Empirical Dynamic Models: A Method for Detecting Causality in Complex Deterministic Systems

Nicola Rennie

Supervisor: Dr. Florian Dost

April 4, 2018

1 Introduction

One of the first concepts many statisticians learn about when it comes to the relationship between two (or more) variables is the concept of correlation: a statistic which is used to quantify a relationship between variables in terms of the strength and direction of that relationship. Although correlation is a valuable method for detecting a relationship between variables, the problem is that **correlation does not imply causation**. For example, Feldman (2017) details that sales of online TV subscriptions have increased over the past few years, and Doward (2017) describes how sales of 'plant-based alternatives to dairy products' have also increased in recent years, and so there is a positive correlation between them. However, based on correlation alone, it is not possible to conclude that purchasing a Netflix subscription would cause someone to also buy a bottle of almond milk, or vice versa. This is where the notion of causality comes in. The benefits of being able to identify a causal relationship between variables include the ability to detect whether a relationship is meaningful or spurious; and therefore being able to make better decisions based on the information available.

Standard methods of detecting causality, such as Granger Causality as described by Granger (1980), are based on the principle of linear correlation i.e. correlation which is constant over time. The problem with this is that a single variable can cause multiple effects (termed first-, second-, and third-order effects) which all occur at different rates. This can lead to there being periods of time when the correlation is positive, periods where it is negative, and periods where it is zero. That is, the correlation is changing over time, and is not well-represented by an average correlation across the whole time series and hence would not be a true representation of the extent of the relationship between two variables. For example, just as the presence of correlation does not imply causation, a lack of correlation does not necessarily suggest a lack of causality, as there may be positive and negative correlation occuring at the same time. This means that dynamic linear models are not suitable for this type of data.

Therefore a method which takes into account this nonlinearity is needed. This can be done by considering the variables of interest to be part of a dynamical system, which interact with each other. To find the 'true' equations which underlie a dynamic system in the real world is essentially impossible as there is such a large number of potentially causal variables and the relationships between those variables are extremely complicated. Rather than using a set of governing equations to predict the state of a multivariate system, first inspired by predator-prey dynamics in ecosystems, Dost and Maier (2017) state that empirical dynamic models (EDMs) are based on the idea of being able to 'infer the system dynamics from any single variable in the system'. This nonparametric, equation-free methodology is able to address the nonlinearity (and potential high dimensionality) in the system, and Deyle

et al (2016) have shown it to be 'robust to observational noise.'

This report seeks to explore this relatively new method and its use in identifying causality in multiple parallel time series data. An in-depth review of empirical dynamic models will be given, with a specific application which looks at identifying a relationship between online retail sales and energy consumption to assess the effects of e-commerce on the environment, as detailed in the article by Dost and Maier (2017).

2 Some Basic Concepts of Empirical Dynamic Models

Before a formal discussion of the theory behind empirical dynamic models can be given, some background terminology and definitions are needed.
Manifold: An n-dimensional manifold is a topological space that, for each point in the space, there is a neighborhood around that point which is topologically homeomorphic to $\mathbb{R}^n$, the Euclidean space. For example, any open subset of $\mathbb{R}^n$. A manifold is a **compact manifold** if it is 'compact as a topological space', as defined by Renze (2018). Weisstein (2018) explains that by this, it is meant that if a topological space, X, is the 'union of a family of open sets, then there is a finite subfamily whose union is X'.
Embedding: Insall et al (2018) describe an embedding as 'a representation of a topological object, manifold, graph, field, etc. in a certain space in such a way that its connectivity or algebraic properties are preserved. In the case of the embedding of a topological space this means the preservation of open sets'.
Diffeomorphism: 'A diffeomorphism is a map between manifolds which is differentiable and has a differentiable inverse', as detailed by Weisstein (2018).

2.1 Taken's Embedding Theorem

Taken's Embedding Theorem, detailed in Takens (1981), is the idea which underpins empirical dynamic models, allowing their use in time series forecasting and detecting causality between variables.

Theorem 1. *Let M be a compact manifold of dimension n. For pairs (φ, y), $\varphi : M \to M$ a smooth diffeomorphism and $y : M \to \mathbb{R}$ a smooth function, it is a generic property that the map $\Phi_{(\varphi,y)} : M \to \mathbb{R}^{2n+1}$, defined by:*

$$\Phi_{(\varphi,y)}(x) = \big(y(x), y(\varphi(x)), \ldots, y(\varphi^{2n}(x)\big)$$

is an embedding, where by 'smooth' it is meant at least C^2.

Proof Omitted.
Tsonis et al (2015) explain that, essentially, Taken's Embedding Theorem states 'the essential information of a multidimensional dynamical system is retained in the time series of any single variable of that system'. This means that omitted or unobserved variables, which could have an effect on the behaviour of the system, can be implicitly taken into account by considering a time series of just one of the observed variables of the system.

2.2 Representing a Time Series as a Manifold

Rather than thinking of the data as a set of time series observations, Sugihara et al (2012) suggest that it can be thought of as 'an **attractor manifold** in a multi-dimensional system state space'. Essentially, this means that each time series variable can be thought of as another dimension. These variables, which are thought (and being tested) to be causally linked, then form the cor-ordinate axis of a multivariate coordinate space, or state space. As described by Sugihara et al (2012), 'the state

of a dynamical system is a specific location in this multivariate state space'. This multi-dimensional manifold approach to time series mandates the application of Taken's Embedding Theorem, which allows the consideration of one variable at a time. This approach is used by Deyle and Sugihara (2011) who explore the idea that 'lagged variables of a single time series can be used as proxy variables to reconstruct an attractor for an underlying dynamic process'.

An **E-dimensional Delay Coordinate Embedding** is a manifold of vectors of length E which come from a variable and the time lags (the 'delay coordinates') up to lag $E-1$ of that variable. For example, for a time series X_t, the shadow manifold of dimension 4 would be formed of vectors of length 4 which take the form $(X_t, X_{t-1}, X_{t-2}, X_{t-3})$. The value E is known as the **embedding dimension**. By Taken's Embedding Theorem, if the embedding dimension is chosen correctly, then this embedding of a time series variable X represents the dynamics of the complete multivariate system (which includes the variable X). Hence, delay coordinate embeddings are also known as **shadow manifolds**, as described by Deyle and Sugihara (2011) as they are a 'shadow' of the true system from which inference can be drawn. (See Section 3.3 for an example of the construction of a shadow manifold from a single time series.)

3 Empirical Dynamic Models for Univariate Prediction and Detecting Nonlinearity

Sugihara (1994) put forth three potential sources of complexity in dynamic systems: observational error; the high dimensional nature of the system with many potentially unobserved variables; and the 'nonlinear interaction of these variables'. In practice, time series will often exhibit a combination of these three sources of complexity. Empirical dynamic models are equipped to deal with all three issues, though largely focus on the element of nonlinearity. The main issue with the application of empirical dynamic models is that they apply only to variables which exhibit behaviour that could come from a dynamic system.

Before fitting a multivariate model to detect causality between variables, it is essential to check whether this assumption that variables are part of a dynamic system is valid. One method of doing so is to fit a univariate empirical dynamic model to each of the variables to check whether the individual variable exhibits behaviour which could come from a dynamic system, and therefore whether fitting an empirical dynamic model is a suitable option. This is done by considering the individual variable, constructing a shadow manifold of the single variable for a training set of the observations, making predictions of future values, and then evaluating the predictions based on observations in the test set of observations. Note that in EDM terminology, the set of all shadow manifold vectors, constructed for each time point, is known as the **library**. Choosing the size of the library which will be used to make predictions is analogous to choosing the sample size of the time series which will be used as the training set. Alongside a more in-depth discussion as to the uses of univariate predictions and the quantification of what classifies as a 'good' prediction; two such methods of obtaining predictions of future (or out-of-sample) points from a single time series are discussed here: simplex projection, and S-maps.

