Erasmus University Rotterdam

Msc Data Science & Marketing Analytics

A Complexity Theory Approach to Causal Estimations for Marketing Time Series Data

Author

Maurice Korf

Student ID

476550

Supervisor

Prof.Dr.Philip Hans Franses

Second Assessor

Dr.NM Almeida Camacho

Date

August, 16, 2021

Abstract

This study argues that complexity theory based methods can be leveraged for marketing-rooted causal questions. Specifically, within marketing, reductionist approaches are traditionally applied where elements are isolated and only then causally evaluated. However, such an approach fails to capture the underlying dynamics that might be present even though a setting might at first seem straightforwardly to interpret. In other words, only a part of the system is evaluated and might not give an accurate representation of what is truely happening. Furthermore, complex phenomena like non-linearity or feedback effects among others also obstruct the use of commonly deployed methods and oftentimes leads to a situation that a complex setting is being linearized, which may lead to unrealistic estimations. Therefore, this study discusses in general how complexity theory based methods adds value to the marketeer's toolbox and specifically discusses the Cross Convergent Mapping (CCM) method. This is a (non-linear) state space reconstruction method that leverages the properties of delay embedding to determine causality between time series. Hence, this study shows that by means of CCM a complete causal network can be mapped even when complex phenomena are observed. Moreover, it is shown that by means of CCM complex causal interactions can be revealed that cannot be identified with the traditional deployed methods. To conclude, complexity theory based methods can be leveraged to enhance causal estimations which subsequently aids management in making better informed strategic decisions.

Contents

1 Introduction **3**

2 Literature Review **6**

2.1 Dynamic Effects in Marketing 6

2.2 Estimating Dynamic Causal Effects 8

2.2.1 The Distributed Lag Model 8

2.2.2 Granger Causality 10

2.3 Complex Systems Theory 11

2.3.1 The Principles of Complex Systems 11

2.3.2 Complexity Theory in Marketing 27

2.3.3 Case Study: Complexity Theory in Advertising 31

3 Methods **35**

3.1 Cross Convergent Mapping 35

3.1.1 Cross Convergent Mapping Algorithm 40

4 Data **48**

5 Results **51**

5.1 Cross Convergent Mapping Results 51

5.2 Robustness Test for Cross Convergent Mapping 62

6 Discussion **66**

7 Conclusion **67**

A Appendix **72**

1 Introduction

If the price is increased by 10%, what will be the impact on the number of sales? If the advertising expenditures are increased by 15%, how will this affect the gross revenue? These are just some of the many possible questions a marketeer might seek answers for. However, an element that is of crucial importance for most of such questions is that an answer is sought for in terms of a causal effect estimation. More specifically, it is desired to understand how a particular variable affects a targeted outcome because, after all, correlation does not imply causation.

The notion of causality is within the field of marketing research commonly associated with conducting experiments, such as setting up randomized controlled experiments where the expected difference in the variable of interest denotes the causal effect. In such experimental settings and more generally when observational data is available, causal interpretations are oftentimes based on the Neyman-Rubin causal framework otherwise known as the potential outcomes framework (Holland, 1986; Maldonado and Greenland, 2002; Neyman et al., 1923; Rubin, 1974). Particularly, within the field of marketing research a large body of literature exists on utilizing observational data with methods underlying the potential outcomes framework as a means of estimating causal effects (e.g.,Bertrand et al., 2010; Huang et al., 2020; Pochun et al., 2018).

However, for certain marketing related research questions the utilization of cross-sectional data to estimate causal effects is not desired as it might encounter potential limitations, most notably the limitation that dynamic effects cannot be captured. More precisely, when cross-sectional data is utilized to estimate for example the causal impact of advertising on sales then the estimation does not control for the fact that advertising has carry-over effects. In other words, the effects of the given advertising expenditures on sales are distributed over time which imply that estimations based on data from one specific point in time do not capture the full effect (Bruce et al., 2017; Tull, 1965). Moreover, another prominent example that demonstrates the limitations of using cross-sectional data for causal inference rooted questions is when one is interested in the effect of pricing on sales. More specifically, pricing has been shown to poses dynamic effects stemming from, among others, competition, consumer patient levels and the word-of-mouth effect and thus again implies that an estimation from one point in time does not capture the full effect (Ajorlou et al., 2018; Lobel, 2020; Schlosser and Boissier, 2018).

Hence, for certain marketing mix related questions it would make sense to utilize time series data to be able to capture relevant dynamic effects, which subsequently contributes to a more realistic estimation of a particular effect. Within the field of time series analysis various methods exist to estimate causal effects, where commonly a distributed lag model with exogenous regressors is deployed as a baseline to estimate such dynamic causal effects. Such an approach and its estimations are otherwise known as dynamic multipliers or cumulative dynamic multipliers (Stock, Watson, et al., 2012). Nonetheless, still one of the most widely deployed methods for estimating causal effects in time series is by means of Granger Causality. Granger Causality is a probabilistic concept of causality and in essence does not estimate *true* causality but whether one time series provides statistically significant information about the future values of another time series. Formulated differently, Granger Causality constitutes a statistical hypothesis test to test whether one time series contains information to forecast another one and thus merely provides *predictive* causality. Nevertheless, the Granger Causality test is the basis for numerous other causality identifying methods and is still one of the most popular methods deployed in practice (Eichler, 2012; Granger, 1969).

However, a crucial limitation of the aforementioned methods and various related methods is that it assumes a form of linearity and thus does not work well in settings where non-linearity is present or when the interrelated causal relationships are not so straightforward. Specifically, an issue that arises is when the assumption of separability is not fulfilled which is crucial for Granger Causality and implies that the causal variables should be independent of the variables it influences. Additionally, it could occur that shared causal variables are present resulting into synergistic effects that cannot be captured by traditional (linear) methods (Čenys et al., 1991). More generally, settings where various of such complex phenomena occur undermine the crucial assumptions of commonly deployed (linear) causal methods in both time series analysis and marketing and thus demands for a different approach. In addition, this study argues that the occurrence of such complex phenomena, like feedback loops, tend to be more likely within marketing than initially thought of and thereby confirms the need for a different approach. More fundamentally, within marketing there is a tendency to deploy reductionist methods where single components are isolated and only then evaluated. For instance, the traditional focus is on for example evaluating the causal effects of the advertising efforts of product one on the sales of product one, while there might actually be a more complex underlying system present. For example, there might be complementary products where its advertising and pricing might also influence both the sales and advertising efforts of product one. In other words, at first a particular setting might seem to be straightforward to evaluate while in reality it is much more complex due the many components and interactions that are present which are likely to give rise to phenomena like feedback effects and thus non-linearity.

Hence, this study proposes a complexity theory based approach by treating a marketing setting as a complex system. Specifically, there are various key elements that characterizes a complex system and can be used as guideline to determine whether a particular setting seems plausible to be regarded as complex. Consequently, applying such complexity theory based methods allows to capture the dynamics of the system and provides insights in all the causal interrelationships among the variables. More precisely, this study discusses the Cross Convergent Mapping (CCM) method which is a (non-linear) state space reconstruction method that investigates whether two time series variables are causally influencing each other by means of leveraging the properties of delay embedding.

In short, this study is pioneering in two ways, first because this study explores the theory of complex systems in relation to marketing from a causality perspective and secondly on how causal methods designed for complex systems can be leveraged empirically for marketing rooted questions. So, the key idea is to approach marketing causally-rooted questions from a complexity theory based approach instead of the traditional deployed reductionist approach with the goal that complex marketing settings can be more accurately captured. Therefore, this study first argues how complexity theory is relevant for causal estimations within marketing and subsequently how such methods from complexity theory like Cross Convergent Mapping can be empirically leveraged. Hence, this research aims to answer the following research question:

"How can complexity theory based methods be leveraged to enhance causal estimations from time series data for marketing mix related questions?"

Based on this research question, it is by no means argued that complexity theory based methods should substitute all the traditional deployed methods or approaches, but this study argues that the field of complexity theory broadens the possibilities for marketeers. More specifically, marketing settings that may be characterized by complex phenomena like non-linearity, feedback effects or many interacting variables could be more accurately captured by complexity based methods. In other words, within marketing it is oftentimes observed that settings are being *linearized* which is not always necessary nor the best choice and could even result into findings that are very different from reality. By being able to map all causal relations, especially in complex settings which is oftentimes the case with real data, provides management with more accurate and valuable insights. More precisely, the consequences of a particular strategic decision can be better understood and thus helps management in making better informed decisions. So, this study first discusses the key characteristics of complex systems and how and why this is relevant for marketing. Additionally, a case study is provided to demonstrate complexity theory's relevancy. Thereafter, the specific method Cross Convergent Mapping (CCM) is discussed which is a (non-linear) state space reconstruction method. Subsequently, empirical evidence is provided where CCM is applied to real marketing data and is followed by a discussion.

2 Literature Review

It is first discussed how time series data adds value for certain marketing mix related questions. More specifically, it is discussed how time series data can be utilized to capture dynamic effects where two prominent examples from the literature are provided, namely: dynamic effects of advertising and the dynamic effects of pricing. Next, causal methods that are commonly deployed in time series analysis are briefly discussed including their limitations. Thereafter, a brief general overview of the theory of complex systems is provided and subsequently discussed in the context of marketing. Lastly, a case study is provided to argue why complexity based approaches might add value for causality rooted questions related to marketing.

2.1 Dynamic Effects in Marketing

Within marketing different phenomena might show dynamic behavior where the two most prominent examples are advertising effects and pricing and are therefore briefly discussed.

Dynamic effects of Advertising

One of the first studies that extensively examined the phenomenon of carry-over effects of advertising was conducted by Tull (1965). Tull (1965) provides empirical evidence and a model, which in essence shows that the advertising expenditures has effect on sales distributed over time, implying that the effect of advertising is delayed and the effect is spread over a longer period. One possible driver for such carry-over effects is brand loyalty, where advertising is a means to introduce a particular brand and therewith can initiate brand loyalty. A second driver for carry-over effects originates from the field of psychology referred to as the cummulated impression model. More specifically, this argues that the advertising impressions cumulate gradually which implies that by repeatedly showing an advertisement it reinforces the impression shown to the audience and therewith the impression becomes stronger. This subsequently means that the awareness of the brand and the products grow over time and eventually lead to a purchase but not necessarily directly after having been exposed to an advertisement (Tull, 1965). In line with Tull (1965) is the empirical evidence from Simester et al. (2009) whom conducted a field experiment. More precisely, a controlled field experiment was conducted for a large durable goods retailer where the specific interest was in the dynamic effects of advertising on the long-run. Two competing effects were found where the first was intertemporal substitution which leads to large short-term positive effects but negative effects on long-term. The other contrasting effect is that advertising tends to significantly increase future demand which can be attributed to goodwill effect. Which of these two effects predominates depends on the type of consumers, where specifically the most ordering customers, referred to as *best customers*, the intertemporal substitution effect clearly dominates. Hence, this study was the first to show that advertising might also have negative long-term effects on sales and again confirms the dynamic behavior of advertising (Simester et al., 2009). Next to these studies, the dynamic nature of advertising by means of demonstrating carry-over effects was again confirmed by the meta-marketing study conducted by Köhler et al. (2017).

Dynamic effects of pricing

Within marketing a commonly asked question is what the effect of a price reduction is on the number of sales. On short-term it is plausible to expect an increase in sales, especially for elastic goods. However, when measuring the effect of a price reduction on sales it is important to be aware of the dynamic characteristics pricing possess and that the long-term effects may be different than short-term. More specifically, Mela et al. (1997) provide empirical evidence on that consumers also become on the long-term more price sensitive, especially non-loyal customers where this group tend to grow when reducing prices. However, it is important to be aware of competition and that it is even a possible that price wars might erupt which can be harmful on the long-term (Heil and Helsen, 2001). In other words, when reducing prices and measuring only the effect at one specific moment in time does not provide an estimations of the full effect, since it might be underestimated as Mela et al. (1997) argue or overestimated when it erupts a price war for example. Moreover, the effect is also greatly dependent on the market structure whether it is a monopolistic or perfectly competitive market and the price elasticities of demand among other factors. In short, by changing the price and measuring the effects at one specific moment in time does not seem to provide an unbiased estimate due to the dynamic effects that play a key role in pricing (Rao, 2009).

2.2 Estimating Dynamic Causal Effects

2.2.1 The Distributed Lag Model

In order to estimate dynamic causal effects in time series analysis, one of the most intuitive methods to do this is by means of deploying a distributed lag model. The idea of estimating causal effects based on randomized controlled experiments can be adapted to a time series setting. More specifically, in a time series setting the same subject will act as the treatment and control subject but differing over time. In other words, in some periods the subject acts as the treatment unit while in other periods of time it acts as the control unit. In this way the dynamic causal effects can be estimated, since the effect over time on the variable of interest can be captured. In its simplest form, the distributed lag model can be represented as follows:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \ldots + \beta_{r+1} X_{t-r} + \epsilon_t \tag{1}$$

In equation 1 the variable X_t and its r lags are related to Y_t whereas the ϵ denotes the error term which captures all the effects of the omitted variables on Y_t. The β_1 denotes the immediate effect of a unit change in X_t on Y_t whereas in general X_{t-k} captures the effect of a unit change in X on Y, k periods later. Hence, the dynamic causal effect is the change in X_t on the current and future values of Y, so in other words the sequence of $\beta_1, \ldots, \beta_{r+1}$. The aforementioned approach mimics the concept of an experiment, except that in a time series setting the same subject receives repeatedly different treatment levels. However, this approach underlies two assumptions, firstly that the time series X_t and Y_t are jointly stationary, which in essence requires that the probability distribution of the time series variables do not change over time, allowing to use historical relationships to forecast the future. Secondly, the regressors (X in equation 1) should be uncorrelated with the error term ϵ_t, formally knows as exogeneity. This can be either exogeneity or strict exogeneity where the former implies that the conditional mean of the error term ϵ_t is zero given the current and past values of X_t, whereas the latter implies that the conditional mean of the error term ϵ_t is zero given current, past and future values of X_t. This exogeneity assumption intuitively implies that all the coefficients for the r lags included in equation 1 are non-zero and that lags beyond r are zero. Moreover, when the included regressors fulfill the exogeneity assumption, Ordinary least Squares (OLS) can be deployed as a means of estimating equation 1. Nonetheless, besides these two assumptions there are three more assumptions related to the distributed lag model that should be met, firstly that (Y_t, X_t) and (Y_{t-j}, X_{t-j}) becomes independent when j is large, implying that the distributions become independently distributed. Secondly, large outliers are unlikely and thirdly that no perfect multicollinearity is present.

The dynamic causal effect is otherwise known as the dynamic multiplier and refers to the respective estimated coefficients $\beta_1, \ldots, \beta_{r+1}$ in equation 1. In addition, the cumulative dynamic multiplier refers to the cumulative dynamic effect, which is the cumulative sum of the dynamic multipliers where for example the r-period cumulative dynamic multiplier corresponds to the sum of $\beta_1 + \ldots + \beta_{r+1}$. The cumulative dynamic multipliers can be estimated directly, which can be denoted, given equation 1, as follows:

$$Y_t = \phi_0 + \phi_1 \Delta X_t + \phi_2 \Delta X_{t-1} + \phi_3 \Delta X_{t-2} + \ldots + \phi_r \Delta X_{t-r+1} + \phi_{r+1} X_{t-r} + \epsilon_t \tag{2}$$

The cumulative multipliers are denoted by the coefficients, $\phi_1, \ldots, \phi_{r+1}$, in equation 2. Equation 1 and equation 2 are equivalent, where for example ϕ_0 corresponds to β_0, ϕ_1 corresponds to β_1 and ϕ_2 corresponds to the sum of β_1 and β_2 and so on. However, important to note is that the error terms in the distributed lag regression model as shown in equation 1 and 2 can be autocorrelated, which has the implication that the standard errors estimated by OLS are inconsistent and may therefore result into misleading statistical inferences. Hence, the heteroskedasticity and autocorrelation consistent (HAC) errors should be used such as the Newey-West variance estimator. In essence, given a simple distributed lag model with no lags the variance of the coefficient $\hat{\beta_1}$ can be estimated by multiplying the OLS standard error by a factor f_T. This factor f_T is defined as:

$$f_T = 1 + 2 \sum_{j=1}^{T-1} \left(\frac{T-j}{T} \right) \rho_j \tag{3}$$

This factor f_T adjusts the usual OLS standard error for the serial correlation in the error term. However, this factor f_T is not known, since it depends on the autocorrelation of $v_t = (X_t - \mu_x)\epsilon_t$ which is unknown, where $\rho_j = corr(v_j, v_{t-j})$. Nonetheless, Newey and West (1987) proposed to estimate equation 3 by:

$$\hat{f_T} = 1 + 2 \sum_{j=1}^{m-1} \left(\frac{m-j}{m} \right) \widetilde{\rho_j} \tag{4}$$

Equation 4 aims to seek a consistent estimator of f_T which is not straightforward, because if too many autocorrelations are included to be estimated then the result will be an estimator with a large variance while when too few autocorrelations are estimated it tends to ignore autocorrelations at the higher lags. This implies that in both cases the estimator tends to be inconsistent. Hence, it is proposed to include the number of autocorrelations depending on the sample size T, where the parameter m is the truncation parameter and depends on T such that in small sample sizes fewer autocorrelations are estimated and in larger sample sizes more autocorrelations are estimated, but much fewer than T. One rule of thumb proposed by Newey-West (1987) for m is as follows:

$$m = 0.75T^{\frac{1}{3}} \tag{5}$$

The resulting number from equation 5 is rounded to the nearest integer. In short, this study follows the Newey-West (1987) HAC estimators although this is just one of the possible guidelines that can be followed. However, besides estimating dynamic causal effects by means of the distributed lag model, a commonly deployed method in practice to test for causality is by deploying the Granger Causality test. Although Granger Causality does not estimate dynamic causal effects as the distributed lag model does, it remains to be one of the most widely used methods for identifying causality, hence it is discussed further.

2.2.2 Granger Causality

Granger Causality is a probabilistic concept of causality and is a widely adopted method for cause-and-effect inferences. It is in essence a statistical hypothesis test by means of using the F-statistic that tests whether the coefficients on all the lags of one of the regressors are zero. In other words, the null-hypothesis states that a particular regressor has no predictive information for the variable of interest, Y_t, beyond which the other regressors already contain. However, noteworthy is that the Granger Causality test does not account for *true* causality but rather *predictive* causality. More specifically, it tests whether one time series contains statistically significant information about the future value of another time series. Hence, if X Granger causes Y this implies that X seems to be a useful predictor, given all the other regressors, but this does not imply causality as it is commonly understood like in an experimental design. In its simplest forms, it can be tested whether X Granger causes Y and whether Y Granger causes X as shown in equation 6 and 7.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_q X_{t-q} + \epsilon_t \tag{6}$$

$$X_t = \gamma_0 + \gamma_1 X_{t-1} + \gamma_2 X_{t-2} + \ldots + \gamma_p X_{t-p} + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \ldots + \delta_q Y_{t-q} + \mu_t \tag{7}$$

Granger Causality subsequently tests the null-hypotheses of whether $\phi_1 = \phi_2 = \ldots = \phi_q = 0$ and $\delta_1 = \delta_2 = \ldots = \delta_q = 0$, corresponding to equation 6 and 7 respectively. As with the distributed lag model, it is assumed that the time series are stationary. Additionally, Granger Causality solely captures information about linear causal relationships and thus cannot capture non-linear effects. Lastly, another critical assumption is separability which implies that the cause and effect must be separable which is generally satisfied in linear stochastic systems (Granger, 1969; Yang et al., 2018). Nonetheless, it has been shown by Sugihara et al. (2012) that separability does not hold in dynamic systems that are not completely stochastic, which are systems where through time some information about past states is carried forward. Formulated differently, such a setting would occur when for example X is causal to Y while simultaneously information about X is captured in Y implying that the causal variable X does not have completely unique information. Subsequently, based on Taken's theorem (1981) if X influences Y then past values of X can be recovered from Y, also referred to as cross mapping. In such settings the Granger Causality test is not valid, hence Sugihara et al. (2012) propose the Convergent Cross Mapping method to still be able to identify causality in time series. More generally, methods like the CCM and its extensions are suitable for complex systems which tend to have the characteristic of being dynamic and not completely stochastic. However, the term complex systems is rather more general and constitutes of different complex phenomena that might occur which would undermine the more traditional deployed methods. Hence, the concept of complex systems and argumentation why it might be relevant to study in a marketing setting are further discussed.

2.3 Complex Systems Theory

2.3.1 The Principles of Complex Systems

The notion of complex systems can be understood broadly and does not entail one universal definition since it has seen numerous applications in various different disciplines, ranging from meteorology to physics and from sociology to economics. Nonetheless, more generally a complex system consists of numerous individual components that interact with each other, where the collective behavior in the system cannot be straightforwardly inferred by a reductive study of its single components. In other words, the behaviour in a complex system is generally difficult to model due to the presence and interactions of (causal) relationships, dependencies, emergence, feedback loops, feed forward and non-linearity among others. In essence, complex systems are generally called complex due to the fact that the relationships between the single components matters as much as the components themselves. However, the field of complex system theory is primarily focused on understanding such relationships rather than focusing on the single components. The concept of causality demonstrates well how a complex system differs from a non-complex system. More specifically, in figure 1 on the right it can be observed that A causes B which in turn causes C and is referred to as a causal chain. The concept of the causal chain is the primary focus for most research fields where chains are usually detected by means of discovering correlations. However, the concept of partial causality otherwise known as circular causality plays a key role in complex systems. A looped causal system is shown on the left in figure 1 where C could also causally affect A. Nevertheless, when a causal loop is encountered in experimental settings the researcher usually decides (implicitly) to cut the loop at a specific location in order to transform the causal loop into a linear causal chain. The main reason for cutting the loop is that analyzing a linear causal chain is much easier and correlations can be detected easier. However, the specific location of the cut can alter the conclusions, such as when for example the cut is made between A and B which would lead to analyzing the causal chain of B on C on A and could yield different results.

Nevertheless, in general there is no universal theory or framework that can explain complex systems in its entirety. The theory of complex systems has its roots in numerous different fields which imply that the deployed tools and methods are to a great extent domain-specific dependent. Nonetheless, some simple rules were discovered that characterize most, if not all, complex systems. More precisely, complexity can arise from these simple rules in two ways. The first one is referred to as *less is more* which means that the iteration of simple rules over time can lead to complex behaviour. Additionally, the second one is referred to as *more really is more* and refers to the fact that the interactions between many-bodied systems results into complex behaviour due to the fact that unpredictable behaviour is emerging.

