Master's Thesis

Applying Causal Discovery to Financial Data

Luigi Bassani Antivari

Applying Causal Discovery to Financial Data

Master's Thesis

at Frankfurt School of Finance and Management

Supervised by

Prof. Dr. Grigory Vilkov

Prof. Dr. Florian Ellsäßer

Submitted

by Luigi Bassani Antivari

Master of Science in Finance

8467249

Frankfurt am Main, Adalbert Strasse 44-48

+49-176-812-88-883

luigi.bassani_antivari@fs-students.de

Frankfurt am Main, September 2023

Abstract

Causal discovery, the task of identifying the causal relationships between variables, has been a long-standing challenge for many domains such as climate science, econometrics, and healthcare. One of the most advanced causal discovery methods is the PCMCI+, which uses conditional independence tests to reconstruct the causal relationships between time series and predict future values. Despite the changing nature of real-world relationships, PCMCI+ assumes constant causal relationships through time. This assumption is shared among most causal discovery methods and makes them inapplicable to predict financial markets, which are known to be a highly evolving dynamical system. To model the time-varying causal relationships of financial data we propose EVO-PCMCI+, a novel multivariate forecasting framework which iterates a variation of the PCMCI+ combined with various predictive functions. We evaluated EVO-PCMCI+ using synthetic and real-world datasets. The empirical results demonstrate the effectiveness of the proposed method in both capturing time-varying causal relationships as well as forecasting with discrete accuracy short-term financial data.

Contents

1. Introduction

The main objective of this paper is to forecast non-stationary and autocorrelated financial data by capturing time-varying causal relationships between variables. The pursuit of causality-driven predictions holds substantial importance due to its potential to unravel the underlying mechanisms governing intricate systems (Eichler, 2013). In the realm of time series forecasting, the accurate prediction of various variables from observational data, particularly through the lens of causal relationships, has posed an enduring challenge across disparate domains. Similarly to metereology, forecasting financial time-series data is notoriously challenging due to the intricacies and dynamism of real-world relationships (Nauta, 2019). However, current causal predictive techniques fall short in capturing the evolving causal relationships of many complex temporal systems. Founded on the iteration of existing graphical causal models, we propose an innovative multivariate forecasting framework that detects time-varying causal relationships and applies them for time series prediction.

What are the causal relationships between currencies? What causes the water level of a river to rise? Machine learning models and deep learning methods have reached outstanding predictive capabilities, however, they are not able to answer these causal questions Similarly to traditional statistical methods, the main predictive factor of machine learning methods is correlation, but correlation does not imply causation. (Soo, 2020). Being an asymmetrical measure, correlation does not inform on the directionality of the relationship, namely if variable A causes B or vice versa. Even when two time series show strong correlation, it does not necessarily indicate a causal relationship between them . An illustrative example is the correlation between the number of storks and the number of births, with a p-value of 0.008 (Matthews, 2000). This p-value suggests that there is only a 1 in 125 chance that this correlation is spurious, seemingly implying that storks deliver babies. The use of correlation lead to other similar misconceptions, such that the lack of pirates causes global warming or that sending more firefighters to a fire results in more damage (Vigen, 2015). Because we have at least partial knowledge of the physical causes behind the phenomenon of birth, global warming, and fire extinction, it is evident that these are examples of spurious correlations. On the other hand, detecting spurious correlations from large datasets whose underlying causal relationships are unknown is a way more challenging task (Moraffah, 2021). Undetected spurious correlation is one of the main causes that hinders the accuracy of statistical forecasting models. In the financial industry, trading strategies that are based purely on statistical measures are known to have short life if they are not repeatedly adjusted (Börjesson, 2020). In most of cases where stability of correlation is not enforced, predictive models based on correlation cannot guarantee robust relationships, making it impossible to predict when they will stop working (Eichler, 2013). In order to distinguish spurious from non-spurious correlation, researchers proposed various approaches to uncover the underlying causal mechanisms governing variables' relationships (Holland, 1986).

Causality, as opposed to correlation, is a symmetrical measure that precisely aims at describing directional causal relationships between variables. One of the first causality measures ever developed is Granger causality which is based on the idea that if variable A helps improving the prediction of variable B, then A Granger-causes B (Shojaie & Fox, 2021). In addition to Granger causality, several other causal frameworks have been proposed, such as transfer entropy, graph theory based methods, and dynamic Bayesian networks. Together, these methods form the discipline referred as causal inference.

Causal inference methods facilitate the identification of causality between variables in complex dynamical systems and explore the physical mechanisms (Kretschmer et al. 2016). Current causal inference methods have solid theoretical basis and are increasingly applied to multiple disciplines, such as meteorology and ecology field (Sugihara et al. 2012) . Being first applied for causal treatment effect estimation to improve the results of randomized controlled trials, causality is now being used to study observational data thanks to latest advancements in causality-driven methods. The field of causal inference can be divided into two main areas: casual treatment effect estimation and causal discovery.

In this paper we focus on causal discovery, which is defined as the process of inferring causal relationships between observed time series variables. The goal of causal discovery methods is to derive a causal model that describes a given dataset by identifying the causal structure underlying the dataset. Among the wide range of causal discovery methods developed over time, most are limited to qualitative causal knowledge exploration and only a minority are designed for time series forecasting (Moraffah, 2021). One of the methods that allows for forecasting is the Conditional Selection and

Momentary Conditional Independence (PCMCI) algorithm, proposed by Runge (Runge, 2019).

The PCMCI framework is a graphical causal discovery model designed to address high-dimensionality and non-linearity of time series data. Runge also developed the PCMCI+, which in addition to the features of the PCMCI allows for better accommodation of autocorrelated time series (Runge, 2022). As they are, both PCMCI and PCMCI+ assume the underlying causal structure to stay constant over time. This assumption is often referred as causal stationarity and is shared by most causal discovery frameworks (Glymour, 2019). The application of models that assume causal stationarity is inappropriate for forecasting financial data, whose underlying causal structure is commonly assumed to evolve over time (Ravivanpong, 2022). Because financial markets are a highly evolving system, the analysis of their time-series demands for causal discovery methods that model the temporal evolution of causal relationships.

In response to this challenge, we introduce a PCMCI-based multivariate forecasting framework that aims at predicting non-stationary financial time series by detecting the underlying evolving causal patterns. To detect evolving causal structures and use those to make predictions, our approach iterates over time the combination of a variation of the PCMCI+ predictive method with various linear and machine learning forecasting models. We call our causality-driven forecasting framework EVO-PCMCI+. The overarching objective of this paper is to propose a seminal approach to modelling non-stationary time-series with evolving causal relationships. By presenting the capabilities and the limits of our approach, we hope to introduce causality as an informative data analysis tool for financial data.

2. Background

In this section, first we provide an overview of the concepts of causality, causal inference, and causal discovery, while mentioning some of their successful real-world applications. Second, we present the basics of graphical causal models covering some of the data challenges they aim to solve. Third, we introduce the PCMCI causal discovery method and the graphical causal method on which it is based, the PC algorithm. The wide range of causal discovery and forecasting methods will be discussed in the literature review section. Finally, we introduce EVO-PCMCI+, our proposed multivariate causal forecasting framework.

2.1 Causality

The search for ways to measure causality is a long standing quest that has been approached from different perspectives. Researches proposed various methods to identify causality from various types of data (Nogueira, 2022). To illustrate the concept of causality we mention here two of the most established and used approaches: Granger causality and Transfer entropy. The statistical concept of Granger causality represents one of the first attempt to quantify causality and suggests that "A Granger-causes B" if the prediction of B given the past of B is worse than the prediction of B given the past of both B and A. Born as a measure of pairwise causality, Granger causality has been extended in several ways (Shojaie & Fox, 2021). One of such extensions is the Multivariate Granger Causality (MVGC) which has been successfully applied in macroeconomics and climate sciences (Barnett & Seth, 2014).

Comparable to Granger causality, Transfer entropy is another of the most used approaches that aim at detecting causality. Transfer entropy is an information-theoretic measure that identifies causality by assessing whether information about the past of A can reduce the uncertainty about the future of B. Transfer entropy advances Granger causality by removing the need of predicting variable B. Since its introduction, Transfer entropy has been applied in various fields, including neurosciences, physiology, climatology, finance, and social sciences (Duan, 2021). In addition to Granger Causality and Transfer entropy, the range of causal frameworks is wide and will be discussed in more detail in the literature review section.

2.2 Causal Inference

Built on the various measures of causality and methods to search for it, causal inference is defined as a framework that integrates statistical and machine learning methods to identify the causal relationships between variables (Nogueira et al., 2022). Causal inference transcends the limitations of mere statistical correlation, striving to unravel the underlying mechanisms that govern the relationships between variables (Holland, 1986). In earth sciences, causal inference has unveiled the interconnectedness of geological processes, elucidating how seismic activity in one region can trigger tsunamis in another (Beroza & Ide, 2011).

Causal inference saw its first applications in medical research where is commonplace that statistical methods based on correlation are not able to give a definite answer on whether a new medicine is beneficial of not to cure a certain disease. In front of these challenges, researchers developed accurate predictive methods based on causality measures to estimate the effect of a medication on a set of the patient's health variables. These methods are called causal treatment effect estimation methods and have been proven to reconstruct underlying causal relationships that accurately describe the actual mechanics behind complex dynamical systems (Malinsky & Spirtes, 2018).

Causal treatment effect estimation methods are performed on experimental data, a type of time series data where the value of one or more variables are directly manipulated by the researchers to test their effects on the other variables. In experimental settings, randomized controlled trials are employed to isolate the effect of a treatment or intervention variable, thus enabling accurate assessment of the causal impact on the outcome. This approach, exemplified by Rubin's potential outcomes framework, quantifies the difference between the potential outcomes with and without the treatment. In healthcare, causal treatment effect estimation has transformed treatment strategies by revealing the true impacts of medical interventions on patient outcomes, facilitating evidence-based medical decisions (Rubin, 2005). In essence, causal treatment effect estimation methods focus on establishing the causal influence of a pre-defined intervention on an outcome variable.

2.3 Causal Discovery

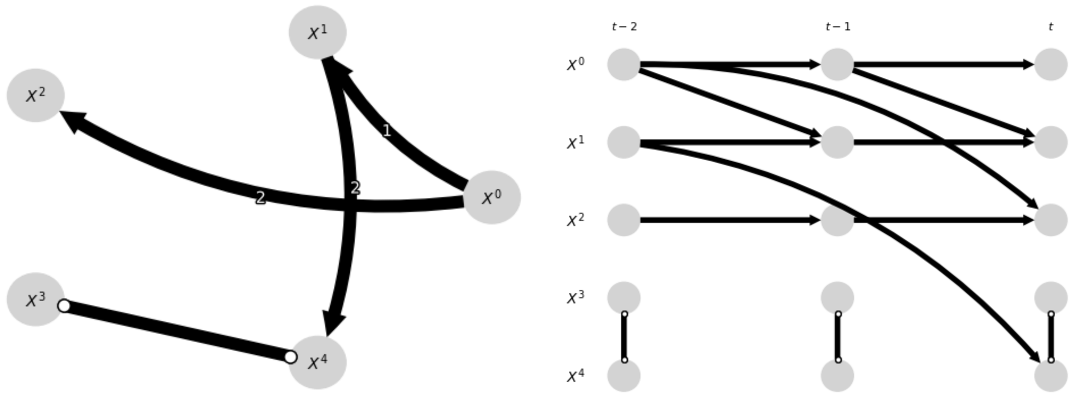
In many fields, such as climate science and capital markets, an experimental approach to learning causal relations is infeasible owing to ethical, practical, or economic reasons (Shanmugam, 2018). For instance, it is infeasible to artificially modify the temperature of the tropical area to test how this affects tropical winds. Similarly to climate science, in capital markets would be difficult to manipulate the value of financial indices such as the S&P500 or the interest rates just for experimental purposes. Regarding interest rates, one could argue that the national fixed income indices represent a "treatable" dataset from the perspective of central banks, where the variable of monetary policy corresponds to the tested treatment, but we will reserve such an hypothesis for future research. As opposed to experimental data, "untreatable" data is called observational data and is defined to contain no time series variables that have been explicitly manipulated. The set of assumptions that build the theoretical foundations of causal treatment effect estimations methods makes them inapplicable to observational data. Inspired by causal treatment methods, researchers developed various frameworks to perform causal inference on observational time series dataset, called causal discovery methods.

Causal discovery methods are designed to identify the underlying causal structure from observed time series, and together with causal treatment effect estimation methods, they define the two main fields of causal inference (Glymour at al., 2019). Causal discovery methods have been successfully applied to observational time series data to infer causal relationships in earth science and neurology. In the financial landscape, causal discovery techniques have delineated the intricate web of relationships between economic indicators, enhancing our ability to forecast market trends and mitigate systemic risks (Durbin & Koopman, 2012).

Once the variables under examination has been determined, causal discovery methods aim at identifying which variables cause the other variables to change. Among the many challenges of causal discovery, identifying time-varying causal relationships has been approached only by a few methods, each of them suffering of specific limitations. For most causal discovery methods, the assumption of causal stationarity, namely that the underlying causal structure does not change through time, is a fundamental element of their theoretical foundations (Jiang, 2017). The capability to identify evolving causal structures would extend the applicability of causal discovery methods to many real-world complex systems.

2.4 Graphical Causal Models

To structure and understand the causal relationships between variables, many causal discovery methods are framed using graphical causal models (Glymour, 2019). Graphical causal models owe their origins to Graph Theory, which can be defined as the study of connections between multi-dimensional points (Harary, 1969). Graphical models map the examined set of variables and the set of their assumed causal relationships respectively into the nodes and the links of the discovered causal graph (Runge, 2023). A causal graph, or more generally a graph, is a mathematical structure used to model pairwise relations between multiple objects. Visually, a graph consists of a set of points which some pairs are connected by a link and others are not. The mathematical foundations of most graphical causal discovery methods requires them to adopt a specific type of graph, called Directed Acyclical Graph (DAG). DAGs are graphs whose links are directed, meaning that links have an arrow-headed shape indicating the direction of the relationship, and they are acyclical, meaning that do not have any path of sequentially adjacent links that creates a cycle. The reason for causal discovery methods for assuming graph acyclicality is often to simplify many calculations while ensuring the theoretical validity of the model (Runge, 2023). Withing causal DAGs, every causal relationship is represented by a directed link from the “causing” variable to the “caused” variable, which are often referred as causal parent and causal child. Graphical causal models share the same mathematical foundations of structural equation models (SEM), also called functional causal models (FCM), and they can be interpreted as general systems of equations whose functions can include and are not limited to regression models, factor models, ARM time series models, latent class models, and other functions (Glymour, 2019). Because different causal models assume different causal functions, the identification of causality is not unique but varies from method to method. More specifically, while some methods aim at quantifying deterministic causal functions, such as Granger causality and Transfer entropy, other methods do not quantify causal relationships but only aim at identifying their presence and their directionality, the PCMCI causal discovery method is one of them. Among the most representative graphical causal discovery frameworks we find constraint-based casual methods and score-based casual methods.



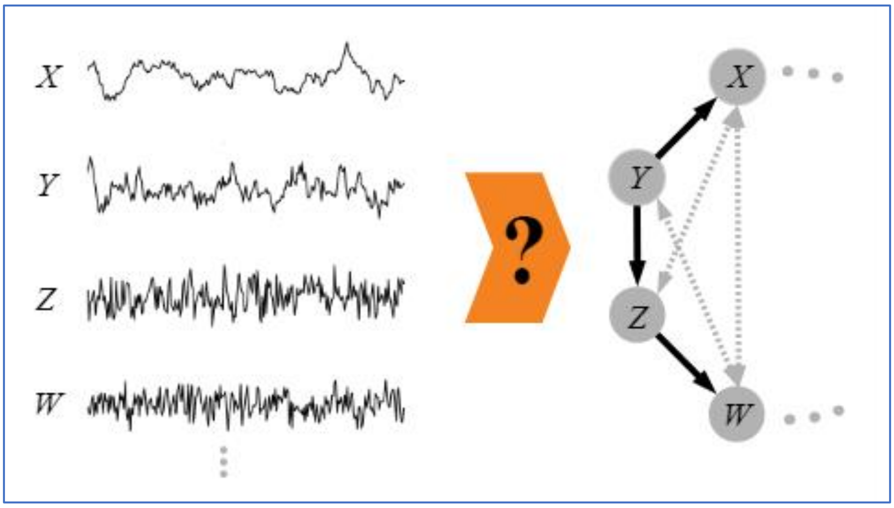


Figure 1: Shows the causal discovery task of identifying causal relationships from observed time series (TO BE REPLACED WITH YOUR GRAPHS!!)

Constraint-based causal methods integrate graphical and statistical techniques to investigate the presence and the directionality of causal relationships between variables (Glymour, 2019). The way in which these methods identify causal relationships follows a set of constraints that derive directly from the graphical study of spurious associations between multiple variables. Constraint-based causal methods start from assuming that each variable is a potential cause of all other variables, which in graphical terms refers to a fully connected graph whose every pair of variables is connected by bidirected link. Next, they test all potential causal relationships and only the ones that meet the predetermined causal constraints will be included in the finally discovered causal graph.

