Statistica Sinica Preprint No: SS-2025-0236	
Title	Quantification and Inference of Asymmetric Relations Under Generative Exposure Mappings
Manuscript ID	SS-2025-0236
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0236
Complete List of Authors	Soumik Purkayastha and Peter Xuekun Song
Corresponding Authors	Soumik Purkayastha
E-mails	soumik@pitt.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Quantification and Inference of Asymmetric Relations Under Generative Exposure Mappings

Soumik Purkayastha — *University of Pittsburgh*

Peter X.-K. Song — *University of Michigan*

Abstract: Learning directionality between variables is crucial yet challenging, especially for mechanistic relationships without a priori ordering assumptions. We propose a coefficient of asymmetry to quantify directional asymmetry using Shannon's entropy within a generative exposure mapping (GEM) framework. GEMs arise from experiments where a generative function g maps exposure X to outcome Y through $Y = g(X)$, extended to noise-perturbed GEMs as $Y = g(X) + \epsilon$. Our approach considers a rich class of generative functions while providing statistical inference for uncertainty quantification—a gap in existing bivariate causal discovery techniques. We establish large-sample theoretical guarantees through data-splitting and cross-fitting techniques, implementing fast Fourier transformation-based density estimation to avoid parameter tuning. The methodology accommodates contamination in outcome measurements. Extensive simulations demonstrate superior performance compared to competing causal discovery methods. Applied to epigenetic data examining DNA methylation and blood pressure relationships, our method unveils novel pathways for cardiovascular disease genes *FGF5* and *HSD11B2*. This framework serves as a discovery tool for improving scientific research rigor, with GEM-induced asymmetry representing a low-dimensional imprint of underlying causality.

Key words and phrases: cross-fitting, data-splitting, differential entropy, directionality, information theory.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

1. Introduction

In many statistical applications, *ordering* among variables is prefixed according to a certain scientific hypothesis, scientific knowledge, or a problem of interest. However, when the ordering itself is of scientific interest or if statistical analyses are sensitive to the choice of ordering, an inevitable challenge lies in inferring a sense of order in a given set of variables. We posit that the notion of order, or *asymmetry* (if it exists), is reflective of an underlying generative mechanism that maps an exposure X to an outcome Y. This paper aims to develop statistical methods that quantify and infer said asymmetry in a general setup for continuous data with or without contamination of outcomes. We term this general setup as the *generative exposure mapping* (GEM) that defines an underlying ordering. Conceptually, GEMs arise from experiments where a mechanistic procedure, or generative function (GF) g outputs an outcome Y for an input exposure X through the mapping $Y = g(X)$. To accomodate contaminated, or noise-perturbed outcomes, the GEM framework can be expanded to *noise-perturbed GEM* (NPGEM), denoted by $Y = g(X) + \epsilon$, where ϵ is treated as the outcome contaminant. In this work, we show that an effective strategy for the assessment of asymmetry in GEMs (with mild structural assumptions on GFs) is to prove or disprove the ordering induced from a hypothesized GEM using Shannon's information theory. We propose a conceptually easy and computationally manageable statistical methodology to quantify and provide inference for relational asymmetry. As evidenced in this paper, the utility of Shannon's information theoretic measures allows us to study asymmetry with no need of estimating the GF g.

The simplest occasions of asymmetry involve placing spatial or temporal ordering conditions between the two (Cox, 1992). Establishing a presumed causal ordering of variables

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

often requires using specific subject-matter knowledge in connection to external or *a priori* information (Cox, 1990). Alternatively, distributional asymmetries are studied by factorizing the joint density f_{XY} as the product of the marginal f_X and the conditional $f_{Y|X}$; see Choi et al. (2020); Tagasovska et al. (2020); Ni (2022), among others. Recently, some asymmetric measures of association have been proposed, including the generalized measure of correlation (Zheng et al., 2012) and a rank-based asymmetry measure (Chatterjee, 2020), but none have meaningful connections to directionality implied by GEMs. For example, in Zheng et al. (2012) the authors propose generalized measures of correlation for asymmetry and nonlinearity but they cannot track directionality implied by GEMs. Similarly, Chatterjee (2020) proposes a rank-based measure of association that is asymmetric, but fails to reflect any generative mechanism. The role of ordering in causal inference is apparent where the existence of *a priori* direction of causation is inherently hypothesized. This prompts the need for *causal discovery* on learning underlying causal structures from observational data (Pearl, 2009).

We establish a self-contained theoretical framework that significantly broadens the existing Information Geometric Causal Inference (IGCI) (Daniušis et al., 2010; Janzing et al., 2012) framework for examining asymmetry in exposure-outcome pairs. In the IGCI framework, the authors propose methodology to examine directionality in the GEM $Y = g(X)$ for bijective g. The GEM framework broadens the scope of this framework by considering a much richer class of GFs, including bijective functions. Moreover, we provide statistical inference to quantify the uncertainty in the determination of underlying directionality, while most existing bivariate causal discovery techniques (Hoyer et al., 2008; Fonollosa, 2019; Blöbaum et al., 2019) do not. Moreover, our method does not assume any sort of dependence struc-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

ture between the exposure X and contaminant ϵ in NPGEMs, providing a generalization of another causal discovery framework proposed by Hoyer et al. (2008).

We organize the paper as follows: Section 2 introduces the setup of GEMs and some basic information theoretic concepts. Section 3 provides a recap of IGCI and provides a population-level measure of asymmetry under a hypothesized GEM. Section 4 introduces NPGEMs to allow for errors in the outcome Y with technical justifications for both feasibility and robustness of the proposed coefficient of asymmetry with random samples from the GEM. Section 5 presents estimation and inference details; we implement a fast Fourier transformation-based density estimation technique to estimate key estimands of interest under a proposed GEM, followed by a cross-fitting technique to quantify estimation uncertainty while improving statistical efficiency. Finally, Sections 6 and 7 exhibit the performance of the proposed framework, estimation, and inference through simulation studies and real data applications respectively.

2. Preliminaries

We begin by outlining basic concepts and notation used throughout our paper. We first define the settings of our GEM and then introduce some elementary information theoretic concepts that are key to our framework.

2.1 The generative exposure mapping setup

Let us consider a population $\mathcal{U}$ of units indexed by $i = 1, \ldots n$ on which a randomized experiment is performed. Since an experimental design details the selection of a particular exposure value x from the support $\mathcal{X}$ of continuous exposure X, we can say the probability

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

density function (PDF) of X, given by f_X, governs the experiment. Let the $i-$th population unit independently get exposure X_i. According to Cox (1990, 1992), a causal ordering between X and outcome Y must be explained through (i) a mechanism governing the exposure and (ii) a generative process yielding the outcome, with exposure as the input. We consider a structured generative mechanism specified by the GEM of the form with an ordering between two variables:

$$Y = g(X). \tag{2.1}$$

Let $Y_i = g(X_i)$ be the outcome of the i-th population unit in $\mathcal{U}$, which are assumed to be independent. In a GEM, note that the behaviour of Y, described by its PDF f_Y, is dictated entirely by f_X and g. The exposure mapping model of Aronow and Samii (2017) is noted to be a special case of the GEM in (2.1) for discrete X and is further equivalent to the "effective treatments" setup described by Manski (2013). The overall challenges of examining asymmetry in GEMs involve inferring whether exposure X yields outcome Y or *vice-versa* from paired observations $\{(x_i, y_i)\}_{i=1}^n$. We focus exclusively on assessing asymmetry in GEMs by leveraging information-theoretic notions of *entropy*. The subsequent section provides a brief review of entropy, which will play a key role in quantifying asymmetry in GEMs. It is important to note that our framework presupposes that X and Y are associated, which must be established before testing for directionality. If needed, one can use the methodology proposed by Purkayastha and Song (2024) to test for association using mutual information (Cover, 2005).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

2.2 Basic information theoretic concepts

Let X and Y be two random variables with joint density function f_{XY}. Let f_X and f_Y be the marginal densities of X and Y, respectively. Further, let $\mathcal{X}$ and $\mathcal{Y}$ denote the respective support sets of X and Y. The *joint differential entropy* of (X, Y) and *marginal differential entropy* of X are given by $H(X,Y) = E_{XY}\{-\log f_{XY}(X,Y)\}$ and $H(X) = E_X\{-\log f_X(X)\}$ respectively (Shannon, 1948). Differential entropy measures the randomness of a continuous random variable (Orlitsky, 2003) and is a limiting case of Shannon's entropy, which was originally described for discrete random variables. For the remainder of this paper, we omit the word differential, although our focus is always continuous random variables. In the next section, we demonstrate the flexibility and capacity of Shannon's entropy measure to quantify asymmetry under GEMs.