There are two main uses for using univariate predictions besides making more accurate forecasts: i) detecting nonlinearity in a time series which will mandate the further application of empirical dynamic models; ii) selecting the optimal embedding dimension to determine how complex the system is. Both S-maps and simplex projection can be used to test for nonlinearity in a time series though Ye et al (2016) suggest that simplex projection is usually applied (with a varying embedding dimension) in order to find an optimal embedding dimension for the time series data, whilst S-maps are most often applied (with fixed embedding dimension and varying θ) to test for nonlinear dynamics in the time series data.

For simplex projection, given that the acceptance or rejection of the assumption that the data could come from a deterministic dynamic system (and the selection of the optimal embedding dimension) is based on whether the model has 'good' predictability, it is essential to define what is meant by 'good'. As discussed in Section 3.1, due to the short-term forecast horizon, most predictions are made using a 'leave-one-out' approach. These one-step-ahead forecasts can be compared with the true observations as a method of quantifying the forecast skill of a particular shadow manifold. For univariate predictions made using a simplex projection, this leads to the **simplex correlation**. This is usually taken to be the Pearson correlation coefficient between the simplex projection forecasts and the true observations, for one-step-ahead forecasts over a range of different time indexes. The closer ρ is to 1, the higher the forecasting ability of the model. Therefore, if the values of ρ are close to 1 then those variables can be said to exhibit the behaviour of a dynamic system and empirical dynamic models can be applied. The use of simplex projection predictions in obtaining an optimal embedding dimension is discussed further in Section 3.1.

As will be discussed in Section 3.2, in terms of using S-maps to detect nonlinearity, if the varying parameter $\theta > 0$, this is an indication that there is nonlinearity in the data. For a range of different values of θ, predictions can be made using S-maps and then compared with the observed values. S-map correlation is defined analogous to the simplex correlation i.e. it is simply the correlation between the S-map forecasts and the true observations. Again, the closer ρ is to 1, the better the forecasting ability of the model and hence, if the S-map correlation is higher when $\theta > 0$, this is an indication of nonlinearity. Due to the fact that S-maps are usually carried out with fixed embedding dimension, they are not usually applied to find an optimal embedding dimension.

Simplex projection can also be used to test for nonlinearity as an alternative to (or in combination with) S-maps. Using simplex projection to test for both nonlinearity and select an embedding dimension may lead to the problem of being unable to tell between whether the data does not come from a deterministic dynamic system or whether the embedding dimension has been poorly chosen. By convention, if the simplex correlation between the predicted values and the observed values is $\rho \geq 0.3$ for any value of E, this is seen as justification of the assumption of a deterministic dynamic system. The optimal embedding dimension, as discussed further in Section 3.1, balances a parsimonious choice of vector length with a higher value of ρ which indicates good forecasting ability.

3.1 Univariate Prediction Using Simplex Projection

One method of obtaining predictions in the univariate case is **simplex projection**. For a chosen embedding dimension E, the shadow manifold is built as described above by constructing a vector of length E for each observation of the time series, x_t, consisting of the observation and its lagged observations. Then the forecast of the shadow manifold vector at time $t+1$ is given by $\hat{S_{t+1}}$:

$$\hat{S_{t+1}} = \frac{\sum_{i=1}^{E+1} w_{t,i} S_{t+1,i}}{\sum_{i=1}^{E+1} w_{t,i}}$$

where are a number of different choices for the weights, $w_{t,i}$. This can be a simplistic as equal weights for each of the nearest neighbours, that is:

$$w_{t,i} = \begin{cases} \frac{1}{E+1} & \forall\ i = 1, \ldots, E+1 \\ 0 & \text{otherwise} \end{cases}$$

or more complex as in Dost and Maier (2017) who define the weights, $w_{t,i}$, as a function of the distance of the neighbour from the point scaled by the closest vector, as follows:

$$w_{t,i} = exp\left(\frac{-D\left(S_t, S_{t,i}\right)}{D\left(S_t, S_{t,1}\right)}\right)$$

where $D(S_t, S_{t,i})$ is the distance (see below) between the vector S_t and its i^{th} nearest neighbour. The forecast of the time series at time $t+1$ is then the first element in the vector $\hat{S_{t+1}}$. The method of nearest neighbours could potentially allow future values to be used in the predictions of past values which does not make sense, and so care must be taken when constructing algorithms to only allow past observations to be used. Using the univariate model in this way is similar to using a 'univariate autoregressive (AR) linear model' in that it uses previous observations of a time series variable to predict future values of that variable. However, simplex projection differs from AR models as the observations used in prediction are those which are closest in space rather than closest in time.

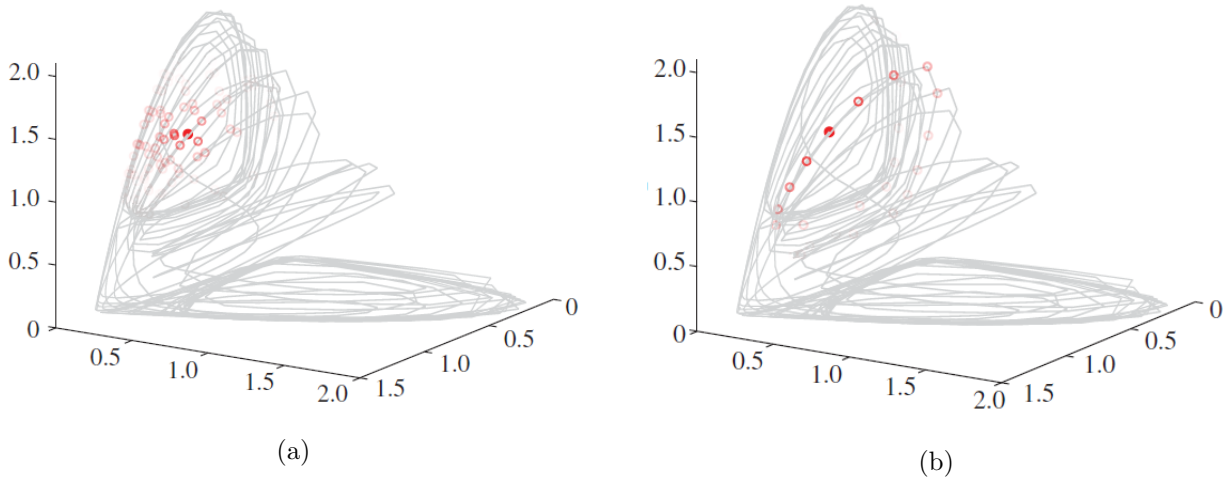


Figure 1: *Figure 1 shows the difference in weights given to each point in the forecast using (a) simplex projection (closest in space) (b) dynamic linear models (closest in time)* Deyle et al (2016)

Choice of Embedding Dimension:
The optimal choice of embedding dimension, E, should be a balanced choice such that the dimension of the embedding is minimised but the information contained about the whole system within that embedding is maximised. Let N be 'dimension of the whole system i.e. the number of causal variables in the system', then an upper bound for the minimum value of E is, as described in the Supplementary Information by Dost and Maier (2017):

$$E_{min} \leq 2N + 1$$

This upper bound comes from the dimension of the range of the mapping, $\Phi_{(\varphi,y)}$, described in Taken's Embedding Theorem. The proof is omitted here, though it is discussed further by Deyle and Sugihara (2011).