One the other hand, score-base causal methos usually rank the potential causal relationships or causal graphs based on a score and select the ones with the highest score (Moraffah, 2021). Many current graphical causal models implement a combination of constraint-based and score-based methods, with some of them leveraging machine learning methods as well. More generally, graphical causal models represent a flexible, interpretable, and visually intuitive framework that has been adapted to model a wide range of assumptions and causal functions.

2.5 PC Algorithm

As one of the first and most used graphical causal methods, the PC algorithm takes the name of its inventor, Peter and Clark, and represents the backbone of many advanced causal discovery methods (Spirtes, at al., 1993). The PC algorithm is a constraint-based causal discovery method that aims at reconstructing the underlying causal graph given multiple observed time-series. The PC algorithm is based on a series of conditional independence tests and is structured in two steps.

First, the PC algorithm systematically tests different combinations of variables' pairs to determine their independence conditioned on the set of other variables. When the relationship between two variables does not meet the predetermined causal constraints, the relationship is removed from the set of possible causal relationships. Second, once the potential causal relationships have been selected, the PC algorithm implements a causal discovery process to find the orientation of the assumed causal links. This process integrates conditional independence tests with a set of rules that directly derive from the concept of d-separation (Glymour, 2019). D-separation is a concept used to distinguish spurious correlations in the multi-variate case that defines the conditional independence relationships between variables based on specific graphical structures. Chains, forks, and colliders are a few examples of these graphical structures. A chain occurs when three variables are connected in a linear sequence, such as $A \rightarrow B \rightarrow C$, and indicates that A and C are conditionally independent given B. A fork is represented by a variable point toward other two variables, such as $A \leftarrow C \rightarrow B$, and also indicates that A and C are conditionally independent given B. On the other hand, a collider is defined when two variables point toward the same variable, such as $A \rightarrow C \leftarrow B$, and indicates that A and B are independent but become conditionally dependent given C. In addition to these rules, d-separation applies other similar criteria based on more complex graphical structures (Wahl, 2023), however, an overview of such rules is beyond the scope of this paper. Based on these rules, the PC algorithm is able to estimate the orientation of the causal relationships and reconstruct the underlying causal structure.

Even though the original PC algorithm was proposed for general random variables without time order, the algorithm has been vastly extended to model scenarios where variables is a time series of values. One of these time series extensions of the PC algorithm is the PCMCI casual discovery method.

2.6 PCMCI Casual Discovery Method

Inspired by the PC algorithm, Runge proposed the PCMCI causal discovery method that aims to solve the challenges of high-dimensionality and nonlinearity of time series data (Runge, 2019). The fundamental process implemented by the PCMCI consists of two steps. First, the PCMCI applies the PC algorithm or other variations to identify the potential temporally-lagged causal parents of each variable. Second, it applies momentary conditional independence (MCI) tests to select the estimated causal parents from the potential causal parents. Because the PCMCI causal discovery method represents the causal discovery engine of our framework, a more detailed explanation of its functioning is presented in the data and methodology section.

Successful real-world applications of the PCMCI include climate science, where it allowed to detect causal relationships between surface air temperature anomalies, and cardiovascular analysis, where it discovered the causal structure between heart beat frequency, diastolic blood pressure, and systolic blood pressure (Runge, 2023). To address different types of data challenges, Runge advanced the PCMCI in several ways. One these advancement is the PCMCI+, which is better suited for autocorrelated time series (Runge, 2022). On the other hand, the LPMCI algorithm is described as a high-recall causal discovery method designed for time series with latent confounders (Gerhardus & Runge, 2021). Said differently, the LPMCI allows to drop the causal sufficiency assumption by modelling and discovering the effects of causal parents that are external to the set of variables analysed. The PCMCI causal discovery method as well as all of its current variations are based on the assumption of causal stationarity, namely that the causal relationships are assumed to stay constant over time.

2.7 Our Framework

One limitation of existing causal discovery methods is their assumption of stationary causal relationships over time, which does not reflect the dynamic nature of real-world systems. While the assumption of causal stationarity might be acceptable for some climate, earth or biological studies, it cannot hold true when studying financial markets, as they are known to be a highly evolving dynamical system (Ravivanpong, 2022). For instance, if causal relationships between global currencies exist, they can be expected to change over time due to external factors, such as technical innovations, geopolitical tensions, or new regulations. Given that forecasting models for financial data are generally not able to predict such precise external events, they should at least attempt to detect when the causal structure has changed and reconstruct the newly formed causal graph, which is what our proposed methods aims to achieve.

To address the challenge of causal non-stationarity of financial time series, we introduce EVO-PCMCI+, our proposed multivariate causal forecasting framework based on the PCMCI+ causal discovery method that aims at detecting evolving causal relationships between variables and provide short-term predictions. We structure the problem like Runge, however, our framework distinguishes itself from the standard PCMCI+ forecasting method for two main reasons. First, while Runge proposes to use the causal graph descriptive of the whole training dataset as the graph predictor, we use the most recent causal graph estimated by the sliding window PCMCI as we believe it to be more representative for short term predictions. Second, instead of predicting the target variables for multiple time steps in the future, we predict only one step into the future and reiterate the whole process for every time step by shifting the training set to include new available data. In the forecasting phase our framework tests three different types of predictive functions: linear regression, gaussian process regression, and KNeighbours regression. To sum up, our proposed iterative causal discovery approach is designed to learn the evolving causal relationships and apply them for short-term time series forecasting of non-stationary financial data.

We put three research questions into this paper and we will try to analyze them thoroughly. First, how can we discover evolving causal relationships from non-stationary auto-correlated time series? Second, How can we apply the learned evolving causal structures to short-term time series forecasting? Third, how accurate is the proposed framework at forecasting financial data?

Our motivation is to provide empirical evidence by analysing multivariate time-series forecasting of financial data in a new frontier. Through a combination of synthetic experiments and real-world benchmarking, we empirically validate the performance of our framework against pertinent baseline methods. The results obtained provide insights into the framework's efficacy in both causal discovery and multivariate time series forecasting, highlighting its potential applicability across diverse financial domains.

Our objectives are the following: we investigate the causal discovery and forecasting capabilities of applying our proposed method to financial time-series data. We apply a sliding window approach of the PCMCI causal discovery method as presented by Runge (2021?). We also modify the original algorithm to accommodate iterations as well as the use of the most recent causal graph as graph predictor. Further, we apply various regressors on the learned causal knowledge for short-term time series prediction. Finally we evaluate our predictions with current benchmark forecasting models.

The remainder of the paper is structured as follows. The second section of the paper summarizes past and current developments in the literature. The third section presents data and methodology and introduces training and testing principles. The fourth section presents the causal discovery and forecasting results. The fifth section discusses challenges and remarks relevant to our proposed framework. The sixth section summarizes findings and concludes the study with future recommendations.

3. Literature Review

3.1 Forecasting financial data

Predicting financial time-series, such as stock prices or market indices, is a complex and challenging task. Academics have long debated whether is possible to predict financial time-series from historical returns. While some studies have suggested the presence of short-term patterns and trends in financial data, the Efficient Market Hypothesis (EMH) states that asset prices fully reflect all available information, making them essentially unpredictable. One concept that counters the concept of EMH is the delay in the incorporation of new information into asset prices. This phenomenon is often referred to as "information lag" or "information diffusion" (Huang, 2008). Information lag suggests that financial markets do not instantaneously and fully reflect all available information, therefore advocating for a weaker form of EMH. Starting from an information being observed and expressed, a series of physical steps is required before the information becomes “available”, as stated by the EMH. Even if current high-frequency trading strategies are able to detect market inefficiencies within microseconds, this does not mean that all relevant information for asset prices is transmitted at that speed. Financial reports and many macroeconomic indicators are made “available” with a frequency of multiple months. Even at the moment of the publication, the respective asset prices often fluctuate and do not agree on a single price. This is due to the fact that asset prices are the recorded results of a monetary debate between various traders around the world. Even if all these traders are rational and share the same “available” information, which is unlikely, they all have different weights, different incentives, and different interpretations of the same information that come together to shape the momentary debate that defines asset pricing. Many trading strategies aim to understand patterns and behaviours of how these monetary debates are conducted. For instance, high-frequency trading strategies search for prices' patterns on a very-short term temporal range.

Many financial forecasting models implicitly leverage the concept of information lag to predicting the effect that the not yet transmitted information will have on asset prices. Traditional statistical forecasting methods as well as most of deep and machine learning forecasting methods are based on correlation but do not distinguish between spurious and non-spurious cases (Nauta, 2019). On the contrary, causal discovery forecasting methods aim precisely at distinguishing spurious associations from actual causal relationships and use the latter to predict future variables. Temporal causal discovery forecasting methods have found applications in different domains and have shown promising results in forecasting time-series data (Nogueira, 2022). However, current causal discovery time-series forecasting models often assume constant causal relationships over time, neglecting the evolving nature of real-world relationships. This literature review explores existing time series forecasting methods, leading to the proposal of EVO-PCMCI+, a novel multivariate time series forecasting framework that detects evolving causal structures based on to the iterative application of the PCMCI+ causal discovery method. This approach explicitly considers time-varying causal relationships, addressing causal non-stationarity inherent of financial data.

3.2 Traditional statistical forecasting methods

Traditional statistical methods have long been employed to analyse time series data and predict future values from past observations. Statistical forecasting methods such as ARIMA, VAR, and GARCH are among the most commonly used predictive methods in the financial industry (Makridakis at al.,2018). The Autoregressive Integrated Moving Average (ARIMA) approach consists of a linear univariate regressive model that allows for non-stationary time series data. Vector Autoregression (VAR) is the multivariate extension of the AR model. Thanks to its versatility, this approach has found applications in macroeconomic forecasting, financial markets analysis, and climate modeling. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, on the other hand, are designed to account for time-varying volatility and heteroskedasticity of financial time series. GARCH models are widely used in risk management, portfolio optimization, and option pricing thanks to their ability to accommodate changing levels of volatility. These statistical methods represent the backbone of several other forecasting frameworks that have been developed over time, however, as they are, they do not account for the possibility of time-varying relationships. Moreover, being essentially founded on the study of correlation, statistical methods do not allow to answer causal questions from time series data.

3.3 Machine and deep learning forecasting methods

In addition to traditional methods, machine and deep learning methods have been increasingly applied to time series forecasting, with recurrent neural networks (RNNs) and long short-term memory (LSTM) networks being the most popular. RNNs are a class of neural networks that can process sequential data by maintaining a hidden state that captures information about the past inputs. LSTMs are a type of RNN that can selectively remember or forget information from the past, allowing them to distinguish short-term from long-term dependencies. Gated Recurrent Units (GRUs) are another type of RNN that have been used for time series forecasting. GRUs are similar to LSTMs but have fewer parameters, making them faster to train (Lai, 2018).

Convolutional neural networks (CNNs) have also been applied to time series forecasting, but are less commonly used than RNNs and LSTMs. LSTNet is a recent deep learning architecture that combines the strengths of CNNs and LSTMs for time series forecasting (Lai at al., 2017) . Attention-mechanism based forecasting models are another type of deep learning model that have been used for time series forecasting. These models use an attention mechanism to focus on the most relevant parts of the input sequence when making predictions.

Despite the high computational costs, challenges of interpretability, and the large amount of data required, deep learning methods have shown such promising results that they allowed to advance the study of numerous fields. Statistical and machine learning methods represent the current reference of time series forecasting methods, however, most of them are founded on the study of correlation alone, which makes them unable to answer the causal questions posed in this paper. Machine learning methods that are built around the measure of causality exist and we will discuss these frameworks later.

3.4 Causal discovery methods

Causal discovery has been recently applied for time series forecasting tasks. With the objective to understand the causal relationships between observed time series, researchers developed several temporal causal discovery methods. Causal discovery methods for time series data can be classified in the following overlapping categories: asymmetry-based causal models, restricted functional causal models, nonlinear state-space methods, graphical causal models, constraints based causal methods, score based causal methods, hybrid causal machine learning forecasting models, other hybrid methods: state-space methods, dynamic Bayesian networks, and Gaussian processes.

Asymmetry-based causal models

Asymmetry-based causal methods represent the majority of temporal causal frameworks and they include Granger causality and Transfer entropy (Glymour, 2019). Asymmetry-based causal methods are founded on the concept of causality's temporal precedence and they assume that the cause must temporally precede its effects. This assumption is useful as it allows to limit the set of potential causal parent of each temporal target variable to the temporal variables that precede the specific target variable. Granger causality is based on the idea that variable A is the causal parent of variable B if the past of variable A enhances the prediction of variable B. Multivariate Granger Causality (MVGC) is an extension of the Granger causality test for the multivariate time series case, however, both versions assumes linear and stationary relationships (Barnett & Seth, 2014). On the other hand, Transfer entropy search for causality by assessing whether information about the past of A can reduce the uncertainty about the future of B (Duan, 2021).

Restricted functional causal models

Restricted functional causal models aim to define specific causal function between variables. The most prominent restricted functional causal models is the Linear Non-Gaussian Acyclic Model (LiNGAM) which models causal relationships as linear non-Gaussian causal functions (Shimizu at al., 2006). The Vector Autoregressive Linear Non-Gaussian Acyclic Model (VAR-LiNGAM) algorithm extends the LiNGAM framework for multivariate time series data (Shojaie & Fox, 2021). The more advanced Just-In-Time Linear Non-Gaussian Acyclic Model (JIT-LiNGAM) aims to detect slowly evolving causal structures by iterating the LiNGAM method each time new data is available (Fujiwara, 2023). Because these methods assume linear causal functions, they are not ideal to forecast financial data, which can be assumed to be governed by non-liner relationships. On the other hand, Nonlinear State-space methods aim at capture the evolving causal structure over time by modelling distinct systems of causal functions at different states (Runge, 2023). However, their successful implementation necessitates a profound understanding of the system's functions and states, making them somewhat reliant on domain knowledge.

Constraints based causal methods

Constraint-based causal methods leverage a set of constraints derived from the graphical study of spurious associations among multiple variables. Constraints-based causal frameworks include the PC algorithm, the PCMCI framework, the Fast Causal Inference (FCI) algorithm, and the Causal Discovery from Nonstationary/heterogeneous Data (CD-NOD) algorithm (Slavov, 2022). The PC algorithm represents the first attempt to identify causal relationships by systematically testing conditional independence between variables. Because the PC algorithm does not assume specific causal functions but only aims at detecting the presence and orientation of causal relationships, the PC is able to account for non-linear causal relationships. Proposed by Runge, the PCMCI advances the PC algorithm by significantly decreasing its computational cost via a constraint-based process that is more efficient at removing non-causal parents (Runge, 2019). In its simplest form, the PCMCI allows to discover non-linear causal structures from high-dimensional time series datasets. The PCMCI has been extended in various ways: the PCMCI+ is designed for autocorrelated time series, the Latent-PCMCI (LPCMCI) accommodates the assumption of a latent cofounder (hidden variable (Gerhardus & Runge, 2021)), the Joint-PCMCI (JPCMCI) is designed for multiples datasets with the same variables (system-variables) (Günther & Runge, 2023), and finally the Regime-PCMCI (RPCMCI) accounts for regime-based non-stationary causal structures (Saggioro, 2020).

Another notable constraint-based causal discovery method is Fast Causal Inference (FCI) algorithm, which instead of testing all potential causal links like the PC algorithm, the FCI algorithm iteratively searches for causal links in a reasoned order (Spirtes, 2001). At the cost of overlooking some true causal relationships, the FCI algorithm converges to the discovered causal graph much faster than the PC algorithm. The Time Series Fast Causal Inference (tsFCI) and the Structural Vector Autoregressive Fast Causal Inference (SVAR-FCI) represent two different adaptions of the FCI to the time series case, however, both variations assume fixed linear causal functions through time (Assaad, 2022).

The Causal Discovery from Nonstationary/heterogeneous Data (CD-NOD) algorithm allows to model distributional shifts of the time series, while the Causal Discovery from Autocorrelated and Non-Stationary Time Series Data (CDAN) algorithm extends the CD-NOD to account for autocorrelation (Zhang, 2017).

Temporal regime-based and changing-module-based causal discovery methods for non-stationary time series data, such as the RPCMCI, the CD-NOD, and the CDAN algorithms, allow causal structures to changes only at predetermined points in time. However, researchers studying the financial markets have no way to estimate at what point in time the underlying causal structures will change.

While constraint-based causal methods are effective in detecting time-varying causal relationships from time series data, they may struggle to capture complex dependencies if the form of dependence is not well-suited for the chosen conditional independence tests. In such cases, other causal discovery methods, such as score-based or deep learning extensions, may be more appropriate.

Score based causal methods

Score-based models estimate causal relationships by fitting a causal model to the data and evaluating how well it fits. This approach typically begins with specifying the structural equations describing causal relationships among the variables, often requiring expert knowledge or assumptions about the causal functions governing the dataset. In score-based methods, potential estimated causal graphs are compared based on a score, often representing the probability of such causal graphs to exist (Assaad, 2022).

