AAAI Technical Report FS-12-03
Discovery Informatics: The Role of AI Research in Innovating Scientific Processes

On Causality Inference in Time Series

Mohammad Taha Bahadori and Yan Liu
Department of Computer Science
University of Southern California

Abstract

Causality discovery has been one of the core tasks in scientific research since the beginning of human scientific history. In the age of data tsunami, the task could involve millions of variables, which cannot be achieved feasibly by human. However, the causal discovery using artificial intelligence and statistical techniques in non-experimental settings faces several challenges. In this work, we address three practical challenges regarding Granger causality, one of the most popular causality inference techniques for time series data. First, we analyze the consistency of two most popular Granger causality techniques and show that the significance test is not consistent in high dimensions. Second, we review the nonparametric generalization of the Lasso-Granger technique called Generalized Lasso Granger (GLG) to uncover Granger causality relationships among irregularly sampled time series. Finally, we describe two techniques to uncover the casual dependence in non-linear datasets. Extensive experiments on the climate datasets are provided to show the significant advantages of the proposed algorithms over their state-of-the-art counterparts.

1 Introduction

Discovering the causal relationships among multiple natural processes constitutes the core of the scientific method which is one of the top achievements of the human being throughout the history. Once Democritus, an ancient Greek philosopher (460-370 BC), stated "I would rather discover one causal law than be the king of Persia". This means since the dawn of the civilization, human being was fully aware of the significance of the identification of causal relationships in the nature. With the development of new technologies in many domains, we are now in an era of data tsunami, in which massive amount of time series data become available for analysis and mining. Uncovering the underlying causal graphs from large amount of data could have great benefits to many areas ranging from scientific discovery to practical applications, but achieving this is an extremely challenging task in artificial intelligence and machine learning.

Granger Causality (Granger 1969) is one of the earliest methods developed to quantify the causal effect among multiple time series. It is based on the common conception that the cause usually occurs prior to its effect. Formally, X Granger causes Y if its past value can help to predict the future value of Y beyond what could have been done with the past value of Y only. It has gained tremendous success across many domains due to its simplicity, robustness, and extendability (Panchenko and Valentyn 2004; Brovelli et al. 2004; Hiemstra and Jones 1994; Asimakopoulos, Ayling, and Mansor Mahmood 2000; Marinazzo, Pellicoro, and Stramaglia 2008).

Granger causality, like many other data-driven causality analysis approaches, is confronted with three major challenges: The computational challenges which stem from the characteristics that many practical applications involve a large number of variables but very few observations (compared with the much larger number of variables involved), which makes it difficult for statistical methods to effectively infer the causality relationships. The second challenge originates from imperfectness of the data collection procedures. In many real-world applications the data has not been collected on uniformly spaced time intervals. Finally, the original Granger causality is based on vector auto-regressive (VAR) model assumption for the data. In many real-world scientific applications the data significantly deviates from the VAR model and non-linear extensions are on demand. We test our algorithms in two climatology datasets to highlight their superior performance in comparison with the state of the art algorithms.

To address the first challenge we analyze the consistency of two popular Granger causality inference techniques; the significance tests and Lasso-Granger. We show that while both methods are consistent in low dimensions, the significance tests become inconsistent in high dimensions; thus the Lasso-Granger is the preferred method in high dimensions. Confident about the robust performance of Lasso-Granger, we develop a non-parametric generalization of the Lasso-Granger for irregular time series called Generalized Lasso-Granger (GLG) (Bahadori and Liu 2012) and show its superior performance over existing algorithms in uncovering the Granger causality patterns in irregular time series. For non-linear extensions, we describe two extensions of the Granger causality (Liu, Bahadori, and Li 2012), one based on the copula approach and the other one based on Transfer Entropy technique.

In the rest of the paper, we first review the Granger causality and the existing approaches to uncover it in Section 2, and then we discuss the theoretical consistency analysis in Section 3. Sections 4 and 5 are dedicated to description of GLG and non-linear extensions of Granger causality, respectively. In Section 6, we show experiment results on application datasets to show the significant advantage of the proposed algorithms over the state of the art ones, and finally in Section 7 we summarize the paper and hint on future work.

