Dynamical Systems Theory for Causal Inference with Application to Synthetic Control Methods

Supplementary Material

Yi Ding and Panos Toulis

1 Proof Theorem 1

Proposition 1. *Let $d_Y(t,t') = ||\widetilde{Y}_t - \widetilde{Y}_{t'}||$ and $r_Y(t,t_i) = d_Y(t,t_i)/d_Y(t,t_1)$, where $\widetilde{Y}_t = [Y_t, Y_{t-\tau}, \ldots, Y_{t-(d-1)\tau}]$ is the delayed coordinate embedding of Y_t, and $\{\widetilde{Y}_{t_i} : i = 1, \ldots, d+1\}$ are the $d+1$ nearest neighbors to $\widetilde{Y}_t$. The CCM score of Y_t on X_t based on MAE metric is equal to:*

$$\mathrm{CCM}(X_t \mid Y_t) = \Big|\Big(X_0 - \frac{\mu}{1-\alpha}\Big)\Big(\alpha^t - \sum_{i=1}^{d+1} w_Y(t_i,t)\alpha^{t_i}\Big) + \Big(E_t - \sum_{i=1}^{d+1} w_Y(t_i,t)E_{t_i}\Big)\Big|,$$

where $E_t = \sum_{s=1}^{t} \alpha^{t-s}\epsilon_s$, and $w_Y(t_i,t) = e^{-r_Y(t,t_i)}/\sum_{j=1}^{d+1} e^{-r_Y(t,t_j)}$. Similarly, let $d_X(t,t') = ||\widetilde{X}_t - \widetilde{X}_{t'}||$ and $r_X(t,t'_i) = d_X(t,t'_i)/d_X(t,t'_1)$, where $\widetilde{X}_t = [X_t, X_{t-\tau}, \ldots, X_{t-(d-1)\tau}]$ is the delayed coordinate embedding of X_t, and $\{\widetilde{X}_{t'_i} : i = 1, \ldots, d+1\}$ are the $d+1$ nearest neighbors to $\widetilde{X}_t$. The CCM score of X_t on Y_t based on MAE metric is equal to:

$$\mathrm{CCM}(Y_t \mid X_t) = \Big|\beta\Big(X_0 - \frac{\mu}{1-\alpha}\Big)\Big(\alpha^{t-1} - \sum_{i=1}^{d+1} w_X(t'_i,t)\alpha^{t'_i-1}\Big) + \beta\Big(E_t - \sum_{i=1}^{d+1} w_X(t'_i,t)E_{t'_i}\Big) + \Big(\zeta_t - \sum_{i=1}^{d+1} w_X(t'_i,t)\zeta_{t'_i}\Big)\Big|,$$

where $w_X(t'_i,t) = e^{-r_X(t,t'_i)}/\sum_{j=1}^{d+1} e^{-r_X(t,t'_j)}$.

Proof. Recall that the AR model is as follows:

$$X_t = \alpha X_{t-1} + \mu + \epsilon_t,$$
$$Y_t = \beta X_{t-1} + \mu + \zeta_t,$$

where α, β are fixed, with $|\alpha| < 1$, $\beta \geq 0$, μ is the drift, $\epsilon_t \sim \mathcal{N}(0, \sigma_X^2)$ and $\zeta_t \sim \mathcal{N}(0, \sigma_Y^2)$ are zero-mean and constant-variance normal errors. We can solve the recursion to obtain:

$$X_t = \alpha^t X_0 + P_t(\alpha)\mu + E_t,$$
$$Y_t = \beta\alpha^{t-1} X_0 + (\beta P_{t-1}(\alpha) + 1)\,\mu + \beta E_t + \zeta_t,$$

where $P_t(\alpha) = 1 + \alpha + \ldots + \alpha^{t-1} = (1-\alpha^t)/(1-\alpha)$ and E_t is the summation of weighted error terms $E_t = \sum_{s=1}^{t} \alpha^{t-s}\epsilon_s$.

