Causality in financial markets : time reversal asymmetry and multi-scale lead-lag networks

Marcus Cordi

▶ **To cite this version:**

Marcus Cordi. Causality in financial markets : time reversal asymmetry and multi-scale lead-lag networks. Other. Université Paris Saclay (COmUE), 2019. English. NNT : 2019SACLC013 . tel-02063380

HAL Id: tel-02063380
https://theses.hal.science/tel-02063380

Submitted on 11 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

université PARIS-SACLAY

CentraleSupélec

NNT : 2019SACLC013

Thèse de doctorat

Causality in Financial Markets: Time Reversal Asymmetry and Multi-Scale Lead-Lag Networks

Thèse de doctorat de l'Université Paris-Saclay
préparée à CentraleSupélec

Ecole doctorale n°573 INTERFACES
Spécialité de doctorat: Mathématiques appliquées

Thèse présentée et soutenue à Gif-sur-Yvette, le 7 mars 2019, par

Marcus Cordi

Composition du Jury :

Rosario N MANTEGNA Università degli Studi di Palermo	Président
Fabio CACCIOLI University College London	Rapporteur
Damien CHALLET CentraleSupélec, Université Paris-Saclay	Directeur de thèse
Emmanuel BACRY École Polytechnique	Examinateur
Jyrki PIILO University of Turku	Examinateur
Sarah LEMLER CentraleSupélec, Université Paris-Saclay	Examinateur

Abstract

This thesis aims to uncover the underlying causality structure of financial markets by focusing on the inference of investor causal networks at multiple timescales in two trader-resolved datasets.

The first part of this thesis is devoted to the causal strength of Hawkes processes. These processes describe in a clearly causal way how the activity rate of e.g. an investor depends on his past activity rate; its multivariate version also makes it possible to include the interactions between the agents, at all time scales. The main result of this part is that the classical MLE estimation of the process parameters does not vary significantly if the arrow of time is reversed in the univariate and symmetric multivariate case. This means that blindly trusting univariate and symmetric multivariate Hawkes processes to infer causality from data is problematic. In addition, we find a dependency between the level of causality in the process and its endogeneity. For long time series of synthetic data, one can discriminate between the forward and backward arrows of time by performing rigorous statistical tests on the processes, but for empirical data the situation is much more ambiguous, as it is entirely possible to find a better Hawkes process fit when time runs backwards compared to forwards.

Asymmetric Hawkes processes do not suffer from very weak causality. Fitting them to the individual traders' actions found in our datasets is unfortunately not very successful for two reasons. We carefully checked that traders actions in both datasets are highly non-stationary, and that local stationarity cannot be assumed to hold as there is simply not enough data, even if each dataset contains about one million trades. This is also compounded by the fact that Hawkes processes encode the pairwise influence of traders for all timescales simultaneously.

In order to alleviate this problem, the second part of this thesis focuses on causality between specific pairs of timescales. Further filtering is achieved by reducing the effective number of investors; Statistically Validated Networks are applied to cluster investors into groups based on the statistically high synchronisation of their actions (buy, sell or neutral) in time intervals of a given timescale. This part then generalizes single-timescale lead-lag SVNs to lead-lag networks between two timescales and introduces three slightly different methods

These methods make it possible to characterize causality in a novel way. We are able to compare the time reversal asymmetry of trader activity and that of

price volatility, and conclude that the causal structure of trader activity is considerably more complex than that of the volatility for a given category of traders. Expectedly, institutional traders, whose impact on prices is much larger than that of retail clients, have a causality structure that is closer to that of volatility. This is because volatility, being a macroscopic quantity, aggregates the behaviour of all types of traders, thereby hiding the causality structure of minor players.

Acknowledgements

I would first of all like to thank my thesis director Damien Challet who has supported me immensely during these years. Your enthusiasm, technical expertise and consistently positive attitude have been an enormous help to me.

Secondly, I would also like to thank Fréderic Abergel, who as then director of the MICS laboratory welcomed me.

Thirdly, I would like to thank Stanislao Gualdi, who as a senior researcher greatly supported me, especially during my first year at CentraleSupélec.

I am grateful to all the past and current members of the MICS laboratory who have made my day-to-day life at CentraleSupélec a pleasant experience.

Lastly, I would like to thank my family, who have always supported me

Résumé

Cette thèse a pour but d'explorer la structure de causalité qui sous-tend les marchés financiers. Elle se concentre sur l'inférence multi-échelle de réseaux de causalité entre investisseurs dans deux bases de données contenant les identifiants des investisseurs.

La première partie de cette thèse est consacrée à l'étude de la causalité dans les processus de Hawkes. Ces derniers définissent la façon dont l'activité d'un investisseur (par exemple) dépend du passé; sa version multivariée inclut l'interaction entre séries temporelles, à toutes les échelles. Les résultats principaux de cette partie est que l'estimation avec le maximum de vraisemblance des paramètres du processus changent remarquablement peu lorsque la direction du temps est inversée, tant pour les processus univariés que pour les processus multivariés avec noyaux d'influence mutuelle symétriques, et que la causalité effective de ces processus dépend de leur endogénéité. Cela implique qu'on ne peut pas utiliser ce type de processus pour l'inférence de causalité sans précautions. L'utilisation de tests statistiques permet la différentiation des directions du temps pour des longues données synthétiques. Par contre, l'analyse de données empiriques est plus problématique: il est tout à fait possible de trouver des données financières pour lesquelles la vraisemblance des processus de Hawkes est plus grande si le temps s'écoule en sens inverse.

Les processus de Hawkes multivariés avec noyaux d'influence asymétriques ne sont pas affectés par une faible causalité. Il est malheureusement difficile de les calibrer aux actions individuelles des investisseurs présents dans nos bases de données, pour deux raisons. Nous avons soigneusement vérifié que l'activité des investisseurs est hautement non-stationaire et qu'on ne peut pas supposer que leur activité est localement stationaire, faute de données en nombre suffisant, bien que nos bases de données contiennent chacune plus de 1 million de transactions. Ces problèmes sont renforcés par le fait que les noyaux dans les processus de Hawkes codent l'influence mutuelle des investisseurs pour toutes les échelles de temps simultanément.

Afin de pallier ce problème, la deuxième partie de cette thèse se concentre sur la causalité entre des échelles de temps spécifiques. Un filtrage supplémentaire est obtenu en réduisant le nombre effectif d'investisseurs grâce aux Réseaux Statistiquement Validés. Ces derniers sont utilisés pour catégoriser les investisseurs, qui sont groupés selon leur degré de la synchronisation de leurs actions (achat, vente, neutre) dans des intervalles déterminés à une échelle temporelle donnée. Cette partie propose une méthode pour l'inférence de réseaux de

meneurs et suiveurs déterminés à une échelle de temps donnée dans le passé et à une autre dans le futur. Trois variations de cette méthode sont étudiées.

Cette méthode permet de caractériser la causalité d'une façon novatrice. Nous avons comparé l'asymétrie temporelle des actions des investisseurs et celle de la volatilité des prix, et conclure que la structure de causalité des investisseurs est considérablement plus complexe que celle de la volatilité. De façon attendue, les investisseurs institutionnels, dont l'impact sur l'évolution des prix est beaucoup plus grand que celui des clients privés, ont une structure causale proche de celle de la volatilité: en effet, la volatilité, étant une quantité macroscopique, est le résultat d'une aggrégation des comportements de tous les investisseurs, qui fait disparaître la structure causale des investisseurs privés.

Contents

1 Introduction **6**

1.1 Context 6

1.2 Outline 16

2 Hawkes Processes **18**

2.1 General introduction to Hawkes processes 18

2.2 Testing the causality of Hawkes processes with time reversal . . . 21

2.2.1 Introduction 21

2.2.2 Univariate processes 23

2.2.3 Multivariate processes 34

2.2.4 Application to Data 39

2.2.5 Discussion 42

2.2.6 Appendix: Univariate Hawkes processes with power-law kernels 43

2.2.7 Appendix: Log-likelihood of the univariate HP with a sum of exponentials 44

2.2.8 Appendix: Asymmetric multivariate case 45

2.3 Application of Hawkes processes to the inference of investor activity interaction networks 49

3 Lead-Lag Statistically Validated Networks **54**

3.1 Introduction 54

3.2 Method 56

3.2.1 SVNs and LL-SVNs 56

3.2.2 LL-SVNs with two timescales 58

3.3 Dataset 62

3.4 Results 63

3.4.1 $\Delta t_1 = \Delta t_2$ 63

3.4.2 $\Delta t_1 \neq \Delta t_2$ 68

3.5 Conclusions 75
3.6 Appendix: Additional figures 76

4 Conclusions and Outlook **91**

Chapter 1

Introduction

1.1 Context

The main goal of this thesis is to determine investor causality networks in financial markets and to link them to price dynamics.

Causality is a fundamental concept in Science: do the actions of agent A affect the actions of agent B, are their actions simply correlated or are the actions of agents A and B actually influenced by the actions of a completely different agent C? A particularly useful framework to encode and visualize causality is directed networks whose nodes are agents and directed edges are causal relationships. Thus this thesis contributes to the financial network literature.

Correlation is frequently used to discern statistical relationships between various observed variables, but what we are often looking for, from a practical point of view, is causal dependencies. The famous lesson "correlation does not imply causation" that most students of statistics have heard is as ever relevant in this context.

Causality is a rather interesting concept, both from a philosophical and methodological perspective. How do we define causality? How do we go about if we want to detect it? These are old questions debated in philosophy, for example by Hume (2016) and Kant (1999), and the intuitive notion we have of causality is surprisingly difficult to define in a mathematically precise way (Chattopadhyay, 2014). These questions become of course immensely more difficult in a time dependent complex network with many interacting units, such as financial networks.

As an example of the difficulties in distinguishing causality from correlation, Fig. 1.1 displays two pairs of stochastic processes, where for one pair the processes are only negatively correlated, whereas for the other pair the pro-

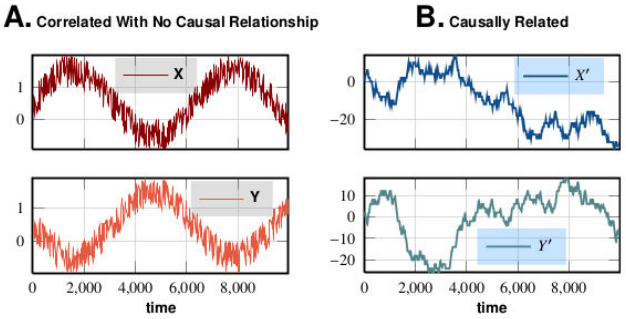


Figure 1.1: Illustration of causality and correlation. In plate A the two stochastic processes A and B are negatively correlated, but there is no discernable causal dependence. In plate B, on the other hand, the two stochastic processes X' and Y' are causally related (in the sense of Granger causality) from Y' to X' (taken from Chattopadhyay (2014)).

cesses are causally related in the sense that one of the processes carries unique information that improves future prediction of the other process, even though the two processes themselves are negatively correlated (Chattopadhyay, 2014). In this specific instance, the definition of causality used here is referred to as 'Granger causality' (or 'Wiener-Granger causality'), and is based on an idea introduced by Norbert Wiener which states that one variable, or time series, could be considered 'causal' with regards to another if the predictability of the second variable is improved if information from the first variable is incorporated (Wiener, 1956). Wiener did not, however, have a practical implementation of his idea (Bressler and Seth, 2011). In the context of linear autoregressive models of stochastic processes, such an implementation was introduced by Granger (1963, 1969). This definition of causality might not however necessarily mean true causality, since the Granger causality test only fulfils the Humean definition of causality (Maziarz, 2015). A clear limitation of the practical implementation to test if Granger causality is present between time series is that since it is based on linear correlation it does not take into account non-linear causal relationships, and is therefore best suited for 'non-wild' (Gaussian) variables.

There are of course other ways of mathematically defining causality between time series, for example transfer entropy, where one studies the amount of directed transfer of information, by using Shannon's entropy, between two stochastic processes (Schreiber, 2000). Transfer entropy reduces to Granger causality for vector auto-regressive processes (Barnett et al., 2009), and can thus be useful when the assumptions for Granger causality are not valid (e.g. non-linearity).

Another method, which has been used extensively in this thesis work, is the method of Statistically Validated Networks (SVNs), which is an unsupervised

method used to validate links in complex networks with heterogeneous elements, against a null hypothesis that takes into account spurious relationships found simply because of the heterogeneity of the system. Using this method we thus have an implicit (or underlying) causality structure, but we only observe the outcome of it.

As previously mentioned, correlation should not be confused with causation. However, lagged correlation (between two variables) can be useful to provide us with information about the predictability of the system, even though it does not reveal completely and definitively the causal structure. Another important aspect to take into account is that one needs to go beyond lagged correlation defined at equal time intervals, i.e. the time slices used to determine the shifted correlations are of equal length. This is related to the length of the typical time scales, which will be discussed later.

We will now present some examples from the literature where lagged correlation has been used to study causal relationships in financial markets.

A common phenomena, especially in economics, is the so-called lead-lag effect, where one (leading) variable is cross-correlated with another (lagging) variable at later times. Even though the question of causality still remains, it might still be, from a practical point of view, valuable to establish these lead-lag relationships. Many different forms of the lead-lag effect in economic data have been studied extensively, but the results, as we will see, depend on the tools used and their continual improvement. The results also change because the market dynamics are non-static over time, it has been claimed, for example, that the market has become more efficient[1], and this would partly be due to technological progress, e.g. Moore's law (Present, 2000)[2].

An example of the lead-lag effect phenomena is the cross-correlation of stock returns, studied by Tóth and Kertész (2006). More specifically, the cross-correlations $C(\tau)$ of stock returns with a time shift τ between the pairs' price return time series. In Fig. 1.2 some example plots are displayed, for high-frequency data obtained from two different years, and in Fig. 1.3 the evolution over eleven years of the average time shift which has the maximum estimated correlation, is displayed. The authors report that lead-lag relationships between returns of stock have significantly decreased over the years, which they interpret as a sign that

[1] The efficient market hypothesis (EMH) (Tobin, 1969; Malkiel and Fama, 1970) states that all available information is already reflected in the current price, which means it is not possible to predict future prices based on past records.

[2] See Hardiman et al. (2013) for an example of how market dynamics have changed because of Moore's law and the emergence of high-frequency trading.

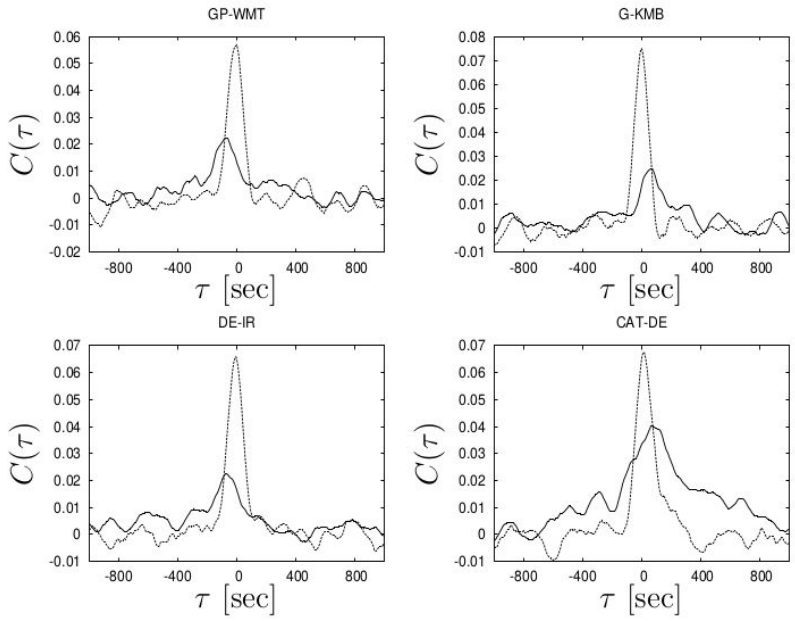


Figure 1.2: Example plots of the changes in the cross-correlation for the years 1993 (solid) and 2003 (dashed) (taken from Tóth and Kertész (2006))

the market is becoming increasingly efficient.

Trading is however asynchronous, something that the estimator used by Tóth and Kertész (2006) does not completely take into account. In order to fairly compare two assets, especially at high-frequency trading, another estimator is needed, and a more suitable estimator is the Hayashi-Yoshida estimator (Hayashi et al., 2005), which deals with the issue of asynchronicity and makes use of all available tick-by-tick data. Huth and Abergel (2014) exploit this estimator and use tick-by-tick data to show that some assets follow the path of others with a small time lag (no more than five minutes), by observing strongly asymmetric cross-correlation functions. In order to determine which one is the leading and which one is the lagging asset, the Lead-Lag Ratio LLR between two assets, defined as

$$LLR = \frac{\int_0^T C(\tau)^2 \mathrm{d}\tau}{\int_0^T C(-\tau)^2 \mathrm{d}\tau} \tag{1.1}$$

is used, and where $C(\tau)$ here corresponds to the Hayashi-Yoshida cross correlation estimator. In Fig. 1.4 the clearly asymmetric correlation structure between CAC40 future (FCE) and Total (TOTF.PA) is displayed, and in Fig. 1.5 the lead-lag network of stocks in the CAC40 is obtained by considering the LLR between stocks.

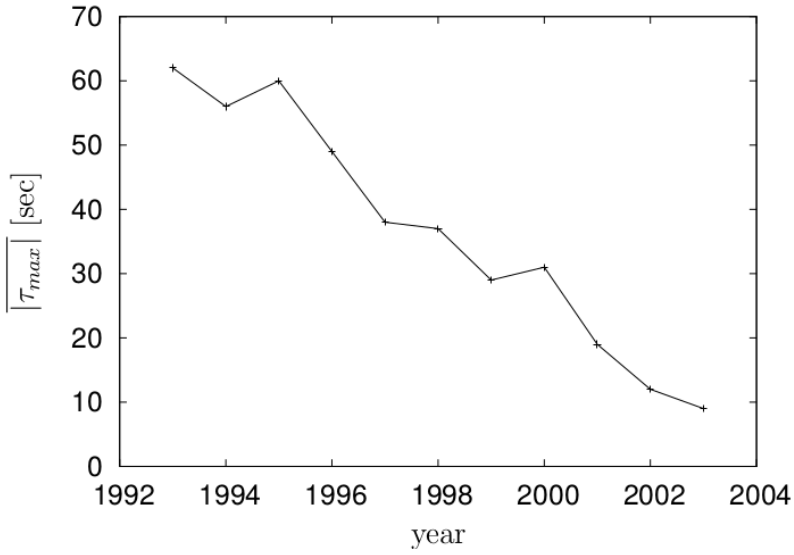


Figure 1.3: Average time shift τ which yields the maximum estimated correlation (taken from Tóth and Kertész (2006))

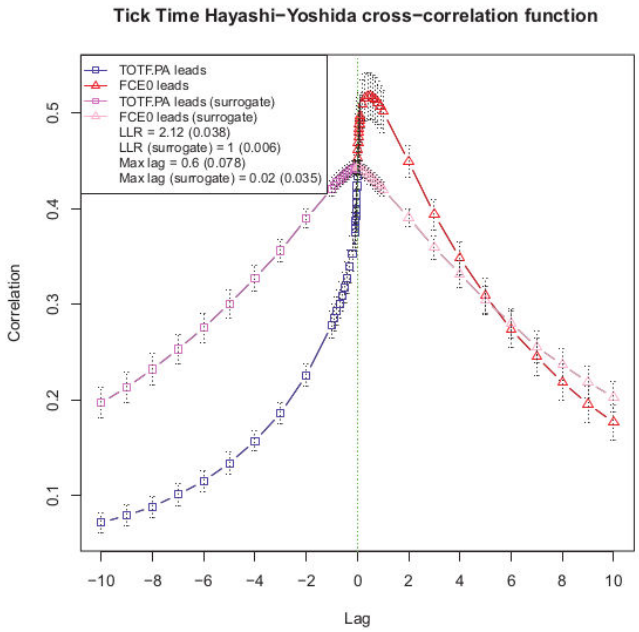


Figure 1.4: Example of a tick time cross-correlation function with the Hayashi-Yoshida cross-correlation estimator for asynchronous data (taken from Huth and Abergel (2014))

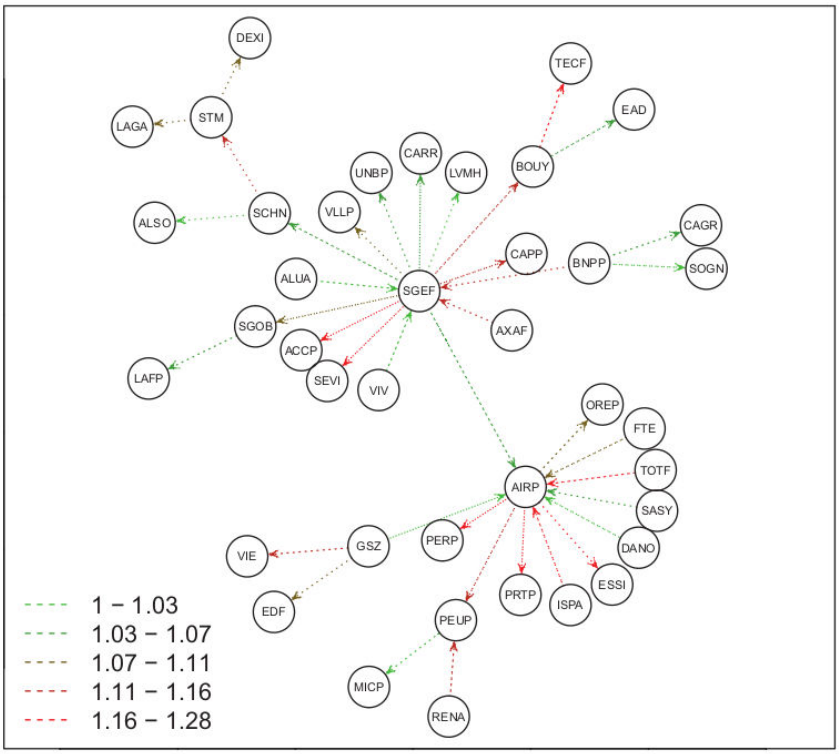
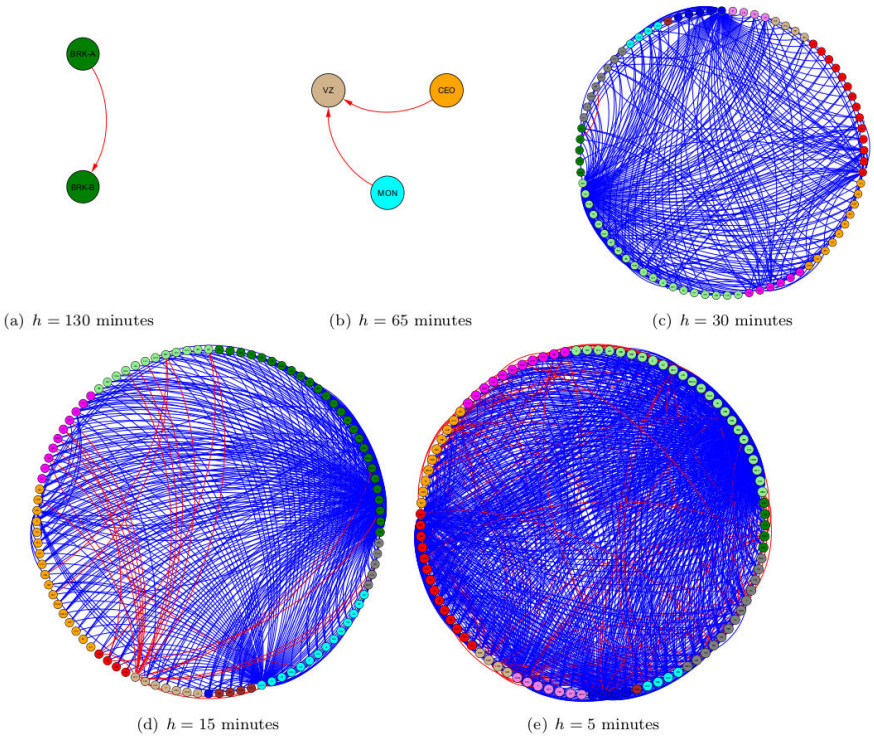


Figure 1.5: Lead-lag network on the CAC40 (taken from Huth and Abergel (2014))

Curme et al. (2015) introduce a statistical validation of the lead-lag relationships with a multiple hypothesis testing correction (Bonferroni and False Discovery Rate (FDR)). Specifically, the links are filtered according to a threshold of statistical significance by using a shuffling technique introduced by Efron and Tibshirani (1994). The lead-lag links are established at different time horizons h, and the resulting networks at different h are presented in Fig. 1.6. It is worth noting that links may represent a positive or negative correlation. In Fig. 1.7 the number of validated links are plotted for two different dataset, one for the years 2002-2003 and one for the years 2011-2012. For a given time horizon h more links are usually validated in the 2002-2003 dataset than in the 2011-2012 dataset. The authors suggest that this might be an indication that there has been an increase in market efficiency over this decade.

A final example of a lead-lag relationship in financial data is the so called 'leverage effect', which refers to the observed tendency of an asset's volatility to be negatively correlated with its returns (Ait-Sahalia et al., 2013)[3]. If $S_i(t)$

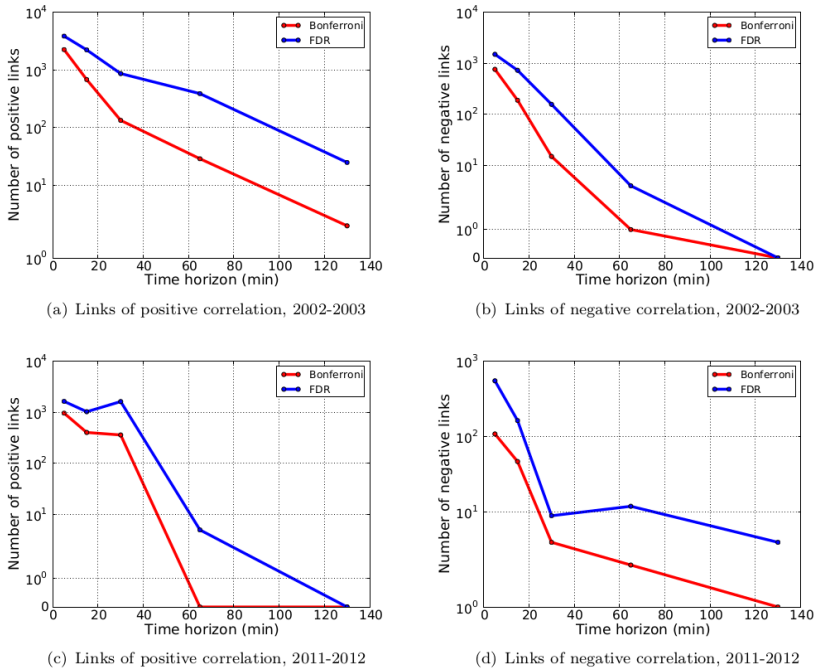
[3]The term 'leverage' refers to a possible economic interpretation of this phenomenon, namely that as the price of an asset declines, companies become more leveraged since the relative value of their debt rises relative to their equity. Therefore, it would be natural to assume that their stock becomes riskier (more volatile) (Black, 1976; Christie, 1982).



(a) $h = 130$ minutes (b) $h = 65$ minutes (c) $h = 30$ minutes

(d) $h = 15$ minutes (e) $h = 5$ minutes

Figure 1.6: Networks of statistically validated (Bonferroni) lagged correlations for various time horizons h, with data from large market-capitalization companies on the NYSE in 2011-2012. Blue links represent positive correlations and red links negative correlations (taken from (Curme et al., 2015)).



(a) Links of positive correlation, 2002-2003

(b) Links of negative correlation, 2002-2003

(c) Links of positive correlation, 2011-2012

(d) Links of negative correlation, 2011-2012

Figure 1.7: Number of positive and negative statistically validated link for both Bonferroni and FDR lagged correlations networks,for both datasets of 2002-2003 and 2011-2012 (taken from (Curme et al., 2015)).

denotes the price of stock i at time t, and $\delta S_i(t) = S_i(t+1) - S_i(t)$ the daily price change, then $\delta x_i(t) = \frac{\delta S_i(t)}{S_i(t)}$ denotes the relative price change. The leverage correlation function, calculated over a time horizon T, may then be defined as

$$\mathcal{L}_i(\tau) = \frac{1}{Z} \left\langle \left[\delta x_i(t+\tau)\right]^2 \delta x_i(t) \right\rangle , \tag{1.2}$$

where $Z = \left\langle \delta x_i(t)^2 \right\rangle^2$ is a normalization. In other words, $\mathcal{L}_i(\tau)$ measures the correlation between the price change at time t and the square volatility at time $t+\tau$.

In Fig. 1.8 the (mostly) negative correlation between return and volatility for some US stocks is displayed. For the sake of our general discussion, it is interesting to note that the causality of the effect has been debated (Bekaert and Wu, 2000); is the volatility increase caused by the price drop, or do prices tend to fall when volatility increases?

The notion of causality is intimately tied to the concept of the arrow of time, i.e. the one-way direction or asymmetry of time. A cause always precedes its effect, which means that the causal event occurs before the event it affects. If we take a sequence of events, and then view this sequence from the end, i.e. with the arrow of time reversed, are we able to measure a significant difference com-

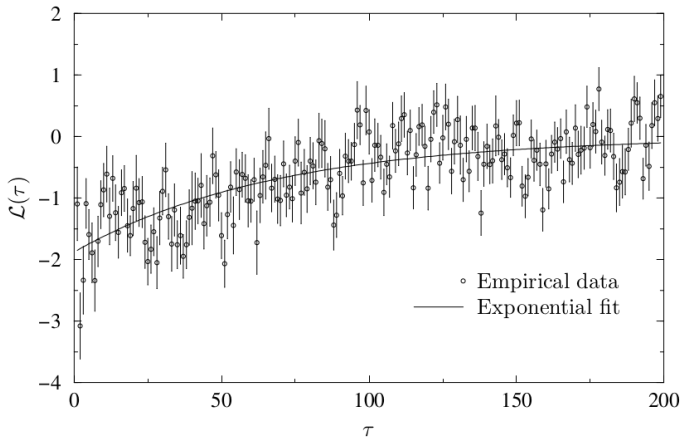


Figure 1.8: Empirical correlation between return and volatility, averaged over 437 US stocks with an exponential fit (taken from (Bouchaud et al., 2001)).

pared with the actual series of events? If we do observe a noticeable difference, then we can be quite sure that there is some kind of underlying causality in the time series (but we do not know necessarily know why or how). In physics, many basic laws display what is referred to as Time Reversal Invariance (TRI), e.g. the laws of Newtonian mechanics, Maxwell's equations of electromagnetism and the Dirac equation for quantum mechanics. Formally speaking, this means that the transformation $t \rightarrow -t$ produces an exact symmetry of the system in question. It is, however, apparent that the real world offers many examples where the direction of the arrow of time matters significantly; the process of an ice-cube melting into water is very different from a puddle of water turning into ice (Taleb, 2007)! A solution to this problem was eventually found in thermodynamics, with the introduction of entropy - the entropy of an isolated system never decreases and moves towards thermodynamic equilibrium, i.e. the state with maximum entropy.

Time series involving financial data, much like thermodynamic systems, do display Time Reversal Asymmetry (TRA), i.e. the direction of the arrow of time does indeed matter. If one only stays at the level of prices or returns it might seem difficult to notice a difference between the directions, due to the (apparently) highly random and complex nature of many financial times series. The objective is then to develop a method to spot the difference, or, using the terminology employed by Lynch et al. (2003), to find the 'mugshots' which immediately tell us, simply with the naked eye, if TRI or TRA is present. Many models do not incorporate this empirical phenomenon (e.g. Brownian motion, GARCH, Heston stochastic volatility model etc.), and the challenge is then to

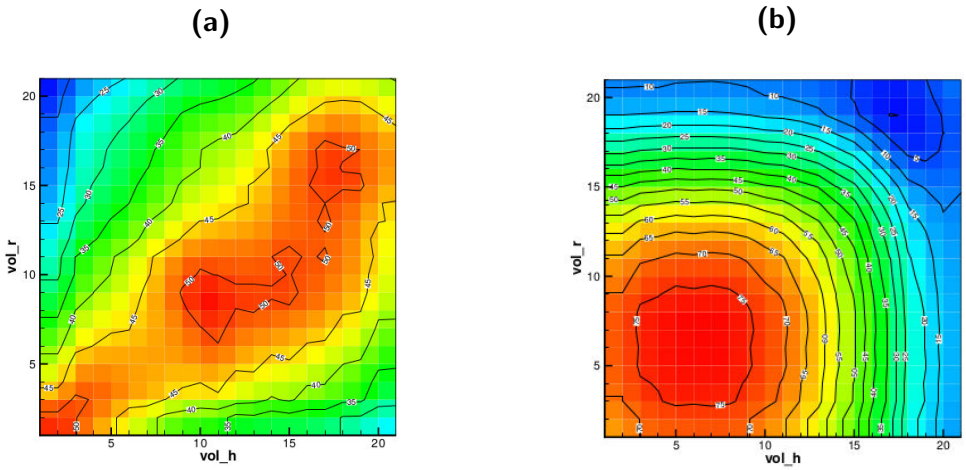


Figure 1.9: Correlation between past historical volatilities (horizontal axis) and future realized volatilities (vertical axis) for (a) the USD/CHF foreign exchange and (b) a theoretical volatility cascade with Ornstein-Uhlenbeck partial log volatilities (taken from Borland et al. (2005)). Note that the asymmetry in (a) with regards to the diagonal indicates TRA, whereas the symmetry found in (b) indicates TRI.

develop models which do. As an example of these mugshots, in Fig. 1.9 two of these mugshots are displayed. These mugshots show how past volatilities affect future volatilities on different timescales. In the mugshot of the empirical data, one clearly sees a strong asymmetry (thus the presence of TRA), and in the mugshot of the theoretical process an almost perfect symmetry (thus the presence of TRI). Recently, models have been proposed which do account for these kinds of asymmetries, see for example Borland and Bouchaud (2005), Zumbach et al. (2014), Blanc et al. (2017) and Euch et al. (2018).

