Causally Appropriate Graphical Modelling for Time Series with applications to Economics, Ecology and Environmental Science

A thesis submitted in partial fulfilment of the requirements for the Degree of Master of Science in Mathematics and Statistics in the University of Canterbury

By Carla Meurk

2005

Contents

1 Introduction

In 1967 Clive Granger proposed a definition to decide if a variable in a time series was causal or not. This definition has come to be known as Granger-Causality, or G-Causality for short. Over the years, there has been much debate over whether this procedure should be deemed "causal" at all (Granger, C., 1988). About 30 years later, Marco Reale and Granville Tunnicliffe-Wilson, developed a method - Graphical Modelling for Time Series (GMTS) for providing causal models of multivariate time series (Reale, M., 1998). Originally, this method was applied to economics, and here I extend its application to ecology and environmental science.

The aims of this thesis split roughly into two parts. The first aim is in relation to causality, where I ask what, if any, place does causality have in statistics? In particular, I will discuss the notion of G-Causality and whether I believe it can legitimately be deemed causal. Finally, I will discuss GMTS and whether it provides a causally appropriate modelling strategy.

The second aim relates to the practical application of graphical modelling in time series, in particular its application to ecological and environmental time series. I explore whether GMTS produces improved models of the data compared with a traditional approach.

Chapter 2 briefly outlines some common concepts used in time series analysis as well as introducing some graphical modelling terms that will be referred to subsequently. Chapter 3 aims to provide some background to the causal debate and where I stand *vis* the relationship that should exist between statisticians and the topic of causation. In addition, the GMTS modelling strategy to be adopted for subsequent analysis is explained along with some remarks on its suitability as a causally appropriate strategy.

Chapters 4, 5 and 6 provide three very different case studies from which to explore and assess the GMTS methodology. The first, is an example from economics, where this methodology was originally applied. This example contains a number of variables and hence posed a combinatorial challenge in terms of possible contemporaneous relationships. Through this case study however, not only was the combinatorial problem resolved, but ways of possibly extending the automation of this process discovered.

Chapter 5 contains an environmental example of hydrological processes where, due to automated recording of data, there are vast numbers of observations (in the order of 10^4) for a number of different variables. This chapter is interesting both because it provided evidence that the GMTS approach consistently generates better models than a traditional approach, but also provides some interesting insight into the causal problem GMTS faces. Finally, I adapted the newly developed SINful approach to graphical modelling, proposed by Drton and Perlman (2004), to time series to see if it improved on GMTS.

Chapter 6 takes us to the other extreme, an ecological dataset built up over more than 20 years, which consists of only 90 observations. This dataset strongly violates all the requirements a multivariate time series needs for sensible analysis and required some manipulation of the original GMTS method in order to create a functioning model. This case study indicated that while GMTS was developed for time series with a few adjustments, it may have relevance to other types of modelling.

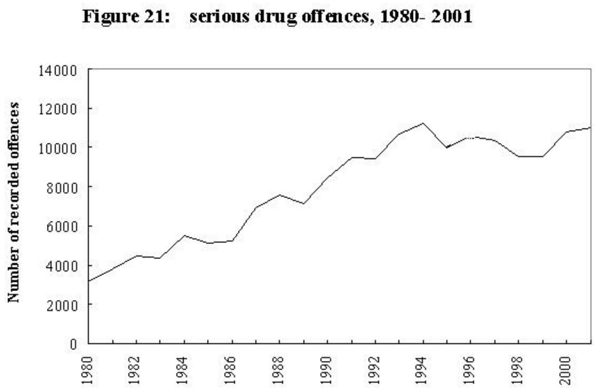


Figure 1: A time series showing the number of serious drug offences committed in New Zealand from 1981-2000.

2 Introduction to Time Series and Graphical Modelling

This thesis is concerned with the causally appropriate graphical modelling of multivariate time series - vector autoregressive (VAR(p)) models in particular. In this chapter some basic concepts are introduced. These are divided into three sections: time series modelling, graphical modelling and model assessment methods.

2.1 Time Series

A time series is simply a set of data collected over time. For example, figure 1 gives a time series showing the number of serious drug crimes committed in New Zealand from 1981-2000.

Time series can model either discrete or continuous processes. In addition, the sampling of the process can also be either discrete or continuous. Figure 2 gives examples of the different sorts of time series. The All Blacks win/loss record is an example of a time series with discrete sample space and discrete state space, a signal passing through a cable has continuous sample space and discrete state space. The time series of temperature, is an example of a process with continuous state space sampled discretely and an example of a series with continuous state and sample space is the plot of seismic activity.

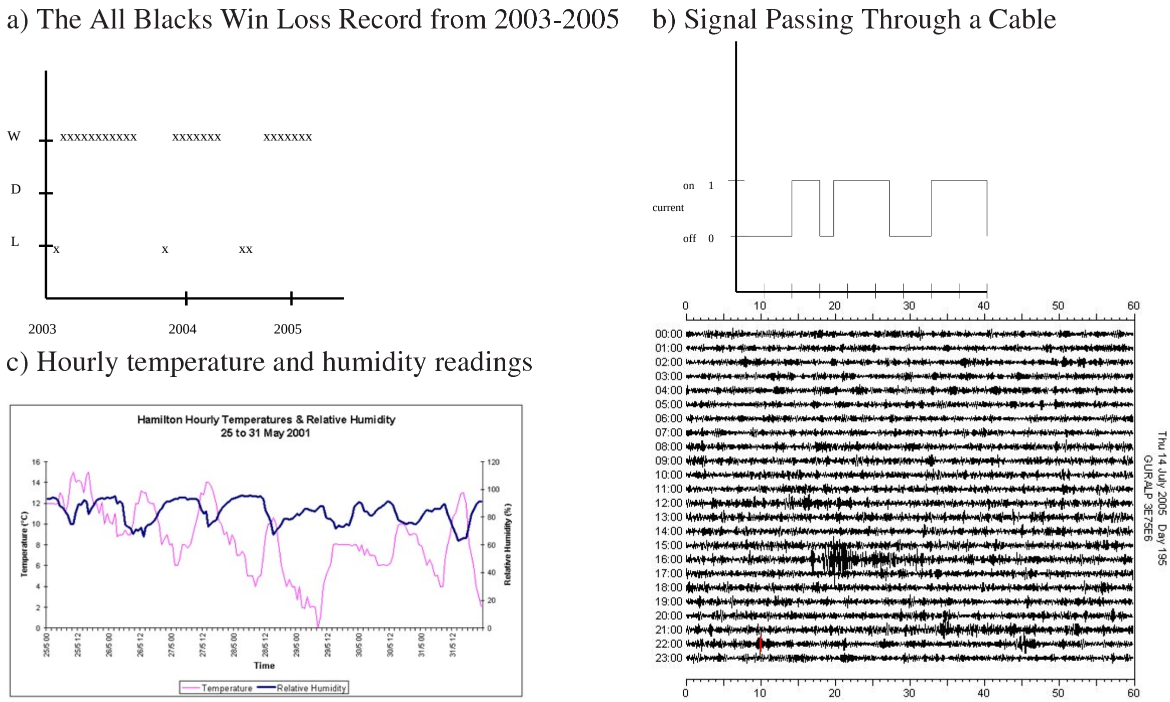


Figure 2: Examples of different types of time series

In this thesis we will focus on time series with continuous state-space and discrete sample space. Hence, all definitions provided will be the discrete formulation.

2.1.1 Stochastic Processes, Stationarity and Ergodicity

Time series are finite realisations of *stochastic processes*, that is, the data are a collection of random variables that are ordered in time. This means that each observed data point is only one *realisation* in a distribution, and hence, each time series (a collection of data points) is only one realisation of an infinite number of possible time series (Chatfield, 1989). As figure 3 shows, at each time point the observed value is drawn from a different distribution.

The task of modelling, given just one observation from each random variable might seem like an impossible task. But, by making a few key assumptions, a sophisticated discipline of time series analysis has developed. These key assumptions are stationarity and ergodicity.

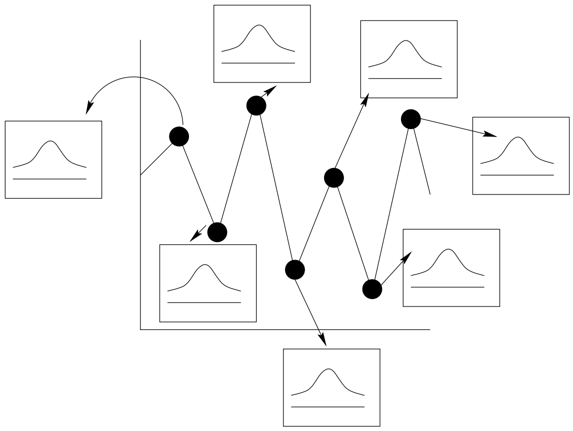


Figure 3: Diagram of time series thought of as a sequence of random variables each from a different distribution

Stationarity

Stationarity is defined with respect to the parameters of interest and in time series this is usually the mean and variance. If a series is stationary it implies that there is no change in the mean and variance over time (Chatfield, C., 1989, 10).

There are two definitions of stationarity: a strict definition, which is very thorough but difficult to assess and a weaker one which is commonly used.

A series is said to be strictly stationary if:

The joint distribution of $X_t, X_{t+1}, \ldots, X_{t+n}$ and $X_{t+\tau}, \ldots, X_{t+n+\tau}, \forall t, \tau, n$ are the same (1)

A series is said to be weakly stationary if:

$$E(X_t) = \mu \quad \forall t \quad \& \quad \text{Cov}(X_t, X_{t-\tau}) = \gamma(\tau) \quad \forall t, \tau. \tag{2}$$

(Chatfield, C., 1989, 28-29), where X_t refers to the observed variable of a series at time t. If a series satisfies the condition of weak stationarity we say that its first two moments, the mean, $E(X_t)$, and covariance, $\text{Cov}(X_t, X_{t-\tau})$, are *time invariant*.

Satisfying the requirement of stationarity allows the observed values to be considered as points in the **same** distribution and this means the series can be considered to be a "proper" sample to which we can apply standard statistical techniques.

In the development of time series theory, the condition of stationarity is very important, and, where time series do not satisfy this requirement, effort is often spent trying to create a stationary series from the data by methods such as differencing a series to create $\nabla X_t = X_t - X_{t-1}$, or removing any deterministic trends and modelling the residuals.

Having said this, in practise whether a series is considered to be stationary or not is not an exact science. There are formal tests for stationarity such as the Dickey-Fuller (Dickey, D *et al.*, 1979) and Phillips-Perron (Phillips, P *et al.*, 1988) tests. In particular, if one is interested in an $AR(p)$ model the augmented Dickey-Fuller test can be used. This is a hypothesis test where the null hypothesis is that the process is I(1) (Hamilton,1994,475-531). If the process is I(1) (integrated of order 1) the series will be stationary when differenced once. However, because time series data are only one realisation of an infinite number of series, it is sometimes argued that we may assume the series is stationary even if this is not what we observe, and hence tests such as this are not always helpful (Hamilton,1994, 444-447). In addition, some modelling strategies can in fact model I(1) processes, which most non-stationary processes are, and hence, the lack of stationarity is not necessarily a major concern.

Ergodicity

Ergodicity like stationarity is defined with respect to parameters of interest. The property of a series being ergodic means that the expectation of observed variables up to time $n = t$ (where t is finite) is representative of the expectation of the entire process, i.e. as $n \to \infty$ (Reale:1998, Chatfield:1989). For a discrete stochastic processes this is formally stated as:

If $\{x_n, n \geq 0\}$ is a strictly stationary stochastic process with $E\{|x_0|\} \leq \infty$ and the sequence $\{x_n\}$ mutually independent then:

$$\lim_{n \to \infty} \frac{x_0 + x_1 \ldots x_n}{n+1} = E\{x_0\} \tag{3}$$

(Doob, 1953, 464-466).

Equation 1.3 means that if a sequence of observations ordered in time from 0 to t, $\{x_0, x_1, \ldots, x_t\}$, is a strictly stationary stochastic process with finite mean, and, the observations are mutually independent, then, as the sequence of x_n extends to ∞, the expected value of $\{x_n\}$ remains the same.

2.1.2 A Univariate Example: AR(p) Models

Depending on the nature of the data and the objectives of the research there are many different types of time series models. One of the simplest is the generic autoregressive model of order p, or AR(p). An AR(p) model predicts the current value X_t as dependent on past values of the series X, i.e. $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$ up to a specified order p, exploiting the serial correlation often present in time series data.

An AR(p) is defined as:

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + Z_t \quad \text{where}\{Z_t\} \sim (0, \sigma^2). \tag{4}$$

Here, α refers to the coefficients of the dependent variables and $\{Z_t\} \sim (0, \sigma^2)$ refers to the sequence of error terms with mean 0 variance σ^2, with Cov(Z_t,Z_{t-k})=0, $\forall k \neq 0$.

To model a process as an AR(p) model, the appropriate α_i values and lag p must be found.

2.1.3 Finding p

Finding the order of an AR model is done by using the sample partial autocorrelation function (PACF) denoted $\pi(\cdot)$. The order p is found as the point at which the condition $\pi_\tau = 0$ for $\tau \geq p$ is satisfied.

The partial autocorrelation function π_τ is defined as:

$$\begin{aligned}
\pi_\tau &= \text{Corr}(X_t, X_{t-\tau}|X_{t-1}, \ldots, X_{t-\tau+1}) \\
&= \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{m-1} \\ \gamma_1 & \gamma_2 & \cdots & \gamma_{m-2} \\ \vdots & \ldots & & \vdots \\ \gamma_{m-1} & \gamma_{m-2} & \cdots & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{bmatrix}
\end{aligned} \tag{5}$$

$$\begin{aligned}
\text{Where, } \gamma_\tau &= \text{Cov}\left[X_t, X_{t+\tau}\right] \\
&= E\{[X_t - E(X_t)][X_{t+\tau} - E(X_{t+\tau}))]\}
\end{aligned}$$

and, $\gamma(\cdot)$ is known as the autocovariance function.

(Hamilton,J.,1994,111).

Example

To illustrate this, consider the dataset `lh`, provided by `R` version 2.1.0 (`R` Development Core Team, 2005). This univariate time series is data collected at 10 minute intervals, for levels of luteinizing hormones in a human female. For this dataset the PACF can be calculated and plotted as in figure 4.

The dotted lines are the boundaries of the confidence interval around zero provided by the Ljung-Box statistic.

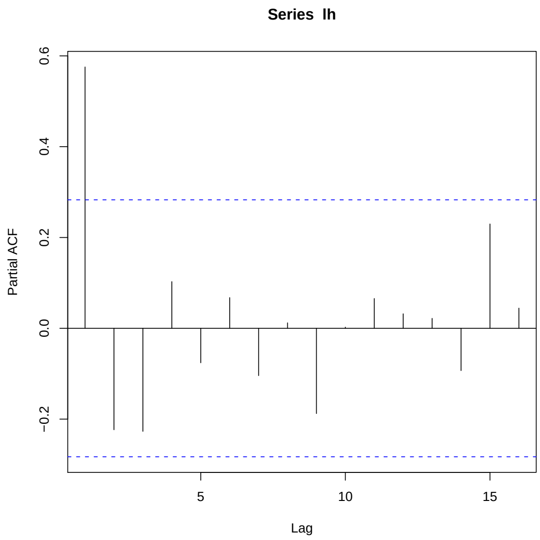


Figure 4: Plot of PACF up to lag 15.

From inspection of figure 4 there appears to be only one significant value, and hence this equation is defined as an AR(1) process.

In some cases however, p is very large and this is where other models such as moving average (MA) and autoregressive moving average (ARMA) models may be appropriate. MA models represent $\{X_t\}$ in terms of the lagged errors in $\{Z_t\}$, while an ARMA models a process in terms of both the lagged values in $\{X_t\}$ and $\{Z_t\}$. MA and ARMA models are not discussed further as the focus of this thesis is on multivariate AR models.

2.1.4 Finding α: Likelihood Estimation

Continuing with the luteinizing hormone example, on deciding that this is an AR(1) model, it will have the form: $X_t = \alpha X_{t-1} + Z_t$. Now we need to find α. Maximum likelihood estimation can be used to find the value of α_1 for AR models.

Maximum likelihood methods (MLMs) take a model g which describes a probability distribution of the data, with model parameters θ and a model form $model$. We can write this $g(x|\theta, model)$, where x is the data. The likelihood function is then denoted as $\mathcal{L}(\theta|data, model)$ - or in our case $\mathcal{L}(\theta|x, g)$. With both x and g known, the likelihood function calculates the likelihood of θ being a certain value. The assumption is made that the best estimate is the most likely estimate, and hence θ is sought which maximises $\mathcal{L}$ (Burnham, K. *et al.*, 2002, 7-8). This function is solved numerically using any one of a variety of optimization methods (for a selection of these see Hamilton, 1994, 133-142).

One of the simplest ways of estimating is the method of ordinary least squares, and this can be thought of as a special case of a MLM (Burnham, K *et al.*, 1998, 9).

An ordinary least squares fit takes an input matrix A, in this case $A = [X_{t-1}]$, where X_{t-1} is a vector of length n, a vector x of unknown regression coefficients (in our case α_i), and, the resulting series b (X_t). This is a system $Ax = b$, where x is unknown. To find x, $A^{-1}b$ must be evaluated.

