
Sign Identifiability of Causal Effects in Stationary Stochastic Dynamical Systems

Gijs van Seeventer¹

Saber Salehkaleybar¹

¹LIACS, Leiden University, Netherlands

Abstract

We study identifiability in continuous-time linear stationary stochastic differential equations with a known causal structure. Unlike existing approaches, we relax the assumption of a known diffusion matrix, thereby respecting the model’s intrinsic scale invariance. Therefore, rather than recovering drift coefficients themselves, we introduce edge-sign identifiability: for a given causal structure, we ask whether the sign of a given drift entry is uniquely determined across all observational covariance matrices induced by parametrisations compatible with that structure. This leads to a trichotomy of edge-sign identifiability: identifiable, non-identifiable, and partially identifiable. This trichotomy introduces the new notion of partial identifiability to the literature, which we show is a genuine category in our setting. Under a notion of faithfulness, we derive criteria to identify membership of each category for general graphs. Applying our criteria to specific causal structures, both analogous to classical causal settings (e.g., instrumental variables) and novel cyclic settings, we determine their edge-sign identifiability and, in some cases, obtain explicit expressions for the sign of a target edge in terms of the observational covariance matrix.

1 INTRODUCTION

Learning dynamical systems from observational data via parametrised models is central across scientific domains, from systems biology [Marbach et al., 2012] to economics [Hamilton, 2020]. The aim of such modelling is often to answer a causal question: *what happens to a target variable Y if we intervene on a variable X ?*

Recently, causal modelling of stationary diffusions has

emerged to address settings in which time trajectories are unavailable or unobservable [Fitch, 2019, Lorch et al., 2024]. In such cases, observations can be viewed as samples collected from a stationary process. These processes are often described by stationary stochastic differential equations (SDEs) [Øksendal, 2003]. While stationary SDEs induce time-invariant observational distributions, they internally encode temporal causal dependencies, allowing for a natural causal interpretation [Lorch et al., 2024, Améndola et al., 2025]. This interpretation admits a graphical representation analogous to structural causal models (SCMs) [Sokol and Hansen, 2014, Pearl, 2009]. Unlike acyclic SCMs, graphical models induced by stationary SDEs naturally allow for cycles and self-loops, features ubiquitous in real-world dynamical systems.

In the literature on graphical models, it is often assumed that the causal structure, namely, which variables directly affect others, is known. Given a fixed causal structure, a central question is identifiability, i.e., whether the value of a causal effect (or a property of it, such as its sign) can be determined from the observational distribution. While determining the magnitude of a causal effect is preferable, the sign alone is often informative, e.g., whether an intervention helps or harms. Therefore, sign identifiability has been and remains an active research topic [Manski, 1990, Lanners et al., 2026].

Identifiability is also a central question in the literature on causal SDE models, which studies causal SDEs under various assumptions (see Section 3). In the absence of path data, we focus on the most tractable and interpretable SDE: continuous-time, linear, time-homogeneous stationary SDEs, i.e., the stationary Ornstein-Uhlenbeck (OU) process. For OU processes, existing notions of identifiability assume a known causal structure and a fixed part of the model parameters, specifically the diffusion matrix. However, fixing the diffusion matrix imposes a strong restriction since the OU process is invariant under positive rescaling (see Section 2.1.1). Therefore, respecting this scale invariance, we assume that only the causal structure is known, while the diffusion matrix is allowed to be rescaled. Our

main contributions are as follows:

- **Sign identifiability.** Only assuming the causal structure to be known limits what can be learned about the causal effect. Therefore, we introduce a notion of edge-sign identifiability (see Section 2.2), focusing on the sign rather than the magnitude of a causal effect.
- **Categories of sign identifiability.** Our analysis distinguishes three categories: identifiable, non-identifiable, and partially identifiable. While the term *partial identifiability* is often used as a name for the broader research area of bounding causal effects [Lanterns et al., 2026], we use it here to denote a new identifiability category for edge signs. We also show that in the confounding structure, the sign of the causal effect is partially identifiable with positive measure (for more details, see Section 4.1), making it a genuine category. Moreover, numerical experiments in Section 5 suggest that partial identifiability constitutes an important intermediate category besides identifiability and non-identifiability.
- **General criteria and applications to specific structures.** We derive general criteria for edge-sign identifiability and apply them to specific graph structures in Section 4. General criteria include, for instance, a graphical criterion for determining whether an edge is identifiable. We apply the general criteria in some special graph structures, including causal structures analogous to bivariate cause–effect and instrumental variable settings, for which we obtain an explicit expression for the sign of the causal effect in terms of the covariance matrix.

2 PRELIMINARIES AND PROBLEM SETUP

2.1 BACKGROUND AND NOTATION

We denote random vectors in $\mathbb{R}^d$ as $X \in \mathbb{R}^d$. Its i th entry is $X_i \in \mathbb{R}$ for $i \in \{1, \dots, d\}$. In addition, we denote the set of positive definite matrices as PD_d and the set of positive definite diagonal matrices as PDD_d , where d is the dimension. Note that $PDD_d \subset PD_d$.

2.1.1 Stochastic Differential Equations

Stochastic differential equations (SDEs) describe stochastic processes $\{X(t)\}$, $X(t) \in \mathbb{R}^d$, which is a collection of random vectors $X(t)$ for a time t . We will consider *continuous stationary* processes, which means that the probability density $f_t(X(t))$ is the same for all considered times, i.e., $f_t(X(t)) = f_{t'}(X(t'))$ with $t, t' \geq 0$. A general SDE has the form $dX(t) = g(X(t), t)dt + h(X(t), t)d\beta(t)$, where $g : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^d$ is the drift function, $h : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \rightarrow$

$\mathbb{R}^{d \times d}$ is the diffusion function and β is a random noise term. We consider linear, time-homogeneous SDEs driven by a Wiener process, i.e., SDEs with linear, time-invariant drift and diffusion functions and Gaussian noise. We will assume the commonly used *Itô* interpretation, which is a choice in what it means to integrate a random noise term [Øksendal, 2003].

The only non-trivial continuous stationary, linear and time-homogeneous SDE is known as the multivariate Ornstein-Uhlenbeck (OU) process [Doob, 1942].¹ The OU process is described by

$$dX(t) = A(X(t) - b)dt + CdW(t), \quad (1)$$

where $A \in \mathbb{R}^{d \times d}$ is the drift matrix, $b \in \mathbb{R}^d$ is a constant vector, $C \in \mathbb{R}^{d \times d}$ is the diffusion matrix and $W(t)$ is the d -dimensional Wiener process at time t . Due to being a Wiener process, the OU process is characterised by the first two moments, i.e., the mean $m(t)$ and the covariance matrix $\Sigma(t)$. Since we consider a stochastic *stationary* process, $\frac{d}{dt}m(t) = 0$ and $\frac{d}{dt}\Sigma(t) = 0$. Therefore the mean $m = b$ and every entry of the covariance matrix $\Sigma_{ij} = \mathbb{E}[(X_i(t) - m)(X_j(t) - m)]$. Without loss in generality, we take the mean $m = 0$ such that $X(t) \sim \mathcal{N}(0, \Sigma)$. Moreover, the stationarity condition $\frac{d}{dt}\Sigma = 0$ is equivalent to the Lyapunov equation

$$A\Sigma + \Sigma A^T = -D, \quad (2)$$

where $D = CC^T$ [Särkkä and Solin, 2019]. In addition, stationarity requires the drift matrix A to be Hurwitz stable, i.e., the real parts of the eigenvalues of A must be negative. Note that, given $D \in PD_d$, we have that A is Hurwitz stable in the Lyapunov equation (Eq. (2)) if and only if $\Sigma \in PD_d$ [Frommer and Hashemi, 2012, Theorem 1.1]. Furthermore, if we choose an uncorrelated noise setting, i.e., the diffusion matrix C being diagonal, we will always have $D \in PDD_d \subset PD_d$. If we have correlated noise, C is no longer diagonal, and it needs to be verified that $D \in PD_d$. Unless otherwise noted, we will assume $D \in PDD_d$.

A Wiener process W is scale invariant, i.e., $W(t) = W(at)/\sqrt{a}$ for $a \in \mathbb{R}_+$. Furthermore, given that we are considering a stationary process, we have that two times $t, at \in [0, \infty)$ are indistinguishable in the sense that $X(t) \stackrel{d}{=} X(at)$ in distribution. This means that we can always rescale the OU process with $\sqrt{a}$ to rewrite the original Eq. (1) into

$$dX(t) = aA(X(t) - m)dt + \sqrt{a}CdW(t). \quad (3)$$

Note that scaling $C \mapsto \sqrt{a}C$ implies $D = CC^T \mapsto aD$. To preserve the Lyapunov equation (Eq. (2)), we need that $A \mapsto$

¹If the OU process is not initialized at its stationary distribution, it converges to stationarity only asymptotically. We therefore assume that the process is either initialized at stationarity or observed after a sufficiently long period. After this period, we relabel time so that the stationary regime starts at $t = 0$ and consider $t \geq 0$.

aA . This means that we can always rescale the $(A, D) \mapsto (aA, aD)$. Therefore, from the covariance matrix, we can only identify the parameters of the drift matrix A up to some global scaling $a > 0$.

2.1.2 Graphs

Let $G = (V, E)$ be a directed graph with node set $V = \{V_0, \dots, V_d\}$ and edges E where an edge $e := (V_i, V_j) \in E$ is a directed edge from V_i to V_j . Unless stated otherwise, all graphs are directed.

A directed path $V_i, \dots, V_j$ is a sequence such that there is a directed edge $V_k \rightarrow V_{k+1}$ for all $k = i, \dots, j-1$, where we call V_i the ancestor of V_j . We denote the set of ancestors of V_i in graph G by $\text{An}_G(V_i)$. Furthermore, if a directed path starts and ends at the same node, we call it a cycle, i.e. $V_i \rightarrow \dots \rightarrow V_i$. If a cycle has n distinct nodes, then the cycle has length n . A special example of a cycle is the self-loop. A self-loop is a directed path from a node to itself, i.e., $V_i \rightarrow V_i$. Note that if a node V_i is in a cycle, it is in its own ancestral set, i.e., $V_i \in \text{An}_G(V_i)$. If there are no cycles in a graph, we call it a directed acyclic graph (DAG).

2.1.3 Causal Interpretation of SDEs

Stochastic processes described by SDEs admit a causal interpretation in which interventions correspond to modifications of the equations [Sokol and Hansen, 2014, Lorch et al., 2024]. In particular, for the OU process (Eq. (1)), we interpret a non-zero drift matrix entry $A_{ij} \neq 0$ as a direct causal effect of process X_j on X_i , with a causal strength given by A_{ij} , in line with [Améndola et al., 2025, Varando and Hansen, 2020]. Under this perspective, several concepts from the structural causal model (SCM) literature are useful. One such concept is the representation of direct causal relations by a directed graph $G = (V, E)$, where the nodes $V_i \in V$ correspond to the process X_i and edges $e \in E$ to the presence of a direct causal effect. When considering only the orientation of direct causal effects, i.e., ignoring their strength, we refer to this as the causal structure [Peters et al., 2017]. Accordingly, a drift matrix entry $A_{ij} \neq 0$ is represented by an edge $V_j \rightarrow V_i$ with edge weight A_{ij} . With slight abuse of notation, we will represent an edge $e = (V_j, V_i) \in E$ with its corresponding drift entry A_{ij} , writing $A_e := A_{ij}$. Then, $\text{supp}(A) \subseteq E$ means that $e \notin E \implies A_e = 0$. If equality holds, i.e., $\text{supp}(A) = E$, then $e \notin E \iff A_e = 0$. The latter corresponds to structural minimality in the SCM literature [Peters et al., 2017]. We will always assume structural minimality.

In addition, since Hurwitz stable matrices may have non-zero diagonal entries and need not be triangular, the graphs representing SDEs can contain cycles and self-loops. For triangular matrices, however, the eigenvalues coincide with the diagonal entries. Hence, Hurwitz stability in the triangular

case requires all diagonal entries to be negative.²

A useful distinction in the SCM literature is between causal discovery (learning the causal graph from data) and causal inference (causal effect identification given a specified graph). We consider the latter setting and assume the directed graph $G = (V, E)$ is known.

Transferring concepts from the existing SCM literature to SDE-based models requires care because the underlying data-generating mechanisms differ. In an SCM, data are generated by structural assignments, and conditional-independence constraints are typically characterised via graphical separation criteria (e.g., d -separation for DAGs and σ -separation for cyclic SCMs) [Peters et al., 2017, Bongers et al., 2021]. In contrast, an SDE generates data through continuous-time stochastic dynamics, and in our setting, we observe samples from the stationary distribution induced by the SDE. Moreover, for SDEs describing stationary processes, such as the OU process, marginal independences are the only source of independence relations [Boege et al., 2025]. This contrasts sharply with the SCM literature in which additional conditional independences arise and are characterised via d - and σ -separation.

2.2 DEFINITIONS

This section introduces the definitions used throughout the paper. We begin by defining a set of covariance matrices Σ that encode the marginal independences implied by the assumed directed graph G . As discussed in the previous subsection, in our setting, marginal independences are the only source of independence constraints coming from a graph $G = (V, E)$. Under the assumption of a diffusion matrix $D \in PDD_d$ and $\text{supp}(A) \subseteq E$, [Boege et al., 2025, Theorem 1.2] gave a graphical criterion for such marginal independencies based on ancestral relationships. In particular, $X_i \perp\!\!\!\perp X_j$ in any random vector X distributed according to a stationary distribution respecting the causal structure, if and only if, $\text{An}_G(V_i) \cap \text{An}_G(V_j) = \emptyset$.

Definition 2.1 (T-Faithfulness). We define the set F_G of t -faithful covariance matrices Σ for a graph $G = (V, E)$ as

$$F_G := \{\Sigma : \Sigma \in PDD_d; \forall V_i, V_j \in V, \text{An}_G(V_i) \cap \text{An}_G(V_j) = \emptyset \iff \Sigma_{ij} = 0\}. \quad (4)$$

Remark 2.2. In other words, t -faithfulness assures that the marginal independences captured by Σ are exactly those that can be read off from the graph by checking common ancestral relationships of pairs of variables. The latter can be encoded more densely represented in a trek graph (see [Boege et al., 2025, below Definition 1.1]), hence the term t -faithful. The definition further guarantees that $\Sigma \in PDD_d$,

²Ignoring diagonal entries (self-loops), a triangular matrix corresponds to a directed acyclic graph (DAG).

thereby ensuring valid solutions to the Lyapunov equation (Eq. (2)) (see Section 2.1.1).

The following four definitions focus on the possible $\text{sign}(A_e) \in \{+, -, 0\}$ associated with edges e in the directed graph $G = (V, E)$ under the t -faithfulness assumption. Our focus on signs is motivated by the scaling invariance discussed in Section 2.1.1 where, for a given Σ , if (A, D) satisfies the Lyapunov equation (Eq. (2)), then (aA, aD) also satisfies it for any $a > 0$. Hence, the drift matrix is only identifiable up to a global positive rescaling, and for any given edge e , we treat its sign as the primary information that can be recovered from Σ . The first two Definitions 2.3 and 2.5 characterise which covariance matrices Σ are compatible with a given $\text{sign}(e)$ under a graph G . The third Definition 2.6 introduces a new notion of identifiability. When fixing the covariance matrix Σ , the third definition reduces to the fourth Definition 2.8.

Definition 2.3 (Edge Signature Set). For a graph $G = (V, E)$ and edge $e \in E$, we define the edge signature set $\mathcal{M}_{G,e}^k$ as

$$\mathcal{M}_{G,e}^k := \{\Sigma \in F_G : \exists A, D \text{ s.t. } A\Sigma + \Sigma A^T = -D; \\ D \in PDD_d; \text{supp}(A) = E; \text{sign}(A_e) = k\}, \quad (5)$$

where $k \in \{+, -\}$ and

$$\mathcal{M}_{G,e}^0 := \{\Sigma \in F_G : \exists A, D \text{ s.t. } A\Sigma + \Sigma A^T = -D; \\ D \in PDD_d; \text{supp}(A) \subset E; \text{sign}(A_e) = 0\}. \quad (6)$$

Remark 2.4. We interpret this definition as all t -faithful covariance matrices Σ that could generate a $\pm$ sign for edge $e \in E$ from graph G when the drift matrix matches the causal structure of the graph $G = (V, E)$. While we focus on studying the $\mathcal{M}_{G,e}^+$ and $\mathcal{M}_{G,e}^-$, the $\mathcal{M}_{G,e}^0$ edge signature set will be a useful theoretical tool.

Since we are only interested in covariance matrices compatible with the graph under structural minimalism, we define a set of possible covariance matrices.

Definition 2.5 (Possible Set). For a graph $G = (V, E)$ and edge $e \in E$, define the possible set as

$$\mathcal{M}_{G,e}^p := \mathcal{M}_{G,e}^+ \cup \mathcal{M}_{G,e}^-. \quad (7)$$

Next, we introduce the notion of edge-sign identifiability.

Definition 2.6 (Edge-Sign Identifiability). The sign of edge $e \in E$ in graph $G = (V, E)$ with signature-sets $\mathcal{M}_{G,e}^k$ and $k \in \{+, -\}$, is:

- non-identifiable if $\mathcal{M}_{G,e}^+ = \mathcal{M}_{G,e}^-$ while $\mathcal{M}_{G,e}^- \neq \emptyset$,
- partially identifiable if $\mathcal{M}_{G,e}^+ \neq \mathcal{M}_{G,e}^-$ and $\mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^- \neq \emptyset$,

- identifiable if $\mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^- = \emptyset$ while $\mathcal{M}_{G,e}^- \neq \emptyset$ or $\mathcal{M}_{G,e}^+ \neq \emptyset$.

Remark 2.7. Intuitively, Definition 2.6 formalizes whether the sign of the edge weight A_e is determined by the covariance matrix. If e is *identifiable*, then all the covariance matrices fix the sign. If e is *non-identifiable*, then no covariance matrix ever resolves the sign in the sense that whenever the covariance matrix is compatible with one sign, it is also compatible with the other. If e is *partially identifiable*, then there exist covariance matrices for which the sign is uniquely determined, and there exist others that are compatible with both signs. Considering the edge weight A_e as the direct causal strength, the edge-sign identifiability thus formalises whether we can learn its sign value.

When considering a fixed covariance matrix $\Sigma \in \mathcal{M}_{G,e}^p$, Definition 2.6 reduces to the following definition. For a proof, see Appendix A.

Definition 2.8 (Pointwise Edge-Sign Identifiability). Let $e \in E$ be an edge in graph $G = (V, E)$ with signature-sets $\mathcal{M}_{G,e}^k$ and $k \in \{+, -\}$. For a covariance matrix $\Sigma \in \mathcal{M}_{G,e}^p$, the sign of e is:

- non-identifiable if $\Sigma \in \mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^-$,
- identifiable if $\Sigma \notin \mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^-$.

2.3 SIGN IDENTIFICATION PROBLEM

Assume the data-generating process is an OU process (Eq. (1)) with uncorrelated noise such that $D \in PDD_d$ under the assumption of structural minimality. Given a directed graph $G = (V, E)$ and an edge $e \in E$, determine whether the sign of e is non-identifiable, partially identifiable, or identifiable according to the Definition 2.6. If identifiable, determine its $\text{sign}(e)$.

3 RELATED WORK

The use of SDEs as a causal modelling tool is an active field of research. SDEs can model both dynamic and stationary processes. Many works focus on dynamic processes [Mogensen et al., 2018, Stippinger et al., 2023, Cinquini et al., 2025]. This requires access to sample paths (time trajectories), which is not always feasible in practice. For example, most single-cell RNA sequencing techniques destroy the cell being sampled [Liu et al., 2024]. On the other hand, works on stationary processes in general do not require access to sample paths, although there are exceptions [Manten et al., 2024]. Without requiring a sample path, typically obtained via discrete measurements, most works (discussed below) focus on continuous-time models. However, there are exceptions that consider discrete time [Recke et al., 2026].