2 Preliminaries

Granger Causality is one of the most popular approaches to quantify causal relationships among time series observations. It is based on two major principles: (i) The cause happens prior to the effect and (ii) The cause makes unique changes in the effect (Granger 1969; 1980). There have been extensive debates on the validity and generality of these principles. In this paper, we omit the lengthy discussion and simply assume their correctness for the rest of the discussion.

Given two stationary time series $X = \{X(t)\}_{t\in\mathbb{Z}}$ and $Y = \{Y(t)\}_{t\in\mathbb{Z}}$, we can consider the following information sets: (i) $\mathcal{I}^\star(t)$, the set of all information in the universe up to time t, and (ii) $\mathcal{I}^\star_{-X}(t)$, the set of all information in the universe excluding X up to time t. Under the two principles of Granger causality, the conditional distribution of future values of Y given $\mathcal{I}^\star_{-X}(t)$ and $\mathcal{I}^\star(t)$ should differ. Therefore X is defined to Granger cause Y (Granger 1969; 1980) if

$$\mathbb{P}[Y(t+1) \in A|\mathcal{I}^\star(t)] \neq \mathbb{P}[Y(t+1) \in A|\mathcal{I}^\star_{-X}(t)], \quad (1)$$

for some measurable set $A \subseteq \mathbb{R}$ and all $t \in \mathbb{Z}$. As we can see, the original definition of Granger causality is very general and does not have any assumptions on the data generation process. However, modeling the distributions for multivariate time series could be extremely difficult while linear models are a simple yet robust approach, with strong empirical performance in practical applications. As a results, Vector Auto-regression (VAR) models have evolved to be one of the dominate approaches for Granger causality.

Up to now, two major approaches based on VAR model have been developed to uncover Granger causality for multivariate time series. One approach is the *significance test* (Marinazzo, Pellicoro, and Stramaglia 2008): given multiple time series $X_1, \ldots, X_V$, we run a VAR model for each time series X_j, i.e.,

$$X_j(t) = \sum_{i=1}^{V} \boldsymbol{\beta}_{j,i}^\top \mathbf{X}_i^{t,Lagged} + \epsilon_j, \quad (2)$$

where $\mathbf{X}_i^{t,Lagged} = [X_i(t-L), \ldots, X_i(t-1)]$ is the history of X_i up to time t, L is the maximal time lag, and $\boldsymbol{\beta}_{j,i} = [\beta_{j,i}(1), \ldots, \beta_{j,i}(L)]$ is the vector of coefficients modeling the effect of time series X_i on the target time series. We can determine that time series X_i Granger causes X_j if at least one value in the coefficient vector $\boldsymbol{\beta}_{j,i}$ is nonzero by statistical significant tests. The second approach is the *Lasso-Granger approach* (Valdés-Sosa et al. 2005; Arnold, Liu, and Abe), which applies lasso-type VAR model to obtain a sparse and robust estimate of the coefficient vectors for Granger causality tests. Specifically, the regression task in Eq. (2) can be achieved by solving the following optimization problem:

$$\min_{\{\boldsymbol{\beta}\}} \sum_{t=L+1}^{T} \left\| X_j(t) - \sum_{i=1}^{V} \boldsymbol{\beta}_{j,i}^\top \mathbf{X}_i^{t,Lagged} \right\|_2^2 + \lambda \left\|\boldsymbol{\beta}\right\|_1, \quad (3)$$

where λ is the penalty parameter, which determines the sparsity of the coefficient vector $\boldsymbol{\beta}$.