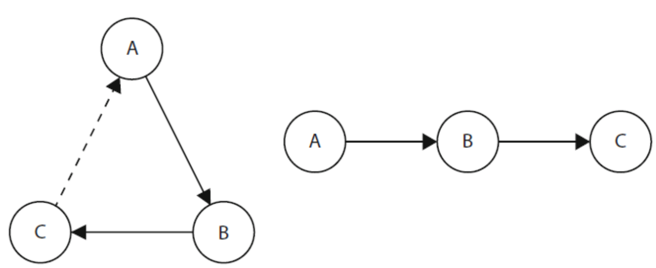


Figure 1: The right shows a causal chain as measured in an experiment, whereas the left shows a causal loop system which transforms into the causal chain if the line between A and C is not measured (Tranquillo, 2019)

One of the most prominent simple rules that were discovered shows that a change in a single parameter can lead to a wide variety of behaviours, where understanding such behaviours is one of the fundamental aims of complexity theory. The essence of this simple rule is famously illustrated by the exponential growth equation of the population. More specifically, the growth of the population can be represented by the differential equation shown in equation 8 where the x denotes the number of species and α represents the growth rate, which is in this setting exponential. In addition, the parameter K represents the carrying capacity which indicates that due to the presence of limited resources the population growth will eventually plateau down. Visually, equation 8 would be like a S-curve where initially for low values of x exponential growth can be observed but this growth is slown down to the carrying capacity value K which acts as the plateau value.

$$\frac{dx}{dt} = \alpha x(1 - x/K) \tag{8}$$

However, to investigate the dynamics of a differential equation it can be transformed into its discrete form so that it becomes dimensionless. This allows to specifically study the behaviour rather than the actual values. Hence, equation 9 denotes the discrete form of equation 8, where parameter r is a combination of α and K. So, equation 9 yields the time course of x where this time course depends on the initial parameter settings r and the starting value x. By varying the starting values for r and x different sequences can be discovered and thus shows that different behaviors might occur while being in the same system. This has the implication that depending on the initial starting conditions, x might be trending to a different static equilibrium.

$$x_{t+1} = rx_t(1 - x_t) \tag{9}$$

Another simple rule has been deduced from agent-based systems, which shows that a purely reductionistic approach can never completely describe a complex system. More specifically, when studying just one agent in isolation would be very simplistic and just uncovers changes in internal states but misses all the rules controlling the interactions between agents. A third simple rule shows that complex systems are oftentimes sensitive to the initial conditions and that by means of iteration in both space and time the dynamics of how patterns change over time can be uncovered. These three prominent simple rules are not the only simple rules but show the key features of complex systems (Tranquillo, 2019).

Differential Equations

Besides these simple rules, one of the most important tools to actually model real-world systems or to study system dynamics are by means of differential equations. Differential equations can be applied to any system that changes, where for example in some models there are several interdependent (coupled) differential equations as shown in equation 10. The term $\frac{dx}{dt}$ refers to how x changes over time t whereas the functions f, g, h are the rules stating by how much it changes. In system dynamics the term *state* refers to a snapshot of the variables in a system at a particular moment in time, where in equation 10 the snapshot constitutes of three variables at a particular moment in time t and is therefore labelled as a third-order system. Moreover, the functions f, g, h denote how this snapshot will subsequently change in the future.

$$\begin{aligned} \frac{dx}{dt} &= f(x, y, z) \\ \frac{dy}{dt} &= g(x, y, z) \\ \frac{dz}{dt} &= h(x, y, c) \end{aligned} \tag{10}$$

A general distinction between differential equations is whether it is a linear or non-linear equation. A linear system is in essence a superposition system, which implies that the system has homogeneity and additivity. More specifically, homogeneity refers to the fact that if the input is tripled then the output should also tripled and additivity indicates that when multiple different linear systems are combined that a larger linear system will be the result. However, this also shows that a non-linear system is labelled as non-linear when at least one term or function is non-linear. As opposed to linear systems, finding an analytic solution to non-linear systems is not so straightforward, hence two approaches exist that can be deployed to learn about the dynamics of the system while no analytical solution has to be found. The first approach is by means of numerical simulations where the differential equation is transformed into numerical approximations in an iterative procedure. However, oftentimes differential equations are denoted with time being the continuous variable, t, implying that there exists infinitely many time instants between two moments. Nonetheless, it is not feasible to compute infinitely number of data points, hence the differential equation is discretized which allows to solely compute the state variables at a given time instant. One commonly deployed numerical method is Euler's iteration:

$$\begin{aligned} \frac{dx}{dt} &= f(x) \\ x_{t+\Delta x} &= x_t + \Delta t f(x) \end{aligned} \tag{11}$$

This in essence shows how much x will change, given the current value of x and how much farther in time the prediction should be made for, Δt. In other words, based on the current value x_t, other future values of x can be found denoted as $x_{t+\Delta x}$. Graphically, this means that the slope of current point is used to predict the next point (Fieguth, 2017; Tranquillo, 2019).

Geometric approach and perturbations

Besides the numerical simulation approach, the second approach deploys geometric methods to understand the key terms within system theory and additionally provides more intuition. In its simplest forms, one could have a first-order non-linear system where $f(x)$ might take the form of $-x^3$. The key step is to find the equilibria which equals to finding the roots of $f(x)$. In such points it means that x does not change, since the differential equation equals zero ($\frac{dx}{dt} = 0$). Graphically, a phase plot can be constructed by plotting $\frac{dx}{dt}$ against x where subsequently the equilibria are denoted by the points where $f(x)$ intersects the x-axis as illustrated in figure 2. Furthermore, the behaviour near the equilibrium within a system can also be closely investigated by means of perturbation theory. Essentially, the idea is to examine what happens around the equilibrium when the system is slightly perturbed by an external force fiction I for a a particular moment in time. So, generally stated it can be written as:

$$\frac{dx}{dt} = f(x) + I \tag{12}$$

To illustrate the idea of perturbation theory, suppose again that $f(x) = -x^3$ and the perturbation equals I=1, then the effect of this perturbation on the equilibrium point is illustrated in figure 2. Given that the system is located in the equilibrium point, $x = 0$, it shows that if the perturbation causes x to be positive as indicated by *arrow 1*, then the change in x, as shown by $f(x)$, will be negative and therewith reduces the value of x as indicated by *arrow 2*. Moreover, if the perturbation has the result that x will be negative then the change in x will be positive and thus tends to move back to the equilibrium point where $x = 0$. Put differently, despite the effect of the perturbation, either positive or negative, the system moves back to the equilibrium point $x = 0$. Such equilibria points are referred to as stable equilibria or otherwise known as attractors of the system.

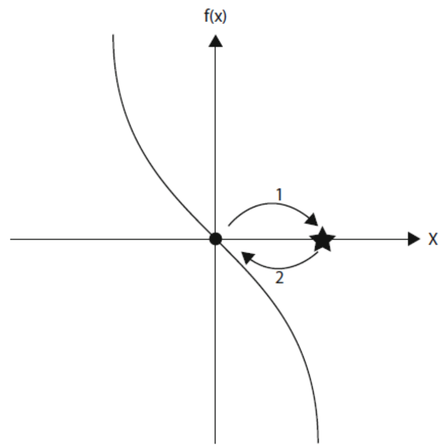


Figure 2: The plot of $\frac{dx}{dt} = -x^3 + 1$, where 1 denotes the perturbation. The slightly positive perturbation has a negative effect on the change of x given that x is positive and vice versa, which implies it has a stable equilibrium

However, depending on $f(x)$ a perturbation could also reveal an unstable equilibrium which is referred to as a *repellor* of the system where the perturbation has the result of moving the x away from the equilibria, which could be observed when for example $f(x) = x^3$. Additionally, a third option would be that the equilibrium is a saddle point, which can however only occur in non-linear systems such as when $f(x) = x^2$. Furthermore, the idea of attractors and repellors may be deployed as decision boundaries within a system, since states move towards or away from the equilibria in the system (Fieguth, 2017; Tranquillo, 2019).

Bifurcations, Velocity Vectors and Attractor Basins

A further approach to examine the equilibria points is by means of a Bifurcation plot, where two related key concepts are velocity vectors and the attractor basins. These key concepts can be illustrated by assuming that $f(x) = x^2$, where its equilibrium point is a sadle point and indicates in this example that if a perturbation decreases x then it will be pulled towards the equilibrium ($x = 0$), while when the perturbation increases x it will be pulled away from the equilibrium. Hence, to investigate this behaviour in more detail, velocity vectors can be used to evaluate each point of $f(x)$ to quantify the strength of how x moves towards or is pulled away from the equilibrium point, respectively for when x is negative and positive. The velocity vectors are denoted by arrows where its length corresponds to its strength, as illustrated in figure 3. The velocity vectors in figure 3 indicate that on the left of the -1 equilibrium the system pulls away from this equilibrium while between the equilibria -1 and +1 the system is pulled towards the +1 equilibrium. However, on the right of the +1 equilibrium the systems pulls again back to the equilibrium. Hence, +1 can be characterized as a stable equilibrium while -1 is characterized as an unstable equilibrium. The two equilibria combined define the basin of attraction, which in this illustration is defined by the interval -1 to $+\infty$ where the system in this interval will be attracted to the stable +1 equilibrium. More formally, a basin of attraction can be defined as a stable equilibrium that is being surrounded by unstable points, which subsequently regulate the basin size.

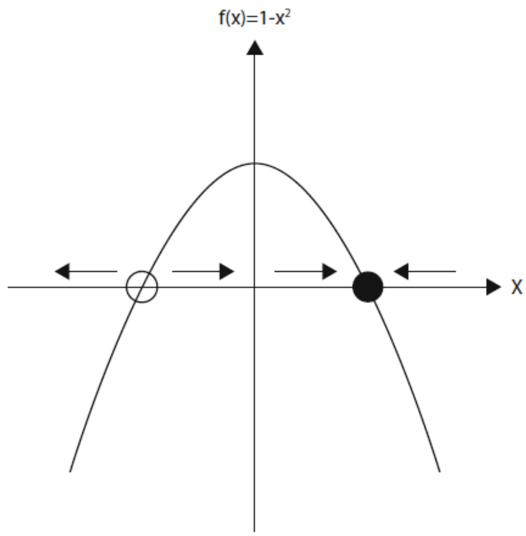


Figure 3: The plot of $\frac{dx}{dt} = 1 - x^2$, where the roots are +1 and -1. The velocity vectors are represented by arrows and indicate that on the left of -1 the system pulls away from the equilibrium, while between -1 and 1 the system pulls towards the +1 equilibrium. However, on the right of the +1 equilibrium the system is also pulled back to the equilibrium. Additionally, filled circles denote stable equilibria while open circles denote unstable ones.

After introducing the key ideas of velocity vectors and the attractor basins, the idea of bifurcation can be further elaborated. Given that $f(x) = x^2$, one can add a so-called bifurcation parameter, r, which in essence generalizes the previous example. So, the simple illustration would change to equation 13. The effect of this bifurcation parameter, r, can be observed in figure 4 where the parabola is vertically shifted. The main idea is that the bifurcation parameter determines how many equilibria points are present and that it can change its type, for example from unstable to stable. In other words, this shows that systems are sensitive to changes in the parameters and may impact its stability and therewith the behaviour of the entire system .

$$\frac{dx}{dt} = x^2 + r \tag{13}$$

When the bifurcation parameter r is lower than 0, two equilibria points can be observed as shown on the left in figure 4 where there is both a stable and an unstable equilibrium. Moreover, when $r = 0$ there will be a saddle equilibrium as before, but when r is larger than 0 there will be no equilibria in the system and x is pushed towards infinity.

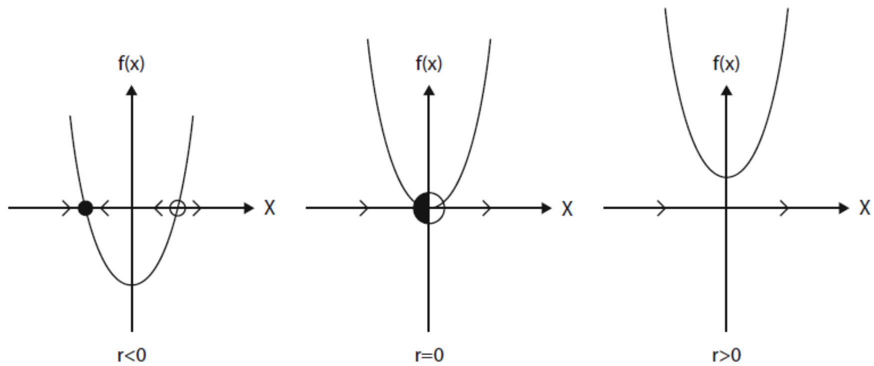


Figure 4: The phase plots for $\frac{dx}{dt} = x^2 + r$, where r denotes the perturbation parameter and differs per plot.

The behaviour of a system as a result of bifurcations can be effectively shown in a bifurcation plot. In this way, the locations of the stable, saddle and unstable equilibria are plotted for different values of r. As shown in previous example (figure 4), if the bifurcation parameter r is lower than 0 both a stable and an unstable equilibrium appear, which is illustrated in the bifurcation plot on the left of the the vertical axis (figure 5). The solid line denotes all the stable equilibria whereas the dashed line denotes all the unstable equilibria. With the same rationale, when r equals zero a saddle point emerges and is therefore referred to as a saddle-node bifurcation. However, when r is greater than 0 a different scenario occurs. More specifically, in such a setting the velocity vectors are of different strength since x tends to increase slowly around the point where $x = 0$ but increases much more rapidly when the initial value is very negative or very positive. This behaviour can be explained by the fact that the point $x = 0$ can be treated as a near equilibrium point or ghost equilibrium and thus implies that the velocity vectors around this point are smaller, which is referred to as the critical slow down. Hence, when a critical slow down phase can be observed it is an indication of a bifurcation which implies that a change in a (bifurcation) parameter changes the stability of the equilibria and therewith changing the behaviour of the system (Fieguth, 2017; Tranquillo, 2019).

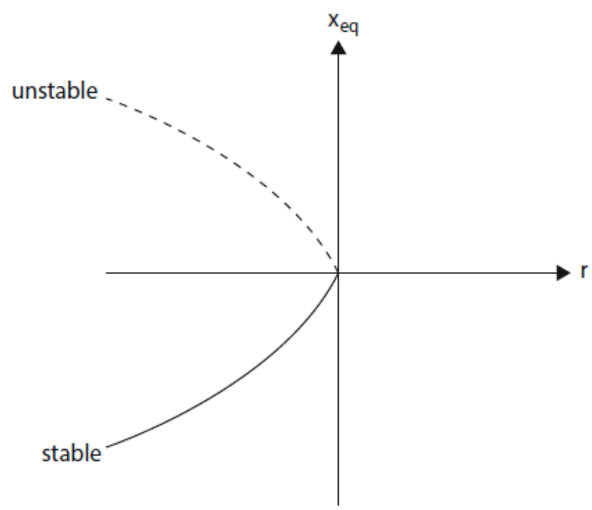


Figure 5: The bifurcation plot for $\frac{dx}{dt} = x^2 + r$, where r denotes the general perturbation parameter and varies

Hysteresis and Tipping Points

Next to the concepts of bifurcations, velocity vectors and attractor basins, two other important phenomena that occur in systems are: hysteresis and tipping points. Both concepts can be well illustrated by the system defined as

$$\frac{dx}{dt} = rx + x^3 - x^5 \tag{14}$$

Such a system results into a subcritical pitchfork bifurcation and is illustrated in figure 6. In such a subcritical pitchfork bifurcation, the equilibrium value for x is changed depending on the value for r. More precisely, between $-r_s$ and 0 three stable equilibria appear on the branches and x-axis. Additionally, two unstable equilibria appear where again the solid lines represent the stable ones and the dashed lines the unstable ones. Nevertheless, suppose the initial value of r is highly positive and x is located on the upper bifurcation branch. If the r value is decreased sufficiently it will move the x value to the left indicated by arrow 1. Subsequently, if r is decreased even further all the way down to r_s then the x value drops to the origin indicated by arrow 2. Moreover, if r is then increased again then x will remain on the horizontal axis until r equals 0, indicated by arrow 3. Thereafter, if r is increased further then the x value will increase again to the stable upper bifurcation branch. So, this example demonstrates the idea of hysteresis which relates to the memory of the system given that various stable equilibria occur. Specifically, hysteresis refers to that a certain state within the system is dependent on its history or past trajectories and thus implies that the system's output depends on where it initially came from. Hence, the past trajectory impacts which equilibrium eventually is selected in the system. The concept of hystersis also appears in economics, referring to the delayed effect of a certain event such as in unemployment. More precisely, it oftentimes happens that the unemployment rate still increases even though the economy is already recovering from a recession and thus shows the delayed effects of unemployment and is referred to as hysteresis (Jaeger and Parkinson, 1994). Nonetheless, no universal mathematical approach exists for the phenomenon of hysteresis and is considered as a very complex phenomenon to model.

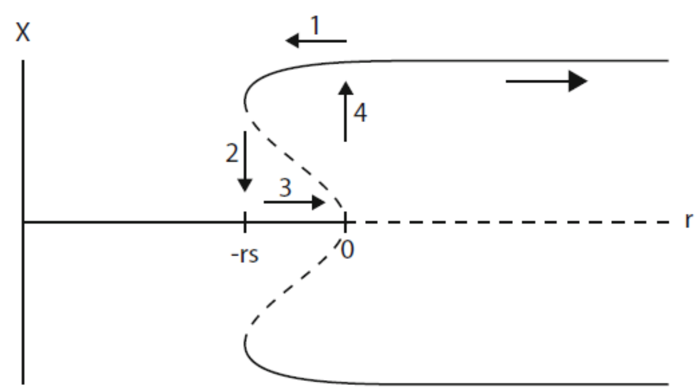


Figure 6: The bifurcation plot for $\frac{dx}{dt} = rx + x^3 - x^5$, where r denotes the general perturbation parameter and varies. Specifically, a subcritial pitchfork bifurcation is shown with three stable equilibria and two unstable equilibria as represented by the solid and dashed lines, respectively.

The second phenomenon that is observed in the example illustrated in figure 6 is that of a tipping point. It can occur that a certain parameter is changed, like r, but that the value of x does not change and thus it seems that the result does not change, as indicated by arrow 3. However, just after the point of $r = 0$ there is a sudden large increase in x hence this point is referred to as the tipping point. Such points show a sudden change but are oftentimes in real-world systems hard to predict and are perceived as to be out of the blue when they do occur.

Phase Space

Up until now various key concepts from system theory have been discussed, particularly in a one-dimensional system where nonetheless the same concepts also hold for higher-dimensional systems. However, when moving to higher-dimensional systems the possible behaviors within a system can change substantially and is therefore further discussed by starting with a two-dimensional setting followed by a three-dimensional setting. So, a general two-dimensional system can be denoted as

$$\begin{aligned} \frac{dx}{dt} &= g(x,y) \\ \frac{dy}{dt} &= h(x,y) \end{aligned} \tag{15}$$

In equation 15 the functions x and y are dependent on each other and is referred to as a coupled system, which imply that if one equation changes then the other one will change too. Naturally, a decoupled system would refer to a setting where $\frac{dx}{dt}$ would solely depend on x and $\frac{dy}{dt}$ solely on y. The variables x and y are oftentimes plotted against the variable time t. However, when the variables are plotted against each other, specifically excluding time variables, it is referred to as a *phase space*. Based on a phase space plot it can be observed how different variables, excluding time, interact or vary with each other and therewith revealing a certain trajectory. Plotting the trajectory of x and y in a real system can be challenging if possible at all, hence an alternative method is available where the trajectory of x and y can be reconstructed in a two-dimensional phase space by just a single variable. The essence is to deploy Taken's method and to construct a surrogate second variable from $x(t)$. Then the variable $x(t)$ will be plotted against its delayed version $x(t+\pi)$ where π denotes the delay. This idea can be generalized to higher-order systems where naturally also a higher-order surrogate variable needs to be created (Takens, 1981; Tranquillo, 2019).

Nullclines

So, based on a phase space plot one can identify the trajectories of certain variables, such as x and y in a two-dimensional phase space. However, the behaviour of a system can be further examined by means of so-called nullclines. The idea of nullclines can be illustrated by the following coupled two-dimensional system:

$$\begin{aligned} \frac{dx}{dt} &= x^3 - y + 3 \\ \frac{dy}{dt} &= y^3 + x \end{aligned} \tag{16}$$

As before, the interest is in finding the equilibria which could be found by solving the equations but another approach is by means of the nullclines. Each of the differential equations in 16 should be set to zero which allows to rewrite the equations into $y = x^3 + 3$ and $x = -y^3$, respectively. These rewritten equations are respectively known as the x-nullcline and y-nullcline and are illustrated in figure 7.

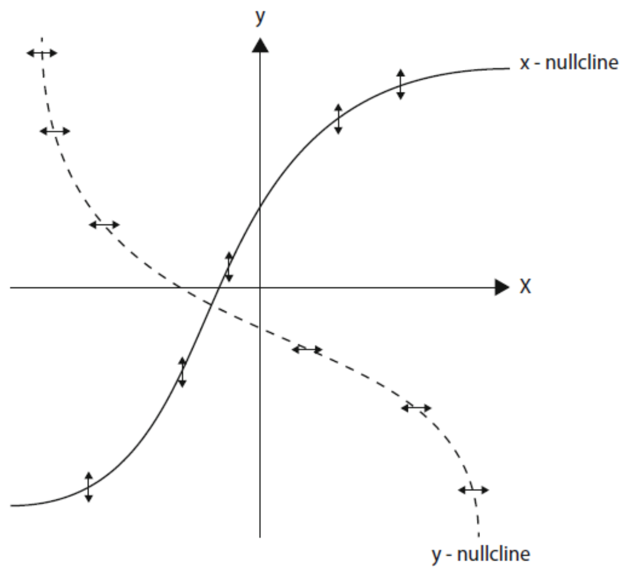


Figure 7: A two-dimensional phase space of variable x and y, where the solid line denotes the x-nullcline and the dashed line the y-nullcline.