An exemplary score-based causal method is the Greedy Equivalence Search (GES) algorithm (Chickering, 2020). The GES algorithm starts with an empty graph (no links) and iteratively adds, deletes, and reverses causal links, comparing potential causal graphs based on a score. The scoring include statistical tests or information-theoretic criteria like the Bayesian Information Criterion (BIC) and the Minimum Description Length (MDL). The algorithm converges when no further improvements can be made to the causal graph and the graph with the highest score is selected as the estimated causal graph underlying the dataset.

Other score-based frameworks are the Greedy Randomized Search Procedure (GRaSP) and the Gaussian Structural Perturbation (GSP) algorithm (Lam et al., 2022). While the GRaSP applies a very fast randomized search algorithm to find the causal graph that best fits a high-dimensional dataset, the GSP algorithm applies perturbation to the dataset and estimate the underlying causal relationships by assessing changes in the covariance matrix.

Machine learning based causal models

Some machine learning based causal forecasting models have been recently proposed, showcasing the potential synergies between machine learning and causal discovery. The Transfer Entropy Graph Neural Networks (TEGNN) method leverages graph neural networks to estimate transfer entropy-based causal relationships and predict the time series (Duan, 2021). Another hybrid forecasting method is the Temporal Causal Discovery Framework (TCDF), it adopts attention-based convolutional neural networks combined with a causal validation step (Nauta, 2019). Finally, the Amortized Causal Discovery (ACD) method implements neural networks to capture the shared dynamics across temporally adjacent causal graphs and applies a score-based algorithm to evaluate the quality of the discovered causal structures (Löwe, 2022).

Despite the field of causal discovery and forecasting has significantly expanded in recent years to accommodate disparate temporal challenges, one key limitation of existing causal forecasting models is their inability to effectively detect evolving causal relationships any point in time. More specifically, current causal methods designed to model time-varying causal structures, such as the RPCMCI, the CD-NOD, and the CDAN algorithms, consist of regime-based or changing-module-based frameworks that allow causal structures to change only at predetermined points in time. However, researchers studying the financial markets have no way to estimate at what point in time the underlying causal structures will change. In real-world financial systems, causal structures can evolve over time due to various factors, such as political tensions, regulatory updates, or economic events, that cannot be predicted.

3.5 Our Framework

In light of this limitation, we propose EVO-PCMCI+, a novel multivariate time series forecasting framework that detects evolving causal structures based on the iterative application of the PCMCI+ causal discovery method. This approach explicitly considers time-varying causal relationships, addressing causal non-stationarity inherent in financial data. The proposed framework is evaluated on real-world financial data and compared to state-of-the-art methods. The results show that EVO-PCMCI+ effectively capture evolving causal structures that are applicable for forecasting.

In conclusion, forecasting financial time-series is a challenging task, and traditional statistical methods as well as most of deep and machine learning forecasting methods are based on correlation but do not distinguish between spurious and non-spurious cases. Causal discovery forecasting methods aim to distinguish spurious associations from actual causal relationships and use the latter to predict future variables. However, current causal discovery time-series forecasting models often assume constant causal relationships over time, neglecting the evolving nature of real-world relationships. Moreover, the causal models that search for time-varying changing relationships limit the changes to happen only at predetermined points in time. To address this crucial gap we present a novel multivariate time series forecasting framework that detects evolving causal structures based on the iterative application of the PCMCI+ causal discovery method, explicitly considering time-varying causal relationships, and addressing causal non-stationarity inherent of financial data.

4. Data and Methodology

In this section, first we describe the datasets used for our analysis, second we present the methodology of EVO-PCMCI+, our proposed time series forecasting framework based on estimated evolving causal structures.

4.1 Data

In our study we focus on three different datasets: “Synthetic”, “Rivers”, “Currencies”. The “Synthetic” and “Rivers” datasets are used to test the causal discovery capabilities of the model. Given that the ground truth causal structures of these two datasets are known, we can compare the causal structures resulting from the model with the actual ones and evaluate the model’s causal discovery performance. On the other hand, the ground truth causal structures of the “Currencies” is not known. In fact, the “Currencies” dataset is used first to explore qualitative causal knowledge, and second to evaluate the forecasting accuracy of the model.

4.1.1. The "Synthetic" dataset

The "Synthetic" dataset is the only artificial dataset of this study and it is generated from a numerical process based on a predefined structural equation model. The "Synthetic" dataset is recreated following the example from Runge repository of tutorials (Runge, 2022a). The data generation process is designed so that the variables are subject to contemporaneous dependencies as well as noises with different variances. In our study, we generated X_t^i , i = 1, ..., 5 time series of length 5000 based on the following structural causal process:

$$\begin{cases} X_t^0 = 0.7X_{t-1}^0 + \eta_t^0 \\ X_t^1 = 0.8X_{t-1}^1 + 0.3X_{t-1}^0 + \eta_t^1 \\ X_t^2 = 0.5X_{t-1}^2 - 0.5X_{t-2}^0 + \eta_t^2 \\ X_t^3 = 0.X_{t-1}^3 + \eta_t^3 \\ X_t^4 = 0.X_{t-1}^4 + 0.3X_t^3 + \eta_t^4 \end{cases}$$

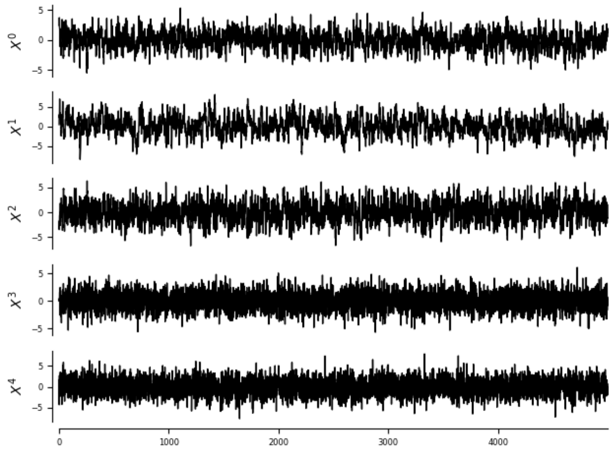


Figure 2: Shows the time series synthetically generated from the adjacent structural equation model

4.1.2 The "Rivers" dataset

The "Rivers" dataset contains daily discharges of rivers in the upper Danube basin, provided by the Bavarian Environmental Agency. Similarly to Mhalla, we use the "Rivers" dataset to test the causal discovery capabilities of the proposed framework on a real-world dataset (Mhalla, 2020). We use measurements from 2017 to 2019 from the Iller at Kempten Kt, the Danube at Dillingen Dt, and the Isar at Lenggries Lt. The Iller flows into the Danube within a day, which implies an instantaneous causal link Kt → Dt and no direct causal links between the pairs Kt, Lt and Dt, Lt. Moreover all variables may be confounded by rainfall or other weather conditions (Gerhardus & Runge, 2020), which challenges the ability of the applied methods to distinguish between spurious associations and true causal relationships.

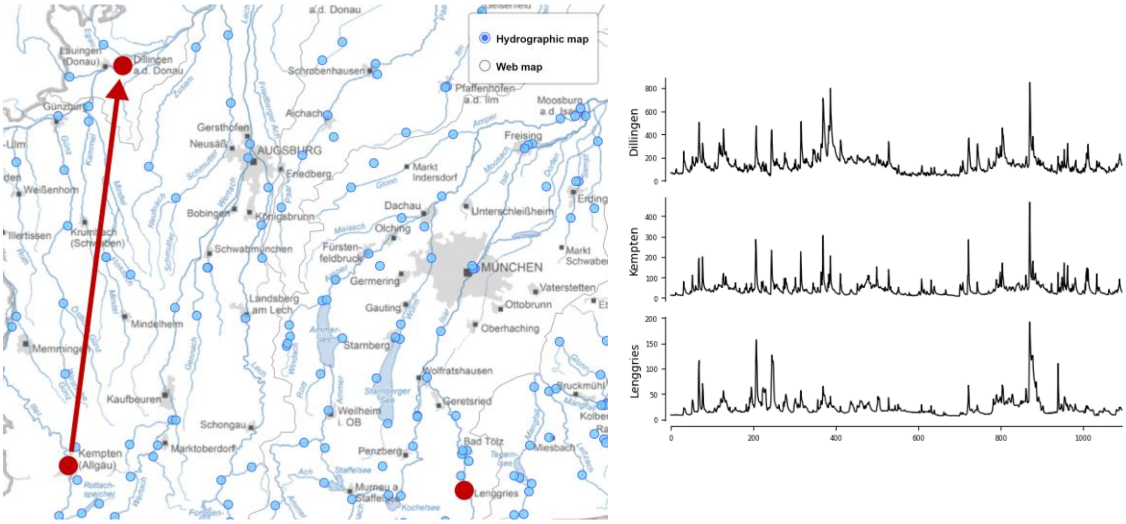


Figure 3: Shows the river stations: https://www.gkd.bayern.de/en/rivers/waterlevel

4.1.3 The “Currencies” dataset

The “Currencies” dataset is composed by the time series of daily logarithmic returns of seven different currencies. The currencies time series are downloaded from the international monetary fund website (IMF, 2023). The currencies analysed are the following currencies: AUD, CNY, JPY, EUR, CHF, GBP, USD. The values are measures as currency units per Special Drawing Rights (SDR), meaning that every currency is priced in relation not to another single currency but to a basket of all other currencies (IMF, 2023).

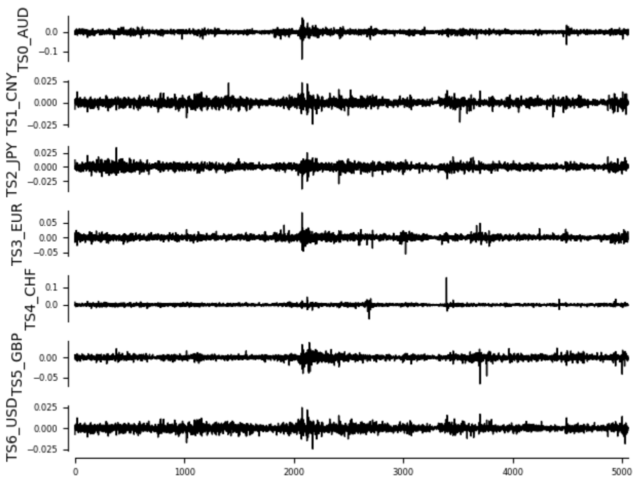


Figure 4: Shows

4.2 Methodology

The methodology of our analysis is structured in five main parts. First, we frame our method’s problem statement. Second, we specify the method’s assumptions. Third, we present the model’s architecture. Fourth, we briefly illustrate the functioning of the

PCMCI+ causal discovery method, on which our framework is based. Finally, we present the four steps that build EVO-PCMCI+: sliding window causal discovery, prediction, iteration, and evaluation.

4.2.1 Problem Statement

We structure the problem like Runge (Runge, 2023). Consider an underlying discrete-time structural causal model (SCM) $X_t = (X_t^1, \ldots, X_t^N)$ with

$$X_t^j = f_j(\mathcal{P}(X_t^j), \eta_t^j) \tag{1}$$

where f_j represent the underlying nonlinear functional dependency, $\mathcal{P}(X_t^j)$ stands for the set of temporally-specified variables that are causal parents of X_t^j , and η_t^j represents mutually $(i \neq j)$ and serially $(t' \neq t)$ independent dynamical noise.

For all datasets, we represent the causal relationships described by the SCM in a form of directed acyclical graph (DAG) that:

1. has one node per variable X_t^j and
2. has a directed edge $X_{t-\tau}^i \rightarrow X_t^j$ if and only if $X_{t-\tau}^i \in \mathcal{P}(X_t^j)$

The notation τ specifies the time-lag of each variable. With $\tau > 0$, the edge $X_{t-\tau}^i \rightarrow X_t^j$ represents a time-lagged relationship; while with $\tau = 0$, the edge $X_t^j \rightarrow X_t^j$ represents a contemporaneous relationships.

4.2.2 Model Assumptions

The core idea of our model is that there may exist causal relationships between observed historical time series that partly explain the future changes of the same time series. For example, in the case of the "River" dataset, we assume that daily river discharges measured at one station will affect in some way the daily river discharges measured at the next station if we were to follow the river's flow of water. Similarly, we assume that financial time series variables may causally affect each other values, such as a high return for one currency on one day might partly explain the return of another currency. In this section we explain the causal assumptions of our forecasting method. Except for causal stationarity, our method is based on the set of assumptions assumed by the PCMCI causal discovery method (Runge, 2019).

The first assumption of our method is faithfulness, also called stableness. Faithfulness assumes that if two variables are independent given some other subset of variables, then there exist no causal relationship between the two variables. Faithfulness is crucial for causal discovery as it provides an unambiguous criteria to test the presence of causal relationships (Runge, 2023).

The Causal Markov Condition represents the second assumption of our method. We assume that, given a variable's set of causal parents, the variable is conditionally independent of all variable's past historical values except the variable's set of causal parents. Therefore, for each variable, the set of its causal parents includes only variables that are not independent to the variable examined.

Another assumption of the standard PCMCI is causal sufficiency, also called unconfoundedness, and it assumes that all common drivers are among the observed variables. Therefore, if there exist external variables that have any effect on the system under analysis causal sufficiency does not hold. Such external drivers are also called latent variables or hidden cofounders. Uncofoundness clearly does not hold when studying financial systems due to the high interconnectedness of financial variables. For each financial variable, the number of possible causes is too high to be managed by a realistic causal discovery algorithm. Or said differently, it is almost impossible to isolate a set of financial variables for which one can confidently say that all the causes of the systems are among the variables considered. The assumption of causal sufficiency represents on of the main challenges to financial data for causal discovery methods. In front of this challenge, Runge proposes the Latent-PCMCI, which is designed to account for hidden cofounder (Gerhardus & Runge, 2021). However, the computational complexity of the LPCMCI limits its applicability only to datasets with a small number of both variables and assumed hidden variables. In our framework we assume causal sufficiency even if doea

Additionally, our method assume non-linearity allowing to model for non-linear causal relationships between time-series variables. Furthermore, we assume autocorrelation between time series by explicitly modelling every variable to be a potential time-lagged cause of itself.

Finally, our method assume the underlying causal graph to be partially-stationary or, expressed differently, to slowly change over time. A slowly-changing causal graph can be intended as a causal graph of which most of its causal links do not change from one period to the other.

The way in which our model detect time-varying causal graphs is by iterating the PCMCI causal discovery method over time while shifting its training window with the new available data. The attentive reader will have noticed that we are applying the PCMCI causal discovery method, which assumes stationary causal relationships, to data that we assume to have non-stationary causal relationships. Even though this approach may sound unproper, its validity can be intended by the comparison of measuring causal change to measuring speed. In the same way the speed of an object cannot be measured at a one precise instant with one single photography or observation, but necessitates minimum two consecutive spatial-temporal observations, the change of a causal relationship cannot be measured by one single causal discovery application but minimum two consecutive causal-temporal observations are needed. Applying the PCMCI iteratively through time to detect changes in causality can be intended as combining temporally consecutive photos to detect movement. In the next section we illustrate more specifically the mechanics of our framework.

4.2.3 Model Architecture

In this section we present the model architecture that allows our framework to search for time-varying causal structures from non-stationary observational time series data. The architecture of our method is divided into 3 main phases: sliding window causal discovery, prediction, and iteration. A schematic visualization of our model architecture can be seen in Figure ???.

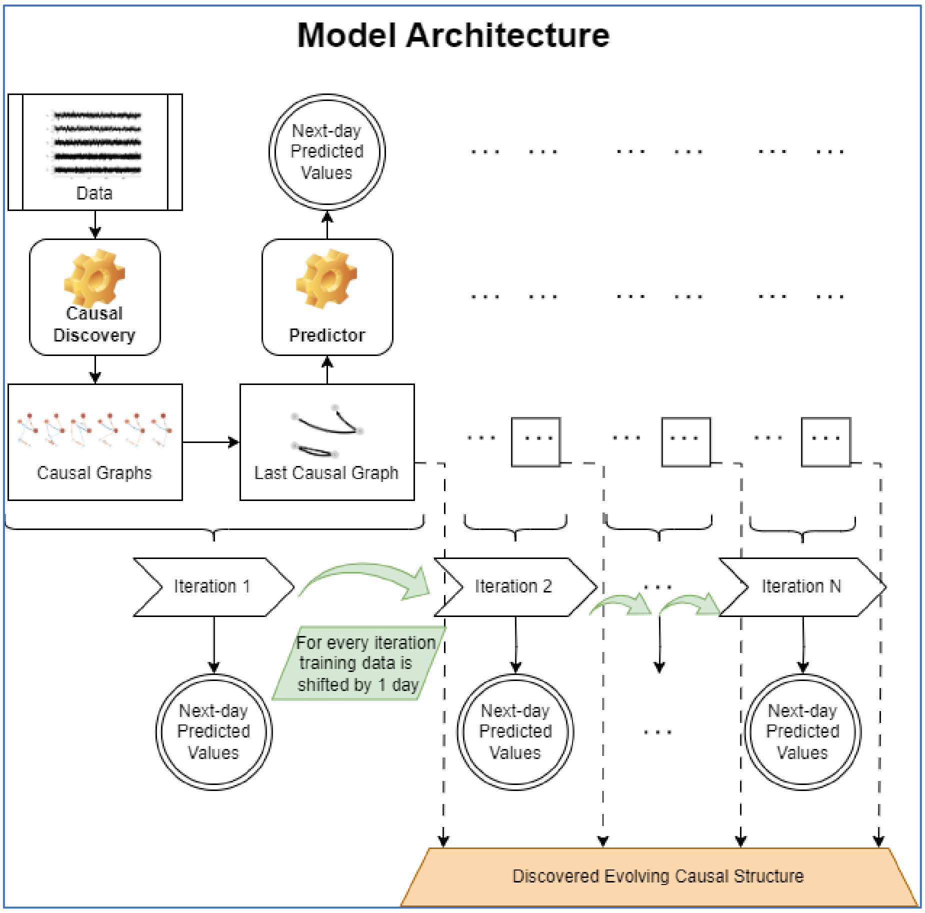


Figure 5; Show a conceptual diagram of our method's architecture

Our model starts by processing the dataset so that is readable by the causal discovery algorithm. Then, we split the data between training and testing subsamples. Next, we set the parameters for the sliding window PCMCI+ causal discovery algorithm. Before explaining how we treat the results of the sliding window version of the PCMCI+ , we will next describe the process by which the PCMCI+ estimates the underlying causal structure.