The research on causal modelling of stationary processes with continuous time has so far mainly focused on the study of linear SDEs with Gaussian noise, i.e., on stationary OU processes. Since the main interest is the direct causal effects and the associated sparsity structure, the research has focused on the drift matrix A . The drift matrix A is constrained by the Lyapunov equation (Eq. (2)), which may explain why some works on *causal* stationary OU processes refer to it as *graphical continuous Lyapunov models* (GCLM), first coined in [Varando and Hansen, 2020]. In this line of research, initial works focused on causal discovery [Fitch, 2019, Dettling et al., 2024].

More recently, identifiability has received increased attention. Dettling et al. [2023] introduces a notion of (generic) identifiability based on uniqueness: for a given graph G and $D \in PD_d$, identifiability holds if all stable drift matrices A are in one-to-one correspondence (almost surely) with covariance matrices Σ . While they derive results under $D \in PD_d$, stronger and more comprehensive characterisations are obtained under the stricter assumption $D \in PDD_d$. Their characterisations combine conditions derived from the covariance matrix Σ with structural constraints given by the sparsity pattern of A . Building on this work, Améndola et al. [2025] considers $D \in PDD_d$ and graphs G that are acyclic when self-loops are ignored. They provide a graphical criterion for model equivalence together with a polynomial-time algorithm to decide whether a model is unique in a given equivalence class and whether two models are equivalent.

Our work differs from these approaches by relaxing the strong assumption that D is known. Requiring a fixed D in addition to the graph, ignores the scale invariance inherent to the stationary OU process, as reflected both in the Lyapunov equation (Eq. (2)) and in the driving Wiener process. For example, as a consequence, the identifiability notion in [Dettling et al., 2023] does not capture *partial sign identifiability*: for a given graph, there may exist admissible covariance matrices Σ for which the sign of an edge weight is identifiable, and others for which it is not. As an example, see Remark 4.6.

Beyond linear SDEs, recent advances have addressed causal discovery for general drift and diffusion parametrisations in stationary continuous-time processes. Lorch et al. [2024] propose a method based on a kernel objective that quantifies the deviation of an SDE parametrisation from empirical observations. Bleile et al. [2026] improved this approach in terms of computational efficiency. Our sign identifiability results could be of interest for linear parametrisations learned by these methods to assess whether they can correctly recover edge signs in cases where the signs are identifiable. In addition, in such settings, these methods also do not account for the existence of partial and non-identifiable edges.

4 EDGE-SIGN IDENTIFIABILITY RESULTS

This section presents theoretical results on edge-sign identifiability. All proofs rely on the Lyapunov equation (Eq. (2)). Section 4.1 establishes theorems that hold for arbitrary graphs G . Section 4.2 uses these theorems to analyse specific causal structures. We present results both with and without latent variables.

4.1 SIGN IDENTIFIABILITY IN GENERAL GRAPHS

We begin by presenting two lemmas and a theorem ($\mathcal{M}_e^0$ criterion) that establish sign identifiability results for a fixed covariance matrix Σ . We continue by presenting a theorem (graphical criterion) and a supporting lemma. The theorem establishes a graphical rule to determine identifiability for a given graph G and edge e , i.e., all covariance matrices entailed by the graph G in the possible edge set. This theorem is valid only for graphs without latent variables (see Remark 4.2). We end by providing two examples of applying the graphical criterion to general structures.

Lemma 4.1. *Let $e \in E$ be an edge in a graph $G = (V, E)$. Then*

$$\Sigma \in \mathcal{M}_{G,e}^+ \text{ and } \Sigma \in \mathcal{M}_{G,e}^- \implies \Sigma \in \mathcal{M}_{G,e}^0, \quad (8)$$

$$\Sigma \in \mathcal{M}_{G,e}^+ \text{ and } \Sigma \in \mathcal{M}_{G,e}^0 \implies \Sigma \in \mathcal{M}_{G,e}^-, \quad (9)$$

$$\Sigma \in \mathcal{M}_{G,e}^- \text{ and } \Sigma \in \mathcal{M}_{G,e}^0 \implies \Sigma \in \mathcal{M}_{G,e}^+. \quad (10)$$

Proof sketch. The proof proceeds analogously for all implications. We select a covariance matrix Σ in the intersection of the two signature sets on the left-hand side of the implication, which also ensures $\Sigma \in F_G$. For this Σ , we obtain two Lyapunov equations (Eq. (2)). Each can be rescaled by an arbitrary scalar $a \in \mathbb{R}$, and their sum remains a valid Lyapunov equation. Such a rescaling, however, need not correspond to a valid OU model. We therefore choose the rescaling so that $D \in PDD_d$. To satisfy Definition 2.3 for the signature set on the right-hand side of the implication, we additionally ensure that the rescaling preserves $\text{supp}(A) = E$ and yields the required $\text{sign}(A_e)$. The full proof is provided in Appendix D.1.

Remark 4.1. We emphasise that the key mechanism in the proof is the scale invariance of the Lyapunov equation: the drift matrix A and the diffusion matrix D can be rescaled while preserving both the OU model and the induced covariance matrix Σ . Whereas existing approaches eliminate this freedom by fixing the scale, we exploit it. The rescaling is therefore not merely a nuisance but a structural feature that is utilised in our sign identifiability results.

Theorem 4.2 ($\mathcal{M}^0$ Criterion). *Let $e \in E$ be an edge in*

graph $G = (V, E)$ and let $\Sigma \in \mathcal{M}_{G,e}^p$. Then

$$\Sigma \in \mathcal{M}_{G,e}^0 \iff e \text{ is non-identifiable for } \Sigma. \quad (11)$$

Proof. In the proof, we always refer to the same graph $G = (V, E)$. For brevity, we will suppress the subscript G and only write $\mathcal{M}_e^k$. There are eight possible combinations of edge signature set memberships for a given Σ :

1. $\mathcal{M}_e^0 \cap \neg \mathcal{M}_e^+ \cap \neg \mathcal{M}_e^-$, 5. $\neg \mathcal{M}_e^0 \cap \neg \mathcal{M}_e^+ \cap \neg \mathcal{M}_e^-$,
2. $\mathcal{M}_e^0 \cap \neg \mathcal{M}_e^+ \cap \mathcal{M}_e^-$, 6. $\neg \mathcal{M}_e^0 \cap \neg \mathcal{M}_e^+ \cap \mathcal{M}_e^-$,
3. $\mathcal{M}_e^0 \cap \mathcal{M}_e^+ \cap \neg \mathcal{M}_e^-$, 7. $\neg \mathcal{M}_e^0 \cap \mathcal{M}_e^+ \cap \neg \mathcal{M}_e^-$,
4. $\mathcal{M}_e^0 \cap \mathcal{M}_e^+ \cap \mathcal{M}_e^-$, 8. $\neg \mathcal{M}_e^0 \cap \mathcal{M}_e^+ \cap \mathcal{M}_e^-$,

where we use the shorthand notation $\neg \mathcal{M}_e^k := \{\Sigma \in \mathcal{M}_e^p : \Sigma \notin \mathcal{M}_e^k\}$. Since $\Sigma \in \mathcal{M}_e^p$, we have that $\Sigma \notin \neg \mathcal{M}_e^0 \cap \neg \mathcal{M}_e^+ \cap \neg \mathcal{M}_e^-$ and $\Sigma \notin \mathcal{M}_e^0 \cap \neg \mathcal{M}_e^+ \cap \neg \mathcal{M}_e^-$. Moreover, Eq. (8) in Lemma 4.1 yields $\Sigma \notin \neg \mathcal{M}_e^0 \cap \mathcal{M}_e^+ \cap \mathcal{M}_e^-$. Similarly, Eq. (9), implies $\Sigma \notin \mathcal{M}_e^0 \cap \mathcal{M}_e^+ \cap \neg \mathcal{M}_e^-$, and Eq. (10), implies $\Sigma \notin \mathcal{M}_e^0 \cap \neg \mathcal{M}_e^+ \cap \mathcal{M}_e^-$. This leaves three possible edge signature set combinations for Σ ,

- $\neg \mathcal{M}_e^0 \cap \neg \mathcal{M}_e^+ \cap \mathcal{M}_e^-$ • $\mathcal{M}_e^0 \cap \mathcal{M}_e^+ \cap \mathcal{M}_e^-$
- $\neg \mathcal{M}_e^0 \cap \mathcal{M}_e^+ \cap \neg \mathcal{M}_e^-$

Among the remaining combinations, if $\Sigma \in \mathcal{M}_e^0$, then $\Sigma \in \mathcal{M}_e^+ \cap \mathcal{M}_e^-$. In addition, if $\Sigma \in \neg \mathcal{M}_e^0$, then $\Sigma \notin \mathcal{M}_e^+ \cap \mathcal{M}_e^-$. Therefore, based on the pointwise edge-sign identifiability Definition 2.8, e is non-identifiable for $\Sigma \in \mathcal{M}_e^p$ if and only if $\Sigma \in \mathcal{M}_e^0$. $\square$

Lemma 4.3. *If Definition 2.3 is modified by replacing $D \in PDD_d$ with $D \in PD_d$, then Lemma 4.1 and Theorem 4.2 remain valid.*

Proof sketch. We redefine the edge signature sets by allowing $D \in PD_d$ instead of $D \in PDD_d$ (see Definition D.1). Since Boege et al. [2025] require D to be diagonal, we can no longer use their result. Therefore, we impose $\Sigma \in PD_d$ rather than $\Sigma \in F_G$. The proof of Lemma 4.1 is analogous, except that establishing the existence of a rescaling with $D \in PD_d$ is slightly more involved. With this redefinition and the corresponding extension of Lemma 4.1 to $D \in PD_d$, the proof of Theorem 4.2 proceeds unchanged. The full proof is provided in Appendix D.2.

To verify the $\mathcal{M}^0$ criterion (Theorem 4.2) in practice, it is useful to note that testing membership of an edge signature set can be formulated as a linear optimisation problem. See Remark 5.1 for a more detailed description. In addition, the $\mathcal{M}^0$ criterion is also a powerful theoretical tool, as will be showcased in the proof of Theorem 4.5. Before we can proceed to the theorem, we introduce the following lemma, which is essential to its proof.

Lemma 4.4. *Let $G = (V, E)$, $G' = (V, E')$ and $G'' = (V, E'')$ with $E'' \subset E' \subset E$, if $F_G \neq F_{G'}$ then $F_G \neq F_{G''}$.*

Proof sketch. Observe that the number of marginally independent pairs of nodes can only increase due to edge deletion. Therefore, if a graph G' is a graph with deleted edges with respect to a graph G , then it can only have the same or more marginally independent pairs of nodes. Furthermore, the t -faithful sets F_G and $F_{G'}$ are only equal if they have the same marginally independent pairs of nodes. Using these observations twice, once for G and G' and once for G' and G'' , we obtain Lemma 4.4. For the detailed proof, see Appendix D.3.

Theorem 4.5 (Graphical Criterion). *Without latent variables, let $G = (V, E)$ be a graph with an edge $e \in E$. Define $G' = (V, E \setminus \{e\})$. Then the edge e is identifiable if the entailed marginal independencies by G and G' differ, i.e., if there exist $V_i, V_j \in V$ such that*

$$\text{An}_{G'}(V_i) \cap \text{An}_{G'}(V_j) = \emptyset \neq \text{An}_G(V_i) \cap \text{An}_G(V_j). \quad (12)$$

Proof. Let $G = (V, E)$ be a graph with an edge e and $G' = (V, E \setminus \{e\})$ be a graph such that the entailed marginal independencies of the graphs G and G' are different. Then setting edge $e = 0$, i.e., removing it from graph G , results in graph G' and the edge signature set $\mathcal{M}_{G,e}^0$ (see Definition 2.3). Let $\Sigma \in \mathcal{M}_{G,e}^0$, then by definition of $\mathcal{M}_{G,e}^0$, we have that $\Sigma \in F_G$ and Σ satisfies the Lyapunov equation for some drift matrix A . Since this drift matrix A used to generate $\mathcal{M}_{G,e}^0$ only requires $\text{supp}(A) \subset E$, it follows that $\text{supp}(A) \subseteq E'$. Starting with $\text{supp}(A) = E'$, we know that the marginal independencies are not the same. Therefore, if Σ is generated via $\text{supp}(A) = E'$ then $\Sigma \notin F_{G'}$. Using Lem. 4.4, we then know that for all graphs $G'' = (V, E'')$ with $E'' \subseteq E'$ we have that $F_G \neq F_{G''}$. Hence, for any Σ generated via $\text{supp}(A) \subseteq E'$ we have that $\Sigma \notin F_{G'}$. Therefore, by the definition of $\mathcal{M}_{G,e}^0$ requiring $\Sigma \in F_G$, we have that $\mathcal{M}_{G,e}^0 = \emptyset$. Hence, using the $\mathcal{M}^0$ -criterion (Theorem 4.2), we have that for all $\Sigma \in F_G$, edge e in G is identifiable. Since $\mathcal{M}_{G,e}^p \subseteq F_G$, we have that e is identifiable in graph G and the proof is complete. $\square$

Remark 4.2. If a graph G contains latent variables, the covariance matrix can be written as $\Sigma = [\Sigma_{hh}, \Sigma_{ho}, \Sigma_{oh}, \Sigma_{oo}]$, where o denotes observable and h denotes hidden. In this case, the observed block Σ_{oo} constrains only part of the full covariance matrix. The observed covariance induces the set $\Sigma_{\text{set}} := \{\Sigma' \in \mathcal{M}_{G,e}^p; \Sigma'_{oo} = \Sigma_{oo}\}$ consisting of all covariance matrices compatible with G that agree on the observable block. It may then occur, even if all $\Sigma' \in \Sigma_{\text{set}}$ are identifiable, that $\Sigma_{\text{set}} \cap \mathcal{M}_{G,e}^+ \neq \emptyset$ and $\Sigma_{\text{set}} \cap \mathcal{M}_{G,e}^- \neq \emptyset$. Therefore, in that scenario Σ_{set} is not restricted to a single sign. In the sense of Definition 2.6, this implies non-identifiability. Hence, Σ_{oo} alone, even together with the observation that each $\Sigma \in \mathcal{M}_{G,e}^p$ is identifiable, is

insufficient to conclude that the edge e is sign identifiable. For example, we refer to the proof in Appendix D.5.1.

To show the applicability of the graphical criterion (Theorem 4.5), we provide two examples of general applications of the theorem.

Example 4.3 (Directed Tree). Any directed tree graph $G = (V, E)$ has the property that while G is connected, removing any edge $e \in E$ makes the graph disconnected. Therefore, the deletion of any edge e , will change the marginal independences. Hence, using the graphical criterion, we have that all edges are identifiable. For examples of directed trees, see Figures 1a and 1b.

Example 4.4 (General Instrumental Variable). Let $G = (V, E)$ be a graph with distinct nodes $i, j, k \in V$. For any edge $i \rightarrow j \in E$, if there exists a node k that is an ancestor of i and it has only one directed path to j (through i), then removing edge $i \rightarrow j$ will create a new marginal independence. Hence, by the graphical criterion, we have that the edge $i \rightarrow j$ is identifiable. Note that k functions similarly to an instrumental variable, but in contrast to an instrumental variable (IV), it can have a distance of more than one edge to the node i . For standard IV examples, see Figure 1e and 1f.

4.2 CLASSICAL AND NOVEL GRAPH STRUCTURES

In this section, we study edge-sign identifiability for specific causal graphs. These graphs are analogous to those common in the acyclic SCM literature (e.g., instrumental variable and confounding settings), as well as novel graphs that allow cycles beyond self-loops. The graphs we study are shown in Fig. 1. Each variable has a self-loop, but this has been suppressed in the figures for readability. We are always interested in the sign of the (red) edge α (also indicated in the figures). For the graphs in Figs. 1a–1f, we provide theoretical guarantees on whether the sign of the red edge α is identifiable. The last three graphs, Fig. 1g–1i, are studied numerically in the next section.

Note that in the case of the instrumental variable (IV) and the (one) proxy variables, the edge of interest corresponds to the common edge of interest in the literature. In Section 4.2.2 (the latent variable case), we consider the variable H in the graphs of Fig. 1 to be hidden.

4.2.1 Without Latent Variables

In Theorem 4.6, we characterize the edge-sign identifiability of α for the graphs in Fig. 1 (Subfigures 1a–1f) under t -faithfulness. Using the Lyapunov equation (Eq. (2)) and the $\mathcal{M}^0$ -criterion (Theorem 4.2), we derive algebraic constraints that characterise the sign identifiability in terms of covariance matrices $\Sigma \in \mathcal{M}_{G,\alpha}^p$. Specifically, this analysis separates the three identifiability categories (see Theo-

rem 4.6), specifies when a partially identifiable α becomes identifiable (see Lemma 4.7), and yields an explicit formula for $\text{sign}(\alpha)$ as a function of Σ (see Lemma 4.8) for some causal graphs. The drawback of using the $\mathcal{M}^0$ -criterion is that it can be algebraically involved to show whether $\Sigma \in \mathcal{M}_{G,\alpha}^0$.

Theorem 4.9 establishes identifiability of α for the same graphs via a purely graphical argument. In contrast to Theorem 4.6, it neither distinguishes between non- and partial identifiability nor characterises the conditions in terms of $\Sigma \in \mathcal{M}_{G,\alpha}^p$. However, its proof follows directly from the graphical criterion (Theorem 4.5) and is therefore substantially simpler than the proof of Theorem 4.6.

Theorem 4.6 (Edge-Sign Identifiability without Latent Variables). *For the red edge α in the graphs in Figs. 1a up to 1f under the t -faithfulness assumption, the sign of α is:*

- *partially identifiable for 1c and 1d,*
- *identifiable for 1a, 1b, 1e and 1f.*

Proof sketch. For each graph, we use the Lyapunov equation (Eq. (2)) to obtain a system of equations in the unknown drift and diffusion parameters. Imposing $D \in PDD_d$ and $\Sigma \in F_G$, for the graphs 1a, 1b, 1e and 1f, we obtain an explicit dependence of $\text{sign}(\alpha)$ on entries of Σ , which yields sign identifiability. For the remaining graphs, we set $\alpha = 0$ and determine for which covariance matrices Σ this leads to a contradiction. Such Σ 's cannot lie in $\mathcal{M}_{G,\alpha}^0$. By the $\mathcal{M}^0$ -criterion (Theorem 4.2), this separates covariance matrices for which the sign is identifiable from those for which both signs remain compatible. Applying this analysis yields partial identifiability for both 1c and 1d.

Remark 4.5. For the graph in Fig. 1c, we show in the proof in Appendix D.4.3, that partial identifiability holds with positive measure. Showing that partial identifiability constitutes a genuine edge-sign identifiability category.

Remark 4.6. The (red) edge α in the graph in Fig. 1c is an example where assuming a fixed diffusion matrix leads to different results. According to Theorem 4.6, the (red) edge α is partially identifiable. In contrast, when considering a fixed diffusion matrix C , then the same (red) edge α is identifiable [Dettling et al., 2023, Theorem 5.3].

The following two lemmas are obtained in the course of the proof of Theorem 4.6.