The cross estimation for X_t is

$$\hat{X}_t = \sum_{i=1}^{d+1} w_Y(t_i,t) X_{t_i} = \sum_{i=1}^{d+1} w_Y(t_i,t)\left(\alpha^{t_i} X_0 + P_{t_i}(\alpha)\mu + E_{t_i}\right),$$

where $w_Y(t_i,t) = e^{-r_Y(t,t_i)}/\sum_{j=1}^{d+1} e^{-r_Y(t,t_j)}$. Then, the CCM score of Y_t on X_t based on MAE can be expressed as:

$$\begin{aligned}
|X_t - \hat{X}_t| &= \Big|\alpha^t X_0 + P_t(\alpha)\mu + E_t - \Big(\sum_{i=1}^{d+1} w_Y(t_i,t)(\alpha^{t_i} X_0 + P_{t_i}(\alpha)\mu + E_{t_i})\Big)\Big| \\
&= \Big|X_0\Big(\alpha^t - \sum_{i=1}^{d+1} w_Y(t_i,t)\alpha^{t_i}\Big) + \mu\Big(P_t(\alpha) - \sum_{i=1}^{d+1} w_Y(t_i,t)P_{t_i}(\alpha)\Big) + \Big(E_t - \sum_{i=1}^{d+1} w_Y(t_i,t)E_{t_i}\Big)\Big| \\
&= \Big|X_0\Big(\alpha^t - \sum_{i=1}^{d+1} w_Y(t_i,t)\alpha^{t_i}\Big) + \mu\Big(\frac{1-\alpha^t}{1-\alpha} - \sum_{i=1}^{d+1} w_Y(t_i,t)\frac{1-\alpha^{t_i}}{1-\alpha}\Big) + \Big(E_t - \sum_{i=1}^{d+1} w_Y(t_i,t)E_{t_i}\Big)\Big| \\
&= \Big|\Big(X_0 - \frac{\mu}{1-\alpha}\Big)\Big(\alpha^t - \sum_{i=1}^{d+1} w_Y(t_i,t)\alpha^{t_i}\Big) + \Big(E_t - \sum_{i=1}^{d+1} w_Y(t_i,t)E_{t_i}\Big)\Big|. \qquad (1)
\end{aligned}$$

Similarly, the cross estimation for Y_t is

$$\hat{Y}_t = \sum_{i=1}^{d+1} w_X(t'_i,t) Y_{t'_i} = \sum_{i=1}^{d+1} w_X(t'_i,t)(\beta\alpha^{t'_i-1} X_0 + (\beta P_{t'_i-1}(\alpha)+1)\mu + \beta E_{t'_i} + \zeta_{t'_i}), \qquad (2)$$

where $w_X(t'_i,t) = e^{-r_X(t,t'_i)}/\sum_{j=1}^{d+1} e^{-r_X(t,t'_j)}$. Then, the CCM score of X_t on Y_t based on MAE can be expressed as:

$$\begin{aligned}
|Y_t - \hat{Y}_t| &= \Big|\beta\alpha^{t-1}X_0 + \Big(\beta P_{t-1}(\alpha)+1\Big)\mu + \beta E_t + \zeta_t - \Big(\sum_{i=1}^{d+1} w_X(t'_i,t)(\beta\alpha^{t'_i-1}X_0 + (\beta P_{t'_i-1}(\alpha)+1)\mu + \beta E_{t'_i} + \zeta_{t'_i})\Big)\Big| \\
&= \Big|\beta X_0\Big(\alpha^{t-1} - \sum_{i=1}^{d+1} w_X(t'_i,t)\alpha^{t'_i-1}\Big) + \beta\mu\Big(P_{t-1}(\alpha) - \sum_{i=1}^{d+1} w_X(t'_i,t)P_{t'_i-1}(\alpha)\Big) + \beta\Big(E_t - \sum_{i=1}^{d+1} w_X(t'_i,t)E_{t'_i}\Big) \\
&\quad + \Big(\zeta_t - \sum_{i=1}^{d+1} w_X(t'_i,t)\zeta_{t'_i}\Big)\Big| \\
&= \Big|\beta X_0\Big(\alpha^{t-1} - \sum_{i=1}^{d+1} w_X(t'_i,t)\alpha^{t'_i-1}\Big) + \beta\mu\Big(\frac{1-\alpha^{t-1}}{1-\alpha} - \sum_{i=1}^{d+1} w_X(t'_i,t)\frac{1-\alpha^{t'_i-1}}{1-\alpha}\Big) + \beta\Big(E_t - \sum_{i=1}^{d+1} w_X(t'_i,t)E_{t'_i}\Big) \\
&\quad + \Big(\zeta_t - \sum_{i=1}^{d+1} w_X(t'_i,t)\zeta_{t'_i}\Big)\Big| \\
&= \Big|\beta\Big(X_0 - \frac{\mu}{1-\alpha}\Big)\Big(\alpha^{t-1} - \sum_{i=1}^{d+1} w_X(t'_i,t)\alpha^{t'_i-1}\Big) + \beta\Big(E_t - \sum_{i=1}^{d+1} w_X(t'_i,t)E_{t'_i}\Big) + \Big(\zeta_t - \sum_{i=1}^{d+1} w_X(t'_i,t)\zeta_{t'_i}\Big)\Big|. \\
&\qquad (3)
\end{aligned}$$

