

# On Granger causality and the effect of interventions in time series

Michael Eichler · Vanessa Didelez

Received: 8 December 2008 / Accepted: 6 November 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** We combine two approaches to causal reasoning. Granger causality, on the one hand, is popular in fields like econometrics, where randomised experiments are not very common. Instead information about the dynamic development of a system is explicitly modelled and used to define potentially causal relations. On the other hand, the notion of causality as effect of interventions is predominant in fields like medical statistics or computer science. In this paper, we consider the effect of external, possibly multiple and sequential, interventions in a system of multivariate time series, the Granger causal structure of which is taken to be known. We address the following questions: under what assumptions about the system and the interventions does Granger causality inform us about the effectiveness of interventions, and when does the possibly smaller system of observable times series allow us to estimate this effect? For the latter we derive criteria that can be checked graphically and are in the same spirit as Pearl's back-door and front-door criteria (Pearl 1995).

**Keywords** Causality · Causal inference · Decision theory · Graphical models · Time series

---

M. Eichler  
Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht,  
The Netherlands

V. Didelez (✉)  
Department of Mathematics, University of Bristol, University Walk, BS8 1TW Bristol, UK  
e-mail: vanessa.didelez@bristol.ac.uk

## 1 Introduction

In epidemiology and related fields, causal inference is predominantly regarded as a decision problem: one of several treatments can potentially be administered and the causal question is whether the choice of treatment makes any difference to some outcome, or typically to the average outcome, and if yes, which is the best treatment. The word ‘treatment’ is used as a generic term here, it comprises also, for instance, the manner in which a drug is given (e.g. pills or injection, by a nurse or by a relative etc.) and does not need to be of medical nature at all (e.g. an educational programme). [Rubin \(1974, 1978\)](#) stresses that the various treatments to be compared need to be well defined in the sense that it must be clear how an individual or object under investigation can potentially receive *any* of the treatments. This implies that we must at least in principle be able to conceive of a randomised study (even if impractical or unethical) where each subject receives a randomly chosen treatment out of those to be compared. An illuminating discussion of this issue in the context of a public health application can be found in [Hernán and Taubman \(2008\)](#). Restricting the definition of a treatment in this way precludes, in most contexts, investigation of variables like sex or birth weight as potential causes of a disease, for instance. [Robins \(1986\)](#) also emphasises the importance of a well defined treatment that could in principle be applied to anyone in the target population (cf. also [Robins et al. 2004](#)), and even formulates the causal target of inference in terms of a hypothetical randomised study. A decision theoretic approach to causality is further strongly advocated by [Dawid \(2000, 2002\)](#), pointing out that typical assumptions underlying causal inference concern not only the system under investigation but also the contemplated interventions. The approaches chosen by these authors are clearly influenced by the kind of applications they have in mind, for example complex treatment strategies for chronically ill patients, public health or policy interventions.

In the computer science literature, a ‘treatment’ is instead called an ‘intervention’, formalised as an external manipulation, possibly triggered by the history of the system, that sets a variable to a specific value independently of the values of other variables; we will henceforth regard the words ‘treatment’ and ‘intervention’ as interchangeable. Although it is not always explicitly required that such interventions should actually be feasible, they are central to the notion of causality underlying causal diagrams ([Spirtes et al. 2001](#); [Pearl 2000](#)). The qualifier ‘causal’ means specifically that such a diagram, and hence the associated model, remains invariant under certain interventions except for those nodes that represent the variables that are being manipulated. Defining causality in terms of such treatments or interventions facilitates a formal distinction between association and causation, and hence helps to make the assumptions underlying causal inference explicit. For instance, to assess whether we can estimate the effect of an intervention from observational (especially non-experimental) data, we can use the well-known back-door and front-door criteria proposed by [Pearl \(1995\)](#), which derive their names from the graphical rules used to check them on causal diagrams.

The above view of causality has been criticised in at least two respects ([Granger 1986](#); [Aalen and Frigessi 2007](#); [Commenges and Gegout-Petit 2007](#); [Heckman 2008](#)). For one, it can be argued that some systems, like e.g. cells, economies, climate, or the planetary system, are driven by causal mechanisms regardless of whether humans can

intervene in them, let alone subject the system to (randomised) experiments. Once we understand their functioning it is up to us to devise suitable ways to exploit the causal relations to achieve our ends, be it to prevent global warming or future financial crises, or to eradicate the HI-virus etc. Secondly, much of the above literature does not involve any dynamic modelling even though it is almost universally agreed that a cause has to precede the effect. Notable exceptions are Robins' work who has pioneered dynamic treatments (Robins 1986) and, in econometrics, the work by Heckman and coworkers (Heckman and Navarro e.g. 2007).

It is therefore not a surprise that in non-experimental subjects like econometrics, and especially in typical application of time series methodology, a different approach to causality is more popular. Granger (1969, 1980, 1988) proposed a notion of causality that is based on the following important points: the cause occurs before the effect and the cause must contain unique information that is otherwise not available and helps to predict the effect. Formalising this, he defines that a time series  $X_a(t)$  is causal for another time series  $X_b(t)$  if the prediction of  $X_b(t + 1)$  based on *all the information in the universe* up to time  $t$ ,  $\mathcal{I}^*(t)$ , is better than the prediction of  $X_b(t + 1)$  based on all that information but without  $\{X_a(s) : s \leq t\}$ , denoted by  $\mathcal{I}_{-a}^*(t)$  (note that it is required that  $\mathcal{I}^*(t)$  does not contain redundant information, e.g. the same measurements just on different scales). As  $\mathcal{I}^*(t)$  is not available in practice, the corresponding operational definition is based on a multivariate time series  $X(t) = X_V(t) = (X_1(t), \dots, X_d(t))'$ ,  $V = \{1, \dots, d\}$ , and the information sets are given by  $\mathcal{I}(t) = \{X_V(s) : s \leq t\}$  and  $\mathcal{I}_{-a}(t) = \{X_{V \setminus \{a\}}(s) : s \leq t\}$ . Granger then says that the series  $X_a(t)$  is *noncausal* for the series  $X_b(t)$  if the prediction of  $X_b(t + 1)$  is the same for  $\mathcal{I}(t)$  and  $\mathcal{I}_{-a}(t)$ . If the prediction is in fact better based on  $\mathcal{I}(t)$ , then  $X_a(t)$  is called a *prima facie* cause of  $X_b(t)$ . We find it noteworthy, and often overlooked, that according to Granger, a positive definition of causality relies on 'all information in the universe' (and this is meant literally, not just as the model universe); without this we only have a definition of *noncausality* and of *prima facie* causality (these will henceforth also be called Granger (non)causality). Hence we could ask, when is the reduced information set  $\mathcal{I}(t)$  used to replace the impractical  $\mathcal{I}^*(t)$  sufficient to be confident that a *prima facie* cause is a cause. This requires to think about latent (unobservable) processes or time series and how their presence may induce *prima facie* causal relations among the observable processes—some suggestions have been made by Hsiao (1982) and Eichler (2007, 2009).

It is worth mentioning that the essential idea behind Granger's approach is of course not restricted to time series. Analogous definitions have also been given for dynamic systems in continuous time (Schweder 1970; Aalen 1987; Florens and Fougère 1996; Comte and Renault 1996). These are particularly relevant to survival analysis, or more generally to event history analysis.

As Granger's definition of causality rests entirely on prediction (even if based on  $\mathcal{I}^*(t)$ ), it does not imply that replacing the 'natural' variation in  $X_a(t)$  by some external manipulation, *when possible*, will have any effect on the distribution of  $X_b(t + 1)$ . In this paper we therefore investigate a combination of the decision theoretic approach and Granger's ideas on causality, building on and expanding some of our earlier work (Eichler and Didelez 2007). As starting point, we assume that the structure among the components of a multivariate time series  $X_V(t)$  is given in terms of *prima*

facie causes, or noncausal relations in Granger's sense. Furthermore, this structure is represented graphically as suggested by Eichler (2001, 2006, 2007). It can be thought of as describing the natural or unperturbed behaviour of the time series, which we want to contrast with changes induced by external manipulations of the system. Hence, the quantity of interest is the effect of an intervention in one or more components of a multivariate time series at one or more points in time on the development of these series at a later point in time. We address various aspects of the connection between Granger causal relations and effects of interventions. For instance, we formulate conditions under which Granger noncausality implies no effect of an intervention, as well as criteria that allow us to compute the effect of an intervention possibly based on a subset  $X_S(t)$  of the original multivariate series,  $S \subset V$ . Using the graphical representation mentioned above, it becomes obvious that these criteria are comparable to the back-door and front-door criteria of Pearl (1995). The assumptions we make correspond to (extended) stability of Dawid and Didelez (2005) and are twofold: they concern the system under consideration, i.e. presence or absence of certain Granger causal relations, as well as the contemplated interventions.

The outline of the paper is as follows. We start in Sect. 2 by formalising Granger's approach to causality for multivariate time series, as well as the concept of interventions and effects thereof. For the latter we follow closely the framework proposed by Dawid (2002) and Dawid and Didelez (2005) as it makes the role of interventions particularly explicit. Section 3 then presents our versions of the back-door and front-door criteria which enable us to compute effects of interventions from a subset of the multivariate series, as would be relevant when some of the processes are in fact latent. How these criteria relate to Granger (non)causality becomes clear in Sect. 4, where we show how they can be verified using an intuitive graphical representation of Granger (non)causal structures. We conclude with a discussion of our results and provide technical details as well as all proofs in an appendix.

## 2 Causality in time series

In this section we formalise the two basic approaches to causality that will be combined later, Granger (non)causality on the one hand and the effect of intervention on the other hand. Throughout the paper we consider a multivariate stationary time series  $X = \{X(t), t \in \mathbb{Z}\}$  with  $X(t) = (X_1(t), \dots, X_d(t))'$ . Let  $V$  be the index set  $\{1, \dots, d\}$ . For any  $A \subseteq V$  we define  $X_A = \{X_A(t)\}$  as the multivariate subprocess with components  $X_a, a \in A$ . Furthermore,  $\bar{X}_A(t)$  denotes the history of  $X_A$  up to and including  $t$ , i.e. the set  $\{X_A(s), s \leq t\}$ . Furthermore, let  $\mathcal{X}_A(t)$  be the  $\sigma$ -algebra generated by  $\bar{X}_A(t)$ . The following regularity assumptions will be imposed throughout.

### Assumption

- (T1)  $X = \{X(t), t \in \mathbb{Z}\}$  is a stationary stochastic process on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .
- (T2) The conditional distribution of  $X(t+1)$  given the past  $\bar{X}(t)$ , denoted by  $\mathbb{P}^{X(t+1)|\bar{X}(t)}$ , has a regular version that is almost surely absolutely continuous with respect to some product measure  $\nu$  on  $\mathbb{R}^d$  with  $\nu$ -a.e. positive density.

(T3) For all disjoint  $A, B, C \subseteq V$ , the subprocesses  $X_A$  and  $X_B$  are measurably separated conditionally on  $X_C$  in the sense that

$$\mathcal{X}_{AUC}(t) \cap \mathcal{X}_{BUC}(t) = \mathcal{X}_C(t)$$

for all  $t \in \mathbb{Z}$ . For this notion of conditional measurable separability we refer to Florens et al. (1990).

The assumption of stationarity is made for the sake of simplicity. Following the remarks in Eichler (2001), the concepts and results can also be generalized to non-stationary time series. The remaining conditions (T2) and (T3) are technical and essentially ensure that the intersection property (C5) of conditional independence (Lauritzen 1996, pp. 29; see also our appendix) holds, as shown in Eichler (2001).

### 2.1 Granger noncausality

Granger noncausality was introduced by Granger (1969) and has become a popular concept not only in econometrics. As mentioned in the introduction, the actual definition addresses causality but requires to condition on ‘all the information in the universe’. In line with most of the literature we therefore define Granger noncausality for a specific information set given by a multivariate time series. When a series is not Granger noncausal we will call it Granger causal or prima facie cause. The following definition also formalises contemporaneous (in)dependence, which is not cast in causal terms as we agree with Granger (1980) that there cannot be contemporaneous causality.

Further, we use the notion of strong Granger noncausality, which is formulated in terms of conditional independence (instead of uncorrelation) and  $\sigma$ -algebras (Florens and Mouchart e.g. 1982), where the symbol  $\perp\!\!\!\perp$  denotes independence (Dawid 1979).

**Definition 2.1** (*Granger noncausality*) Let  $A$  and  $B$  be disjoint subsets of  $V$  and let  $X_A, X_B$  be the corresponding subprocesses of  $X = X_V$  as defined above.

- (i) Then  $X_A$  is (strongly) Granger noncausal for  $X_B$  up to horizon  $h, h \in \mathbb{N}$ , with respect to the process  $X_V$  if

$$X_B(t+k) \perp\!\!\!\perp \overline{X}_A(t) \mid \overline{X}_{V \setminus A}(t)$$

for all  $k = 1, \dots, h$  and  $t \in \mathbb{Z}$ .