3 Consistency Analysis

In this section, we first show that both the significance tests and the Lasso-Granger approach (subject to some conditions) are consistent in identifying the underlying Granger causality relationships, in lower dimensions. We will also show that the significance tests are inconsistent in high dimensions; thus the Lasso-Granger is the preferred method in high dimensions. First, let us define the consistency by introducing the probability of errors for VAR-type models as follows:

$$\begin{aligned}\mathbb{P}[\text{Error}] = \mathbb{P}[\exists \ell:\ |\widehat{\boldsymbol{\beta}}_{i,j}(\ell)| > \alpha_0 | \boldsymbol{\beta}_{i,j} = \mathbf{0}]\mathbb{P}[\boldsymbol{\beta}_{i,j} = \mathbf{0}] \\ + \mathbb{P}[\forall \ell:\ |\widehat{\boldsymbol{\beta}}_{i,j}(\ell)| < \alpha_0 | \boldsymbol{\beta}_{i,j} \neq \mathbf{0}]\mathbb{P}[\boldsymbol{\beta}_{i,j} \neq \mathbf{0}].\end{aligned}$$

We say that a method is consistent if its probability of errors goes to zero when the number of observations increases.

Theorem 1. *In a VAR system, suppose all the confounders in a system have been observed (i.e., there exist no hidden variables). If $T/L - 1 > V$, the significance tests are consistent with probability of error decaying with rate $\mathbb{P}[\text{Error}] \leq 2cL\sqrt{T-L}\exp\left(-\frac{c^2}{2}(T-L)\right)$ for some constant c, where T is the length of time series, and L is the maximal lag; otherwise the significance tests are inconsistent. The Lasso-Granger subject to conditions in (Meinshausen and Yu 2009) is always consistent with model selection error rate $o(L\exp(-T^\nu))$ for some $0 \leq \nu < 1$.*

The proof will be provided in the extended version.

Remarks 1. The result in Proposition 1 states that the error decreases exponentially as the length of the time series increase for both approaches. Also it states that when $L \ll T$ large value of L linearly degrades the performance, whereas in the case of $L \sim T$ the exponential term will be dominant and the error will increase exponentially with L.
2. The consistency results also imply that learning linear Granger causal relationships is a simpler task than learning undirected graphical models (Meinshausen and Bühlmann 2006). This is intuitive since learning the edges for one node is a variable selection process isolated from that for other nodes and therefore no constraint on the neighborhood nodes is required.

4 GLG for Irregular Time Series

After showing the advantages of Lasso-Granger, we intend to extend the power of Lasso-Granger to address the practical challenge of irregularity of time series in real-world applications. Consider the following definition of irregular time series:

Definition 1. *An irregular time series* $\mathbf{x}$ *of length* N *is denoted by* $\mathbf{x} = \{(t_n, x_n)\}_{n=1}^{N}$ *where time-stamp sequence* $\{t_n\}$ *are strictly increasing, i.e.,* $t_1 < t_2 < \ldots < t_N$ *and* x_n *are the value of the time series at the corresponding time stamps.*

The major challenge to uncover temporal causal networks for irregular time series is how to effectively capture the temporal dependence without directly estimating the values of missing data or making restricted assumptions about the generation process of the time series.

Generalized Lasso Granger (GLG)

The key idea of our model is as follows: if we treat $\boldsymbol{\beta}_{i,j}$ in Eq. (3) as a time series, $\boldsymbol{\beta}_{i,j}^{\top}\mathbf{x}^{(j)}$ can be considered as its inner product with another time series $\mathbf{x}^{(j)}$. If we generalize the inner product operator to irregular time series, the temporal causal models for regular time series can be easily extended to handle irregular cases.

Let us denote the generalization of dot product between two irregular time series $\mathbf{x}$ and $\mathbf{y}$ by $\mathbf{x} \odot \mathbf{y}$, which can be interpreted as a (possibly non-linear) function that measures the unnormalized similarity between them. Depending on the target application, one can define different similarity measures, and thus inner product definitions. For example, we can define the inner product as a linear function with respect to the first time series components as follows:

$$\mathbf{x} \odot \mathbf{y} = \sum_{n=1}^{N_x} \frac{\sum_{m=1}^{N_y} x_n y_m w\left(t_n^x, t_m^y\right)}{\sum_{m=1}^{N_y} w\left(t_n^x, t_m^x\right)}, \tag{4}$$

