature on dynamical models for sparse weighted networks by extending the well known fitness model for static binary networks. We consider a zero augmented generalized linear model to handle the weights and a state of the art econometric approach to describe time varying parameters. This results in a flexible model that allows us to decouple the probability of a link to exist from its expected weight, and to explore the influence of external regressors on the network's dynamics. We then exploit such flexibility to investigate how the relevance of EONIA rates on the e-Mid interbank market changed over time.

7.1 Introduction

In the last two decades, networks, or graphs, have attracted an enormous amount of attention as an effective way of describing pairwise relations in complex systems Barabási, 2002. The ever increasing abundance, and variety, of graph data has motivated a great deal of applications of statistical models to graphs (see, for example Newman, 2010, for a review). More recently, the availability of time varying networks' data has stimulated the development of models for temporal networks (Hanneke, Fu, Xing, et al., 2010; Sewell and Chen, 2015; Giraitis, Kapetanios, Wetherilt, and Žikeš, 2016; Mazzarisi, Barucca, Lillo, and Tantari, 2017; Di Gangi, Bormetti, and Lillo, 2019). While the vast majority of this literature focuses on *binary graphs*, i.e. graphs that are defined solely by a set of nodes and a set of links between pairs of nodes, often we can associate a weight to each link. Links' weights are typically positive, discrete or continuous, numbers and can be associated, for example

to the strength of the relation described by each link. In standard binary descriptions, such relevant information is completely lost. For example, in a network of exposures among financial institutions the weight could be the value of the credit. In this case, a binary network would describe in the same way a link associated to an exposure of 1 million to that associated to an exposure of 1 billion. Indeed it is very common for network data to have also informative weights associated with their links. Some additional examples are: the International Trade Network (Leamer and Levinsohn, 1995; Fagiolo, Reyes, and Schiavo, 2010), migration flows (Fagiolo and Santoni, 2016), scientific collaborations (Newman, 2001), transportation networks (Barrat, Barthelemy, Pastor-Satorras, and Vespignani, 2004), just to mention a few. In these cases the data is better described by a weighted, or valued, network, that can be associated with a positive, real valued matrix $\mathbf{Y}$ with elements $Y_{ij} \in \mathcal{R}^+$. Y_{ij} is the value of the link between node i and node j, and $Y_{ij} = 0$ when the link is not present. It is well known that real world networks, both binary and weighted, are very often found to be sparse, i.e. their adjacency matrices have an abundance of zero entries. That is the case, for example, of interbank networks (Anand et al., 2017), a class of weighted temporal networks of paramount importance , that are known to be extremely relevant to financial stability (Allen and Babus, 2011; Haldane and May, 2011), and have motivated the application and development of a number of statistical models for networks (Bargigli et al., 2015b; Mazzarisi, Barucca, Lillo, and Tantari, 2017). In spite of the their relevance, networks' weights have received less attention in the literature on models for temporal networks. Indeed there are only a few models for temporal networks that take them into account. In this chapter we propose a novel model for sparse weighted temporal networks, that also accommodates for the dependency of the network dynamics on external variables. Our efforts are originally motivated by the need to properly model weighted temporal network data, describing overnight exposures in the European interbank market. The resulting modeling framework is nevertheless very general and extremely flexible, as it allows to decouple the probability of a link to exist from its expected weight, and to explore the influence of external covariates on the network's dynamics. We achieve that by extending the well known fitness model (Holland and Leinhardt, 1981a; Caldarelli, Capocci, De Los Rios, and Muñoz, 2002; Garlaschelli and Loffredo, 2008; Chatterjee, Diaconis, Sly, et al., 2011; Yan, Leng, Zhu, et al., 2016; Yan, Jiang, Fienberg, and Leng, 2018) for static binary networks, combining it with a simple generalized linear model to handle the weights and a state of the art econometric approach to describe time varying parameters. In doing so we explicitly account, by means of zero augmentation, for the abundance of zeros that follow from the sparse nature of real world networks. We then exploit such flexibility to investigate the relevance of interbank rates on the e-Mid interbank market, and to contribute to the literature on link persistence in financial networks (Hatzopoulos et al., 2015), exploring also the persistence of links' weights.

7.2 Definition

In order to introduce our *score driven weighted fitness model*, let us describe a sequence of networks with a set of random variables $Y_{ij}^{(t)}$, one for each link. We propose to use Zero Augmentation to model separately the probability of observing a link, $Y_{ij}^{(t)} > 0$, and the probability to observe a specific weight $Y_{ij}^{(t)}$. We choose Zero Augmentation over censoring, as we believe the former to be much more flexible in the context of

network data. With this choice the probability distribution for link ij is

$$P\left(Y_{ij}^{(t)}=y\right)=\begin{cases}1-p_{ij}^{(t)} & for \quad y=0\\ \\ p_{ij}^{(t)}g_{ij}^{(t)}\left(y\right) & for \quad y>0\,.\end{cases} \tag{7.1}$$

where $g_{ij}^{(t)}\left(y\right)$ is the distribution for the positive continuous weight for link ij, conditional on that link being present.

We then model the binary temporal network and its weights by means of time varying fitness and allowing also for the dependency on external covariates X_{ij}. In our model the probability of a link to exist is described by

$$p_{ij}^{(t)}=\frac{1}{1+e^{-(\overleftarrow{\theta}_i^{(t)}+\overrightarrow{\theta}_j^{(t)}+X_{ij}^{(t)}\beta_{bin})}}, \tag{7.2}$$

where, for simplicity we consider only one external covariate $X_{ij}^{(t)}$ with its own associated parameter β_{bin}, while in general nothing prevents us to have multiple covariates. With this choice, the log-likelihood for the single observation $\left(\mathbf{Y}^{(t)}\right)$ in (7.1) is

$$\begin{aligned}\log P\left(\mathbf{Y}^{(t)}|\overleftarrow{\theta}^{(t)},\overrightarrow{\theta}^{(t)},\beta_{bin},\beta_w,\mathbf{X}^{(t)}\right)=\sum_{ij}\left(\Theta\left(Y_{ij}^{(t)}\right)-1\right)\left(\overleftarrow{\theta}_i^{(t)}+\overrightarrow{\theta}_j^{(t)}+\beta_{bin}X_{ij}^{(t)}\right)\\ -\log\left(1+e^{-\left(\overleftarrow{\theta}_i^{(t)}+\overrightarrow{\theta}_j^{(t)}+\beta_{bin}X_{ij}^{(t)}\right)}\right)+\Theta\left(Y_{ij}^{(t)}\right)\log g_{ij}^{(t)},\end{aligned} \tag{7.3}$$

where the function Θ, defined in (2.1), is zero if its argument is less or equal than zero and one it's greater. As in the standard fitness model for binary networks, the time varying *binary fitness* parameters $\left(\overleftarrow{\theta}^{(t)},\overrightarrow{\theta}^{(t)}\right)$ describe the tendency of nodes at time t to form links, that are not explained by the external covariate $\mathbf{X}^{(t)}$. In order to model the weights of the observed links, we consider a generalized version of the fitness model where we associate to each node i, at time t, also the parameters $\overleftarrow{\lambda}_i^{(t)}$, $\overrightarrow{\lambda}_i^{(t)}$, that we call *weighted fitness*. They describe the propensity of a node to have more or less heavy weights in incoming and outgoing links respectively, and are related to the distribution of the weights of present links $g_{ij}^{(t)}$ by

$$E\left[Y_{ij}{}^{(t)}|Y_{ij}^{(t)}>0\right]=e^{\left(\overleftarrow{\lambda}_i^{(t)}+\overrightarrow{\lambda}_j^{(t)}+X_{ij}^{(t)}\beta_w\right)}, \tag{7.4}$$

where we considered the dependency on a single external covariate $X_{ij}^{(t)}$ and indicated the associated regression coefficient with β_w to distinguish it from the coefficient for the binary part in (7.2). This choice of linking the weighted fitness to the mean of the distribution g provides dynamics and heterogeneity only to one parameter of the conditional distribution, as shown in the following with a concrete example. Our score driven weighted fitness model can be defined for a generic distribution g, for both continuous and discrete data, as we discuss in Appendix 7.6.1.

Nevertheless in the following, for concreteness, we will focus on the gamma distribution to model links' weights

$$g_{ij}(y) = \frac{(\mu_{ij})^{-\sigma} y^{(\sigma-1)}}{\Gamma(\sigma)} e^{-\frac{y}{\mu_{ij}}}. \tag{7.5}$$

Given a sequence of observed weighted adjacency matrices $\left\{\mathbf{Y}^{(t)}\right\}_{t=1}^{T}$, we denote by $f^{(t)}$ a, K dimensional vector, where $K = 4 \times N$, containing all the time varying fitness parameters $\overleftarrow{\theta}^{(t)}$, $\overrightarrow{\theta}^{(t)}$, $\overleftarrow{\lambda}^{(t)}$, $\overrightarrow{\lambda}^{(t)}$. With this notation, the model's distribution takes the following form

$$P\left(Y_{ij}^{(t)} = y | f^{(t)}, \beta_{bin}, \beta_w, \sigma\right) = \begin{cases} \frac{e^{-(\overleftarrow{\theta}_i^{(t)} + \overrightarrow{\theta}_j^{(t)} + X_{ij}^{(t)}\beta_{bin})}}{1+e^{-(\overleftarrow{\theta}_i^{(t)} + \overrightarrow{\theta}_j^{(t)} + X_{ij}^{(t)}\beta_{bin})}} & for \quad y = 0 \\ \frac{\left(\mu_{ij}^{(t)}\right)^{-\sigma}\Gamma(\sigma)^{-1}}{1+e^{-(\overleftarrow{\theta}_i^{(t)} + \overrightarrow{\theta}_j^{(t)} + X_{ij}^{(t)}\beta_{bin})}} y^{(\sigma-1)} e^{-\frac{y}{\mu_{ij}^{(t)}}} & for \quad y > 0\,. \end{cases} \tag{7.6}$$

with

$$\mu_{ij}^{(t)} = \sigma^{-1} e^{\left(\overleftarrow{\lambda}_i^{(t)} + \overrightarrow{\lambda}_j^{(t)} + X_{ij}^{(t)}\beta_w\right)}.$$

We let the fitness, both binary and weighted, evolve in time, following the score-driven recursive update rule in (5.2), that in this case takes the form

$$f^{(t+1)} = w + b f^{(t)} + a\boldsymbol{\mathcal{I}}^{(t)} \frac{\partial \log P\left(\mathbf{Y}^{(t)} | f^{(t)}\right)}{\partial f^{(t)\prime}}, \tag{7.7}$$

where w, a and b are three K dimensional vectors of static parameters [1]. $\boldsymbol{\mathcal{I}}^{(t)}$ is a $K \times K$ scaling matrix. Hence, conditionally on the value of the parameters $f^{(t)}$ at time t and the observed adjacency matrix $\mathbf{Y}^{(t)}$, the parameters at time $t+1$ are deterministic. The element k of the score for the in and out binary fitness takes the following form

$$\frac{\partial \log P\left(\mathbf{Y} | f^{(t)}, \beta_{bin}, \beta_w, \mathbf{X}^{(t)}\right)}{\partial \overleftarrow{\theta}_k^{(t)\prime}} = \sum_j \left(\Theta\left(Y_{kj}^{(t)}\right) - \frac{1}{1 + e^{-(\overleftarrow{\theta}_k^{(t)} + \overrightarrow{\theta}_j^{(t)} + X_{kj}^{(t)}\beta_{bin})}}\right)$$

$$\frac{\partial \log P\left(\mathbf{Y} | f^{(t)}, \beta_{bin}, \beta_w, \mathbf{X}^{(t)}\right)}{\partial \overrightarrow{\theta}_k^{(t)\prime}} = \sum_i \left(\Theta\left(Y_{ik}^{(t)}\right) - \frac{1}{1 + e^{-(\overleftarrow{\theta}_i^{(t)} + \overrightarrow{\theta}_k^{(t)} + X_{ik}^{(t)}\beta_{bin})}}\right) \tag{7.8}$$

[1] Hence, in our definition we have three static parameters for each time varying fitness. While in the general formulation of score driven models (as given, for example, in Creal, Koopman, and Lucas, 2013), the parameters b and a are defined as matrices, it is very common to impose some restriction on them in order to limit the number of parameters to be estimated, as we do here.

and does not depend on the choice of g. The element k of the score for the weighted in and out fitness are

$$\frac{\partial \log P\left(\mathbf{Y}|f^{(t)},\beta_{bin},\beta_w,\mathbf{X}^{(t)}\right)}{\partial \overleftarrow{\lambda}_k^{(t)\prime}} = \sum_j \Theta\left(Y_{kj}^{(t)}\right)\frac{\partial \log g_{kj}}{\partial \overleftarrow{\lambda}_k} = \sum_j \Theta\left(Y_{kj}^{(t)}\right)\left(\frac{Y_{kj}^{(t)}}{\mu_{kj}^{(t)}}-\sigma\right)$$

$$\frac{\partial \log P\left(\mathbf{Y}|f^{(t)},\beta_{bin},\beta_w,\mathbf{X}^{(t)}\right)}{\partial \overrightarrow{\lambda}_k^{(t)\prime}} = \sum_i \Theta\left(Y_{ik}^{(t)}\right)\frac{\partial \log g_{ik}}{\partial \overrightarrow{\lambda}_k} = \sum_i \Theta\left(Y_{ik}^{(t)}\right)\left(\frac{Y_{ik}^{(t)}}{\mu_{ik}^{(t)}}-\sigma\right). \tag{7.9}$$

In (7.8) and (7.9), as scaling matrix $\boldsymbol{S^{(t)}}$ we use the Hessian of the log-likelihood.

As any score driven model, our model can be regarded both as a DGP or as a filter of a misspecified dynamics. When used as a DGP, it describes a stochastic dynamics because, at each time t, the adjacency matrix is not known in advance. It is randomly sampled from $P\left(\mathbf{Y}^{(t)}|f^{(t)}\right)$ and then used to compute the score that, as a consequence, becomes itself stochastic. When the sequence of networks $\left\{\mathbf{Y}^{(t)}\right\}_{t=1}^{T}$ is observed, and the model is applied as a filter of the time varying parameters, the static parameters (w,b,a), that best fit the data, can be estimated maximizing the log-likelihood of the whole time series.

Taking into account that each network $\mathbf{Y}^{(t)}$ is independent from all the others *conditionally* on the value of $f^{(t)}$, the log-likelihood can be written as

$$\log P\left(\left\{\mathbf{Y}^{(t)}\right\}_{t=1}^{T}|w,b,a\right) = \sum_{t=1}^{T}\log P\left(\mathbf{Y}^{(t)}|f^{(t)}\left(w,b,a,\left\{\mathbf{Y}^{(t')}\right\}_{t'=1}^{t-1}\right)\right). \tag{7.10}$$

Then the filtered time varying fitness $\widehat{f}^{(t)}$ are obtained by an iterative application of (7.7) using as parameter values the maximum likelihood estimates.

For ease of exposition, so far we introduced a baseline version of our model where the probabilities depend on a single external covariate through the scalar β_{bin} and β_w. This implies uniform, i.e. equal across all links dependency on the covariates. In order to explore node specific dependency on the external variables, we could easily consider also the specification $X_{ij}\left(\beta_{bin_i}+\beta_{bin_j}\right)$, where we associate two covariate coefficients to each node, one for outgoing links and one for incoming links.

7.3 Numerical Simulations in Misspecified Settings

In this section we discuss the results of extensive numerical simulations[2] that we run to evaluate the score driven Weighted Fitness Model as a misspecified filter, i.e. when the true DGP of the simulated data is not the same as the Score Driven filter. This is the typical situation in practical applications, where the true DGP is unknown.

[2]The python code used for the simulations is available at https://github.com/domenicodigangi/DynWGraphsPaper

When modeling a sequence of observed weighted networks with time varying fitness as in (7.6), we can consider two simple alternatives to the score driven update rule. The first one is to specify the fitness, both binary and weighted, as static. This amounts to consider a zero augmented generalized linear model, accounting for node specific effects by means of the fitness. The probability of observing a link is that of the standard fitness model (7.2), where the parameters $\left(\overleftarrow{\theta}, \overrightarrow{\theta}\right)$ are shared for all time steps $T = 1, \ldots, T$, while the distribution of the weights depends on the constant weighted fitness $\left(\overleftarrow{\lambda}, \overrightarrow{\lambda}\right)$ such that

$$E\left[Y_{ij}{}^{(t)} | Y_{ij}^{(t)} > 0\right] = e^{\left(\overleftarrow{\lambda}_i + \overrightarrow{\lambda}_j + X_{ij}^{(t)}\beta_w\right)}. \tag{7.11}$$

The second alternative consists in estimating a static fitness model for each snapshot observed. This procedure results in a sequence of single snapshot estimates for the fitness, where clearly the number of parameters to be estimated grows with the number of networks observed. This sequence of single snapshot estimates provides a filter for the time varying fitness, and has already been discussed, for the binary case, in Di Gangi, Bormetti, and Lillo, 2019, and in Chapter 6.

In the rest of this section we compare the score driven weighted fitness model, when appropriate, with these two alternatives in three numerical experiments.

7.3.1 Filter Time Varying Fitness

In our first numerical experiment we focus on filtering the time varying fitness that evolve in time following a DGP different from the score driven dynamics that we propose. As mentioned in Section 7.2, our model can be regarded as a DGP. Nevertheless, we are more inclined to interpret it as a misspecified filter and intend to validate its applicability as such. Specifically we assess how accurately the score driven methodology allows us to filter the fitness when no external covariates are present, and compare it with the sequence of single snapshot standard MLE estimates. Since the focus here in on filtering time varying fitness, we do not consider the alternative version with constant fitness. In Di Gangi, Bormetti, and Lillo, 2019, and in Chapter 6, we already carried out a similar comparison for the binary part of the model. There we showed that, for the binary part, the sequence of cross sectional estimates clearly under-performs the Score Driven fitness model, both in numerical simulations and empirical applications. Here we repeat a similar exercise considering also the weighted fitness.

Specifically, we consider sequences of networks sampled from (7.6) where each fitness, both binary and weighted, evolves according to an auto-regressive process of order one, $f_i^{(t+1)} = b_0 + b_1 f_i^{(t)} + \epsilon^{(t)}$ where $\epsilon \sim N(0, 0.1)$, $b_1 = 0.98$ and b_0 is chosen for each parameter in order to sample networks that resemble a real world realization. In practice, in order to obtain realistic parameters values for the DGPs, we first estimate the fitness models, both binary and weighted on the first observation available for the temporal interbank network data that we describe in Section 7.4.2. [3]. Given this DGP for the time varying fitness, we sample time series, of 150 observations, from the AR(1) processes for each fitness independently. We use the simulated paths for the fitness as parameters to sample from the model specification

[3] The numerical values for the fitness used are available at https://github.com/domenicodigangi/dynwgraphs. We have checked that similar results hold true using different numerical values for the fitness parameters.

in (7.6). Then, using only the sequence of sampled matrices as input we use the two methods to filter the time varying evolution of the fitness. For each fitness we compute the mean squared error (MSE) over the time series and then average it across nodes, keeping separate the MSE for binary and weighted fitness. Finally, we repeat the sampling and filtering sequence 50 times and average the obtained MSE across the repetitions. The results showed in Table 7.1 confirm that, also for the weighted

	SS Fit.	SD Fit.
Avg. MSE $\left(\overleftarrow{\theta}, \overrightarrow{\theta}\right)$	0.36	0.11
Avg. MSE $\left(\overleftarrow{\lambda}, \overrightarrow{\lambda}\right)$	0.54	0.25

TABLE 7.1: Results from the first experiment: MSE of the filtered fitness averaged across all nodes

fitness, the score driven update rule is a clear better choice in filtering misspecified paths for the time varying fitness.

Having established that as a misspecified filter of the fitness, the sequence of single snapshot estimates under-performs the score driven alternative, also for the weighted fitness, let us mention a second limitation of the single snapshot approach. In the usual empirical setting, only one network is observed in each time step. In such a setting, we might not be able to jointly estimate the sequence of single snapshots estimates and the coefficients β_{bin}, or β_w, due to the low number of observations per parameters. This is the case, for example, if we consider as covariate a variable that is uniform across all links i.e. $X_{ij}^{(t)} = x^{(t)}\ \forall i, j$, a case that we consider in the empirical application of Section 7.4.2. In this case, an identification issue arises for the single snapshot estimates for both the binary and weighted parameters. In fact, the probability of the sequence of observations, given the sequence of weighted fitness, remains unchanged under the following transformation

$$\begin{aligned} \overleftarrow{\lambda}^{(t)} &\to \overleftarrow{\lambda}^{(t)} + c_1 x^{(t)} \\ \overrightarrow{\lambda}^{(t)} &\to \overrightarrow{\lambda}^{(t)} + c_2 x^{(t)} \\ \beta_w &\to \beta_w + c_3, \end{aligned} \tag{7.12}$$

for any choice of (c_1, c_2, c_3) such that $c_1 + c_2 + c_3 = 0$, since for each t such a transformation does not change the sum $\overleftarrow{\lambda}_i^{(t)} + \overrightarrow{\lambda}_j^{(t)} + \beta_w x^{(t)}$. Hence, in a model specification with a uniform external variable, we cannot identify the parameter of interest β_w and this prevents us to use sequences of single snapshot estimators. With a simple change of notation we can see that the same issue arises for the binary parameters. We point out that the model with static fitness, that we use for comparison, and our score driven version, do not suffer from this identification issue, as in both cases the number of static parameters to be estimated does not increase with the number of time steps observed. For instance, in the score driven model, the sequence of parameters $\left(\overleftarrow{\lambda}^{(t)}, \overrightarrow{\lambda}^{(t)}\right)$ for $t = 1, \dots, T$, is not estimated directly but follows from the score driven update rule (7.7) that is uniquely identified given the sequence of observations and the static parameters (w, b, a). For this reason, in the rest of this section, we compare the score driven weighted fitness model only with the alternative having constant fitness.

7.3.2 External Covariates and Omitted Variables Misspecification

In the rest of this section we discuss two numerical experiments that focus on assessing how effective our score driven weighted fitness model can be in estimating the dependency of the network dynamics on external covariates. We show that, in synthetically generated data-sets, introducing the binary and weighted fitness, and allowing them to vary in time, reduces errors due to unobserved variables.

Our second experiment is designed to show that the score driven model allows us to estimate the effect of external covariates, even when the time varying fitness are generated by a DGP that is not score driven. Moreover it highlights the importance of taking into account time varying node specific effects. In fact, assuming the fitness to be constant, when they are actually time varying, results in poor estimates of the dependency on external covariates. In order to show that, we consider samples from the model in (7.6), where we let the fitness evolve with the AR(1) DGP of the previous experiment, and suppose that the network dynamics depends also on the realization of two independent, predetermined, external variables, $\mathbf{X}_{bin}$ and $\mathbf{X}_w$. The first covariate enters in the binary part of the DGP

$$p_{ij}^{(t)} = \frac{1}{1 + e^{-(\overleftarrow{\theta}_i^{(t)} + \overrightarrow{\theta}_j^{(t)} + X_{bin_{ij}}^{(t)} \beta_{bin})}},$$

and the second one influences the expected weights in 7.6 as follows

$$\mu_{ij}^{(t)} = \sigma^{-1} e^{\left(\overleftarrow{\lambda}_i + \overrightarrow{\lambda}_j + X_{w_{ij}}^{(t)} \beta_w\right)}.$$

We fix $\beta_{bin} = 1$, $\beta_w = 1$ and consider two possible specifications for the DGP of the synthetic external covariates. In the first one both external variables are scalar, $X_{bin_{ij}}^{(t)} = x_{bin}^{(t)}$ and $X_{w_{ij}}^{(t)} = x_w^{(t)} \forall i, j$, and follow an AR(1) process equal to the one followed by the fitness in the previous example[4]. In the second specification for the external variables, we set $X_{bin_{ij}}^{(t)} = \Theta\left(Y_{ij}^{(t-1)}\right)$ and $X_{w_{ij}}^{(t)} = \log\left(Y_{ij}^{(t-1)}\right) \forall i, j$. The latter DGP, due to the explicit dependency of the network at time t from its realization at previous time $t-1$, simulates a temporal network with link persistence, i.e. a higher probability of observing at time t a link if this was present at time $t-1$.

We sample 50 sequences of networks, and external covariates, each of 150 time steps. For each sampled sequence, we estimate our score driven weighted fitness model maximizing the likelihood of the static parameters. We then compute the MSE between simulated and estimated values of β_{bin} and β_w across the sample. In order to test the effects of completely disregarding time varying effects in estimating the dependency of the external covariates, we repeat the same procedure also for a version of the model with static fitness and report the results in Table (7.2). It emerges clearly that not taking into account the dynamics of the fitness can severely deteriorate our estimation of the effect of external covariates.

In order to motivate our third and last numerical experiment, let us mention that the topic of estimation errors due to omitted variables has been discussed widely in the econometric literature (Greene, 2000; Barreto and Howland, 2006; Wooldridge, 2010). The standard approach to mitigate it consists in using control variables. This approach has some known downsides (see for example Griliches, 1977; Yatchew and Griliches, 1985; Clarke, 2005) and, most importantly, is not always viable since the

[4]With parameter b_0 set such that the unconditional mean of the AR(1) process is equal to 1.

DGP	AR Fit. & Scalar AR Ext. Var.		AR Fit. Link persistence	
Filter	Const. Fit.	SD-Fit.	Const. Fit.	SD-Fit.
MSE β_{bin}	0.14	0.06	0.23	0.02
MSE β_w	0.34	0.02	0.18	0.015

TABLE 7.2: Results for the second experiment: MSE of the estimated regression coefficients over 50 replicates of the numerical experiment. In both DGPs the fitness follow an AR(1) process. Two columns on the left: results from the DGP with scalar external regressors following two AR(1) processes. Two columns on the right: results for the DGP that simulates a persistent dynamics in both links and weights.

data on appropriate controls might not be available for a number of reasons. Indeed, this is very common for financial networks, as the one that we consider in Section 7.4.2, where the variables that one would like to use as controls are likely to be privacy protected and often unavailable to researchers. Considering, for example the case of interbank networks, we could very well expect the current leverage of a bank to have a significant influence on the decision to borrow or lend, i.e. create interbank links. Nevertheless it is very unlikely for this information to be available, at the frequency required. This issue is even more frequent when the data-sets are anonymous, and the identity of the nodes is not known. For this reasons, here we present a numerical experiment with the narrow objective of motivating the application of time varying fitness when estimating the effect of external covariates on temporal networks. In short, in our third experiment we show that allowing the fitness to vary in time is extremely beneficial to mitigate the errors that arise due to omitted variables, at least in the context of controlled numerical simulations. We assume the network's dynamics, both that of the links and that of the weights, to be determined by two external covariates. We then assume that, for whatever reason, only one of them is observed. Specifically the DGP considered is similar to the one in (7.6) but with parameters defined as

$$p_{ij}^{(t)} = \frac{1}{1 + e^{-(\beta_{1,bin} x_1^{(t)} + \beta_{2,bin} x_2^{(t)})}},$$

and

$$\mu_{ij}^{(t)} = \sigma^{-1} e^{\left(\beta_{1,w} x_1^{(t)} + \beta_{2,w} x_2^{(t)}\right)}.$$

We assume that the $x_2^{(t)}$ is not observed and assess the effect of disregarding it in estimating the effect of the available one. Furthermore, we show that introducing the score driven fitness, in this controlled numerical experiment, compensates the impact of the neglected variable on the estimates. In this setting, as before, we consider external variables that follow an AR(1) model with high persistence, $b_1 = 0.98$. We sample the DGP and estimate the coefficients 50 times for a sequence of networks long 150 time steps. We then compare the MSE for the estimates of parameters $\beta_{1,bin}$ and $\beta_{1,w}$, obtained using three different specifications of the model in (7.6). One without node heterogeneity, hence no fitness, where the probability of observing a link depends only on the observed covariate, one with constant fitness, and our model with score driven fitness, also depending on the observed external covariate. From the results that we report in Table (7.3), it follows that considering a model with time varying fitness is extremely beneficial even when the DGP is not characterized

DGP	No Fit. 2 regressors		
Filter 1 Reg.	No Fit.	Const.Fit.	SD-Fit.
MSE β_{bin_1}	5.66	0.53	0.01
MSE β_{w_1}	133	0.39	0.08

TABLE 7.3: Misspecified filtering of a DGP with two covariates and no fitness. Results from our third experiment: MSE of the estimated regression coefficients for the observed variable. The DGP has no fitness and depends on two covariates each following an AR(1) process. The first column (No Fit.) shows the MSE for the estimates of β_{bin} and β_w, when no fitness is used. The second column (Const. Fit.) when the estimates are obtained using a model with constant fitness. In the third column (SD Fit.) we show the same results using a model with score driven fitness.

by explicit node specific effects, i.e. the DGP itself depends only on the realizations of the external variables and does not include node specific fitness. We believe that the results presented in this section strongly support our choice to describe temporal weighted networks by means of time varying fitness. Moreover, they give us clear insights to interpret the results of Section 7.4, where we find a clear advantage, in terms of goodness of fit, in using our score driven weighted fitness model to describe empirical data.

