Quantification and Inference of Asymmetric Relations Under Generative Exposure Mappings

Soumik Purkayastha and Peter X.-K. Song

University of Pittsburgh and University of Michigan

Supplementary Material

This supplement contains proofs of all theoretical results stated in the main manuscript (Section S1), additional simulation results on the behaviour of $\hat{C}_{X\to Y}$ in noise-perturbed GEMs (Section S2), additional material for the methylation data application (Section S3), details on data-splitting and cross-fitting inference including estimation of $\hat{\sigma}^2_C$ (Section S4), proofs and discussion resolving the ambiguity in causal direction when the nature of the generative function is unknown (Section S5), and diagnostics for determining generative function dynamics (Section S6).

S1 Proofs

I. Proof of Theorem 1

Proof. Note that de Bruijn's identity (Barron, 1986, Proof of Lemma 1) states that $\frac{\delta H(Y+\sqrt{t}Z)}{\delta t} = \frac{I(Y+\sqrt{t}Z)}{2}$. In other words, $H(Y+\sqrt{t}Z)$ is an increasing function of t as $I(Y+\sqrt{t}Z) > 0$ for all $t > 0$. Hence, there exists

σ', such that $H(Y+\epsilon) \leq H(Y+\sqrt{\sigma'}Z)$. Further, from the fundamental theorem of calculus, note that

$$H(Y+\sqrt{\sigma'}Z) - H(Y) = \frac{1}{2}\int_0^{\sigma'} I(Y+\sqrt{t'}Z)dt.$$

We use convolution inequality of Fisher information (Zamir, 1998, Theorem 1) given by:

$$I(Y+\sqrt{t}Z) \leq \frac{I(Y)I(\sqrt{t}Z)}{I(Y)+I(\sqrt{t}Z)},$$

and note that $I(\sqrt{t}Z) = 1/t$ to obtain the following inequality:

$$H(Y+\sqrt{\sigma'}Z) \leq H(Y) + \frac{1}{2}\log(\sigma' I(Y)+1).$$

□

II. Proof of Theorem 3

Proof. From Theorem 2, for any small ϵ, there exists sufficiently large n such that

$$\left|\hat{f}_{X;1}(x) - f_X(x)\right| < \epsilon, \text{ and } \left|\hat{f}_{X;2}(x) - f_X(x)\right| < \epsilon, \ x \in \mathbb{R}.$$

Next, note that if $|x-a| < a$ for $a > 0$, i.e., if $|x/a - 1| < 1$, the following Taylor series expansion holds:

$$\begin{aligned}\ln(x) - \ln(a) &= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{na^n}(x-a)^n \\ &= \frac{1}{a}(x-a) - \frac{1}{2a^2}(x-a)^2 + \frac{1}{3a^3}(x-a)^3 + \cdots\end{aligned}$$

Since, from Theorem 2, $\hat{f}_X$ is uniformly consistent for f_X, we neglect quadratic and higher terms in the above expression and write the following approximations for $j \in 1, \ldots, n$:

$$\begin{aligned}\ln\left(\hat{f}_{X;1}(X_{n+j})\right) &= \ln\left(f_{X;1}(X_{n+j})\right) + \frac{\hat{f}_{X;1}(X_{n+j}) - f_{X;1}(X_{n+j})}{f_{X;1}(X_{n+j})}, \\ \ln\left(\hat{f}_{X;2}(X_j)\right) &= \ln\left(f_{X;2}(X_j)\right) + \frac{\hat{f}_{X;2}(X_j) - f_{X;2}(X_j)}{f_{X;2}(X_j)}.\end{aligned}$$

By Assumption 5, f_X is bounded below by a constant, say B^{-1}, on its support, it is easy to show that

$$\begin{aligned}\left|\hat{H}_1(X) - H_{0;1}(X)\right| &\leq \frac{1}{n}\sum_{j=1}^{n}\left|\frac{\hat{f}_{X;2}\left(X_j\right) - f_X\left(X_j\right)}{f_X\left(X_j\right)}\right| \leq \frac{\epsilon}{B}, \\ \left|\hat{H}_2(X) - H_{0;2}(X)\right| &\leq \frac{1}{n}\sum_{j=1}^{n}\left|\frac{\hat{f}_{X;1}\left(X_{n+j}\right) - f_X\left(X_{n+j}\right)}{f_X\left(X_{n+j}\right)}\right| \leq \frac{\epsilon}{B},\end{aligned}$$

which together imply $\left|\hat{H}(X) - H_0(X)\right| \overset{a.s.}{\to} 0$, as $n \to \infty$. Further, the strong law of large numbers implies as $n \to \infty$, we have $H_0(X) - H(X) \overset{a.s.}{\to} 0$.

Hence, using the inequality obtained above in conjunction with the strong law of large numbers yields $\hat{H}(X) - H(X) \xrightarrow{a.s.} 0$ as $n \to \infty$. Similar results hold for $\hat{H}_Y$; thus, invoking the continuous mapping theorem, we are able to show $\hat{C}_{X\to Y} \xrightarrow{a.s.} C_{X\to Y}$ as $n \to \infty$. This concludes the proof. □

III. Proof of Theorem 4

Proof. We consider the following Taylor series expansion:

$$\begin{aligned}\sqrt{n}\left\{\hat{H}_2(X) - H_{0;2}(X)\right\} &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{\hat{f}_{X;1}(X_{n+j}) - f_{X;1}(X_{n+j})}{f_{X;1}(X_{n+j})}\right\} + o_{\mathcal{P}}(n^{-1/2}) \\ &= \frac{S_n(2)}{\sqrt{n}} + o_{\mathcal{P}}(n^{-1/2}).\end{aligned}$$

We now show that the leading term $S_n(2)/\sqrt{n} \xrightarrow{\mathcal{P}} 0$ as $n \to 0$, which will establish

$$\sqrt{n}\left\{\hat{H}_2(X) - H_{0;2}(X)\right\} \xrightarrow{\mathcal{P}} 0, \quad \text{as } n \to 0.$$