where w is the kernel function. For example w can be the Gaussian kernel defined as following:

$$w(t_1, t_2) = \exp\left(-\frac{(t_2 - t_1)^2}{\sigma^2}\right). \tag{5}$$

Given the generalization of the inner product operator, we can now extend the regression in Eq. (3) to obtain the desired optimization problem for irregular time series. Formally, suppose P number of irregular time series $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(P)}$, are given. Let Δt denote the average length of the sampling intervals for the target time series (e.g. $\mathbf{x}^{(i)}$) and $\boldsymbol{\beta}'_{i,j}(t)$ be a pseudo time series, i.e.:

$$\boldsymbol{\beta}'_{i,j}(t) = \{(t_l, \beta_{i,j,l}) | l = 1, \ldots, L, t_l = t - l\Delta t\},$$

which means that for different value of t, $\boldsymbol{\beta}'_{i,j}(t)$ share the same observation vectors (i.e, $\{\boldsymbol{\beta}_{i,j}\}$), but the time stamp vectors vary according to the value of t. We can perform the causality analysis by generalized Lasso Granger (GLG) method that solves the following optimization problem:

$$\min_{\{\mathbf{a}_{i,j}\}} \sum_{n=\ell_0}^{N_i} \left\| x_n^{(i)} - \sum_{j=1}^{P} \boldsymbol{\beta}'_{i,j}(t_n^{(i)}) \odot \mathbf{x}^{(j)} \right\|_2^2 + \lambda \left\|\boldsymbol{\beta}_i\right\|_1, \tag{6}$$

where ℓ_0 is the smallest value of n that satisfies $t_n^{(i)} \geq L\Delta t$.

The above optimization problem is not convex in general and the convex optimization algorithms can only find a local minimum. However if the generalized inner product is defined to make Problem 6 convex, there are efficient algorithms such as FISTA (Beck and Teboulle 2009) to solve optimization problems of the form $f(\boldsymbol{\theta}) + \|\boldsymbol{\theta}\|_1$ where $f(\boldsymbol{\theta})$ is convex. In this paper, we use the linear generalization of the inner product given by Eq. (4) with which Problem (6) can be reformulated as linear prediction of $x_n^{(i)}$ using parameters $\boldsymbol{\beta}'_{i,j}(t_n^{(i)})$ subject to norm-1 constraint on the value of the parameters. Thus, the problem is a Lasso problem and can be solved more efficiently by optimized Lasso solvers such as Coordinate Descent (Friedman, Hastie, and Tibshirani 2010).

5 Non-Linear Granger Causality Discovery

The Lasso-Granger and Significance test methods are based on the vector autoregressive model for the data. In real-world there are many cases in which the distribution of the data significantly deviates from the VAR model. In this section we describe two methods to uncover Granger causality with non-linear temporal-causal dependency: The copula approach which is based on a non-linear mapping using the marginal distribution of the data and the Transfer Entropy which relies on the concept of entropy as a measure of information.

Copula Solution The Copula approach has been proposed for dependency analysis of time series with non-Gaussian marginal distributions, (Embrechts, Mcneil, and Straumann 2002). It has been used for prediction of time series (Leong and Valdez 2005) and learning the dependency graph among time series (Liu, Lafferty, and Wasserman 2009). In the copula framework, first the marginal distribution of the time series $\mathbf{x}^i$ are estimated as $\tilde{F}_i$. Next the observations are transformed to the copula domain as $u_t^i = \Phi^{-1}(\tilde{F}_i(x_t^i))$, where Φ is the cdf of the unit Gaussian distribution. Finally the temporal causal graph can be uncovered by analysis of dependency among u_t^i using algorithms such as `glasso` algorithm (Friedman, Hastie, and Tibshirani 2008). We report an edge from node i to node j if in the precision matrix has is at least one non-zero element from lagged $u_{t-\ell}^j$ to u_t^i, for $\ell \geq 1$. The method in (Leong and Valdez 2005) is used for prediction of the future values of the time series.