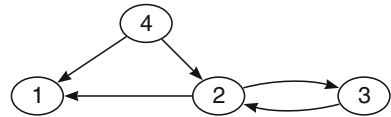
If the above holds for  $h = 1$  we simply say that  $X_A$  is (strongly) Granger noncausal for  $X_B$  with respect to  $X_V$ , and this will be denoted by  $X_A \not\rightarrow X_B [X_V]$ . Similarly, if the above holds for all  $h \in \mathbb{N}$  we say that  $X_A$  is (strongly) Granger noncausal for  $X_B$  at all horizons, and this will be denoted by  $X_A \overset{(\infty)}{\not\rightarrow} X_B [X_V]$ .

- (ii) The processes  $X_A$  and  $X_B$  are contemporaneously independent with respect to the process  $X_V$  if

$$X_A(t+1) \perp\!\!\!\perp X_B(t+1) \mid \overline{X}_V(t)$$

for all  $t \in \mathbb{Z}$ . This will be denoted by  $X_A \approx X_B [X_V]$ .

**Fig. 1** Mixed graph associated with the processes  $X$  and  $Z$  in Example 2.2



Strong Granger noncausality means that the past of  $X_A$  up to time  $t$  does not help to predict the distribution of  $X_B$  at the next point in time  $t + 1$  given information about the past of all the remaining components of  $X_V$  (including  $X_B$ 's past). In contrast, strong Granger noncausality at *all horizons* implies that this holds not only for the one-step prediction but for any time in the future. This is more restrictive and if not stated otherwise we will only deal with strong Granger noncausality at horizon  $h = 1$ . Note that for ease of notation we usually drop the ‘strong’.

*Example 2.2* Consider the following multivariate Gaussian process  $X$  with

$$\begin{aligned} X_1(t) &= \alpha_1 X_4(t - 2) + \beta_{12} X_2(t - 1) + \varepsilon_1(t), \\ X_2(t) &= \alpha_2 X_4(t - 1) + \beta_{23} X_3(t - 1) + \varepsilon_2(t), \\ X_3(t) &= \beta_{32} X_2(t - 1) + \varepsilon_3(t), \\ X_4(t) &= \varepsilon_4(t) \end{aligned}$$

where  $\varepsilon_i, i = 1, 2, 3, 4$ , are independent Gaussian white noise processes with mean 0 and variance  $\sigma^2$ . It is immediately obvious that with respect to the set  $\{X_1, X_2, X_3, X_4\}$ , the component  $X_1$  is Granger noncausal for all other variables, and for instance  $X_3$  is Granger noncausal for  $X_1$ . However, with respect to the reduced set  $\{X_1, X_3, X_4\}$  it cannot be assumed anymore that  $X_3$  is Granger noncausal for  $X_1$ . It is less intuitive, but also clear from the full model that  $X_3(t)$  is also not Granger noncausal anymore for  $X_1$  with respect to  $\{X_1, X_2, X_3\}$  due to the selection effect when conditioning on the past of  $X_2$  which induces an association between  $X_3$  and  $X_4$ . This shows that the Granger noncausal structure of a multivariate time series depends on the components chosen to make up the multivariate series  $X$ .

In Sect. 4, we introduce a graphical representation for Granger noncausal relations, where the absence of an arrow from a node  $a$  to a node  $b$  means that  $X_a$  is Granger noncausal for  $X_b$  with respect to the whole multivariate series. The presence of an arrow from  $a$  to  $b$  means that  $X_a$  is a prima facie cause of  $X_b$ . For the above example the corresponding graph is shown in Fig. 1.

## 2.2 Effects of interventions

While Granger noncausality describes the ‘natural’ behaviour of a multivariate time series, we now develop the idea of defining a causal effect as the effect of an intervention in such a system as has first been proposed by [Eichler and Didelez \(2007\)](#). We start by introducing the concept of intervention indicators, which allows us to distinguish formally between the ‘natural’ behaviour of a system and its behaviour

under an intervention (Pearl 1993, 2000; Lauritzen 2001; Dawid 2002; Spirtes et al. 2001).

**Definition 2.3 (Regimes)** Let  $X$  be a multivariate stationary time series. Consider a set of indicators  $\sigma = \{\sigma_a(t); a \in A, t \in \tau\}$  denoting interventions in  $X_a(t)$ ,  $a \in A \subseteq V$ , at points  $t \in \tau$  in time. Each  $\sigma_a(t)$  takes values in some set  $\mathcal{S}_a$  augmented by one additional state  $\emptyset$ . Different values of  $\sigma$  indicate different distributions of the time series  $X$  in the following way.

- (i) *Idle Regime:* When  $\sigma_a(t) = \emptyset$  we let  $X_a(t)$  arise naturally without intervention. With  $\mathbb{P}_{\sigma_a(t)=\emptyset}$  (which will often be abbreviated to  $\mathbb{P}_\emptyset$  or even just  $\mathbb{P}$ ) we denote the distribution of the time series  $X$  under the idle regime. We also call this the *observational regime*.
- (ii) *Atomic interventions:* Here  $\mathcal{S}_a = \mathcal{X}_a$ , the domain of  $X_a(t)$ , such that  $\sigma_a(t) = x^*$  means we intervene and force  $X_a(t)$  to assume the value  $x^*$ . Hence  $\mathbb{P}_{\sigma_a(t)=x^*}$  (or shorter  $\mathbb{P}_{x^*}$ ) denotes the distribution of the time series  $X$  under such an atomic intervention, with

$$\mathbb{P}_{\sigma_a(t)=x^*}(X_a(t) = x \mid \bar{X}_V(t - 1)) = \delta_{\{x^*\}}(x),$$

where  $\delta_D(x)$  is one if  $x \in D$  and zero otherwise.

- (iii) *Conditional intervention:* Here  $\mathcal{S}_a$  consists of functions  $g(\bar{X}_C(t-1)) \in \mathcal{X}_a$ ,  $C \subset V$ , such that  $\sigma_a(t) = g$  means  $X_a(t)$  is forced to take on a value that depends on past observations of  $\bar{X}_C(t - 1)$ . With  $\mathbb{P}_{\sigma_a(t)=g}$  denoting the distribution of the time series  $X$  under such a conditional intervention, we have

$$\begin{aligned} \mathbb{P}_{\sigma_a(t)=g}(X_a(t) = x \mid \bar{X}_V(t - 1)) &= \mathbb{P}_{\sigma_a(t)=g}(X_a(t) = x \mid \bar{X}_C(t - 1)) \\ &= \delta_{\{g(\bar{X}_C(t-1))\}}(x). \end{aligned}$$

- (iv) *Random intervention:* Here  $\mathcal{S}_a$  consists of distributions meaning that  $X_a(t)$  is forced to arise from such a distribution, i.e. the conditional distribution  $\mathbb{P}_{\sigma_a(t)=s}(X_a(t) \mid \bar{X}_V(t - 1))$  is *known* and possibly a function of  $\bar{X}_C(t - 1)$ ,  $C \subseteq V$ .

With this definition, we are considering a family of probability measures  $\mathbb{P}_\sigma$  on  $(\Omega, \mathcal{F})$  indexed by the possible values that  $\sigma$  can take. While  $\mathbb{P} = \mathbb{P}_{\sigma=\emptyset}$  describes the natural behaviour of the time series under the observational regime, any implementation of an intervention strategy  $\sigma = s$  will change the probability measure to  $\mathbb{P}_{\sigma=s}$ . We note that the assumption of stationarity is only made for the natural probability measure  $\mathbb{P}$ . In the following, we write  $\mathbb{E}_\emptyset$  (or shorter just  $\mathbb{E}$ ) and  $\mathbb{E}_{\sigma=s}$  to distinguish between expectations with respect to  $\mathbb{P}_\emptyset$  and  $\mathbb{P}_{\sigma=s}$ ; the shorthand  $\mathbb{P}_s$  and  $\mathbb{E}_s$  is used when it is clear from the context what variables are intervened in. In contrast, we write  $\mathbb{E}^{\mathcal{H}} X$  or  $\mathbb{E}^Y X$  to denote the conditional expectation of a random variable  $X$  given a  $\sigma$ -algebra  $\mathcal{H}$  or another random variable  $Y$ , respectively (Kallenberg e.g. 2001).

Our aim in this paper is to predict the effects of an intervention  $\sigma = s$  from data collected under the observational regime. This is only possible if the multivariate system as well as the intervention of interest are such that the intervention only changes the

conditional distribution of the variable at the point in time it targets according to Definition 2.3 (ii–iv), while all other conditional distributions remain the same as under the idle regime. To formalize this, we say that a random variable  $Y$  is independent of  $\sigma_a(t)$  conditionally on some  $\sigma$ -algebra  $\mathcal{H} \subseteq \mathcal{F}$  if

$$\mathbb{E}_{\sigma_a(t)=s}^{\mathcal{H}} g(Y) = \mathbb{E}_{\emptyset}^{\mathcal{H}} g(Y)$$

for all  $s \in \mathcal{S}_a$  and  $y \in \mathbb{R}$  and all measurable functions  $g$ , which we abbreviate by  $Y \perp\!\!\!\perp \sigma_a(t) \mid \mathcal{H}$ . If, for example,  $\mathcal{H} = \mathcal{X}_V(t)$ , we also write  $Y \perp\!\!\!\perp \sigma_a(t) \mid \bar{X}_V(t)$ . It is important to note that  $\sigma_a(t)$  is not a random variable and that the above notion of independence—although being one of statistical independence—is not symmetric. For a more detailed discussion of conditional independence involving non-random quantities we refer to Dawid (2002).

With this notation, the required link between the probability measures  $\mathbb{P}$  and  $\mathbb{P}_\sigma$  under the observational and the interventional regime, respectively, is established by the following assumptions. They characterise situations where the multivariate system as well as the intervention of interest are such that the intervention only changes the conditional distribution of the variable at the point in time it targets according to Definition 2.3 (ii–iv), while all other conditional distributions remain the same as under the idle regime. These assumptions are analogous to those of (extended) stability in Dawid and Didelez (2005).

**Assumption (Stability)** Let  $X$  be a multivariate time series that is stationary under the idle regime. The interventions  $\sigma = \{\sigma_a(t), a \in A, t \in \tau\}$  from Definition 2.3 are assumed to have the following properties, where we use  $\sigma_A(t) = \{\sigma_a(t); a \in A\}$ .

- (I1) for all  $t \notin \tau: X_V(t) \perp\!\!\!\perp \sigma \mid \bar{X}_V(t - 1)$ ;
- (I2) for all  $t \in \tau: X_V(t) \perp\!\!\!\perp \{\sigma_A(t'); t' \in \tau, t' \neq t\} \mid (\bar{X}_V(t - 1), \sigma_A(t))$ ;
- (I3) for all  $t \in \tau: X_{V \setminus A}(t) \perp\!\!\!\perp \sigma_A(t) \mid \bar{X}_V(t - 1)$ ;
- (I4) for all  $t \in \tau, a \in A: X_a(t) \perp\!\!\!\perp (X_{V \setminus \{a\}}(t), \sigma_{A \setminus \{a\}}(t)) \mid (\bar{X}_V(t - 1), \sigma_a(t) = s)$ .

With the above assumptions the distribution of the time series  $X$  under the interventional regime is fully specified by its natural distribution described by  $\mathbb{P} = \mathbb{P}_\emptyset$ , for instance Example 2.2, and the conditional distributions given in Definition 2.3 (ii–iv) under the chosen intervention. Therefore, whenever we make model assumptions in the following (or like stationarity, earlier), especially regarding Granger noncausal structures, these are supposed to be under the idle regime.

Note that (I1)–(I3) imply that  $(\bar{X}_V(t - 1), \bar{X}_{V \setminus A}(t)) \perp\!\!\!\perp \sigma$  for  $t = \min\{s \in \tau\}$ , as can be seen by iterative application of property (C4, Appendix) for conditional independence. By a similar reasoning we also obtain for any  $t \in \mathbb{Z}$

$$X_{V \setminus A}(t) \perp\!\!\!\perp \sigma \mid \bar{X}_V(t - 1). \tag{2.1}$$

In the special case of a single intervention in  $X_a(t)$ , i.e.  $A = \{a\}$ ,  $\tau = \{t\}$ , the assumptions (I1)–(I3) therefore simplify to

$$(\bar{X}_V(t - 1), X_{V \setminus \{a\}}(t)) \perp\!\!\!\perp \sigma_a(t) \tag{2.2}$$



and

$$\{X_V(t + j); j \in \mathbb{N}\} \perp\!\!\!\perp \sigma_a(t) \mid \overline{X}_V(t). \tag{2.3}$$

Whether the assumptions of stability (I1)–(I4) hold with regard to a given system  $X_V$  depends on the considered or practically feasible interventions as well as on the set  $V$  of components. It may for example be difficult, in practice, to subject a patient to a certain treatment without changing other aspects of her life. As an example for the second issue, assume that an isolated intervention in  $X_a(t)$  is possible, but that a variable  $X_c$  that predicts  $X_a$  and  $X_b$  under the idle regime has been ignored (we could loosely call  $X_c$  a confounder). Then, under  $\sigma_a(t) = x^*$ , the variable  $X_a(t)$  does not carry any information about  $\overline{X}_c(t)$ , whereas under the idle regime it does. As  $X_c$  predicts  $X_b$  we can therefore *not* assume that  $\mathbb{P}_{\sigma_a(t)=x^*}(X_b(t + h) \mid \overline{X}_a(t - 1), X_a(t) = x^*, \overline{X}_b(t))$  is the same as  $\mathbb{P}_{\emptyset}(X_b(t + h) \mid \overline{X}_a(t - 1), X_a(t) = x^*, \overline{X}_b(t))$ , as would be required for (2.3) with  $V = \{a, b\}$ . Consequently, in the following sections, we will often use the distinction between the system  $X_V$  which is such that the assumptions of stability are plausible, and a sub-series  $X_S, S \subset V$ , which could be the smaller set of actually observable components. For a discussion of the role of latent variables in the context of Granger noncausality we refer again to Hsiao (1982) and Eichler (2007, 2009). We also note that our approach to interventions in a system of multivariate time series is somewhat related to, but less restrictive than, the settable systems in White (2006).

