



# Use of prior knowledge to discover causal additive models with unobserved variables and its application to time series data

Takashi Nicholas Maeda<sup>1,2</sup> · Shohei Shimizu<sup>2,3</sup>

Received: 16 January 2024 / Accepted: 2 August 2024  
© The Behaviormetric Society 2024

## Abstract

This paper proposes two methods for causal additive models with unobserved variables (CAM-UV). CAM-UV assumes that the causal functions take the form of generalized additive models and that latent confounders are present. First, we propose a method that leverages prior knowledge for efficient causal discovery. Then, we propose an extension of this method for inferring causality in time series data. The original CAM-UV algorithm differs from other existing causal function models in that it does not seek the causal order between observed variables, but rather aims to identify the causes for each observed variable. Therefore, the first proposed method in this paper utilizes prior knowledge, such as understanding that certain variables cannot be causes of specific others. Moreover, by incorporating the prior knowledge that causes precedes their effects in time, we extend the first algorithm to the second method for causal discovery in time series data. We validate the first proposed method by using simulated data to demonstrate that the accuracy of causal discovery increases as more prior knowledge is accumulated. Additionally, we test the second proposed method by comparing it with existing time series causal discovery methods, using both simulated data and real-world data.

**Keywords** Causal discovery · Latent confounders · Time-series data · Causal additive models

---

Communicated by Kohei Adachi.

✉ Takashi Nicholas Maeda  
takashi.maeda@gakushuin.ac.jp

<sup>1</sup> Gakushuin University, Tokyo, Japan

<sup>2</sup> RIKEN, Tokyo, Japan

<sup>3</sup> Shiga University, Shiga, Japan

## 1 Introduction

Causal discovery refers to a special class of statistical and machine learning methods that infer causal relationships. These studies propose inferential methods deductively derived from assumptions about the data generation process, and the methods enable us to create causal graphs between observed variables without additional experiments. The assumptions of existing causal methods include acyclicity of causal graphs, absence of latent confounders, and independence and identical distribution of exogenous variables (Spirtes and Glymour 1991; Shimizu et al. 2006, 2011; Peters et al. 2014; Zheng et al. 2018). The methods have been applied to various types of data including economic data (Lai and Bessler 2015), meteorological data (Ebert-Uphoff and Deng 2012), fMRI data (Smith et al. 2011).

This paper proposes a causal discovery method for time-series data assuming the presence of latent confounders. Most existing methods for time-series data assume the absence of a latent confounder (Chu and Glymour 2008; Hyvärinen et al. 2010). However, most data do not satisfy such assumption. A causal discovery method for time-series data, latent Peter-Clark momentary conditional independence (LPCMCI) (Gerhardus and Runge 2020), assumes the presence of latent confounders. However, since LPCMCI is a constraint-based method, it cannot distinguish causal structures that entail the same set of conditional independence between variables. This paper aims to propose a causal functional model-based method for time-series data assuming the presence of latent confounders. We extend the causal additive models with unobserved variables (CAM-UV) algorithm (Maeda and Shimizu 2021a, b) to propose time-series CAM-UV (TS-CAM-UV), a method for causal discovery from time-series data with latent confounders. The original CAM-UV algorithm assumes that: (1) data are independently and identically distributed, (2) causal functions take the form of a generalized additive model of nonlinear functions, and (3) latent confounders are present. TS-CAM-UV, being a causal function model-based method, can identify causal relationships, provided the data fulfills its assumptions.

Causal discovery methods for time-series data represent the state of variable  $X_i$  at time point  $t$  as  $X_i^t$  treating the states of  $X_i$  at different points such as  $X_i^t, X_i^{t-1}, \dots, X_i^s$  as separate variables. This allows for representing causal relationships between variables at different time points.

Time series causal discovery methods can be described as causal discovery methods that utilize the prior knowledge that effects do not precede their causes in time. Therefore, before proposing the TS-CAM-UV algorithm, this paper proposes a method called CAM-UV with prior knowledge (CAM-UV-PK), which applies prior knowledge to CAM-UV. TS-CAM-UV is proposed as a method that introduces the knowledge that variables representing future states cannot be the cause of variables representing past states. To the best of our knowledge, this is the first method for time series causal discovery that adopts a causal function model approach assuming the presence of latent confounders.

The contributions of this paper are as follows:

- This paper proposes a method called the CAM-UV-PK algorithm, which can introduce prior knowledge in the form of statements such as  $X_i$  cannot be a cause of

- $X_j$ . The performance of the CAM-UV-PK algorithm is verified using simulation data.
- We propose a time-series causal discovery method called the TS-CAM-UV algorithm, which applies the prior knowledge that variables representing future states cannot be causes of variables representing past states. The performance of the TS-CAM-UV algorithm is verified using both simulation data and real-world data.

The remainder of this paper comprises the following. Section 2 reviews previous studies on causal discovery methods for i.i.d. data and time-series data. Section 3 introduces the models of the data generation processes of CAM-UV and TS-CAM-UV, followed by Sect. 4 which shows the identifiability of those models. Section 5 introduces the two proposed methods, the CAM-UV-PK algorithm and the TS-CAM-UV algorithm. Section 6 shows and discusses the results of the experiments of the proposed methods. Section 7 brings the paper to a conclusion.

## 2 Related studies

Causal discovery methods often assume that the causal structures form directed acyclic graphs (DAGs), that there is no latent confounders, and that data are independently and identically distributed (Chickering 2002; Peters et al. 2014; Shimizu et al. 2006, 2011; Spirtes and Glymour 1991). The constrained-based methods including the Peter-Clark (PC) algorithm (Spirtes and Glymour 1991) and the fast causal inference (FCI) algorithm (Spirtes et al. 1999) infer causal relationships on the basis of conditional independence in the joint distribution. FCI identifies the presence of latent confounders whereas PC assumes the absence of unobserved common causes. PC and FCI cannot distinguish between the two causal graphs that entail exactly the same sets of conditional independence. Compared to constrained-based methods, causal functional model-based methods can identify the entire causal models under proper assumptions. Linear non-Gaussian acyclic models (LiNGAM) (Shimizu et al. 2006, 2011) assume that causal relationships are linear and the external effects are non-Gaussian. Additive noise models (ANMs) and causal additive models (Peters et al. 2014) assume the causal relationships are nonlinear. Both LiNGAM and ANMs assume the absence of unobserved variables. Causal additive models with unobserved variables (CAM-UV) (Maeda and Shimizu 2021a) are extended models of causal additive models (CAMs) (Bühlmann et al. 2014) and assume that the causal functions take the form of generalized additive models (GAMs) (Hastie and Tibshirani 1990) and that unobserved variables are present.

Time-series causal discovery methods have been proposed as extensions of the above methods. The time-series FCI (tsFCI) algorithm (Entner and Hoyer 2010) and a structural vector autoregression FCI (SVAR-FCI) (Malinsky and Spirtes 2018) adapt FCI algorithm and use time order and stationarity to infer causal relationships. VAR-LiNGAM (Hyvärinen et al. 2010) is based on LiNGAM and assumes the linearity of causal relationships, non-Gaussianity of external effects, and the absence of unobserved common causes. Time series models with independent noise (TiMINo) (Peters et al. 2013) adapts ANMs, and it assumes the absence of latent confounders. The Peter-

Clark momentary conditional independence (PCMCI) algorithm (Runge et al. 2019) is an adaptation of the conditional independence-based PC algorithm that addresses strong autocorrelations in time series via the use of a momentary conditional independence (MCI) test. Latent PCMCI (LPCMCI) (Gerhardus and Runge 2020) is an extension of PCMCI to include unobserved variables. However, to the best of our knowledge, no causal functional model-based method has been proposed for time-series data under the assumption that causal relationships are nonlinear and latent confounders are present.