In order to uncover causality relationship among time series we can either estimate the marginals with a non-parametric density estimator or we can fit a particular distribution as marginal distribution. In the time-varying networks the latter is preferred; since the non-parametric approximation of the marginal distributions leads to over-fitting when the number of observations is scarce.

Transfer Entropy Solution Transfer entropy is usually employed when the data do not follow the auto-regressive model and a nonlinear generalization of the Granger causality framework is desirable. In the Transfer entropy framework (Schreiber 2000), time series X is thought to be a cause of another time series Y if the values of X in the past significantly decrease the uncertainty in the future values of

Y given its past. The amount of decrease in the uncertainty can be quantified as

$$T_{X \to Y} = H(Y^t | Y^{t-L:t-1}) - H(Y^t | Y^{t-L:t-1}, X^{t-L:t-1}),$$

where $H(X)$ is the Shannon entropy of the random variable X. Since the transfer entropy is a pairwise quantity, we can use it in a temporal dependency graph learning framework such as IAMB (Tsamardinos, Aliferis, and Statnikov 2003) to uncover the temporal dependency among multiple time series.

6 Expriments

In this section we present two sets of experiments: the first set on the Paleo dataset with irregular sampling times to discover the monsoon climate patterns in Asia. The second set of experiments are done on two sets of climate and social media datasets. In all the experiments, we use implementation of Lasso in GLMnet package (Friedman, Hastie, and Tibshirani 2010) and tune the penalization parameter of Lasso via AIC (Akaike 1974).

Performance of GLG

Now we apply our method to a Climate dataset to discover the weather movement patterns. Climate scientist usually rely on models with enormous number of parameters that are needed to be measured. The alternative approach is the data-centric approach which attempts to find the patterns in the observations.

The Paleo dataset which is studied in this paper is the collection of density of $\delta^{18}O$, a radio-active isotope of Oxygen, in four caves across China and India, see Figure 1. The inter-sampling time varies from high resolution 0.5 ± 0.35 to low resolution 7.79 ± 9.79; however there is no large gap between the measurement times. The density of $\delta^{18}O$ in all the datasets is linked to the amount of precipitation which is affected by the Asian monsoon system during the measurement period. Asian monsoon system, depicted in Figure 1, affects a large share of world's population by transporting moisture across the continent. The movement of monsoonal air masses can be discovered by analysis of their $\delta^{18}O$ trace.

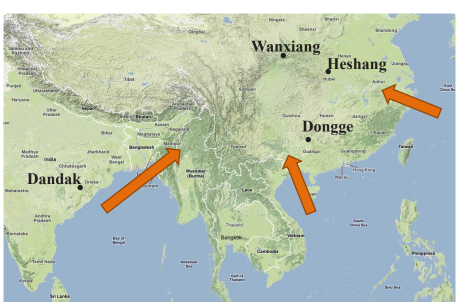

Figure 1: Map of the locations and the monsoon systems in Asia.

In order to analyze the spatial transportation of the moisture we normalize all the datasets by subtracting the mean and divide them by their standard deviation. We use GLG with the Gaussian kernel with bandwidth equal to $0.5(y)$ and maximum lag of $25(y)$; i.e. $L = 50$. In order to compare our results with the results produced by the slotting method in (Rehfeld et al. 2011) we analyze the spatial relationship among the locations in three age intervals. Figure 2 compares the graphs produced by GLG with the ones reported by (Rehfeld et al. 2011).

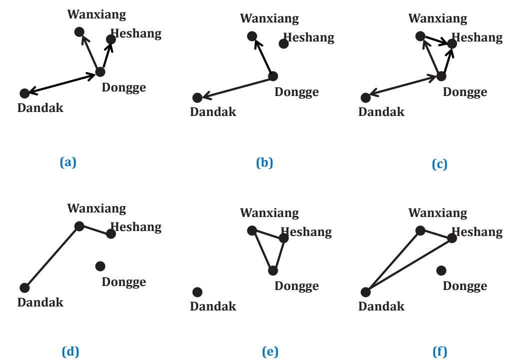

Figure 2: Comparison of the results on the Paleo Dataset: (a) GLG in period 850AD-1563AD. (b) GLG in the period 1250AD-1564AD. (c) GLG in the period 850AD-1250AD. (d) Slotting technique in period 850AD-1563AD. (e) Slotting technique in the period 1250AD-1564AD. (f) Slotting technique in the period 850AD-1250AD.