Example

In our case the model is an AR(1) model. Given the data `lh`, the system $Ax = b$ is:

$$\begin{bmatrix} 2.4 \\ \vdots \\ 3.4 \\ 3.0 \end{bmatrix} \alpha_1 = \begin{bmatrix} 2.4 \\ \vdots \\ 3.0 \\ 2.9 \end{bmatrix} \tag{6}$$

Solving $A^{-1}b = 0.98364$. Hence, in this example the resulting model is
$X_t = 0.98364X_{t-1} + Z_t$. Once this process is modelled, residuals should be checked for unexplained variation and any anomalous trends. Errors are shown in figure 5. Each plot gives different information. The residuals vs fitted plot is used to check for unexplained variance and any trends in variance. In this example, other than the three labelled outliers there are no obvious trends. The QQplot is a quantile-quantile plot, used to check for normality, and should display a straight line except at each end where it curves slightly to

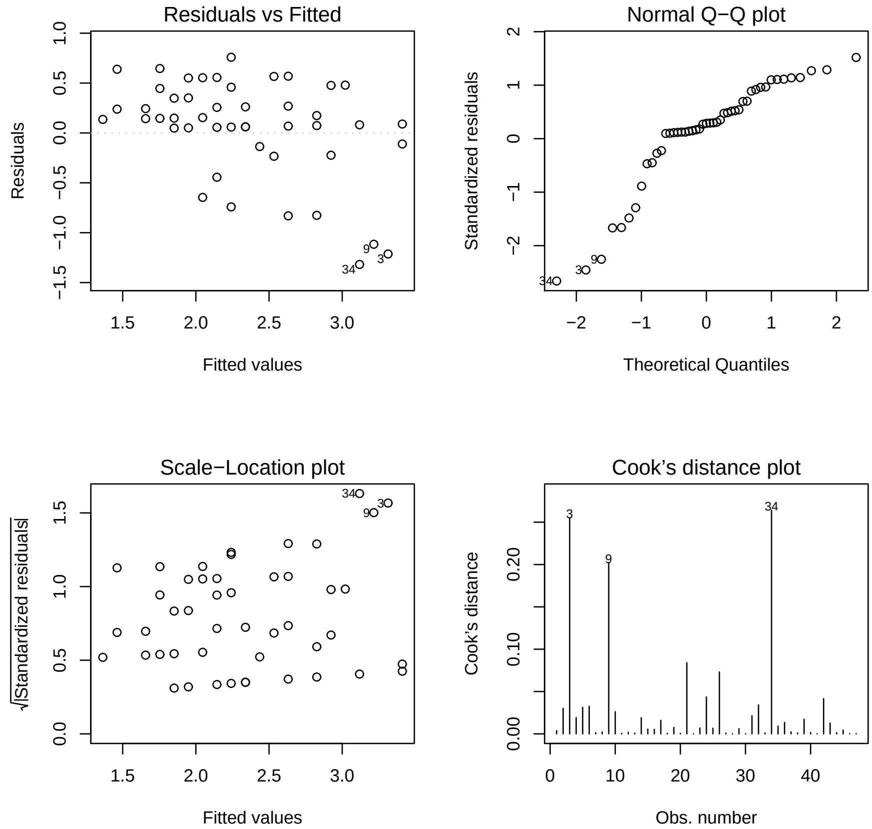


Figure 5: Plot of errors after fitting of AR(1) process to dataset `lh`

form a slight S shape. In this case this plot shows some lack of normality. The studentized residuals provide a different view of the data which can sometimes give a better idea of outliers. The Cook's distance plot measures the influence each variable has on the model, in this case observations 3,9 and 34 had a very high influence. Depending on the objective, one may try to remodel this process without these points, or, look at these points in more detail given that they are so different.

2.1.5 The Multivariate Case: Vector Autoregressive Models

The extension of univariate time series to multivariate time series occurs when we wish to analyse multiple time series where the individual time series may be related in some way. For example, consider the three time series (where observations are recorded simultaneously) X_t, Y_t, Z_t are series on mortality rates, smoking rates and cancer rates respectively.

These series could each be modelled as an AR process, but given their possible interrelatedness it may be more meaningful to model these series together. This is where the AR models extend to VAR models.

There are various different types of VAR models. The formulation of a canonical VAR(p) (cVAR) looks much the same as an AR(p) except with the variables represented as vectors and matrices rather than single values.
A cVAR is defined as:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} ... \alpha_p X_{t-p} + Z_t \tag{7}$$

where, $X_t = \begin{bmatrix} X_{1,t} \\ \vdots \\ X_{n,t} \end{bmatrix}$, $\alpha_m = \begin{bmatrix} \alpha_{1,1,q} & \dots & \alpha_{1,n,q} \\ \vdots & \vdots & \vdots \\ \alpha_{n,1,q} & \dots & \alpha_{n,n,q} \end{bmatrix}$, $1 \leq q \leq p$, $Z_t = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$,
$Z \sim$ i.i.d., where we are modelling n time series.

The other type of VAR is a structural VAR (sVAR) which has the form:

$$\alpha_0^* X_t = \alpha_1^* X_{t-1} + \alpha_2^* X_{t-2} ... \alpha_p^* X_{t-p} + Z_t, \quad X, Z, \alpha^* \quad \text{defined as above.} \tag{8}$$

The difference between the two is that the vector X_t has a matrix of coefficients in the case of the sVAR. The structural VAR is more general and can provide a better understanding of the system under study. In addition, there are two types of sVAR models, simultaneous equation models and recursive, or causal, models. The difference relates to the matrix of coefficients α_0^*. In a simultaneous equation model, the matrix of contemporaneous coefficients cannot be reduced to a matrix in upper triangular form, whereas in a recursive model it can. Simultaneous equation models are tricky to deal with and for the duration of this thesis, when I refer to an sVAR model I will be referring to a recursive sVAR model.

2.1.6 Modelling VAR Processes

A cVAR model, having the same form as a univariate AR model can be parametrised in a similar way as the univariate AR. The sVAR however, is more complex because the coefficient matrix α_0^* prevents solving the equation 8 by the method described in section 2.1.4. In addition, although for a given sVAR model a unique cVAR can be found, the converse is not true and, hence, trying to transform an sVAR model, which cannot be solved, into a cVAR model, which can, is not an option.

For a given sVAR there is a unique cVAR:

Given a recursive sVAR as in equation 8:

α_0^{*-1}exists and is unique. Multiplying both sides of equation 8 gives

$$X_t = \alpha_0^{*-1}\alpha_1^* X_{t-1} + \alpha_0^{*-1}\alpha_2^* X_{t-2} \ldots \alpha_0^{*-1}\alpha_p^* X_{t-p} + \alpha_0^{*-1} Z_t \quad (9)$$

this is a cVAR model with$V$is the variance-covariance matrix of residuals □

For a given cVAR there exists more than one sVAR:

Set $\alpha_0^{*-1}\alpha_k^* X_{t-k} = \beta_{t-k} X_{t-k},$ then equation 9 can be written as:

$X_t = \beta_1 X_t + \beta_2 X_{t-1} + ... + \beta_{t-k} X_{t-k} + E_t$

Where, $E_t = \alpha_0^{*-1} Z_t \quad V = D$ can be written as $\alpha^{*-1} V \alpha^* \neq D$

Multiplying through by α^* the cVAR in equation 9 can be recovered (10)

however $\alpha^*$is not unique, as multiple factorisations exist.

hence for each cVAR there will be more than one sVAR□.

(Reale, M., 1998, 42). One method of solving sVAR models is by orthogonalization of residuals. This method models a cVAR, takes the residual matrix E_t, drawing on the results given above, and uses a method such as Choleski decomposition to express the

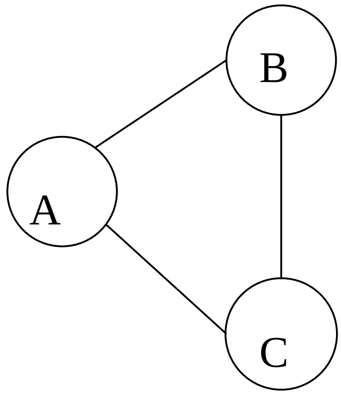


Figure 6: An undirected graph (UG)

residuals $V = ADA^T$. The value of A is then an estimate of the matrix of contemporaneous coefficients. However, this method is not unique as there may be several alternative orthogonalizations, which would result in different values for A. Despite these difficulties, there are other advantages in modelling data as an sVAR model. This will be discussed further in chapter 2.

2.2 Graphical Models

2.2.1 Language

Graphical models are diagrams which contain nodes and edges. Figure 6 is an example of a graphical model and is an undirected graph (UG). In this model the circles containing the letters A,B and C are referred to as *nodes*, and in our case are possible variables in a model. The lines which join each node are referred to as *edges*.

If a direction is added to each edge such that a cycle is not created, the resulting graph is a directed acyclic graph or DAG, as in figure 7. In a DAG a node with an incoming arrow is referred to as a *child*, the *parent* is the node with the associated outgoing arrow. For example, in figure 7 B is a parent of both A and C. In our case a DAG specifies a system of equations of one variable in terms of others. This DAG corresponds to the

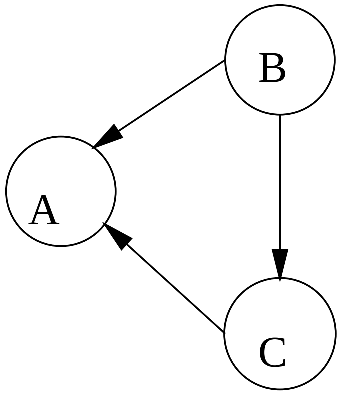


Figure 7: A directed acyclic graph (DAG)

system of equations 11.

$$
\begin{aligned}
A &= \alpha B + \beta C \\
C &= \gamma B
\end{aligned}
\tag{11}
$$

2.2.2 Conditional Independence Graphs

If we wish to use graphical models in this way, then the important question is, given a set of variables (nodes) what criteria must be satisfied for an edge to be drawn between two nodes. One possibility is to use the criteria of conditional independence.

Independence is defined as:

$$P(A|B) = P(A) \tag{12}$$

If this condition is satisfied then we say A is independent of B or $A \perp\!\!\!\perp B$, and interpret it by saying that the occurrence of event B does not affect the probability of event A.

This principal of independence can be extended to an idea of conditional independence:

$$P(A|B, C) = P(A|C) \tag{13}$$

If this condition is satisfied then we may say that A is *conditionally* independent of B, or $A \perp\!\!\!\perp B|C$. The interpretation is that given C, the occurrence of event B does not affect the probability of event A.

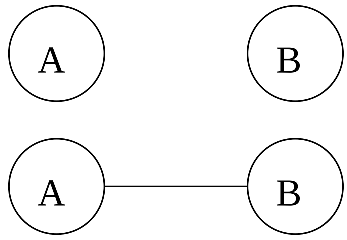


Figure 8: Top: A and B are conditionally independent. Bottom: A and B are conditionally dependent.

Graphically, if two variables are conditionally independent, then there is no edge between two nodes (figure 8, top). If two variables are conditionally dependent then there is an edge drawn between each node (figure 8, bottom). A conditional independence graph (CIG) is a graphical model which represents these relationships between all variables under study.

Another term that will be used is that of subgraphs. Subgraphs are parts of graphs, for example figure 8 (bottom) is a subgraph of the graph in figure 6.

2.2.3 Turning CIGs into DAGs

To add direction to a CIG there are two cases to consider. Where there is a clear "arrow of time", the direction can be added automatically. In other cases subject matter knowledge or methods of exhaustion need to be employed.

There is one exception to this. Referring to figures 6 and 7, suppose events B and C occur at the same time (contemporaneously) and both occur before A, adding the arrows directed from B to A and C to A is straightforward. Further, suppose we have *a priori* knowledge that event C could not cause event B. It would be natural to assume we have created a DAG equivalent to the following CIG but this is not quite correct. The model is also equivalent to the CIG in figure 9. This brings us to the principal of moralisation.

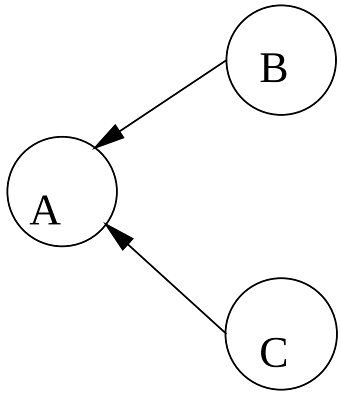


Figure 9: Another DAG equivalent to CIG in figure 6

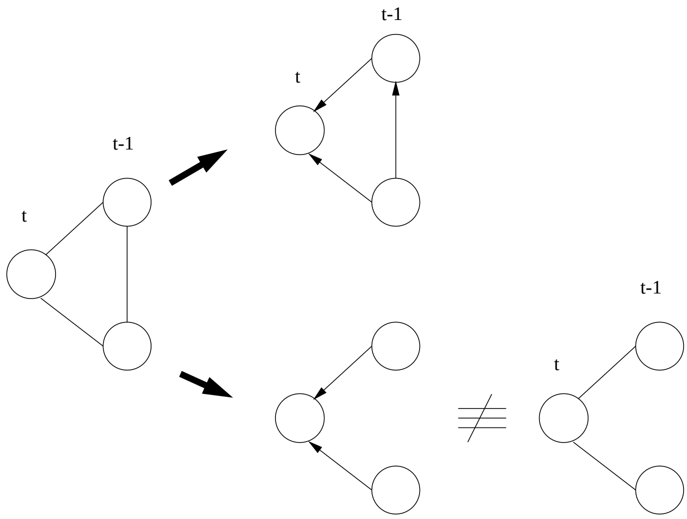


Figure 10: A diagram which shows the principal of moralisation when transforming a CIG into a DAG. The DAG at the top is one example of a possible DAG permutation. The DAG at the bottom with a missing link between parents, is also equivalent. A corollary of this is that the DAG shown at the bottom is not equivalent the CIG formed simply by removing arrows.

2.2.4 Moralisation

Theoretically, in transforming a CIG into a DAG it is important that the dependence relations are unchanged. In most cases a CIG and its corresponding DAG i.e. the CIG with arrows added, can be shown to be equivalent. However, in cases of subgraphs such as in figure 10, there are multiple DAGs which correspond to one CIG.

The language used in this process is defined as follows: When moving from a CIG with a subgraph of the type in figure 6 a graph is *demoralised* by identifying the links which join two parents without underlying conditional dependence. When moving from a DAG to a CIG, a graph is *moralised* by adding a link between two parents of the same node. For example, in figure 9 both B and C are parents of A hence, to recover the CIG we add an edge between the parents B and C.

2.3 Model Assessment, AIC, HIC and SIC

The final concept to be introduced is that of model comparison. Here, I present the Akaike, Hannan-Quinn and Schwarz information criterion (AIC, HIC and SIC respectively). There are various ways of assessing the "goodness of fit" of a model. When one wishes to comparatively assess models constructed from the same dataset the AIC and related criteria are good methods to use, (Burnham *et al.*, 1998) the formal definition of the AIC is:

$$AIC = -2\log(\mathcal{L}(\hat{\theta}|y)) + 2K, \tag{14}$$

(Akaike, H., 1973) where, K denotes the number of estimable parameters in the model and $\mathcal{L}(\hat{\theta}|y)$ is the likelihood function for a model y with estimated parameters θ which maximise $\mathcal{L}$.

The AIC value is an estimate of "the expected, relative distance between the fitted model and the unknown true mechanism that generated the observed data". (Burnham, K. *et al.*, 2002,61).

In the case of least squares regression the AIC value is given as:

$$AIC = n\log(\hat{\sigma^2}) + 2K, \tag{15}$$

Where K is the number of regression parameters fitted. $\hat{\sigma^2}$ is estimated as:

$$\hat{\sigma^2} = \frac{\sum \epsilon_i^2}{n}. \tag{16}$$

Here, ϵ_i^2 are the residual errors from the least squares fit and n refers to the sample size.

The AIC criteria can be thought of as picking the model which displays optimal trade-off between fit and parsimony. A good model will have small errors i.e. $\hat{\sigma^2}$ will be close to 0, taking $\log(\hat{\sigma^2})$ means that this value will be even smaller, and, if $\hat{\sigma^2} < 1$, negative. AIC then penalises for the number of parameters a model has, by adding 2 for each parameter included, the "best" model will be the model with the lowest AIC score.

This criteria does fail sometimes. If there are a large number of possible model parameters, the AIC is a less reliable estimator i.e. it has a tendency to overfit. In this case other information criteria can be used. The two other information criteria I will refer to are the HIC (Hannan,*et al.*,1979) and SIC (Schwarz,G.,1978).

In the case of least squares regression the HIC value is given as:

$$HIC = n\log(\hat{\sigma^2}) + 2\log(\log(n))K \tag{17}$$

For sample sizes larger than 15, $(\log(\log(n))$ HIC method penalises the addition of parameters more harshly than the AIC.

In the case of least squares regression the SIC value is given as:

$$SIC = n\log(\hat{\sigma^2}) + \log(n)K \tag{18}$$

This penalises more severely than AIC for sample sizes over about 8, and more than the HIC criteria.

These criteria are very helpful however, it is important to remember that these criteria can only give the best comparative model and it is the statistician's responsibility to ensure that the set of models being tested are a set of suitable candidate models (Burnham, *et al*,1998,72).

3 Causality as a Statistical Concern

In the previous chapter, I developed a framework of basic information that will be utilised in the following study of methodology for a causally appropriate graphical modelling approach to time series. However, there are two more concepts to discuss -the concept of causality and the GMTS approach.

3.1 Motivation

Statisticians, whether looking to describe or make some inference from data are often concerned with looking for "*relationships*" between variables. This term is often considered synonymous with terms like association, correlation and even causation. However, causation between two variables seems to be a much stronger sort of relation than simple correlation. Unfortunately, whereas correlation is a well defined and understood concept, definitions of causality will rarely be the same between two people. It is perhaps this inability to tie-down a concrete definition of what causation is, especially compared with concepts of correlation, that leads many statisticians to believe that causality should not, or could not be, the concern of the statistician but rather a concern for the subject matter specialist. For example, it is not a job of the statistician to make causal statements about ecological systems, it is the job of the ecologist. Alternatively, those that do attempt to provide a causal framework often make the the mistake of needing to "define" what causation is. I argue that, while not all statisticians need to be concerned with causality, those that do, need to be concerned with finding causal links not defining what it is, the latter being best left to the philosophers.