3. Asymmetry in GEMs

In this section we derive a legitimate population-level measure that will be used to confirm ordering or asymmetry under a hypothesized GEM.

Let us begin with the case of a GEM as described by (2.1); intuitively, we consider the exposure X is collected from an experiment that is governed by a density f_X. This setup is often used in Fisher's fiducial inference (Hannig et al., 2016) in which errors, although unobservable in practice, may be simulated from a certain pivotal distribution. With X serving as an input, the *GF* g generates the population-level outcome $Y = g(X)$, which has density f_Y. We will consider a specific classification of *GF*s, denoted by $\mathcal{G}_-$, as defined below.

Definition 1. Let $\mathcal{G}_-$ be a class of differentiable functions on a compact set $\gamma \subset \mathbb{R}$. For any

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

$g \in \mathcal{G}_-$, let the following integral be well-defined and satisfy the inequality

$$\exp\left(|\gamma|^{-1}\int_{\gamma}\log|\nabla g(x)|\,dx\right) < 1,$$

where $|\gamma|$ denotes the Lebesgue measure of γ, and ∇ denotes the gradient of g. We interpret any function $g \in \mathcal{G}_-$ to exhibit *contracting dynamics* over the support γ, meaning that the geometric mean of $|\nabla g(x)|$ over γ is strictly less than one.

Remark 1. Alternatively, we may have a class of functions $\mathcal{G}_+$ with 'expanding dynamics' over support γ; for any $g^\star \in \mathcal{G}_+$, we have

$$\exp\left(|\gamma|^{-1}\int_{\gamma}\log|\nabla g^\star(x)|dx\right) > 1,$$

i.e., the geometric mean of $|\nabla g(x)|$ over γ is strictly more than one.

An identifiability assumption required to unearth the induced asymmetry from a GEM is that the distribution of exposure X (or the law of X) and the mechanism of the *GF* g do not influence each other when yielding the outcome Y, which we quantify through functional orthogonality, as defined below. In the literature of functional analysis, orthogonality is appropriate to characterize the notion of "no influence", which will be adopted in this paper. Similar assumptions are found in Janzing and Schölkopf (2010); Daniušis et al. (2010); Mooij et al. (2016).

Assumption 1. *Let g be a differentiable function and f_X be the density that governs an experiment with exposure X with compact support $\mathcal{X}$. We say that g and f_X are functionally*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

orthogonal if they satisfy the equality

$$\int_{\mathcal{X}} \log\left(|\nabla g(x)|\right) f_X(x)dx = |\mathcal{X}|^{-1} \int_{\mathcal{X}} \log\left(|\nabla g(x)|\right) dx,$$

where the ∇ *operator denotes the gradient of* g *with respect to its argument.*

Remark 2. Assumption 1 automatically holds when $X \sim \mathcal{U}(0,1)$. This is analogous to the assumption of randomization or no confounding in the school of Neyman-Rubin causality; the uniform distribution on X is analogous to randomization in an experiment that leads to no bias in operating exposure X.

Remark 3. Assumption 1 may also hold under non-uniform distributions such as Beta(a, b), provided there is a functional balance between the shape of f_X and the transformation g. For example, if $g(x) = x^k, k \neq 1$ or $g(x) = \log(x)$, Assumption 1 holds for specific choices of (a, b). Intuitively, this resembles balancing moments or information content: the shape of f_X counteracts the local expansion or compression induced by g, yielding an overall uncorrelatedness. As a concrete case, when $g(x) = x^k$, Assumption 1 holds for Beta(a, b) distributions with $\mathbb{E}[\log X] = -1$, which includes asymmetric shapes such as Beta$(1, 1.5)$. This highlights that functional orthogonality is not limited to uniform inputs, but rather depends on a compatibility between the distribution of X and the dynamics of g.

Remark 4. To interpret Assumption 1, consider the quantities $\log\left(|\nabla g(X)|\right)$ and $f_X(X)$ as random variables on support $\mathcal{X}$. Then, their covariance with respect to the uniform distribution on $\mathcal{X}$ is proportional to

$$\int_{\mathcal{X}} \log\left(|\nabla g(x)|\right) f_X(x)dx - |\mathcal{X}|^{-1} \int_{\mathcal{X}} \log\left(|\nabla g(x)|\right) dx \int_{\mathcal{X}} f_X(x)dx,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

where the last integral is equal to one. Consequently, the equality in 1 expresses that the density of exposure, given by $f_X(x)$, and the (log-transformed) dynamics of the generative mechanism, given by $\nabla g(x)$, are uncorrelated mechanisms of nature, linking with the notion of algorithmic independence (Janzing and Schölkopf, 2010). Geometrically, Assumption 1 requires that the parts of the input domain where the data are densely distributed are not systematically related (or, correlated) to the parts of the function g that are most expansive or contractive (i.e., where $|\nabla g(x)|$ is large or small). This holds naturally in a randomized experiment where X is, for example, uniformly distributed.

The validity of Assumption 1 in the absence of unmeasured confounding may be checked; we provide an a practical guide in Algorithm 1 below. In brief, we outline steps to numerically estimate the left and right side quantities in Assumption 1 and obtain the difference between them. We call this estimate the "orthogonality deviation score" and denote it by D. In addition to the the point estimate $\hat{D}$, we also provide a bootstrap confidence interval. A confidence interval that is narrow and contains to zero suggests Assumption 1 is plausible for the specified direction. A confidence interval that is far from zero indicates a likely violation of the assumption. Please see section 5 of supplement for an example illustrating two contrasting scenarios, one satisfying the assumption and one violating it.

We now propose the decision rule that will detect the direction of causation $X \to Y$ induced by the GEM $Y = g(X)$. The decision rule uses the following asymmetry coefficient between X and Y, as defined below.

Definition 2. Let $g \in \mathcal{G}_-$ be a *GF* in (2.1). Then, we define the *asymmetry coefficient* between X and Y as follows

$$C_{X \to Y} := H(X) - H(Y),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Algorithm 1 Orthogonality Deviation Score (D) for Assumption 1

input: dataset of paired observations $\{(x_i, y_i)\}_{i=1}^n$.
output: D and its 95% bootstrap confidence interval.
fit a smoothing spline to obtain an estimate of $\hat{g}(x)$, numerically compute $\nabla\hat{g}(x)$.
estimate $\text{LHS}_{\text{est}} \leftarrow n^{-1}\sum_{i=1}^n \log(|\nabla\hat{g}(x_i)|)$
estimate $\text{RHS}_{\text{est}} \leftarrow |\hat{\mathcal{X}}|^{-1}\int_{\min(X)}^{\max(X)} \log(|\nabla\hat{g}(x)|)dx$
calculate $D \leftarrow \text{LHS}_{\text{est}} - \text{RHS}_{\text{est}}$.
for $b = 1$ to B (e.g., $B = 500$) **do**
    generate a bootstrap sample $\{(x_i^*, y_i^*)\}_{i=1}^n$.
    repeat steps 3-7 on the bootstrap sample to compute a score D_b^*.
end for
construct a 95% CI from the empirical distribution of scores
return $\hat{D}$ and its 95% bootstrap CI.

where $H(X)$ and $H(Y)$ are the entropy functions of X and Y respectively.