Using identical arguments, we establish similar results for $\hat{H}_2(Y)$ as well. It is sufficient to show $\mathbb{E}\left[\{S_n(2)/\sqrt{n}\}^2\right] \to 0$ as $n \to \infty$. Note that $\mathbb{E}\left[\{S_n(2)/\sqrt{n}\}^2\right] = \mathbb{E}^2\left[\{S_n(2)/\sqrt{n}\}\right] + \mathbb{V}\left[\{S_n(2)/\sqrt{n}\}\right]$. First, we prove

$\mathbb{E}\left[\{S_n(2)/\sqrt{n}\}\right]$:

$$\begin{aligned}
\mathbb{E}\left\{\frac{S_n(2)}{\sqrt{n}}\right\} &= \mathbb{E}_{\mathcal{D}_1}\left[\mathbb{E}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\frac{S_n(2)}{\sqrt{n}}\middle|\mathcal{D}_1\right\}\right] \\
&= \mathbb{E}_{\mathcal{D}_1}\left[\mathbb{E}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{\hat{f}_{X;1}(X_{n+j})-f_X(X_{n+j})}{f_X(X_{n+j})}\right)\middle|\mathcal{D}_1\right\}\right] \\
&= \sqrt{n}\,\mathbb{E}_{\mathcal{D}_1}\left[\mathbb{E}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\left(\frac{\hat{f}_{X;1}(X)-f_X(X)}{f_X(X)}\right)\middle|\mathcal{D}_1\right\}\right],
\end{aligned}$$

where the inner expectation term is evaluated as follows:

$$\mathbb{E}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\left(\frac{\hat{f}_{X;1}(X)-f_X(X)}{f_X(X)}\right)\middle|\mathcal{D}_1\right\} = \int_{\mathbb{R}}\left(\frac{\hat{f}_{X;1}(x)-f_X(x)}{f_X(x)}\right)f_X(x)dx = 0,$$

where the equality holds since $\hat{\phi}_1$ is the Fourier transform associated with the optimal density function estimator $\hat{f}_{X;1}$, and we know $\int_{\mathbb{R}}\hat{f}_{X;1}(x)dx = \hat{\phi}_1(0) = 1$ and consequently, $\mathbb{E}\left[\{S_n(2)/\sqrt{n}\}\right] = 0$. Next, we consider the term $\mathbb{V}\left[\{S_n(2)/\sqrt{n}\}\right]$:

$$\mathbb{V}\left[\left\{\frac{S_n(2)}{\sqrt{n}}\right\}\right] = \mathbb{E}_{\mathcal{D}_1}\left[\mathbb{V}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\left(\frac{S_n(2)}{\sqrt{n}}\right)\middle|\mathcal{D}_1\right\}\right] + \mathbb{V}_{\mathcal{D}_1}\left[\mathbb{E}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\left(\frac{S_n(2)}{\sqrt{n}}\right)\middle|\mathcal{D}_1\right\}\right],$$

where the second term is zero, as per our calculations above. Conditional on $\mathcal{D}_1$, the terms $\hat{f}_{X;1}(X_{n+j})$ are independent and identically distributed for

$1 \leq j \leq n$. We have:

$$\begin{aligned}
\mathbb{V}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\left(\frac{S_n(2)}{\sqrt{n}}\right)\middle|\mathcal{D}_1\right\} &= \mathbb{V}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{\hat{f}_{X;1}(X_{n+j})-f_X(X_{n+j})}{f_X(X_{n+j})}\right)\middle|\mathcal{D}_1\right\} \\
&= \mathbb{V}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\left(\frac{\hat{f}_{X;1}(X)-f_X(X)}{f_X(X)}\right)\middle|\mathcal{D}_1\right\} \\
&= \mathbb{E}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\left(\frac{\hat{f}_{X;1}(X)-f_X(X)}{f_X(X)}\right)^2\middle|\mathcal{D}_1\right\},
\end{aligned}$$

The last equality follows from

$$\mathbb{E}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\left(\hat{f}_{X;1}(X)-f_X(X)\right)/f_X(X)\middle|\mathcal{D}_1\right\}=0.$$

Moreover,

$$\mathbb{E}_{\mathcal{D}_2|\mathcal{D}_1}\left\{\left(\frac{\hat{f}_{X;1}(X)-f_X(X)}{f_X(X)}\right)^2\middle|\mathcal{D}_1\right\} \leq B\int_{\mathbb{R}}\left(\hat{f}_{X;1}(x)-f_X(x)\right)^2dx,$$

where B is a (positive) lower bound for the density f_X over its support. Consequently, we get

$$\begin{aligned}
\mathbb{V}\left[\left\{S_n(2)/\sqrt{n}\right\}\right] &\leq B\times\mathbb{E}_{\mathcal{D}_1}\left\{\int_{\mathbb{R}}\left(\hat{f}_{X;1}(x)-f_X(x)\right)^2dx\right\} \\
&= B\times MISE(\hat{f}_{X;1},f_X).
\end{aligned}$$

Bernacchia and Pigolotti (2011) present an expression of $MISE$ in terms of the optimal kernel and prove that the last expression goes to zero as

sample size increases, i.e., $MISE(\hat{f}_{X;1}, f_X) \to 0$ as $n \to \infty$. This allows us to claim $\sqrt{n}\left(\hat{H}_2(X) - H_{0;2}(X)\right) \xrightarrow{\mathcal{P}} 0$ as $n \to \infty$. Note that the arguments presented above are generally valid for any true density function that is bounded away from zero and infinity on its support. Hence, they can also be used to establish similar results involving $\sqrt{n}\left(\hat{H}_2(Y) - H_{0;2}(Y)\right) \xrightarrow{\mathcal{P}} 0$ as $n \to \infty$.