4.2.4 PCMCI+ Causal Discovery Method

Specifically tailored for time-series data, the PCMCI+ causal discovery method estimates the underlying causal structure $\hat{\mathcal{G}}$ by combining a conditional selection algorithm derived from the PC algorithm with a momentary conditional independence (MCI) test. Therefore, the PCMCI+ method is structured in two stages: the PC and the MCI stages, and both of them consist of conditional independence tests (Runge, 2022).

PC Stage

The task of PC stage is to find the sets S of the strongest causal parents p for each variable; these sets will be used as conditions by the MCI stage. For each variable X_t^j, the PC stage starts by assuming a set S of preliminary parents $\hat{P}(X_t^j)$ that includes all the time-lagged variables $X_{t-\tau}^i$, with $i, j \in \{1, \dots, N\}$ and time delays $\tau \in \{1, \dots, \tau_{max}\}$, where the maximum time delay τ_{max} is set by the user.

First, the PC stage conducts unconditional independence tests between all the current and the lagged variables pairs $(X_{t-\tau}^i, X_t^j)$. If the null hypothesis $X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \hat{P}(X_t^j) \setminus \{X_{t-\tau}^i\}$ that a lagged variable $X_{t-\tau}^i$ is conditionally independent on X_t^j cannot be rejected at a significance level $\alpha PC \in [0, 1]$ (also set by the user), the PC stage removes the lagged variable $X_{t-\tau}^i$ from the set S of preliminary parents [Kretschmer et al. 2018].

$$\text{PC: } X_{t-\tau}^i \not\perp\!\!\!\perp X_t^j \mid S \text{ for any } S \text{ with } |S| = p \tag{1}$$

Next, the PC stage sorts the preliminary parents $\hat{P}(X_t^j)$ by their absolute test statistic value. The type of test statistic can also be specified by the user between linear and non-linear conditional independence tests. Finally, the PC stage iteratively conduct conditional independence tests to find S, the set of strongest parents p of each variable.

MCI Stage

The MCI stage performs momentary conditional independence (MCI) [Pearl 2013] tests between all variable pairs $(X_{t-\tau}^i, X_t^j)$, where S, the set of strongest parents p discovered by the PC stage, is used as the conditioning set of the MCI tests. At every MCI test, the causal relationship $X_{t-\tau}^i \rightarrow X_t^j \in \hat{\mathcal{G}}$ is established if and only if

$$\text{MCI: } X_{t-\tau}^i \not\perp\!\!\!\perp X_t^j \mid \hat{P}(X_t^j) \setminus X_{t-\tau}^i, \hat{P}_{pX}(X_{t-\tau}^j) \tag{1}$$

where $\hat{P}_{pX}(X_{t-\tau}^j) \subseteq \hat{P}(X_{t-\tau}^j)$ denotes the p_X strongest parents according to the sorting in the PC stage. The discovered causal graph is comprehensive of all detected directed lagged links and directed/undirected contemporaneous links that represent the causal parentships between the variables.

PCMCI+ Parameters

The PCMCI+ requires the user to specify a three main parameters: the maximum time-lag of causal relationships τ_{max}, the type of conditional independence test, and the significance level αPC . In this section we are going to present what these parameters are and how our method makes use of them.

The first parameter of the PCMCI+ is τ_{max} , which specifies the assumed maximum time delay of causal relationships and it can be set as high as 30 time periods. In our framework we set τ_{max} equal to two, which means that we limit the PCMCI+ to search for causal relationships that take no longer than two days to have effect.

The second parameter is the type of conditional independence tests to be implemented by both the PC and the MCI stage. The selection of tests statistics offered by PCMCI+ are shown in table in the Appendix. In our model we use Linear Partial Correlation (ParCorr) and two types of nonlinear independence tests: Gaussian process and distance regression (GPDC), and Conditional Mutual Information (CMI). ParCorr assumes linear additive noise models and GPDC only additivity. GPDC is based on a distance correlation test on the residuals of a Gaussian process regression, which makes it a suitable test for a large class of nonlinear dependencies with additive noise []. CMI is a fully nonparametric test based on a K-Nearest Neighbors (KNN) estimator of conditional mutual information that does not require additivity accommodating almost any type of dependency (Runge, 2023).

An additional hyperparameter of the PCMCI+ is αPC and it represents the minimum significance level for a casual relationship to be considered such. Too small values of αPC result in many true links not being included in the condition set for the MCI tests and, hence, increase false positives. Too high levels of αPC lead to high dimensionality of the condition set, which reduces detection power and increases the runtime. In our method we set αPC equal to 0.02, as suggested by Runge (Runge, 2019). In case the causal graph discovered results to be too sparse or too crowded, meaning that it displays too few or too many causal links, αPC can be manually adjusted by the user to increase or decrease the number of causal relationships searched by the PCMCI+ for each causal graph.

Now that the functioning of the PCMCI+ causal discovery method has been briefly illustrated, in the next section we will explain how our method utilizes the sliding window version of the PCMCI+ to estimate a temporal sequence of causal graphs.

4.2.5 Sliding Window Causal Discovery

In our method we apply the sliding window version of the PCMCI+ causal discovery method, which simply consists in running the PCMCI+ on consecutive time windows of the same multivariate time series dataset. Naturally, this approach leads our framework to have two additional parameters: the length of the time window w , and the length of the time step s between the starting points of two consecutive time windows. Formally, as proposed by Runge, we run PCMCI+ sequentially on the samples in each time window

$$\{\mathbf{X}_t\}_{t=s\cdot i}^{s\cdot i+w-1} \text{ for } i = 0, 1, \dots . \quad (1)$$

In our experiments we set window size w equal to window step s so that the time windows do not overlap adopted do not overlap. We set value of these temporal parameters differently for each dataset so that the number of causal graphs estimated is equal to six.

Once the parameters of the sliding window PCMCI+ causal discovery have been set and the algorithm has been applied to the training dataset, the model finally outputs the discovered temporal sequences of causal graphs. Under the assumption that causal relationships are stationary, meaning that they are fixed through time, Runge proposes to use an “aggregated” causal graph as graph predictor []. The way Runge calculates the “aggregated” graph can be intended as the weighted average between the temporal causal graphs of the discovered sequence. However, because we assume evolving/non-stationary causal relationships, and because our goal is to perform next-day forecasting, the “aggregated” graph cannot be a good approximation of the next-day causal graph, given that it assigns too much weight on observations that are too far in the past.

In response to this challenge, our framework utilizes the most recent causal graph of the estimated sequence of causal graphs as the graph predictor. The reasoning for this being that we assume that most recent causal graph to be a better approximation of the next-day causal structure than the "aggregated" graph. Further research can be extended to estimate the next day causal structure with different techniques.

Once our framework has determined the graph predictor to use, the last step of the sliding window causal discovery phase is to extract the link coefficients of the graph predictor. The link coefficients contain the information of the causal parentships of each variable in a format of a list of arrays that is processable by the predictor. In the next section we illustrate how our framework uses the extracted link coefficients to estimate the variables' next-day predictions.

4.2.6 Prediction

In this section we explain how our framework uses the results of the causal discovery stage to forecast the variables' next-day values. Our prediction scheme is derived from the one proposed by Runge, with the only difference being that we use the most recent causal graph as the initial causal predictor (Runge, 2015). Specifically, for the prediction of Y_{t+h} given the multivariate time series X, our prediction consists of the following steps:

1) Use the link coefficients of the most recent causal graph discovered as the initial causal predictors $\mathcal{P}_{Y_{t+h}}$ of $Y_{t+h} = f_j(\mathcal{P}_{Y_{t+h}}, \eta_t^j)$.
2) Find the optimal set of predictors $\mathcal{P}_{Y_{t+h}}^{(p)} \subseteq \mathcal{P}_{Y_{t+h}}$ that have the highest multivariate mutual information (MMI) $\hat{I}\,(\mathcal{P}_{Y_{t+h}}^{(p)}\ ;\ Y_{t+h})$ values.
3) Apply the KNN linear predictor to the optimal set of predictors $\mathcal{P}_{Y_{t+h}}^{(p)}$ and forecast the next-day variables.

The KNN linear predictor is an advanced nearest-neighbor estimator designed for continuous data that flexibly adapts to any function f_j. The only free parameter of this estimator is the number of the nearest neighbors k , which determines the size of the hyper-cubes around each sample point (Runge, 2017). Once k has been determined, to predict each variable the KNN linear predictor applies the following procedure:

1) calculate the distances $d_{t,s} = \left\| \mathcal{P}_{t+h}^{(p)} - \mathcal{P}_s^{(p)} \right\|$ for all $s \in \mathcal{T}$ with $s > \mathcal{T}_{max} + h$

2) sort the distances in increasing order $d_{t,s_1} < d_{t,s_2} < \cdots$ yielding an index sequence $s_1, s_2, \dots$.
3) estimate Y_{t+h} by calculating the conditional expectation of the average of the k nearest neighbours

$$\hat{Y}_{t+h} = \frac{1}{k} \sum_{j=1}^{k} Y_{s_j} \tag{1}$$

The dataset on which we train the KNN linear predictor corresponds to the time window from which we extracted the causal predictors.

4.2.7 Iteration

The iteration phase represents the final step of our multivariate forecasting framework. Instead of predicting the variables' future values for the next 30 days, as proposed by Runge, we only predict the future values for the next day and reiterate the whole process thirty times, day after day.

For every iteration, we slide the training dataset of both the causal discovery and the prediction phases by one day to the future. In this way, the causal graph used as predictor graph updates itself everyday based on the new available data. This iterative technique brings two benefits. First, the model bases its predictions on the most recent available data. Second, the causal graph used as predictor slightly changes at every iteration allowing the model to detect time-varying causal relationships. Formally, we run both the rolling window causal discovery and KNN prediction phases sequentially on the sample in each time window

$$\{\mathbf{X}_t\}_{t=i}^{i+w-1} \text{ for } i = [0, 29] \, , \tag{1}$$

where w corresponds to size of the time window from which we extracted the causal predictors. At every iteration our method outputs the next-day forecasts for all variables. After all the iterations have been performed, our method collects the predicted values in a single dataframe for each variable. In this way, the forecasted time series can be easily plotted and compared to the actual data.

4.2.8 Evaluation

In this section we present the evaluation metrics utilized by our framework to assess the accuracy of both the causal discovery results and the forecasting results. To evaluate the causal graphs discovered by our framework we employ the Structural Humming Distance (SHD) measure. SHD aims at quantifying the dissimilarity between two graphical structures and is calculated as the minimum number of operations required to transform one causal graph into another. The operations include adding a link, removing a link, and reverting a link's directionality. A lower SHD value indicates closer similarity between the estimated and the true causal graphs, and therefore refers to better causal discovery performance.

On the other hand, to evaluate the time series predictions for each currency we calculate the Normalized Mean Absolute Error (NMAE). This metric is used to facilitate the comparison regarding the MAE of datasets with different scales. As a mean of normalization, the model performance evaluation tool uses the mean of the measured data. []

$$NMAE(\boldsymbol{y}, \widehat{\boldsymbol{y}}) = \frac{MAE(\boldsymbol{y}, \widehat{\boldsymbol{y}})}{\frac{1}{n}\sum_{i=1}^{n}|y_i|} = \frac{MAE(\boldsymbol{y}, \widehat{\boldsymbol{y}})}{mean(|\boldsymbol{y}|)}$$

Finally, we calculate the overall model's prediction performance by taking the average between the NMAE values of each forecasted time series. In the next section we present the results of applying our framework to four different datasets.

5. Experimental Results

This section is divided into two main parts. In the first part we assess the causal discovery capabilities of the most recent causal graph estimated by the sliding window PCMCI and PCMCI+. In the second part we evaluate the forecasting accuracy of our framework EVO-PCMCI+.

5.1 Causal Discovery Results

The causal discovery phase of our framework replicates the sliding window approach of the PCMCI and PCMCI+ methods with the only difference being that it considers the most recent causal graph to be the most representative one. Once the temporal sequence of causal graphs has been estimated, we evaluate the causal discovery performance of our framework by comparing the estimated most recent causal graph with the true underlying causal graph. To quantify the level of dissimilarity between the estimated and the actual graph the framework calculates the Structural Humming Distance (SHD) between the two graphs. We can perform this calculation only for the "Synthetic" and "Rivers" datasets, as they are the only datasets of our study of which the true underlying causal structure is known, or at least assumed. For the "Currencies" dataset, whose the true causal graphs is unknown, we limit our analysis to a qualitative description of the causal discovery results. In the next section we begin the analysis of the causal discovery results of our framework.

5.1.1 Causal Discovery Results - "Synthetic" Dataset

The first dataset on which we apply our framework is the "Synthetic" dataset. The "Synthetic" dataset is generated from a pre-defined structural causal graph that is fixed through time. In Figure ???, one can observe the generated time series data studied with the relative sequence of causal graphs discovered by the sliding window PCMCI+.

Figure ??? shows the generated time series data studied with the relative sequence of causal graphs discovered by the PCMCI. Regarding the causal discovery sliding window parameters we choose the training data to be of 1081 observations, with each consecutive sliding window analyse to include 180 observations, for a total of six discovered causal graphs. We chose the sliding window size of the PCMCI to be 180 days long, so that each causal graph discovered spans a time range of 6 months. Moreover, we chose the length of the training dataframe to be 1081 in order to train the model on 3 years of historical data.

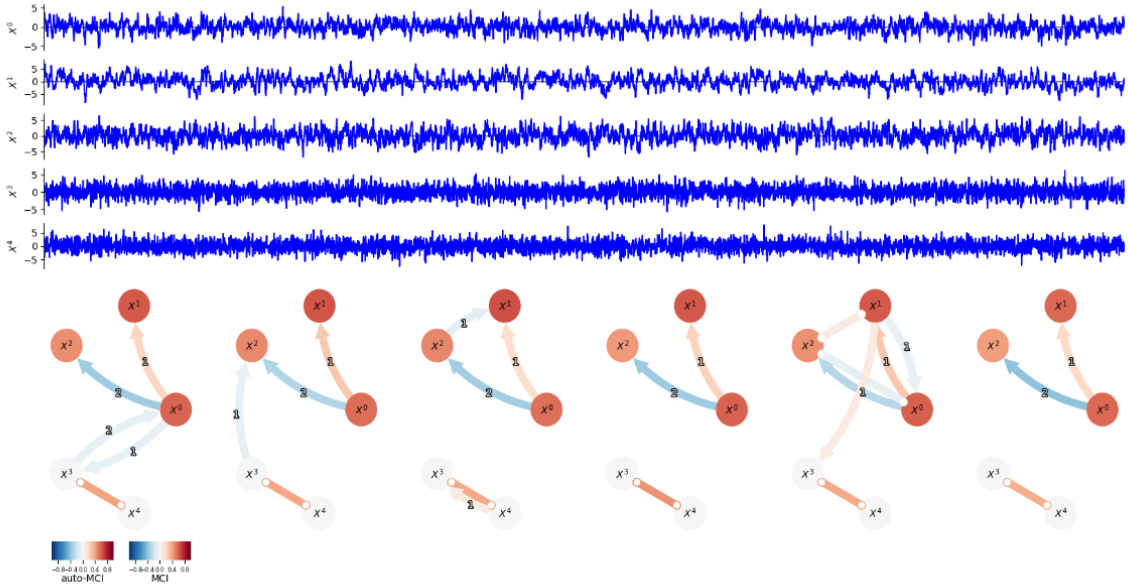
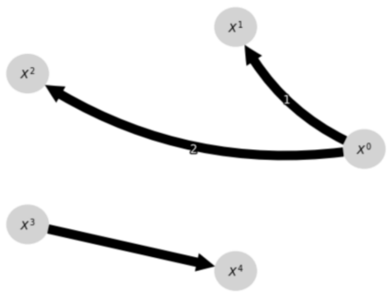
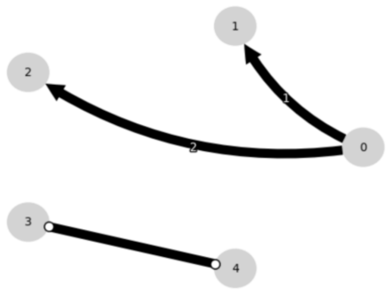


Figure 6: Shows the estimated sequence of causal graphs discovered by our framework using the sliding window PCMCI+ on the "Synthetic" dataset with with a conditional independence test of partial correlation (ParCorr).