Let us now consider effects of interventions. In general, this can be any function of the post-intervention distribution of  $\{X_V(t + j); j \in \mathbb{N}\}$  given an individual intervention  $\sigma_a(t) = s$ , for instance. It will often involve the comparison of setting  $X_a(t)$  to different values, e.g. setting  $X_a(t) = x^*$  as compared to setting it to  $X_a(t) = x^0$  which could be a baseline value in some sense. One may also want to use the idle case as baseline for the comparison. Typically, one is interested in the mean difference between interventions or between an intervention and the idle case, where we note that due to stationarity of  $X$  we can assume without loss of generality that  $\mathbb{E}o(X(t)) = 0$ . Hence, the average causal effect defined below can be regarded as the average difference between a strategy  $s$  and no intervention.

**Definition 2.4** (*Average causal effect*) The *average causal effect (ACE)* of an individual intervention according to strategy  $s$  in  $X_a(t)$  on  $X_b(t + h)$ ,  $a, b \in V, h > 0$  is given by

$$ACE_s = \mathbb{E}_{\sigma_a(t)=s} X_b(t + h).$$

The effect of *multiple* interventions  $\sigma_a(t) = s_a, a \in A \subseteq V$ , is given analogously by  $ACE_{s_A} = \mathbb{E}_{\sigma_A(t)=s_A} X_b(t + h)$ , where  $s_A$  and  $\sigma_A$  denote the corresponding sets.

The effect of *sequential* interventions  $\sigma_a(t_1) = s_1, \dots, \sigma_a(t_K) = s_K$ , for  $t_1 < \dots < t_K < t + h$  is further given by  $ACE_s = \mathbb{E}_{\sigma_a=s} X_b(t + h)$ , where  $s$  and  $\sigma_a$  denote the sets across the points in time when interventions take place.

Note that different strategies can be compared by considering, e.g.  $ACE_{s_1} - ACE_{s_2}$ . Even though we will mostly focus on the ACE, the results presented in this paper hold

more generally for  $\mathbb{E}_s f(X_b(t+h))$  for any measurable function  $f$  and thus for the post-intervention distribution  $\mathbb{P}_s(X_b(t+h))$ .

With the above definition of a causal effect in terms of interventions a first connection between Granger noncausality and the effect of an intervention can be established as follows.

**Corollary 2.5** *Consider a multivariate time series  $X = X_V$  and an individual intervention  $\sigma_a(t) = s$  satisfying (2.2) and (2.3). If  $X_a$  is Granger noncausal for  $X_b$  with respect to  $X_V$ , i.e.  $X_a \not\rightarrow X_b \ [X_V]$ , then there is no causal effect of intervening in  $X_a(t)$  on  $X_b(t+1)$ .*

We cannot say anything analogous for the effect of intervening in  $X_a(t)$  on  $X_b(t+h)$  for  $h > 1$  because there might be an ‘indirect’ effect through variables at points in time between  $t$  and  $t+h$ . Note that in the proof of Corollary 2.5 we rely on both conditions (2.2) and (2.3), underpinning that Granger noncausality on its own is not enough to make statements about the effect of interventions. However, we do not need the whole  $X_V$  to be observable in the above corollary—the system  $X_V$  with respect to which  $X_a \not\rightarrow X_b \ [X_V]$  can therefore include latent time series if this helps to justify the stability assumptions.

### 3 Identification of effects of intervention

In this section we will state some sufficient conditions that enable us to estimate the ACE of a specific intervention or strategy  $\sigma = s$  from data on the observable (i.e. non latent) components of the multivariate time series that has been collected under the observational regime  $\sigma = \emptyset$ ; if this is possible then we say that the ACE is *identifiable*. Formally, identifiability means that we can express the ACE in terms of observable or known quantities alone. We will not refer to Granger (non)causality here, but show in Sect. 4 how the latter can help us to answer the question of identifiability by providing a graphical check of the conditions.

#### 3.1 Simple back-door criterion

The back-door criterion for an individual intervention in a time series, as given below, has been established by Eichler and Didelez (2007) and reflects what is known in epidemiology as adjusting for confounding. The name is due to the graphical way of checking this criterion (cf. Sect. 4). As we show further below, it can be extended to the case of sequential interventions.

**Theorem 3.1** (Back-door criterion) *Let  $a, b \in S \subseteq V$ , where  $a = b$  is possible. If a conditional intervention is considered as in Definition 2.3 (iii, iv), we also assume for the conditioning set  $C$  that  $C \subseteq S$ . Suppose that assumptions (2.2) and (2.3) hold and that*

$$X_b(t+h) \perp\!\!\!\perp \sigma_a(t) \mid \overline{X}_S(t) \quad \forall h \in \mathbb{N}. \quad (3.1)$$

Then  $S$  identifies the effect of  $\sigma_a(t) = s$  on  $X_b(t+h)$  for all  $h \in \mathbb{N}$ , and the average causal effect  $ACE_s$  is given by

$$\mathbb{E}_s X_b(t+h) = \mathbb{E}_\emptyset \mathbb{E}_s^{\bar{X}_a(t-1), \bar{X}_{S \setminus \{a\}}(t)} \mathbb{E}_\emptyset^{\bar{X}_S(t)} X_b(t+h). \tag{3.2}$$

In 3.2 we can estimate  $\mathbb{E}_O^{\bar{X}_S(t)} X_b(t+h)$  from observational data, while the second expectation is with respect to the interventional distribution, which is fully known. The outer expectation is again observational. Hence, provided that  $X_S$  has been observed, we can use the above to estimate the causal effect ignoring  $X_{V \setminus S}$ . Dawid (2002) calls such a set  $\bar{X}_S(t)$  ‘sufficient covariates’ or ‘de-confounder’. Note that under the stability assumptions,  $V$  always identifies the causal effect due to condition (2.3). In this sense we could say that the whole system  $V$  contains all ‘relevant’ variables or components to identify an individual causal effect of  $X_a(t)$  on  $X_b(t+h)$ . Note however, that if an intervention  $\sigma_d = s$  in a different variable, say  $X_d$ ,  $d \neq a$ , is considered, a different system  $V'$  might be required to justify the stability assumptions with respect to this  $\sigma_d$ .

An example for Theorem 3.1 is given in Eichler and Didelez (2007). We will provide a more complex example for sequential interventions further below.

### 3.2 Multiple and sequential back-door criterion

Consider the effect of a joint intervention  $s_A = \{s_a; a \in A\}$  in more than one variable at a given point in time, i.e.  $\sigma_A(t) = \{\sigma_a(t), a \in A\}$ ,  $A \subset V$  on a set  $X_B(t+h)$ ,  $B \subset V$ . The causal effect  $\mathbb{E}_{s_A} X_B(t+h)$  is given componentwise by  $\mathbb{E}_{s_A} X_b(t+h)$ ,  $b \in B$ . Assuming that (I1)–(I4) are valid, we replace condition (3.1) by  $X_B(t+h) \perp\!\!\!\perp \sigma_A(t) \mid \bar{X}_S(t)$  and obtain the causal effect in complete analogy to (3.2) as

$$\mathbb{E}_{s_A} X_b(t+h) = \mathbb{E}_\emptyset \mathbb{E}_{s_A}^{\bar{X}_A(t-1), \bar{X}_{S \setminus A}(t)} \mathbb{E}_\emptyset^{\bar{X}_S(t)} X_b(t+h).$$

More interesting is the following theorem that addresses the identifiability of sequential interventions, i.e. those at different points in time.

**Theorem 3.2** (Sequential back-door criterion) *Let  $a, b \in S \subseteq V$ , where  $a = b$  is possible. If conditional interventions are considered as in Definition 2.3 (iii, iv), we also assume for the conditioning set  $C$  that  $C \subseteq S$ . Consider interventions  $\sigma = \{\sigma_a(t_k), k = 1, \dots, K, t_1 < t_2 < \dots < t_K\}$  and suppose that assumptions (I1)–(I4) hold. Let  $\sigma^{>k} = \{\sigma_a(t_{k+1}), \dots, \sigma_a(t_K)\}$  and  $\sigma^{<k} = \{\sigma_a(t_1), \dots, \sigma_a(t_{k-1})\}$  and  $s^{>k}$  analogously. Assume that*

$$X_b(t+h) \perp\!\!\!\perp \sigma_a(t_k) \mid (\bar{X}_S(t_k), \sigma^{<k} = \emptyset, \sigma^{>k} = s^{>k}) \tag{3.3}$$

Then  $S$  identifies the effect of sequential interventions  $s = \{s_1, \dots, s_K\}$  in  $X_a(t_1), \dots, X_a(t_K)$  on  $X_b(t+h)$ , and the average causal effect  $ACE_s$  is given by

$$ACE_s = \mathbb{E}_{\emptyset} \prod_{k=1}^K \left( \mathbb{E}_{s_k}^{\bar{X}_S(t_{k-1}), X_{S \setminus \{a\}}(t_k)} \mathbb{E}_{\emptyset}^{\bar{X}_S(t_k)} \right) X_b(t+h), \tag{3.4}$$

Similar to the simple back-door criterion, (3.4) consists of conditional expectations that can either be estimated if  $X_S$  is observable, or that are known by the chosen intervention. For longitudinal and survival settings, formula (3.4) is well known as G-formula (Robins 1986). The conditions (3.3) are in the same spirit as those given in Dawid and Didelez (2005) for the longitudinal case.

*Example 3.3* Let  $X$  be a stationary and purely nondeterministic Gaussian process. In order to satisfy the technical conditions (T1) to (T3), we furthermore assume that  $X$  has spectral matrix  $f(\lambda)$ ,  $\lambda \in [-\pi, \pi]$ , with eigenvalues that are bounded and bounded away from zero uniformly for all  $\lambda \in [-\pi, \pi]$  (Eichler 2007).

Suppose now that, for time points  $t_1 < t_2 < t_3$ , we are interested in the average causal effect of setting  $X_a(t_1)$  and  $X_a(t_2)$  to the values  $x_{1,a}^*$  and  $x_{2,a}^*$ , respectively, on  $X_b(t_3)$ , and that the effect is identified by the variables in  $S$ . By the assumptions on the spectral matrix, the subprocess  $X_S$  has a mean-square convergent autoregressive representation

$$X_S(t) = \sum_{j=1}^{\infty} \Phi(j) X_S(t-j) + \varepsilon_S(t), \tag{3.5}$$

where  $\varepsilon(t)$ ,  $t \in \mathbb{Z}$ , are independent and identically normally distributed with mean zero and non-singular covariance matrix  $\Sigma$ . Moreover, the best  $h$ -step predictor  $\mathbb{E}^{\bar{X}_S(t)} X(t+h)$  is equal to the best linear  $h$ -step predictor, that is,

$$\mathbb{E}^{\bar{X}_S(t)} X(t+h) = \sum_{j=1}^{\infty} \Phi^{(h)}(j) X(t-j+1). \tag{3.6}$$

Here, the coefficients  $\Phi^{(h)}(k)$  of the multi-step predictor can be computed recursively from the coefficients of the autoregressive representation in 3.5 using the relations

$$\Phi^{(h)}(k) = \sum_{j=1}^{k-1} \sum_{s \in S} \Phi^{(1)}(j) \Phi^{(h-j)}(k) + \Phi^{(1)}(h+k-1), \tag{3.7}$$

where  $\Phi^{(1)}(j) = \Phi(j)$  (Box et al. e.g. 1994, Section 5.3).

From Theorem 3.2, we find that the causal effect of an intervention at  $X_a(t_1)$  and  $X_a(t_2)$  on  $X_b(t_3)$  is given by

$$\mathbb{E}_{\emptyset} \mathbb{E}_{s_1}^{X_S(t_1-1), X_{S \setminus \{a\}}(t_1)} \mathbb{E}_{\emptyset}^{X_S(t_1)} \mathbb{E}_{s_2}^{X_S(t_2-1), X_{S \setminus \{a\}}(t_2)} \mathbb{E}_{\emptyset}^{X_S(t_2)} X_b(t_3). \tag{3.8}$$

Letting  $h_i = t_{i+1} - t_i, i = 1, 2$  and applying 3.6 twice, we obtain

$$\begin{aligned} \mathbb{E}_s(X_b(t_3)) &= \phi_{ba}^{(h_2)}(1)x_{2,a}^* + \phi_{ba}^{(h_2)}(h_1 + 1)x_{1,a}^* + \sum_{c \in S \setminus \{a\}} \phi_{bc}^{(h_2)}(1)\phi_{ca}^{(h_1)}(1)x_{1,a}^* \\ &\quad + \sum_{j=2}^{h_1} \sum_{c \in S} \phi_{bc}^{(h_2)}(j)\phi_{ca}^{(h_1-j+1)}(1)x_{1,a}^*. \end{aligned}$$

Here, the first term represents the causal effect of the intervention in  $X_a(t_2)$  on  $X_b(t_3)$ , the second term gives the effect of the intervention on  $X_a(t_1)$  that is not mediated by any of the variables  $X(t_2), \dots, X(t_1 + 1)$  while the remaining terms collect the indirect effects mediated by all variables in  $X(t_2), \dots, X(t_1 + 1)$  but  $X_a(t_2)$ .