Lemma 4.7 (Conditions for Partial Sign Identifiability). *For the red edge α in the graphs shown in Figs. 1c and 1d, the edge α becomes identifiable under additional conditions on the covariance matrix $\Sigma \in \mathcal{M}_{G,\alpha}^p$. These conditions are stated in Appendix B.*

Lemma 4.8 (Sign Expressions for Sign Identifiable Edges). *For the sign identifiable red edge α from the graphs shown in Figs. 1a, 1b, 1e, and 1f, the sign of α can be expressed in terms of $\Sigma \in \mathcal{M}_{G,\alpha}^p$ as follows:*

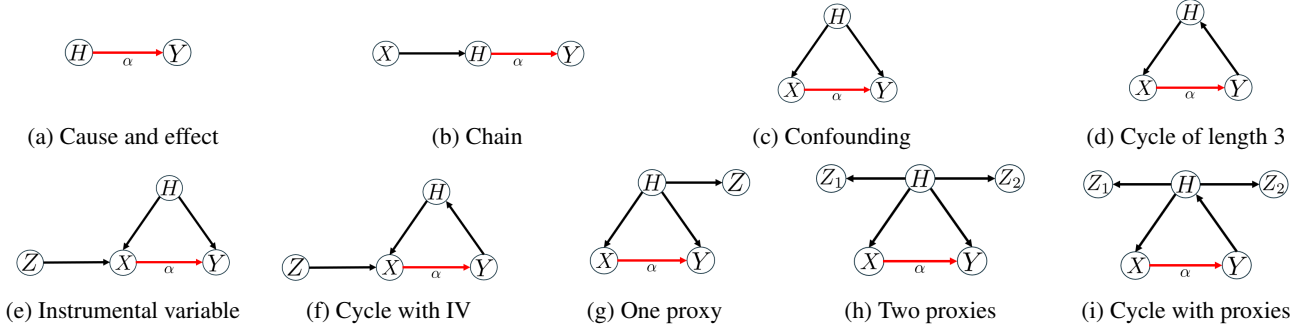


Figure 1: Nine considered causal structures. The red edge α under consideration is indicated by α . Figures (a)-(g) are discussed in both Section 4.2 and Section 5. The node H is considered observable and latent in Section 4.2.1 and Section 4.2.2, respectively. Figures (h) and (i) are only studied numerically in Section 5.

- for 1a: $\text{sign}(\alpha) = \text{sign}(\sigma_{hy})$,
 - for 1b: $\text{sign}(\alpha) = \text{sign}(\sigma_{hy}) / \text{sign}(\sigma_{hx})$,
 - for 1e: $\text{sign}(\alpha) = \text{sign}(\sigma_{zy}) / \text{sign}(\sigma_{zx})$.
 - for 1f:
- $$\text{sign}(\alpha) = \begin{cases} \text{sign}(\sigma_{zy}/\sigma_{zx}) & , \text{ if } \rho_{zy}\rho_{xy}/\rho_{zx} < 1, \\ -\text{sign}(\sigma_{zy}/\sigma_{zx}) & , \text{ if } \rho_{zy}\rho_{xy}/\rho_{zx} > 1. \end{cases}$$

Theorem 4.9 (Graphical Edge-Sign Identifiability). *For the red edge α in the graphs in Figs. 1a, 1b, 1e and 1f, the sign of α is identifiable.*

Proof. Let the graphs shown in Figs. 1a, 1b, 1e and 1f be the original graphs G_0, G_1, G_2 and G_3 and let G'_0, G'_1, G'_2 and G'_3 denote the corresponding graphs obtained by removing the red edge α . Comparing G_0 with G'_0 and G_1 with G'_1 , we see that H and Y no longer share a common ancestor in G'_0 and G'_1 , respectively. Comparing G_2 with G'_2 and G_3 with G'_3 , we see that Z and Y no longer share a common ancestor in G'_2 and G'_3 , respectively. Hence, the marginal independences implied by G and G' differ in each case, and the graphical criterion (Theorem 4.5) yields identifiability of α for all four graphs. $\square$

4.2.2 Latent Variables

Theorem 4.10. *For the red edge α in the graphs in Fig. 1a, 1c, 1e and 1f (with H being latent) under the t -faithfulness assumption, the sign of α is:*

- non-identifiable for 1a and 1c,
- identifiable for 1e and 1f.

Proof sketch. We build on the proofs from the no latent case (Theorem 4.6). The only change is that H is now latent, so covariance entries involving H are unobserved. We therefore treat the corresponding blocks (e.g., $\Sigma_{ho}, \Sigma_{oh}, \Sigma_{hh}$) as free variables that can vary subject to $\Sigma \in \mathcal{M}_{G,\alpha}^p$. This

additional freedom allows us to choose latent-dependent covariance entries so that some scenarios that were (partially) identifiable in the fully observed case become non-identifiable. Note that ensuring $\Sigma \in \mathcal{M}_{G,e}^p$ is one of the main challenges in this proof.

Remark 4.7. In particular, among the sign expressions in Lemma 4.8, only for the (cycle with) instrumental variable case (Fig. 1e and 1f), we still have sign-identifiability in the latent variable setting.

5 NUMERICAL RESULTS

This section reports numerical results on sign-identifiability of the red edge α for the graphs in Fig. 1 (no latent variables). The results are summarized in Table 1. Each entry is the empirical fraction in $[0, 1]$ of identifiable instances of α across 1000 independently generated samples produced by the algorithm described below. If the fraction equals 1 (resp. 0), α is identifiable (resp. non-identifiable). If the fraction lies in $(0, 1)$, α is partially identifiable.³

5.1 METHOD

Throughout this section, we repeatedly use that for $D \in PDD_d \subset PD_d$ in the Lyapunov equation (Eq. (2)), $\Sigma \in PD_d$ if and only if A is Hurwitz stable [Frommer and Hashemi, 2012].

For a fixed graph $G = (V, E)$, samples are generated as follows. We define a symbolic drift matrix A_{sym} , diffusion matrix D_{sym} and covariance matrix Σ , where $\text{supp}(A_{sym}) = E$, D_{sym} is diagonal and Σ_{sym} respects the marginal independences of the graph. We first draw a Hurwitz stable drift matrix A with $\text{supp}(A) = E$ and a diagonal matrix $D \in PDD_d$. Each non-zero entry from (A_{sym}, D_{sym}) is

³For the implementation, see the following link: repository. The repository includes a guiding example to test any desired custom causal structure.

Graph in Fig.	1a	1b	1c	1d	1e	1f	1g	1h	1i ,
Edge-sign identifiable	1.0	1.0	0.44	0.64	1.0	1.0	0.85	1.0	1.0 ,

Table 1: Empirical fraction in $[0, 1]$ of sampled covariance matrices $\Sigma \in \mathcal{M}_{G,\alpha}^p$ for which the red edge α is sign identifiable (graphs in Fig. 1, no latent variables). For a discussion on these numerical results, see Section 5.2.

sampled uniformly from a bounded interval (e.g., $A_{ij} \sim U(-10, 10)$), restricting to the appropriate domain when the sign is known (e.g., $D_{ii} \sim U(0, 10)$). Since (A, D) can be rescaled by any $a \in \mathbb{R}_+$ without changing Σ , this sampling effectively explores the parameter space. We resample A until it is Hurwitz stable. Given (A, D) , we solve the Lyapunov equation to obtain $\Sigma \in PD_d$. We then verify if Σ respects the marginal independences by comparison with the zero pattern in Σ_{sym} , if this fails we reject the sample. If it passes, $\Sigma \in \bigcup_{e \in E} \mathcal{M}_{G,e}^p$.

For a fixed edge $e \in E$, we then test sign-identifiability by searching for (A', D') with $D' \in PDD_d$ and $\text{sign}(A'_e) = -\text{sign}(A_e)$ that satisfies the Lyapunov equation with Σ . This feasibility problem is solved numerically (see remark below). Since $\Sigma \in PD_d$ and $D' \in PDD_d$, any feasible A' is necessarily Hurwitz stable. If such (A', D') exists, e is declared non-identifiable; otherwise it is identifiable (Definition 2.8). This procedure is repeated for 1000 independent samples. See Appendix C for the pseudocode.

Remark 5.1. Without latent variables, the Lyapunov equation induces a linear system in the unknowns (A', D') for fixed Σ . Hence, testing feasibility reduces to a linear optimisation (or feasibility) problem, which admits sound and complete polynomial-time algorithms.⁴ Consequently, if a Hurwitz stable matrix A can be sampled in polynomial time, the overall procedure runs in polynomial time.

In contrast, with latent variables, the Lyapunov constraints become bilinear: the unobserved covariance entries $\sigma_{hh} \in \Sigma$ interact multiplicatively with drift coefficients A_{ij} . The resulting feasibility problem is bilinear and therefore NP-hard [Petrik and Zilberstein, 2011]. For this reason, we restrict our experiments to graphs without latent variables.

5.2 EDGE-SIGN IDENTIFIABILITY

Table 1 reports the sign identifiability results for the red edge α in Fig. 1. The second row shows the empirical fraction of samples in which the edge is identifiable. The numerical results yield three observations. First, the empirical fractions are fully consistent with Theorem 4.6: sign-identifiable graphs have fraction 1, non-identifiable graphs have fraction 0, and partially identifiable graphs yield fractions in $(0, 1)$. Second, for the graphs in Figs. 1g–1i (not analysed in Section 4.2.1), the sign appears identifiable in Figs. 1h

and 1i, and partially identifiable in Fig. 1g. Third, in the partially identifiable category, both outcomes occur with substantial frequency (identifiable fractions 0.44, 0.64, and 0.85 for Figs. 1c, 1d, and 1g), indicating the need to verify sign identifiability for the specific covariance matrix under consideration for those structures.

6 CONCLUSION

We studied identifiability in continuous linear stationary SDEs given the causal graph, where we relaxed the assumption of knowing the diffusion matrix C . In this setup, the linear SDE is scale invariant (when C is not fixed), therefore, we aimed to identify the sign of a given edge. We introduced edge-sign identifiability and characterised three notions of sign identifiability, namely, identifiable, non-identifiable, and partially identifiable. Of which, partial identifiability forms a new category in the literature. Using these categories, we derived general criteria characterising when the sign of an edge can be determined from the observational covariance matrix Σ , given the causal graph. We illustrated the applicability of our results on classical structures, including an instrumental variable setting, for which we obtained an explicit sign in terms of Σ . Moreover, we showed that in the confounding setting, partial identifiability has positive measure, thereby showing that partial identifiability is a genuine category. Numerical experiments further indicate that when an edge is partially identifiable, both identifiable and non-identifiable constitute substantial sub-categories. Future directions include extensions to subgraphs and graph-level sign identifiability.

References

- Carlos Améndola, Tobias Boege, Benjamin Hollering, and Pratik Misra. Structural identifiability of graphical continuous lyapunov models. *arXiv preprint arXiv:2510.04985*, 2025.
- Fabian Bleile, Sarah Lumpp, and Mathias Drton. Efficient learning of stationary diffusions with stein-type discrepancies. *arXiv preprint arXiv:2601.16597*, 2026.
- Tobias Boege, Mathias Drton, Benjamin Hollering, Sarah Lumpp, Pratik Misra, and Daniela Schkoda. Conditional independence in stationary distributions of diffusions. *Stochastic Processes and their Applications*, 184:104604, 2025.

⁴To see how the graph size in terms of the number of nodes scales, see Appendix C.1 for a numerical benchmark.

- Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M Mooij. Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5):2885–2915, 2021.
- Martina Cinquini, Isacco Beretta, Salvatore Ruggieri, and Isabel Valera. A practical approach to causal inference over time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14832–14839, 2025.
- Philipp Dettling, Roser Homs, Carlos Améndola, Mathias Drton, and Niels Richard Hansen. Identifiability in continuous lyapunov models. *SIAM Journal on Matrix Analysis and Applications*, 44(4):1799–1821, 2023.
- Philipp Dettling, Mathias Drton, and Mladen Kolar. On the lasso for graphical continuous lyapunov models. In *Causal Learning and Reasoning*, pages 514–550. PMLR, 2024.
- J. L. Doob. The brownian movement and stochastic equations. *Annals of Mathematics*, 43(2):351–369, 1942. ISSN 0003486X, 19398980. URL <http://www.jstor.org/stable/1968873>.
- Katherine Fitch. Learning directed graphical models from gaussian data. *arXiv preprint arXiv:1906.08050*, 2019.
- Andreas Frommer and Behnam Hashemi. Verified stability analysis using the lyapunov matrix equation. *matrix*, 10:20, 2012.
- James D Hamilton. *Time series analysis*. Princeton university press, 2020.
- Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, England, 2 edition, October 2012.
- Quinn Lanners, Cynthia Rudin, Alexander Volfovsky, and Harsh Parikh. Data fusion for partial identification of causal effects. *Advances in neural information processing systems*, 38:150638–150702, 2026.
- Yifei Liu, Kai Huang, and Wanze Chen. Resolving cellular dynamics using single-cell temporal transcriptomics. *Current Opinion in Biotechnology*, 85:103060, 2024.
- Lars Lorch, Andreas Krause, and Bernhard Schölkopf. Causal modeling with stationary diffusions. In *International Conference on Artificial Intelligence and Statistics*, pages 1927–1935. PMLR, 2024.
- Charles F Manski. Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323, 1990.
- Georg Manten, Cecilia Casolo, Emilio Ferrucci, Søren Wengel Mogensen, Cristopher Salvi, and Niki Kilbertus. Signature kernel conditional independence tests in causal discovery for stochastic processes. *arXiv preprint arXiv:2402.18477*, 2024.
- Daniel Marbach, James C Costello, Robert Küffner, Nicole M Vega, Robert J Prill, Diogo M Camacho, Kyle R Allison, Manolis Kellis, James J Collins, et al. Wisdom of crowds for robust gene network inference. *Nature methods*, 9(8):796–804, 2012.
- Søren Wengel Mogensen, Daniel Malinsky, and Niels Richard Hansen. Causal learning for partially observed stochastic dynamical systems. In *UAI*, pages 350–360, 2018.
- Judea Pearl. *Causality*. Cambridge University Press, Cambridge, England, September 2009.
- Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT press, 2017.
- Marek Petrik and Shlomo Zilberstein. Robust approximate bilinear programming for value function approximation. *Journal of Machine Learning Research*, 12(92):3027–3063, 2011. URL <http://jmlr.org/papers/v12/petrik11a.html>.
- Cecilie Olesen Recke, Sarah Lump, Nataliia Kushnerchuk, Janike Oldekop, Jiayi Li, Jane Ivy Coons, and Elina Robeva. Identifiability in graphical discrete lyapunov models. *arXiv preprint arXiv:2601.21818*, 2026.
- Alexander Sokol and Niels Richard Hansen. Causal interpretation of stochastic differential equations. *Electron. J. Probab*, 19(100):1–24, 2014.
- Marcell Stippinger, Attila Bencze, Ádám Zlatniczki, Zoltán Somogyvári, and András Telcs. Causal discovery of stochastic dynamical systems: a markov chain approach. *Mathematics*, 11(4):852, 2023.
- Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2019.
- Gherardo Varando and Niels Richard Hansen. Graphical continuous lyapunov models. In *Conference on Uncertainty in Artificial Intelligence*, pages 989–998. Pmlr, 2020.
- Bernt Øksendal. *Stochastic Differential Equations*. Springer Berlin Heidelberg, 2003. ISBN 9783642143946. doi: 10.1007/978-3-642-14394-6. URL <http://dx.doi.org/10.1007/978-3-642-14394-6>.

Sign Identifiability of Causal Effects in Stationary Stochastic Dynamical Systems, (Supplementary Material)

Gijs van Seeventer¹

Saber Salehkaleybar¹

¹LIACS, Leiden University, Netherlands

A CLARIFICATION POINTWISE EDGE-SIGN IDENTIFIABILITY

For a fixed $\Sigma \in \mathcal{M}_{G,e}^p$, Definition 2.6 amounts to intersecting the corresponding sets with the singleton $\{\Sigma\}$. Hence,

- non-identifiable if $\mathcal{M}_{G,e}^+ \cap \{\Sigma\} = \{\Sigma\} = \mathcal{M}_{G,e}^- \cap \{\Sigma\}$ while $\mathcal{M}_{G,e}^- \cap \{\Sigma\} \neq \emptyset$,
- partially identifiable if $\mathcal{M}_{G,e}^+ \cap \{\Sigma\} \neq \mathcal{M}_{G,e}^- \cap \{\Sigma\}$ and $\mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^- \cap \{\Sigma\} \neq \emptyset$,
- identifiable if $\mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^- \cap \{\Sigma\} = \emptyset$ while $\mathcal{M}_{G,e}^- \cap \{\Sigma\} \neq \emptyset$ or $\mathcal{M}_{G,e}^+ \cap \{\Sigma\} \neq \emptyset$.

Therefore, non-identifiability reduces to $\Sigma \in \mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^-$ and identifiability reduces to $\Sigma \notin \mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^-$. Since $\mathcal{M}_{G,e}^+ \cap \{\Sigma\} \neq \mathcal{M}_{G,e}^- \cap \{\Sigma\}$ implies that $\Sigma \notin \mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^-$, whereas $\mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^- \cap \{\Sigma\} \neq \emptyset$ implies that $\Sigma \in \mathcal{M}_{G,e}^+ \cap \mathcal{M}_{G,e}^-$, we obtain a contradiction. Hence, partial identifiability cannot be meaningfully defined for a fixed Σ . Summarising this gives us the pointwise edge-sign identifiability Definition 2.8.

B COVARIANCE CONDITIONS LEMMA 4.7

For the sign identifiable edges α from the graphs G shown in Fig. 1c and 1d with the matching covariance matrices $\Sigma \in \mathcal{M}_\alpha^p$, α is identifiable when

- for 1c if one of the following conditions holds,

$$\begin{aligned}
 (c.1) \quad & \frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2)(\rho_{hx}\rho_{hy} - \rho_{xy})}{2\rho_{hx}\rho_{hy}(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1)} \geq 1, \\
 (c.2) \quad & \text{sign}(\sigma_{hx}\sigma_{hy}) \neq \text{sign}(\sigma_{xy}), \\
 (c.3) \quad & \frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} \geq 1.
 \end{aligned} \tag{13}$$

- for 1d if one of the following conditions hold

$$\begin{aligned}
 (c.1) \quad & \text{If } d > 0, a < 0 \text{ and } b < 0, \text{ and } (-a + c)/b \leq 1, \\
 (c.2) \quad & \text{If } d > 0, a > 0 \text{ and } b > 0, \text{ and } (-a + c)/b \leq 1, \\
 (c.3) \quad & \text{If } d < 0, a < 0 \text{ and } b > 0, \text{ and } (-a + c)/b \geq 1, \\
 (c.4) \quad & \text{If } d < 0, a > 0 \text{ and } b < 0, \text{ and } (-a + c)/b \geq 1, \\
 (c.5) \quad & d = 1,
 \end{aligned} \tag{14}$$

where

$$\begin{aligned}
a &:= \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} (\rho_{hx} - \rho_{hy}\rho_{xy}), \\
b &:= \frac{\rho_{xy}\rho_{hx}}{\rho_{xy} - \rho_{hy}\rho_{hx}} \left(\rho_{hx} - \frac{\rho_{hy}}{\rho_{xy}} \right), \\
c &:= \frac{\rho_{hy}}{\rho_{xy}} + \rho_{xy}\rho_{hy}, \\
d &:= \frac{\rho_{hx}\rho_{hy}}{\rho_{xy}}.
\end{aligned} \tag{15}$$

C ALGORITHM FOR NUMERICAL EXPERIMENTS

Algorithm 1 Determining Sign Identifiability

Require: Symbolic: drift matrix A_{sym} , diffusion matrix D_{sym} , (t -faithful) covariance matrix Σ_{sym} and edge e

```

1:  $N_{samples} \leftarrow 1000$ 
2: Identifiable  $\leftarrow 0$ 
3: for  $i \leftarrow 1$  to  $N_{samples}$  do
4:    $A \in \text{Hurwitz stable}, D \in PDD \leftarrow \text{drawParam}(A_{sym}, D_{sym})$ 
5:    $\Sigma \in PD \leftarrow \text{calcSigma}(A, D, \Sigma_{sym})$ 
6:   if not  $\Sigma \in F_G$  then
7:      $N_{samples} - = 1$ 
8:   else if not  $\text{oppSol}(e, A, \Sigma, A_{sym}, D_{sym})$  then
9:     Identifiable  $+ = 1$ 
10:  end if
11: end for
12: Fraction  $\leftarrow \text{Identifiable} / N_{samples}$ 
13: return  $-1$ 

```

C.1 NUMERICAL BENCHMARK OF RUNTIME FINDING AN OPPOSITE SOLUTION

To indicate the scalability of linear programming (LP) for determining the feasibility of a sign of a desired edge for a given causal structure G and covariance matrix Σ we ran an experiment. Note that this corresponds to the function $\text{oppSol}(e, A, \Sigma, A_{sym}, D_{sym})$ in Algorithm 1. In this experiment, we determined the runtime as a function of the number of nodes for different average degrees. For each number of nodes n and average degree d , we drew a sample of 100 random Erdős-Rényi graphs $G(n, m)$ with $m = nd$ edges. Per random graph in the sample, we tracked the wall-clock runtime of checking the feasibility of a sign (for a given causal structure G and covariance matrix Σ). For each sample, we calculated the mean and the standard deviation in the runtime. The results of the experiment are shown in Table 2.