7.4 Link and Weight Dynamics in the Italian e-MID

We apply our model to the interbank overnight loans market described as a temporal weighted network. Inter-bank markets are an important point of encounter for banks' supply and demand of extra liquidity, and have received much attention in the literature (see Green et al., 2016, for a review). In particular, e-MID has been investigated in many papers (see, for example Iori et al., 2008; Finger, Fricke, and Lux, 2013; Mazzarisi, Barucca, Lillo, and Tantari, 2017; Barucca and Lillo, 2018, and references therein). We use data from the e-MID, a market where banks can extend loans to one another for a specified term and/or collateral. Our data-set contains the list of all credit transactions in each day from June 6, 2009 to February 27, 2015. In our analysis, we investigate the interbank network of overnight loans, aggregated weekly. The standard approach in the literature to model temporal interbank networks is to disregard the size of the exposures and consider only the presence or absence of links, i.e. consider only the binary network. Thanks to the flexibility of our score driven weighted fitness model, we are able to take into account and explicitly model also the weights of the loans. We thus consider a link from bank j to bank i present at week t if bank j lent money overnight to bank i, and its weight as the total amount lent over that period. This results in a set of $T = 298$ weekly aggregated weighted networks. For a detailed description of the data-set, we refer the reader to Barucca and Lillo, 2018.

As it is evident from the left panel of Figure 7.1, the number of links in e-MID is significantly lower in the second half of the dataset. In particular it started declining in 2011, most likely as a consequence of the european sovereign debt crisis and then fluctuated around a new lower level since the beginning of 2012. As discussed in Barucca and Lillo, 2018, the decreased number of links corresponds to a

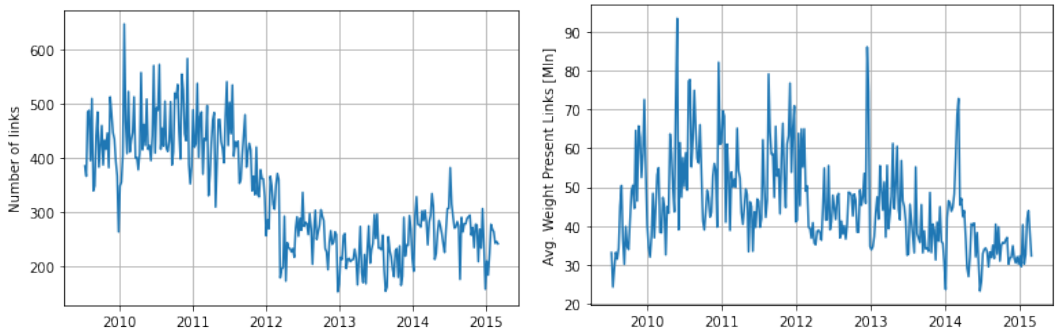

FIGURE 7.1: In the left panel we show the number of links present in the data for each time step. In the right panel we show the average weight, in Millions of Euro, of the present links.

lower number of banks being active in the market and both the network density[5] and the average weight of present links has not followed a similar clear transition to a different level.

7.4.1 Weight Prediction Exercise

As a first empirical application of our model we explore the possibility of using it to predict the weight of future links. The problem of link prediction in binary temporal networks is extremely relevant in practical applications and has been discussed widely in the literature (Lü and Zhou, 2011; Wang, Xu, Wu, and Zhou, 2015; Martínez, Berzal, and Cubero, 2016; Haghani and Keyvanpour, 2017). It can be defined in multiple ways depending on the context, the type of data at hand, and whether we want to predict the presence of a link in the future or the existence of a link in a partially observed network. Weight forecasting in temporal networks has instead received much less attention so far. Among the works mentioned in section 4.3, that propose models for weighted temporal networks, only Giraitis, Kapetanios, Wetherilt, and Žikeš, 2016 discuss a forecasting application. Moreover, they run a forecasting exercise on a very dense and small network[6]. For this reason here we run an exercise using the full dataset described in the previous section. We focus on forecasting the weights of the network at time $t+1$, using only information available at time t. This is very much in line with what we did in Di Gangi, Bormetti, and Lillo, 2019, and in Chapter 6, with the important difference that here we consider the prediction of the weights instead of the prediction of links' presence. Since our focus is on weights prediction, we consider only the links that are actually present at time $t+1$ and do not discuss the prediction of the existence of a link, that we considered in our previous work. Moreover, all the out of sample measures of goodness of fit for the weights considered in the rest of this Chapter are computed on the subset of links observed at time $t+1$.

In our weight forecasting exercise we compare the forecasts obtained using two methods, that we describe in the following. In both approaches, when forecasting

[5]Number of links present divided by the number of possible links, given the number of active banks.

[6]This is due to the structure of the UK interbank market that they consider. Indeed the fraction of links present at every time step for the full temporal network that they consider is always above 40%. Furthermore, they run the link forecasting only on the sub-network composed by the 4 largest banks in the system, thus increasing the density of the actual network considered.

observations at time $t+1$ we use only observations from $t - T_{train}$ to t for training[7]. The first approach that we consider uses our score driven model to forecast the fitness at time $t+1$ with the update rule (7.7). This is very easy to do in practice since, as mentioned in Section 7.2, the score driven fitness at time $t+1$ are deterministic conditionally on the observations at time t. The second approach that we consider is a combination of the sequence of single snapshot estimates, described at the beginning of Section 7.3, and a set of AR(1) models, one for each fitness. In the latter approach, we first obtain the sequence of single snapshot estimates on the training set and then use them to estimate an AR(1) process for each fitness. Finally, we use this set of AR(1) models to forecast one value of the fitness at time $t+1$. Practically, we repeatedly estimate the two models on rolling windows of length $T_{train} = 100$ time steps and, once for each estimate, we forecast the first out of sample observation for each train window. We then compare the weights of present link at time $t+1$ with the expected values obtained from the two models and quantify the error by the mean squared error between the logarithms of observed and predicted weights

$$\text{MSE Log.} = \frac{\sum_{ij} \Theta\left(Y_{ij}^{(t+1)}\right)\left(\log\left(E\left[Y_{ij}|\overleftarrow{\lambda}^{(t+1)}, \overleftarrow{\lambda}^{(t+1)}\right]\right) - \log\left(Y_{ij}^{(t+1)}\right)\right)^2}{\sum_{ij} \Theta\left(Y_{ij}^{(t+1)}\right)},$$

where $\overleftarrow{\lambda}^{(t+1)}$ and $\overleftarrow{\lambda}^{(t+1)}$ are the forecasts for the in and out weighted fitness, obtained using only observations up to time t. Similarly we compute the mean absolute difference (MAD)

$$\text{MAD Log.} = \frac{\sum_{ij} \Theta\left(Y_{ij}^{(t+1)}\right)\left|\log\left(E\left[Y_{ij}|\overleftarrow{\lambda}^{(t+1)}, \overleftarrow{\lambda}^{(t+1)}\right]\right) - \log\left(Y_{ij}^{(t+1)}\right)\right|}{\sum_{ij} \Theta\left(Y_{ij}^{(t+1)}\right)}.$$

We compare the logarithms of predicted and observed weights because the distribution of observed weights is quite heterogeneous, it roughly spans 5 order of magnitudes, and directly comparing the weights would result in measures of goodness of fit mainly describing the fit of the largest weights [8].

Method	Score Driven	AR(1) Single Snap.	Diebold-Mariano (p-value)
MSE Log.	0.859	0.882	1.73×10^{-7}
MAD Log.	0.726	0.737	1.21×10^{-7}

TABLE 7.4: Results of the weight prediction exercise. In the table we show the mean square error and mean absolute difference between the logarithm of observed weights and that of the weights predicted with the two methods discussed. In the first column the results of our score driven approach, in the second column those of the forecasting approach based on single snapshot estimates and AR(1) processes for the fitness. In the third column we show the p-value of a Diebold-Mariano test for the null hypothesis that the two forecasts are equivalent.

[7] Extremes included.

[8] Similar results hold using measures of relative error for the comparison between predicted and observed weights.

From the results reported in Table 7.4 we deduce that, similarly to the binary case, the score driven dynamical evolution is a better choice to predict the weights with respect to a prediction based on sequence of single snapshot estimates, both in terms of MSE and MAD for the logarithms. Moreover we run a Diebold-Mariano (Diebold and Mariano, 2002; Harvey, Leybourne, and Newbold, 1997) test that rejects the hypothesis that the two forecasts are statistically equivalent.

7.4.2 The Effect of Interest Rates on Interbank Lending

In this section we discuss an application of our model to investigate the effect of the interest rates on the dynamics of the interbank network data introduced above. To track average interest rates, we use the Euro Overnight Index Average (EONIA) benchmark. EONIA *is a measure of the effective interest rate prevailing in the euro interbank overnight market. It is computed as a weighted average of the interest rates on unsecured overnight contracts on deposits denominated in euro, as reported by a panel of contributing banks.*[9]

Intuitively we can expect banks' funding rates and the topology of the interbank market to be deeply related. This relation is of clear interest from the point of view of the policymaker and has received much attention in the literature (see, for example, Akram and Christophersen, 2010; Iori, Burcu, and Olmo, 2015; Arciero et al., 2016; Temizsoy, Iori, and Montes-Rojas, 2017; Brunetti, Harris, Mankad, and Michailidis, 2019). Of particular relevance for the results discussed in this section is the work of Akram and Christophersen, 2010 that investigated the effects of banks characteristics and the conditions of the network as a whole on the interest rates that each bank faces on the interbank market. They exploited a remarkable dataset, attained from Norges Bank real time gross settlement system, that allowed them to model bank specific interest rates as dependent on a set of variables and controls, including overall market's liquidity. From basic supply and demand reasoning one would expect that excess liquidity in the market would have a negative pressure on the interest rates on average, and indeed they found that interest rates tend to be lower when the overall liquidity available on the market is higher.

A second work that is relevant for the purposes of this section is that of Brunetti, Harris, Mankad, and Michailidis, 2019, where the authors considered data on the e-Mid interbank market for a period ranging from the beginning of 2006 to the end of 2012, focusing solely on the binary part of the daily temporal network of overnight loans. They computed various aggregated network statistics for each time step, thus obtaining one univariate time series for each statistic. Among other quantities, they computed the density of each network[10] and, using a standard linear regression, found it to be positively related with EONIA.

In the following we explore the impact of interest rates on the probability of observing each link and on the expected weight of observed links by applying our score driven weighted fitness model to the e-Mid dataset, using as external variable the EONIA interest rate. For our estimates, we use a training set comprising the first 80% of time steps and left the last 20% to assess goodness of fit out of sample. Moreover, similarly to what done in the numerical simulations discussed Section 7.3, we compare the score driven dynamical fitness with two alternative specifications: a model without fitness and one with constant fitness, as defined in Section 7.3, both with EONIA as the only external variable. We then compare their goodness of fit both in-sample and out of sample. In Table 7.5 we show the results that clearly

[9] Definition from `https://stats.oecd.org/`.

[10] Defined as the number of connections as a proportion of all possible connections

confirm the importance of including time varying fitness to improve goodness of fit. We report the Bayesian Information Criterion for each model, computed separately for the likelihood of observing links and the likelihood of their weights, to compare goodness of fit in-sample. The binary and weighted parts of each model are evaluated separately out of sample. We quantify out of sample accuracy in predicting link's presence by means of the AUC, while for the weights we compute the MSE of the logs of the weights, only for the present links. Since the model without fitness

Model	No. Fit.	Const. Fit.	SD-Fit.
BIC Bin	1.75×10^6	0.53×10^6	0.45×10^6
BIC Weight	1.898×10^6	1.460×10^6	1.459×10^6
AUC - Test Set	0.48	0.82	0.92
MSE Log. - Test Set	58.15	1.01	0.776

TABLE 7.5: EONIA Effect on e-Mid. We report the in sample Bayesian Information Criterion (BIC), for both the binary and weighted part of model 7.6, the area under the curve (AUC) for our of sample evaluation of the binary part and the MSE of the logarithms for the out of sample evaluation of the weighted part. The first column describes the results for a model without any fitness. The second one for the case of constant fitness and the third one for the case of score driven fitness.

is clearly not a good fit for the data we do not discuss it further, and in Table 7.6 we report the estimated regression coefficients using the models with constant and score driven fitness. The model with score driven fitness is clearly the best fit for the

Model	Const. Fit.	SD-Fit.
β_{bin}	0.69 ± 0.06	0.29 ± 0.05
β_w	0.022 ± 0.029	-0.13 ± 0.02

TABLE 7.6: EONIA Effect on e-Mid. Estimates of the regression coefficients using models with constant or score driven fitness.

data, both in sample and out of sample, as measured by the measures reported in Table 7.5. Moreover, the parameters estimated by the score driven fitness model have always a statistically significant difference with respect to those estimated using a model with constant fitness. We interpret this discrepancy as a sign that disregarding the fitness dynamics can lead us also to miss-guided qualitative interpretations, consistently with our numerical results of section 7.3.2. For example, in this case, the β_w parameter estimated using a model with constant beta is not significantly different from zero.

From the estimate $\beta_{bin} = 0.29 \pm 0.05$, we can deduce that, in the considered period, the probability of observing a link in the network is positively related with the interest rates, hence the lowering of interest rates tends to reduce the overall market interconnectdness, even taking into account bank specific effects captured by the fitness. This result is coherent with the relation between network density and EONIA found in Brunetti, Harris, Mankad, and Michailidis, 2019, although the approach based on standard regression on aggregated network statistics is different from ours. In fact, thanks to the time varying binary fitness in our model, the estimated effect of EONIA is decoupled from bank specific effects that are instead accounted for by

the fitness. Such a separation of bank specific effects from the impact of a covariate is instead not possible when considering the density of the whole network as done in Brunetti, Harris, Mankad, and Michailidis, 2019.

For what concerns the effect on the liquidity exchanged trough the observed links, i.e. the link's weights, the estimated $\beta_w = -0.13 \pm 0.02$, indicates that the weight of the observed overnight loans is negatively related with the average interest rate in the market. Our result is coherent with the work of Akram and Christophersen, 2010 on the Norwegian interbank market, but our methodology allows us to explore a different aspect of the relation between liquidity and interest rates. In fact, their result regards the relation between bank specific rates and aggregated liquidity, while we explore the relation between average rates on the market and the weight of present links, controlling for time varying bank specific effects by means of the time varying fitness. Thanks to zero augmentation and our separate modelling of links and weights, our finding is directly related with the average magnitude of the overnight loans that are actually present, more than with the total liquidity in the market. In summary, the data considered indicate that lower interest rates are related with a reduction of network interconnectdness but an increase of the average liquidity flow for the loans that are present.

We point out that our results on the relation between the dynamics of the e-Mid interbank network and EONIA, are obtained leveraging the full information available in the description of a temporal network as a temporal sequence of matrices, and considering the impact of the covariate both on the probability of each individual link and the expected weight of observed links. Differently from Akram and Christophersen, 2010 and Brunetti, Harris, Mankad, and Michailidis, 2019 we do not need to collapse the matrices into a single network statistic to estimate the effects of external variables. We directly use matrix valued network data and, thanks to the time varying latent fitness parameters, we can decouple the impact of EONIA from unobserved time varying node specific effects. The advantage of using matrix valued network data will become even more evident in the next section where we consider link specific covariates and carry out an analysis that would be completely impossible with standard regression methods on univariate network statistics.

7.4.3 Link and Weight Persistence

As a final application, we use our model to contribute to the literature on the persistence in interbank networks (Weisbuch, Kirman, and Herreiner, 2000; Cocco, Gomes, and Martins, 2009; Hatzopoulos et al., 2015; Mazzarisi, Barucca, Lillo, and Tantari, 2017) by exploring both the persistence of links and that of the weights. The existence of privileged lending relations between pairs of banks is a well known phenomenon and it is often referred to as *preferential trading* (Weisbuch, Kirman, and Herreiner, 2000). The motivations behind it can be explained by the relevance of strong lending relationships between banks as a way to overcome monitoring of credit worthiness and limit the risk of counter-party default (Cocco, Gomes, and Martins, 2009). The existence of preferential trading behaviours has been assessed quantitatively by means of statistical methods specifically developed for the purpose (Hatzopoulos et al., 2015). Additionally, models for binary temporal networks have been proposed that explicitly take it into account (Mazzarisi, Barucca, Lillo, and Tantari, 2017).

In this section we exploit the flexibility of our model and estimate the effect of two predetermined covariates that are meant to capture the persistence of links and weights. For what concerns link persistence of the binary network, we explore how

the presence of a link at time $t-1$ influences the probability of observing a link at time t. That amounts to use $\Theta\left(Y_{ij}^{(t-1)}\right)$ as external covariate in Eq. (7.2). To assess persistence in the links' weights, we estimate the effect of the weight of a link at $t-1$ in determining its weight at time t by using $\log\left(Y_{ij}^{(t-1)}\right)$ as covariate in Eq. (7.4). Let us recall that we have tested numerically the possibility to estimate such effects in synthetically generated data in Section 7.3. As in the previous section, we

Model	No. Fit.	Const. Fit.	SD-Fit.
β_{bin}	0.0644 ± 0.0385	2.875 ± 0.045	2.0478 ± 0.0448
β_w	1.05 ± 0.003	0.073 ± 0.0019	0.0637 ± 0.0018
BIC Bin	5.66×10^6	4.46×10^6	4.16×10^6
BIC Weight	2.61×10^6	1.451×10^6	1.454×10^6
AUC - Test Set	0.748	0.882	0.932
MSE Log. - Test Set	21.22	0.9	0.764

TABLE 7.7: Results on estimates of link persistence in e-Mid. One column for each one of the three alternative model specifications considered. In the third and fourth rows we show the in sample BIC, respectively for the binary and weighted parts of the model in (7.6). The last two rows are out of sample measures of goodness of fit. The fifth one is the out of sample AUC for the binary part. The last row is the out of sample MSE of the logarithms of the weights.

compare three models, a model without fitness, one with constant fitness, and our score driven fitness model, all using the same external covariates and two scalar coefficients, β_{bin} and β_w that quantify the persistence of links and weights respectively. The results in Table 7.7 confirm that neglecting node specific time varying effects results in worst fitting of the data, as is evident by looking at the superior performances, both in sample and out of sample of the models with score driven fitness, with respect to those without or with constant fitness. The three model specifications all result in positive coefficients both for the binary and the weighted covariates. With the best performing model among those three, the model with score driven fitness, we estimate $\beta_{bin} = 2.0478 \pm 0.0448$. This indicates that globally the presence of a link at time $t-1$ positively impacts the probability of observing that same link at time t. This is in agreement with the current consensus in the literature, indicating for the existence of preferential trading behaviours in e-Mid, that has been validated empirically only on the binary part of temporal interbank networks, for example by Hatzopoulos et al., 2015 and Mazzarisi, Barucca, Lillo, and Tantari, 2020. The novel aspect of our analysis lays in the estimated $\beta_w = 0.0637 \pm 0.0018$, that highlights a weights persistence effect. This result complements the analysis of Hatzopoulos et al., 2015, as they considered the weighted networks of the number of loans between each pair of banks, neglecting altogether the amount lent for each loan. By design, our model allows us to highlight the tendency of banks to form links whose weight is positively related with their weights at previous steps. A tendency that we might refer to as weights' persistence.

7.5 Conclusions

In this work, we proposed a model for the description of sparse weighted temporal networks that extends the well known fitness model for static binary networks. In our new score driven weighted fitness model, we model also links weights with an additional set of fitness. Both binary and weighted fitness have a stochastic dynamics driven by the score of the conditional likelihood. Additionally, we considered also the possibility for the network dynamics to depend on a set of external covariates.

Our numerical simulations proved the advantages of the score driven fitness over static fitness and over a sequence of standard cross sectional estimates. As an empirical application, we used our model to explore the determinants of the dynamics of links and weights in the e-MID interbank network. We proved that there is a significant advantage in using score driven parameters to forecast weights, with respect to single snapshot estimates. Moreover we used the flexibility of our model to estimate the impact of the EONIA rate in determining the links and weights dynamics. And used it to inform the discussion on persistence in interbank networks with a first empirical evidence of persistence in the weights.

Our work opens a number of possibilities for future research. First, the possibility to jointly model and predict links' presence and their weights could find relevant applications in the financial stability literature, since weighted financial networks are known to be among the determinants of Systemic Risk, and their dynamical description has so far neglected the weights of the links. Second, score driven weighted fitness model could be applied on multiple instances of real world sparse weighted temporal networks, where the standard approach of neglecting the weights might result in significant information loss. The last future application that we plan to explore is the possibility to use as external covariate a variable related to whether two nodes belong to the same group or community. Community detection has attracted an enormous amount of attention in various streams of literature (Javed et al., 2018), also in the context of temporal networks (Rossetti and Cazabet, 2018). We believe that the dynamical fitness could offer valuable insight in whether a given partition of the nodes into groups is more or less valid across nodes and time.

7.6 Appendix

7.6.1 Fitness Identification

It is easy to see that, already in its static version, the fitness of the fitness model for a directed network are not identified. This fact is well known for the binary fitness model (Yan, Leng, Zhu, et al., 2016) and remains true for the weighted fitness that we consider here. For this reason, let us indicate here with $\overleftarrow{\varphi}$ and $\overrightarrow{\varphi}$ the vectors of in and out fitness respectively, without specifying whether they are binary or weighted. With this notation it is easy to see that the following transformation

$$\overleftarrow{\varphi} \to \overleftarrow{\varphi} + c \tag{7.13}$$
$$\overrightarrow{\varphi} \to \overrightarrow{\varphi} - c \tag{7.14}$$

does not change the sum $\overleftarrow{\varphi}_i + \overrightarrow{\varphi}_j + \beta X_{ij}$ and leaves the PMF unchanged, $\forall c$. The parameters are not uniquely identified and, in order to compare estimates across different time steps, we need to impose an identification restriction, as commonly done in the literature (Yan, Leng, Zhu, et al., 2016; Mazzarisi, Barucca, Lillo, and

Tantari, 2020). In this work we require that

$$\sum_i \overleftarrow{\varphi}_i = \sum_j \overrightarrow{\varphi}_j, \tag{7.15}$$

at every time step.

7.6.2 Alternative Distributions for the Weights

As mentioned in the main text, we considered the gamma distribution to model the weights for concreteness but it is extremely easy to consider different distributions to model the weights. If, for example, we consider the log-normal distribution, we can simply substitute its parametric form

$$g_{ij}(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{\frac{-\left(\ln y - \mu_{ij}\right)^2}{2\sigma^2}} \tag{7.16}$$

in place of the gamma distribution, and obtain as a consequence a different update equation for the time varying fitness, as they would be driven by the score of the log-normal

$$s_i\left(\mathbf{Y}^{(t)}, \overleftarrow{\lambda}, \overrightarrow{\lambda}, \sigma\right) = \sum_j \frac{1}{\sigma^2} \log \frac{E\left[y_{ij} | y_{ij} > 0\right]}{y_{ij}} - K_i. \tag{7.17}$$

Indeed, extending our approach to use a different distribution is also practically straightforward. The interested reader would only need to add the functions to compute the log-likelihood, sample from the distribution and compute its derivatives and scaling matrix. Conveniently the log-likelihood and sampling methods are already available for many distributions in PyTorch (Paszke et al., 2019), upon which we built the core library for this work. `https://github.com/domenicodigangi/dynwgraphs`.

7.6.3 Filtered Fitness Dynamics With External Covariates

In this section we inspect the dynamical behaviour and the role of the fitness filtered with our score driven weighted fitness model, defined in Section 7.2, when we consider the effects of external covariates. In particular, we discuss how the dynamical evolution of the filtered fitness changes when we consider the dependency on EONIA, estimated as described in Section 7.4.2, with respect to their behaviour when no external covariate is considered.

To this end, for a subset of the most active links[11], we compute the sum of the in and out fitness corresponding to their probabilities, and expected weights. For example, for link (i, j) we consider $\overleftarrow{\theta}_i + \overrightarrow{\theta}_j$ and $\overleftarrow{\lambda}_i + \overrightarrow{\lambda}_j$. We compute these quantities using two sets of filtered fitness. The first set is obtained by estimating and running our score driven weighted fitness model as a filter without considering any external covariate, similarly to what we did in Section 7.4. The second set is obtained with a similar approach, now including in the model specification also the dependency on EONIA, exactly as in Section 7.4.2. Then, given these two sets of fitness sums we look at their Spearman rank correlation (Spearman, 1961), over time, with EONIA, thus obtaining four correlations for each of the considered links, for each of the two sets of filtered fitness (with and without EONIA as covariate) we have one

[11] We consider the links that are present at least 5% of the times.

correlation for the binary fitness and one for the weighted fitness. We consider the sum of fitness because, as is evident from (7.2) and (7.4), they are related to the probability of observing a link, or to the expected weight of that link. Moreover, given that the EONIA rate had a substantial drop around the middle of 2012, as showed in Figure 7.2 in computing the correlation with the sums of fitness we split the data set in two periods and consider them separately.

FIGURE 7.2: Weekly average of EONIA rate over the time period considered.

The plots shown in Figure 7.3 highlight the different behaviour of the fitness, with respect to EONIA when the latter is explicitly considered as external covariate. We notice that in the first period, i.e. when EONIA rates are higher, the correlation between fitness sums and EONIA are significantly different when we include EONIA as covariate or we do not. To statistically confirm the difference we run a Kolmogorow-Smirnov test having as null hypothesis the fact that the two distributions are identical. In the first period the distributions result different for both the binary and weighted part, considering a confidence level of 5% to reject the null hypothesis. In the second period instead, we cannot reject the hypotheses that the two distributions are different. Hence the relation of the fitness with EONIA in the second period does not seem to be affected by whether the latter is explicitly included or not as external covariate.

In order to better understand this behaviour we repeated the same analysis of correlation artificially modifying the regression coefficients before filtering the fitness. We keep everything equal except the values of β_{bin} and β_w. Instead of their MLE values, we now artificially set the values of these parameters to be $\beta_{bin} = 3$ and $\beta_w = 3$ in Figure 7.5 and $\beta_{bin} = -3$ and $\beta_w = -3$ in Figure 7.4. In practice, we change the regression coefficients and then filter the time varying fitness using the score driven update rule. We notice that, when the regression coefficient is artificially inflated , Figure 7.5, the fitness sums tend to negatively correlate with EONIA. While, when we set the coefficients to be artificially negative, the fitness positively correlate with EONIA. In both cases the filtered fitness behaviour tends to mitigate the impact of the artificial regression coefficients. These effect is more evident in the first period than in the second, as expected due to the higher values of EONIA in the first period.