The key challenge of this approach lies in the fact that we do not have knowledge of the generative function g. However, we see that the asymmetry coefficient is able to capture generative asymmetry even without explicitly knowing the form of g, since we have the following

$$C_{X\to Y} = -\int_{\mathcal{X}} \log\left(|\nabla g(x)|\right) f_X(x)dx = -|\mathcal{X}|^{-1}\int_{\mathcal{X}} \log\left(|\nabla g(x)|\right) dx > 0,$$

where the inequalities are supported by Assumption 1 and Definition 1. Following Definition 2, we formulate the following hypothesis to confirm or negate the presence of the pathway $X \to Y$ for a given GEM $Y = g(X)$, where $g \in \mathcal{G}_-$ has contracting dynamics:

$$H_0 : C_{X\to Y} \leq 0 \quad \text{vs} \quad H_1 : C_{X\to Y} > 0. \tag{3.1}$$

If H_0 is rejected, we obtain evidence to support presence of the pathway $X \to Y$ induced by the GEM in (2.1).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Remark 5. The positive sign of $C_{X\to Y}$ relies on the 'contracting dynamics' nature of g. However, in the scenario where g has expanding dynamics (see Remark 1), we will have $C_{X\to Y} < 0$. Hence, we must modify hypothesis in (3.1) accordingly, given by

$$H_0 : C_{X\to Y} \geq 0 \quad \text{vs} \quad H_1 : C_{X\to Y} < 0. \tag{3.2}$$

Remark 6. A key challenge from a practical standpoint is that if g has expanding dynamics, its inverse g^{-1} must have contracting dynamics; please see section 5 of the supplement for a formal proof. While one asymmetry coefficient is defined as $C_{X\to Y} = H(X) - H(Y)$, while the coefficient for the reverse direction is $C_{Y\to X} = H(Y) - H(X) = -C_{X\to Y}$. Consider the two hypotheses:

1. Setting A: The true causal direction is $X \to Y$ with an expanding function g. This implies $C_{X\to Y} < 0$.

2. Setting B: The true causal direction is $Y \to X$ with a contracting function h. This implies $C_{Y\to X} > 0$.

The mathematical conditions for these two settings, $C_{X\to Y} = H(X) - H(Y) < 0$ and $C_{Y\to X} = H(Y) - H(X) > 0$, are identical. Hence, just by observing the sign of the asymmetry coefficient, we cannot distinguish between two complementary scenarios. The key to resolving this ambiguity is that Assumption 1 is not symmetric: it can be proved that at most one of the two settings described above will be compatible with Assumption 1. In other words, Assumption 1 will be compatible with at most one of the implied causal directions $X \to Y$ or $Y \to X$. Please see section 5 of the supplement for a formal proof.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Remark 7. If we have prior domain knowledge that the generative mechanism is, for instance, known to be expanding, the ambiguity is resolved immediately. Upon observing a significantly negative asymmetry coefficient ($C_{X\to Y} < 0$), we can confidently conclude the direction is $X \to Y$. If we observe a positive coefficient, we would conclude the direction is $Y \to X$. If we do not have prior domain knowledge, we propose a exploratory tool to detect if the generative function in the GEM $Y = g(X)$ is expanding or contracting; please see section 6 of the supplement for more details.

Further, it is worth pointing out that even when Assumption 1 does not hold, the proposed analytics and algorithms in our framework remain useful and produce meaningful, yet weaker interpretability of the directionality. That is, although without Assumption 1 the interpretability of causality cannot be guaranteed, the detection of relational directionality remains meaningful and useful in the study of asymmetric relationships. This can easily be seen by noting that $H(X) - H(Y) = H(X|Y) - H(Y|X)$. Therefore, $H(X) > H(Y)$ implies $H(X|Y) > H(Y|X)$. This indicates that the uncertainty of Y given X is less than the uncertainty of X given Y, establishing that X provides more information about Y than vice-versa.

4. Asymmetry in the presence of noise perturbation

In a practical data generation scenario, measurement errors or noise perturbations are inevitable. Thus, we now extend the setup of the previously noise-free GEM by adding randomness on the outcome Y, so a noise-perturbed GEM (NPGEM) takes the form:

$$Y^* = g(X) + \epsilon, \text{ such that } \mathbb{E}(\epsilon) = 0, \mathbb{V}(\epsilon) = \sigma, \tag{4.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

where Y^* is the 'contaminated' version of outcome Y.

4.1 Contamination for GFs with contracting dynamics

From Definition 2 and the hypothesis in 3.1, note that the asymmetry coefficient can measure population-level asymmetry between exposure and outcome for GFs with contracting dynamics even with noise perturbation when we encounter Y^* instead of Y in the population-level model, so long as the ordering of $H(Y^*)$ being lower than $H(X)$ is preserved.

We introduce an intermediate NPGEM with normally distributed disturbance Z having mean 0 and variance 1, such that $H(Y^*) \leq H(Y + \sqrt{\sigma'}Z)$ for a certain $\sigma' > 0$. This NPGEM is merely for a technical need; de Bruijn's theorem (Cover, 2005) has shown that $H(Y + \sqrt{\sigma'}Z)$ is an increasing function of σ'. We state the following theorem:

Theorem 1. *We consider a NPGEM given by (4.1) with $g \in \mathcal{G}_-$, f_Y and $H(Y)$ denoting the density and entropy of Y respectively. Then, the upper-bound on the entropy of Y^* is given by $H(Y^*) \leq H(Y) + \frac{1}{2}\log(\sigma' I(Y) + 1)$, where σ' is a value such that $H(Y^*) \leq H(Y + \sqrt{\sigma'}Z)$ with $Z \sim N(0,1)$ and $X \perp Z$, and $I(Y) := \mathbb{E}_Y[\nabla_Y \log(f_Y(Y))]^2$ is the (nonparametric) Fisher information of Y.*

Note that the existence of the σ' is ensured by de Bruijn's identity. We refer the reader to the supplement for the proof of Theorem 1. To ensure the asymmetry coefficient properly measures the induced asymmetry, we must have $H(Y^*) \leq H(Y) + \log(\sigma' I(Y) + 1)/2 < H(X)$, implying

$$I(Y) < \frac{\exp(2C_{X\to Y}) - 1}{\sigma'} = \frac{\exp(2C_{X\to Y}) - \exp(2C^0_{X\to Y})}{\sigma'}, \tag{4.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

where $C^0_{X\to Y} \equiv 0$ is value of $C_{X\to Y}$ in a *balanced* or *symmetric* relation. We argue that the numerator $\exp(2C_{X\to Y}) - \exp(2C^0_{X\to Y})$ is a measure of the deviation from symmetry and serves as the signal that our method must capture. To arrive at (4.2), note that we have (i) $C_{X\to Y} = H(X) - H(Y) > 0$, and (ii) Theorem 1 ensures $\sigma' I(Y) < \exp(2C_{X\to Y}) - \exp(2C^0_{X\to Y})$. Hence, the asymmetry coefficient works in low-noise regimes of the NPGEMs, where (4.2) provides a lower bound on the signal-to-noise ratio that the asymmetry coefficient can tolerate and remains effective to quantify NPGEM-induced asymmetry.

4.2 Contamination for GFs with expanding dynamics

For $Y^\star = g(X) + \epsilon$ with *expanding dynamics* ($g \in \mathcal{G}_+$), we have $H(g(X)) > H(X)$. To ensure that the entropy of the contaminated output $Y^\star$ remains greater than that of the input X, we consider the impact of the transformation g and the added noise ϵ. If g has *expanding dynamics*—that is, $H(g(X)) > H(X)$—then we can guarantee $H(Y^\star) > H(X)$ under the added condition of independence of noise ϵ and exposure X. Using the entropy power inequality (Cover, 2005), we can claim that the differential entropy of the contaminated output satisfies $H(Y^\star) \geq \max\left\{H(g(X)), H(\epsilon)\right\} > H(X)$. Hence, adding independent noise cannot reduce the entropy below that of $g(X)$, which already exceeds $H(X)$ when $g \in \mathcal{G}_+$.

5. Estimation and inference

In the GEM framework, we must estimate and perform inference using $\hat{C}_{X\to Y}$. To do so, we must first estimate the underlying marginal densities f_X and f_Y, which may be thought of as infinite-dimensional nuisance parameters. If the same data that were used for density estimation are also used to provide inference for $C_{X\to Y}$, standard inferential procedures may

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

suffer from bias. To circumvent this, we consider a data-splitting and cross-fitting approach (Chernozhukov et al., 2018). Moreover, we use a fast Fourier transformation (FFT) technique to estimate the underlying density functions and provide plug-in estimators of entropy. This technique is key to providing a stable estimator. To the best of our knowledge, our method is the first to combine FFT-based techniques and data-splitting techniques to estimate and perform inference on entropy. The density estimation technique in Section 5.1 provides an accurate and fast solution without incurring the need for tuning parameters. Based on the estimated densities, we obtain consistent estimates of $\hat{C}_{X\to Y}$. Next, in Section 5.2, we describe the data-splitting and cross-fitting technique that allows us to provide an inference rule for testing for GEM-induced asymmetry using either the hypothesis in 3.1 for GFs with contracting dynamics or the hypothesis in 3.2 for GFs with expanding dynamics.