### 3.3 Front-door criterion

As counterpart to the back-door criterion, Pearl (1995) introduced the so-called front-door criterion to identify causal effects. In brief, it characterises situations where there is unobserved confounding between the cause and effect of interest, but where the causal effect can still be recovered because it is fully mediated by a third variable (or set of variables) which is not affected by the confounding factors. It turns out that the computation of the causal effect can then be decomposed in the effect of the cause of interest on the mediator and the effect of the mediator on the target variable, without the need to actually be able to intervene in the mediating variable (Pearl, 2000, p. 82). We show here that the same principle can be applied to time series. However, we need somewhat more technical preliminaries than for the back-door criterion.

We consider the causal effect of a sequential intervention in all of  $X_a(t - h), \dots, X_a(t - 1)$  on  $X_b(t), a \neq b$ , specified by  $s = \{s^h, \dots, s^1\}$ . The interventions are allowed to be conditional on  $\overline{X}_S(t - h - 1)$  as well as previous values of  $X_a(t - h), \dots, X_a(t - 2)$  (which is only relevant if they are random interventions). Let  $X_C$  be the set of mediating series, then  $S = \{a, b\} \cup C$  is the identifying set and  $U = V \setminus S$  stands for the unobserved set of possible confounders. The assumptions following below characterise the situation in which the front-door criterion applies to time series. The first of the two assumptions is stated with respect to a measure  $\mathbb{Q}_k$ , in which the conditional distribution of  $X_C(t - j)$  given  $X_{V \setminus C}(t - j)$  and  $\overline{X}_V(t - j - 1)$  is replaced by the marginal distribution of  $X_C(t - j)$  for  $1 \leq j \leq k$ . The exact definition of  $\mathbb{Q}_k$  is given in the appendix. Heuristically one can say that the measure  $\mathbb{Q}_k$  mimicks an interventional regime where  $X_C(t - k), \dots, X_C(t - 1)$  is drawn randomly from its marginal distribution. Note however, that it is not necessary to assume that such interventions in  $X_C(t - j)$  are practically feasible, nor that they would indeed result in an intervention distribution corresponding to  $\mathbb{Q}_k$  if they were feasible, as can be seen from the proof of the following theorem in the appendix.

**Assumption** Consider a multivariate time series  $X_V(t)$  with intervention indicator  $\sigma = \{\sigma_a(t-h), \dots, \sigma_a(t-1)\}, a \neq b \in V, C \subset V \setminus \{a, b\}$  and  $U = V \setminus (C \cup \{a, b\})$ . For all  $t \in \mathbb{Z}$  and  $k \geq 1$  we assume

- (F1)  $X_b(t) \perp\!\!\!\perp (\bar{X}_a(t-k), \sigma_a(t-k)) \mid \bar{X}_{V \setminus \{a\}}(t-k) \quad [\mathbb{Q}_{k-1}]$ ;
- (F2)  $X_C(t) \perp\!\!\!\perp \bar{X}_{U \cup \{b\}}(t) \mid \bar{X}_a(t), \bar{X}_C(t-1) \quad [\mathbb{P}]$

A rough interpretation of assumption (F1) is that the effect of  $X_a(t-k)$  on  $X_b(t)$  is fully mediated by  $X_C$  between  $t-k+1$  and  $t-1$ . In addition, (F2) can be thought of as ensuring that  $X_C$  is not affected by confounding. Note that (F2) is with respect to the observational regime  $\sigma = \emptyset$ . The causal effect can now be computed as follows.

**Theorem 3.4** (Front-door criterion) *Suppose that (F1) and (F2) hold for  $X_V$ , as well as (II)–(I4). For any  $h \geq 2$  the causal effect of interventions in  $X_a(t-h), \dots, X_a(t-1)$  on  $X_b(t), a \neq b$ , is identified by  $S = \{a, b\} \cup C$ , and is given by  $ACE_S =$*

$$\mathbb{E}_S X_b(t) = \mathbb{E}_O \prod_{j=1}^h \left[ \mathbb{E}_{S^j}^{\bar{X}_{C \cup \{a\}}(t-j)} \mathbb{E}_O^{\bar{X}_{C \cup \{a\}}(t-j-1), X_a(t-j)} \right] \mathbb{E}_{\mathbb{Q}_h}^{\bar{X}_S(t-h-1), \bar{X}_C(t-1)} X_b(t), \tag{3.9}$$

where  $\mathbb{E}_{\mathbb{Q}_h}^{\bar{X}_S(t-h-1), \bar{X}_C(t-1)} X_b(t)$  for fixed values of  $\bar{X}_S(t-h-1)$  and  $\bar{X}_C(t-1)$  is given by

$$\int x_b(t) dQ_h(x_b(t) \mid \bar{x}_C(t-1), \bar{x}_S(t-h-1)) = \int \dots \int x_b(t) dP(x_b(t) \mid \bar{x}_S(t-1)) \prod_{j=1}^h dP(x_{\{a,b\}}(t-j) \mid \bar{x}_S(t-j-1)).$$

The last expectation in (3.9) is in fact the same as the effect of  $X_C$  on  $X_b$  if an intervention in  $X_C$  were possible, while the product of expectations corresponds to the effect of sequential interventions in  $X_b$  on  $X_C$ . Both can be compared to the G-formula (3.4).

Similar to the back-door criterion, all expected values required to compute the ACE with the front-door criterion are either known or can be estimated from observable quantities if  $X_S$  can be observed. The expectation with respect to  $\mathbb{Q}_h$  relies entirely on  $S$  as can be seen from the last line of the above theorem; note that this is the same expression as given (without proof) in Eichler and Didelez (2007).

*Example 3.5* Consider again the situation in Example 3.3 and suppose that the assumptions of Theorem 3.4 are satisfied. To compute the joint average causal effect of  $X_a(t-2)$  and  $X_a(t-1)$  on  $X_b(t)$ , we first evaluate the conditional expectation with respect to  $\mathbb{Q}_2$ . We have

$$\mathbb{E}_{\mathbb{Q}_2}^{\bar{X}_C(t-1), \bar{X}_S(t-3)} X_b(t) = \Phi_{bC}(1) X_C(t-1) + [\Phi_{bC}(2) + \Phi_{bb}(1)\Phi_{bc}(1) + \Phi_{ba}(1)\Phi_{ac}(1)] X_C(t-2) + \text{terms linear in } \bar{X}_S(t-3).$$

Thus the average causal effect is given by

$$\mathbb{E}_s X_b(t) = \Phi_{bC}(1) \Phi_{Ca}(1) X_a(t - 2).$$

We note that  $X_a(t - 1)$  has no effect on  $X_b(t)$  since any causal effect from  $X_a$  on  $X_b$  must be mediated by  $X_C$  and thus takes at least two time steps to reach  $X_b$ .

For larger lags  $h$ , the average causal effect can be computed similarly. For instance, for  $h = 3$ , we obtain

$$\begin{aligned} \mathbb{E}_s X_b(t) = & \Phi_{bC}(1) \Phi_{Ca}(1) X_a(t - 2) + [\Phi_{bC}(2) + \Phi_{bC}(1) \Phi_{CC}(1) \Phi_{Ca} \\ & + \Phi_{bb}(1) \Phi_{bC}(1) + \Phi_{ba}(1) \Phi_{aC}(1)] X_a(t - 3). \end{aligned}$$

### 4 Graphical representation

In this section we revisit the graphical representation of Granger(non)causal relations in a multivariate time series suggested by [Eichler \(2001\)](#) and [Eichler \(2007\)](#). An example was given in Fig. 1 for the time series in Example 2.2. Using these graphs or path diagrams, we devise graphical rules to check if a set  $S \subset V$  exists that satisfies the back-door or front-door criteria. This will provide us with further interesting insights into the relation between Granger causality and intervention causality.

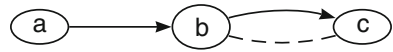
Note that we do not address, here, the question of how to *find* the Granger causal relations among a multivariate time series, but instead we assume that they are given, e.g. due to background knowledge. Causal search algorithms are detailed in [Spirtes et al. \(2001\)](#) (and briefly in [Pearl \(2000\)](#)), some issues concerning their application to multivariate time series are discussed in [Eichler \(2009\)](#).

#### 4.1 Graph notation

The graphs  $G = (V, E)$ , with nodes (or vertices)  $V$  and edges  $E$ , used here are so-called mixed graphs that may contain two types of edges directed edges  $a - - b$  or  $a \leftarrow b$ , and (dashed) undirected<sup>1</sup> edges  $a - - b$  for distinct nodes  $a, b \in V$ . Multiple edges between two nodes are allowed if they are of different type or orientation, i.e. there can be up to three edges between two nodes. Most of the terminology known for directed acyclic graphs can still be applied for these mixed graphs. For instance when  $a \rightarrow b$ , we call  $a$  a *parent* of  $b$ , and the set  $\text{pa}(b)$  is the set of all parents of  $b$ ; and when  $a - - b$  then  $a$  is a *spouse* of  $b$ , where  $\text{sp}(b)$  is the set of all spouses of  $b$ . As in [Frydenberg \(1990\)](#), a node  $b$  is said to be an *ancestor* of  $a$  if either  $b = a$  or there exists a directed path  $b \rightarrow \dots \rightarrow a$  in  $G$ . The set of all ancestors of elements in  $A$  is denoted by  $\text{an}(A)$ , which by definition includes  $A$  itself. Notice that this differs from [Lauritzen \(1996\)](#).

<sup>1</sup> In contrast to [Richardson \(2003\)](#), we use dashed undirected edges  $- - -$  instead of bi-directed edges  $\leftrightarrow$  as we use directed edges exclusively for indicating direction in time; dashed edges with a similar connotation have been used by [Cox and Wermuth \(1996\)](#).

**Fig. 2** Example of a mixed graph



The notion of a path in a mixed graph deserves some special attention. A path in our graphs cannot uniquely be defined by a sequence of nodes as there may be different edges between two nodes. Hence a *path*  $\pi$  from  $a$  to  $b$  is defined as a sequence  $\pi = (e_1, \dots, e_n)$  of *not necessarily distinct* edges  $e_i \in E$ , such that  $e_i$  is an edge between  $v_{i-1}$  and  $v_i$  for some sequence of *not necessarily distinct* vertices  $v_0 = a, v_1, \dots, v_n = b$ . A path  $\pi$  in  $G$  is called a *directed path* if it is of the form  $a \rightarrow \dots \rightarrow b$  or  $a \leftarrow \dots \leftarrow b$ ; in the former case we say it is a directed path from  $a$  to  $b$  and in the latter from  $b$  to  $a$ . Similarly, if  $\pi$  consists only of undirected edges, it is called an *undirected path*. Furthermore, a path between vertices  $a$  and  $b$  is said to be *b-pointing* if it has an arrowhead at the endpoint  $b$ , that is,  $e_n = v_{n-1} \rightarrow b$ . More generally, we call a path a *B-pointing path* if it is *b-pointing* for *some*  $b \in B$ . Similarly, we call a path between vertices  $a$  and  $b$  *bi-pointing* if it has an arrowhead at both endpoints, that is,  $e_1 = a \leftarrow v_1$  and  $e_n = v_{n-1} \rightarrow b$ . In particular we will make use of the following definition.

**Definition 4.1** (*Front- and back-door paths*) Let  $\pi = (e_1, \dots, e_n)$  be a path from  $a$  to  $b$ . We say that  $\pi$  is a *front-door path* from  $a$  to  $b$  if  $e_1 = a \rightarrow v_1$ . Otherwise we call  $\pi$  a *back-door path* from  $a$  to  $b$ .

An intermediate vertex  $c \in \{v_1, \dots, v_{n-1}\}$  on a path  $\pi$  is said to be an *m-collider* if the edges preceding and succeeding  $c$  on the path both have an arrowhead or a dashed tail at  $c$  (e.g.  $\rightarrow c \leftarrow, \text{---}c\text{---}, \text{---}c \leftarrow$ ); otherwise the vertex  $c$  is said to be an *m-noncollider* on the path (e.g.  $\rightarrow c \rightarrow, \text{---}c \rightarrow, \leftarrow c \rightarrow$ ). Notice that at least one of the edges that are adjacent to an *m-noncollider*  $c$  on a path must be a directed edge with a tail at  $c$ . Also notice that the first and last vertices on a path are neither colliders nor noncolliders. Furthermore, if the path passes through a vertex  $c$  more than once, this vertex may be a *m-collider* as well as a *m-noncollider* depending on its position on the path.

**Definition 4.2** (*m-blocked*) With the above definitions, a path  $\pi$  between vertices  $a$  and  $b$  is said to be *m-connecting* given a set  $S$  if

- (i) every *m-noncollider* on the path is not in  $S$ , and
- (ii) every *m-collider* on the path is in  $S$ ,

otherwise we say the path is *m-blocked* given  $S$ .