### 3 Models

#### 3.1 CAM-UV: causal additive models with unobserved variables

Causal additive noise models with unobserved variables (CAM-UV) (Maeda and Shimizu 2021a, b) are defined as the equation below:

$$V_i = \sum_{X_j \in opa(V_i)} f_{i,j}(X_j) + \sum_{U_j \in upa(V_i)} f_{i,j}(U_j) + N_i \quad \text{with } i = 1, \dots, m, \quad (1)$$

where  $V = \{V_i\}$  is the set of observed or unobserved variables,  $X = \{X_i\}$  the set of observed variables,  $U = \{U_i\}$  is the set of unobserved variables,  $N_i$  is the external effect on  $V_i$ ,  $opa(V_i) \subset X$  is the set of observed direct causes (*observed parents*) of  $V_i$ ,  $upa(V_i) \subset U$  is the set of unobserved direct causes (*unobserved parents*) of  $V_i$ , and  $f_{i,j}$  is a nonlinear function. External effects and unobserved variables both refer to variables that are not included in the data being analyzed while observed variables refer to variables that are included in the data. External effect, denoted as  $N_i$ , is a variable that directly influences only  $V_i$ , while unobserved variables, denoted as  $\{U_i\}$ , are variables that affect multiple observed variables. The indices of the observed variables  $\{X_i\}$  and the unobserved variables  $\{U_i\}$  are the same as the indices of  $\{V_i\}$ . For example, in the form of  $\{X_1, X_2, U_3, U_4, U_5, X_6, \dots, U_m\}$ , the indices of  $\{X_i\}$  and  $\{U_i\}$  are mutually exclusive, and when combined, they constitute all natural number sequences less than or equal to  $m$ . If we rewrite all the observed variables  $\{X_i\}$  and unobserved variables  $\{U_i\}$  as  $\{V_i\}$ , Eq. 1 becomes the following:

$$V_i = \sum_{V_j \in pa(V_i)} f_{i,j}(V_j) + N_i \quad \text{with } i = 1, \dots, m, \quad (2)$$

where  $pa(V_i) = opa(V_i) \cup upa(V_i)$  is the set of the direct causes of  $V_i$ . Additionally, Assumption 1 is imposed on CAM-UV.

**Assumption 1** All the causal functions and the external effects in CAM-UV satisfy the following condition: If variables  $V_i$  and  $V_j$  have terms involving functions of the same external effect  $N_k$ , then  $V_i$  and  $V_j$  are mutually dependent (i.e.,  $(N_k \not\perp\!\!\!\perp V_i) \wedge (N_k \not\perp\!\!\!\perp V_j) \Rightarrow (V_i \not\perp\!\!\!\perp V_j)$ ).

Assumption 1 is satisfied in most cases. To begin with, when  $V_i$  and  $V_j$  are independent, Eq. 3 needs to be satisfied.

$$\text{cov}(V_i, V_j) = \sum_{V_k \in pa(V_i), V_l \in pa(V_j)} \text{cov}(f_{i,k}(V_k), f_{j,l}(V_l)) = 0 \tag{3}$$

Since different external variables are independent of each other, this equation always holds if  $V_i$  and  $V_j$  do not have terms with the same external variables. However, if  $V_i$  and  $V_j$  have terms with the same external variables, in order for this equation to be satisfied, all functions  $f$  containing that external variables inside must meet the conditions that make Eq. 3 equal to 0. Such conditions are only met in quite special cases.

### 3.2 TS-CAM-UV: time series causal additive models with unobserved variables

Time-series causal additive noise models with unobserved variables (TS-CAM-UV) are stationary discrete-time structural causal models that can be described as below:

$$V_i^t = \sum_{X_j^s \in opa(V_i^t)} f_{i,t,j,s}(X_j^s) + \sum_{U_j^s \in upa(V_i^t)} f_{i,t,j,s}(U_j^s) + N_i^t \quad \text{with } i = 1, \dots, m, \tag{4}$$

where  $t$  and  $s$  are time indices,  $m$  is a natural number,  $V = \{V_i^t\}$  is the set of observed or unobserved variables,  $X = \{X_i^t\}$  is the set of observed variables,  $U = \{U_i^t\}$  is the set of unobserved variables,  $f_{i,t,j,s}$  is a nonlinear function, the noise variables  $N_i^t$  are jointly independent,  $opa(V_i^t) \subset X$  is the set of observed direct causes of  $V_i^t$ , and  $upa(V_i^t) \subset U$  is the set of unobserved direct causes of  $V_i^t$ . Similar to Eq. 1, the indices  $i$  of  $\{X_i^t\}$  and  $\{U_i^t\}$  do not overlap with each other, and when the indices of both are combined, they form a sequence of natural numbers less than or equal to  $m$ .

The stationarity of time-series causal relationships is assumed as the following: The causal relationship of the variable pair  $(V_i^{t-\epsilon}, V_j^t)$  is the same as that of all the time shifted pairs  $(V_i^{t'-\epsilon}, V_j^{t'})$ . The causal effect of  $V_j^s$  on  $V_i^t$  is called a lagged effect if  $s < t$  holds, and is called a contemporaneous effect if  $t = s$  holds. It is also assumed that there is a natural number  $r$  as the maximum time lag such that the longest time lag of the direct causal effects does not exceed  $r$ . While it is true that the cause precedes the effect in time, if the time slice of the data analyzed are not sufficiently short, the cause and effect may appear to occur simultaneously. This type of causal effect, where the time difference between the cause and effect is shorter than the time slice of data, is referred to as a contemporaneous effect.

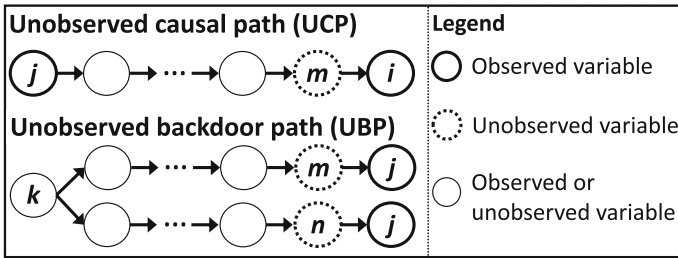


Fig. 1 Definitions of an unobserved causal path (UCP) and an unobserved backdoor path (UBP)

## 4 Identifiability

### 4.1 CAM-UV

The identifiability of CAM-UV is discussed in Maeda and Shimizu (2021a, b), and this section briefly presents it. When the causal relationship is linear, an observed variable  $X_j$  being an indirect cause of an observed variable  $X_i$ , even if there is an unobserved variable  $U_k$  in the causal path such that the causal relationship is  $X_j \rightarrow U_k \rightarrow X_i$ , the residual when regressing  $X_i$  on  $X_j$  becomes independent of  $X_j$ . However, in the case of a non-linear causal relationship, the residual when regressing  $X_i$  on  $X_j$  cannot be independent of  $X_j$ . This is referred to as cascade additive noise models (CANMs) (Cai et al. 2019). Therefore, in the case of non-linear causal relationships, compared to linear ones, there are more instances where causal relationships cannot be identified using only regression and independence tests. Before discussing the cases where causal relationships cannot be identified in CAM-UV, we define unobserved causal paths (UCPs) and unobserved backdoor paths (UBPs) which are illustrated in Fig. 1 and used in the lemmas in this section.