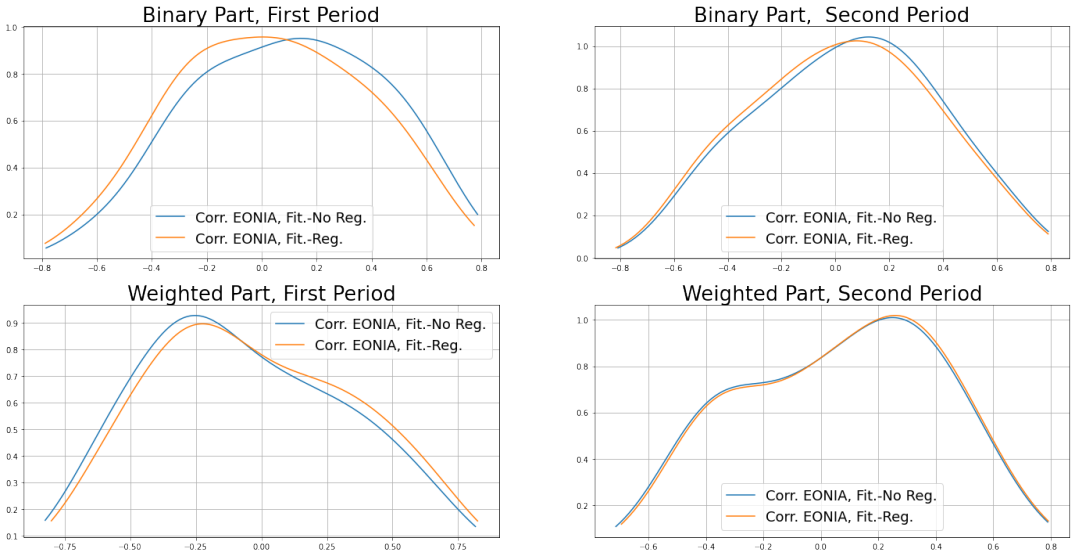

FIGURE 7.3: In the above plots we inspect the correlation between the sums of pairs of fitness, associated to a subset of the links, and the external covariate EONIA. Specifically we show the distribution of the Spearman correlation between the fitness sum and EONIA computed in two periods. The first one ranges from the beginning of the sample up to 15/07/2012, the second one from 22/07/2012 to the end of the sample. The correlations computed for the first period are showed on the left panels, those computed in the last period on the right. In the top panels we show the correlations between EONIA and the sums of binary fitness, while in the bottom one we consider the weighted fitness.

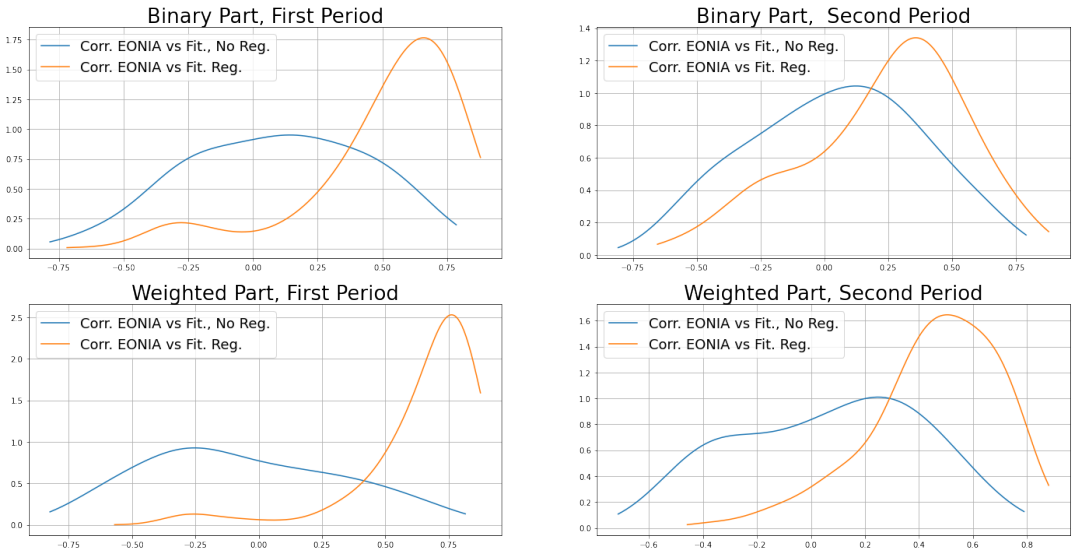

FIGURE 7.4: Similar plots as in Figure 7.3, now repeated after substituting artificial values for the regression coefficients $\beta_{bin} = -3$ and $\beta_w = -3$.

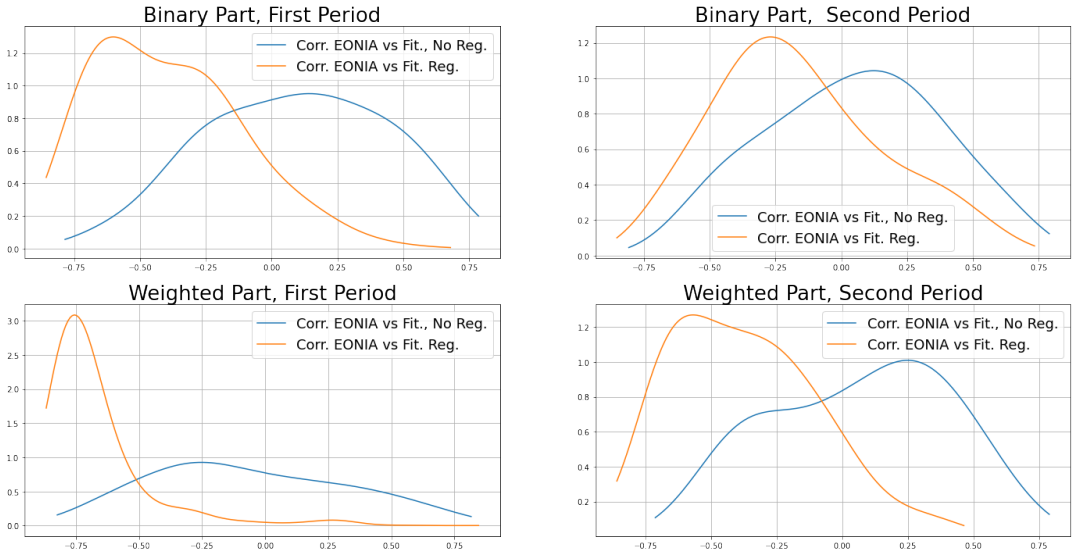

FIGURE 7.5: Similar plots as in Figure 7.3, now repeated after substituting artificial values for the regression coefficients $\beta_{bin} = 3$ and $\beta_w = 3$.

Chapter 8

Score Driven Kinetic Ising Model

Synopsis of the Chapter A common issue when analyzing real-world complex systems is that the interactions between the elements often change over time: this makes it difficult to find optimal models that describe this evolution and that can be estimated from data, particularly when the driving mechanisms are not known. Here we offer a new perspective on the development of models for time-varying interactions introducing a generalization of the well-known Kinetic Ising Model (KIM), a minimalistic pairwise constant interactions model which has found applications in multiple scientific disciplines. Keeping arbitrary choices of dynamics to a minimum and seeking information theoretical optimality, the Score-Driven methodology lets us significantly increase the knowledge that can be extracted from data using the simple KIM. In particular, we first identify a parameter whose value at a given time can be directly associated with the local predictability of the dynamics. Then we introduce a method to dynamically learn the value of such parameter from the data, without the need of specifying parametrically its dynamics. Finally, we extend our framework to disentangle different sources (e.g. endogenous vs exogenous) of predictability in real time. We apply our methodology to several complex systems including financial markets, temporal (social) networks, and neuronal populations. Our results show that the Score-Driven KIM produces insightful descriptions of the systems, allowing to predict forecasting accuracy in real time as well as to separate different components of the dynamics. This provides a significant methodological improvement for data analysis in a wide range of disciplines.

Almost all the contents of this chapter previously appeared in (Campajola, Di Gangi, Lillo, and Tantari, 2020).

8.1 Introdution

Complex systems, characterized by a large number of simple components that interact with each other in a non-linear way, have been an increasingly important field of study over the last decades. Interactions make the whole more than the sum of its parts (Bar-Yam, 2002): for this reason the effort when modeling complex systems is ultimately directed to understand how interactions arise, how to parametrize them into quantitative models and how to estimate them from empirical measurements.

One complication that is ubiquitous to real complex systems, but very rarely considered in modeling, is that interactions change over time: traders in financial markets continuously adapt their strategic decision-making to each other's actions (Challet, Chicheportiche, Lallouache, and Kassibrakis, 2016) and to new information (Lillo et al., 2015); preys change their behavior to avoid predators (Schmitz, 2017); neurons reinforce (or inhibit) connections in response to stimuli (Tavoni et al., 2017). As we show also below, a modeling approach assuming that all the interactions are

constant can be misleading, sometimes leading to spurious estimations of the interactions, which can be avoided only with very strong limitations to sample selection and experimental design (when possible).

In this chapter we propose a novel approach to the development of models for time-varying interactions based on the generalization of a minimalistic constant-interactions model, which is commonly used in many scientific disciplines, the Kinetic Ising Model (Crisanti and Sompolinsky, 1988). In the following we show that this generalization allows to describe conditions where the predictability of the overall dynamics of the observed process is variable, while commonly employed constant interaction models fail in this respect. More importantly, our modeling approach does not assume that the causes or the dynamics of the variable interactions are known, but they are estimated (or filtered) from the data themselves. Thus, different types of time-varying interactions can be present in the investigated system, including non-stationarities of various form (regime-shift, seasonalities, etc.). Indeed it often occurs that the modeler has no insight on the nature of the underlying dynamics of interactions: the dynamics that is given to the time-varying parameters then needs to be as agnostic as possible with respect to the actual generating dynamics, *i.e.* be robust to model misspecification errors.

Since it is generally difficult to determine why and how interactions change over time, it is even riskier to try to model their dynamics with specific external drivers. Conversely we assume a generic Markovian paradigm in which $J_{ij}^{(t)}$ - representing the interaction between the system's variables $s_i^{(t)}$ and $s_j^{(t)}$ at time t - endogenously adapt to the observations of $s^{(t)}$ themselves, i.e.

$$J_{ij}^{(t+1)} = F(\mathbf{J}^{(t)}, s^{(t)}). \tag{8.1}$$

The updating functional F is determined by general assumptions based on information theory principles. First of all, one can assume the interactions variation depends on surprise: the more an observation of the system's state is "unexpected", the more the relations between its components will change. In social systems, for example, friendship relationships can get damaged if not constantly fed or may arise from unexpected gestures of openness. This is also a common principle in biological learning processes and artificial neural networks, where the least expected inputs have the largest impact on the values of the synapses or inter-units weights (Ackley, Hinton, and Sejnowski, 1985). The most widespread measure of surprise is minus the logarithm of the conditional likelihood $p(s^{(t)}|J^{(t)})$ of observing a given pattern with the current level of interactions. As a second principle we assume that the system's reaction to surprise is to adapt to it, making what has been unexpected for that moment, expected for the future. In this sense the interactions change to increase the log likelihood of the last observation i.e.

$$\mathbf{J}^{(t+1)} = w + \mathbf{b}\mathbf{J}^{(t)} + \mathbf{a}^{(t)} \frac{\partial \log p(s^{(t)}|\mathbf{J}^{(t)})}{\partial \mathbf{J}^{(t)}}, \tag{8.2}$$

which can be interpreted as the updating rule of an autoregressive process with a gradient ascent perturbation with given learning rate parameter $A^{(t)}$, which possibly depends on time. This type of observation-driven (Cox et al., 1981) dynamics has been recently introduced (Creal, Koopman, and Lucas, 2013; Harvey, 2013b) in defining the class of *score-driven models*. These have been shown to be an optimal

choice among observation-driven models when minimizing the Kullback-Leibler divergence to an unknown generating probability distribution (Blasques, Koopman, and Lucas, 2015) and have risen in popularity in econometrics (Bernardi and Catania, 2019) as well as network science (Di Gangi, Bormetti, and Lillo, 2019).

The focus of the Chapter is the score-driven generalization of the KIM (Derrida, Gardner, and Zippelius, 1987; Crisanti and Sompolinsky, 1988), which is the dynamical counterpart of the celebrated Ising spin glass model (Kirkpatrick and Sherrington, 1978; Edwards and Anderson, 1975). Ising models in general are known to be among the simplest models of complex systems that have been developed in the field of statistical physics and are at the roots of the theory on collective behavior and phase transitions. This large interest is also due to the fact that they fall into the class of Maximum Entropy models (Jaynes, 1957; Schneidman, Berry, Segev, and Bialek, 2006; Marre, El Boustani, Frégnac, and Destexhe, 2009) when only average values and cross correlations are taken into account. The KIM in particular has been adopted in a variety of fields, such as neuroscience (Cocco, Monasson, Posani, and Tavoni, 2017; Nghiem et al., 2018; Ferrari et al., 2018), computational biology (Tanaka and Scheraga, 1977; Imparato, Pelizzola, and Zamparo, 2007; Agliari, Barra, Guerra, and Moauro, 2011), economics and finance (Bornholdt, 2001; Bouchaud, 2013; Sornette, 2014; Campajola, Lillo, and Tantari, 2020) and has been studied in the literature of machine learning (LeCun, Bengio, and Hinton, 2015; Hornik, Stinchcombe, and White, 1989; Decelle and Zhang, 2015) to understand recurrent neural network models.

The KIM describes the time evolution of a set of N binary variables $s^{(t)} \in \{-1, 1\}^N$ for $t = 1, \dots, T$, typically called "spins", which can influence each other through a time lagged interaction. We focus on its applications to time series analysis and extend it to allow the presence of time-varying parameters with score-driven dynamics. In its standard form the Kinetic Ising Model for time series (Campajola, Lillo, and Tantari, 2019) involves three main sets of parameters: a $N \times N$ interaction or coupling matrix $\mathbf{J}$ and a N-dimensional vector h of variable-specific biases, which we summarize as $\Theta = (\mathbf{J}, h)$. The model is Markovian with synchronous dynamics, characterized by the transition probability

$$p(s^{(t)}|s^{(t-1)}; \beta, \Theta) = \frac{e^{\beta \sum_i s_i^{(t)} g_i^{(t)}}}{K^{(t)}} \tag{8.3}$$

where $K^{(t)}$ is a normalizing constant commonly known as the partition function in statistical mechanics, and β is a parameter that determines the amount of noise in the dynamics, known as the *inverse temperature*. Typically the quantity $g_i^{(t)} \equiv \sum_j J_{ij} s_j^{(t-1)} + h_i$ is called the *effective field* perceived by spin i at time t. Furthermore, it is possible in principle to introduce dependency on any number K of external regressors $x_k^{(t)}$, by adding a term $b_{ik} x_k^{(t)}$ to $g_i^{(t)}$ for each $k \in \{1, \dots, K\}$, as done for instance in (Campajola, Lillo, and Tantari, 2020).

From the standard KIM we use Eq.(5.2) to provide a dynamics to the parameters (β, Θ) thus introducing a Score Driven generalization of the KIM. Notice however that the number of parameters in the KIM is large, $O(N^2)$: as customary in high-dimensional modeling, in the following we will propose two parsimonious and informed parameter restrictions that simplify the treatment and define two kinds of Score-Driven KIM, each tailored to highlight different effects.

As we show in this Chapter, the development of a score-driven KIM addresses three important points: first, introducing a dynamical noise parameter $\beta^{(t)}$ allows

to gain real time insight on the ability of the model to explain the observed dynamics, thus leading to more informed forecasts; second, neglecting time variability of parameters by estimating a standard KIM turn out to produce systematic errors, in particular the estimated values are different from the time-averaged values that generated the sample; third, by introducing a convenient factorization for the model parameters, it is possible to discriminate whether an observation is better explained by endogenous interactions with other variables or by exogenous effects, offering an improved understanding of the dynamics that generated the data even when these effects are not constant over time. We prove the effectiveness of our modeling approach by extensive numerical simulations and by empirical application to different complex systems.

8.2 Preliminary Discussion of the Kinetic Ising Models

Spin systems have been analyzed by physicists since the early 20th century, mostly as models to understand the microscopical foundations of magnetism. A spin is indeed a proxy for an atomic magnetic moment, i.e. the torque the atom is subject to when immersed in a magnetic field. The first spin model is the celebrated 1D Ising Model (Ising, 1925), which mathematically abstracts the problem to an infinite chain of binary variables s (the spins) that can be either in an "up" or "down" state. These perceive the local magnetic field generated by their nearest neighbors as well as any external magnetic field and tend to "align" (i.e. match the state) or "disalign" (i.e. go in the opposite state) with the net field they sense, based on the value of a parameter $\mathbf{J}$. The model was intended to verify the hypothesis that thermal properties of ferromagnetic materials can arise from microscopic interactions between their atoms, but failed to do so because no net magnetic field would be observed at equilibrium. However, it was later shown (Onsager, 1944) that the failure was not due to the mechanism, but to the oversimplification of taking a 1-dimensional system: in fact, if one takes a 2D lattice instead of a 1D chain, this extremely simple model qualitatively reproduces the macroscopic thermal properties of ferromagnets. The success of the Ising Model has led to its extension and refinement to describe exotic materials such as spin glasses (Kirkpatrick and Sherrington, 1978), and its fascinating ability to describe macroscopic properties determined by microscopic coordination posed the foundations to many quantitative models of complex systems, with examples of successful Ising-like models for protein and DNA chains (Tanaka and Scheraga, 1977), neurons (Hopfield, 1982) and financial markets (Bouchaud, 2013).

The appeal of Ising Models comes in part from the fact that they belong to the class of *Maximum Entropy* models, as introduced by (Jaynes, 1957). The principle states that, given a set of constrained quantities from available observations - such as sample averages - a probability distribution that maximizes Shannon's entropy (Shannon, 1948) subject to the constraints is the best distribution to describe the observations, as it is the one that makes the least arbitrary assumptions. In particular Ising Models result from Shannon's entropy maximization constraining means and correlations of the spins, thus making them a popular choice to describe systems that can be encoded in binary strings.

The KIM is the out-of-equilibrium version of the Sherrington-Kirkpatrick (SK) spin glass (Derrida, Gardner, and Zippelius, 1987; Crisanti and Sompolinsky, 1988), developed a few years later and proposed as dynamical model for asymmetric neural networks with discrete time and synchronous sampling. The model's transition probability, describing the probability of observing a future configuration $\{s_i^{(t)}\}$

given a current configuration $\{s_t^{(t-1)}\}$, reads

$$p(\{s_i^{(t)}\}|\{s_i^{(t-1)}\}, \mathbf{J}, h) = \frac{1}{K^{(t)}} \exp\{\sum_{i,j} J_{ij} s_i^{(t)} s_j^{(t-1)} + \sum_i h_i s_i^{(t)}\} \tag{8.4}$$

Differently from its predecessor, which describes the equilibrium properties of spin glasses, this model describes the dynamics of a system of spins which have asymmetric interactions, namely spin i's effect on spin j is different from spin j's effect on i. This difference is incorporated in the structure of the $\mathbf{J}$ matrix, which is symmetric in the SK model and asymmetric in the KIM. Having $J_{ij} \neq J_{ji}$ in fact implies that these coefficients can no longer describe a synchronous interaction, as for instance a correlation coefficient, but need to describe an asynchronous one, specifically in this case a lag one interaction.

Typically, in the physics literature, the $\mathbf{J}$ elements are assumed to be *iid* Gaussian random variables, $J_{ij} \sim \mathcal{N}(J_0/N, J_1^2/N)$ and the properties of the model as data generating process are the object of analysis. As shown in (Crisanti and Sompolinsky, 1988), the KIM loses the so-called "spin glass" phase of the SK model - a phase in which the system "freezes" in a metastable configuration with local order but no global order - and only presents a dynamic phase transition between a paramagnetic phase - where spins do not show preferential alignment - and a ferromagnetic phase - where all spins align in one direction - when the mean of the $\mathbf{J}$ elements, J_0/N, is greater than $1/N$.

A more complete characterization of the model can be found in the physics literature (Crisanti and Sompolinsky, 1988; Derrida, Gardner, and Zippelius, 1987; Coolen, 2001a; Coolen, 2001b), with recent developments contributing to neuroscience (Tyrcha, Roudi, Marsili, and Hertz, 2013), machine learning (Dunn and Roudi, 2013; Decelle and Zhang, 2015; Campajola, Lillo, and Tantari, 2019) and finance (Campajola, Lillo, and Tantari, 2020) literatures. As a last remark, the model has been developed in at least another independent strand of literature with the name of Discrete Auto-Regressive model (DAR) (Jacobs and Lewis, 1978), and an equivalence between these models has been recently shown in (Campajola, Lillo, Mazzarisi, and Tantari, 2021).

8.3 The Dynamical Noise KIM

The first score-driven KIM we propose addresses the first two points made above, namely the real time prediction of forecast accuracy and the correction of systematic estimation errors of a constant parameter model. The Dynamical Noise KIM (DyNoKIM) is defined by letting the noise parameter β in Eq. 8.3 be time-varying, while all other parameters are constant.

To better understand the rationale behind this choice, let us introduce the theoretical Area Under the ROC Curve (AUC) (Hanley and McNeil, 1982; Bradley, 1997), a standard measure of the accuracy of the forecast, and study how it varies as a function of β in the standard KIM. We provide the details of the derivation in Appendix 8.7, where we show how the AUC depends both on β and on the unconditional distribution of the effective fields $g_i^{(t)}$. In Figure 8.1 we display the result assuming that g is Gaussian distributed with mean g_0 and standard deviation g_1. This is the case for instance if the J_{ij} entries are Gaussian distributed with zero mean. We see that the AUC is monotonically increasing with β, but also that the distribution of the static parameters affects the slope with which the curve converges towards 1, namely the smaller the mean and variance of the effective fields g_i, the slower the growth of

AUC. Figure 8.1 tells us that the larger is β the more reliable is the prediction of the model. Hence if we are able to estimate β locally we can assess in real time how good the model is in forecasting the next observation. This is why in the DyNoKIM we consider a time-varying β.

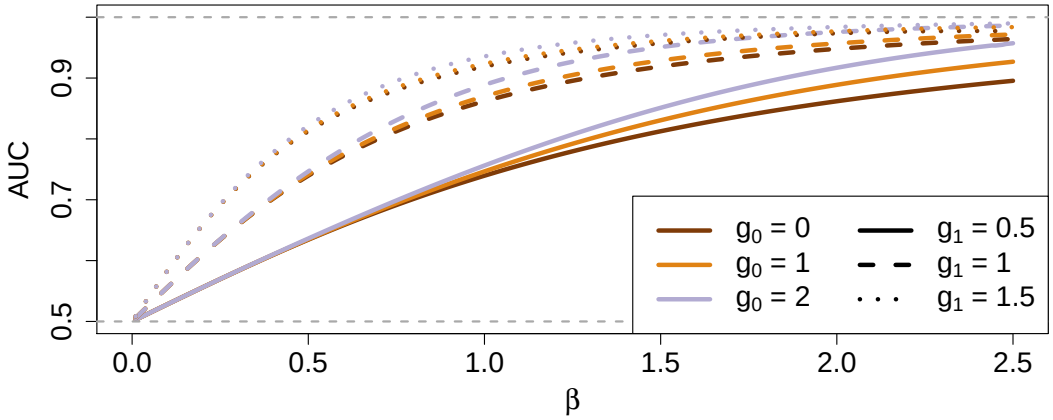

FIGURE 8.1: Theoretical AUC as a function of β assuming g_i is Gaussian distributed with mean g_0 and standard deviation g_1. Different colors correspond to different values of g_0, while line types identify values of g_1. We see that increasing β has the effect of reducing the uncertainty on the random variable $s_i^{(t)}$, keeping g_i unchanged. Grey dashed lines at AUC $= 0.5$ and AUC $= 1$ are guides to the eye.

Specifically, the DyNoKIM is characterized by the transition probability

$$p(s^{(t)}|s^{(t-1)};\mathbf{J},\beta^{(t)}) = \frac{e^{\beta^{(t)}\sum_i s_i^{(t)} g_i^{(t)}}}{K^{(t)}} \tag{8.5}$$

with $K^{(t)} = \prod_i 2\cosh\left[\beta^{(t)} g_i^{(t)}\right]$. We give score-driven dynamics to $f^{(t)} = \log\beta^{(t)}$, as β is positive and inversely related to the noise:

$$\log\beta^{(t+1)} = w + b\log\beta^{(t)} + a\boldsymbol{\mathcal{I}}^{-1/2(t)}\nabla_t \tag{8.6}$$

where w, b and a are scalar parameters and $\boldsymbol{\mathcal{I}}^{(t)}$ is the Fisher Information and $\nabla_t = \frac{\partial p(s^{(t)}|s^{(t-1)},\beta^{(t)})}{\partial\beta^{(t)}}$ is the score.

The interpretation for this model is simple yet extremely useful: the higher the value of β, the smaller the uncertainty over the realization of $s^{(t)}$ or, in other words, the more accurate a prediction of the value of $s^{(t)}$, as we have shown in Fig. 8.1. Operationally, at a given time $t-1$ with an observation $s^{(t-1)}$, it is possible to use the DyNoKIM to produce one-step ahead forecasts for $s^{(t)}$, which we call $\hat{s}_i^{(t)}$. These are obtained as

$$\hat{s}_i^{(t)} = \text{sign}\left[p\left(s_i^{(t)} = 1\middle|s^{(t-1)},\Theta,\beta^{(t)}\right) - \alpha\right] \tag{8.7}$$

where α is an arbitrary threshold level. Sweeping the value of α between 0 and 1 one obtains a ROC curve, which in turn can be used to calculate the AUC. We report simulation results for this procedure in Appendix 8.7. Notice that $\beta^{(t)}$ depends only from past observations $\mathcal{S}^{(t-1)}$ through Eq.8.6, thus the predictions are fully causal.

In the statistical physics literature there have been few attempts to study similar models (Penney, Coolen, and Sherrington, 1993; Beck and Cohen, 2003; Beck,

Cohen, and Swinney, 2005). However these works assume that the sampling of the observations and of the time-varying parameters take place on two separated time scales, meaning that the parameters are locally constant when the observations are sampled. This is not true for score-driven models, which are in fact designed to not require this assumption, intuitively formalized by the values of the parameters b and a. If $b \gg a$ then the evolution of f is indeed slower than the one of observations, while if $b \ll a$ they evolve on the same time scale.