5.1 Self-consistent density estimation

The self-consistent estimator (SCE) was proposed by Bernacchia and Pigolotti (2011); O'Brien et al. (2016) to minimize the mean integrated squared error (MISE) between the estimated density and the true density without incurring any manual parameter tuning. The estimation process relies on fast Fourier transforms (FFT). Utilizing this 'optimal' density estimator, Purkayastha and Song (2024) propose a plug-in estimator of MI, termed as the `fastMI`, which is shown to be a consistent and fast estimator. Extending the usage of the self-consistent density estimator here, we then estimate the marginal entropies $\hat{H}(X)$ and $\hat{H}(Y)$, thereby obtaining an estimate of $\hat{C}_{X\to Y}$.

Let us consider a random sample denoted by $\mathcal{S} = \{X_1, X_2, \ldots, X_n\}$ from an unknown density f with support $\mathcal{X}$ (without loss of generality, $\mathcal{X} = \mathbb{R}$). We assume f belongs to the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Hilbert space of square-integrable functions, given by $\mathcal{L}^2 = \left\{f : \int f^2(x)dx < \infty\right\}$. The SCE is denoted by $\hat{f} \in \mathcal{L}^2$. First, in order to define $\hat{f}$, we require a kernel function $\mathcal{K}$, which belongs to the class of functions given by

$$\mathbb{K} := \left\{\mathcal{K} : \mathcal{K}(x) \geq 0, \mathcal{K}(x) = \mathcal{K}(-x); \int \mathcal{K}(t)dt = 1\right\}.$$

Specifically, $\hat{f}$ is the convolution of a kernel $\mathcal{K}$ and delta functions centered on the dataset:

$$\hat{f}(z) \equiv n^{-1}\sum_{j=1}^{n}\mathcal{K}(x - X_j) = n^{-1}\sum_{j=1}^{n}\int_{\mathbb{R}}\mathcal{K}(s)\delta(x - X_j - s)ds, \quad x \in \mathbb{R} \tag{5.1}$$

where $\delta(x)$ is the Dirac delta function. The optimal $\hat{f}$ is identified by the optimal kernel $\hat{\mathcal{K}}$, where "optimality" is intended as minimising the MISE between the true density f and the estimator $\hat{f}$:

$$\hat{\mathcal{K}} = \underset{\mathcal{K}\in\mathbb{K}}{\operatorname{argmin}}\ MISE(\hat{f}, f) = \underset{\mathcal{K}\in\mathbb{K}}{\operatorname{argmin}}\ \mathbb{E}\left[\int_{\mathbb{R}}\{\hat{f}(x) - f(x)\}^2 dx\right], \tag{5.2}$$

where the $\mathbb{E}$ operator denotes taking expectation over the entire support of f. The SCE in (5.1) may be represented equivalently by its inverse Fourier transform pair, $\hat{\phi} \in \mathcal{L}^2$, given by $\hat{\phi}(t) = \mathcal{F}^{-1}\big(\hat{f}(z)\big) = \kappa(t)\mathcal{C}(t)$, where $\mathcal{F}^{-1}$ represents the multidimensional inverse Fourier transformation from space of data $z \in \mathbb{R}$ to frequency space coordinates $t \in \mathbb{R}$. $\kappa = \mathcal{F}^{-1}\big(\mathcal{K}\big)$ is the inverse Fourier transform of the kernel $\mathcal{K}$ and $\mathcal{C}$ is the empirical characteristic function (ECF) of the data, defined as $\mathcal{C}(t) = n^{-1}\sum_{j=1}^{n}\exp\left(itZ_j\right)$. Bernacchia and Pigolotti (2011) derive the optimal transform kernel $\hat{\kappa}$ minimizing the $MISE$ given by (5.2), given as follows:

$$\hat{\kappa}(t) = \frac{n}{2(n-1)}\left[1 + \sqrt{1 - \frac{4(n-1)}{|n\mathcal{C}(t)|^2}}\right] I_{A_n}(t), \tag{5.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

where A_n serves as a low-pass filter that yields a stable estimator (Purkayastha and Song, 2024). We follow the nomenclature of Bernacchia and Pigolotti (2011) and denote A_n as the set of "acceptable frequencies". The optimal transform kernel $\hat{\kappa}$ in (5.3) may be anti-transformed back to the real space to obtain the optimal kernel $\hat{\mathcal{K}} \in \mathbb{K}$, which yields the optimal density estimator $\hat{f}$ according to Equation 5.1.

Remark 8. The key components are (a) the ECF $\mathcal{C}(t)$: the empirical characteristic function, $\mathcal{C}(t)$, is the Fourier transform of the data and represents the observed "signal" in the frequency domain, and (b) the indicator function $I_{A_n}(t)$: this term acts as a "low-pass filter". It is active (equal to 1) only for frequencies where the signal strength, represented by $|n\mathcal{C}(t)|^2$, is above a certain threshold. In Purkayastha and Song (2024) we define the filter set as follows:

$$A_n(t) = \left\{ t \in \mathbb{R} : |C(t)|^2 \geq C_{\min}^2 = \frac{4(n-1)}{n^2} \right\}.$$

For frequencies where the signal is too weak (i.e., likely dominated by noise), the term inside the square root would be negative; the filter sets $\hat{\kappa}(t)$ to zero for these frequencies, effectively removing them and ensuring a stable estimate. As shown in Purkayastha and Song (2024), this self-consistent, data-driven process determines the optimal level of smoothing automatically, in contrast to traditional kernel methods that rely on user-specified bandwidths.

Theorem 2 presents the sufficient conditions for the estimate $\hat{f}$ to converge to the true density f for $n \to \infty$. First, we state the technical assumptions needed for Theorem 2 to hold.

Assumption 2. *Let the true density f be square-integrable and its corresponding Fourier transform ϕ be integrable.*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Assumption 3. *Let us make the following assumptions about the low-pass filter A_n: $\mathcal{V}(A_n) \to \infty$ $\mathcal{V}(A_n)/\sqrt{n} \to 0$ and $\mathcal{V}(\bar{A}_n) \to 0$ as $n \to \infty$, where $\bar{A}_n$ is the complement of A_n and the volume of A_n is given by $\mathcal{V}(A_n)$.*

Assumption 4. *Let the true density f be continuous on dense support $\mathcal{X}$.*

Theorem 2. *If Assumptions 2 and 3 hold, then the self consistent estimator $\hat{f}$, which is defined by (5.1) - (5.3), converges almost surely to the true density as $n \to \infty$. Further, if Assumption 4 holds, we have uniform almost sure convergence of $\hat{f}$ to f as $n \to \infty$.*

The proof of Theorem 2 is omitted here; see Purkayastha and Song (2024) for the detail.

5.2 Data-splitting and cross-fitting inference

We have noted earlier that if the same data that were used for density estimation are also used to provide inference for $C_{X\to Y}$, standard inferential procedures may suffer from bias (Chernozhukov et al., 2018). Here, we describe a data-splitting and cross-fitting technique to help us circumvent this issue and provide valid inference on $\hat{C}_{X\to Y}$, thereby filling a gap in literature (Daniušis et al., 2010).

Let $\mathcal{D} = \{(X_1, Y_1), \ldots, (X_{2n}, Y_{2n})\}$ be a random sample drawn from a bivariate distribution f_{XY} with marginal f_X for X and f_Y for Y. Since we do not have knowledge of f_X or f_Y, we invoke a data-splitting and cross-fitting technique to estimate the underlying density functions as well as the relevant entropy terms. That is, we first split the available data $\mathcal{D}$ into two equal-sized but disjoint sets denoted by $\mathcal{D}_1 := \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, and $\mathcal{D}_2 := \{(X_{n+1}, Y_{n+1}), \ldots, (X_{2n}, Y_{2n})\}$. Using one data split $\mathcal{D}_1$, we obtain estimates of the marginal density functions $\hat{f}_{X;1}$ and $\hat{f}_{Y;1}$ by the SCE method described in Section 5.1. The estimated density functions are evaluated for data belonging to the second data split $\mathcal{D}_2$ to