D PROOFS

D.1 LEMMA 4.1

Proof. Throughout the proof, let $G = (V, E)$ be a graph and $e \in E$ a fixed edge. To start, let $\Sigma \in \mathcal{M}_e^+ \cap \mathcal{M}_e^-$. Then we obtain the following two solutions to the Lyapunov equation (Eq. (2))

$$\begin{aligned}
aA^+\Sigma + \Sigma aA^{+T} &= -aD^+, \\
bA^-\Sigma + \Sigma bA^{-T} &= -bD^-,
\end{aligned} \tag{16}$$

Table 2: Runtime as a function of the number of nodes for different average degrees. Values are wall-clock runtime in seconds, reported as mean $\pm$ standard deviation.

Nodes	Avg. degree 2	Avg. degree 3	Avg. degree 4
10	$(2.41 \pm 1.74) \times 10^{-3}$	$(2.55 \pm 0.183) \times 10^{-3}$	$(3.12 \pm 0.226) \times 10^{-3}$
25	$(5.52 \pm 0.569) \times 10^{-3}$	$(7.06 \pm 0.740) \times 10^{-3}$	$(9.00 \pm 0.683) \times 10^{-3}$
50	$(1.41 \pm 0.264) \times 10^{-2}$	$(2.16 \pm 0.323) \times 10^{-2}$	$(2.91 \pm 0.397) \times 10^{-2}$
100	$(4.20 \pm 0.0496) \times 10^{-2}$	$(7.43 \pm 1.76) \times 10^{-2}$	$(1.09 \pm 0.176) \times 10^{-1}$
250	$(2.68 \pm 0.0463) \times 10^{-1}$	$(3.14 \pm 0.757) \times 10^{-1}$	$(7.25 \pm 2.69) \times 10^{-1}$
500	1.18 ± 0.0388	1.17 ± 0.00915	2.10 ± 1.18
1,000	5.34 ± 0.248	4.86 ± 0.0989	5.25 ± 0.190
2,500	37.3 ± 2.37	34.2 ± 3.47	33.8 ± 0.746
5,000	159 ± 14.9	162 ± 20.5	144 ± 22.5

with $a, b \in \mathbb{R}$, A^k a Hurwitz stable matrix, and $D^k \in PDD_d$ where $k \in \{+, -\}$. The sum of the two solutions is again a valid equation. Hence,

$$\begin{aligned}
aA^+\Sigma + \Sigma aA^{+T} + bA^-\Sigma + \Sigma bA^{-T} &= -aD^+ - bD^-, \\
(aA^+ + bA^-\Sigma + \Sigma(aA^+ + bA^-)^T &= -aD^+ - bD^-, \\
A\Sigma + \Sigma A^T &= -D,
\end{aligned} \tag{17}$$

where $A = aA^+ + bA^-$ and $D = aD^+ + bD^-$. For any $x \in \mathbb{R}^d$ where $x \neq 0$, with $a = t$ and $b = 1 - t$, $t \in [0, 1]$, we have:

$$\begin{aligned}
x^T(tD^+ + (1-t)D^-)x &> 0, \\
tx^TD^+x + (1-t)x^TD^-x &> 0.
\end{aligned} \tag{18}$$

Since $D^+, D^- \in PDD_d$ this is true for any choice of $t \in [0, 1]$. Therefore, $D \in PDD_d$. In addition, due to $A_{ij} = 0$ for any $(j, i) \notin E$, $\text{supp}(tA^+ + (1-t)A^-) \subseteq E$. Furthermore, we can pick some $t \in (0, 1)$ such that $tA_e^+ + (1-t)A_e^- = 0$. Then, $\text{sign}(A_e) = \text{sign}(tA_e^+ + (1-t)A_e^-) = 0$. Finally, $\Sigma \in \mathcal{M}_e^+ \cap \mathcal{M}_e^- \subseteq F_G$. Therefore, according to Definition 2.3, we have $\Sigma \in \mathcal{M}_e^0$. To summarise,

$$\Sigma \in \mathcal{M}_e^+ \text{ and } \Sigma \in \mathcal{M}_e^- \implies \Sigma \in \mathcal{M}_e^0. \tag{19}$$

Furthermore, let $\Sigma \in \mathcal{M}_e^+ \cap \mathcal{M}_e^0$. Then, we obtain the following two solutions to the Lyapunov equation (Eq. (2))

$$\begin{aligned}
aA^+\Sigma + \Sigma aA^{+T} &= -aD^+, \\
bA^0\Sigma + \Sigma bA^{0T} &= -bD^0,
\end{aligned} \tag{20}$$

with $a, b \in \mathbb{R}$, A^k a Hurwitz stable matrix and $D^k \in PD_d$ where $k \in \{+, 0\}$. The sum of the two solutions is again a valid equation. Hence,

$$\begin{aligned}
aA^+\Sigma + \Sigma aA^{+T} + bA^0\Sigma + \Sigma bA^{0T} &= -aD^+ - bD^0, \\
(aA^+ + bA^0)\Sigma + \Sigma(aA^+ + bA^0)^T &= -aD^+ - bD^0, \\
A\Sigma + \Sigma A^T &= -D,
\end{aligned} \tag{21}$$

where $A = aA^+ + bA^0$ and $D = aD^+ + bD^0$. If we pick $a = -1$ to get

$$\begin{aligned}
x^T(-D^+ + bD^0)x &> 0, \\
x_i(-D_{ii}^+ + bD_{ii}^0)x_i &\stackrel{(a)}{>} 0, \\
(-D_{ii}^+ + bD_{ii}^0)x_i^2 &> 0, \\
(-D_{ii}^+ + bD_{ii}^0) &> 0, \\
bD_{ii}^0 &> D_{ii}^+, \\
b &> \frac{D_{ii}^+}{D_{ii}^0},
\end{aligned} \tag{22}$$

where (a) use that D is diagonal. If $b > \max_i (D_{ii}^+ / D_{ii}^0)$, then $D \in PDD_d$. Since b is unbounded, we can pick b such that $D \in PDD_d$. Moreover, since $A_{ij} = -A_{ij}^+ + bA_{ij}^0$ and b is still unbounded from above, we can choose b sufficiently large such that $\text{supp}(-A^+ + bA^0) = E$. Furthermore, $\text{sign}(A_e) = \text{sign}(-A_e^+ + bA_e^0) = \text{sign}(-A_e^+) = -\text{sign}(A_e^+) = -$ for any choice of b . Finally, $\Sigma \in \mathcal{M}_e^+ \cap \mathcal{M}_e^0 \subseteq F_G$. Therefore, according to Definition 2.3, if we pick b such that $D \in PDD_d$ and $\text{supp}(-A^+ + bA^0) = E$, then $\Sigma \in \mathcal{M}_e^-$. To summarise

$$\Sigma \in \mathcal{M}_e^+ \text{ and } \Sigma \in \mathcal{M}_e^0 \implies \Sigma \in \mathcal{M}_e^-. \quad (23)$$

Finally, we can analogously show,

$$\Sigma \in \mathcal{M}_e^- \text{ and } \Sigma \in \mathcal{M}_e^0 \implies \Sigma \in \mathcal{M}_e^+. \quad (24)$$

□

D.2 LEMMA 4.3

Proof. Throughout the proof, let $G = (V, E)$ be a graph and $e \in E$ a fixed edge. We define the edge signature sets for $D \in PD_d$ as follows:

Definition D.1 (PD Edge Signature Set). For a graph $G = (V, E)$ and edge $e \in E$, we define the edge signature set $\mathcal{M}_{G,e}^k$ as

$$\begin{aligned} \mathcal{M}_{G,e}^k := & \{ \Sigma \in PD_d : \exists A, D \text{ s.t. } A\Sigma + \Sigma A^T = -D; \\ & D \in PD_d; \text{supp}(A) = E; \text{sign}(A_e) = k \}, \end{aligned} \quad (25)$$

where $k \in \{+, -\}$ and

$$\begin{aligned} \mathcal{M}_{G,e}^0 := & \{ \Sigma \in PD_d : \exists A, D \text{ s.t. } A\Sigma + \Sigma A^T = -D; \\ & D \in PD_d; \text{supp}(A) \subset E; \text{sign}(A_e) = 0 \}. \end{aligned} \quad (26)$$

Since we do not know of any known constraints on the (marginal) independences in the literature for a graph G with $D \in PD$, we no longer enforce t -faithfulness F_G .

To start, let $\Sigma \in \mathcal{M}_e^+ \cap \mathcal{M}_e^-$. Then, we obtain the following two solutions to the Lyapunov equation (Eq. (2)):

$$\begin{aligned} aA^+\Sigma + \Sigma aA^{+T} &= -aD^+, \\ bA^-\Sigma + \Sigma bA^{-T} &= -bD^-, \end{aligned} \quad (27)$$

with $a, b \in \mathbb{R}$, A^k a Hurwitz stable matrix, and $D^k \in PD_d$ where $k \in \{+, -\}$. The sum of the two solutions is again a valid equation. Hence,

$$\begin{aligned} aA^+\Sigma + \Sigma aA^{+T} + bA^-\Sigma + \Sigma bA^{-T} &= -aD^+ - bD^-, \\ (aA^+ + bA^-)\Sigma + \Sigma(aA^+ + bA^-)^T &= -aD^+ - bD^-, \\ A\Sigma + \Sigma A^T &= -D, \end{aligned} \quad (28)$$

where $A = aA^+ + bA^-$ and $D = aD^+ + bD^-$. We pick $a = t$ and $b = 1 - t$ with $t \in [0, 1]$. For any $x \in \mathbb{R}^d$ where $x \neq 0$, we have that

$$\begin{aligned} x^T(tD^+ + (1-t)D^-)x &> 0, \\ tx^TD^+x + (1-t)x^TD^-x &> 0. \end{aligned} \quad (29)$$

Since $D^+, D^- \in PD_d$ this is true for any choice of $t \in [0, 1]$, therefore $D \in PD_d$. In addition, due to $A_{ij} = 0$ for any $(j, i) \notin E$, $\text{supp}(tA^+ + (1-t)A^-) \subseteq E$. Furthermore, we can pick t such that $tA_e^+ + (1-t)A_e^- = 0$. Then, $\text{sign}(A_e) = \text{sign}(tA_e^+ + (1-t)A_e^-) = 0$. In addition, $\Sigma \in \mathcal{M}_e^+ \cap \mathcal{M}_e^- \subseteq PD_d$. Therefore, according to Definition D.1, we have: $\Sigma \in \mathcal{M}_e^0$. To summarise,

$$\Sigma \in \mathcal{M}_e^+ \text{ and } \Sigma \in \mathcal{M}_e^- \implies \Sigma \in \mathcal{M}_e^0. \quad (30)$$

Furthermore, let $\Sigma \in \mathcal{M}_e^+ \cap \mathcal{M}_e^0$. Then, we obtain the following two solutions to the Lyapunov equation (Eq. (2))

$$\begin{aligned} aA^+\Sigma + \Sigma aA^{+T} &= -aD^+, \\ bA^0\Sigma + \Sigma bA^{0T} &= -bD^0, \end{aligned} \quad (31)$$

with $a, b \in \mathbb{R}$, A^k a Hurwitz stable matrix and $D^k \in PD_d$ where $k \in \{+, 0\}$. The sum of the two solutions is again a valid equation. Hence,

$$\begin{aligned} aA^+\Sigma + \Sigma aA^{+T} + bA^0\Sigma + \Sigma bA^{0T} &= -aD^+ - bD^0, \\ (aA^+ + bA^0)\Sigma + \Sigma(aA^+ + bA^0)^T &= -aD^+ - bD^0, \\ A\Sigma + \Sigma A^T &= -D, \end{aligned} \quad (32)$$

where $A = aA^+ + bA^0$ and $D = aD^+ + bD^0$. Let $a = -1$. For any $x \in \mathbb{R}^d$ where $x \neq 0$, we have that

$$\begin{aligned} x^T(-D^+ + bD^0)x &> 0, \\ bx^TD^0x &> x^TD^+x, \\ b &> \frac{x^TD^+x}{x^TD^0x}, \\ b &> \frac{x^TD^+x}{x^TD^0x} \frac{x^Tx}{x^Tx}, \\ b &>^{(a)} \frac{R(D^+, x)}{R(D^0, x)}, \\ b &>^{(b)} \frac{\lambda_{max}^+}{\lambda_{min}^0}, \end{aligned} \quad (33)$$

where (a) use the definition of the Rayleigh quotient R , (b) use that $\lambda_{max}^+/\lambda_{min}^0 \geq R(D^+, x)/R(D^0, x)$, and λ^k are the eigenvalues of D^k . Since $D^k \in PD_d$ we have that $\lambda_{min}^k > 0$. In addition, as b is unbounded, we can pick a value of b that satisfies the inequality. Therefore, we can always pick b such that $D \in PD_d$. Moreover, since $A_{ij} = -A_{ij}^+ + bA_{ij}^0$ and b is still unbounded from above, we can also pick b sufficiently large to get $\text{supp}(-A^+ + bA^0) = E$. Furthermore, $\text{sign}(A_e) = \text{sign}(-A_e^+ + bA_e^0) = \text{sign}(-A_e^+) = -\text{sign}(A_e^+) = -$ for any choice of b . Finally, $\Sigma \in \mathcal{M}_e^+ \cap \mathcal{M}_e^0 \subseteq PD_d$. Therefore, according to Definition D.1, if we pick b such that $D \in PD_d$ and $\text{supp}(-A^+ + bA^0) = E$, then $\Sigma \in \mathcal{M}_e^-$. To summarise

$$\Sigma \in \mathcal{M}_e^+ \text{ and } \Sigma \in \mathcal{M}_e^0 \implies \Sigma \in \mathcal{M}_e^-. \quad (34)$$

Moreover, we can analogously show,

$$\Sigma \in \mathcal{M}_e^- \text{ and } \Sigma \in \mathcal{M}_e^0 \implies \Sigma \in \mathcal{M}_e^+. \quad (35)$$

We have now proven that Lemma 4.1 can be extended to $D \in PD_d$.

The proof for the $\mathcal{M}^0$ criterion Theorem 4.2 follows from the same arguments when using the adjusted Definition D.1 and extend form of Lemma 4.1 in the original proof (see below Theorem 4.2). □

D.3 LEMMA 4.4

Proof. Let $G = (V, E)$, $G' = (V, E')$ and $G'' = (V, E'')$ with $E'' \subset E' \subset E$, i.e. G' is obtained by deleting edges from G , and G'' by deleting edges from G' . Denote MI_H as the set of pairs of marginally independent nodes of a graph H . The set of pairs of marginally independent nodes MI_H can only increase or stay the same due to the deletion of edges, since this deletion can only remove ancestors and not add new ones. Therefore, $MI_G \subseteq MI_{G'} \subseteq MI_{G''}$. In addition, note that for two graphs H and H' to have the same t -faithful set F_H , they need to have $MI_H = MI_{H'}$. Therefore, if $F_G \neq F_{G'}$, then $MI_G \subset MI_{G'} \subseteq MI_{G''}$. Hence, $MI_G \subset MI_{G''}$ and we get the desired result $F_G \neq F_{G''}$. □

D.4 THEOREM 4.6

In the proofs, we solve the equations that result from comparing the matrix entries from the left and right hand side in the Lyapunov equation (Eq. (2)). We use that matrices on the left and right hand side are symmetric, such that a $d \times d$ matrix results in $a = d(d-1)/2$ equations. To facilitate reading in and comparisons between the proofs, we use the convention of giving the sets of equations consistent numbers (i) to the Roman number of equations a in each proof.

Furthermore, we use the property of triangular matrices that their eigenvalues are on the diagonal. For a triangular Hurwitz stable drift matrix $A \in \mathbb{R}^{d \times d}$, this therefore means that the diagonals all have to be negative.

In addition, we use that any matrix $B \in PD_d$ has positive diagonal entries, i.e., $B_{ii} > 0$. Therefore the diagonal matrix has strict positive diagonals, i.e., $D = D_{ii} > 0$. In addition, we will always assume that the covariance matrix $\Sigma \in \mathcal{M}_{G,\alpha}^p$ such that Σ is t -faithful. Therefore $\Sigma_{ii} > 0$. Another property of covariance matrices without exact linear dependencies between random variables X_i and X_j , as is the case in our OU process Eq. (1), is that $|\Sigma_{ij}| < \sqrt{\Sigma_{ii}\Sigma_{jj}}$. For the off-diagonals in the covariance matrix Σ we will therefore use the notation that $\Sigma_{ij} = \rho_{ij} \sqrt{\Sigma_{ii}\Sigma_{jj}}$ with $\rho_{ij} \in (-1, 1)$ being the correlation coefficient.

Finally, we can write the covariance matrix as $\Sigma = K R K$, where K is diagonal with $K_{ii} = \sqrt{\sigma_{ii}} > 0$ such that $K \in PDD_d$, and

$$R_{ij} = \begin{cases} 1 & \text{if } i = j, \\ \rho_{ij} & \text{if } i \neq j. \end{cases} \quad (36)$$

This means that if and only if $R \in PD_d$, then $\Sigma \in PD_d$. Therefore, if $\Sigma \in PD_3$ we have

$$R = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}, \quad (37)$$

and, by Sylvester's criterion [Horn and Johnson, 2012], $R \in PD_d$ if and only if $1 - \rho_{12}^2 > 0$ and $1 + 2\rho_{12}\rho_{13}\rho_{23} - (\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2) > 0$. Since $|\rho_{ij}| < 1$, the first condition is always satisfied, we only need to ensure that

$$1 + 2\rho_{12}\rho_{13}\rho_{23} - (\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2) > 0, \quad (38)$$

to show that $R \in PD_3$ and thus $\Sigma \in PD_3$.

D.4.1 Cause and Effect

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1a. The nodes $V = \{H, Y\}$ correspond to the stationary SDE process $X = (H, Y)^T$, then the Hurwitz stable drift matrix A respecting the causal structure of graph G is

$$A = \begin{bmatrix} s_h & 0 \\ \alpha & s_y \end{bmatrix}, \quad (39)$$

the diagonal diffusion matrix is

$$D = \begin{bmatrix} d_h & 0 \\ 0 & d_y \end{bmatrix} \in PDD_2, \quad (40)$$

and the covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_{hh} & \sigma_{hy} \\ \sigma_{hy} & \sigma_{yy} \end{bmatrix} \in \mathcal{M}_{G,\alpha}^p. \quad (41)$$

The resulting set of equations to solve is

$$\begin{aligned} (i) \quad & -d_h = 2s_h\sigma_{hh}, \\ (ii) \quad & 0 = \alpha\sigma_{hh} + s_h\sigma_{hy} + s_y\sigma_{hy}, \\ (iii) \quad & -d_y = 2\alpha\sigma_{hy} + 2s_y\sigma_{yy}. \end{aligned} \quad (42)$$

Eq. (ii) is satisfied if and only if $\alpha = b_1\sigma_{hy}$, where $b_1 = -\frac{s_y+s_h}{\sigma_{hh}}$. Since $s_y, s_h < 0$ and $\sigma_{hh} > 0$, we have that $b_1 > 0$ and thus $\text{sign}(\alpha) = \text{sign}(b_1\sigma_{hy}) = \text{sign}(\sigma_{hy})$. Since $\Sigma \in \mathcal{M}_{G,\alpha}^p$, we have $\sigma_{hy} \neq 0$ meaning that the sign of α is $+$ or $-$.