**Definition 1** A directed path from an observed variable to another is called a *causal path* (CP). A CP from  $X_j$  to  $X_i$  is called an *unobserved causal path* (UCP) if it ends with the directed edge connecting  $X_i$  and its unobserved direct cause (i.e.,  $X_j \rightarrow \dots \rightarrow U_m \rightarrow X_i$  where  $U_m$  is an unobserved direct cause of  $X_i$ ).

**Definition 2** An undirected path between  $X_i$  and  $X_j$  is called a *backdoor path* (BP) if it consists of the two directed paths from a common ancestor of  $X_i$  and  $X_j$  to  $X_i$  and  $X_j$  (i.e.,  $X_i \leftarrow \dots \leftarrow V_k \rightarrow \dots \rightarrow X_j$ , where  $V_k$  is the common ancestor). A BP between  $X_i$  and  $X_j$  is called an *unobserved backdoor path* (UBP) if it starts with the edge connecting  $X_i$  and its unobserved direct cause, and ends with the edge connecting  $X_j$  and its unobserved direct cause (i.e.,  $X_i \leftarrow U_m \leftarrow \dots \leftarrow V_k \rightarrow \dots \rightarrow U_n \rightarrow X_j$ , where  $V_k$  is the common ancestor and  $U_m$  and  $U_n$  are the unobserved direct causes of  $X_i$  and  $X_j$ , respectively). The undirected path  $X_i \leftarrow U_k \rightarrow X_j$  is also a UBPs, as  $V_k$ ,  $U_m$ , and  $U_n$  can be the same variable.

The identifiability of CAM-UV is based on Lemmas 1–3 shown below. They show that it is possible to identify the direct causal relationship between two variables if they do not have a UCP or a UBPs, otherwise it is impossible to identify the direct

direct causal relationship but possible to identify the presence of a UCP or a UBP. This is due to the fact that when the causal relationship is non-linear, if the parent of an observed variable  $X_i$  is an unobserved variable  $U_j$ , the ancestral variables of  $U_j$  cannot be removed from  $X_i$  by regression. Lemma 1 is about the condition of variable pair  $(X_i, X_j)$  having a UCP or a UBP. Lemma 2 is about the condition of variable pair  $(X_i, X_j)$  not having a UBP, a UCP, or a direct causal relationship. Lemma 3 is about the condition that  $X_j$  is a direct cause of  $X_i$ , and they do not have a UCP or a UBP. Assumption 2, which is used in Lemmas 1–3, is presented first, followed by Lemmas 1–3. Please refer to Maeda and Shimizu (2021b) for the proofs of the lemmas.

**Assumption 2** Let  $M_1$  and  $M_2$  denote sets satisfying  $M_1 \subseteq X$  and  $M_2 \subseteq X$  where  $X$  is the set of all the observed variables in CAM-UV defined in Sect. 2. We assume that functions  $G_i$  take the forms of generalized additive models (GAMs) (Hastie and Tibshirani 1990) such that  $G_i(M_1) = \sum_{X_m \in M_1} g_{i,m}(X_m)$  where each  $g_{i,m}(X_m)$  is a nonlinear function of  $X_m$ . In addition, we assume that functions  $G_i$  satisfy the following condition: When both  $(X_i - G_i(M_1))$  and  $(X_j - G_j(M_2))$  have terms involving functions of the same external effect  $N_k$ , then  $(X_i - G_i(M_1))$  and  $(X_j - G_j(M_2))$  are mutually dependent (i.e.,  $(N_k \perp\!\!\!\perp X_i - G_i(M_1)) \wedge (N_k \perp\!\!\!\perp X_j - G_j(M_2)) \Rightarrow ((X_i - G_i(M_1)) \perp\!\!\!\perp (X_j - G_j(M_2)))$ ).

**Lemma 1** Assume the data generation process of the variables is CAM-UV as defined in Sect. 3.1. If and only if Eq. 5 is satisfied, there is a UCP or UBP between  $X_i$  and  $X_j$  where  $G_1$  and  $G_2$  denote regression functions satisfying Assumption 2.

$$\begin{aligned} \forall G_1, G_2, M_1 \subseteq (X \setminus \{X_i\}), M_2 \subseteq (X \setminus \{X_j\}), \\ [(X_i - G_1(M_1)) \perp\!\!\!\perp (X_j - G_2(M_2))] \end{aligned} \tag{5}$$

Equation 5 indicates that the residual of  $X_i$  regressed on any subset of  $X \setminus \{X_i\}$  and the residual of  $X_j$  regressed on any subset of  $X \setminus \{X_j\}$  cannot be mutually independent.

**Lemma 2** Assume the data generation process of the variables is CAM-UV as defined in Sect. 3.1. If and only if Eq. 6 is satisfied, there is no direct causal relationship between  $X_i$  and  $X_j$ , and there is no UCP or UBP between  $X_i$  and  $X_j$  where  $G_1$  and  $G_2$  denote regression functions satisfying Assumption 2.

$$\begin{aligned} \exists G_1, G_2, M \subseteq (X \setminus \{X_i, X_j\}), N \subseteq (X \setminus \{X_i, X_j\}), \\ [((X_i - G_1(M)) \perp\!\!\!\perp (X_j - G_2(N)))] \end{aligned} \tag{6}$$

Equation 6 indicates that there are regression functions such that the residuals of  $X_i$  and  $X_j$  regressed on subsets of  $X \setminus \{X_i, X_j\}$  are mutually independent.

**Lemma 3** Assume the data generation process of the variables is CAM-UV as defined in Sect. 3.1. If and only if Eqs. 7 and 8 are satisfied,  $X_j$  is a direct cause of  $X_i$ , and there is no UCP or UBP between  $X_i$  and  $X_j$  where  $G_1$  and  $G_2$  denote regression functions satisfying Assumption 2.

$$\forall G_1, G_2, M \subseteq (X \setminus \{X_i, X_j\}), N \subseteq (X \setminus \{X_j\}), \tag{7}$$

$$[(X_i - G_1(M)) \perp\!\!\!\perp (X_j - G_2(N))]$$

$$\exists G_1, G_2, M \subseteq (X \setminus \{X_i\}), N \subseteq (X \setminus \{X_i, X_j\}), \tag{8}$$

$$[(X_i - G_1(M)) \perp\!\!\!\perp (X_j - G_2(N))]$$

Equation 7 indicates that the residual of  $X_i$  regressed on any subset of  $X \setminus \{X_i, X_j\}$  and the residual of  $X_j$  regressed on any subset of  $X \setminus \{X_j\}$  cannot be mutually independent. Equation 8 indicates that there are regression functions such that the residual of  $X_i$  regressed on a subset of  $X \setminus \{X_j\}$  and the residual of  $X_j$  regressed on a subset of  $X \setminus \{X_i, X_j\}$  are mutually independent.