More precisely, the x-nullcline denotes all the $x - y$ pairs where the system cannot change in the x direction or in other words where $\frac{dx}{dt} = 0$. With the same rationale, the y-nullcline denotes all $x - y$ pairs where the system cannot change in y direction, implying that $\frac{dy}{dt} = 0$. Moreover, the point where both nullclines intersect with each other denotes the equilibrium point.

Furthermore, to examine whether the observed equilibria are stable or not one can evaluate the velocity vectors as in the one-dimensional setting. More specifically, if the velocity vectors all point towards the respective equilibrium it can be considered as a stable equilibrium while when the velocity vectors are all pointing to different directions then the respective equilibrium would be characterized as unstable. More formally, a Jacobian matrix can be constructed which shows the slopes of the velocity vectors, as shown in equation 17.

$$\mathbf{J}_f(x,y) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial x} \end{bmatrix}_{x_{eq},y_{eq}} \tag{17}$$

The Jacobian matrix shows the first-order partial derivatives with respect to x and y for two functions, since it is a two-dimensional setting but naturally can be generalized to m functions. However, in this case f_1 and f_2 could refer to the functions $g(x,y)$ and $h(x,y)$ in equation 15 respectively. This Jacobian matrix specifically evaluates the equilibrium point referred to as (x_{eq}, y_{eq}). Intuitively, the resulting matrix shows the directions and magnitude of the velocity vectors in the x and y directions for the equilibrium point and contributes to determining the type of equilibrium (Fieguth, 2017; Panfilov, 2018; Tranquillo, 2019).

Separatix, Phase Portrait and Limit Cycles

The concept basins of attractions which was already introduced for one-dimensional systems can naturally be extended to a two-dimensional setting or to even higher dimensional systems. However, in the two-dimensional setting the basins of attractions are areas, as illustrated in figure 8. In this two-dimensional setting, the basins of attraction are separated by a so-called separatix which denotes an unstable line (dashed). This line can intuitively be seen as a border and depending on the initial conditions the starting point will be left or right of this line. Based on the initial starting point, the respective velocity vectors are followed towards the corresponding equilibrium within the respective basin. However, when the system is located in a certain basin it will remain there unless a substantial perturbation occurs that would move the system into another basin, where it essentially moves the system accross the separatix. This once again confirms a system's sensitivity to the initial starting conditions. More generally, figure 8 displays a phase portrait referring to the fact that it captures both the velocity vectors and nullclines into one plot.

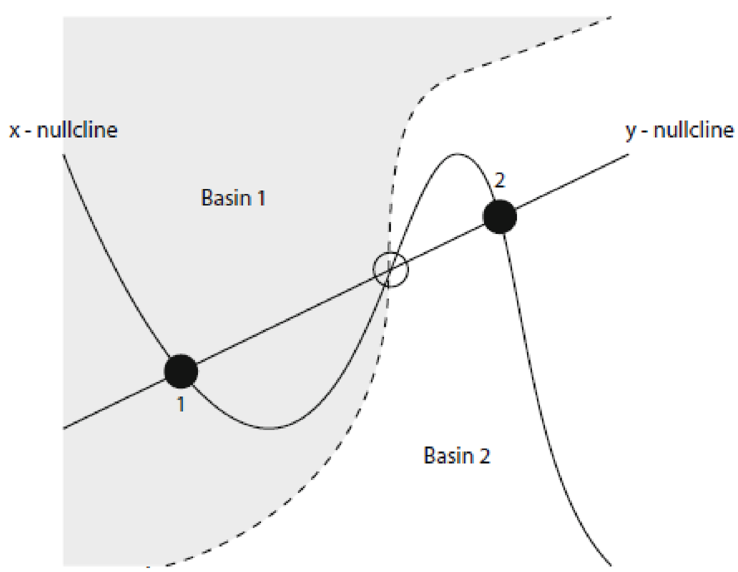


Figure 8: A two-dimensional system where the separatix is denoted by the dashed line and separates the basins of attraction. Additionally, the x-nullcline and y-nullcline are denoted by the solid lines next to the two stable equilibria (filled circle) and the unstable equilibrium (open circle)

Next to these concepts, another commonly observed phenomenon is a so-called limit cycle. More precisely, a limit cycle can be observed in phase space and can be illustrated by any *closed* shape such as shown in figure 9.

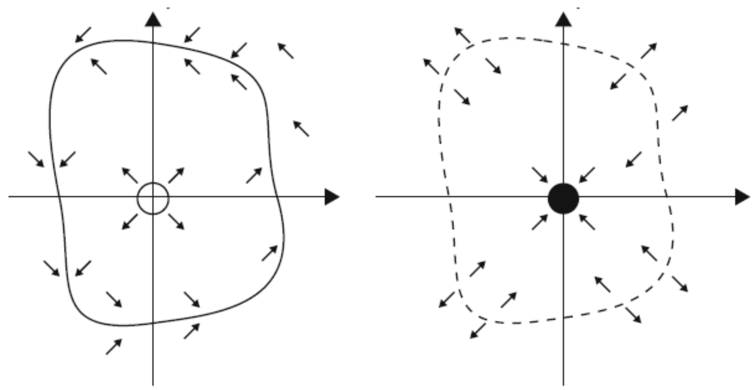


Figure 9: On the left a stable limit cycle is shown (solid line) while the right side shows an unstable limit cycle (dashed line/separatix)

So, as indicated by the velocity vectors in figure 9 on the left all close by trajectories are being attracted to the stable limit cycle, but when the trajectory actually lands on the limit cycle itself it will be looped indefinitely. Formulated differently, the same x-y pair values are continuously repeated for an indefinite period of time and is comparable to an oscillation. Moreover, it has been been mathematically proven that in a stable limit cycle an unstable point must reside. Intuitively this makes sense, because as aforementioned a stable limit cycle attracts all close by trajectories which means that when the initial start condition is within this limit cycle there must be a force that pushes the velocity vectors outwards. Hence, this force can subsequently originate from an unstable equilibrium that pushes the velocity vectors outwards.

Moreover, when a separatix forms a closed shape as illustrated on the right in figure 9, it forms an unstable limit cycle. As indicated by the velocity vectors, all points located nearby the unstable limit cycle will move away from this. So, with the same rationale as before, a stable point must reside within the unstable limit cycle so that the points that are located within this cycle are pushed away from the cycle path (Tranquillo, 2019).

Chaos Theory

After having briefly discussed the two-dimensional setting, it naturally raises the question whether the three-dimensional setting also introduces different phenomena that might occur in systems. Following previous notation, the setting can be extended to a third-order system:

$$\begin{aligned} \frac{dx}{dt} &= f(x, y, z) \\ \frac{dy}{dt} &= g(x, y, z) \\ \frac{dz}{dt} &= h(x, y, z) \end{aligned} \tag{18}$$

All the previously discussed phenomena that could occur in the one and two-dimensional system also appear in higher-order systems. However, a different phenomenon that can occur in third-and-higher-ordered systems is so-called *chaos*. A system with chaotic behaviour is very sensitive to perturbations and its starting conditions and is by definition deterministic. Nonetheless, quite some variables and systems were first thought of to be random, while they were actually deterministic but chaotic. More precisely, the main difference between randomness and chaos is that a a state in a random system cannot be fully accurately predicted while the state in a chaotic system can be predicted well *if* the initial conditions are completely known since small differences in the initial conditions can result into substantially different outcomes. In essence, chaos theory shows that a deterministic system might actually be unpredictable or that it may be very difficult to accurately predict a certain state in such a system. Hence, this also makes the boundary between randomness and a deterministic but chaotic system to some extent less straightforward in first instance.

Remarkable behavior of a chaotic system in phase space is that the trajectories do not stabilize to a certain value and nor does a limit cycle appear. So, a different kind of behavior can be observed which is referred to as the Lorenz attractor more commonly known as the strange attractor, which can only occur in higher than two-dimensional systems. In short, the values in the phase space do not go to infinity, are not repeated and the trajectories do not cross each other where such an example is illustrated in figure 10 (Tranquillo, 2019).

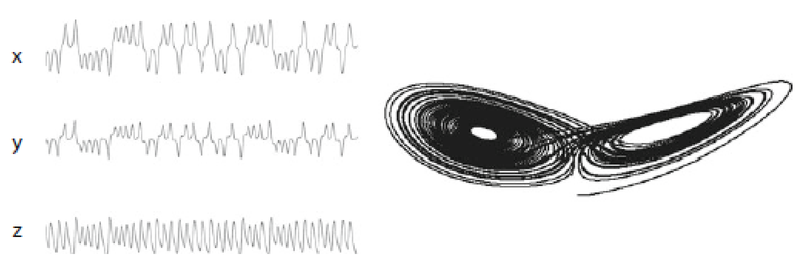


Figure 10: On the right the x, y, z phase space is illustrated and on the left the time courses of the respective variables. This is known as the Lorenz attractor

To further clarify the notion of chaos, a famous example is a bifurcation plot of the discrete logistic function, which was earlier shown in equation 9, and is illustrated in figure 11. In this plot the x demonstrates the state of the system where the state is evolving according to the equations $x_{n+1} = rx_n(1 - x_n)$. However, the dynamics of the system are highly dependent on the parameter r where for low r values (< 3) the system moves to a stable point. Nonetheless, for large r values (> 3.5) the systems begins to oscillate and the x values differ substantially. In essence, it is illustrated that the system is very sensitive to its initial starting conditions. The system first moves towards a single point, thereafter when r increases it oscillates between two points, then between four points then between eight and so forth. This is referred to as period doubling where the value of $r = 3.57$ is considered as critical value and from then onward characterized as chaos. At first, the values of x seem to be random but can actually be attributed to chaos since the system is moving to an infinity period oscillation as r increases. The speed of this period doubling has been proven by Feigenbaum (1978) to be at a constant speed of 4.669... comparable to the number π and is therefore referred to as the Feigenbaum constant. This finding is universally applicable to any equation and system that shows chaotic behavior (Fieguth, 2017; Giglio et al., 1981; Tranquillo, 2019).

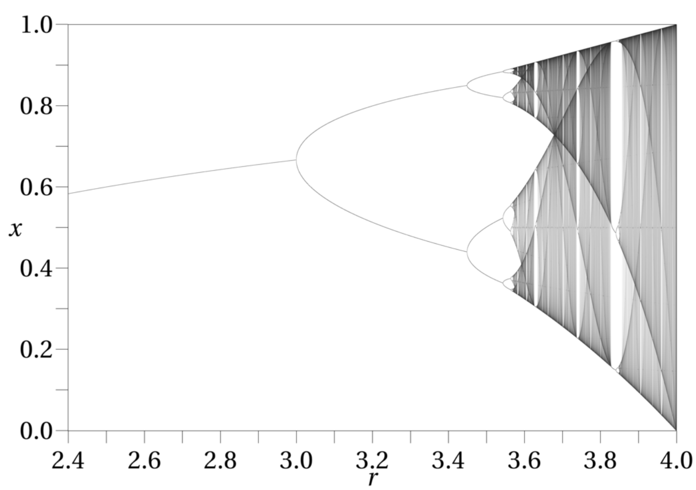


Figure 11: Bifurcation plot of $x_{n+1} = rx_n(1 - x_n)$, which displays chaotic behavior by means of period doubling. The phenomenon of chaos can be observed from the critical value $r = 3.57$ onward.

Additionally, it is argued that there are three fundamental features of chaos theory which are: non-linearity, system states and emergent order (Doherty and Delener, 2001). The *non-linearity* property was extensively earlier discussed, but primarily results from different feedback effects both positive and negative. Where positive feedback reinforces the effects while negative feedback dampens the effects. In other words, even a very small initial change could lead to a significant change in the system as was shown by Lorenz (1963, 1969) whom discovered this property by studying weather forecast models where this phenomenon is also referred to as the *butterfly effect*. The second property, *system states*, refers to structural or behavioural instability of the system. Structural specifically refers to the fact that the initial conditions have significant effects on the systems, which was illustrated in figure 11. Additionally, small changes in the systems may also lead to cumulative effects or butterfly effects in the behavior of the system. The third property, *emergent order*, refers to that the components of the system have an intrinsic self-organizing property. Moreover, Doherty and Delener (2001) argue that the required degree of non-linearity to result into chaos within marketing is relatively low due to all the assumptions regarding demand and advertising elasticities, competitors, interdependencies between marketing mix variables among others. In other words, due to all these components, variables and levels that are interdependent or interact with each other already with a low degree of non-linearity one can observe chaos in a marketing setting.

To conclude, up until now the most common phenomena that occur in systems have been discussed starting from from one-dimensional systems to higher-ordered systems. In general, when moving from one up to higher dimensional systems it is shown that simultaneously also more complex phenomena can be observed such as chaotic behavior. However, noteworthy to mention is that this part only touches upon the tip of the iceberg regarding the research field of complex systems but provides the understanding of relevant key terms. Nevertheless, the idea of complexity theory in relation to marketing, to the extent it has been researched to this date, will be further discussed.

2.3.2 Complexity Theory in Marketing

Applications in marketing

To this date it seems that very limited number of studies have been conducted yet to linking complex system theory to a marketing setting, which can be attributed to the fact that complexity theory and chaos are very difficult to identify and work with (Gregersen and Sailer, 1993). Nonetheless, there are a few studies that conceptualize the ideas from complexity theory to a marketing setting, such as Varnali (2019) where the customer journey is conceptualised from a complexity theory point of view. It is argued that customer service can be considered as a self-organizing system since the service environment continuously changes and no complete set of predefined rules can capture these dynamics. More specifically, it is argued that the customer journey is characterized by three fundamental characteristics of complex systems including being autopoietic, path dependent and non-linear. The customer journey is considered as autopoitic referring to the fact that it is based upon reiterative feedback loops in a network of interactions. For example, the effect of a customer's action at one touchpoint is recursively fed back to the customer and subsequently affects its future behavior. Moreover, a customer journey tends to be non-linear since customer experiences are very sensitive to small perturbations and when the customer service meets all the customer's expectations at all touchpoints, multiplier effects could be observed for the customer experience. Furthermore, the customer journey tends to be path-dependent since the order of events can substantially affect the non-linear effects on customer experiences. Hence, the customer journey tends to fit the definition of a complex system as argued by Varnali (2019) and subsequently offers the possibility to deploy different tools that are specifically designed for complex systems (Varnali, 2019).

Another application of complexity theory in relation to marketing is in the setting of new product diffusion models. More precisely, Goldenberg et al. (2001b) show that the stochastic cellular automata simulation technique, borrowed from complex system analysis, can be leveraged to model consumer's heterogeneity on new product growth. In essence, the assumptions made about individuals and how it influences the aggregate level parameter values can be better understood, which implies that also the limitations of the aggregate analysis can be more accurately interpreted. The key idea of deploying the automata simulation technique is that it can generate data on an individual level, which is generally not available, and therewith allowing to model heterogeneity effects. Besides the new product diffusion models application, the cellular automata simulation technique has also been applied to understand the underlying process of word-of-mouth marketing (Goldenberg et al., 2001a). More specifically, it helps understanding the aggregate effect of the spread of information via word-of-mouth, through so-called *weak* and *strong* social ties. Where weak ties refers to less personal communication an individual makes to acquaintances and strong ties refers to the strong personal communication an individual makes to closer friends. By means of deploying the cellular automata technique, data could be generated to research whether any apparent differences or patterns between strong and weak ties could be observed. More generally, it is argued that methods like the cellular automata technique borrowed from complex system analysis can be leveraged to analyse various marketing phenomena (Goldenberg et al., 2001a, 2001b).

Chaos in Marketing

Next to the above discussed marketing applications there is also more specific literature, although again limited, about applying chaos theory in marketing settings. One article that pioneered in relating chaos theory to marketing was by Hibbert and Wilkinson (1994). Hibbert and Wilkinson (1994) argue that traditional marketing methods overlook the complex dynamics that even occur in the simplest non-linear equations and that in the field of marketing the primary focus is too much on equilibrium approaches and linear approximations. Chaos theory offers managers and policymakers new ways of understanding complex phenomena rather than assuming it to be external shocks or noise and can subsequently be used to better account for certain patterns or behavior. Additionally, chaos theory can be deployed to distinguish purely randomness from chaos, which is of added value since purely random processes cannot be predicted while chaotic systems can be predicted on short-term but just not on long-term. Lastly, chaos theory can be utilized to find new ways of explaining the evolution of certain marketing systems. Hibbert and Wilkinson (1994) show three different models that result into complex behavior and chaos where the three models correspond to three different settings, namely: new product diffusion models, market evolution and brand competition. Besides showing three marketing case studies that result into complex behavior and chaos, it is showed how chaos can be distinguished from randomness by means of deploying one of the following methods: return maps, correlation dimension, the Lyapunov exponent and prediction error. The essence is that chaos theory can be of substantial value to distinguish randomness from chaos because within marketing traditionally all time series that seem to appear random are attributed to random shocks or noise while this is not automatically true and might actually indicate there there are underlying (complex) non-linear dynamics. If so, then various methods can subsequently be deployed to investigate the origin of these dynamics and why this behavior can be observed (Gregersen and Sailer, 1993; Hibbert and Wilkinson, 1994). Although Hibbert and Wilkinson (1994) already pioneered with the idea of applying chaos theory for marketing phenomena and showing all the opportunities it offers, the amount of literature on this subject is to this date is still very scarce. The main explanation for this as stated by Hibbert and Wilkinson (1994) and Gregersen and Sailer (1993) is due to the fact that chaos very difficult to identify and to work with such models/theories.

Nevertheless, about 8 years later one other study appeared on relating chaos to marketing which was done by Doherty and Delener (2001). More specifically, Doherty and Delener (2001) use the laws of chaos to investigate the evolution of high-tech markets, where such markets are characterized by rapid advancements, globalization, fierce competition, risky strategies and freer international trade among others and thus are considered to be a complex system of which behavior can be explained by means of chaos theory. The three fundamental features of chaos are followed by Doherty and Delener (2001), which were earlier discussed. The first property *non-linearity* can be attributed to the presence of both positive and negative feedback effects, such as the growth of complementary products or learning curve effects among others. The second property *system states* can be traced to aspects like the initial structures of firms or the market. Additionally, small changes can occur due to changes in for example policies, regulations or entries/exits of firms. Moreover, the third property *emergent order* can be related to the social order that is emerging from actions of economic agents. Doherty and Delener (2001) extend the two factor NK model, which is based on the laws of chaos and evolution, to a four factor NK model. This four factor NK model accounts for various growth and adaptive activities that are observed in the business environment. The main idea is to model the evolution of the business environment where the four factors capture the interdepencies within and among firms. This model demonstrates that decentralisation regarding firm structure and decision-making is important, since without decentralization it is shown that the company tends to perform suboptimally in the market. A company should divide itself into N number of units where N should be chosen when it is located at the edge of chaos. In other words, the number of interdependencies within a firm should be sufficiently low so that the optimization of each seperate unit results into the optimization of the complete system (company) while being subject to the overall company mission. Besides the organizational structure, this model also shows that innovation is a autocatalytic process implying that it leads to the creation of multiple niche markets where subsequently each niche market again leads to innovation and subsequently to even more niche markets and so forth. Another insight that was obtained through this model is that segmentation, differentiation and positions tends to limit the change to chaos, since a higher degree of independence can be achieved. Lastly, it is shown that no long-term forecasts can be made about a complex system due to the fact that too many variables and interdependencies are present in such systems. Nonetheless, in summary, Doherty and Delener (2001) show that the laws and models of chaos can be utlized to model the evolution of a business environment and subsequently encourage to further research chaos theory in relation to marketing.

Complexity Economics

Besides the aforementioned marketing applications, there is also more general literature available on how complexity theory is relevant for the field of economics, where economics is closely related and interrelated to the field of marketing. Within the field of economics, since last decades, a different way of thinking about the economical science in its entirety has emerged. Namely, complexity economics, which sees the economy not as being in an equilibrium but rather as a dynamical system where economic agents continuously adapt. In other words, it is a different view of the economy where the system is continuously evolving and phenomena occur that have not been observed earlier in the traditional neoclassical approach. This approach does not exclude the idea of the presence of equilibria, but rather that the economic system is not necessarily in equilibrium due to the fact that it is an organic evolving system and thus actually generalizes the traditional approach by including nonequilibrium economics. Within complex systems many different elements interact with each order and co-create certain patterns. This same idea can be observed within economics, where economic agents continually change their market moves like pricing, forecasting, buying decisions and so forth. The key difference between the general equilibrium approach and the complexity economics approach is that instead of asking how prices and quantities are consistent with their overall pattern of pricing and quantities, complexity economics asks how strategies and actions may endogenously change with the patterns they create (Arthur, 2013).

To summarise, despite the fact that there are some applications of complexity theory and chaos theory in marketing, it is still very scarce and almost no recent literature is available on this subject, which can be attributed to the fact that the concepts within complexity and chaos theory are difficult and it is not straightforward to translate such concepts to marketing settings according to Doherty and Delener (2001), Hibbert and Wilkinson (1994), and Wilkinson and Young (2013). Nonetheless, because of this scarcity of literature on this subject, the aim of this study is to provide new evidence. More precisely, this study focuses on how causal effects might more accurately estimated by treating certain marketing settings as complex systems which subsequently allows to deploy more advanced methods to capture the causal relationships.

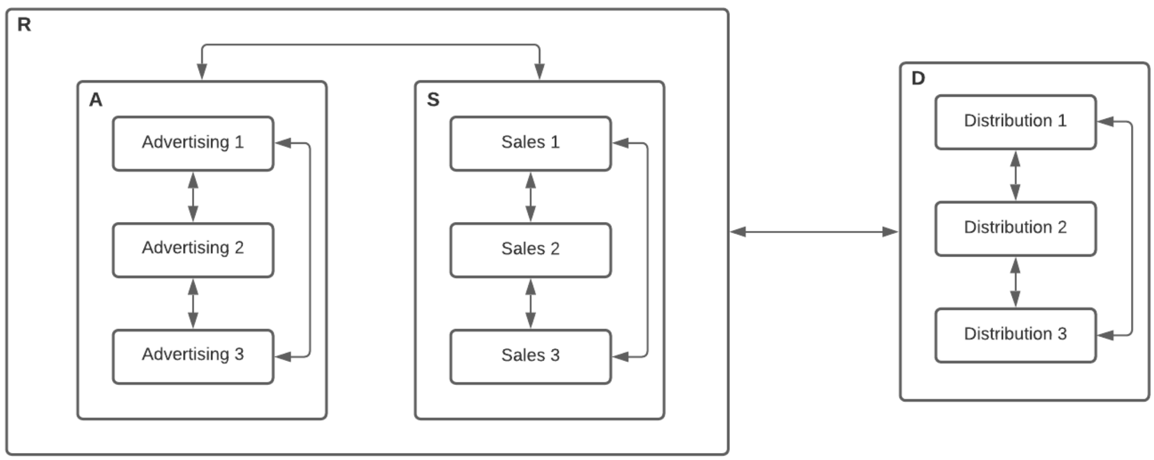
2.3.3 Case Study: Complexity Theory in Advertising

As aforementioned, this study illustrates how complexity based methods might be leveraged to more accurately estimate causal effects for different marketing mix related questions. However, the first step is to evaluate whether a particular marketing setting seems to fit the definition of a complex system, although there is not one universal definition that can be followed to easily label a setting as a complex system or not. Nonetheless, there are various important features that typically characterize complex systems where the most prominent characteristics are: non-linearity, there are numerous single components that interact with each other, might be a self-organizing system and presence of emergence (Varnali, 2019).