The fact that financial time series often display TRA leads us to another important aspect of financial networks, namely, timescales. It should come as no surprise that participants in financial markets have vastly different time horizons. Previous work has mainly considered this fact at a meta level, for example by studying the price dynamics. Parallels have been drawn between turbulence and financial markets (Ghashghaie et al., 1996), because of the highly complex interplay over a wide range of length and time scales witnessed in turbulence (it has been shown however that there are certain limits in the analogies, see for example Arneodo et al. (1996); Lynch et al. (2003)). This thesis work provides a new approach, where the interactions between timescales are defined at a very detailed level (individual traders), and this unveils a new structure of the causality. In order to take into account the various time scales that are present, we consider quantities, e.g. correlation, determined between time slices which are not (necessarily) of equal length.

The Hawkes process (HP hereafter) provides a first promising approach to the modelling of the (causal) interactions of agents in networks. An HP is, at its core, a modified Poisson process with self-excitation. In other words, given that an event has occurred, it is more likely (temporarily) that another event will occur. This characteristic behaviour is provided by a kernel which may assume many different kinds of shapes (exponential, power-law, combination of the two etc.). This should, in theory, be suitable to model the behaviour of an agent active in the financial market. The HP may also be multivariate, with cross-excitation between the different subprocesses, i.e. in our setting between different agents. The fact that these processes are self-excited encodes a causality into them, but as we will see, this is not always completely obvious empirically.

The second approach is based on the method of Statistically Validated Networks (SVNs), which assumes that investors can be clustered into a finite (and relatively small) number of groups and that investment fluxes at time $t+1$ depend on the activity or inactivity of these groups at time t. The fundamental premise here is the idea to group investors according to the similitude of their activity patterns: for each period (5 minutes, 1 hour, one day), one assigns a tag to each investor (e.g. net buy, net sell, neutral, no activity). One then computes the probability that a given pair of traders act in a synchronized way and establish a link if the probability is high enough (with multiple-hypothesis testing correction). Then, community detection algorithms are applied to the resulting network. Existing work has shown that lead-lag links may be established between these groups when they are defined at one timescale. Our contribution here has been to provide a general method to establish lead-lag links between groups defined at different timescales. The results provide us with new insights into the structure of markets at a very detailed level.

1.2 Outline

This thesis is divided as follows:

In Chapter 2 the HP is formally introduced and applied to some data. Certain limitations of the HP related to its TRI and its capacity to model causality are discussed.

In Chapter 3 the SVN method is extended to incorporate interactions between different time scales, and the new method developed is applied to some data. Thanks to this method, new results related to the TRA of financial markets are obtained.

In Chapter 4 the results obtained in previous parts are summarised, and an outlook for future research is provided.

Chapter 2

Hawkes Processes

2.1 General introduction to Hawkes processes

A fundamental question when investigating investor interaction networks is: what determines the activity of agents in financial markets? There are numerous potential factors. The first, and perhaps the most obvious one, is the price itself. Another important factor is the arrival of news. Furthermore, the trader may also follow certain strategies which dictate his activity, and finally the activity of other traders may have an influence (explicitly or implicitly) on his activity. The factors mentioned above may be considered as either exogenous or endogenous, i.e. they are either events (actions) which are generated outside the 'system' (loosely speaking) or they are internal in the sense that they are triggered by the system itself (i.e. the agents themselves). The first two factors may be considered as exogenous, whereas the the two last factors could be considered as endogenous.

As mentioned previously, Hawkes processes (HPs) provide us with a promising theoretical framework that should, at least in theory, be useful to study these types of investor interaction networks.

The HP may be formulated in several different ways. The easiest way is to start with the one-dimensional case, where we let $\{t_i\}_{i\geq 1}$ be a univariate simple point process, t_i be the time of event i, and $N_t = \sum_{i\geq 1} \theta(t - t_i)$ be the associated counting process, where $\theta(t)$ is the Heaviside function. $(N_t)_t$ is then a univariate HP if it has the conditional intensity

$$\lambda(t) = \lambda_0(t) + \int_{-\infty}^{t} K(t-s)\mathrm{d}N_s = \lambda_0(t) + \sum_{t_i<t} K(t-t_i), \tag{2.1}$$

where $K(t)$ is the kernel which models the self-excitation, or the 'memory', of

the process.

Using this formulation of the HP, $\lambda_0(t)$, the baseline intensity, models the arrival of exogenous (or external) events [1], (akin to a Poisson process) and $\sum_{t_i<t} K(t-t_i)$ models the arrival of endogenous events, i.e. events that are generated in a self-excited fashion, or, one might say, events generated by themselves in a cluster-like way. Certain strategies of the agents may thus (assumedly) be modelled by this kind of self-excited behaviour. Due to the branching structure of HPs one may also disentangle these different kinds of events, i.e. separate exogenous events from endogenous ones (Zhuang et al., 2002; Marsan and Lengline, 2008).

A popular choice of kernel is the exponential one, consisting of a sum of a variable number of exponentials, i.e

$$K(t) = \sum_{j=1}^{P} \alpha_j \mathrm{e}^{-\beta_j t}. \tag{2.2}$$

This kind of kernel offers much flexibility in the sense that it can mimic the properties of other kernels with other kinds of decay, for example power-law decay, by using several exponentials, and thus introducing more degrees of freedom. As we will see later, the flexibility offered by this choice of kernel may be somewhat problematic for the (presumed) inherent causality of the process that is modelled. This is especially apparent when non-simulated real-world data is modelled.

Another limitation with kernels of this type is that it is assumed that the maximal impact $(\sum_{j=1}^{P} \alpha_j)$ of an event of the conditional intensity is immediate, when in fact it might be more likely that there is a certain delay (latency) of the maximal impact. In Achab et al. (2018), which we will refer to later, the authors address this issue by developing a useful non-parametric method that estimates the integrated kernel instead.

A scenario one might imagine where HPs could be useful to model the behaviour of agents is when the agents employ the strategy of moving averages. One basic strategy is to use fixed length moving average (FLMA) filters (Gunasekarage and Power, 2001). Two moving averages, one short-run and one long-run, are calculated, and when they cross a buy or sell signal is generated. Specifically, a buy (sell) signal is generated when the short moving average cuts the long moving average from below (above). It would thus seem reasonable

[1] The exogenous baseline intensity may be non-stationary.

to assume that the actions of an agent using this kind of strategy will follow a certain time scale, dictated by his choice of parameters, and this may then be captured by the HP.

It is fairly straightforward to extend the HP to the multidimensional case. If we let $\{(t_i^m)\}_{m=1,\dots,M}$ be an M-dimensional (M components) point process, and $\mathbf{N}_t = (N_1^t, \dots, N_t^M)$ the associated counting process, then the conditional intensity λ^m, $m = 1, \dots, M$ is given by

$$\lambda^m(t) = \lambda_0^m(t) + \sum_{n=1}^{M} \int_0^t K^{mn}(t-s)\mathrm{d}N_s^n = \lambda_0^m(t) + \sum_{n=1}^{M} \sum_{t_i^n < t} K^{mn}(t - t_i^n), \quad (2.3)$$

where, if we assume, typically, that the kernel is a sum of exponentials and that the shape of the kernel is identical for all self and cross-excitation terms, i.e. $K^{mn}(t) = \sum_{j=1}^{P} \alpha_j^{mn} \mathrm{e}^{-\beta_j^{mn} t}$. In vector form, we may write

$$\boldsymbol{\lambda}(t) = \boldsymbol{\lambda_0} + \int_0^t \boldsymbol{K}(t-s)\mathrm{d}\boldsymbol{N_s}. \quad (2.4)$$

The influence of different agents on each other, in this formulation, can thus be modelled as cross-excitation between different components of the process.

A crucial element of the HP, which distinguishes it from the Poisson process, is the causality it models, even in the one-dimensional case. The direction of the arrow of time is thus important in this case, and depending on the kernel we use, the 'memory' of the process may assume different timescales, and thus imply different levels of causality.

Furthermore, there is also a causality structure between components, i.e. which component triggers the activity of another component, and this might not be completely evident when a high-dimensional system is modelled (see Bacry et al. (2016) for a fast and efficient estimation method of the parameters of a high-dimensional HP exploiting a mean-field approximation). Xu et al. (2016) develop a method for learning the Granger causality of an HP by considering the relationship between the impact function of the HP and its causality graph.

Achab et al. (2018) use a more direct and powerful approach, where the matrix of the integrated kernels are estimated, without any parametric modelling and estimation of the kernels themselves, by introducing a moment matching method that fits the second-order and third-order cumulants.

During the course of our work, we found that there were certain properties

that had not apparently been explored, namely the (variable) inherent causality of some of the more basic formulations of the HP. This became apparent when we noticed that the parameter estimations of the process did not differ significantly when the arrow of time was reversed for the HP, which thus provided the impetus to the article 'Testing the causality of Hawkes processes with time reversal'.

2.2 Testing the causality of Hawkes processes with time reversal

Note: This section is published in 'Journal of Statistical Mechanics'.

2.2.1 Introduction

Hawkes processes (HPs hereafter) extend Poisson processes by allowing a modulation of the current event rate as a function of the past events. Such processes are causal and thus provide a simple way to introduce time-reversal asymmetry and self-excitation in the modelling of discrete events. This makes HPs invaluable in the modelling of physical, economic and social phenomena in which the occurrence of one event increases for some time the probability of another event. Examples may be found in seismology, where an earthquake typically is followed by aftershocks (Ogata, 1988; Gardner and Knopoff, 1974; Zhuang et al., 2002; Marsan and Lengline, 2008), criminology, where a fight between rival gangs may trigger various criminal retaliations (Mohler et al., 2011), neurology, where the spiking activity of individual neurons may depend on the neuron's own spiking history (Truccolo et al., 2005; Pillow et al., 2008; London et al., 2010), and credit risk, where the default of one company in a portfolio may lead to the default of other companies (Dassios and Zhao, 2017). Since HPs are causal, one may infer the extent to which a phenomenon is not time-reversible by fitting such processes to some data. However, the strength of the causality of HPs, i.e. the extent to which they allow to discriminate both arrows of time, while being of fundamental importance, is as of yet unexplored. This paper attempts to fill this gap, both for pure HPs and fits of HPs to real data.

HPs are causal by construction. Let $\{t_i\}_{i\geq 1}$ be a univariate simple point process, t_i being the time of event i. Let $N_t = \sum_{i\geq 1} \theta(t - t_i)$, where θ is the Heaviside function, be the associated counting process. $(N_t)_t$ is called a univariate HP if it

admits the conditional intensity

$$\lambda(t) = \lambda_0(t) + \int_{-\infty}^{t} K(t-s)\mathrm{d}N_s = \lambda_0(t) + \sum_{t_i<t} K(t-t_i), \tag{2.5}$$

where $\lambda_0(t)$ is the baseline intensity (hereafter we will assume constant baseline intensity, i.e. $\lambda_0(t) = \lambda_0$), $K(t)$ is the kernel of the process and t_i the time of event i: λ is defined in a causal way from past events, hence the direction of time is well-defined. It would thus seem foolish to fit an HP to the reverted vector of events, i.e. to the events taken in the backward direction. Accordingly, the belief that a time series of events with an inverted arrow of time cannot possibly be mistaken for an HP is widely established (see for example Kirchner (2017)).

We show here that the extent to which true HPs are distinctively causal depends on the method used to assess the fitted model and, when fitting them to data, on the nature of the data. Indeed, a parametric kernel estimation of univariate and symmetric multivariate HPs on synthetic data leads on *average* to almost the same values for both time arrows. Why this may be the case is best gathered from a classic plot that superposes the activity rate $\lambda(t)$ with the event times (Fig. 2.1). The twist is to plot the activity rate from the same sets of events, and to let time run backwards: the activity rate and the clustering of events are visually plausible for both directions of the time arrow.

More often than not in the literature, the goodness of fit of Hawkes processes is not quantitatively assessed, but only qualitatively with Q-Q plots (which often *look* good), probably because HPs are assumed to be useful extensions of Poisson processes that are either totally adequate or cannot possibly describe precisely the data, which amounts to making unverified assumptions about the goodness of fits in either case. However, recent results show that parametric fits of HPs to high-frequency financial data do pass goodness of fit tests provided that a multi-timescale kernel is used and the non-stationary baseline intensity is properly accounted for (Lallouache and Challet, 2016; Omi et al., 2017). Starting from this positive results, we investigate here to what extent goodness of fit tests are able to discriminate between a forward and a backward arrow of time for synthetic data (i.e. in an ideal setting), the latter being very often detected as not HPs. A related issue is found when one infers kernel with time-reversal symmetric quantities, which by definition yield exactly the same kernel for both arrows of time. For example, the non-parametric kernel inference of Bacry et al. (2012) is based on the covariance of event rates, which is symmetric with respect to time

reversal. We show here that such kernels only rarely pass tests of goodness of fit. However, we point out that this method provides a useful approximation of the true kernel shape precisely when causality is weak (i.e. in the limit of small endogeneity), which may then help choosing a parameteric kernel family.

Fitting HPs to real data is more troublesome. For example, data collection may further degrade causality if the time resolution is too coarse. However, the main problem, by far, is that one does not know the shape of the kernel. We show that the more flexible the kernel, the harder it becomes for tests of goodness of fit to discriminate between the forward and backward arrows of time, sometimes yielding statistically significant fits for both time directions of the same set of events. In financial data, fits usually (and reassuringly) favour the forward arrow of time. This in itself complements previous works that quantify the asymmetry of price volatility with respect to time reversal of financial markets (Lynch et al., 2003; Zumbach, 2009; Blanc et al., 2017). However, there are cases when the backward arrow of time yields better fits than the forward one. This is at odds with the reality of financial markets, which shows that a significant fit of a weakly causal HP does not necessarily correspond to physical causality. By extension, inferring from a fit that a system is causal because of the success of a fit of a weakly causal HP should rest on a comparison with a fit of the reverse arrow of time.

2.2.2 Univariate processes

We performed extensive numerical simulations by generating HPs with a single exponential kernel

$$K(t) = \alpha \mathrm{e}^{-\beta t} \tag{2.6}$$

and constant baseline intensity for a variety of parameters with the Ogata thinning method (Ogata, 1981); results for a power-law kernel are reported in Appendix 2.2.6 and are similar to those obtained with a single exponential.

Let us define

$$n = \int_0^\infty K(s)\mathrm{d}s. \tag{2.7}$$

The exponential kernel (2.6) defines a stationary HP if $n = \frac{\alpha}{\beta} < 1$. With this condition, and recalling $E[\mathrm{d}N_s] = E[\lambda(s)\mathrm{d}s]$, the expected intensity satisfies

$$E\left[\lambda(t)\right] = \lambda_0 + \int_{-\infty}^{t} K(t-s)E\left[\lambda(s)\right]\mathrm{d}s = \lambda_0 + E\left[\lambda(t)\right]n,$$

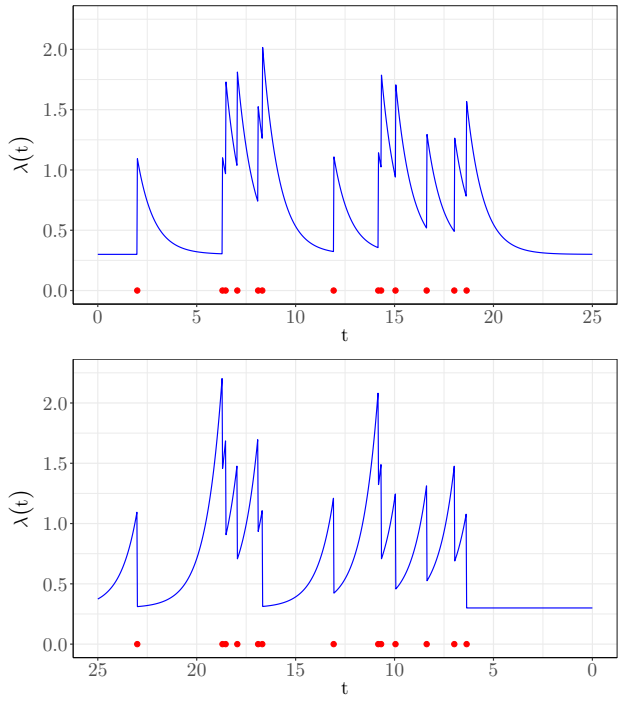


Figure 2.1: Intensity as a function of time of an HP with an exponential kernel $K(t) = \alpha e^{-\beta t}$ and $\lambda_0 = 0.3$, $\alpha = 0.8$ and $\beta = 1.2$ for the true (top) and the time-reversed sequence of events (bottom). The red points indicate events.

hence

$$E[N_t - N_0] = E\left[\int_0^t \lambda(s)\mathrm{d}s\right] = \frac{\lambda_0 t}{1-n}. \tag{2.8}$$

Detailed mathematical treatment of point processes and their intensities can be found in e.g. Brémaud (1981); Daley and Vere-Jones (2003). In our tests data points will be grouped according to the variable n, often called endogeneity of the process (or reflexivity as it might be referred to in the field of finance (Soros, 2003)). The endogeneity quantifies the level of the relative self-excitement of the process (Filimonov and Sornette, 2012).

We have adjusted the time horizon T so that all the simulations have the same expected number of events in order to allow a proper comparison between all the results obtained with different values of n.

In order to avoid calibration issues, we first of all remove ("burn") the non-stationary part of all simulations. The time of stationarity t_0 is defined as the first time the instantaneous intensity is greater or equal to the average (expected) intensity, i.e.

$$t_0 = \inf\{t \in \{t_i\}_{i=1,\dots,n} : \lambda(t) \geq \mu\}, \tag{2.9}$$

where $\mu = E[\lambda(t)]$. The process is then shifted: $t_i' = t_i - t_0, t_i > t_0$ and $T' =$

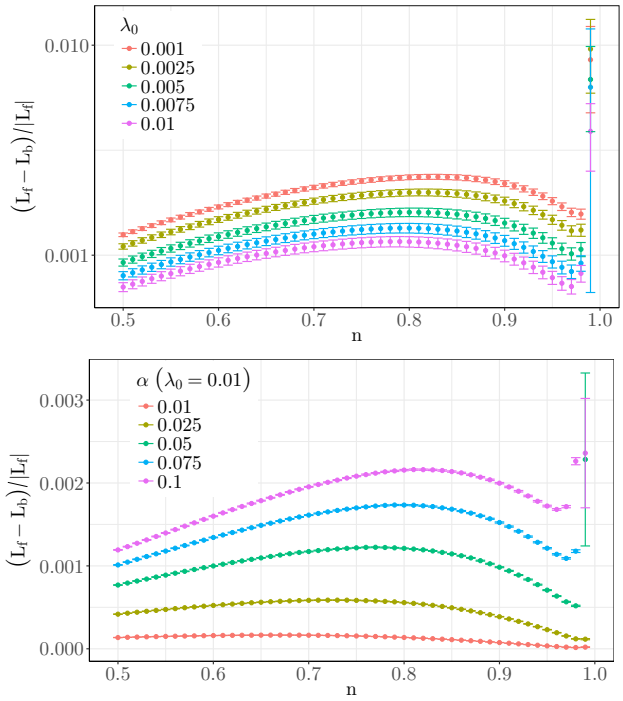


Figure 2.2: Relative difference of the log-likelihood between forward and backward time arrows for an HP with an exponential kernel. All possible permutations of $\lambda_0 = \{0.001, 0.0025, 0.0050, 0.0075, 0.0100\}$ and $\alpha = \{0.010, 0.025, 0.050, 0.075, 0.100\}$, with β chosen according to the desired endogeneity n, are considered. The data points are grouped according to their endogeneity and averaged over 100 runs for each parameter permutation. The expected number of events is set to 10^6.

$T - t_0$. This requires us to modify the usual likelihood estimation, as explained below. We shall henceforth drop the prime symbols for the sake of readability.

The vector of event times obtained from the simulations (or data) correspond by definition to the forward arrow of time and will be denoted henceforth by $\{t_i^{(f)}\}_{i=1,\dots,n}$. The events in the backward arrow of time simply correspond to taking the last event of the forward time arrow as the first one in the backward arrow of time, the second last event as the second one and so on; mathematically, $t_i^{(b)} = T - t_{n+1-i}^{(f)}$.

We compare the adequacy of HPs to both forward and backward event time series with three methods: the likelihood function calculated with the true parameters, Maximum Likelihood Estimation (MLE hereafter) and goodness of fit.

Log-likelihood

The idea here is to compare the true log-likelihood, i.e. computed with the true kernel, of simulations of HPs for the real (forward) and reversed (backward) event time vectors. The log-likelihood of a univariate point process N_t with

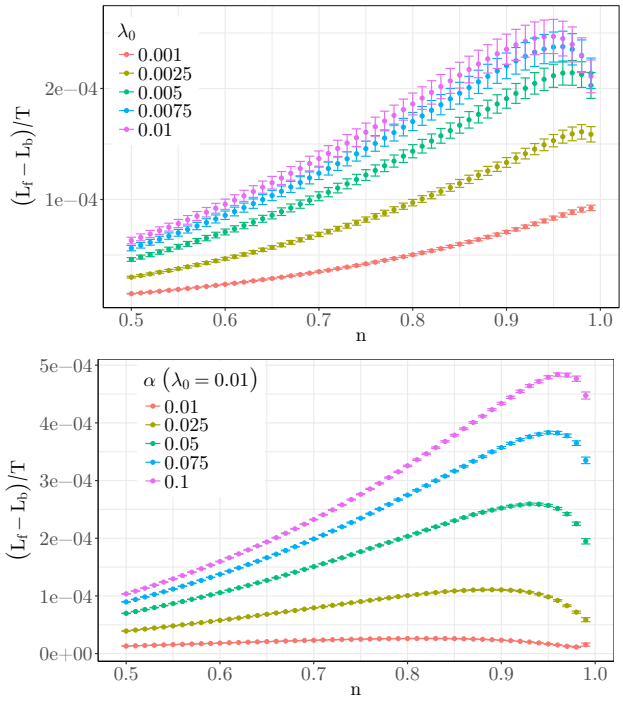


Figure 2.3: Difference of the log-likelihood scaled by T between forward and backward time arrows for an HP with an exponential kernel for $\lambda_0 = \{0.001, 0.0025, 0.0050, 0.0075, 0.0100\}$ and $\alpha = \{0.010, 0.025, 0.050, 0.075, 0.100\}$, while β is adjusted to match the desired n. The data points are grouped according to their endogeneity and averaged over 100 runs for each parameter permutation. The expected number of events is set to 10^6.

intensity $\lambda(t)$ is written as

$$\ln \mathcal{L}\left((N_t)_{t\in[0,T]}\right) = -\int_0^T \lambda(s)\mathrm{d}s + \int_0^T \ln \lambda(s)\mathrm{d}N_s. \tag{2.10}$$

In the case of an HP with an exponential kernel and a constant baseline intensity, the log-likelihood is

$$\begin{aligned}\ln \mathcal{L}\left(\{t_i\}_{i=1,\dots,n}\right) = -\lambda_0 T - \sum_{i=1}^{n} \frac{\alpha}{\beta}\left(1 - \mathrm{e}^{-\beta(T-t_i)}\right) \\ + \sum_{i=1}^{n} \ln\left[\lambda_0 + \sum_{k=1}^{i-1} \alpha \mathrm{e}^{-\beta(t_i - t_k)}\right].\end{aligned} \tag{2.11}$$

This expression, however, takes into account the initial non-stationary part of the process. A fair comparison between the forward and backward processes requires the removal of the non-stationary part of the process, which leads to small modifications of the above mathematical expression.

The general idea behind the modification is that if the simulation has already reached a stationary state, then the (constant) baseline intensity λ_0 should be replaced by a time-dependent baseline intensity $\lambda_0'(t)$, which is given by n-stationary

$$\lambda_0'(t) = \lambda_0 + \left(\frac{\lambda_0}{1-n} - \lambda_0\right)\frac{K(t)}{K(0)}. \tag{2.12}$$

A similar procedure is developed in Roueff et al. (2016). In the case of the exponential kernel we obtain

$$\begin{aligned}\ln \mathcal{L}\left(\{t_i\}_{i=1,\dots,n}\right) = -\lambda_0 T - \left(\frac{\lambda_0}{1-\frac{\alpha}{\beta}} - \lambda_0\right)\frac{1-\mathrm{e}^{-\beta T}}{\beta} \\ - \sum_{i=1}^{n} \frac{\alpha}{\beta}\left(1 - \mathrm{e}^{-\beta(T-t_i)}\right) \\ + \sum_{i=1}^{n} \ln\left[\lambda_0 + \left(\frac{\lambda_0}{1-\frac{\alpha}{\beta}} - \lambda_0\right)\mathrm{e}^{-\beta t_i} + \sum_{k=1}^{i-1} \alpha \mathrm{e}^{-\beta(t_i - t_k)}\right].\end{aligned} \tag{2.13}$$

In order to assess the performance of this correction we content ourselves with comparing the average difference between the MLE estimates and the true values (see Table 2.1), and see that the modified log-likelihood does indeed generally perform slightly better than the standard log-likelihood on truncated HPs.

Fig. 2.2 displays the average relative difference of the log-likelihood calcu-

Table 2.1: Average difference between the true parameter values and the estimations obtained via MLE for the forward (top) and backward process (bottom)) with the standard log-likelihood function (SLL) and the modified log-likelihood function (MLL) for a truncated HP (the same parameter choice as used in Figs. 2.2, 2.3 and 2.4, except the variable number of expected events).

Forward	$\lambda_0^{(f)}$	$\alpha^{(f)}$	$\beta^{(f)}$
SLL. $E[N_T] = 10^4$	9.901%	2.496%	2.751%
MLL $E[N_T] = 10^4$	5.497%	2.294%	2.192%
SLL $E[N_T] = 10^5$	1.205%	0.692%	0.687%
MLL $E[N_T] = 10^5$	1.056%	0.661%	0.656%
SLL $E[N_T] = 10^6$	0.332%	0.220%	0.217%
MLL $E[N_T] = 10^6$	0.328%	0.213%	0.210%

Backward	$\lambda_0^{(b)}$	$\alpha^{(b)}$	$\beta^{(b)}$
SLL $E[N_T] = 10^4$	10.476%	2.674%	2.886%
MLL $E[N_T] = 10^4$	6.100%	2.450%	2.461%
SLL $E[N_T] = 10^5$	1.637%	0.849%	1.020%
MLL $E[N_T] = 10^5$	1.496%	0.817%	0.988%
SLL $E[N_T] = 10^6$	1.088%	0.552%	0.797%
MLL $E[N_T] = 10^6$	1.083%	0.544%	0.790%

lated with the true parameters for both time arrows. It turns out that it is surprisingly small, typically 0.2% on average for a very large number of events, except for near-critical ($n \simeq 1$) processes. Here we see that, as one would expect, the likelihood of the forward event time series is consistently larger than that of the backward event time series.

On average, a lower baseline intensity λ_0 implies a larger difference in the log-likelihood, as one might expect since the Poissonian properties of the process are less prominent. Similarly, a larger α also implies a larger difference because each event carries with it a larger impact on the intensity. The difference of the forward and backward log-likelihood scaled by T has a similar behaviour (see Fig. 2.3).

We have checked the fraction of the simulations for which the true likelihood of the backward process is larger than that of the forward process. Expectedly, since we compute the likelihood with the true kernel, we found $8 \cdot 10^{-6}$, which is to say none. One should however keep in mind that when dealing with empirical data, one faces three additional problems that may change this rosy outcome, as indeed the above situation is an ideal case. First, one does not know the true kernel shape nor its parameters. Second, the number of events in the above simulations is much larger than those of the typical dataset. Third, the question of how to deal with a non-constant baseline intensity is fundamental, but still under active investigation; the issue here is to properly discriminate between exogenous and endogenous events, i.e. to attribute time variations of the intensity to the kernel or to the baseline intensity. Another possibility might be that the kernel itself could be non-stationary, as developed for example in Kobayashi and Lambiotte (2016).

Parameter estimation

The small difference found in the log-likelihood suggests that the estimation of the parameters based on maximum likelihood leads to fairly similar parameter values. We thus perform MLE on synthetic data; we impose that $\alpha < \beta$ in order to fulfil the requirement for a stationary process, both for the original and the time-reversed sequence of events. The few non-convergent estimations were excluded from the analysis. Since we choose as initial points the true parameter values, the optimisation is typically not required to be bound constrained and an algorithm by Nelder and Mead (1965) is used. When working with real-world data as in Section 2.2.4, however, there is a need for a bound constrained optimisation and the L-BFGS-B algorithm (Byrd et al., 1995) is used.

Unsurprisingly, Fig. 2.4 reveals that the estimated parameters only weakly depend on the direction of the time arrow of the event time series. One notes that the baseline intensity is somewhat overestimated for the time-reversed process. One interpretation is that since causality is lost, the fitting process must attribute more events to the Poisson process.

Similarly, the estimates of α, in conjunction with the estimates of β, for the backward process are overestimated. This also suggests that for the backward process too much importance is given to the short term effect or impact of the previous events, and that the memory extends less into the history of the process. It is worth noting that since the estimations of α and β are similarly overestimated, the resulting estimates of the endogeneity n is relatively close to the true value. Finally, closer to criticality there is an apparent tendency of the estimates for both arrows of time to converge.

It is also worth mentioning here that if we compare the forward and backward likelihood calculated with the MLE parameters for medium-size data sets (around 10^4 events) we see that in 1.3% of the cases that the backward likelihood actually is larger, and for even smaller data sets (around 500 events) it is 16%. In practice, available data sets are typically quite small, and therefore the log-likelihood is not a guaranteed way to distinguish between the two arrows of time.

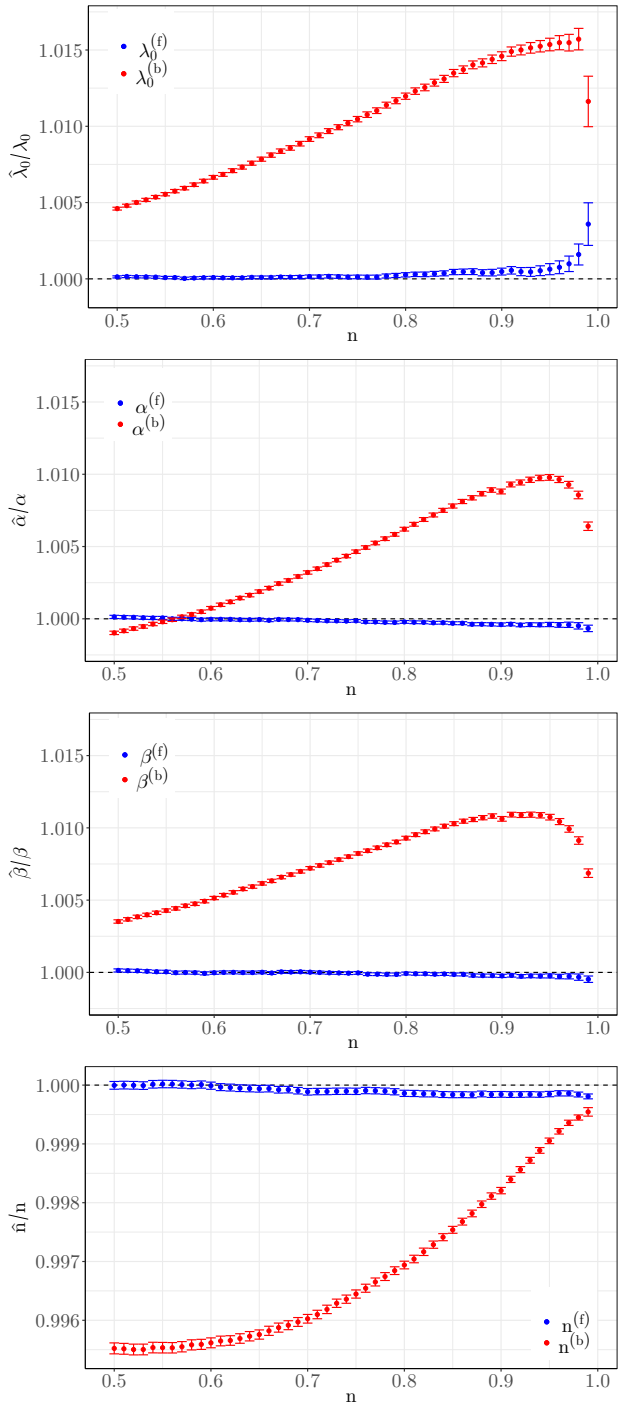


Figure 2.4: Relative difference in the estimation of λ_0, α, β and n in the MLE of the exponential HP for the forward (blue) and the backward process (red). All possible permutations of $\lambda_0 = \{0.001, 0.0025, 0.0050, 0.0075, 0.0100\}$ and $\alpha = \{0.010, 0.025, 0.050, 0.075, 0.100\}$, with β chosen according to the desired endogeneity n, are considered. The data points are grouped according to their endogeneity and averaged over 100 runs for each parameter permutation. The expected number of events is set to 10^6.