3.1.1 Does Every Statistician Need Causation?

A causal explanation is not always needed or desirable, and depends largely on the type of data being dealt with and the sort of answers required. For example, suppose one wanted to know how often a particular river flooded in a certain area. In this case it would not matter what variables were used, or how many, so long as they improved the accuracy of the predictions - a causal explanation is not required. Alternatively, suppose one wants to know where the best place to spend money on stopbanks is, to prevent future floods in an area, in this case a model of the system which provided information on the causal processes would be preferable, and that is what this work attempts to explain. This is a slightly simplistic example to illustrate a point. There will be some duality between prediction and causal efficacy, i.e. a good predictive model should say something about causal mechanisms. However, in practise some compromise is usually required and a decision to prefer one type of explanation over the other needs to be made.

3.2 The Mistake of Defining Causation

There are a number of authors who begin their search for causation in statistics by defining what causation is. This is an understandable thing to do, it certainly *seems* that by defining terms carefully, one is being very scientific and precise in there work. However, there must be a distinction made between metaphysics, philosophy of statistics and statistical philosophy, and, by defining a concept of causality and then basing a statistical framework around this I believe these important distinctions are lost.

3.2.1 Different Job Descriptions for Different Jobs

In our discussion of statistics and causality we are clearly bridging a gap between philosophy and statistics, however, in developing causally appropriate statistical models, we are developing statistical methodology, this is statistical philosophy. There is a related (and very interesting) question, a question for a philosopher of statistics that is, can data provide causal knowledge and can statistical methods justify these causal links? This is an *epistemic* question because it is a question regarding our abilities to gain knowledge, in this case, causal knowledge through statistical method. Related to this is a *metaphysical* question for a philosopher, what is causation?

While these are all interesting questions, in what follows, I hope to argue that a lot of previous work on causality and statistics blurs these distinctions (especially the second and third), and that I consider this to be in error.

3.2.2 Metaphysics and Counterfactuals

First a (brief) definition of some philosophical terms.

Metaphysics

Metaphysics is the study of things *as they are*. For example, a physicist studies certain types of substances - particles, atoms, neutrinos etc., to discover physical laws, etc. A person studying metaphysics asks what sort of thing a substance is. Another example is a psychologist who searches for information about how our minds function, a person studying metaphysics asks what is the mind? In other words, where a scientist or researcher accepts the existence of the things their subject deals with, and works to learn how these things function, a philosopher asks what the nature of these "things" are.

Counterfactual

A counterfactual statement is a statement about something that has not happened and cannot happen. For example, if we take the statement, "A happened and then B happened". Then the counterfactual statement is, "If A did not happen then B did (would or could) not happen". If, in our world event A happened, by making a counterfactual statement we are imagining a *possible world*, i.e. an imaginary place where A did not happen and the imaginary consequence of this.

3.2.3 Counterfactual Accounts of Causation

In the early 18th century the famous philosopher David Hume put forward a number of arguments that are still influential in philosophical thought today. One such argument was his counterfactual definition of causality. This account of causation is surprisingly widespread in statistics. Basically, this view states that if A is a cause of B then if A did not happen then B would not have happened.

Judea Pearl

A leading author on the subject of causality in statistics, Judea Pearl, strongly advances the cause of counterfactual knowledge.

> "...that our scientific, legal, and ordinary languages are loaded with counterfactual utterances indicated clearly that counterfactuals are far from being metaphysical; they must have definite testable implications and must carry valuable substantive information. The analysis of counterfactuals therefore represents an opportunity to anyone ... [to integrate] substantive knowledge with statistical data so as to refine the former and interpret the latter." (Pearl, J., 2000, 34)

Rubin's Causal Model

The idea of causation, commonly referred to as Rubin's Model (Edwards, D., 2000, 225-234) is based on the possibility of counterfactual reasoning. In this model causal effect $C(i)$ is defined as $C(i) = Y(i, a) - Y(i, c)$ where $Y(i, a)$ is the response a unit (i) has to a treatment a and $Y(i, c)$ is the response a unit (i) has to a treatment (or control) c. However, because each unit can only be treated with one or other treatment we need a way to infer about the other, i.e. develop a framework to make an accurate counterfactual claim.

In making a counterfactual claim of causality, from which to build a statistical framework these models make a *metaphysical* assertion, an assertion about what causation is and this assertion is problematic for two reasons.

(1) While some philosophers accept this account of causality, it is by no means universally accepted. One search of the journal database, Philosopher's Index (Webspirs,2005),

with search criteria `counterfactual` generated 829 results equally split both for and against, with publications from 1940 until today. Given that this idea has been around since the 18th century these works would constitute a small subset of the total literature on this subject. Whether causation can be defined counterfactually is a highly contentious topic. Hence, if a statistician wishes to advance a counterfactual account, they need to acknowledge that anyone who does not share this strong attachment will find a causal description of this sort entirely meaningless.

Statistics, along with other scientific disciplines, needs to focus on developing methodology to *find* causal links. This methodology must be robust enough that it is buffered from the developments and changes philosophers make in defining causation because these differ wildly. If this is not done, it is hard to see how statistical methodology for causality will be accepted as a viable analytical tool, because it will always be necessary for the user to have strong metaphysical attachments to a certain philosophical idea, or in fact know what a strong metaphysical attachment is.

(2) On a more practical note - a statistician looks at data. How can data give counterfactual information? How can data give information about something that cannot, by definition, happen? It is impossible. This is quite different from the use of methodology for filling in missing data points. Hence, in subscribing to a counterfactual view, the statistician has to provide a scheme to make decisions about what is not happening and what might not have happened as a result. This introduces the possibility of bias in making causal assertions. Consequently, any objectivity that a concrete definition of causality might have provided is lost.

Overall, statisticians are best to avoid accounts of causality which require strong metaphysical assertions. In addition, even if a counterfactual account of causation is true, the problem of causation shifts to become a problem of finding counterfactual information from data - a problem which is impossible to overcome statistically.

In making this assertion the obvious question to ask is, how cause can be discovered if a definition of causation is not known? The short answer is that everyone knows what causation is. Rain *causes* grass to grow, procrastination *causes* the deadline to be missed and so on, an absolute definition is not required. This may seem very subjective, but there are many philosophical theories which support a "pragmatic" approach. These tend to be (and are not limited to) accounts of knowledge which appeal to human psychology and the functioning of mind as well as deflationist accounts which aim to justify beliefs without the requirement of infinite chains of reasoning. These types of issues are discussed at length in two books written by Alvin Plantinga (1993).

3.2.4 Granger Causality

From the 1960's Granger has been interested in the notion of causation for time series analysis and the notion of Granger-Causality (or G-Causality) has enjoyed wide popularity as a method of analysis, not only in econometrics but in a wide variety of applications. Granger displays a sound understanding of where a statistician's place is *vis* causation. He has tried to create a scheme for finding causal links, without needing to define what causation is.

The concept of causality Granger proposed (Granger, C.,1988, 199-200) is as follows: Given $J_t : x_{t-j}, y_{t-j}, w_{t-j}, j \geq 0$ and $J'_t : x_{t-j}, w_{t-j}, j \geq 0$ then, y_t does not [Granger] cause x_{t+1} w.r.t. J_t If

$$f(x_{t+1}|J_t) = f(x_{t+1}|J'_t)$$
$$\text{If,} f(x_{t+1}|J_t) \neq f(x_{t+1}|J'_t) \quad (19)$$

then, y_t is a 'prima facie' [Granger] cause of x_{t+1} w.r.t. J_t

This definition has two main assumptions:

- Cause occurs before effect,

- That a causal series (of data) contains special information about the series being caused that is not available in other available series, (here w_t).

When $f(x_{t+1}|J_t) \neq f(x_{t+1}|J'_t)$, it is implied that $f(x_{t+1}|J'_t)$ has a smaller variance of forecast error.

This conception of causality is appealing because it does not promote a strong definition of what causality is or is not, but rather it focusses on a way of finding causal links.

Many argue that Granger-causality, is not about causation but rather forecasting efficacy (Granger, C., 1988), and hence the term causation should not be used. It is easy to see why one might form this view, because, central to the definition of Granger-causality is an idea that if including a variable (y_t) betters the explanatory power of the model then y_t is a Granger cause of the process x_t. However, I disagree and believe Granger-causality has the potential to be just that - *causal*.

3.2.5 Why G-Causality is Causal

What I refer to as a sensibly selected subset is, as Granger puts it, "...variables for which the researcher has some prior belief that causation is, in some sense, likely." (Granger, C., 1988, 201). In addition, the term "better forecaster" would undoubtedly be tempered by the preference for more parsimonious models, i.e. the selective inclusion of links according to the degree of improvement they provide. Hence, if a sound statistical method for selecting possibly causal variables, and, a way of comparatively assessing appropriate models on their fit can be found, then there would be no reason to think that this is not at the very least causally sensitive.

3.3 Causal Modelling of Multivariate Time Series

In chapter 1, univariate AR(p) models and the various multivariate, VAR models were introduced. More so than other types of time series models, autoregressive models, where current variables are modelled in terms of past variables, lend themselves to a causal interpretation. In particular, sVAR models are a favoured type of model, especially in econometrics (Hamilton,1994,324-327). There are two main reasons for this. First, sVAR models can provide a better explanation of the process than a cVAR model. Second, although at first glance it may seem as though a cVAR model is more causally appropriate, with present variables explained solely in terms of past variables, there are some cases in which a causal description which includes relationships between contemporaneous variables is superior to one which excludes such variables.

3.3.1 Contemporaneous Causality

What is contemporaneous causality? Hicks (1979) provided an explanation of contemporaneous causality as the intrinsic link of two *processes* so that when one process changes, so does the other. An analogy to this might be a boat on the water, when the water swells the boat goes up, when the swell goes down so does the boat, clearly the swell *causes* the boat to rise, and, there is no time lag between the two events, hence the causality is deemed contemporaneous.

Granger (1988) gives three alternative possible explanations of causality in economics, although these principals appear quite general. The first is similar to that forwarded by Hicks (1979). The second is that there is no true instantaneous causality, but there is a small lag between cause and effect, which is much smaller than the interval over which the data are collected and the third is, that the observed contemporaneous causal link is due to a common cause of variables not included in the model.

3.4 Conditional Independence and Causal Sensitivity

In the previous section I discussed the relationship between conditional independence graphs (CIGs) and directed acyclic graphs (DAGs). Obviously, the idea is that the direction of the DAG is the causal direction of variables in a model. But why would we wish to derive a DAG from conditional independence relationships? In his paper "Conditional Independence in Statistical Theory" Dawid (1979) described the way in which he believed conditional independence could give insight into the underlying processes of situations that a statistician is concerned with. In particular, Dawid noted that conditional independence was a key in attaining causal information from data. This view that, "conditional independence assumptions are the primary vehicle for expressing substantive knowledge" (Pearl, 1998, 79) has been quite pervasive. It is easy to see why this is the case. If $A \perp\!\!\!\perp B$ then, "the occurrence of B does not affect the probability of A", and if there is no affect then there can be no causal affect. Of course, the converse is not necessarily true, if $A \not\perp\!\!\!\perp B$ then it cannot be said that there is definitely a causal effect. Hence, it is quite natural to think of conditionally independent variables as a subset of all non-causal relationships.

3.5 The GMTS approach

One particular approach to causally appropriate graphical modelling of multivariate time series is the approach to modelling first proposed by Marco Reale and Granville Tunnicliffe-Wilson (Reale, M., 1998) known as Graphical Modelling for Time Series (GMTS).

GMTS consists of three steps: (1) creation of the CIG, (2) creation of equivalent directed acyclic graphs (DAGs), and finally, (3) regression analysis and model selection.

3.6 Step 1: Conditional Independence Graph

As stated in chapter 2, creation of the CIG allows for the removal of all non-causal links. To construct this, all possible input variables need to be tested pairwise for conditional independence. In the case of time series this means that p needs to be known, so that a matrix of lagged variables can be constructed. As mentioned in chapter 1, the order of an AR process is defined by the sample PACF. This is interesting because, if we assume that our data follow a multivariate normal distribution then conditional independence is equivalent to the partial correlation of two variables being 0.

Formally,

$$\begin{aligned} &\text{If,} \quad X \sim MVN(\mu, V) \\ &x_i \perp\!\!\!\perp x_j \leftrightarrow \pi_{i,j} = \text{Corr}(x_i, x_j | \mathbf{X}\ \{x_i, x_j\}) = 0 \end{aligned} \tag{20}$$

Hence, the CIG will define the order of the VAR.

As this relationship holds, the CIG can be found by invoking the inverse variance lemma to find the matrix of partial correlations.

Suppose **X** is a vector following a multivariate distribution with mean μ and variance V. If,

$$\begin{aligned} &W = V^{-1} \ \text{and,} \\ &\tau_{i,j} = -\frac{w_{i,j}}{\sqrt{w_{i,i}, w_{j,j}}}, \text{where } \ w_{i,j} \in W \text{then} \\ &\tau_{i,j} = \text{Corr}(x_i, x_j | \mathbf{X}\ \{x_i, x_j\}) \end{aligned} \tag{21}$$

This lemma shows that if the covariance matrix V (or alternatively the matrix of correlation coefficients, which is equivalent to the covariance matrix scaled so that the diagonal contains unit vectors) is known, then the matrix of partial correlations can be found by taking the inverse and scaling row-wise and column-wise, so that the diagonal elements equal 1.

Once the matrix of partial correlations is found, where all values are between 0 and 1, a critical value must be defined, based on a user given t-value. Once this is done each value must be tested against the threshold value, where the correlation is deemed statistically significant the variables are considered conditionally dependent, and where it is below, the variables are considered conditionally independent. In this way a parsimonious structure is defined.

The threshold value of significance is defined as:

$$crit = \frac{t^2}{t^2 + v} \tag{22}$$

(Reale, 1998) where, v is the residual degrees of freedom.

From the CIG, the order of the VAR model can potentially be identified, although the problem of multiple testing should be properly addressed. Alternatively, we could use the information criteria discussed earlier, I refer to this method more fully in chapter 4.

3.7 Step 2: Creating Alternative Models

After the CIG has been found, the process of identifying possible DAG models begins. This is easily done with most edges, simply by adding the “arrow of time” from past to present variables. However, this cannot be done with the relationships among contemporaneous variables.

As mentioned in chapter 1, adding direction to the CIG of contemporaneously linked variables can be carried out by invoking subject matter knowledge. However, when the causal direction is unknown, methods of exhaustion, where models of all combinations are created and assessed are required.

Following the specification of possible contemporaneous DAG models possible moral edges need to be identified because these will be possible candidates for removal from the model. This is because their inclusion in the CIG maybe due to the relationship each parent has to a common child, rather than to each other and hence a possibly moral link may not represent a causal relationship. Once this has been carried out, all alternate possible DAGs corresponding to the CIG will have been found. Each graphical model can then be written down as a system of equations, and regression analysis and model selection performed.

When referring to models I use the terms saturated model and models with saturated moral links. This refers to all possible edges of a model being included. For example, a traditional modelling approach models are completely saturated. In a GMTS model, edges up to or at a certain lag, may be saturated, and if subset selection has not been performed, or, the best model selected preserves all moral links then this is a model with saturated moral links. Occasionally, the term effectively saturated is used to denote the fact that certain directions are omitted on physical grounds but all physically possible edges are included.

3.8 Step 3: Regression and Model Selection

For each alternate model the unknown coefficients are estimated and each model assessed by AIC, HIC and SIC criteria.

In this thesis the process is carried out in `MATLAB` and the system of equations specified as follows.

A matrix `G` of lagged variables is created. For example, a VAR(5) model which models 3 variables would have the associated matrix:

$$G = [X_t, Y_t, Z_t, X_{t-1}, Y_{t-1}, Z_{t-1}, \ldots, X_{t-5}, Y_{t-5}, Z_{t-5}] \tag{23}$$

If one alternative DAG corresponded to the system of equations:

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} && + \alpha_2 X_{t-2} \\ Y_t &= \beta_1 X_t && + \beta_2 Y_{t-4} \\ Z_t &= \gamma_1 X_t && + \gamma_2 X_{t-2} + \gamma_3 Y_{t-4} \quad + \gamma_4 Z_{t-5} \end{aligned} \tag{24}$$

These equations would be represented as `G(:,`$[1,4,7]$`)`, `G(:,`$[2,4,14]$`)`, `G(:,`$[1,7,14,18]$`)` in `MATLAB`. The numbers refer to the column of `G` equal to the variable written in the system of equations.

Once the models are specified least squares regression is performed, coefficients tested for significance using the t test, and AIC, HIC and SIC values calculated. Following this, subset selection is performed on possible moral edges. If the removal of an edge causes a decrease in the information criteria values, then the moral edge is omitted from the model. In chapter 3 a fully worked example of this process is given.

3.9 Is GMTS Causally Appropriate?

It certainly seems plausible that GMTS is a causally sensitive approach and further, it appears that it should select models whose variables satisfy the definition of G-causality. First, in deriving the CIG a subset of non-causal links are removed hence remaining links will be a "sensibly selected subset". Second, the model selected is the model which performs best according to the information criteria. Arguably, given the model selected by the information criteria will provide the smallest reasonable errors, without the inclusion of spurious edges, this would have to at least preserve the causal sensitivity introduced in selection. However, I will reserve any conclusive judgement on the matter until after I have seen how it performs in the case studies to come.

3.10 Other Advantages of GMTS

Apart from providing a framework to derive causally appropriate time series models using graphical modelling, there are two main advantages for using the GMTS methodology.

(1) As previously mentioned in section 3.3 sVAR models are often preferred over cVAR models. However, deriving them is usually done by orthogonalisation of residuals, a process which is not unique. The GMTS approach, however, provides a straightforward way of modelling sVAR processes.