So, suppose a setting of a large company with three different products where for each separate product the following information is known: the price, distribution amount, the advertising efforts, and feature. A question of interest to a marketeer could be for example what the effects of pricing has on sales or what the effect of advertising has on sales. Traditionally, a marketeer might evaluate the causal effect of the advertising efforts of product 1 on the sales of product 1 to better understand the causal impact of advertising. Such a reductionist approach where elements are isolated is commonly deployed within marketing but is not necessarily the most plausible approach. For example, in the setting with three different products, that can be either complementary or substitutes, isolating certain components does not capture the complete picture. More specifically, it is likely that the advertising of the different products will interact with each other and subsequently affect the respective sales of each product. In other words, solely looking at the advertising efforts of one specific product and attributing all the observed effects to that might give a biased view of what is truly happening. By approaching this setting in such a reductionist approach would seem like that there is a linear effect of advertising on sales, while this may be attributed to the fact that all the other dynamics are not taken into account. More precisely, in this setting it seems plausible to assume that the interactions between the sales and advertising result into non-linear effects, which was comparably also argued by Doherty and Delener (2001) whom stated that complementary products are a source of feedback effects and thus result into non-linearity.

To clarify the idea even further of why treating certain marketing settings as a complex system might add value, a simple example is illustrated in figure 12 which might be to some extent oversimplified but demonstrates the main motivation of a complexity theory based approach. So, suppose again the setting as described above where three complementary products are offered and each product has its own advertising efforts and distribution. Moreover, assume that the question of interest here is to find out what the causal effect of advertising 1 is on the sales of product 1, where the numbers 1 to 3 of each variable refer to respective product. As aforementioned, reductionist approaches are traditionally deployed within marketing where this idea in its simplest forms is illustrated in (a). In other words, it is investigated how advertising 1 causally impacts sales 1, where subsequently different models might be explored that also control for the other variables. However, controlling for the other variables does not ignore the fact that there might actually be more complicated dynamics present. More specifically, the setting outlined here might actually behave as illustrated in (b) which shows that deploying reductionist methods is not straightforwardly the best choice and cannot capture the actual dynamics of the system that may be present.

((a)) Simple idea of evaluating causality of advertising on sales



((b)) The system might actually be less straightforward and might involve non-linearity and numerous interdependencies among sales, advertising channels and distribution

Figure 12: Here a very simple idea is showcased of a reductionist approach (a) for evaluating causality compared to how the system might actually behave and could be treated as a complex system

An illustration of likely occurring behavior in this setting is shown in (b), which shows that multiple components are interdependent and that non-linearity seems plausible. More precisely, it is likely that the sales of the products are interdependent, especially if the products are complementary, resulting into feedback effects and thus non-linearity, as argued by Doherty and Delener (2001). Moreover, it is also plausible to assume that advertising efforts for the different products interact with each other, due to different reasons. First, because there will be a budget where this budget has to be allocated and the budget allocation might change over time. Secondly, because when product 2 is advertised heavily then it is likely that, perhaps more indirectly, that product 1 and 3 might also benefit from this due to the branding, which might be even reinforced due to the complementary nature of the products. Next to the fact that between advertising efforts A and sales S such interdependencies are likely, it is also plausible to observe bidirectional (causal) effects between advertising and sales, since earlier research found such a relationship between advertising and sales (e.g. Lee et al., 1996; Sharma and Kapur, 2014). Besides, due to the complementary nature of the products it is likely that for example the advertising of product 2 will also impact the sales of product 1, both directly and indirectly through the sales of product 2. Moreover, another component where both sales and advertising (R) might interact with is the distribution D. Given that distribution is constrained by certain capacities it seems likely to find interdependencies between the different products. Additionally, distribution also interacts with sales and advertising, because if the distribution is low then naturally also the sales can be lower due to out-of-stock scenario's even though when for example the advertising efforts are increased. Furthermore, it can also be argued from the other way around, so sales and advertising can also lead to changes in the distribution. Specifically, the allocation within distribution could shift among the products depending on performance but it can also occur that the sales significantly increase and would lead to a permanent increase in the distribution capacity. The effect of this distribution expansion can flow back to the sales, since an out-of-stock scenario is less likely and therefore might also result into higher sales given that the demand is higher than supply.

To summarize, in a setting with just three products that seem to be straightforward to interpret might actually show more complex underlying dynamics such as non-linearity and numerous components that interact with each other. Here just a simple example was provided to show that most marketing settings can be actually more complex than initially thought of. However, labeling a particular setting as a complex system or not is not that straightforward and there is not one test or something a like to determine this. Nonetheless, domain specific knowledge can be utilized to map a particular situation and although not all interdependencies and relationships can be formally tested it can still give a good overview of whether it would make sense to investigate complexity based methods further for that particular setting. Additionally, direct non-linearity between variables can be evaluated by means of visual inspection or different tests and to identify the presence of chaos different test exist but again the system as a whole, feedback effects or more complicated phenomena alike cannot be formally tested and would also require an expert opinion to judge whether it is worth to investigate the complexity based phenomena further.

3 Methods

In this section the method Cross Convergent Mapping originating from complexity theory is discussed. This method is designed to detect causality between (time-series) variables even when complex phenomena like feedback effects or non-linearity among others are observed. Moreover, this method follows the principle of *"a lack of correlation, does not imply a lack of causation"* since this method is still able to identify causality even though there is a lack of correlation as stated by the authors Sugihara et al. (2012).

3.1 Cross Convergent Mapping

Cross Convergent Mapping (CCM) is a (non-linear) state space reconstruction method and distinguishes correlation from causality by means of leveraging the properties of delay embedding (Sugihara et al., 2012). Within complex systems numerous general complex phenomena can be observed, as the term *complex* logically implies, and CCM is able to deal with such settings. More specifically, a commonly observed phenomenon within complex systems is referred to as *mirage* correlation, which means that the correlation depends on the state of the system and thus that the sign and magnitude changes over time. However, correlation is not a necessity nor sufficient to establish causation which is demonstrated by one of the most popular methods to detect causality in time series, namely Granger Causality. Granger Causality is commonly deployed as a means to detect causality between time series variables, although Granger Causality utilizes predictability for defining causality rather than *true* causality as understood from experimental settings. Despite the fact that Granger Causality is a commonly deployed method for (*predictive*) causality, it is crucially dependent on its separability assumption. As earlier discussed, separability in short refers to the fact that a variable x contains independent unique information about the target variable y. However, this assumption cannot be satisfied when the variable y also captures information about x since this would imply that x is not a unique causative factor to y. In other words, it seems counter intuitive but when x is unilateral causing y ($x \implies y$) then the variable x can be estimated based on y but not vice versa. In essence, non-separability allows to predict past states of the causative factor where this concept plays a fundamental role within CCM, since this method aims to deal with such settings. Specifically, systems that are both linear and fully stochastic generally satisfy the separability assumption, but otherwise it usually fails to satisfy this assumption. To further clarify the idea behind non-separability, a simple example with a logistic model is illustrated.

$$\begin{aligned} X(t+1) &= 3.9X(t)[1 - X(t) - \beta Y(t)] \\ Y(t+1) &= 3.7Y(t)[1 - Y(t) - \beta X(t)] \end{aligned} \tag{19}$$

In this model the X and Y are dependent on each other as long as $\beta \neq 0$, more generally the parameter β determines how sensitive X is to Y where a large value substantially decreases the growth of X. However, the equations in 19 can be rewritten to show that the values $Y(t+1)$ and $Y(t)$ can be used to obtain the value for $X(t)$ and the values $X(t+1)$ and $X(t)$ can be used to obtain the value for $Y(t)$. Hence, after algebraically rewriting the equations in 19 gives:

$$\begin{aligned} \beta Y(t) &= 1 - X(t) - \frac{X(t+1)}{3.9X(t)} \\ 0.2X(t) &= 1 - Y(t) - \frac{Y(t+1)}{3.7Y(t)} \end{aligned} \tag{20}$$

However, a problem of the equations in 20 is that it is dependent on future values, respectively $X(t+1)$ and $Y(t+1)$ hence these equations alone are not sufficient to retrieve the cross mapping dynamics. Therefore, equation 20 can be substituted back into equation 19 to obtain equations with lagged values, as shown in equation 21.

$$\begin{aligned} X(t) &= \frac{3.9}{0.2}\left[(1-\beta Y(t-1))\left(1 - Y(t-1) - \frac{Y(t)}{3.7Y(t-1)}\right) - \frac{1}{0.2}\left(1 - Y(t-1) - \frac{Y(t)}{3.7Y(t-1)}\right)^2\right] \\ Y(t) &= \frac{3.7}{\beta}\left[(1-0.2X(t-1))\left(1 - X(t-1) - \frac{X(t)}{3.9X(t-1)}\right) - \frac{1}{\beta}\left(1 - X(t-1) - \frac{X(t)}{3.9X(t-1)}\right)^2\right] \end{aligned} \tag{21}$$

Equation 21 shows the cross map estimates of $X(t)$ and $Y(t)$, respectively, where $Y(t)$ cannot be solved when $\beta = 0$ while $X(t)$ remains to be defined everywhere. This can also already be observed in equation 20 where $\beta Y(t)$ can clearly not be solved in terms of X given that $\beta = 0$ and implies that the past states of X are not relevant to determine Y. More importantly, in a bidirectional setting ($\beta \neq 0$) the equations of 21 can be substituted back into equations 19 so that $X(t+1)$ can be written as a function solely in terms of $X(t)$ and $X(t-1)$ and thus shows that the variable X can be predicted only based on its past values. However, this touches upon the fundamental issue, because the Granger Causality method would eliminate the variable Y from the model since it does not affect the predictive performance because $x(t)$ can be predicted based on its lagged values and thus y would be classified as redundant. Subsequently, Granger Causality would *incorrectly* conclude that Y does not cause X and this is one of the key points CCM can address.

In short, CCM can deal with systems that tend to be non-separable, CCM can also identify causality even when the variables are low to not correlated and CCM can distinguish among interactions from shared driving variables. Moreover, Granger Causality specifically experiences issues when the system is not purely stochastic as aforementioned which is also addressed by CCM. More specifically, when the system is not purely stochastic then there exists an underlying manifold that governs the system dynamics. So, suppose two variables X and Y then in dynamical systems theory it is stated that these variables are causally related if they are from the same dynamic system which implies that these variables share a common attractor manifold M. The implication of this is that the variables contain information about each other where, as illustrated before in equations 19 to 21, the information of past states of X can be retrieved from variable Y. This already touches upon the fundamental idea of Cross Convergent Mapping, since this method tests to what extent historical values of Y can be used to estimate past states of X which can only occur if X causes Y. More technically, CCM inspects whether there is correspondence between the shadow manifolds M_y and M_x that are derived from the manifold M. More precisely, it is tested whether the time indices of closely located points on M_y can be used to identify closely located points on M_x because if that is possible then this implies that the variable Y can be used to estimate X and the other way around. To clarify this idea and the concept of manifolds further it can be best illustrated by means of an example. More specifically, the famous Lorenz System is commonly used to illustrate such concepts including demonstrating the CCM method, where not the idea behind the Lorenz system in itself is relevant but rather the general rationale of what exactly is happening. So, the Lorenz system is a coupled non-linear third-order system which was originally designed to model the air movement in the atmosphere and was additionally the first one to discover the concept of chaos with this particular system, which was extensively discussed in the literature review. Nevertheless, the Lorenz system is denoted as follows (Lorenz, 1963):

$$\begin{aligned} \frac{dx}{dt} &= \sigma y - \sigma x \\ \frac{dy}{dt} &= rx - y - xz \\ \frac{dz}{dt} &= zy - bz \end{aligned} \tag{22}$$

This system and the idea of the CCM method can subsequently be illustrated graphically, as shown in figure 13 (Sugihara et al., 2012). So, in this system each equation (component) is dependent on the state and dynamics of the other components and thus each component naturally changes when the other two components changes. In other words, a three-dimensional setting is here displayed where each dimension refers to the respective equation, referring to the variables X, Y and Z. The manifold, M, refers to the set of all trajectories and are simply denoted by the solid *lines* in figure 13. Moreover, $m(t)$ refers to a particular point on this manifold and is therefore denoted as $m(t) = [X(t), Y(t), Z(t)]$. Subsequently, for each respective dimension the time series for that particular dimension can be projected. Graphically, this would imply that for each *edge* of the state space, where there are three in this setting (X, Y, Z), a time series can be projected from the manifold to each respective axis. This would yield in total three time series where for each dimension the placement of that respective variable on the manifold is projected, over time. However, the idea of the manifold M can also be viewed from the other way around. So, for each respective variable X, Y and Z there is a time series and when plotting these three time series simultaneously in a state space, the manifold M can be recreated. Subsequently, based on the manifold M the shadow manifolds M_x and M_y can be derived which are constructed based on the lagged values of X and Y, respectively, where the lagged values are denoted by τ in figure 13. The idea of creating shadow manifolds originates from Takens (1981) and was generalized by Deyle and Sugihara (2011), where it is mathematically proven that by looking at one of the projected time series of the original manifold, a shadow manifold can be created. So, the original time series projection and its lagged values act as coordinates and construct a particular shadow manifold, which shows similar trajectories as the original manifold as observable in figure 13. As an example, M_x displays the shadow manifold for the time series X where each point on this shadow manifold captures time segments and essentially shows the historical behavior of the variable X. Furthermore, the shadow manifold remains to capture the essential mathematical properties of the original manifold such as its topology. Hence, the strength of using shadow manifolds is that just a single time series variable and its lags can be used to recover states from the original manifold (Deyle and Sugihara, 2011).

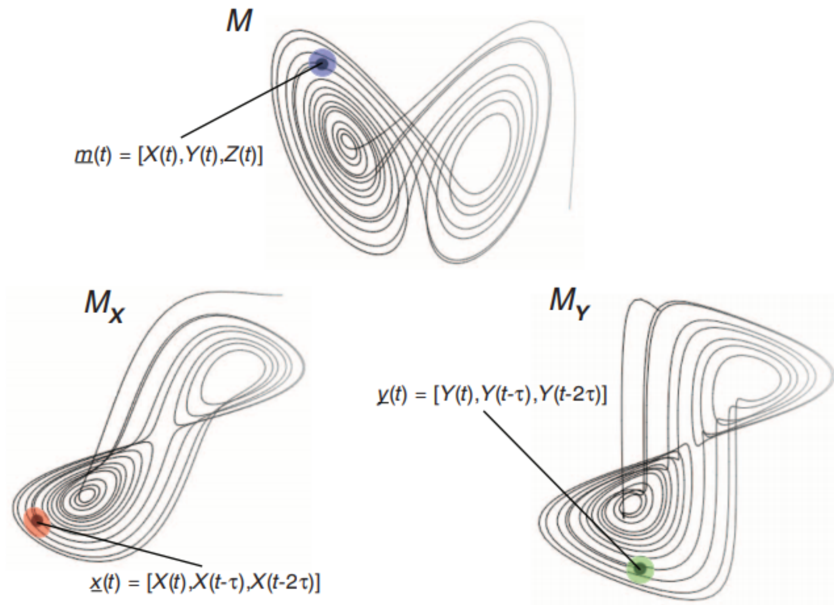


Figure 13: Here the Lorenz system is displayed as M which refers to the attractor Manifold. However, Cross Convergent Mapping investigates the shadow manifolds of variables X and Y, referred to as M_x and M_y respectively. More specifically, the correspondence between M_x and M_y is tested where the states of X might be estimated based on Y and the other way around

The idea of creating shadow manifolds is commonly known as (non-linear) state space reconstruction which allows to work with low-dimensional proxies of the real system. Moreover, the key idea is that points located on the shadow manifold of X (M_x) correspond to close by located points on the shadow manifold of Y (M_y). This would be possible since the variables X and Y form, in this example, a coupled system and refers back to equation 22. More specifically, this idea is further illustrated in figure 13 where for example points located in the red circle on M_x correspond to points located in the green circle on M_y and implies that these points will have the same values for t, so the time indices. Hence, based on this idea the state of each variable can be estimated based on the other, so Y can be utilized to estimate the states of X and vice versa. Next to the idea of cross mapping, convergence also plays a vital role in the CCM method since convergence is the factor that discriminates correlation from causation. More precisely, with more data (longer time series) the cross-mapped estimations are more accurate due to the fact that the estimation error declines. Another feature of the CCM method is that it deals with transitivity. More specifically, causation is transitive and implies in the bidirectional case that $U \Leftrightarrow V \Leftrightarrow W$ and in the unidirectional case that $U \Rightarrow V \Rightarrow W$. In other words, even though U and W do not directly interact with each other they are still causally linked to each other. However, the CCM method can differentiate between coupled variables and variables that share a common driver, where two of such possible scenario's that CCM can deal with are illustrated in figure 14 (Sugihara et al., 2012).

((a)) CCM can identify that W and V are actually decoupled and driven by U

((b)) CCM can also differentiate shared drivers from coupled variables in large interaction networks.

Figure 14: CCM can differentiate between coupled variables and common drivers and can distinguish unidirectional from bidirectional relations. Adapted from Sugihara et al. (2012)

.

3.1.1 Cross Convergent Mapping Algorithm

After having discussed the key ideas and intuition of Cross Convergent Mapping (CCM), it is now further elaborated how CCM is actually applied in an empirical setting. More precisely, in practice *empirical dynamic modelling* is applied since explicit parametric equations cannot be straightforwardly hypothesized due to the complexity of the system or due to the fact that the underlying mechanisms are simply not known. Hence, time series data is used to construct an empirical model which infers the patterns and captures the behavior of the system rather than that equations have to be hypothesized. To actually do this, the *R* package called **rEDM** was developed by the authors Ye et al. (2016), which contains various tools to model empirical dynamic models including the CCM method.

To illustrate how the CCM algorithm works in practice, consider two time series variables X and Y both with length L and are subsequently denoted as $\{X\} = (X(1), X(2), \ldots, X(L))$ and $\{Y\} = (Y(1), Y(2), \ldots, Y(L))$, respectively. The first step is to create a set of vectors with lags, which are referred to as a set of *delayed embedded* vectors. In other words, a set of delayed embedded vectors is constructed for the variable X and Y and are respectively denoted as $\underline{x}_t$ and $\underline{y}_t$ as shown in equation 23 and 24.

$$\underline{x}_t = \{\boldsymbol{X}_t, \boldsymbol{X}_{t-\tau}, \boldsymbol{X}_{t-2\tau}, \ldots, \boldsymbol{X}_{t-(E-1)\tau}\} \tag{23}$$

$$\underline{y}_t = \{\boldsymbol{Y}_t, \boldsymbol{Y}_{t-\tau}, \boldsymbol{Y}_{t-2\tau}, \ldots, \boldsymbol{Y}_{t-(E-1)\tau}\} \tag{24}$$

Each respective set of delayed embedded vectors constructs the (delay) embedding or in other words the so-called shadow manifold. This shadow manifold or equivalently the set of delayed embedded vectors is a proxy of the underlying system hence the term *shadow* and shows that based on just a single variable, such as X, the original system and thus its corresponding manifold can be reconstructed. Hence, the process of creating a shadow manifold is referred to as state space reconstruction and the fact that a single time series variable can be used as a proxy for the underlying system captures the essence of Takens (1981) Theorem. So, the set of delayed vectors $\underline{X}_t$ constructs the shadow manifold (delay embedding) for the variable X and is denoted as M_x. The shadow manifold M_x is constructed as follows: each vector in the set $\underline{X}_t$ represents a dimension and thus actually creates the state space *(also called phase space)* where simultaneously all the points of all the respective vectors are plotted in this state space and would reveal the trajectories of the variable X, which acts as proxy of the original system. Moreover, following exactly the same rationale, the shadow manifold for the variable Y can be constructed and is denoted as M_y, which is again equivalent to plotting the set of delayed embedded vectors $\underline{Y}_t$ in state space.

To clarify this idea further, consider figure 15 which shows the Lorenz system. In this specific example, the variable X is used to reconstruct the original manifold of the Lorenz system. More specifically, it is assumed for the sake of illustration that we have a three-dimensional system so that $E = 3$ and implies that the set of delayed embedded vectors can be denoted as follows $\underline{x}_t = \{\boldsymbol{X}_t, \boldsymbol{X}_{t-\tau}, \boldsymbol{X}_{t-2\tau}\}$. Formulated differently, we have a vector with the present values of X, a vector of X with a lag τ and a third vector of X with lag 2τ. Each delayed embedded vector in the set $\underline{X}_t$ corresponds to a dimension as shown in figure 15 and thereby determines the dimensionality of the state space. In this example, a particular point on the embedding M_x at time t is a three-dimensional point and is denoted as $\underline{X}_t = [X_t, X_{t-\tau}, X_{t-2\tau}]$. In essence, these three delayed embedded vectors capture the coordinates of all points that together construct the shadow manifold (embedding). More generally, a particular point on an E-dimensional embedding is an E-dimensional point, hence this state space is therefore sometimes also referred to as E-space (embedding space) (Sugihara and May, 1990).