Goodness of fit test

For a given kernel K, baseline intensity λ_0 and a time series $\{t_i\}_{i=1,\dots,n}$ one defines the compensators

$$\Lambda(t_{i-1}, t_i) = \int_{t_{i-1}}^{t_i} \lambda(s)\mathrm{d}s = \int_{t_{i-1}}^{t_i} (\lambda_0 + \sum_{t_k<s} K(s - t_k))\mathrm{d}s \tag{2.14}$$

which are exponentially distributed with an average rate of 1 if the data comes from an HP (Papangelou, 1972). Thus we choose here the Kolmogorov-Smirnov test (KS test hereafter) to test the equality between the distribution of the compensators and the exponential distribution. The same test was used to find statistically valid fits of HP to high frequency data both in the foreign exchange market (Lallouache and Challet, 2016) and in the equity market (Omi et al., 2017).

Let us start with parametric estimation. We first test if the estimated kernel corresponds to the true one, i.e. the kernel obtained with the *a priori* known true parameter values (μ^*, α^* and β^*). Fig. 2.5 displays the histogram of the p-values corresponding to this hypothesis. As expected for the forward case (upper plot), a uniform distribution is obtained since the null hypothesis holds. In the backward case, most fits are rejected. In a real-life situation, however, one does not know the true kernel. In this case, as shown by Fig. 2.6 where the parameters obtained via MLE are used ($\widehat{\mu}_{\mathrm{ML}}$, $\widehat{\alpha}_{\mathrm{ML}}$ and $\widehat{\beta}_{\mathrm{ML}}$), the test accepts more samples as being HPs processes, for both arrows of time. This is due to the additional freedom one has to find slightly over-fitting parameters.

Thus, we see that the KS-test performs satisfactorily in the sense that it is clearly able to distinguish between the forward and backward process both for the MLE parameters (where in a sense the MLE "overfits" the parameters to the underlying data) and the true parameters. This emphasizes the need to assess the goodness of fits when fitting HPs to data.

The fact that the KS test is able to discriminate between the two arrows of time has a clear implication for the non-parametric kernel estimation method introduced by Bacry et al. (2012); since it is based on the auto-covariance of the event rate, which is time-symmetric by definition, this methods yields the same kernel for both directions of time. As a consequence, in view of the power of the KS test in that respect, it is understandable that this method does not yield kernels that may be deemed statistically significant, as shown by Fig. 2.7, where the parameters used ($\widehat{\mu}_{\mathrm{NP}}$, $\widehat{\alpha}_{\mathrm{NP}}$ and $\widehat{\beta}_{\mathrm{NP}}$) are estimated from the non-parametrically obtained kernels by linear interpolation. More specifically, by

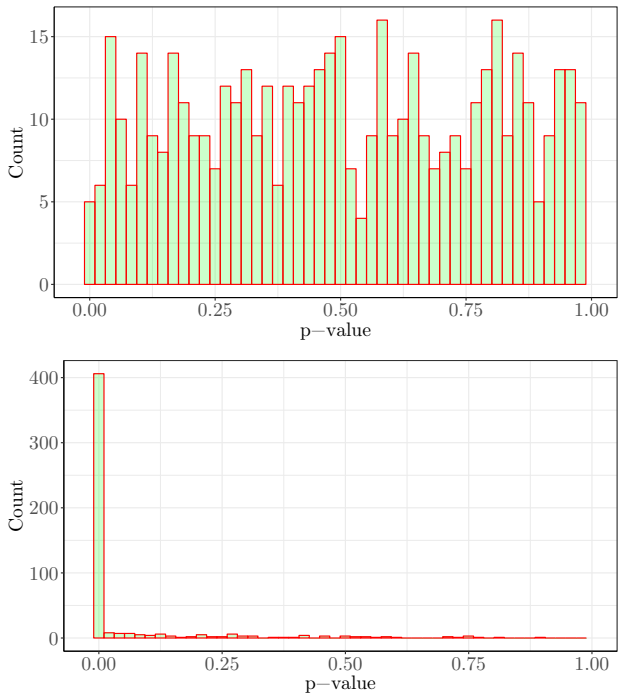


Figure 2.5: Histogram of the p-values from the KS-test obtained for the forward (upper) and backward (lower) exponential HP with the true parameter values. The parameters used for the simulations are fixed to $\lambda_0 = 0.001$ and $\alpha = 0.01$, with β chosen to the desired endogeneity $n = \{0.50, 0.75, 0.90, 0.95, 0.99\}$. The data is collected over 100 runs for each parameter permutation and the expected number of events is set to 10^6.

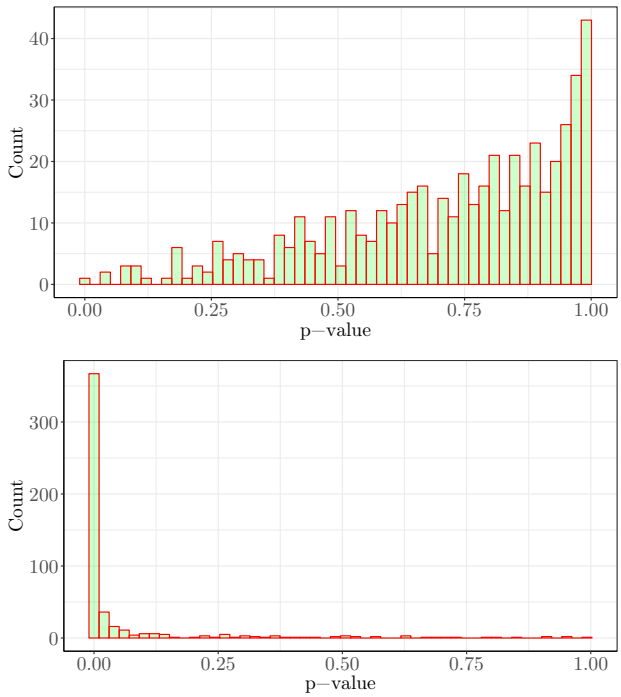


Figure 2.6: Histogram of the p-values from the KS-test obtained for the forward (upper) and backward (lower) exponential HP with the MLE parameter values. The parameters are fixed to $\lambda_0 = 0.001$ and $\alpha = 0.01$, with β chosen to the desired endogeneity $n = \{0.50, 0.75, 0.90, 0.95, 0.99\}$. 100 runs for each parameter combination; expected number of events set to 10^6.

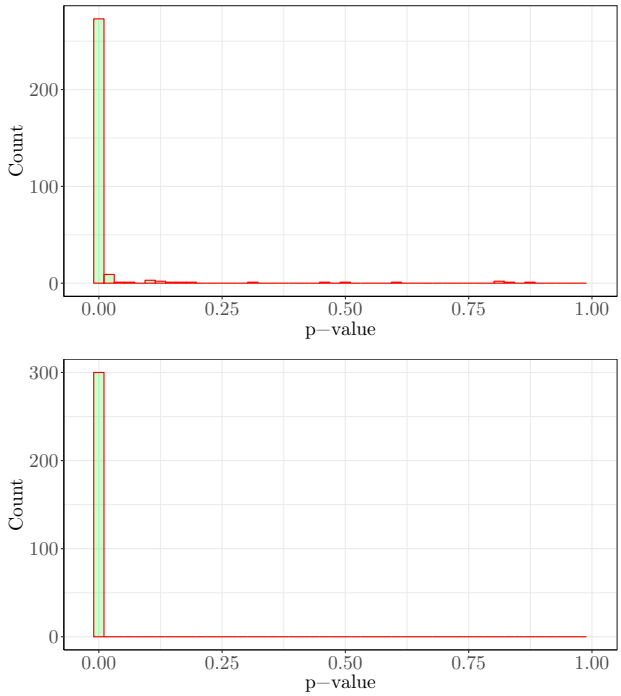


Figure 2.7: Histogram of the p-values from the KS-test obtained for the forward (upper) and backward (lower) exponential HP with parameters extracted from the non-parametrically estimated kernels. The parameters are fixed to $\lambda_0 = 0.001$ and $\alpha = 0.01$, with β chosen to the desired endogeneity $n = \{0.50, 0.75, 0.90, 0.95, 0.99\}$. 100 runs for each parameter combination; expected number of events set to 10^6.

taking the logarithm of the non-negative kernel estimate, estimates of $\widehat{\alpha}_{\mathrm{NP}}$ and $\widehat{\beta}_{\mathrm{NP}}$ are obtained by linear regression, and by Eq. 2.8 we may obtain an estimate of $\widehat{\mu}_{\mathrm{NP}}$. It is worth noting here that this method, which is quite crude, produces a considerable amount of invalid results for the higher endogeneities (hence the smaller sample in Fig. 2.7).

We stress nevertheless that the non-parametric method is an invaluable tool to assess the global shape of HPs in a preliminary exploration, and to choose a suitable parametric family which itself may pass goodness of fit tests.

2.2.3 Multivariate processes

The above findings generalize to multivariate HPs, in which several univariate HPs may also mutually excite each other. More precisely, an M-dimensional (M components) HP is defined as

$$\boldsymbol{\lambda}(t) = \boldsymbol{\lambda_0} + \int_0^t \boldsymbol{K}(t-s)\mathrm{d}\boldsymbol{N_s} \tag{2.15}$$

where the (exponential) kernel is given by

$$K^{mn}(t) = \left(\alpha^{mn} \mathrm{e}^{-\beta^{mn}(t-s)}\right)_{m,n=1,\dots,M}. \tag{2.16}$$

The intensity may thus be written as (with a constant baseline intensity)

$$\lambda^m(t) = \lambda_0^m + \sum_{n=1}^{M} \sum_{t_i^n < t} \alpha^{mn} \mathrm{e}^{-\beta^{mn}(t-t_i^n)}. \tag{2.17}$$

The expected number of events is

$$E\left[\boldsymbol{N}(t)\right] = \boldsymbol{\mu} t, \tag{2.18}$$

where

$$\boldsymbol{\mu} = \left(\boldsymbol{I} - \int_0^\infty \boldsymbol{K}(u)\mathrm{d}u\right)^{-1} \boldsymbol{\lambda_0}. \tag{2.19}$$

Here we define $N_t = \sum_{m=1}^{M} N_t^m$.

For the multidimensional HP, denoting $\{t_i\}_{i=1,\dots,N}$ the ordered pool of all events $\left\{\{t_i^m\}_{m=1,\dots,M}\right\}$, the log-likelihood can be computed as the sum of the likelihood of each coordinate, namely

$$\ln \mathcal{L}\left(\{t_i\}_{i=1,\dots,N}\right) = \sum_{m=1}^{M} \ln \mathcal{L}^m\left(\{t_i\}\right), \tag{2.20}$$

where

$$\ln \mathcal{L}^m\left(\{t_i\}\right) = -\int_0^T \lambda^m(s)\mathrm{d}s + \int_0^T \ln \lambda^m(s)\mathrm{d}N_s^m. \tag{2.21}$$

Equation (2.21) may be written as

$$\begin{aligned} \ln \mathcal{L}^m\left(\{t_i\}\right) = -\lambda_0^m T - \sum_{n=1}^{M} \sum_{t_i^n} \frac{\alpha^{mn}}{\beta^{mn}} \left(1 - \mathrm{e}^{-\beta^{mn}(T-t_i^n)}\right) \\ + \sum_{t_i^m} \ln \left[\lambda_0^m + \sum_{n=1}^{M} \sum_{t_k^n < t_i^m} \alpha^{mn} \mathrm{e}^{-\beta^{mn}(t_i - t_k^n)}\right]. \end{aligned} \tag{2.22}$$

If we, as in the one-dimensional case, remove the non-stationary part of the

process, we obtain

$$\begin{aligned}\ln \mathcal{L}^m\left(\{t_i\}\right) = -\lambda_0^m T + \frac{\mu^m - \lambda_0^m}{\sum_{n=1}^{M} \alpha^{mn}} \sum_{n=1}^{M} \frac{\alpha^{mn}}{\beta^{mn}} \left(\mathrm{e}^{-\beta^{mn} T} - 1\right) \\ - \sum_{n=1}^{M} \sum_{t_i^n} \frac{\alpha^{mn}}{\beta^{mn}} \left(1 - \mathrm{e}^{-\beta^{mn}(T - t_i)}\right) \\ + \sum_{t_i^m} \ln \left[\lambda_0^m + (\mu^m - \lambda_0^m) \frac{\sum_{n=1}^{M} \alpha^{mn} \mathrm{e}^{-\beta^{mn} t_i}}{\sum_{n=1}^{M} \alpha^{mn}} \right. \\ \left. + \sum_{n=1}^{M} \sum_{t_k^n < t_i} \alpha^{mn} \mathrm{e}^{-\beta^{mn}(t_i - t_k^n)} \right],\end{aligned} \tag{2.23}$$

Analogously to the univariate case, we can find an appropriate limit to the non-stationary period by considering

$$t_b = \inf \left\{ t \in \{t_i\}_{i=1,\ldots,N} : \sum_{m=1}^{M} \lambda^m(t) \geq \sum_{m=1}^{M} \mu^m \right\}. \tag{2.24}$$

The process is then shifted $t_i^{m'} = t_i^m - t_b, t_i^m > t_b$ and $T' = T - t_b$.

A sufficient condition for stationarity is

$$\rho(\Gamma) = \max_{a \in S(\Gamma)} |a| < 1, \tag{2.25}$$

where $S(\Gamma)$ denotes the set of all eigenvalues of Γ and

$$\Gamma = \int_0^\infty \boldsymbol{K}(u) \mathrm{d}u = \left(\frac{\alpha^{mn}}{\beta^{mn}} \right)_{m,n=1\ldots,M} \tag{2.26}$$

Here we focus on symmetric multivariate HPs, where the mutual excitation matrix can be written as

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 & \alpha_m \\ \alpha_m & \alpha_0 \end{pmatrix}.$$

For the sake of simplicity, we fix the baseline intensities and timescales to the same values for both components of the process, i.e. $\boldsymbol{\lambda}_0 = (\lambda_0, \lambda_0)$ and $\boldsymbol{\beta} = (\beta, \beta)$.

In the symmetric case $\rho(\Gamma) = \frac{\alpha_0 + \alpha_m}{\beta}$. For the presentation of the results in

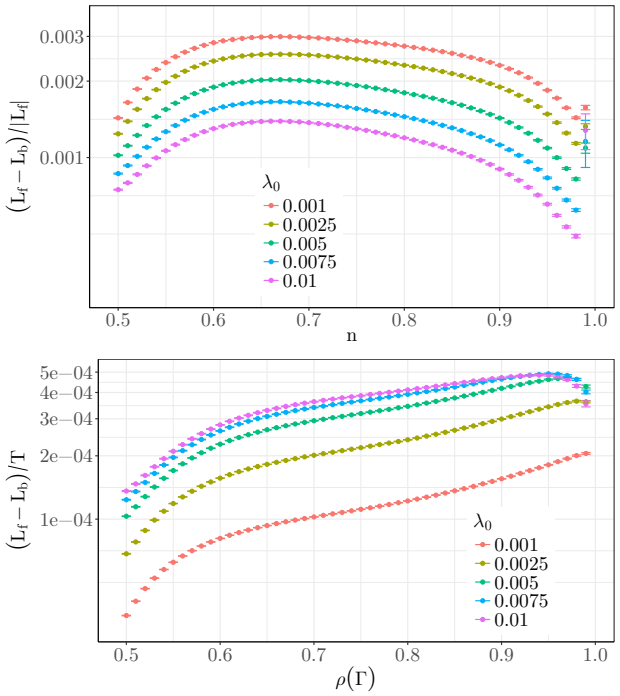


Figure 2.8: Relative difference of the log-likelihood between forward and backward time arrows (top) and difference of the log-likelihood between forward and backward time arrows with regards to T (bottom) for a multidimensional HP with a symmetric excitation kernel. All possible permutations of $\lambda_0 = \{0.0010, 0.0025, 0.005, 0.0075, 0.0100\}$, $\alpha_0 = \{0.049\}$, with α_m chosen according to the desired maximum eigenvalue $\rho(\Gamma)$, and $\beta = 0.1$ are considered. The data points are grouped according to maximum eigenvalue and averaged over 100 runs for each parameter permutation. The expected total number of events is set to 10^6.

the multivariate setting, the largest eigenvalue $\rho(\Gamma)$ was chosen as the control parameter instead of the endogeneity n since it is directly linked to the expected total number of events in the process (i.e. summed over all components).

Figs. 2.8, 2.9, and 2.10 show that the results for symmetric multidimensional HP are in many ways analogous to those of the univariate HP, i.e. the log-likelihood plots display a similar behaviour and the parametric estimations do not deviate significantly from each other in the forward and backward case. Our remarks regarding the non-parametric method of Bacry et al. (2012), which is only valid for symmetric HPs, still hold.

The above findings are however not true for asymmetric multivariate HPs, in which changing the direction of time leads to clearly different log-likelihoods and parameters (see Appendix 2.2.8), hence significantly increases the effective causality of such processes.

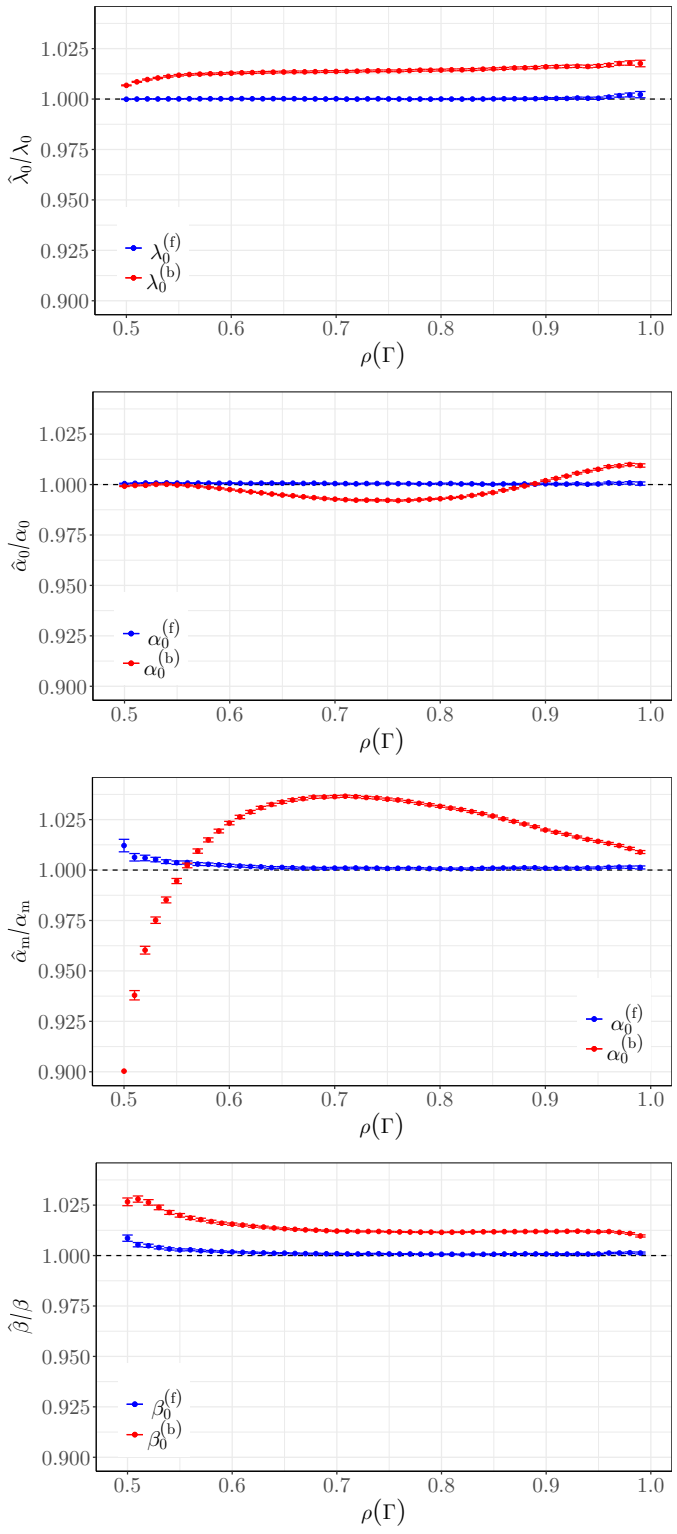


Figure 2.9: Relative difference in the estimation of the various parameters in the MLE of the multidimensional HP with a symmetric excitation matrix for the forward (blue) and the backward process (red). All possible permutations of $\lambda_0 = \{0.0010, 0.0025, 0.005, 0.0075, 0.0100\}$, $\alpha_0 = \{0.049\}$, with α_m chosen according to the desired maximum eigenvalue $\rho(\Gamma)$, and $\beta = 0.1$ are considered. The data points are grouped according to maximum eigenvalue and averaged over 100 runs for each parameter permutation. The expected total number of events is set to 10^6.

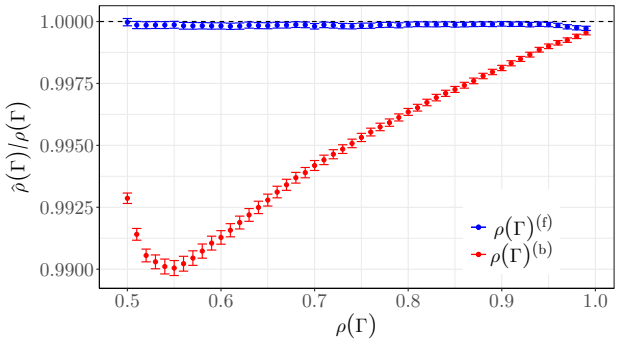


Figure 2.10: (Continued) Relative difference in the estimation of the various parameters in the MLE of the multidimensional HP with a symmetric excitation matrix for the forward (blue) and the backward process (red). All possible permutations of $\lambda_0 = \{0.0010, 0.0025, 0.005, 0.0075, 0.0100\}$, $\alpha_0 = \{0.049\}$, with α_m chosen according to the desired maximum eigenvalue $\rho(\Gamma)$, and $\beta = 0.1$ are considered. The data points are grouped according to maximum eigenvalue and averaged over 100 runs for each parameter permutation. The expected total number of events is set to 10^6.

2.2.4 Application to Data

Since the difference between forward and backward estimates is related to the endogeneity of the process, it is worth discussing some typical values found empirically. As an example of an application of the HP, we studied some fits of the HP to price data of the Exchange-Traded Fund SPDR S&P 500 ETF. We follow largely the methods developed in Lallouache and Challet (2016), and we focus on shorter time intervals where the authors find that the HP excel at fitting (one hour or less), i.e. time-intervals where we may assume that the baseline intensity is constant.

The data set encompasses price data over one day, 15-12-2015 from 9:30 to 16:00, and consists of approximately 950 000 data points. We tried the same fitting procedure for 23 other days and the results are consistent. We focus on bid prices and only consider changes in the bid price, which effectively reduces the dataset to only 33 000 data points. The coarse nature of the time has the interesting consequence that several bid price changes may occur during the same millisecond (the temporal resolution of the data). In order to address this issue, we assume that the order in which the price changes is correct and draw at random the times of the event within the same millisecond with a uniform distribution. One thus expects to lose some causality because of the coarse temporal resolution.

The fits are done with constant baseline intensity λ_0 and a kernel defined as

a sum of exponentials, i.e. $K(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t}$, where $P = 1, 2, 3$. This kind of kernel offers a lot of flexibility in terms of fitting. $P = 2$ has been found to be a suitable choice for shorter time scales. We compare the results for both the non-stationary and stationary assumption (see Appendix 2.2.7 for details), with, of course, a particular emphasis on what happens when the process is reversed. Unlike the case with the synthetic data, the time horizon is not known in the empirical data. It is thus assumed that the last event in the calibrated data set is the time horizon.

The goodness of fits is not only assessed with KS test, but also with the Ljung-Box test (LB test), which checks if the compensators of HPs introduced in Eq. (2.14) are not auto-correlated. We use the slight modification of the test introduced in Lallouache and Challet (2016) in order to take into account the data cleaning procedure. The Aikaike Information Criterion (AIC) is also used to compare the merit of different kernels. The results for $P = 1, 2, 3$ are presented in Table 2.2.

It is clear that assuming that the process is stationary, and using the slightly modified methods, does not significantly improve the fits, and does not merit much attention. However, if we compare the forward and backward cases, we see that unlike when the synthetic data was considered, it is is not as clear cut and it is not sufficient to consider just the p-value obtained in the KS test to determine the arrow of time. The forward case does indeed consistently perform better than the backward case but the values obtained for the backward case are still acceptable, and when the degrees of freedom in the model are increased, we generally get a better p-value.

If we turn our attention to the LB test the situation is similar, but here the difference between the two cases is even smaller. In fact, we sometimes see that the backward process occasionally, when $P = 3$, performs better than the the forward process. On the other hand the log-likelihood is consistently larger for the forward case, but not significantly. Finally, the AIC favours a kernel with a larger number of degrees of freedom.

The whole picture makes sense in the light of the results on synthetic data: the estimated endogeneity n depends on the kernel and time window chosen and varies between approximately 0.30 and 0.70, hence far from criticality, a region in which the difference between forward and backward results is small, hence causality is weak.

Table 2.2: Comparison of the ability of the exponential HP with $P = 1, 2, 3$ to fit the empirical data (forward and backward) with different time windows. n is the estimated endogeneity, pKS and pLB are the Kolmogorov-Smirnov and Ljung-Box test p-values, $\log(\mathcal{L})$ is the log-likelihood, AIC is the Akaike Information Criterion and N is the average number of events in a time window. Results obtained where the process is assumed to be stationary from the start are put in parenthesis. Values are averaged over all non-overlapping windows with more than 150 events.

$P = 1$ (forward)

	$n^{(f)}$	$pKS^{(f)}$	$pLB^{(f)}$	$\log \mathcal{L}^{(f)}$	$AIC^{(f}$	N
1h	0.33 (0.33)	6.2e-6 (6.2e-6)	8.3e-7 (8.3e-7)	2282.04 (2282.08)	-4558.08 (-4558.15)	4675
30m	0.32 (0.32)	1.3e-3 (1.3e-3)	0.03 (0.03)	1237.49 (1237.52)	-2468.99 (-2469.03)	2518
15m	0.32 (0.32)	0.02 (0.02)	0.06 (0.06)	628.29 (628.31)	-1250.59 (-1250.61)	1259
10m	0.32 (0.32)	0.07 (0.07)	0.15 (0.15)	420.01 (420.02)	-834.02 (-834.04)	839
5m	0.32 (0.32)	0.24 (0.24)	0.30 (0.30)	212.95 (212.95)	-419.90 (-419.91)	420

$P = 1$ (backward)

	$n^{(b)}$	$pKS^{(b)}$	$pLB^{(b)}$	$\log \mathcal{L}^{(b)}$	$AIC^{(b)}$	N
1h	0.32 (0.32)	2.6e-7 (2.6e-7)	2.6e-6 (2.6e-6)	2262.78 (2262.78)	-4519.55 (-4519.57)	4675
30m	0.32 (0.32)	3.9e-4 (3.9e-4)	0.03 (0.03)	1227.17 (1227.17)	-2448.35 (-2448.35)	2518
15m	0.31 (0.31)	0.01 (0.01)	0.05 (0.05)	623.29 (623.28)	-1240.57 (-1240.57)	1259
10m	0.31 (0.31)	0.04 (0.04)	0.16 (0.16)	416.61 (416.60)	-827.22 (-827.21)	839
5m	0.31 (0.31)	0.19 (0.19)	0.30 (0.30)	211.29 (211.29)	-416.59 (-416.58)	420

$P = 2$ (forward)

	$n^{(f)}$	$pKS^{(f)}$	$pLB^{(f)}$	$\log \mathcal{L}^{(f)}$	$AIC^{(f}$	N
1h	0.59 (0.59)	0.02 (0.02)	0.17 (0.17)	2438.63 (2438.69)	-4867.26 (-4867.38)	4675
30m	0.57 (0.57)	0.17 (0.17)	0.34 (0.34)	1317.87 (1317.92)	-2625.75 (-2625.85)	2518
15m	0.53 (0.53)	0.40 (0.40)	0.35 (0.35)	665.25 (665.28)	-1320.51 (-1320.56)	1259
10m	0.54 (0.54)	0.47 (0.47)	0.38 (0.37)	444.61 (444.58)	-879.22 (-879.16)	840
5m	0.51 (0.51)	0.64 (0.63)	0.43 (0.44)	225.42 (224.72)	-440.83(-439.43)	420

$P = 2$ (backward)

	$n^{(b)}$	$pKS^{(b)}$	$pLB^{(b)}$	$\log \mathcal{L}^{(b)}$	$AIC^{(b)}$	N
1h	0.59 (0.59)	3.8e-4 (3.9e-4)	0.18 (0.18)	2386.57 (2386.59)	-4763.14 (-4763.18)	4675
30m	0.55 (0.55)	0.03 (0.03)	0.32 (0.32)	1290.59 (1290.60)	-2571.18 (-2571.21)	2518
15m	0.49 (0.49)	0.16 (0.17)	0.31 (0.30)	651.78 (652.14)	-1293.56 (-1294.27)	1259
10m	0.50 (0.50)	0.26 (0.26)	0.36 (0.36)	435.90 (436.16)	-861.80 (-862.32)	840
5m	0.47 (0.47)	0.48 (0.48)	0.40 (0.41)	220.40 (220.37)	-430.79 (-430.73)	420

$P = 3$ (forward)

	$n^{(f)}$	$pKS^{(f)}$	$pLB^{(f)}$	$\log \mathcal{L}^{(f)}$	$AIC^{(f}$	N
1h	0.69 (0.69)	0.01 (0.03)	0.22 (0.21)	2468.46 (2480.59)	-4922.92 (-4947.19)	4675
30m	0.67 (0.66)	0.15 (0.15)	0.38 (0.38)	1333.28 (1339.70)	-2652.57 (-2665.41)	2518
15m	0.60 (0.59)	0.40 (0.39)	0.40 (0.40)	674.12 (674.11)	-1334.25 (-1334.22)	1259
10m	0.60 (0.61)	0.50 (0.49)	0.44 (0.44)	449.38 (448.05)	-884.77 (-882.09)	839
5m	0.55 (0.55)	0.72 (0.71)	0.46 (0.46)	227.80 (227.80)	-441.59 (-441.60)	419

$P = 3$ (backward)

	$n^{(b)}$	$pKS^{(b)}$	$pLB^{(b)}$	$\log \mathcal{L}^{(b)}$	$AIC^{(b)}$	N
1h	0.67 (0.67)	4.4e-03 (0.01)	0.29 (0.28)	2412.88 (2412.51)	-4811.76 (-4811.03)	4675
30m	0.64 (0.63)	0.07 (0.05)	0.40 (0.41)	1303.68 (1302.34)	-2593.35 (-2590.68)	2518
15m	0.57 (0.57)	0.20 (0.20)	0.39 (0.38)	658.42 (658.41)	-1302.85 (-1302.82)	1259
10m	0.55 (0.55)	0.30 (0.31)	0.40 (0.42)	438.79 (439.00)	-863.58 (-863.99)	839
5m	0.52 (0.52)	0.54 (0.53)	0.44 (0.44)	222.10 (222.15)	-430.20 (-430.30)	419

2.2.5 Discussion

The above findings for both the univariate and symmetric multivariate cases have several consequences. First, their causality is much weaker than previously implicitly assumed, even with synthetic data whose kernel family is known. This in turn makes it sometimes difficult to distinguish between the forward and backward event time vectors and thus strongly emphasises the importance of using goodness of fit tests, even for synthetic data. Since the kernel is the only causal term in HPs, weak causality contributes to the difficulties in fitting HPs, especially when the baseline intensity varies with time. This is one of the reasons why we have accounted for the potential lack of initial non-stationary part when calibrating HPs in some cases with a modified log-likelihood function, such as the one we propose, a point which has received little attention to our knowledge.

A practical consequence of this work is that fitting weakly causal HPs to real data rests on even shallower ground because of data imperfection, for two main reasons. First, data cannot be assumed to be perfect; for example the time resolution of the data may be coarse enough to allow several events to take place during the same data time and the event times may be affected by non-negligible noise, as it happens sometimes in financial tick-by-tick data. These two time-related problems further weaken the causality of HPs. Second, the shape of the kernel is *a priori* unknown.

This raises an important issue: simple kernels with very few degrees of freedom are seldom satisfactory, thus more complex models with more degrees of freedom are introduced until satisfactory results are achieved. The same is true for the backward arrow of time, and reassuringly, the results are often worse, but not systematically and certainly not in a manner as convincing as when one knows the kernel shape. In other words, the larger the number of degrees of freedom of a kernel, the more successful the fits, but at the cost of weakening the difference between forward and backward arrows of time because more degrees of freedom also allow a more precise fit of the backward event times vector. At all rates, our results suggest an additional test for HPs: one should reject the hypothesis that HPs describe the data if the forward event time vector leads to worse goodness of fit tests than the backward event time vector. Indeed, in such cases, it makes little sense to trust the causality that HPs introduce.

2.2.6 Appendix: Univariate Hawkes processes with power-law kernels

A power-law kernel for the univariate HP may be defined as

$$K(t) = u(t+v)^w. \tag{2.27}$$

The endogeneity thus equals

$$n = -\frac{u}{w+1}v^{w+1} \tag{2.28}$$

and the log-likelihood is given by

$$\begin{aligned} \ln \mathcal{L}\left(\{t_i\}_{i=1,\dots,n}\right) &= -\lambda_0 T \\ -\sum_{i=1}^{n} \frac{u}{w+1}\left((T-t_i+v)^{w+1} - v^{w+1}\right) &. \\ +\sum_{i=1}^{n} \ln\left[\lambda_0 + \sum_{k=1}^{i-1} u(t_i - t_k + v)^w\right] &. \end{aligned} \tag{2.29}$$

If the initial non-stationary part of the process is removed, the mathematical expression above must be modified;

$$\begin{aligned} \ln \mathcal{L}\left(\{t_i\}_{i=1,\dots,n}\right) &= -\lambda_0 T \\ -\frac{v}{w+1}\left(\frac{\lambda_0}{1+\frac{u}{w+1}v^{w+1}} - \lambda_0\right)\left((\frac{T}{v}+1)^{w+1} - 1\right) &\\ -\sum_{i=1}^{n} \frac{u}{w+1}\left((T-t_i+v)^{w+1} - v^{w+1}\right) &\\ +\sum_{i=1}^{n} \ln\left[\lambda_0 + \left(\frac{\lambda_0}{1+\frac{u}{w+1}v^{w+1}} - \lambda_0\right)\left(\frac{t_i}{v}+1\right)^w\right. &\\ \left. + \sum_{k=1}^{i-1} u(t_i - t_k + v)^w\right] &. \end{aligned} \tag{2.30}$$

Fig. 2.11 displays the relative difference of the log-likelihood as a function of the endogeneity.