(2) One of the major assumptions required for time series analysis to take place is an assumption of stationarity i.e. the series must be an $I(0)$ process. However, it has

been shown that GMTS can model non-stationary (integrated of order 1 - $I(1)$) processes effectively (Reale, M. and Tunnicliffe-Wilson, G., 2002).

4 Case Study I: Economics

> "The great thing about being a statistician" Mr Tukey once told a colleague, "is that you get to play in everyone's backyard". -Davis Leonhardt of John Tukey

The purpose of this case study is to present a worked example of the GMTS methodology in the area in which it was first applied i.e. economic time series. This dataset posed a difficulty in that it contained a large number of variables. This presented a combinatorial problem in the specification of alternate models. In the process of modelling these data I began to notice some interesting patterns, which lead me to conclude that more of the GMTS process is algorithmic, and hence, potentially computational, than was previously recognised.

4.1 The Data

Monthly financial data were collected from April 1987 to April 2002 (180 observations) by the Reserve Bank of New Zealand. These data were retrieved from their financial statistics database, `http://www.rbnz.govt.nz/statistics/`.

Data on nine variables were collected:

- Interest rates on money at call (R_t). This value was found as weighted average of \$10,000 at call. Institutions surveyed were weighted according to their NZ dollar funding.
- 90 day bank bills (S_t). These are "IOUs" issued by banks, usually in denominations of \$1 million or more, to raise capital until longer term financing can be found. 90 day bank bills are considered the benchmark for pricing other interest rates. These are averaged at 11am daily among banks, and in this case, the value averaged monthly.

- Yield of government stock at 1 year (T_t). Government stock is issued by the government at a fixed interest rate for a fixed term, although stock can be traded on the stock exchange. Government stock is issued as a way for the government to borrow money and is a low risk investment.

- Yield of government stock at 3 years (U_t)

- Yield of government stock at 5 years (V_t)

- Base lending rates (W_t). This is the weighted average of lending rates offered to new business borrowers, weighted by each surveyed institution's total NZ dollar claims.

- First Mortgage housing rates (X_t). This is a weighted average of first mortgage interest rates offered to new borrowers for residential properties. Weighted by each surveyed institution's total lending outstanding for housing property.

- United States of America interest rates (Y_t). This is a weighted average of interest rates from surveyed US institutions.

- Uncovered Interest Rate Parity (Z_t). The UIP can be written as:

$$E_t(s_{t+1}) - s_t = i_t^* - i_t + u. \tag{25}$$

(Stephens, D., 2004,3). The UIP is derived based on the theory that if the interest rates in one country increase compared to another then the two countries exchange rates should vary accordingly, this is so a person cannot borrow in one country to invest in another and hence make an instantaneous risk-free profit. Here i_t^* refers to the foreign interest rate (Australian interest rate in this example), i_t is the comparitive domestic interest rate and u is the risk premium associated with holding NZ dollar assets.

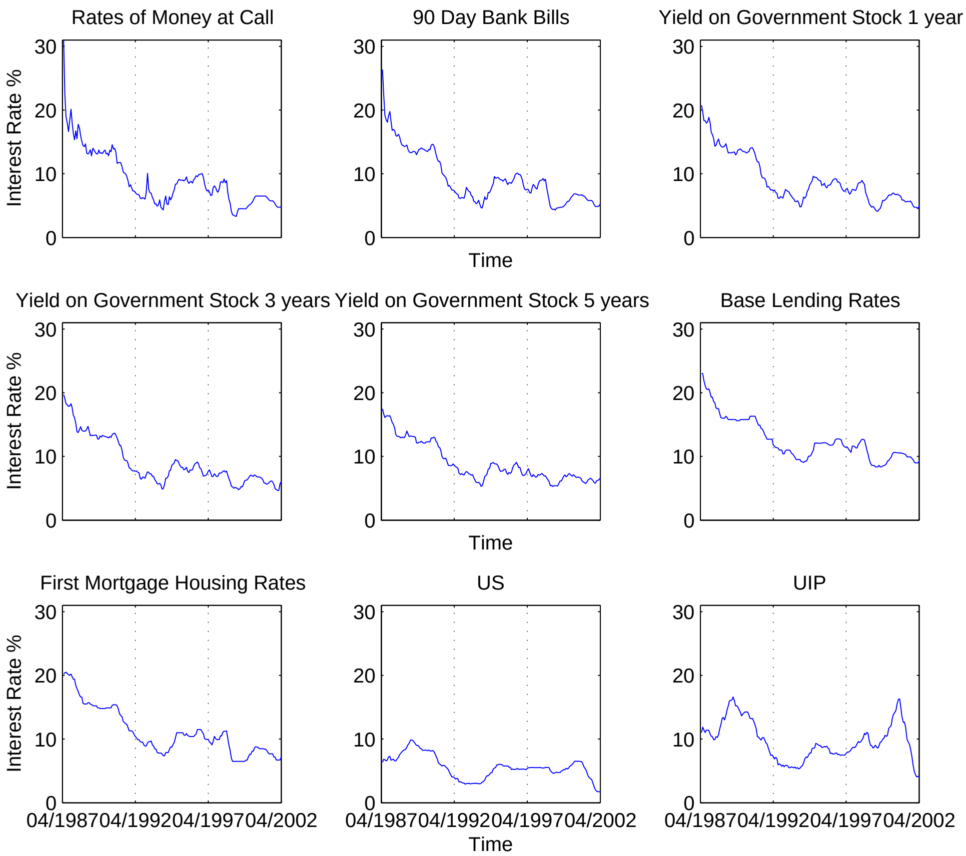


Figure 11: Plot of 9 variables in the dataset.

The plots in figure 4.1 display two trends. Plots 1-7 (reading left to right, top to bottom) display an overall decay with a local minimum between 1993 and 1994. This is followed by a small recovery over the next two years but then further decay. The first plot, of rates of money at call, contains the maximum interest rate of 31% as well as the maximum range from 5-31%, this is as expected because money at call is the most volatile of interest rates. The plot of 90 day bank bills has only a slightly smaller range than money at call.

The plots of US interest rates and the UIP display a different behaviour to the other 7 plots. These plots appear more stable over time with two peaks at about 1989 and 2001. The plot of US interest rates has the smallest range from 2-10%.

The number of observations (180) provides reasonable degrees of freedom for analysis of 9 variables, so there was no need to reduce the number of variables in the dataset to be modelled.

4.2 Step 1: Deriving the CIG

To derive the CIG the function `cigts` in `MATLAB`, written by M.Reale (appendix A.1) was executed with the input arguments being: X-the full data set as described above, $lags$=5. The t-value was defined as tv=1.98 which corresponds to an α level probability of 0.05 for 160 degrees of freedom. From this set of input values the table 1 is derived (a complete table is given in appendix C). The definition of a lag of 5 was arrived at following exploration of the behaviour of different CIGs, starting with a high lag specified and reducing the lag systematically, until I was satisfied that the CIG had terminated at $p + 1$. Other methods for finding the order, based on information criteria, would have selected a VAR(2) model, however, as the interest here was with a causal model, this was considered too small. The t-value varied according to the degrees of freedom, but in effect, with 180 observations and 9×6 variables being modelled, the t-distribution is approximately a normal distribution and hence a value of 1.96 would have sufficed.

variables	lags	pc	pc	pc	pc	pc	pc	pc	pc	pc
1	0	1	x	x	x	x	x	x	x	x
2	0	1	1	x	x	x	x	x	x	x
3	0	0	1	1	x	x	x	x	x	x
4	0	0	0	1	1	x	x	x	x	x
5	0	0	0	0	1	1	x	x	x	x
6	0	0	0	0	0	0	1	x	x	x
7	0	0	1	0	0	0	1	1	x	x
8	0	0	0	0	0	0	0	0	1	x
9	0	0	0	0	0	0	0	0	1	1
1	1	1	1	0	1	0	0	0	0	0
2	1	1	1	1	1	0	0	0	0	0
3	1	0	1	1	1	0	0	0	0	0
4	1	0	0	1	1	1	0	0	0	0
5	1	0	0	0	1	1	0	0	0	0
6	1	1	0	0	0	0	1	0	0	0
7	1	0	0	0	0	0	0	1	1	1
8	1	0	0	0	0	0	0	0	1	1
9	1	0	0	0	0	0	0	0	1	1
1	2	0	0	0	0	0	0	0	0	0
2	2	1	1	0	0	0	0	0	0	0
3	2	0	1	1	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 1: `MATLAB` output of CIG, an entry of 0 indicates conditional independence and an entry of 1 indicates conditional dependence i.e. that an edge can be drawn between two nodes. Full table of values contained in appendix C

This table (1) provides all the information required to create the CIG. The first two columns in each row refer to a potential variable in the model, while columns 3-11 refer to the 9 variables at time t we wish to find equations for. For example, row 4 of the table contains the information 4 0 0 0 1 1 x x x x x. This means that the node of variable 4 at lag 0, U_t in this case, has an edge which links to variable 3, (T_t), and trivially to itself. The x's simply prevent doubling up of information regarding contemporaneous variables. Row 10 of the table is 1 1 1 1 0 1 0 0 0 0 0. In this case variable 1 at lag 1 (R_{t-1}) has an edge linking it to variables 1, 2 and 4 (R_t, S_t and U_t). Reading the entire table in this way the CIG in figure 12 can be derived. As previously mentioned, the order of the CIG specifies the order of the VAR and hence in this case we have a VAR(5) model.

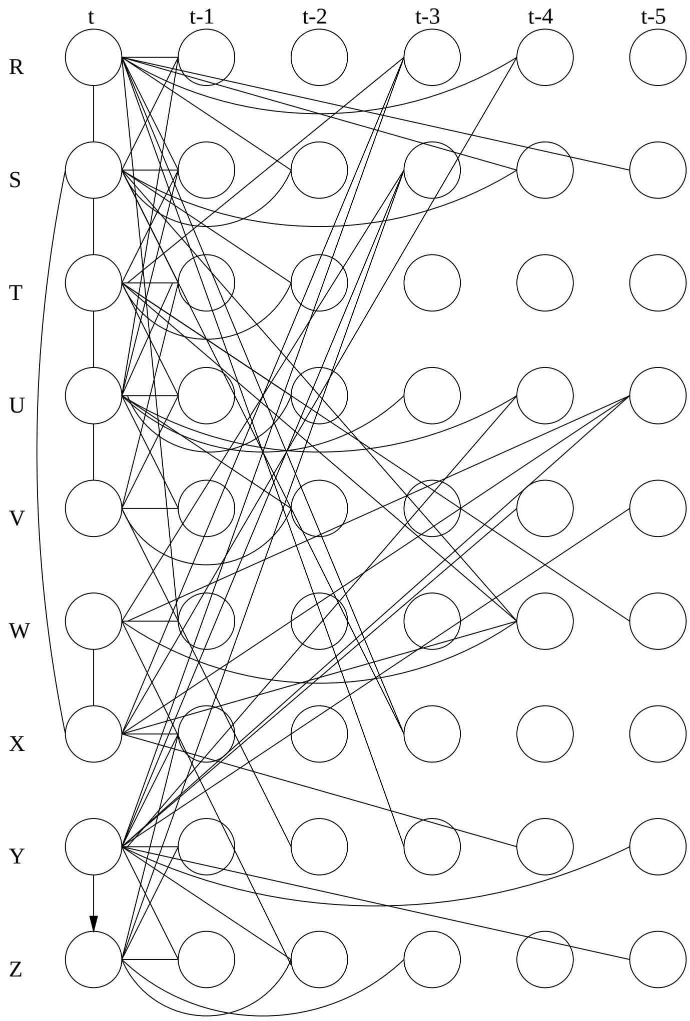


Figure 12: Conditional independence graph of economic data as read from table 1. One edge has been directed from Y_t to Z_t

4.3 Step 2: Creation of Alternative DAGs

To create alternative DAG models, and hence alternative models, the first step is to define the two "causal" directions, and add direction to the CIG in accordance to these. With time series, the first step is to add the "arrow of time". The vertical direction (the contemporaneous direction), is not always known. In this case, there is one known relationship, the relationship between Y_t and Z_t. This edge must be directed from Y_t to Z_t because it does not make sense to think that the UIP (a parity measure between New Zealand and Australia) would be a cause of interest rate changes in the United States of America. Beyond this however, all other causal directions are unknown. To find the possible DAGs in this case will require a method of exhaustion. This arrow has been inserted in figure 12.

4.3.1 Step 2(a): Finding Possible Contemporaneous Models

The problem of specifying alternative models when large numbers of variables were involved appeared to be quite tricky because it was not simply a matter of specifying alternate models but, specifying alternate models which were acyclic and consistent with the rules of moralisation. However, after some thought it became clear that because the specification of all possible models was equivalent to generating a list of binary words it was computationally efficient to carry out and hence systematic methods of elimination from an exhaustive list were easily implemented.

Computationally, specifying alternate contemporaneous models is straightforward. Each edge only has two possible outcomes, oriented up or down. Hence, to find all possible configurations of the contemporaneous variables simply requires the generation of a list of the 2^6 binary words of length 6, with an additional row for the known edge (Y_t, Z_t) which is oriented down. There are 2^6 words because there are 6 contemporaneous edges each with two orientations. In these words the digits 0 and 1 are assigned an orientation. For example, a 1 refers to a downwards orientation and 0 refers to upwards orientation. If we consider a graph with only 2 possible edges there are 2^2 possible combinations of ori-

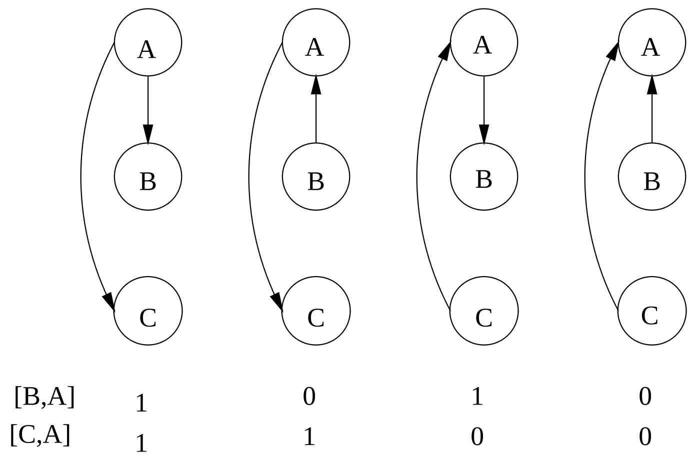


Figure 13: A simple example of a graph with two edges. There are 2^2 binary words corresponding to a different set of possible orientations.

entations as given in figure 13. In order for this method to work, it is important that each edge is specified in a consistent way. In this case the default specification is downwards. For example in figure 13 [B,A] refers to the variable A being directed into B. This way when the orientation is upwards the pairs can simply be flipped to reflect this. Once all combinations are generated, these need to be checked for cycles and DAG representations inconsistent with the stated CIG (code written by C Meurk written for `MATLAB` contained in appendix A.3). In this case, there are no cycles, however there are a number of disallowed DAG configurations which correspond to the subgraph in figure 14. Removing such inconsistencies reduces the number of possible contemporaneous DAGs from 64 to 13, a more manageable number to deal with. The possible contemporaneous DAG structures for this case study are given in figure 15. Once each of these models are specified is found these can be written as systems of equations. For example, Model A corresponds

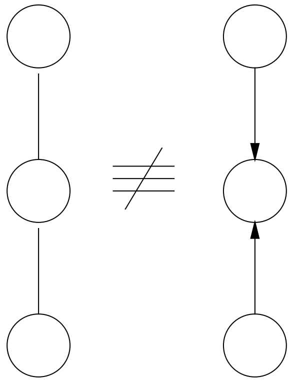


Figure 14: DAG representation inconsistent with corresponding CIG. Any graph containing this subgraph is removed from the set of possible models.

to the equations:

$$
\begin{aligned}
R_t &= \alpha S_t \\
S_t &= \beta_1 T_t + \beta_2 X_t \\
T_t &= \gamma U_t \\
U_t &= \delta V_t \\
V_t &= 0 \\
W_t &= \varepsilon X_t \\
X_t &= 0 \\
Y_t &= 0 \\
Z_t &= \zeta Y_t
\end{aligned} \tag{26}
$$

For MATLAB to perform regression, these equations need to be written in terms of columns of the matrix of lagged variables $G = [R_t, S_t, \ldots, Z_t, R_{t-1}, \ldots, Z_{t-1}, \ldots, Z_{t-5}]$. The column numbers of each variable are given inside their corresponding nodes. Hence, all alternate contemporaneous models corresponding to these subgraphs can be specified as in table 2.