□

We now proceed to the proof of Theorem 1 using the results in Equation (1) and Equation (3). We repeat the Assumption 1 in the main paper.

Assumption 1. *For the CCM scores in Proposition 1, fix* t *and let* $L \to \infty$*, and suppose that:*
(a) $\min_{i=1,\dots,d+1} \min\{t_i', t_i\} \to \infty$*;*
(b) $\limsup_{i=1,\dots,d+1} |w_Y(t_i,t) - \frac{1}{d+1}| = 0,$ *and* $\limsup_{i=1,\dots,d+1} |w_X(t_i',t) - \frac{1}{d+1}| = 0$*;*
(c) $\min_{i\neq j} |t_i - t_j| \to \infty$ *and* $\min_{i\neq j} |t_i' - t_j'| \to \infty$*.*

Theorem 1. *Suppose that Assumptions 1(a)-(c) hold for the CCM scores in Proposition 1, then*

$$\mathrm{CCM}(X_t \mid Y_t) \xrightarrow{d} \mathrm{FN}\bigg((X_0 - \frac{\mu}{1-\alpha})\alpha^t, \frac{2-\alpha^{2t}}{1-\alpha^2}\sigma_X^2\bigg),$$
$$\mathrm{CCM}(Y_t \mid X_t) \xrightarrow{d} \mathrm{FN}\bigg(\beta(X_0 - \frac{\mu}{1-\alpha})\alpha^{t-1}, \frac{2-\alpha^{2t}}{1-\alpha^2}\beta^2\sigma_X^2 + 2\sigma_Y^2\bigg),$$

where $\mathrm{FN}(\mu,\sigma^2) = |N(\mu,\sigma^2)|$ *is the folded normal distribution with mean* μ *and variance* σ^2*.*

Proof. From $\alpha < 1$ and Assumption 1(a) we get:

$$\sum_{i=1}^{d+1} w_Y(t_i,t)\alpha^{t_i} \le \sum_{i=1}^{d+1} w_Y(t_i,t)\alpha^{\min_i \min\{t_i,t_i'\}} \le \alpha^{\min_i \min\{t_i,t_i'\}} \sum_{i=1}^{d+1} w_Y(t_i,t) = \alpha^{\min_i \min\{t_i,t_i'\}} \to 0.$$

Since $E_t = \sum_{s=1}^t \alpha^{t-s}\epsilon_s = \alpha^{t-1}\epsilon_1 + \alpha^{t-2}\epsilon_1 + \cdots + \epsilon_t$ and $\epsilon_t \sim \mathcal{N}(0,\sigma_X^2)$, we have

$$\mathrm{E}(E_t) = 0, \text{ and } \mathrm{var}(E_t) = \frac{1-\alpha^{2t}}{1-\alpha^2}\sigma_X^2.$$

For $t_i \neq t_j$, it is straightforward to show that $\mathrm{cov}(E_{t_i}, E_{t_j}) = O(\alpha^{|t_i - t_j|}) \to 0$, where the limit follows from Assumption 1(c). Similarly, the results hold for t_i', t_j'.