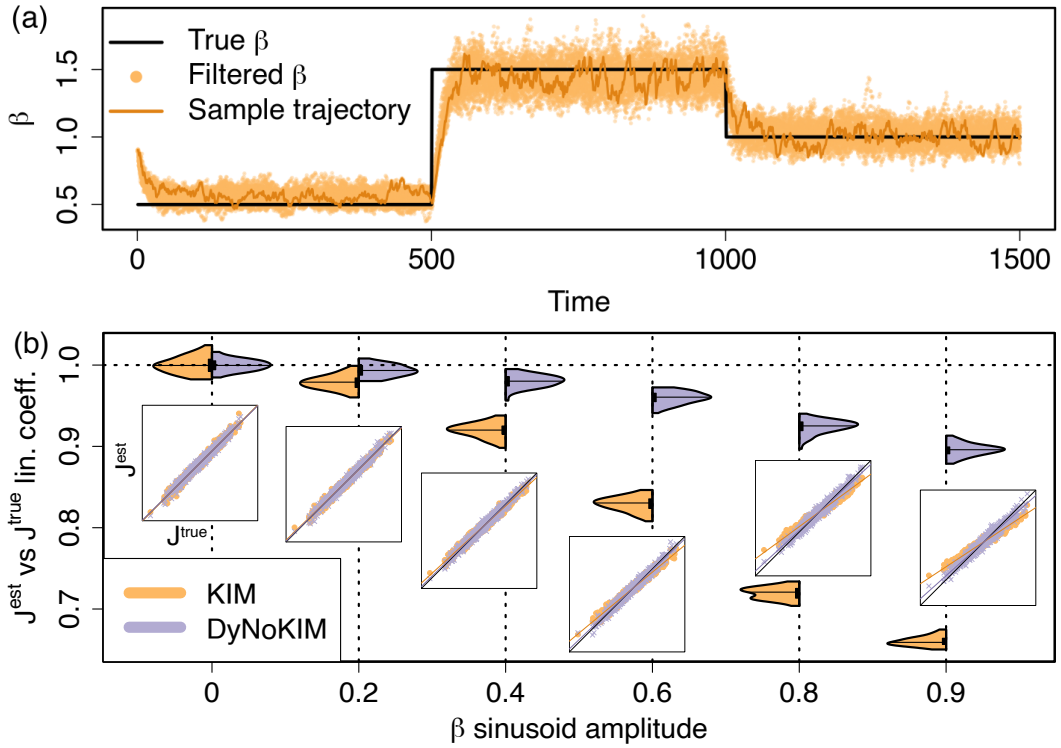

FIGURE 8.2: (a) Example of a filtered β trajectory with a piecewise constant generating process over 30 simulations. (b) Estimation of $\mathbf{J}$ under model misspecification with a time-varying $\beta^{(t)} = 1 + \mathcal{K}\sin(\omega t)$, comparing the KIM and the DyNoKIM. On the x axis we plot the amplitude $\mathcal{K}$, on the y axis the distribution of the coefficient b of the linear regression over 60 simulations. Insets show example scatter plots of the true $\mathbf{J}$ values (x axis) and the estimated values (y axis) using the standard KIM (yellow points) or the DyNoKIM (purple crosses). Simulation parameters are $\omega = 2\pi/300$, $T = 3000$, $N = 30$, $J_{ij} \sim \mathcal{N}(0, 1/\sqrt{N})$, $h_i = 0\ \forall\, i$

The estimation of the DyNoKIM requires some restrictions. It is known (Mézard and Sakellariou, 2011) that, given a set of observations, the parameter β in the standard KIM of Eq. 8.3 is not identifiable. In fact, for any two values β_1 and β_2 there are also two sets of parameters Θ_1 and Θ_2 such that $p(s^{(t)}|s^{(t-1)}; \beta_1, \Theta_1) = p(s^{(t)}|s^{(t-1)}; \beta_2, \Theta_2)$ for all $s^{(t)}$. For this reason in inference problems it is typically assumed that $\beta = 1$ incorporating its effect in the size of the other parameters. When β is made time-varying though, the identification problem is limited to its time average value $\langle\beta\rangle$ (which still needs to be assumed equal to 1), while its local value can be inferred from the data. This result has implications particularly for forecasting applications: a forecast should be considered more or less reliable by looking at the value of $\beta^{(t)}$ at the previous instant in time and considering how much above 0.5 the corresponding expected AUC is, according to the relation shown in Figure

8.1. Finally, the parameters of Eq. 8.6 are inferred by Maximum Likelihood Estimation as follows. The KIM static parameters Θ are inferred via Maximum Likelihood Estimation using a known Mean Field technique (Mézard and Sakellariou, 2011) or, when this is not possible, via standard Gradient Descent methods. Given Θ we estimate w, b, a by performing a targeted estimation (Francq, Horvath, and Zakoïan, 2011) through ADAM stochastic Gradient Descent (Kingma and Ba, 2014). Targeted estimation, which is common in observation-driven models such as the GARCH (Bollerslev, 1986b), first fits the mean value of the time-varying parameter $\langle f \rangle = w/(1-b)$ and then fits the (w, b, a) parameters keeping this ratio constant. This procedure significantly reduces the estimation time and produces accurate estimates in our simulations. Further details on the process can be found in Appendix 8.7.

Our main focus here is to study the model's ability to retrieve the correct parameters also when the data generating process is not score-driven. Indeed there is little reason to believe that this sort of dynamics is an actual data generating process for real-world complex systems, where β might follow exogenous and unknown dynamics. The power of score-driven models lies also in the capability of estimating time-varying parameters, such as $\beta^{(t)}$, without actually requiring any assumption on their true dynamical laws. In this sense they behave as filters for the underlying unknown dynamics of the parameters. To show that this is the case also for the DyNoKIM, Fig. 8.2a displays an example of misspecified $\beta^{(t)}$ dynamics, a deterministic double step function, that is correctly recovered by the score-driven approach. We simulate 30 time series of length T using the given values of $\beta^{(t)}$ to generate the $s^{(t)}$; given only the simulated $s^{(t)}$ time series, the inference algorithm determines the optimal static parameters a, b and $\mathbf{J}$ and filters the optimal value of $\beta^{(t)}$ at each time. The resulting $\beta^{est\,(t)}$ values are well localized around the simulated ones.

One could argue that a KIM with a time varying $\beta^{(t)}$ has similar performances to a standard KIM with a constant β equal to $\langle \beta \rangle$. This is not the case. Fig. 8.2b shows the results for a set of simulations where $\beta^{(t)}$ follows a deterministic sinusoidal dynamics, $\beta^{(t)} = 1 + \mathcal{K} \sin \omega t$, varying the amplitude $\mathcal{K}$, and the time evolution of $s^{(t)}$ is given by Eq. 8.5. For each value of $\mathcal{K}$ we simulate 60 time series of T observations and fit both the constant parameters KIM and the score-driven DyNoKIM, then comparing the inferred J^{est} with the one that was used to generate the data, J^{true}, by means of a linear regression model $J_{ij}^{est} = a + b J_{ij}^{true} + \epsilon$. We see from Figure 8.2b that when β is not constant, the KIM underestimates the absolute value of the parameters, highlighted by the fact that $b < 1$ (and $a \approx 0$, not shown). The error is greatly reduced in the DyNoKIM thanks to the way in which we solve the indetermination of $\langle \beta \rangle$: after the model parameters are estimated and a filtered $\beta^{est\,(t)}$ is found, we normalize its mean to 1 and multiply the estimated J^{est} by the same factor, leaving the likelihood of the model unchanged. This result supports our argument that using a KIM on data where parameters of the data generating process are time varying can be misleading and leads to significant errors, something that can be overcome by adopting the score driven models proposed here.

Finally, let us point out that the Lagrange Multiplier (LM) test, described in Section 5.6, can be used to reject the hypothesis of constant β.

8.3.1 Forecasting stock price activity with DyNoKIM

Our first application of DyNoKIM is to financial markets. Measuring high-frequency price volatility in financial markets is a non-trivial task that has been at the core of

research in quantitative finance over the last two decades (Aït-Sahalia, Mykland, and Zhang, 2011). Volatility is in fact a latent process which is hard to measure for reasons that range from price staleness to microstructural effects like price discretization and bid-ask bounce. Price activity, namely the binary time series marking the events of price changes, is a proxy for high-frequency price volatility that has been used recently to quantify the endogeneity in the price formation (Filimonov and Sornette, 2012; Hardiman, Bercot, and Bouchaud, 2013; Filimonov and Sornette, 2015; Hardiman and Bouchaud, 2014; Wheatley, Wehrli, and Sornette, 2019; Rambaldi, Pennesi, and Lillo, 2015; Rambaldi, Filimonov, and Lillo, 2018).

Here we propose the DyNoKIM as an effective tool to forecast stock price activity at high frequency. The advantages with respect to standard methods is twofold: first, we are able to model the dynamics of a large panel of assets, hence considering volatility spillovers between them; second, the score driven approach allows us to measure the local predictability of price activity in real time. We study the 100 largest capitalization stocks in the NASDAQ and NYSE over 11 trading days. Price activity is defined as a binary variable $s_i^{(t)}$ for each stock i, taking value $+1$ if the stock price has changed in the interval $(t-1, t]$ and -1 otherwise, with time discretized at 5 seconds. The choice of time scale is largely arbitrary: we choose 5 seconds to obtain a set of variables that have unconditional mean as close to 0 as possible to have a balanced dataset. We focus our attention on the lagged interdependencies among different stocks, by applying the DyNoKIM to the multivariate time series $s^{(t)}$.

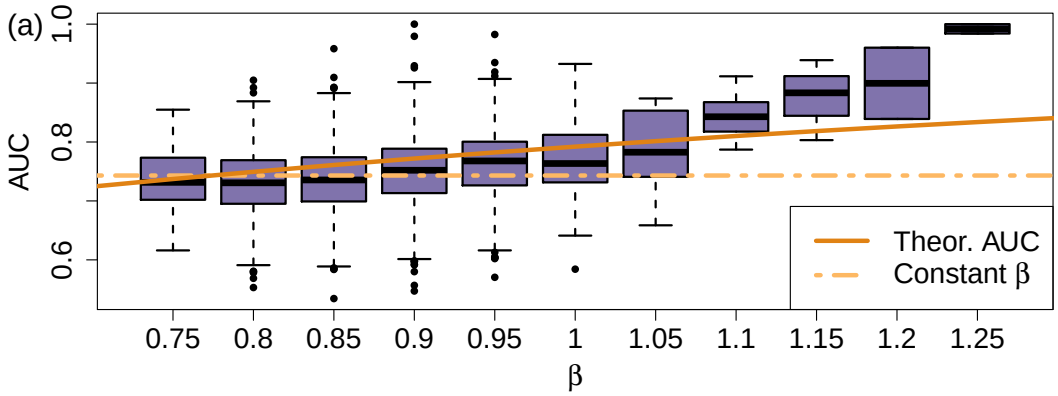

FIGURE 8.3: AUC statistics compared to $\beta^{(t)}$ for stock price activity on November 19, 2019 aggregated for different values of $\beta^{(t)}$, compared to the theoretical expected AUC with Gaussian $g_i^{(t)}$ and to the performance with constant β;

Our theoretical results from Figure 8.1 suggest to use β to quantify the reliability of forecasts of price activity using this model. We thus estimate the model parameters once per day and use them to filter $\beta^{(t)}$ on the following day, while measuring the out of sample accuracy of the predicted price movements using the AUC metric. To ensure that there is reason to model the system with time-varying β, we apply a Lagrange Multiplier (LM) test (Calvori, Creal, Koopman, and Lucas, 2017) with a null hypothesis of constant β, finding strong rejections of the null at the $p < 0.001$ level for every day in the sample. Further information on the test can be found in Section 5.6.

We show an example of the results of this analysis in Fig. 8.3 where we consider a single day. We empirically observe that when the filtered value of $\beta^{(t)}$ is large, the

subsequent forecast of activity is systematically more reliable because AUC is larger. We find a good agreement between the empirical results and the theoretical values for AUC under the assumption of Gaussian effective fields g_i, even if some discrepancy is observable due to the non-Gaussianity of actual fields. Thus we conclude that the DyNoKIM can be effectively used to model high frequency volatility of a large portfolio of stocks and to measure in real time its level of predictability.

The empirical example presented show that our theoretical results for the DyNoKIM are indeed verified in realistic applications and that using this method - which we believe could be applied even to more sophisticated models - can result in a significant gain in the use of forecasting models, giving a simple criterion to discriminate when to trust (or not) the forecasts.

8.4 The Dynamic Endogeneity KIM

A more general specification of the score-driven KIM is the Dynamic Endogeneity Kinetic Ising Model (DyEnKIM), where we assume that each parameter $\mathbf{J}$ and h has its own specific time-varying factorization. Going back to Eq. 8.3, we now impose the following structure to the parameters:

$$\begin{aligned} \beta &= 1 \\ J_{ij}^{(t)} &= \beta_{diag}^{(t)} J_{ij}\delta_{ij} + \beta_{off}^{(t)} J_{ij}(1-\delta_{ij}) \\ h_i^{(t)} &= \beta_h^{(t)}(h_i + h_0^{(t)}) \end{aligned} \tag{8.8}$$

where δ_{ij} is the Kronecker symbol which is 1 if $i = j$ and 0 otherwise and we will call $\boldsymbol{\beta}^{(t)} = (\beta_{diag}, \beta_{off}, \beta_h)$ in the following.

With this choice we want to be able to discriminate between different components of the observed system dynamics: one associated with the idiosyncratic properties of variable i (β_h), with general trends (h_0), with autocorrelations (β_{diag}), and finally with lagged cross-correlations among variables (β_{off}). In this formulation each of these time-varying parameters β measures the relative importance of one term over the others in the generation of the data, highlighting periods of higher endogeneity of the dynamics (when correlations have higher importance) rather than periods where the dynamics is more idiosyncratic or exogenously driven. We report a consistency analysis for the DyEnKIM in Section 8.7 of the Appendix, where we show that even under model misspecification this approach correctly separates the different components of the dynamics and captures their relative importance. Additionally, when using the test of Section 5.6, in the DyEnKIM, having multiple time-varying parameters, we test each parameter against two null hypotheses, one where all parameters are constant and one where all other parameters are score-driven, applying FDR correction for multiple tests.

8.4.1 Role of non stationarity in neural data

As a first example of the application of the DyEnKIM, we consider the firing dynamics of a set of neurons. Inferring the network of connections between neurons by observing the correlated dynamics of firing has received a lot of attention in the last two decades (Cocco, Leibler, and Monasson, 2009; Schneidman, Berry, Segev, and Bialek, 2006) and the KIM has been extensively used for this purpose (Hertz et al., 2010; Zeng, Aurell, Alava, and Mahmoudi, 2011; Hoang, Song, Periwal, and Jo, 2019). The underlying idea is that the (lagged) correlation in the firing of two time

series suggests the existence of a physical connection between the two corresponding neurons.

However, as pointed out in (Tyrcha, Roudi, Marsili, and Hertz, 2013), correlated behavior can also be generated by the fact that neurons are subject to a common non-stationary input, for example driven by the external environment. Disentangling the contributions to correlations coming from external drivers from those coming from genuine interactions is critical to reliably identify the network structure between neurons.

To this end (Tyrcha, Roudi, Marsili, and Hertz, 2013) proposes an inferential method to achieve this result by considering a KIM with a time dependent external field $h_i^{(t)}$ representing the contribution of the external stimuli and of all the non recorded neurons to the activity of neuron i at time t. However the inference method requires many "trials" or repetitions of the experiment, under the strong methodological assumption that all the repetitions are obtained under identical conditions, an hypothesis that might be difficult to control in such type of complex experiments.

We now show that DyEnKIM can be used for this purpose *on a single experiment*. We use the data of (Tkačik et al., 2014) obtained from a multichannel experiment recording firing patterns of 160 salamander retina neurons, stimulated by a film clip of a swimming fish. The 20s experiment is sampled with time binning of 20ms, corresponding to $T = 944$ and we considered the $N = 40$ most active neurons. Finally the experiment is repeated 297 times.

The DyEnKIM of Eq. 4 is estimated and for each experiment we perform an LM test. We find that while for $\beta_{off}^{(t)}$, $\beta_h^{(t)}$, and $h_0^{(t)}$ we reject the null hypothesis of constant parameter in 99.3%, 76.8%, and 100% of the experiments respectively, this percentage drops to 43.1% for $\beta_{diag}^{(t)}$. For this reason we consider a simplified model where $\beta_{diag}^{(t)}$ is constant[1]. Fig. 8.4a shows the temporal dynamics of the three filtered parameters. Since we are able to filter the dynamics for each experiment, in the figure we show the mean and the 90% confidence interval. It is evident that the three parameters show significant variations, likely in response to the external stimulus provided by the film clip and by unobserved neurons.

In order to evaluate how well our model describes the empirical data we consider two statistics: (i) the distribution of the number of synchronous (i.e., within the same time bin) spikes and (ii) the Zipf plot, obtained as the rank plot of the frequency of each spiking pattern. Both quantities depend on the many body synchronous correlations among spins, thus are not automatically explained by KIM-type models which fit the pairwise correlations. As a benchmark model we consider a constant parameter KIM estimated on the whole dataset. In Fig. 8.4b-c we show these statistics. We observe that the DyEnKIM reproduces both quite well, while the constant parameter KIM largely fails in describing the distribution of the number of synchronous spikes and in predicting the frequency of the most frequent patterns (rank between 2 and ~ 100) where the underestimation is up to an order of magnitude. We also considered a sparse version of the KIM obtaining similar results in the Appendix.

The above results are very interesting because they show that a pairwise dynamic interaction model is able to reproduce higher-order correlations, *if one takes into account the time varying dynamics of the global interactions* (see also (Schneidman, Berry, Segev, and Bialek, 2006) for the static Ising model). It is important to stress once more that, while an approach as in Ref. (Tyrcha, Roudi, Marsili, and Hertz, 2013)

[1]The following results are essentially unchanged when considering a time varying $\beta_{diag}^{(t)}$.

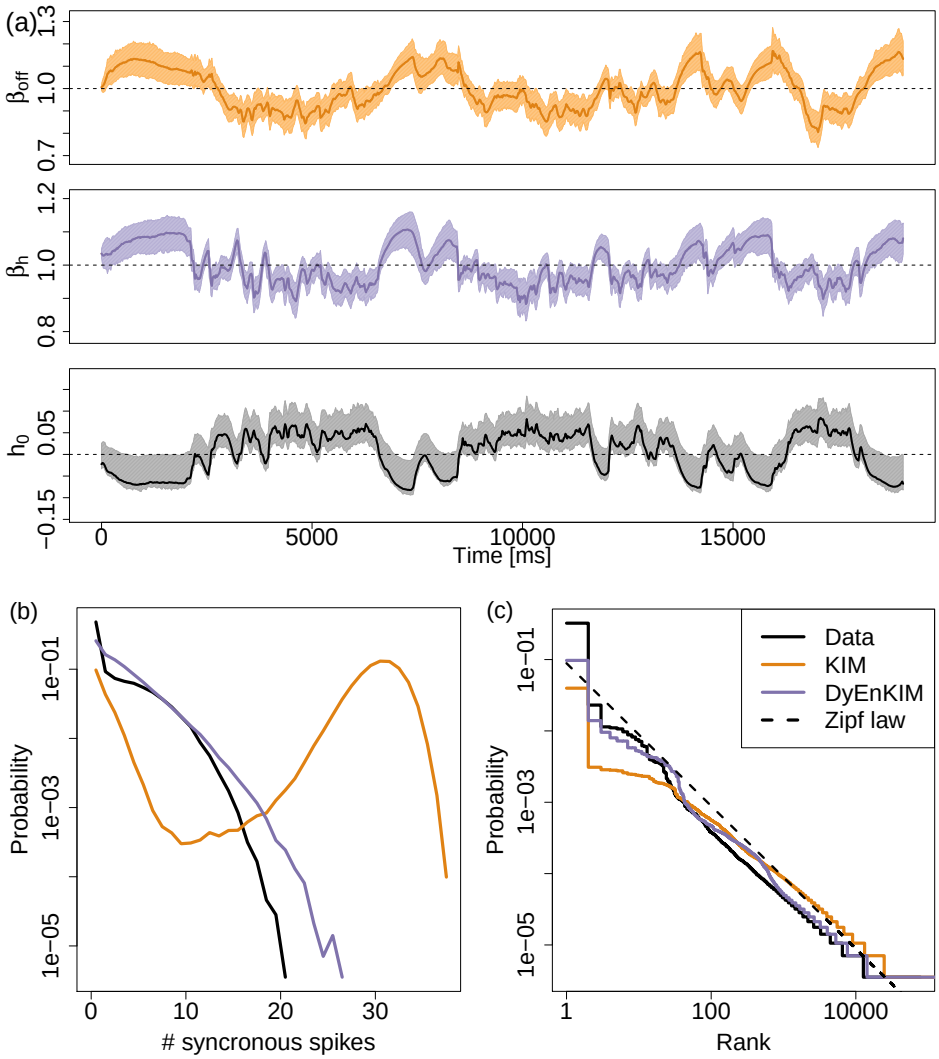

FIGURE 8.4: (a) Filtered values of β_{off}, β_h, and $h_0^{(t)}$ for salamander retina data. The continuous line is the mean value across the 297 experiments and the 90% confidence interval (i.e. 268/297 of the filtered values stay within the bands). (b) Estimated probability density function of the number of synchronous spikes. (c) Zipf plot of the frequency of observed patterns. In (b-c) the probability densities are obtained as average across the experiments, but a small variability is observed when considering individual experiments.

requires many experiments and the strong methodological assumption that these are identical realizations of the same process, our method to measure time-varying interactions can be performed on a single experiment. Incidentally, one can then use the estimation to test whether the different experiments are statistically equivalent by comparing the estimates across replicas. Moreover our model has only three time dependent scalars, while the model of (Tyrcha, Roudi, Marsili, and Hertz, 2013) requires a time dependent field for each of the N neurons, thus the latter is highly parametrized with a modeled dynamics strongly constrained by the data.

8.4.2 Disentangling endogenous and exogenous price dynamics

As a second application of the DyEnKIM we consider the problem of quantifying the contribution to stock price changes due to exogenous events (e.g. news, announcements) and to endogenous feedbacks. A vast literature (Filimonov and Sornette, 2012; Filimonov and Sornette, 2015; Hardiman, Bercot, and Bouchaud, 2013; Hardiman and Bouchaud, 2014; Rambaldi, Pennesi, and Lillo, 2015; Rambaldi, Filimonov, and Lillo, 2018; Wheatley, Wehrli, and Sornette, 2019) has tackled this point, but almost invariably this has been done by assuming that the relation between price and external drivers, as well as those driving the internal feedback, is constant in time. The DyEnKIM allows us to test this hypothesis, by considering time varying parameters whose dynamic can be filtered from data. Understanding the role of exogenous or endogenous drivers in market volatility is very important, also to devise possible policy measures able to avoid their occurrences and DyEnKIM, being able to identify them in real time, could provide valuable tools for market monitoring.

For this application we focus on two events that caused huge turmoil in the stock markets at the intraday level. The first one is the May 6, 2010 Flash Crash, when a seemingly unjustifiable sudden drop in the price of E-mini S&P 500 futures contracts caused all major stock indices to plummet in a matter of a few minutes, recovering most of the lost value when circuit breakers came into place. Multiple explanations of what happened have been offered by a large number of academics, regulators and practitioners: responsibility has been attributed to careless algorithmic trading (Commission and Commission, 2010), deteriorated market liquidity which quickly vanished when price volatility increased (Easley, De Prado, and O'Hara, 2011), market fragmentation (Madhavan, 2012; Menkveld and Yueshen, 2019), predatory trading strategies by high-frequency traders (Kirilenko, Kyle, Samadi, and Tuzun, 2017; Aquilina, Budish, and O'Neill, 2020).

The second event we analyze is the announcement following the Federal Open Market Committee (FOMC) meeting of July 31, 2019. In this meeting the Federal Reserve operated its first interest rate cut in over a decade, the last one dating back to the 2008 financial crisis, encountering mixed reactions in both the news and the markets. In particular an answer to a question in the Q&A press conference by the Fed Chairman Powell has been highlighted by news agencies, when being asked whether further cuts in the future meetings were an option, he answered "we're thinking of it essentially as a midcycle adjustment to policy" (Powell, July 31, 2019). This answer triggered turmoil in the equity markets, with all major indices dropping around 2% in a few minutes.

Like in the previous section, we construct our dataset for both events taking price movements for the then S&P100-indexed stocks at the 5 seconds time scale and constructing the associated price activity time series. Differently from the previous example, here we apply the DyEnKIM methodology to study variations in the relative importance of different sets of parameters as events unfold. In this case the LM test rejects the null of constant parameters for all βs and all datasets. To better interpret the results we introduce the value of the components of the effective fields $g_i^{(t)}$, each related to one of the time-varying parameters

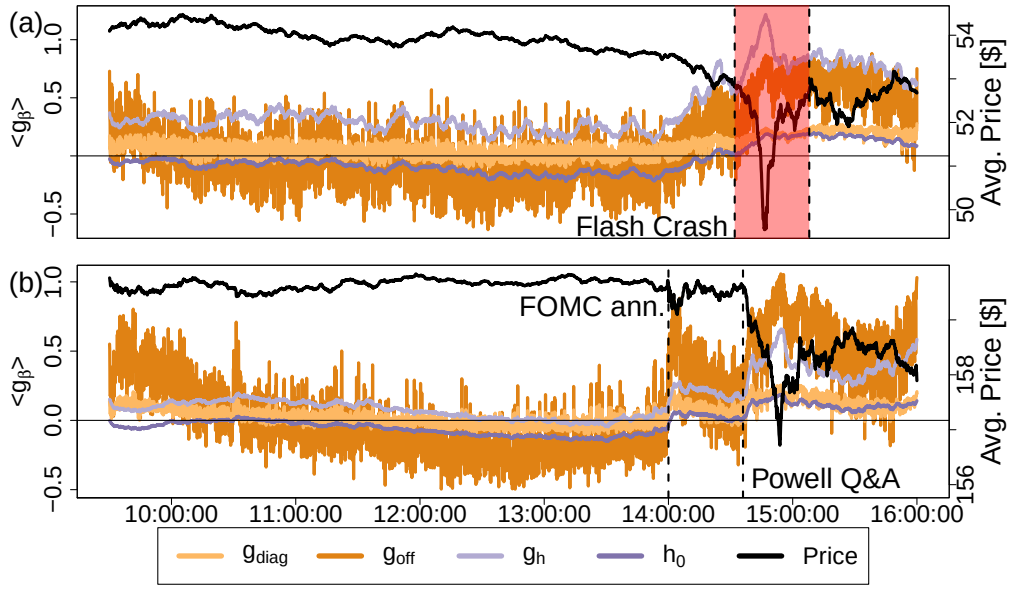

FIGURE 8.5: Values of $\langle g_{diag} \rangle^{(t)}$, $\langle g_{off} \rangle^{(t)}$, $\langle g_h \rangle^{(t)}$ and $h_0^{(t)}$ during the day of May 6, 2010 Flash Crash (a) and during the day of FOMC announcement on July 31, 2019 (b). The black lines are the the average midprice across the S&P100 stocks. The red area in the top panel highlights the time window (14:32:00 to 15:08:00 EST) where the Flash Crash takes place.

$$
\begin{aligned}
g_i^{(t)} &= g_{i,diag}^{(t)} + g_{i,off}^{(t)} + g_{i,h}^{(t)} \\
g_{i,diag}^{(t)} &= \beta_{diag}^{(t)} J_{ii} s_i^{(t)} \\
g_{i,off}^{(t)} &= \beta_{off}^{(t)} \sum_j J_{ij} s_j^{(t)} \\
g_{i,h}^{(t)} &= \beta_h^{(t)} (h_i + h_0^{(t)})
\end{aligned}
$$

which we then average at each time across all indices i, obtaining the quantities $\langle g_{diag} \rangle^{(t)}$ and so on.

Since the model is applied to price activity, which can be thought of as a proxy of high-frequency volatility (Filimonov and Sornette, 2012; Hardiman, Bercot, and Bouchaud, 2013), the financial interpretation of these time-varying parameters relates to volatility clustering in the case of β_{diag}, to volatility spillovers for β_{off}, to higher or lower market-wise volatility for h_0 and the relevance of exogenous effects is given by β_h. Thus the $\langle g_\cdot \rangle^{(t)}$ quantities can be intuitively related to what the explained sum of squares means for linear regression models, in the sense that the more a $\langle g_\cdot \rangle^{(t)}$ is far from 0 relative to others the more the data are affected at time t by that subset of parameters and the corresponding variable. We choose to show these quantities as a simple way of assessing the relevance of the components, a problem that is not easily solved in this kind of models.

The top panel of Figure 8.5 shows the components of the fields during the Flash Crash of May 6, 2010. Here the parameters show a very significant variation around the crash, with a large increase of $\langle g_h \rangle$ in the 45 minutes preceding the crash together with a similar increase of the endogeneity field $\langle g_{diag} \rangle$ and $\langle g_{off} \rangle$ during the event,

which then stay large until market close. This indicates that the turmoil induced by the Flash Crash reverberated for the remainder of the trading hours, even after the prices had recovered at pre-crash levels. The intraday pattern is overshadowed by the effect of the crash, but the picture at the beginning of the day is similar to normal trading days[2]. These results indicate an exogenous increase in activity before the crash, which is accompanied by the endogenous mechanism of volatility spillovers between stocks, as evidenced by large value of $\langle g_{off} \rangle$ during and after the Flash Crash. In conclusion our analysis indicates that both exogenous and endogenous drivers were important for the onset of Flash Crash.

In the bottom panel of Figure 8.5 we show the values of the effective fields on July 31, 2019. The FOMC announcement went public at 14:00:00 EST and is followed by a press conference at 14:30:00 EST, with a Q&A starting at around 14:36:00 EST. Again we see that the usual intraday pattern is interrupted by the news, which however, differently from the Flash Crash, was a scheduled event. This difference leads to the complete absence of any sort of "unusual" effect in the earlier hours of the day, as typically analysts provide forecasts regarding these announcements in the previous days and this information is already incorporated in the prices. What then happens is that, if the news does not meet market expectations, a correction in prices will occur as soon as the information is made public, leading to higher market volatility in the minutes and hours following the announcement (Chuliá, Martens, and Dijk, 2010; Hautsch, Hess, and Veredas, 2011). In this specific case, forecasts were mixed between a 0.25% and a 0.50% interest rates cut scenario.