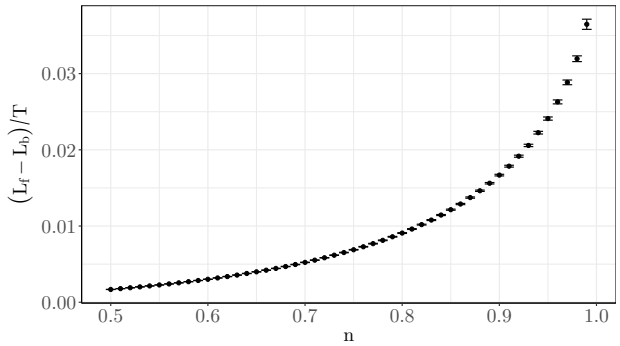


Figure 2.11: Relative difference of the log-likelihood with regards to T (the time horizon of the simulations) between forward and backward time arrows for an HP with a power-law kernel. The selection of parameters is limited to $\lambda_0 = 0.05$, $u = 0.06$,$w = -2.5$ and with a varying v chosen according to the desired endogeneity n. The data points are grouped according to their endogeneity and averaged over 100 runs. The expected number of events is set to 10^5.

2.2.7 Appendix: Log-likelihood of the univariate HP with a sum of exponentials

If the HP kernel consists of a sum of P exponentials, i.e.

$$K(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t}, \tag{2.31}$$

then the associated log-likelihood is given by

$$\begin{aligned} \ln \mathcal{L}\left(\{t_i\}_{i=1,\dots,n}\right) = -\lambda_0 T - \sum_{i=1}^{n} \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} \left(1 - e^{\beta_j (T - t_i)}\right) \\ + \sum_{i=1}^{n} \ln \left[\lambda_0 + \sum_{k=1}^{i-1} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t_i - t_k)}\right]. \end{aligned} \tag{2.32}$$

Consequently, the modified log-likelihood, where it is assumed that there is no initial non-stationary part, is given by

$$\begin{aligned}
&\ln \mathcal{L}\left(\{t_i\}_{i=1,\ldots,n}\right) = \\
&-\lambda_0 T + \frac{\left(\frac{\lambda_0}{1-n} - \lambda_0\right)}{\sum_{j=1}^{P} \alpha_j} \left(\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(\mathrm{e}^{-\beta_j T} - 1\right)\right) \\
&\qquad - \sum_{i=1}^{n}\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}\left(1 - \mathrm{e}^{\beta_j (T - t_i)}\right) \\
&+ \sum_{i=1}^{n} \ln \left[\lambda_0 + \left(\frac{\lambda_0}{1-n} - \lambda_0\right)\left(\frac{\sum_{j=1}^{P} \alpha_j \mathrm{e}^{-\beta_j t_i}}{\sum_{j=1}^{P} \alpha_j}\right)\right. \\
&\qquad \left. + \sum_{k=1}^{i-1}\sum_{j=1}^{P} \alpha_j \mathrm{e}^{-\beta_j (t_i - t_k)}\right].
\end{aligned} \tag{2.33}$$

2.2.8 Appendix: Asymmetric multivariate case

The mutual influence in the asymmetric case is defined as

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 & \alpha_m^1 \\ \alpha_m^2 & \alpha_0 \end{pmatrix},$$

where $\alpha_m^1 \neq \alpha_m^2$ (specifically $\alpha_m^1 < \alpha_m^2$ by convention here). The largest eigenvalue is now given by

$$\rho(\Gamma) = \frac{\alpha_0 + \sqrt{\alpha_m^1 \alpha_m^2}}{\beta}. \tag{2.34}$$

In the asymmetric case, we see that the parameter estimates of the backward arrow of time are very significantly different from those of the forward arrow of time. Even more, in about 75% cases the backward-arrow fits produce almost nonsensical results or simply do not converge. When these runs are removed, there is a bias towards keeping the runs which are significantly longer than the desired number of events, which might help explain the misestimation of certain parameters.

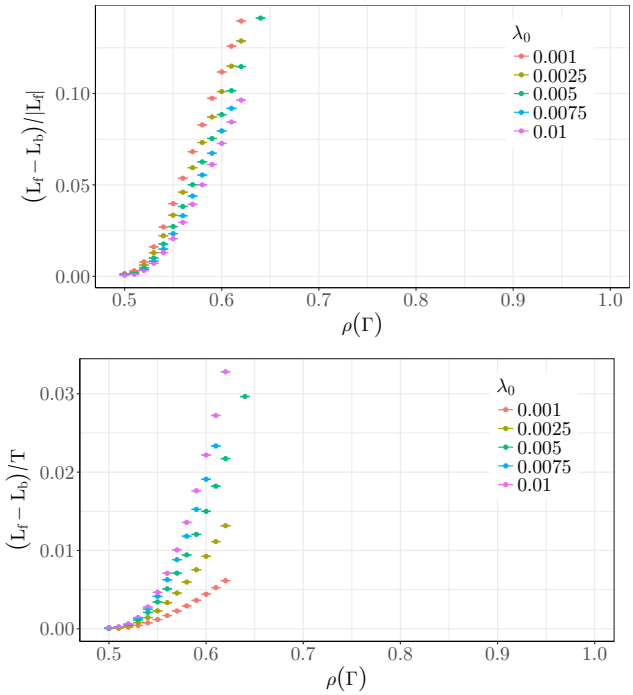


Figure 2.12: Relative difference of the log-likelihood between forward and backward time arrows (top) and difference of the log-likelihood between forward and backward time arrows with regards to T (bottom) for a multidimensional HP with an asymmetric excitation kernel. All possible permutations of $\lambda_0 = \{0.0010, 0.0025, 0.0050, 0.0075, 0.100\}$, $\alpha_m^1 = \{0.049\}$, with α_m^2 chosen according to the desired maximum eigenvalue $\rho(\Gamma)$ and $\alpha_m^1 < \alpha_m^2$, and $\beta = 0.1$ are considered. The data points are grouped according to maximum eigenvalue and averaged over 100 runs for each parameter permutation. The expected total number of events is set to 10^6.

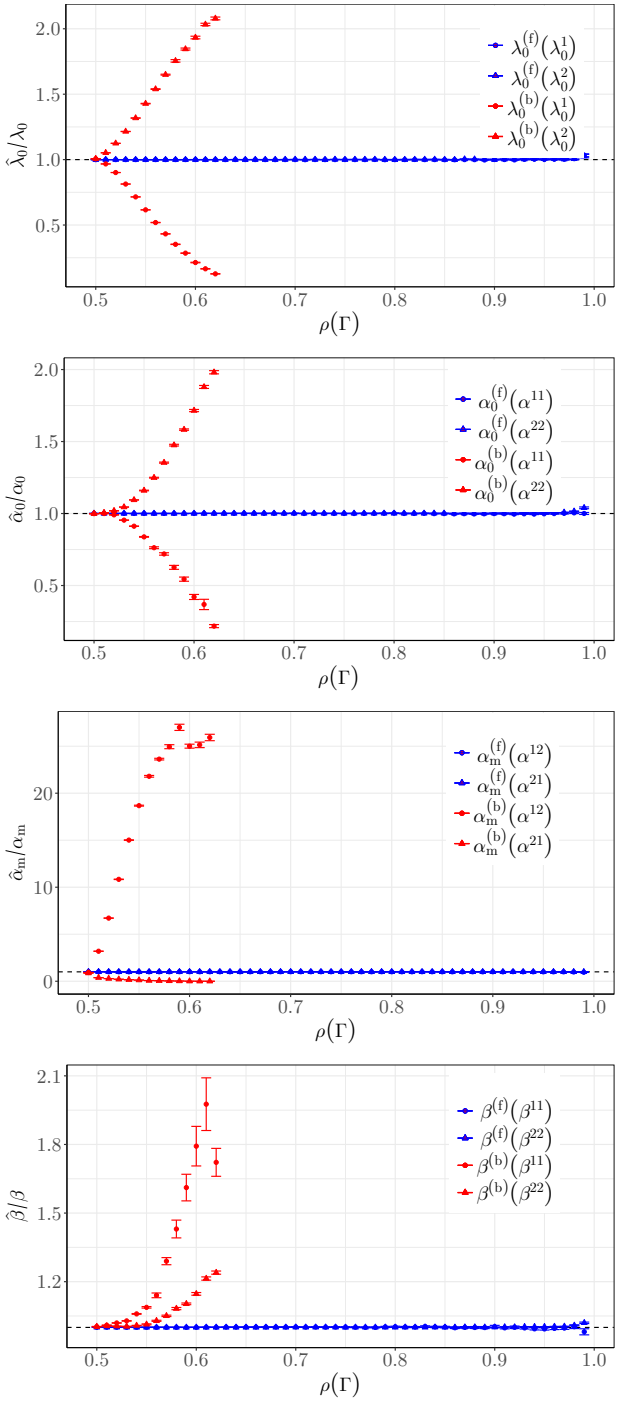


Figure 2.13: Relative difference in the estimation of the various parameters in the MLE of the multidimensional HP with an asymmetric excitation matrix for the forward (blue) and the backward process (red). All possible permutations of $\lambda_0 = \{0.0010, 0.0025, 0.0050, 0.0075, 0.100\}$, $\alpha_m^1 = \{0.049\}$, with α_m^2 chosen according to the desired maximum eigenvalue $\rho(\Gamma)$ and $\alpha_m^1 < \alpha_m^2$, and $\beta = 0.1$ are considered. The data points are grouped according to maximum eigenvalue and averaged over 100 runs for each parameter permutation. The expected total number of events is set to 10^6.

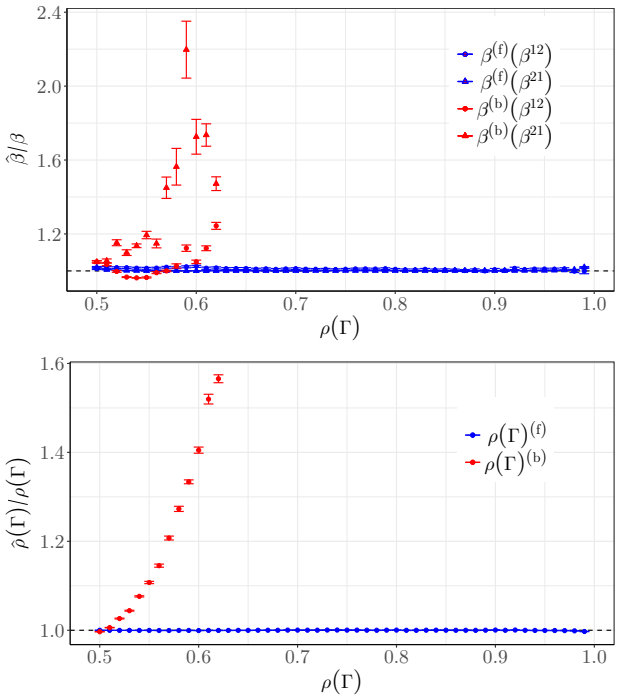


Figure 2.14: (Continued) Relative difference in the estimation of the various parameters in the MLE of the multidimensional HP with an asymmetric excitation matrix for the forward (blue) and the backward process (red). All possible permutations of $\lambda_0 = \{0.0010, 0.0025, 0.0050, 0.0075, 0.100\}$, $\alpha_m^1 = \{0.049\}$, with α_m^2 chosen according to the desired maximum eigenvalue $\rho(\Gamma)$ and $\alpha_m^1 < \alpha_m^2$, and $\beta = 0.1$ are considered. The data points are grouped according to maximum eigenvalue and averaged over 100 runs for each parameter permutation. The expected total number of events is set to 10^6.

2.3 Application of Hawkes processes to the inference of investor activity interaction networks

The causality of asymmetric multivariate Hawkes processes makes it in principle possible to infer causal networks between the actions of investors. In an ideal case, the influence kernel will not only reveal the strength of the influence of the actions of a given investor on another one, but also contain the timescales of influence. In practice, two important problems arise. First, fitting even a univariate Hawkes process requires a fair amount of data, hence fitting a multivariate Hawkes processes to many traders requires even more data. Second, the computation time is likely prohibitive.

Achab et al. (2018) introduce a non-parametric method, the Non-Parametric Hawkes Cumulant (NPHC) which does not require to fit all self and mutual influence kernels to infer causality networks between trader activity. The idea is to estimate the matrix of integrated kernels in a multivariate HP from a small set of moments. In other words, unlike many other available methods, the aim is not not to estimate the kernels associated with each component (i.e. the nodes or the agents in the framework of an investor interaction network), but to take a shortcut by directly estimating the integrated kernels. In this way, one can infer what underlying causality structure HPs can capture. Specifically, one estimates the matrix $\boldsymbol{G}$ of the integrated kernels, i.e.

$$G^{ij} = \int_0^\infty K^{ij}(s)\mathrm{d}s \tag{2.35}$$

Element G^{ij} encodes the mean total number of events of type i directly triggered by an event of type j. Therefore matrix $\boldsymbol{G}$ encodes the causality that HPs can infer. Eichler et al. (2017) show indeed that for the multivariate HP $\boldsymbol{N_t}$, N_t^j does not Granger cause N_t^i if and only if $K^{ij}(u) = 0$ for $u \in \mathbb{R}^+$, which is equivalent to $G^{ij} = 0$. It is thus straightforward to observe the causal relationships between different components in the multivariate HP if the elements in $\boldsymbol{G}$ are significantly different from zero.

The goal of the NPHC method is to estimate the matrix $\boldsymbol{G}$ using a matching cumulants (or moments) method. More precisely, this method is based on the first three integrated cumulants, and if $1 \leq i, j, k, \leq d$, where d is the dimension of the HP, and one assumes that the process is stationary (i.e. the spectral norm

$||\boldsymbol{G}||$ satisfies $||\boldsymbol{G}|| < 1$), one has

$$\Lambda^i \mathrm{d}t = \mathbb{E}(\mathrm{d}N_t^i) \tag{2.36}$$

$$C^{ij}\mathrm{d}t = \int\limits_{\tau\in\mathbb{R}} \Big(\mathbb{E}(\mathrm{d}N_t^i \mathrm{d}N_{t+\tau}^j) - \mathbb{E}(\mathrm{d}N_t^i)\mathbb{E}(\mathrm{d}N_{t+\tau}^j)\Big) \tag{2.37}$$

$$\begin{aligned}\Gamma^{ijk}\mathrm{d}t = \iint\limits_{\tau,\tau'\in\mathbb{R}^2} \Big(&\mathbb{E}(\mathrm{d}N_t^i \mathrm{d}N_{t+\tau}^j \mathrm{d}N_{t+\tau'}^k) + 2\mathbb{E}(\mathrm{d}N_t^i)\mathbb{E}(\mathrm{d}N_{t+\tau}^j)\mathbb{E}(\mathrm{d}N_{t+\tau'}^k) \\ &-\mathbb{E}(\mathrm{d}N_t^i \mathrm{d}N_{t+\tau}^j)\mathbb{E}(\mathrm{d}N_{t+\tau'}^k) - \mathbb{E}(\mathrm{d}N_t^i \mathrm{d}N_{t+\tau'}^k)\mathbb{E}(\mathrm{d}N_{t+\tau}^j) - \mathbb{E}(\mathrm{d}N_{t+\tau}^j \mathrm{d}N_{t+\tau'}^k)\mathbb{E}(\mathrm{d}N_t^i)\Big),\end{aligned} \tag{2.38}$$

where Eq.(2.36) is the mean intensity of the HP, Eq.(2.37) is the integrated covariance density matrix and Eq.(2.38) is a measure of the skewness of $\boldsymbol{N}_t$.

In order to fix the matrix $\boldsymbol{G}$ uniquely, one needs the third cumulant since the integrated covariance only contains symmetric information, and thus no causality, whereas the skewness provided by the third cumulant breaks the symmetry between past and future information.

One thus, first of all, computes an estimation $\widehat{\boldsymbol{M}}$ of these centered moments $M(\boldsymbol{G})$, and then we look for a matrix $\boldsymbol{G}$ that minimizes the L^2 error $||M(\widehat{\boldsymbol{G}})||^2$.

If one sets $\boldsymbol{R} = (\boldsymbol{I_d} - \boldsymbol{G})^{-1}$, we obtain the following identities

$$\Lambda^i = \sum_{m=1}^{d} R^{im}\mu^m \tag{2.39}$$

$$C^{ij} = \sum_{m=1}^{d} \Lambda^m R^{im} R^{jm} \tag{2.40}$$

$$\Gamma^{ijk} = \sum_{m=1}^{d} \Big(R^{im}R^{jm}C^{km} + R^{im}C^{jm}R^{km} + C^{im}R^{jm}R^{km} - 2\Lambda^m R^{im}R^{jm}R^{km}\Big). \tag{2.41}$$

Achab et al. (2018) derive an estimator of $\boldsymbol{R}$ and thus of $\boldsymbol{G}$ and use this method to analyse the mutual influence of the various types of events that may occur in best prices of order book data (new limit order, new market order, cancellation that changes or does not change the best price). Rambaldi et al. (2018) study labelled orders on the CAC40 index future, and can thus explore how the 16 most active market participants contribute to price volatility. A large-dimensional HP is introduced where eight different kinds of actions for each agent: orders which immediately move the mid-price up (down), aggressive orders which are executed immediately at the best ask (bid) price and do not move the mid-price, new limit orders which arrive at the best ask (bid) price

and do not move the mid-price and cancellations of orders at the best ask (bid) price which do not empty the queue and do not move the mid-price.

Specifically, if $\mathcal{A}$ denotes a set of M agents and if one considers the eight different kinds of orders as a set $\mathcal{T}$, the counting process $N_{i,\alpha}(t)$ is associated with agent $i \in \mathcal{A}$ and orders of type $\alpha \in \mathcal{T}$. In total, in the HP, one therefore has $8 \times M$ components, which means that the conditional intensity of $N_{i,\alpha}$ may be defined as

$$\lambda_t^{i,\alpha} = \lambda_0^{i,\alpha} + \sum_{j \in \mathcal{A}} \sum_{\beta \in \mathcal{T}} \int_0^t K^{i,\alpha;j,\beta}(t-s) \mathrm{d}N_s^{j,\beta}, \tag{2.42}$$

where $\lambda_0^{i,\alpha}$ is the baseline intensity and the interaction kernel $K^{i,\alpha;j,\beta}(s)$ represents the impact of an event of type α of agent j on the occurrence likelihood of an event of type α of agent i. Eq.(2.42) is then parametrized as

$$K^{i,\alpha;j,\beta}(s) = \sum_{l=1}^{L} \theta_l^{i,\alpha;j,\beta} g_l(s), \tag{2.43}$$

where L is set to $L = 10$.

An important assumption made in this model is that the way agent i reacts to the action β of agent j does not depend on j unless $i = j$, which amounts to

$$\theta_l^{i,\alpha;j,\beta} = \begin{cases} \alpha_l^{i,\alpha;\beta}, & \text{if } i = j \\ \beta_l^{i,\alpha;\beta}, & \text{otherwise.} \end{cases} \tag{2.44}$$

This is motivated by the argument that any agent perceives the activity of other agents only through their anonymous orders in the book (the data is labelled, but the real-time feed is not).

Instead of labelled orders, we have comprehensive datasets detailing the actions of all the clients on the foreign exchange market of two types of brokers, one with retail clients and one with institutional clients only (see Table 3.1 on p. 63). Whereas there are $O(10^4)$ clients in each dataset, the activity rate is very heterogeneous and best described as a power-law (see Challet et al. (2018)). In addition, most clients are not active every day, or not during the same typical time window, and some of them act by bursts and then stay quiet for a long while. In other words, our datasets are less amenable to the above method. Nevertheless, we used the `tick` package (Bacry et al., 2017) and applied it to the 30 most active clients of our institutional dataset over the whole period. Their kernel norms determined by NPHC method are reported in Fig. 2.15, which

must be read as follows: $G^{i \leftarrow j}$ is the average number of trades of trader i caused by trader j. One sees that the self-excitation (diagonal) is typically larger than mutual influence, but is clearly larger for some traders than others at the chosen timescales. More interestingly, there are also horizontal lines for traders 9, 17, 21, and 26. This means that these traders are influenced by other traders and it is a good indication of the causality in this population. This is already known from Lead-Lag Statistically Validated Networks (LL-SVNs) determined with another method (Challet et al., 2018), which is more appropriate than HPs for the trader-resolved datasets we have.

In principle, having access to the full mutual excitation kernels makes it possible to have a very minute understanding of the timescales at which one trader, say i, exerts his influence on trader j. The point is that nothing prevents trader j to extert its influence on trader i at different timescales. Timescales are lost when computing the norms of kernels (although time still has to be discretized). Assuming sparse mutual excitation kernels and using a Gaussian basis is an alternative way to infer causal activity networks and the timescales at which causality acts (Zhou et al., 2013). The next chapter introduces a much simpler method, better adapted to the non-stationarity of our data and the fact that the activity of the traders in our datasets is very irregular.

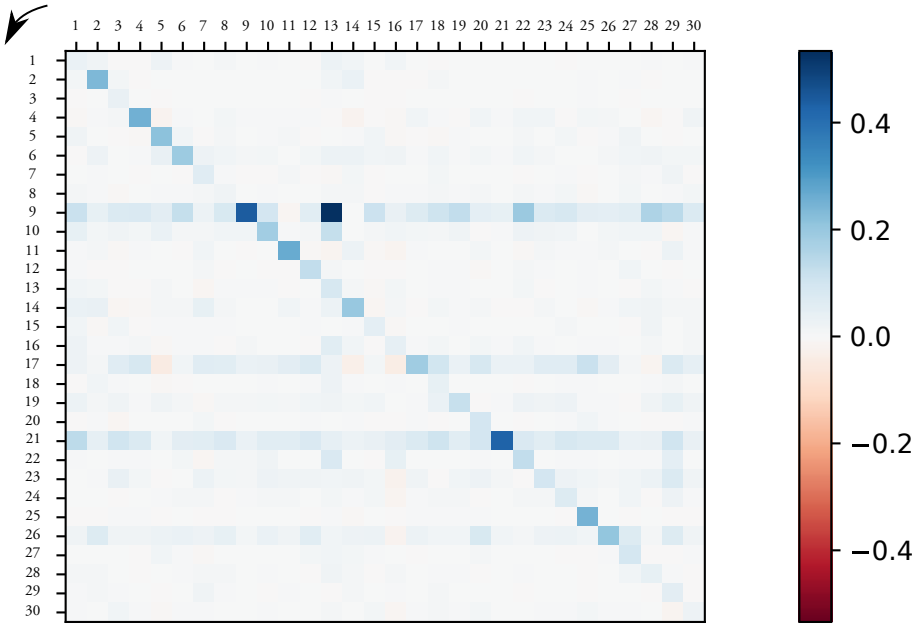


Figure 2.15: Kernel norms $G^{i \leftarrow j}$, determined by the NPHC method, of institutional traders of the 30 most active traders. Time slices of 30s, kernel integrated over 4 hours.

Chapter 3

Lead-Lag Statistically Validated Networks

Note: This chapter is submitted to 'Quantitative Finance' under the title of 'Multi-timescale lead-lag networks and the market nanostructure origin of asset price time reversal asymmetry'.

3.1 Introduction

The collective behaviour of investors in financial markets plays a major part in shaping the complexity of price dynamics. A major challenge in the analysis and modelling of market dynamics comes from the very large heterogeneity of market participants, particularly with respect to their activity rate and feedback speed. Most agent-based models of financial markets omit timescale heterogeneity, usually focusing on strategy heterogeneity (fundamentalists, trend-followers or noise traders) and the way they learn to use them (see Hommes (2006) for a review; see however Marsili and Piai (2002); Mosetti et al. (2006); Kroujiline et al. (2016)).

The typical time-horizon of trader activity ranges from a fraction of a second to a few months (Dacorogna et al., 1998; Zumbach, 2009). A fundamental question is thus how to characterize the causal structure of market activity across timescales. In other words, is there a hierarchical (or more complex) structure in which activity propagates? Since trader-resolved data is hard to obtain, past works focused on price dynamics and volatility propagation. Intuitively, the price dynamics should reflect in some way heterogeneous trader time horizons (see e.g. Müller et al. (1993)). Early works exploit the intuitive analogy between turbulent flows and price changes (Ghashghaie et al., 1996), and simple cascade

models of the price dynamics have been proposed (Lux et al., 2001). Heterogeneous trader timescales may also explain why multi-scale GARCH models are generally much better than plain GARCH ones (see e.g. Lynch et al. (2003); Borland and Bouchaud (2005); Chicheportiche and Bouchaud (2014)). In particular, Müller et al. (1997) argue that since coarsely-defined volatility predicts finely-defined volatility significantly better than the other way around, the behaviour of long-term traders should influence the behaviour of short-term traders.

The above discussion implicitly assumes that prices are time reversal asymmetric (TRA). Zumbach and Lynch (2001), and Zumbach (2009) show indeed that financial time series are significantly asymmetric with respect to the reversal of the arrow of time. While classical models of price and volatility dynamics are not TRA, GARCH processes that incorporate price returns defined over several time scales are TRA (Zumbach and Lynch, 2001; Zumbach, 2009; Chicheportiche and Bouchaud, 2014). The same holds for Hawkes processes, which are causal processes by definition and hence ideal candidates for financial modelling (Bacry et al., 2015), although their univariate and symmetric multivariate versions are surprisingly weakly TRA (Blanc et al., 2017; Cordi et al., 2018).

While there are many ways to define the timescale of a given agent, we take a more global approach here and rely instead on the notion of groups of agents determined at various timescales (seconds, minutes, hours, etc.), and investigate how the activity of one group at a given timescale influences the activity of some other group at another timescale. This opens up the possibility of inferring multi-timescale causal networks of trader activity directly instead of relying on analogies. Note that the framework which we introduce here is generic and applies to any system in which the state of one of its elements over a given time window may be summarized by a discrete state, from a small set of possible states.

Groups of traders are determined with Statistically Validated Networks (SVNs); SVNs were introduced by Tumminello et al. (2011) and have been applied to e.g. mobile communication networks (Li et al., 2014), clusters of orthologous genes, and the relationship between actors and movies (Tumminello et al., 2011). They were then used to cluster Finnish investors (Tumminello et al., 2012) and more recently to understand their long-term ecology (Musciotto et al., 2018). The main idea is that a group of similar traders should act in a similar way. The trick is to define networks of interaction according to the degree of pairwise synchronization between the actions of traders and use community detection of the resulting network to define groups of traders. Crucially, since the actions

of all the members of a group are remarkably similar, the action of the group is representative of the action of each of its members, which is very helpful to reduce the dimension of trader datasets.

SVNs rely on time coarsening at a given timescale (e.g. 1 day, the best available resolution of the dataset analyzed in Tumminello et al. (2012), or 1 hour in Challet et al. (2018)). Which timescale to choose is not obvious, all the more since traders have widely different activity rates. As we shall see below, the answer depends on the type of traders (retail or institutional) and most probably on the clientele composition of a broker.

Recently, Challet et al. (2018) introduced Lead-Lag SVNs (LL-SVNs) to infer lead-lag networks between the states of agents in complex systems and applied them to trader-resolved data. The persistence in LL-SVNs is large enough to make it possible to predict the sign of the order flow and the VWAP of a broker clients over the next hour. A reason why these lead-lag networks exist and persist is that investors consistently react with different speeds to common information (Boudoukh et al., 1994; Jegadeesh and Titman, 1995).

Here, we extend the LL-SVN method to lead-lag networks between states determined at two different timescales. This is needed to infer how information flows from one timescale to another and to find asymmetric reciprocal influence, as is the case in trader-resolved data. Causality with respect to these two timescales is then well defined, and the ensemble of causality relationships between many pairs of timescales provides a fine picture of how information propagates in a complex system. We also discuss how the TRA of the activity of the two types of traders in our dataset compares with that of the volatility, i.e. to relate macroscopic price properties to nanoscopic decisions[1].

3.2 Method

3.2.1 SVNs and LL-SVNs

Assume that one has N time series, e.g. the transaction history of N traders. The SVN method works as follows: one first chooses a time resolution Δt and splits the time into slices of length Δt. Here, $(t, \Delta t)$ denotes the timeslice $[t, t + \Delta t[$, and for the sake of simplicity, we shall write it as t when no ambiguity arises.

For each timeslice, one summarizes the activity of each time series by a discrete state taken from a small number of possible states. For traders, it is natural

[1] Market microstructure focuses on price formation from anonymous orders sent by traders.

to define four different states: mostly buying (+1), mostly selling (-1), neutral (0) and inactive (NA). The imbalance ratio of time series i for each timeslice t is then

$$\rho_i(t) = \frac{v_i(t)}{a_i(t)} \tag{3.1}$$

where $v_i(t) = v_i(t, \Delta t)$ is the total signed transaction volume of trader i during timeslice t, and, similarly, $a_i(t)$ is the sum of the absolute trading volume during this timeslice t. The state of agent i during timeslice t is

$$\sigma_i(t) = \begin{cases} 1 & \text{if } \rho_i(t) > \rho_0 \\ -1 & \text{if } \rho_i(t) < -\rho_0 \\ 0 & \text{if } \rho_i(t) < |\rho_0| \\ \text{NA} & \text{if } v_i(t) = a_i(t) = 0. \end{cases}, \tag{3.2}$$

As in previous works, we use $\rho_0 = 0.01$, but the specific choice of this parameter does not have much influence on the results provided that it is small.

The level of synchronicity between two given states of two given traders is determined by assuming that the occurrence of each state follows a Poissonian process in discrete time (the timeslices). Then, using an exact expression for the probability of synchronicity of two independent process, it is straightforward to compute the p-value of these states for these traders. The computation is performed for all possible pairs of traders and all allowed pairs of states. Here, since one wishes to group traders, the set of allowed pairs is $(\{(1,1),(-1,-1),(0,0)\}$; we drop the inactive state by focusing on the most active traders. Testing all the pairs of traders for each possible state pair yields a large number of tests, thus multiple hypothesis testing correction is needed: we use the False Discovery Rate (FDR) (Benjamini and Hochberg, 1995), with an FDR rate set to $p_0 = 0.05$. An SVN network is obtained by keeping links whose p-values are smaller than the FDR-adjusted threshold (see Tumminello et al. (2012) for more details).

The resulting network may then be decomposed into groups (communities) by using the InfoMap method (Rosvall and Bergstrom, 2008), which is one of the most efficient methods of community detection in networks (Lancichinetti and Fortunato, 2009). The multi-links are converted into weighted links by assigning a weight equal to the number of validated links between two traders. Since links are only allowed between traders who take similar actions, the state of each group of traders is well defined and mirrors those of the traders that it includes.

Let us introduce some more mathematical notations. Mathematically, one

can define the state of group $g \in G$, where G is the set of all groups, during timeslice t, by

$$\sigma_g(t) = \begin{cases} 1 & \text{if } \rho_g(t) > \rho_0 \\ -1 & \text{if } \rho_g(t) < -\rho_0 \\ 0 & \text{if } \rho_g(t) < |\rho_0| \\ \text{NA} & \text{if } V_g(t) = 0, \end{cases} \tag{3.3}$$

where $\rho_g(t) = \frac{V_g(t)}{|V_g(t)|}$ and $V_g(t) = \sum_{i \in g} v_i(t)$ is the aggregate signed volume of the traders belonging to group g during timeslice t. We also define $A_g(t) = \sum_{i \in g} a_i(t)$ as the aggregate absolute volume of the traders belonging to group g during timeslice t. Grouping traders is surprisingly efficient and significantly decreases the dimensionality of the problem, i.e. the effective number of time series to track in a population of clients of a broker. As it reduces the dimension of the data set, using groups tremendously helps to simplify and speed up the determination of lead-lag networks, and we shall keep this procedure.

The easiest case is of course when the states of traders who lead and who lag are determined with the same coarse resolution Δt, as in Challet et al. (2018).

3.2.2 LL-SVNs with two timescales

The main methodological contribution of our work is to introduce a general framework to infer lead-lag relationships between groups whose states are determined at two (possibly) different timescales.

The general principle is simple (see Fig. 3.1 for a graphical illustration):

1. Let Δt_1 and Δt_2 be two timeslice durations.
2. Apply the SVN method to both Δts in order to determine two sets of groups G_1 and G_2 (optional but recommended)[2].
3. Find SVNs between the suitably lagged values of group (or agent) states.

By convention, in the following, Δt_1 is the timescale at which the leading states of agents are determined and Δt_2 the timescale of the lagging states of agents. When $\Delta t_1 = \Delta t_2$, the segmentations of a time series for both the leading and lagging states coincide and no special caution regarding their alignment is needed. However, when $\Delta t_1 \neq \Delta t_2$, for a given time $t = k_1 \Delta t_1$, $k_1 \in \mathbb{N}$,

[2]When the number of agents is not too large, this step may be skipped.