Note that a path is also *m-blocked* when the same node is a *m-collider* and a *m-noncollider* at different stages on the path because then (i) and (ii) cannot both be satisfied.

*Example 4.3* In the simple graph given in Fig. 2 we find that on some paths between  $a$  and  $c$  the node  $b$  is an *m-collider* like  $a \rightarrow b \text{---} c$  and on others it is an *m-noncollider* like  $a \rightarrow b \rightarrow c$ . Hence, the latter paths are *m-blocked* given  $\{b\}$  but the former are not. Further, for example, any path between  $a$  and  $a$  itself is only blocked given the empty set. This is because on any such path either  $c$  or  $b$  is a *m-collider*.



**Fig. 3** Illustration of global Markov property



### 4.2 Graphical time series models

In order to represent Granger noncausality graphically we identify the components  $X_1, \dots, X_d$  of the time series with the nodes  $V = \{1, \dots, d\}$  of a mixed graph  $G = (V, E)$ . The case that  $X_a$  is Granger causal (at horizon  $h = 1$ ) for  $X_b$  with respect to  $X_V$  is represented by a directed edge  $a \rightarrow b$ , and the case that both time series are contemporaneously dependent with respect to  $X_V$  is represented by an undirected edge  $a - - b$ . Vice versa, the absence of a directed (undirected) edge implies Granger noncausality (contemporaneous independence) with respect to  $X_V$ . When the graph  $G$  is constructed in this way for a given multivariate time series  $X_V$ , and assuming (T1)–(T3), then it satisfies the following Markov properties, as shown in Eichler (2001, 2007).

**Definition 4.4** (*Global Markov properties*) Let  $X$  be a multivariate stationary time series and  $G = (V, E)$  be a mixed graph. Then  $X$  satisfies the *global Granger causal Markov property* with respect to  $G$  if the following two conditions hold for all disjoint subsets  $A, B$ , and  $C$  of  $V$ .

- (i) If every  $B$ -pointing path between  $A$  and  $B$  is  $m$ -blocked given  $B \cup C$ , then  $X_A$  is Granger noncausal for  $X_B$  with respect to  $X_{A \cup B \cup C}$ , i.e.  $X_A \nrightarrow X_B [X_{A \cup B \cup C}]$ .
- (ii) If every bi-pointing path between  $A$  and  $B$  is  $m$ -blocked given  $A \cup B \cup C$  and there is no undirected edge between  $A$  and  $B$ , then  $X_A$  and  $X_B$  are contemporaneously independent with respect to  $X_{A \cup B \cup C}$ , i.e.  $X_A \approx X_B [X_{A \cup B \cup C}]$ .

To confirm the construction of these graphs, we note that if there is no directed edge  $a \rightarrow b$ , then every  $b$ -pointing path between  $a$  and  $b$  is  $m$ -blocked given  $V \setminus \{a\}$  because every such path has an edge  $v \rightarrow b$  with  $v \neq a$ . Hence the above implies that the absence of an individual directed edge  $a \rightarrow b$  indeed means that  $X_a \nrightarrow X_b [X_V]$ ; more generally we have for  $B \subset V$

$$X_{V \setminus (\text{pa}(B) \cup B)} \nrightarrow X_B [X_V]. \tag{3.10}$$

With a similar reasoning we obtain

$$X_{V \setminus (\text{sp}(B) \cup B)} \approx X_B [X_V]. \tag{3.11}$$

*Example 4.5* For the graph in Fig. 3 it follows from (3.10) that  $X_{\{a,d\}}$  is noncausal for  $X_b$  with respect to  $X_{\{a,b,c,d\}}$ , and from (3.11) it follows that  $X_{\{a,c,d\}}$  is contemporaneously independent of  $X_b$ . With the Global Markov properties we can, however, also determine what the relations are in subsets of  $\{a, b, c, d\}$ . For instance, it is obvious that every path between  $a$  and  $b$  is  $b$ -pointing, and  $c$  but not  $d$   $m$ -separates these two nodes. This is because every path has to go through  $c$  which is always a  $m$ -noncollider

due to the directed edge to  $b$ . In contrast,  $d$  can be a  $m$ -collider or  $m$ -noncollider on different paths and it is hence not enough to condition on  $d$ . Consequently,  $X_a$  is Granger noncausal for  $X_b$  with respect to  $X_{\{a,b,c\}}$  but not with respect to  $X_{\{a,b,d\}}$ . In contrast,  $X_b \not\rightarrow X_a[X_{\{a,b,c\}}]$  as well as  $X_b \not\rightarrow X_a[X_{\{a,b,d\}}]$  because on every  $a$ -pointing path  $c$  as well as  $d$  are always  $m$ -noncolliders. For the same reason  $X_a \approx X_b[X_{\{a,b,c\}}]$  as well as  $X_a \approx X_b[X_{\{a,b,d\}}]$ .

In particular cases we can read off a time series graph when Granger noncausality holds at all horizons. These are characterised as follows (Eichler cf. 2007, Thm 4.5).

**Theorem 4.6** (Noncausality at all horizons) *Suppose that  $X$  satisfies the global Markov properties with respect to a mixed graph  $G$ . Let  $A, B, C$  be disjoint subsets of  $V$ . If every  $\text{an}(B)$ -pointing path between  $A$  and  $\text{an}(B)$  is  $m$ -blocked given  $B \cup C$ , then  $X_A$  is Granger noncausal at all horizons with respect to  $X_{A \cup B \cup C}$ , i.e.  $X_A \not\rightarrow^{(\infty)} X_B [X_{A \cup B \cup C}]$ .*

As every  $\text{an}(B)$ -pointing path between  $V \setminus \text{an}(B)$  and  $\text{an}(B)$  is  $m$ -blocked by  $\text{an}(B)$ , an immediate implication of the above is that

$$X_{V \setminus \text{an}(B)} \not\rightarrow^{(\infty)} X_B [X_V], \quad (3.12)$$

i.e., as one would expect intuitively, those variables that are not a graphical ancestor of  $X_B$  are Granger noncausal for  $X_B$  at all horizons with respect to  $X_V$ . In Fig. 3, for instance,  $X_b$  is Granger noncausal for  $X_{\{a,c,d\}}$  at all horizons. The following corollary addresses the question if a similar statement can be made about the effect of an intervention in a node  $X_a$  that is not an ancestor of  $X_b$ .

**Corollary 4.7** *Consider a multivariate time series  $X$  that obeys the global Markov properties for a graph  $G$  under the observational regime, and a single intervention  $\sigma_a(t)$  satisfying (2.2) and (2.3). For  $a, b \in V$ , whenever  $a \notin \text{an}(b)$ , then intervening in  $X_a(t)$  has no causal effect on  $X_b(t+h)$  for all  $h \in \mathbb{N}$ .*

Like Corollary 2.5, the above relies on the stability assumptions (2.2) and (2.3), and again  $X_V$  may be chosen to contain suitable latent series if this helps to justify these assumptions. Note that the converse of Corollary 4.7 is not necessarily the case:  $a \in \text{an}(b)$  does not necessarily imply that an intervention in  $X_a(t)$  has an effect on  $X_b(t+h)$  as the overall effect is a combination of direct and indirect effects that may cancel each other.

A special case arises when the graph  $G$  associated with a multivariate time series  $X_V$  contains no undirected edges, i.e. when there is no contemporaneous dependence among the variables. The above graphical representation and properties are then analogous to the local independence graphs for continuous time processes proposed by Didelez (2007, 2008).

### 4.3 Graphical criteria for identifiability

Let us now address the question of how we can see from the Granger causal structure of a multivariate time series whether any of the criteria that permit identification of

causal effects are satisfied. The following result has been shown by Eichler and Didelez (2007) for the case of an individual intervention; we show here that it also applies to multiple and sequential interventions and we specify how to construct the minimal identifying set. Due to Corollary 4.7 we only need to consider the case  $a \in \text{an}(b)$ .

**Theorem 4.8** (Back-door criterion) *Consider a multivariate time series  $X$  that obeys the global Markov properties for a graph  $G$ . Assume that  $a \in \text{an}(b)$ .*

- (i) *The assumptions (3.1) and (3.3) of Theorems 3.1 and 3.2 are satisfied if all  $\text{an}(b)$ -pointing back-door paths between  $a$  and  $\text{an}(b)$  are  $m$ -blocked given  $S$ .*
- (ii) *The minimal set  $S$  satisfying (i) is given by  $S = \{a, b\} \cup \text{pa}(a) \cup D$ , where  $D$  is the set of all nodes  $v$  such that there is a back-door path from node  $a$  to  $v$  for which all intermediate nodes are  $m$  colliders and all intermediate nodes as well as  $v$  itself are ancestors of  $b$ .*

**Example 4.9** Consider the graph in Fig. 3 and assume we are interested in the (individual or sequential) effect of  $X_c$  on  $X_b$ . According to Theorem 4.8 part (ii), the minimal set  $S$  required to identify this is  $S = \{a, b, c, d\}$ , i.e. the whole system. Note that  $a$  has to be included because it is connected to  $c$  via a back-door path  $c - - d \leftarrow a$  such that the intermediate node  $d$  is an  $m$ -collider and both  $a, d$  are ancestors of  $b$ . It may seem counterintuitive that we need to take  $X_a$  and  $X_d$  into account as neither have a directed edge into  $X_b$ . However, they both predict  $X_c$  at points in time between  $t$  and  $t + h$  which in turn predicts  $X_b(t + h)$  and therefore act as confounders. If, in contrast, we are interested in the (individual or sequential) effect of  $X_a$  on  $X_b$  then the minimal identifying set is  $S = \{a, b, d\}$ . The node  $c$  is not included because any back-door path from  $a$  to  $c$  starts with  $a \leftarrow d$ , so the node  $d$  is an  $m$ -noncollider. Again, this might seem counterintuitive as  $c$  has a directed path into  $a$  as well as  $b$  and may therefore be suspected to be a confounding process. However, the confounding is, in this case, fully accounted for by including  $X_d$ .

When more specific knowledge about the time series model is available, e.g. at what lags dependencies occur, it may be possible to reduce the identifying set  $S$  given in Theorem 4.8 (ii). However, at the general level considered here, where we do not impose any other structure than the Granger (non)causal relations shown in the graph, this is not possible.

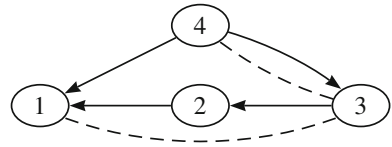
We mentioned earlier the special case of graphs without any undirected edges. For such graphs the following is obvious from the definition of the set  $S$  in Theorem 4.8 (ii).

**Corollary 4.10** *If the graph  $G$  contains no undirected edges, then the set  $S$  in Theorem 4.8 (ii) is given by  $S = \{a, b\} \cup \text{pa}(a)$ .*

**Example 4.11** This is a continuation of Example 2.2. Suppose that we are interested in the effect of an intervention setting  $X_3(t)$  to  $x_3^*$  on  $X_1(t + 2)$ . From the full model, we obtain immediately that

$$\mathbb{E}_{\sigma_3(t)=x_3^*} X_1(t + 2) = \beta_{12}\beta_{23}x_3^*.$$

**Fig. 4** Path diagram associated with the process described in Example 4.13



Using Corollary 4.10 (or Theorem 4.8 part (ii)), however, we see that the minimal set required for identification of the above causal effect is given by  $S = \{1, 2, 3\}$  as  $pa(3) = 2$  and there are no undirected edges. Simple calculations show that  $X_{\{1,2,3\}}$  has the autoregressive representation

$$\begin{aligned} X_1(t) &= \left( \frac{\alpha_1\alpha_2}{1 + \alpha_2^2} + \beta_{12} \right) X_2(t - 1) - \frac{\alpha_1\alpha_2\beta_{23}}{1 + \alpha_2^2} X_3(t - 2) + \tilde{\varepsilon}_1(t), \\ X_2(t) &= \beta_{23} X_3(t - 1) + \tilde{\varepsilon}_2(t), \\ X_3(t) &= \beta_{32} X_2(t - 1) + \tilde{\varepsilon}_3(t), \end{aligned} \tag{3.13}$$

where  $\tilde{\varepsilon}_i, i = 1, 2, 3$ , are again independent zero mean Gaussian white noise processes. Thus, the effect of setting  $X_3(t) = x_3^*$  on  $X_1(t + 2)$ , as identified by  $S = \{1, 2, 3\}$ , can be obtained from the autoregressive representation 3.13 as

$$\mathbb{E}_{\sigma_3(t)=x_3^*} X_1(t + 2) = \phi_{13}^{(2)}(1) = \phi_{12}(1)\phi_{23}(1) + \phi_{13}(2) = \beta_{12}\beta_{23}x_3^*.$$

Further using Corollary 4.10, we see that identification of the effect of an intervention in  $X_2(t)$  on  $X_1(t + h)$  would require all variables as  $pa(2) = \{3, 4\}$ , while identification of the effect of an intervention in  $X_2(t)$  on  $X_3(t + h)$  requires  $S = \{2, 3, 4\}$ .

The following theorem provides a graphical check for the assumptions (F1) and (F2) of Theorem 3.4.