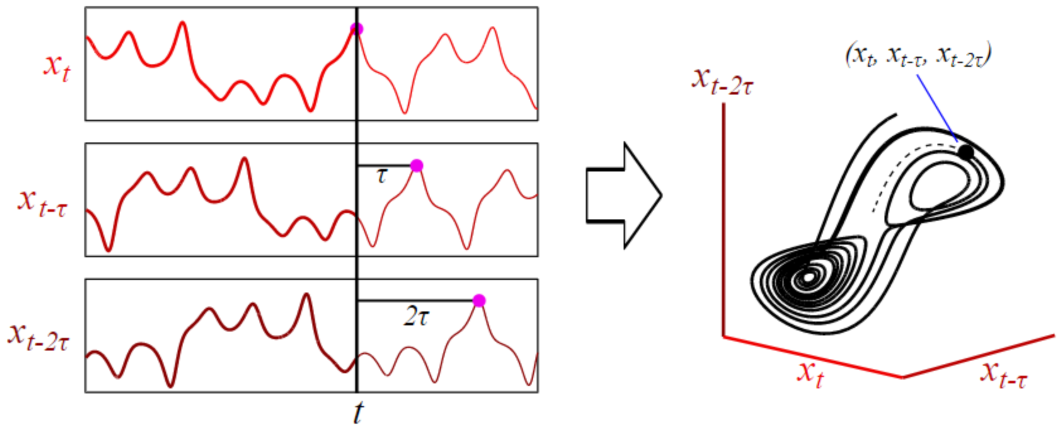


Figure 15: Here the Lorenz attractor is illustrated as an example and shows how a single time series variable can be used to retrieve the information of the system behavior following Takens (1981). Instead of using all three variables X, Y and Z, this example shows that a single variable X with its lags can be used to reconstruct the underlying system or in other words the original manifold. In other words, information about all the other variables in the system can be recovered by just looking at a single variable. Illustration is by Ye et al. (2016).

However, as can be observed from equation 23 and 24 there are two parameters, namely τ and E. As aforementioned, E refers to the embedding dimension or simply the dimension of the state space which is sometimes also called *E-Space*. An embedding with too few dimensions (lags) results into having too many singularities. This means that the shadow manifold will have points that are not well-behaved and have the implication that some points on the manifold actually relate to different system states. For example, it could be the case that for a given value of today (x_t) numerous different values might be predicted for tomorrow (X_{t+1}) and thus relates to different states and thereby creates ambiguity. On the other side, having too many dimensions will impact the prediction accuracy negatively since useful information is diluted by unuseful information. In layman terms, using the weather forecast as an example, using the weather of two weeks ago to forecast the weather of tomorrow will dilute useful information such as what the weather was yesterday. Therefore, the embedding dimension parameter E needs to be tuned, which is done by means of *simplex projection*. Simplex projection is a (non-linear) nearest neighbor forecasting method which was originally developed to distinguish chaotic time series from random noise (Sugihara and May, 1990). Based on simplex projection, points of interest on the embedding within state space can be forecast by means of taking a weighted average of its nearest neighbors. In other words, a particular point of interest can be forecast based on similar past events/points. The crucial element here is the dimension of the embedding, E, since this determines what events/points are considered as *nearest* neighbors. Moreover, important to note is that *nearest* is used in the context of state space and not in time and thus implies that a nearest neighbor is not necessarily close in time to the point of interest. More technically, the first step is to construct the embedding with E dimensions based on the variable of interest, such as y. A point of interest that is located on this embedding is now an E-dimensional point and the goal is to find its nearest neighbors in order to compute a prediction value for this point. Specifically, the neighborhood of points should be minimal which is defined as finding the smallest *simplex* that is formed from its $E+1$ nearest neighbors which contains the point of interest. The $E+1$ nearest neighbors are the vertices of the simplex and where this $E+1$ number is argued by Sugihara and May (1990) to result into the smallest simplex that can still have an E-dimensional interior point. Subsequently, to actually obtain a prediction the domain of the simplex is projected into its range which in essence means it is keeping track of where all the points in the simplex are positioned after p time steps. Then, it is computed where this point of interest has moved within the range of this simplex and exponential weights are assigned to the original distances of its nearest neighbors. Consequently, the exponential weighted average will yield a prediction value for the point of interest. So, for a particular point of interest y_t on the embedding of x, simplex projection will provide a prediction y_{t+1} given that the interest is in the prediction for one period later for this particular point. Formally, the prediction for a particular point of interest y_t one period later can be denoted as

$$\hat{Y}_{t+1} = \sum_{i=1}^{E+1} W_i X_t(j) \tag{25}$$

The parameter W_i denotes the weights and as aforementioned it utilizes an exponential weighting scheme. Specifically, it depends on the distance between $\underline{x}(t)$ and its i^{th} nearest neighbor on M_x as shown in equations 26 and 27

$$W_i = \frac{u_i}{\sum_{j=1}^{E+1} u_j} \tag{26}$$

where

$$u_i = exp\left(\frac{-d[x(t), x(t_i)]}{d[x(t), x(t_1)]}\right) \tag{27}$$

In other words, equation 27 shows that the distances are exponentially weighted, where $d[x(q), x(r)]$ denotes the euclidean distance between vectors q and r. In short, for a given E, simplex projection can be deployed to obtain estimates of the points located in the corresponding embedding space.

Consequently, for a given E, all the estimates can be compared to their corresponding observed values to evaluate how accurate the predictions actually are. Specifically, this is done by computing the correlation between the estimates and observed values which gives the so-called forecast skill (ρ). Hence, we can apply simplex projection for different embedding dimensions E and plot these embedding dimensions against their corresponding forecast skill. Then, the optimal embedding dimension corresponds to highest forecast skill (ρ).

Next to tuning the embedding dimension E, the time delay (lag) τ parameter can also be investigated, although generally the default value of 1 is followed. The effect of τ on the shadow manifold (delay embedding) depends on the nature and sampling frequency of the time series variable. For example, consider a time series variable U that measures every second how many students are sitting in the library of the Erasmus University. This would imply that many observations will be similar, since it is not expected that the number of students sitting in the library will change substantially over each second. Subsequently, if simultaneously a small time delay (lag) is considered of say $\tau = 1$ then this has the implication that the system collapses to a one dimensional diagonal line within state space. More specifically, if for example $E = 3$ then this means we have a three-dimensional state space where the three respective axis are U_t, U_{t-1} and U_{t-2}. The essence here is that the observations between these three lagged vectors are not very different because as argued before, it is not expected to observe substantial changes over each second since the value of 2 seconds ago will not be different from now. In other words, these three (lagged) vectors will have quite some overlap and thus the same coordinates in state space and implies that it will collapse to almost one dimension, a diagonal line. In such a setting, if τ is increased to for example 1 hour then the points in state space will be more spread out throughout the space. In contrast, if the time series variable would have been sampled very sparsely then a large τ is not desirable as the points will already be relatively spread out in the state space and by increasing τ will aggravate this, which is not desirable. The τ parameter is explicitly investigated in the *Extended Cross Convergent Mapping* method by Ye et al. (2015) and is briefly discussed in the discussion but for now this study follows the default value of $\tau = 1$, since this also seems justifiable based on the nature of the data which is further elaborated in the results section.

So, after discussing the key ideas of how shadow manifolds are constructed and how the parameters τ and E are estimated, it is now discussed how cross mapping and convergence are used in CCM to investigate causality. Each shadow manifold is a one-to-one map of the original manifold and subsequently implies that each shadow manifold is a one-to-one map of the other respective shadow manifolds. Therefore, the mapping between shadow manifolds can be tested to investigate whether the respective variables, in this case X and Y, are interacting in the same system because in dynamic systems theory two variables are causally related if the times series variables are coupled and belong to the same dynamic system. Empirically, this can be tested by evaluating the prediction accuracy of the mappings M_y to X and M_x to Y. Specifically, CCM tests how well the local neighborhood on one shadow manifold corresponds to the local neighborhood on another shadow manifold. So, the goal is to evaluate the prediction accuracy of $\hat{Y}(t) \mid M_x$ and $\hat{X}(t) \mid M_y$ where these notations respectively correspond to the cross-map estimates of Y and X.

Suppose the goal is to evaluate whether y causes x ($y \implies x$), then this implies that the goal is to obtain the cross map estimate of y denoted as $\hat{Y}(t) \mid M_x$. Then the first step is to construct the shadow manifold for variable x (M_x). Further suppose that there is a point of interest at time t denoted as y_t. The key idea here is that *if* the variable y causes x then this information must be encoded in the shadow manifold of x (M_x). In other words, the value of x at a particular moment in time in this state space should provide information about the variable y, since the state space of x captures all the factors that are driving x. This implies that a particular position within x's state space should say something about the variable y if y causally drives x. Consequently, this implies for the point of interest that each nearest neighbor point corresponds to a value of y that is close to y_t. Hence, a weighted average of these nearest neighbors would yield an estimate $\hat{y}_t$ that is close to the observed value y_t. However, if all the nearest neighbor points are located further away in state space from the point of interest, then this means that the estimate $\hat{y}_t$ will be far off from the observed value y_t, which is likely when the variable y barely influences the variable x.

$$\hat{Y}(t) \mid M_x = \sum_{i=1}^{E+1} W_i Y(t_i) \tag{28}$$

In equation 28 the parameter W_i denotes the weighting, which depends on the distance between $\underline{x}(t)$ and its i^{th} nearest neighbor on M_x. Moreover, $Y(t_i)$ denotes the simultaneous values of Y. The weights W_i are specifically defined as

$$W_i = \frac{u_i}{\sum_{j=1}^{E+1} u_j} \tag{29}$$

where

$$u_i = exp\left(\frac{-d[x(t), x(t_i)]}{d[x(t), x(t_1)]}\right) \tag{30}$$

In other words, equation 30 shows that the distances are exponentially weighted, where $d[x(q), x(r)]$ denotes the euclidean distance between vectors q and r. However, another way to explain exactly the same idea is that if X and Y would be causally related then the nearest neighbors located on M_x should identify corresponding nearest neighborhood points on M_y since they are one-to-one maps. Nonetheless, the above equations 28 to 30 display the setting of cross mapping X to Y but this is analogously defined for the cross mapping of Y to X. For the sake of completeness, the cross mapping of Y to X is shown in equation 31. Naturally, also the equations 32 and 30 changes correspondingly.

$$\hat{X}(t) \mid M_y = \sum_{i=1}^{E+1} W_i X(t_i) \tag{31}$$

However, important to note is that the explanation above demonstrates the key idea with just a single point that is of interest, while this procedure is actually performed for all the points that are present in the state space. More specifically, there are two criteria before it can be concluded that there is causality. The first one is that there should be significant correlation between the observed and predicted value. Secondly, there should be convergence. This concept of convergence is crucial and distinguishes between a simple correlation and causation. In the beginning it was stated that two time series were considered x and y of length L, but this length L plays an important role. More precisely, the complete analysis as above discussed is applied for different time series lengths L. This L is called the library size and is just a subsequence of the time series variable or in other words it represents a certain window size. Suppose $L = 3$ then this means that for example for the variable x only three observations from its time series are considered, which could be the first three observations, the last three or any other subsequence. Given $L = 3$, then the CCM analysis is performed for all three points, so for all the points in the shadow manifold, and would therefore give three estimated values. Simply stated, the result would be a column vector with three entries corresponding to the estimated values. Moreover, this column vector can be compared to the column vector containing the corresponding observed values. Subsequently, the correlation between these vectors is computed and would give a correlation (ρ) between the predicted and observed values for the given time series length L. Additionally, for each time series length L about 100 samples are drawn where each sample would give a certain correlation and thus an average correlation is computed for the given L. This is thereafter done for numerous different time series length L and per L the mean correlation between predicted and observed values is obtained.

However, the crucial concept here is convergence and relates to all these different time series lengths L. Specifically, when the time series lengths L are plotted against their corresponding mean correlations (ρ), which acts as an accuracy measure, convergence must be observed. This means that when the time series length L increases, then the forecast skill ρ should increase as well. Only if this can be observed in addition to a significant correlation, then it can be concluded that the variables are causally driving each other. The idea behind convergence can be explained by the fact that when the length of the time series L increases, then the manifolds become denser and the neighborhood becomes smaller, implying that the distances among the nearest neighborhood points also decreases. Hence, $\hat{Y}(t) \mid M_x$ converges towards $Y(t)$ and $\hat{X}(t) \mid M_y$ converges towards $X(t)$. Due to this convergence, it can be tested if the states of M_x correspond to the states on M_y.

4 Data

This chapter introduces the data to be used to demonstrate the Cross Convergent Mapping method in a marketing setting. A case study was already presented in chapter 2.3.3, but the related data is presented in this part. So, the used data is from a large department store where three products are considered, referred to as *product01*, *product02* and *product03*. Subsequently, for each product the following information is available: *sales*, *price*, *advertising efforts* and *distribution*. The time series variables *sales*, *price* and *advertising efforts* are shown for each respective product. Figure 16 shows the time series related to product one.

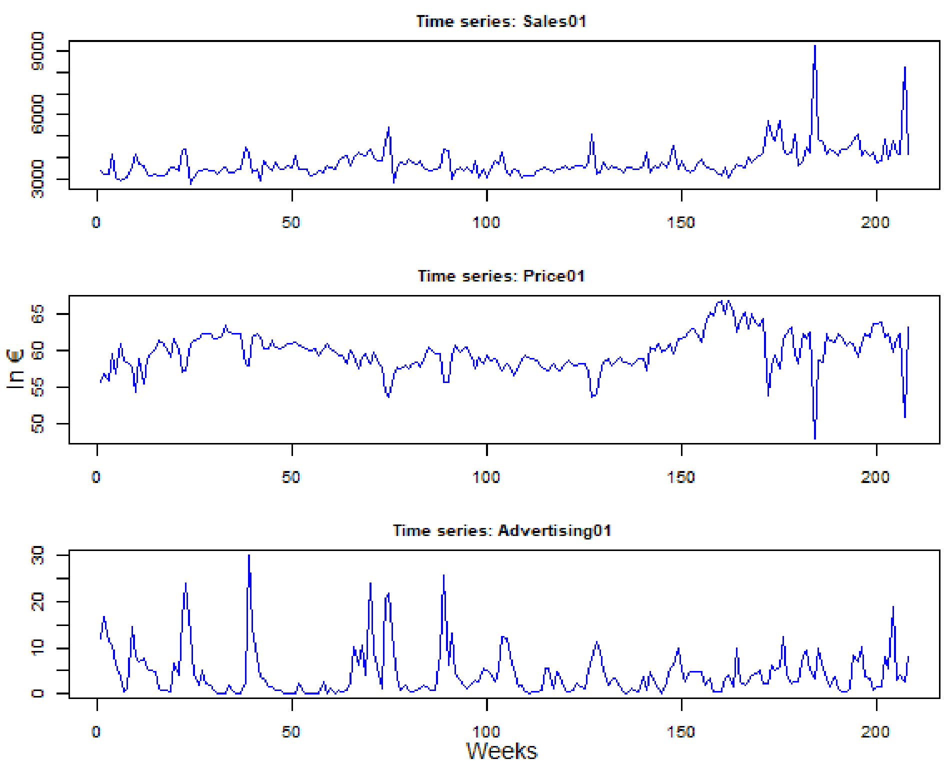


Figure 16: Here the sales, price and advertising budget are shown on a weekly basis for product one. Noticeable are the spikes around week 200 for both the sales and price which seems to be, a priori, a logic relation where a lower price and higher sales are correlated

Next to product one, figure 17 shows the relevant time series for product two. The course of the time series for product two is to a great extent similar to that of for product one. Hence, a priori, it is expected that these products are correlated and might even be causally related.

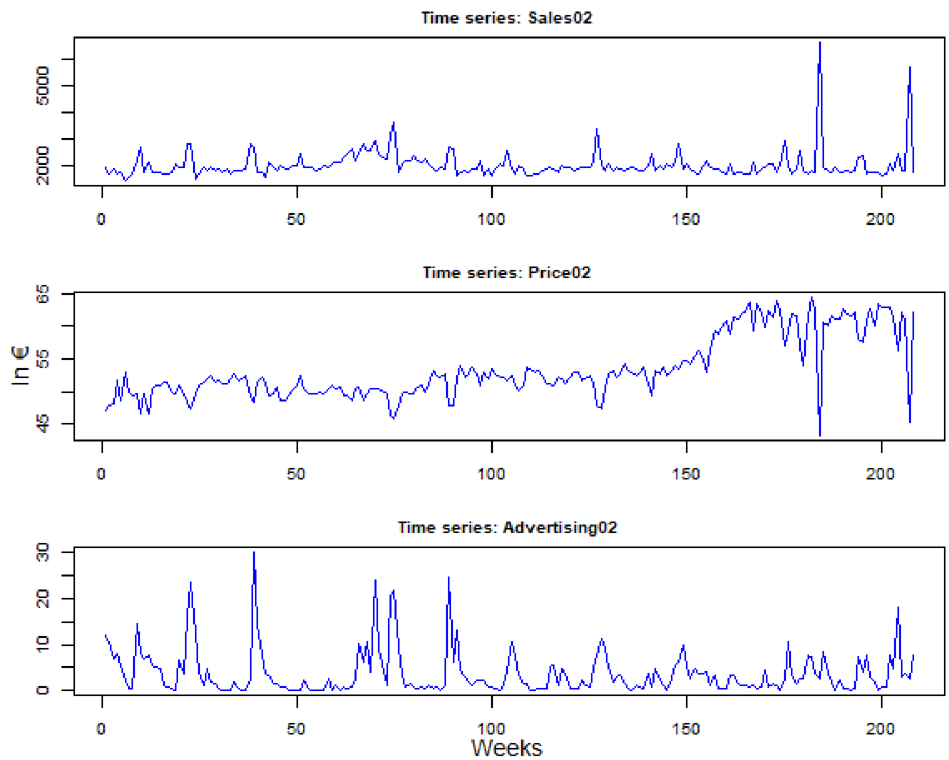


Figure 17: Here the sales, price and advertising budget are shown on a weekly basis for product two. Again, remarkable are the spikes around week 200 which, as for product one, makes sense that sales and price are correlated.

Furthermore, figure 18 shows the respective time series for product three. However, for product three there are no advertising efforts hence this time series variable is not shown for product three. Another noteworthy observation is that the price of product three seems to follow a *step function* compared to the prices of product one and two. The price of product three clearly decreases up until around week 100 and again increases from week 140 onward.

Moreover, before conducting any analyses the correlations among the time series have been inspected as shown in table 1. As expected, the correlations between sales and price are mainly negative although for product three they are solely positive which seems counter intuitive at first. Two other noteworthy observations are the high correlations between *sales01* and *sales02* with $\rho = 0.81$ and the variables *ads01* and *ads02* with $\rho = 0.96$. For these respective correlations it seems, a priori, that a causal relation would be plausible.

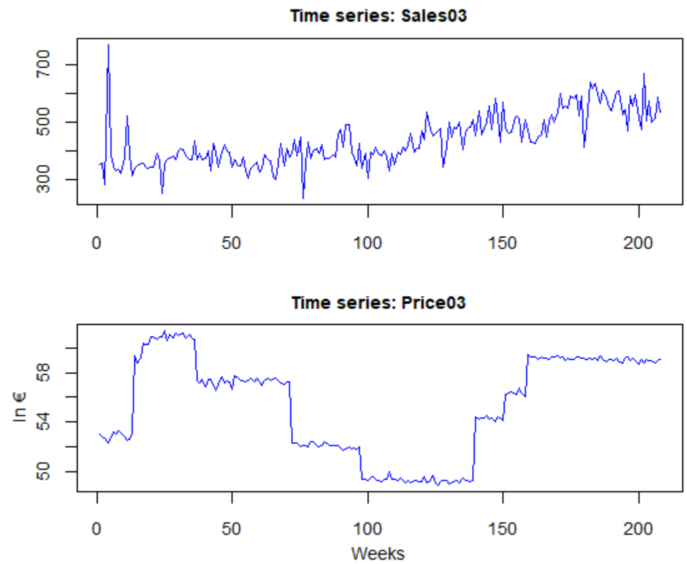


Figure 18: Here the sales, price and advertising budget are shown on a weekly basis for product three. Remarkable is the step function course of the price and also the sales are slowly increasing over time.

Table 1: Correlation Table

Variable	Sales01	Sales02	Sales03	Ads01	Ads02	Price01	Price02	Price03
Sales01	1							
Sales02	**0.81**	1						
Sales03	0.58	0.14	1					
Ads01	0.21	0.25	-.01	1				
Ads02	0.18	0.28	-0.05	**0.96**	1			
Price01	-0.38	-0.58	0.15	-0.29	-0.29	1		
Price02	0.13	-0.38	0.63	-0.15	-0.23	0.65	1	
Price03	0.26	0.05	0.17	-0.01	0	0.53	0.38	1

5 Results

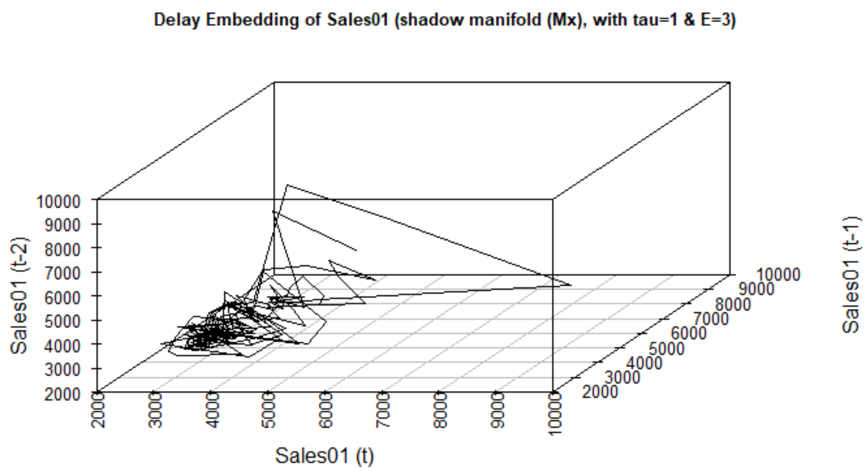
The idea is to map the causal relations among the given variables *sales*, *price* and *advertising efforts* by means of applying Cross Convergent Mapping. In the data there are three products, which means that there are 9 variables in total where for example for product one there are the variables *sales01*, *price01* and *Ads01* and the same holds for product two and three. For each general variable, *sales*, *price* and *advertising efforts*, its causal network is mapped first. Thereafter, all maps are merged into one figure to illustrate the complete causal network among all variables. However, before showing all the causal maps, one example is first step-by-step discussed in order to illustrate in detail how Cross Convergent Mapping works in practice.