### 4.2 TS-CAM-UV

The identifiability of causality in TS-CAM-UV is the same as in CAM-UV. Lemmas 4–6 on identifiability in TS-CAM-UV correspond to Lemmas 1–3 on identifiability in CAM-UV.

**Lemma 4** Assume the data generation process of the variables is TS-CAM-UV as defined in Sect. 3.2. If and only if Eq. 9 is satisfied, there is a UCP or UBP between  $X_i^t$  and  $X_j^s$  where  $G_1$  and  $G_2$  denote regression functions satisfying Assumption 2.

$$\forall G_1, G_2, M \subseteq (X \setminus \{X_i^t\}), N \subseteq (X \setminus \{X_j^s\}), \tag{9}$$

$$[(X_i^t - G_1(M)) \perp\!\!\!\perp (X_j^s - G_2(N))]$$

**Proof** The relationships between  $X_i^t$  and  $X_j^s$  in TS-CAM-UV are the same as those of  $X_i$  and  $X_j$  in CAM-UV defined in Sect. 3.1. Therefore, Lemma 4 holds because of Lemma 1. □

**Lemma 5** Assume the data generation process of the variables is TS-CAM-UV as defined in Sect. 3.2. If and only if Eq. 10 is satisfied, there is no direct causal relationship between  $X_i^t$  and  $X_j^s$ , and there is no UCP or UBP between  $X_i^t$  and  $X_j^s$  where  $G_1$  and  $G_2$  denote regression functions satisfying Assumption 2.

$$\exists G_1, G_2, M \subseteq (X \setminus \{X_i^t, X_j^s\}), N \subseteq (X \setminus \{X_i^t, X_j^s\}), \tag{10}$$

$$[((X_i^t - G_1(M)) \perp\!\!\!\perp (X_j^s - G_2(N)))]$$

**Proof** The relationships between  $X_i^t$  and  $X_j^s$  in TS-CAM-UV are the same as those of  $X_i$  and  $X_j$  in CAM-UV defined in Sect. 3.1. Therefore, Lemma 5 holds because of Lemma 2. □

**Lemma 6** Assume the data generation process of the variables is TS-CAM-UV as defined in Sect. 3.2. If and only if Eqs. 11 and 12 are satisfied,  $X_j^s$  is a direct cause



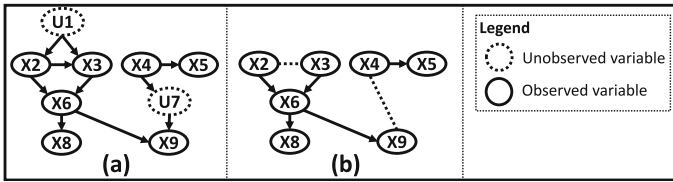


Fig. 2 a True causal graph. b Causal graph generated by the CAM-UV algorithm

of  $X_i^t$ , and there is no UCP or UBP between  $X_i^t$  and  $X_j^s$  where  $G_1$  and  $G_2$  denote regression functions satisfying Assumption 2.

$$\forall G_1, G_2, M \subseteq (X \setminus \{X_i^t, X_j^s\}), N \subseteq (X \setminus \{X_i^t\}), \quad (11)$$

$$\left[ (X_i^t - G_1(M)) \perp\!\!\!\perp (X_j^s - G_2(N)) \right]$$

$$\exists G_1, G_2, M \subseteq (X \setminus \{X_i^t\}), N \subseteq (X \setminus \{X_i^t, X_j^s\}), \quad (12)$$

$$\left[ (X_i^t - G_1(M)) \perp\!\!\!\perp (X_j^s - G_2(N)) \right]$$

**Proof** The relationships between  $X_i^t$  and  $X_j^s$  in TS-CAM-UV are the same as those of  $X_i$  and  $X_j$  in CAM-UV defined in Sect. 3.1. Therefore, Lemma 6 holds because of Lemma 3.  $\square$

## 5 Methods

### 5.1 CAM-UV-PK: causal additive models with unobserved variables using prior knowledge

This section proposes a method called CAM-UV using prior knowledge (CAM-UV-PK). This method is for discovering causal additive models with unobserved models defined in Sect. 3.1. In addition to the arguments of the CAM-UV algorithm, the CAM-UV-PK algorithm requires an argument  $\mathbf{T}$  that is a list of ordered variable pairs. If an ordered variable pair  $(X_i, X_j)$  is included in  $\mathbf{T}$ , it means that it is assumed that  $X_i$  cannot be a direct or indirect cause of  $X_j$ .

The CAM-UV algorithm and CAM-UV-PK algorithm output causal graphs with directed edges and undirected dashed edges. Directed edges indicate variable pairs having direct causal relationships, and undirected dashed edges indicate variable pairs having UCPs or UBPs. For example, Fig. 2a shows a true causal graph, and Fig. 2b shows the causal graph generated by the CAM-UV algorithm.  $X_2$  and  $X_3$  have a UBP ( $X_2 \leftarrow U_1 \rightarrow X_3$ ), so they are connected with an undirected dashed path in Fig. 2b.  $X_4$  and  $X_9$  have a UCP ( $X_4 \rightarrow U_7 \rightarrow X_9$ ), so they are also connected with an undirected dashed path in Fig. 2b.

The CAM-UV-PK algorithm incorporates restriction using prior knowledge  $\mathbf{T}$  into the process of causal inference in the CAM-UV algorithm. The CAM-UV algorithm