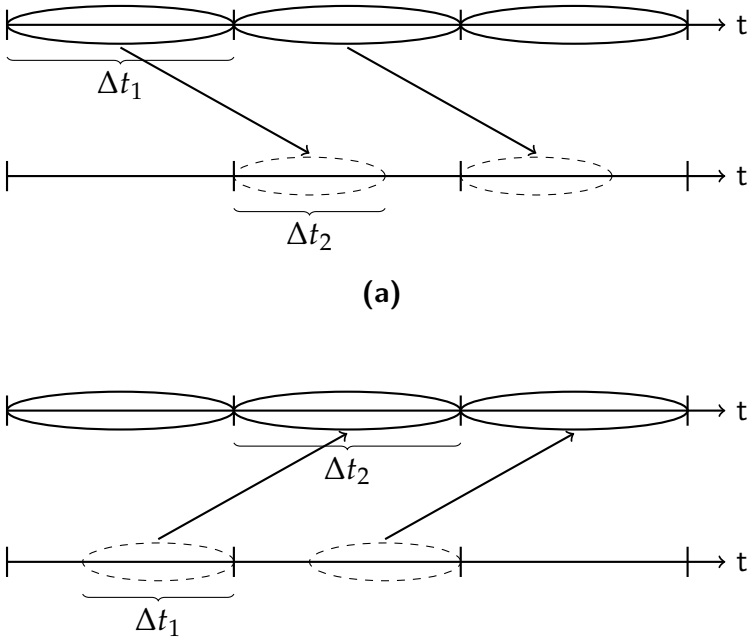


Figure 3.1: Schematic diagram showing how lead-lag links are established when (a) $\Delta t_1 > \Delta t_2$ and (b) $\Delta t_1 < \Delta t_2$. The dotted lines indicate that the state of the group has been recalculated if $t/\Delta t_2$ is not an integer, which corresponds to a time-shift with regards to the time-interval where the groups were determined.

the boundaries of timeslices for both timescales are generally not aligned, i.e. there is generally no integer k_2 such that $k_2\Delta t_2 = k_1\Delta t_1$. This is a problem when inferring LL-SVNs, as non-aligned slices induce a lag between the end of the leading slice and the lagging one, which would then reduce the strength of causality relationships. In addition, one needs to avoid computing the states of agents or groups on partially overlapping timeslices for the longest timescale. This is why we align the computation of the states at times $t = k\max(\Delta t_1, \Delta t_2)$, $k \in \mathbb{N}$ (see Fig. 3.1).

This alignment problem suggests several possible variations, both in terms of how and when groups and their states are determined. Methods I, II, and III introduced below each define a set of lead-lag links.

Method I

This method only uses a single grouping of agents, and is thus both faster and simpler. While agent grouping (clustering) is done with respect to one of the two timescales (see below), the state of each group is computed in timeslices of length Δt_1 and Δt_2 which are aligned as in Fig. 3.1. Only clustering with

respect to a single timescale may sometimes miss subtle differences of group membership, especially if the timescales are very different.

Method I works as follows when $\Delta t_1 > \Delta t_2$ (it is assumed here that $t = k\Delta t_1$, where $k = 0, 1, 2, \dots$):

1. Time is discretized at timescale Δt_1 in order to obtain the group set G.
2. For each group $g \in G$ and timeslices $(t, \Delta t_1)$, the states $\sigma_g(t, \Delta t_1) = \sigma_g^{(1)}(t)$ are determined.
3. For each group $h \in G$ and timeslices $(t, \Delta t_2)$, the states $\sigma_h(t, \Delta t_2) = \sigma_h^{(2)}(t)$ are determined.
4. For each possible pair (g, h), g and $h \in G$, the p-value of the synchronicity between $\sigma_g^{(1)}(t)$ and $\sigma_h^{(2)}(t + \Delta t_1)$ is calculated.

When $\Delta t_1 < \Delta t_2$, one needs to consider $t = k\Delta t_2$, where $k = 0, 1, 2, \dots$:

1. Time is discretized at timescale Δt_2 in order to obtain the group set G.
2. For each group $g \in G$ and timeslices $(t, \Delta t_2)$, the states $\sigma_g(t, \Delta t_2) = \sigma_g^{(2)}(t)$ are determined.
3. For each group $h \in G$ and timeslices $(t - \Delta t_1, \Delta t_1)$, the states $\sigma_h(t - \Delta t_1, \Delta t_1) = \sigma_h^{(1)}(t - \Delta t_1)$ are determined.
4. For each possible pair (g, h), g and $h \in G$, the p-value of the synchronicity between $\sigma_g^{(2)}(t)$ and $\sigma_h^{(1)}(t - \Delta t_1)$ is calculated.

Since there is only one set of groups, defining self-referential lead-lag links (from one group to itself) is straightforward.

In the implementation of the method above it is clear that the time discretization used for the group classification is always based on the longer timescale, regardless of whether it acts as lead or lag. We have also implemented the method above with the shorter timescale as basis for time discretization used for the group classification, and we checked that the results did not differ significantly.

Method II

This method defines two groups, one for each time scale, over the whole calibration window, denoted by G_1 and G_2. Method II works as follows when $\Delta t_1 > \Delta t_2$, assuming that $t = k\Delta t_1$, where $k = 0, 1, 2, \dots$:

1. Time is discretized at timescale Δt_1 and Δt_2 in order to obtain G_1 and G_2.
2. For each group $g \in G_1$ and timeslices $(t, \Delta t_1)$, the states $\sigma_g(t, \Delta t_1) = \sigma_g^{(1)}(t)$ are determined.
3. For each group $h \in G_2$ and timeslices $(t, \Delta t_2)$, the states $\sigma_h(t, \Delta t_2) = \sigma_h^{(2)}(t)$ are determined.
4. For each possible pair (g, h), $g \in G_1$ and $h \in G_2$, the p-value of the synchronicity between $\sigma_g^{(1)}(t)$ and $\sigma_h^{(2)}(t + \Delta t_1)$ is calculated.

When $\Delta t_1 < \Delta t_2$ the method works as follows, assuming that $t = k\Delta t_2$, where $k = 0, 1, 2, \ldots$:

1. Time is discretized at timescale Δt_1 and Δt_2 in order to obtain G_1 and G_2.
2. For each group $g \in G_2$ and timeslices $(t, \Delta t_2)$, the states $\sigma_g(t, \Delta t_2) = \sigma_g^{(2)}(t)$ are determined;
3. For each group $h \in G_1$ and timeslices $(t - \Delta t_1, \Delta t_1)$, the states $\sigma_h(t - \Delta t_1, \Delta t_1) = \sigma_h^{(1)}(t - \Delta t_1)$ are determined.
4. For each possible pair (g, h), $g \in G_2$ and $h \in G_1$, the p-value of the synchronicity between $\sigma_g^{(2)}(t)$ and $\sigma_h^{(1)}(t - \Delta t_1)$ is calculated.

Since the alignment follows the time slices of the longer timescale, we avoid any overlap (and thus unnecessary correlation) between two adjacent time slices.

Method III

Finally, we introduce Method III which ensures that group inference and group states are computed in the same time slices. More specifically, what is different in this method is that the groups for the shorter time-interval are determined 'in place' with regards to how their trade volumes are aggregated, depending on if the shorter timescale acts as lead or lag. We therefore have two different sets of groups for the shorter timescale in order to avoid overlap. The advantage of this method is thus that we avoid the re-calculation of the group states (which is necessary in the other two methods) and that clustering fully corresponds to the states used to determine the LL-SVN. The disadvantage is that clustering is performed with fewer events for the shorter timescale.

The method works as follows when $\Delta t_1 > \Delta t_2$ assuming that $t = k\Delta t_1$, where $k = 0, 1, 2, \ldots$:

1. Time is discretized at timescale Δt_1 in order to obtain G_1.
2. Time is discretized as $[t, t + \Delta t_2[$ in order to obtain G_2.
3. For each group $g \in G_1$, the states $\sigma_g(t, \Delta t_1) = \sigma_g^{(1)}(t)$ are determined.
4. For each group $h \in G_2$, the states $\sigma_h(t, \Delta t_2) = \sigma_h^{(2)}(t)$ are determined.
5. For each possible pair (g, h), $g \in G_1$ and $h \in G_2$, the p-value of the synchronicity between $\sigma_g^{(1)}(t)$ and $\sigma_h^{(2)}(t + \Delta t_1)$ is calculated.

When $\Delta t_1 < \Delta t_2$ the method works as follows assuming that $t = k\Delta t_2$, where $k = 0, 1, 2, \ldots$:

1. Time is discretized at timescale Δt_2 in order to obtain G_2.
2. Time is discretized as $[t - \Delta t_1, t[$ in order to obtain G_1.
3. For each group $g \in G_2$, the states $\sigma_g(t, \Delta t_2) = \sigma_g^{(2)}(t)$ are determined.
4. For each group $h \in G_1$, the states $\sigma_h(t - \Delta t_1, \Delta t_1) = \sigma_h^{(1)}(t - \Delta t_1)$ are determined.
5. For each possible pair (g, h), $g \in G_2$ and $h \in G_1$, the p-value of the synchronicity between $\sigma_g^{(2)}(t)$ and $\sigma_h^{(1)}(t - \Delta t_1)$ is calculated.

3.3 Dataset

Our datasets contain trader-resolved transactions of the EUR/USD currency pair and come from two independent sources: Swissquote Bank SA (SQ hereafter), a Swiss broker-dealer with a large market share in foreign exchange (FX) transactions in Switzerland, and a large anonymous dealer bank which serves major institutional clients. Both datasets list all the trades of their clients: traded currency pair, anonymous client identification number, trade time (at a millisecond resolution), signed volume, and the FX transaction rate. We focus on the EUR/USD pair as it is one of the most traded pairs in both datasets. A summary of the datasets structure and contents is provided in Table 3.1.

While FX markets never close, transactions are quite rare during nights and weekends. We thus focus on active hours, i.e. from 9:00 to 17:00 on weekdays. We only look for links between adjacent timeslices on the same day in order to avoid spurious boundary effects or overnight lead-lag links.

Dataset	Timespan	Traders	Trades
LB	01 Jan. 2013 - 15 Sep. 2014	$> 10^3$	$> 10^5$
SQ	01 Jan. 2014 - 30 Jun. 2014	$> 10^3$	$> 10^5$

Table 3.1: Basic statistics of the datasets studied for EUR/USD currency pair

3.4 Results

Since the active population in both datasets evolves much faster than the total duration of the datasets, one cannot use the whole datasets to infer lead-lag networks. We use here rolling calibration time windows of $T_{\text{in}} = \{30, 60, 90, 120\}$ business days[3]. For each time window, we apply Methods I, II, and III to each pair of timescales Δt_1 and Δt_2 belonging to the arithmetic sequence from 5 minutes to 240 minutes (4 hours) with a step of 5 minutes (which corresponds to 1176 unique pairs of timescales). Computations over the whole length of a single dataset last for about a day for each T_{in} and each dataset using 72 cores, for all pairs of timescales. In order to speed-up computations, we only keep the 500 most active traders in each calibration windows.

3.4.1 $\Delta t_1 = \Delta t_2$

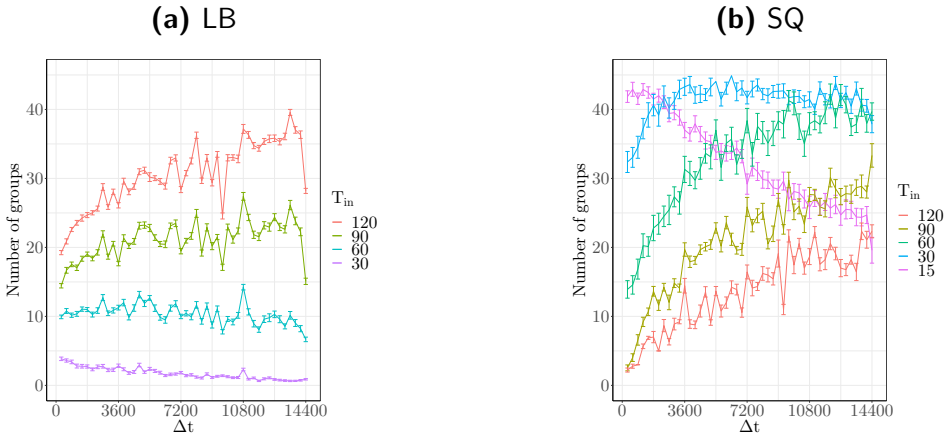


Figure 3.2: Average number of groups as a function of Δt and T_{in}. $\Delta t = \Delta t_1 = \Delta t_2$.

We first focus on the diagonal $\Delta t_1 = \Delta t_2 = \Delta t$. In this case, determining lead-lag networks does not require any special care and indeed the three methods defined above are identical and correspond to the single-timescale method of Challet et al. (2018). A systematic investigation of global properties of lead-

[3]We have also used $T_{\text{in}} = 15$ for the SQ dataset.

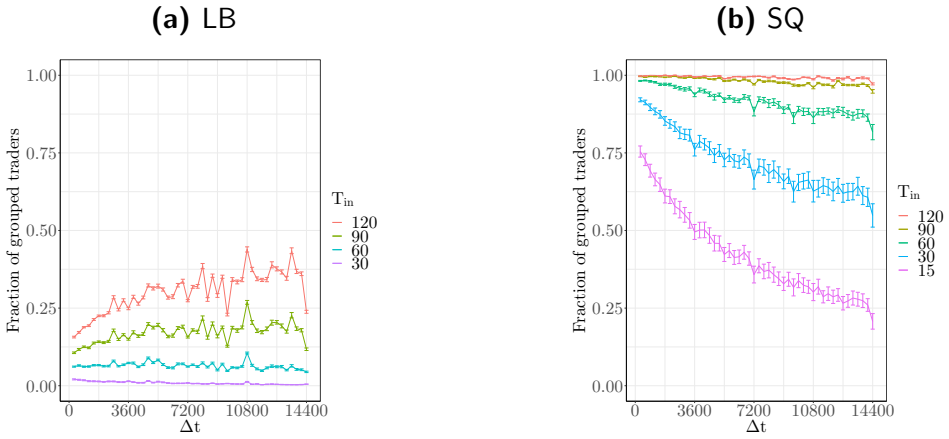


Figure 3.3: Average fraction of traders grouped by SVNs as a function of Δt and T_{in}. $\Delta t = \Delta t_1 = \Delta t_2$.

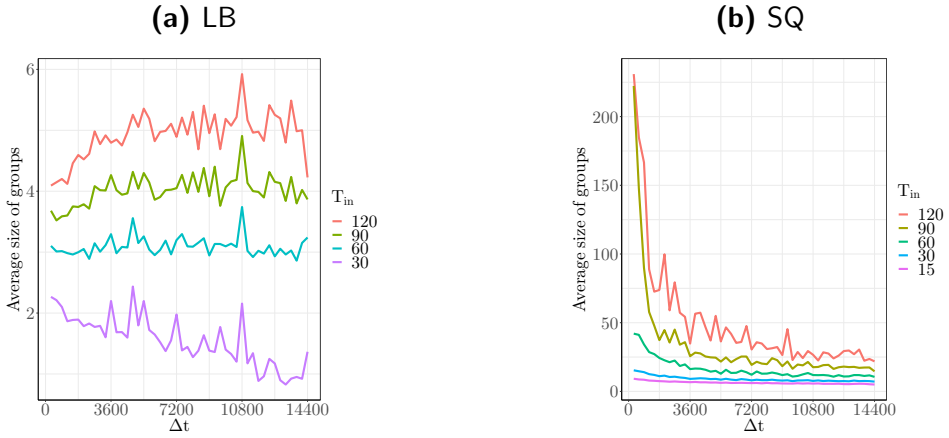


Figure 3.4: Average size of groups as a function of Δt and T_{in}. $\Delta t = \Delta t_1 = \Delta t_2$.

lag networks as a function of Δt and the window calibration length T_{in} in our datasets is necessary, as it indeed reveals fundamental differences between retail and institutional clients (at least in our datasets), which in turn will help to understand the results with two different timescales.

Fig. 3.2 plots the number of groups averaged over all calibration windows as a function of Δt for all T_{in}, for both LB and SQ. The number of groups found by the LL-SVNs and InfoMap is a measure of the statistically validated diversity of behaviour and of the potential richness of connectivity. For example, only a few groups of LB clients for $T_{\text{in}} = 30$ and large Δt are detected, while the largest value of $T_{\text{in}} = 120$ yields the most groups for LB. One also sees a sudden drop of the number of groups for $\Delta t = 14400s = 4h$, which is likely a by-product of the fact that we keep 8 hours of trading each day.

The number of groups of SQ retail clients behaves in the exactly opposite way unless T_{in} is small: the smaller T_{in}, the larger the number of groups. The

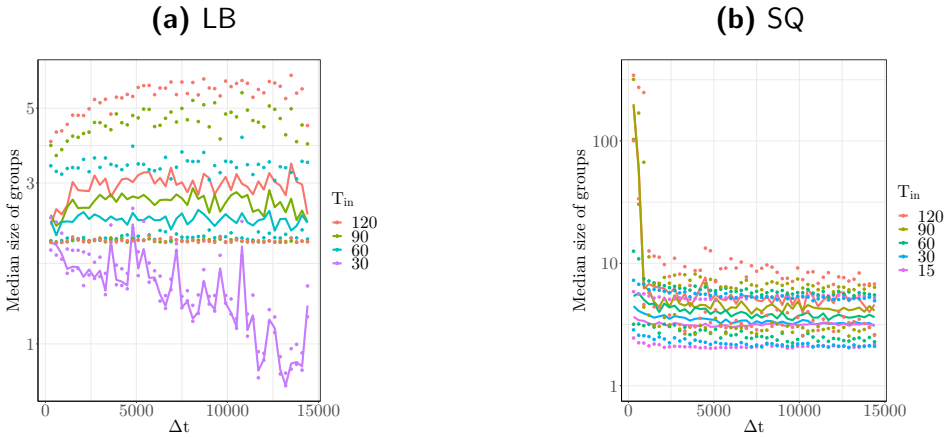


Figure 3.5: Average median size of groups as a function of Δt and T_{in} (points represent the first quartile and third quartile). $\Delta t = \Delta t_1 = \Delta t_2$.

case $T_{\text{in}} = 15$ for SQ shows that the effective number of points, proportional to $T_{\text{in}}/\Delta t$, must be large enough for the method to be powerful enough.

The group size distribution is very skewed: Fig. 3.5 plots the median size of the groups as a function of Δt, which is much smaller that the average group size. In fact, one often sees the emergence of a very large group for small Δt, while other groups are typically very small. We will thus focus on $T_{\text{in}} = 120$ for LB and $T_{\text{in}} = 30$ for SQ[4].

The *raison d'être* of calibration in sliding windows is *a priori* the non-stationarity not only of the population of traders, but also of their behaviour. If both are roughly stationary, a longer T_{in}, at fixed Δt, should give more precise and richer results, and inversely. This is likely a major cause of the difference between SQ and LB traders, the latter behaving in a much more stationary manner.

Let us now turn to the links themselves. Since we deal with lead-lag networks, they are directed. Links can be of two types: either from one group to another one, or to the same group, which we call a self-referential link. Occasionally, some groups only link to themselves, which would happen if they use an effective strategy whose activity does not systematically lead another one, but whose activity, on average, occurs at a scale comparable to Δt.

Fig. 3.7 plots the average total number of lead-lag links, and distinguishes within these links the average number of self-referential and 'only' self-referential lead-lag links. The lead-lag networks of the two types of traders are clearly different: the typical fraction of groups with only self links is small for LB traders, but much larger for SQ traders. The timeslice length Δt influences the number of non-self-referential links for both populations: their number decreases sharply

[4] Additional figures, for other values of T_{in}, are provided in Appendix 3.6

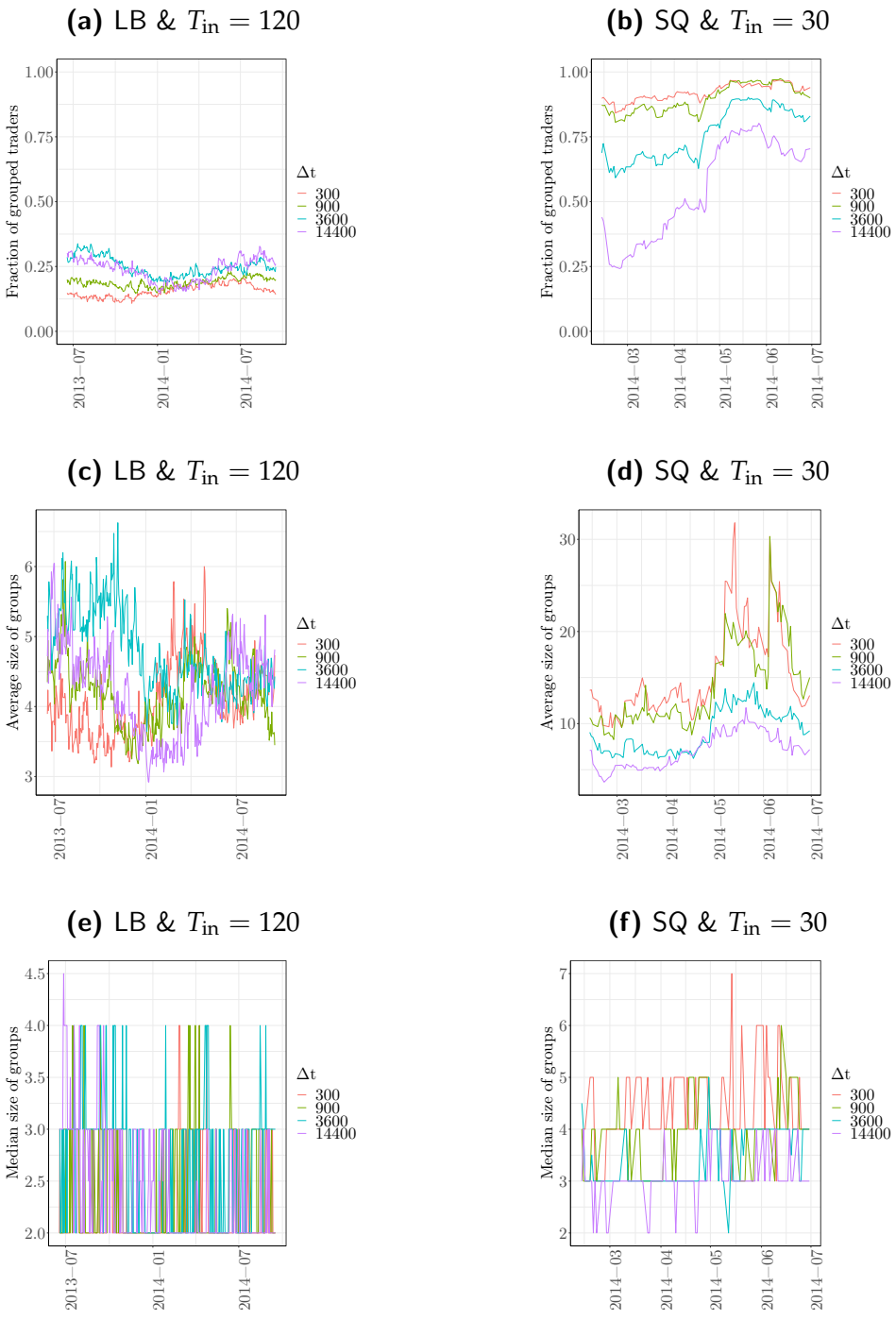


Figure 3.6: Fraction of traders grouped by SVNs, average size of groups and median size of groups as a function of time for LB and SQ. $\Delta t = \Delta t_1 = \Delta t_2$.

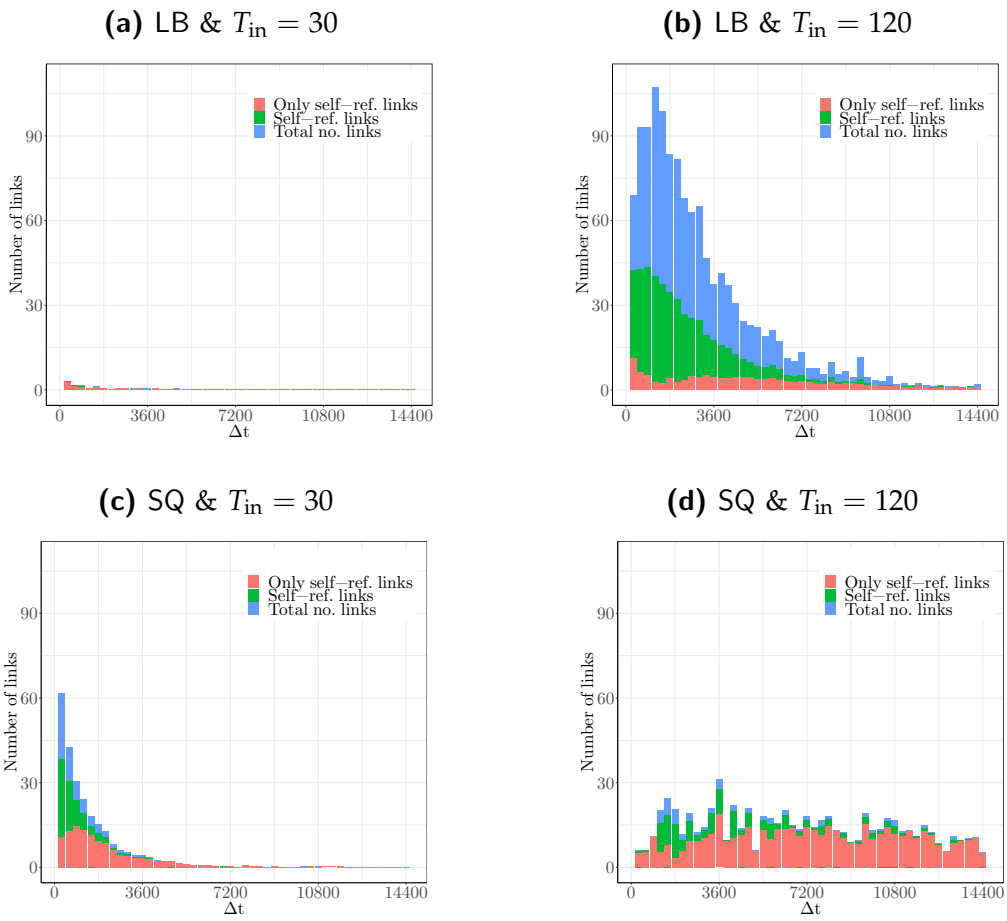


Figure 3.7: Total number of links as a function of the timeslice duration (in seconds) $\Delta t = \Delta t_1 = \Delta t_2$, for groups with only self-referential links, self-referential links and links to other groups and only links to other groups.

when $\Delta t > 1$ hour and are negligible at resolutions coarser than 2 hours for SQ and 3 hours for LB.

3.4.2 $\Delta t_1 \neq \Delta t_2$

Links

When $\Delta t_1 \neq \Delta t_2$, both timescales may influence each other in an asymmetric way. Our strategy is to capture such an asymmetry by using several quantities related to both the directed network structure and the rate of trading. Each quantity is estimated for each pair $(\Delta t_1, \Delta t_2)$, each of them ranging from 5 minutes to 4 hours (14440 seconds) by steps of 5 minutes, which gives 1176 unique pairs. Since we measure these quantities over many calibration windows, we obtain a time series for each quantity and for each pair.

Let us first start with the number of links. The left hand side plots of Figs. 3.8 and 3.9 show the average number of links for each pair of timescales. Let us clarify the convention: Δt_1 (on the x-axis) leads on Δt_2 (on the y-axis): as a consequence, points above the $y = x$ line correspond to smaller timescales leading on longer timescales, and inversely. It is useful to keep in mind that on the diagonal $\Delta t_1 = \Delta t_2$ (Fig. 3.7) the number of links is maximal for small values of Δt for both LB and SQ.

The three methods give qualitatively similar results, although it is more difficult for Method III to detect links for timescales very far from the diagonal for LB. In accordance with Fig. 3.7, there are more links for smaller values of Δt_1 and Δt_2 around the diagonal. In addition, one generally finds that the number of links has a local maximum on the diagonal. There are also more links for some particular values of either Δt_1 or Δt_2, e.g. multiples of full hours. This may indicate that some traders have a typical activity change over 1 hour, e.g. a trading strategy that depends on the time of the day, or that they trade between, say, 9:00 and 10:00, 10:00 and 11:00, and so on.

At least for LB, it is obvious that there are more links above than below the diagonal, which implies that there are on average more links from shorter timescales to longer timescales. The statistical significance of this difference is assessed in the following way: let us denote the number of links of the pair $(\Delta t_1, \Delta t_2)$, the first timescale of the pair leading on the second one, in calibration window i by $W_i(\Delta t_1, \Delta t_2)$. One then applies a t-statistics to the time series of the difference $\delta W_i(\Delta t_1, \Delta t_2) = W_i(\Delta t_1, \Delta t_2) - W_i(\Delta t_2, \Delta t_1)$, In order to avoid too many false positives, we use a false discovery rate (FDR) correction for multiple

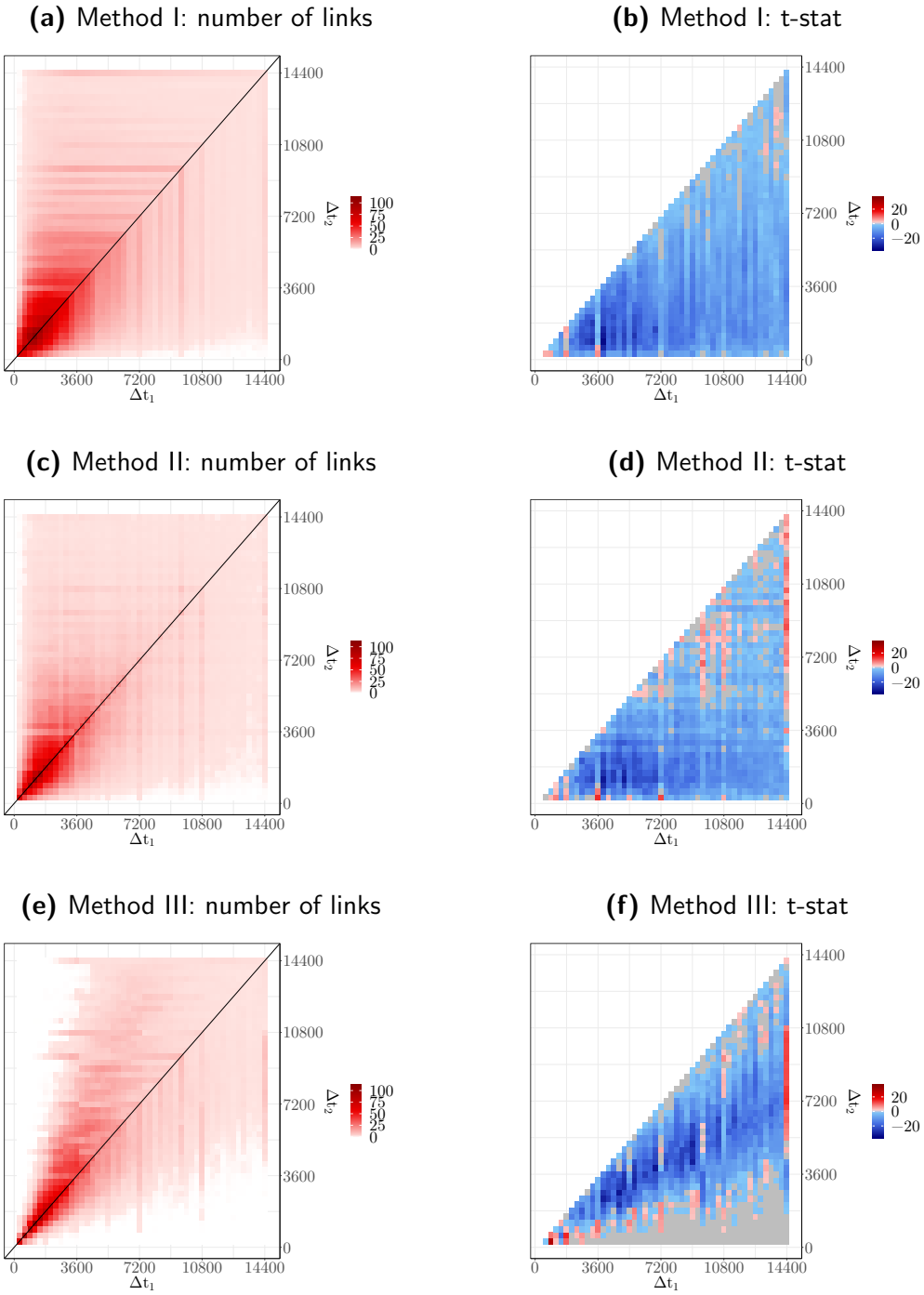


Figure 3.8: Left hand-side plots: average number of lead-lag links for LB (Δt_1 leads on Δt_2). Right hand-side plots: t-statistics of the difference between the number of links of the pairs $(\Delta t_1, \Delta t_2)$ and $(\Delta t_2, \Delta t_1)$; negative values indicate that shorter timescales link significantly more to longer timescales. $T_{\text{in}} = 120$

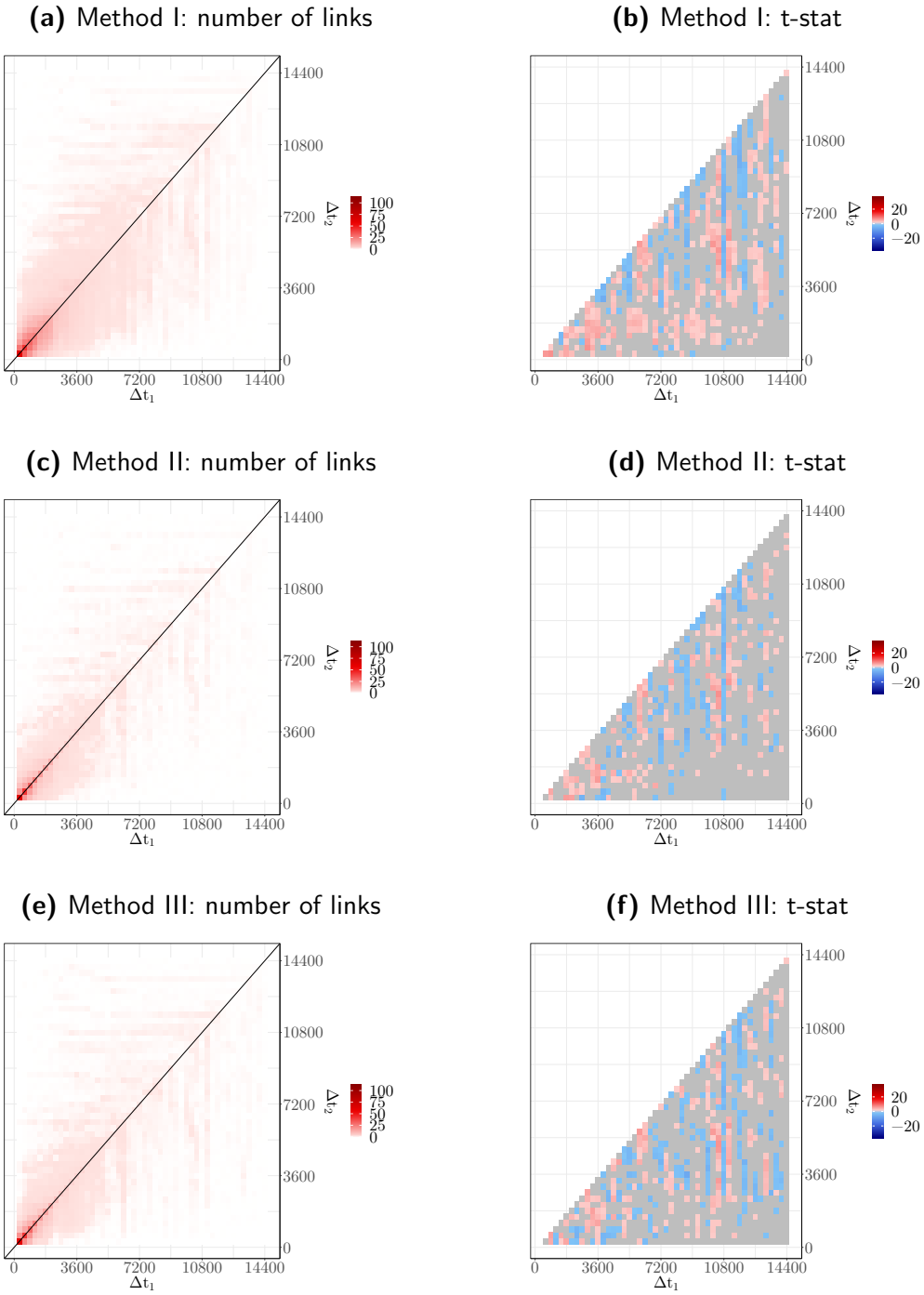


Figure 3.9: Left hand-side plots: average number of lead-lag links for SQ (Δt_1 leads on Δt_2). Right hand-side plots: t-statistics of the difference between the number of links of the pairs $(\Delta t_1, \Delta t_2)$ and $(\Delta t_2, \Delta t_1)$; negative values indicate that shorter timescales link significantly more to longer timescales. $T_{\text{in}} = 30$.

hypotheses made in this plot, setting the rate at 0.2. Right columns of Figs 3.8 and 3.9 plots the selected t-stats of $\delta W_i(\Delta t_1, \Delta t_2$: blue zones correspond to lead-lag links from shorter to longer timescales, and reversely for red zones.