Therefore there exists no $\Sigma \in \mathcal{M}_{G,\alpha}^0$, such that by virtue of the $\mathcal{M}^0$ criterion Theorem 4.2, for any $\Sigma \in \mathcal{M}_{G,\alpha}^p$, the sign of edge α in graph G is identifiable. $\square$

D.4.2 Chain

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1b. The nodes $V = \{H, X, Y\}$ correspond to the SDE process $X = (H, X, Y)^T$, then the Hurwitz stable drift matrix A respecting the causal structure of graph G is

$$A = \begin{bmatrix} s_x & 0 & 0, \\ \beta & s_h & 0, \\ 0 & \alpha & s_y \end{bmatrix}, \quad (43)$$

the diagonal diffusion matrix is

$$D = \begin{bmatrix} d_h & 0 & 0, \\ 0 & d_x & 0, \\ 0 & 0 & d_y \end{bmatrix} \in PDD_3, \quad (44)$$

and the t -faithful covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_{hh} & \sigma_{hx} & \sigma_{hy}, \\ \sigma_{hx} & \sigma_{xx} & \sigma_{xy}, \\ \sigma_{hy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \in \mathcal{M}_{G,\alpha}^p. \quad (45)$$

The resulting set of equations to solve is

$$\begin{aligned} (i) \quad & -d_h = 2s_h\sigma_{hh}, \\ (ii) \quad & 0 = \beta\sigma_{hh} + s_x\sigma_{hx} + s_h\sigma_{hx}, \\ (iii) \quad & -d_x = 2\beta\sigma_{hx} + 2s_x\sigma_{xx}, \\ (iv) \quad & 0 = \alpha\sigma_{hx} + s_y\sigma_{hy} + s_h\sigma_{hy}, \\ (v) \quad & 0 = \alpha\sigma_{xx} + \beta\sigma_{hy} + s_x\sigma_{xy} + s_y\sigma_{xy}, \\ (vi) \quad & -d_y = 2\alpha\sigma_{xy} + 2s_y\sigma_{yy}. \end{aligned} \quad (46)$$

Analogous to the proof in D.4.1, Eq. (iv) is satisfied if and only if $\alpha = b_1\sigma_{hy}/\sigma_{hx}$, where $b_1 = -(s_y + s_h)$. Since $s_y, s_h < 0$, we have that $b_1 > 0$. Therefore, $\text{sign}(\alpha) = \text{sign}(b_1\sigma_{hy})/\text{sign}(\sigma_{hx}) = \text{sign}(\sigma_{hy})/\text{sign}(\sigma_{hx})$. Since $\Sigma \in \mathcal{M}_{G,\alpha}^p$, we have $\sigma_{hy} \neq 0$ meaning that the sign of α is $+$ or $-$. Therefore there exists no $\Sigma \in \mathcal{M}_{G,\alpha}^0$. According to the $\mathcal{M}^0$ criterion Theorem 4.2, the sign of edge α in graph G is identifiable. $\square$

D.4.3 Confounding

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1c. The nodes $V = \{H, X, Y\}$ correspond to the SDE process $X = (H, X, Y)^T$, then the Hurwitz stable drift matrix A respecting the causal structure of graph G is

$$A = \begin{bmatrix} s_h & 0 & 0, \\ \gamma & s_x & 0, \\ \delta & \alpha & s_y \end{bmatrix}, \quad (47)$$

the diagonal diffusion matrix is

$$D = \begin{bmatrix} d_h & 0 & 0, \\ 0 & d_x & 0, \\ 0 & 0 & d_y \end{bmatrix} \in PDD_3, \quad (48)$$

and the t -faithful covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_{hh} & \sigma_{hx} & \sigma_{hy}, \\ \sigma_{hx} & \sigma_{xx} & \sigma_{xy}, \\ \sigma_{hy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \in \mathcal{M}_{G,\alpha}^p. \quad (49)$$

In the numerical Section 5.2 we find examples of $\Sigma, \Sigma' \in \mathcal{M}_{G,\alpha}^p$ where Σ is identifiable, and Σ' is non-identifiable. In other words, we show there exist covariance matrices in both $\mathcal{M}_{G,\alpha}^\pm$ and only in either $\mathcal{M}_{G,\alpha}^+$ or $\mathcal{M}_{G,\alpha}^-$. Hence, by Definition 2.6, the sign of edge α for graph G is partially identifiable.

To exclude that these examples are some degenerate cases, we show that the set of covariance matrices Σ yielding identifiability (respectively, non-identifiability) is not a measure-zero subset in $\mathcal{M}_{G,\alpha}^p$. To that end, we first characterize $\Sigma \in \mathcal{M}_{G,\alpha}^p$. This means that we want to show that there exists a Hurwitz drift matrix A with $\text{supp}(A) = E$ and a diagonal diffusion matrix $D \in PDD_3$ such that the Lyapunov equation (Eq. (2)) is satisfied.

From Dettling et al. [2023] Corollary 5.4, we know that for the set

$$\mathcal{M}_{G,D} := \{\Sigma \in PD_d : \exists A \text{ such that } A\Sigma + \Sigma A^T = -D; \text{supp}(A) \subseteq E\}, \quad (50)$$

we have $\mathcal{M}_{G,D} = PD_d$ for the graph G and any given $D \in PD_d$. Comparing $\mathcal{M}_{G,D}$ with the edge signature set (Definition 2.3) and the resulting possibility set (Definition 2.5), we observe that they are structurally similar. If the result $\mathcal{M}_{G,D} = PD_d$ could be strengthened to require $\text{supp}(A) = E$, then, since we allow any $D \in PDD_d$ and $F_G \subset PD_d$, it would follow that

$$\mathcal{M}_{G,\alpha}^p = F_G. \quad (51)$$

To show this, we start from the set of equations resulting from the Lyapunov equation.

$$\begin{aligned} (i) \quad & -d_h = 2s_h\sigma_{hh}, \\ (ii) \quad & 0 = \gamma\sigma_{hx} + (s_h + s_x)\sigma_{hx}, \\ (iii) \quad & -d_x = 2\gamma\sigma_{hx} + 2s_x\sigma_{xx}, \\ (iv) \quad & 0 = \alpha\sigma_{hx} + \delta\sigma_{hh} + (s_h + s_y)\sigma_{hy}, \\ (v) \quad & 0 = \alpha\sigma_{xx} + \delta\sigma_{hx} + \gamma\sigma_{hy} + (s_x + s_y)\sigma_{xy}, \\ (vi) \quad & -d_y = 2\alpha\sigma_{xy} + 2\delta\sigma_{hy} + 2s_y\sigma_{yy}. \end{aligned} \quad (52)$$

Since $d_h, \sigma_{hh} > 0$ and $s_h < 0$, we have that Eq. (i) is always satisfied. Eq. (ii) is satisfied if and only if, $\gamma = -(s_h + s_x)\sigma_{hx}/\sigma_{hh}$. Eq. (iii) is satisfied in and only if,

$$\begin{aligned} & 2\frac{-(s_h + s_x)}{\sigma_{hh}}\sigma_{hx}^2 + 2s_x\sigma_{xx} = -d_x, \\ \iff & 2\frac{-(s_h + s_x)}{\sigma_{hh}}\sigma_{hx}^2 + 2s_x\sigma_{xx} \stackrel{(a)}{<} 0, \\ & \frac{-(s_h + s_x)}{\sigma_{hh}}\sigma_{hx}^2 + s_x\sigma_{xx} < 0, \\ & \frac{-(s_h + s_x)}{\sigma_{hh}}\rho_{hx}^2\sigma_{hh}\sigma_{xx} + s_x\sigma_{xx} \stackrel{(b)}{<} 0, \\ & -(s_h + s_x)\rho_{hx}^2\sigma_{xx} + s_x\sigma_{xx} < 0, \\ & -(s_h + s_x)\rho_{hx}^2 + s_x < 0, \\ & s_x(1 - \rho_{hx}^2) < s_h\rho_{hx}^2, \\ & s_x < \frac{s_h\rho_{hx}^2}{1 - \rho_{hx}^2} = -\frac{s_h\rho_{hx}^2}{\rho_{hx}^2 - 1}, \end{aligned} \quad (53)$$

where (a) we use that $d_x > 0$ and (b) we substitute $\sigma_{hx}^2 = \rho_{hx}^2\sigma_{hh}\sigma_{xx}$. Since $s_x, s_h < 0$ and $\rho_{hx}^2 < 1$, we have $1 - \rho_{hx}^2 > 0$, and hence $\rho_{hx}^2/(\rho_{hx}^2 - 1) > 0$. Therefore, the inequality demands that

$$\left| -\frac{s_h\rho_{hx}^2}{\rho_{hx}^2 - 1} \right| < |s_x|, \quad (54)$$

and we can rewrite it as

$$s_x = -b_1 s_h \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}, \quad b_1 > 1. \quad (55)$$

Eq. (iv) is satisfied if and only if

$$\alpha = \frac{-\delta\sigma_{hh} - (s_h + s_y)\sigma_{hy}}{\sigma_{hx}}. \quad (56)$$

Eq. (v) is satisfied if and only if

$$\begin{aligned} -\delta\sigma_{hx} &= \alpha\sigma_{xx} + \gamma\sigma_{hy} + (s_x + s_y)\sigma_{xy}, \\ -\delta\sigma_{hx} &\stackrel{(a)}{=} \frac{-\delta\sigma_{hh} - (s_h + s_y)\sigma_{hy}}{\sigma_{hx}}\sigma_{xx} + \gamma\sigma_{hy} + (-b_1s_h\frac{\rho_{hx}^2}{\rho_{hx}^2 - 1} + s_y)\sigma_{xy}, \\ \delta(\sigma_{hh}\sigma_{xx}/\sigma_{hx} - \sigma_{hx}) &= \frac{-(s_h + s_y)\sigma_{hy}}{\sigma_{hx}}\sigma_{xx} + \gamma\sigma_{hy} + (-b_1s_h\frac{\rho_{hx}^2}{\rho_{hx}^2 - 1} + s_y)\sigma_{xy}, \\ \delta &= \left(\frac{-(s_h + s_y)\sigma_{hy}}{\sigma_{hx}}\sigma_{xx} + \gamma\sigma_{hy} + (-b_1s_h\frac{\rho_{hx}^2}{\rho_{hx}^2 - 1} + s_y)\sigma_{xy}\right) / (\sigma_{hh}\sigma_{xx}/\sigma_{hx} - \sigma_{hx}), \\ \delta &= \left(- (s_h + s_y)\sigma_{hy}\sigma_{xx} + \gamma\sigma_{hy}\sigma_{hx} + (-b_1s_h\frac{\rho_{hx}^2}{\rho_{hx}^2 - 1} + s_y)\sigma_{xy}\sigma_{hx}\right) / (\sigma_{hh}\sigma_{xx} - \sigma_{hx}^2), \end{aligned} \quad (57)$$

where (a) we substituted the values for α and s_x .

Since $d_y > 0$, Eq. (vi) is satisfied if and only if

$$\begin{aligned} 0 &> 2\alpha\sigma_{xy} + 2\delta\sigma_{hy} + 2s_y\sigma_{yy}, \\ 0 &> \alpha\sigma_{xy} + \delta\sigma_{hy} + s_y\sigma_{yy}, \\ 0 &\stackrel{(a)}{>} \frac{-\delta\sigma_{hh} - (s_h + s_y)\sigma_{hy}}{\sigma_{hx}}\sigma_{xy} + \delta\sigma_{hy} + s_y\sigma_{yy}, \\ 0 &> \delta\left(\sigma_{hy} - \frac{\sigma_{hh}\sigma_{xy}}{\sigma_{hx}}\sigma_{xy}\right) - (s_h + s_y)\frac{\sigma_{xy}\sigma_{hy}}{\sigma_{hx}} + s_y\sigma_{yy}, \\ 0 &\stackrel{(b)}{>} \left(- (s_h + s_y)\sigma_{hy}\sigma_{xx} + \gamma\sigma_{hy}\sigma_{hx} + (-b_1s_h\frac{\rho_{hx}^2}{\rho_{hx}^2 - 1} + s_y)\sigma_{xy}\sigma_{hx}\right) \left(\sigma_{hy} - \frac{\sigma_{hh}\sigma_{xy}}{\sigma_{hx}}\right) / (\sigma_{hh}\sigma_{xx} - \sigma_{hx}^2) \\ &\quad - (s_h + s_y)\sigma_{hy} + s_y\sigma_{yy}, \\ 0 &\stackrel{(c)}{>} f(\Sigma, b_1, s_h, s_y). \end{aligned} \quad (58)$$

where (a) we substituted the values for α , (b) we substituted the value of δ and (c) we introduced the function f to denote the lengthy the right hand side.

In addition, we can write the expressions for α , γ and δ in a different way when starting from Eq. (ii, iv) and (v). We begin by substituting s_x in to γ , then

$$\begin{aligned} \gamma &= -(s_h + -b_1s_h\frac{\rho_{hx}^2}{\rho_{hx}^2 - 1})\frac{\sigma_{hx}}{\sigma_{hh}}, \\ &= -(1 + -b_1\frac{\rho_{hx}^2}{\rho_{hx}^2 - 1})s_h\frac{\sigma_{hx}}{\sigma_{hh}}, \\ &\stackrel{(a)}{=} (1 - t\frac{\rho_{hx}^2}{\rho_{hx}^2 - 1})t\frac{\sigma_{hx}}{\sigma_{hh}}, \end{aligned} \quad (59)$$

where (a) we pick $b_1 = t$ and $s_h = -t$, with $t > 1$ satisfying both $s_h < 0$ and $b_1 > 1$, then γ only depends on t and we can write $\gamma(t)$. Furthermore, since $t, \sigma_{hh} > 0$ and $\sigma_{hx} \neq 0$, $\gamma = 0$ if and only if $t = (\rho_{hx}^2 - 1)/\rho_{hx}^2$. With this choice of $b_1 = t$ and $s_h = -t$, s_x also only depends on t , i.e.,

$$s_x = t^2\frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}. \quad (60)$$

We then pick $s_y = -(t + \varepsilon)$, where $\varepsilon > 0$.

Starting from Eq. (52), we can write Eq. (iv) and (v) as a system of linear equations

$$\begin{pmatrix} \sigma_{hh} & \sigma_{hx} \\ \sigma_{hx} & \sigma_{xx} \end{pmatrix} \begin{pmatrix} \delta \\ \alpha \end{pmatrix} = \begin{pmatrix} -(s_h + s_y)\sigma_{hy} \\ -\gamma\sigma_{hy} - (s_x + s_y)\sigma_{xy} \end{pmatrix}. \quad (61)$$

Moreover, Eq. in (61) the coefficient matrix

$$M := \begin{pmatrix} \sigma_{hh} & \sigma_{hx} \\ \sigma_{hx} & \sigma_{xx} \end{pmatrix} \quad (62)$$

does not depend on ε , whereas the right-hand side does. Using our choices of s_h , s_x and s_y we have

$$-(s_h + s_y)\sigma_{hy} = (2t + \varepsilon)\sigma_{hy}, \quad -(s_x + s_y)\sigma_{xy} = \left(-t^2 \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1} + t + \varepsilon\right)\sigma_{xy}. \quad (63)$$

Moreover, for $\gamma(t)$ we obtained

$$\gamma = \left(1 - t \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}\right) t \frac{\sigma_{hx}}{\sigma_{hh}} = \left(t - t^2 \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}\right) \frac{\sigma_{hx}}{\sigma_{hh}}, \quad (64)$$

which is independent of ε . Hence the right-hand side of Eq. (61) equals

$$\begin{aligned} r(\varepsilon) &= \begin{pmatrix} -(s_h + s_y)\sigma_{hy} \\ -\gamma\sigma_{hy} - (s_x + s_y)\sigma_{xy} \end{pmatrix} \\ &= \begin{pmatrix} (2t + \varepsilon)\sigma_{hy} \\ -\left(t - t^2 \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}\right) \frac{\sigma_{hx}}{\sigma_{hh}} \sigma_{hy} + \left(t - t^2 \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1} + \varepsilon\right)\sigma_{xy} \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} 2t\sigma_{hy} \\ \left(t - t^2 \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}\right) \left(\sigma_{xy} - \frac{\sigma_{hx}\sigma_{hy}}{\sigma_{hh}}\right) \end{pmatrix}}_{=: r_0} + \varepsilon \underbrace{\begin{pmatrix} \sigma_{hy} \\ \sigma_{xy} \end{pmatrix}}_{=: r_1}. \end{aligned} \quad (65)$$

Therefore,

$$r(\varepsilon) = r_0 + \varepsilon r_1, \quad (66)$$

for fixed vectors $r_0, r_1 \in \mathbb{R}^2$ (depending on t and Σ but not on ε). Since $\Sigma \succ 0$, M is positive definite. Therefore,

$$\begin{pmatrix} \delta(\varepsilon) \\ \alpha(\varepsilon) \end{pmatrix} = M^{-1}r(\varepsilon) = M^{-1}r_0 + \varepsilon M^{-1}r_1, \quad (67)$$

so both $\delta(\varepsilon)$ and $\alpha(\varepsilon)$ are affine functions of ε .

Now suppose $\delta(\varepsilon)$ were identically zero for all ε . Then both its constant and linear coefficients would be zero, i.e., $e_1^\top M^{-1}r_0 = e_1^\top M^{-1}r_1 = 0$, which would force $(M^{-1}r_1)_1 = 0$. But $r_1 = (\sigma_{hy}, \sigma_{xy})^\top$, and since M is invertible this would imply $(\sigma_{hy}, \sigma_{xy})^\top = 0$, contradicting $\sigma_{hy} \neq 0$ and $\sigma_{xy} \neq 0$ (which hold for $\Sigma \in F_G$ in the confounding graph). Hence $\delta(\varepsilon)$ is a non-constant affine function and can vanish for at most one value of ε . The same argument applies to $\alpha(\varepsilon)$.

Consequently, there are at most two values of ε for which either $\delta(\varepsilon) = 0$ or $\alpha(\varepsilon) = 0$. Choosing any $\varepsilon > 0$ different from these values yields

$$\delta(\varepsilon) \neq 0 \quad \text{and} \quad \alpha(\varepsilon) \neq 0. \quad (68)$$

Therefore, to summarise, if we choose

$$t \neq (\rho_{hx}^2 - 1)/\rho_{hx}^2 \quad (69)$$

and ε different from the two corresponding conflicting values, then $\text{supp}(A) = E$.

Since $f(\Sigma, -t, t, t + \varepsilon)$ is a strict inequality and a continuous function (in fact, a multivariate polynomial) in (b_1, s_h, s_y) , we may choose t and ε at positive distance $\varepsilon' > 0$ from their conflicting values, while still satisfying $f(\Sigma, -t, t, t + \varepsilon) < 0$. Hence, whenever $f(\Sigma, -t, t, t + \varepsilon)$ holds with $\text{supp}(A) \subseteq E$, we can perturb t and ε slightly so that the inequality remains satisfied while ensuring $\text{supp}(A) = E$.

Using the aforementioned result that $\mathcal{M}_{G,D} = PD_d$, it follows that for every $\Sigma \in PD_3$ there exists a feasible choice of parameters satisfying

$$f(\Sigma, -t, t, t + \varepsilon) < 0, \quad (70)$$

with $\text{supp}(A) \subseteq E$.