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

obtain the following estimates of marginal entropies $\widehat{H_2}(X) = -n^{-1}\sum_{j=1}^{n}\ln\left(\hat{f}_{X;1}\left(X_{n+j}\right)\right)$ and $\widehat{H_2}(Y) = -n^{-1}\sum_{j=1}^{n}\ln\left(\hat{f}_{Y;1}\left(Y_{n+j}\right)\right)$. Interchanging the roles of data splits $\mathcal{D}_1$ and $\mathcal{D}_2$, by a similar procedure, we obtain the estimated densities $\hat{f}_{X;2}$ and $\hat{f}_{Y;2}$. The estimated density functions are evaluated for data belonging to data split $\mathcal{D}_1$ to obtain the estimated entropies $\widehat{H_1}(X) = -n^{-1}\sum_{j=1}^{n}\ln\left(\hat{f}_{X;2}\left(X_j\right)\right)$ and $\widehat{H_1}(Y) = -n^{-1}\sum_{j=1}^{n}\ln\left(\hat{f}_{Y;2}\left(Y_j\right)\right)$. Taking an average of the two sets of estimates, we obtain the so-called "cross-fitted" estimates of the marginal entropies $\hat{H}(X) = \left\{\widehat{H_1}(X) + \widehat{H_2}(X)\right\}/2$ and $\hat{H}(Y) = \left\{\widehat{H_1}(Y) + \widehat{H_2}(Y)\right\}/2$. Next, we define oracle estimators of H_X, and H_Y for each of the data splits:

$$H_{0;1}(X) = -\frac{1}{n}\sum_{j=1}^{n}\ln\left(f_X\left(X_j\right)\right), \text{ and } H_{0;2}(X) = -\frac{1}{n}\sum_{j=1}^{n}\ln\left(f_X\left(X_{n+j}\right)\right), \tag{5.4}$$

with respect to $\mathcal{D}_1$ and

$$H_{0;1}(Y) = -\frac{1}{n}\sum_{j=1}^{n}\ln\left(f_Y\left(Y_j\right)\right), \text{ and } H_{0;2}(Y) = -\frac{1}{n}\sum_{j=1}^{n}\ln\left(f_Y\left(Y_{n+j}\right)\right), \tag{5.5}$$

with respect to $\mathcal{D}_2$. The quantities so obtained are averaged to obtain the "cross-fitted oracle estimates" given by $H_0(X) = \left\{H_{0;1}(X) + H_{0;2}(X)\right\}/2$, $H_0(Y) = \left\{H_{0;1}(Y) + H_{0;2}(Y)\right\}/2$. Please see Remark 9 for justification of our use of a 50/50 data-splitting ratio as opposed to unbalanced (i.e., 60/40, or 70/30) splits.

The following two theorems establish both consistency and asymptotic normality of the cross-fitted estimates. In addition to Assumptions 2 – 4, we impose the following assumption that is needed for Theorems 3 and 4 to hold.

Assumption 5. *Let the densities f_X and f_Y be bounded away from zero and infinity on their support.*

Similar assumptions are made by Moon et al. (1995); Audibert and Tsybakov (2007).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Theorem 3. *Let Assumptions 2 – 5 hold when estimating $\hat{f}_X$ or $\hat{f}_Y$ using data splits $\mathcal{D}_1$ and $\mathcal{D}_2$. We implement the data-splitting and cross-fitting procedure described in Section 5.2. Then, the cross-fitted estimate is strongly consistent, i.e., $\hat{C}_{X\to Y} \xrightarrow{a.s.} C_{X\to Y}$ as $n \to \infty$.*

Theorem 4. *Let Assumptions 2 – 5 hold when estimating $\hat{f}_X$ or $\hat{f}_Y$ using data splits $\mathcal{D}_1$ and $\mathcal{D}_2$. We implement the data-splitting procedure described in Section 5.2 to obtain cross-fitted estimates $\hat{H}(X)$ and $\hat{H}(Y)$ using $\mathcal{D}_1$ and $\mathcal{D}_2$. Then, we have the following:*

$$\sqrt{n}\begin{pmatrix}\hat{H}(X) - H_0(X)\\ \hat{H}(Y) - H_0(Y)\end{pmatrix} \xrightarrow{\mathcal{P}} 0, \text{ as } n \to \infty. \tag{5.6}$$

We refer the reader to the supplement for proofs of Theorem 3 and Theorem 4.

Lemma 1. *By the multivariate central limit theorem, assuming $\mathbb{V}[\log(f_X(X))] < \infty$ and $\mathbb{V}[\log(f_Y(Y))] < \infty$, we have $(H_0(X), H_0(Y))'$ jointly converge in distribution, given by*

$$\sqrt{n}\begin{pmatrix}H_0(X) - H(X)\\ H_0(Y) - H(Y)\end{pmatrix} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma), \text{ as } n \to \infty,$$

where Σ is the 2×2 dispersion matrix of $(H_0(X), H_0(Y))'$.

Using Theorem 4 and Lemma 1, we get the following corollary.

Corollary 1. *Let Assumptions 2 – 5 hold when obtaining $\hat{C}_{X\to Y}$ using cross-fitted estimates from data splits $\mathcal{D}_1$ and $\mathcal{D}_2$. We make use of Lemma 1 to note that $\sqrt{n}\left(\hat{C}_{X\to Y} - C_{X\to Y}\right) \xrightarrow{\mathcal{D}} N(0, \sigma_C^2)$, as $n \to \infty$, where σ_C^2 denotes the asymptotic variance of the estimate $\hat{C}_{X\to Y}$ and can be estimated by the Monte Carlo technique, given density estimates $\hat{f}_X$ and $\hat{f}_Y$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

We refer the reader to the supplement for the proof of Corollary 1. See the supplement on how to estimate $\hat{\sigma}^2_C$ using data-splitting and cross-fitting (Chernozhukov et al., 2018).

5.3 Testing for asymmetry in GEMs using $C_{X\to Y}$

The one-sided hypothesis in 3.1 postulates a putative directionality from X to Y for $g \in \mathcal{G}_-$. This null is "protected" with a high (*say*, 95%) confidence and will be rejected if $\hat{C}_{X\to Y}$ is significantly bigger than zero. The asymptotic normality established in Corollary 1 is the theoretical basis for the proposed hypothesis testing method based on $\hat{C}_{X\to Y} = \hat{H}(X) - \hat{H}(Y)$. We construct a one-sided 95% asymptotic confidence interval (CI) of $\hat{C}_{X\to Y}$, denoted by $(\hat{L}, \infty)$ where $\hat{L}$ is the estimated lower bound. Consequently, if $\hat{L} \leq 0$, we would reject the null hypothesis and conclude that the postulated directionality is disproved by the data at hand. For $g \in \mathcal{G}_+$, we would test the one-sided hypothesis in 3.2, given by $H_0 : C_{X\to Y} < 0$.

6. Simulation studies

We now report the findings of extensive simulation studies that provide empirical evidence supporting the validity of $\hat{C}_{X\to Y}$ to infer GEM-induced asymmetry in bivariate (X, Y).

6.1 Behaviour of $C_{X\to Y}$ under GEM

In this simulation, we assess the capacity of $C_{X\to Y}$ for measuring the induced asymmetry from a hypothesized GEM in the strong asymmetry framework when the postulate of Assumption 1 is satisfied. We generate Monte Carlo simulated datasets from GEM models with or without noise disturbance, each consisting of 1000 *i.i.d.* samples of pairs, where *GF* g satisfies Assumption 1 under $X \sim U(0, 1)$. We set $g(x)$ to be one of the following functions: $x^{1/3}, x^{1/2}, x^2, x^3, \exp(x)$, and $\sin(\pi x/2)$. Note that the first four GFs have contracting

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

dynamics on $[0, 1]$, while the last two have expanding dynamics. For each setting, the experiments were repeated $R = 250$ times to gauge Monte Carlo approximation errors from a set of empirical values $\{\tilde{C}^r_{X\to Y}\}_{r=1}^R$. We report the mean along with bottom and upper 2.5^{th} percentiles of the empirical $\tilde{C}_{X\to Y}$ in Table 1. Note that here the empirical distribution

Table 1: Examining mean $\tilde{C}_{X\to Y}$ and (2.5%, 97.5%) percentiles in the GEM framework. We consider uniformly distributed X and six choices of g. The first row represents the behaviour of $C_{X\to Y}$ in the noise-free GEM given by $Y = g(X)$ whereas the subsequent rows examine shifts in the behaviour of $C_{X\to Y}$ in the NPGEM given by $Y^* = g(X) + \epsilon$, where $\mathbb{V}(\epsilon) = \sigma$.