The published announcement at 14:00 EST mostly matched these forecasts, with the FOMC lowering the interest target rate by 0.25%, and we indeed see that the price levels are not particularly affected by the news. However a transient increase in volatility, and in particular the endogenous components, can still be observed in the few minutes following the announcement, quickly returning to average levels. It is interesting to see the reaction to the press conference held 30 minutes after the release, and in particular to the answers the Chairman of the Fed Jerome H. Powell gives to journalists in the Q&A. As soon as the Q&A starts, around 14:36 EST, prices begin to plummet in response to the Chairman's answers, possibly reacting to the statement that this interest rates cut was only intended as a "midcycle adjustment to policy" rather than as the first of a series. Expectations of further rates cuts in the later months of the year could be a reason for this adjustment in the prices when these forecasts are not met, as usually lower interest rates push the stock prices up. We see however that this unexpected event causes a behavior in the estimated time-varying parameters resembling what we have seen in the Flash Crash, albeit the endogenous components are even more significant here.

8.5 Score Driven KIMs for Temporal Networks

Networks are a paradigmatic tool to describe pairwise relations in complex systems (Newman, Barabási, and Watts, 2006; Cohen and Havlin, 2010; Barabási, 2013; Newman, 2018) and applications include human mobility (Gao, Wang, Gao, and Liu, 2013), migration (Fagiolo and Mastrorillo, 2013), disease spreading (Draief and Massoulie, 2010), international trade (Bhattacharya et al., 2008) and financial stability (Gai, Haldane, and Kapadia, 2011; Cimini, Squartini, Garlaschelli, and Gabrielli, 2015), to mention a few. More recently, the increasing availability of time varying relational data stimulated a widespread and fast growing interest in the analysis

[2] See Section 8.7 of the appendix.

of temporal networks (Holme and Saramäki, 2012). It also motivated the development of a number of models to describe the dynamics of temporal networks (Sewell and Chen, 2015; Sewell, 2018; Mazzarisi, Barucca, Lillo, and Tantari, 2020; Hanneke, Fu, Xing, et al., 2010) A network, defined by a set of M nodes and a set of links between pairs of nodes, can be described by an $M \times M$ binary adjacency matrix $\mathbf{A} \in \{0,1\}^{M\times M}$, where $A_{ij} = 1$ if a link between nodes i and j is present and $A_{ij} = 0$ otherwise. When the relation described by the links is not directional, $A_{ij} = A_{ji}$ and the network is said to be undirected. We consider temporal networks where the number of nodes M is fixed across multiple time steps and indicate the adjacency matrix of the graph at time t by $\mathbf{A}^{(t)}$.

In order to use the KIM to model a temporal network, we map the elements of the adjacency matrix into spins, associating a present link to a spin $+1$ and an absent link to a spin -1. In this way we represent each adjacency matrix $\mathbf{A}\,(t)$ as a vector $s^{(t)} \in \{-1,1\}^N$ where $N = M(M-1)/2$, assuming the network to be undirected and without self loops. In light of this mapping, the matrix $\mathbf{J}$ now captures the tendency of links to influence each other at lag one - for example the diagonal terms can be interpreted as measuring link persistence - while the elements of h are associated with the idiosyncratic probability to observe a given link.

More formally, indicating by $vec(\mathbf{A})$ the vectorized version of matrix $\mathbf{A}$, the mapping can be summarized by the relation

$$vec(\mathbf{A}^{(t)})_k = 1/2 + s_k^{(t)}/2$$

and the KIM is equivalent to the following version of the TERGM

$$P(s^{(t)}|s^{(t-1)},\theta) = \frac{e^{\sum_{ij} J_{ij} s_i^{(t-1)} s_j^{(t)} + \sum_i h_i s_i^{(t)}}}{K(J,h)} \tag{8.9}$$

$$= \frac{e^{\sum_{ab} \theta_{ab}^{(1)} vec(\mathbf{A}^{(t-1)})_a vec(\mathbf{A}^{(t)})_b + \sum_a \theta_a^{(2)} vec(\mathbf{A}^{(t)})_a - \sum_a \theta_a^{(3)} vec(\mathbf{A}^{(t-1)})_a}}{K(\theta)}, \tag{8.10}$$

where we omitted the constant terms, not depending on the adjacency matrix, that have been absorbed in the normalization constant.

Interestingly, such a mapping highlights that (standard) KIM can be seen as belonging to the Temporal Exponential Random Graph Model (TERGM) (Hanneke, Fu, Xing, et al., 2010) family. Indeed, KIM is equivalent to a TERGM having three kinds of network statistics: first, the set of all possible lagged interactions $q_{ab}^{(1)} = vec(\mathbf{A}^{(t-1)})_a vec(\mathbf{A}^{(t)})_b$ between pairs of links, each appearing in this specification of Eq. 4.1 with a parameter $\theta_{ab}^{(1)} = 4J_{ab}$; second, a term associated to the probability of each link to be observed, $q_a^{(2)} = vec(\mathbf{A}^{(t)})_a$, with $\theta_a^{(2)} = 2(h_a - 1)$; and the last group of statistics, $q_a^{(3)} = vec(\mathbf{A}^{(t-1)})_a$, is related with the presence or absence of each link at the previous time step with parameters $q_a^{(3)} = 2\sum_b J_{ab}$.

The DyNoKIM allows us to consider time varying θ_i and to estimate the forecast accuracy of the model at each time step, as showed with the link prediction example presented in the main text.

Interestingly, a wide range of TERGM specifications can be mapped to the KIM. As a simple example, let us consider a TERGM with two terms only

$$P(\mathbf{A}^{(t)}|\mathbf{A}^{(t-1)},\theta) = \frac{e^{\sum_{ij}\left[\theta_{dens}\mathbf{A}_{ij}^{(t)} + \theta_{stab}(\mathbf{A}_{ij}^{(t)}\mathbf{A}_{ij}^{(t-1)} + (1-\mathbf{A}_{ij}^{(t)})(1-\mathbf{A}_{ij}^{(t-1)}))\right]}}{K(\theta)}. \tag{8.11}$$

This can be rewritten as

$$P(s^{(t)}|s^{(t-1)},\theta) = \frac{e^{N\theta_{dens}/2+\sum_k\left(s_k^{(t)}\theta_{dens}/2+s_k^{(t)}s_k^{(t-1)}\theta_{stab}/2\right)}}{K(\theta)},$$

which is exactly equivalent to a KIM restricted to have just two parameters $J_{ij} = J_{diag} = \theta_{stab}/2\ \forall i,j$ and $q_i = q_0 = \theta_{dens}/2\ \forall i$, and, absorbing the constants in the normalization function, we have

$$P(s^{(t)}|s^{(t-1)},J_{diag},q_0) = \frac{e^{\sum_k s_k^{(t)}q_0+s_k^{(t)}s_k^{(t)}J_{diag}}}{K(\theta)}.$$

If we consider the DyNoKIM extension of such a restricted KIM we obtain

$$P(s^{(t)}|s^{(t-1)},J_{diag},h_0,\beta^{(t)}) = \frac{e^{\beta^{(t)}\sum_k\left(s_k^{(t)}h_0+s_k^{(t)}s_k^{(t-1)}J_{diag}\right)}}{K(\theta)} \tag{8.12}$$

that is effectively an extended version of the initial TERGM. Moreover, it is easy to see that the DyEnKIM results in the following

$$P(s^{(t)}|s^{(t-1)},J_{diag},h_0,\beta^{(t)}) = \frac{e^{\sum_k\left(s_k^{(t)}\beta_h^{(t)}h_0+s_k^{(t)}s_k^{(t-1)}\beta_{diag}^{(t)}J_{diag}\right)}}{K(\theta)}. \tag{8.13}$$

that maps to a version of (8.11) with dynamical parameters $\theta_{dens}^{(t)}$ and $\theta_{stab}^{(t)}$ evolving independently. We believe this observation is very relevant as it is an extension of the TERGM at hand to its version where each parameter is allowed to follow its own evolution, potentially unrelated to the others. This is a different evolution from the DyNoKIM's one, as the latter is driven by a single $\beta^{(t)}$ and maps into comoving TERGM parameters. Indeed, also in this context, the two models have different purposes and different applicability. TERGM extensions resulting from DyNoKIM allow us to quantify forecast accuracy, similarly to what showed in the main text, while DyEnKIM, similarly to what we discussed in the applications presented and in numerical simulations, allows for a decoupling of the temporal relevance of different network statistics.

As a final remark, we point out that the class of TERGMs that can be mapped into KIMs, and benefit of the corresponding score driven extensions, is not restricted to cases with linear dependency on the lagged adjacency matrix. In fact, we can also consider network's statistics depending on products of lagged matrix elements, e.g. $q(\mathbf{A}^{(t)},\mathbf{A}^{(t-1)}) = \sum_{ijk}\mathbf{A}_{ik}^{(t)}\mathbf{A}_{ij}^{(t-1)}\mathbf{A}_{ik}^{(t-1)}$, as long as they depend linearly on $\mathbf{A}^{(t)}$. A TERGM with such statistics can be mapped in a KIM with the addition of predetermined regressors and is easily extended, for example, to the corresponding DyNoKIM version. For example, all the statistics discussed as explicit examples in (Hanneke, Fu, Xing, et al., 2010) take this form, and can be mapped into a KIM. Although a full characterization of the set of TERGM's specifications that can be mapped into a KIM lies outside the scope of this work, we suspect it to be very large, and potentially include all statistics commonly used in practice.

In summary, as we have shown that KIM belongs to the TERGM family, and its score driven extensions result in extensions of the corresponding TERGM. That is the case for both DyNoKIM and DyEnKIM, as we showed explicitly for a simple TERGM specification. We believe that our findings open up a vast space of potential

applications of score driven KIM to temporal networks, and raise interesting theoretical questions on the possibility of mapping a generic TERGM into a KIM, that we leave for future explorations.

Moreover, it turns out that a large subset of possible TERGM specifications can be mapped into a KIM. Hence, the score driven KIM that we propose here is an extension of the TERGM allowing its parameters to evolve in time. This frames DyNoKIM also as a contribution to the literature on network models with time varying parameters, alongside with a recent extension of a different, but related, family called Exponential Random Graphs (Holland and Leinhardt, 1981b) to its score driven version (Di Gangi, Bormetti, and Lillo, 2019).

8.5.1 Link Prediction in Temporal Networks with DyNoKIM

As a final contribution we discuss an empirical application to temporal networks data. Specifically, we show that DyNoKIM can be used to model temporal networks. In particular we show that DyNoKIM dynamically provides the level of predictability of links of the network by exploiting again the relation between $\beta^{(t)}$ and AUC.

The problem of link prediction in networks is very important and can be framed in different ways (Wang, Xu, Wu, and Zhou, 2015; Martínez, Berzal, and Cubero, 2016). For discrete time temporal networks, link prediction amounts to forecasting the presence of a link at time $t+1$ given the observations available up to time t. This is easily done with the KIM defining the forecast exactly as in Eq 8.7.

We apply DyNoKIM to a real world temporal network describing close proximity between workers at the Institut National de Veille Sanitaire in Saint-Maurice (Génois et al., 2015). The data was collected with the sensing platform developed by the SocioPatterns (SocioPatterns Research Collaboration, 2008) collaboration and describe situations of face-to-face proximity between pairs of workers lasting at least 20 seconds. The observations cover 10 working days, from June 24 to July 3, 2013. For each day, we construct the time series of adjacency matrices, at a frequency of 20 seconds between 7:30 am and 5:30 pm. A link between two workers is present if they face each other at a distance less than 1.5 meters and is absent otherwise. As is often the case in real temporal networks, a large number of links is never, or very rarely, observed. Since for such trivial links the prediction problem is not interesting, and to keep the computational complexity to a reasonable level, we consider only the subset of the 100 most active links in each day. For each day, we estimate the DyNoKIM on a training set consisting of the first 75% of observations and then use the remaining 25% for out of sample validation. For each t we compute the AUC and report in Fig. 8.6b the aggregated results for all days. As in the financial application, we observe a monotonically increasing relation between $\beta^{(t)}$ and AUC, indicating that DyNoKIM is a reliable tool to dynamically quantify forecast accuracy also in applications to temporal networks data. Also in this case, we observe a good agreement with the theoretical prediction, with differences explainable by the non Gaussianity of the estimated matrix $\mathbf{J}$.

8.6 Discussion

We have applied the score-driven methodology to extend the Kinetic Ising Model to a time-varying parameters formulation, introducing two new models for complex

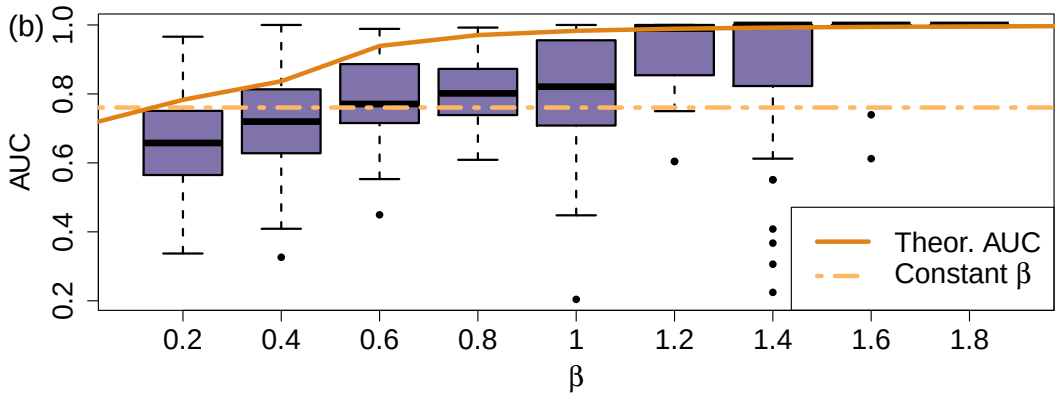

FIGURE 8.6: AUC statistics compared to $\beta^{(t)}$ for link prediction in the SocioPatterns dataset, compared with the theoretical expected AUC and the constant β benchmark.

systems: the Dynamical Noise Kinetic Ising Model and the Dynamic Endogeneity Kinetic Ising Model. We showed that the DyNoKIM, characterized by a time-varying noise level parameter $\beta^{(t)}$, has a clear utility in forecasting applications, as the Area Under the ROC Curve can be showed to be a growing function of $\beta^{(t)}$, while the DyEnKIM can be used to discriminate between endogenous and exogenous effects in the evolution of a multivariate time series.

We then provided example applications of the two models. We successfully employed the DyNoKIM to quantify the real-time forecasting accuracy of stock price activities in the US stock market. The result, largely matching the predictions from theory and simulations, is a methodological breakthrough for the real-world application of time-varying parameter models of complex systems, opening to the possibility of implementing real-time indicators quantifying the accuracy of model-based predictions.

We have then applied the DyEnKIM to model a population of salamander retina neurons and describe the high-frequency volatility of US stocks in proximity of extreme events such as the Flash Crash of May 6, 2010 or around scheduled announcements as the FOMC report of July 31, 2019. We designed the DyEnKIM to disentangle the effects of interactions from the ones of exogenous sources on the observed collective dynamics, a task that is typically non-trivial but nonetheless fundamental in the modeling of complexity. Our results show that this distinction can be made regardless of the underlying system, providing a detailed description and insight on the dynamics, and most importantly without requiring multiple controlled experiments, as is common practice in previous applications of the KIM on neuron populations, thus opening to the adoption of the model in contexts where running repeated experiments is costly or impossible.

Finally, we discussed the possibility to use KIM and our score driven extensions to model temporal networks and discussed an empirical application of the DyNoKIM to the real-time link prediction in a temporal social network.

In conclusion, the Score-Driven KIM poses the foundations for a new modeling paradigm in complex systems. We foresee several relevant extensions such as the modeling of non binary data, for example extending to a Potts-like model (Binder,

1981), or to non-Markovian settings. The key advantages provided by the score-driven methodology in terms of ease of estimation and minimization of model misspecification errors open to the implementation of more accurate and versatile models, interesting a wide range of disciplines that look to describe and unravel complexity from empirical observations.

8.7 Appendix

Data

US stock prices data provided by LOBSTER academic data - powered by NASDAQ OMX. The data consists of the reconstructed Limit Order Book (LOB) for each US stock with timestamps at millisecond precision. We take the mid-price (*i.e.* the average between the best ask and the best bid prices in the LOB) as a real-time proxy of the price, as done in (Rambaldi, Pennesi, and Lillo, 2015). Press reports about the analyzed market events can be found on financial media outlets. FOMC meeting reports are publicly available at *federalreserve.gov*. The salamander retina neuron data has been collected by Prof. Michael J. Berry II and made publicly available at *doi:10.15479/AT:ISTA:61*. It consists of measurements from 160 salamander retina ganglion cells collected through a multi-electrode array. The cells are responding to a light stimulus in the form of a 20 s naturalistic movie and the experiment is repeated 297 times. The electrical signal has been preprocessed to obtain a binary time series for each neuron with time resolution of 20 ms, identifying time intervals where the neuron has produced at least one spike with a 1, and 0 otherwise. From the public dataset we selected the 40 neurons with highest average spike rate over the 297 repeats of the experiment.

The data describing situations of face to face proximity between individuals in the workplace, is provided by the SocioPatterns (SocioPatterns Research Collaboration, 2008) collaboration. It was collected, over a period of two weeks, in one of the two office buildings of the Institut National de Veille Sanitaire (InVS), located in Saint Maurice near Paris, France. Two thirds of the total staff agreed to participate to the data collection. They were asked to wear a sensor on their chest, that allow exchange of radio packets only when the persons are facing each other at a range closer than 1.5 m. By design, any contact that lasted at least 20 seconds was recorded with a probability higher than 99%. In our temporal network application, we associate a node to each individual, and assign a link between two workers if they face each other at a distance less than 1.5 meters. We then consider only the subset of the 100 most active links in each day.

Details on the inference method

As mentioned in the main text our estimation procedure is done in steps, starting by estimating the parameters $\Theta = (J, h)$ of the standard KIM and then running a targeted estimation for the w, b and a parameters. In this Appendix we provide some further details about this procedure.

The whole process can be summarized as the maximization of the log-likelihood $\mathcal{L}(\Theta, b^{(t)}, w, b, a)$ of the model in question, which in the case of the DyNoKIM reads (setting as usual $h_i = 0\ \forall i$)

$$\mathcal{L}(\Theta, \beta^{(t)}, w, b, a) = \sum_{t=1}^{T} \left\{ \sum_i \left[b^{(t)} \sum_j s_i^{(t)} J_{ij} s_j^{(t-1)} \right] - \log K^{(t)} \right\} \tag{8.14a}$$

$$\text{with} \quad \log \beta^{(t+1)} = w + b \log \beta^{(t)} + a \boldsymbol{\mathcal{I}}^{-1/2^{(t)}} \nabla_t \tag{8.14b}$$

and the definitions of the various quantities are given in the main text. The log-likelihood shown above has a recursive form, as each term in the sum of Eq. 8.14a depends on $\beta^{(t)}$, which is determined recursively through Eq. 8.14b from a starting condition $\beta(1)$. This means that, if one were to maximize $\mathcal{L}$ with respect to all the parameters by applying a standard Gradient Descent method, at each computation of $\mathcal{L}$ and its gradient it would be necessary to compute the recursion, resulting in a slow and computationally cumbersome process. In order to make the estimation quicker we implement our multi-step procedure, relying on existing methods for the estimation of the standard KIM and of observation-driven models.

Our first step consists of maximizing $\mathcal{L}$ with respect to the standard KIM parameters Θ. This is done adopting the Mean Field approach of Mézard and Sakellariou (Mézard and Sakellariou, 2011), which is both fast and accurate in the estimation of fully connected models. We refer the interested readers to the original publication for further details on the method itself. In the specific case of neuron spike data, the Mean Field method fails numerically and we resort to standard Gradient Descent methods. The main reason to detach this step from the optimization of the complete log-likelihood is that Θ contains a large number of parameters: if one can get an estimate for those without recurring to slow and hard to tune Gradient Descent methods the computational cost of the inference reduces significantly.

Given the values of Θ obtained in the first step, we then move to the targeted estimation of w, b and a. This consists in first estimating a target value $\bar{f}$ for the unconditional mean of $f^{(t)} = \log \beta^{(t)}$ and then optimize w, b and a maintaining the ratio $w/(1-b) = \bar{f}$ fixed. To estimate $\bar{f}$ we maximize the log-likelihood of Eq. 8.14a temporarily imposing $a = b = 0$, hence Eq. 8.14b becomes $\log \beta^{(t)} = \bar{f} = \text{const}$. Finally, given this target value we optimize $\mathcal{L}$ with respect to w, b and a maintaining the ratio $w/(1-b) = \bar{f}$ fixed and setting $f(1) = \bar{f}$ to start the recursion of Eq. 8.14b. During these last two steps we use the ADAptive Momentum (ADAM) (Kingma and Ba, 2014) Stochastic Gradient Descent method as optimization algorithm, as we found in our case it had better performance with respect to other available methods.

This targeted estimation is not necessary - one could directly estimate w, b and a together - but it is a standard procedure in the estimation of observation-driven models like the GARCH (Francq, Horvath, and Zakoïan, 2011), as it typically reduces the total number of iterations of gradient descent.

We point out one last remark concerning the indetermination of $\langle \beta \rangle$ in the model of Eq. 3 in the main text (and of $\langle \boldsymbol{\beta} \rangle$ for the DyEnKIM), which is crucial to understand the results of our simulations. The fact that these values cannot be identified is not problematic *per se*, but requires caution when comparing models and filtered parameters across different samples, or when comparing estimates with simulations. To avoid misleading results, one needs to enforce the sample mean of the filtered $\beta^{(t)}$ (or of each of the elements of $\boldsymbol{\beta}^{(t)}$ in the DyEnKIM) to be equal to a reference value, which without loss of generality we pick to be $\langle \beta \rangle = 1$. This is easily done by running the estimation and filtering, then measuring $\langle \beta \rangle$ and rescaling $\beta'^{(t)} = \beta^{(t)} / \langle \beta \rangle$. To leave the model unchanged an opposite rescaling is needed for the parameters $\mathbf{J}$ and h, each having to be multiplied by $\langle \beta \rangle$ themselves. This transformation does not change the log-likelihood, thus the model parameters are still MLE, but crucially allows to set a reference value for β that solves the indetermination.

Given this remark, in all the simulations we show where the data generating process of $\beta^{(t)}$ is misspecified we generate its values making sure that their sample mean is 1. By doing so we do not lose any generality in our results, as the indetermination

needs to be solved for the data generating process too if one wants to obtain meaningful results, and we are able to correctly compare the simulated values of $\mathbf{J}$, h and $\beta^{(t)}$ with the ones that are estimated by the score-driven model. Notably, since the model is misspecified, this cannot be achieved during estimation by enforcing the targeted unconditional mean to be equal to 1, as the score in that case is not a martingale difference and thus the unconditional mean of the score-driven parameter is ill-defined itself, as shown by Creal et al. (Creal, Koopman, and Lucas, 2013).

Derivation of the theoretical AUC

Here we expand on the derivation of the theoretical Area Under the ROC Curve shown in Fig. 1 in the main text. A ROC curve is a set of points $(FPR(\alpha), TPR(\alpha))$, with $\alpha \in [0,1]$ being a free parameter determining the minimum value of $p(s_i^{(t)} = +1|s^{(t-1)}; \beta, \Theta)$ which is considered to predict $\hat{s}_i^{(t)} = 1$. If the prediction $\hat{s}_i^{(t)}$ matches the realization $s_i^{(t)}$ then the classification is identified as a True Positive (or Negative, if $p < \alpha$), otherwise it is identified as a False Positive (Negative). The True Positive Rate (TPR) is the ratio of True Positives to the total number of realized Positives, that is True Positives plus False Negatives. Similarly the False Positive Rate (FPR) is the ratio of False Positives to the total number of realized Negatives. Summarizing

$$TPR = \frac{TP}{TP + FN}; \qquad FPR = \frac{FP}{FP + TN}$$

We can explicitly derive the analytical form of the theoretical Area Under the Curve, that is the area that lies below the set of points $(FPR(\alpha), TPR(\alpha))$, assuming the data generating process is well specified and performing some assumptions on the distribution of the model parameters. As a reminder, a classifier having $AUC = 0.5$ is called an *uninformed classifier*, meaning it makes predictions statistically indistinguishable from random guessing, while values of AUC greater than 0.5 are a sign of good forecasting capability.

Following the definition of TPR and FPR one can compute their expected values

$$TPR_\phi(\alpha, \beta) = \frac{1}{K_\phi^+(\beta)} \int_{g_i : p^+ > \alpha} dg_i \phi(g) p^+(\beta, g_i) \tag{8.15a}$$

$$FPR_\phi(\alpha, \beta) = \frac{1}{K_\phi^-(\beta)} \int_{g_i : p^+ > \alpha} dg_i \phi(g) p^-(\beta, g_i) \tag{8.15b}$$

where $K_\phi^\pm(\beta) = p(s_i = \pm 1)$ is a normalization function, $\phi(g)$ is the unconditional distribution of the effective fields g_i and we have abbreviated the probability of sampling a positive or negative value as

$$p^\pm(\beta, g_i) = \frac{e^{\pm \beta g_i}}{2\cosh(\beta g_i)}$$

The definition of the theoretical AUC then reads as

$$AUC_\phi(\beta) = \int_1^0 TPR_\phi(\alpha, \beta) \frac{\partial FPR_\phi(\alpha, \beta)}{\partial \alpha} d\alpha$$

that is the area below the set of points $(FPR(\alpha), TPR(\alpha))$. The lower limit to the integration in Eqs. 8.15 is $g_{min} : p^+(g_{min}) = \alpha$, which is found to be

$$g_{min}(\alpha,\beta) = \frac{1}{2\beta}\log\frac{\alpha}{1-\alpha}$$

Then applying the partial derivative to the definition of FPR it follows that

$$\frac{\partial FPR}{\partial \alpha} = -\frac{1}{K_\phi^-(\beta)}\frac{\partial g_{min}}{\partial \alpha}\phi(g_{min})(1-\alpha)$$

where we have substituted $p^-(\beta, g_{min}) = 1-\alpha$. Plugging all the above results in the definition of AUC_ϕ we then find

$$AUC_\phi(\beta) = \frac{1}{K_\phi^+(\beta)K_\phi^-(\beta)}\int_0^1 d\alpha\left[\int_{g_{min}(\alpha,\beta)}^{+\infty} dg\phi(g)\frac{e^{\beta g}}{2\cosh\beta g}\right]\left[\frac{1}{2\alpha\beta}\phi(g_{min}(\alpha,\beta))\right] \tag{8.16}$$

From an operational perspective $\phi(g)$ is the distribution that the effective fields show cross-sectionally across the whole sample, that is $g_i^{(t)} \sim \phi(g)\ \forall i,t$, but it can also be calculated by giving a prior distribution to the static parameters of the model, $\Theta = (J,h,b)$. Finding this distribution can be useful to provide an easier and more accurate evaluation of the expected AUC of a forecast at a given β value, as it provides a bridge from the model parameters to the $AUC(\beta)$ we derived in Eq. 8.16 and shown in Fig. 1 in the main text.