Figure ??? shows the results of applying the sliding window version of the PCMCI to the "Synthetic" dataset. We can observe, that even though the discovered causal graphs change through time, they are all relatively similar to the true causal graph shown in Figure ???. Since we choose to use the most recent discovered causal graph as the predictor graph, the following table offers a more detailed comparison.

True Causal Graph	Estimated Causal Graph
 *Figure 7: Shows the true causal graph underlying the Structural Causal Model (SCM) of the "Synthetic" dataset.*	 *Figure 8: Shows the temporally most recent causal graph discovered by the PCMCI*

To compare the graphs we use the Structural Hamming Distance measure, which is calculated by counting the number of different edges between a pair of graphs. Naturally, the lower the SHD value, the less precise is the causal discovery method.

Table 1: Shows the Structural Hamming Distance (SHD) measure for the different combination of causal discovery methods and conditional independence tests. The lower the SHD, the more precise is the discovered causal graph.

Causal Discovery Method	Conditional Independence Test	SHD
PCMCI	ParCorr	3
	RobustParCorr	3
	ParCorrWLS	5
	GPDC	3
	CMIknn	5
PCMCIplus	ParCorr	2
	RobustParCorr	4
	ParCorrWLS	3
	GPDC	2
	CMIknn	5

As shown in table ???, we calculated the SHD for a range of conditional independence tests, both for the PCMCI and the PCMCIplus causal discovery methods. Notably, the combinations of the PCMCIplus causal discovery engine with the ParCorr and GPDC conditional independence test yield the most accurate causal graph with a SHD as low as 2. Having said this, however, we cannot generalize stating that these two methods are the most accurate for any type of observational time series dataset. Every dataset is different and can have a different set of specific functions of possible causal relationships among variables. To evaluate the causal discovery performance of the PCMCI on a different and real-world dataset, in the next section we investigate possible causal relationships between the daily waterflow discharges of three German river stations.

5.1.2 Causal Discovery Results - “River” Dataset

Next, we examine the results of applying the sliding window PCMCI causal discovery method to the “River” datasets. The “Rivers” dataset contains the observations of the daily discharges of rivers in the upper Danube basin taken at the following stations: Kempten Kt, Dillingen Dt, and Lenggries Lt. In Figure ??? we can see the time series of daily river discharges for the three German river stations. The estimated sequence of causal graphs discovered by the PCMCI is shown below.

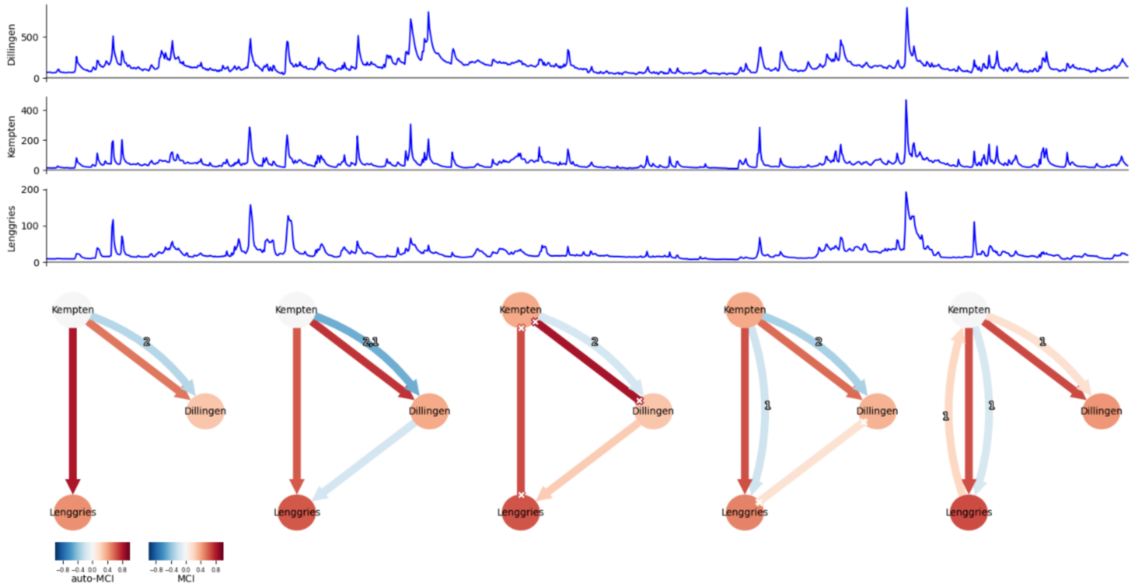
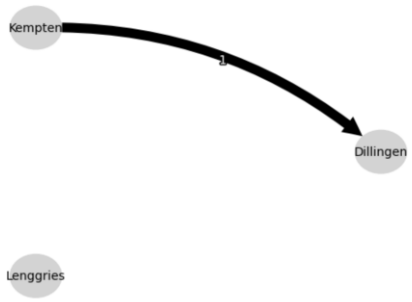
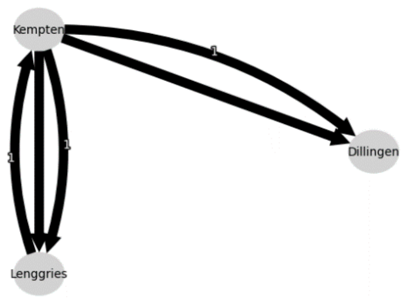


Figure 9: Shows the daily river discharges time series (above) and the estimated sequence of causal graphs discovered by the PCMCI on the “Rivers” dataset

As opposed to the “Synthetic” dataset, for the “River” dataset we do not have the exact underlying causal structure. Therefore, in order to evaluate the causal discovery accuracy of PCMCI on this dataset, we will compare our results with specific domain knowledge which is informative on the causal directionality between the variables examined. For causal discovery sliding window parameters we choose the training data to be of 1000 observations, with each consecutive sliding window analyse to include 180 observations for a total of five discovered causal graphs.

While the Iller flows into the Danube upstream of Dillingen with the water from Kempten reaching Dillingen within a day approximately, the Isar reaches the Danube downstream of Dillingen. Based on the stated scenario, we expect a contemporaneous link Kt → Dt and no direct causal links between the pairs Kt, Lt and Dt, Lt. [Causal

Inference in Non-linear Time-series using Deep Networks and Knockoff Counterfactuals]. These variables may be confounded by rainfall or other weather conditions that might change the causal structures through time, this choice allows testing the ability of the methods to detect and distinguish directed and bidirectional links

Presumed Causal Connection	Estimated Causal Graph
 *Figure 10: Shows the true causal graph underlying the Structural Causal Model (SCM) of the "Synthetic" dataset.*	 *Figure 11: Shows the temporally most recent causal graph discovered by the PCMCI*

In table ??? we evaluate how the causal graphs estimated by applying the sliding window PCMCI on the dataset differ from the knowledge-based causal structures. Because the true causal graph has only one link, we lower the parameter alphaPC equal from 0.2 to 0.1.

Table 2: Shows the Structural Hamming Distance (SHD) measure for the different combination of causal discovery methods and conditional independence tests. The lower the SHD, the more precise is the discovered causal graph.

Causal Discovery Method	Conditional Independence Test	SHD
PCMCI	ParCorr	13
	RobustParCorr	7
	ParCorrWLS	11
	GPDC	6
	CMIknn	6
PCMCIplus	ParCorr	4
	RobustParCorr	1
	ParCorrWLS	6
	GPDC	2
	CMIknn	4

The results indicate that the PCMCIplus achieve much better discovery performance with any conditional independence test. This difference could be explained by the fact that the PCMCIplus as been designed to consider variable's autocorrelation. In the next sections we will show the results of the causal discovery methods when applied to financial datasets.

In this section we present the results of applying the causal discovery phase of our framework to the "Currencies dataset. Because no underlying causal structure is known, we cannot test the validity of the our estimated causal graphs, and will limit our analysis to qualitative causal exploration. However, we will comment whether our results are in line with current financial theories.

5.1.3 Causal Discovery Results - "Currencies" Dataset

Similarly as the previous datasets, we start by applying the sliding window PCMCI to the "Currencies" dataset. For the sliding window parameters we choose the training data to consist of 4500 observations, with each consecutive sliding window to include 720 observations, for a total of six discovered causal graphs. For causal discovery we use the parameters that led to the most accurate results in the previous experiments: causal discovery class PCMCIplus with the RobustParrCorr or GDPC independence test. Unfortunately, however, we could not implement the GPDC conditional independence test because of it high computational costs.

Moreover, because the number of edges discovered was initially too high, we changed α to 0.05 instead of 0.1 to limit the number of spurious relationships. Too small values of α result in many true links not being included in the condition set for the MCI tests and, hence, increase false positives. Too high levels of α lead to high dimensionality of the condition set which reduces detection power and increases the runtime []. Figure ??? shows the time series as well as the relative sequence of causal graphs discovered.

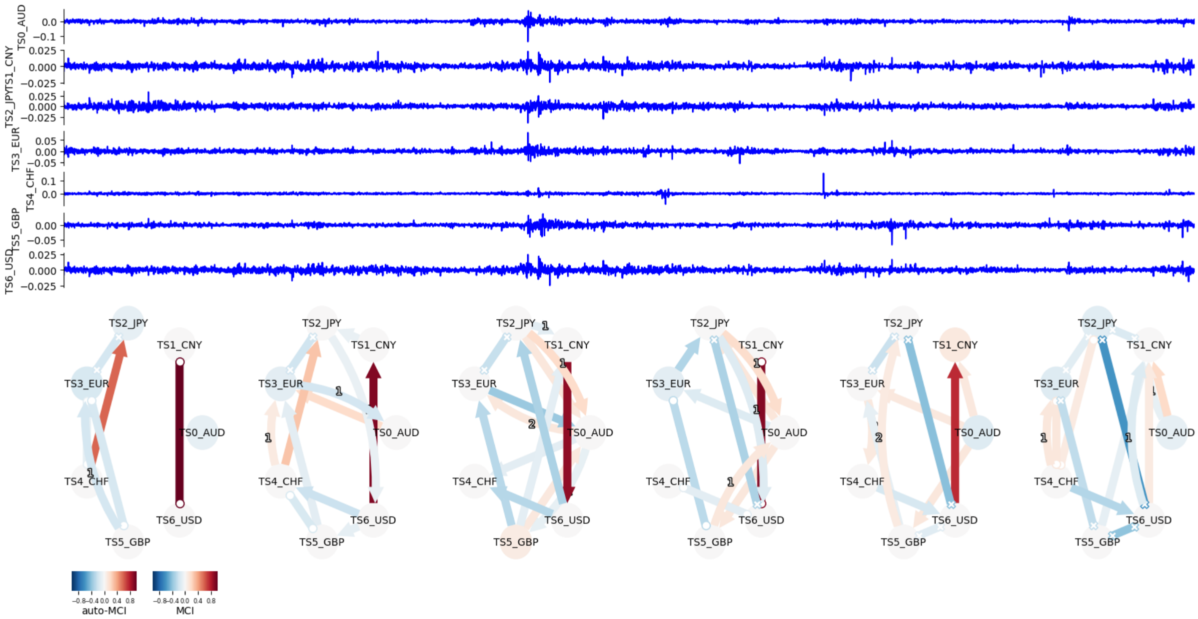


Figure 12: Shows the currencies' SDR daily returns from 19??? to ?? (above) and the estimated sequence of causal graphs discovered by the PCMCIplus with the ParrCorr conditional independent test.

Here, as opposed to the "Synthetic" and the "Rivers" datasets, we can notice the temporal evolution of the causal graphs with some causal relationships grow stronger or weaker over time. For example, looking the edges between CNY and USD, in the first graphs the colour of the edge is very dark, almost brown, then it slowly fades away until disappearing in the most recent graph. The colour of an edge represents a statistic measure indicating the frequency or the expectancy of a certain edge to manifest. The darker the colour of the edge, the more frequent or expectable is to find it, therefore, a possible interpretation is that the causal relationship between CNY and USD gradually lost frequency or strength over time. Moreover, we can see that the edge is detected to be a contemporaneous link in the first, third, and fourth graph, but is detected as a directed link suggesting that USD is a causal parent of CNY.

Another example is the edge between JPY and USD, which is lacking in the first two graphs, but it appears in the third graph in light blue colour, and it gets darker in the remaining graphs. Even though is detected as a contemporaneous link most of the time, the fifth graph suggests a causal effect from USD to JPY.

Estimated Causal Graph

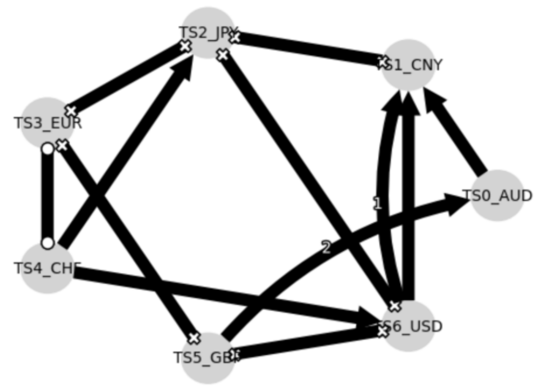


Figure 13: Shows the temporally most recent causal graph discovered by the PCMCI+ with conditional independence test

In addition, one may observe from the most recent graph that the causal relationships follow an interesting geopolitical patterns. EUR is connected to the geographically close CHF, GBP, but interestingly also to JPY, which is maybe attributable to strong commercial relationships between the regions. GBP is connected to EUR, USD, and AUD, somewhat representing the English-speaking part of the world. USD is connected to CHF, GBP, CNY, and JPY. While it is hard to think that the price behaviour of CHF is a causal parent of USD, it is interesting to notice that USD is more connected to the Asian currencies rather than to EUR or AUD.

In the next section, we will show the accuracy of our day-by-day predictive model when applied to financial datasets.

5.2 Forecasting Results

In this third part of the experimental results, we apply our entire forecasting framework to all four datasets and we evaluate its predictions compared to the actual data. Specifically, our forecasting framework finally consists in iterating the combination of the PCMCI+ rolling window causal discovery method and various types of predictive functions.

We kept the same sliding window parameters of the previous experiments, with the training data of 4500 observations, every sliding window include 720 observations, for a total of six discovered causal graphs. PCalpha is set equal to 0.05.

Next, we apply the same causal discovery methods to the “Currencies” and “Interest Rates” datasets, whose underlying causal structure is unknown. For these datasets, we cannot test the causal discovery performance of our model, but we will evaluate if the resulting causal graphs are in line with some financial theory. Finally, we apply our time series forecasting method to all datasets and evaluate the resulting predictions with the actual data.

5.2.1 Forecasting Results - “Synthetic” Dataset

In this section we evaluate the forecasting performance of EVO-PCMCI+ on the “Synthetic” dataset. In figure ???? we can observe the sequence of thirty successive causal graph used to predict the values of all variables for the next day. Each causal graph represents the most recent graph estimated by the sliding window PCMCI+. For every iteration we shift the training dataset of the PCMCI+ to include the newly available data. Therefore, the first graph is estimated with the first iteration, the second graph with the second iteration, an so forth.