Note that we can interchange the roles of $\mathcal{D}_1$ and $\mathcal{D}_2$ in the proof above to arrive at

$$\sqrt{n}\begin{pmatrix} \hat{H}_1(X) - H_{0;1}(X) \\ \hat{H}_1(Y) - H_{0;1}(Y) \end{pmatrix} \xrightarrow{\mathcal{P}} 0, \text{ as } n \to \infty,$$

which can be combined with the result we have obtained, given by:

$$\sqrt{n}\begin{pmatrix} \hat{H}_2(X) - H_{0;2}(X) \\ \hat{H}_2(Y) - H_{0;2}(Y) \end{pmatrix} \xrightarrow{\mathcal{P}} 0, \text{ as } n \to \infty,$$

and the continuous mapping theorem to yield the following:

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

This concludes the proof. □

IV. Proof of Corollary 1

Proof. Using Theorem 4 and Lemma 2 in conjunction with Slutsky's theorem, we get

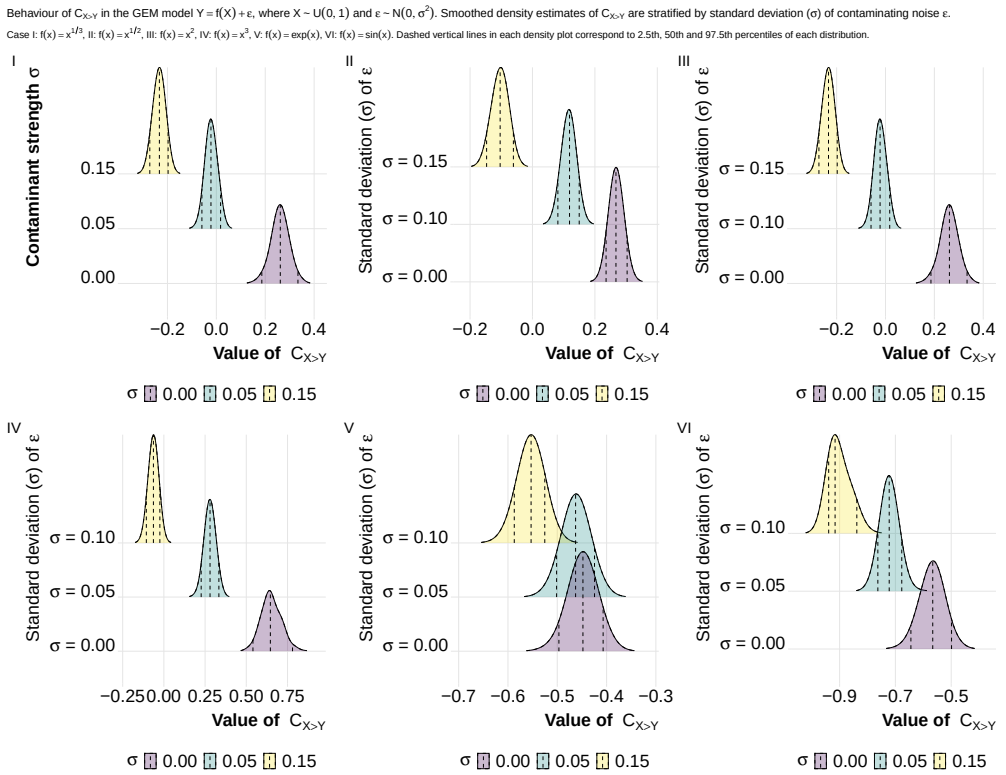
$$\sqrt{n}\begin{pmatrix}\hat{H}(X)-H(X)\\ \hat{H}(Y)-H(Y)\end{pmatrix} \xrightarrow{\mathcal{D}} N\left(\mathbf{0},\Sigma\right), \text{ as } n\to\infty,$$

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

$$\sqrt{n}\left(\hat{C}_{X\to Y}-C_{X\to Y}\right) \xrightarrow{\mathcal{D}} N\left(0,\sigma_C^2\right), \text{ as } n\to\infty.$$

This concludes the proof. □

S2 Behaviour of $\hat{C}_{X\to Y}$ in *NPGEM*s



Supplementary Figure 1: Examining the behaviour of $\hat{C}_{X\to Y}$ under possibly contaminated *GEM*s.

S3 Additional material for application to methylation data

See Supplementary Figure 2 for cross-inference results on all six genes.