5.1 Cross Convergent Mapping Results

So, the example that is step-by-step elaborated to show how CCM works with empirical data, is concerned with the question whether the sales of product one causally drive the sales of product two and vice versa. The first step is to create a set of vectors with lags, which are referred to as a set of *delayed embedded* vectors. For this example, such a set is created for the variable *Sales01* and denoted as $\underline{X}_t$ and the variable *Sales02* denoted as $\underline{Y}_t$ as shown in equation 32 respectively.

$$\underline{X}_t = \begin{bmatrix} \mathbf{Sales01}_t \\ \mathbf{Sales01}_{t-\tau} \\ \mathbf{Sales01}_{t-2\tau} \\ \vdots \\ \mathbf{Sales01}_{t-E_1\tau} \end{bmatrix} \qquad \underline{Y}_t = \begin{bmatrix} \mathbf{Sales02}_t \\ \mathbf{Sales02}_{t-\tau} \\ \mathbf{Sales02}_{t-2\tau} \\ \vdots \\ \mathbf{Sales02}_{t-E_2\tau} \end{bmatrix} \tag{32}$$

Each respective set of delayed embedded vectors constructs its delay embedding or in other words the shadow manifold. So, the set of delayed vectors $\underline{X}_t$ constructs the shadow manifold (delay embedding) for the variable *Sales01* and is denoted as M_x. Furthermore, the set $\underline{Y}_t$ constructs the shadow manifold for the variable *Sales02* and is denoted as M_y. Subsequently, both shadow manifolds are respectively plotted in figure 19 to provide a notion of intuition. However, important to note is that for the sake of illustration it has been assumed that $E = 3$ and $\tau = 1$, while the embedding dimension of three is not necessarily the optimal one. In this plotted example, the set of delayed embedded vectors would contain three vectors with up to lag two and are subsequently plotted to construct the shadow manifold in the state space.



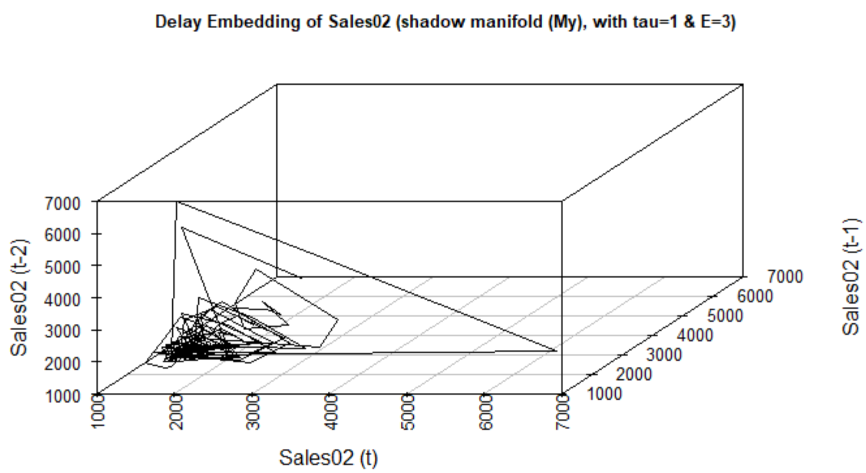
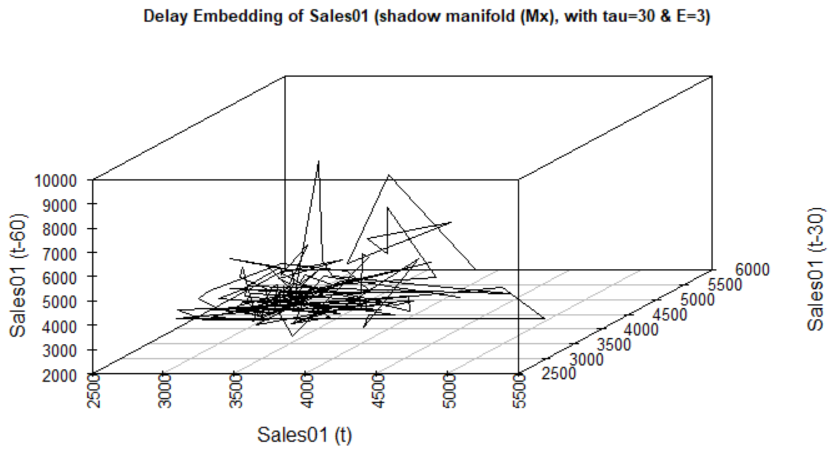


Figure 19: Two shadow manifolds are shown, the top figure corresponds to the shadow manifold for the variable *Sales01* and is denoted as M_x whereas the lower figure corresponds to the shadow manifold of the variable *Sales02* and is denoted as M_y. Note that $E = 3$ for the sake of illustration.

At first glance, the shadow manifolds in figure 19 seem not to be very clear, which can be partly attributed to the scaling of the figure itself. However, with real data it should not be expected to obtain a result as elegant as for example the famous Lorenz Attractor as was shown in figure 15. Nonetheless, figure 19 gives a more intuitive idea of the concept of constructing a shadow manifold in this case specifically for the variables *Sales01* and *Sales02*. However, to get a sense of how the time delay τ impacts the shadow manifold of the sales variables from figure 19, these plots have again been plotted but now with a time delay equalling $\tau = 30$ as shown in figure 20. As can be observed in figure 20, in both cases it seems that the points are more spread out through state space. This behavior can be explained by the fact that the sales (observations) are measured on a weekly basis and thus already contain quite some difference, implying it is already relatively sparsely sampled. Hence, increasing τ just makes the points spread out even more in state space and results into even *stranger* shapes.



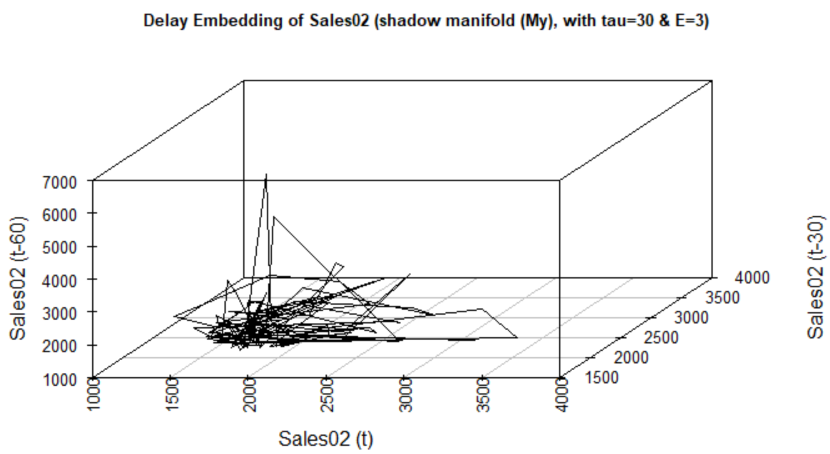


Figure 20: Two shadow manifolds are shown, the top figure corresponds to the shadow manifold for the variable *Sales01* and is denoted as M_x whereas the lower figure corresponds to the shadow manifold of the variable *Sales02* and is denoted as M_y. For both shadow manifolds a time delay of $\tau = 30$ is considered to demonstrate the effect of the time delay (τ). The top and bottom figure correspond respectively to the top and bottom of figure 20. Note that $E = 3$ for the sake of illustration.

For this study the parameter τ will be set to its default value 1 following Sugihara et al. (2012), but also due to the fact that the time series data in this study are not sampled very frequently, just once a week and would therefore mean that a lower τ is desirable. This implies that only E has to be tuned. The optimal embedding dimension is found by means of *simplex projection* as extensively earlier discussed. Optimal is defined as the embedding dimension that corresponds to the highest forecast skill ρ, which is the correlation between the predicted and observed values. As an example, figure 21 shows the plot for the variable *sales01* and shows that E=8 is considered as optimal since it corresponds to the highest forecast accuracy.

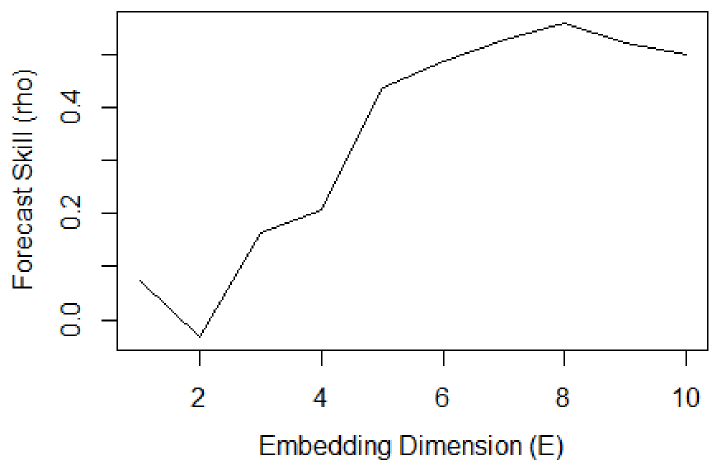


Figure 21: Here the embedding dimension E is plotted against the ρ which measures the forecast skill or in other words measures the correlation between observed and predicted values. Here it is shown for the *sales01* variable where E=8 is considered as optimal.

Specifically, figure 21 demonstrates the simplex projection results for different embedding dimensions E. In this case, when using an embedding dimension of 8 the most accurate predictions can be obtained. Moreover, the optimal embedding dimension has to be found per variable, because simplex projection is just based on a single variable. In other words, there are in total 8 variables (Ads03 is zero hence left out) each with a corresponding optimal embedding dimension which is illustrated in table 2.

Table 2: Optimal Embedding Dimension per Variable

Variable	Optimal Embedding Dimension E
Sales01	8
Sales02	9
Sales03	6
Ads01	5
Ads02	4
Price01	3
Price02	3
Price03	1

After tuning the parameters, the actual analysis can be conducted. The CCM analysis is conducted pair-wise and yields a plot showing the interactions between two time series variables. Two of such CCM resulting plots are shown in figure 22 and 23. Figure 22 shows the plot for the example that investigated whether the sales of product one causally drives the sales of product two.

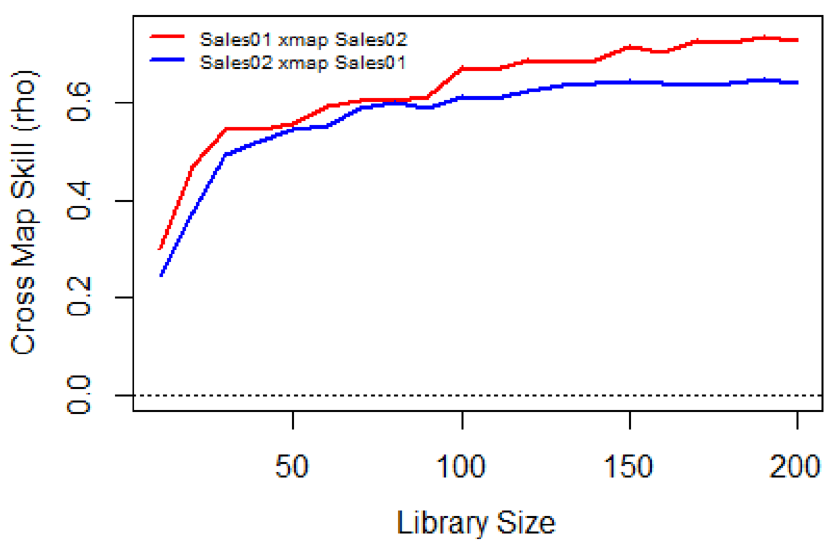


Figure 22: Here the prediction accuracy (ρ) is plotted against the time series length L for the variables *sales01* and *sales02*. This plot shows that there is both interaction and convergence as L increases.

Regarding figure 22, there is a *red* and *blue* line where the red line line corresponds to the cross map estimate of *Sales02* and the blue line to the cross map estimate of *Sales01*. More precisely, the term *"Sales01 xmap Sales02"* implies that is tested whether the variable *Sales02* causally drives the variable *Sales01*. In other words, a shadow manifold for the variable *Sales01* is constructed (M_x) and utilized to predict the values for *Sales02*. If the variable *Sales02* indeed causally drives the variable *Sales01* then the predicted values should be close to the observed values and would result into a sufficient high correlation ρ (cross map skill). However, a crucial element here is that also convergence should be observed since this convergence distinguishes a simple correlation from causality. This implies that as the library size L increases that also the cross map skill ρ should increase. Recall that the library size is another word for window size or taking a subsequence of the original time series data variable. So, in this case the original variables *Sales01* and *Sales02* have 200 observations hence the maximum library size is set to 200. However, suppose $L = 50$, then this means that a subsequence of 50 observations is sampled from the original time series variable *Sales01* or *Sales02* which can be for example the first 50 observations, the last 50 observations or any other subsequence. Furthermore, important to note is that 100 samples are drawn for each library size L and thus implies that the correlation ρ that corresponds to a given L is actually the average correlation for that L.

Analogously to the red line, the blue line corresponding to the statement *"Sales02 xmap Sales01"* implies that it is tested whether the variable *Sales01* causally drives the variable *Sales02*. In other words, a shadow manifold is constructed for *Sales02* which was denoted as M_y and is used to predict values for the variable *Sales01*. Subsequently, the correlation between the observed and predicted values is computed for a given L and denoted as the cross map skill ρ. However, important to note is that the blue line lies below the red line and implies that the causal relation $Sales02 \Longrightarrow Sales01$ is stronger than $Sales01 \Longrightarrow Sales02$. Nonetheless, there is a bidirectional causal relationship between *Sales01* and *Sales02* ($Sales01 \Longleftrightarrow Sales02$).

In contrast to the example shown in figure 22, figure 23 shows an example where there is no causality between two variables. Specifically, there is no causal relationship between the variables *Sales02* and *Sales03* since the the cross map skill (ρ) is zero and logically there is also no convergence.

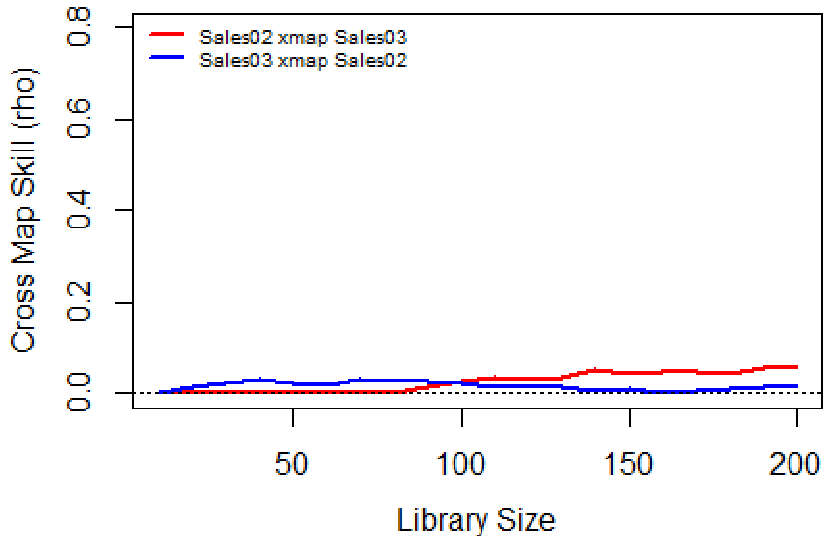


Figure 23: Here the prediction accuracy (ρ) is plotted against the time series length L for the variables *sales02* and *sales03*. This plot shows that there is neither interaction nor convergence as L increases.

To conclude, it has been shown that the causal relation between the variables *Sales01* and *Sales02* is bidirectional where the effect of *Sales02* on *Sales01* is stronger than vice versa. Moreover, it has also been demonstrated that there is no causal relation between the variables *Sales03* and *Sales02*. However, to provide a clear summary, figure 24 maps all the causal relations among the *sales* variables where additionally also the relation between *Sales01* and *Sales03* has been investigated and is mapped in this figure. Specifically, the arrows indicate how the causal relationship flows and correspond to a particular number. This number represents the cross map skill (ρ) at the full library size ($L = 200$) and thus shows the strength of the causal relation. Technically, this number denotes the correlation between the observed and predicted values and logically a higher number is an indication of a stronger causal relation since the predicted values are closer to the observed value. However, recall that CCM follows a two-step causality condition, where the cross map skill (ρ) should be significant and that convergence should be observed as the library size increases. The significance of the found cross map skills for each relation are extensively discussed in the next section, while all the CCM resulting plots that are used to determine if convergence is present or not are presented in Appendix A. Next to the causal map for the *sales* variables as shown in figure 24, also two causal maps have been respectively created for the variables *advertising efforts* and *prices*. The causal map for the variable *advertising efforts* is shown in figure 25 and provides an overview on how the advertising efforts causally affect the different sales streams. Furthermore, figure 26 shows how the *pricing* affects the sales of the three different products. Subsequently, all three causal maps are integrated into one plot and is illustrated in figure 27.

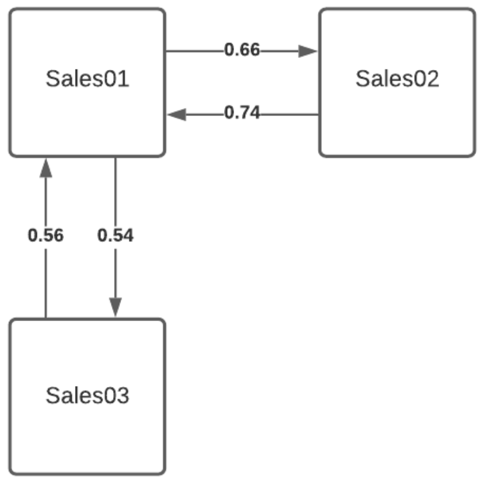


Figure 24: Mapping the causal relationships between the three sales streams, sales01, sales02 and sales03. The numbers denote the cross map skill (ρ) at its full library size ($L = 200$) and is an indication for the strength of the causal relation. Here sales02 and sales03 do not causally force each other. However, between the other variables there is bidirectional causality.

Next to the sales variables, the causal interactions between sales and advertising has been investigated, as aforementioned, and is shown in figure 25. Most noteworthy and as expected, a priori, is that advertising causally drives sales but it does not straightforwardly apply to the other way around, so that sales also drive advertising. Specifically, in the case of product one there is bidirectional causality between sales01 and ads01, although the effect of sales on advertising is weaker than the effect advertising has on sales. Another remarkable observation is that the advertising efforts for product one (*Ads01*) also impacts the sales of product two and subsequently the advertising of product two (*Ads02*) causally influences the sales of product one. More precisely, *Ads01* has a stronger effect on *Sales02* than on *Sales01* which is remarkable. Furthermore, the effect of *Ads02* on *Sales01* is equally strong as *Ads01* on *Sales01*. In other words, there seems to be some sort of spillover effect taking place in the advertising efforts with regards to the different products and could possibly be due to the nature of the products. Another interesting phenomenon that can be observed is that there is a very strong causal relation between *Ads01* and *Ads02* and implies that the different advertising efforts are strongly interacting with each other. This strong interaction could potentially be an explanation for the observed causal relations between *Ads01* and *Sales02* and the relation *Ads02* and *Sales01*. Specifically, the strong interaction between *Ads01* and *Ads02* could be a common driver for the other relations. However, this is further investigated in the next section. Additionally, as noted before, for product three the advertising efforts are zero and therefore left out. Nonetheless, *Sales03* still has some weak impact on the advertising efforts of product one (*Ads01*).

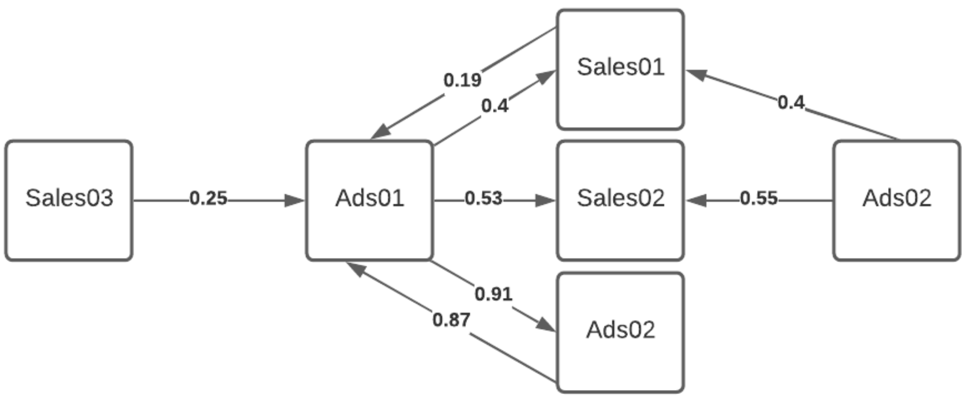


Figure 25: Mapping the causal relationships between the sales and advertising efforts variables. The numbers denote the cross map skill (ρ) at its full library size ($L = 200$) and is an indication for the strength of the causal relation. However, note that there are no advertising efforts for product three hence it is not included.

Besides the sales and advertising efforts, it is further investigated how the sales interact with prices as shown in figure 26. A remarkable observation is that there is bidirectional causality between sales and pricing, but particularly remarkable is that sales have a stronger causal effect on the prices than vice versa. This seems counter-intuitive at first, but as shown in table 1 the correlation between sales and pricing is negative for product one and two and thus implies that higher amount of sales also results into lower prices. One possible (economic) explanation could be due to economies of scale. Moreover, the sales of product two affects the prices of product one and the sales of product one subsequently also impact the prices of product two. This could potentially be explained by the fact that the sales of product one and two are causally related as shown in figure 24 and therewith influence each others prices. The sales of product one and two are highly positive correlated ($\rho = 0.81$, table 1) and thus seems to be complementary in nature and could be an explanation for the found causal relation between the prices and sales of the two different products. Furthermore, no causal relations were found between the price of product three and the sales of the other two products. As shown in 1 the correlation among the sales and prices of products one and two are negative, which is as expected.