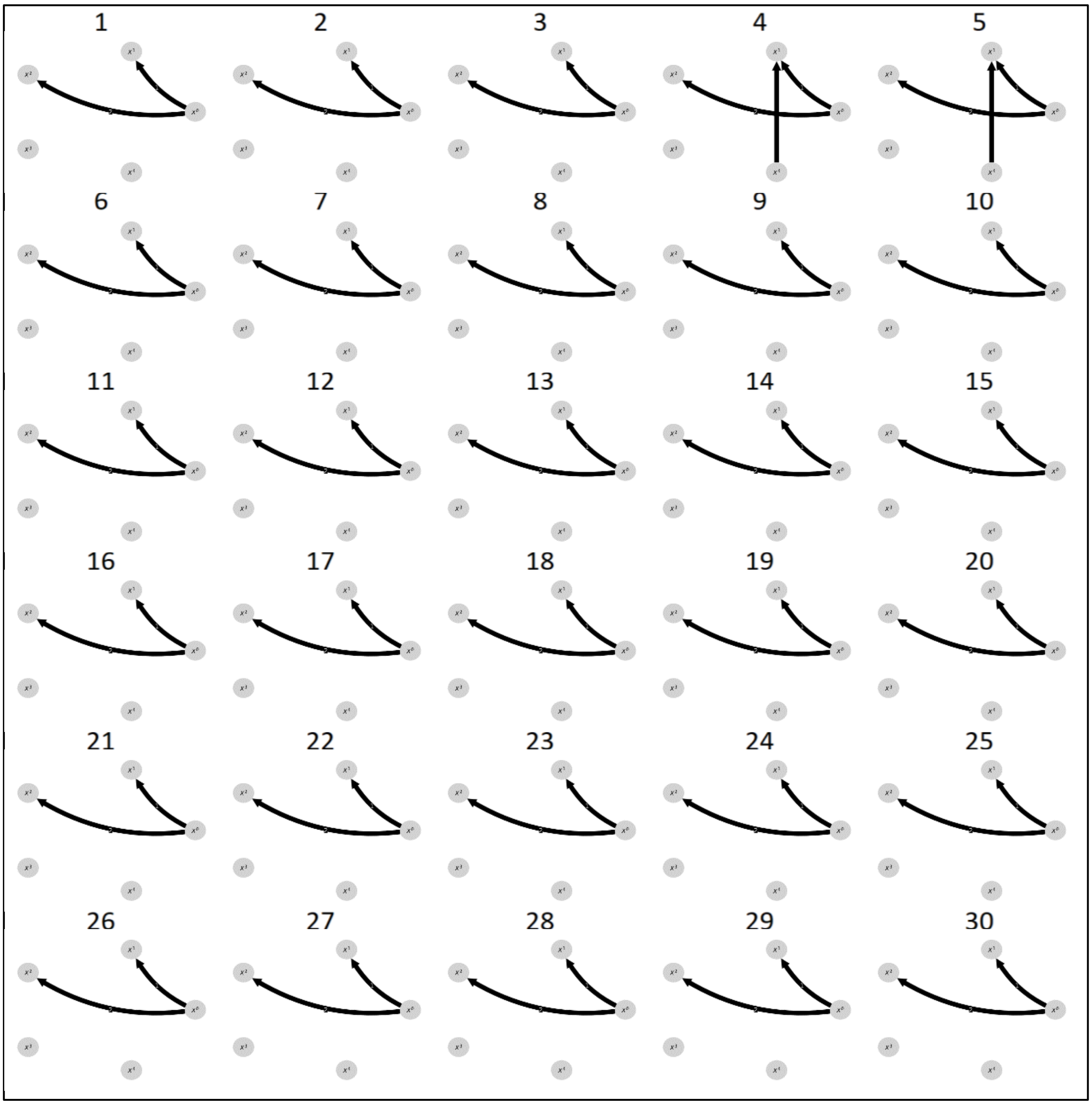
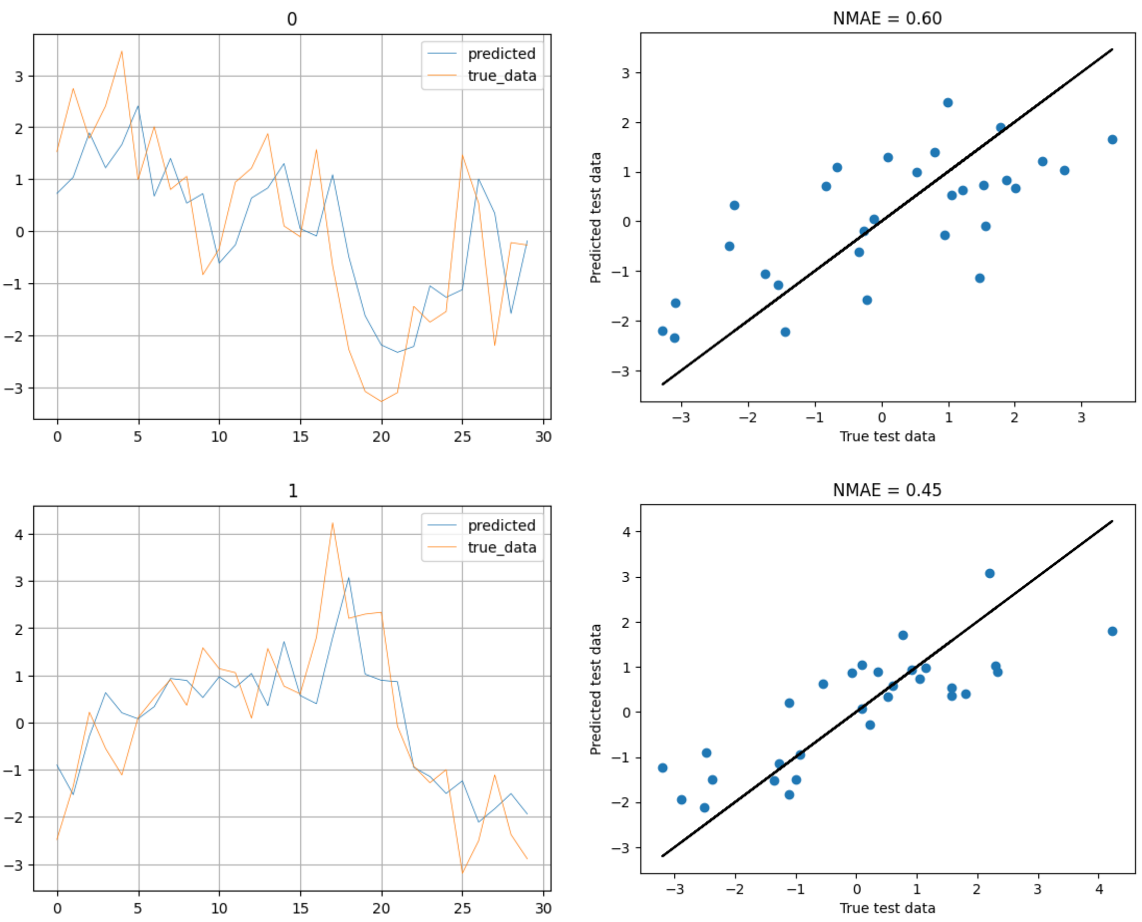


Figure 14: Shows

From Figure ??? we can observe that, except for the fourth and fifth graph, the causal graphs estimated are all the same and equal to the true underlying causal graph of the “Synthetic” dataset (Figure ???). The fact that most of the estimated graph precisely describe the true causal structures confirms the robustness of the PCMCI+ causal discovery phase and indicates the viability of its iterative application.

The forecasting results of EVO-PCMCI+ on the “Synthetic” dataset are illustrated in Figure ????. The first row of plots shows to the prediction for the target variable X_0, the second row refers to target variable X_1, and so forth up to variable X_4. The plots on the left illustrate the predicted values relative to the true values. The plots on the right shows the accuracy of the predictions calculated using the Normalized Mean Absolute Error (NMAE). The closer are the blue points to cartesian bisector, the more accurate are the predictions, and therefore, lower NMAE values reflect better predictions.



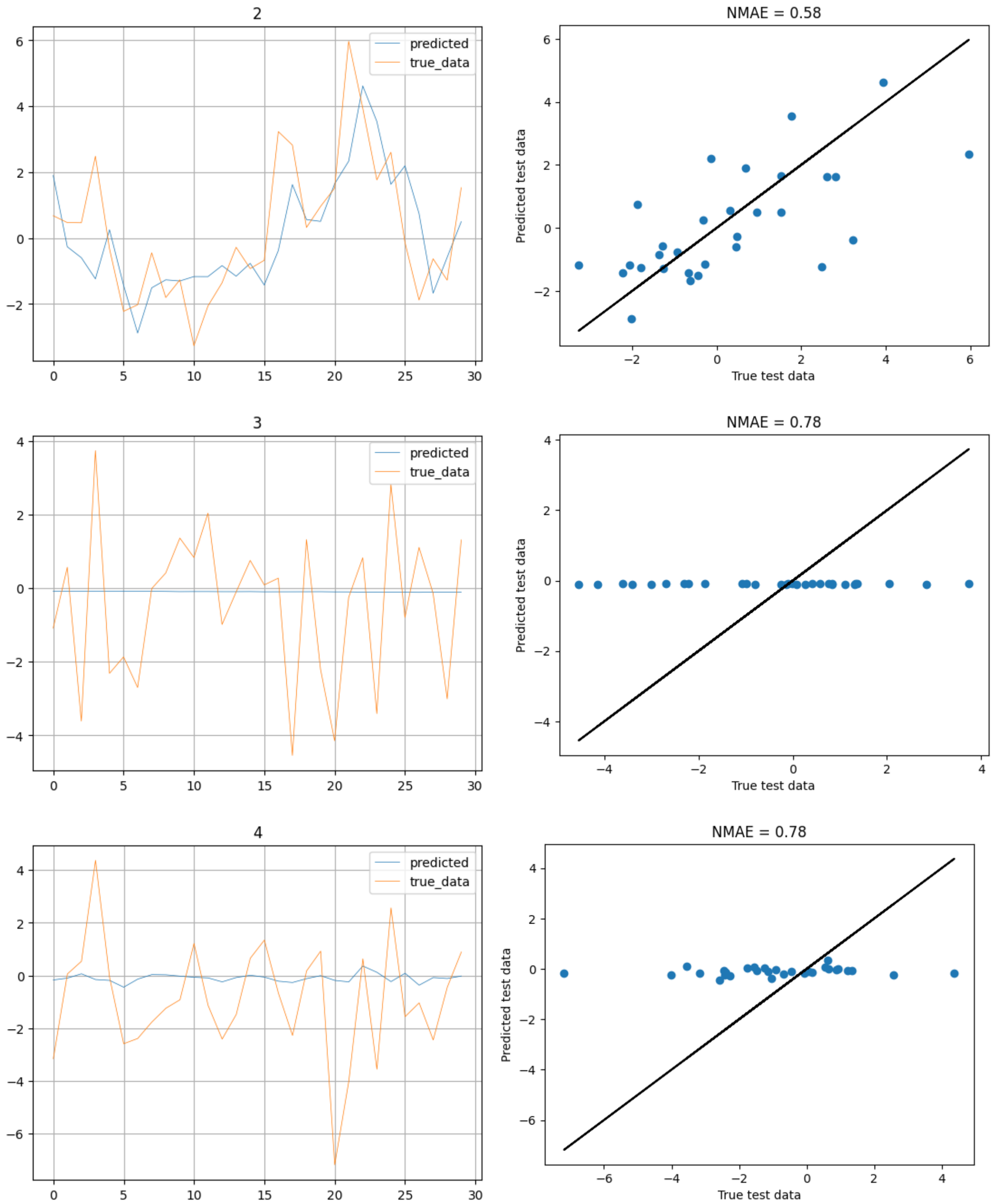


Figure 15: Shows

Average NMAE: 0.6386720206991013

RobustParrCOrr - LinearRegression

Notably, the forecasting results for the first three variables are much better than the ones for the last two variables. Additionally, Figure ??? shows that in each estimated graph the variables X_4 and X_5 are not connected to any other variable. This makes sense and is due to the fact that the PCMCI is not designed to discover at least one causal parent for each target variable. The variables that are discovered to have no causal parents are intended as if they were caused by nothing else other than themselves, and this makes them difficult to predict. In such cases, our framework estimate all values to be close to mean. If we consider the accuracy of forecasting X_4, X_5 (NMAE = 0.78) as representative for the case in which no causal information is used for forecasting, by comparing this accuracy with the accuracy relative to forecasting $X_{1,} X_2, X_3$, we can claim that causal information partly explains future values for this specific dataset. The average NMAE achieved by the EVO_PCMCI+ on the "Synthetic" dataset is of 0.76

5.2.2 Forecasting Results - "River" Dataset

In this section we evaluate the forecasting performance of EVO-PCMCI+ on the "River" dataset. In figure ???? shows the sequence of the estmated causal graph used to predict the values of the "River" dataset. Also for this dataset, we can see that most graphs are similar, indicating a causal connection from Kempten to Dillingen. As we have seen, this is line with domain knowledge and reassure us on applying these graphs for forecasting.

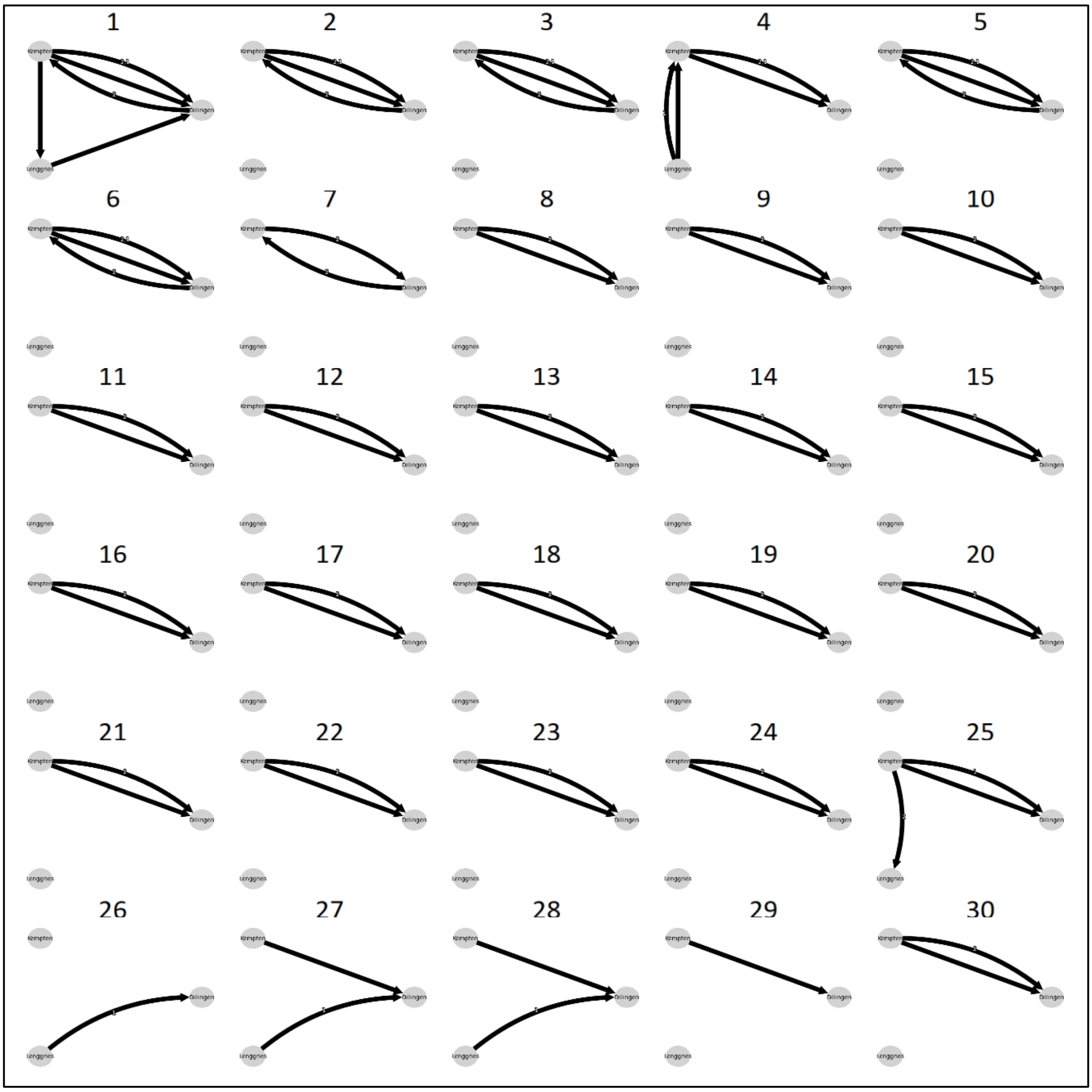


Figure 16: Shows

The forecasting results of EVO-PCMCI+ on the "River" dataset are illustrated in Figure ????. The first row of plots shows to the prediction for water discharge levels measured at Dillingen (target variable X_0), the second row refers to Kempten (target variable X_1), the third row refers to Lenggries (target variable X_2). We can se that the water discharge levels measured at Dillingen are fairly well predicted, with an accuracy as low as 0.31 NMAE.