**Algorithm 1:** Determine the directed edges

```

Input:  $X$ : Samples of  $p$  observed variables  $\{X_1, \dots, X_p\}$ ,  $T$ : A list of ordered pairs of
variables where it is assumed that the first variable cannot be the direct or indirect cause
of the second variable,  $d$ : maximal number of variables to examine causality for each
step, significance level for independence test  $\alpha$ .

Output: the sets of the parents  $\{M_1, \dots, M_p\}$ .

1 function getDirectedEdges( $X, d, \alpha$ )
2   # PHASE 1: Extracting the candidates of the parents of each variable.
3   for  $i = 1$  to  $p$  do
4     Initialize  $M_i \leftarrow \emptyset$ .
5   Initialize  $t \leftarrow 2$ .
6   while  $t \leq d$  do
7     Initialize  $noChange \leftarrow \text{True}$ .
8     foreach  $K \in \{K | K \subseteq X, |K| = t\}$  do
9       # Finding the most endogenous variable  $X_b$  in  $K$ 
10      foreach  $K \in \{K | K \subseteq X, |K| = t\}$  do
11         $maxIndependence \leftarrow 0$ 
12         $maxIndependenceVariable \leftarrow \text{NULL}$ 
13        foreach  $X_i \in K$  do
14          # Checking whether there exists variable  $X_j \in K \setminus \{X_i\}$  that cannot be a
15          # cause of  $X_i$  according to prior knowledge  $T$ .
16          if  $\exists X_j \in K \setminus \{X_i\}, [(X_j, X_i) \in T]$  then
17            continue
18           $indep_e \leftarrow \widehat{p\text{-HSIC}}(X_i - G_1(M_i \cup K \setminus \{X_i\}), \{X_j - G_2(M_j) | X_j \in K \setminus \{X_i\}\})$ 
19          if  $maxIndependence < indep_e$  then
20             $maxIndependence \leftarrow indep_e$ 
21             $maxIndependenceVariable \leftarrow X_i$ 
22           $X_b \leftarrow maxIndependenceVariable$ 
23          # Computing the independence between the residuals
24           $e \leftarrow \widehat{p\text{-HSIC}}(X_b - G_1(M_b \cup K \setminus \{X_b\}), \{X_j - G_2(M_j) | X_j \in K \setminus \{X_b\}\})$ 
25           $h \leftarrow \max_{x_j \in K \setminus \{X_b\}} \widehat{p\text{-HSIC}}(X_b - G_1(M_b), X_j - G_2(M_j))$ 
26          # Checking whether  $X_b$  is really a sink of  $K$ 
27          if  $(\alpha < e) \wedge (\alpha > h)$  then
28            # When  $X_b$  is a sink of  $K$ , add each variable in  $K \setminus \{X_b\}$  to  $M_b$ 
29             $M_b \leftarrow M_b \cup (K \setminus \{X_b\})$ 
30             $noChange \leftarrow \text{False}$ 
31          # If each  $M_i$  remains unchanged, increment  $t$  by one. If not, substitute 2 for  $t$ .
32          if  $noChange = \text{True}$  then
33             $t \leftarrow t + 1$ 
34          else
35             $t \leftarrow 2$ 
36          # PHASE 2: Determining the parents of each variable.
37          for  $i = 1$  to  $p$  do
38            foreach  $X_j \in M_i$  do
39              # Checking whether  $X_j$  is parent of  $X_i$ 
40              if  $\alpha < \widehat{p\text{-HSIC}}(X_i - G_1(M_i \setminus \{X_j\}), X_j - G_2(M_j))$  then
41                # When  $X_j$  is not a parent, remove it from  $M_i$ 
42                 $M_i \leftarrow M_i \setminus \{X_j\}$ 
43          return  $\{M_1, \dots, M_p\}$ 

```

has two-step algorithm (Maeda and Shimizu 2021a, b). The first step determines the directed edges, and the second one determines the undirected dashed edges. There is no difference in the second step between the CAM-UV-PK algorithm and the CAM-UV algorithm. The first step of the CAM-UV-PK algorithm is listed in Algorithm 1. Lines 14–16 in Algorithm 1 are added to the CAM-UV algorithm. This part of the algorithm refers to the prior knowledge  $\mathbf{T}$  to avoid considering unnecessary causal candidates. The method extracts the candidates of the direct causes (parents) of each variable (lines 2–34) and determines the direct causes of each variables (lines 35–41). The method identifies the most endogenous variable  $X_b$  in each  $K \in \{K | K \subseteq X, |K| = t\}$ . When  $X_i = X_b$  is satisfied,  $X_i$  maximizes  $\widehat{p\text{-HSIC}}(X_i - G_1(M_i \cup K \setminus \{X_i\}), \{X_j - G_2(M_j) | X_j \in K \setminus \{X_i\}\})$ .  $G_1$  and  $G_2$  are determined by the GAM regression method proposed in Wood (2004).  $\widehat{p\text{-HSIC}}(A, B)$  is the the  $p$ -value of Hilbert-Schmidt Independence Criteria (HSIC) (Gretton et al. 2008). HSIC is a metric that captures the nonlinear dependencies between variables, and higher values of  $\widehat{p\text{-HSIC}}$  indicate a stronger level of independence between the variables. In lines 14–16 which are newly added in CAM-UV-PK, the method checks whether there exists  $X_j \in K \setminus \{X_i\}$  that cannot be a direct or indirect cause of  $X_i$  according to the prior knowledge  $\mathbf{T}$ . If  $(X_j, X_i) \in \mathbf{T}$  is satisfied, the method stops checking whether  $X_i$  is endogenous to  $K \setminus \{X_i\}$ . Therefore, this check prevents incorrect inference of causal relationships.

### 5.2 TS-CAM-UV: time series causal additive models with unobserved variables

This section proposes a method called the time-series CAM-UV (TS-CAM-UV) algorithm. The TS-CAM-UV algorithm uses as prior knowledge the assumption, called time priority, that effect does not precede its cause in time. The TS-CAM-UV algorithm uses the CAM-UV-PK algorithm, and the prior knowledge of time priority is used for the argument of the CAM-UV-PK,  $\mathbf{T}$ .

The TS-CAM-UV algorithm first creates data with  $q \times (r + 1)$  variables where  $q$  is the number of the variables of original data, and  $r$  is the maximal considered time lag given as an argument. Let  $\mathbf{X}_t = \{X_1^t, \dots, X_q^t\}$  denote the variables in original data. The TS-CAM-UV algorithm creates data with variables  $\mathbf{X}^{\text{new}} = \{X_1^t, \dots, X_q^t, X_1^{t-1}, \dots, X_q^{t-1}, \dots, X_1^{t-r}, \dots, X_q^{t-r}\}$ . Equations 13 and 14 represent the original data and the new data in matrix form, respectively. The matrix of the original data is named  $D_{\text{original}}$ , and the matrix of the new data is named  $D_{\text{new}}$ . Each row of these matrices corresponds to an observation, and each column corresponds to a variable. If the number of observations in the original data is  $n$ , the number of rows in  $D_{\text{new}}$  cannot exceed  $n - r$ . This is because each row stores the values of the same variable from time point  $t$  to time point  $t - r$ . The TS-CAM-UV algorithm creates data with  $n - r$  rows.

$$D_{\text{original}} = \left. \begin{bmatrix} x_1^1 & \cdots & x_q^1 \\ \vdots & & \vdots \\ x_1^n & \cdots & x_q^n \end{bmatrix} \right\} n \text{ rows} \tag{13}$$

$$D_{\text{new}} = \left. \begin{bmatrix} x_1^{r+1} & \cdots & x_q^{r+1} & \cdots & x_1^1 & \cdots & x_q^1 \\ \vdots & & \vdots & & \vdots & & \vdots \\ x_1^n & \cdots & x_q^n & \cdots & x_1^{n-r} & \cdots & x_q^{n-r} \end{bmatrix} \right\} n - r \text{ rows} \tag{14}$$

The TS-CAM-UV algorithm also creates a list of ordered variables  $K = \{(X_i^{t'}, X_j^{t'}) | t > t', 1 \leq i \leq q, 1 \leq j \leq q, \}$ .

The TS-CAM-UV algorithm uses  $\mathbf{X}^{\text{new}}$  and  $K$  for the arguments of CAM-UV-PK  $\mathbf{X}$  and  $\mathbf{T}$ , respectively. Then, CAM-UV-PK outputs a causal graph of the  $q$  variables with  $r$  time lag.