In practice, the value of N is often unknown due to the fact that many variables will be omitted or unobserved. Then this upper bound cannot be computed so the choice of embedding dimension is essentially free to be specified by the modeller. In the literature, when selecting E, a range of different values are tested, and the best performing value is chosen as the embedding dimension. The values are tested in a similar way to the suitability of EDMs as a method for testing causality between time series variables i.e. for each of the potential choices of E, a shadow manifold is constructed, predictions are made using the simplex projection of a training set of observations, these predictions are compared to the test observations, the value of E which makes the best predictions is chosen. In most cases, the predictability of the model increases as E increases then levels off. This often leads to choosing E as the value on the 'elbow' on the graph (of E against the predictability correlations for each value), as this obtains close to the best predictions but avoids overfitting by not choosing E too high. The other disadvantage of choosing a high value of E is data loss. For an embedding dimension of E, the first $E-1$ observations are discarded from the shadow manifold as there are not enough lagged

observations to construct the vectors. With large data sets, this is rarely a problem, however if the number of observations is limited in number, then this is a factor which must be considered.

Choice of Distance Metric:
In the calculation of the weights in the nearest neighbours prediction, there is the choice of which distance metric should be used. In particular the Euclidean distance ($D(a,b) = \sqrt{(\sum_i (a_i - b_i)^2)}$), the Manhattan distance ($D(a,b) = \sum_i |a_i - b_i|$), or the LP norm ($D(a,b) = (\sum_i (a_i - b_i)^P)^{\frac{1}{P}}$) can be used, which are the three options in the R package 'rEDM'. The Euclidean distance is by far the most commonly used, though for some particular data, alternative distance metrics may provide superior predictions. There is no single distance metric which will perform optimally for all time series, and so in practice different metrics should be tested and the best performing one chosen. There has been some limited research into the effect of the choice of distance metric on the performance of nearest-neighbours algorithms carried out by Hu et al (2016), but in general the choice is highly data dependent.

Choice of Sampling Rate:
In general, the sampling rate is chosen to be 1. This means that every observation in the time series could be used, and therefore potentially selected as a nearest neighbour when it comes to simplex projection. This is largely chosen as a matter of convenience. An alternative could be to reduce the sampling rate, for example considering only every second observation in the time series. Then for a time series X_t, the shadow manifold of dimension 4 would be formed of vectors of length 4 which take the form $(X_t, X_{t-2}, X_{t-4}, X_{t-6})$. Clearly, a disadvantage of reducing sampling rate is the fewer number of observations which can be used in predictions. However, in contrast to using all observations, if using the same number of nearest neighbours, it would likely also result in those nearest neighbours being further away from the current point. Therefore information from further away in the state space could be utilised in the predictions. The optimal choice of sampling rate is therefore highly dependent on the data, and the rates at which different effects are occuring.

Choice of Number of Nearest Neighbours:
An alternative method to increase the distance in the state space which is used in predictions, is to increase the number of nearest neighbours used in the predictions. Simplex projection typically uses $E+1$, i.e. the smallest number of nearest neighbours, as described in the Supplementary Information by Dost and Maier (2017). Increasing the number of nearest neighbours above this minimum value would include more points, from further away, in the forecast which could potentially increase predictability by reducing the effect of noise in the data but could also lead to overfitting, or to two different points which are close together producing the same forecast. The number of nearest neighbours is another parameter which can be tuned to the data, if the modeller wishes to do so.

Choice of Forecast Horizon:
As with any method of forecasting, the forecast horizon i.e. how far ahead to forecast given the current data, must be chosen. Due to the nonlinear nature of a system being described by using empirical dynamic models, only short-term forecasts are appropriate. If the accuracy of the forecasts does not decrease as the forecast horizon increases, this would mean that the system dynamics are not changing over time which suggests some linearity in the data, in which case other modelling approaches, such as AR models, may be more appropriate. In practice, this nonlinearity usually results in making one-step-ahead forecasts. In terms of detecting linearity, and selecting the embedding dimension, this leads to a 'leave-one-out' approach, where all past vectors (in the shadow manifold) 'except those that share values with the point being forecasted' can be used, described in the Supplementary Information provided by Dost and Maier (2017).

3.2 Univariate Prediction Using S-Maps

Another method of obtaining predictions in the univariate setting is by using **S-maps**. S-maps are very similar to simplex projection in that they use a delay coordinate embedding on a single time series to generate an reconstructed attractor manifold. This means that predictions are also made using points which are close in space as opposed to close in time. The difference between the two is in the algorithm used to make the predictions. S-map stands for 'sequential locally weighted global linear map' as defined by Sugihara (1994). The name comes from the fact that, unlike simplex projection which uses a fixed number of nearest neighbours surrounding the target point $\mathbf{x}(t^*)$, S-maps make use of all points in the library, with different weights which depend on how far away they are from $\mathbf{x}(t^*)$.

In the univariate case, for a single selected embedding dimension, E, the one-step-ahead forecast from a target point $\mathbf{x}(t^*)$ using S-maps is given by:

$$\hat{\mathbf{x}_{t+1}} = \sum_{i=1}^{n} w_i \mathbf{x}_{t+1,i}$$

where n is the size of the library (which, for a sampling rate of 1, is essentially the length the (past) time series minus $E-1$); and the weights, w_i, are:

$$w_i = exp\left(\frac{-\theta D\left(\mathbf{x}(t_i), \mathbf{x}(t^*)\right)}{D_n}\right)$$

where $D\left(\mathbf{x}(t_i), \mathbf{x}(t^*)\right)$ is the distance between the target point and each library point (see Section 3.1) , and D_n is 'the mean distance of all paired library points', as detailed by Chang et al (2017). That is:

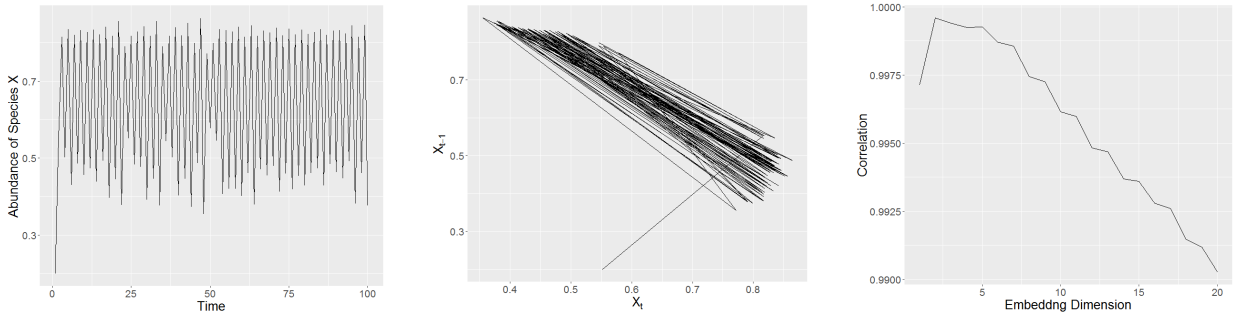
$$D_n = \frac{1}{n}\sum_{i=1}^{n} D\left(\mathbf{x}(t_i), \mathbf{x}(t^*)\right)$$

The parameter θ controls the 'degree of state dependency' i.e. how much weight is given to neighbours which are close compared with neighbours which are far. Choosing $\theta = 0$ results in all vectors in the library being given equal weight i.e. it reduces the model to a linear autoregressive model, as there is no priority given to points which are closer in space. If $\theta = 0$ is the optimal choice of this tuning parameter, then this suggests that a linear model provides superior forecasts, rejecting the assumption of nonlinearity meaning that empirical dynamic models need not be applied. Therefore, if the optimal choice of θ is greater than zero, this indicates nonlinearity in the system. Finding the optimal choice of θ in the S-map algorithm is easier than finding the optimal embedding dimension for the simplex projection method as there is no need to balance parsimony with forecasting ability: θ does not change the complexity or dimension of the model. Howver, it should be noted that choosing θ to be too small can underestimate the degree of nonlinearity resulting in poor forecasts, whereas choosing it to be too large results in excessive weight being given to closer points making the model more sensitive to observation error, see Section 5. Simply, the value of θ which returns the highest value of S-map correlation, ρ, is chosen. See Section 4.2 for further inofrmation on the theory behind S-maps.