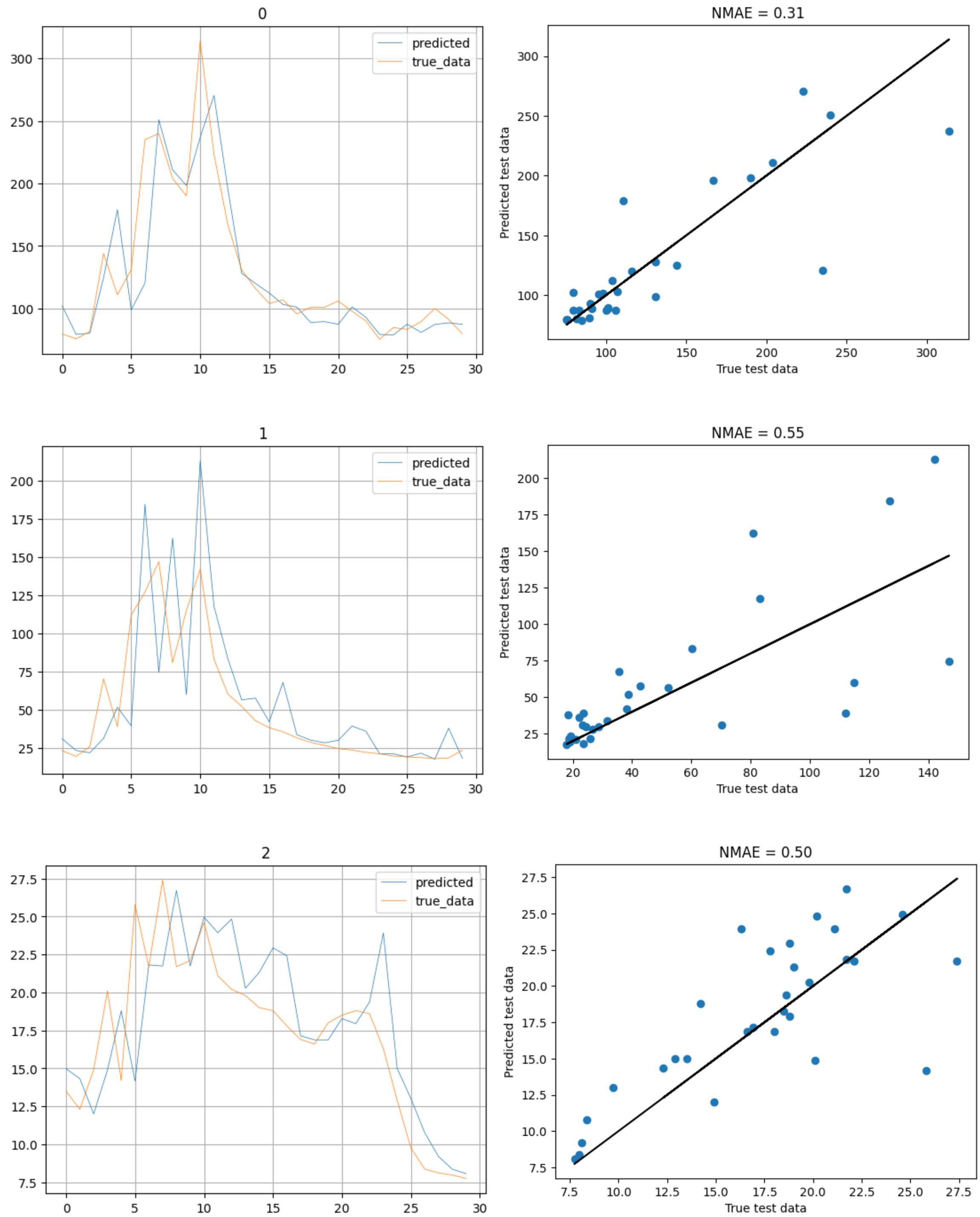


Figure 17: Shows

The PCMCI parameters that we applied are the following: PCMCI+ as causal discovery method, Robust Partial Correlation as conditional independence test, alphaPC equal to 0.2, and the KNeighbors regressor as our predictive function.

The average NMAE achieved by the EVO_PCMCI+ on the “River” dataset is of 0.45.

5.2.3 Forecasting Results - “Forex” Dataset

In this section we evaluate the forecasting performance of EVO-PCMCI+ on the “Currencies” dataset. Figure ???? shows the sequence of the estimated causal graph used to predict the values of the “River” dataset. The estimated causal graph show similar patterns between them, with one or two causal links changing from graph to the next one. The similarity and the slowly-varying feature of the discovered graphs suggest the effectiveness of EVO-PCMCI+ in capturing evolving causal relationships from non-stationary datasets such as the “Forex” dataset.

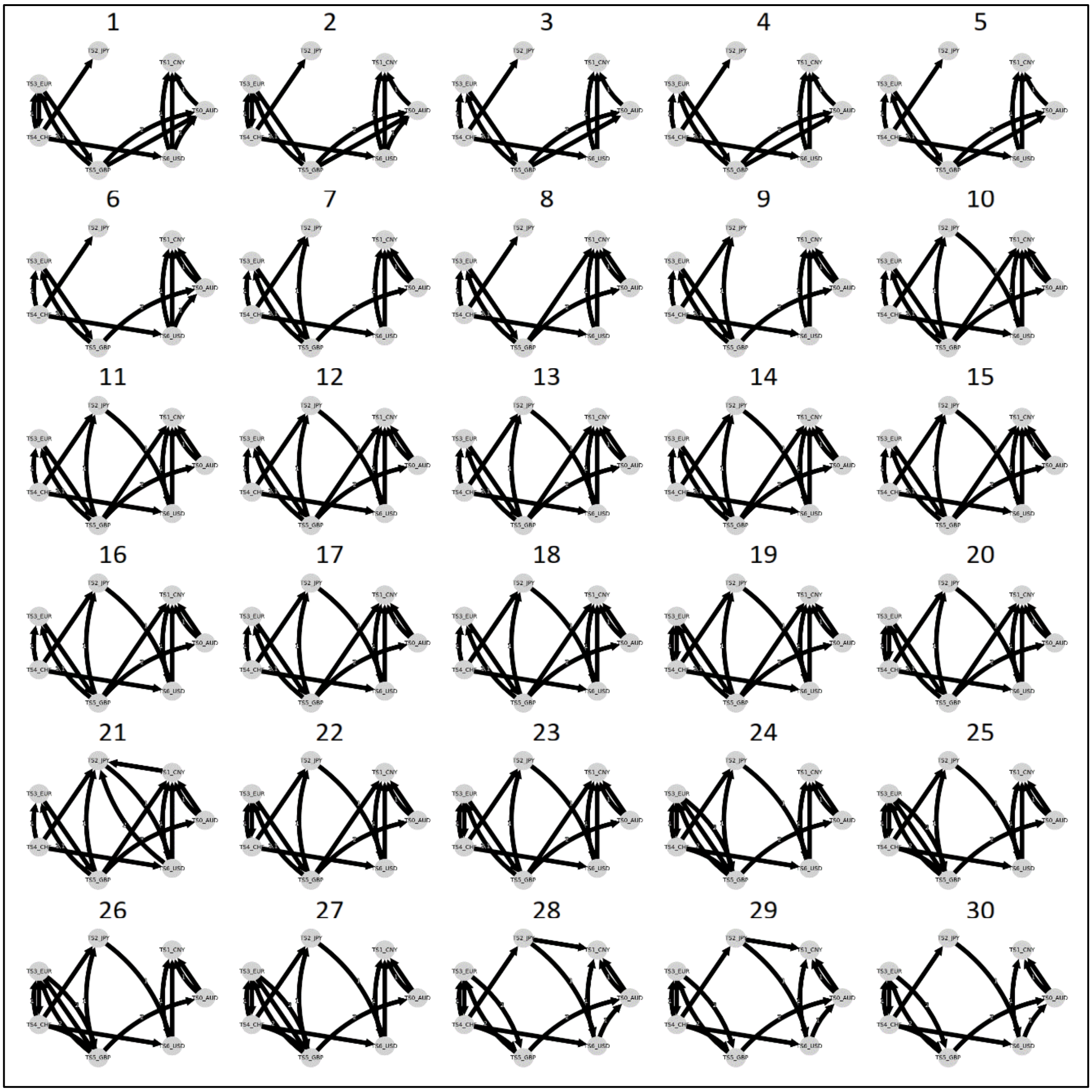
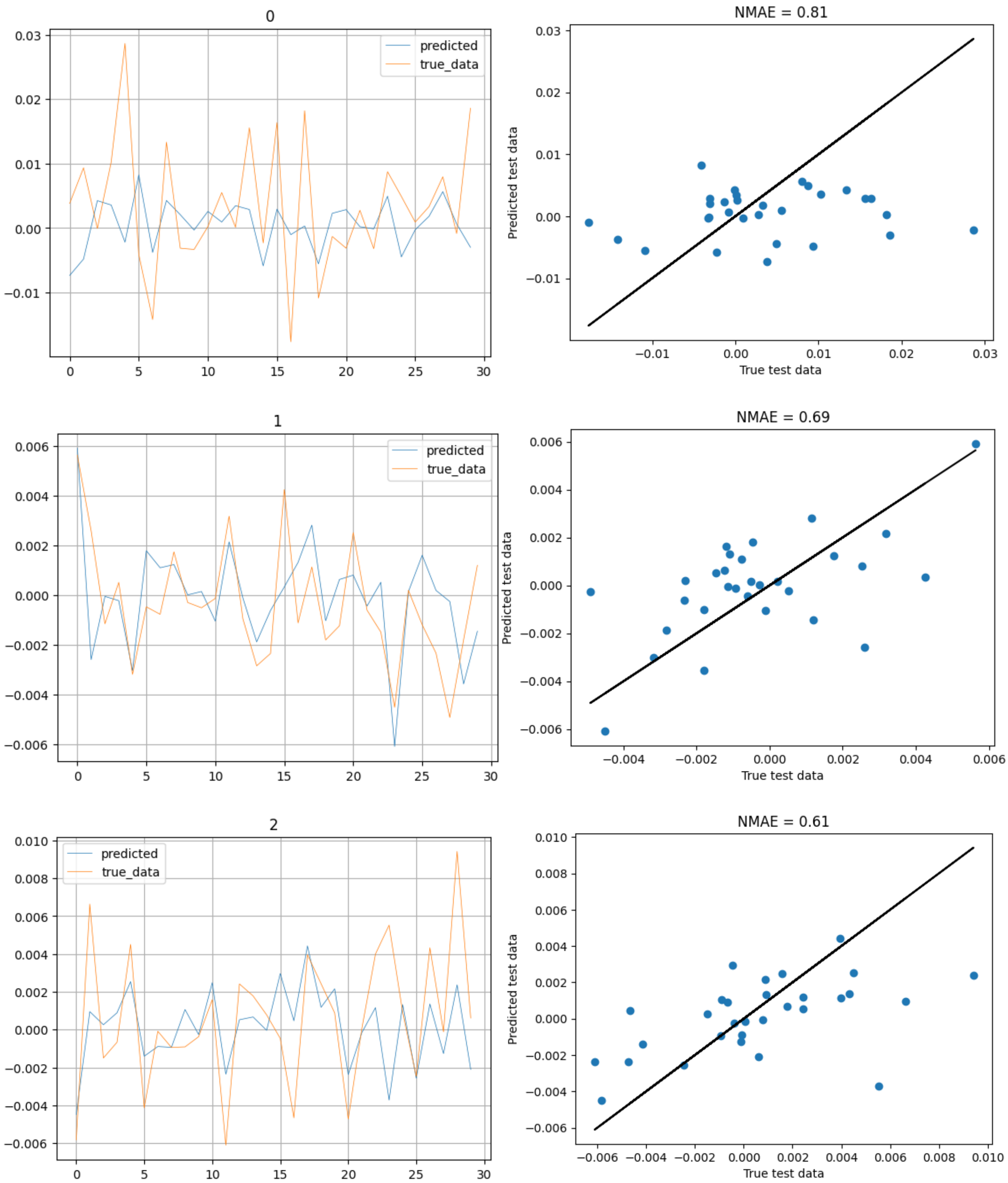
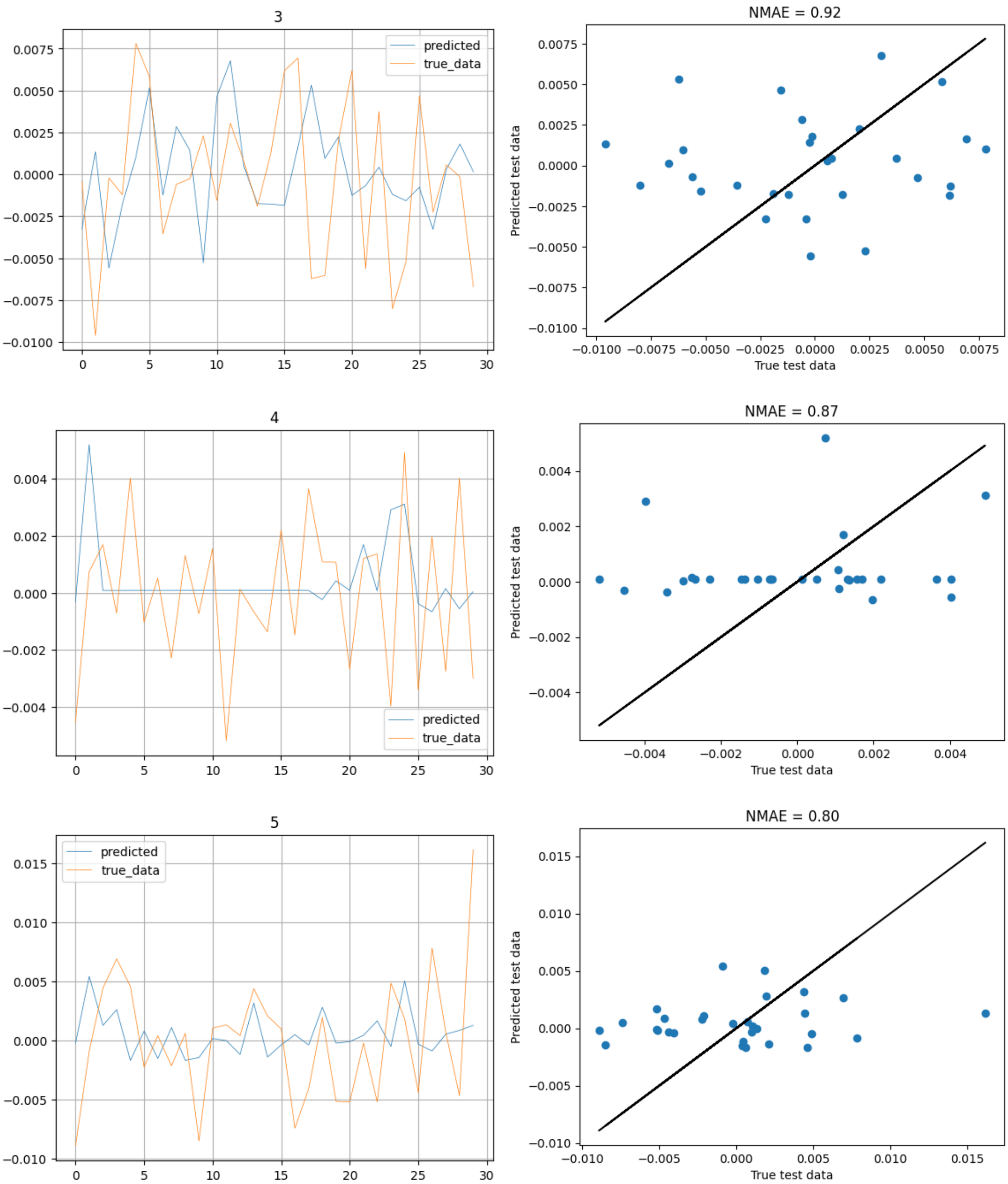


Figure 18: Shows

The forecasting results of EVO-PCMCI+ on the “Forex” dataset are illustrated in Figure ????.

The rows of plots refer to the target currencies in the following order: AUD, CNY, JPY, EUR, CHF, GBP, USD. The average NMAE achieved by the EVO_PCMCI+ on the “Forex” dataset is of 0.76, with the best predicted currency being JPY with a NMAE as low as 0.61.




3
predicted
true_data
NMAE = 0.92
Predicted test data
True test data
4
predicted
true_data
NMAE = 0.87
Predicted test data
True test data
5
predicted
true_data
NMAE = 0.80
Predicted test data
True test data

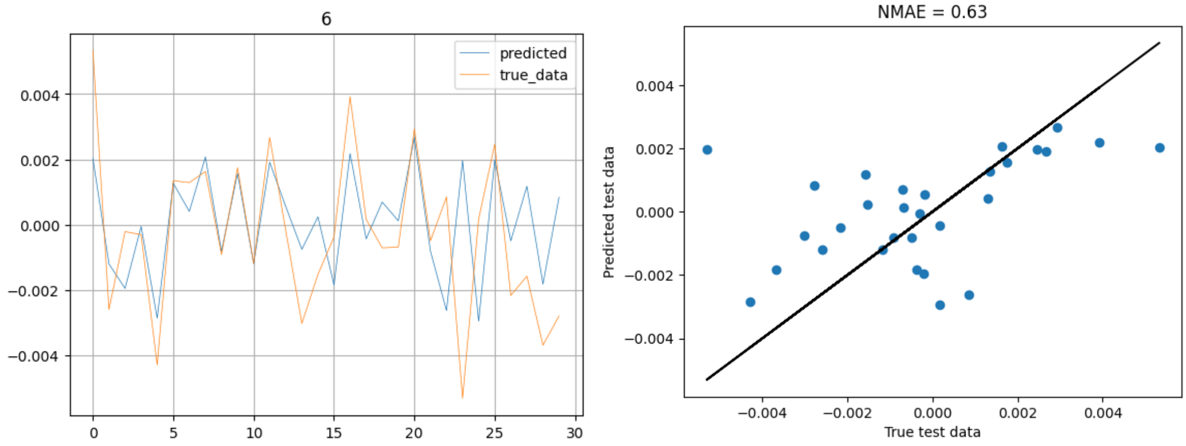


Figure 19: Shows

The average NMAE achieved by the EVO_PCMCI+ on the “Forex” dataset is of 0.76,

5.3 Summary

5.4 Discussion

We can interpret the usability of general time series causal predictive methods in two different scopes: one which focuses on the accuracy of a single event and the other which focuses on the accuracy of the overall system. Let’s take the example of climate science, some studies try to predict when, where, and how strong, a climatic catastrophe will manifest. In such studies, the accuracy of the model is critical and the model needs be specifically designed to accommodate the most advanced domain knowledge in both terms of assumptions and complexity of the causal model. Similarly, when studying the financial markets, some studies try to predict the timing and strength of the next crisis or bubble. To do this, specific industry knowledge of the asset class that is going to crash is required. For example, only the ones who were experts in the US mortgage industry have been able to forecast the subprime crisis of 2008. Therefore, if one were to apply causal models for a scope that requires great accuracy on a single event, the model should be precisely tailored with the support of an industry expert to understand the causal structure within the sector.