$$\mathrm{E}(E_{t_i}) \text{ and } \mathrm{Cov}(E_{t_i}, E_{t_i}) = \frac{1-\alpha^{2t_i}}{1-\alpha^2}\sigma_X^2 \to \frac{1}{1-\alpha^2}\sigma_X^2.$$

From Assumption 1(b) we have:

$$\sum_{i=1}^{d+1} w_Y(t_i,t)E_{t_i} \xrightarrow{\mathrm{d}} \frac{1}{d+1}\sum_{i=1}^{d+1} E_{t_i} \ \to N(0, \frac{\sigma_X^2}{1-\alpha^2}),$$

from which it follows that

$$E_t - \sum_{i=1}^{d+1} w_Y(t_i,t)E_{t_i} \sim N(0, \frac{2-\alpha^{2t}}{1-\alpha^2}\sigma_X^2).$$

Hence, $\mathrm{CCM}(X_t \mid Y_t)$ converges in a distribution to

$$\mathrm{CCM}(X_t \mid Y_t) \xrightarrow{\mathrm{d}} \mathrm{FN}\bigg((X_0 - \frac{\mu}{1-\alpha})\alpha^t, \frac{2-\alpha^{2t}}{1-\alpha^2}\sigma_X^2\bigg),$$

where $\mathrm{FN}(\cdot,\cdot)$ is the folded normal distribution.

Similarly, since $\zeta_t \sim \mathcal{N}(0,\sigma_Y^2)$ and $\limsup_{i=1,\dots,d+1} |w_X(t_i',t) - \frac{1}{d+1}| = 0$, we have

$$\sum_i w_X(t_i',t)\zeta_{t_i'} = \frac{1}{d+1}\sum_{i=1}^{d+1} w_X(t_i',t)\zeta_{t_i'} + o_P(1) \xrightarrow{\mathrm{d}} N(0,\sigma_Y^2).$$

It follows that $\zeta_t - \sum_{i=1}^{d+1} w_X(t'_i, t)\zeta_{t'_i} \sim \mathcal{N}(0, 2\sigma_Y^2)$, and that $\mathrm{CCM}(Y_t \mid X_t)$ converges in a distribution to

$$\mathrm{CCM}(Y_t \mid X_t) \xrightarrow{\mathrm{d}} \mathrm{FN}\Big(\beta(X_0 - \frac{\mu}{1-\alpha})\alpha^{t-1}, \frac{2-\alpha^{2t}}{1-\alpha^2}\beta^2\sigma_X^2 + 2\sigma_Y^2\Big).$$

□

2 Data

2.1 California's Tobacco Control Program

California's tobacco control program Abadie et al. (2010) uses the annual state-level per-capita cigarette sales panel data from 1970 to 2000. Artificial control units $A_{s,t}$ are created in our simulated study, where s denotes a hypothetical state and t indicates time, and then are added in the donor pool. Then, we perform the standard synthetic control analysis, and check whether CCM+SCM or SCM select the artificial units to construct synthetic California.

We use time series templates to generate artificial control units. In particular, we create 39 artificial states $A_{s,t}$ (the same number of states in original study) with corresponding panel data and four predictors of the outcome variable. The panel data are generated from multiple sets of noisy copies from the template and the predictors are from original tobacco data but with permuted indices for each artificial state. We add $A_{s,t}$ to the original pool to construct a new pool including 77 control units. We also apply moderate data transformations to ensure these adversaries sizable but unrelated to original data, and multiple sets of artificial control units are generated. We run simulations for each set of adversaries and display result distributions with box plots. The template data are described as follows.

Unemployment. The unemployed percent of US labor force data include annual average employment status of the civilian population from 1976 to 2016, giving 41 years of data for 51 states. [1] We define artificial units as $A_{s,t} = kU_{s,t}$, where k is a scalar, and $U_{s,t}$ is unemployment for state s at time t. To make the data sizable with the tobacco data, k is set to be 6. To generate multiple sets of artificial control units, we select the starting year between 1976 to 1986 and take the following 31 data points as 31 years of data for each state, which gives 11 sets of artificial control units. The simulation results are obtained over 11 runs.