4.3.2 Step 2(b): Adding Edges with Definite Lagged Variables

Once the possible contemporaneous relationships are found, these need to be added to the lagged variables and any possible moral links identified. At this point, any edges judged

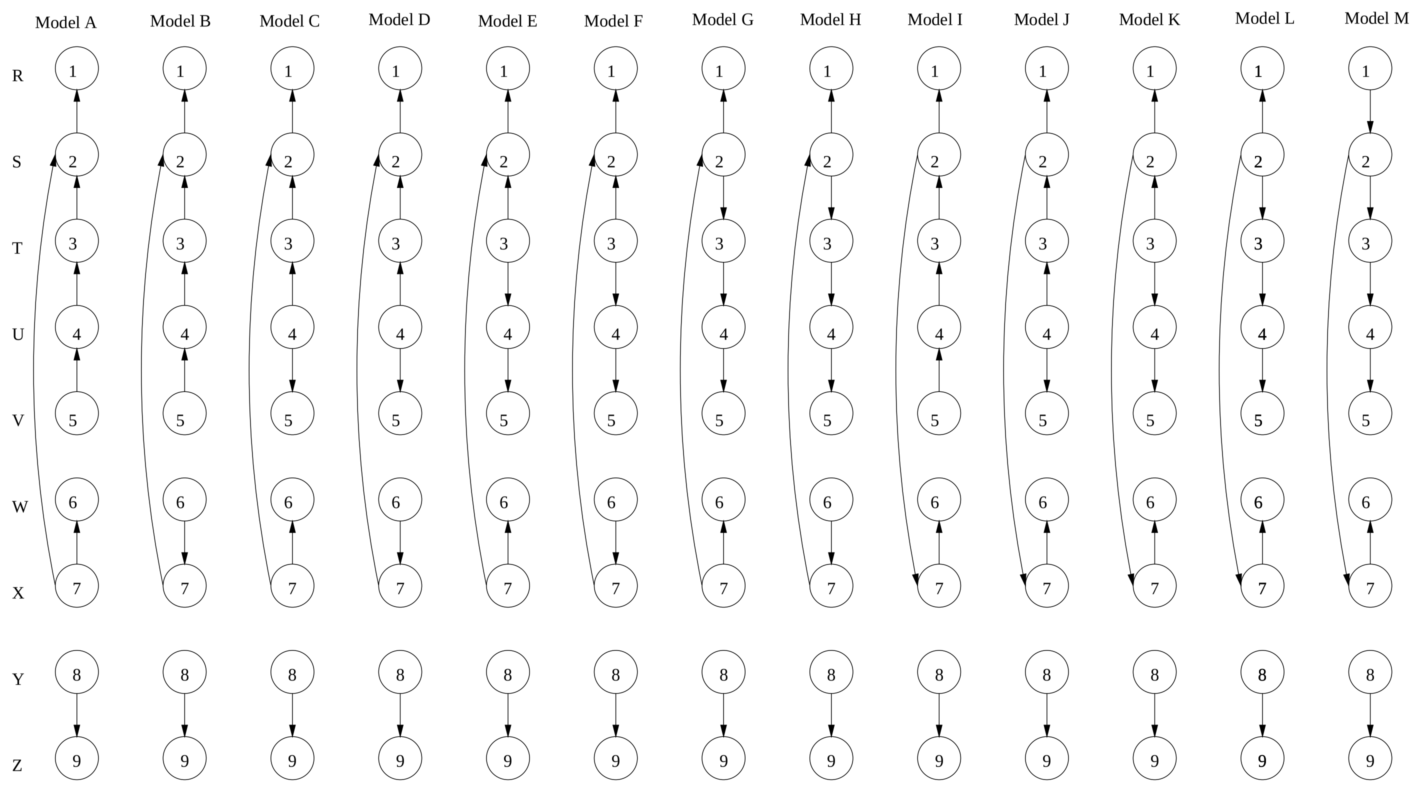


Figure 15: Table of all 13 possible contemporaneous DAG configurations which correspond to the contemporaneous CIG. Numbers inside each node refer the column of the matrix `G` they correspond to.

to be non-existent on subject-theoretic grounds can be removed. In this case all edges which lead to Y_t (US interest rates), except Y_{t-k} values, were removed.

The lagged edges fall into two categories, those which are possibly moral, and those which are real. Edges which are definitely real are identified if they fulfill one of two criteria:

Either

A node contains only one outgoing edge.

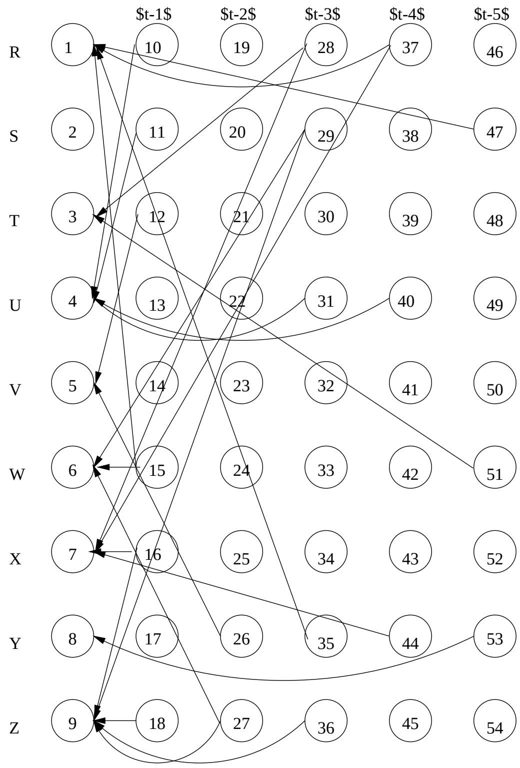
Or

A node contains multiple outgoing edges, but the outgoing edges do not meet contemporaneous variables, which themselves are joined, i.e. the edge is NOT part of a subgraph of the form in figure 6.

The subgraph containing edges which satisfy either of these conditions is referred to as the *generic part*. This is because it will, in its entirety, be part of all the possible models specified in section 4.3.1. The generic part of our model, with its corresponding representation as a system of equations is given in figure 16, where $\mathrm{variables_t}$, refers to the contemporaneous relationships and $\mathrm{moral_{t-k}}$ refer to the possible moral links, to be found in section 4.3.3.

4.3.3 Step 2(c): Identification of Possible Moral Edges

The final step in model specification is the existence of possible moral edges. This involves the edges of our original CIG from section 12, not included in the model figure 16. Again, this step has been considered a difficult step, because it needs to be carried out by hand. But, as I discovered, there is a relationship between different variables and their possible moral links, and this allows for alternate models to be specified algorithmically and hence can be automated.



$$R_t = \mathrm{variables_t} + \alpha \mathrm{W_{t-1}} + \alpha \mathrm{Y_{t-3}} + \alpha \mathrm{R_{t-4}} + \alpha \mathrm{S_{t-5}} + \mathrm{moral_{t-k}}$$
$$[1, 15, 35, 37, 47, v, m]$$
$$S_t = \mathrm{variables_t} + 0 + \mathrm{moral_{t-k}}$$
$$[2, v, m]$$
$$T_t = \mathrm{variables_t} + \gamma \mathrm{R_{t-3}} + +\mathrm{moral_{t-k}}$$
$$[3, 28, 51, v, m]$$
$$U_t = \mathrm{variables_t} + \delta \mathrm{R_{t-1}} + \delta \mathrm{S_{t-1}} + \delta \mathrm{T_{t-1}} + \ldots + \mathrm{moral_{t-k}}$$
$$[4, 10, 11, 31, 40, v, m]$$
$$V_t = \mathrm{variables_t} + \mathrm{generic_{t,t-k}} + \mathrm{moral_{t-k}}$$
$$[5, 26, v, m]$$
$$W_t = \mathrm{variables_t} + \mathrm{generic_{t,t-k}} + \mathrm{moral_{t-k}}$$
$$[6, 15, 27, 29, v, m]$$
$$X_t = \mathrm{variables_t} + \mathrm{generic_{t,t-k}} + \mathrm{moral_{t-k}}$$
$$[7, 16, 28, 37, 44, v, m]$$
$$Y_t = \mathrm{variables_t} + \mathrm{generic_{t,t-k}} + \mathrm{moral_{t-k}}$$
$$[8, 53, v, m]$$
$$Z_t = \mathrm{variables_t} + \mathrm{generic_{t,t-k}} + \mathrm{moral_{t-k}}$$
$$[9, 16, 18, 27, 29, 36, v, m]$$

Figure 16: Generic DAG and corresponding equations.

Consider figure 17. There are 16 different subgraphs containing a possible moral edge, and hence each model will have 16 possible moral edges, corresponding to the number of subgraphs of the type given in figure 6. The identification of which is the moral edge will depend on the contemporaneous link between variables. For example, in figure 18, the lagged input variable is the same for the equations of both A and B, but which edge is the possible moral edge swaps when the orientation between A and B changes. Hence, once one model specification is known all others can be derived automatically. This is shown in table 3 where bracketed () expressions denote possible moral edges and then concatenated with each possible contemporaneous model.

4.3.4 Alternative Models

Finally, the information from each of these three parts, the contemporaneous part, the generic part and the moral part can be combined to give the specification of all possible

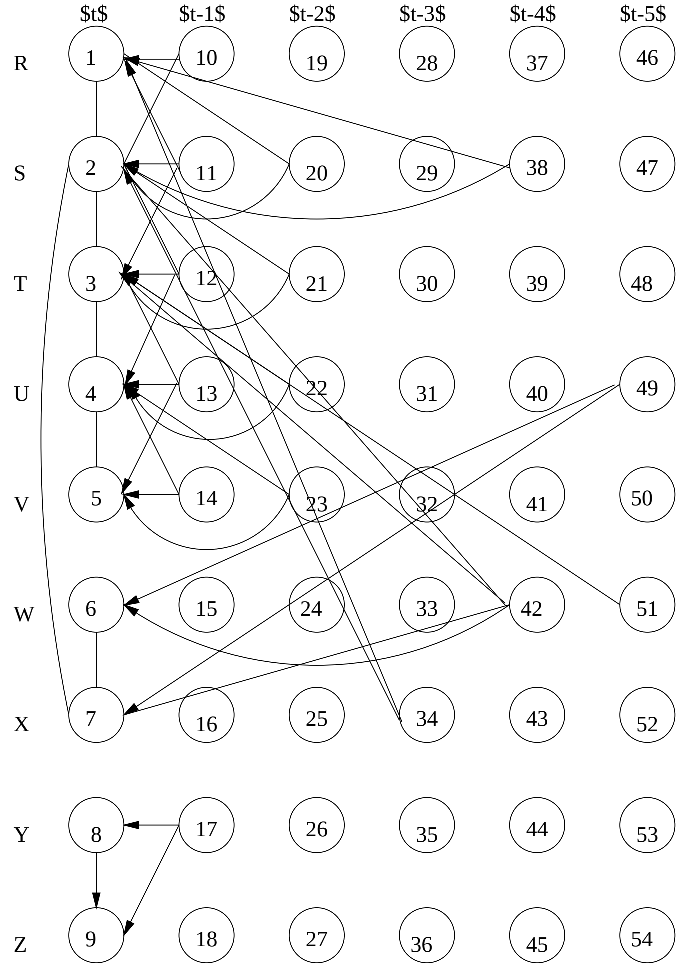


Figure 17: Subgraph containing all possible moral edges. This subgraph + the generic part represent the full model.

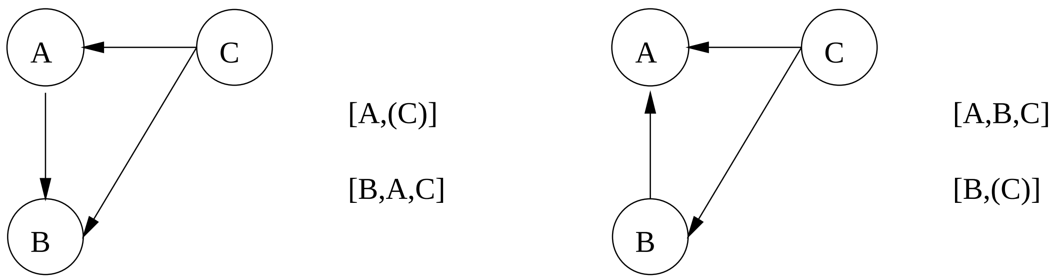


Figure 18: This figure shows the way in possible moral edges can be found from a graph. Bracketed () values denote possible moral edges.

models (table 4). Notice that, in specifying the saturated forms, i.e. with all possible moral links included, the number of models is specified by the number of viable contemporaneous models. Moral links are subsequently removed through subset selection.

4.4 Step 3: Regression and Best Model Selection

Using the `MATLAB` code `DAGfit` (written by M Reale, appendix A.2) regression can be performed on all models as specified in table 4. Each model was assessed by three criteria - AIC, HIC and SIC, as discussed in section 2.3. The results for the 13 alternate models, with comparisons to the saturated case are given in table 4.4.

By both AIC and HIC criteria the saturated sVAR was selected as best model, but using the SIC it was the second worst. It was decided that the saturated sVAR should not be selected in this case because its rankings were inconsistent among the criteria, a known problem for these criteria (refer chapter 1). Ignoring the sVAR model, the first, second, third and fourth models respectively are Models I,J,K and L. These models maintain their ranking in all three criteria and hence were selected for further subset selection. This decision to look at these four models was taken with the understanding that should models show large improvement when subsets were selected, a new strategy of selection would need to be used. Also, it is important to note that the 13 models, which at this stage include all possible moral links were identical except for their contemporaneous links. It was important to check all specified models to see if there were any who had SIC values that were not too much larger than Model L and that had very different moral configurations because this could cause a dramatic change in subset selection. Referring again to table 4 models I,J,K and L are of a very similar form, compared with Model M for example. This difference refers to the layout of possible moral links. This method of narrowing down possible models, found by GMTS, means that the models in table 4 are the *set of possible candidate models* (Burnham, K., *et al.*, 1998, 19) from which the "best models" are taken, and further subset selection carried out, to see if initial ranking changes. If rankings are unchanged the best of the original candidate models is selected and thorough subset selection undertaken. This method of model selection differs from that proposed by Reale (1998) where following the specification of the CIG, alternate models for contemporaneous variables were found and assessed, and subsequently subset selection, carried out *prior* to regression analysis. I believe the method that I have outlined is an improvement because it requires less manual analysis in early phases (not to be mistaken with an individual statistician's active analysis of the model at each stage), allowing for models with more variables and a higher lag to be assessed.

Initial subset selection showed very little change in information criteria values, and no change in rankings. Hence, model I was selected for thorough subset selection and the model in figure 19 corresponding to equation 27 was selected. This subset selection only looked to remove possible moral links. Obviously, in conjunction with more knowledge about the economic study it could be possible to remove more links. Adding the regression coefficients some links were considered insignificant according to the t-value of their regression coefficient. Note that thorough subset selection resulted in very little model improvement.

$$
\begin{aligned}
R_t &= 1.1342S_t + 0.5104R_{t-1} - 0.7148S_{t-1} \\
S_t &= 0.9471T_t + 1.1225S_{t-1} - 1.1482T_{t-1} - 0.3535S_{t-2} + 0.4360T_{t-2} \\
T_t &= 0.9364U_t + 0.1146S_{t-1} + 1.0095T_{t-1} - 1.0125U_{t-1} - 0.3383T_{t-2} + 0.3161U_{t-2} - 0.0203W_{t-5} \\
U_t &= 0.9722V_t + 1.0920U_{t-1} - 1.0099V_{t-1} - 0.2786U_{t-2} + 0.1394U_{t-3} - 0.0789U_{t-4} \\
V_t &= 1.3445V_{t-1} - 0.3682V_{t-2} + 0.0287Y_{t-2} \\
W_t &= 0.4582X_t + 0.6783W_{t-1} + 0.0333Z_{t-2} - 0.1505S_{t-3} + 0.0633W_{t-4} - 0.0898U_{t-5} \\
X_t &= 0.3123S_t + 0.7906X_{t-1} - 0.1197R_{t-4} + 0.1017W_{t-4} - 0.0527U_{t-5} \\
Y_t &= 1.1710Y_{t-1} - 0.1741Y_{t-5} \\
Z_t &= 1.6392Y_t - 1.5177Y_{t-1} + 1.1364Z_{t-1} - 0.2908Z_{t-2} + 0.1138Z_{t-3}
\end{aligned}
\tag{27}
$$

Finally, the errors were plotted and appear stable (figure 20), with the possible exception of the UIP where the errors appear to diverge over time. Although, the maximum error is

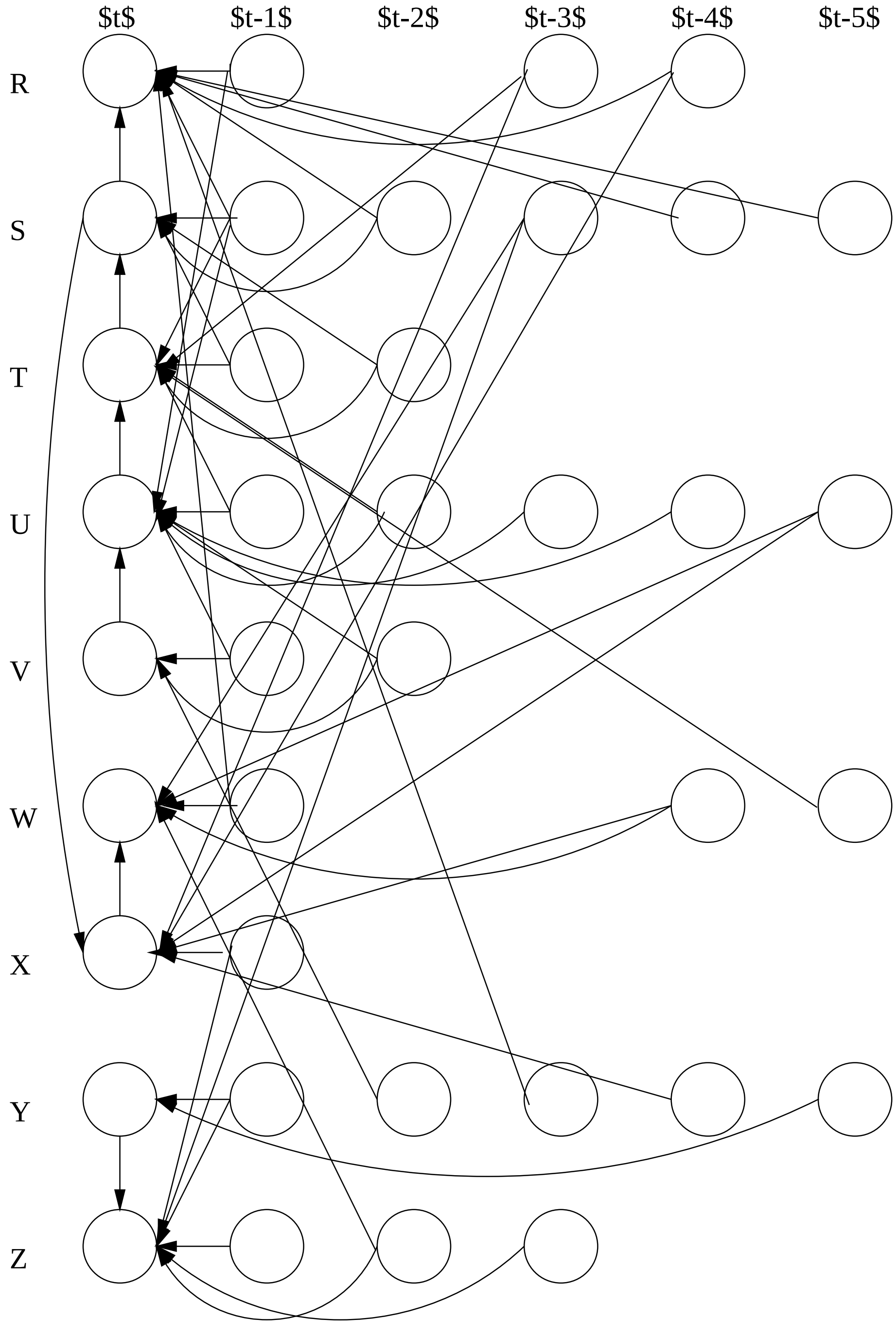


Figure 19: The information criteria values corresponding to this model are -4.6934+e003 (AIC), -4.6214e+003 (HIC) and -4.5159e+003 (SIC)

Figure 20: Plot showing residuals errors after VAR modelling

1, I would be careful in trying to predict the UIP from this model, because it is possible that the divergence continued.