**Theorem 4.12** (Front-door criterion) *Consider a multivariate time series  $X$  that obeys the global Markov properties for a graph  $G$ . Assumptions (F1) and (F2) for the front-door criterion hold if*

- (G1) every directed path from  $a$  to  $b$  is  $m$ -blocked given  $C$ ;
- (G2) there are no directed edges  $v \rightarrow c$  for all  $v \in V \setminus (C \cup \{a\})$  and  $c \in C$ ;
- (G3) there are no undirected edges  $v - - c$  for all  $v \in V \setminus C$  and  $c \in C$ .

*Example 4.13* Consider the graph in Fig. 4 and assume we are interested in the effect of intervening in  $X_3$  on  $X_1$ . The back-door criterion requires the minimal set  $S = \{1, 3, 4\}$ . However, assume that  $X_4$  is a latent process, so we have to check if, instead, the front-door criterion can be applied with  $C = \{2\}$ . The only directed path from  $X_3$  to  $X_1$  is  $3 \rightarrow 2 \rightarrow 1$  which is clearly  $m$ -blocked by 2. Further  $V \setminus (C \cup \{3\}) = \{1, 4\}$  has no directed edges into  $C = \{2\}$ , and  $V \setminus C = \{1, 3, 4\}$  has no undirected edges with  $\{2\}$ . Hence, the front-door theorem can be used to compute the effect of an intervention in  $X_3(t - h), \dots, X_3(t - 1)$  on  $X_1(t)$  if  $\bar{X}_{\{1,2,3\}}(t)$  is observable. This illustrates nicely that the back-door principle works by adjusting for confounding, while the front-door principle is based on obtaining a total effect by exploiting mediating variables.

## 5 Discussion and conclusions

In this paper we have proposed a way of combining Granger's ideas on causality with an intervention based approach. Granger (1980) himself points out that controllability, as aimed at by interventions, is a deeper concept than Granger causality, and he gives an example for a situation where the structure of a system is altered by changing a previously uncontrolled variable to one that is controlled. This latter case is exactly what is excluded by our assumptions of stability. While Granger noncausal and *prima facie* causal relations describe the 'natural' behaviour of a multivariate system of time series, we have to believe that certain aspects of this system remain the same (i.e. stable) under intervention (i.e. control) in order to be able to draw any inference for the latter case. The stability assumptions essentially demand that the system as well as the intervention can be chosen so that all conditional distributions, except for the ones of the variables targeted by the intervention, remain the same. Then, we can deduce that if a process is Granger noncausal for another process (with respect to this system) an intervention in the former will not affect the latter one time lag later.

Related to the previous comments is the observation that the notion of Granger noncausality describes the absence of a *direct* causal relation, as we can only infer the absence of a causal effect at one time lag. Hence, Granger causality (or *prima facie* cause) implies the possibility of a direct as well as indirect effect over several lags. We showed that it is in fact the absence of a directed path from  $X_a$  to  $X_b$ , i.e. of a chain of Granger causal relations, that permits to exclude a causal effect of an intervention at any lag  $h$ . Vice versa, one may say that when we investigate the causal effect of an intervention in  $X_a(t)$  on  $X_b(t+h)$  we do not care whether this is a direct effect or whether it is mediated by other (or the same) components at the times between  $t$  and  $t+h$ .

A problem with Granger's original definition is that it relies on conditioning on all the information in the universe up to time  $t$ . Similarly, our assumptions of stability will also be easier to justify when the system  $X_V$  is 'large' enough, in practice this will often mean that it contains latent time series. The global Markov properties can then be used to infer what the Granger (non)causal relations among  $a$ , say observable, subset  $X_S$  of the whole system are; and the identifying back-door and front-door criteria characterise when the effect of interventions can be computed from  $X_S$ . The graphical rules to find a minimal set  $S$  illustrate that the notion of 'confounding' is less obvious for the time series case, due to the potential for indirect relations via several time lags. However, interestingly, the graphical check of the back-door criterion is the same for single, multiple or sequential interventions.

Our approach is complicated by the possible presence of contemporaneous dependencies, which allow for correlated errors at a given point in time that are independent across time points. It is clear that these are not relevant to the effects of interventions. However, they can affect identifiability by giving rise to what is known as selection effect (Hernán et al. 2004) when conditioning, for instance, on the past of certain components. The proposed identifying criteria take this into account. The special case where there is no contemporaneous dependence may often be unrealistic for multivariate time series, but it is plausible for certain multivariate stochastic processes in continuous time, like marked point processes. The corresponding graphical representation

(Didelez 2007, 2008) is based on the continuous time analogue of Granger noncausality called local independence (Schweder 1970). As local independence graphs correspond essentially to the graphs for time series without undirected edges, we suspect that our results about the effects of interventions, here, can be transferred to the continuous time case. A formal investigation of this conjecture will, however, necessitate some subtle measure-theoretic considerations.

**Acknowledgements** We acknowledge financial support from the London Mathematical Society.

## Appendix

### A. Conditional independence

The following properties of conditional independence are heavily used for the proofs and are therefore stated here. They go back to Dawid (1979) (see also Lauritzen (1996, pp. 29)).

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and let  $\mathcal{F}_i, i = 1, \dots, 4$  be sub- $\sigma$ -algebras of  $\mathcal{F}$ . Recall that two sub- $\sigma$ -algebras  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are said to be independent conditionally on  $\mathcal{F}_3$  if  $\mathbb{E}(X|\mathcal{F}_2 \vee \mathcal{F}_3) = \mathbb{E}(X|\mathcal{F}_3)$  a.s. for all real-valued, bounded,  $\mathcal{F}_1$ -measurable random variables  $X$ . Then the basic axioms of conditional independence are:

- (C1)  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_3 \Rightarrow \mathcal{F}_2 \perp\!\!\!\perp \mathcal{F}_1 | \mathcal{F}_3$  (symmetry)
- (C2)  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 \vee \mathcal{F}_3 | \mathcal{F}_4 \Rightarrow \mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_4$  (decomposition)
- (C3)  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 \vee \mathcal{F}_3 | \mathcal{F}_4 \Rightarrow \mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 \vee \mathcal{F}_3 | \mathcal{F}_3 \vee \mathcal{F}_4$  (weak union)
- (C4)  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_4$  and  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_3 | \mathcal{F}_2 \vee \mathcal{F}_4 \Rightarrow \mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 \vee \mathcal{F}_3 | \mathcal{F}_4$  (contraction)
- (C5)  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 | \mathcal{F}_3 \vee \mathcal{F}_4$  and  $\mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_3 | \mathcal{F}_2 \vee \mathcal{F}_4 \Rightarrow \mathcal{F}_1 \perp\!\!\!\perp \mathcal{F}_2 \vee \mathcal{F}_3 | \mathcal{F}_4$  (intersection property)

Properties (C1)–(C4) are always satisfied, while property (C5) requires additional assumptions that essentially ensure that there are no logical redundancies. In our case, assumptions (T1)–(T3) ensure that (C5) holds for the multivariate time series we consider. Dawid (1979) discusses more general notions of the above properties.

### B. Proofs

*Proof of Corollary 2.5* As  $X_a \rightsquigarrow X_b [X_V]$  (under the idle regime), we know that  $X_b(t+1) \perp\!\!\!\perp X_a(t) | \bar{X}_{V \setminus \{a\}}(t)$ . From (2.3) it follows that  $X_b(t+1) \perp\!\!\!\perp \sigma_a(t) | \bar{X}_V(t)$ . Both together imply with (C4) of the properties of conditional independence that  $X_b(t+1) \perp\!\!\!\perp \sigma_a(t) | (X_{V \setminus \{a\}}(t), \bar{X}_V(t-1))$ . With (2.2) and again (C4) it follows in turn that  $X_b(t+1) \perp\!\!\!\perp \sigma_a(t)$  which implies  $\mathbb{E}_s X_b(t+1) = \mathbb{E}_o X_b(t+1)$ , i.e. the intervention has no effect on  $X_b(t+1)$ .  $\square$

*Proof of Theorems 3.1 and 3.2* By the law of iterated conditional expectation

$$\mathbb{E}_s X_b(t+h) = \mathbb{E}_s \mathbb{E}_s^{\bar{X}_S(t_1-1), X_{S \setminus \{a\}}(t_1)} \mathbb{E}_s^{\bar{X}_S(t_1)} X_b(t+h).$$

The last expectation in the above line can be modified exploiting (3.3). The middle expectation can be modified due to Assumption (I4) and the first one is equal to the expectation under the idle regime because using (I1)–(I3) we obtain  $\overline{X}_S(t_1 - 1), X_{S \setminus \{a\}}(t_1) \perp\!\!\!\perp \sigma$ . Hence

$$\mathbb{E}_s X_b(t + h) = \mathbb{E}_\emptyset \mathbb{E}_{s_1}^{\overline{X}_S(t_1-1), X_{S \setminus \{a\}}(t_1)} \mathbb{E}_{s_{>1}}^{\overline{X}_S(t_1)} X_b(t + h),$$

where  $s^{>k} = \{s_{k+1}, \dots, s_k\}$ . Similarly, the last expectation can now be written as

$$\mathbb{E}_{s_{>1}}^{\overline{X}_S(t_1)} X_b(t + h) = \mathbb{E}_\emptyset^{\overline{X}_S(t_1)} \mathbb{E}_{s_2}^{\overline{X}_S(t_2-1), X_{S \setminus \{a\}}(t_2)} \mathbb{E}_{s_{>2}}^{\overline{X}_S(t_2)} X_b(t + h).$$

Substituting back into the expression for  $\mathbb{E}_s X_b(t + h)$ , replacing the inner expectation iteratively for  $k = 1, \dots, K$ , and noting that  $\mathbb{E}_{s^{>K}}^{\overline{X}(t_K)} = \mathbb{E}_\emptyset^{\overline{X}(t_K)}$  we obtain

$$\mathbb{E}_s X_b(t + h) = \mathbb{E}_\emptyset \prod_{k=1}^K \left[ \mathbb{E}_{s_k}^{\overline{X}_S(t_k-1), X_{S \setminus \{a\}}(t_k)} \mathbb{E}_\emptyset^{\overline{X}(t_k)} \right] X_b(t + h)$$

which was the claim.

For  $K = 1$ , this proves in particular Theorem 3.1. □

*Preliminaries for the proof of Theorem 3.4* For ease of notation, we use the following abbreviations

$$Y_t = X_b(t) \quad X_t = X_a(t) \quad Z_t = X_C(t) \quad U_t = X_U(t),$$

furthermore, let  $W_t = X_S(t) = (X_a(t), X_b(t), X_C(t))$  and let  $V_t = X_V(t)$ .

Under assumptions (T1) to (T3) on the time series, the conditional distribution of  $Z_t$  given  $\overline{V}_{t-1}, U_t, X_t, Y_t$  and the marginal distribution of  $Z_t$  have densities  $f(z_t | \overline{v}_{t-1}, u_t, x_t, y_t)$  and  $f(z_t)$ , respectively, with respect to a product measure  $\nu$  and are related by

$$\int h(z_t) dP(z_t) = \int h(z_t) \frac{f(z_t)}{f(z_t | \overline{v}_{t-1}, u_t, x_t, y_t)} dP(z_t | \overline{v}_{t-1}, u_t, x_t, y_t)$$

for every measurable function  $h(z_t)$ .

Let  $\mathbb{Q}_j$  be measures on  $(\Omega, \mathcal{F})$  obtained iteratively from  $\mathbb{P}$  by setting  $\mathbb{Q}_0 = \mathbb{P}$  and

$$\mathbb{Q}_j(A) = \mathbb{E}_{\mathbb{Q}_{j-1}} \left( \frac{f(Z_{t-j})}{f(Z_{t-j} | \overline{V}_{t-j-1}, U_{t-j}, X_{t-j}, Y_{t-j})} 1_A \right)$$

for  $j = 1, \dots, h$ , and let  $\mathbb{Q}_j = \mathbb{Q}_j^{\overline{V}_t}$  be the measures on  $\mathbb{R}^{\mathbb{N}^V}$  induced by  $\overline{V}_t$ . Similarly, we define measures  $\mathbb{Q}_{j,s}$  for experimental regimes with interventions in  $X_t$  obtained from  $\mathbb{P}_s$  instead of  $\mathbb{P}$ . Since the construction of  $\mathbb{Q}_j$  leaves the marginal distribution of  $\overline{V}_{t-j-1}$  unchanged, we have  $\mathbb{Q}_j(A) = \mathbb{P}(A)$  and  $\mathbb{Q}_{j,s}(A) = \mathbb{P}_s(A)$  for

all  $A \in \sigma\{\bar{V}_{t-j-1}\}$ . Further, under  $\mathbb{Q}_j$  (or  $\mathbb{Q}_{j,s}$ ),  $Z_{t-1}, \dots, Z_{t-j}$  are independent of any past or contemporaneous variables or interventions. Hence condition (F1) is equivalent to  $Y_t \perp\!\!\!\perp \bar{X}_{t-j}, \sigma_{t-j} \mid \bar{Z}_{t-1}, \bar{Y}_{t-j}, \bar{U}_{t-j} \ [\mathbb{Q}_{j-1}]$

Note also that under (F2),  $f(z_t \mid \bar{v}_{t-1}, u_t, x_t, y_t) = f(z_t \mid \bar{w}_{t-1}, x_t, y_t)$ . It follows that the marginal distribution of  $\bar{W}_t$  under  $\mathbb{Q}_h$  can be determined from its marginal distribution under  $\mathbb{P}$ .