If the corresponding choice of t and ε yields $\text{supp}(A) \subset E$, we can perturb them to $\bar{t}, \bar{\varepsilon}$ such that $\text{supp}(A) = E$ while preserving the strict inequality, by continuity of f . Consequently, we obtain

$$\mathcal{M}_{G,\alpha}^p = F_G. \quad (71)$$

Next we characterize models in $\mathcal{M}_{G,\alpha}^0$. Note that now, since $\alpha = 0$, we only require $\text{supp}(A) \subset E$. We set $\alpha = 0$, which simplifies Eq. (iv), (v) and (vi) to

$$\begin{aligned} (iv) \quad & \delta\sigma_{hh} + (s_h + s_y)\sigma_{hy} = 0, \\ (v) \quad & \delta\sigma_{xx} + \gamma\sigma_{hy} + (s_x + s_y)\sigma_{xy} = 0, \\ (vi) \quad & 2\delta\sigma_{hy} + 2s_y\sigma_{yy} = -d_y, \end{aligned} \quad (72)$$

The other Eq. (i – iii) remain the same as before, meaning that Eq. (i) is always satisfied, Eq. (ii) is satisfied if and only if $\gamma = -(s_h + s_x)/\sigma_{hh}\sigma_{hx}$ and Eq. (iii) is satisfied if and only if $s_x = -b_1 s_h \rho_{hx}^2 / (\rho_{hx}^2 - 1)$ with $b_1 > 1$.

From the new set of equations with $\alpha = 0$, we see that Eq. (iv) is satisfied if and only if $\delta = b_2 \sigma_{hy}$ where $b_2 = -(s_h + s_y)/\sigma_{hh} > 0$.

Therefore, it remains to satisfy the two Eq. (v) and (vi). In these equations, we substitute $\delta = b_2 \sigma_{hy}$, $\gamma = b_1 \sigma_{hx}$.

Eq. (v) becomes

$$\begin{aligned} & b_2 \sigma_{hy} \sigma_{hx} + b_1 \sigma_{hx} \sigma_{hy} + (s_x + s_y) \sigma_{xy} = 0, \\ & -(s_h + s_y) \rho_{hx} \rho_{hy} \sqrt{\sigma_{xx} \sigma_{yy}} - (s_h + s_x) \rho_{hx} \rho_{hy} \sqrt{\sigma_{xx} \sigma_{yy}} + (s_x + s_y) \rho_{xy} \sqrt{\sigma_{xx} \sigma_{yy}} \stackrel{(a)}{=} 0, \\ & \left[-(2s_h + s_x + s_y) \rho_{hx} \rho_{hy} + (s_x + s_y) \rho_{xy} \right] \sqrt{\sigma_{xx} \sigma_{yy}} = 0, \\ & -(2s_h + s_x + s_y) \rho_{hx} \rho_{hy} + (s_x + s_y) \rho_{xy} = 0, \end{aligned} \quad (73)$$

where (a) $b_2 = -(s_h + s_y)/\sigma_{hh}$, $b_1 = -(s_h + s_x)/\sigma_{hh}$ and $\sigma_{ij} = \rho_{ij} \sqrt{\sigma_{ii} \sigma_{jj}}$. Note that in the final line, due to $-(2s_h + s_x + s_y) > 0$ and $(s_x + s_y) < 0$, $\text{sign}(\rho_{hx} \rho_{hy}) \neq \text{sign}(\rho_{xy})$ leads to a contradiction. Therefore, to be in $\mathcal{M}_{G,\alpha}^0$, we should have $\text{sign}(\sigma_{hx} \sigma_{hy}) = \text{sign}(\sigma_{xy})$.

Eq. (vi) becomes

$$\begin{aligned} & 2b_2 \sigma_{hy}^2 + 2s_y \sigma_{yy} = -d_y, \\ & -(s_h + s_y) \rho_{hy}^2 + s_y \stackrel{(a)}{<} 0, \end{aligned} \quad (74)$$

where (a) uses the same steps as in Eq. (53), $b_2 = -(s_h + s_y)/\sigma_{hh}$ and $\sigma_{ij} = \rho_{ij} \sqrt{\sigma_{ii} \sigma_{jj}}$. Which analogously yields

$$s_y = -b_3 s_h \frac{\rho_{hy}^2}{\rho_{hy}^2 - 1}, \quad (75)$$

with $b_3 > 1$.

Summarising,

s_y and s_x can be expressed as

$$s_y = -b_3 s_h \frac{\rho_{hy}^2}{\rho_{hy}^2 - 1}, \quad \text{and} \quad s_x = -b_1 s_h \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}, \quad (76)$$

with $b_1, b_3 > 1$

Substituting this into Eq. (v) yields

$$\begin{aligned} 0 &= -\left(2s_h - b_3 s_h \frac{\rho_{hy}^2}{\rho_{hy}^2 - 1} - b_1 s_h \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}\right) \rho_{hx} \rho_{hy} + \left(-b_3 s_h \frac{\rho_{hy}^2}{\rho_{hy}^2 - 1} - b_1 s_h \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}\right) \rho_{xy}, \\ 0 &= -2\rho_{hx} \rho_{hy} + \left(b_3 \frac{\rho_{hy}^2}{\rho_{hy}^2 - 1} + b_1 \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}\right) (\rho_{hx} \rho_{hy} - \rho_{xy}), \\ \frac{2\rho_{hx} \rho_{hy}}{\rho_{hx} \rho_{hy} - \rho_{xy}} &= b_3 \frac{\rho_{hy}^2}{\rho_{hy}^2 - 1} + b_1 \frac{\rho_{hx}^2}{\rho_{hx}^2 - 1}. \end{aligned} \quad (77)$$

Since the right hand side is negative ($\rho_{ij}^2 - 1 < 0$), the left hand side must also be negative.

We can solve this equation exactly. For notational simplicity let $a := \rho_{hy}^2/(\rho_{hy}^2 - 1) < 0$, $b := \rho_{hx}^2/(\rho_{hx}^2 - 1) < 0$ and $c := 2\rho_{hx}\rho_{hy}/(\rho_{hx}\rho_{hy} - \rho_{xy}) < 0$. Then from the above we have that

$$\begin{aligned}
c &= b_3a + b_1b, \\
\frac{-b_1b + c}{a} &= b_3, \\
\frac{-b_1b + c}{a} &\stackrel{(a)}{>} 1, \\
-b_1b + c &\stackrel{(b)}{<} a, \\
b_1 &< \frac{-a + c}{b}, \\
1 &\stackrel{(c)}{<} \frac{-a + c}{b},
\end{aligned} \tag{78}$$

where (a) enforces $b_3 > 1$, i.e., satisfying Eq. (vi), (b) since $a < 0$ the inequality $>$ flips to $<$ and (c) is the tightest way to allow a choice $b_1 > 1$, i.e., satisfying Eq. (iii). Substituting the original definitions for a, b and c back the inequality gives

$$\begin{aligned}
1 &< \frac{-\frac{\rho_{hy}^2}{\rho_{hy}^2-1} + \frac{2\rho_{hy}\rho_{hx}}{\rho_{hy}\rho_{hx}-\rho_{xy}}}{\frac{\rho_{hx}^2}{\rho_{hx}^2-1}}, \\
\frac{\rho_{hx}^2}{\rho_{hx}^2-1} &\stackrel{(a)}{>} -\frac{\rho_{hy}^2}{\rho_{hy}^2-1} + \frac{2\rho_{hy}\rho_{hx}}{\rho_{hy}\rho_{hx}-\rho_{xy}}, \\
\frac{\rho_{hy}^2}{\rho_{hy}^2-1} + \frac{\rho_{hx}^2}{\rho_{hx}^2-1} &> \frac{2\rho_{hy}\rho_{hx}}{\rho_{hy}\rho_{hx}-\rho_{xy}}, \\
\frac{\rho_{hy}^2}{\rho_{hy}^2-1} + \frac{\rho_{hx}^2}{\rho_{hx}^2-1} &\stackrel{(b)}{>} \frac{2d\rho_{xy}}{d\rho_{xy}-\rho_{xy}}, \\
\frac{\rho_{hy}^2}{\rho_{hy}^2-1} + \frac{\rho_{hx}^2}{\rho_{hx}^2-1} &> \frac{2d}{d-1},
\end{aligned} \tag{79}$$

where (a) we use that since $\rho_{hx}^2 < 1$, we have that the inequality flips, and (b) we use the notation $\rho_{hy}\rho_{hx} = d\rho_{xy}$. Since $\rho_{hy}^2, \rho_{hx}^2 < 1$, the left-hand side of Eq. (79) is less than zero. Therefore, $2d/(d-1) < 0$ meaning that $0 < d < 1$, i.e., if

$$\frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} \geq 1 \quad \text{or} \quad \frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} \leq 0, \tag{80}$$

then there is a contradiction. Note that $\rho_{xy}/(\rho_{hy}\rho_{hx}) < 0$ was already excluded by the requirement that $\text{sign}(\rho_{hy}\rho_{hx}) = \text{sign}(\rho_{xy})$. Continuing with the derivation of Eq. (79),

$$\begin{aligned}
\frac{\rho_{hy}^2}{\rho_{hy}^2-1} + \frac{\rho_{hx}^2}{\rho_{hx}^2-1} &> \frac{2d}{d-1}, \\
\frac{\rho_{hy}^2(\rho_{hx}^2-1) + \rho_{hx}^2(\rho_{hy}^2-1)}{(\rho_{hy}^2-1)(\rho_{hx}^2-1)} &> \frac{2d}{d-1}, \\
\frac{2\rho_{hy}^2\rho_{hx}^2 - \rho_{hx}^2 - \rho_{hy}^2}{(\rho_{hy}^2-1)(\rho_{hx}^2-1)} &> \frac{2d}{d-1}, \\
\frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hx}^2 - \rho_{hy}^2)(d-1)}{2d(\rho_{hy}^2-1)(\rho_{hx}^2-1)} &\stackrel{(a)}{<} 1, \\
\frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hx}^2 - \rho_{hy}^2)(\rho_{hy}\rho_{hx} - \rho_{xy})}{2\rho_{hy}\rho_{hx}(\rho_{hy}^2-1)(\rho_{hx}^2-1)} &\stackrel{(b)}{<} 1,
\end{aligned} \tag{81}$$

where in (a) we use that since $2d/(d-1) < 0$, the inequality flips, and in (b) we take the reverse of the final two steps of Eq. (79).

Eq. (v) has become a single inequality that enforces Eq. (iii) and (vi) as well. Therefore, satisfying this final inequality is a necessary condition for membership in $\mathcal{M}_{G,\alpha}^0$.

To collect the results, there are three conditions on Σ for membership in $\mathcal{M}_{G,\alpha}^0$:

$$\begin{aligned} (c.1) \quad & \frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2)(\rho_{hx}\rho_{hy} - \rho_{xy})}{2\rho_{hx}\rho_{hy}(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1)} < 1, \\ (c.2) \quad & \text{sign}(\sigma_{hx}\sigma_{hy}) = \text{sign}(\sigma_{xy}), \\ (c.3) \quad & \frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} < 1. \end{aligned} \tag{82}$$

Hence, a covariance matrix Σ belongs to $\mathcal{M}_{G,\alpha}^0$ if and only if it satisfies all the conditions in Eq. (82). By the $\mathcal{M}^0$ -criterion Theorem 4.2, we have that covariance matrices $\Sigma \in \mathcal{M}_{G,\alpha}^p$ violating one of the conditions in Eq. (82) are identifiable.

To summarise what we have shown up to this point, $\mathcal{M}_{G,\alpha}^p = F_G$, and verifying if $\Sigma \in \mathcal{M}_{G,\alpha}^0$ reduces to three conditions. Hence, any $\Sigma \in F_G$ that satisfies the three conditions is in $\mathcal{M}_{G,\alpha}^0$ and is thus non-identifiable. Conversely, any $\Sigma \in F_G$ that violates one of the three conditions is identifiable. We will use this to show that the sets of non-identifiable and identifiable covariance matrices both have non-zero measure. Throughout the rest of the proof, we will use the notation that for a $\rho_{ij} \in (\underline{\rho}_{ij}, \bar{\rho}_{ij})$ we have that $\underline{\rho}_{ij}$ is the infimum and $\bar{\rho}_{ij}$ is the supremum.

To begin, let

$$\Sigma_{id} := \{\Sigma \in F_G : \Sigma \text{ does not satisfy one of the conditions in Eq.(82)}\}. \tag{83}$$

Let Σ_{set} be a set of covariance matrices for which the correlation coefficients are

$$\rho_{hx} \in (0.001, 0.002), \quad \rho_{hy} \in (-0.001, 0), \quad \rho_{xy} \in (0.001, 0.002). \tag{84}$$

Then

$$\begin{aligned} 1 + 2\rho_{hx}\rho_{hy}\rho_{xy} - (\rho_{hx}^2 + \rho_{hy}^2 + \rho_{xy}^2) &\geq 1 + 2\bar{\rho}_{hx}\underline{\rho}_{hy}\bar{\rho}_{xy} - (\bar{\rho}_{hx}^2 + \underline{\rho}_{hy}^2 + \bar{\rho}_{xy}^2), \\ &= 1 + 2 \cdot 0.002 \cdot (-0.001) \cdot 0.002 - (0.002^2 + (-0.001)^2 + 0.002^2), \\ &\approx 1.00 > 0. \end{aligned} \tag{85}$$

Hence, Sylvester's criterion Eq. (38) holds. Therefore, since $|\rho_{ij}| < 1$, we have that $\Sigma \in PD_3$. Moreover, the non-zero correlations ρ_{ij} respect the marginal independences of G , hence $\Sigma_{set} \in F_G$.

Furthermore, since $\text{sign}(\sigma_{ij}) = \text{sign}(\rho_{ij})$, we have

$$\text{sign}(\sigma_{hx}\sigma_{hy}) = - \neq + = \text{sign}(\sigma_{xy}), \tag{86}$$

so Σ_{set} violates Condition (c.2) from Eq. (82). Hence, $\Sigma_{set} \subseteq \Sigma_{id}$.

Finally, the Lebesgue measure satisfies

$$m(\Sigma_{set}) = 0.001^3 > 0. \tag{87}$$

By monotonicity, we conclude

$$m(\Sigma_{id}) \geq m(\Sigma_{set}) > 0, \tag{88}$$

and hence Σ_{id} is not a measure-zero set.

Now let

$$\Sigma_{non} := \{\Sigma \in F_G : \Sigma \text{ satisfies the conditions in Eq.(82)}\}, \tag{89}$$

and let Σ_{set} be a set of covariance matrices with the correlation coefficients

$$\rho_{hx} \in (0.326, 0.338), \quad \rho_{hy} \in (0.29, 0.305), \quad \rho_{xy} \in (0.734, 0.842). \tag{90}$$

Then

$$1 + 2\rho_{hx}\rho_{hy}\rho_{xy} - (\rho_{hx}^2 + \rho_{hy}^2 + \rho_{xy}^2) \geq 1 + 2\underline{\rho}_{hx}\underline{\rho}_{hy}\underline{\rho}_{xy} - (\bar{\rho}_{hx}^2 + \bar{\rho}_{hy}^2 + \bar{\rho}_{xy}^2), \approx 0.22 > 0. \quad (91)$$

Hence, Sylvester's criterion Eq. (38) holds. Therefore, since $|\rho_{ij}| < 1$, we have that $\Sigma \in PD_3$. Moreover, the non-zero correlations ρ_{ij} respect the marginal independences of G , hence $\Sigma_{\text{set}} \in F_G$.

To satisfy the conditions in Eq. (82), we require that in Condition (c.1) the left-hand side is smaller than one on the whole domain, i.e.,

$$\begin{aligned} \max \left(\frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2)(\rho_{hx}\rho_{hy} - \rho_{xy})}{2\rho_{hx}\rho_{hy}(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1)} \right) &< 1, \\ \max \left(\frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2)(d - 1)}{2d(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1)} \right) &< 1, \end{aligned} \quad (92)$$

where we use the same step as in Eq. (79) with the notation that $\rho_{hy}\rho_{hx} = d\rho_{xy}$ and $0 < d < 1$. We will show that the numerator and denominator are always positive. Therefore, to find the maximum, we need to get the numerator as large as possible and the denominator as small as possible.

To begin with the first term in the numerator. Since $\rho_{hy}^2, \rho_{hx}^2 \in (0, 1)$, we have that $\rho_{hy}^2 - 1 < 0$, and $\rho_{hx}^2 - 1 < 0$. Hence,

$$2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2 = \rho_{hy}^2(\rho_{hx}^2 - 1) + \rho_{hx}^2(\rho_{hy}^2 - 1) < 0. \quad (93)$$

Therefore,

$$|2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2| < |2\underline{\rho}_{hy}^2\underline{\rho}_{hx}^2 - \bar{\rho}_{hy}^2 - \bar{\rho}_{hx}^2|. \quad (94)$$

For the second term in the numerator, since $0 < d < 1$, we have that

$$d - 1 < 1. \quad (95)$$

Therefore, both terms in the numerator are negative, making the whole numerator positive. In addition, since $\rho_{hy}\rho_{hx} = d\rho_{xy}$, we have that

$$|d - 1| \leq |\min(d) - 1| = \left| \frac{\rho_{hx}\rho_{hy}}{\bar{\rho}_{xy}} - 1 \right|. \quad (96)$$

To continue with the first term in the denominator. We have that

$$d \geq \min(d) = \frac{\rho_{hx}\rho_{hy}}{\bar{\rho}_{xy}} \approx 0.11 > 0. \quad (97)$$

For the second and third term, since $|\rho_{ij}| < 1$, we have that

$$\rho_{ij}^2 - 1 < 0. \quad (98)$$

Therefore,

$$(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1) > 0. \quad (99)$$

Hence, the denominator is positive. Furthermore, since

$$|\rho_{ij}^2 - 1| \geq |\bar{\rho}_{ij}^2 - 1|, \quad (100)$$

we have that

$$|(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1)| \geq |(\bar{\rho}_{hx}^2 - 1)(\bar{\rho}_{hy}^2 - 1)|. \quad (101)$$

Combining it all together, we have

$$\max \left(\frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2)(d - 1)}{2d(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1)} \right) < \left(\frac{(2\underline{\rho}_{hy}^2\underline{\rho}_{hx}^2 - \bar{\rho}_{hy}^2 - \bar{\rho}_{hx}^2)(\min(d) - 1)}{2\min(d)(\bar{\rho}_{hx}^2 - 1)(\bar{\rho}_{hy}^2 - 1)} \right) \approx 0.93 < 1, \quad (102)$$

hence, satisfying Condition (c.1).

Finally, we verify the other two conditions in Eq. (82). For Condition (c.2), we note that all the correlation coefficients of the covariance matrices in Σ_{set} are positive, and hence Condition (c.2) is satisfied. For Condition (c.3) we note that,

$$\frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} \leq \frac{\bar{\rho}_{hy}\bar{\rho}_{hx}}{\bar{\rho}_{xy}} \approx 0.14 < 1. \quad (103)$$

Meaning that Condition (c.3) is also satisfied.

Thus Σ_{set} satisfies the conditions in Eq. (82), and therefore $\Sigma_{\text{set}} \subseteq \Sigma_{\text{non}}$.

Finally, the Lebesgue measure satisfies

$$m(\Sigma_{\text{set}}) \approx 0.01 \cdot 0.02 \cdot 0.11 > 0. \quad (104)$$

By monotonicity, we conclude

$$m(\Sigma_{\text{non}}) \geq m(\Sigma_{\text{set}}) > 0, \quad (105)$$

and hence Σ_{non} is not a measure-zero set.