4.5 Interpretation

The interpretation of this model is that the long term interest rates drive the short term interest rates, i.e. government stock at 5 years, determines government stock at 3 years, all the way down to the interest rates of money at call.

The US interest rates have a lagged effect on both short term and long term interest rates consistent with the previous statement. For example, there is a 3 month lag between US interest rates and its affect on interest rates on money at call, while only a 2 month lag to affect government stock at 5 years.

There is some possible "feedback" with the edge from 90 day bank bills to first mortgage lending rates, because it is only due to a missing link between base lending rates and first mortgage rates that a cycle is not created. Base lending rates appear to affect government stock with lag of 5 months, and it is possible that a higher order model may have shown links with other lending rates.

There is a repeated pattern between 90 day bank bills, government stock at 1 year, 3 years and 5 years which is obvious when viewing the graphical model in figure 19, implying that these are all highly similar processes, or alternatively, driven by the same processes. It would be interesting to model this system where these four variables are represented by one process to see if this changes the model. In addition, the US interest rates, with no input from other variables, may be best left unmodelled and only included in terms of its affect on other variables.

4.6 Conclusion

Initially, this case study seemed to pose an intractable combinatorial problem. But, by splitting up the problem into three parts, it quickly became algorithmic, as can be seen in figure 21. However, once this was done, it turned out that the removal of possible moral links made little improvement to the model, and no change in rankings. The stability of rankings may be due to the fact that possible moral edges are determined by contemporaneous relationships, both mathematically and causally. That is, if there is a link between X_t and Y_t it is more likely that there will be a link between X_t and Y_{t-k} values. Hence, the initial ranking candidate models is likely to be preserved in subset selection. In addition, if this is the case then the strategy of model selection used by Reale (1998) where subset selection occurred prior to regression following analysis of contemporaneous models, and the strategy used here, where all candidate models are created and then regression analysis performed, will both select the same best model. This is because even if the relationship between lagged variables is included, model ranking is dependent on the contemporaneous configuration. However, as mentioned, the method described here has the advantage that it can be automated.

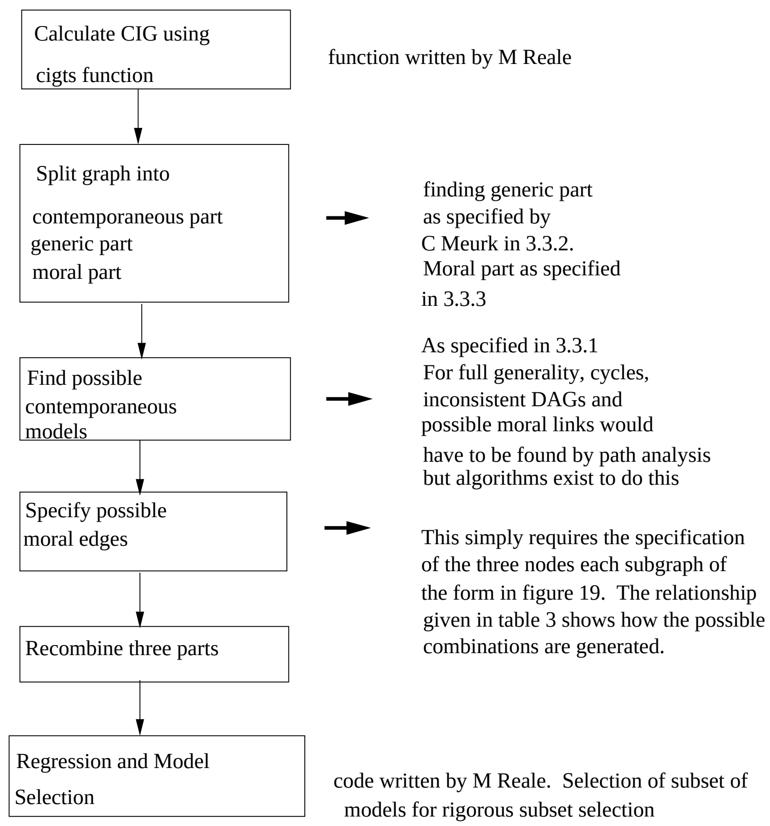


Figure 21: Flowchart showing strategy for automating GMTS.

5 Case Study II: Hydrological Data

These data were collected from the Pukemanga catchment near Hamilton New Zealand. The study was "part of a larger project ... designed to investigate sediment and nutrient loss from steep hill country sheep and cattle farmland" (Davie, T., 2004). Thanks to Tim Davie from Manaaki Whenua, Landcare Research, Lincoln, for providing these data and a working paper.

5.1 Data Description

The data consisted of soil moisture readings, rainfall data and water depth measurements from weirs along the catchment. Recording devices for soil moisture were located in the upper, mid and lower basins of the catchment (figure 22).

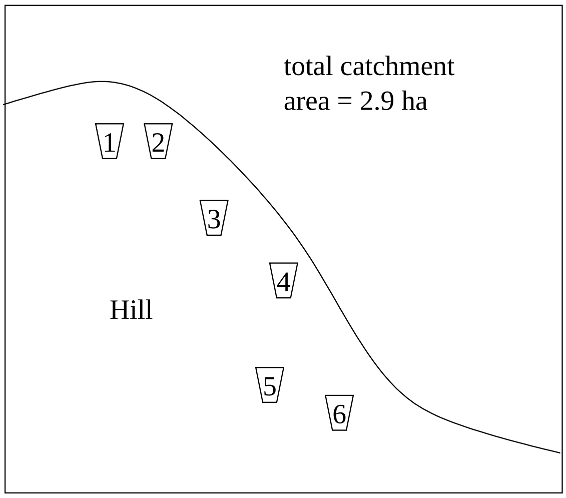


Figure 22: Simplified diagram showing the layout of recording devices in the catchment area.

Soil Moisture

Six runoff plots were instrumented using three Time Domain Reflectometers to record soil moisture at three depths 5cm, 10cm and 15cm. The six runoff plot sites provide two replicates over three different types of soil and slope. Data loggers (Campbell Scientific Instruments CR10X) at the plot sites took soil moisture readings every minute from which the mean was calculated every 30 minutes. The resulting time series was created from these means.

These data were collected electronically over the period from the 1st of September 2002 to the 1st of March 2004. For each of the six soil moisture plots there are approximately $26,000$ observations at each of three depths.

Rainfall

Rainfall data were collected continuously at the site. Two gauges were used, a Texas Electronics rain gauge (0.1mm/tip), and an OTA tipping bucket rain gauge (0.5mm/tip). A tipping bucket filled and emptied automatically whenever the water depth in the bucket was 0.2mm. During non-storm events data (summed over the 30 minute period) were recorded every 30 minutes, however over a certain threshold of rainfall, data were collected every minute, and in pre and post-storm periods data were collected every 10 minutes. Consequently, the rainfall data are unevenly spaced. Due to the spacing of the other series, it was necessary to group the data into even intervals. This was done by running the R code `spacedata` written by C. Meurk(B.1). To be consistent with the soil moisture data the smallest interval of time that could be modelled was 30 minutes. Data were grouped by taking the sum of rainfall in each 30 minute interval. In total, the rainfall data consisted of about 16,300 observations prior to grouping.

5.2 Preliminary Data Analysis

Given that there were so many series (six runoff plots at three depths and rainfall), containing such a large number of observations, it was not computationally possible to model all the series. Hence, exploratory data analysis was carried out to identify possible trends and variability in the data so that a subset could be selected which would provide an appropriate and interesting series for GMTS modelling.

5.2.1 Analysis by Plotting

To begin, simple descriptive techniques were used to explore the data. The quickest and easiest way to get a feel for the data was to plot it (figure 23).

Soil moisture at sites 1,2,5 and 6 showed very high correlation both among sites and depths. The soil moisture series at 5cm depth showed the highest variability while the soil moisture series at 15cm depth showed the lowest variability and overall soil moisture levels, but with a similar pattern of fluctuation as the 5cm series. There were few exceptions to this, the most noticeable being between February 2003 and May 2003 where there was a period of low variability and reduction in soil moisture levels and a change in the order of comparative moisture at each site depth. Overall, at the sites 1 and 2 soil moisture fluctuated between 30% and 75%, while at sites 5 and 6 it fluctuated between 22% and 63%.

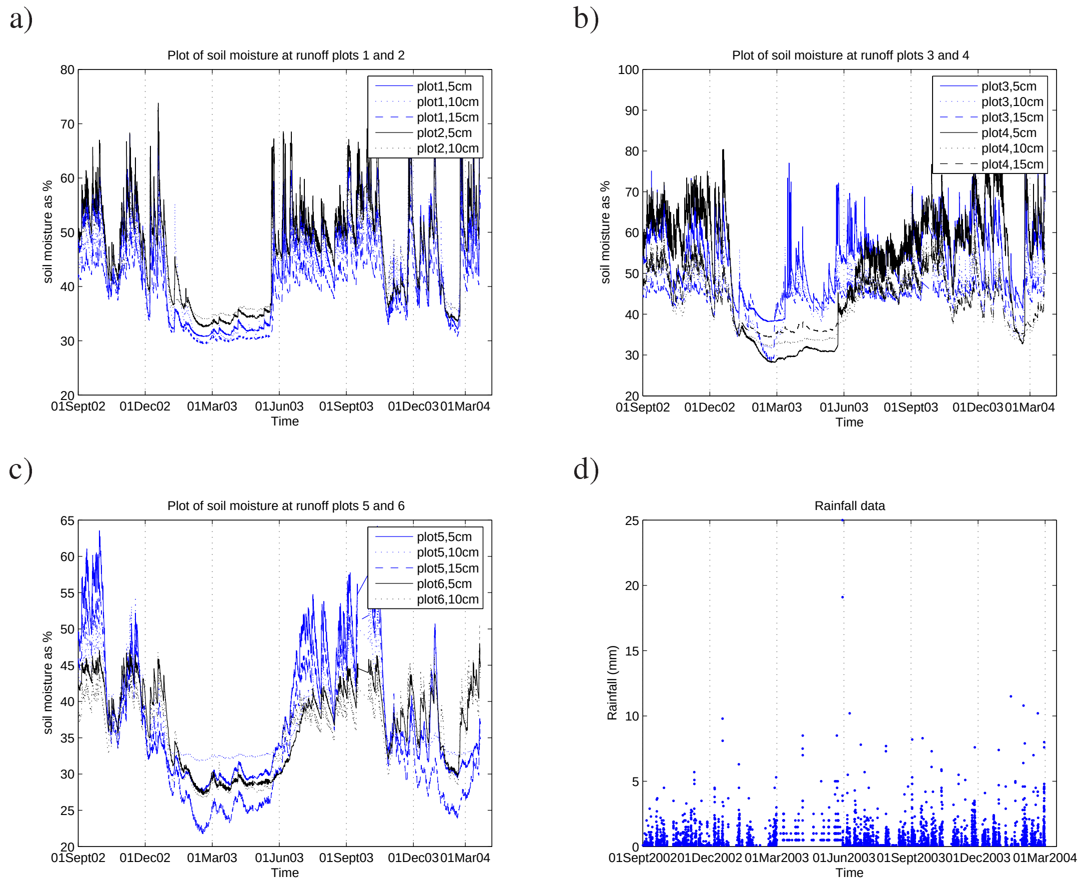


Figure 23: Plots showing soil moisture levels at all sites and rainfall data, plotted as scatter plot with 0 values removed for ease of inspection.

Soil moisture at sites 3 and 4 behaved differently from the others, though high correlation remained. There was also high variability however the overall behaviour appeared more stable than with sites 1,2,5 and 6. Between February 2003 and May 2003 there is some unusual behaviour displayed with a spike from 45%-72% occurring around late March-early April not matched in any other plots. Soil moisture at sites 3 and 4 fluctuated between 29% and 80%.

Rainfall showed approximately the same distribution throughout the study with the notable exception from February 2003 to May 2003. During most rainfall events 0-0.5mm rain fell per half hour, with the highest recorded at 2.5mm per half hour.

5.2.2 Analysis of Correlation

The series in this dataset were further explored using the `MATLAB` commands `autocorr`, which finds the sample autocorrelation function for a series up to a specified lag and `crosscorr`, which finds the sample cross-correlation function between two series up to a specified lag (figure 5.2.2). The `autocorr` analysis shows that correlation within site series (results for site 1 and site 6 at 10cm depth shown in figure 5.2.2) was very high, i.e. none less than 0.97 up to lag 20 ($t - 10$ hours). The rainfall series was not so highly autocorrelated between 0.056 and 0.565 (3dp) up to lag 20.

The `crosscorr` analysis showed very high correlation of series between sites, with correlation being between about 0.922 and 0.950, for both series shown. The exception to this was the cross correlation of the rainfall series with soil moisture which ranged from 0.027 to 0.1. Unlike the autocorrelations which appeared to decrease linearly, the cross correlations displayed a bell shape with the highest correlation occurring between $t - 9$ and $t - 11$. These behaviours are not atypical in hydrological time series (HTS) display. HTS often display high serial correlation (or partial correlation), and in practise hydrological series are often “pre-whitened”, i.e. some of the correlation removed, prior to modelling (Yue, S. *et al.*, 2003, 51).

--Selected MATLAB Output--

********AUTOCORRELATION TO LAG 20************

plot1(10cm)	plot6(10cm)	rainfall
1	1	1
0.999542	0.999611	0.565194
0.998655	0.998837	0.36045
0.997579	0.997924	0.281237
0.996382	0.996944	0.240073
0.995093	0.99592	0.216629
0.993721	0.994855	0.196854
0.99228	0.993757	0.173251
0.990789	0.99263	0.158722
0.989261	0.991495	0.142035
0.987709	0.990368	0.126093
0.986142	0.989254	0.116902
0.984568	0.988167	0.0993248
0.982993	0.987115	0.0856055
0.981416	0.98609	0.0796165
0.979841	0.985093	0.0753873
0.978268	0.984124	0.070396
0.976701	0.983177	0.0670891
0.975144	0.982245	0.0703455
0.973594	0.98132	0.0604151
0.972046	0.980402	0.0559149

****

********CROSSCORRELATION TO LAG 20************

Plot1,5(10cm)	Plot1,2(10cm)	Plot6(10cm),Rainfall
1	1	1
0.928497	0.926254	0.0272509
0.93142	0.928918	0.0295769
0.934337	0.931562	0.0317559
0.93723	0.93422	0.0341546
0.940048	0.93689	0.0369983
0.94276	0.939551	0.0411666
0.945314	0.94214	0.0466204
0.947556	0.944596	0.0529891
0.949402	0.946845	0.0608685
0.950391	0.948364	0.0744209
0.948914	0.947546	0.0976419
0.946541	0.945823	0.100851
0.943839	0.943658	0.101281
0.9409	0.941131	0.101174
0.937808	0.938357	0.100516
0.934671	0.935325	0.0992596
0.931524	0.932169	0.0976477
0.928367	0.928952	0.0961152
0.925181	0.925689	0.0953346
0.921972	0.922414	0.0950495

****

Figure 24: Printout of MATLAB code for the autocorr and crosscorr analysis for soil moisture and rainfall. Note that, when executing `crosscorr`, MATLAB evaluates the CCF at $\pm 1..p$. For time series analysis we are only interested in correlation with past values and so all future values have been omitted.

As mentioned in The GMTS methodology can model both $I(0)$ and $I(1)$ processes (refer section 3.10). Nevertheless, checks for stationarity of the series were done using the augmented Dickey-Fuller test (ADF) in R with the order p estimated at 200 following Hamilton (Hamilton, 1994, 530) and with `type` specified as `none` for all tests. The ADF tested the hypothesis H_ϕ :The series is $I(1)$ vs H_A :The series is $I(0)$.

5.2.3 Selection of Variables

Given the high correlation at each site among different depths it was a relatively easy decision to reduce the number of variables being studied to just the 10cm depth series. This reduced 6×3 series to 6. In addition, due to the high correlation between replicates the number of series to study was reduced further from 6 (3 sites $\times 2$ replicates) to 3 sites excluding replicates. Finally, because this study was interested in the causal implications of the GMTS methodology it was desirable to include the rainfall series. Hence, the decision was made to look at only 3 variables in total, rainfall (X_t), the series of soil moisture data at runoff plot 1 (Y_t) and at runoff plot 6 (Z_t) at 10cm depths. Choosing these variables, allowed the added advantage of defining a vertical direction of causation from from $X_t \to Y_t \to Z_t$ i.e. water always flows downhill.