We applied our three-step analytical pipeline to investigate the directional

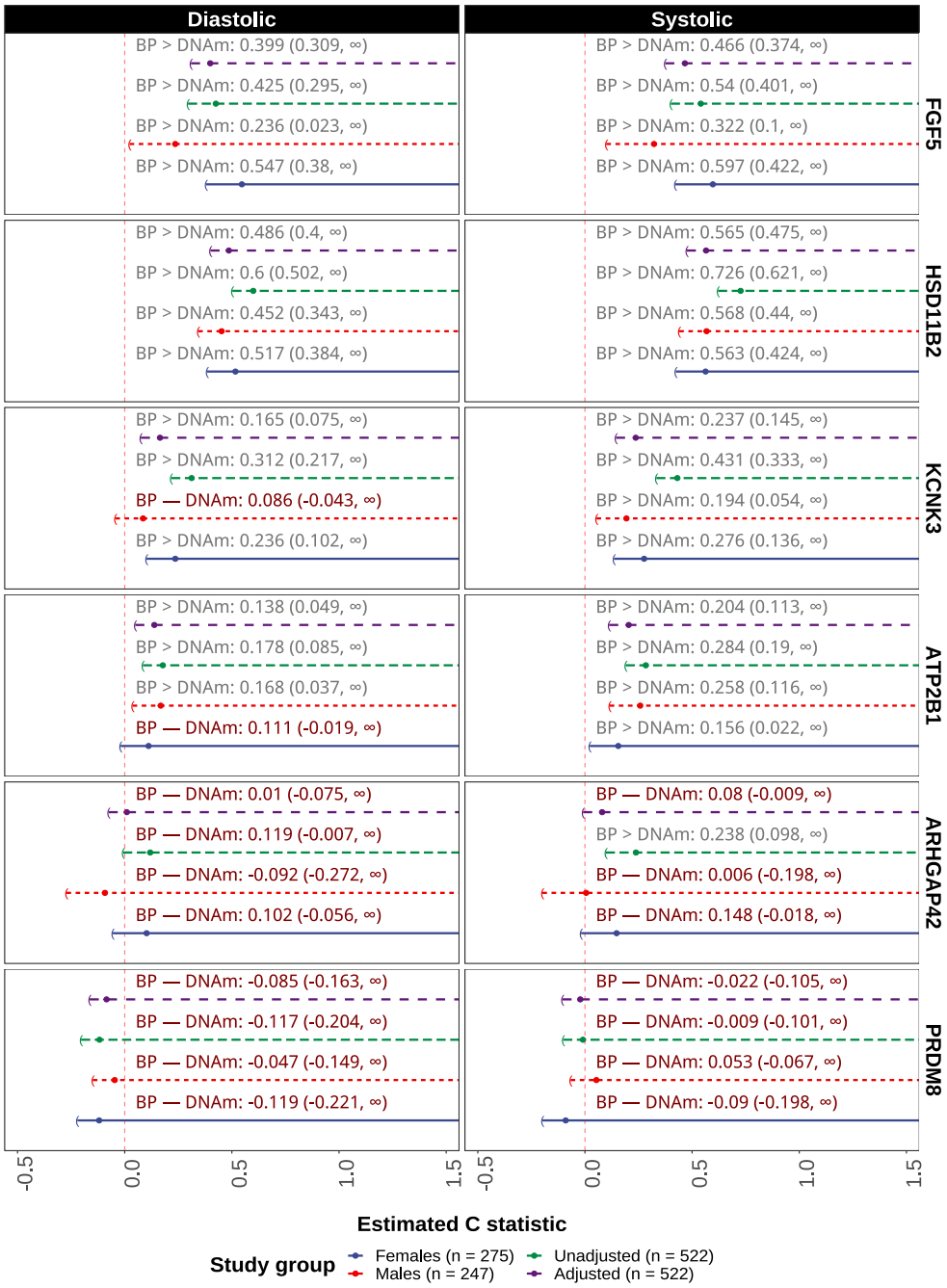
relationship between *DNAm* and blood pressure (systolic *SBP* and diastolic *DBP*) for six candidate genes, stratifying the analysis by sex. Our findings reveal distinct directional pathways for several genes. We summarise our major findings here:

1. 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 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 $\hat{C}_{\text{BP}\rightarrow\text{DNAm}}$ being strictly positive.

2. 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.

3. Finally, 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.

To empirically check if the tolerance of normally distributed noise contamination on the *BP* measurements is controlled by the bound of Equation (4.5) in the main paper, we fitted a linear generalised additive model (*GAM*) with linear splines to estimate $\hat{g}$. Using the fitted model and residuals, we can estimate $\hat{I}(\hat{Y})$ and $\hat{C}_{X\to Y} = \hat{H}(X) - \hat{H}(\hat{Y})$, and then obtained the estimated critical value $\hat{\sigma}_{CRIT} = \left(\exp(2\hat{C}_{X\to Y}) - 1\right)/\hat{I}(\hat{Y})$ as well as the residual variance $\hat{\sigma}$. We present a tabular summary of the bootstrapped distribution of $\hat{\sigma} - \hat{\sigma}_{CRIT}$ in Table 1. According to Section 4 of the main article, our method correctly captures the direction induced by the *NPGEM* so long as $\hat{\sigma}$ and $\hat{\sigma}_{CRIT}$ are comparable. We implement a bootstrap approach to test whether σ and σ_{CRIT} are comparable and present our findings in Table 1. This approach is *not* a rigorous test, but rather a diagnostic tool to examine noise contamination relative to signal in the data. From Table 1 we see that in most cases the estimated $\hat{\sigma}$ and $\hat{\sigma}_{CRIT}$ are comparable since most bootstrap-based 95% CIs of $\hat{\sigma}_{CRIT} - \hat{\sigma}$ contain zero. There are



Supplementary Figure 2: Cross-fitting inference results for asymmetry in BP and $DNAm$ include unadjusted $\hat{C}_{\mathrm{BP}\to\mathrm{DNAm}}$ along with sex-adjusted $\hat{C}_{\mathrm{BP}\to\mathrm{DNAm}|\mathrm{sex}}$ for six candidate genes.

only a few cases where there is some evidence to reverse the inequality of $\sigma \leq \sigma_{CRIT}$, namely in $ARHGAP42$ for males, $KCNK3$ for males, and finally in $PRDM8$ for males. Although the significant 95% CIs are very close to zero, our findings imply that we exercise caution with examining asymmetry in an $NPGEM$ framework using the strong asymmetry coefficient.

S4 Data-splitting, cross-fitting inference, and estimation of $\hat{\sigma}_C^2$

Using Theorem 4 and Lemma 2, we get

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

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

$$\sqrt{n}\left(\hat{C}_{X\to Y} - C_{X\to Y}\right) \xrightarrow{\mathcal{D}} N\left(0, \sigma_C^2\right), \text{ as } n \to \infty.$$

Note that $\sigma_C^2 := \sigma_{11} + \sigma_{22} - 2\sigma_{12}$, where σ_{ij} is the $(i,j)^{th}$ element of the covariance matrix given by Σ. To obtain $\hat{\Sigma}$ we use the same data-splitting and cross-fitting technique described in Section 5.2 of the main article. For