The plots for LB are overwhelmingly blue: there are more links from short timescales to long timescales, for the three methods. There is a clear exception for $\Delta t_1 = 4h$, which once again is probably a by-product of keeping exactly 8 hours of data each day. One notes however small red regions when two groups of traders are used (Methods II and III): at around (3h, 2h) for Method II and III, for relatively small values of the lagging timescales for Method III, and (1h, 5m) and (2h, 5m) for Method II.

For the SQ traders, the link structure is much more complex. Focusing on the common results between the three methods, one finds a zone where longer timescales have more links to shorter timescales when $\Delta t_1 <$ 1h, and also around (3h, 1.5h). One also notes an alternance of positive and negative vertical stripes.

The number of links themselves are not sufficient to characterize the lead-lag between timescales for traders. For example, how a given group links to other ones may also be surprising. Indeed, it is quite possible that a group has more than one link to another group, even for the same initial state. For example, group 1 may have links $+1 \to +1$ and $+1 \to -1$ with group 2. This happens quite often but is not as strange as it may appear at first: such dual links mean, in this case, that the mostly buying activity of group 1 triggers either $+1$ or -1 in group 2. In other words, it triggers a directional activity of group 2, whose sign is undetermined. In a prediction setting, dual links of course reduce the prediction power, but as long as enough single links do exist, order flow prediction is possible, as shown by Challet et al. (2018).

Activity

Being able to account for two timescales makes it possible to connect Time Reversal Asymmetry (TRA) at the level of trader behaviour to that of the price itself. TRA of prices, while being totally intuitive in financial markets, is not totally trivial to measure owing to the amount of noise in financial data. Zumbach and Lynch (2001) proposed to measure the asymmetry between historical volatility measured over Δt_h in the past and realized volatily, estimated over Δt_r. More precisely, for a given t, one estimates the historical volatility $v_h(t)$ in the interval $]t - \Delta t_h, t]$ and the realized volatility $v_r(t)$ in the interval $[t, t + \Delta t_r[$; then one estimates the correlation of v_h and v_r for all chosen ts, denoted by $\rho(\Delta t_h, \Delta t_r)$. This results in volatility correlation mugshots in which one clearly sees the asym-

metry of $\rho_v(\Delta t_h, \Delta t_r)$ with respect to the diagonal $\Delta t_r = \Delta t_h$. Zumbach (2009) investigates further the TRA of volatility and proposes two more measures of TRA by noticing that the price returns in the time intervals over which volatility is estimated can be defined according to their own timescale, whose fine structure is investigated, e.g. in Chicheportiche and Bouchaud (2014).

Connecting agent activity and volatility is natural if time subordination holds (Clark, 1973). In other words, if the volatility per trade is locally constant, then the volatility in a time interval depends on the number of trades occurring in that period of time assuming that prices are diffusive. While this neglects jumps of various origins, e.g. microstructural noise due to heavy-tailed distribution gaps in limit order books (Gillemot et al., 2006), we only need to assume that there is a monotonic average relation between the number of trades and volatility to connect trader activity and volatility.

Therefore, we can estimate the correlation between the activity rate of traders in leading groups and lagging groups, determined at two different timescales, as above. Let us therefore denote the total number of trades of agents in group g during timeslice $(t, \Delta t)$ by $N^{(g)}(t, \Delta t)$; in addition, we denote by

$$N_1(t) = \sum_{g \in G_1} N^{(g)}(t, -\Delta t_1)$$

the total activity of the leading groups at time t, and similarly

$$N_2(t) = \sum_{g' \in G_2} N^{(g')}(t, \Delta t_2)$$

the total activity of the agents in the lagging groups (note that with Method I, $G_1 = G_2$). We then can compute the correlation between activity rates $N_1(t)/\Delta t_1$ and $N_2(t)/\Delta t_2$, denoted by $\rho(\Delta t_1, \Delta t_2)$.

Figs. 3.10 and 3.11 plot $\rho(\Delta t_1, \Delta t_2$ for the LB and SQ datasets respectively (left-hand side plots), and correspond to the mugshots of Zumbach and Lynch (2001) but for activity rates. In the case of LB, the asymmetry is clear and is confirmed by the right-hand side plots which report the t-statistics of $\delta\rho(\Delta t_1, \Delta t_2) = \rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; only the values validated by FDR are in color, the unvalidated ones being reported in gray. As the three methods point to the same conclusion: activity on shorter timescales in the past is more correlated with future activity on longer timescales than the opposite (blue zone), this globally mirrors the dependence between the number of links and the correlation. Note that this is an anti-Zumbach effet. Methods I and II however suggest a subtler

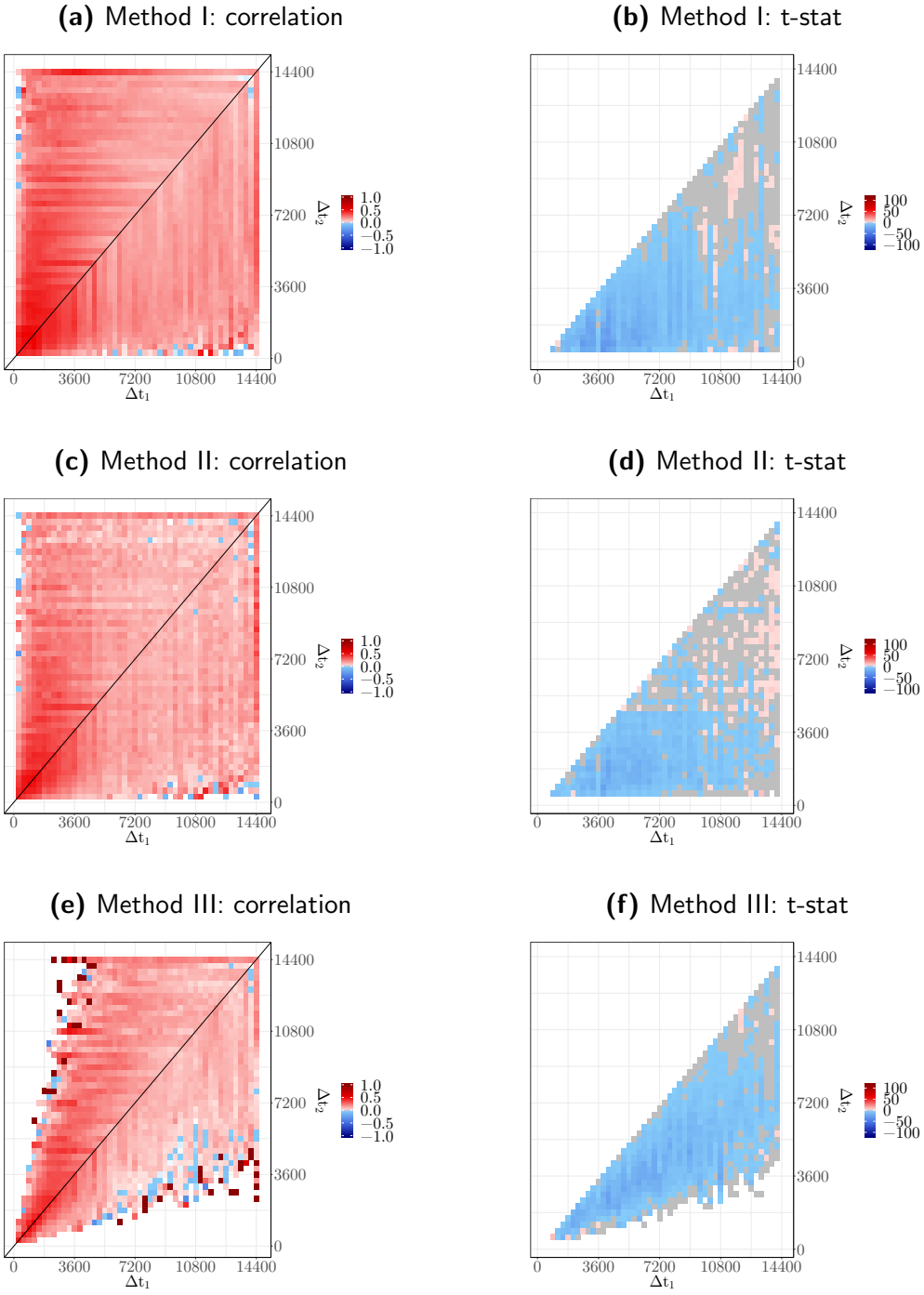


Figure 3.10: Left hand-side plots: average correlation between the leading (at timescale Δt_1) and lagging (at timescale Δt_2) activity rates, $\rho(\Delta t_1, \Delta t_2)$. Right hand-side plots: t-statistics of the difference $\rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; negative value correspond to activity at small timescales being more correlated to future activity at larger timescale than reversely. LB dataset.

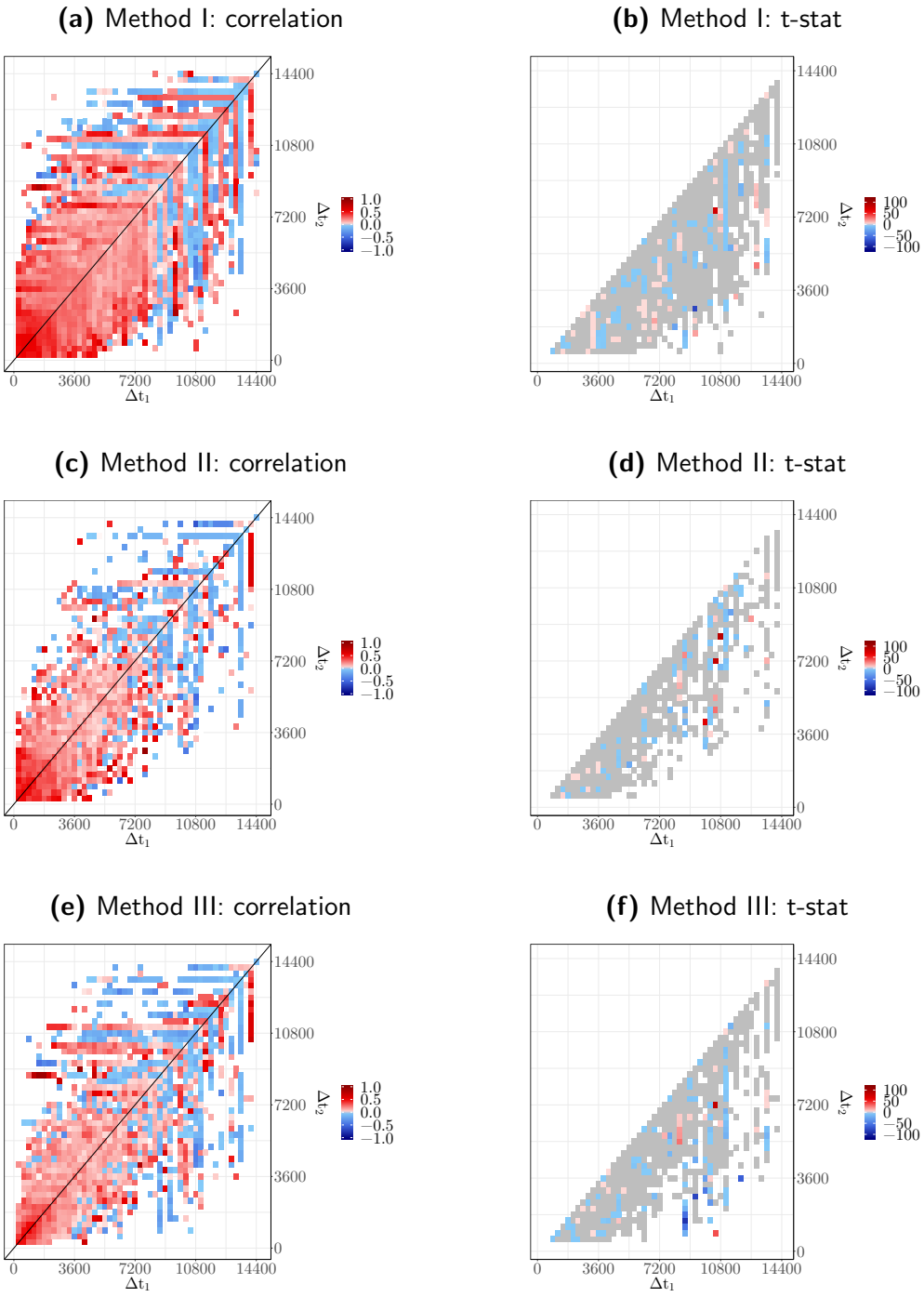


Figure 3.11: Left hand-side plots: average correlation between the leading (at timescale Δt_1) and lagging (at timescale Δt_2) activity rates, $\rho(\Delta t_1, \Delta t_2)$. Right hand-side plots: t-statistics of the difference $\rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; negative values correspond to activity at small timescales being more correlated to future activity at larger timescale than reversely. SQ dataset.

picture: Zumbach effect emerges (red zone) when $\Delta t_2 > 2$hours. Our dataset is not sufficiently long to report results for Δt_1 or $\Delta t_2 > 4$hours. The SQ dataset only leads to statistically validated TRA in a few vertical stripes.

Our results on trader activity TRA suggest a quite more complex picture of the interaction between timescales than that of volatility TRA, as reported in Zumbach and Lynch (2001) and Zumbach (2009). Volatility TRA in FX markets is the result of the interaction between all the categories of traders from all brokers-dealers, and seems to be always in the same direction, from large to small timescales (the red areas in Figs. 3.10 and 3.11). At a more detailed level, which requires trader-resolved data, and which we here call market nanostructure, the sign of TRA is not unique and is trader and timescale dependent. We also note that in the kernel estimates of Chicheportiche and Bouchaud (2014) on daily equity data, the sign of influence of some timescales may have either sign, while fitting ZHawkes processes to intraday equity data yields only coefficients of the same sign (Blanc et al., 2017).

3.5 Conclusions

Extending the SVNs to two timescales opens up the possibility of accessing a much finer picture of the causal structure of activity in complex systems. Beyond the structure of the inferred networks, it also allows one to quantify the time reversal asymmetry of the system, which leads to some surprises, and which underlines the fundamental greater phenomenological richness brought by agent-resolved data.

First, we found that the causal structure of activity of institutional traders is best investigated over longer periods in comparison to retail traders, which points to a greater stability of institutional traders in their behaviour. Then, the direction of the asymmetric influence of timescale on activity depends on the timescales and may have a sign opposite to that of volatility TRA.

Since the latter is measured on anonymous data that encompasses a much greater variety of traders and broker-dealers than our datasets do, we can only speculate at the moment that a more complete dataset, e.g. for equities, will reveal how these various trader activity TRAs, once combined, yield that of the volatility.

3.6 Appendix: Additional figures

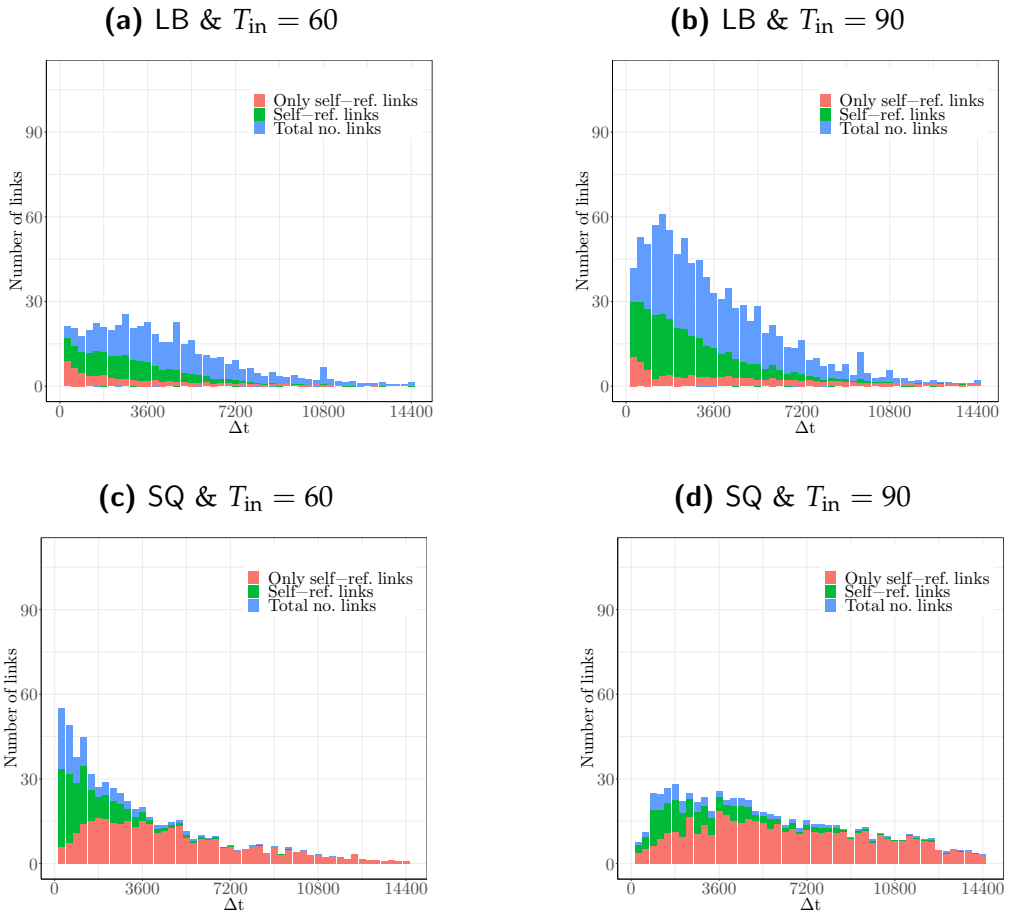


Figure 3.12: Total number of links as a function of the timeslice duration (in seconds) $\Delta t = \Delta t_1 = \Delta t_2$, for groups with only self-referential links, self-referential links and links to other groups and only links to other groups.

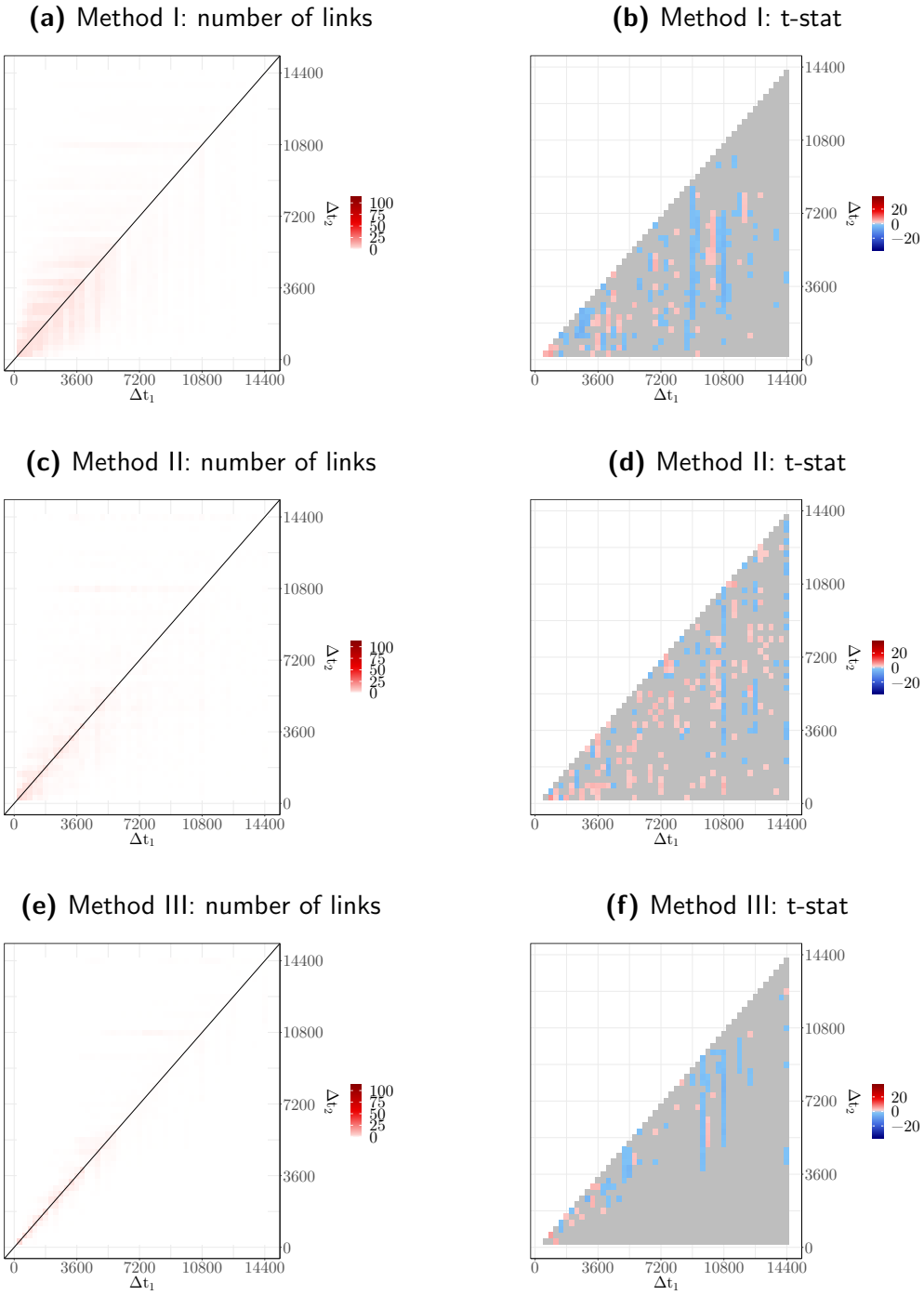


Figure 3.13: Left hand-side plots: average number of lead-lag links for LB (Δt_1 leads on Δt_2). Right hand-side plots: t-statistics of the difference between the number of links of the pairs $(\Delta t_1, \Delta t_2)$ and $(\Delta t_2, \Delta t_1)$; negative values indicate that shorter timescales link significantly more to longer timescales. $T_{\text{in}} = 30$

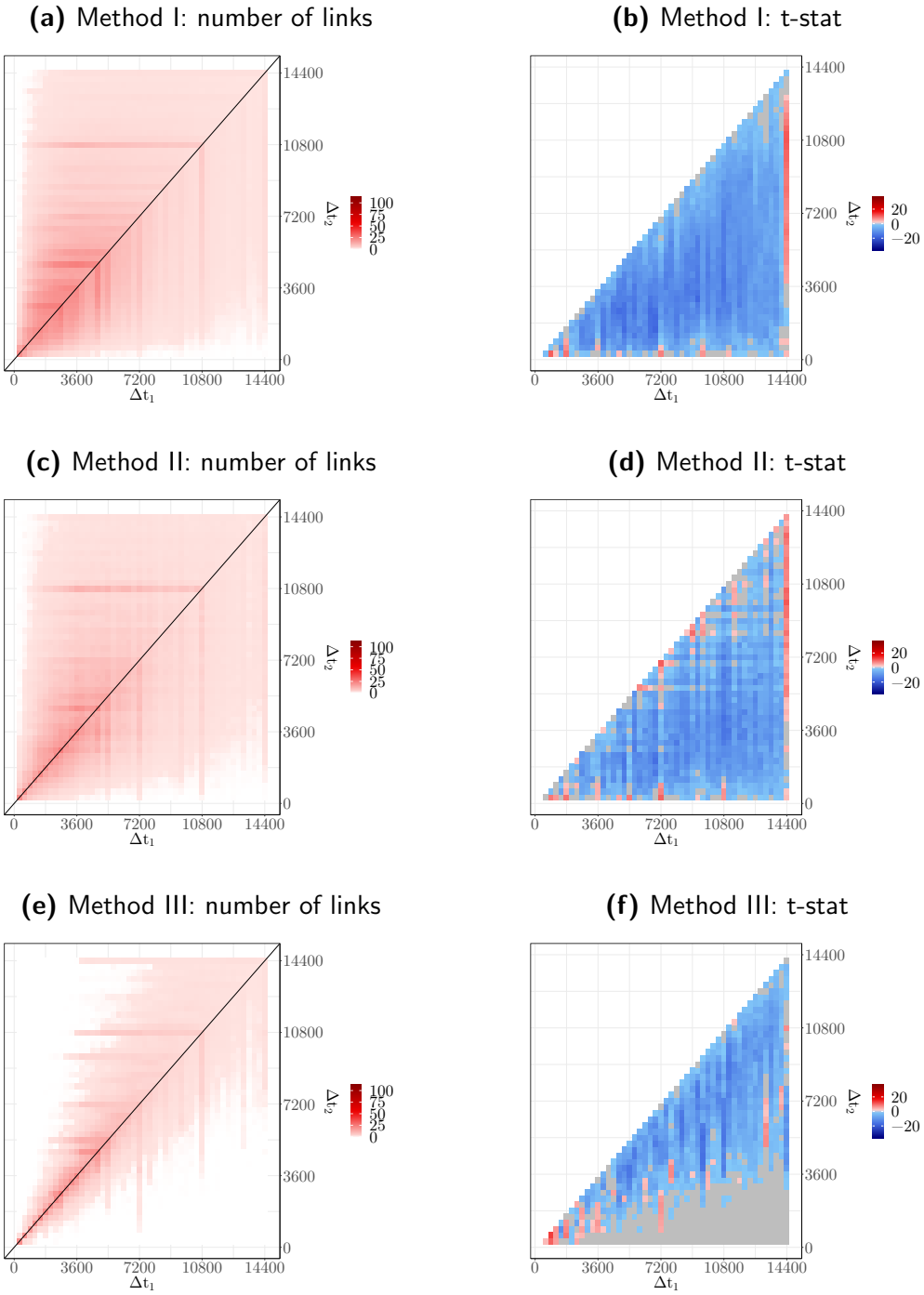


Figure 3.14: Left hand-side plots: average number of lead-lag links for LB (Δt_1 leads on Δt_2). Right hand-side plots: t-statistics of the difference between the number of links of the pairs $(\Delta t_1, \Delta t_2)$ and $(\Delta t_2, \Delta t_1)$; negative values indicate that shorter timescales link significantly more to longer timescales. $T_{\text{in}} = 60$

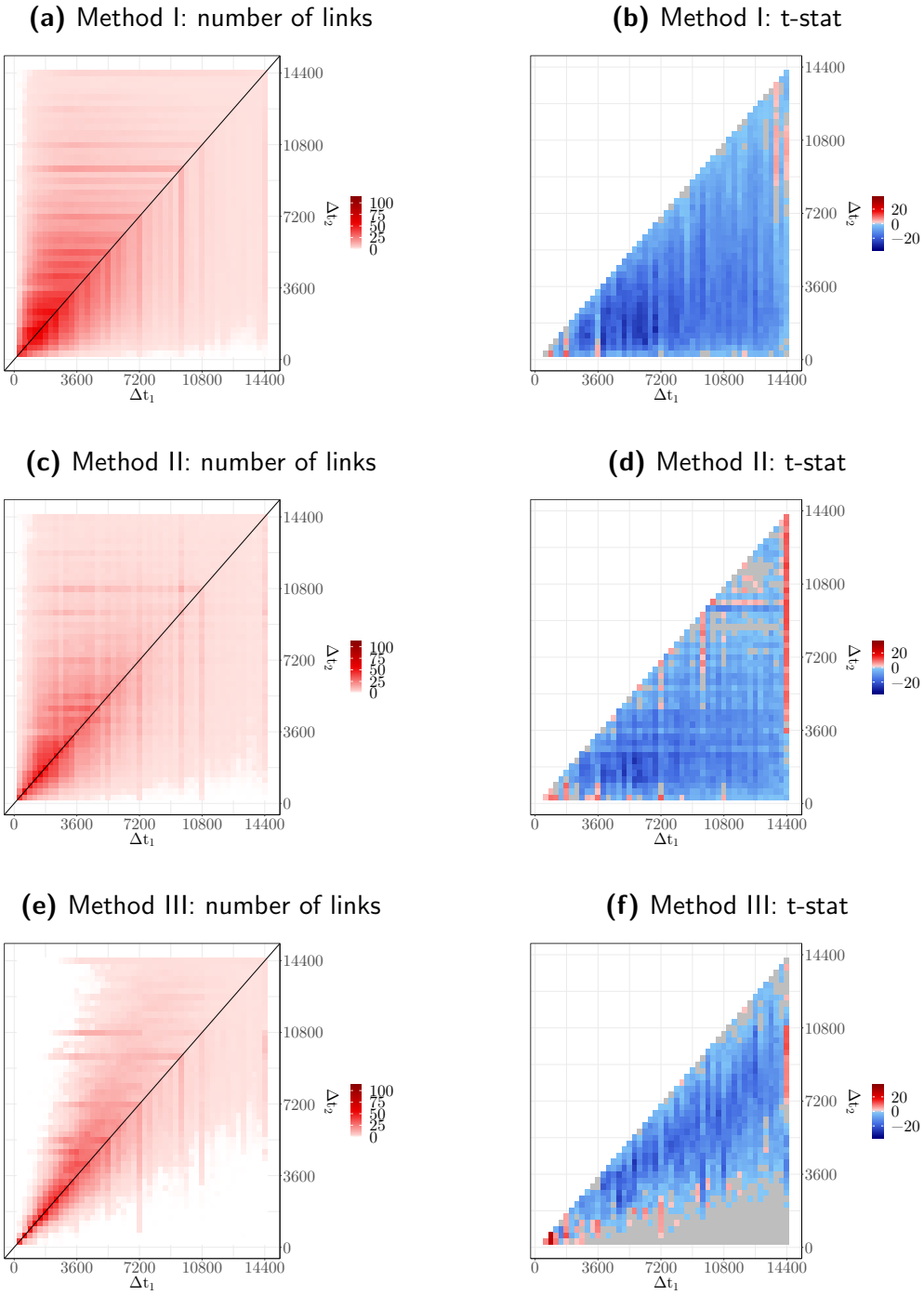


Figure 3.15: Left hand-side plots: average number of lead-lag links for LB (Δt_1 leads on Δt_2). Right hand-side plots: t-statistics of the difference between the number of links of the pairs $(\Delta t_1, \Delta t_2)$ and $(\Delta t_2, \Delta t_1)$; negative values indicate that shorter timescales link significantly more to longer timescales. $T_{\text{in}} = 90$