Let us assume, as is standard in the literature (Crisanti and Sompolinsky, 1988; Roudi and Hertz, 2011; Mézard and Sakellariou, 2011), that the parameters Θ are structured in such a way that

$$J_{ij} \overset{iid}{\sim} \mathcal{N}(J_0/N, J_1^2/N - J_0^2/N^2)$$
$$h_i \overset{iid}{\sim} \mathcal{N}(h_0, h_1^2)$$

If that is the case then the distribution of $g_i^{(t)}$ is itself a Gaussian, as $g_i^{(t)}$ is now a sum of independent Gaussian random variables J_{ij} and h_i with random coefficients $s_j^{(t)}$. Let us also define two average operators: the average $\langle\cdot\rangle$ over the distribution p, also called the *thermal* average (which, the system being ergodic, coincides with a time average for $T\to\infty$), and the average $\overline{\cdot}$ over the distribution of parameters, also known as the *disorder* average. Following Mézard and Sakellariou (Mézard and Sakellariou, 2011) we can then find the unconditional mean of s_i which reads

$$m_i = \langle s_i^{(t)}\rangle = \left\langle \tanh\left[\beta g_i^{(t)}\right]\right\rangle \tag{8.17}$$

where we have substituted the conditional mean value of $s_i^{(t)}$ inside the brackets. This depends from the distribution of $g_i^{(t)}$: assuming stationarity and calling $g_i^0 = \langle g_i^{(t)}\rangle$ and $\Delta_i^2 = \langle g_i^{2(t)}\rangle - \langle g_i^{(t)}\rangle^2$ we find that they are

$$g_i^0 = \langle \sum_j J_{ij}s_j^{(t)} + h_i\rangle = \sum_j J_{ij}m_j + h_i \tag{8.18a}$$

$$\Delta_i^2 = \left\langle\left(\sum_j J_{ij}s_j^{(t)} + h_i\right)^2\right\rangle - \left\langle\sum_j J_{ij}s_j^{(t)} + h_i\right\rangle^2 = \sum_{j,k} J_{ij}J_{ik}\left[\left\langle s_j^{(t)}s_k^{(t)}\right\rangle - m_jm_k\right] \tag{8.18b}$$

In Eq. 8.18b spins $s_j^{(t)}$ and $s_k^{(t)}$ are mutually conditionally independent under distribution p: this means that the only surviving terms are the ones for $j = k$, and thus we find

$$\Delta_i^2 = \sum_j J_{ij}^2 (1 - m_j^2) \tag{8.19}$$

Having determined the value of the mean and variance of the effective field of spin i we can now proceed to average over the disorder and find the unconditional distribution of effective fields at any time and for any spin, $\phi(g)$. First we realize that the average of Eq. 8.17 can now be substituted by a Gaussian integral

$$m_i = \int Dx \tanh\left[\beta\left(g_i^0 + x\Delta_i\right)\right] \tag{8.20}$$

where Dx is a Gaussian measure of variable $x \sim \mathcal{N}(0,1)$. Then we can see that the unconditional mean of the fields distribution $\phi(g)$ is

$$g_0 = \overline{\langle g_i^{(t)} \rangle} = \overline{\sum_j J_{ij} m_j + h_i} \tag{8.21}$$

Given the above results and the definition of $\mathbf{J}$, the dependency between J_{ij} and m_j vanishes like $O(1/N)$, which means that the two can be averaged over the disorder separately in the limit $N \to \infty$. This results in the following expression for the unconditional mean of $g_i^{(t)}$

$$g_0 = J_0 \overline{m_j} + h_0 = J_0 m + h_0 \tag{8.22}$$

where

$$m = \overline{m_i} = \overline{\int Dx \tanh\left[\beta(g_i + x\Delta_i)\right]}$$

both the integral and the average here are of difficult solution and results have been provided by Crisanti and Sompolinsky (Crisanti and Sompolinsky, 1988): they show that in the limit $N \to \infty$ and with $h_i = 0\ \forall i$ the system can be in one of two phases, a paramagnetic phase where $m = 0$ if β is smaller than a critical threshold $\beta_c(J_0)$ and $J_0 < 1$, and a ferromagnetic phase where $m \neq 0$ otherwise. In the following we report results for simulations in the paramagnetic phase, as the inference is not possible in the ferromagnetic phase. To give better intuition let us consider the integral above in the limit $\beta \to 0$: then we can expand the hyperbolic tangent around 0 to find (since x has zero mean)

$$m \approx \overline{\beta g_i} = \overline{\beta\left(\sum_j J_{ij} m_j + h_0\right)} = \beta(J_0 m + h_0) \tag{8.23}$$

which in turn leads to an approximated solution for g_0 in the limit $\beta \to 0$

$$g_0 \approx h_0\left(\frac{\beta J_0}{1 - \beta J_0} + 1\right)$$

Moving on to the variance of g the calculation is straightforward. Adding the mean over the disorder to Eq. 8.18b we find

$$
\begin{aligned}
g_1^2 &= \overline{\left\langle \left[\sum_j J_{ij} s_j^{(t)} + h_i \right]^2 \right\rangle} - \overline{\left\langle \sum_j J_{ij} s_j^{(t)} + h_i \right\rangle}^2 = \\
&= \overline{\sum_j J_{ij}^2 + h_i^2 + 2h_i \sum_j J_{ij} m_j} - \overline{\sum_j J_{ij} m_j + h_i}^2 = \\
&= J_1^2 + h_1^2 - J_0^2 m^2
\end{aligned} \tag{8.24}
$$

Equations 8.22 and 8.24 can then be used to calculate, given the parameters of the distribution generating Θ, the values of g_0 and g_1 that are to be plugged in the distribution $\phi(g)$ of Eq. 8.16

We simulated a Kinetic Ising Model with $N = 100$ spins for $T = 2000$ time steps at different constant values of β and then measured the AUC of predictions assuming the parameters are known. In Fig. 8.7 we report a comparison between these simulated values and the theoretical ones provided by Eq. 8.16 varying β and the hyperparameters J_0, J_1, h_0 and h_1 in the Gaussian setting we just discussed and adopting the expansion for $\beta \to 0$. We see that the approximation for small β of Eq. 8.23 does not affect the accuracy of the theoretical prediction for larger values of β and that the mean is correctly captured by Eq. 8.16. The only exception to this is found for $\beta > 1$ and $J_0 = 1$, which according to the literature is close to the line of the ferromagnetic transition: in this case the small β approximation fails to predict the simulated values. Larger values of N and T (not shown here) produce narrower error bars.

The general effect we see from Fig. 8.7 is that higher variance of the $\mathbf{J}$ and h parameters leads to higher AUC values leaving all else unchanged (orange squares and yellow circles), while moving the means has little effect as long as the system is in its paramagnetic phase.

These results are easy to obtain thanks to the assumption that the model parameters $\mathbf{J}$ and h have Gaussian distributed entries, but in principle the distribution $\phi(g)$ can be derived also for other distributions, albeit probably requiring numerical solutions rather than the analytical ones we presented here.

Further details on the DyEnKIM

There are a couple of subtleties that need to be pointed out regarding the structure of the b and a parameters and of the Fisher Information $\boldsymbol{\mathcal{I}}$ of the DyEnKIM, which are matrices rather than scalars as in the case of the DyNoKIM.

In order to make the estimation less computationally demanding in our example applications we assume a, b and $\boldsymbol{\mathcal{I}}$ diagonal, disregarding the dependencies between time-varying parameters: this will likely make our estimates less precise, but it also reduces the number of static parameters to be inferred, letting us bypass model selection decisions which are outside the scope of this chapter.

As previously discussed there is also in this case the problem of identification for the averages of the components of $\boldsymbol{\beta}$, which we solve in the exact same way as we did for the DyNoKIM by dividing the values of each component by their sample mean while multiplying the associated static parameter by the same factor, again leaving the likelihood of the model unchanged, but setting a reference level for $\boldsymbol{\beta}$.

As a last remark, notice that the DyNoKIM and the DyEnKIM are equivalent when $h_0^{(t)} = 0 \; \forall \, t$ and $\beta_{diag} = \beta_{off} = \beta_h = \beta$. For this reason we mainly present simulation results for the DyNoKIM alone to keep the manuscript concise, as we found

no significant differences between the two models when it comes to the reliability of the estimation process.

Consistency analysis for estimation

We perform a consistency test on simulated data, aimed at understanding whether the two-step estimation procedure we outlined above is able to recover the values of the parameters of the model when the model itself generated the data.

Here we report results for simulations run with parameters $N = 50$, $T = 750$ or $T = 1500$, $J_{ij} \sim \mathcal{N}(0, 1/\sqrt{N})$, $h_i = 0 \ \forall i$, $b = 0.95$ and $a = 0.01$. We see from Fig. 8.8 that the estimation of the elements of $\mathbf{J}$ is indeed consistent: we estimate a linear regression model between the estimated and the true values of J_{ij}, namely $J_{ij}^{est} = c_0 + c_1 J_{ij}^{true} + \epsilon$, and plot the histogram of the values of c_1 and of the coefficient of determination R^2 of the resulting model from 250 simulations and estimations (c_0 is consistently found to be very close to 0 in all our simulations and for this reason we omit it). In the ideal case where for any i, j $J_{ij}^{est} = J_{ij}^{true}$ one would have $c_1 = R^2 = 1$, which is what we aim for in the limit $T \to \infty$. We see from our results that there is indeed a convergence of both values towards 1 when increasing sample size, reducing both the bias and the variance of the regression parameters.

Turning to the score-driven dynamics parameters a and b, the situation does not change significantly. In Fig. 8.9 we show the histograms of estimated values of b and a over 250 simulations of $N = 50$ variables for both $T = 750$ and $T = 1500$. It again appears clearly that when increasing the sample size the bias and variance of the estimators converge towards 0, with the estimated parameter converging towards its simulated value. Thanks to these results we are able to confidently apply the two-step estimation method without the need to estimate all the parameters at once.

To add further evidence to what we presented in the main text, here we also report two additional figures regarding the filtering of misspecified $\beta^{(t)}$ for the DyNoKIM and the DyEnKIM. In Fig. 8.10 we show two examples of misspecified $\beta^{(t)}$ dynamics that are correctly recovered by the score-driven approach: the first is a deterministic sine wave function and the second is an AutoRegressive model of order 1 (AR(1)) which follows the equation

$$\beta^{est\,(t+1)} = c_0 + c_1 \beta^{est\,(t)} + \epsilon^{(t)}$$

where $\epsilon^{(t)} \sim \mathcal{N}(0, \Sigma^2)$ with parameters $c_0 = 0.005$, $c_1 = 0.995$, $\Sigma = 0.01$ so to have $\langle \beta^{est} \rangle = 1$ and we select a simulation where $\beta^{(t)} > 0 \ \forall \ t$. In both cases we simulate 30 time series of length T using the given values of $\beta^{(t)}$ to generate the $s^{(t)}$; given only the simulated $s^{(t)}$ time series, the inference algorithm determines the optimal static parameters a, b and $\mathbf{J}$ and filters the optimal value of $\beta^{(t)}$ at each time. We see that regardless of whether the underlying true dynamics is deterministic, stochastic, or more or less smooth the filter is rather accurate in retrieving the simulated values.

Regarding the DyEnKIM we want to show that different effects are correctly separated and identified when estimating the model on a misspecified data generating process. In fact while the consistency analysis largely resembles the one we reported for the DyNoKIM in Figures 8.8 and 8.9 and for this reason we omit it, the effect of filtering multiple time-varying parameters is something that cannot be predicted by the simulations on the DyNoKIM alone.

In Figure 8.11 we show the results when estimating the DyEnKIM on a dataset generated by a Kinetic Ising Model with time-varying $\beta_{diag}^{(t)}$, $\beta_{off}^{(t)}$ and $\beta_h^{(t)}$ but where

the dynamics of the parameters is predetermined instead of following the score-driven update rule. We arbitrarily choose to take a constant $\beta^{(t)}_{diag} = 1$, a piecewise constant $\beta^{(t)}_{off}$ and an exponentiated sinusoidal $\beta^{(t)}_{h} = \exp[\sin(\omega t)]$, with $\omega = 5\frac{2\pi}{T}$, $T = 1500$ and $N = 30$. The results show that the filter works correctly and that the different time-varying parameters are consistently estimated, regardless of the kind of dynamics given to each of them.

Additional results on neural population data

In this section we provide further results on the application of the KIM to the neural population data. In particular we discuss the possibility to use the DyEnKIM to test the significance of the elements of **J** fitted over multiple experiments and expand on the comparison between the static parameters KIM and the score-driven version.

In Figure 8.12a we show an example of the analyzed data in the form of a raster plot of one of the experiments. It appears clear that the dynamics of spikes is bursty and there are non-negligible auto- and cross-correlation effects among neurons, likely driven by the external stimulus of the video (Tkačik et al., 2014).

As mentioned in the section on the DyNoKIM in the main text, it is possible to use the filtered time-varying parameters to correct the values of **J** from misspecification error. The same can be done with the DyEnKIM, correcting the elements of **J** by a factor given by the sample mean of β_{diag} and β_{off} and the external fields h by the sample mean of β_h. In Figure 8.12b we show a scatterplot of the values of J_{ij}, comparing the parameters fitted with a KIM on all available data (x-axis) with the average value $\bar{J}_{ij} = \frac{1}{M}\sum_k^M J^{(k)}_{ij}$, where $J^{(k)}_{ij}$ is the value fitted with the DyEnKIM correction on experiment k. It is clear that, as in the case of DyNoKIM shown in Figure 2 of the main text, the correction - which in this case only affects off-diagonal elements as $\beta_{diag} = 1$ - tends to increase the absolute value of J_{ij} with respect to the static KIM version.

Pruning irrelevant parameters is central to the definition of meaningful statistical models. Here we propose two alternative methods, Decimation and t-testing; the first is standard in the literature on Kinetic Ising Models (Decelle and Zhang, 2015), whereas the second exploits the repeated experiments in the data to compare parameters fitted on different samples and assess their significance by means of a t-test. For Decimation we refer the interested readers to the original paper introducing it (Decelle and Zhang, 2015). As an alternative in cases where multiple repetitions of the experiment are available, as is the case for our neuron spike dataset, it is possible to fit a DyEnKIM for each of the M experiments and then use a Student's t-test on the set of values J^k_{ij}, $k = 1, \dots, M$ to test whether their average is significantly different from 0. In Figure 8.12c we show a comparison between these two methods, with Decimation applied to the KIM **J** and the t-test used to validate the DyEnKIM result, adopting a Bonferroni correction for multiple hypothesis testing at the $p < 0.01$ level. In our case, the Decimation approach selects less elements of J_{ij} as significant, whereas the t-test appears to be less specific or more sensitive. This difference is possibly related to Decimation being a likelihood-based method, which may suffer from misspecification in case the data generating process is not a KIM, but answering this question goes beyond the scope of this work. Finally, in Figure 8.12d we report a visualization of the t-tested DyEnKIM $\bar{J}$ matrix. The diagonal elements are largely positive, indicating significant autocorrelation in spiking dynamics (as would be expected by a visual inspection of the raster plot), whereas

off-diagonal elements are generally smaller in absolute value and both positive and negative, albeit significantly different from 0.

Results on the application of DyEnKIM on regular trading days

As a comparison with the observations reported in the main text regarding particular events affecting stock markets, such as the Flash Crash and the FOMC announcement of July 31st 2019, here we briefly discuss observations for a regular trading day where nothing as exceptional happened. In Figure 8.13 we replicate the plot shown in Figure 5 of the main text for six days in November 2019. Here we see that the **J**-related components of effective fields $\langle g_{diag} \rangle^{(t)}$ and $\langle g_{off} \rangle^{(t)}$ show a U-shaped pattern throughout the trading day, having higher values at the opening and closing, while the h-related $\langle g_h \rangle$ only shows an increase towards the end of the day. The h_0 parameter, which captures the average exogenous price activity across all stocks, shows itself a U-shaped pattern which is more pronounced at closing, consistent with the intraday pattern typical of traded volume. The consistency of this result throughout these relatively uneventful days thus reinforces the qualitative description provided by the DyEnKIM for the turbulent events analyzed in the main text.

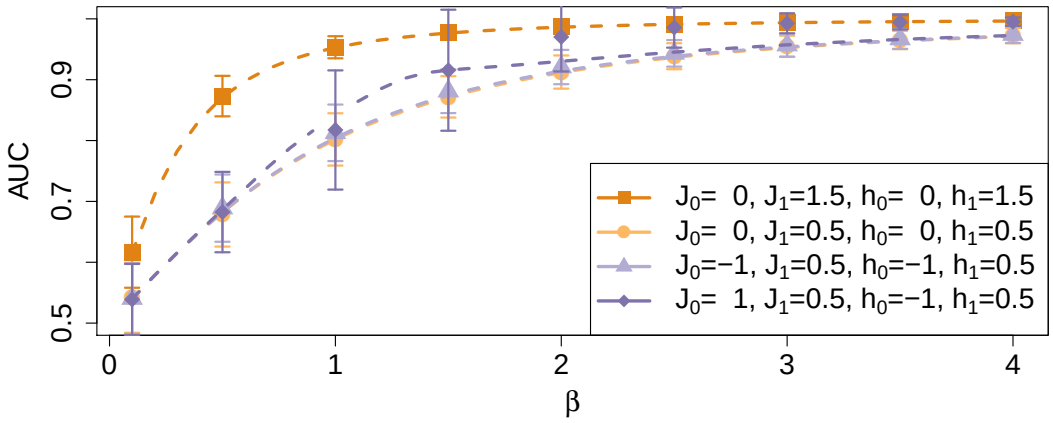

FIGURE 8.7: Comparison between the AUC estimated on data simulated from a Kinetic Ising Model and the theoretically derived AUC with Gaussian distribution of the **J** and h parameters, varying β and the hyperparameters J_0, J_1, h_0 and h_1. Plot points report average simulated values for a given β with error bars at ± 1 standard deviation, dashed lines report theoretical values predicted by Eq. 8.16.

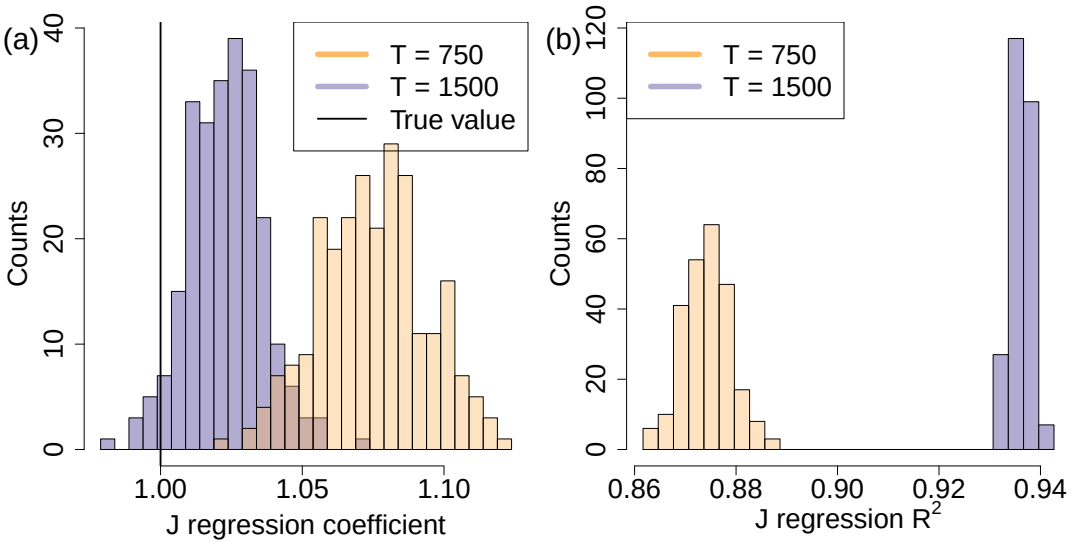

FIGURE 8.8: Consistency of the **J** matrix estimation. (a) Histogram of linear regression coefficients b between inferred and true values of J_{ij} over 250 samples for $N = 50$, $T = 750$ and $T = 1500$; (b) Histogram of coefficients of determination (R^2) for the same set of models. The convergence of both values towards 1 when increasing T is a sign of consistency of the estimation.

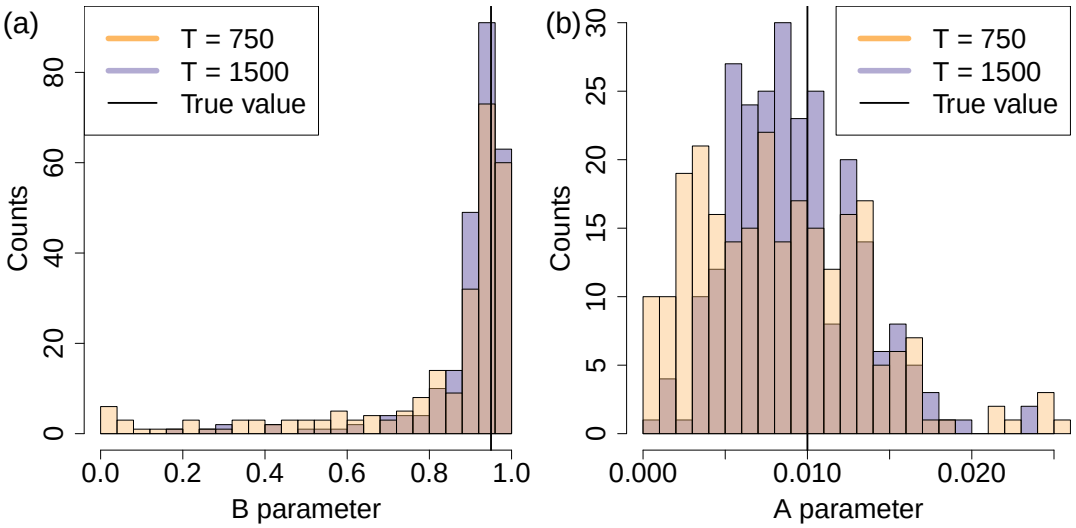

FIGURE 8.9: Consistency of the score-driven dynamics parameters. (a) Histogram of estimated values of b over 250 samples for $N = 50$, $T = 750$ and $T = 1500$; (b) Histogram of estimated values of a over 250 samples for the same set of models. The convergence towards the true value by increasing T is a sign of consistency of the estimation.

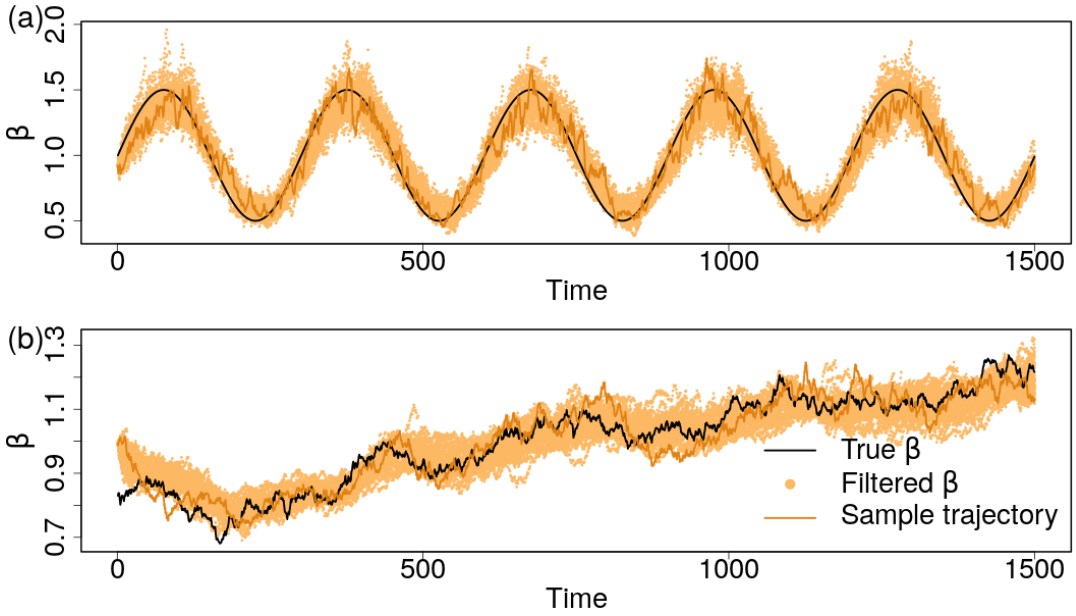

FIGURE 8.10: Simulation and estimation of a misspecified score-driven model over 30 simulations, with sample trajectories highlighted. (a) Deterministic β following a sinusoidal function; (b) Stochastic $\beta^{(t)}$ following an AutoRegressive model of order 1.

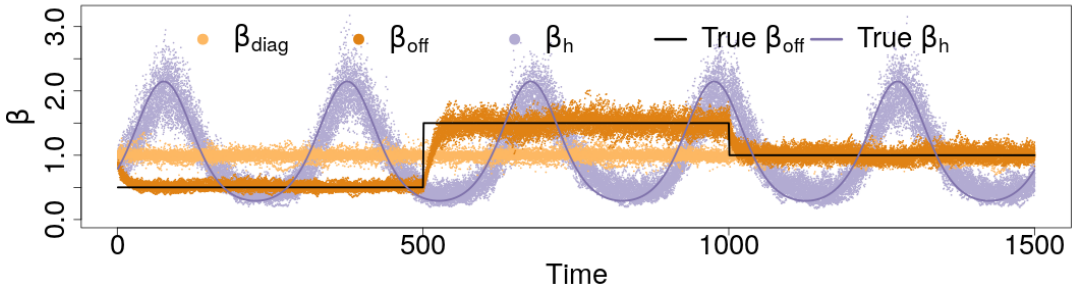

FIGURE 8.11: Estimation of $\beta^{(t)}_{diag}$, $\beta^{(t)}_{off}$ and $\beta^{(t)}_{h}$ under model misspecification. The model was simulated with a constant $\beta^{(t)}_{diag} = 1$, a piece-wise constant $\beta^{(t)}_{off}$ and an exponentiated sinusoidal $\beta^{(t)}_{h} = \frac{1}{\mathcal{J}_0(1)} \exp[\sin(\omega t)]$, with $\omega = 5\frac{2\pi}{T}$ and $\mathcal{J}_0$ the Bessel function of first kind of order 0 to normalize the mean. The points are the result of 30 different simulations and estimations, the lines show the values of β_{off} and β_h used to generate the data.

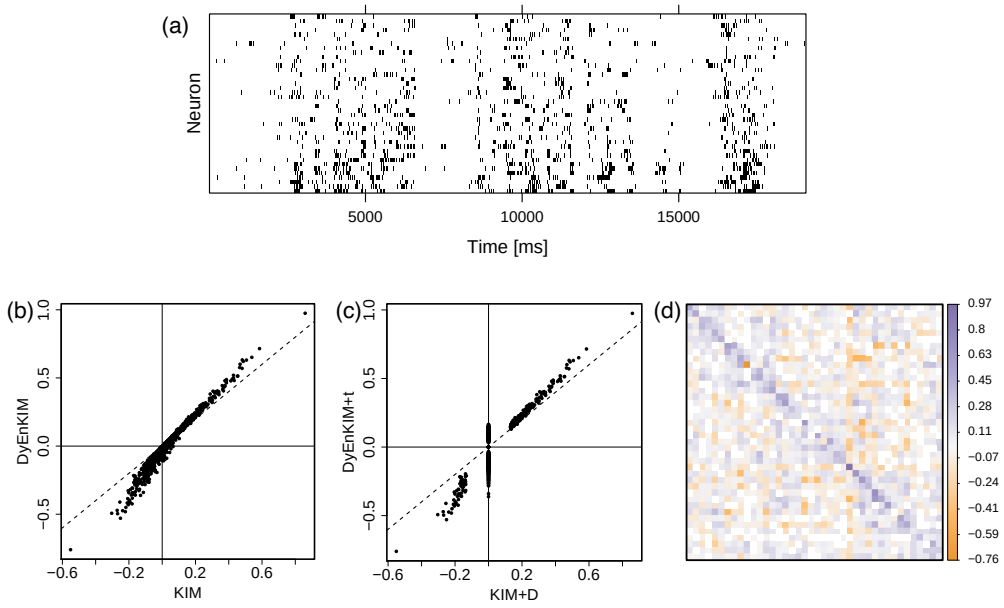

FIGURE 8.12: Additional results on the neuron spike data. (a) Raster plot for one sample in the data. Time is on the x-axis, neurons on the y-axis, a black dot at (t,i) indicates a spike from neuron i at time t; (b) Comparison between fitted values of J_{ij} using a KIM over the complete dataset (x-axis) and the average fitted values of J_{ij} with DyEnKIM correction, $\overline{J}_{ij} = \frac{1}{M}\sum_i^M J_{ij}^{(k)}$, where $J_{ij}^{(k)}$ is the $\mathbf{J}$ matrix fitted on sample k and M is the total number of experiments (y-axis). A dashed line is traced on the diagonal as guide to the eye; (c) The same as panel b after the Decimation pruning technique has been applied to the KIM and the t-test pruning has been applied to the DyEnKIM; (d) Visualization of the DyEnKIM $\overline{J}$ after the t-test pruning has been applied. White squares correspond to non-validated interactions.