*Proof of Theorem 3.4* Writing the ACE in integral notation,  $\mathbb{E}_s(Y_t) = \int y_t dP_s(y_t)$ , we will first show that

$$dP_s(y_t) = \int \dots \int dQ_{k,s>k}(y_t \mid \bar{z}_{t-1}, \bar{v}_{t-k-1}) \prod_{j=1}^k [dP(z_{t-j} \mid \bar{x}_{t-j-1}, \bar{z}_{t-j-1}) \times dP_{s^j}(x_{t-j} \mid \bar{x}_{t-j-1}, \bar{w}_{t-h-1})] dP_{s>k}(\bar{v}_{t-k-1}) \tag{B.1}$$

for  $0 \leq k \leq h$ , where  $s^{>k} = \{s^h, \dots, s^{k+1}\}$ . We proceed by induction over  $k$ . For  $k = 0$ , the relation follows immediately from  $dP_s(y_t) = \int dP_s(y_t \mid \bar{v}_{t-1}) dP_s(\bar{v}_{t-1})$  and  $Q_{0,s} = P_s$ . For the induction step, assume that B.1 holds for  $k - 1$ . First, we note that by assumption (F1)

$$dQ_{k-1,s \geq k}(y_t \mid \bar{z}_{t-1}, \bar{v}_{t-k}) = dQ_{k-1,s>k}(y_t \mid \bar{z}_{t-1}, \bar{y}_{t-k}, \bar{u}_{t-k}). \tag{B.2}$$

Next, the last factor in B.1 can be factorized as

$$dP_{s \geq k}(\bar{v}_{t-k}) = dP_{s \geq k}(v_{t-k} \mid \bar{v}_{t-k-1}) dP_{s>k}(\bar{v}_{t-k-1}).$$

Since with assumption (F2), intervention in  $X_{t-k}$  and exploiting (2.1), the variables  $Z_{t-k}, X_{t-k}$ , and  $(U_{t-k}, Y_{t-k})$  are independent conditionally on  $\bar{V}_{t-k-1}$ ,  $dP_{s \geq k}(v_{t-k} \mid \bar{v}_{t-k-1})$  can further be factorised into

$$dP_{s \geq k}(y_{t-k}, u_{t-k} \mid \bar{v}_{t-k-1}) dP_{s \geq k}(z_{t-k} \mid \bar{v}_{t-k-1}) dP_{s \geq k}(x_{t-k} \mid \bar{v}_{t-k-1}). \tag{B.3}$$

For the first factor of (B.3), we find

$$dP_{s \geq k}(y_{t-k}, u_{t-k} \mid \bar{v}_{t-k-1}) = dP_{s>k}(y_{t-k}, u_{t-k} \mid \bar{v}_{t-k-1})$$

since the distribution of  $U_{t-k}$  and  $Y_{t-k}$  is not affected by the intervention in  $X_{t-k}$  (due to (I3)), while the third factor of (B.3) equals the intervention distribution due to (I4)

$$dP_{s \geq k}(x_{t-k} \mid \bar{v}_{t-k-1}) = dP_{s^k}(x_{t-k} \mid \bar{x}_{t-k-1}, \bar{w}_{t-h-1}).$$

For the second factor of (B.3), we obtain again by assumption (F2), and using (2.1) to drop the subscript of  $P$ ,

$$dP_{s \geq k}(z_{t-k} \mid \bar{v}_{t-k-1}) = dP(z_{t-k} \mid \bar{x}_{t-k-1}, \bar{z}_{t-k-1}).$$



Substituting the factors and B.2 back into B.1, we obtain  $dP_s(y_t) =$

$$\int \cdots \int \left[ \int dQ_{k-1,s>k}(y_t|\bar{z}_{t-1}, \bar{y}_{t-k}, \bar{u}_{t-k}) dP_{s>k}(y_{t-k}, u_{t-k}|\bar{v}_{t-k-1}) \right] \prod_{j=1}^k [dP(z_{t-j}|\bar{x}_{t-j-1}, \bar{z}_{t-j-1}) dP_{s^j}(x_{t-j}|\bar{x}_{t-j-1}, \bar{w}_{t-h-1})] \times dP_{s>k}(\bar{v}_{t-k-1}). \tag{B.4}$$

Evaluating the first term in brackets, using (F1), as

$$\begin{aligned} & \int dQ_{k-1,s>k}(y_t|\bar{z}_{t-1}, \bar{y}_{t-k}, \bar{u}_{t-k}) dP_{s>k}(y_{t-k}, u_{t-k}|\bar{v}_{t-k-1}) \\ &= \iint dQ_{k-1,s>k}(y_t|\bar{z}_{t-1}, \bar{y}_{t-k}, \bar{u}_{t-k}) dP_{s>k}(y_{t-k}, u_{t-k}, x_{t-k}|\bar{v}_{t-k-1}) \\ &= \iint dQ_{k-1,s>k}(y_t|\bar{z}_{t-1}, \bar{v}_{t-k}) dP_{s>k}(y_{t-k}, u_{t-k}|x_{t-k}, \bar{v}_{t-k-1}) dP_{s>k}(x_{t-k}|\bar{v}_{t-k-1}) \\ &= \int dQ_{k-1,s>k}(y_t|\bar{z}_{t-1}, x_{t-k}, \bar{v}_{t-k-1}) dP_{s>k}(x_{t-k}|\bar{v}_{t-k-1}) \end{aligned}$$

Since, under  $Q_{k-1,s>k}$ ,  $Z_{t-k}$  depends only on  $X_{t-k}$  and  $V_{t-k-1}$ , the conditional distribution of  $Y_t$  given  $\bar{Z}_{t-1}$ ,  $X_{t-k}$ , and  $\bar{V}_{t-k-1}$  is the same under  $Q_{k-1,s>k}$  and  $Q_{k,s>k}$ . Similarly, the conditional distribution of  $X_{t-k}$  is the same under  $P_{s>k}$  and  $Q_{k,s>k}$ . Noting furthermore that  $\bar{Z}_{t-1} \perp\!\!\!\perp X_{t-k} | \bar{V}_{t-k-1}$  under  $Q_k$ , we finally obtain in (B.4) for the first term in brackets

$$\begin{aligned} & \int dQ_{k,s>k}(y_t|\bar{z}_{t-1}, x_{t-k}, \bar{v}_{t-k-1}) dQ_{k,s>k}(x_{t-k}|\bar{z}_{t-1}, \bar{v}_{t-k-1}) \\ &= dQ_{k,s>k}(y_t|\bar{z}_{t-1}, \bar{v}_{t-k-1}), \end{aligned}$$

This proves B.1 by induction.

It now follows for  $k = h$ , and with  $P_{s>h} = P$  and  $Q_{h,s>h} = Q_h$

$$\begin{aligned} dP_s(y_t) &= \int \cdots \int \left[ \int dQ_h(y_t|\bar{z}_{t-1}, \bar{w}_{t-h-1}, \bar{u}_{t-h-1}) dP(\bar{u}_{t-h-1}|\bar{w}_{t-h-1}) \right] \\ &\quad \times \prod_{j=1}^h [dP(z_{t-j}|\bar{x}_{t-j-1}, \bar{z}_{t-j-1}) dP_{s^j}(x_{t-j}|\bar{x}_{t-j-1}, \bar{w}_{t-h-1})] \\ &\quad \times dP(\bar{w}_{t-h-1}) \\ &= \int \cdots \int dQ_h(y_t|\bar{z}_{t-1}, \bar{w}_{t-h-1}) \prod_{j=1}^h [dP(z_{t-j}|\bar{x}_{t-j-1}, \bar{z}_{t-j-1}) \\ &\quad \times dP_{s^j}(x_{t-j}|\bar{x}_{t-j-1}, \bar{w}_{t-h-1})] dP(\bar{w}_{t-h-1}), \end{aligned}$$

which yields the asserted expression for the ACE.

To conclude the proof of the front-door criterion, we need to show that the expectation with respect to  $\mathbb{Q}_h$  in the expression of the ACE is indeed of the stated form. We have

$$\mathbb{E}_{\mathbb{Q}_h}^{\bar{Z}_{t-1}, \bar{W}_{t-h-1}} Y_t = \int dQ_h(y_t | \bar{w}_{t-1}) \prod_{j=1}^h dQ_h(y_{t-j}, x_{t-j} | \bar{w}_{t-j-1}).$$

Here, the conditional distribution  $dQ_h(y_t | \bar{w}_{t-1})$  has the density

$$\begin{aligned} f_{\mathbb{Q}_h}(y_t | \bar{w}_{t-1}) &= \frac{f_{\mathbb{Q}_h}(y_t, \bar{w}_{t-1} | \bar{w}_{t-h-1})}{f_{\mathbb{Q}_h}(\bar{w}_{t-1} | \bar{w}_{t-h-1})} \\ &= \frac{\int f_{\mathbb{Q}_h}(y_t | \bar{v}_{t-1}) f_{\mathbb{Q}_h}(\bar{v}_{t-1} | \bar{v}_{t-h-1}) d\bar{u}_{t-1} \cdots d\bar{u}_{t-h} dP(\bar{u}_{t-h-1})}{\int f_{\mathbb{Q}_h}(\bar{v}_{t-1} | \bar{v}_{t-h-1}) d\bar{u}_{t-1} \cdots d\bar{u}_{t-h} dP(\bar{u}_{t-h-1})}. \end{aligned} \tag{B.5}$$

The density  $f_{\mathbb{Q}_h}(\bar{v}_{t-1} | \bar{v}_{t-h-1})$  can be further factorized as

$$f_{\mathbb{Q}_h}(\bar{v}_{t-1} | \bar{v}_{t-h-1}) = \prod_{j=1}^h f_{\mathbb{Q}_h}(z_{t-j}) f_{\mathbb{Q}_h}(u_{t-j}, y_{t-j}, x_{t-j} | \bar{v}_{t-j-1}),$$

where  $f_{\mathbb{Q}_h}(u_{t-j}, y_{t-j}, x_{t-j} | \bar{v}_{t-j-1}) = f_{\mathbb{P}}(u_{t-j}, y_{t-j}, x_{t-j} | \bar{v}_{t-j-1})$  by definition of  $\mathbb{Q}_h$ . Since the factors  $f_{\mathbb{Q}_h}(z_{t-j}), j = 1, \dots, h$ , do not depend on  $\bar{u}_{t-1}$ , they cancel in the ratio B.5. Using (F2), we can also replace them by  $f_{\mathbb{P}}(z_{t-j} | x_{t-j}, y_{t-j}, u_{t-j}, \bar{v}_{t-j-1}) = f_{\mathbb{P}}(z_{t-j} | \bar{z}_{t-j-1}, \bar{x}_{t-j})$ . Following the same steps backwards, we obtain that (B.5) is equal to

$$\frac{f_{\mathbb{P}}(y_t, \bar{w}_{t-1} | \bar{w}_{t-h-1})}{f_{\mathbb{P}}(\bar{w}_{t-1} | \bar{w}_{t-h-1})} = f_{\mathbb{P}}(y_t | \bar{w}_{t-1}),$$

which shows that  $dQ_h(y_t | \bar{w}_{t-1}) = dP(y_t | \bar{w}_{t-1})$ . Similarly, it can be shown that  $dQ_h(y_{t-j}, x_{t-j} | \bar{w}_{t-j-1}) = dP(y_{t-j}, x_{t-j} | \bar{w}_{t-j-1})$  for  $1 \leq j \leq h$ , which yields the desired expression for the expectation  $\mathbb{E}_{\mathbb{Q}_h}^{\bar{Z}_{t-1}, \bar{W}_{t-h-1}} Y_t$ .  $\square$

*Proof of Corollary 4.7* As  $a$  is not an ancestor of  $b$ , any  $\text{an}(b)$ -pointing path between  $a$  and  $\text{an}(b)$  must contain nodes  $v, w \in \text{an}(b), v, w \neq a$ , in the constellation  $v \rightarrow w$ , because otherwise it cannot be  $\text{an}(b)$ -pointing. Hence there is always the  $m$ -non-collider  $v$  on any such path so that it must be blocked by  $V \setminus \{a\}$ . With Theorem 4.6 this implies that  $X_b(t+h) \perp\!\!\!\perp \bar{X}_a(t) | \bar{X}_{V \setminus \{a\}}(t)$  and we can prove the claim in exactly the same way as Corollary 2.5.  $\square$

*Proof of Theorem 4.8* We start by constructing the minimal set  $S$  described in part (ii); note that by definition it has to contain  $\{a, b\}$ . Furthermore,  $S$  has to contain the graph parents of  $a$  because otherwise the back-door path  $a \leftarrow v_1 \rightarrow a$  is  $m$ -connecting and  $\text{an}(b)$ -pointing (as  $a \in \text{an}(b)$ ). Any path starting  $a \leftarrow v_1 \cdots$  is blocked by the parents of  $a$  as  $v_1 \in \text{pa}(a)$  is a non-collider. Thus, we now only need to consider back-door paths that start with  $a \leftarrow v_2 \cdots$ . If  $v_2 \in \text{an}(b)$  it has to be contained in  $S$  otherwise  $a \leftarrow v_2 \rightarrow \cdots b$  is  $m$ -connecting. If  $v_2 \notin \text{an}(b)$  then

any an(*b*)-pointing path  $a \dashrightarrow v_2 \dashrightarrow \dots$  or  $a \dashrightarrow v_2 \leftarrow \dots$  is *m*-blocked if  $v_2 \notin S$ ; and every an(*b*)-pointing path starting  $a \dashrightarrow v_2 \rightarrow \dots$  must contain a constellation  $\rightarrow v_3 \dashrightarrow$  or  $\rightarrow v_3 \leftarrow$  with  $v_3 \notin \text{an}(b)$ , so is *m*-blocked as  $v_3 \notin S$ . The same reasoning can be applied iteratively to show that any  $v \in D$  has to be in *S*. It is clear from this reasoning that whenever one node of  $\text{pa}(a)$  or of *D* is not contained in *S* an *m*-connecting back-door path to an(*b*) is created, so the set *S* is minimal.