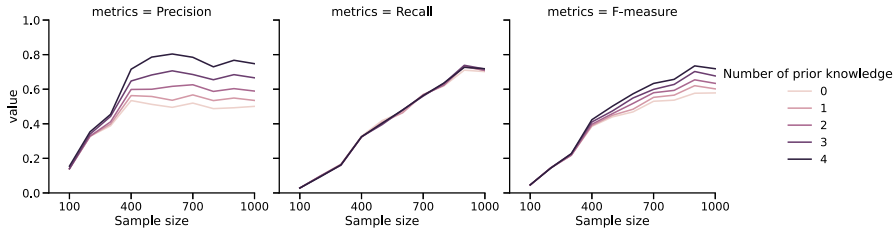
## 6 Experiments

We conducted experiments to examine the performance of the CAM-UV-PK algorithm and the TS-CAM-UV algorithm. The CAM-UV-PK algorithm is compared with that of CAM-UV. The TS-CAM-UV algorithm is compared with VarLiNGAM and LPCMCI. Here, we primarily compare the accuracy of directed edges. This is because, in other methods, there are no approaches that consider the effects of unobserved intermediate variables (unobserved variables on the causal paths between observed variables), and also because CAM-UV aims to ensure that the inference of directed edges is not biased due to latent confounders.

### 6.1 CAM-UV-PK: causal additive models with unobserved variables using prior knowledge

We examined the performance of CAM-UV-PK compared to CAM-UV using simulated data. We compared and evaluated the performance of CAM-UV-PK with prior knowledge ranging from 0 to 4. The CAM-UV algorithm is the same as the CAM-UV-PK algorithm with no input of prior knowledge. We performed 100 experiments using artificial data with each sample size  $n \in \{100, 200, \dots, 900, 1000\}$  to compare our method to existing methods. In each experiment, the samples are created as follows:

- The number of observed variables is 10.
- The number of the observed variable pairs having unobserved common causes is 4.
- The number of observed variable pairs having unobserved causal intermediate variables is 2.
- The number of the observed variable pairs having direct causal effects is 10.
- Variable pairs having unobserved common causes, unobserved intermediate causal variables, or direct causal relationships were randomly selected under the restriction that the set of variable pairs with unobserved common causes, the set of



**Fig. 3** The performance of the CAM-UV-PK and CAM-UV algorithms: the CAM-UV algorithm is equivalent to the CAM-UV-PK algorithm with no prior knowledge

variable pairs with unobserved intermediate causal variables, and the set of variable pairs with direct causal relationships were mutually disjoint.

- The causal effect of  $V_j$  on  $V_i$  is determined as follows:

$$(\sin(a_1(V_j + b_1)))^3 c_1 + \left( \frac{1}{1 + \exp(-a_2(V_j + b_2))} - 0.5 \right) c_2 \quad (15)$$

where  $a_1, a_2, b_1, b_2, c_1,$  and  $c_2$  are constants that take random value for each  $(i, j)$ . Constants  $a_1$  and  $a_2$  are taken from  $U(9, 11)$ ,  $b_1$  and  $b_2$  are taken from  $U(-0.1, 0.1)$ , and  $c_1$  and  $c_2$  are taken from  $U(3, 5)$ . This function is also used in experiments to validate the TS-CAM-UV algorithm in the next section so that causal effects do not converge or diverge over time.

The arguments of TS-CAM-UV,  $\alpha$  (significance level for independence test) and  $d$  (maximal number of variables to examine causality for each step) are set to 0.01 and 2, respectively.

We compared the performance of the identification of direct causal relationships. We used precision, recall, and F-measure as the evaluation measures. True positive (TP) is the number of true directed edges that a method correctly infers in terms of their positions and directions. Precision represents the TP divided by the number of estimations, and recall represents the TP divided by the number of all the true directed edges. Furthermore, F-measure is defined as  $F\text{-measure} = 2 \cdot \text{precision} \cdot \text{recall} / (\text{precision} + \text{recall})$ . In each experiment, out of the ten variable pairs with direct causal relationships, four were excluded from the evaluation. These four causal relationships were used as prior knowledge in CAM-UV-PK.

Figure 3 shows the results of the identification of direct causal relationships. The figure plots the average of precision, recall, and F-measure. It can be seen that precision and F-measure increase with the number of prior knowledge. The CAM-UV algorithm is the CAM-UV-PK algorithm without prior knowledge, and this has the lowest precision and F-measure. The number of prior knowledge does not significantly affect recall. When the sample size increases from 900 to 1000, all metric values decrease. This may be attributed to reaching the upper limit of performance around this sample size range.

The above experimental results of the CAM-UV-PK algorithm confirm that the number of prior knowledge improves the precision and F-measure of the identification of direct causal relationships.

## 6.2 TS-CAM-UV: time series causal additive models with unobserved variables

We examined the performance of TS-CAM-UV compared to LPCMCI and VarLiNGAM using simulated data and real-world data. For LPCMCI, two methods of conditional independence test were used for the comparison: Partial correlation test (ParCorr) and Gaussian process regression and a distance correlation test on the residuals (GPDC). ParCorr assumes linear additive noise models, and GPDC assumes nonlinear additive noise models.

### 6.2.1 Simulated data

We performed 100 experiments using artificial data with each sample size  $n \in \{100, 200, \dots, 1900, 2000\}$  to compare our method to existing methods. In each experiment, the samples are created as follows:

- The number of observed variables and the maximum time lag are 3 and 2, respectively. Therefore, the number of the variables representing different time lags of all the observed variables is 9 (e.g.  $|\{X_i^t\}| = 9$ ).
- The number of observed variable pairs having unobserved common causes is 2.
- The number of observed variable pairs having unobserved intermediate variables is 2.
- The number of observed variable pairs having direct causal relationships is 5.
- Variable pairs having unobserved common causes, unobserved intermediate causal variables, or direct causal relationships were randomly selected under the restriction that the set of variable pairs with unobserved common causes, the set of variable pairs with unobserved intermediate causal variables, and the set of variable pairs with direct causal relationships were mutually disjoint.
- The causal effect of  $V_j^s$  on  $V_i^t$  is determined as below:

$$\left(\sin\left(a_1\left(V_j^s + b_1\right)\right)\right)^3 c_1 + \left(\frac{1}{1 + \exp(-a_2(V_j^s + b_2))} - 0.5\right) c_2 \quad (16)$$

where  $a_1, a_2, b_1, b_2, c_1,$  and  $c_2$  are constants that take random value for each  $(i, j, t, t')$ . Constants  $a_1$  and  $a_2$  are taken from  $U(9, 11)$ ,  $b_1$  and  $b_2$  are taken from  $U(-0.1, 0.1)$ , and  $c_1$  and  $c_2$  are taken from  $U(3, 5)$ .

In this experiment, we compared the performance of the identification of direct causal relationships. That is, the edges with arrows in causal graphs ( $\rightarrow$ ).