Gene	BP	Group	95% bootstrap CI for $\hat{\sigma}_{CRIT} - \hat{\sigma}$
$FGF5$	Systolic	Females	(-0.138, 0.268)
		Males	(-0.170, 0.001)
		Combined	(-0.145, 0.071)
	Diastolic	Females	(-0.140, 0.549)
		Males	(-0.177, 0.060)
		Combined	(-0.162, 0.173)
$HSD11B2$	Systolic	Females	(-0.105, 0.528)
		Males	(-2.611, 1.671)
		Combined	(-3.400, 3.498)
	Diastolic	Females	(-0.101, 0.691)
		Males	(-0.162, 0.277)
		Combined	(-3.400, 3.498)
$ARHGAP42$	Systolic	Females	(-0.827, 0.262)
		Males	(-0.191, -0.101)*
		Combined	(-0.175, -0.008)*
	Diastolic	Females	(-0.896, 0.523)
		Males	(-0.226, -0.026)*
		Combined	(-0.566, 0.186)
$ATP2B1$	Systolic	Females	(-0.284, 0.163)
		Males	(-0.184, 0.160)
		Combined	(-1.156, 0.121)
	Diastolic	Females	(-0.893, 0.104)
		Males	(-1.063, 0.352)
		Combined	(-2.279, 0.261)
$KCNK3$	Systolic	Females	(-0.195, 0.094)
		Males	(-0.195, -0.094)*
		Combined	(-0.161, 0.261)
	Diastolic	Females	(-0.353, 0.924)
		Males	(-0.186, -0.004)*
		Combined	(-1.287, 0.625)
$PRDM8$	Systolic	Females	(-0.750, 0.305)
		Males	(-0.214, -0.091)*
		Combined	(-0.253, -0.118)*
	Diastolic	Females	(-0.285, 0.361)
		Males	(-0.240, -0.024)*
		Combined	(-0.364, 0.028)

Table 1: Comparing empirical distributions of bootstrapped $\hat{\sigma}_{CRIT} - \hat{\sigma}$ estimates. Cases marked by * indicate $\hat{\sigma} > \hat{\sigma}_{CRIT}$.

ease of exposition, we repeat the details of our estimation and inference technique below.

Data-splitting

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 know 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:

$$\mathcal{D}_1 := \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \text{ and } \mathcal{D}_2 := \{(X_{n+1}, Y_{n+1}), \ldots, (X_{2n}, Y_{2n})\}.$$

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 of the main article. The estimated density functions are evaluated for data belonging to the second data split $\mathcal{D}_2$ to obtain the following estimates of marginal entropies:

$$\widehat{H_2}(X) = -\frac{1}{n}\sum_{j=1}^{n} \ln\left(\hat{f}_{X;1}\left(X_{n+j}\right)\right), \text{ and } \widehat{H_2}(Y) = -\frac{1}{n}\sum_{j=1}^{n} \ln\left(\hat{f}_{Y;1}\left(Y_{n+j}\right)\right).$$

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:

$$\widehat{H_1}(X) = -\frac{1}{n}\sum_{j=1}^{n} \ln\left(\hat{f}_{X;2}\left(X_j\right)\right), \text{ and } \widehat{H_1}(Y) = -\frac{1}{n}\sum_{j=1}^{n} \ln\left(\hat{f}_{Y;2}\left(Y_j\right)\right).$$

In anticipation of estimating the variance of $\sqrt{n}\left(\hat{C}_{X\to Y} - C_{X\to Y}\right)$ later we define the following quantities:

$$\hat{\Sigma}(1) := \begin{Bmatrix} \hat{\sigma}_{11}(1) & \hat{\sigma}_{12}(1) \\ & \hat{\sigma}_{22}(1) \end{Bmatrix}, \qquad \hat{\Sigma}(2) := \begin{Bmatrix} \hat{\sigma}_{11}(2) & \hat{\sigma}_{12}(2) \\ & \hat{\sigma}_{22}(2) \end{Bmatrix},$$

where the elements of $\hat{\Sigma}(1)$ are given by:

$$\hat{\sigma}_{11}(1) := \frac{1}{n}\sum_{j=1}^{n}\left\{-\log\left(\hat{f}_{X;2}\left(X_j\right)\right) - \hat{H}_1(X)\right\}^2,$$
$$\hat{\sigma}_{22}(1) := \frac{1}{n}\sum_{j=1}^{n}\left\{-\log\left(\hat{f}_{Y;2}\left(Y_j\right)\right) - \hat{H}_1(Y)\right\}^2,$$
$$\hat{\sigma}_{12}(1) := \frac{1}{n}\sum_{j=1}^{n}\left\{-\log\left(\hat{f}_{X;2}\left(X_j\right)\right) - \hat{H}_1(X)\right\}\left\{-\log\left(\hat{f}_{Y;2}\left(Y_j\right)\right) - \hat{H}_1(Y)\right\}.$$

Similarly, we define the elements of $\hat{\Sigma}(2)$ as follows:

$$\hat{\sigma}_{11}(2) := \frac{1}{n}\sum_{j=1}^{n}\left\{-\log\left(\hat{f}_{X;1}\left(X_{n+j}\right)\right) - \hat{H}_1(X)\right\}^2,$$
$$\hat{\sigma}_{22}(2) := \frac{1}{n}\sum_{j=1}^{n}\left\{-\log\left(\hat{f}_{Y;1}\left(Y_{n+j}\right)\right) - \hat{H}_1(Y)\right\}^2,$$
$$\hat{\sigma}_{12}(2) := \frac{1}{n}\sum_{j=1}^{n}\left\{-\log\left(\hat{f}_{X;1}\left(X_{n+j}\right)\right) - \hat{H}_1(X)\right\}\left\{-\log\left(\hat{f}_{Y;1}\left(Y_{n+j}\right)\right) - \hat{H}_1(Y)\right\}.$$

Cross-fitting

Taking an average of the two sets of estimates, we obtain the so-called "cross-fitted" estimates of the marginal entropies:

$$\hat{H}(X) = \frac{\widehat{H_1}(X) + \widehat{H_2}(X)}{2}, \text{ and } \hat{H}(Y) = \frac{\widehat{H_1}(Y) + \widehat{H_2}(Y)}{2}.$$

Using these cross-fitted estimates we obtain $\hat{C}_{X\to Y} := \hat{H}(X) - \hat{H}(Y)$. For variance estimation purposes, we use $\hat{\Sigma} := \left(\hat{\Sigma}(1) + \hat{\Sigma}(2)\right)/2$.