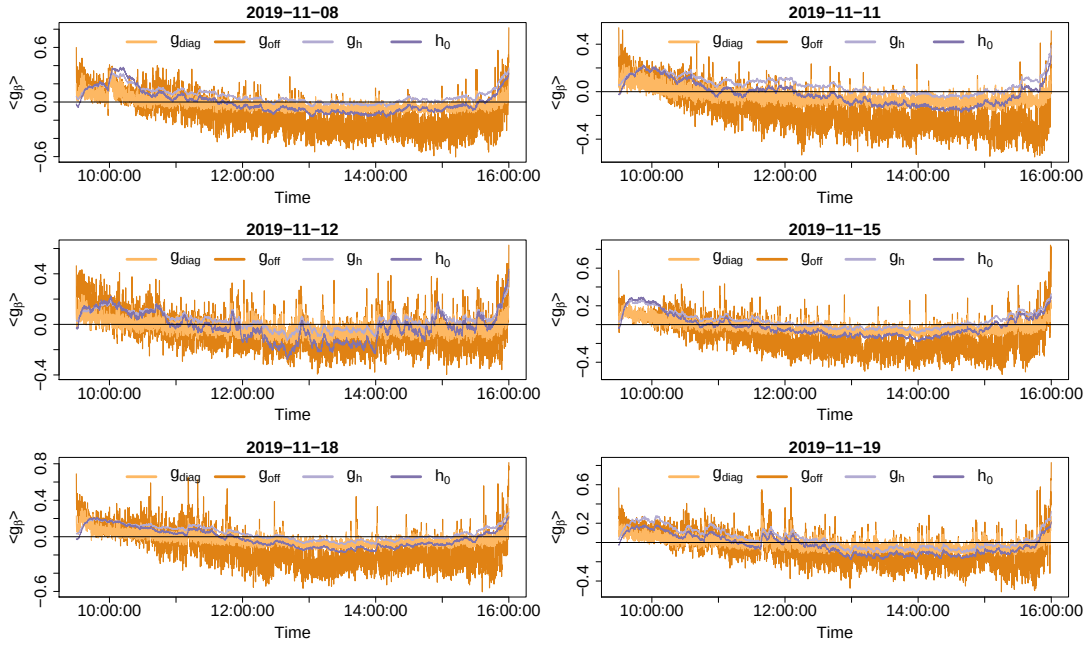

FIGURE 8.13: Values of $\langle g_{diag} \rangle^{(t)}$, $\langle g_{off} \rangle^{(t)}$, $\langle g_h \rangle^{(t)}$ and $h_0^{(t)}$ during six days in November 2019, when no abnormal event was recorded. The usual U-shaped pattern of intraday volatility and volume is observed.

Chapter 9

Conclusions

In this thesis we presented contributions to the literature on statistical models for both static and temporal networks. All our works propose novel methodologies with clear applications to multiple fields that are validated trough extensive numerical simulations and are then applied to a number of different real world systems.

In Chapter 3 we focused on the problem of estimating metrics of systemic risk due to fire sale spillover in presence of limited information on the composition of portfolios of financial institutions. A full knowledge of the portfolio holdings of each institution in the economy is generally required to have a precise estimate of any the risk metrics by Greenwood, Landier, and Thesmar, 2015 that is based on the mechanism of portfolio rebalancing through fire sales. Nevertheless, such a huge and detailed information may not be available at the required frequency. We circumvented the problem by providing accurate estimates of systemic risk metrics that are based on a partial knowledge of the system, more precisely only on the sizes of balance sheets and the capitalization of assets (or asset classes), which are much easier to trace. In this respect, we showed that the method of Cross-Entropy minimization does a very good job in estimating aggregate vulnerability and individual bank systemicness without requiring any knowledge of the underlying matrix of bank portfolio holdings. Furthermore, we compared the results with a Max Entropy ensembles. Specifically we introduced a new ensemble (MECAPM), which reproduces, on average, the CECAPM and performs quite well in estimating systemicness and indirect vulnerability of single institutions, outperforming standard Max Entropy competitors. The contributions presented in this chapter fit well within the literature on the reconstruction of networks and systemic risk from partial information. Like most works in the field (Anand et al., 2017), we considered only the common setting where aggregate information at the node level is available. In the real world different situations where regulators have incomplete knowledge on financial networks often arise. The first example that comes to mind is the setting where a regulator, e.g. a central bank, only has knowledge of the portion of the network that falls under its supervision. Such a setting is not uncommon as often regulators have no data on the financial connections between entities outside their jurisdiction. This situation might lead to inaccurate estimations of systemic risk that still need to be addressed properly using methods similar to the ones presented in Chapter 3.

In Chapter 6 we proposed a framework for the description of temporal networks that extends the well known Exponential Random Graph Models. In the new approach, the parameters of the ERGM have a score driven dynamics. If the log-likelihood is not available in closed-form, we showed how to adapt the score-driven updating rule to a generic ERGM by resorting to the conditional pseudo-likelihood. In this way, our approach can describe the dynamic dependence of the PMF from virtually all the network statistics usually considered in ERGM applications. We investigated two specific ERGM instances by means of an extensive Monte Carlo

analysis of the SD-ERGM reliability as a filter for time-varying parameters. The chosen examples allowed us to highlight the applicability of our method to models with a large number of parameters and to models for which the normalization of the PMF is not available in closed form. The numerical simulations proved the clear superior performance of the SD-ERGM over a sequence of standard cross sectional ERGM estimates. This is not only true in the sparse network regime, but also in the dense case when the number of nodes is far from the asymptotic limit. Finally, we run two empirical exercises on real networks data. The first application to e-MID interbank network showed that the SD-ERGM provides a quantifiable advantage in a link forecasting exercise over different time horizons. The second example on the U.S. Congress co-voting political network enlightened that the ERGM and the SD-ERGM could provide a significantly different picture in describing the parameter dynamics.

Following a similar approach, in Chapter 7 we focused on describing also links' weights. We proposed a novel model for sparse weighted temporal networks that also accommodates for the dependency of the network dynamics on external variables, and its application to weighted temporal network data, describing overnight exposures in the European interbank market. Our work contributes to the extremely scarce literature on dynamical models for sparse weighted networks by extending the very well known fitness model for static binary networks. We considered a zero augmented generalized linear model to handle the weights and the score driven framework to describe time varying parameters. This results in a flexible model that allows us to decouple the probability of a link to exist from its expected weight, and to explore the influence of external covariates on the network's dynamics. We then exploit such flexibility to investigate how the relevance of EONIA rates on the e-Mid interbank market changed over time. We started from the well known *fitness model* Caldarelli, Capocci, De Los Rios, and Muñoz, 2002; Garlaschelli and Loffredo, 2008, also known as *beta model* or *configuration model*, and considered its extension to describe the weights and the dependency on external covariates. Alongside the standard, *binary fitness* parameters, we associate to each node i two new parameters that we call *weighted fitness*. They describe the propensity of a node to have more or less heavy weights in incoming and outgoing links respectively, and use them to model the weighted adjacency matrix using a zero augmented distribution. We then extended this model to the dynamical context by allowing the fitness, both binary and weighted, and the regression parameters to change over time, following the Score Driven approach. We run extensive numerical simulations to make sure that this update rule defines an effective way to *filter* the time varying parameters, also when their temporal evolution is governed by a different DGP. We empirically applied our model running an exercise in weight forecasting and to explore the relation between the dynamics of a portion of the European interbank market, e-Mid, and the EONIA rate. Finally, we explored the persistence of weights in the e-Mid market.

In both Chapter 6 and Chapter 7 we introduced novel methodological contributions that are extremely flexible. We proved that they are very effective in a range of numerical experiments, and tested them in three empirical applications. We believe that more empirical applications have yet to be explored to exploit the full potential of our methodologies and validate the added value that they can bring in empirical analysis. Probably the most relevant empirical application that we would foresee is to the literature of network reconstruction from partial information and its extension to take into account the longitudinal dimension of temporal networks. A systematic

exploration of network forecasting from partial information represents an interesting avenue of research that we have yet to investigate.

Finally, in Chapter 8 we applied the score-driven methodology to extend the Kinetic Ising Model to a time-varying parameters formulation, introducing two new models for complex systems: the Dynamical Noise Kinetic Ising Model and the Dynamic Endogeneity Kinetic Ising Model. We showed that the DyNoKIM, characterized by a time-varying noise level parameter $\beta^{(t)}$, has a clear utility in forecasting applications, as the Area Under the ROC Curve can be showed to be a growing function of $\beta^{(t)}$, while the DyEnKIM can be used to discriminate between endogenous and exogenous effects in the evolution of a multivariate time series. We then provided example applications of the two models. We successfully employed the DyNoKIM to quantify the real-time forecasting accuracy of stock price activities in the US stock market. The result, largely matching the predictions from theory and simulations, is a methodological breakthrough for the real-world application of time-varying parameter models of complex systems, opening to the possibility of implementing real-time indicators quantifying the accuracy of model-based predictions. We have then applied the DyEnKIM to model a population of salamander retina neurons and describe the high-frequency volatility of US stocks in proximity of extreme events such as the Flash Crash of May 6, 2010 or around scheduled announcements as the FOMC report of July 31, 2019. We designed the DyEnKIM to disentangle the effects of interactions from the ones of exogenous sources on the observed collective dynamics, a task that is typically non-trivial but nonetheless fundamental in the modeling of complexity. Our results show that this distinction can be made regardless of the underlying system, providing a detailed description and insight on the dynamics, and most importantly without requiring multiple controlled experiments, as is common practice in previous applications of the KIM on neuron populations, thus opening to the adoption of the model in contexts where running repeated experiments is costly or impossible. Last but not least, we discussed the possibility to use KIM and our score driven extensions to model temporal networks and discussed an empirical application of the DyNoKIM to the real-time link prediction in a temporal social network. Although very general the spin like description of a complex systems allows each entity to be in only one of two states. This might be limiting and result in information losses for systems that would be better described by more than two states. Hence, the application of a similar approach, as the one presented in Chapter 8, to Potts like models (Potts, 1952; Wu, 1982) could result in a even wider realm of potential applications with respect to our score driven KIM.

Bibliography

Acharya, Viral, Robert Engle, and Matthew Richardson (2012). "Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks". In: *American Economic Review*.

Ackley, David H, Geoffrey E Hinton, and Terrence J Sejnowski (1985). "A learning algorithm for Boltzmann machines". In: *Cognitive Science*.

Adrian, Tobias and Markus K. Brunnermeier (2011). "CoVar". In: *SSRN, Working Paper 1269446*.

Adrian, Tobias and Hyun Song Shin (2008). "Liquidity and financial contagion". In: *Financial Stability Review*.

— (2010). "Liquidity and leverage". In: *Journal of Financial Intermediation*.

— (2014). "Procyclical Leverage and Value-at-Risk". In: *Review of Financial Studies*.

Agliari, Elena, Adriano Barra, Francesco Guerra, and Francesco Moauro (2011). "A thermodynamic perspective of immune capabilities". In: *Journal of Theoretical Biology*.

Aït-Sahalia, Yacine, Per A Mykland, and Lan Zhang (2011). "Ultra high frequency volatility estimation with dependent microstructure noise". In: *Journal of Econometrics*.

Akram, Q Farooq and Casper Christophersen (2010). *Interbank Overnight Interest Rates-Gains from Systemic Importance*. Working Paper. Norges Bank.

Albert, Réka and Albert-László Barabási (2002). "Statistical mechanics of complex networks". In: *Reviews of Modern Physics*.

Allen, Franklin and Ana Babus (2011). "Networks in Finance". In: *The network challenge: strategy, profit, and risk in an interlinked world*. Ed. by Paul R JKleindorfer, Yoram Jerry R Wind, and Robert E Gunther. Social Science Research Network. Chap. 21.

Allen, Franklin and Douglas Gale (2000). "Financial contagion". In: *Journal of Political Economy*.

Allen, Linda, Turan G. Bali, and Yi Tang (2012). "Does Systemic Risk in the Financial Sector Predict Future Economic Downturns?" In: *Review of Financial Studies*.

Almog, Assaf, Tiziano Squartini, and Diego Garlaschelli (2017). "The double role of GDP in shaping the structure of the International Trade Network". In: *International Journal of Computational Economics and Econometrics*.

Amini, Hamed, Rama Cont, and Andreea Minca (2013). "Resilience to contagion in financial networks". In: *Mathematical Finance*.

Anand, Kartik, Ben Craig, and Goetz Von Peter (2015). "Filling in the blanks: Network structure and interbank contagion". In: *Quantitative Finance*.

Anand, Kartik et al. (2013). "A network model of financial system resilience". In: *Journal of Economic Behavior and Organization*.

Anand, Kartik et al. (2017). "The missing links: A global study on uncovering financial network structures from partial data". In: *Journal of Financial Stability*.

Aquilina, Matteo, Eric Budish, and Peter O'Neill (2020). "Quantifying the high-frequency trading "arms race": A simple new methodology and estimates". In: *United Kingdom Financial Conduct Authority Occasional Paper*.

Arciero, Luca et al. (2016). "How to Measure the Unsecured Money Market: The Eurosystem's Implementation and Validation Using TARGET2 Data". In: *International Journal of Central Banking*.

Arnold, Bruce, Claudio Borio, Luci Ellis, and Fariborz Moshirian (2012). "Systemic risk, macroprudential policy frameworks, monitoring financial systems and the evolution of capital adequacy". In: *Journal of Banking & Finance*.

Banulescu, Georgiana-Denisa and Elena-Ivona Dumitrescu (2015). "Which are the SIFIs? A Component Expected Shortfall approach to systemic risk". In: *Journal of Banking & Finance*.

Bar-Yam, Yaneer (2002). "General features of complex systems". In: *Encyclopedia of Life Support Systems (EOLSS), UNESCO, EOLSS Publishers, Oxford, UK*.

Barabási, Albert-László (2002). *Linked: The New Science Of Networks*. Basic Books.

— (2013). "Network science". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*.

Bargigli, L. et al. (2015a). "The multiplex structure of interbank networks". In: *Quantitative Finance*.

Bargigli, Leonardo and Mauro Gallegati (2011). "Random digraphs with given expected degree sequences: A model for economic networks". In: *Journal of Economic Behavior & Organization*.

Bargigli, Leonardo et al. (2015b). "The multiplex structure of interbank networks". In: *Quantitative Finance*.

Barndorff-Nielsen, Ole (2014). *Information and exponential families in statistical theory*. John Wiley & Sons.

Barrat, Alain, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani (2004). "The architecture of complex weighted networks". In: *Proceedings of the national academy of sciences*.

Barrell, Ray, E. Philip Davis, Dilruba Karim, and Iana Liadze (2010). "Bank regulation, property prices and early warning systems for banking crises in {OECD} countries". In: *Journal of Banking & Finance*.

Barreto, Humberto and Frank Howland (2006). *Introductory econometrics: using Monte Carlo simulation with Microsoft excel*. Cambridge University Press.

Barucca, Paolo and Fabrizio Lillo (2018). "The organization of the interbank network and how ECB unconventional measures affected the e-MID overnight market". In: *Computational Management Science*.

Bauwens, Luc and David Veredas (2004). "The stochastic conditional duration model: a latent variable model for the analysis of financial durations". In: *Journal of Econometrics*.

Beck, Christian and Ezechiel GD Cohen (2003). "Superstatistics". In: *Physica A: Statistical mechanics and its Applications*.

Beck, Christian, Ezechiel GD Cohen, and Harry L Swinney (2005). "From time series to superstatistics". In: *Physical Review E*.

Bera, Anil K. and Sung Y. Park (2008). "Optimal Portfolio Diversification Using the Maximum Entropy Principle". In: *Econometric Reviews*.

Bernardi, Mauro and Leopoldo Catania (2019). "Switching generalized autoregressive score copula models with application to systemic risk". In: *Journal of Applied Econometrics*.

Bhattacharya, Kunal et al. (2008). "The international trade network: weighted network analysis and modelling". In: *Journal of Statistical Mechanics: Theory and Experiment*.

Billio, Monica, Roberto Casarin, and Matteo Iacopini (2018). "Bayesian Tensor Regression Models". In: *Mathematical and Statistical Methods for Actuarial Sciences and Finance*. Springer.

Binder, K (1981). "Static and dynamic critical phenomena of the two-dimensional q-state Potts model". In: *Journal of Statistical Physics*.

Bisias, Dimitrios, Mark Flood, Andrew W. Lo, and Stavros Valavanis (2012). "A Survey of Systemic Risk Analytics". In: *Annual Review of Financial Economics*.

Blasques, Francisco, Siem Jan Koopman, Katarzyna Łasak, and André Lucas (2016). "In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models". In: *International Journal of Forecasting*.

Blasques, Francisco, Siem Jan Koopman, and Andre Lucas (2014). *Maximum likelihood estimation for generalized autoregressive score models*. Tech. rep. Tinbergen Institute Discussion Paper.

— (2015). "Information-theoretic optimality of observation-driven time series models for continuous responses". In: *Biometrika*.

Blasques, Francisco, André Lucas, and Andries C van Vlodrop (2020). "Finite Sample Optimality of Score-Driven Volatility Models: Some Monte Carlo Evidence". In: *Econometrics and Statistics*.

Bollerslev, Tim (1986a). "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics*.

— (1986b). "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics*.

Bornholdt, Stefan (2001). "Expectation bubbles in a spin model of markets: Intermittency from frustration across scales". In: *International Journal of Modern Physics C*.

Bouchaud, Jean-Philippe (2013). "Crises and Collective Socio-Economic Phenomena: Simple Models and Challenges". In: *Journal of Statistical Physics*.

Bradley, Andrew P (1997). "The use of the area under the ROC curve in the evaluation of machine learning algorithms". In: *Pattern Recognition*.

Bradley, Ralph Allan and Milton E Terry (1952). "Rank analysis of incomplete block designs: I. The method of paired comparisons". In: *Biometrika*.

Brunetti, Celso, Jeffrey H Harris, Shawn Mankad, and George Michailidis (2019). "Interconnectedness in the interbank market". In: *Journal of Financial Economics*.

Buccheri, Giuseppe, Giacomo Bormetti, Fulvio Corsi, and Fabrizio Lillo (2018). "Filtering and Smoothing with Score-Driven Models". In: *SSRN, Working Paper 3139666*.

Bullmore, Ed and Olaf Sporns (2009). "Complex brain networks: graph theoretical analysis of structural and functional systems". In: *Nature Reviews Neuroscience*.

Butts, Carter T (2008). "A Relational Event Framework for Social Action". In: *Sociological Methodology*.

Caccioli, Fabio, Munik Shrestha, Cristopher Moore, and J. Doyne Farmer (2014). "Stability analysis of financial contagion due to overlapping portfolios". In: *Journal of Banking & Finance*.

Caldarelli, G., A. Capocci, P. De Los Rios, and M. A. Muñoz (2002). "Scale-Free Networks from Varying Vertex Intrinsic Fitness". In: *Physical Review Letters*.

Caldarelli, Guido (2007). *Scale-free networks: complex webs in nature and technology*. Oxford University Press.

Calvori, Francesco, Drew Creal, Siem Jan Koopman, and André Lucas (2017). "Testing for parameter instability across different modeling frameworks". In: *Journal of Financial Econometrics*.

Campajola, Carlo, Domenico Di Gangi, Fabrizio Lillo, and Daniele Tantari (2020). "Modelling time-varying interactions in complex systems: the Score Driven Kinetic Ising Model". In: *arXiv:2007.15545*.

Campajola, Carlo, Fabrizio Lillo, Piero Mazzarisi, and Daniele Tantari (2021). "On the equivalence between the Kinetic Ising Model and discrete autoregressive processes". In: *Journal of Statistical Mechanics: Theory and Experiment*.

Campajola, Carlo, Fabrizio Lillo, and Daniele Tantari (2019). "Inference of the kinetic Ising model with heterogeneous missing data". In: *Physical Review E*.

— (2020). "Unveiling the relation between herding and liquidity with trader lead-lag networks". In: *Quantitative Finance*.

Campbell, LL (1970). "Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities". In: *The Annals of Mathematical Statistics*.

Casteigts, Arnaud, Paola Flocchini, Walter Quattrociocchi, and Nicola Santoro (2012). "Time-varying graphs and dynamic networks". In: *International Journal of Parallel, Emergent and Distributed Systems*.

Challet, Damien, Rémy Chicheportiche, Mehdi Lallouache, and Serge Kassibrakis (2016). "Trader lead-lag networks and order flow prediction". In: *SSRN, Working Paper 2839312*.

Chatterjee, Sourav, Persi Diaconis, Allan Sly, et al. (2011). "Random graphs with a given degree sequence". In: *The Annals of Applied Probability*.

Chen, Yi-Ting (2015). "Modeling Maximum Entropy Distributions for Financial Returns by Moment Combination and Selection". In: *Journal of Financial Econometrics*.

Chuliá, Helena, Martin Martens, and Dick van Dijk (2010). "Asymmetric effects of federal funds target rate changes on S&P100 stock returns, volatilities and correlations". In: *Journal of Banking & Finance*.

Cimini, Giulio, Tiziano Squartini, Diego Garlaschelli, and Andrea Gabrielli (2015). "Systemic risk analysis on reconstructed economic and financial networks". In: *Scientific Reports*.

Cimini, Giulio et al. (2014). "Reconstructing topological properties of complex networks using the fitness model". In: *Social Informatics*. Springer International Publishing.

Clarke, Kevin A (2005). "The phantom menace: Omitted variable bias in econometric research". In: *Conflict Management and Peace Science*.

Cocco, Joao F, Francisco J Gomes, and Nuno C Martins (2009). "Lending relationships in the interbank market". In: *Journal of Financial Intermediation*.

Cocco, Simona, Stanislas Leibler, and Rémi Monasson (2009). "Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods". In: *Proceedings of the National Academy of Sciences*.

Cocco, Simona, Rémi Monasson, Lorenzo Posani, and Gaia Tavoni (2017). "Functional networks from inverse modeling of neural population activity". In: *Current Opinion in Systems Biology*.

Cohen, Reuven and Shlomo Havlin (2010). *Complex networks: structure, robustness and function*. Cambridge university press.

Commission, US Securities & Exchange and Commodity Futures Trading Commission (2010). "Findings regarding the market events of May 6, 2010". In: *Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*.

Cont, Rama and Lakshithe Wagalath (2014). "Fire sales forensics: measuring endogenous risk". In: *Mathematical Finance*.

Coolen, ACC (2001a). "Statistical mechanics of recurrent neural networks I: Statics". In: *Handbook of Biological Physics*.

— (2001b). "Statistical mechanics of recurrent neural networks IIdynamics". In: *Handbook of Biological Physics*.

Corsi, Fulvio, Fabrizio Lillo, Davide Pirino, and Luca Trapin (2018). "Measuring the propagation of financial distress with Granger-causality tail risk networks". In: *Journal of Financial Stability*.

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

Cox, David R et al. (1981). "Statistical analysis of time series: Some recent developments". In: *Scandinavian Journal of Statistics*.

Craig, Ben and Goetz Von Peter (2014). "Interbank tiering and money center banks". In: *Journal of Financial Intermediation*.

Cranmer, Skyler J and Bruce A Desmarais (2011). "Inferential network analysis with exponential random graph models". In: *Political Analysis*.

Creal, Drew, Siem Jan Koopman, and André Lucas (2013). "Generalized autoregressive score models with applications". In: *Journal of Applied Econometrics*.

Crisanti, A and Haim Sompolinsky (1988). "Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics". In: *Physical Review A*.

Davidson, Russell, James G MacKinnon, et al. (2004). *Econometric theory and methods*. Oxford University Press New York.

Decelle, Aurélien and Pan Zhang (2015). "Inference of the sparse kinetic Ising model using the decimation method". In: *Physical Review E*.

Demirgüç-Kunt, Asli and Enrica Detragiache (1998). "The Determinants of Banking Crises in Developing and Developed Countries". In: *Staff Papers - International Monetary Fund*.

Derrida, Bernard, Elizabeth Gardner, and Anne Zippelius (1987). "An exactly solvable asymmetric neural network model". In: *Europhysics Letters*.

Desmarais, Bruce A and Skyler J Cranmer (2012). "Statistical mechanics of networks: Estimation and uncertainty". In: *Physica A: Statistical Mechanics and its Applications*.

Di Gangi, Domenico, Giacomo Bormetti, and Fabrizio Lillo (2019). "Score-Driven Exponential Random Graphs: A New Class of Time-Varying Parameter Models for Dynamical Networks". In: *arXiv:1905.10806*.

— (2021). "Score-Driven Fitness Model for Sparse Weighted Temporal Networks". In: *preparation*.

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

Diebold, Francis X and Robert S Mariano (2002). "Comparing predictive accuracy". In: *Journal of Business & Economic Statistics*.

Dorogovtsev, S.N., J.F.F. Mendes, and A.N. Samukhin (2003). "Principles of statistical mechanics of uncorrelated random networks". In: *Nuclear Physics B*.

Draief, Moez and Laurent Massoulie (2010). *Epidemics and rumours in complex networks*. Cambridge University Press Cambridge.

Duarte, F. and T. M. Eisenbach (2013). "Fire-Sale Spillovers and Systemic Risk". In: *Federal Reserve Bank of New York Staff Reports,*

Dunn, Benjamin and Yasser Roudi (2013). "Learning and inference in a nonequilibrium Ising model with hidden nodes". In: *Physical Review E*.

Duttagupta, Rupa and Paul Cashin (2011). "Anatomy of banking crises in developing and emerging market countries". In: *Journal of International Money and Finance*.

Easley, David, Marcos M Lopez De Prado, and Maureen O'Hara (2011). "The microstructure of the "flash crash": flow toxicity, liquidity crashes, and the probability of informed trading". In: *The Journal of Portfolio Management*.

Easley, David, Jon Kleinberg, et al. (2010). *Networks, crowds, and markets*. Cambridge University Press Cambridge.

Edwards, Samuel Frederick and Phil W Anderson (1975). "Theory of spin glasses". In: *Journal of Physics F: Metal Physics*.

Ehrmann, Michael and Paul Schure (2020). "The European Systemic Risk Board–Governance and Early Experience". In: *Journal of Economic Policy Reform*.

Engle, Robert (2002). "New frontiers for ARCH models". In: *Journal of Applied Econometrics*.

Engle, Robert F (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica*.

Engle, Robert F and Jeffrey R Russell (1998). "Autoregressive conditional duration: a new model for irregularly spaced transaction data". In: *Econometrica*.

Erdős, Paul and Alfréd Rényi (1959). "On random graphs I." In: *Publ. Math. Debrecen*.

Erdős, Paul and Alfred Rényi (1961). "On the strength of connectedness of a random graph". In: *Acta Math. Hungar.*

Fache Rousová, Linda, Audrius Jukonis, Elisa Letizia, Marios Gravanis, et al. (2020). "Derivatives-related liquidity risk facing investment funds". In: *Financial Stability Review*.

Fagiolo, Giorgio and Marina Mastrorillo (2013). "International migration network: Topology and modeling". In: *Physical Review E*.

Fagiolo, Giorgio, Javier Reyes, and Stefano Schiavo (2010). "The evolution of the world trade web: a weighted-network analysis". In: *Journal of Evolutionary Economics*.

Fagiolo, Giorgio and Gianluca Santoni (2016). "Revisiting the role of migrant social networks as determinants of international migration flows". In: *Applied Economics Letters*.

Fagiolo, Giorgio, Tiziano Squartini, and Diego Garlaschelli (2013). "Null models of economic networks: the case of the world trade web". In: *Journal of Economic Interaction and Coordination*.

Fan, Jianqing and Wenyang Zhang (2008). "Statistical methods with varying coefficient models". In: *Statistics and its Interface*.

Faust, Katherine and John Skvoretz (2002). "Comparing networks across space and time, size and species". In: *Sociological Methodology*.

Ferrari, Ulisse et al. (2018). "Separating intrinsic interactions from extrinsic correlations in a network of sensory neurons". In: *Physical Review E*.