5.3 Modelling the Data with GMTS

This dataset was used to thoroughly explore the GMTS approach to time series models. This exploration split roughly into three questions. The first question posed was how does modelling of these processes with GMTS compare with a traditional modelling approach? Second, how do the GMTS models change as the data are aggregated and/or when different seasons are modelled. Finally, the question is asked whether this can be thought of as a causally sensitive modelling procedure, and an indepth discussion on the application of the principal of moralisation in time series analysis is undertaken.

5.3.1 Specifying the Order of the VAR

In chapter 4, it was seen that the order of the VAR is given by the order of the CIG. However, after extensive investigation of the CIGs that were generated, I found it very difficult to decide on a value for the order, because models were all initially saturated followed by a small but steady number of subsequent edges with no obvious termination up to lag 96. It was not possible to generate CIGs any larger than this because even generating a CIG up to lag 96 could take half a working day to compute. It was decided that another method for determining the order of the VAR must be devised. The method for finding the order involved fitting successively higher order (saturated) models using least squares regression and finding the model's associated AIC, SIC and HIC values. The order was defined at the lag at which the suitable information criteria values converged within a reasonable range. Two values for the order, one generous the other conservative, were often produced by the different criteria. In all cases I took p to be the generous value. This was because the conservative estimates of order did not give sufficient lag for an

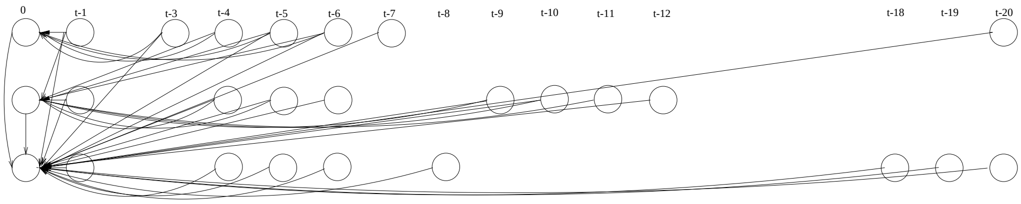


Figure 25: Best model selected for data spaced at 30 minute intervals

interesting causal interpretation to be given and by consistently using the more generous estimate not only were subtle patterns observed in the graphical model, but there was a consistency among models which gave some basis for giving comparative interpretations.

5.4 Results for Data Spaced 30 Minutely

The resulting best DAG for the data spaced 30 minutely is given in figure 25 and corresponds to the system of equations in 28.

$$\begin{aligned}
X_t &= 0.5258X_{t-1} + 0.0054X_{t-3} + 0.0323X_{t-4} + 0.0336X_{t-5} + 0.0403X_{t-6} + 0.0312X_{t-11} \\
Y_t &= 0.1959X_{t-1} + 1.1512Y_{t-1} - 0.0689X_{t-4} - 0.1825Y_{t-4} - 0.0088X_{t-5} + 0.0321Y_{t-5} \\
&\quad + 0.0321Y_{t-5} - 0.0065Y_{t-12} + 0.0083Y_{t-20} \\
Z_t &= 0.005X_t + 0.3809Y_t + 0.0659X_{t-1} - 0.4397Y_{t-1} + 1.1311Z_{t-1} - 0.0327X_{t-3} \\
&\quad + 0.0995Y_{t-4} - 0.2053Z_{t-4} - 0.0091X_{t-5} - 0.0606Y_{t-5} + 0.1128Z_{t-5} - 0.0033X_{t-6} \\
&\quad + 0.0208Y_{t-6} - 0.0315Z_{t-6} - 0.0069X_{t-7} - 0.0207Z_{t-8} + 0.0132Y_{t-10} - 0.0244Y_{t-11} \\
&\quad + 0.0166Y_{t-12} - 0.0054X_{t-16} + 0.0257Z_{t-18} - 0.0296Z_{t-19} - 0.0031X_{t-20} - 0.0031Y_{t-20} \\
&\quad + 0.0146Z_{t-20}
\end{aligned} \tag{28}$$

The chosen "best" model was the contemporaneous causal model (CCM) with all moral links included. In this case the contemporaneous DAG did not correspond to the CIG generated i.e. the CIG did not have an edge joining X_t with Y_t (refer figure 14). Under most circumstances this would imply that the orientation of at least one of the edges is directed upwards, but in this application the causal direction is fixed, hence in selecting this model I had to make the decision that the edge (X_t, Y_t) was omitted due to error,

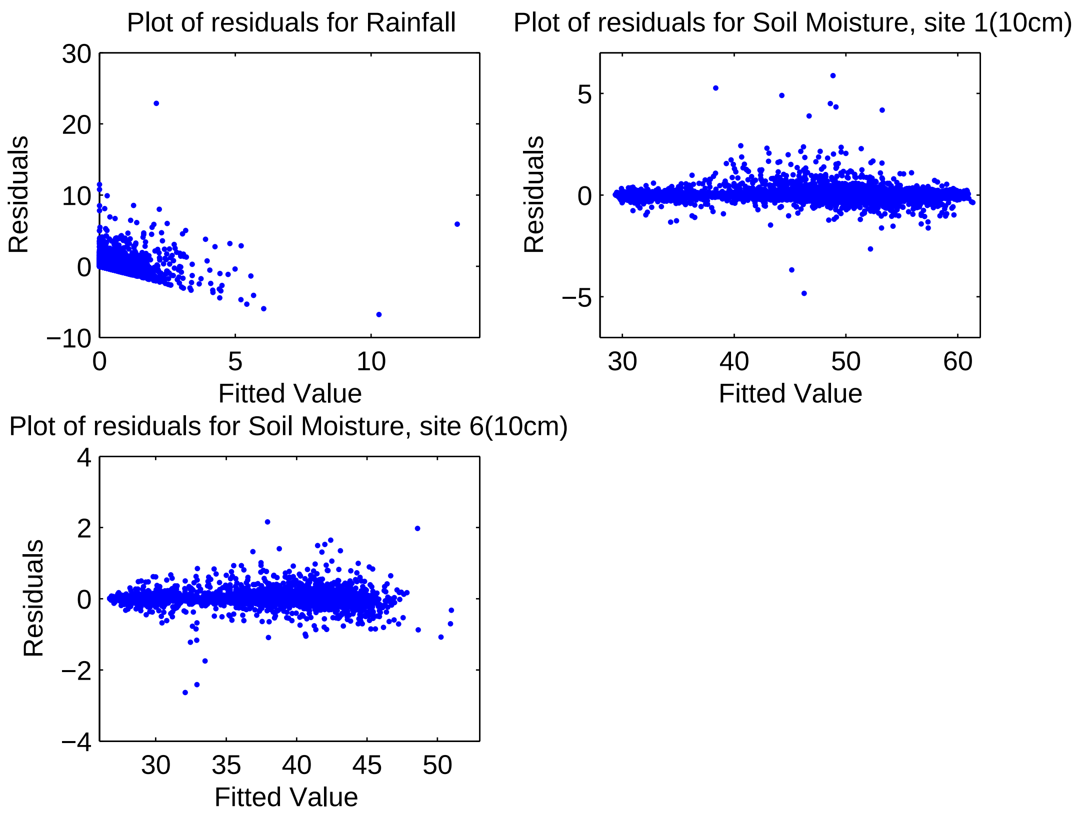


Figure 26: Fitted values plotted against residuals for the data aggregated 30 minutely.

rather than being truly conditionally independent.

This model performed quite a lot better than a traditional saturated model as assessed by all information criteria (table 6). Further, plotting the residuals (figure 26) against fitted values showed that the errors were stable and the absolute errors were quite low (mostly within approximately 3%-6% of the observed data), for both soil moisture plots. The modelling of the rainfall data were not so good, clearly the residuals were not independent and identically distributed. This was not unexpected because the rainfall data contained many zero's and in fact was probably "best" modelled by $X_t = 0$. It is also not of particular concern, a good model of rainfall would be desirable, but in reality the main interest in rainfall was only in its affect on soil moisture levels.

5.4.1 Interpretation

The causal interpretation is that soil moisture levels at the bottom of the hill are contemporaneously affected by both rainfall and soil moisture levels in the top of the hill. Between t and $t-1$ (30 minute lag) the model is *effectively* fully saturated, i.e. all causally plausible links are filled. Between $t-3$ and $t-6$ ($1\frac{1}{2}$, 3 hour lag) the model is nearly saturated. Interestingly, there are no links from variables at $t-2$. From $t-7$ to $t-20$ the behaviour becomes more distinctive. There is a $4\frac{1}{2}-6$ hour lagged effect of soil moisture at the top of the hill and soil moisture at the bottom. At the same lag soil moisture levels at the top of the hill have an affect on their current value. At a lag of $9-10$ hours soil moisture levels at the bottom of the hill affect their current state.

For the most part this interpretation seems very plausible from a causal point of view, with one exception. Does it make sense for rainfall and soil moisture at the bottom of the hill to be causally linked contemporaneously, and not soil moisture levels at the top? My suspicion is that this is probably not the case. In addition, the omission of links at $t-2$ is surprising. A possible interpretation of this is that the observed relationships of variables at t and $t-1$, are non-causal relationships, while the actual causal affects have a lag of $1\frac{1}{2}$ or more.

5.5 Aggregated Models

Hydrological processes (along with other sorts of processes) are likely to display different causal patterns on different scales. In addition, given the problem of initial model saturation it was hoped that data aggregation would create more sparse models.

5.5.1 Data Aggregation

As the rainfall data were collected continuously, aggregation could be done simply by summing successive 30 minute observations. The soil moisture observations however,

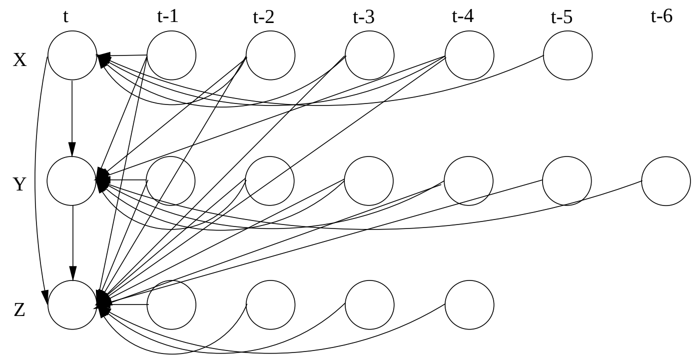


Figure 27: The best model selected by all information criteria for data aggregated hourly.

were sampled minutely but averaged for each 30 minute interval by the recording device. To aggregate these data averages from the relevant number of successive 30 minute observations were calculated. In doing this the form of the data remained the same. To aggregate data `R` code `spacedata` and `CREATEX` appendix B.2,B.1 (written by C Meurk) were used. Models for data aggregated hourly and 2 hourly are presented.

5.5.2 Best Model when Data Aggregated Hourly

$$
\begin{aligned}
X_t &= 0.4174X_{t-1} + 0.0648X_{t-2} + 0.0740X_{t-3} + 0.0326X_{t-4} + 0.0436X_{t-5} \\
Y_t &= 0.0582X_t + 0.1786X_{t-1} + 1.4383Y_{t-1} - 0.0989X_{t-2} - 0.5091Y_{t-2} + 0.1347Y_{t-3} \\
&\quad - 0.0193X_{t-4} - 0.0762Y_{t-4} + 0.0221Y_{t-6} \\
Z_t &= 0.0208X_t + 0.4044Y_t + 0.0472X_{t-1} - 0.5959Y_{t-1} + 1.4016Z_{t-1} - 0.0388X_{t-2} \\
&\quad + 0.2594Y_{t-2} - 0.5028Z_{t-2} - 0.0094X_{t-3} - 0.1089Y_{t-3} + 0.1755Z_{t-3} - 0.0071X_{t-4} \\
&\quad + 0.0626Y_{t-4} - 0.1119Z_{t-4} - 0.0294Y_{t-5}
\end{aligned}
\tag{29}
$$

The chosen “best” model was the CCM with all moral links included (figure 27). The lag has shrunk dramatically to 6 (a lag of 6 hours). This relates strongly to the model in figure 25 which became very sparse at $t - 12$ (a lag of 6 hours). This model is one link away from being fully saturated between t and $t - 4$. Unfortunately, saturated models do not provide too much causal insight.

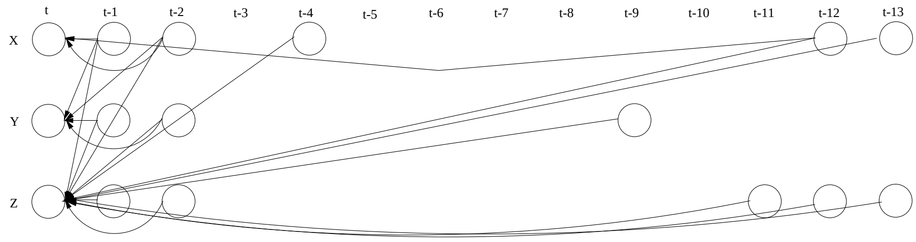


Figure 28: Best Model as selected by all information criteria for data aggregated 2 hourly.

5.5.3 Best Model For Data Aggregated 2 hourly

The model aggregating the data 2 hourly produced a much sparser model, and a model with a larger lag. The model is near saturated from t to $t-2$ (consistent with previous models) but then becomes very sparse. In addition, this model was unusual because unlike most other models, regression removed a large number of links included in the original CIG (figure 29). Arguably, this could be because it is actually a VAR(2) process, but to model it as such provides no causal insight.

This was the only model where the best model was not the CCM. It does not appear that this model provides too much causal insight. At $t-9$ (18 hours) there is a link between soil moisture at the top of the hill and soil moisture at the bottom, this is plausible. Between $t-12$ and $t-13$ (24-26 hours) there is a link between rainfall and soil moisture at the bottom (which seems unlikely given there is no close link to soil moisture at the top) and, at the same point, z_t links with itself, but there is no similar process for y_t with itself. It would be expected that one of the main differences in soil moisture processes would be the time delay between a variable having an affect at one position on the hill over another.

$$
\begin{aligned}
X_t &= 0.4435X_{t-1} + 0.0089X_{t-2} + 0.0393X_{t-12} \\
Y_t &= 0.1642X_{t-1} + 1.3444Y_{t-1} - 0.0696X_{t-2} - 0.3825Y_{t-2} \\
Z_t &= 0.1034X_{t-1} - 0.0377Y_{t-1} + 1.3113Z_{t-1} - 0.0595X_{t-2} + 0.0609Y_{t-2} - 0.3567Z_{t-2} \\
&\quad - 0.0062X_{t-4} + 0.0969Z_{t-11} - 0.0146X_{t-12} + 0.0751Z_{t-12} - 0.0091X_{t-13} - 0.1243Z_{t-13}
\end{aligned} \tag{30}
$$

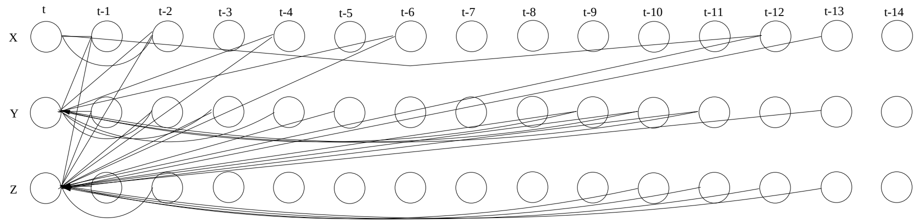


Figure 29: CIG corresponding to the DAG of 2 hourly data. This model was unusual because regression removed a number of links.

Overall, aggregation of the data did not give as much insight into the causal processes of this system as was hoped. However, it did shed light on some of the intricacies of GMTS.

5.6 Models for Different Seasons

So far, modelling has focussed on the data in its entirety. However, it is possible that there is significant seasonal variation. In particular the plot (figure 23) showed a distinctive period of calm behaviour over Autumn 2003 followed by a highly variable period over the following Winter and Spring. Subsets of the data (at 30 minute intervals) which correspond to seasons Autumn and Spring were used to create models of these seasons. These subsets contained approximately 4350 observations each, of data grouped at 30 minute intervals.

5.6.1 Autumn

The chosen model for Autumn was the CCM model with saturated contemporaneous causal links. This model was highly saturated and hence not conducive to a causal interpretation except perhaps to say that the model reflects the stability of the system as seen in the plots.