σ	$g(x)$ with contracting dynamics				$g(x)$ with expanding dynamics	
	$x^{1/3}$	$x^{1/2}$	x^2	x^3	$\exp(x)$	$\sin(\pi x/2)$
0	0.51 (0.46, 0.55)	0.27 (0.24, 0.30)	0.26 (0.19, 0.33)	0.65 (0.54, 0.78)	-0.45 (-0.50, -0.41)	-0.57 (-0.64, -0.50)
0.10	0.28 (0.24, 0.32)	0.12 (0.08, 0.15)	-0.02 (-0.06, 0.02)	0.08 (0.04, 0.13)	-0.50 (-0.54, -0.47)	-0.82 (-0.92, -0.67)
0.20	-0.00 (-0.05, 0.04)	-0.10 (-0.15, -0.06)	-0.23 (-0.27, -0.20)	-0.18 (-0.23, -0.14)	-0.61 (-0.64, -0.58)	-0.86 (-0.91, -0.81)

of $\tilde{C}_{X\to Y}$ may be used to study behaviour of the asymmetry coefficient under each given GEM. Further, density estimates of $\tilde{C}_{X\to Y}$ in Supplementary Figure 1 show the behaviour of the asymmetry coefficient under each NPGEM described in Section 4. In all six cases, the average $\tilde{C}_{X\to Y}$ reflects the induced asymmetry under each noise-free GEM model. These numerical results confirm the theoretical properties and insights in Section 4.

6.2 Coverage probability, bias and standard error of $\hat{C}_{X\to Y}$

We use a simulation experiment to show that our cross-fitting estimation and inference technique can help estimate the proposed asymmetry coefficient with small error and appropriate coverage probability. We simulate data from $X \sim f_X$ with a present entropy $H(X)$ and generate $Y = g(X)$ in a GEM via a bijective function g, with density f_Y and entropy

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

$H(Y)$. We intentionally set f_X and g in a way that the induced population-level parameter $C_{X\to Y} = H(X) - H(Y)$ is positive. We consider two cases that have both closed-form distributions so that the exact true asymmetry coefficient is known. (i) $X \sim$ Lognormal$(5, 1)$ and $g(t) := \log(t)$, which implies $Y \sim$ N$(5, 1)$ and the true coefficient $= 5$; and (ii) $X \sim$ Exp$(\mu = 1)$ and $g(t) := t^{2/3}$, which implies $Y \sim$ Weibull$(\text{scale} = 1, \text{shape} = 3/2)$ and the true coefficient $= 0.213$. Our simulation is set up as follows: we vary the sample size

Table 2: Examining $\hat{C}_{X\to Y}$: absolute bias (A.Bias), empirical standard error (ESE), the asymptotic SE (ASE) and coverage probability (CP) for different sample sizes $n \in \{250, 500, 750\}$ under two cases: (i) $X \sim$ Lognormal$(5, 1)$ and $Y \sim$ N$(5, 1)$ and (ii) $X \sim$ Exp$(\text{mean} = 1)$ and $Y \sim$ Weibull$(\text{scale} = 1, \text{shape} = 3/2)$.

	Case (I)			Case (II)		
	$n = 250$	$n = 500$	$n = 750$	$n = 250$	$n = 500$	$n = 750$
A.Bias	0.095	0.058	0.048	0.082	0.062	0.049
ESE	0.112	0.073	0.060	0.087	0.064	0.051
ASE	0.114	0.084	0.070	0.102	0.074	0.061
CP	0.945	0.980	0.965	0.935	0.935	0.960

$n \in \{250, 500, 750\}$; in each case, we obtain both estimate $\hat{C}_{X\to Y}$ and asymptotic variance $\hat{\sigma}^2_C$ according to the formula given in Corollary 1. Then, we construct a 95% asymptotic confidence interval, say $\left(\hat{L}_C, \hat{U}_C\right)$. Repeating this procedure over $R = 200$ times, we obtain a set of estimates $\left\{\hat{C}_{X\to Y}(r), \hat{\sigma}^2_C(r)\right\}_{r=1}^{R}$ as well as the corresponding 95% asymptotic confidence intervals, $\left\{\left(\hat{L}_C(r), \hat{U}_C(r)\right)\right\}_{r=1}^{R}$. Table 2 lists the empirical mean, absolute bias, and standard error of the asymmetry coefficient estimates as well as the average of the estimated asymptotic variances. In addition, we report the empirical coverage probability, defined as the proportion of 95% asymptotic confidence intervals covering the true asymmetry coefficient parameter. In all the cases considered, our proposed methodology yields an estimate that has low estimation error. The empirical standard error and average asymptotic stan-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

dard error are close to each other, confirming the validity of the asymptotic normality given in Corollary 1. More importantly, the coverage probability is close to the nominal 95% level in both scenarios, so the proposed cross-fitting inference is numerically stable and trustful. In summary, these two simulation experiments clearly confirms the large-sample theoretical results given Theorem 3 and Corollary 1.

Remark 9. We adopt a 50/50 split for our data, a common and robust choice recommended in the cross-fitting literature (Chernozhukov et al., 2018). This approach provides a balance, ensuring both data folds are sufficiently large for stable nuisance density estimation and subsequent parameter inference. To justify the use of a 50/50 data-splitting ratio, we conducted a simulation study for Cases (I) and (II) under varying sample sizes to assess the performance of the estimator under various unbalanced splits. The results, presented in supplementary Table 2, show that the 50/50 split is optimal in the sense that it yields the lowest absolute bias and a coverage probability that is closest to the nominal 95% level. As the split becomes more unbalanced (e.g., 70/30, 90/10), the bias steadily increases and the coverage probability deteriorates significantly. This provides strong empirical support for using the 50/50 split in our main analysis.

6.3 Comparison of performance of $\hat{C}_{X\to Y}$ with other causal discovery methods

We compare the accuracy power of our method against three competing methods. The competing methods are the Additive Noise Model (ANM) (Mooij et al., 2016), the Conditional Distribution Similarity (CDS) (Fonollosa, 2019), and the Regression Error-based Causal Inference (RECI) (Blöbaum et al., 2019) approaches. We use the Causal Discovery Toolbox (Kalainathan et al., 2020) to implement these methods, which are used with default

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

parameters supplied by the toolbox.

6.3.1 Simulated data

We consider uniformly distributed X and six choices of bijective g. We generate $Y = g(X)+\epsilon$, where $\mathbb{V}(\epsilon) = \sigma$ and $\text{Cov}(X, \epsilon) = \rho$. We draw samples of size $n = 1000$ repeatedly for $r = 250$ times and report the proportion of times each method is able to detect the correct directionality $X \to Y$ and present our findings in Table 3, which reveals that our method consistently outperform all other methods across all but one simulated data setting.

6.3.2 Benchmark data

We compared the accuracy of the four competing methods using the publicly available `CauseEffectPairs` benchmark that consists of data for $k = 99$ different cause-effect pairs selected from 37 data sets from various domains (Mooij et al., 2016). Of the 99 pairs, ANM was able to correctly detect 51 pairs and returned inconclusive findings for 18 pairs, while $\hat{C}_{X\to Y}$ was able to detect 58 pairs (under contracting dynamics) and 35 pairs (under expanding dynamics), and was inconclusive for 7 pairs. The CDS and RECI approaches proved more successful, correctly detecting 67 and 62 pairs respectively with no inconclusive findings.

7. Application

7.1 Introduction

We apply the above asymmetry analytic to study the epigenetic relation between DNA methylation ($DNAm$) and blood pressure (BP) that is one of the most important risk factors for cardiovascular disease (CVD). We use data from $n = 522$ children (including 247 boys

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Table 3: Comparing performance of $\hat{C}_{X\to Y}$ with competing pairwise causal discovery methods. We consider uniformly distributed X and six choices of bijective g. We generate $Y = g(X) + \epsilon$, where $\mathbb{V}(\epsilon) = \sigma$ and $\mathrm{Cov}(X, \epsilon) = \rho$. We draw samples of size $n = 1000$ repeatedly for $r = 250$ times and report the proportion of times each method is able to detect the correct directionality $X \to Y$.