2.2 Brexit Vote

We picked 30 OECD-member countries as controls, and UK as the treated unit. We collected quarterly real GDP data of these countries from the OECD Economic Outlook database (June 2017) from 1995Q1 to 2018Q4 [2], where data from 2017Q4 till 2018Q4 are forecasts. The whole quarterly GDP data has 96 data points, and the first 86 points are before Brexit vote. It is assumed that the treatment took form after 2016Q2, and the countries in the donor pool are not affected by the treatment. We also collected predictors of outcome variable such as private consumption, investment, inflation rate, interest rate, and exchange rate.

We normalized the time series for each country by dividing the time series by its 1995 average and then taking logarithm of that time series to generate the approximately zero starting point in 1995. The predictors of outcome variable include:

[1] Data collected from the Local Area Unemployment Statistics (LAUS) program of the Bureau of Labor Statistics (BLS) (Bureau of Labor Statistics, 2018) https://www.bls.gov/lau/staadata.txt

[2] https://stats.oecd.org/index.aspx?DataSetCode=EO

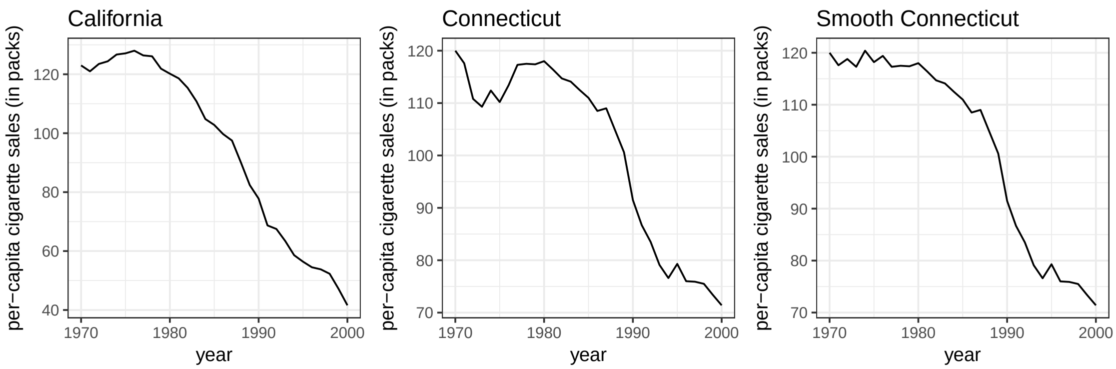

Figure 1: Complementary to Figure 3 in the main paper: Per-capita cigarette sales trends for California, Connecticut, and Smooth Connecticut.

(a) Real private consumption: the sum of real final consumption expenditure of both households and non-profit institutions serving households, from 1997Q1 to 2017Q2.
(b) Real investment: total gross fixed capital formation, from 1995Q1 to 2017Q2.
(c) Net exports: the external balance of goods and services, from 1997Q1 to 2017Q2.
(d) Inflation series: the change in the Consumer Price Index (CPI), from 1998Q1 to 2017Q3.
(e) Quarterly short-term nominal interest rates: quarterly averages of monthly values, from 2002Q1 to 2017Q4.
(f) Nominal exchange rates: from 1997Q1 to 2018Q4.

The artificial control units are generated in the same way as the example of California's tobacco control program. We use time series template and create 31 artificial countries $A_{s,t}$ (same number of control countries in the original study) with corresponding panel data and six predictors of the outcome variable. We add $A_{s,t}$ to the original pool to construct a new pool including 61 control units. We also apply moderate data transformations to ensure these adversaries sizable and unrelated to original panel data. The details on how to generate the adversaries are described as follows.

Calls. We use the calls data collected from the Monthly average daily calls to directory assistance from Jan 1962 to Dec 1976 [3] as the template for our adversarial attack. We choose the first 106 data points from this series due to its similar trend with the brexit data, which gives 11 different sets of templates by choosing different starting points. For each template, we fit an autoregressive model and create 31 noisy copies as 31 artificial countries by adding Gaussian noise to them. The simulation results are obtained over 11 runs.