Inference

From Corollary 1 in the main article, the asymptotic behaviour of $\sqrt{n}\hat{C}_{X\to Y}$ is given by

$$\sqrt{n}\left(\hat{C}_{X\to Y} - C_{X\to Y}\right) \xrightarrow{\mathcal{D}} N(0, \sigma_C^2), \text{ as } n \to \infty,$$

where σ_C^2 denotes the asymptotic variance of the estimate $\hat{C}_{X\to Y}$ and can be estimated using $\hat{\Sigma}$ as follows: $\hat{\sigma}_C^2 := \hat{\sigma}_{11} + \hat{\sigma}_{22} - 2\hat{\sigma}_{12}$, where $\hat{\sigma}_{ij}$ is the $(i,j)^{th}$ element of the covariance matrix given by $\hat{\Sigma}$.

Optimal split: 50/50

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. 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.

Table 2: Examining $\hat{C}_{X \to Y}$: absolute bias (A.Bias) and coverage probability (CP) for different sample sizes $n \in \{250, 500, 1000\}$ under two cases: (i) $X \sim$ Lognormal$(5, 1)$ and $Y \sim$ N$(5, 1)$ and (ii) $X \sim$ Exp(mean $= 1$) and $Y \sim$ Weibull(scale $= 1$, shape $= 3/2$) under different data-splitting ratios.

		Case (I)			**Case (II)**		
		$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
A.Bias	50/50	0.095	0.058	0.045	0.082	0.062	0.040
	60/40	0.097	0.059	0.046	0.085	0.064	0.042
	70/30	0.101	0.065	0.050	0.094	0.065	0.040
	80/20	0.118	0.078	0.052	0.108	0.075	0.044
	90/10	0.183	0.101	0.073	0.149	0.101	0.063
CP	50/50	0.945	0.980	0.957	0.935	0.935	0.940
	60/40	0.925	0.970	0.950	0.920	0.915	0.955
	70/30	0.845	0.900	0.925	0.880	0.850	0.920
	80/20	0.765	0.815	0.840	0.765	0.795	0.865
	90/10	0.475	0.660	0.645	0.550	0.600	0.705

S5 Resolving ambiguity in causal direction $X \to Y$ or $Y \to X$ when nature of generative function is unknown

The Inverse of an Expanding Function is Contracting

A key insight for our practical workflow is that for a monotonic function, if g has expanding dynamics, its inverse g^{-1} must have contracting dynamics. This can be proven formally. Recall that a function g has expanding dynamics if the geometric mean of its gradient's magnitude is greater than one. Taking the logarithm, this is equivalent to the average log-gradient

being positive:

$$|\mathcal{X}|^{-1} \int_{\mathcal{X}} \log(|\nabla g(x)|)dx > 0.$$

For a function h with contracting dynamics, we must have its average log-gradient be negative:

$$|\mathcal{Y}|^{-1} \int_{\mathcal{Y}} \log(|\nabla h(y)|)dy < 0.$$

Proof. Let U be a random variable uniformly distributed on the support $\mathcal{X}$. The condition for expanding dynamics can be written as an expectation $E[\log(|\nabla g(U)|)] > 0$. Let us define a new random variable $Z = \log(|\nabla g(U)|)$, where $E[Z] > 0$. Using a change of variables ($y = g(x)$, $dy = |\nabla g(x)|dx$) and the inverse function rule ($\nabla h(y) = 1/\nabla g(x)$), we get:

$$\begin{aligned}\int_{\mathcal{Y}} \log(|\nabla h(y)|)dy &= \int_{\mathcal{X}} \log\left(\left|\frac{1}{\nabla g(x)}\right|\right) \cdot |\nabla g(x)|dx \\ &= \int_{\mathcal{X}} (-\log|\nabla g(x)|) \cdot |\nabla g(x)|dx.\end{aligned}$$

Since $|\nabla g(x)| = \exp(\log(|\nabla g(x)|))$, this is equivalent to:

$$|\mathcal{X}| \cdot E[-\log(|\nabla g(U)|) \cdot \exp(\log(|\nabla g(U)|))] = |\mathcal{X}| \cdot E[-Ze^Z].$$

Our goal is to prove that this expression is negative, which requires showing

that $E[Ze^Z] > 0$. We use the formula for covariance, $\text{Cov}(A, B) = E[AB] - E[A]E[B]$, which gives $E[Ze^Z] = \text{Cov}(Z, e^Z) + E[Z]E[e^Z]$. We can now evaluate each term on the right-hand side:

1. $\mathbf{E[Z] > 0}$ since g has expanding dynamics.

2. $\mathbf{E[e^Z] > 0}$: Since e^z is always positive, its expectation must be positive.

3. $\text{Cov}(\mathbf{Z}, \mathbf{e^Z}) \geq \mathbf{0}$: This is a known property of covariance.

Combining these facts, we see that $E[Ze^Z]$ is the sum of a non-negative term and a strictly positive term. Therefore, $E[Ze^Z] > 0$. This implies that $|\mathcal{X}| \cdot E[-Ze^Z] < 0$, which in turn proves that $\int_{\mathcal{Y}} \log(|\nabla h(y)|)dy < 0$. Thus, the inverse function h must have contracting dynamics. □

Indistinguishability of direction via the asymmetry coefficient alone

Just by observing the sign of the asymmetry coefficient, we cannot distinguish between two complementary scenarios. This creates an ambiguity we must resolve. The asymmetry coefficient is defined as $C_{X\to Y} = H(X) - H(Y)$. The coefficient for the reverse direction is $C_{Y\to X} = H(Y) - H(X) = -C_{X\to Y}$. Consider the two hypotheses that could explain an observation where the estimated coefficient is significantly negative:

1. *Hypothesis A:* The true causal direction is $X \to Y$ with an **expanding**

function g. According to the paper, this implies $C_{X\to Y} < 0$.