Results on the Paleo Dataset Figure 2 parts (a) and (d) show the results of influence analysis with GLG and slotting technique, respectively. Our results identify two main transportation patterns. First, the edges from Dongge to other locations which can be interpreted as the effect of movement of air masses from southern China to other regions via the East Asian Monsoon System (EAMS). Second, an edge from Dandak to Dongge which shows the Indian Monsoon System (IMS) significantly affects Dongge in southern China. The graph in the period 1250AD-1563AD is sparser than the graph in 850AD-1250AD which can be due to the fact that the former age period is a cold period, in which air masses do not have enough energy to move from India to China, while in contrast the latter age period is a warm phase and the air masses initiated in India impact southern China regions. During the warm period we can see that other branches of EAMS are also more active which result in denser graph in the warm period. The differences between our results and the results from Slotting technique can be because of the fact that in the Slotting technique an edge is positively identified even if two time series have significant correlation at zero lag. However, by the definition of Granger causality, only past values of one time series should help prediction of the other one in order to be considered as a cause of it.

Non-linear Causation

The study of extreme value of wind speed and gust speed is of great interest to the climate scientists and wind power engineers. A collection of wind observations is provided by

AWS Convergence Technologies, Inc. of Germantown, MD. It consists of the observations of surface wind speed (mph) and gust speed (mph) every five minutes. We choose 153 weather stations located on a grid laying in the $35N - 50N$ and $70W - 90W$ block. Following the traditions in this domain, we generated extreme value time series observations, i.e, daily maximum values, at different weather stations. The objective is examine how the wind speed (or gust speed) at different locations affects each other and how well we can make predictions on future wind speed. It has been shown that the daily maximum values follow the extreme value distribution which is very different from the Gaussian distribution. Thus, this dataset can be used for testing the performance of the algorithms in non-linear datasets.

Prediction Performance In the prediction task, we conduct experiments via the sliding window approach: given time series observations of length T and a window size S, we train a model on observations of $x_s, \ldots, x_{T-S+s-1}$ and test it on the $(T - S + s)^{th}$ sample, for $s = 1, \ldots, S$. We set S to be 10 for all datasets. For evaluation, we use the root mean squared error (RMSE) measure averaged over S experiments and all nodes.

Table 1: Comparison of RMSE by different algorithms in the prediction tasks.

	Wind	Gust
Granger	.0695	.0943
Transfer entropy	.0692	.0983
Copula	**.0678**	**.0934**

Table 1 shows the prediction accuracy of different algorithms on all datasets. The superior performance of the Copula based approach can be attributed to its ability to capture the non-linear dependencies in the data. Transfer entropy requires many observations to perform well, which could be a potential issue in the real applications.

Graph Learning Performance Figure 3 shows the inferred temporal dependencies from the time series of wind speedby Granger-Copula. Given the limited space, we limit our discussion on the new york region. The main observation is that the weather in the inland regions are heavily influenced by the coastline region. The graph indicates two clusters. One is coastal cluster centered around New York city and another inland cluster centered at Troy, which have strong temporal relationship with neighboring cities confirmed by the climatologists.

7 Conclusion and Future Work

In this paper after establishing the consistency results for the significance tests and Lasso-Granger we described several generalizations to be used in different real applications. The GLG method is designed to uncover the temporal causal dependencies among irregular time series and copula based approach and Transfer Entropy are designed to uncover temporal causal relationships in non-linear datasets.