The arguments of the TS-CAM-UV algorithm, VarLiNGAM, and LPCMCI were set as follows:

- TS-CAM-UV
  - Significance level for independence test: 0.01.
  - Maximal number of causal variables to examine causality for each step: 2.
  - Maximal number of time lags: 2.
- VarLiNGAM

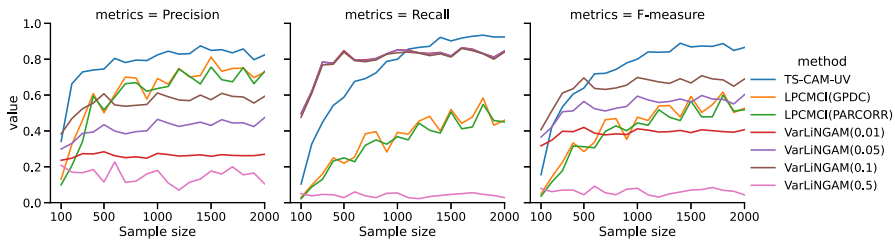


Fig. 4 The performance of the TS-CAM-UV compared to LPCMCI and VarLiNGAM

- Maximal number of time lags: 2.
  - Threshold value for the strength of the causal effects (i.e. the absolute values of coefficients): 0.01, 0.05, 0.1, and 0.5.
- LPCMCI
- Significance level for independence test: 0.01.
  - Maximal number of time lags: 2.
  - Methods of conditional independence test: GPDC and ParCorr.

The results are shown in Fig. 4. The figure plots the average of precision, recall, and F-measure. The values in the brackets for VarLiNGAM indicate threshold values for the strength of causal effects. TS-CAM-UV showed the highest precision for  $n \geq 200$ , the highest recall for  $n \geq 1200$ , and the highest F-measure for  $n \geq 600$  compared to other methods.

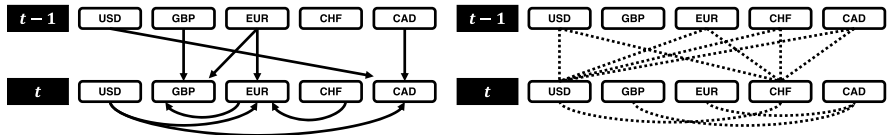
### 6.2.2 Real world data

We also conducted an experiment using official foreign exchange quotation data for the Japanese yen at Mizuho Bank.<sup>1</sup> The data consist of daily quotes for USD, GBP, EUR, CHF, and CAD from the 26th October 2021 to the 8th November 2023. The total sample size is 500.

We set the maximal lag length of every method to 1. The threshold value for causal effects for VarLiNGAM was set to 0.1 which gave the best result in experiments using simulated data shown in Sect. 6.2.1. All other arguments were kept the same as in Sect. 6.2.1.

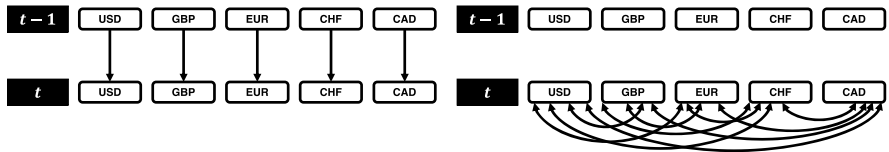
Figure 5 shows the results: (a) the causal graph only with directed edges generated from TS-CAM-UV, (b) the causal graph with edges other than directed edges generated from TS-CAM-UV, (c) the causal graph only with directed edges generated from LPCMCI using ParCorr, (d) the causal graph with edges other than directed edges generated from LPCMCI using ParCorr, (e) the causal graph only with directed edges generated from LPCMCI using GPDC, (f) the causal graph with edges other than directed edges generated from LPCMCI using GPDC, and (g) the causal graph only with directed edges generated from VarLiNGAM. The dashed lines in Fig. 5b show the variable pairs estimated to have UBPs or UCPS. The bidirected edges in Fig. 5d,

<sup>1</sup> Mizuho Bank: <https://www.mizuho.co.jp/market/historical.html> (in Japanese).



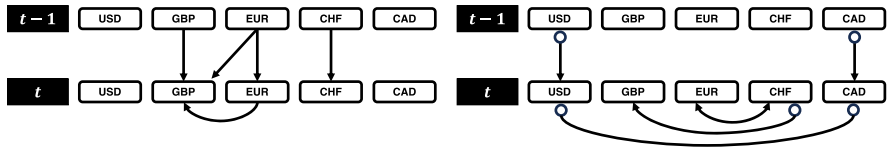
(a) The causal graph only with directed edges generated from TS-CAM-UV.

(b) The causal graph only with edges other than directed edges generated from TS-CAM-UV.



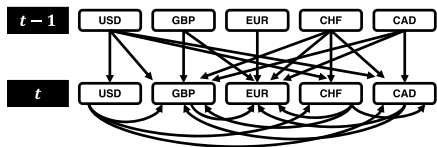
(c) The causal graph only with directed edges generated from LPCMCI (ParCorr).

(d) The causal graph with edges other than directed edges generated from LPCMCI (ParCorr).



(e) The causal graph only with directed edges generated from LPCMCI (GPDC).

(f) The causal graph with edges other than directed edges generated from LPCMCI (GPDC).



(g) The causal graph generated from TS-VarLiNGAM.

**Fig. 5** Causal graphs generated using foreign exchange data

g indicate the presence of unobserved common causes. The circles in Fig. 5f indicate that they can be tails or arrows.

We do not compare the performance of the methods based on the results because there is no ground truth for the relationships among the variables. Patton (2006) demonstrated that exchange rates between currencies have an asymmetric structure, which can change given a certain trigger. The structure may not satisfy the assumption of time stationarity if such a trigger occurs within the period of the data. In this study, we conduct experiments under the assumption that time stationarity holds. However, if an extended method that incorporates non-stationarity models can be developed, further experiments will be necessary when that extension is realized in the future.



We compare TS-CAM-UV with other methods to see the types and number of variable pairs that are connected. Figure 5c shows that LPCMCI (ParCorr) draws an edge from the variable representing the state at time  $t - 1$  of each currency to the variable representing the state at time  $t$  of the currency (e.g.  $X_i^{t-1} \rightarrow X_i^t$ ), but does not draw edges between variables of different currencies. Compared to this, Fig. 5a shows that TS-CAM-UV connects the variables of different currencies with directed edges. This may be due to the fact that ParCorr assumes linear causal relationships. The causal relationship between the previous and current values of the same currency may be linear, while other causal relationships may be nonlinear. Figure 5e shows LPCMCI (GPDC) connects the variables of different currencies with directed edges. The number of variable pairs connected by LPCMCI (GPDC) is less than the number of variable pairs connected by TS-CAM-UV. This may be due to the fact that LPCMCI is a constraint-based method and cannot distinguish between all graphs with the same set of conditional independence between observed variables. Constraint-based methods infer causal relationships from conditional independence. By their very nature, even if all the tests conducted by these methods make accurate inferences, there may still be pairs of variables with causal relationships that cannot be determined, depending on the true underlying causal graph. Figure 5g shows that VarLiNGAM connects more variable pairs with directed edges than TS-CAM-UV. This may be due to the fact that VarLiNGAM assumes the absence of latent confounders.

To summarize, the TS-CAM-UV algorithm is based on a causal functional model, which enables it to identify the direction of causality in variable pairs that LPCMCI could not orient. Furthermore, by assuming the presence of unobserved variables, it can avoid incorrect orientations, similar to what occurs with VarLiNGAM.