2. *Hypothesis 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 hypotheses, $C_{X\to Y} < 0$ and $C_{Y\to X} > 0$, are identical. The key to resolving the ambiguity is that Assumption 1 is not symmetric.

Proof. We assume the true direction is $X \to Y$ and that g and f_X are functionally orthogonal, satisfying Assumption 1:

$$\int_{\mathcal{X}} \log(|\nabla g(x)|) f_X(x) dx = \frac{1}{|\mathcal{X}|} \int_{\mathcal{X}} \log(|\nabla g(x)|) dx.$$

We now check if the inverse relationship, $X = h(Y)$, also satisfies the assumption. This would require:

$$\int_{\mathcal{Y}} \log(|\nabla h(y)|) f_Y(y) dy \stackrel{?}{=} \frac{1}{|\mathcal{Y}|} \int_{\mathcal{Y}} \log(|\nabla h(y)|) dy.$$

Using a change of variables, we express the left-hand side (LHS) and right-

hand side (RHS) in terms of g and f_X:

$$\begin{aligned}\text{LHS} &= \int_{\mathcal{X}} \log\left(\left|\frac{1}{\nabla g(x)}\right|\right) \cdot \frac{f_X(x)}{|\nabla g(x)|} \cdot |\nabla g(x)| dx \\ &= -\int_{\mathcal{X}} \log(|\nabla g(x)|) f_X(x) dx \\ &= -|\mathcal{X}|^{-1} \int_{\mathcal{X}} \log(|\nabla g(x)|) dx\end{aligned}$$

$$\begin{aligned}\text{RHS} &= |\mathcal{Y}|^{-1} \int_{\mathcal{X}} \log\left(\left|\frac{1}{\nabla g(x)}\right|\right) |\nabla g(x)| dx \\ &= -|\mathcal{Y}|^{-1} \int_{\mathcal{X}} \log(|\nabla g(x)|) |\nabla g(x)| dx.\end{aligned}$$

For the assumption to be symmetric, we would need LHS = RHS. However, these two expressions are generally not equal due to the extra weighting term $|\nabla g(x)|$ in the RHS integral and the different normalization constants ($|\mathcal{X}|$ vs. $|\mathcal{Y}|$). This proves the asymmetry. The rationale is that the distribution of an *effect* (f_Y) is fundamentally shaped by the interaction of the *cause's* distribution (f_X) and the *mechanism* (g). It is therefore highly improbable that this derived distribution f_Y would also be functionally orthogonal to the inverse mechanism $h = g^{-1}$. □

To provide a practical method for assessing the plausibility of Assumption 1, we developed a data-driven diagnostic check. This procedure esti-

mates the two sides of the functional orthogonality equality and calculates an "Orthogonality Deviation Score," as described in Algorithm 1 of the manuscript. A score close to zero suggests the assumption is plausible. We use a bootstrap procedure to construct a 95% confidence interval (CI) for this score to assess its statistical significance. We demonstrate the effectiveness of this diagnostic with two simulation cases using data generated from the model $Y = X^3 + \epsilon$, where $X \sim \text{Uniform}(0, 2)$ and $n = 500$.

- *Case 1 (Assumption Satisfied):* We run the diagnostic on the true causal direction, $X \to Y$. Since the input X is drawn from a uniform distribution, Assumption 1 is satisfied by design.

- *Case 2 (Assumption Violated):* We use the *same data* but run the diagnostic on the incorrect, anti-causal direction, $Y \to X$. In this direction, the input (now Y) has a non-uniform distribution derived from X^3. This is expected to violate the functional orthogonality assumption.

Simulation Results

The results of the bootstrap analysis ($B = 500$ iterations) are summarized in Supplementary Table 3.

The results clearly demonstrate the diagnostic's ability to distinguish

Table 3: Bootstrap results for the Assumption 1 diagnostic check.

Causal Direction	**Orthogonality Deviation Score**	**95% Bootstrap CI**
Correct ($X \to Y$)	-0.0135	[-0.1481, 0.1276]
Incorrect ($Y \to X$)	0.6738	[0.5888, 0.7552]

between the two scenarios.

- In the correct causal direction (Case 1), the mean deviation score is very close to zero, and the 95% confidence interval clearly contains zero. This provides no evidence against Assumption 1, indicating that it is plausible.

- In the incorrect causal direction (Case 2), the mean score is large and positive, and the 95% confidence interval is strictly positive and far from zero. This serves as a strong indicator, correctly identifying that Assumption 1 is likely violated for this direction.

This simulation confirms that our diagnostic procedure is a valuable and effective tool for researchers to assess the plausibility of Assumption 1 when determining the correct causal direction. The visual separation of the two bootstrap distributions, as shown in Supplementary Figure 3, provides further compelling evidence.

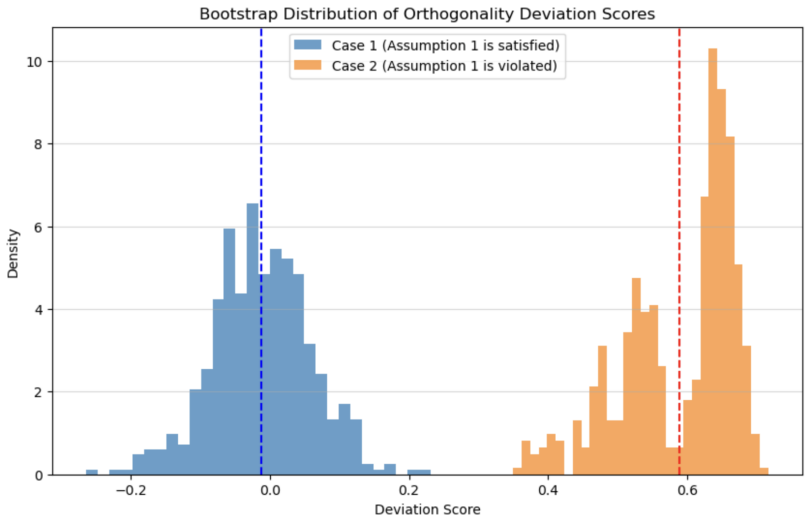


Figure 3: Bootstrap distributions of the deviation score for the correct and incorrect causal directions. The distribution for the correct direction is centered at zero, while the distribution for the incorrect direction is clearly shifted away from it.