Therefore, both the identifiable and non-identifiable covariance matrices form subsets of $\mathcal{M}_{G,\alpha}^p$ with positive Lebesgue measure. Hence, the edge α in graph G is partially identifiable with positive measure. $\square$

D.4.4 Cycle of Length 3

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1d. The nodes $V = \{H, X, Y\}$ correspond to the SDE process $X = (H, X, Y)^T$, then the Hurwitz stable drift matrix A respecting the causal structure of graph G is

$$A = \begin{bmatrix} s_x & 0 & \gamma, \\ \beta & s_h & 0, \\ 0 & \alpha & s_y \end{bmatrix}, \quad (106)$$

the diagonal diffusion matrix is

$$D = \begin{bmatrix} d_h & 0 & 0, \\ 0 & d_x & 0, \\ 0 & 0 & d_y \end{bmatrix} \in PDD_3, \quad (107)$$

and the t -faithful covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_{hh} & \sigma_{hx} & \sigma_{hy}, \\ \sigma_{hx} & \sigma_{xx} & \sigma_{xy}, \\ \sigma_{hy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \in \mathcal{M}_{G,\alpha}^p. \quad (108)$$

In the numerical Section 5.2 we find examples of $\Sigma, \Sigma' \in \mathcal{M}_{G,\alpha}^p$ where Σ is identifiable, and Σ' is non-identifiable. Therefore we show there exist covariance matrices in both $\mathcal{M}_{G,\alpha}^\pm$ and only in either $\mathcal{M}_{G,\alpha}^+$ or $\mathcal{M}_{G,\alpha}^+$ such that the edge α for graph G is partially-identifiable.

For completeness, we want to show when $\Sigma \in \mathcal{M}_{G,\alpha}^p$ is (non-)identifiable. The resulting set of equations to solve is

$$\begin{aligned} (i) \quad & -d_x = 2\gamma\sigma_{xy} + 2s_x\sigma_{xx}, \\ (ii) \quad & 0 = \beta\sigma_{xx} + \gamma\sigma_{hy} + s_h\sigma_{xh} + s_x\sigma_{xh}, \\ (iii) \quad & -d_h = 2\beta\sigma_{xh} + 2s_h\sigma_{hh}, \\ (iv) \quad & 0 = \alpha\sigma_{xh} + \gamma\sigma_{yy} + s_y\sigma_{xy} + s_x\sigma_{xy}, \\ (v) \quad & 0 = \alpha\sigma_{hh} + \beta\sigma_{xy} + s_h\sigma_{hy} + s_y\sigma_{hy}, \\ (vi) \quad & -d_y = 2\alpha\sigma_{hy} + 2s_y\sigma_{yy}. \end{aligned} \quad (109)$$

We note that since the drift matrix A is not triangular, the self loops s_x, s_y and s_h are unconstrained.

In order to characterize models in $\mathcal{M}_{G,\alpha}^0$, we set $\alpha = 0$, which simplifies Eq. (iv), (v) and (vi) to

$$\begin{aligned}(iv) \quad 0 &= \gamma\sigma_{yy} + s_y\sigma_{xy} + s_x\sigma_{xx}, \\(v) \quad 0 &= \beta\sigma_{xy} + s_h\sigma_{hy} + s_y\sigma_{hy}, \\(vi) \quad -d_y &= 2s_y\sigma_{yy}.\end{aligned}\tag{110}$$

Due to $d_y, \sigma_{yy} > 0$, Eq. (vi) is satisfied if and only if $s_y < 0$. Moreover, Eq. (v) is satisfied if and only if

$$\begin{aligned}\beta &= -\frac{(s_h + s_y)\sigma_{hy}}{\sigma_{xy}}, \\&\stackrel{(a)}{=} -(s_h + s_y)\frac{\rho_{hy}}{\rho_{xy}}\sqrt{\frac{\sigma_{hh}\sigma_{yy}}{\sigma_{xx}\sigma_{yy}}}, \\&= -(s_h + s_y)\frac{\rho_{hy}}{\rho_{xy}}\sqrt{\frac{\sigma_{hh}}{\sigma_{xx}}},\end{aligned}\tag{111}$$

and Eq. (iv) is satisfied if and only if

$$\begin{aligned}\gamma &= -\frac{(s_y + s_x)\sigma_{xy}}{\sigma_{yy}}, \\&\stackrel{(a)}{=} -(s_y + s_x)\rho_{xy}\frac{\sqrt{\sigma_{xx}\sigma_{yy}}}{\sigma_{yy}}, \\&= -(s_y + s_x)\rho_{xy}\sqrt{\frac{\sigma_{xx}}{\sigma_{yy}}},\end{aligned}\tag{112}$$

where (a), in both, $\sigma_{ij} = \rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}$.

Since $d_x > 0$ and can be chosen arbitrarily, Eq. (i) is satisfied if and only if

$$\begin{aligned}0 &> \gamma\sigma_{xy} + s_x\sigma_{xx}, \\&\stackrel{(a)}{>} -(s_y + s_x)\rho_{xy}\sqrt{\frac{\sigma_{xx}}{\sigma_{yy}}}\rho_{xy}\sqrt{\sigma_{xx}\sigma_{yy}} + s_x\sigma_{xx}, \\0 &> -(s_y + s_x)\rho_{xy}^2\sigma_{xx} + s_x\sigma_{xx}, \\0 &> -(s_y + s_x)\rho_{xy}^2 + s_x, \\&\rho_{xy}^2 s_y > (1 - \rho_{xy}^2)s_x, \\&\frac{\rho_{xy}^2}{1 - \rho_{xy}^2} s_y \stackrel{(b)}{>} s_x,\end{aligned}\tag{113}$$

where (a) $\rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}$ and we substitute γ and (b) $\rho_{xy}^2 < 1$ such that $1 - \rho_{xy}^2 > 0$ and the division doesn't flip the inequality. In addition, since $1 - \rho_{xy}^2 > 0$ and $s_y < 0$, $\rho_{xy}^2/(1 - \rho_{xy}^2)s_y < 0$. Therefore $s_x < 0$. Hence, let

$$s_x = b_1 \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} s_y,\tag{114}$$

with $b_1 > 1$.

Since $d_h > 0$ and can be chosen arbitrarily, Eq. (iii) is satisfied if and only if

$$\begin{aligned}0 &> \beta\sigma_{xh} + s_h\sigma_{hh}, \\&\stackrel{(a)}{>} -(s_h + s_y)\frac{\rho_{hy}}{\rho_{xy}}\sqrt{\frac{\sigma_{hh}}{\sigma_{xx}}}\rho_{hx}\sqrt{\sigma_{xx}\sigma_{hh}} + s_h\sigma_{hh}, \\0 &> -(s_h + s_y)\frac{\rho_{hy}\rho_{hx}}{\rho_{xy}}\sigma_{hh} + s_h\sigma_{hh}, \\0 &> -(s_h + s_y)\frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} + s_h, \\&\frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} s_y > \left(1 - \frac{\rho_{hy}\rho_{hx}}{\rho_{xy}}\right)s_h,\end{aligned}\tag{115}$$

where (a) $\rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}$ and we substitute β .

Note that $\rho_{hx}\rho_{hy}/\rho_{xy}$ is unconstrained except for the requirement that $\Sigma \in F_g$, which implies $\sigma_{ij} \neq 0$ and hence $\rho_{hx}\rho_{hy}/\rho_{xy} \neq 0$. Let $d := \rho_{hx}\rho_{hy}/\rho_{xy}$. We have four scenarios:

1. If $d < 0$, then

$$\begin{aligned} s_h &\stackrel{(a)}{<} \frac{d}{1-d}s_y, \\ s_h &\stackrel{(b)}{=} b_2 \frac{d}{1-d}s_y \quad \text{with } b_2 < 1 \end{aligned} \tag{116}$$

2. if $0 < d < 1$, then

$$\begin{aligned} s_h &\stackrel{(a)}{<} \frac{d}{1-d}s_y, \\ s_h &\stackrel{(c)}{=} b_2 \frac{d}{1-d}s_y \quad \text{with } b_2 > 1 \end{aligned} \tag{117}$$

3. if $d = 1$, then

$$0 \cdot s_h = 0 < s_y \stackrel{(d)}{<} 0, \tag{118}$$

which is a contradiction.

4. if $d > 1$, then

$$\begin{aligned} s_h &\stackrel{(e)}{>} \frac{d}{1-d}s_y, \\ s_h &\stackrel{(f)}{=} b_2 \frac{d}{1-d}s_y \quad \text{with } b_2 > 1, \end{aligned} \tag{119}$$

where (a) due to $d < 1$, we have that $1 - d > 0$ and the inequality isn't flipped, (b) due to $d \leq 0$, $1 - d > 0$ and $s_y < 0$, we have that $d/(1-d)s_y > 0$ such that s_h being smaller than $d/(1-d)s_y$ requires $b_2 < 1$, (c) due to $d > 0$, $1 - d > 0$ and $s_y < 0$, we have that $d/(1-d)s_y < 0$ such that s_h being smaller than $d/(1-d)s_y$ requires $b_2 > 1$, (d) we use $s_y < 0$, (e) due to $d > 1$, we have that $1 - d < 0$ such that the inequality is flipped and (f) due to $d > 0$, $1 - d < 0$ and $s_y < 0$, we have that $d/(1-d)s_y > 0$ such that s_h being bigger than $d/(1-d)s_y$ requires $b_2 > 1$. Summarised this gives,

$$s_h = b_2 \frac{d}{1-d}s_y, \quad \text{with } \begin{cases} b_2 < 1 & \text{if } d \leq 0, \\ b_2 > 1 & \text{if } d > 0. \end{cases} \tag{120}$$

In addition we can write s_h as

$$\begin{aligned} s_h &= b_2 \frac{d}{1-d}s_y, \\ &\stackrel{(a)}{=} b_2 \frac{\rho_{hx}\rho_{hy}/\rho_{xy}}{1 - \rho_{hx}\rho_{hy}/\rho_{xy}}s_y, \\ &= b_2 \frac{\rho_{hx}\rho_{hy}}{\rho_{xy} - \rho_{hx}\rho_{hy}}s_y, \end{aligned} \tag{121}$$

where (a) substitute $d = \rho_{hx}\rho_{hy}/\rho_{xy}$.

Using all of the above in Eq. (ii), we get

$$\begin{aligned}
0 &= \beta\sigma_{xx} + \gamma\sigma_{hy} + (s_h + s_x)\sigma_{xh}, \\
0 &\stackrel{(a)}{=} -(s_h + s_y)\frac{\rho_{hy}}{\rho_{xy}}\sqrt{\frac{\sigma_{hh}}{\sigma_{xx}}}\sigma_{xx} - (s_y + s_x)\rho_{xy}\sqrt{\frac{\sigma_{xx}}{\sigma_{yy}}}\rho_{hy}\sqrt{\sigma_{hh}\sigma_{yy}} + (s_h + s_x)\rho_{xh}\sqrt{\sigma_{xx}\sigma_{hh}}, \\
0 &= -(s_h + s_y)\frac{\rho_{hy}}{\rho_{xy}}\sqrt{\sigma_{hh}\sigma_{xx}} - (s_y + s_x)\rho_{xy}\rho_{xy}\sqrt{\sigma_{xx}\sigma_{hh}} + (s_h + s_x)\rho_{xh}\sqrt{\sigma_{xx}\sigma_{hh}}, \\
0 &= -(s_h + s_y)\frac{\rho_{hy}}{\rho_{xy}} - (s_y + s_x)\rho_{xy}\rho_{xy} + (s_h + s_x)\rho_{xh}, \\
0 &\stackrel{(b)}{=} -(b_2\frac{\rho_{hx}\rho_{hy}}{\rho_{xy} - \rho_{hx}\rho_{hy}}s_y + s_y)\frac{\rho_{hy}}{\rho_{xy}} - (s_y + b_1\frac{\rho_{xy}^2}{1 - \rho_{xy}^2}s_y)\rho_{xy}\rho_{xy} + (b_2\frac{\rho_{hx}\rho_{hy}}{\rho_{xy} - \rho_{hx}\rho_{hy}}s_y + b_1\frac{\rho_{xy}^2}{1 - \rho_{xy}^2}s_y)\rho_{xh}, \quad (122) \\
0 &= \left[b_1\frac{\rho_{xy}^2}{1 - \rho_{xy}^2}(\rho_{hx} - \rho_{hy}\rho_{xy}) + b_2\frac{\rho_{xy}\rho_{hx}}{\rho_{xy} - \rho_{hy}\rho_{hx}}\left(\rho_{hx} - \frac{\rho_{hy}}{\rho_{xy}}\right) - \left(\frac{\rho_{hy}}{\rho_{xy}} + \rho_{xy}\rho_{hy}\right) \right] s_y, \\
0 &= b_1\frac{\rho_{xy}^2}{1 - \rho_{xy}^2}(\rho_{hx} - \rho_{hy}\rho_{xy}) + b_2\frac{\rho_{xy}\rho_{hx}}{\rho_{xy} - \rho_{hy}\rho_{hx}}\left(\rho_{hx} - \frac{\rho_{hy}}{\rho_{xy}}\right) - \left(\frac{\rho_{hy}}{\rho_{xy}} + \rho_{xy}\rho_{hy}\right), \\
0 &\stackrel{(c)}{=} b_1a + b_2b - c,
\end{aligned}$$

where (a) substitute the expression for β , γ and $\sigma_{ij} = \rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}$, (b) substitute s_x and s_h and (c) define $a := \frac{\rho_{xy}^2}{1 - \rho_{xy}^2}(\rho_{hx} - \rho_{hy}\rho_{xy})$, $b := \frac{\rho_{xy}\rho_{hx}}{\rho_{xy} - \rho_{hy}\rho_{hx}}\left(\rho_{hx} - \frac{\rho_{hy}}{\rho_{xy}}\right)$ and $c := \frac{\rho_{hy}}{\rho_{xy}} + \rho_{xy}\rho_{hy}$.

Note that a , b and c are unconstrained while b_2 depends on the sign of $d = \rho_{hx}\rho_{hy}/\rho_{xy}$. Since the sign of c doesn't matter, this gives us in total 8 outcomes to check. We provide the derivation of two possible outcomes. The other outcomes follow analogously. Let $a, b < 0$, then

$$\begin{aligned}
0 &= b_1a + b_2b - c, \\
-b_1a &= b_2b - c, \\
-a &\stackrel{(a)}{<} b_2b - c, \\
-a + c &< b_2b, \\
\frac{-a + c}{b} &\stackrel{(b)}{>} b_2,
\end{aligned} \quad (123)$$

where (a) $b_1 > 1$ and $a < 0$ such that $-b_1a > -a$ and (b) $b < 0$ such that the inequality flips. If $d < 0$, then $b_2 < 1$, such that b_2 is unconstrained from below and is always possible. If $d > 0$ we have that $b_2 > 1$, such that $(-a + c)/b > b_2 > 1$ if and only if $(-a + c)/b > 1$. Therefore if $a, b < 0$ and $d > 0$ we have a contradiction when $(-a + c)/b \leq 1$.

Listing all the conditions that lead to a contradiction, we obtain:

$$\begin{aligned}
(c.1) \quad &\text{If } d > 0, a < 0 \text{ and } b < 0, \text{ and } (-a + c)/b \leq 1, \\
(c.2) \quad &\text{If } d > 0, a > 0 \text{ and } b > 0, \text{ and } (-a + c)/b \leq 1, \\
(c.3) \quad &\text{If } d < 0, a < 0 \text{ and } b > 0, \text{ and } (-a + c)/b \geq 1, \\
(c.4) \quad &\text{If } d < 0, a > 0 \text{ and } b < 0, \text{ and } (-a + c)/b \geq 1, \\
(c.5) \quad &d = 1.
\end{aligned} \quad (124)$$

Hence, a covariance matrix Σ does not belongs to $\mathcal{M}_{G,\alpha}^0$ if and only if it satisfies one of the conditions in Eq. (124). By the $\mathcal{M}^0$ -criterion Theorem 4.2, we have that $\Sigma \in \mathcal{M}_{G,\alpha}^p$ satisfying one of the conditions in Eq. (124) are identifiable. $\square$

D.4.5 Instrumental Variable (IV)

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1e. The nodes $V = \{H, X, Y\}$ correspond to the SDE process $X = (H, X, Y)^T$, then the Hurwitz stable drift matrix A

respecting the causal structure of graph G is

$$A = \begin{bmatrix} s_z & 0 & 0 & 0, \\ 0 & s_h & 0 & 0, \\ \beta & \gamma & s_x & 0, \\ 0 & \delta & \alpha & s_y \end{bmatrix}, \quad (125)$$

the diagonal drift matrix is

$$D = \begin{bmatrix} d_z & 0 & 0 & 0, \\ 0 & d_h & 0 & 0, \\ 0 & 0 & d_x & 0, \\ 0 & 0 & 0 & d_y \end{bmatrix} \in PDD_4, \quad (126)$$

and the t -faithful covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_{zz} & 0 & \sigma_{zx} & \sigma_{zy}, \\ 0 & \sigma_{hh} & \sigma_{hx} & \sigma_{hy}, \\ \sigma_{zx} & \sigma_{hx} & \sigma_{xx} & \sigma_{xy}, \\ \sigma_{zy} & \sigma_{hy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \in \mathcal{M}_{G,\alpha}^p. \quad (127)$$

The resulting set of equations to solve is

$$\begin{aligned} (i); -d_z &= 2s_z\sigma_{zz}, \\ (ii); -d_h &= 2s_h\sigma_{hh}, \\ (iii); 0 &= \beta\sigma_{zz} + s_x\sigma_{zx} + s_z\sigma_{zx}, \\ (iv); 0 &= \gamma\sigma_{hh} + s_x\sigma_{hx} + s_h\sigma_{hx}, \\ (v); -d_x &= 2\beta\sigma_{zx} + 2\gamma\sigma_{hx} + 2s_x\sigma_{xx}, \\ (vi); 0 &= \alpha\sigma_{zx} + s_y\sigma_{zy} + s_z\sigma_{zy}, \\ (vii); 0 &= \alpha\sigma_{hx} + \delta\sigma_{hh} + s_y\sigma_{hy} + s_h\sigma_{hy}, \\ (viii); 0 &= \alpha\sigma_{xx} + \beta\sigma_{zy} + \delta\sigma_{hx} + \gamma\sigma_{hy} + s_x\sigma_{xy} + s_y\sigma_{xy}, \\ (ix); -d_y &= 2\alpha\sigma_{xy} + 2\delta\sigma_{hy} + 2s_y\sigma_{yy}. \end{aligned} \quad (128)$$

Analogous to the proof in D.4.1 we see that Eq. (vi) is satisfied if and only if $\text{sign}(\alpha) = \text{sign}(\sigma_{zy})/\text{sign}(\sigma_{zx})$. Since $\Sigma \in \mathcal{M}_{G,\alpha}^p$, we have $\sigma_{zy} \neq 0$ and $\sigma_{zx} \neq 0$ meaning that the sign of α is + or -. Therefore there exists no $\Sigma \in \mathcal{M}_{G,\alpha}^0$, such that by virtue of the $\mathcal{M}^0$ criterion Theorem 4.2, for any $\Sigma \in \mathcal{M}_{G,\alpha}^p$, the sign of edge α in graph G is identifiable. $\square$

D.4.6 Cycle with IV

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1f. The nodes $V = \{H, X, Y\}$ correspond to the SDE process $X = (H, X, Y)^T$, then the Hurwitz stable drift matrix A respecting the causal structure of graph G is

$$A = \begin{bmatrix} s_z & 0 & 0 & 0, \\ 0 & s_h & 0 & \delta, \\ \beta & \gamma & s_x & 0, \\ 0 & 0 & \alpha & s_y \end{bmatrix}, \quad (129)$$

the diagonal drift matrix is

$$D = \begin{bmatrix} d_z & 0 & 0 & 0, \\ 0 & d_h & 0 & 0, \\ 0 & 0 & d_x & 0, \\ 0 & 0 & 0 & d_y \end{bmatrix} \in PDD_4, \quad (130)$$

and the t -faithful covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_{zz} & \sigma_{hz} & \sigma_{zx} & \sigma_{zy} \\ \sigma_{hz} & \sigma_{hh} & \sigma_{hx} & \sigma_{hy} \\ \sigma_{zx} & \sigma_{hx} & \sigma_{xx} & \sigma_{xy} \\ \sigma_{zy} & \sigma_{hy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \in \mathcal{M}_{G,\alpha}^p. \quad (131)$$

The resulting set of equations to solve is

$$\begin{aligned} (i); -d_z &= 2s_z\sigma_{zz}, \\ (ii); 0 &= \delta\sigma_{zy} + s_h\sigma_{hz} + s_z\sigma_{hz}, \\ (iii); -d_h &= 2\delta\sigma_{hy} + 2s_h\sigma_{hh}, \\ (iv); 0 &= \beta\sigma_{zz} + \gamma\sigma_{hz} + s_x\sigma_{zx} + s_z\sigma_{zx}, \\ (v); 0 &= \beta\sigma_{hz} + \delta\sigma_{xy} + \gamma\sigma_{hh} + s_h\sigma_{hx} + s_x\sigma_{hx}, \\ (vi); -d_x &= 2\beta\sigma_{zx} + 2\gamma\sigma_{hx} + 2s_x\sigma_{xx}, \\ (vii); 0 &= \alpha\sigma_{zx} + s_y\sigma_{zy} + s_z\sigma_{zy}, \\ (viii); 0 &= \alpha\sigma_{hx} + \delta\sigma_{yy} + s_h\sigma_{hy} + s_y\sigma_{hy}, \\ (ix); 0 &= \alpha\sigma_{xx} + \beta\sigma_{zy} + \gamma\sigma_{hy} + s_x\sigma_{xy} + s_y\sigma_{xy}, \\ (x); -d_y &= 2\alpha\sigma_{xy} + 2s_y\sigma_{yy}. \end{aligned} \quad (132)$$

We note that since the drift matrix A is not triangular, the self loops s_x , s_y and s_h are not constrained from the outset of the proof.