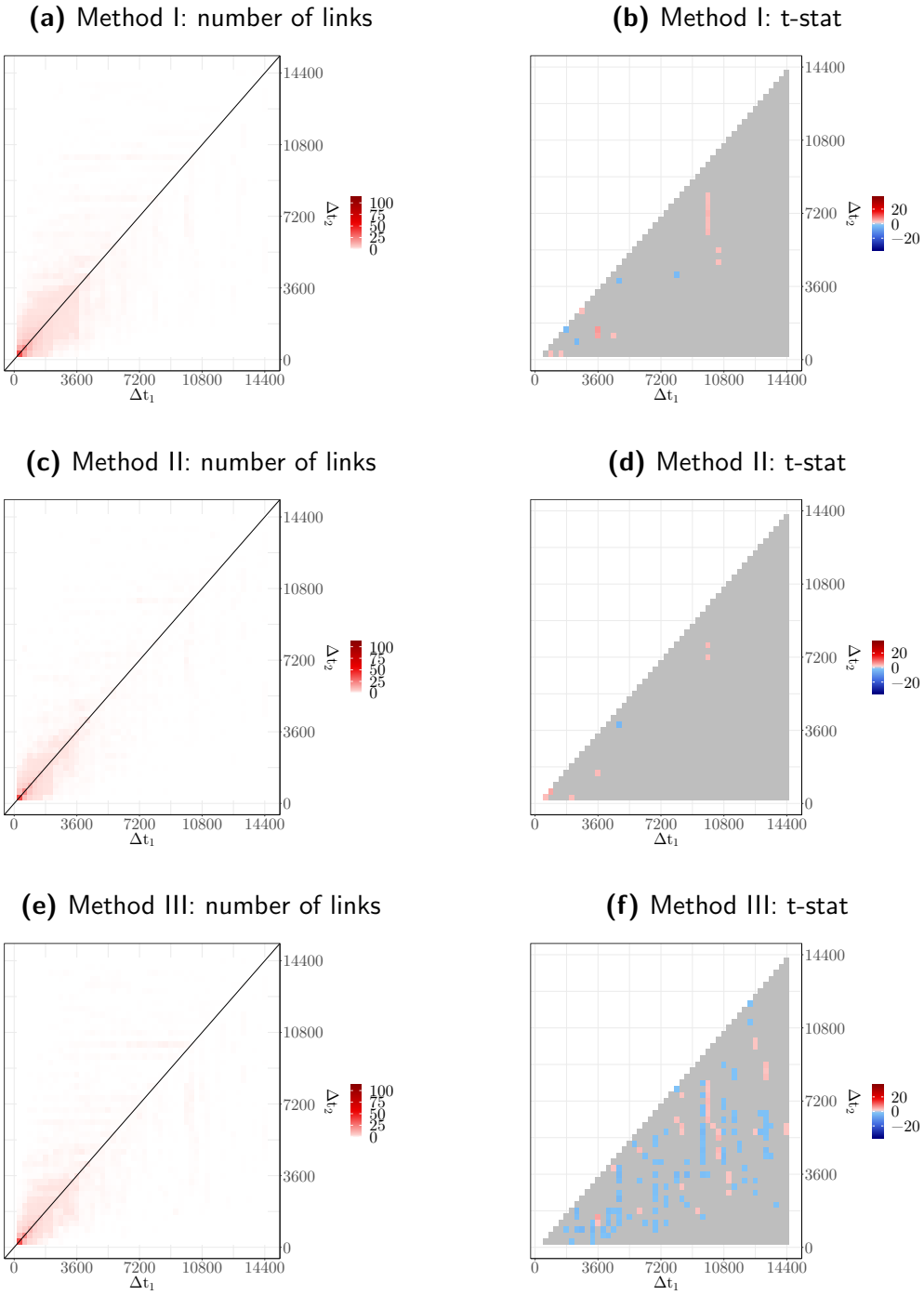


Figure 3.16: Left hand-side plots: average number of lead-lag links for SQ (Δt_1 leads on Δt_2). Right hand-side plots: t-statistics of the difference between the number of links of the pairs $(\Delta t_1, \Delta t_2)$ and $(\Delta t_2, \Delta t_1)$; negative values indicate that shorter timescales link significantly more to longer timescales. $T_{\text{in}} = 15$

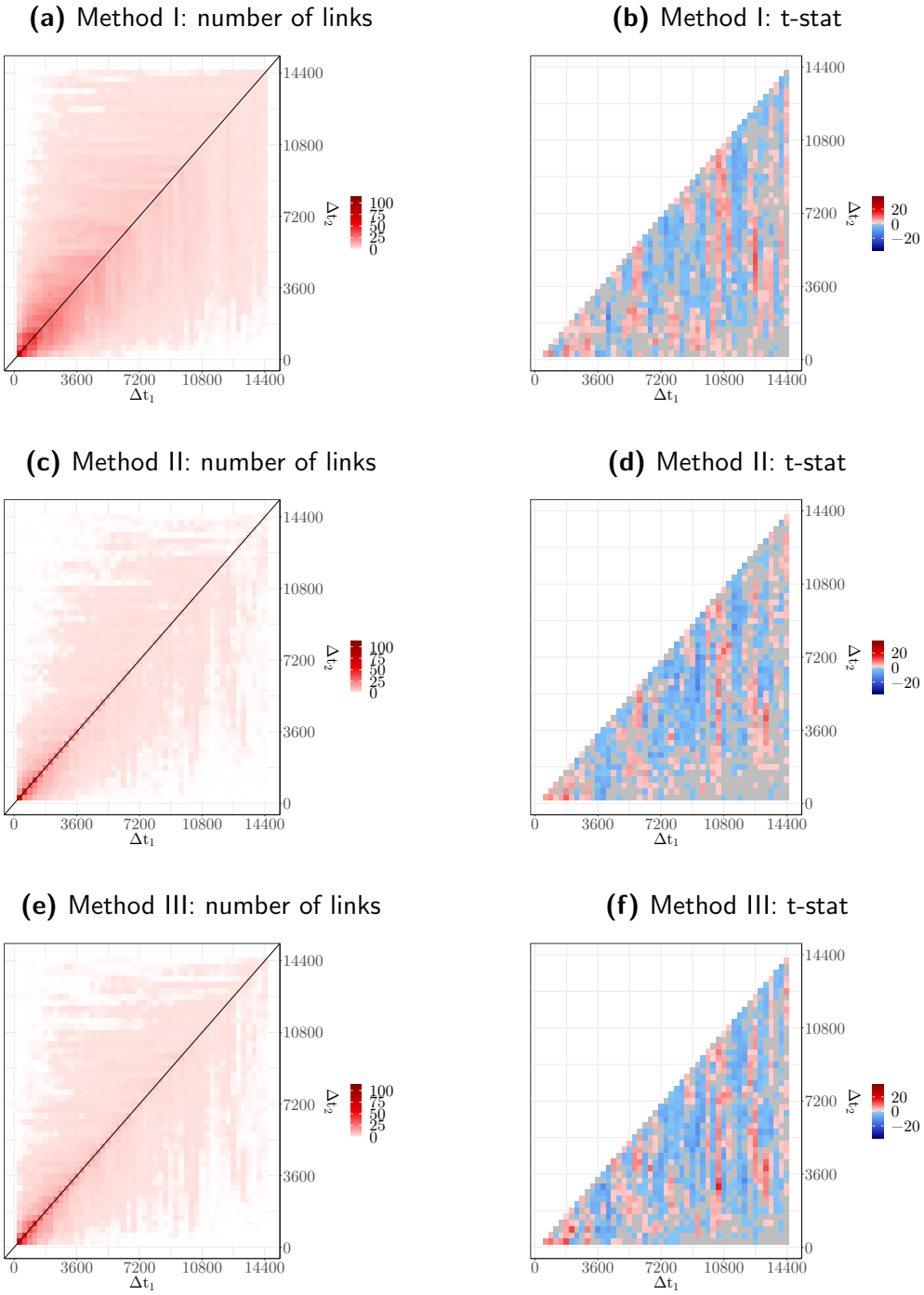


Figure 3.17: Left hand-side plots: average number of lead-lag links for SQ (Δt_1 leads on Δt_2). Right hand-side plots: t-statistics of the difference between the number of links of the pairs $(\Delta t_1, \Delta t_2)$ and $(\Delta t_2, \Delta t_1)$; negative values indicate that shorter timescales link significantly more to longer timescales. $T_{\text{in}} = 60$

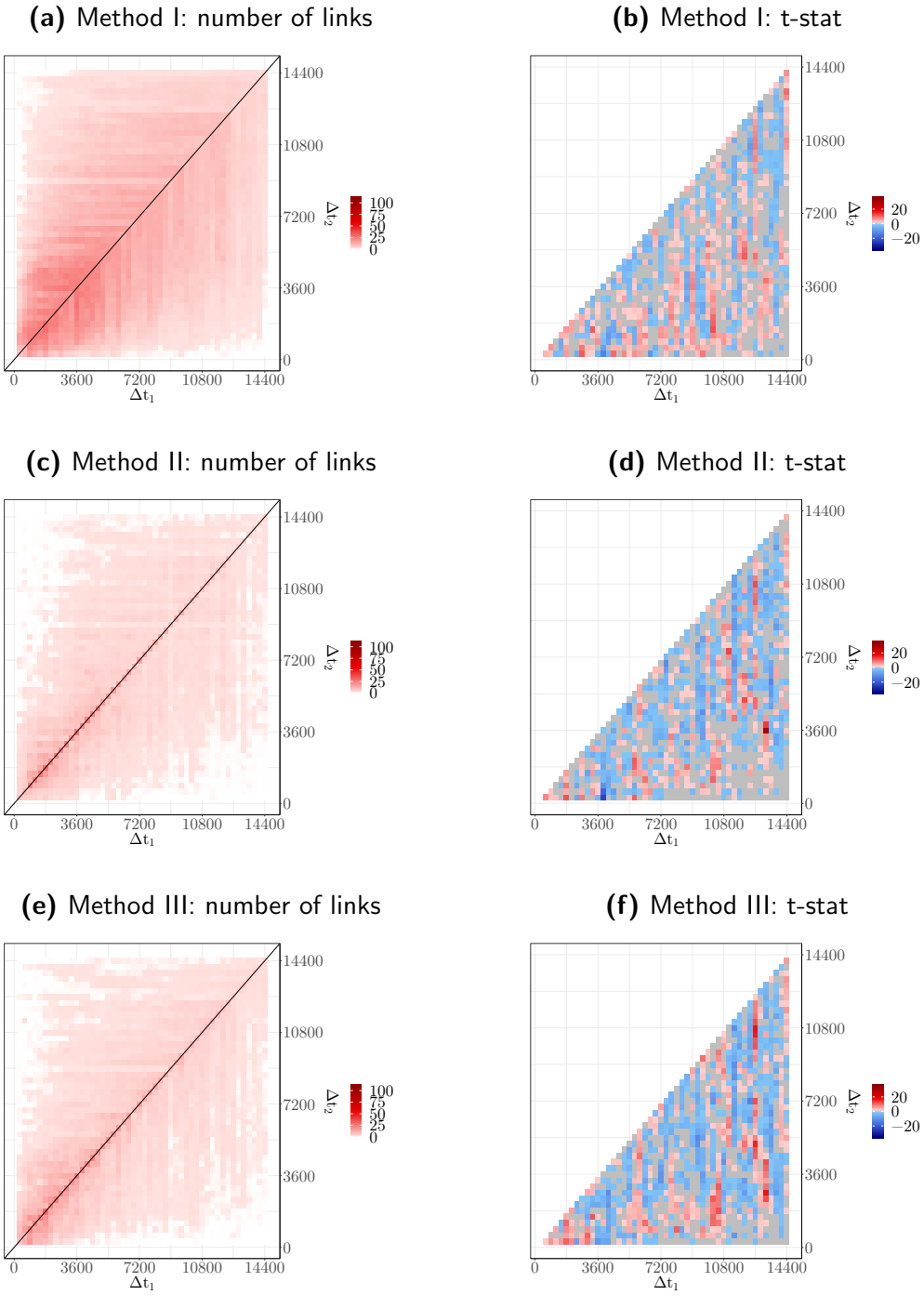


Figure 3.18: Left hand-side plots: average number of lead-lag links for SQ (Δt_1 leads on Δt_2). Right hand-side plots: t-statistics of the difference between the number of links of the pairs $(\Delta t_1, \Delta t_2)$ and $(\Delta t_2, \Delta t_1)$; negative values indicate that shorter timescales link significantly more to longer timescales. $T_{\text{in}} = 90$

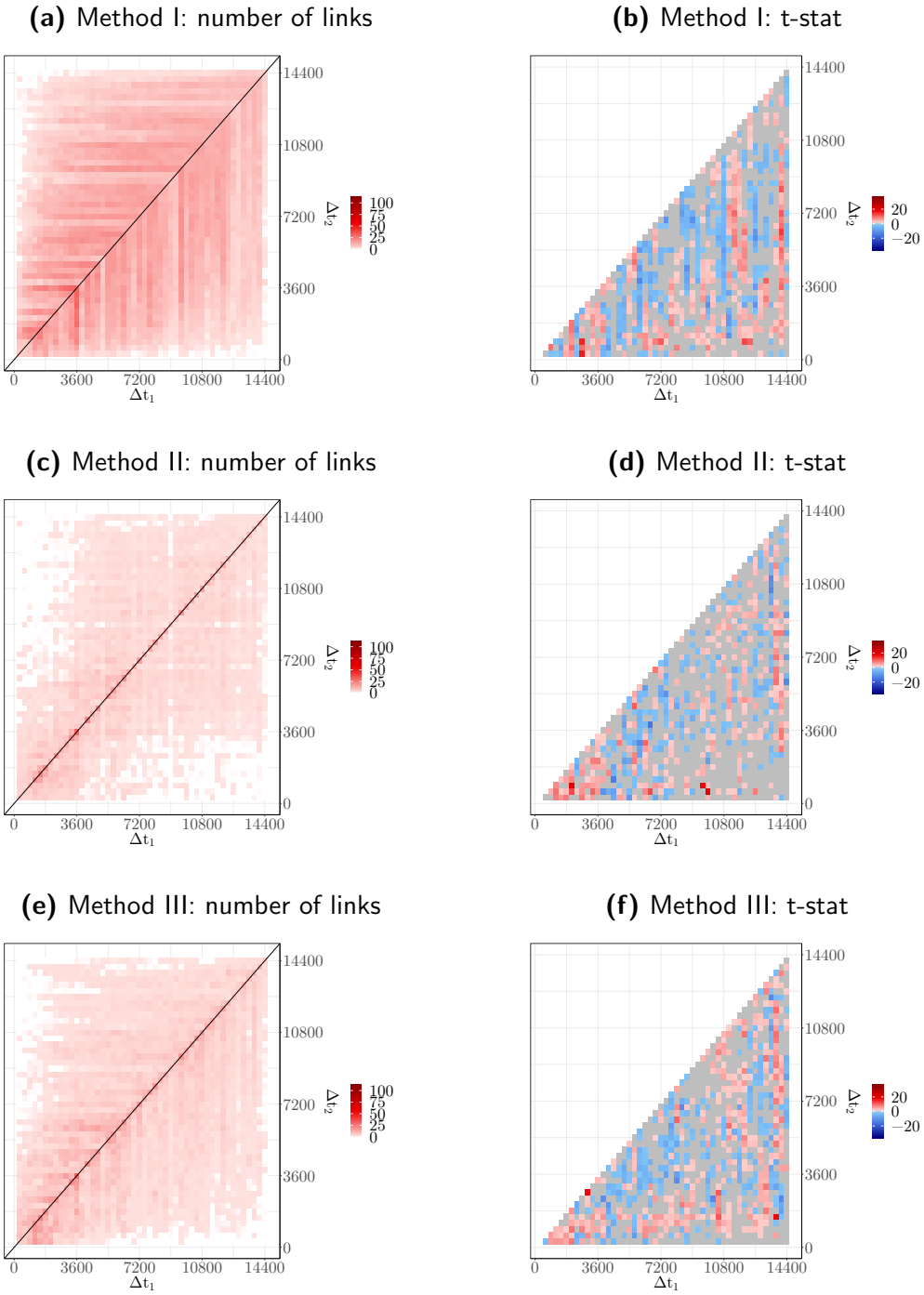


Figure 3.19: Left hand-side plots: average number of lead-lag links for SQ (Δt_1 leads on Δt_2). Right hand-side plots: t-statistics of the difference between the number of links of the pairs $(\Delta t_1, \Delta t_2)$ and $(\Delta t_2, \Delta t_1)$; negative values indicate that shorter timescales link significantly more to longer timescales. $T_{\text{in}} = 120$

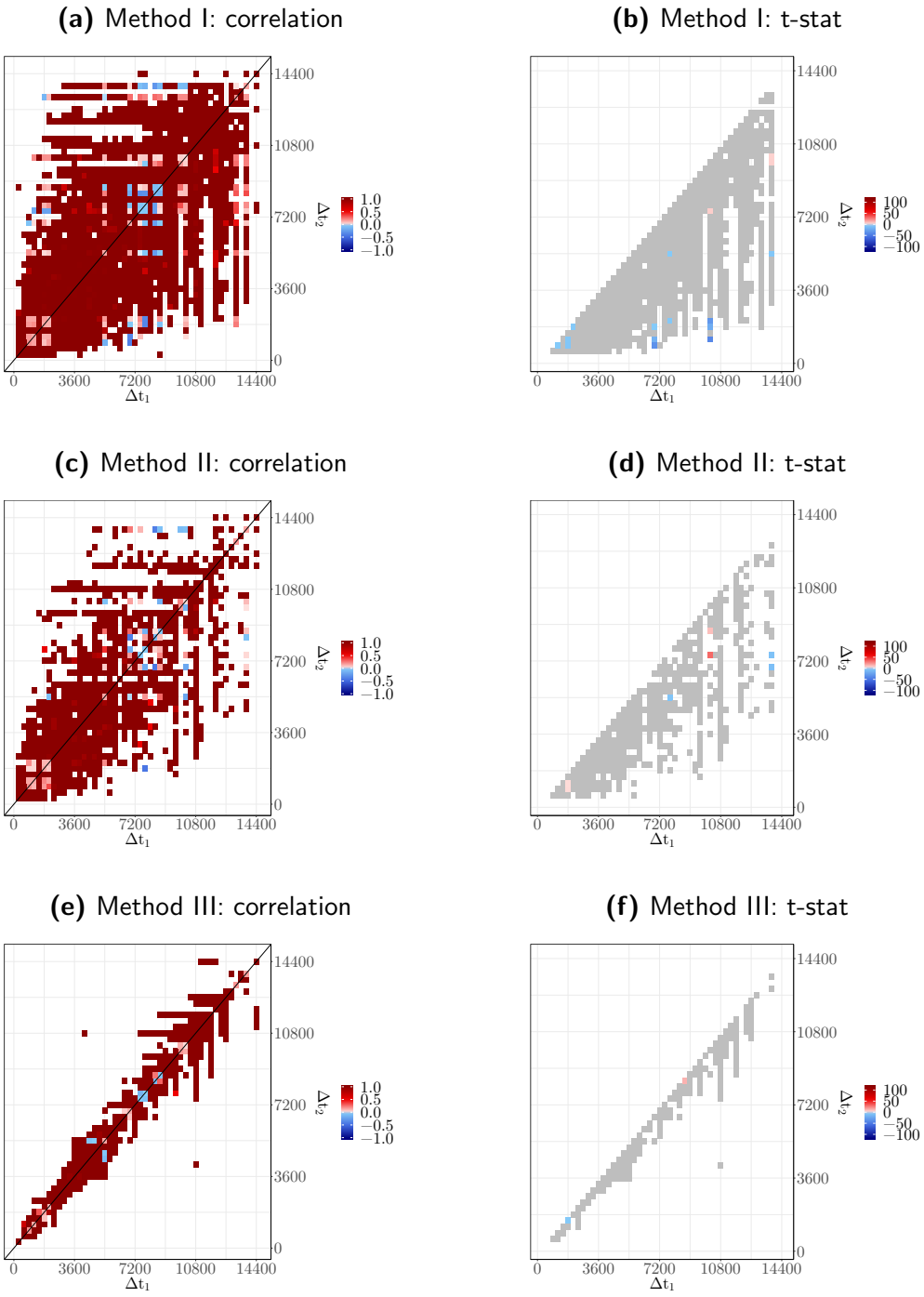


Figure 3.20: Left hand-side plots: average correlation between the leading (at timescale Δt_1) and lagging (at timescale Δt_2) activity rates, $\rho(\Delta t_1, \Delta t_2)$. Right hand-side plots: t-statistics of the difference $\rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; negative value correspond to activity at small timescales being more correlated to future activity at larger timescale than reversely. LB dataset, $T_{\text{in}} = 30$.

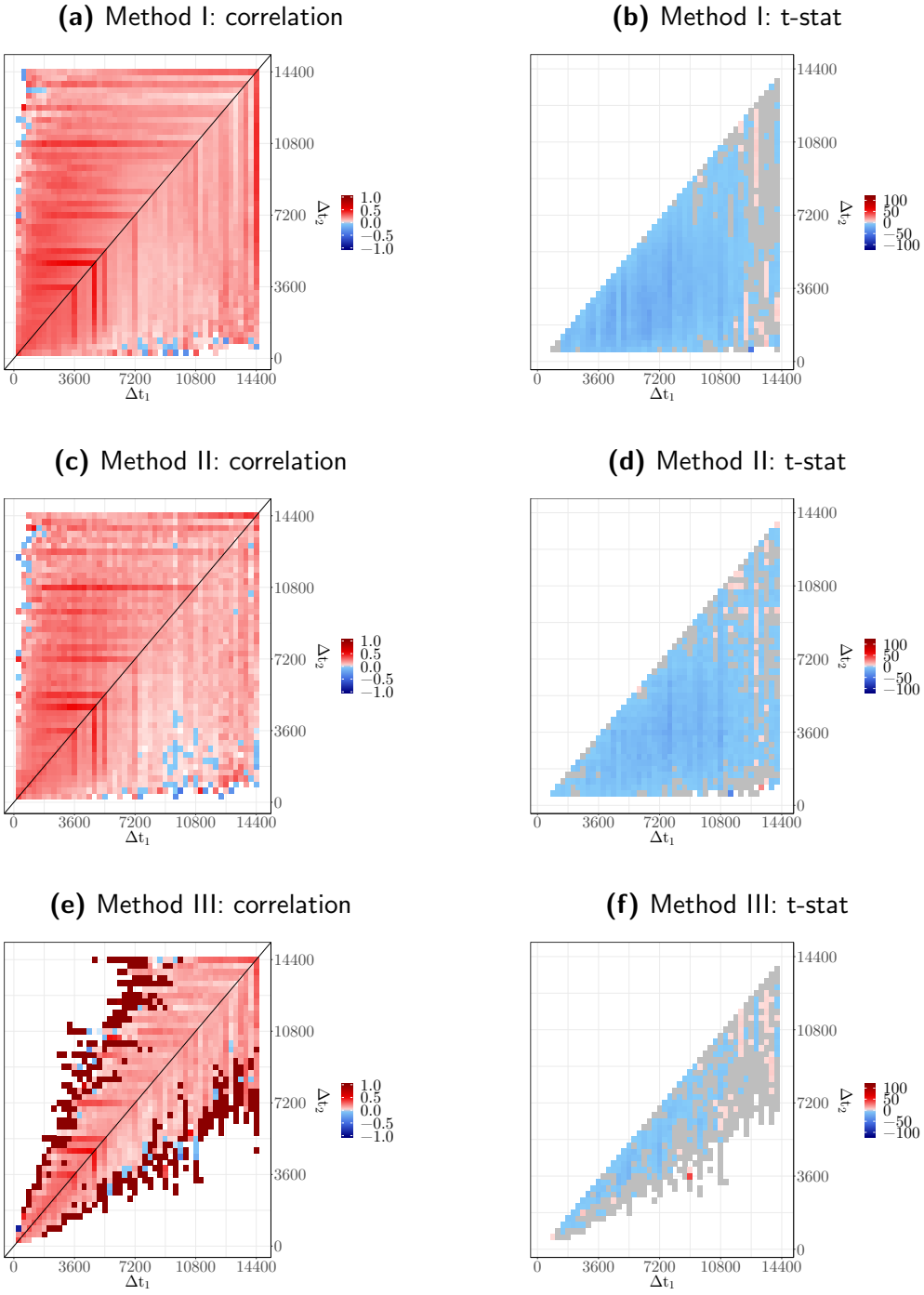


Figure 3.21: Left hand-side plots: average correlation between the leading (at timescale Δt_1) and lagging (at timescale Δt_2) activity rates, $\rho(\Delta t_1, \Delta t_2)$. Right hand-side plots: t-statistics of the difference $\rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; negative value correspond to activity at small timescales being more correlated to future activity at larger timescale than reversely. LB dataset, $T_{\text{in}} = 60$.

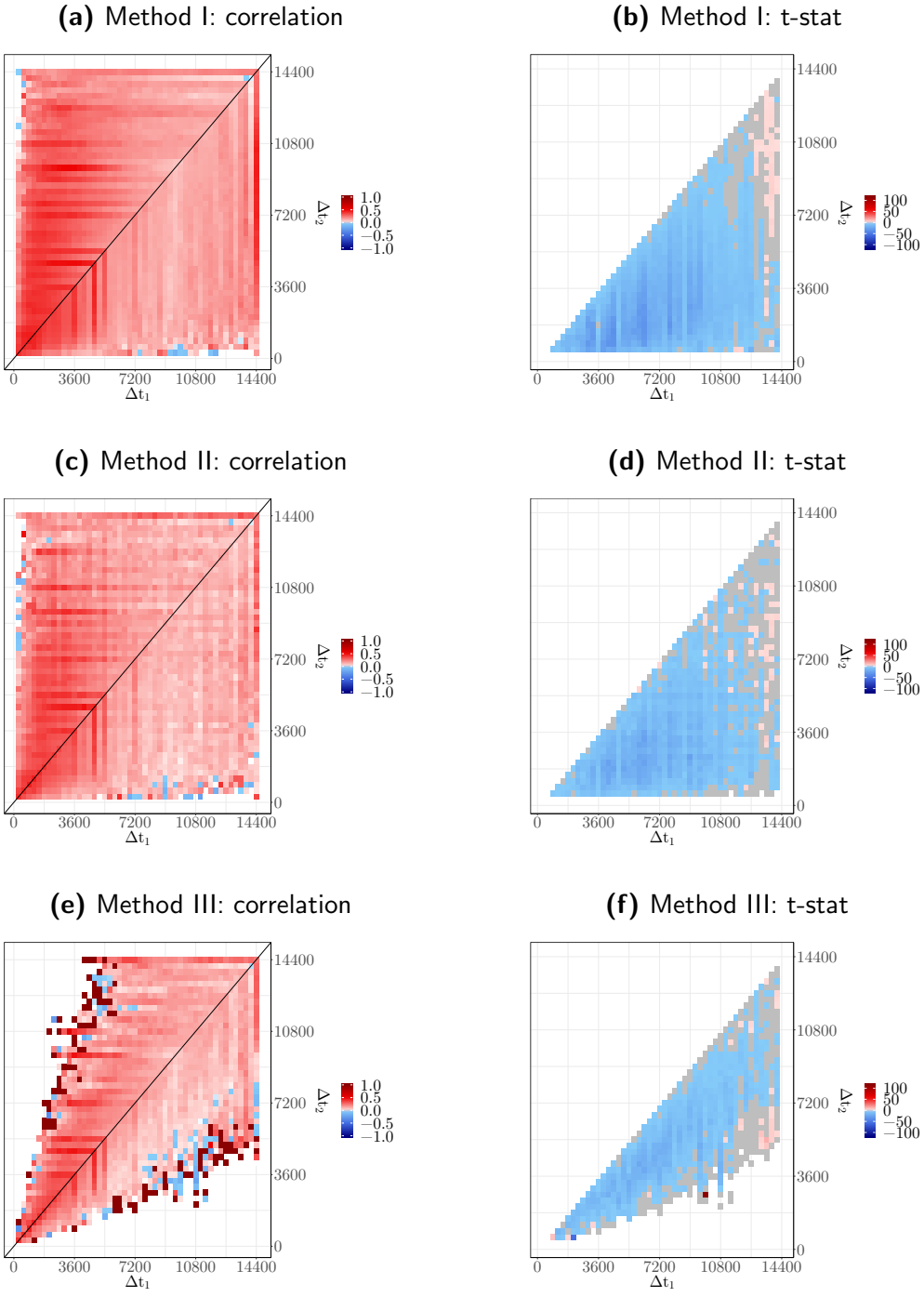


Figure 3.22: Left hand-side plots: average correlation between the leading (at timescale Δt_1) and lagging (at timescale Δt_2) activity rates, $\rho(\Delta t_1, \Delta t_2)$. Right hand-side plots: t-statistics of the difference $\rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; negative value correspond to activity at small timescales being more correlated to future activity at larger timescale than reversely. LB dataset, $T_{\text{in}} = 90$.

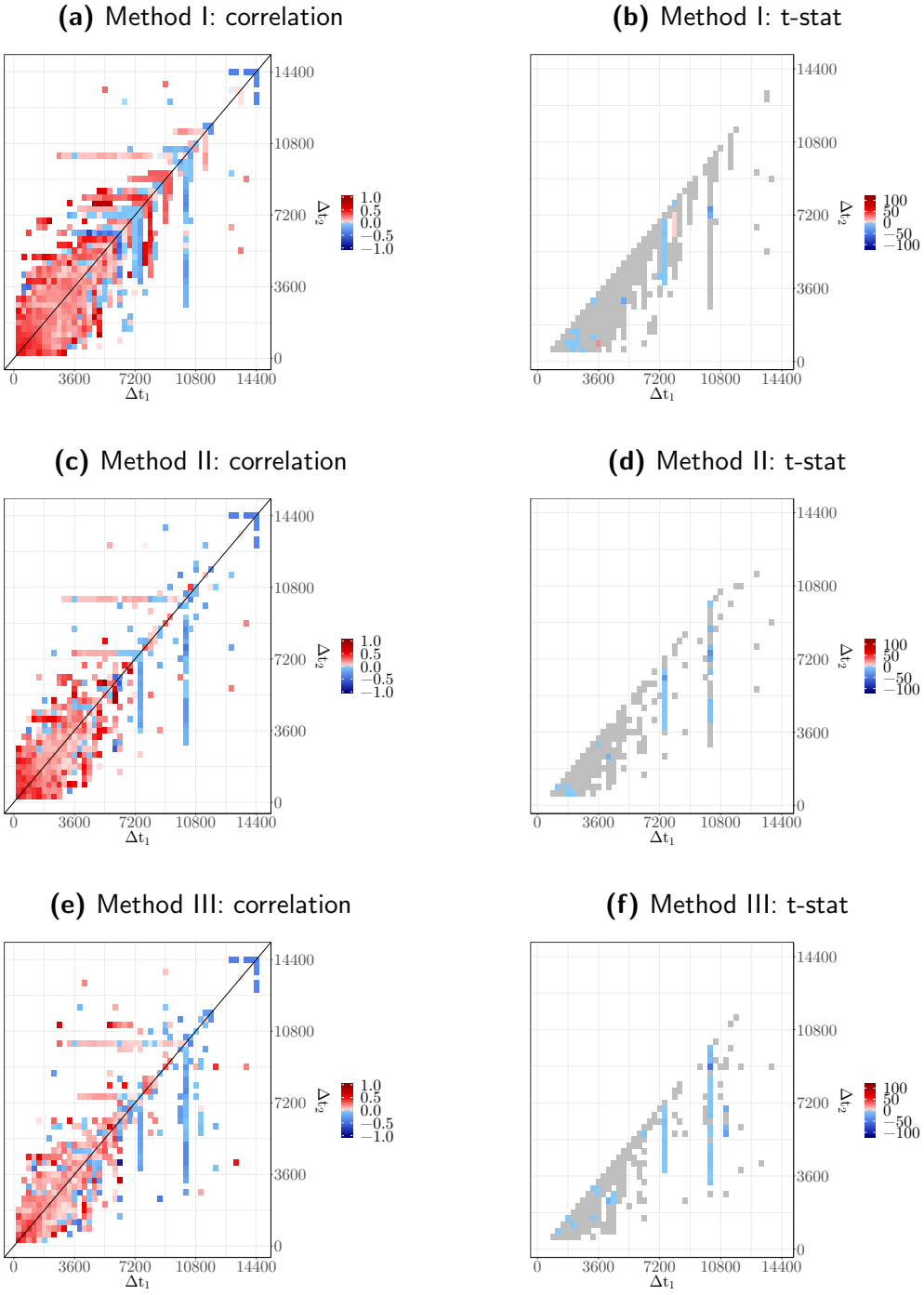


Figure 3.23: Left hand-side plots: average correlation between the leading (at timescale Δt_1) and lagging (at timescale Δt_2) activity rates, $\rho(\Delta t_1, \Delta t_2)$. Right hand-side plots: t-statistics of the difference $\rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; negative values correspond to activity at small timescales being more correlated to future activity at larger timescale than reversely. SQ dataset, $T_{\text{in}} = 15$.