$g(x)$	σ	ρ	Accuracy			
			$\hat{C}_{X\to Y}$	ANM	CDS	REI
$x^{1/3}$	0	0	**1.00**	0.22	0.00	0.00
	0.05	0.10	**1.00**	0.24	0.00	0.00
	0.05	0.60	**1.00**	0.21	0.00	0.00
$x^{1/2}$	0	0	**1.00**	0.18	0.00	0.00
	0.05	0.10	**1.00**	0.22	0.00	0.00
	0.05	0.60	**1.00**	0.18	0.00	0.00
x^2	0	0	**1.00**	0.24	**1.00**	**1.00**
	0.05	0.10	**1.00**	0.18	**1.00**	**1.00**
	0.05	0.60	**1.00**	0.20	**1.00**	**1.00**
x^3	0	0	**1.00**	0.20	0.00	**1.00**
	0.05	0.10	**1.00**	0.20	0.02	**1.00**
	0.05	0.60	**1.00**	0.18	0.00	**1.00**
$\exp(x)$	0	0	**1.00**	0.22	0.00	**1.00**
	0.05	0.10	**1.00**	0.20	0.44	**1.00**
	0.05	0.60	0.98	0.21	0.05	**1.00**
$\sin(\pi x/2)$	0	0	**1.00**	0.24	**1.00**	0.00
	0.05	0.10	**1.00**	0.26	**1.00**	0.00
	0.05	0.60	**1.00**	0.26	**1.00**	0.00

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

and 275 girls) aged 10-18 years in the Early Life Exposures in Mexico to Environmental Toxicants (*ELEMENT*) study (Hernandez-Avila et al., 1996). The primary task of scientific interest is to understand potential *CVD* causal pathways, part of which involves examining asymmetry between *DNAm* alterations and *CVD* risks. In this analysis, we examine six candidate genes found to be significantly associated with systolic blood pressure (*SBP*) and diastolic blood pressure (*DBP*) in at least 20 independent studies according to the NHGRI-EBI GWAS Catalog. These six target genes include **FGF5, HSD11B2, KCNK3, ATP2B1, ARHGAP42** and **PRDM8**. To investigate their epigenetic roles in *CVD*, we investigate whether β values of *DNAm* (specifically, cytosine-phosphate-guanine (CpG) methylation) influence change in *BP* or if the converse is true (DiCorpo et al., 2018; Hong et al., 2023).

7.2 Methods

To perform a gene-level analysis, we average beta-values of *DNAm* over CpG sites within each gene, which is then normalized across subjects by an affine transformation described in Section 2.1 to minimize any undue influence of location or scale changes in individual *DNAm* measurements. The same normalization procedure is applied to *SBP* and *DBP* measurements. Following normalization, we apply a rigorous three-step analysis to each gene-sex-*BP* triplet to infer the directional relationship between *DNAm* and *BP* (either *DBP* or *SBP*). First, we determine the plausible causal direction by verifying the functional orthogonality assumption (Assumption 1) for either potential GEMs: $DNAm = g(BP)$ and $\mathrm{BP} = h(DNAm)$ using bootstrap-based orthogonality deviation score described in Algorithm 1 for this purpose. The analysis proceeds only if the bootstrap confidence interval for the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

deviation score contains zero for exactly one of the two directions. If neither or both directions satisfy the assumption, the result is deemed inconclusive. Second, for the single direction that satisfies functional orthogonality, we characterize the dynamics of its generative function using a bootstrap procedure (see Remark 7 for more details). In brief, we estimate the average log-gradient of the function to classify it as either 'contracting' or 'expanding' dynamics. This step is crucial as it determines the specific null hypothesis to be tested; a contracting function implies a positive asymmetry coefficient ($C_{X\to Y} > 0$), whereas an expanding function implies a negative one ($C_{X\to Y} < 0$). Finally, having established a valid direction and the nature of the generative function, we proceed to the formal inference step of our hypothesized GEM. We employ the data-splitting and cross-fitting procedure to compute the asymmetry coefficient ($\hat{C}_{X\to Y}$) and its corresponding one-sided 95% confidence interval. This allows us to test the directionality hypothesis formulated in the previous step and draw a conclusion about the epigenetic pathway while performing diagnostic checks to verify if the underlying assumptions are satisfied.

7.3 Results

We applied our three-step analytical pipeline to investigate the directional relationship between *DNAm* and blood pressure (systolic - *SPB* and diastolic - *DBP*) for six candidate genes, stratifying the analysis by sex. Our findings reveal distinct directional pathways for several genes.

For genes **FGF5** and **HSD11B2**, the analysis yielded strong, consistent evidence for a directional influence from blood pressure to *DNAm*. In all tested scenarios—for both males and females, and for both *SBP* and *DBP*—the functional orthogonality assumption was

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

uniquely satisfied for the BP $\rightarrow$ *DNAm* direction. The generative functions were consistently identified as having contracting dynamics, and the resulting asymmetry coefficient was statistically significant, with the one-sided 95% confidence interval for $\hat{C}_{BP\rightarrow DNAm}$ being strictly positive.

The results for **KCNK3** and **ATP2B1** were more varied. For **KCNK3**, we identified a significant pathway from *SBP* to *DNAm* in males and from *DBP* to *DNAm* in females. However, the relationship between *SBP* and *DNAm* in females was inconclusive, as the orthogonality assumption was violated. For **ATP2B1**, significant pathways from both *SBP* and *DBP* to *DNAm* were found in males. In females, the *SBP*-*DNAm* relationship was inconclusive due to a similar assumption violation, and no significant directionality was found with *DBP*.

Next, for genes **ARHGAP42** and **PRDM8**, our analysis did not detect any significant directional asymmetry. Across all sex and blood pressure combinations for these two genes, the estimated asymmetry coefficient was not statistically distinguishable from zero, providing no evidence to support a directional relationship in either direction.

Finally, given that noise contamination can particularly affect inference for GFs with contracting dynamics, we also performed a diagnostic check to ensure that the estimated signal-to-noise ratio for each significant pathway was well within the theoretical tolerance bound established in Section 4. Our investigations reveal that the noise perturbation has little influence on the positive results drawn from our inference method.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

7.4 Discussion

Our inference method unveils novel pathways from *BP* to *DNAm* for two genes **FGF5** and **HSD11B2**. **FGF5** is reported in Domouzoglou et al. (2015) as a key gene in several animal studies that increases myocardial blood flow and function, decrease myocyte apoptosis, and increase myocyte number after gene transfer of the growth factor. However, there are no available results of its effects in human cardiovascular disease. Our analysis is the first to unveil a pathway from changes in *BP* to regulate *DNAm* of this key *CVD* gene in a human cohort study. Gene **HSD11B2** was reported in Rahman et al. (2011) to be associated with obesity-related cardiovascular risk factors, e.g. type II diabetes and hypertension. Hypertension is often associated with chronic inflammation and oxidative stress, both of which can trigger an imbalance in the immune system.This can, in turn, modulate a person's methylation patterns. In summary, these new findings confer an added sense of directionality in the study of *BP* variation and epigenetic biomarkers, paving the way for future advancements in individualized risk assessments and even therapeutic targets.

8. Concluding remarks

Asymmetry is an inherent property of bivariate associations and therefore must not be ignored (Zheng et al., 2012). In this paper, we present a new methodology in the framework of generative exposure mapping (GEM) models so that the induced asymmetry between two random variables X and Y is captured by an information-theoretic coefficient $C_{X\to Y}$. Utilizing this asymmetry measure, we develop a cross-fitting inference that enables us to examine a certain hypothesized directionality with proper uncertainty quantification. The intrinsic linkage between information theory and an experiment with a uniformly distributed

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

exposure provides an interesting insight into asymmetry from the perspective of causality, where we argue that the induced asymmetry from a GEM model may be regarded as a low-dimensional imprint of causality. Further, simulation studies reveal that the asymmetry measure $C_{X\to Y}$ can detect different dynamics of outcomes under different GEMs without actually estimating the mapping functions, and such capacity prevails in noise-perturbed GEM models. It is worth noting that such an extension requires one assumption of the contamination error being independent of the exposure, although empirical studies show robustness of the asymmetry coefficient to endogeneity in contaminated GEM models. A potential area of future work would involve addressing endogeneity in contaminated GEM models with correlations between the errors and exposure (Breunig and Burauel, 2024), in which the instrumental variable technique is worth exploring. This extension would require identifying a valid instrument W that is correlated with X but independent of the error ϵ. In such a setup, we would first use the instrument W to model X, analogous to the "first stage" of a Two-Stage Least Squares regression (Imbens and Angrist, 1994). Hence, one could isolate the "clean" component of the exposure (i.e., $X_{IV} = f(W)$) and redefine the asymmetry coefficient using this "clean" component, for instance, as $C^{IV}_{X\to Y} = H(X_{IV}) - H(Y)$. This IV-based approach would, however, require significant new theoretical development to extend our current framework, and to derive the statistical properties of this new coefficient.