As mentioned previously, due to the fact that E is generally fixed, S-maps are not typically applied as a method for finding an optimal embedding dimension. In the univariate case, they are used to accept or reject the assumption of nonlinearity, by assessing the forecast ability based on different values of θ. In general, simplex projection should be applied first to obtain the optimal embedding dimension, then S-maps to confirm if, given that the optimal E has been selected, the system exhibits nonlinear behaviour. This eliminates the problem of using simplex projection for both selecting the embedding dimension and assessing the assumption of nonlinearity (and therefore potentially being unable to tell whether the system does not exhibit nonlinearity or whether the embedding dimension has been poorly chosen), described above, as there is now a method which can be used to test for noninearity separately. The choice of forecast horizon, distance metric, and sampling rate are as

with the simplex projection method, however the choice of number of nearest neighbours is addressed directly by S-maps as by defintion they use all points in the library.

3.3 Example: Two-Species Model Data



(a) (b) (c)

Figure 2: *(a) Time series of the abundance of species X in a two species coupled model (b) Shadow manifold constructed from the time series consisting of the series and its own lag(s) (c) Correlations of univariate simplex projections and observed values over selected embedding dimensions*

Using the 'two_species_model' dataset available in the 'rEDM' package in R, the process of constructing a shadow manifold, selecting an embedding dimension and making unvariate predictions can be demonstrated. The data consists of two time series of the abundance of two species X and Y, which have been generated from a discrete-time coupled Lotka-Volterra model exhibiting chaotic dynamics, with a total of 1000 observations. Due to this generation process, it is known that empirical dynamic models are a suitable approach for obtaining inference, however this can be confirmed by the above described methods.

For clarity, only the first 100 observations have been plotted in Figure 2 (a) and (b), however all 1000 observations are used in any calculations and predictions i.e. a sampling rate of 1 is used. Figure 2 (a) is a time series plot of the abundance of species X, which appears to be stationary but with high variablility. Figure 2 (b) is the constructed 2-dimensional shadow manifold. It consists of 99 vectors: (X_t, X_{t-1}) for $t = 2, ..., 100$ with X_t plotted on the x-axis, and X_{t-1} plotted on the y-axis. It can be seen that there are areas of both positive and negative correlations displayed in this shadow manifold. Figure 2 (c) shows the plot of each of the possible embedding dimensions, from 1 to 20, against the simplex correlation. This simplex correlation, even at its lowest value for the selected embedding dimensions, always exceeds 0.99 which is far above the convential threshold of 0.3. This mandates the further application of empirical dynamic models. In terms of selecting the optimal embedding dimension, $E = 2$ is clearly the best choice as it has the highest forecast skill, is a low dimension avoiding overfitting, and is on the 'elbow' of the graph. Due to the fact that the data has been generated from a known process with only two variables, the upper bound discussed in Section 3.1 is therefore $2 \times 2 = 5$. However, to demonstrate the idea that predictability does not necessarily increase by including an inreasing number of dimensions, the correlations that result from embedding dimensions up to 20 are included.

To highlight the difference between selecting an embedding dimension under the framework of empirical dynamic models, and selecting the order of an auto-regressive model under a linear framework, an AR model was fitted to the data. Using the 'ar' function in R, which returns the optimal order (i.e. specifies which time lags are in the prediction) and the coefficients, the following model was selected based on all 1000 observations:

$$\begin{aligned} X_t &= -0.4222X_{t-1} - 0.1534X_{t-2} - 0.2273X_{t-3} + 0.1894X_{t-4} - 0.0753X_{t-5} - 0.1660X_{t-6} \\ &- 0.1936X_{t-7} - 0.0436X_{t-8} - 0.1329X_{t-9} + 0.0485X_{t-10} - 0.0685X_{t-11} + \epsilon_t \end{aligned}$$

t	AR $\hat{X}_t$	AR Order	SP $\hat{X}_t$	SP E	Observed x_t	SP ρ	S-Map θ	S-Map ρ
20	0.4974235	1	0.4780849	2	0.44644	0.9974002	8	0.9942790
200	0.4529932	7	0.4821248	3	0.48574	0.9990626	8	0.9973820
500	0.4658124	9	0.4463868	2	0.44662	0.9988847	4	0.9980870
1000	0.4451992	11	0.4454644	2	0.44553	0.9996825	4	0.9982177

Table 1: *Analysis of 'two_species_model' data using autoregressive (AR) models, simplex projections (SP), and S-Maps*

This model has order 11, i.e. uses 11 past values of the time series in predictions, which is in contrast to the empirical dynamic model, whose simplex projection method selects 3 as the optimal number of past observations used in prediction.

To further demonstrate the power of empirical dynamic models over their linear counterparts, the both an autoregressive model and a simplex projection are used to make one-step ahead forecasts for a range of time indexes. Table 1 clearly shows that in all cases examined, the simplex projection exceeds the autoregressive model in terms of predictability with forecasts closer to the observed values. It also shows that, as the number of observations in the time series increases, the order of the AR model increases, as does predictability. Note that the coefficients and order of the AR model change for each time point as only the past data is used in predictions.Conversely, the optimal embedding dimension in the simplex projection method changes very little, meaning that more accurate forecasts are being obtained using fewer data points in each prediction. In terms of using S-maps to detect nonlinearity, the optimal values of θ (for all points tested below) are well above 0 indicating nonlinearity in the data, which agrees with the high values of the S-map correlation which suggest good forecasting ability when taking the nonlinearity into account. It is interesting to note that in all below cases, the simplex correlation is slightly higher than the S-map correlation, suggesting that for this data simplex projection provides superior short-term forecasts. The R code used to carry out the above analysis and produce Figure 2 is available in the appendix.

4 Empirical Dynamic Models for Multivariate Prediction and Detecting Causality

Empirical dynamic models can be used to explore the possibility of a causal relationship between two (or more) variables. Chang et al (2017) describe two variables as being 'causally linked if they interact in the same dynamical system'. There are two methods, closely related to the univariate simplex projection and S-maps methods, which can be used to make multivariate predictions and hence, detect causality between those variables: convergent cross-mappings (Section 4.1), and an extension of the univariate S-maps algorithm (Section 4.2). Both methods are essentially extensions of the univariate EDM methods described in the previous section, and rely on the same basic principles in terms of how predictions are made and how the forecasting ability is quantified.

4.1 Multivariate Prediction Using Convergent Cross Mappings

The method of using simplex projection to make forecasts can be extended to the multivariate case as a method for empirically testing for causality between variables. This process is known as **Convergent Cross-Mapping** (CCM), by Sugihara et al (2012), and the name comes from the idea of using an embedding of one variable to make cross-predictions of another then looking for convergence between those predictions and the observed values.

Supppose that there are two time series of variables X and Y. The basic concept is to use the shadow manifold of the variable X to predict vectors in the shadow manifold of Y_t, and therefore make forecasts of the variable Y. The first step is to construct the univariate shadow manifolds for X_t and Y_t under an optimal embedding dimension, E (see Section 3.1). Using simplex projection, the nearest neighbours to the vector X_t are found and the weights relating to each of the neighbours are determined using the same process as in the univariate case. These weights are then applied to the corresponding (in time) vectors in the shadow manifold of Y_t, in order to make a forecast of future values of Y_t. For example, if the nearest neighbours to X_t are X_1, X_{11} and X_{25} in a two-dimensional embedding. The weights are calculated according to the distance of those vectors from X_t and then applied to the vectors Y_1, Y_{11} and Y_{25}, even if these vectors are not closest in space to Y_t. Therefore, the algorithm to make a cross-prediction of the variable Y at time t is as follows:

$$\hat{Y_t} = \frac{\sum_{i=1}^{E+1} w_{t,i} X_{t,i}}{\sum_{i=1}^{E+1} w_{t,i}}$$

where, as specified previously, the weights, $w_{t,i}$ are defined as:

$$w_{t,i} = exp\left(\frac{-D\left(X_t, X_{t,i}\right)}{D\left(X_t, X_{t,1}\right)}\right)$$

where $D\left(X_t, X_{t,i}\right)$ is the distance between the vector X_t and its i^{th} nearest neighbour. The forecast of the time series of Y at time $t+1$ is then the first element in the vector $\hat{Y_{t+1}}$. The same issues as in the univariate case, such as choosing the sampling rate, distance metric, and the number of nearest neighbours, are present in the use of CCM.