3 Discussion on CCM

Due to the success of CCM in quantifying dynamical relationships (Sugihara et al., 2012; Deyle et al., 2013), it may be tempting to consider CCM as a method for causal inference. We recommend putting more thoughts before applying this idea. To illustrate why, we apply CCM on causal relationship detection tasks from the benchmark dataset `CauseEffectPairs` (Mooij et al., 2016), which contains time series pairs that are known a priori to be causal or not. In practice, time series are normalized before applying CCM on them to ensure all series have the same magnitude for comparison and avoid constructing a distorted state space (Chang

[3] https://datamarket.com/data/set/22yq

et al., 2017).

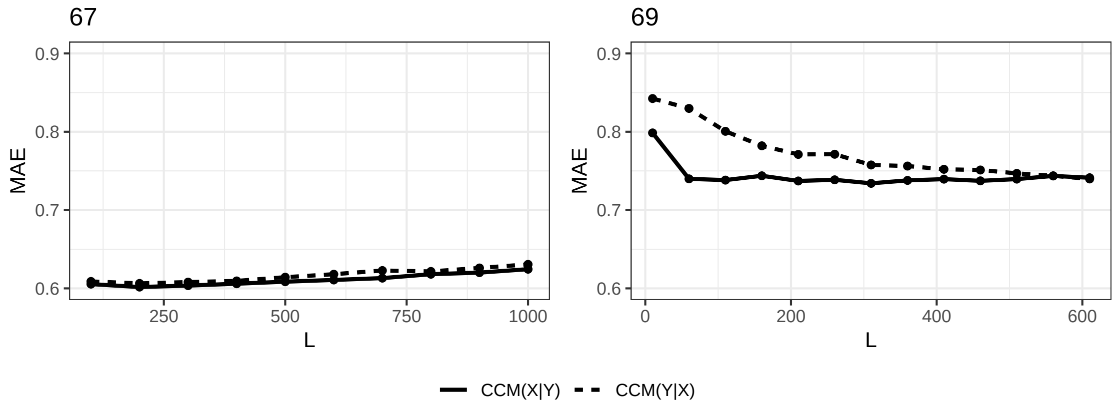

Figure 2: CCM results from two pairs. The number in the title of each figure corresponds to the pair index. The ground truths are: 67 ($X \to Y$), 69 ($Y \to X$).

Two cases where CCM fails to detect the true direction of causality are shown in Figure 2. Pair 67 is the financial time series about stock returns from two companies in which one stock is believed to depend on the other. We can see that the CCM score fails to visually converge as library size L increases. By inspection, the time series are close to random walks. Since CCM theory mainly applies on deterministic or chaotic dynamical systems, it is not reliable as a standalone causal inference method in systems dominated by noise. Another example is Pair 69 in the data of indoor and outdoor temperature. Here, the ground truth is that outdoor temperature variable Y drives the indoor temperature variable X, indicating that the dotted curve should converge faster than the solid curve in the right subplot of Figure 2. However, CCM gives the opposite causal direction result. A possible explanation might be that temperature is periodic since it has been suggested that strong periodicity could undermine the effectiveness of CCM (Chang et al., 2017).

Another practical aspect is that hyperparameters, such as the embedding dimension d and time delay τ, should be carefully chosen. To illustrate this, we consider Pair 68 in the data of internet connections and traffic, where X is bytes sent and Y is number of http connections. Figure 3 shows CCM results for this pair with simple data transformations and with varying the embedding dimension d.

Although CCM uncovers the correct causal detection with the original data under embedding dimension $d = 3$, the result is not strong enough. Moreover, CCM detects a wrong causal direction when the embedding dimension is set to $d = 4$. We note that the results improve with transformations, say, log transforms. Optimality of embedding methods and parameter tuning are currently active research areas (Rosenstein et al., 1994; Small and Tse, 2004; Garland and Bradley, 2015).

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

Bureau of Labor Statistics, U. D. o. L. (2018). *Local Area Unemployment Statistics*.