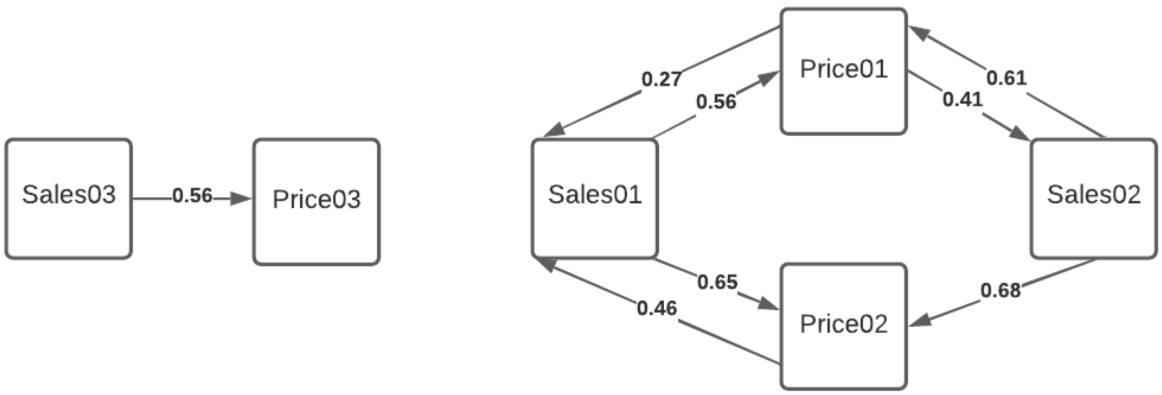


Figure 26: Mapping the causal relationships between the sales and Pricing variables. The numbers denote the cross map skill (ρ) at its full library size ($L = 200$) and is an indication for the strength of the causal relation. Noteworthy is that for all products the sales have a stronger causal impact on advertising than the other way around.

So, figure 24 to 26 show different mappings of the causal relations among the different variables. However, to provide a more concise overview of all the causal interrelationships, figure 27 contains all the relations from figure 24 to 26 in one map and to some extent simplified. Specifically, an open arrow indicates a weaker relation compared to a filled arrow. The most noteworthy observation is that especially products one and two are interacting with each other in terms of sales, prices and advertising efforts and that product three is actually less interacting with these products. Another interesting observation is that the sales have a stronger impact on the price than the other way around.

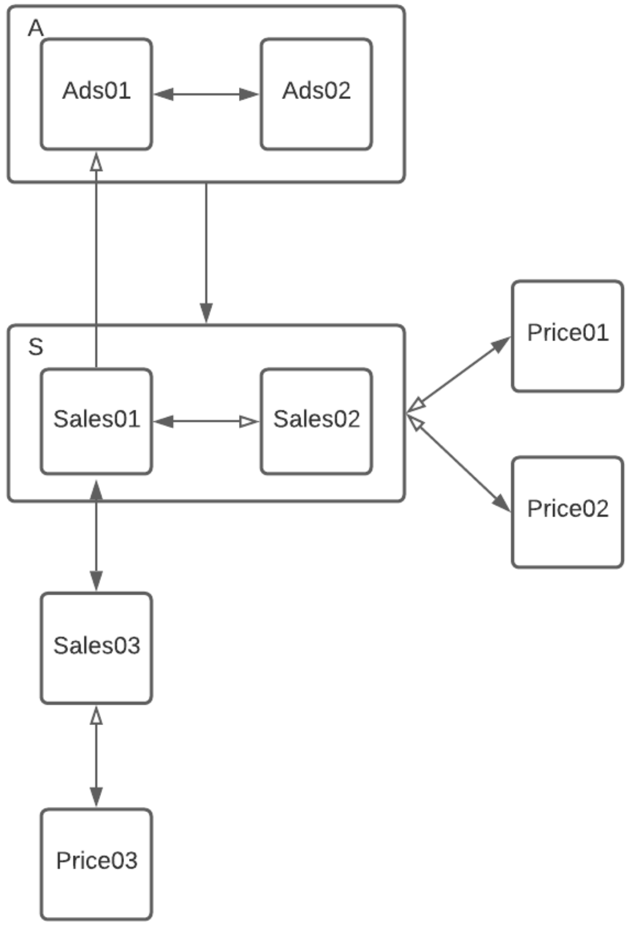


Figure 27: Mapping all the causal relationships between the variables sales, prices and advertising efforts for three products. No arrow implies no causal relationship while a filled arrow shows a strong relationship and an open arrow refers to a weaker relation compared to a filled arrow.

Based on figure 27, it can be observed from the top that the advertising efforts for product one and two causally force each other. Subsequently, these advertising efforts drive the sales where for example the advertising efforts for product one impact the sales of product one but also the sales of product two. This observation also applies for the advertising efforts of product two, which impacts the sales of both product one and product two. Additionally, the sales of product one weakly drive the advertising efforts of product one, while this relationship does not hold for the other products. Furthermore, it can be observed that the sales causally drive the prices for all three products, while the reverse effect is weaker as indicated by the open arrows. Moreover, the sales of product three causally impact the sales of product one.

5.2 Robustness Test for Cross Convergent Mapping

In the afore discussed figures, 24 to 27, all the causal interactions among the variables *sales*, *advertising efforts* and *pricing* are showed and specified. However, although 100 samples are drawn for each library size and thus the cross map skill constitutes of an average rather than it is based on a single correlation, an additional robustness test can be desirable as it provides more confidence in the found causal relationships. Specifically, a question that may arise is to what extent these cross map estimates are observed by chance, or formulated differently are these estimates statistically significant? To provide evidence for statistical significance, the actual cross map skill estimates at its full library size are compared to corresponding cross map estimates that are based on a null distribution which are referred to as null models. In this way it can be evaluated whether the observed cross map skill ρ is statistically different from its corresponding cross map skill that is essentially based on random data. So, after obtaining all the CCM results which were extensively discussed in previous section (5.1), the first key step is to generate *random* data. More specifically, surrogate time series are generated to preserve all the key statistical properties and seasonality while the data are sufficiently randomized to be able to create a null distribution. Subsequently, the actual cross map skill estimates (ρ) are compared to their corresponding cross map skill estimates that are based on the surrogate time series data and provides an indication whether the observed cross map estimate is statistically different from the null model.

To illustrate this idea in more detail, consider again the step-by-step example from previous section where it is evaluated whether *Sales01* causally drives *Sales02* and vice versa. The first step is to obtain the actual cross map estimates at the maximum library size. For this example, it was found that the cross map skill estimate ρ for the variable *Sales02* when $L = 200$ is $\rho = 0.74$ and for the variable *Sales01* it is $\rho = 0.64$. Thereafter, surrogate time series data needs to be generated where the **rEDM** R package by Ye et al. (2016) offers a wrapper function for this task which specifically also preserves seasonality. Subsequently, for each time series variable there are 1000 surrogate time series variables generated and thus for this example there will be 1000 surrogate time series for both *Sales01* and *Sales02*. Then, the CCM analysis is performed between *Sales01* and the 1000 surrogate time series for *Sales02* and also vice versa, so that CCM is also performed between *Sales02* and the 1000 surrogate time series for *Sales01*. This implies that there will be two surrogate cross map skill estimate vectors with 1000 rows each that respectively correspond to the variables *Sales01* and *Sales02*. Subsequently, the initial *observed* cross map skill is compared to these 1000 surrogate cross map skill estimates to determine whether they are statistically significant where $\alpha = 0.05$ is used as threshold. Tables 3 to 5 summarise the findings respectively for the variables *Sales*, *Advertising efforts* and *Pricing*.

Table 3: Significance of cross map estimates at full library size among *sales* variables

Causal Relation	Optimal E	Cross Map Skill ρ ($L = 200$)	P-value	Convergence[1]
Sales01 $\Longrightarrow$ Sales02	9	0.66	0.001^{***}	Yes
Sales02 $\Longrightarrow$ Sales01	8	0.74	0.001^{***}	Yes
Sales01 $\Longrightarrow$ Sales03	6	0.54	0.001^{***}	Yes
Sales03 $\Longrightarrow$ Sales01	8	0.56	0.001^{***}	Yes
Sales02 $\Longrightarrow$ Sales03	6	0	0.24	No
Sales03 $\Longrightarrow$ Sales02	9	0	0.07^{*}	No

Note: [1] See Appendix (A), ***p$<$ 0.01,**p$<$ 0.05,*p$<$ 0.10

As can be observed in table 3, clearly the causal relation $Sales02 \Longrightarrow Sales03$ is not significant (p=0.24) and $Sales03 \Longrightarrow Sales02$ is slightly above the 0.05 threshold (p=0.07). However, the cross maps skills (ρ) are both zero and logically also implies that no convergence can be observed. This evidently indicates that no causal relation is present between the variables *Sales02* and *Sales03*. Hence, these relations were also not mapped in figure 24.

Next to the *Sales* variables, the significance of the causal relations between the *Advertising* and *Sales* variables has been summarised in table 4. All causal relations are significant at the 0.05 level except for two, which are $Ads01 \Longrightarrow Sales03$ and $Ads02 \Longrightarrow Sales03$. In other words, the effects of the advertising efforts for product one and product two on the sales of product three do not seem to differ significantly from random data.

Furthermore, important to note is that the relation $Sales02 \Longrightarrow Ads02$ is significant ($p = 0.013$) but that no convergence could be observed and thus fails to meet the two-step causality condition and subsequently implies that it cannot be concluded that there is a causal relation. This exact same idea also holds for the relations $Sales01 \Longrightarrow Ads02$, $Sales02 \Longrightarrow Ads01$ and $Sales03 \Longrightarrow Ads02$. Next to these, also the two clearly non-significant relations $Ads01 \Longrightarrow Sales03$ and $Ads02 \Longrightarrow Sales03$ ($p = 1$) have logically no convergence since the corresponding cross map skills are zero. Hence, these aforementioned relations that fail to meet the two conditions for causality are not mapped into figure 25.

Noteworthy are also the very strong causal relations between the advertising efforts of product one (*Ads01*) and these of product two (*Ads02*). At first, this relation could be an explanation for the fact that it was observed that *Ads01* causally influences *Sales02* and that *Ads02* causally influences *Sales01*, since the advertising efforts of both products strongly interact with each other. However, the causal relations between *Ads01* and *Sales02* and also between *Ads02* and *Sales01* are both highly significant and thus are not driven by a common driver. In other words, the advertising efforts are in itself causally interacting with each other, but also the advertising efforts with different products and the one is not an explanation for the other. This implies that the strong relation between $Ads01 \iff Ads02$ is not a common driver and thus is not an explanation for the observed relation between respectively *Ads01* and *Sales02* and also between *Ads02* and *Sales01*.

Table 4: Significance of cross map estimates at full library size among *Sales* and *Advertising* variables

Causal Relation	Optimal E	Cross Map Skill ρ ($L = 200$)	P-value	Convergence[1]
Ads01 $\Longrightarrow$ Sales01	8	0.40	0.001***	Yes
Sales01 $\Longrightarrow$ Ads01	5	0.19	0.004***	Yes
Ads02 $\Longrightarrow$ Sales02	9	0.55	0.001***	Yes
Sales02 $\Longrightarrow$ Ads02	4	0.07	0.013**	No
Ads02 $\Longrightarrow$ Sales01	8	0.4	0.001***	Yes
Sales01 $\Longrightarrow$ Ads02	4	0.02	0.03**	No
Ads01 $\Longrightarrow$ Sales02	9	0.53	0.001***	Yes
Sales02 $\Longrightarrow$ Ads01	5	0.15	0.027**	No
Ads01 $\Longrightarrow$ Sales03	8	0.00	1	No
Sales03 $\Longrightarrow$ Ads01	5	0.25	0.001***	Yes
Ads02 $\Longrightarrow$ Sales03	8	0.00	1	No
Sales03 $\Longrightarrow$ Ads02	4	0.06	0.002***	No
Ads01 $\Longrightarrow$ Ads02	4	0.91	0.001***	Yes
Ads02 $\Longrightarrow$ Ads01	8	0.87	0.001***	Yes

Note: [1] See Appendix (A), ***p< 0.01,**p< 0.05,*p< 0.10

Besides the *Sales* and *Advertising* variables, table 5 summarises the significance of cross map estimates among the *Sales* and *Pricing* variables. As can be seen, all cross map relations are significant at the 0.01 level except for two, which is the relation between the price of product three and the sales of product two. However, not all significant cross map estimates show convergence which is a necessary condition for causality besides the significant relationship. The lack of convergence is observed for: $Price02 \implies Sales02$, $Price03 \implies Sales03$, $Price03 \implies Sales01$ and $Sales01 \implies Price03$. Additionally, also for the clearly non-significant cross map estimates $Price03 \implies Sales02$ and $Sales02 \implies Price03$ no convergence is observed as the cross map estimates are zero.

Table 5: Significance of cross map estimates at full library size among *Pricing* and *Sales* variables

Causal Relation	Optimal E	Cross Map Skill ρ ($L = 200$)	P-value	Convergence[1]
Price01 $\implies$ Sales01	8	0.27	0.001***	yes
Sales01 $\implies$ Price01	3	0.56	0.001***	Yes
Price02 $\implies$ Sales02	9	0.17	0.002***	No
Sales02 $\implies$ Price02	3	0.68	0.001***	Yes
Price03 $\implies$ Sales03	8	0.39	0.001***	No
Sales03 $\implies$ price03	1	0.56	0.001***	Yes
Price02 $\implies$ Sales01	8	0.46	0.001***	Yes
Sales01 $\implies$ Price02	3	0.65	0.001***	Yes
Price03 $\implies$ Sales01	8	0.21	0.001***	No
Sales01 $\implies$ price03	1	0.20	0.001***	No
Price01 $\implies$ Sales02	9	0.41	0.001***	Yes
Sales02 $\implies$ Price01	3	0.61	0.001***	Yes
Price03 $\implies$ Sales02	9	0.00	0.50	No
Sales02 $\implies$ Price03	1	0.00	0.27	No

Note: [1] See Appendix (A), ***p< 0.01,**p< 0.05,*p< 0.10

6 Discussion

Regarding the specific case study that was discussed, it was shown that specifically the sales of product one and product two are causally interacting with each other. However, to a slightly lesser extent, also the sales of product one and of product three interact with each. In other words, any factor that would change the sales of product one would also affect the sales of product three and product two. This is especially relevant when management has to make a strategic decision that affects the sales of product one, since this impact cannot be limited to solely the sales of product one but is also spilt over to product two and three. Moreover, the correlation between *Sales01* and *Sales02* and *Sales01* and *Sales03* is both positive with respectively 0.81 and 0.58. This means that if a decision impacts the sales of product one negatively, then this will also impact the sales of product two and three negatively and thus could have to some extent a snowball effect.

Furthermore, the advertising efforts of product one and two are interacting with both the sales of product one and two and shows again that especially product one and two are to some degree dependent on each other. The advertising efforts of product one and product two showed a very high correlation (0.96) and also displayed a very strong causal relation implying that they strongly influence each other. So, a change in one of the advertising efforts budgets will result into a chain reaction. More precisely, a change in the budget of the advertising efforts of product two will affect the advertising efforts of product one, this subsequently affects the sales of product one but *Ads01* also influences the sales of product two and thus will also be impacted. Additionally, *Ads02* also has a direct causal relation with *Sales01* and *Sales02* and thus these sales streams will again be impacted. In other words, a change in the advertising efforts or sales seems to trigger a chain reaction which can amplify the resulting effects and shows that a change in one variable cannot be limited to that variable but results into a chain of reactions. Another key result is that also the prices and sales of product one and two interact strongly with each other, but it is specifically remarkable that for all three products the sales have a stronger effect on the prices than vice versa. The correlations between prices and sales are negative and this could indicate that due to economies of scale the prices can be lowered when sales increase.

After all, this case study demonstrated how variables are causally interacting with each other and thereby creating a more complex system than initially thought of. Specifically, it shows that only considering for example the relation $Ads01 \implies Sales01$ is not sufficient when a strategic decision needs to be made, since a change in one variable actually leads to a chain of reactions and the impact cannot be limited to just one variable. Hence, mapping all the causal relations reveals the underlying system that is present and allows to be better acted upon. However, important to note for this case study is that the degree of non-linearity in the data is limited while CCM can especially be exploited when non-linearity is present. Hence, this study just showed the possibilities of a complexity theory approach but it is encouraged to further research this with marketing data where strong non-linearity is present, where there are sufficient examples in the marketing literature that show the presence of strong non-linearity.

Next to this, another potential limitation is regarding the CCM method. More specifically, a problem of the CCM method is referred to as *generalized synchrony*. This specifically means that CCM cannot distinguish well between a very strong unidirectional relation and a bidirectional relation. In other words, if there is an exceptional strong unidirectional relation between two variables then it might seem like if there is a bidirectional causal relation while this is not true. Hence, the method Extended Cross Convergent Mapping (ECCM) solves this by means of explicitly considering different time lags (τ). By plotting the CCM cross map skills against the time lags, it can be observed whether a driving variable impacts a response variable with a delay and subsequently is used to determine whether *generalized synchrony* is observed or a true bidirectional causal relation (Ye et al., 2015). So, for further research it is interesting and advised to investigate this further to ensure that *generalized synchrony* is not observed. Furthermore, next to CCM and ECCM, it is recommended and interesting to look into the Reservoir Computing Causality Framework (RCC) method by Huang et al. (2020), which argues to be able to detect causality between time series by means of a machine learning framework while being more efficient than CCM and ECCM. Additionally, RCC does not need to estimate the different parameters, the embedding dimension (E) and time delay (τ). Also, it is argued that RCC is much more efficient and more robust against noisy-data and works well with high-dimensional data.

7 Conclusion

This study first argued that obtaining causal estimations for all kind of different marketing-rooted questions is of great interest but that classic approaches such as conducting experiments and subsequently basing on the potential outcomes framework is a limited approach. Besides the fact that it is time-consuming and expensive, such approaches cannot capture all the dynamic effects that might be present, which are for example notable in pricing and advertising. Hence, it was argued that utilizing time series data can offer a solution to this limitation, but that at the same time methods used for causal estimations for time series data also have their limitations. More specifically, the most commonly applied methods like Granger Causality are linear methods and do not work in settings that are characterized by more complex phenomena. Such complex phenomena could be feedback effects, non-linearity or non-separability among others. Hence, instead of *linearizing* complex settings this study argues that methods from complexity theory can be leveraged to more accurately model such complex settings and subsequently yield more accurate causal estimations. Additionally, complexity theory methods are designed to capture complex systems and thus allows to map the complete network instead of following a reductionist approach, as is traditionally followed within marketing. Therefore, this research aimed to answer the question of *how complexity theory based methods can be leveraged to enhance causal estimations from time series data for marketing mix related questions*. In particular, this study extensively discussed the Cross Convergent Mapping (CCM) method, which originates from complexity theory and is a (non-linear) state space reconstruction method that utilizes the properties of delay embedding to estimate causal relations between time series. Furthermore, to demonstrate how CCM can be utilized for marketing time series data, a case study was presented. More precisely, a data set with three products and for each product there was a variable regarding the sales, advertising efforts and pricing. This study showed that by means of CCM, the complete causal network can be mapped even when complex phenomena like non-linearity are present. Moreover, the strength of the causal relations can be observed and the significance of these relations can also be evaluated. Subsequently, by being able to map all causal relations even when the marketing setting seems complex helps management in making more informed strategic decisions, since the consequences of decisions can be better understood. In addition, the case study presented in this paper specifically showed that even in a relatively straightforward setting with just three products that it can actually be more complex than initially thought of. Specifically, it was shown that a change in particular variables like advertising results into a chain of reactions. Traditionally, a marketeer would solely consider the isolated effects and neglect the underlying system which might result into making different decisions that are not optimal.

In short, this study showed how the toolbox of marketeers can be expanded by looking beyond traditional approaches and to borrow methods from complexity theory. This allows marketeers to model complex settings without the need to linearize these settings and subsequently result into more realistic estimations, which helps in making better informed strategic decisions.

References

Ajorlou, A., Jadbabaie, A., & Kakhbod, A. (2018). Dynamic pricing in social networks: The word-of-mouth effect. *Management Science*, *64*(2), 971–979.

Arthur, W. B. (2013). Complexity economics. *Complexity and the Economy*.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & Zinman, J. (2010). What's advertising content worth? evidence from a consumer credit marketing field experiment. *The Quarterly Journal of Economics*, *125*(1), 263–306.

Bruce, N. I., Murthi, B., & Rao, R. C. (2017). A dynamic model for digital advertising: The effects of creative format, message content, and targeting on engagement. *Journal of Marketing Research*, *54*(2), 202–218.

Čenys, A., Lasiene, G., & Pyragas, K. (1991). Estimation of interrelation between chaotic observables. *Physica D: Nonlinear Phenomena*, *52*(2-3), 332–337.

Deyle, E. R., & Sugihara, G. (2011). Generalized theorems for nonlinear state space reconstruction. *Plos One*, *6*(3), e18295.

Doherty, N., & Delener, N. (2001). Chaos theory: Marketing & management implications. *Journal of Marketing Theory and Practice*, *9*(4), 66–75.

Eichler, M. (2012). *Causal inference in time series analysis*. Wiley Online Library.

Feigenbaum, M. J. (1978). Quantitative universality for a class of nonlinear transformations. *Journal of Statistical Physics*, *19*(1), 25–52.

Fieguth, P. (2017). An introduction to complex systems. *Complex Systems and Archaeology*, *10*, 978–83.

Giglio, M., Musazzi, S., & Perini, U. (1981). Transition to chaotic behavior via a reproducible sequence of period-doubling bifurcations. *Physical Review Letters*, *47*(4), 243.

Goldenberg, J., Libai, B., & Muller, E. (2001a). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, *12*(3), 211–223.

Goldenberg, J., Libai, B., & Muller, E. (2001b). Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, *9*(3), 1–18.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424–438.

Gregersen, H., & Sailer, L. (1993). Chaos theory and its implications for social science research. *Human Relations*, *46*(7), 777–802.

Heil, O. P., & Helsen, K. (2001). Toward an understanding of price wars: Their nature and how they erupt. *International Journal of Research in Marketing*, *18*(1-2), 83–98.

Hibbert, B., & Wilkinson, I. F. (1994). Chaos theory and the dynamics of marketing systems. *Journal of the Academy of Marketing Science*, *22*(3), 218–233.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960.

Huang, Y., Fu, Z., & Franzke, C. L. (2020). Detecting causality from time series in a machine learning framework. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *30*(6), 063116.

Jaeger, A., & Parkinson, M. (1994). Some evidence on hysteresis in unemployment rates. *European Economic Review*, *38*(2), 329–342.

Köhler, C., Mantrala, M. K., Albers, S., & Kanuri, V. K. (2017). A meta-analysis of marketing communication carryover effects. *Journal of Marketing Research*, *54*(6), 990–1008.

Lee, J., Shin, B., & Chung, I. (1996). Causality between advertising and sales: New evidence from cointegration. *Applied Economics Letters*, *3*(5), 299–301.