S6 Diagnostics

Determining if a generative function is expanding or contracting

To answer the practical question of whether an unknown function g in the relationship $Y = g(X)$ is expanding or contracting, we can directly probe the definition provided in the main manuscript. A function is contracting if its average log-gradient is negative, and expanding if it is positive. The procedure involves estimating this average log-gradient from the observed data and quantifying its uncertainty:

1. **Estimate the function $\hat{g}(x)$:** We fit a non-parametric smoothing spline to the data pairs (x_i, y_i).

2. **Estimate the average log-gradient:** We numerically integrate the logarithm of the spline's derivative over the support of X given by $\hat{\mathcal{X}} = [\min(X), \max(X)]$, and average by the length of the support:

$$\text{AvgLogGrad}_{\text{est}} = |\hat{\mathcal{X}}|^{-1} \int_{\min(X)}^{\max(X)} \log(|\nabla \hat{g}(x)|) dx.$$

3. **Quantify uncertainty:** We perform a bootstrap analysis by repeatedly resampling the data, re-calculating the average log-gradient, and constructing a 95% confidence interval (CI) from the resulting empirical distribution.

A CI entirely below zero provides strong evidence for contracting dynamics, while a CI entirely above zero indicates expanding dynamics.

Simulation Results

We demonstrate this procedure on two simulated cases ($n = 500$). The results are summarized in Supplementary Table 4.

Table 4: Bootstrap results for generative function dynamics classification.

Case	**True Dynamic**	**Mean Avg. Log-Gradient**	**95% Bootstrap CI**
1: $Y = \sqrt{X} + \epsilon$	Contracting	-0.189	$[-0.203, -0.176]$
2: $Y = e^X + \epsilon$	Expanding	0.501	$[0.497, 0.506]$

The simulation confirms the diagnostic's effectiveness. In both the contracting and expanding cases, the bootstrap confidence interval correctly

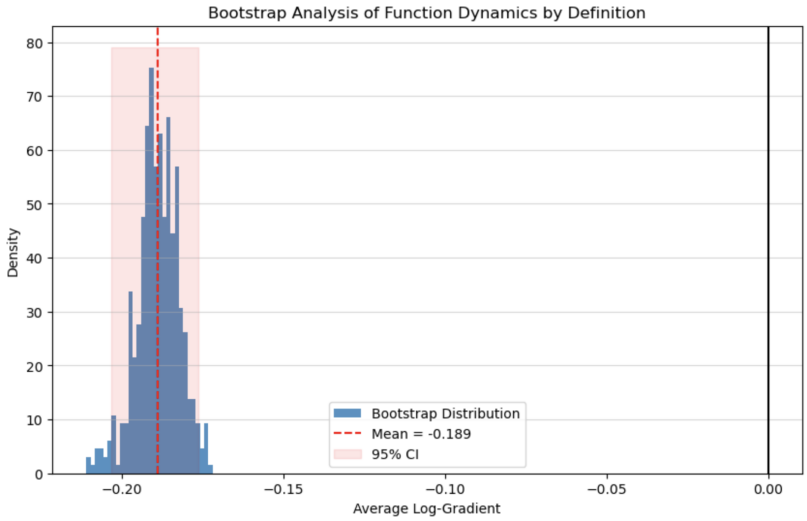


Figure 4: Bootstrap distribution of the estimated average log-gradient for a contracting generative function ($Y = \sqrt{X} + \epsilon$). The distribution lies entirely to the left of zero, allowing for a definitive conclusion.

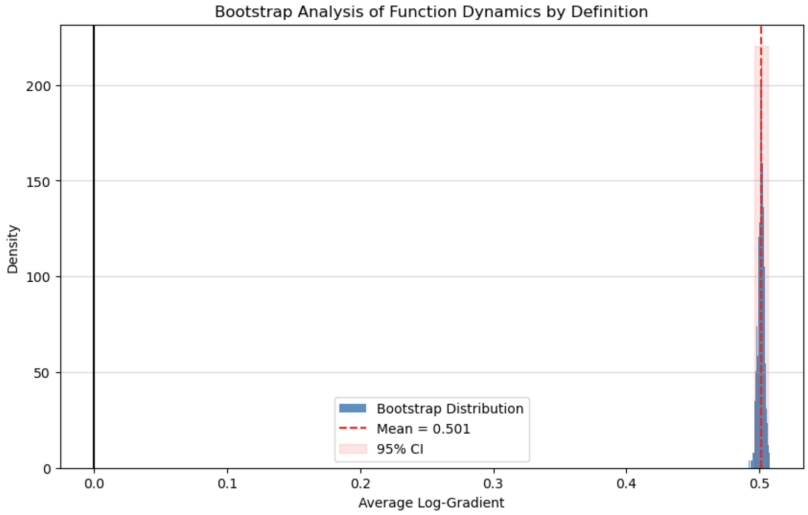


Figure 5: Bootstrap distribution of the estimated average log-gradient for an expanding generative function ($Y = e^X + \epsilon$). The distribution lies entirely to the right of zero, allowing for a definitive conclusion.

identifies the sign of the average log-gradient, providing a reliable, data-driven method to determine the function's dynamics.

Bibliography

Barron, A. R. (1986). Entropy and the central limit theorem. *Ann. of Prob. 14*, 336–342.

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

Zamir, R. (1998). A proof of the fisher information inequality via a data processing argument. *IEEE Trans. on Info. Theo. 44*, 1246–1250.