Analogous to the univariate case, the forecasting ability is measured by the correlation (termed **cross-correlation**) of the cross-predictions and the observed values of Y, when a 'leave-one-out' method, for example, of cross-validation is applied. Again, the closer that this correlation is to 1, the higher the forecasting ability of the model. If the correlation is close enough to 1, this is an indication that some information about Y can be recovered from the shadow manifold of X. Sugihara et al (2012) proved that as the size of the library tends towards infinity, the cross-correlation will converge to 1 if there is 'even a weak deterministic causal impact and perfectly measured noise-free data'. However, in practice, library sizes are always finite and so, if it occurs, convergence will be to some value less than 1. If the cross-prediction converges to a **significantly positive** value, this is evidence that there is a causal relationship, termed 'coupling' by Sugihara et al (2012), between variables. If there is no convergence, then no identifiable causal relationship exists between those two variables. The issue with this method is in deciding what constitutes a significantly positive value, with convention again being values of cross-correlation above 0.3 being accepted as an indication of a causal relationship. Dost and Maier (2017) note that there is 'no formal statistical test' to make this decision, see Section 7.1 for a discussion on the potential future development of this methodology. Although this lack of statistical test makes it more difficult to be confident in the conclusions inferred, bootstrapping can be used to generate confidence intervals for the cross-correlation which can provide some reassurance as to whether there is in fact a coupling between variables.

4.2 Multivariate Prediction Using S-Maps

The issue with using simplex projection is that it does not measure the direction or strength of the effect of one variable on another. In the same way that univariate S-maps can be used to detect the extent of nonlinearity in a system, multivariate S-maps can be used to detect the extent of the effects of variables on one another, termed the 'state-dependent marginal effects' by Dost and Maier (2017). Deyle et al (2016) explain that, when a system is at equilibrium, the **community matrix** is defined as the 'matrix of partial derivatives of the system' i.e. the Jacobian. The community matrix can then be used to consider the interactions between pairs of variables i.e. the marginal effects. However,

when a system is dynamic it is not at equilibrium and so this matrix needs to be recalculated at each point in space. This may seem compuationally intensive, but by using S-maps it can be fairly easily accomplished. Deyle et al (2016) explain that the underlying idea of the S-maps method is to sequentially calculate interaction matrices as the system moves along the attractor manifold, described in Section 2.

As in the univariate S-maps case, multivariate S-maps involve using all of the vectors in the library and weighting them locally around the point $\mathbf{x}(t^*)$, the target point on the attractor manifold. Deyle et al (2016) hence describe S-maps as 'a locally weighted multivariate linear regression scheme', with larger weights given to points which are close to $\mathbf{x}(t^*)$. For an optimally chosen (as described above) embedding dimension E, in a process similar to other regression models, S-maps approximates the following equation, as defined by Deyle et al (2016):

$$\hat{x_i}(t^* + 1) = C_0 + \sum_{j=1}^{E} C_{ij} x_j(t^*)$$

by determining the coefficents of the linear model, $\mathbf{C}$. In a linear model, this would involve estimating it across the whole time series, in terms of how the observations occur in time. However, in S-maps, this linear approximation occurs **locally** around each location $\mathbf{x}(t^*)$. As in the univariate case, Deyle et al (2016) define the weights as:

$$w_i = exp\left(\frac{-\theta D\left(\mathbf{x}(t_i), \mathbf{x}(t^*)\right)}{D_n}\right)$$

where $D\left(\mathbf{x}(t_i), \mathbf{x}(t^*)\right)$ is the distance between the target point and each library point (see Section 3.1) , and D_n is defined analogously to the univariate case, that is:

$$D_n = \frac{1}{n} \sum_{i=1}^{n} D\left(\mathbf{x}(t_i), \mathbf{x}(t^*)\right)$$

This is almost identical to the univariate case, except here the vectors contain observations of multiple variables rather than the lags of a single variable. The choice of how to define the distance is as before in both the univariate simplex projection and s-maps methods, with the Euclidean distance still being the most common. Again, analogous to the univariate case, the tuning parameter θ represents how mch more weight is given to points close to the target point compared to this which are further away. Here, if the optimal choice of θ is greater than zero, this means that the coefficients C change at each location in the attractor manifold. This compares to choosing an optimal $\theta = 0$ which means that the coefficients are constant throughout the attractor manifold, reducing the model to a multivariate autoregressive model. According to Deyle et al (2016), the coefficients, $\mathbf{C}$, 'approximate the Jacobian or interaction elements at successive points along the attractor', and these can be taken as the 'state-dependent marginal interaction effects between variables', explained by Dost and Maier (2017). As noted by Deyle et al (2016), using 'leave-one-out' cross validation is a common method of evaluating the performance of the model, which also helps to avoid overfitting.

Using a locally weighted linear regression as described above, Dost and Maier (2017) and Deyle et al (2016) describe the S-map model as being the singular value decomposition solution for the equation:

$$\mathbf{B} = \mathbf{A} \cdot \mathbf{C}$$

where:

$$A_{ij} = w_i x_j(t_i)$$

gives the matrix $\mathbf{A}$ to have dimension $L \times E$ where L is the length of the library in the shadow manifold for the target variable, and E is the pre-selected optimal embedding dimension. This means that $\mathbf{A}$ is simply a matrix of the weighted attractor manifold vectors. Deyle et al (2016) describe $\mathbf{B}$ as being the 'n-element vector of the predicted variable', that is, the 'future values of the target variable x_j':

$$B_i = w_i x_j(t_i + 1)$$

5 Critical Discussion of EDM Methodology

As described in Section 2 and Section 7, one of the benefits of the use of Taken's Embedding Theorem in emprical dynamic models in the fact that unobserved or omitted variables can implicitly be taken into account. However, when evaluating marginal effects of variables using S-maps, explicitly accounting for these variables will only increase the quality of the estimates of these effects. As will be discussed in Section 7, when previous studies have returned conflicting results, for example say Study 1 concludes that A has a positive effect on B, whilst Study 2 concludes that A has a negative effect on B, EDMs may be of particular use in this case. This could be because there are both positive and negative effects occuring as a result of the same variable with linear models in the studies only picking up on one or other, or due to the effects of unobserved variables, both of which EDMs can deal with.

Though obtaining a sufficiently large sample size is an issue for most methods of time series analysis, this becomes more of an issue in the application of empirical dynamic models. This is due the assumption of a deterministic system which, since there is stochasticity present in most real world systems, contains error in the observed values. Chang et al (2017) note that the minimum number of observations needed in order to apply EDMs increases with the optimal embedding dimension, potentially due to the data lost when constructing the shadow manifold vectors, as discussed in Section 3.2. Sugihara et al (2012) recommended that between 35 and 40 observations of the time series are required for the application of empirical dynamic models. Due to the fact that, by definition, empirical models do not make use of any underlying distribution, it is difficult to state exactly what this minimum number of observations is for any given set of variables. This assumption of a deterministic system, for the above reasons, highlights the need for reasonably reliable measurements, as there is already error present in the model. Unreliable measurements in the time series may also cause problems with endogeneity, as discussed by Dost (2018), that is, the presence of correlation between a variable and the error term. Endogeneity can result in bias in predictions, which, depending on the type of error, may result in either concluding there is a causal relationship when none is present, or being unable to detect a genuine causal relationship, which defeats the purpose of applying empirical dynamic models.