Now we show part (i), assuming that *S* has been chosen according to (ii). Define the set *F* as the vertices *v* in *S* such that  $v = a$  or there exists an undirected path  $v \dashrightarrow \dots \dashrightarrow a$  with all intermediate vertices lying in *S*. From the construction of *S* it follows that  $S = \{b\} \cup F \cup \text{pa}(F)$  and as *S* contains all parents of *F* (see (3.10))

$$X_F(t) \perp\!\!\!\perp \bar{X}_{V \setminus S}(t-1) \mid \bar{X}_S(t-1). \tag{B.6}$$

(The above, as well as all following conditional independencies that do not explicitly mention  $\sigma_a(t)$  all hold under both values of the indicator.) Furthermore, it also follows that  $\text{pa}(F) \cap \text{sp}(F) \subseteq F$ , which implies that  $X_F$  and  $X_{\text{pa}(F) \setminus F}$  are contemporaneously independent wrt.  $X_V$  (see (3.11)). If *b* is a spouse of a vertex in *F* then  $b \in F$ , and we obtain from B.6 and the contemporaneous independence, using (C4),

$$X_F(t) \perp\!\!\!\perp (\bar{X}_{V \setminus S}(t-1), X_{\text{pa}(F) \setminus F}(t)) \mid \bar{X}_S(t-1)$$

and further, since in this case  $S = F \cup \text{pa}(F)$ ,

$$X_a(t) \perp\!\!\!\perp \bar{X}_{V \setminus S}(t-1) \mid \bar{X}_S(t-1), X_{S \setminus \{a\}}(t). \tag{B.7}$$

If  $b \notin \text{sp}(F)$  then  $X_F$  and  $X_{\{b\} \cup \text{pa}(F) \setminus F}$  are contemporaneously independent, which together with B.6 yields again B.7.

Next, we note that from 2.2 we have  $\bar{X}_{V \setminus S}(t-1) \perp\!\!\!\perp \sigma_a(t) \mid (X_{S \setminus \{a\}}(t), \bar{X}_S(t-1))$ . By B.7 and (C4), this can be extended to

$$\bar{X}_{V \setminus S}(t-1) \perp\!\!\!\perp \sigma_a(t) \mid \bar{X}_S(t), \tag{B.8}$$

that is, when additionally conditioning on  $X_a(t)$ .

From the definition of *F* and the global Markov properties, it also follows that

$$X_a(t) \perp\!\!\!\perp X_{V \setminus (S \cup \text{sp}(F))}(t) \mid X_{S \setminus \{a\}}(t), \bar{X}_V(t-1). \tag{B.9}$$

Due to (I4), the intervention distribution of  $X_a(t)$  only depends on  $\bar{X}_C(t-1)$  with  $C \subseteq S$ , the relation holds under both the observational and the experimental regime. From (I1) it also follows using (C3) that

$$X_{V \setminus (S \cup \text{sp}(F))}(t) \perp\!\!\!\perp \sigma_a(t) \mid X_{S \setminus \{a\}}(t), \bar{X}_V(t-1). \tag{B.10}$$

Properties B.9 and B.10 together imply with (C4) and (C3) that

$$X_{V \setminus (S \cup \text{sp}(F))}(t) \perp\!\!\!\perp \sigma_a(t) \mid \bar{X}_S(t), \bar{X}_{V \setminus S}(t-1). \tag{B.11}$$

Now (B.8) and (B.11) together imply with (C4) that

$$X_{V \setminus (S \cup \text{sp}(F))}(t), \bar{X}_{V \setminus S}(t-1) \perp\!\!\!\perp \sigma_a(t) \mid \bar{X}_S(t). \quad (\text{B.12})$$

Further it follows from assumption 2.3 that  $X_b(t+h) \perp\!\!\!\perp \sigma_a(t) \mid \bar{X}_V(t)$  and from (3.12) that  $X_b(t+h) \perp\!\!\!\perp \bar{X}_{V \setminus \text{an}(b)}(t) \mid (\bar{X}_{\text{an}(b)}(t), \sigma_a(t))$ , which together imply

$$X_b(t+h) \perp\!\!\!\perp \sigma_a(t), \bar{X}_{V \setminus \text{an}(b)}(t) \mid \bar{X}_{\text{an}(b)}(t). \quad (\text{B.13})$$

We note that  $\text{an}(b) \setminus S \subseteq V \setminus (S \cup \text{sp}(F))$ : every vertex in  $F \cup \text{sp}(F)$  is  $m$ -connected to  $a$  given  $S$  and thus by definition of  $S$  must be in  $S$  if it is an ancestor of  $b$ . Consequently, B.13 can be rearranged (with (C2) and (C3)) to give

$$X_b(t+h) \perp\!\!\!\perp \sigma_a(t) \mid \bar{X}_S(t), \bar{X}_{V \setminus S}(t-1), X_{V \setminus (S \cup \text{sp}(F))}(t).$$

Together with B.12 the latter implies

$$X_b(t+h) \perp\!\!\!\perp \sigma_a(t) \mid \bar{X}_S(t),$$

which means that Theorem 3.1 holds for the minimal choice of  $S$ . If  $S$  is not minimal then the proof gets more complicated but follows essentially the same lines. Finally, to see that Theorem 3.2 also holds we simply note that all relations up to B.12 concern only the distribution of  $\bar{X}_V(t)$  and thus are not affected by interventions at later time points. Similarly, assumptions (I1) to (I4) on interventions imply that

$$X_b(t+h) \perp\!\!\!\perp \sigma_a(t) \mid \bar{X}_V(t), \sigma^{>t} = s,$$

from which the final result can be obtained similarly as above.  $\square$

*Proof of Theorem 4.12* Conditions (G2) and (G3) imply that  $X_{V \setminus (C \cup \{a\})}$  is Granger noncausal for, and contemporaneously independent of,  $X_C$  with respect to  $X_V$ . With (C4) this immediately implies (F2), noting that  $U \cup \{b\} = V \setminus (C \cup \{a\})$ .

For condition (F1), we note that  $X_C(t-1), \dots, X_C(t-k+1)$  and  $\bar{X}_V(t-k)$  are independent under the measure  $\mathbb{Q}_{k-1}$ . Thus the dependencies of the process over the period from  $t-k$  to  $t$  can be described by the graph  $\tilde{G}$  obtained from the original graph by omitting all directed edges with an arrowhead at  $C$  and all undirected edges with one endpoint in  $C$ . Then assumption (G1) implies that there is no directed path from  $a$  to  $b$  in  $\tilde{G}$ . It follows that

$$X_b(t) \perp\!\!\!\perp \bar{X}_a(t-k) \mid \bar{X}_{V \setminus \{a\}}(t-k) \quad [\mathbb{Q}_{k-1}].$$

Furthermore, it follows from assumptions (I1) to (I3) (and under either of  $\mathbb{P}$  or  $\mathbb{Q}_{k-1}$ ) that

$$X_b(t) \perp\!\!\!\perp \sigma_a(t-k) \mid \bar{X}_V(t-k).$$

Combining the two relations by (C4) we obtain (F1).  $\square$

## References

- Aalen OO (1987) Dynamic modeling and causality. *Scand Actuar J* :177–190
- Aalen OO, Frigessi A (2007) What can statistics contribute to a causal understanding. *Scand J Stat* 34: 155–168
- Box GEP, Jenkins GM, Reinsel GC (1994) *Time series analysis, forecasting and control*. 3. Prentice Hall, Englewood Cliffs
- Commenges D, Gegout-Petit A (2007) A general dynamical statistical model with possible causal interpretation
- Comte F, Renault E (1996) Noncausality in continuous time models. *Econometr Theory* 12:215–256
- Cox DR, Wermuth N (1996) *Multivariate dependencies—models analysis and interpretation*. Chapman & Hall, London
- Dawid AP (1979) Conditional independence in statistical theory (with discussion). *J R Stat Soc B* 41:1–31
- Dawid AP (2000) Causal inference without counterfactuals. *J Am Stat Assoc* 86:9–26
- Dawid AP (2002) Influence diagrams for causal modelling and inference. *Int Stat Rev* 70:161–189
- Dawid AP and Didelez V (2005) Identifying the consequences of dynamic treatment strategies, Technical report 262. Department of Statistical Science, University College London
- Didelez V (2007) Graphical models for composable finite markov processes. *Scand J Stat* 34:169–185
- Didelez V (2008) Graphical models for marked point processes based on local independence. *J R Stat Soc B* 70:245–264
- Eichler M (2001) Graphical modelling of multivariate time series, Technical report. Universität Heidelberg
- Eichler M (2006) Graphical modelling of dynamic relationships in multivariate time series. In: Winterhalder M, Schelter B, Timmer J (eds) *Handbook of time series analysis*. Wiley-VCH, pp 335–372
- Eichler M (2007) Granger causality and path diagrams for multivariate time series. *J Econometr* 137: 334–353
- Eichler M (2009) Causal inference from multivariate time series: what can be learned from granger causality. In: Glymour C, Wang W, Westerståhl D (eds) *Proceedings from the 13th international congress of logic, methodology and philosophy of science*. King's College Publications, London
- Eichler M, Didelez V (2007) Causal reasoning in graphical time series models. *Proceedings of the 23rd conference on uncertainty in artificial intelligence*
- Florens JP, Fougère D. (1996) Noncausality in continuous time. *Econometrica* 64:1195–1212
- Florens JP, Mouchart M (1982) A note on noncausality. *Econometrica* 50:583–591
- Florens JP, Mouchart M, Rolin JM (1990) *Elements of Bayesian statistics*. Marcel Dekker, New York
- Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37:424–438
- Granger CWJ (1980) Testing for causality, a personal viewpoint. *J Econ Dyn Control* 2:329–352
- Granger CWJ (1986) Comment on *Statistics and causal inference* by P. Holland. *J Am Stat Assoc* 81: 967–968
- Granger CWJ (1988) Some recent developments in a concept of causality. *J Econometr* 39:199–211
- Heckman JJ (2008) Econometric causality. *Int Stat Rev* 76:1–27
- Heckman JJ, Navarro S (2007) Dynamic discrete choice and dynamic treatments effect. *J Econon* 136: 341–396
- Hernán MA, Hernández-Díaz S, Robins JM (2004) A structural approach to selection bias. *Epidemiology* 15:615–625
- Hernán MA, Taubman SL (2008) Does obesity shorten life? the importance of well defined interventions to answer causal questions. *Int J Obes* 32:8–14
- Hsiao C (1982) Autoregressive modeling and causal ordering of econometric variables. *J Econ Dyn Control* 4:243–259
- Kallenberg O (2001) *Foundations of modern probability*. 2nd edn. Springer, New York
- Lauritzen SL (1996) *Graphical models*. Oxford University Press, Oxford
- Lauritzen SL (2001) Causal inference from graphical models. In: Barndorff-Nielsen OE, Cox DR, Klüppelberg C (eds) *Complex stochastic systems*. CRC Press, London pp 63–107
- Pearl J (1993) Graphical models, causality and interventions. *Stat Sci* 8:266–269
- Pearl J (1995) Causal diagrams for empirical research (with discussion). *Biometrika* 82:669–710
- Pearl J (2000) *Causality*. Cambridge University Press, Cambridge, UK
- Richardson T (2003) Markov properties for acyclic directed mixed graphs. *Scand J Stat* 30:145–157
- Robins JM (1986) A new approach to causal inference in mortality studies with sustained exposure periods — application to control for the healthy worker survivor effect. *Math Model* 7:1393–1512

- Robins JM, Hernán MA, Siebert U (2004) Effects of multiple interventions. In: Ezzati M, Lopez AD, Rodgers A, Murray CJL (eds) Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors, vol 1. World Health Organization, Geneva
- Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J Edu Psychol* 66:688–701
- Rubin DB (1978) Bayesian inference for causal effects: the role of randomization. *Ann Stat* 6:34–68
- Schweder T (1970) Composable markov processes. *J Appl Prob* 7:400–410
- Spirtes P, Glymour C, Scheines R (2001) Causation, prediction, and search, 2nd edn. [MIT Press, Cambridge, MA. With additional material by David Heckerman, Christopher Meek, Gregory F Cooper and Thomas Richardson]
- White H (2006) Time series estimation of the effects of natural experiments. *J Econometr* 135:527–566