## 7 Conclusion

In this paper, we propose two methods as extensions of CAM-UV: CAM-UV-PK and TS-CAM-UV. The CAM-UV-PK algorithm employs a method that introduces prior knowledge in the form that a certain variable is not a cause of a certain other variable. This is based on the CAM-UV algorithm, which infers causal variables for each observed variable. TS-CAM-UV uses time priority as prior knowledge for CAM-UV-PK, indicating that variables occurring later in time cannot be the cause of earlier variables. To the best of our knowledge, this is the first method for time series causal discovery that adopts a causal function model approach assuming the presence of latent confounders. If the data being analyzed satisfy the assumption that the causal function takes the form of a generalized additive model, then this proposed method can accurately infer causal relationships even in the presence of latent confounders.

Future research will extend our approach to models where the causal graph contains cycles. If the time for the causal effect from the cause variable to the effect variable is shorter than the time slice of the data being analyzed, this causal effect becomes a contemporaneous effect. When there is a causal relationship such as  $X_i^{t-2} \rightarrow X_j^{t-1} \rightarrow X_i^t$ , and the time slice of the data is longer than this causal effect, it results in a contemporaneous effect with cycles. Therefore, future research explore causal discovery methods that allow for cycles in contemporaneous effects. As a reviewer pointed out, TS-

CAM-UV may also be able to be extended to handle time series data from multiple subjects, i.e., longitudinal data. However, distinguishing between time-varying and time-invariant hidden confounders would generally be difficult.

**Acknowledgements** Takashi Nicholas Maeda has been partially supported by Grant-in-Aid for Scientific Research (C) from Japan Society for the Promotion of Science (JSPS) #23K16951. Shohei Shimizu has been partially supported by ONR N00014-20-1-2501, JST CREST JPMJCR22D2, and Grant-in-Aid for Scientific Research (C) from Japan Society for the Promotion of Science (JSPS) #20K11708.

**Funding** Takashi Nicholas Maeda has been partially supported by Grant-in-Aid for Scientific Research (C) from Japan Society for the Promotion of Science (JSPS) #23K16951. Shohei Shimizu has been partially supported by ONR N00014-20-1-2501, JST CREST JPMJCR22D2, and Grant-in-Aid for Scientific Research (C) from Japan Society for the Promotion of Science (JSPS) #20K11708.

**Data availability** We used the official foreign exchange rates for the Japanese yen from Mizuho Bank. This data can be accessed through Mizuho Bank's website at: <https://www.mizuhoibank.co.jp/market/historical.html> (available in Japanese).

## Declarations

**Conflict of interest** The authors declare no Conflict of interest associated with this manuscript.

## References

- Bühlmann P, Peters J, Ernest J (2014) CAM: causal additive models, high-dimensional order search and penalized regression. *Ann Stat* 42(6):2526–2556
- Cai R, Qiao J, Zhang K, Zhang Z, Hao Z (2019) Causal discovery with cascade nonlinear additive noise model. In: Proceedings of the twenty-eighth international joint conference on artificial intelligence Organization, IJCAI-19, pp 1609–1615
- Chickering DM (2002) Optimal structure identification with greedy search. *J Mach Learn Res* 3:507–554
- Chu T, Glymour C (2008) Search for additive nonlinear time series causal models. *J Mach Learn Res* 9(32):967–991
- Ebert-Uphoff I, Deng Y (2012) Causal discovery for climate research using graphical models. *J Clim* 25(17):5648–5665
- Entner D, Hoyer PO (2010) On causal discovery from time series data using FCI. In: Proceedings of the 5th European workshop on probabilistic graphical models
- Gerhardus A, Runge J (2020) High-recall causal discovery for autocorrelated time series with latent confounders. In: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H (eds) *Advances in neural information processing systems*, vol 33. Curran Associates, Inc., pp 12615–12625
- Gretton A, Fukumizu K, Teo CH, Song L, Schölkopf B, Smola AJ (2008) A kernel statistical test of independence. In: Platt JC, Koller D, Singer Y, Roweis ST (eds) *Advances in neural information processing systems* 20. Curran Associates, Inc., pp 585–592
- Hastie TJ, Tibshirani RJ (1990) *Generalized additive models*. Chapman and Hall/CRC, Boca Raton
- Hyvärinen A, Zhang K, Shimizu S, Hoyer PO (2010) Estimation of a structural vector autoregression model using non-Gaussianity. *J Mach Learn Res* 11(56):1709–1731
- Lai P-C, Bessler DA (2015) Price discovery between carbonated soft drink manufacturers and retailers: a disaggregate analysis with PC and LiNGAM algorithms. *J Appl Econ* 18(1):173–197
- Maeda TN, Shimizu S (2021) Causal additive models with unobserved variables. In: de Campos C, Maathuis MH (eds) *Proceedings of the thirty-seventh conference on uncertainty in artificial intelligence*, Proceedings of machine learning research, PMLR, vol 161, pp 97–106
- Maeda TN, Shimizu S (2021) Discovery of causal additive models in the presence of unobserved variables. arXiv preprint [arXiv:2106.02234](https://arxiv.org/abs/2106.02234)
- Malinsky D, Spirtes P (2018) Causal structure learning from multivariate time series in settings with unmeasured confounding. In: Le TD, Zhang K, Kiciman E, Hyvärinen A, Liu L (eds) *Proceedings*

- of 2018 ACM SIGKDD workshop on causal discovery, Proceedings of Machine Learning Research, PMLR, vol 92, pp 23–47
- Patton AJ (2006) Modelling asymmetric exchange rate dependence. *Int Econ Rev* 47(2):527–556
- Peters J, Mooij JM, Janzing D, Schölkopf B (2014) Causal discovery with continuous additive noise models. *J Mach Learn Res* 15(1):2009–2053
- Peters J, Janzing D, Schölkopf B (2013) Causal inference on time series using restricted structural equation models. In: Burges C, Bottou L, Welling M, Ghahramani Z, Weinberger K (eds) *Advances in neural information processing systems*, vol 26. Curran Associates, Inc
- Runge J, Nowack P, Kretschmer M, Flaxman S, Sejdinovic D (2019) Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci Adv* 5:11
- Shimizu S, Hoyer PO, Hyvärinen A, Oca Kerminen A (2006) A linear non-Gaussian acyclic model for causal discovery. *J Mach Learn Res* 7:2003–2030
- Shimizu S, Inazumi T, Sogawa Y, Hyvärinen A, Kawahara Y, Washio T, Hoyer PO, Apr Bollen K (2011) DirectLiNGAM: a direct method for learning a linear non-Gaussian structural equation model. *J Mach Learn Res* 12:1225–1248
- Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Ramsey JD, Woolrich MW (2011) Network modelling methods for FMRI. *Neuroimage* 54(2):875–891
- Spirtes P, Glymour C (1991) An algorithm for fast recovery of sparse causal graphs. *Soc Sci Comput Rev* 9(1):62–72
- Spirtes P, Meek C, Richardson T (1999) Causal discovery in the presence of latent variables and selection bias. In: Cooper GF, Glymour CN (eds) *Computation, causality, and discovery*. AAAI Press, pp 211–252
- Wood SN (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J Am Stat Assoc* 99(467):673–686
- Zheng X, Aragam B, Ravikumar PK, Xing EP (2018) DAGs with NO TEARS: continuous optimization for structure learning. In: Bengio S, Wallach S, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R (eds) *Advances in neural information processing systems*, vol 31. Curran Associates, Inc

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.