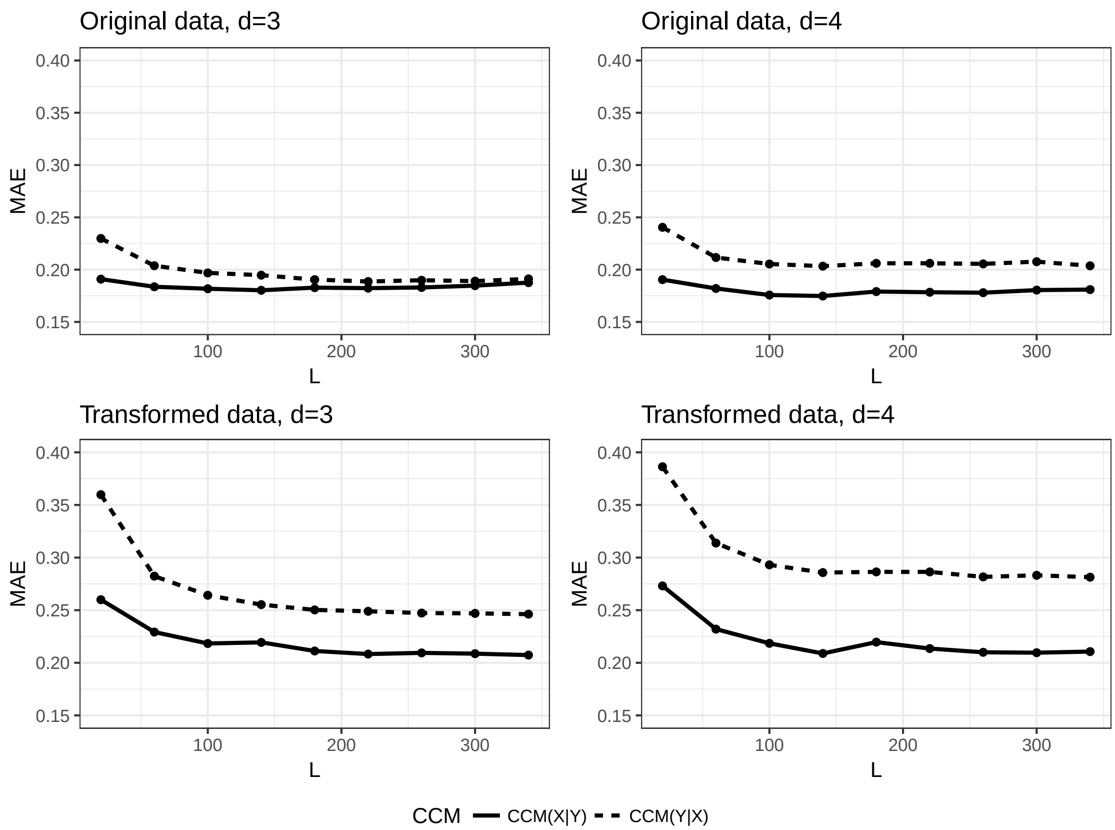

Figure 3: CCM results on pair 68. The ground truth is $Y \to X$.

Chang, C.-W., Ushio, M., and Hsieh, C.-h. (2017). Empirical dynamic modeling for beginners. *Ecological Research*, 32(6):785–796.

Deyle, E. R., Fogarty, M., Hsieh, C.-h., Kaufman, L., MacCall, A. D., Munch, S. B., Perretti, C. T., Ye, H., and Sugihara, G. (2013). Predicting climate effects on pacific sardine. *Proceedings of the National Academy of Sciences*, 110(16):6430–6435.

Garland, J. and Bradley, E. (2015). Prediction in projection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(12):123108.

Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., and Schölkopf, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *Journal of Machine Learning Research*, 17(32):1–102.

Rosenstein, M. T., Collins, J. J., and Luca, C. J. D. (1994). Reconstruction expansion as a geometry-based framework for choosing proper delay times. *Physica D: Nonlinear Phenomena*, 73(1):82 – 98.

Small, M. and Tse, C. (2004). Optimal embedding parameters: a modelling paradigm. *Physica D: Nonlinear Phenomena*, 194(3):283 – 296.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., and Munch, S. (2012). Detecting causality in complex ecosystems. 338.