On the other hand, other climate science studies try to predict river water flow with the scope to improve irrigation systems[]. In such studies, the accuracy of a single event is less important, what plays a major role is the accuracy of the overall system. Similarly, some tasks like trading and portfolio management are also more focused on the accuracy and consistency of the overall method, rather than on the accuracy of predicting a single event. Within multi-asset portfolio management, for example, where industry knowledge is still puzzled about the relationships between the price behavior of asset classes, the synergies between causal and financial studies can help advance current domain knowledge. For such multivariate time series forecasting tasks, everything that is learned is gained. For example, a deep learning method can be applied to the temporal sequence of causal graphs to predict the future causal structure of the data.

7. Conclusion

In this paper we presented EVO-PCMCI+, a novel multivariate forecasting framework that leverages causal discovery to predict non-stationary and autocorrelated time series. Founded on the PCMCI+ causal discovery method, our method effectively distinguishes causation from spurious correlation by performing a series of conditional independence tests between selected variables. The iterative combination of the PCMCI+ with various predictive functions allows our framework to capture time-varying causal relationships and to apply them for time series forecasting. EVO-PCMCI+ summarizes its findings by constructing a temporal causal graph that shows the discovered time-varying causal relationships between time series and their corresponding time delays. The ability to model evolving causal relationships is especially important in finance where the causal structures between variables may vary over time due to, for example, new regulations, sudden events, or other structural changes that are not readily visible from prices. Causal discovery methods can be enhanced and refined when integrated with time series forecasting tasks. Given the lack of domain knowledge for the true underlying causal relationships of many complex real-world systems, causal forecasting models offer the opportunity to evaluate the adopted causal discovery method on the basis of the accuracy of generated predictions.

To assess the causal discovery and forecasting capabilities of our framework we applied EVO-PCMCI+ to different artificial and real-world datasets. Our analysis on datasets whose underlying causal mechanisms are partly known shows that our framework reconstructs the ground truth causal graph with a good level of accuracy. On the other hand, because our analysis suggests that the causal relationships estimated by PCMCI+ improve the accuracy of our framework, the discovered causal graphs can be said to contain information that partly explain future time series values.

For future research, there are several directions in extending the work. First, since EVO-PCMCI+ aims at discovering evolving causal structures, testing our framework on a non-stationary synthetic dataset would bring additional insight. Second, for a complete evaluation of the forecasting capabilities of EVO-PCMCI+, on a comparison with non-stationary time series forecasting baseline models would be needed. Third, the application of EVO-PCMCI+ could be extended to various datasets outside the financial landscape. Fourth, future research could aim to apply our iterative approach to the other variants of the PCMCI prosposed by Runge. Because our method assumes causal sufficiency, namely that all causal parents are within the dataset analysed, which cannot hold when studying financial markets, applying our method to the Latent-PCMCI version to account for hidden variables would be especially relevant for financial data. Lastly, our framework utilizes the most recent causal graph discovered as the predictor for the forecasting task. However, one could try to estimate the graph predictor in different way and an integration with machine learning methods to forecast the next temporal causal graph to be used as graph predictor would be interesting.

Finally, our study provides positive empirical evidence in support that EVO-PCMCI+ performs as expected by successfully discovering evolving causal structures and using such learned pattern to forecast non-stationary financial data. We hope this paper will stimulate further research using causal discovery methods for the analysis and forecast of financial data.

8. Appendix

Bibliography

Ahmad, W., Shadaydeh, M., & Denzler, J. (2022, July 8). Causal Discovery using Model Invariance through Knockoff Interventions. https://doi.org/10.48550/arXiv.2207.04055

Assaad, C. K., Devijver, E., & Gaussier, E. (2022). Survey and Evaluation of Causal Discovery Methods for Time Series. *Journal of Artificial Intelligence Research, 73*, 767–819. https://doi.org/10.1613/jair.1.13428

Barnett, L., & Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, *223*. https://doi.org/10.1016/j.jneumeth.2013.10.018

Beroza, G. C., & Ide, S. (2011). Slow Earthquakes and Nonvolcanic Tremor. *Annual Review of Earth and Planetary Sciences*, *39*(1). https://doi.org/10.1146/annurev-earth-040809-152531

Börjesson, L., & Singull, M. (2020). Forecasting Financial Time Series through Causal and Dilated Convolutional Neural Networks. *Entropy*, *22*(10), 1094. https://doi.org/10.3390/e22101094

Chickering, M. (2020, August 27). Statistically Efficient Greedy Equivalence Search. Retrieved September 23, 2023, from proceedings.mlr.press website: https://proceedings.mlr.press/v124/chickering20a.html

Duan, Z., Xu, H., Huang, Y., Feng, J., & Wang, Y. (2021, December 14). Multivariate Time Series Forecasting with Transfer Entropy Graph. https://doi.org/10.48550/arXiv.2005.01185

Durbin, J., & Koopman, S. J. (2012). Time Series Analysis by State Space Methods. In *Google Books*. OUP Oxford. Retrieved from https://books.google.de/books?hl=en&lr=&id=lGyshsfkLrIC&oi=fnd&pg=PP1&dq=%5BKoopman+et+al.

Eichler, M. (2013). Causal inference with multiple time series: principles and problems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and*

Engineering Sciences, *371*(1997), 20110613. https://doi.org/10.1098/rsta.2011.0613

Entner, D., & Hoyer, P. (2010). *On Causal Discovery from Time Series Data using FCI COMET K-Project -Advanced Engineering Design Automation (AEDA) View project On Causal Discovery from Time Series Data using FCI*.

Federal Reserve Bank of St. Louis. (2023). FRED Economic Data. Retrieved from Stlouisfed.org website: https://fred.stlouisfed.org/

Ferkingstad, E., Løland, A., & Wilhelmsen, M. (2011). Causal modeling and inference for electricity markets. *Energy Economics*, *33*(3), 404–412. https://doi.org/10.1016/j.eneco.2010.10.006

Fujiwara, D., Koyama, K., Kiritoshi, K., Okawachi, T., Izumitani, T., & Shimizu, S. (2023, March 17). Causal Discovery for Non-stationary Non-linear Time Series Data Using Just-In-Time Modeling. Retrieved September 22, 2023, from openreview.net website: https://openreview.net/forum?id=aGmh4Y8wtl

Gerhardus, A., & Runge, J. (2021). LPCMCI: Causal Discovery in Time Series with Latent Confounders. *Elib (German Aerospace Center)*. https://doi.org/10.5194/egusphere-egu21-8259

Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00524

Gopinathan, R., & Durai, S. R. S. (2019). Stock market and macroeconomic variables: new evidence from India. *Financial Innovation*, *5*(1). https://doi.org/10.1186/s40854-019-0145-1

Günther, W., Ninad, U., & Runge, J. (2023, June 22). Causal discovery for time series from multiple datasets with latent contexts. https://doi.org/10.48550/arXiv.2306.12896

Günther, W., Ninad, U., Wahl, J., & Runge, J. (2023, June 20). Conditional Independence Testing with Heteroskedastic Data and Applications to Causal Discovery. https://doi.org/10.48550/arXiv.2306.11498

Harary, F. (1969). Graph Theory. In *Amazon.com*. Retrieved from https://www.amazon.com/Theory-Demand-Printing-Advanced-Program/dp/0201410338

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396). https://doi.org/10.2307/2289064

Huang, W. (2008). Information lag and dynamic stability. *Journal of Mathematical Economics*, *44*(5). https://doi.org/10.1016/j.jmateco.2007.09.001

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*(3), 439–454. https://doi.org/10.1016/s0169-2070(01)00110-8

IMF. (2023). SDR Interest Rate Calculation. Retrieved from www.imf.org website: https://www.imf.org/external/np/fin/data/sdr_ir.aspx

Jiang, M., Gao, X., An, H., Li, H., & Sun, B. (2017). Reconstructing complex network for characterizing the time-varying causality evolution behavior of multivariate time series. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-10759-3

Kainen, P. C. (2023, May 29). Construction numbers: How to build a graph? https://doi.org/10.48550/arXiv.2302.13186

Kocoglu, M., Kyophilavong, P., Awan, A., & Lim, S. Y. (2022). Time-varying causality between oil price and exchange rate in five ASEAN economies. *Economic Change and Restructuring, 56*(2), 1007–1031. https://doi.org/10.1007/s10644-022-09457-6

Kretschmer, M., Coumou, D., Donges, J. F., & Runge, J. (2016). Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation. *Journal of Climate*, *29*(11). https://doi.org/10.1175/jcli-d-15-0654.1

Lai, G., Chang, W.-C., Yang, Y., & Liu, H. (2017). Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1703.07015

Lai, G., Chang, W.-C., Yang, Y., & Liu, H. (2018). *Modeling Long-and Short-Term Temporal Patterns with Deep Neural Networks Multivariate Time Series, Neural Network, Autoregressive models ACM Reference Format*. Retrieved from https://arxiv.org/pdf/1703.07015.pdf

Lam, W.-Y., Andrews, B., & Ramsey, J. (2022, August 17). Greedy relaxations of the sparsest permutation algorithm. Retrieved September 22, 2023, from proceedings.mlr.press website: https://proceedings.mlr.press/v180/lam22a.html

Li, F., Ding, P., & Mealli, F. (2022, October 23). Bayesian Causal Inference: A Critical Review. https://doi.org/10.48550/arXiv.2206.15460

Löwe, S., Madras, D., Zemel, R., Welling, M., Schölkopf, B., Uhler, C., & Zhang, K. (2022). Amortized Causal Discovery: Learning to Infer Causal Graphs from Time-Series Data. *Proceedings of Machine Learning Research, 140*, 1–24.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE, 13*(3). https://doi.org/10.1371/journal.pone.0194889

Malinsky, D., & Danks, D. (2017). Causal discovery algorithms: A practical guide. *Philosophy Compass*, *13*(1), e12470. https://doi.org/10.1111/phc3.12470

Malinsky, D., & Spirtes, P. (2018, August 9). Causal Structure Learning from Multivariate Time Series in Settings with Unmeasured Confounding. Retrieved September 22, 2023, from proceedings.mlr.press website: https://proceedings.mlr.press/v92/malinsky18a.html

Mastakouri, A. A., Schölkopf, B., & Janzing, D. (2021, July 1). Necessary and sufficient conditions for causal feature selection in time series with latent common causes. Retrieved September 22, 2023, from proceedings.mlr.press website: https://proceedings.mlr.press/v139/mastakouri21a.html

Matthews, R. (2000). Storks Deliver Babies (p = 0.008). *Teaching Statistics*, *22*(2), 36–38. https://doi.org/10.1111/1467-9639.00013

Mayberger, M. (2014). *What is the effect of monetary policy on market behavior?* The Park Place Economist: Vol. 22 . Retrieved from The Park Place Economist: Vol. 22 website: https://digitalcommons.iwu.edu/cgi/viewcontent.cgi?article=1415&context=parkplace

Mhalla, L., Chavez-Demoulin, V., & Dupuis, D. J. (2020). Causal Mechanism of Extreme River Discharges in the Upper Danube Basin Network. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *69*(4), 741–764. https://doi.org/10.1111/rssc.12415

Moraffah, R., Sheth, P., Karami, M., Bhattacharya, A., Wang, Q., Tahir, A., … Liu, H. (2021). Causal inference for time series analysis: problems, methods and evaluation. *Knowledge and Information Systems*, *63*(12), 3041–3085. https://doi.org/10.1007/s10115-021-01621-0

Nauta, M., Bucur, D., & Seifert, C. (2019). Causal Discovery with Attention-Based Convolutional Neural Networks. *Machine Learning and Knowledge Extraction*, *1*(1), 312–340. https://doi.org/10.3390/make1010019

Nogueira, A. R., Pugnana, A., Ruggieri, S., Pedreschi, D., & Gama, J. (2022). Methods and tools for causal discovery and causal inference. *WIREs Data Mining and Knowledge Discovery*, *12*(2). https://doi.org/10.1002/widm.1449

Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. (2020). Evaluation of Methods for Causal Discovery in Hydrometeorological Systems. *Water Resources Research*, *56*(7). https://doi.org/10.1029/2020wr027251

Pavlyshenko, B. M. (2022). Forecasting of Non-Stationary Sales Time Series Using Deep Learning. *Arxiv.org*. https://doi.org/10.48550/arXiv.2205.11636

Ravivanpong, P., Riedel, T., & Stock, P. (2022). Towards Extracting Causal Graph Structures from Trade Data and Smart Financial Portfolio Risk Management. Retrieved September 22, 2023, from publikationen.bibliothek.kit.edu website: https://publikationen.bibliothek.kit.edu/1000147758

Rubin, D. B. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, *100*(469). Retrieved from https://www.jstor.org/stable/27590541

Runge, J. (2017). Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. *ArXiv:1709.01447 [Cs, Math, Stat]*. Retrieved from https://arxiv.org/abs/1709.01447

Runge, J. (2022a). tigramite_tutorial_sliding_window_analysis.ipynb. Retrieved from GitHub website: https://github.com/jakobrunge/tigramite/blob/master/tutorials/causal_discovery/tigramite_tutorial_sliding_window_analysis.ipynb

Runge, J. (2022b, January 5). Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. https://doi.org/10.48550/arXiv.2003.03685

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., … Zhang, K. (2019). Inferring causation from time series in Earth system sciences. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10105-3

Runge, J., Donner, R. V., & Kurths, J. (2015). Optimal model-free prediction from multivariate time series. *Physical Review E*, *91*(5). https://doi.org/10.1103/physreve.91.052909

Runge, J., Gerhardus, A., Gherardo Varando, Eyring, V., & Gustau Camps-Valls. (2023). Causal inference for time series. *Nature Reviews Earth & Environment*, *4*(7), 487–505. https://doi.org/10.1038/s43017-023-00431-y

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, *5*(11). https://doi.org/10.1126/sciadv.aau4996

Saggioro, E., de Wiljes, J., Kretschmer, M., & Runge, J. (2020). Reconstructing regime-dependent causal relationships from observational time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *30*(11), 113115. https://doi.org/10.1063/5.0020538

Shanmugam, R. (2018). Elements of causal inference: foundations and learning algorithms. *Journal of Statistical Computation and Simulation*, *88*(16), 3248–3248. https://doi.org/10.1080/00949655.2018.1505197

Shimizu, S., Hoyer, P. O., Hyvä, A., rinen, & Kerminen, A. (2006). A Linear Non-Gaussian Acyclic Model for Causal Discovery. *Journal of Machine Learning Research*, *7*(72). Retrieved from https://jmlr.org/papers/v7/shimizu06a.html

Shojaie, A., & Fox, E. B. (2021). Granger Causality: A Review and Recent Advances. *ArXiv:2105.02675 [Cs, Stat]*. Retrieved from https://arxiv.org/abs/2105.02675

Silini, R., Tirabassi, G., Barreiro, M., Ferranti, L., & Masoller, C. (2022). Assessing causal dependencies in climatic indices. *Climate Dynamics*. https://doi.org/10.1007/s00382-022-06562-0

Slavov, S. (2022). *Causal Inference: Heterogeneous Effects and Non-stationary Environments*. Massachusetts Institute of Technology .

Soo, K. W., & Rottman, B. M. (2020). Distinguishing causation and correlation: Causal learning from time-series graphs with trends. *Cognition*, *195*, 104079. https://doi.org/10.1016/j.cognition.2019.104079

Spirtes, P. (2001, January 4). An Anytime Algorithm for Causal Inference. Retrieved September 23, 2023, from proceedings.mlr.press website: https://proceedings.mlr.press/r3/spirtes01a.html#:~:text=The%20Fast%20Casual%20Inference%20(FCI

Spirtes, P., Glymour, C., & Scheines, R. (1993). Causation, Prediction, and Search. In *Lecture notes in statistics*. Springer New York. https://doi.org/10.1007/978-1-4612-2748-9

Stavroglou, S., Pantelous, A. A., Soramaki, K., & Zuev, K. (2016). Causality Networks of Financial Assets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2888783

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting Causality in Complex Ecosystems. *Science*, *338*(6106), 496–500. https://doi.org/10.1126/science.1227079

Vigen, T. (2015). Spurious Correlations. *Science*, *348*(6238). https://doi.org/10.1126/science.aac5518

Wahl, J., Ninad, U., & Runge, J. (2023, June 12). Foundations of Causal Discovery on Groups of Variables. https://doi.org/10.48550/arXiv.2306.07047

Zhang, K., Huang, B., Zhang, J., Glymour, C., & Schoelkopf, B. (2017). Causal Discovery from Nonstationary/Heterogeneous Data: Skeleton Estimation and Orientation Determination. *Europe PMC (PubMed Central)*. https://doi.org/10.24963/ijcai.2017/187

I hereby confirm that this thesis constitutes my own work, produced without aid and support from persons and/or materials other than the ones listed. Quotation marks indicate direct language from another author. Appropriate credit is given where I have used ideas, expressions or text from another public or non-public source. The paper in this or similar form has never been submitted as an assessed piece of work in or outside of Germany. It also has not yet been published.

Frankfurt am Main, May, 2023 Luigi Bassani Antivari

City, Date Author's Signature