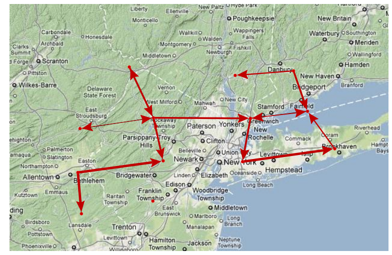

Figure 3: The temporal dependence graph learned by Granger-Copula on the time series of Wind in NY. Thicker edges imply stronger dependency.

This work is going to be extended in several directions: (i) Identification and compensation of the spurious causal effects due to unobserved variables in Granger networks, (ii) Identification of the effects of hidden variables in Granger networks and (iii) Design of efficient non-parametric tests for discovering causality in non-linear systems are some of the exciting future directions to pursue.

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*.

Arnold, A.; Liu, Y.; and Abe, N. Temporal causal modeling with graphical granger methods. In *KDD '07*.

Asimakopoulos, I.; Ayling, D.; and Mansor Mahmood, W. 2000. Non-linear Granger causality in the currency futures returns. *Economics Letters* 68(1):25–30.

Bahadori, M. T., and Liu, Y. 2012. Granger causality analysis in irregural time series. In *SDM'12*.

Beck, A., and Teboulle, M. 2009. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences* 2(1):183.

Brovelli, A.; Ding, M.; Ledberg, A.; Chen, Y.; Nakamura, R.; and Bressler, S. L. 2004. Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *the National Academy of Sciences of the USA*.

Embrechts, P.; Mcneil, A.; and Straumann, D. 2002. Correlation and dependence in risk management: properties and pitfalls. In Dempster, M. H. A., ed., *Risk Management: Value at Risk and Beyond*. Cambridge: Cambridge University Press.

Friedman, J.; Hastie, T.; and Tibshirani, R. 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)* 9(3):432–41.

Friedman, J. H.; Hastie, T.; and Tibshirani, R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*.

Granger, C. W. J. 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37(3):424–438.

Granger, C. W. J. 1980. Testing for causality: A personal viewpoint. *Jour. of Econ. Dyn. and Contr.*

Hiemstra, C., and Jones, J. D. 1994. Testing for Linear and Nonlinear Granger Causality in the Stock Price- Volume Relation. *The Journal of Finance* 49.

Leong, Y. K., and Valdez, E. A. 2005. Using, Claims Prediction with Dependence Models, Copula. *Insurance: Mathematics and Economics*.

Liu, Y.; Bahadori, M. T.; and Li, H. 2012. Sparse-gev: Sparse latent space model for multivariate extreme value time series modeling. In *ICML'12*.

Liu, H.; Lafferty, J. D.; and Wasserman, L. A. 2009. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *JMLR*.

Marinazzo, D.; Pellicoro, M.; and Stramaglia, S. 2008. Kernel Granger causality and the analysis of dynamical networks. *Physical review. E*.

Meinshausen, N., and Bühlmann, P. 2006. High-Dimensional Graphs and Variable Selection with the Lasso. *The Annals of Statistics* 34(3):1436–1462.

Meinshausen, N., and Yu, B. 2009. Lasso-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics* 37(1):246–270.

Panchenko, C. D., and Valentyn. 2004. Modified hiemstra-jones test for granger non-causality. Computing in economics and finance.

Rehfeld, K.; Marwan, N.; Heitzig, J.; and Kurths, J. 2011. Comparison of correlation analysis techniques for irregularly sampled time series. *Nonlinear Processes in Geophysics* 18(3):389–404.

Schreiber, T. 2000. Measuring Information Transfer. *Physical Review Letters* 85(2):461–464.

Tsamardinos, I.; Aliferis, C. F.; and Statnikov, A. 2003. Algorithms for Large Scale Markov Blanket Discovery. In *FLAKE*.

Valdés-Sosa, P. A.; Sánchez-Bornot, J. M.; Lage-Castellanos, A.; Vega-Hernández, M.; Bosch-Bayard, J.; Melie-García, L.; and Canales-Rodríguez, E. 2005. Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 360(1457):969–81.