As mentioned by Chang et al (2017), if a system is nonlinear, that is 'driven mainly by low-dimensional, deterministic processes', then building an empirical dynamic model which incorporates this behaviour and has fairly good forecasting ability is possible, confirming the ideas described by Sugihara (1994). However, the reverse is not true. If a system, is linear and/or driven by a stochastic process, building an empirical dynamic model which deals with this behaviour is not possible. Sugihara (1994) also notes that 'it is impossible to distinguish high-dimensional nonlinear systems from linear stochastic systems given time series data'. This means that quantifying the nonlinearity of a system, usingn S-maps as described in Section 3.3 and Section 4.2, is particluarly useful. Comparing EDMs with dynamic linear models, Deyle et al (2016) note that if a system is changing very slowly in time i.e. 'to produce a nearly linear case' the empirical dynamic models, such as S-maps, and dynamic linear models, such as autoregressive models, produce similar results. However, in ecological systems, and other real-world applications, systems often change rapidly in time. In this case, empirical dynamic models provide superior results, as they recognise that states which occur close together in time may in fact be very different.

EDMs were developed to deal with nonlinear or state-dependent relationships between variables, and there will be some time series variables which do not exhibit this behaviour. EDMs are by no means meant to be blindly applied to any time series variable, hence why testing their suitability using the methods described in Section 3 is of great importance. Just as with any method for modelling time series and making forecasts, or indeed any statistical model, there are times when it appropriate to apply the method, and times when it is not.

6 An Application of Empirical Dynamic Models

6.1 Effects of e-Commerce

In terms of considering the effects of e-commerce, first order effects relate to energy use by hardware, for example server farms which house IT equipment and keep the retailer's business online. Second order effects relate to the changes in the retail process as people choose online retail over other alternatives. These effects include the increase in the number of cardboard packaging boxes as a result of the increasing number of deliveries, but also extends to the fewer number of people going out shopping in supermarkets. As mentioned in Dost and Maier (2017), most studies into the effect of e-commerce on the environment have concentrated on these second order effects, neglecting potential third order effects. For example, though fewer people may be using energy to drive to supermarkets, they may be using energy in a different way e.g. by using electricity watching television at home. Third order effects encapsulate indirect feedback and changes to the system. For example, one of the reasons many shoppers choose online retail, is not only convenience but also price. Online retail can create a more efficient supply chain by reducing waste, which decreases prices, and therefore increases demand, leading to a long-term change in the system.

6.2 Application of EDMs to e-Commerce Effects

In the paper by Dost and Maier (2017), the effects of e-commerce on the environment are explored by considering five different variables using monthly U.S. national data: online retail share (percentage of all retail sales (by value) which were made online); transportation energy consumption; residential energy consumption; commercial energy consumption; and industrial energy consumption. The time series of the online retail share, and residential energy sector variables are shown in Figure 2 (a).

Univariate simplex projection was used both to test the suitability of the data to emprical dynamic models, and to select the optimal embedding dimension. This application highlights another issue when it comes to selecting the embedding dimension: different values of E are optimal for different variables. The optimal E for the online reatil share variable is 12, whilst the optimal E for each of the energy variables varies between 6 and 7, although there is are also peaks in forecast skill at $E = 12$ for these variables. An optimal embedding dimension of 12 may be expected from data which was collected on a monthly basis, as each point in the state space then contains information about an entire year, allowing annual patterns to be taken into account. To balance retaining as simple a model as possible without losing quality of the represntation of the dynamic system, the optimal E was chosen to be the lower values. The difference predictability between choosing $E = 6$ or 7, was shown to be small, meaning the model is robust to changes in embedding dimension, however, this may not be true for all time series data. In terms of testing for nonlinearity, all five variables returned a simplex correlation ρ close to 0.9, which is above the threshold for accepting the nonlinearity assumption of 0.3. Hence, this mandates the application of empirical dynamic models to test for causality between these variables, as they all exhibit behaviour which comes from a deterministic dynamic system.

The nonlinear nature of the variables is further confirmed by visual inspection of the 2-dimensional empirical system attractor manifold of the online retail share and residential energy consumption variables, shown in Figure 2 (b). This figure shows both the negative and positive correlation which occurs between these two variables. In terms of evaluating whether there is a causal relationship between online retail share and energy consumption, convergent cross-mappings were applied. Figure 2 (c) shows the cross-correlations for each of the four categories of energy consumption when cross-mapped with the online retail share. It shows that, as the length of the library increases the cross-correlation converges (to different values for each of the four categories). From this, two inferences can be drawn: i) there is a causal relationship between each of the four variables and the online retail share as the value that the cross-correlation converges to is always above 0.6; ii) exhibited by the different

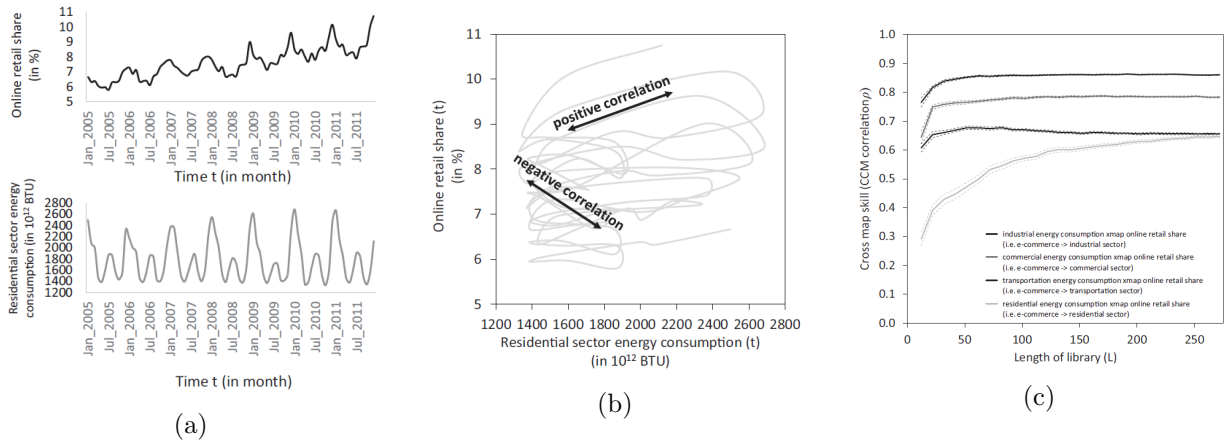


Figure 3: *(a) Time series variables of online retail share, and residential energy consumption (b) Empirical system attractor manifold showing correlation changing over time resulting in nonlinear trajectories (c) Correlations of cross-mapped and observed values over the length of the time series* Dost and Maier (2017)

values that the cross-correlation converges to, the effect of online retail share on energy consumption differs depending on which category the energy use falls into.

7 Areas of Further Research

In terms of further research in this area, there are three main approaches that can be taken: i) developing theory and methodology relating to EDMs in order to make it more efficent and statistically valid; ii) investigating other systems where EDMs could be applied to in order to obtain better inference; iii) extending the application of EDMs to the effects of e-commerce by considering additional variables. These three approaches are each discussed further below.

7.1 Development of Methodology in the Use of Empirical Dynamic Models

As empirical dynamic models are a relatively new method for detecting causality between variables, there is huge potential for the development of new methodology to support the application of such models.

Optimal Embedding Dimension Selection:
Given that many systems are complex and there are often many variables, using the current trial-and-error method to chose E could potentially take a long time. Since the correct choice of E is essential for applying Taken's Embedding Theorem, the EDM methodology hinges on the choice of embedding dimension. One potential future area of research is to develop a method for selecting the embedding dimension for the simplex projection based on the time series of the observations of that variable. A few of potential approaches are outlined here. As discussed in the example in Section 3.3, though choosing E is analogous to specifying the order of an autoregressive model, it is not the same process, nor value. The partial auto-correlation function is a graphical indicator of the order of an autoregressive model, and it measures correlations between points a specific time lag apart. Though this means that the partial auto-correlation function cannot immediately be applied as a to the simplex projection method, there is, however, potential to adapt the methodology to consider correlations between points a specific distance in space apart.