Fienberg, Stephen E and Stanley S Wasserman (1981). "Categorical data analysis of single sociometric relations". In: *Sociological Methodology*.

Filimonov, Vladimir and Didier Sornette (2012). "Quantifying reflexivity in financial markets: Toward a prediction of flash crashes". In: *Physical Review E*.

— (2015). "Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data". In: *Quantitative Finance*.

Finger, Karl, Daniel Fricke, and Thomas Lux (2013). "Network analysis of the e-MID overnight money market: the informational value of different aggregation levels for intrinsic dynamic processes". In: *Computational Management Science*.

Fowler, James H (2006). "Connecting the Congress: A study of cosponsorship networks". In: *Political Analysis*.

Francq, Christian, Lajos Horvath, and Jean-Michel Zakoïan (2011). "Merits and drawbacks of variance targeting in GARCH models". In: *Journal of Financial Econometrics*.

Frank, Ove and David Strauss (1986). "Markov graphs". In: *Journal of the American Statistical Association*.

Gai, Prasanna, Andrew Haldane, and Sujit Kapadia (2011). "Complexity, concentration and contagion". In: *Journal of Monetary Economics*.

Gai, Prasanna and Sujit Kapadia (2010). "Contagion in financial networks". In: *Proceedings of the Royal Society A*.

Gandy, Axel and Luitgard AM Veraart (2016). "A Bayesian methodology for systemic risk assessment in financial networks". In: *Management Science*.

Gao, Song, Yaoli Wang, Yong Gao, and Yu Liu (2013). "Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality". In: *Environment and Planning B: Planning and Design*.

Garlaschelli, Diego, Frank den Hollander, and Andrea Roccaverde (2016). "Ensemble nonequivalence in random graphs with modular structure". In: *arXiv:1603.08759*.

Garlaschelli, Diego and Maria I. Loffredo (2008). "Maximum likelihood: Extracting unbiased information from complex networks". In: *Physical Review E*.

— (2009). "Generalized Bose-Fermi Statistics and Structural Correlations in Weighted Networks". In: *Physical Review Letters*.

Gatheral, Jim, Alexander Schied, and Alla Slynko (2012). "Transient Linear Price Impact and Fredholm Integral Equations". In: *Math. Finance*.

Génois, Mathieu et al. (2015). "Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers". In: *Network Science*.

Giannone, Domenico, Michele Lenza, Huw Pill, and Lucrezia Reichlin (2012). "The ECB and the interbank market". In: *The Economic Journal*.

Gilboa, Itzhak and Massimo Marinacci (2016). "Ambiguity and the Bayesian Paradigm". In: *Readings in Formal Epistemology*. Springer.

Giraitis, Liudas, George Kapetanios, Anne Wetherilt, and Filip Žikeš (2016). "Estimating the dynamics and persistence of financial networks, with an application to the Sterling money market". In: *Journal of Applied Econometrics*.

Goodreau, Steven M (2007). "Advances in exponential random graph (p*) models applied to a large social network". In: *Social networks*.

Graham, Bryan S (2017). "An econometric model of network formation with degree heterogeneity". In: *Econometrica*.

Green, Christopher et al. (2016). "Overnight interbank markets and the determination of the interbank rate: A selective survey". In: *International Review of Financial Analysis*.

Greene, William H (2000). "Econometric analysis 4th edition". In: *International edition, New Jersey: Prentice Hall*.

Greenwood, Robin, Augustin Landier, and David Thesmar (2015). "Vulnerable Banks". In: *Journal of Financial Economics.*

Griliches, Zvi (1977). "Estimating the returns to schooling: Some econometric problems". In: *Econometrica*.

Hafner, Christian M and Hans Manner (2012). "Dynamic stochastic copula models: Estimation, inference and applications". In: *Journal of Applied Econometrics*.

Haghani, Sogol and Mohammad Reza Keyvanpour (2017). "A systemic analysis of link prediction in social network". In: *Artificial Intelligence Review*.

Haldane, Andrew G and Robert M May (2011). "Systemic risk in banking ecosystems". In: *Nature*.

Haldane, Andrew G et al. (2009). "Rethinking the financial network". In: *Speech delivered at the Financial Student Association, Amsterdam, April*.

Hamilton, James D. (1986). "A standard error for the estimated state vector of a state-space model". In: *Journal of Econometrics*.

Handcock, Mark S (2003a). "Assessing degeneracy in statistical models of social networks". In:

— (2003b). "Statistical Models for Social Networks: Inference and Degeneracy". In: *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. National Academies Press.

Hanley, James A and Barbara J McNeil (1982). "The meaning and use of the area under a receiver operating characteristic (ROC) curve." In: *Radiology*.

Hanneke, Steve, Wenjie Fu, Eric P Xing, et al. (2010). "Discrete temporal models of social networks". In: *Electronic Journal of Statistics*.

Hansen, Lars Peter and Thomas J. Sargent (2001). "Robust Control and Model Uncertainty". In: *The American Economic Review*.

Hardiman, Stephen J, Nicolas Bercot, and Jean-Philippe Bouchaud (2013). "Critical reflexivity in financial markets: a Hawkes process analysis". In: *The European Physical Journal B*.

Hardiman, Stephen J and Jean-Philippe Bouchaud (2014). "Branching-ratio approximation for the self-exciting Hawkes process". In: *Physical Review E*.

Harrington, Scott E. (2009). "The Financial Crisis, Systemic Risk, and the Future of Insurance Regulation". In: *Journal of Risk and Insurance*.

Harvey, Andrew C. (2013b). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monographs. Cambridge University Press.

Harvey, Andrew C (2013a). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*. Cambridge University Press.

Harvey, David, Stephen Leybourne, and Paul Newbold (1997). "Testing the equality of prediction mean squared errors". In: *International Journal of Forecasting*.

Hatzopoulos, Vasilis et al. (2015). "Quantifying preferential trading in the e-MID interbank market". In: *Quantitative Finance*.

Hautsch, Nikolaus, Dieter Hess, and David Veredas (2011). "The impact of macroeconomic news on quote adjustments, noise, and informational volatility". In: *Journal of Banking & Finance*.

Hertz, John A et al. (2010). "Inferring network connectivity using kinetic Ising models". In: *BMC Neuroscience*.

Hoang, Danh-Tai, Juyong Song, Vipul Periwal, and Junghyo Jo (2019). "Network inference in stochastic systems from neurons to currencies: Improved performance at small sample size". In: *Physical Review E*.

Hoff, Peter D, Adrian E Raftery, and Mark S Handcock (2002). "Latent space approaches to social network analysis". In: *Journal of the American Statistical Association*.

Holland, Paul W and Samuel Leinhardt (1981a). "An Exponential Family of Probability Distributions for Directed Graphs". In: *Journal of the American Statistical Association*.

Holland, Paul W. and Samuel Leinhardt (1981b). "An Exponential Family of Probability Distributions for Directed Graphs". In: *Journal of the American Statistical Association*.

Holme, Petter and Jari Saramäki (2012). "Temporal networks". In: *Physics Reports*.

Hopfield, John J (1982). "Neural networks and physical systems with emergent collective computational abilities". In: *Proceedings of the national academy of sciences*.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989). "Multilayer feedforward networks are universal approximators". In: *Neural networks*.

Huang, Kerson (2008). *Statistical Mechanics*. Wiley.

Huang, Xuqing, Irena Vodenska, Shlomo Havlin, and H. Eugene Stanley (2013). "Cascading Failures in Bi-partite Graphs: Model for Systemic Risk Propagation". In: *Nature Scientific Reports*.

Huber, Peter J et al. (1967). "The behavior of maximum likelihood estimates under nonstandard conditions". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. University of California Press.

Hunter, David R, Steven M Goodreau, and Mark S Handcock (2008). "Goodness of fit of social network models". In: *Journal of the American Statistical Association*.

Hunter, David R and Mark S Handcock (2006). "Inference in curved exponential family models for networks". In: *Journal of Computational and Graphical Statistics*.

Hunter, David R et al. (2008). "ergm: A package to fit, simulate and diagnose exponential-family models for networks". In: *Journal of Statistical Software*.

Imparato, A, A Pelizzola, and M Zamparo (2007). "Ising-like model for protein mechanical unfolding". In: *Physical Review Letters*.

Iori, Giulia, Kapar Burcu, and Jose Olmo (2015). "Bank characteristics and the interbank money market: a distributional approach". In: *Studies in Nonlinear Dynamics & Econometrics*.

Iori, Giulia et al. (2008). "A network analysis of the Italian overnight money market". In: *Journal of Economic Dynamics and Control*.

Ising, Ernst (1925). "Beitrag zur theorie des ferromagnetismus". In: *Zeitschrift für Physik A Hadrons and Nuclei*.

Jackson, Matthew O (2010). *Social and economic networks*. Princeton university press.

Jacobs, Patricia A and Peter AW Lewis (1978). *Discrete Time Series Generated by Mixtures. III. Autoregressive Processes (DAR (p))*. Tech. rep. Naval Postgraduate School Monterey Calif.

Javed, Muhammad Aqib et al. (2018). "Community detection in networks: A multidisciplinary review". In: *Journal of Network and Computer Applications*.

Jaynes, Edwin T (1957). "Information theory and statistical mechanics". In: *Physical Review*.

Jochmans, Koen (2018). "Semiparametric analysis of network formation". In: *Journal of Business & Economic Statistics*.

Kaminsky, Graciela L. and Carmen M. Reinhart (1999). "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems". In: *American Economic Review*.

Karahanoğlu, Fikret Işık and Dimitri Van De Ville (2017). "Dynamics of large-scale fMRI networks: Deconstruct brain activity to build better models of brain function". In: *Current Opinion in Biomedical Engineering*.

Kim, Bomin, Kevin H Lee, Lingzhou Xue, Xiaoyue Niu, et al. (2018). "A review of dynamic network models with latent variables". In: *Statistics Surveys*.

Kingma, Diederik P and Jimmy Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv:1412.6980*.

Kirilenko, Andrei, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun (2017). "The flash crash: High-frequency trading in an electronic market". In: *The Journal of Finance*.

Kirkpatrick, Scott and David Sherrington (1978). "Infinite-ranged models of spin-glasses". In: *Physical Review B*.

Kolaczyk, Eric D. (2009a). *Statistical Analysis of Network Data*. Springer.

Kolaczyk, Eric D. (2009b). *Statistical Analysis of Network Data: Methods and Models*. 1st. Springer Publishing Company, Incorporated.

Koopman, Siem Jan, Andre Lucas, and Marcel Scharth (2016). "Predicting time-varying parameters with parameter-driven and observation-driven models". In: *Review of Economics and Statistics*.

Kouskoulas, Y., L.E. Pierce, and F.T. Ulaby (2004). "A computationally efficient multivariate maximum-entropy density estimation (MEDE) technique". In: *Geoscience and Remote Sensing, IEEE Transactions on*.

Kritzman, M., Y. Li, S. Page, and R. Rigobon (2011). "Principal Components as a Measure of Systemic Risk". In: *The Journal of Portfolio Management*.

Krivitsky, Pavel N and Mark S Handcock (2014). "A separable model for dynamic networks". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Kullback, Solomon (1997). *Information theory and statistics*. Courier Corporation.

Latapy, Matthieu, Tiphaine Viard, and Clémence Magnien (2018). "Stream graphs and link streams for the modeling of interactions over time". In: *Social Network Analysis and Mining*.

Leamer, Edward E and James Levinsohn (1995). "International trade theory: the evidence". In: *Handbook of International Economics*.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). "Deep learning". In: *Nature*.

Lee, Jihui, Gen Li, and James D Wilson (2020). "Varying-coefficient models for dynamic networks". In: *Computational Statistics & Data Analysis*.

Lee, John HH (1991). "A Lagrange multiplier test for GARCH models". In: *Economics Letters*.

Lenoci, Francesca D and Elisa Letizia (2021). "Classifying Counterparty Sector in EMIR Data". In: *Data Science for Economics and Finance*. Springer, Cham.

Lewis, Jeffrey B. et al. (2019). *Voteview: Congressional Roll-Call Votes Database*. https://voteview.com/.

Lillo, Fabrizio and Davide Pirino (2015). "The impact of systemic and illiquidity risk on financing with risky collateral". In: *Journal of Economic Dynamics and Control*.

Lillo, Fabrizio et al. (2015). "How news affects the trading behaviour of different categories of investors in a financial market". In: *Quantitative Finance*.

Lintner, John (1965). "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets". English. In: *The Review of Economics and Statistics*.

Lü, Linyuan and Tao Zhou (2011). "Link prediction in complex networks: A survey". In: *Physica A: statistical mechanics and its Applications*.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006). "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences". In: *Econometrica*.

Madhavan, Ananth (2012). "Exchange-traded funds, market structure, and the flash crash". In: *Financial Analysts Journal*.

Mancini, Loriano, Angelo Ranaldo, and Jan Wrampelmeyer (2016). "The euro interbank repo market". In: *The Review of Financial Studies*.

Marre, Olivier, Sami El Boustani, Yves Frégnac, and Alain Destexhe (2009). "Prediction of spatiotemporal patterns of neural activity from pairwise correlations". In: *Physical Review Letters*.

Martínez, Víctor, Fernando Berzal, and Juan-Carlos Cubero (2016). "A survey of link prediction in complex networks". In: *ACM computing surveys (CSUR)*.

Mastrandrea, Rossana, Tiziano Squartini, Giorgio Fagiolo, and Diego Garlaschelli (2014). "Enhanced reconstruction of weighted networks from strengths and degrees". In: *New Journal of Physics*.

Mastromatteo, Iacopo, Elia Zarinelli, and Matteo Marsili (2012). "Reconstruction of financial networks for robust estimation of systemic risk". In: *Journal of Statistical Mechanics: Theory and Experiment*.

Mazzarisi, Piero, Paolo Barucca, Fabrizio Lillo, and Daniele Tantari (2017). "A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market". In: *arXiv:1801.00185*.

— (2020). "A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market". In: *European Journal of Operational Research*.

Menkveld, Albert J and Bart Zhou Yueshen (2019). "The Flash Crash: A cautionary tale about highly fragmented markets". In: *Management Science*.

Merton, R. C. et al. (2013). "On a New Approach for Analyzing and Managing Macrofinancial Risks". In: *Financial Analysts Journal*.

Mézard, Marc and J Sakellariou (2011). "Exact mean-field inference in asymmetric kinetic Ising systems". In: *Journal of Statistical Mechanics: Theory and Experiment*.

Milgram, Stanley (1967). "The small world problem". In: *Psychology Today*.

Mistrulli, Paolo Emilio (2011). "Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns". In: *journal of banking & Finance*.

Mitchell, Melanie (2006). "Complex systems: Network thinking". In: *Artificial Intelligence*.

Moody, James and Peter J Mucha (2013). "Portrait of political party polarization". In: *Network Science*.

Mossin, Jan (1966). "Equilibrium in a Capital Asset Market". English. In: *Econometrica*.

Musmeci, Nicoló et al. (2013). "Bootstrapping Topological Properties and Systemic Risk of Complex Networks Using the Fitness Model". In: *Journal of Statistical Physics*.

Nelson, D. B. (1996). "Asymptotically optimal smoothing with arch models". In: *Econometrica*.

Nelson, Daniel B (1991). "Conditional heteroskedasticity in asset returns: A new approach". In: *Econometrica*.

— (1992). "Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model". In: *Journal of Econometrics*.

Newman, Mark (2010). *Networks: an introduction*. Oxford University Press.

— (2018). *Networks*. Oxford university press.

Newman, Mark Ed, Albert-László Ed Barabási, and Duncan J Watts (2006). *The structure and dynamics of networks.* Princeton university press.

Newman, Mark EJ (2001). "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality". In: *Physical Review E*.

— (2011). "Complex systems: A survey". In: *arXiv:1112.1440*.

Nghiem, Trang-Anh et al. (2018). "Maximum-entropy models reveal the excitatory and inhibitory correlation structures in cortical neuronal activity". In: *Physical Review E*.

Nguyen, H Chau, Riccardo Zecchina, and Johannes Berg (2017). "Inverse statistical problems: from the inverse Ising problem to data science". In: *Advances in Physics*.

Obizhaeva, Anna A (2008). "The Study of Price Impact and Effective Spread: New Approach, New Data and New Results". In: *EFA 2008 Athens Meetings Paper*.

Oet, Mikhail V., Timothy Bianco, Dieter Gramlich, and Stephen J. Ong (2013). "SAFE: An early warning system for systemic banking risk". In: *Journal of Banking & Finance*.

Onsager, Lars (1944). "Crystal statistics. I. A two-dimensional model with an order-disorder transition". In: *Physical Review*.

Park, Juyong and M. E. J. Newman (2004a). "Statistical mechanics of networks". In: *Physical Review E*.

— (2004b). "Statistical mechanics of networks". In: *Physical Review E*.

Park, Sung Y. and Anil K. Bera (2009). "Maximum entropy autoregressive conditional heteroskedasticity model". In: *Journal of Econometrics*.

Paszke, Adam et al. (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc.

Penney, RW, ACC Coolen, and D Sherrington (1993). "Coupled dynamics of fast spins and slow interactions in neural networks and spin systems". In: *Journal of Physics A: Mathematical and General*.

Potts, Renfrey Burnard (1952). "Some generalized order-disorder transformations". In: *Mathematical proceedings of the cambridge philosophical society*. Cambridge University Press.

Powell, Jerome (July 31, 2019). "Transcript of Chair Powell's Press Conference". In: *Federal Open Market Committee*.

Ramadiah, Amanah, Fabio Caccioli, and Daniel Fricke (2020). "Reconstructing and stress testing credit networks". In: *Journal of Economic Dynamics and Control*.

Rambaldi, Marcello, Vladimir Filimonov, and Fabrizio Lillo (2018). "Detection of intensity bursts using Hawkes processes: An application to high-frequency financial data". In: *Physical Review E*.

Rambaldi, Marcello, Paris Pennesi, and Fabrizio Lillo (2015). "Modeling foreign exchange market activity around macroeconomic news: Hawkes-process approach". In: *Physical Review E*.

Robins, Garry and Philippa Pattison (2001). "Random graph models for temporal processes in social networks". In: *Journal of Mathematical Sociology*.

Robins, Garry et al. (2007). "Recent developments in exponential random graph (p*) models for social networks". In: *Social networks*.

Rossetti, Giulio (2015). "Social Network Dynamics". PhD thesis. Ph. D. thesis, Computer Science Dept. University of Pisa.

Rossetti, Giulio and Rémy Cazabet (2018). "Community discovery in dynamic networks: a survey". In: *ACM Computing Surveys (CSUR)*.

Roudi, Yasser and John Hertz (2011). "Mean field theory for nonequilibrium network reconstruction". In: *Physical Review Letters*.

Roy, Sandipan, Yves Atchadé, and George Michailidis (2017). "Change point estimation in high dimensional Markov random-field models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Ryu, Hang K. (1993). "Maximum entropy estimation of density and regression functions". In: *Journal of Econometrics*.

Sachs, Angelika (2014). "Completeness, interconnectedness and distribution of interbank exposures-a parameterized analysis of the stability of financial networks". In: *Quantitative Finance*.

Saracco, Fabio, Riccardo Di Clemente, Andrea Gabrielli, and Tiziano Squartini (2015). "Randomizing bipartite networks: the case of the World Trade Web". In: *Scientific Reports*.

Sarkar, Purnamrita and Andrew W Moore (2005). "Dynamic social network analysis using latent space models". In: *Acm Sigkdd Explorations Newsletter*.

Scheffer, Marten et al. (2009). "Early-warning signals for critical transitions". In: *Nature*.

Scheffer, Marten et al. (2012). "Anticipating Critical Transitions". In: *Science*.

Schmid, Christian S and Bruce A Desmarais (2017). "Exponential random graph models with big networks: Maximum pseudolikelihood estimation and the parametric bootstrap". In: *Big Data (Big Data), 2017 IEEE International Conference on*. IEEE.

Schmitz, Oswald (2017). "Predator and prey functional traits: understanding the adaptive machinery driving predator–prey interactions". In: *F1000Research*.

Schneidman, Elad, Michael J Berry, Ronen Segev, and William Bialek (2006). "Weak pairwise correlations imply strongly correlated network states in a neural population". In: *Nature*.

Schweinberger, Michael (2011). "Instability, sensitivity, and degeneracy of discrete exponential families". In: *Journal of the American Statistical Association*.

Schweinberger, Michael, Pavel N Krivitsky, Carter T Butts, and Jonathan R Stewart (2020). "Exponential-Family Models of Random Graphs: Inference in Finite, Super and Infinite Population Scenarios". In: *Statistical Science*.

Serrano, M Ángeles, Marián Boguñá, and Alessandro Vespignani (2009). "Extracting the multiscale backbone of complex weighted networks". In: *Proceedings of the National Academy of Sciences of the United States of America*.

Sewell, Daniel K (2018). "Simultaneous and temporal autoregressive network models". In: *Network Science*.

Sewell, Daniel K and Yuguo Chen (2015). "Latent space models for dynamic networks". In: *Journal of the American Statistical Association*.

Sewell, Daniel K. and Yuguo Chen (2016). "Latent space models for dynamic networks with weighted edges". In: *Social Networks*.

Shannon, Claude E (1948). "A mathematical theory of communication". In: *Bell system technical journal*.

Sharpe, William F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk". English. In: *The Journal of Finance*.

Sheldon, George and Martin Maurer (1998). "Interbank Lending and Systemic Risk: An Empirical Analysis for Switzerland". In: *Swiss Journal of Economics and Statistics*.

Shephard, Neil (2005). *Stochastic volatility: selected readings*. Oxford University Press on Demand.

Shore, Jesse and Benjamin Lubin (2015). "Spectral goodness of fit for network models". In: *Social Networks*.

Simpson, Sean L, Satoru Hayasaka, and Paul J Laurienti (2011). "Exponential random graph modeling for complex brain networks". In: *PloS one*.

Snijders, Tom AB (1996). "Stochastic actor-oriented models for network change". In: *Journal of Mathematical Sociology*.

— (2002). "Markov chain Monte Carlo estimation of exponential random graph models". In: *Journal of Social Structure*.

Snijders, Tom AB, Philippa E Pattison, Garry L Robins, and Mark S Handcock (2006). "New specifications for exponential random graph models". In: *Sociological Methodology*.

SocioPatterns Research Collaboration (2008). *http://www.sociopatterns.org/*.

Song, Chaoming, Shlomo Havlin, and Hernan A Makse (2005). "Self-similarity of complex networks". In: *Nature*.

Soramäki, K, A Wetherilt, and P Zimmerman (2010). "The sterling unsecured loan market during 2006–08: insights from network theory". In: *Unpublished working paper. Bank of England*.

Sornette, Didier (2014). "Physics and financial economics (1776–2014): puzzles, Ising and agent-based models". In: *Reports on Progress in Physics*.

Spearman, Charles (1961). "The proof and measurement of association between two things." In:

Squartini, Tiziano, Giorgio Fagiolo, and Diego Garlaschelli (2011). "Randomizing world trade. II. A weighted network analysis". In: *Physical Review E*.

Squartini, Tiziano, Iman van Lelyveld, and Diego Garlaschelli (2013). "Early-warning signals of topological collapse in interbank networks". In: *Scientific Reports*.

Strauss, David and Michael Ikeda (1990). "Pseudolikelihood estimation for social networks". In: *Journal of the American Statistical Association*.

Tanaka, Seiji and Harold A Scheraga (1977). "Model of protein folding: incorporation of a one-dimensional short-range (Ising) model into a three-dimensional model". In: *Proceedings of the National Academy of Sciences*.

Tauchen, George E and Mark Pitts (1983). "The price variability-volume relationship on speculative markets". In: *Econometrica*.

Tavoni, Gaia et al. (2017). "Functional coupling networks inferred from prefrontal cortex activity show experience-related effective plasticity". In: *Network Neuroscience*.

Temizsoy, Asena, Giulia Iori, and Gabriel Montes-Rojas (2017). "Network centrality and funding rates in the e-MID interbank market". In: *Journal of Financial Stability*.

Tkačik, Gašper et al. (2014). "Searching for collective behavior in a large network of sensory neurons". In: *PLoS Computational Biology*.

Travers, Jeffrey and Stanley Milgram (2011). "An experimental study of the small world problem". In: *The structure and dynamics of networks*. Princeton University Press.

Tucci, Marco P (1995). "Time-varying parameters: a critical introduction". In: *Structural Change and Economic Dynamics*.

Tumminello, Michele et al. (2011). "Statistically Validated Networks in Bipartite Complex Systems". In: *PLoS ONE*.

Tyrcha, Joanna, Yasser Roudi, Matteo Marsili, and John Hertz (2013). "The effect of nonstationarity on models inferred from neural data". In: *Journal of Statistical Mechanics: Theory and Experiment*.

Upper, Christian and Andreas Worms (2004). "Estimating bilateral exposures in the German interbank market: is there a danger of contagion?" In: *European Economic Review*.

Usta, Ilhan and Yeliz Mert Kantar (2011). "On the performance of the flexible maximum entropy distributions within partially adaptive estimation". In: *Computational Statistics & Data Analysis*.

Van Duijn, Marijtje A.J., Krista J. Gile, and Mark S. Handcock (2009). "A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models". In: *Social Networks*.

Varin, Cristiano, Nancy Reid, and David Firth (2011). "An overview of composite likelihood methods". In: *Statistica Sinica*.

Wang, Peng, BaoWen Xu, YuRong Wu, and XiaoYu Zhou (2015). "Link prediction in social networks: the state-of-the-art". In: *Science China Information Sciences*.

Wasserman, Stanley and Philippa Pattison (1996). "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp". In: *Psychometrika*.

Watts, Duncan J and Steven H Strogatz (1998). "Collective dynamics of small-world networks". In: *Nature*.

Weisbuch, Gerard, Alan Kirman, and Dorothea Herreiner (2000). "Market organisation and trading relationships". In: *The Economic Journal*.

Wells, Simon (2004). *Financial interlinkages in the United Kingdom's interbank market and the risk of contagion*. Tech. rep. Sterling Markets Division, Bank of England.

Wheatley, Spencer, Alexander Wehrli, and Didier Sornette (2019). "The endo–exo problem in high frequency financial price fluctuations and rejecting criticality". In: *Quantitative Finance*.

White, Halbert (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". In: *Econometrica*.

Wilson, James D, Nathaniel T Stevens, and William H Woodall (2016). "Modeling and estimating change in temporal networks via a dynamic degree corrected stochastic block model". In: *arXiv:1605.04049*.

Wooldridge, Jeffrey M (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wu, Fa-Yueh (1982). "The potts model". In: *Reviews of Modern Physics*.

Wu, Ximing (2003). "Calculation of maximum entropy densities with application to income distribution". In: *Journal of Econometrics*.

Yan, Ting, Binyan Jiang, Stephen E Fienberg, and Chenlei Leng (2018). "Statistical inference in a directed network model with covariates". In: *Journal of the American Statistical Association*.

Yan, Ting, Chenlei Leng, Ji Zhu, et al. (2016). "Asymptotics in directed exponential random graph models with an increasing bi-degree sequence". In: *The Annals of Statistics*.

Yatchew, Adonis and Zvi Griliches (1985). "Specification error in probit models". In: *The Review of Economics and Statistics*.

Zellner, Arnold and Richard A. Highfield (1988). "Calculation of maximum entropy distributions and approximation of marginalposterior distributions". In: *Journal of Econometrics*.

Zeng, Hong-Li, Erik Aurell, Mikko Alava, and Hamed Mahmoudi (2011). "Network inference using asynchronously updated kinetic Ising model". In: *Physical Review E*.

Zermelo, Ernst (1929). "Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung". In: *Mathematische Zeitschrift*.

Zhang, Yan et al. (2008). "Community structure in Congressional cosponsorship networks". In: *Physica A: Statistical Mechanics and its Applications*.

Zhou, R. X., S. C. Liu, and W. H. Qiu (2008). "Survey of applications of entropy in decision analysis". In: *Control and Decisions*.

Zhou, Rongxi, Ru Cai, and Guanqun Tong (2013). "Applications of Entropy in Finance: A Review". In: *Entropy*.