Due to $d_z, \sigma_{zz} > 0$, Eq. (i) is satisfied if and only if $s_z < 0$. In addition, Eq. (vii) is satisfied if and only if

$$\alpha = -(s_y + s_z) \frac{\sigma_{zy}}{\sigma_{zx}}. \quad (133)$$

Since $d_y > 0$, Eq. (x) is satisfied if and only if

$$\begin{aligned} 0 &> 2\alpha\sigma_{xy} + 2s_y\sigma_{yy}, \\ 0 &> \alpha\sigma_{xy} + s_y\sigma_{yy}, \\ 0 &\stackrel{(a)}{>} -(s_y + s_z) \frac{\sigma_{zy}}{\sigma_{zx}} \sigma_{xy} + s_y\sigma_{yy}, \\ 0 &\stackrel{(b)}{>} -(s_y + s_z) \frac{\rho_{zy}\sqrt{\sigma_{zz}\sigma_{yy}}\rho_{xy}\sqrt{\sigma_{xx}\sigma_{yy}}}{\rho_{zx}\sqrt{\sigma_{zz}\sigma_{xx}}} + s_y\sigma_{yy}, \\ 0 &> -(s_y + s_z) \frac{\rho_{zy}\rho_{xy}}{\rho_{zx}} \sigma_{yy} + s_y\sigma_{yy}, \\ 0 &> -(s_y + s_z) \frac{\rho_{zy}\rho_{xy}}{\rho_{zx}} + s_y, \\ \frac{\rho_{zy}\rho_{xy}}{\rho_{zx}} s_z &> \left(1 - \frac{\rho_{zy}\rho_{xy}}{\rho_{zx}}\right) s_y. \end{aligned} \quad (134)$$

where (a) we substituted $\alpha = (s_y + s_z)\sigma_{zy}/\sigma_{zx}$ and (b) we substituted $\sigma_{ij} = \rho_{ij}\sqrt{\Sigma_{ii}\Sigma_{jj}}$. Let $d = \rho_{zy}\rho_{xy}/\rho_{zx}$ for a simpler notation. Then the inequality can be written as

$$ds_z > (1 - d)s_y. \quad (135)$$

We have five scenarios

1. If $d < 0$, then

$$s_z \frac{d}{(1 - d)} > s_y. \quad (136)$$

Since $d < 0$, $d/(1-d) < 0$. Therefore $s_y < a$, where $a > 0$, such that y can be both positive and negative. Since $d < 0$, $|1-d| > |d|$ and thus $|d/(1-d)| < 1$. Therefore, if $s_y > 0$,

$$\begin{aligned} |s_y| &< |s_z \frac{d}{(1-d)}|, \\ &= |s_z| \cdot |\frac{d}{(1-d)}|, \\ &< |s_z|. \end{aligned} \tag{137}$$

In addition, $s_z < 0$, therefore $\text{sign}(s_z + s_y) = \text{sign}(s_z) = -$. If $s_y < 0$, then $\text{sign}(s_z + s_y) = -$

2. If $d = 0$, then

$$0 > s_y. \tag{138}$$

In addition, $s_z < 0$, therefore $\text{sign}(s_z + s_y) = -$.

3. If $0 < d < 1$, then

$$s_z \frac{d}{(1-d)} > s_y. \tag{139}$$

Since $0 < d < 1$, $d/(1-d) > 0$. Moreover, $s_z < 0$, therefore $s_z d/(1-d) < 0$ and thus $s_y < 0$. Hence, $\text{sign}(s_z + s_y) = -$.

4. If $d = 1$, then

$$s_z > 0. \tag{140}$$

Since $s_z < 0$, this is a contradiction.

5. If $d > 1$, then

$$s_z \frac{d}{(1-d)} < s_y, \tag{141}$$

since $1-d < 0$ the sign of the inequality has flipped. In addition, since $d/(1-d) < 0$ and $s_z < 0$, $s_z \frac{d}{(1-d)} > 0$ and therefore $s_y > 0$. Furthermore, since $d > 0$, $|d| > |1-d|$ and thus $|d/(1-d)| > 1$. Hence,

$$\begin{aligned} |s_y| &> |s_z \frac{d}{(1-d)}|, \\ &= |s_z| \cdot |\frac{d}{(1-d)}|, \\ &> |s_z|. \end{aligned} \tag{142}$$

Therefore, $\text{sign}(s_z + s_y) = \text{sign}(s_y) = +$.

To summarise the result,

$$\text{sign}(s_z + s_y) = \begin{cases} - & , \text{ if } d < 1, \\ + & , \text{ if } d > 1, \end{cases} \tag{143}$$

and if $d = 1$ there is no valid solution for the set of equations and hence $\Sigma \notin \mathcal{M}_{G,\alpha}^p$. Therefore the sign of $\alpha = -(s_y + s_z)\sigma_{zy}/\sigma_{zx}$ is

$$\text{sign}(\alpha) = \begin{cases} \text{sign}(\sigma_{zy}/\sigma_{zx}) & , \text{ if } \rho_{zy}\rho_{xy}/\rho_{zx} < 1, \\ -\text{sign}(\sigma_{zy}/\sigma_{zx}) & , \text{ if } \rho_{zy}\rho_{xy}/\rho_{zx} > 1. \end{cases} \tag{144}$$

Since $\Sigma \in \mathcal{M}_{G,\alpha}^p$, we have $\sigma_{zy} \neq 0$ meaning that the sign of α is $+$ or $-$. Therefore there exists no $\Sigma \in \mathcal{M}_{G,\alpha}^0$, such that by virtue of the $\mathcal{M}^0$ criterion Theorem 4.2, for any $\Sigma \in \mathcal{M}_{G,\alpha}^p$, the sign of edge α in graph G is identifiable.

□

D.5 THEOREM 4.10

We use results and steps from the proofs for that same structures without latent variables detailed in D.4. The only difference in the setup will be that the variable H is hidden now. This means that the covariance matrix values that depend on H , i.e., Σ_h , will be unknown. Therefore these variables are treated as unknown variables and can be chosen within the bounds of $\Sigma \in F_G$.

D.5.1 Cause and Effect

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1a, where node H is hidden. By Lemma 4.8, $\text{sign}(\alpha) = \text{sign}(\sigma_{hy})$. Since we can choose σ_{hy} , the sign of α can always be chosen both positive and negative. Thus for any $\Sigma \in \mathcal{M}_{G,\alpha}^p$, we have $\Sigma \in \mathcal{M}_{G,\alpha}^+$ and $\Sigma \in \mathcal{M}_{G,\alpha}^-$. Therefore, $\mathcal{M}_{G,\alpha}^+ = \mathcal{M}_{G,\alpha}^-$, using Definition 2.6, α is non-identifiable in graph G

□

D.5.2 Confounding

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1c, where node H is hidden. By Lemma 4.7, any covariance matrix Σ satisfying one of the the following conditions renders α identifiable:

$$\begin{aligned} (c.1) \quad & \frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2)(\rho_{hx}\rho_{hy} - \rho_{xy})}{2\rho_{hx}\rho_{hy}(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1)} \geq 1, \\ (c.2) \quad & \text{sign}(\sigma_{hx}\sigma_{hy}) \neq \text{sign}(\sigma_{xy}), \\ (c.3) \quad & \frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} \geq 1. \end{aligned} \tag{145}$$

Hence, a covariance matrix Σ satisfying all the following conditions renders α non-identifiable:

$$\begin{aligned} (c'.1) \quad & \frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2)(\rho_{hx}\rho_{hy} - \rho_{xy})}{2\rho_{hx}\rho_{hy}(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1)} < 1, \\ (c'.2) \quad & \text{sign}(\sigma_{hx}\sigma_{hy}) = \text{sign}(\sigma_{xy}), \\ (c'.3) \quad & \frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} < 1. \end{aligned} \tag{146}$$

We will try to construct such a covariance matrix Σ satisfying the conditions in Eq. (146) by choosing the values of the hidden part of the covariance matrix, i.e., in

$$\Sigma = \begin{bmatrix} \sigma_{hh} & \sigma_{hx} & \sigma_{hy} \\ \sigma_{hx} & \sigma_{xx} & \sigma_{xy} \\ \sigma_{hy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \in \mathcal{M}_{G,\alpha}^p \tag{147}$$

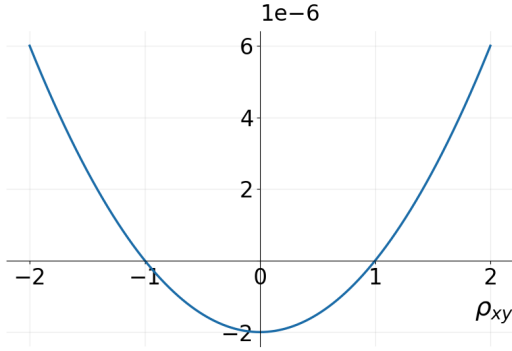
we set the values for

$$\begin{bmatrix} \sigma_{hh} & \sigma_{hx} & \sigma_{hy} \\ \sigma_{hx} & \cdot & \cdot \\ \sigma_{hy} & \cdot & \cdot \end{bmatrix} \in \mathcal{M}_{G,\alpha}^p. \tag{148}$$

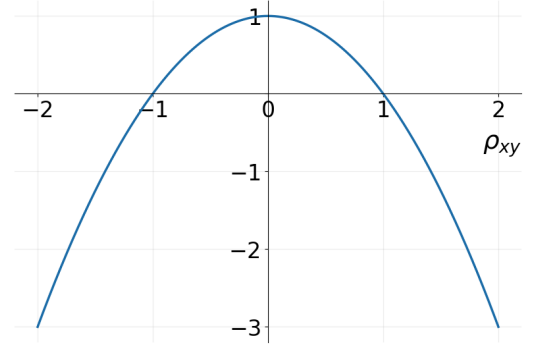
We set $\rho_{hx} = a|\rho_{xy}|$ and $\rho_{hy} = a\rho_{xy}$ with $0 < a < 1$.

To begin with Condition (c'.1), this becomes

$$\begin{aligned} & \frac{(2\rho_{hy}^2\rho_{hx}^2 - \rho_{hy}^2 - \rho_{hx}^2)(\rho_{hx}\rho_{hy} - \rho_{xy})}{2\rho_{hx}\rho_{hy}(\rho_{hx}^2 - 1)(\rho_{hy}^2 - 1)} < 1, \\ & \frac{(2a^4\rho_{xy}^4 - a^2\rho_{xy}^2 - a^2\rho_{xy}^2)(a^2|\rho_{xy}| - 1)}{2a^2|\rho_{xy}|(a^2\rho_{xy}^2 - 1)(a^2\rho_{xy}^2 - 1)} < 1, \\ & (2a^4\rho_{xy}^4 - a^2\rho_{xy}^2 - a^2\rho_{xy}^2)(a^2|\rho_{xy}| - 1) \stackrel{(a)}{<} 2a^2|\rho_{xy}|(a^2\rho_{xy}^2 - 1)(a^2\rho_{xy}^2 - 1), \end{aligned} \tag{149}$$



(a) The left-hand side of Eq. (150).



(b) The right-hand side of Eq. (154).

Figure 2: The sub-figures show the shape of two functions that are used in the proof of Appendix D.5.2.

where in (a) we use that the denominator is positive (see below Eq. (92)). Since Eq. (149) contains only contributions ρ_{xy}^2 and $|\rho_{xy}|$, we can without loss of generality only consider the domain $\rho_{xy} \in (0, 1)$ for the derivation. Eq. (149) then becomes

$$\begin{aligned}
 (2a^4\rho_{xy}^4 - a^2\rho_{xy}^2 - a^2\rho_{xy}^2)(a^2\rho_{xy} - 1) &< 2a^2\rho_{xy}(a^2\rho_{xy}^2 - 1)(a^2\rho_{xy}^2 - 1), \\
 2a^6\rho_{xy}^5 - 2a^4\rho_{xy}^3 - 2a^4\rho_{xy}^4 + 2a^2\rho_{xy}^2 - 2a^2\rho_{xy}(a^2\rho_{xy}^2 - 1)^2 &< 0, \\
 2a^6\rho_{xy}^5 - 2a^4\rho_{xy}^3 - 2a^4\rho_{xy}^4 + 2a^2\rho_{xy}^2 - 2a^2\rho_{xy}(a^4\rho_{xy}^4 - 2a^2\rho_{xy}^2 + 1) &< 0, \\
 2a^6\rho_{xy}^4 - 2a^4\rho_{xy}^2 - 2a^4\rho_{xy}^3 + 2a^2\rho_{xy} - 2a^2(a^4\rho_{xy}^4 - 2a^2\rho_{xy}^2 + 1) &< 0, \\
 2a^6\rho_{xy}^4 - 2a^4\rho_{xy}^2 - 2a^4\rho_{xy}^3 + 2a^2\rho_{xy}^2 - 2a^6\rho_{xy}^4 + 4a^4\rho_{xy}^2 - 2a^2 &< 0, \\
 -2a^4\rho_{xy}^3 + 2(a^2 + a^4)\rho_{xy}^2 - 2a^2 &< 0.
 \end{aligned} \tag{150}$$

Figure 2a shows the shape of the left-hand side of Eq. (150). The shape is symmetric around 0, and on the domain $\rho_{xy} \in [0, 2]$ the shape is convex. Therefore, on the domain $\rho_{xy} \in (0, 1)$,

$$-2a^4\rho_{xy}^3 + 2(a^2 + a^4)\rho_{xy}^2 - 2a^2 < -2a^41^3 + 2(a^2 + a^4)1^2 - 2a^2 = 0. \tag{151}$$

Hence, Condition (c'.1) is satisfied.

To continue with Condition (c'.2). We have that

$$\text{sign}(\sigma_{hx}\sigma_{hy}) = \text{sign}(a|\rho_{xy}|a\rho_{xy}) = \text{sign}(\rho_{xy}) = \text{sign}(\sigma_{xy}), \tag{152}$$

where we have used that $e, |\rho_{xy}| > 0$. Hence, Condition (c'.2) is satisfied.

Finally, we consider Condition (c'.3). We have that

$$\frac{\rho_{hy}\rho_{hx}}{\rho_{xy}} = \frac{a|\rho_{xy}|a\rho_{xy}}{\rho_{xy}} = a^2|\rho_{xy}| < 1, \tag{153}$$

where we have used that $0 < a < 1$ and $0 < |\rho_{xy}| < 1$. Hence, Condition (c'.3) is satisfied. Therefore, all conditions are satisfied and the choice $\rho_{hx} = a|\rho_{xy}|$ and $\rho_{hy} = a\rho_{xy}$ with $0 < a < 1$ makes the covariance matrix Σ non-identifiable.

Next we verify if the choice is t -faithful. We begin by verifying that our choice $\Sigma \in PD_3$. We note that, if we set $\rho_{hx} = a|\rho_{xy}|$ and $\rho_{hy} = a\rho_{xy}$ with $0 < a < 1$. Then, since, $\rho_{xy} \in (-1, 1) \setminus \{0\}$, we have that $\rho_{hx} \in (0, 1)$ and $\rho_{hy} \in (-1, 1) \setminus \{0\}$. Thereby, satisfying part of the constraints. Therefore, only verifying if Sylvester's criterion holds remains. This means we need to verify Eq. (38), i.e.,

$$\begin{aligned}
 0 &< 1 + 2\rho_{xy}\rho_{hx}\rho_{hy} - (\rho_{xy}^2 + \rho_{hy}^2 + \rho_{hx}^2), \\
 0 &\stackrel{(a)}{<} 1 + 2a^2|\rho_{xy}|\rho_{xy}^2 - (a^2\rho_{xy}^2 + a^2\rho_{xy}^2 + \rho_{xy}^2), \\
 0 &< 1 + 2a^2\rho_{xy}^2(|\rho_{xy}| - 1) - \rho_{xy}^2,
 \end{aligned} \tag{154}$$

where in (a) we set $\rho_{hx} = a|\rho_{xy}|$ and $\rho_{hy} = a\rho_{xy}$. Figure (2b) shows the shape of the right-hand side of Eq. (154). The shape is symmetric around $\rho_{xy} = 0$, and on the domain $\rho_{xy} \in [0, 2)$ is concave. Hence, for $\rho_{xy} \in (-1, 1) \setminus \{0\}$ we have that

$$1 + 2a^2\rho_{xy}^2(|\rho_{xy}| - 1) - \rho_{xy}^2 > 1 + 2a^2(\pm 1)^2(|\pm 1| - 1) - (\pm 1)^2 = 0. \quad (155)$$

Hence, Sylvester's criterion holds.

In addition, since $\rho_{hx} \in (0, 1)$ and $\rho_{hy} \in (-1, 1) \setminus \{0\}$, the choice respects the marginal independences.

Finally, combining that our choice $\Sigma \in PD_3$ and Σ respects the marginal independences, we have that $\Sigma \in F_G$. Therefore, by the proof in Appendix D.4.3 $\mathcal{M}_{G,\alpha}^p = F_G$, we have that for any observable block Σ_{oo} , we can always construct a valid covariance matrix Σ such that α is non-identifiable in graph G .

□

D.5.3 Instrumental Variable (IV)

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1e, where node H is hidden. By Lemma 4.8, $\text{sign}(\alpha) = \text{sign}(\sigma_{zy})/\text{sign}(\sigma_{zx})$. This is still constrained by the observed part of Σ . Therefore, we can use the same conclusion as in the case without latent variables. By Theorem 4.6, the sign of edge α in graph G is identifiable.

□

D.5.4 Cycle with IV

Proof. We adopt the assumptions and conventions stated at the start of this section. Let $G = (V, E)$ be the graph of Fig. 1f, where node H is hidden. By Lemma 4.8,

$$(\alpha) = \begin{cases} \text{sign}(\sigma_{zy}/\sigma_{zx}) & , \text{ if } \rho_{zy}\rho_{xy}/\rho_{zx} < 1, \\ -\text{sign}(\sigma_{zy}/\sigma_{zx}) & , \text{ if } \rho_{zy}\rho_{xy}/\rho_{zx} > 1. \end{cases} \quad (156)$$

This is still constrained by the observed part of Σ . Therefore, we can use the same conclusion as in the case without latent variables. By Theorem 4.6, the sign of edge α in graph G is identifiable.

□