Further, as noted in Section 6.2, an embedding dimension of 12 is not unexpected for data which

has been recorded monthly. This suggests there is potential to exploit any qualitative knowledge that the modeller has about the data, either its generating process or collection method; as well as exploring whether seasonality is There is also an opportunity to explore whether aggregating data to the bi-monthly level would allow an embedding dimension of 6 to be selected optimally i.e. retains the forecasting ability of the model but has a more parsimonious length of vector in the library.

Accounting for Stochastisity:
One of the main criticisms of empirical dynamic models is that they are deterministic models which do not take stochasticity into account and produce point estimates as forecasts. A better approach would be to return a distribution for the forecast, centred around this point estimate, which would allow uncertainty to be taken into account. This would allow for confidence intervals to be calculated around this point estimate, which in turn opens up new avenues for testing the forecasting ability of the model. Rather than measuring the quality of the forecast by how correlated the predictions are with observed values, there is potential to measure this ability by what fraction of the observed values lie within a confidence interval of the forecasted value. This would also give a better indication as to which forecasts are causing the overall predictability to be low, in cases where the correlation is low. Obviously, stochastic models are at the opposite end of the spectrum from empirical dynamic models, in terms of which assumptions are being made, which makes combining their methodologies difficult. As yet little work has been done in this area, making it one of the key areas where further research is necessary. When Perretti et al (2013) applied empirical dynamic models to insect population data from Dennis et al (2001), they found that 'the optimal forecasting model for age-structured populations may be one that combines a mechanistic model for the upper age classes with a time series model for the recruit dynamics'. Combining two different types of models in predictions, is another potential way of including stochasticity into empirical dynamic models.

Detecting Changes in the System Dynamics:
Section 3 discussed the 'leave-one-out' method of cross validation used to assess the forecasting ability. An alternative approach described by Chang et al (2017) is to split the entire library in half, using the first half 'as the library set for out-of-sample forecasting of the reserved other half'. This leads to problems if there is a change in the system dynamcis between the first and second half of the time series. In order to determine whether poor forecasting ability is due to the data not exhibiting the behaviour of a dynamic system, or whether there is a change in the system dynamics between the training set and the test set, it is essential to develop a method which can detect a change. There is little literature on this topic, but a starting point could potentially be to consider whether any time series changepoint detection methods could be adapted to deal with shadow manifolds.

Statistical Test for Convergence of Cross-Correlation in CCM:
As mentioned in Section 4.1, there is no formal statistical test for whether the cross-correlation is significantly large enough to conclude that there is a causal relationship between variables. The development of such a test, would allow better conclusions and more accurate inference to be drawn from the convergent cross-mappings.

Similarly, as described in Section 3.2, if the optimal value of θ in the univariate S-map is greater than zero, this is an indication of nonlinearity. This could lead to problems if the optimal value is very close to, but not equal to, zero. If the optimal value of *theta* was found to be, for example, 0.00001, this would still be taken, perhaps somewhat cautiously, as evidence of nonlinearity. Again, a formal statistical test would allow the modeller to make a conclusion with some level of confidence. For example, a hypothesis test such as the follwoing:

$$H_0 : \theta = 0 \text{ vs } H_A : \theta > 0$$

based on either on some underlying distribution, or another non-parametric test would prove more statistically rigorous than a hard line drawn at zero.

7.2 Further Applications of Empirical Dynamic Models

Although intially used in ecology and epidemiology applications, empirical dynamic models have more recently been used to model the effects of e-commerce by Dost and Maier (2017) (see Section 6), detecting causality between galactic cosmic rays and interannual variation in global temperature by Luo et al (2015), and modelling coupled financial systems as described by Deyle and Sugihara (2011), amongst others. Due to the empirical nature of the model, it is widely applicable to many different data sets. However, there are also many data sets to which it cannot be applied as they do not exhibit the behaviour of a dynamic system. Testing the data using the methods described in Section 3 is the only way to conclude whether or not empirical dynamic models are suitable. However, EDMs generally work best when the time series observations are based on (reasonably) reliable measurements, or when current research is returning conflicting models (potentially as a result of both negative and positive correlation occuring at the same time). As described in Section 5, one of the limitations of empirical dynamic models is the requirement for a sufficiently large sample of data. With the rise in popularity of 'big data', small sample sizes are becoming much less of an issue and so there are likely many more possible applications, though small sample size may still be an issue for some specific applications.

The benefits of identifying additonal areas where EDMs can be applied are two-fold: i) inference and information about causal relationships between variables can be found when previous methods have failed to do so; ii) it will increase the awareness amongst researchers of this relatively new method and encourage further work to be done in the area to improve the methodology.

7.3 Further Investigation into e-Commerce Effects Using EDMs

Further use of empirical dynamic models in the exploration of the effect of e-commerce on the environment could reveal more about the relationships between the variables already considered and, potentially, whether other omitted variables could have caused changes in the online retail share variable. For example economic changes such as the 2008 financial crash, or the 'relocation of production to other countries' as suggested by Dost and Maier (2017), could not only have had an impact on how shoppers are buying their products, but also on how online retailers use energy to minimise the impact these factors have on their businesses. As metioned in Section 2, the application of Taken's Embedding Theorem means that omitted (or unobserved) variables are implicitly taken into account. However, by explicitly accounting for factors such as economic markers or global shifts in production locations, inference made about the causal relationships between variables may be made clearer.

Additonally, energy consumption is used as the measure for the impact on the environment. Though, as explained by Dost and Maier (2017), this is because it is 'physically more fundamental than greenhouse gas emissions and less likely to differ structurally over time'. Though the differing dynamics of the system over time may be an issue in using empirical dynamic models (as described in Section 7.1), if this can be overcome and suitable data can be obtained, examinng the existence of a causal relationship between online retail share and greenhouse gas emissions may be of interest. Particularly in terms of reporting the effects of e-commerce to members of the public who, according to Clark (2012), when it comes to understanding environmental impact tend to think of carbon footprints and greenhouse gases as relevant measures.

8 Concluding Remarks

The use of empirical dynamic models as a means of testing for causality is a relatively new area of research and, as such, there are still areas which need to be developed further. Though not applicable to all time series data, they have been shown to outperform alternative models, particularly when there is nonlinearity present. The limitations of such models have been discussed, and some potential

starting points for further research moving forward have been explored.

References

[1] Chang, C. W., Ushio, M., & Hsieh, C. hao. (2017). Empirical dynamic modeling for beginners. Ecological Research, 32(6), 785796. https://doi.org/10.1007/s11284-017-1469-9

[2] Clark, D. (2012) What's a carbon footprint and how is it worked out? The Guardian. 4 April. https://www.theguardian.com/environment/2012/apr/04/carbon-footprint-calculated. Accessed 30.03.2018.

[3] Dennis B, Desharnais RA, Cushing JM, Henson SM, Costantino RF (2001) Estimating chaos and complex dynamics in an insect population. Ecol Monogr 71(2):277303

[4] Deyle, E. R., & Sugihara, G. (2011). Generalized theorems for nonlinear state space reconstruction. PLoS ONE, 6(3). https://doi.org/10.1371/journal.pone.0018295

[5] Deyle, E. R., R. M. May, S. B. Munch, & G. Sugihara. (2016). *Tracking and forecasting ecosystem interactions in real time.* Proceedings of the Royal Society B 283(1822): 20152258.