Another important extension is the incorporation of covariates (or modifiers), $\boldsymbol{Z}$, into the GEM framework. This presents several challenges that future work must address. First, it will require theoretical effort to extend the concept of functional orthogonality (as formalized by Assumption 1) to a conditional setting. The simple independence between the input density and the generative function must be re-conceptualized given $\boldsymbol{Z}$. Consequently, a

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

new diagnostic tool would be needed to verify this conditional orthogonality. The current "orthogonality deviation score" (see Algorithm 1) is only defined for the bivariate case and would be insufficient. Finally, the asymmetry coefficient itself would need to be refined. A promising approach would be to develop a "conditional asymmetry coefficient", likely based on conditional entropies, to properly isolate the "low-dimensional imprint of causality" after accounting for the influence of $\boldsymbol{Z}$.

Several further applications and extensions of the proposed $C_{X\to Y}$ measure and cross-fitting inference method are worth exploring, including the extension of GEM-induced asymmetry in functional or longitudinal variables; for example, the found asymmetries between $DNAm$ and BP in our data analysis would become more reliable if repeated measurements of BP are available. In conclusion, our framework may be used either as a discovery tool, thereby aiding practitioners in improving the rigor and reproducibility of scientific research. We have made our implementation available as an open-source Python package with complete documentation at https://github.com/soumikp/causalGEM.

Supplementary Materials

The online Supplementary Material contains the following sections:

- Section I: Proofs.
- Section II: Behavior of $\hat{C}_{X\to Y}$ in NPGEMs.
- Section III: Methylation Data Application.
- Section IV: Data-splitting and Cross-fitting.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

- Section V: Resolving ambiguity in causal direction $X \to Y$ or $Y \to X$ when nature of generative function is unknown.

- Section VI: Diagnostics.

Acknowledgements

Peter X.-K. Song's research was supported by NSF DMS-2113564 and NIH R01ES033656, and Soumik Purkayastha's research was supported by the University of Michigan Rackham Predoctoral Fellowship. This research was further supported in part by the University of Pittsburgh Center for Research Computing and Data, RRID:SCR022735, through the resources provided. Specifically, the use of the HTC cluster was supported by NIH award number S10OD028483.

References

Aronow, P. M. and C. Samii (2017). Estimating average causal effects under general interference, with application to a social network experiment. *Ann. of App. Stat. 11*, 1912–1947.

Audibert, J.-Y. and A. B. Tsybakov (2007, April). Fast learning rates for plug-in classifiers. *The Annals of Statistics 35*(2), 608 – 633.

Bernacchia, A. and S. Pigolotti (2011). Self-consistent method for density estimation. *J. Roy. Stat. Soc.: Series B 73*, 407–422.

Blöbaum, P., D. Janzing, T. Washio, S. Shimizu, and B. Schölkopf (2019, January). Analysis of cause-effect inference by comparing regression errors. *PeerJ Computer Science 5*, e169.

Breunig, C. and P. Burauel (2024). Testability of reverse causality without exogenous variation.

Chatterjee, S. (2020). A new coefficient of correlation. *J. Amer. Stat. Assoc. 116*, 2009–2022.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*, 1–68.

Choi, J., R. Chapkin, and Y. Ni (2020). Bayesian causal structural learning with zero-inflated poisson bayesian networks. In *Proc. of the 33rd Int. Conf. on Neur. Inf. Proc. Sys.*, pp. 5887–5897.

Cover, T. M. (2005). *Elements of Information Theory*. India: John Wiley & Sons.

Cox, D. R. (1990). Role of models in statistical analysis. *Stat. Science 5*, 169–174.

Cox, D. R. (1992). Causality: some statistical aspects. *J. Roy. Stat. Soc.: Series A 155*, 291–301.

Daniušis, P., D. Janzing, J. Mooij, J. Zscheischler, B. Steudel, K. Zhang, and B. Schölkopf (2010). Inferring deterministic causal relations. In *Proc. of the 26th Conf. on Uncertainty in AI*, pp. 143–150.

DiCorpo, D. A., S. Lent, W. Guan, M.-F. Hivert, J. S. Pankow, C. E. W. Group, and C. D. W. Group (2018). Mendelian randomization suggests causal influence of glycemic traits on dna methylation. *Diabetes 67*, 1707–P.

Domouzoglou, E. M., K. K. Naka, A. P. Vlahos, M. I. Papafaklis, L. K. Michalis, A. Tsatsoulis, and E. Maratos-Flier (2015). Fibroblast growth factors in cardiovascular disease: The emerging role of fgf21. *Am. J. Physiol. Heart. Circ. Physiol. 309*, 1029–1038.

Fonollosa, J. A. R. (2019). *Conditional Distribution Variability Measures for Causality Detection*, pp. 339–347. Springer International Publishing.

Hannig, J., H. Iyer, R. C. S. Lai, and T. C. M. Lee (2016). Generalized fiducial inference: A review and new results. *J. Amer. Stat. Assoc. 111*, 1346–1361.

Hernandez-Avila, M., T. Gonzalez-Cossio, E. Palazuelos, I. Romieu, A. Aro, E. Fishbein, K. E. Peterson, and H. Hu (1996). Dietary and environmental determinants of blood and bone lead levels in lactating postpartum women living in mexico city. *Env. Health Persp. 104*, 1076–1082.

Hong, X., K. Miao, W. Cao, J. Lv, C. Yu, T. Huang, D. Sun, C. Liao, Y. Pang, Z. Pang, et al. (2023). Association between dna methylation and blood pressure: a 5-year longitudinal twin study. *Hypertension 80*, 169–181.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Hoyer, P., D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf (2008). Nonlinear causal discovery with additive noise models. In *Proc. of the 21st Int. Conf. on Neur. Inf. Proc. Sys.*, pp. 689–696.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Janzing, D., J. Mooij, K. Zhang, J. Lemeire, J. Zscheischler, P. Daniusis, B. Steudel, and B. Scholkopf (2012). Information-geometric approach to inferring causal directions. *Artif. Int. 182*, 1–31.

Janzing, D. and B. Schölkopf (2010). Causal inference using the algorithmic markov condition. *IEEE Transactions on Information Theory 56*(10), 5168–5194.

Kalainathan, D., O. Goudet, and R. Dutta (2020). Causal discovery toolbox: Uncovering causal relationships in python. *Journal of Machine Learning Research 21*(37), 1–5.

Manski, C. F. (2013, February). Identification of treatment response with social interactions. *The Econometrics Journal 16*, S1–S23.

Mooij, J. M., J. Peters, D. Janzing, J. Zscheischler, and B. Schölkopf (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *J. Mach. Lear. Res. 17*, 1–102.

Moon, Y.-I., B. Rajagopalan, and U. Lall (1995, September). Estimation of mutual information using kernel density estimators. *Physical Review E 52*(3), 2318–2321.

Ni, Y. (2022). Bivariate causal discovery for categorical data via classification with optimal label permutation. In *Proc. of the 35th Int. Conf. on Neur. Inf. Proc. Sys.*, pp. 10837–10848.

O'Brien, T. A., K. Kashinath, N. R. Cavanaugh, W. D. Collins, and J. P. O'Brien (2016). A fast and objective multidimensional kernel density estimation method: fastKDE. *Comp. Stat. & Data Anal. 101*, 148–160.

Orlitsky, A. (2003). Information theory. In *Encyclopedia of Physical Science and Technology*, pp. 751–769. Elsevier.

Pearl, J. (2009). *Causality: Models, reasoning and inference*. Cambridge University Press, England.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0236

Purkayastha, S. and P. X.-K. Song (2024). fastMI: A fast and consistent copula-based nonparametric estimator of mutual information. *Journal of Multivariate Analysis 201*, 105270.

Rahman, T. J., B. M. Mayosi, D. Hall, P. J. Avery, P. M. Stewart, J. M. C. Connell, H. Watkins, and B. Keavney (2011). Common variation at the 11-β hydroxysteroid dehydrogenase type 1 gene is associated with left ventricular mass. *Circulation: Cardiovascular Genetics 4*, 156–162.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. J. 27*, 379–423.

Tagasovska, N., V. Chavez-Demoulin, and T. Vatter (2020). Distinguishing cause from effect using quantiles: Bivariate quantile causal discovery. In *Proc. of the 37th Int. Conf. on Mach. Lear.*, pp. 9311–9323.

Zheng, S., N.-Z. Shi, and Z. Zhang (2012). Generalized measures of correlation for asymmetry, nonlinearity, and beyond. *J. Amer. Stat. Assoc. 107*, 1239–1252.