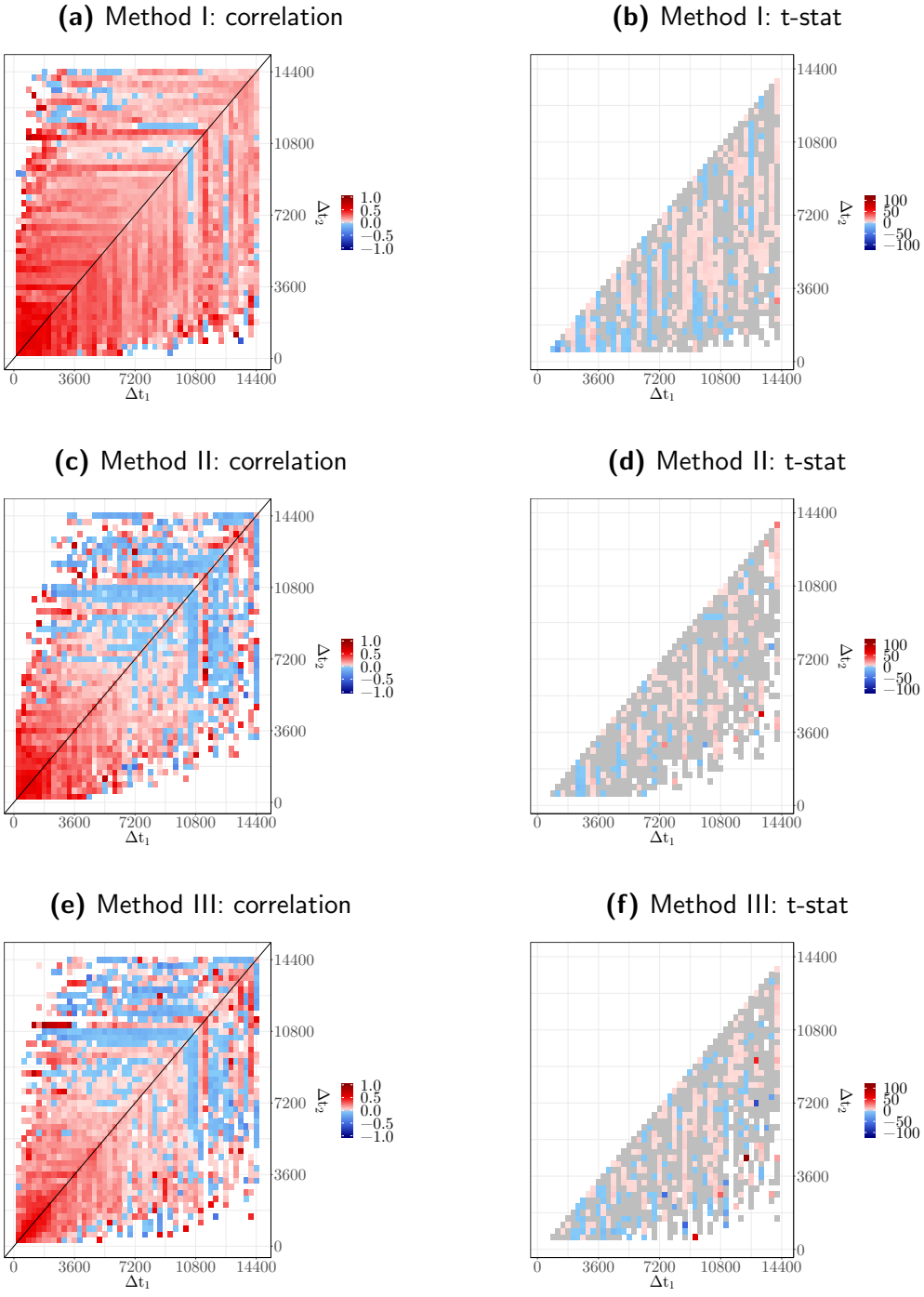


Figure 3.24: Left hand-side plots: average correlation between the leading (at timescale Δt_1) and lagging (at timescale Δt_2) activity rates, $\rho(\Delta t_1, \Delta t_2)$. Right hand-side plots: t-statistics of the difference $\rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; negative values correspond to activity at small timescales being more correlated to future activity at larger timescale than reversely. SQ dataset, $T_{\text{in}} = 60$.

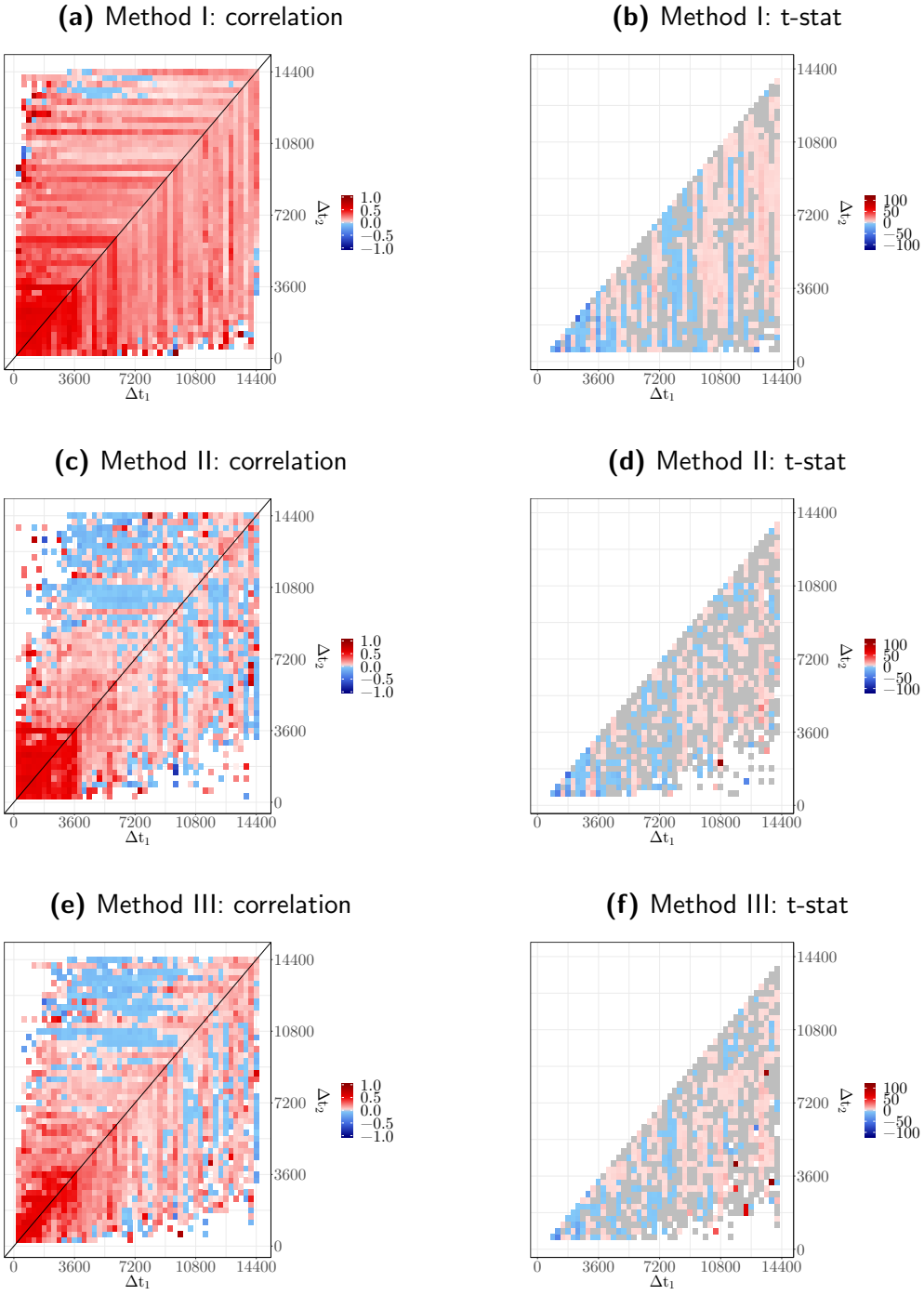


Figure 3.25: Left hand-side plots: average correlation between the leading (at timescale Δt_1) and lagging (at timescale Δt_2) activity rates, $\rho(\Delta t_1, \Delta t_2)$. Right hand-side plots: t-statistics of the difference $\rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; negative values correspond to activity at small timescales being more correlated to future activity at larger timescale than reversely. SQ dataset, $T_{\text{in}} = 90$.

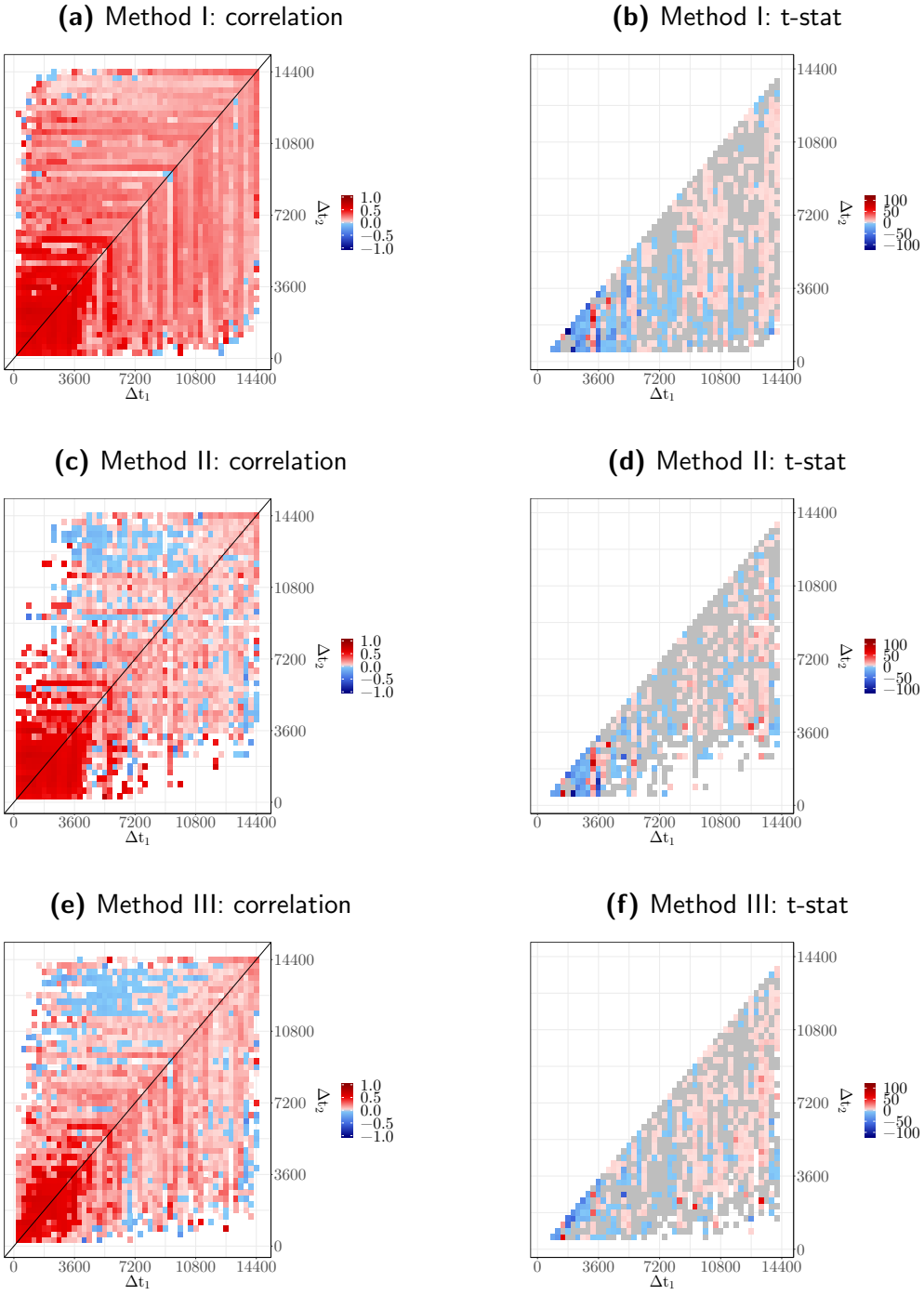


Figure 3.26: Left hand-side plots: average correlation between the leading (at timescale Δt_1) and lagging (at timescale Δt_2) activity rates, $\rho(\Delta t_1, \Delta t_2)$. Right hand-side plots: t-statistics of the difference $\rho(\Delta t_1, \Delta t_2) - \rho(\Delta t_2, \Delta t_1)$; negative values correspond to activity at small timescales being more correlated to future activity at larger timescale than reversely. SQ dataset, $T_{\text{in}} = 120$.

Chapter 4

Conclusions and Outlook

This thesis has focused on how to determine causality in financial markets from trader-resolved data. In other words, how the macroscopic dynamics of prices is shaped by the decisions of traders (which we have referred to as the nanoscopic level).

The dynamics of causal systems depends on whether time runs forward or backward. As a consequence, the system is asymmetric with respect to the inversion of the arrow of time (TRA). This implies that a method used to capture causality must also be sensitive to the inversion of the time arrow $t \to -t$. The large heterogeneity of the timescales of market participants raises the fascinating question of how these timescales interact, or, if one can analyse trader-resolved data, how agents operating at different timescales interact.

An *a priori* ideal method to determine causality and the mutual influence of all timescales at once is to fit Hawkes processes (HPs hereafter) to trader-resolved data. A meticulous investigation of the most active traders in our data, one by one, showed that their activity patterns are too non-stationary, irregular, and globally too rare for HPs to produce meaningful fits.

Working with HP with that purpose in mind lead to quantifying the effective causal strength of the simplest HPs (univariate and multivariate with symmetric mutual influence kernels). Fitting causal models on the data whose arrow of time has been inverted, i.e. changing the index of the N events from $\{1, \cdots, N\}$ to $\{N, \cdots, 1\}$, is a deceptively simple, yet efficient, method to assess the intrinsic ability of a model to infer TRA. It turns out that the estimated parameters of the simplest HPs vary remarkably little when the arrow of time is inverted. Unless one has sufficiently long time series and one performs proper statistical tests, it is difficult to distinguish the arrow of time in the estimations in the univariate and symmetric multivariate case, whereas it is clearly apparent in the asymmetric

multivariate case.

The situation with empirical data is worse, first of all because there is no proof that they have been generated by an HP, and also because data collection does not guarantee the precision of timestamps. These results emphasize the fact that fits of HPs to financial data must be backed up by proper statistical tests. We have also found that fits of HPs are more likely to be significant when the level of endogeneity is high.

Grouping traders and considering the group-level activity solved both the lack of data for a given trader and the irregular behavior of traders. Using sliding windows solved another problem, namely that the population of active traders is not stationary even for a single year. The method known as Statistically Validated Networks (SVNs) groups traders in a pairwise way according to the statistically proven (and surprising) coordination between them. Crucially, it requires choosing a timescale at which the activity of agents is summarized. It was known that SVNs provide an effective and robust way to cluster agents. A systematic investigation shows that the most meaningful timescale depends on the type of investors (retail or institutional), the behaviour of the latter being better captured by longer timescales (at least in our datasets).

We then introduced a method (with three variants) to infer the lead-lag structure of trader activity when the activity in the past is determined at one timescale and the activity in the future at another. It was shown to be effective to infer strong asymmetries between timescales for institutional FX traders, and much weaker ones for retail clients. This is yet another proof of the very strong heterogeneity of market participants and of the amount of detail lost in the aggregation process that determines the price evolution.

Possible extensions to and applications of the methods and results of this thesis include:

1. Is prediction of agent-behaviour improved by feeding machine learning methods with states of groups which are determined at several timescales? What does one learn about causality from the respective importance of each timescale? For example, is there a relationship between the timescales of the most relevant predictors and the pair of timescales which display significant asymmetric mutual influence?

2. The fact that clustering with SVNs consistently produces a large group may be detrimental to the determination of lead-lag links. Lead-lag relationships could be investigated directly from agent to agent, if computational resources are sufficient.

3. Larger and longer datasets would be a boost to extend the time-domain in which TRA can be detected.

4. Finally, the domain of application of the multi-scale lead-lag inference method introduced here is not limited to traders. Applying it to other types of datasets, e.g. of consumer behaviour, is a natural extension, and could be used e.g. to determine which early adopters of new products are influential.

Bibliography

Massil Achab, Emmanuel Bacry, Stéphane Gaïffas, Iacopo Mastromatteo, and Jean-François Muzy. Uncovering causality from multivariate Hawkes integrated cumulants. *Journal of Machine Learning Research*, 18(192):1–28, 2018.

Yacine Ait-Sahalia, Jianqing Fan, and Yingying Li. The leverage effect puzzle: Disentangling sources of bias at high frequency. *Journal of Financial Economics*, 109(1):224–249, 2013.

Alain Arneodo, Jean-Philippe Bouchaud, Rama Cont, Jean-François Muzy, Marc Potters, and Didier Sornette. Comment on"Turbulent cascades in foreign exchange markets". *arXiv preprint cond-mat/9607120*, 1996.

E. Bacry, M. Bompaire, S. Gaïffas, and S. Poulsen. tick: a Python library for statistical learning, with a particular emphasis on time-dependent modeling. *ArXiv e-prints*, July 2017.

Emmanuel Bacry, Khalil Dayri, and Jean-François Muzy. Non-parametric kernel estimation for symmetric hawkes processes. Application to high frequency financial data. *The European Physical Journal B*, 85(5):157, 2012.

Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

Emmanuel Bacry, Stéphane Gaïffas, Iacopo Mastromatteo, and Jean-François Muzy. Mean-field inference of Hawkes point processes. *Journal of Physics A: Mathematical and Theoretical*, 49(17):174006, 2016.

Lionel Barnett, Adam B Barrett, and Anil K Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Physical review letters*, 103(23): 238701, 2009.

Geert Bekaert and Guojun Wu. Asymmetric volatility and risk in equity markets. *The review of financial studies*, 13(1):1–42, 2000.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

Fischer Black. Proceedings of the 1976 meetings of the american statistical association. American Statistical Association, 1976.

Pierre Blanc, Jonathan Donier, and J-P Bouchaud. Quadratic Hawkes processes for financial prices. *Quantitative Finance*, 17(2):171–188, 2017.

Lisa Borland and J-Ph Bouchaud. On a multi-timescale statistical feedback model for volatility fluctuations. *arXiv preprint physics/0507073*, 2005.

Lisa Borland, Jean-Philippe Bouchaud, Jean-François Muzy, and Gilles Zumbach. The dynamics of financial markets–mandelbrot's multifractal cascades, and beyond. *arXiv preprint cond-mat/0501292*, 2005.

Jean-Philippe Bouchaud, Andrew Matacz, and Marc Potters. Leverage effect in financial markets: The retarded volatility model. *Physical review letters*, 87(22): 228701, 2001.

Jacob Boudoukh, Matthew P Richardson, and RE Whitelaw. A tale of three schools: Insights on autocorrelations of short-horizon stock returns. *Review of financial studies*, 7(3):539–573, 1994.

Pierre Brémaud. Point processes and queues: martingale dynamics. 1981.

Steven L Bressler and Anil K Seth. Wiener–Granger causality: a well established methodology. *Neuroimage*, 58(2):323–329, 2011.

Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

Damien Challet, Rémy Chicheportiche, Mehdi Lallouache, and Serge Kassibrakis. Statistically validated lead-lag networks and inventory prediction in the foreign exchange market. *Advances in Complex Systems*, page 1850019, 2018.

Ishanu Chattopadhyay. Causality networks. *arXiv preprint arXiv:1406.6651*, 2014.

Rémy Chicheportiche and Jean-Philippe Bouchaud. The fine-structure of volatility feedback I: Multi-scale self-reflexivity. *Physica A: Statistical Mechanics and its Applications*, 410:174–195, 2014.

Andrew A Christie. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, 10(4):407–432, 1982.

Peter K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–155, 1973. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/1913889`.

Marcus Cordi, Damien Challet, and Ioane Muni Toke. Testing the causality of Hawkes processes with time reversal. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(3):033408, 2018.

Chester Curme, Michele Tumminello, Rosario N Mantegna, H Eugene Stanley, and Dror Y Kenett. Emergence of statistically validated financial intraday lead-lag relationships. *Quantitative Finance*, 15(8):1375–1386, 2015.

Michel Dacorogna, Ulrich Müller, Olivier Pictet, and Richard Olsen. Modelling short-term volatility with GARCH and HARCH models. 1998.

DJ Daley and D Vere-Jones. Basic Properties of the Poisson Process. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, pages 19–40, 2003.

Angelos Dassios and Hongbiao Zhao. A generalized contagion process with an application to credit risk. *International Journal of Theoretical and Applied Finance*, 20(01):1750003, 2017.

Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

Michael Eichler, Rainer Dahlhaus, and Johannes Dueck. Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *Journal of Time Series Analysis*, 38(2):225–242, 2017.

Omar El Euch, Jim Gatheral, Radoš Radoičić, and Mathieu Rosenbaum. The Zumbach effect under rough Heston. *arXiv preprint arXiv:1809.02098*, 2018.

Vladimir Filimonov and Didier Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.

JK Gardner and Leon Knopoff. Is the sequence of earthquakes in southern california, with aftershocks removed, Poissonian? *Bulletin of the Seismological Society of America*, 64(5):1363–1367, 1974.

Shoaleh Ghashghaie, Wolfgang Breymann, Joachim Peinke, Peter Talkner, and Yadollah Dodge. Turbulent cascades in foreign exchange markets. *Nature*, 381 (6585):767, 1996.

Laszlo Gillemot, J Doyne Farmer, and Fabrizio Lillo. There's more to volatility than volume. *Quantitative Finance*, 6(5):371–384, 2006.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.

C.W.J. Granger. Economic processes involving feedback. *Information and Control*, 37:28–48, Mars 1963.

Abeyratna Gunasekarage and David M Power. The profitability of moving average trading rules in South Asian stock markets. *Emerging Markets Review*, 2 (1):17–33, 2001.

Stephen J Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. Critical reflexivity in financial markets: a Hawkes process analysis. *The European Physical Journal B*, 86(10):442, 2013.

Takaki Hayashi, Nakahiro Yoshida, et al. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 2005.

Cars H Hommes. Heterogeneous agent models in economics and finance. *Handbook of computational economics*, 2:1109–1186, 2006.

David Hume. An enquiry concerning human understanding. In *Seven Masterpieces of Philosophy*, pages 191–284. Routledge, 2016.

Nicolas Huth and Frédéric Abergel. High frequency lead/lag relationships—empirical facts. *Journal of Empirical Finance*, 26:41–58, 2014.

Narasimhan Jegadeesh and Sheridan Titman. Overreaction, delayed reaction, and contrarian profits. *The Review of Financial Studies*, 8(4):973–993, 1995.

Immanuel Kant. *Critique of pure reason*. Cambridge university press, 1999.

Matthias Kirchner. An estimation procedure for the Hawkes process. *Quantitative Finance*, 17(4):571–595, 2017.

Ryota Kobayashi and Renaud Lambiotte. TiDeH: Time-Dependent Hawkes Process for Predicting Retweet Dynamics. In *ICWSM*, pages 191–200, 2016.

Dimitri Kroujiline, Maxim Gusev, Dmitry Ushanov, Sergey V Sharov, and Boris Govorkov. Forecasting stock market returns over multiple time horizons. *Quantitative Finance*, 16(11):1695–1712, 2016.

Mehdi Lallouache and Damien Challet. The limits of statistical significance of Hawkes processes fitted to financial data. *Quantitative Finance*, 16(1):1–11, 2016.

Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117, 2009.

Ming-Xia Li, Vasyl Palchykov, Zhi-Qiang Jiang, Kimmo Kaski, János Kertész, Salvatore Miccichè, Michele Tumminello, Wei-Xing Zhou, and Rosario N Mantegna. Statistically validated mobile communication networks: the evolution of motifs in European and Chinese data. *New Journal of Physics*, 16(8):083038, 2014.

Michael London, Arnd Roth, Lisa Beeren, Michael Häusser, and Peter E Latham. Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, 466(7302):123, 2010.

Thomas Lux et al. Turbulence in financial markets: the surprising explanatory power of simple cascade models. *Quantitative finance*, 1(6):632–640, 2001.

Paul E Lynch, Gilles O Zumbach, et al. Market heterogeneities and the causal structure of volatility. *Quantitative Finance*, 3(4):320–331, 2003.

Burton G Malkiel and Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

David Marsan and Olivier Lengline. Extending earthquakes reach through cascading. *Science*, 319(5866):1076–1079, 2008.

Matteo Marsili and Maurizio Piai. Colored minority games. *Physica A: Statistical Mechanics and its Applications*, 310(1-2):234–244, 2002.

Mariusz Maziarz. A review of the Granger-causality fallacy. *The journal of philosophical economics: Reflections on economic and social issues*, 8(2):86–105, 2015.

George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, 2011.

Giancarlo Mosetti, Damien Challet, and Yi-Cheng Zhang. Minority games with heterogeneous timescales. *Physica A: Statistical Mechanics and its Applications*, 365(2):529–542, 2006.

Ulrich A Müller, Michel M Dacorogna, Rakhal D Davé, Olivier V Pictet, Richard B Olsen, and J Robert Ward. Fractals and intrinsic time: A challenge to econometricians. *Unpublished manuscript, Olsen & Associates, Zürich*, 1993.

Ulrich A Müller, Michel M Dacorogna, Rakhal D Davé, Richard B Olsen, Olivier V Pictet, and Jacob E Von Weizsäcker. Volatilities of different time resolutions—analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2-3):213–239, 1997.

Federico Musciotto, Luca Marotta, Jyrki Piilo, and Rosario N Mantegna. Long-term ecology of investors in a financial market. *Palgrave Communications*, 4(1): 92, 2018.

John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

Yosihiko Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401): 9–27, 1988.

Takahiro Omi, Yoshito Hirata, and Kazuyuki Aihara. Hawkes process model with a time-dependent background rate and its application to high-frequency financial data. *Physical Review E*, 96(1):012303, 2017.

F Papangelou. Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society*, 165:483–506, 1972.

Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207): 995, 2008.

I Present. Cramming more components onto integrated circuits. *Readings in computer architecture*, 56, 2000.

Marcello Rambaldi, Emmanuel Bacry, and Jean-François Muzy. Disentangling and quantifying market participant volatility contributions. *arXiv preprint arXiv:1807.07036*, 2018.

Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.

François Roueff, Rainer Von Sachs, and Laure Sansonnet. Locally stationary Hawkes processes. *Stochastic Processes and their Applications*, 126(6):1710–1743, 2016.

Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2): 461, 2000.

George Soros. *The alchemy of finance*. John Wiley & Sons, 2003.

Nassim Nicholas Taleb. *The black swan: The impact of the highly improbable*, volume 2. Random house, 2007.

James Tobin. A general equilibrium approach to monetary theory. *Journal of money, credit and banking*, 1(1):15–29, 1969.

Bence Tóth and János Kertész. Increasing market efficiency: Evolution of cross-correlations of stock returns. *Physica A: Statistical Mechanics and its Applications*, 360(2):505–515, 2006.

Wilson Truccolo, Uri T Eden, Matthew R Fellows, John P Donoghue, and Emery N Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology*, 93(2):1074–1089, 2005.

Michele Tumminello, Salvatore Micciche, Fabrizio Lillo, Jyrki Piilo, and Rosario N Mantegna. Statistically validated networks in bipartite complex systems. *PloS one*, 6(3):e17994, 2011.

Michele Tumminello, Fabrizio Lillo, Jyrki Piilo, and Rosario N Mantegna. Identification of clusters of investors from their real trading activity in a financial market. *New Journal of Physics*, 14(1):013041, 2012.

Norbert Wiener. The theory of prediction. *Modern mathematics for engineers*, 1956.

Hongteng Xu, Mehrdad Farajtabar, and Hongyuan Zha. Learning Granger causality for Hawkes processes. In *International Conference on Machine Learning*, pages 1717–1726, 2016.

Ke Zhou, Hongyuan Zha, and Le Song. Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes. In *Artificial Intelligence and Statistics*, pages 641–649, 2013.

Jiancang Zhuang, Yosihiko Ogata, and David Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380, 2002.

Gilles Zumbach. Time reversal invariance in finance. *Quantitative Finance*, 9(5): 505–515, 2009.

Gilles Zumbach and Paul Lynch. Heterogeneous volatility cascade in financial markets. *Physica A: Statistical Mechanics and its Applications*, 298(3-4):521–529, 2001.

Gilles Zumbach, Luis Fernández, and Caroline Weber. Processes for stocks capturing their statistical properties from one day to one year. *Quantitative Finance*, 14(5):849–861, 2014.

université PARIS-SACLAY

ÉCOLE DOCTORALE
INTERFACES
Approches Interdisciplinaires : Fondements, Applications et Innovation

Titre: Causalité des marchés financiers: asymétrie temporelle et réseaux multi-échelles de meneurs et suiveurs

Mots clés: processus de Hawkes, asymétrie temporelles, réseaux validés statistiquement, réseaux de meneurs et suiveurs

Résumé: Cette thèse a pour but d'explorer la structure de causalité qui sous-tend les marchés financiers. Elle se concentre sur l'inférence multi-échelle de réseaux de causalité entre investisseurs dans deux bases de données contenant les identifiants des investisseurs.

La première partie de cette thèse est consacrée à l'étude de la causalité dans les processus de Hawkes. Ces derniers définissent la façon dont l'activité d'un investisseur (par exemple) dépend du passé; sa version multivariée inclut l'interaction entre séries temporelles, à toutes les échelles. Les résultats principaux de cette partie est que l'estimation avec le maximum de vraisemblance des paramètres du processus changent remarquablement peu lorsque la direction du temps est inversée, tant pour les processus univariés que pour les processus multivariés avec noyaux d'influence mutuelle symétriques, et que la causalité effective de ces processus dépend de leur endogénéité. Cela implique qu'on ne peut pas utiliser ce type de processus pour l'inférence de causalité sans précautions. L'utilisation de tests statistiques permet la différentiation des directions du temps pour des longues données synthétiques. Par contre, l'analyse de données empiriques est plus problématique: il est tout à fait possible de trouver des données financières pour lesquelles la vraisemblance des processus de Hawkes est plus grande si le temps s'écoule en sens inverse.

Les processus de Hawkes multivariés avec noyaux d'influence asymétriques ne sont pas affectés par une faible causalité. Il est malheureusement difficile de les calibrer aux actions individuelles des investisseurs présents dans nos bases de données, pour deux raisons. Nous avons soigneusement vérifié que l'activité des investisseurs est hautement non-stationaire et qu'on ne peut pas supposer que leur activité est localement stationaire, faute de données en nombre suffisant, bien que nos bases de données contiennent chacune plus de 1 million de transactions. Ces problèmes sont renforcés par le fait que les noyaux dans les processus de Hawkes codent l'influence mutuelle des investisseurs pour toutes les échelles de temps simultanément.

Afin de pallier ce problème, la deuxième partie de cette thèse se concentre sur la causalité entre des échelles de temps spécifiques. Un filtrage supplémentaire est obtenu en réduisant le nombre effectif d'investisseurs grâce aux Réseaux Statistiquement Validés. Ces derniers sont utilisés pour catégoriser les investisseurs, qui sont groupés selon leur degré de la synchronisation de leurs actions (achat, vente, neutre) dans des intervalles déterminés à une échelle temporelle donnée. Cette partie propose une méthode pour l'inférence de réseaux de meneurs et suiveurs déterminés à une échelle de temps donnée dans le passé et à une autre dans le futur. Trois variations de cette méthode sont étudiées.

Cette méthode permet de caractériser la causalité d'une façon novatrice. Nous avons comparé l'asymétrie temporelle des actions des investisseurs et celle de la volatilité des prix, et conclure que la structure de causalité des investisseurs est considérablement plus complexe que celle de la volatilité. De façon attendue, les investisseurs institutionnels, dont l'impact sur l'évolution des prix est beaucoup plus grand que celui des clients privés, ont une structure causale proche de celle de la volatilité: en effet, la volatilité, étant une quantité macroscopique, est le résultat d'une aggrégation des comportements de tous les investisseurs, qui fait disparaître la structure causale des investisseurs privés.

Title: Causality in Financial Markets: Time Reversal Asymmetry and Multi-Scale Lead-Lag Networks

Keywords: Hawkes process, time reversal asymmetry, statistically validated networks, lead-lag networks

Abstract: This thesis aims to uncover the underlying causality structure of financial markets by focusing on the inference of investor causal networks at multiple timescales in two trader-resolved datasets.

The first part of this thesis is devoted to the causal strength of Hawkes processes. These processes describe in a clearly causal way how the activity rate of e.g. an investor depends on his past activity rate; its multivariate version also makes it possible to include the interactions between the agents, at all time scales. The main result of this part is that the classical MLE estimation of the process parameters does not vary significantly if the arrow of time is reversed in the univariate and symmetric multivariate case. This means that blindly trusting univariate and symmetric multivariate Hawkes processes to infer causality from data is problematic. In addition, we find a dependency between the level of causality in the process and its endogeneity. For long time series of synthetic data, one can discriminate between the forward and backward arrows of time by performing rigorous statistical tests on the processes, but for empirical data the situation is much more ambiguous, as it is entirely possible to find a better Hawkes process fit when time runs backwards compared to forwards.

Asymmetric Hawkes processes do not suffer from very weak causality. Fitting them to the individual traders' actions found in our datasets is unfortunately not very successful for two reasons. We carefully checked that traders actions in both datasets are highly non-stationary, and that local stationarity cannot be assumed to hold as there is simply not enough data, even if each dataset contains about one million trades. This is also compounded by the fact that Hawkes processes encode the pairwise influence of traders for all timescales simultaneously.

In order to alleviate this problem, the second part of this thesis focuses on causality between specific pairs of timescales. Further filtering is achieved by reducing the effective number of investors; Statistically Validated Networks are applied to cluster investors into groups based on the statistically high synchronisation of their actions (buy, sell or neutral) in time intervals of a given timescale. This part then generalizes single-timescale lead-lag SVNs to lead-lag networks between two timescales and introduces three slightly different methods

These methods make it possible to characterize causality in a novel way. We are able to compare the time reversal asymmetry of trader activity and that of price volatility, and conclude that the causal structure of trader activity is considerably more complex than that of the volatility for a given category of traders. Expectedly, institutional traders, whose impact on prices is much larger than that of retail clients, have a causality structure that is closer to that of volatility. This is because volatility, being a macroscopic quantity, aggregates the behaviour of all types of traders, thereby hiding the causality structure of minor players.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France