[6] Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S., & Sugihara, G. (2016). Global environmental drivers of influenza. Proceedings of the National Academy of Sciences, 113(46), 1308113086. https://doi.org/10.1073/pnas.1607747113

[7] Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S., & Sugihara, G. (2016). Global environmental drivers of influenza. Proceedings of the National Academy of Sciences, 113(46), 1308113086. SI : Step-by-step EDM Analysis. http://www.pnas.org/content/suppl/2016/10/26/1607747113.DCSupplemental/pnas Accessed 22.03.2018.

[8] Didelez, V. (2009). Statistical causality. in: W. Ostreng (ed), Consilience. Interdisciplinary Communications, Centre for Advanced Study, Oslo, 2007

[9] Dost, F. Is e-commerce good or bad for the environment? Complex deterministic system challenges and how to model them. STOR601 Research Talk Presentation January 2018

[10] Dost, F., & Maier, E. (2017). E-Commerce Effects on Energy Consumption: A Multi-Year Ecosystem-Level Assessment. Journal of Industrial Ecology, 0(0), 114. https://doi.org/10.1111/jiec.12639

[11] Dost, F., & Maier, E. (2017). E-Commerce Effects on Energy Consumption: A Multi-Year Ecosystem-Level Assessment. Journal of Industrial Ecology, 0(0), 114. https://doi.org/10.1111/jiec.12639 Supplementary Information

[12] Doward, J. (2017) Organic food sales soar as shoppers put quality before price. The Guardian. 19 February. https://www.theguardian.com/environment/2017/feb/19/sales-of-organic-food-soar-fruit-vegetables-supermarkets. Accessed 02.03.2018.

[13] Eichler, M. (2012). Causal inference in time series analysis. Causality: Statistical Perspectives and Applications., 327354. https://doi.org/10.1002/9781119945710.ch22

[14] Feldman, D. (2017) Netflix Is On Track To Exceed $11B In Revenue This Year. Forbes. 16 October. https://www.forbes.com/sites/danafeldman/2017/10/16/netflix-is-on-track-to-exceed-11b-in-revenue-this-year. Accessed 02.03.2018.

[15] Granger, C. W. J. (1980). Testing for causality. A personal viewpoint. Journal of Economic Dynamics and Control, 2(C), 329352. https://doi.org/10.1016/0165-1889(80)90069-X

[16] Insall, Matt; Rowland, Todd; and Weisstein, Eric W. "Embedding." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/Embedding.html. Accessed 26.02.2018.

[17] Hu, L.-Y., Huang, M.-W., Ke, S.-W., & Tsai, C.-F. (2016). The distance function effect on k-

nearest neighbor classification for medical datasets. SpringerPlus, 5(1), 1304. https://doi.org/10.1186/s40064-016-2941-7

[18] Ma, H., K. Aihara, and L. Chen. (2014). Detecting causality from nonlinear dynamics with short-term time series. Scientific Reports 4: 7464.

[19] Perretti, C. T., S. B. Munch, and G. Sugihara. (2013). Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. Proceedings of the National Academy of Sciences 110(13): 52535257.

[20] Pleven, L. 2016. Whats more eco-friendly: Going to the mall or shopping online. The Wall Street Journal. 22 March. www.wsj.com/articles/whats-more-eco-friendly-going-to-the-mall-or-shopping-online-1458667153. Accessed 02.03.2018.

[21] Renze, John; Rowland, Todd; and Weisstein, Eric W. "Compact Manifold." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/CompactManifold.html. Accessed 26.02.2018.

[22] Rowland, Todd. "Manifold." From MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein. http://mathworld.wolfram.com/Manifold.html. Accessed 26.02.2018.

[23] Sugihara, G., & May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature. https://doi.org/10.1038/344734a0

[24] Sugihara, G. (1994). Nonlinear forecasting for the classification of natural time series, 477495.

[25] Sugihara, G., R. M. May, H. Ye, C. Hsieh, E. Deyle, M. Fogarty, and S. Munch. (2012). Detecting causality in complex ecosystems. Science 338(6106): 496500.

[26] Takens, F. (1981). Detecting strange attractors in turbulence. Dynamical Systems and Turbulence, (Springer Lecture Notes in Mathematics, vol. 898) edited by D.A. Rand and L.S. Young, Springer-Verlag, Berlin.

[27] Tsonis, A. A., E. R. Deyle, R. M. May, G. Sugihara, K. Swanson, J. D. Verbeten, and G. Wang. (2015). Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. Proceedings of the National Academy of Sciences 112(11): 32533256.

[28] Weisstein, Eric W. "Compact Space." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram
Accessed 26.02.2018.

[29] Weisstein, Eric W. "Diffeomorphism." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram
Accessed 26.02.2018.

[30] Ye, H., Beamish, R. J., Glaser, S. M., Grant, S. C. H., Hsieh, C., Richards, L. J., Sugihara, G. (2015). Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proceedings of the National Academy of Sciences, 112(13), E1569E1576. https://doi.org/10.1073/pnas.1417063112

[31] Ye, H., Clark, A., Deyle, E., Sugihara, G. (2016) rEDM: an R package for empirical dynamic modeling and convergent cross-mapping

Appendix

```
install.packages("rEDM")
library(rEDM)
data("two_species_model")

#Fig 2(a)
x <- two_species_model$x[1:100]
y <- 1:100
df <- data.frame(x=x,y=y)
ggplot(data=df, aes(x=y, y=x, group=1)) + geom_path() + labs(x = "Time") +
  labs(y = "Abundance of Species X") + theme(text = element_text(size=20))
#Fig 2(b)
X <- two_species_model$x
x <- X[1:99]
y <- X[2:100]
df <- data.frame(x=y,y=x)
ggplot(data=df, aes(x=x, y=y, group=1)) + geom_path() + labs(x = expression(X[t])) +
  labs(y = expression(X[t-1])) + theme(text = element_text(size=20))
#Fig 2(c)
mod <- simplex(X, lib = c(1, 500), pred = c(501, 1000),E=1:20)
x <- mod$E
y <- mod$rho
df2 <- data.frame(x=x,y=y)
ggplot(df2,aes(x=x, y=y)) + geom_path() + labs(x = "Embeddng Dimension") +
  labs(y = "Correlation") + theme(text = element_text(size=20))

#Table 1
ar(X)
dat1 <- X[1:19]
mod1 <- ar(dat1)
predict(mod1)
dat1a <- X[1:20]
mod1e <- simplex(dat1a, lib = c(1, 10), pred = c(11,20),E=1:10)
mod1e <- simplex(dat1a, lib = c(1, 10), pred = c(11,20),E=2, stats_only=FALSE)
mod1e$model_output
s_map(dat1a, E=2)
dat2 <- X[1:199]
mod2 <- ar(dat2)
predict(mod2)
dat2a <- X[1:200]
mod2e <- simplex(dat2a, lib = c(1, 100), pred = c(101,200),E=1:10)
mod2e <- simplex(dat2a, lib = c(1, 100), pred = c(101,200),E=3, stats_only=FALSE)
mod2e$model_output
s_map(dat2a, E=3)
dat3 <- X[1:499]
mod3 <- ar(dat3)
predict(mod3)
dat3a <- X[1:500]
mod3e <- simplex(dat3a, lib = c(1, 250), pred = c(251,500),E=1:10)
mod3e <- simplex(dat3a, lib = c(1, 250), pred = c(251,500),E=2, stats_only=FALSE)
```

```
mod3e$model_output
s_map(dat3a, E=2)
dat4 <- X[1:999]
mod4 <- ar(dat4)
predict(mod4)
dat4a <- X[1:1000]
mod4e <- simplex(dat4a, lib = c(1, 500), pred = c(501,1000),E=1:10)
mod4e <- simplex(dat4a, lib = c(1, 800), pred = c(801,1000),E=2, stats_only=FALSE)
mod4e$model_output
s_map(dat4a, E=2)
```