Lobel, I. (2020). Dynamic pricing with heterogeneous patience levels. *Operations Research*, *68*(4), 1038–1046.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, *20*(2), 130–141.

Lorenz, E. N. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of Atmospheric Sciences*, *26*(4), 636–646.

Maldonado, G., & Greenland, S. (2002). Estimating causal effects. *International Journal of Epidemiology*, *31*(2), 422–429.

Mela, C. F., Gupta, S., & Lehmann, D. R. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research*, *34*(2), 248–261.

Neyman, J., Dabrowska, D. M., & Speed, T. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 465–472.

Panfilov, A. V. (2018). Qualitative analysis of differential equations. *ArXiv Preprint ArXiv:1803.05291*.

Pochun, T., Brennan, L., & Parker, L. (2018). Advertising effects? an elemental experiment. *Australasian Marketing Journal (AMJ)*, *26*(4), 338–349.

Rao, V. R. (2009). *Handbook of pricing research in marketing*. Edward Elgar Publishing.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688.

Schlosser, R., & Boissier, M. (2018). Dynamic pricing under competition on online marketplaces: A data-driven approach. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 705–714.

Sharma, S., & Kapur, D. (2014). Causality relationship between advertising expenditure and sales: A study of indian service sector. *Management and Labour Studies*, *39*(3), 275–292.

Simester, D., Hu, Y., Brynjolfsson, E., & Anderson, E. T. (2009). Dynamics of retail advertising: Evidence from a field experiment. *Economic Inquiry*, *47*(3), 482–499.

Stock, J. H., Watson, M. W. et al. (2012). *Introduction to econometrics* (Vol. 3). Pearson New York.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, *338*(6106), 496–500.

Sugihara, G., & May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, *344*(6268), 734–741.

Takens, F. (1981). Detecting strange attractors in turbulence. *Dynamical systems and turbulence, warwick 1980* (pp. 366–381). Springer.

Tranquillo, J. V. (2019). *An introduction to complex systems*. Springer.

Tull, D. S. (1965). The carry-over effect of advertising. *Journal of Marketing*, *29*(2), 46–53.

Varnali, K. (2019). Understanding customer journey from the lenses of complexity theory. *The Service Industries Journal*, *39*(11-12), 820–835.

Wilkinson, I. F., & Young, L. C. (2013). The past and the future of business marketing theory. *Industrial Marketing Management*, *42*(3), 394–404.

Yang, A. C., Peng, C.-K., & Huang, N. E. (2018). Causal decomposition in the mutual causation system. *Nature Communications*, *9*(1), 1–10.

Ye, H., Clark, A., Deyle, E., & Sugihara, G. (2016). Redm: An r package for empirical dynamic modeling and convergent cross-mapping. *Cran. r-project. org.*

Ye, H., Deyle, E. R., Gilarranz, L. J., & Sugihara, G. (2015). Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific Reports*, *5*(1), 1–9.

A Appendix

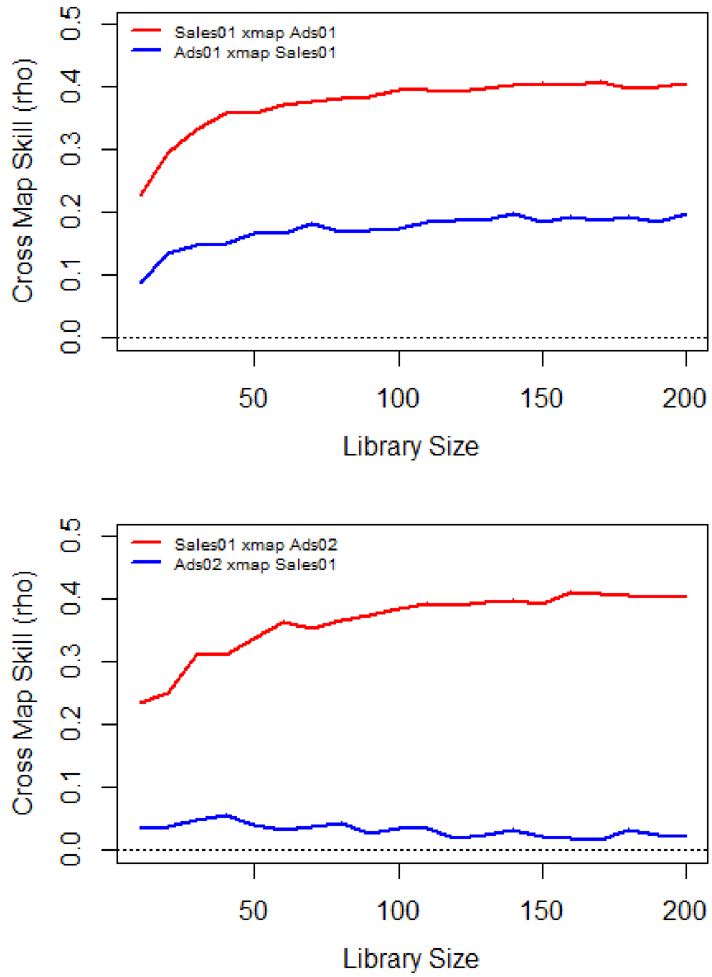


Figure 28: Cross Convergent Mapping (CCM) resulting plots for the variables *Sales01* and *Advertising efforts*, respectively *Ads01* and *Ads02*. These plots are used to determine whether convergence is present, implying that an upward trend should be observed given the increase in the library size.

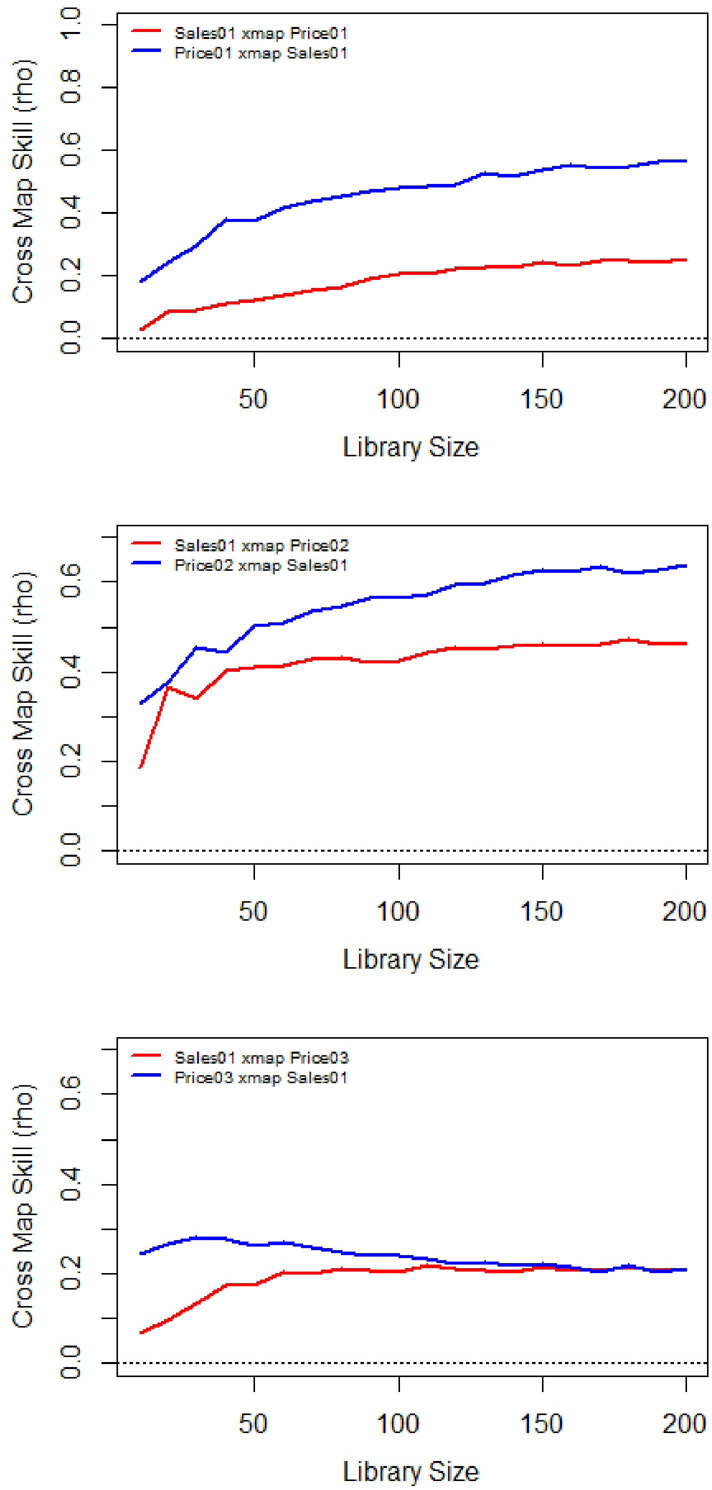


Figure 29: Cross Convergent Mapping (CCM) resulting plots for the variables *Sales01* and *pricing*, respectively *Price01*, *Price02* and *Price03*. These plots are used to determine whether convergence is present, implying that an upward trend should be observed given the increase in the library size.

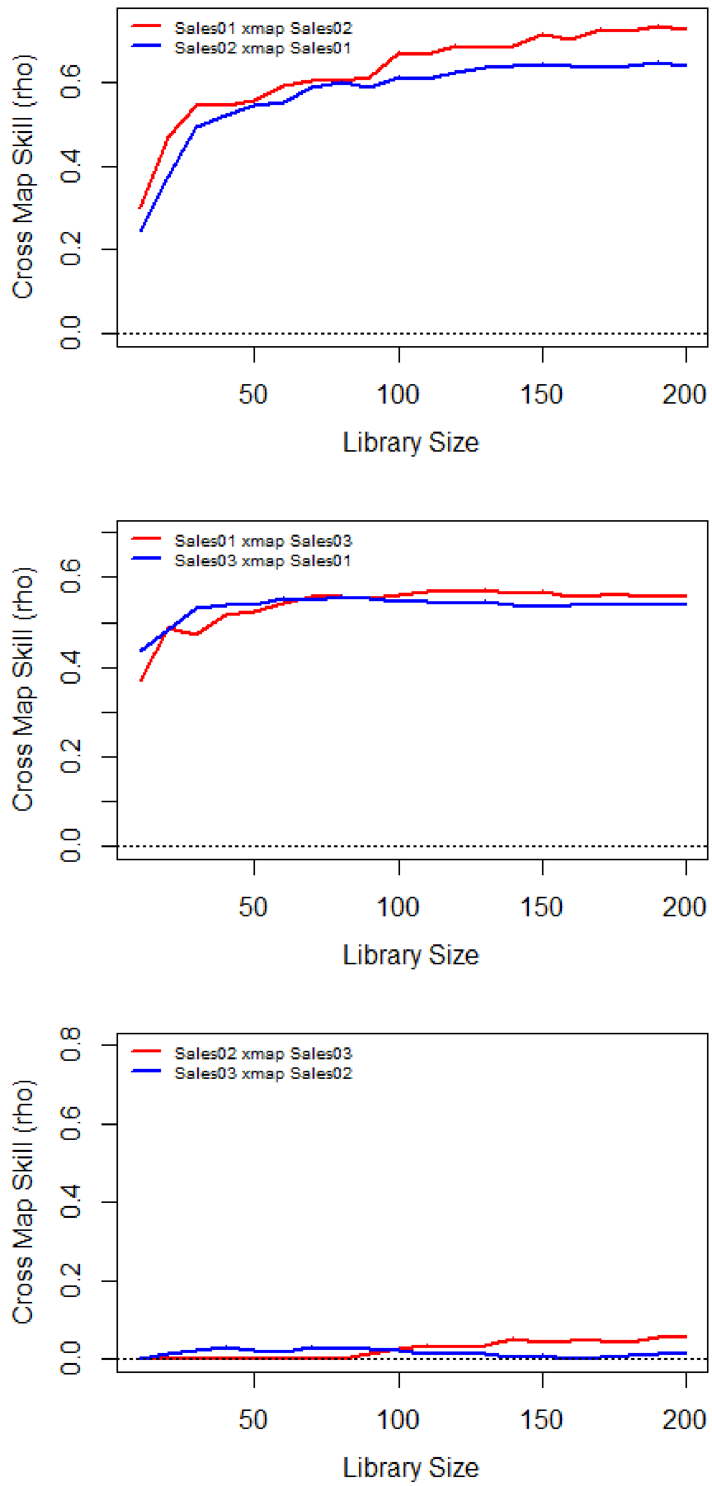


Figure 30: Cross Convergent Mapping (CCM) resulting plots for the variables *Sales01*, *Sales02* and *Sales03*. These plots are used to determine whether convergence is present, implying that an upward trend should be observed given the increase in the library size.

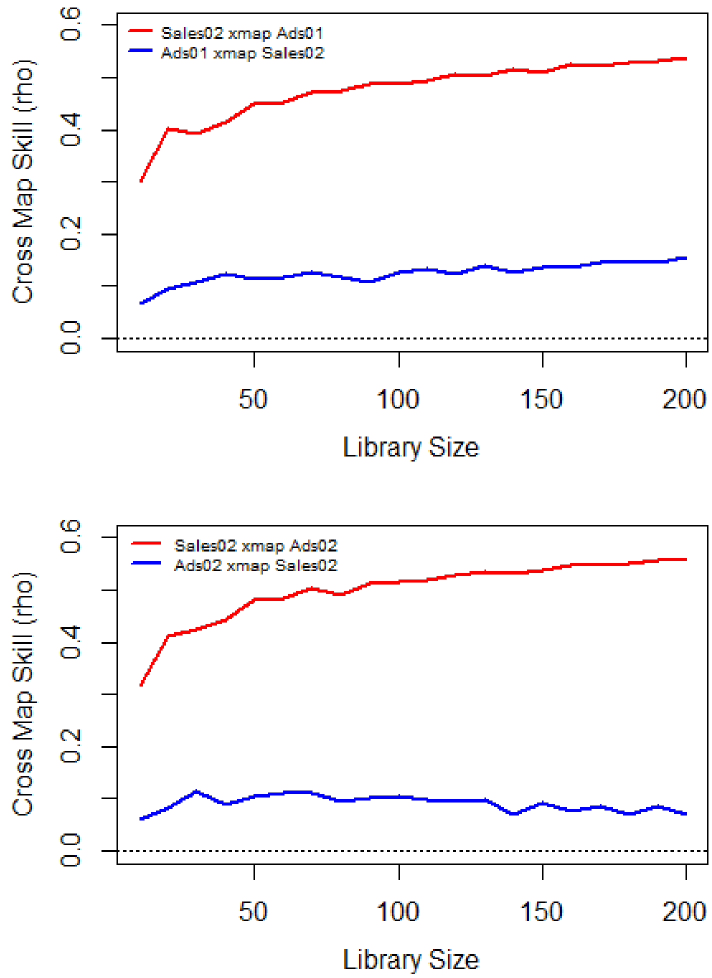


Figure 31: Cross Convergent Mapping (CCM) resulting plots for the variables *Sales02*, *Advertising efforts*, respectively *Ads01* and *Ads02*. These plots are used to determine whether convergence is present, implying that an upward trend should be observed given the increase in the library size.

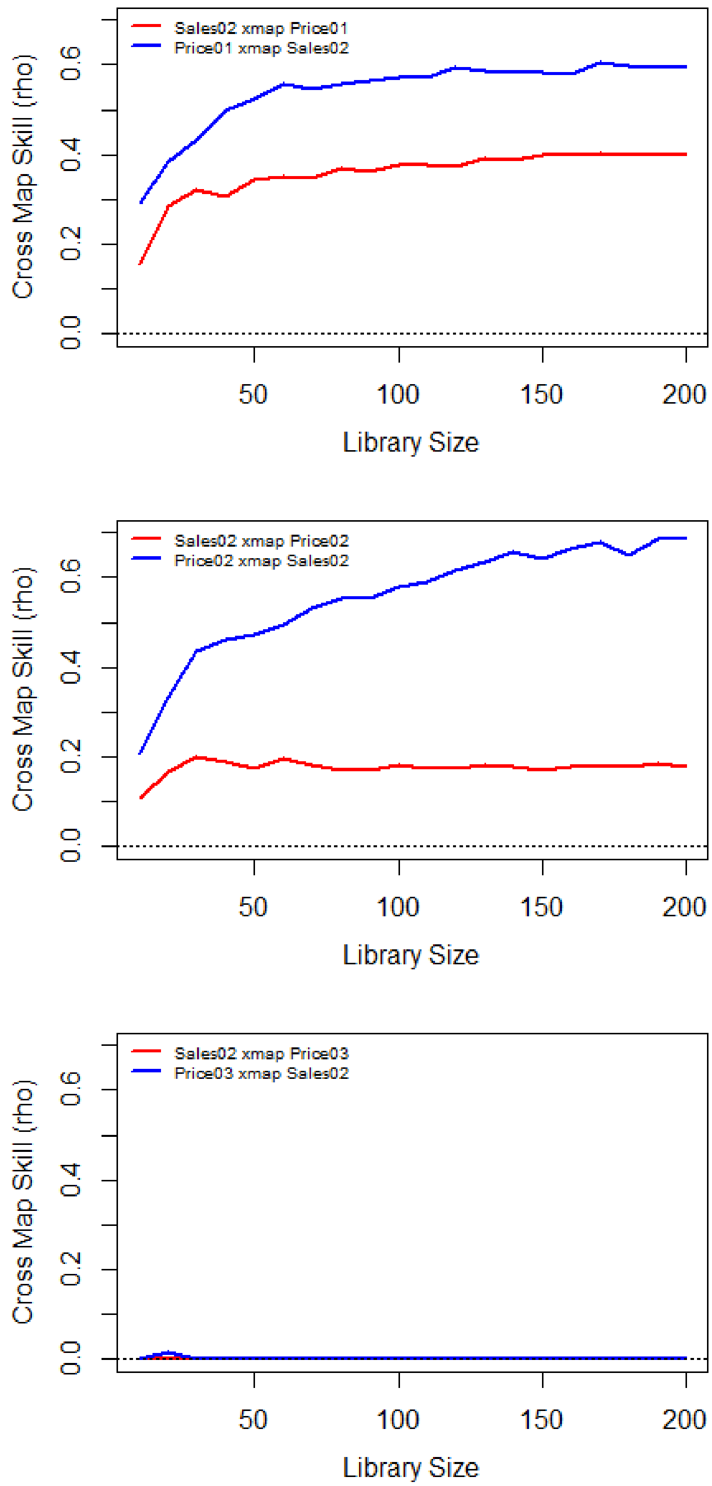


Figure 32: Cross Convergent Mapping (CCM) resulting plots for the variables *Sales02*, *Pricing*, respectively *Price01*, *Price02* and *Price03*. These plots are used to determine whether convergence is present, implying that an upward trend should be observed given the increase in the library size.

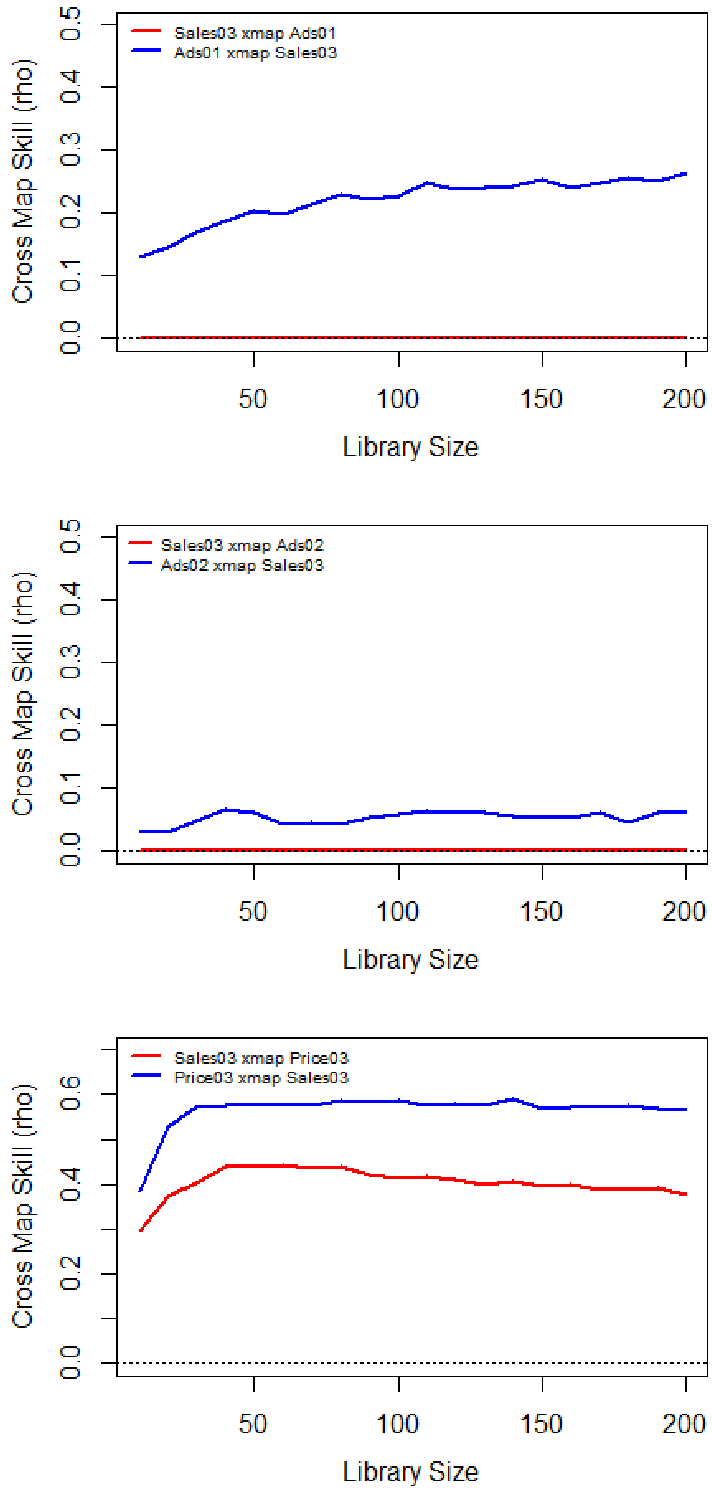


Figure 33: Cross Convergent Mapping (CCM) resulting plots for the variables *Sales03*, *Ads02*, *Ads01* and *Price03*. These plots are used to determine whether convergence is present, implying that an upward trend should be observed given the increase in the library size.