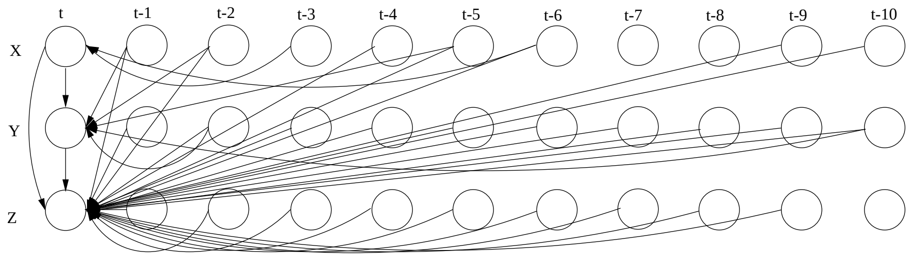


Figure 30: Resulting Best Model DAG for Autumn 2003

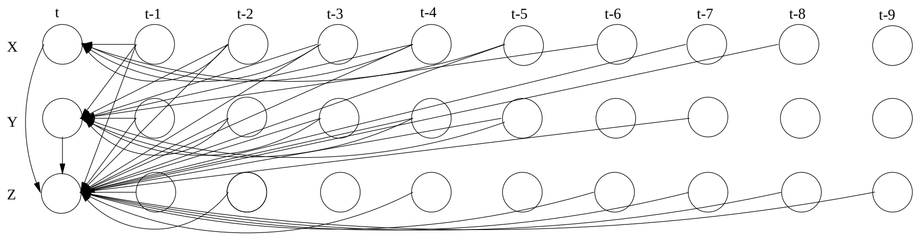


Figure 31: Resulting Best Model DAG for Spring 2003

$$
\begin{aligned}
X_t &= 0.4672X_{t-3} + 0.0584X_{t-6} \\
Y_t &= 0.0330X_t + 0.0410X_{t-1} + 0.00629X_{t-2} + 1.1944Y_{t-2} - 0.0463X_{t-5} - 0.1946Y_{t-10} \\
Z_t &= 0.0082X_t + 0.6669Y_t + 0.0051X_{t-1} - 1.0919Y_{t-1} + 1.8346Z_{t-1} - 0.0057X_{t-2} \\
&\quad + 0.4775Y_{t-2} - 1.2789Z_{t-2} - 0.1245Y_{t-3} + 0.8325Z_{t-3} - 0.0080X_{t-4} + 0.3066Y_{t-4} \\
&\quad - 0.8744Z_{t-4} + 0.0101X_{t-5} - 0.5179Y_{t-5} + 0.9102Z_{t-5} - 0.0080X_{t-6} + 0.5023Y_{t-6} \\
&\quad - 0.7566Z_{t-6} - 0.4037Y_{t-7}0.5268Z_{t-7} + 0.0042X_{t-8} + 0.2266Y_{t-8} - 0.3536Z_{t-8} \\
&\quad + 0.0033X_{t-9} - 0.0412Y_{t-9} + 0.1881Z_{t-9} - 0.0029X_{t-10} - 0.0654Y_{t-10}
\end{aligned}
\tag{31}
$$

5.6.2 Spring

The model selected for spring is also a CCM. This model is virtually saturated from t to $t-5$. Between $t-6$ and $t-9$ Z_t links to itself. Rainfall appears to affect Y_t at a lag of 6 hours while it affects Z_t at a lag of 7-8 hours.

$$
\begin{aligned}
X_t &= 0.4133X_{t-1} + 0.0429X_{t-2} + 0.0742X_{t-3} + 0.0685X_{t-4} + 0.0643X_{t-5} \\
Y_t &= 0.1530X_{t-1} + 1.6098Y_{t-1} + 0.0553X_{t-2} - 0.8633Y_{t-2} - 0.1012X_{t-3} + 0.4111Y_{t-3} \\
&\quad + 0.0089X_{t-4} - 0.2150Y_{t-4} + 0.0573Y_{t-5} \\
Z_t &= -0.0076X_t + 0.4994Y_t + 0.0783X_{t-1} - 0.7693Y_{t-1} + 1.5034Z_{t-1} - 0.0303X_{t-2} \\
&\quad + 0.4262Y_{t-2} - 0.6333Z_{t-2} - 0.0258X_{t-3} - 0.2146Y_{t-3} + 0.1644X_{t-4} - 0.0946Y_{t-4} \\
&\quad - 0.0193X_{t-5} - 0.0650Y_{t-5} - 0.0134X_{t-6} + 0.0644Y_{t-6} - 0.0704Z_{t-6} + 0.0076X_{t-7} \\
&\quad - 0.0333Y_{t-7} + 0.0891Z_{t-7} - 0.0128X_{t-8} - 0.1053Z_{t-8} + 0.0493Z_{t-9}
\end{aligned} \tag{32}
$$

5.6.3 Model Comparison Between Seasons

In environmental and ecological applications often difference is more important than similarity. For example, it is of interest to know what the difference is between the full season model, the model of Spring and the model of Autumn.

Comparing the figures 25,30 and 31, the full season model with the omission of the $t-2$ links is virtually saturated up to $t-6$, similar to the model of Autumn, although the Autumn model includes the links between X_t and Y_t. The model for Spring is also virtually saturated up to about $t-5$. The exception is the links between Z_t and its lagged variables. In this way the Spring model appears more closely related to the full season model. In addition, the contemporaneous subgraph is the same for both Spring and the full season model, which is unusual because both these subgraph DAGs resulted from a similar problem of CIG configuration (refer section 5.4).

After $t-6$, all models become very sparse quite quickly. In the full season model there is a progression from Y affecting Z to Y affecting itself to Z affecting itself. The Autumn model shows that from about $t-5$ Y has a continued affect on Z, followed by X with Z. Throughout this model Z affects itself. This model is half the order of the full season model with an order of 10. Again, Spring shows a similar behaviour to the full season model displaying a slight progression of effect. For example, at $t-5$ and $t-7$ Y affects Z and at $t-7$ and $t-8$ X affects Z. The model of Spring displays the lowest order of 9. Viewing the dataset in light of the subtle similarities the model for Spring has with the full season model it was noted that the full dataset spanned 18 months or 6 seasonal changes. Of these 6 seasons Spring and Summer were studied twice. It may be possible that this observed similarity is due to a bias towards Spring and Summer. It would be interesting to model all the seasons in light of this and look at the model progression that occurs, unfortunately, as with this assessment it would be highly subjective. In order to undertake a study, which could be of great use to the hydrologists, a way of measuring the difference or distance between models would be required.

5.7 Causal Relevance of These Models

Interestingly, in all but one case (the 2 hour model) the saturated CCM was selected as the best model. Is this because hydrological processes are contemporaneously linked, or could there be another reason? Through this case study I came to an interesting conclusion about the place of moralisation and the bias that arises with testing using the information criteria.

5.7.1 Moralisation in Practise

In chapter 3 the best model selected including all possible moral links had an SIC value of -4502.4. After performing subset selection this had "improved" to -4515.9 and a total of 4 moral edges had been removed. In the hydrological case the saturated models were almost always favoured, there were no possible moral edges whose removal caused an improvement. From a practical point of view, it appears that, in these cases at least, simply adding a temporal arrow and deciding on contemporaneous configurations from a CIG model represents a close to optimal fit/parsimony trade-off. For the small improvement that the removal of moral links may provide is does not really seem worth the effort to

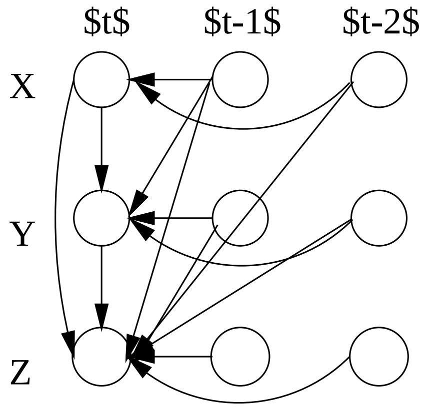


Figure 32: An example of an effectively saturated model to lag 2, possible moral links are given by the dashed lines

carry out extensive subset selection.

In fact, in these models, where moral edges are known in the graphical model is not where possible edge removal might help. For example, if we refer to the subgraph in figure 32 the list of possible moral links are:(X_t, Y_t),(X_t, X_{t-1}),(Y_t, Y_{t-1}),(X_t, X_{t-2}) and (Y_t, Y_{t-2}).

But what about Z? Z is at the bottom of the causal chain, and as a result there are no possible moral links. This is because any value of Z_{t-k} can only be a parent to one node, Z_t, and as shown in section 4.3.3, for a moral edge to exist, a lagged variable must be the parent of two joined contemporaneous variable as in figure 18.

Referring to the equations for the best models given in this chapter each equation for Z is very lengthy. Surely some of these edges could be removed, however, unlike X and Y there is no theoretical justification for doing so. Practically, edges which feed into Z_t could be removed if they caused model improvement. However, this would weaken the causal appropriacy of the model. In addition, in a case like this where there is a well defined causal direction, the removal of moral edges amounts to the removal of serial links in the X and Y direction only, e.g. X_t, X_{t-k} and Y_t, Y_{t-k} links, and this systematic removal of a certain type of edge, undermines the validity of moralisation because it

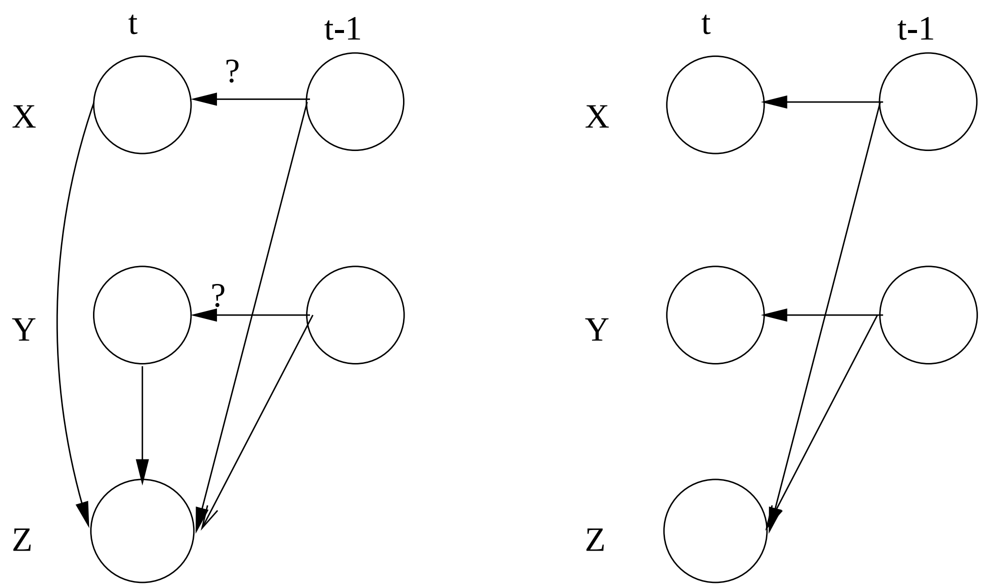


Figure 33: A demonstration of the possible causal bias due to moralisation

introduces a bias towards the selection of certain sorts of relations over others. Hence, in this case the principal of moralisation has not helped in the removal of edges where needed i.e. in the case of Z_t, and further, it is doubtful the extent to which moral links should be removed as an edge selection bias has been introduced.

5.7.2 The Contemporaneous Boundary

Theoretically, there is another problem which concerns the combination of contemporaneous causality, moral edges and the use of information criteria to assess alternative causal models. Looking solely at the contemporaneous part of figure 32 the rule of demoralisation only allows for the potential removal of one edge - the edge from (X_t, Y_t). It is conceivable however, that contemporaneous causality is not present. As Granger states, the decision whether to include contemporaneous links or not "...cannot be achieved by purely statistical means" (Granger,C.,1988,208). With GMTS it appears that, if it is known that contemporaneous causality is definitely present or definitely not present, then there is not too much problem in concluding the best model selected is the best causal model. However, if we are unsure whether contemporaneous causality is present it appears the contemporaneous causal model will be favoured by the information criteria. Figure 33 describes the possible bias of information criteria to select contemporaneous causal models as “best” models. This diagram indicates why a model with saturated contemporaneous causal links will be favoured, it has the maximum “choice” of links for explanatory power, and the maximum ability to remove them to increase parsimony. In contrast, the model on the right does not have this choice and hence all edges must be included. Therefore, regardless of causal superiority there is a potential bias of the information criteria to accept the CCM. For example, the link between X_t and X_{t-1} in figure 33 is dependent on the existence of an edge between X_t and Z_t, a contemporaneous edge. If this edge does not exist, then the link X_t and X_{t-1} must be real. In this case, the same relationship will occur with Y_{t-1} and Z_t. Supposing that this model was a VAR(1) process, it may not make too much difference, because the inclusion of only 2 possible edges are at stake, however, referring to figure 32 which goes to lag 2, contemporaneous saturation allows for maximum of 6 possible moral edges (not 7, if the edge between X_t and Y_t is removed, then the possible moral links Y_t, X_{t-k} will become real), and as the order increases this trend will continue.

5.8 A SINful Alternative

In many of these cases causal interpretations of the model could not be given, or were limited, because models were highly saturated. Hence, a method which removed more edges and hence produce sparser graphs was desired. In 2004 Drton and Perlman proposed a method for creating undirected graphs UG. This method takes a covariance or correlation matrix and finds the partial correlation matrix and then uses Fisher’s z-transform and Sidak’s correlation inequality (1967) and develops a simultaneous testing procedure which aims to reduce errors and produce sparser models. The term SINful refers to the fact that this method separates edges into Signficant, Indeterminate and Non-significant edges, in quite a clear manner. This method has not been used in time series analysis. Given the problem of model saturation in these case studies it seemed worthwhile to try and apply this method to VAR modelling.

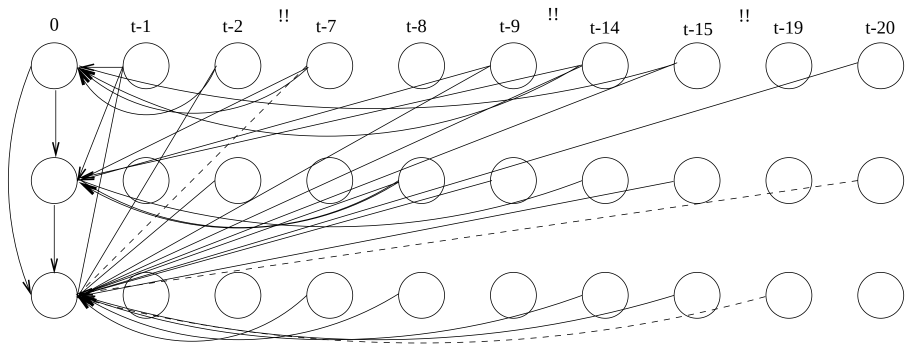


Figure 34: A SINful model. The dashed lines refer to the Indeterminate edges.

The data used was the full dataset spaced at 30 minute intervals hence this model can be compared to the model derived in figure 25. Drton and Perlman have written a suite of programmes available in the latest release (2.1.0) of R (Drton,M., 1994). The appropriate procedure was modified so that the output contained only the relationships between current variables, and current with lagged variables. In non-temporal models, SINful will produce a graph with relationships between all pairs of variables which is undesirable for time series analysis.

The resulting model in figure 34 had an AIC value of -164980 (other information criteria values were consistent with this) compared to the GMTS model for 30 minute data which had an AIC of -272840. The SINful model has removed more edges however, this does not correspond with a more interpretable model. In addition, much explanatory power has been lost as is evident in the AIC value and the error plots 35. In fact, the AIC for the saturated model was -238790 in light of this it was concluded that the SINful approach

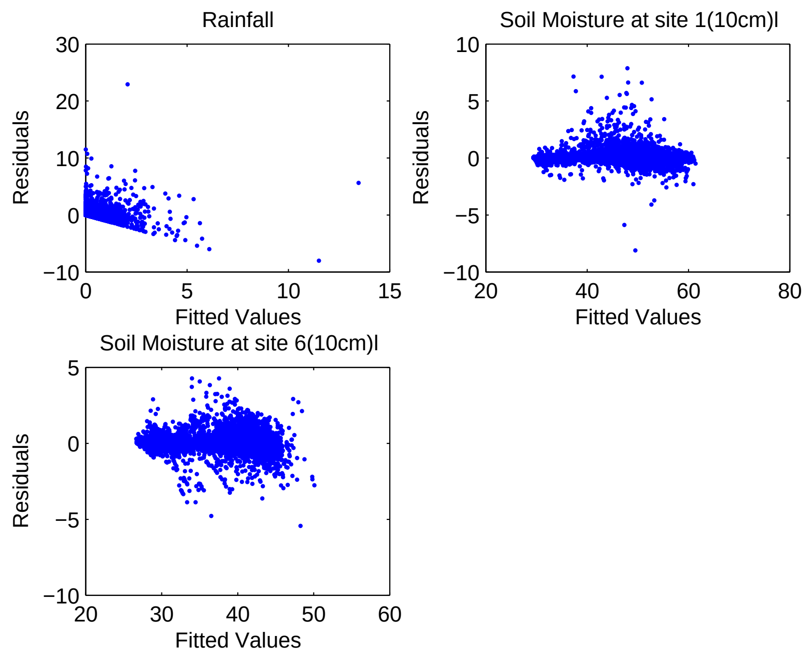


Figure 35: Plot of residuals for SINful model of 30 minute data

has not provided a viable model in this case.

$$
\begin{aligned}
X_t &= 0.53X_{t-1} + 0.0555X_{t-2} + 0.0622X_{t-7} + 0.0184X_{t-14} + 0.0198X_{t-15} \\
Y_t &= 0.0245X_t + 0.3065X_{t-1} + 1.1298Y_{t-2} - 0.055X_{t-7} - 0.1340Y_{t-7} - 0.055X_{t-8} \\
&\quad - 0.134Y_{t-8} + 0.0141Y_{t-14} \\
Z_t &= 0.0401X_t + 0.4462Y_t + 0.0708X_{t-1} + 0.1266X_{t-1} - 0.0572Y_{t-1} + 1.0776Z_{t-7} \\
&\quad - 0.5819Y_{t-8} - 0.0648Z_{t-8} - 0.1185X_{t-9} + 0.1619X_{t-14} - 0.0289X_{t-12} + 0.1107Y_{t-13} \\
&\quad - 0.1690Z_{t-14} - 0.0154X_{t-15} - 0.0375Y_{t-16} + 0.1187Z_{t-19} - 0.0453X_{t-20} - 0.0292Y_{t-20}
\end{aligned}
\tag{33}
$$

5.9 Conclusion and Problems for Future Research

This case study highlighted some unexpected behaviours in the applying the principle of moralisation and further research needs to be undertaken to establish under what conditions (if any) it provides a significant improvement in modelling. In addition, in cases where the removal of moral links can improve the model it is apparent that contemporaneous causal models will be preferred over non contemporaneous causal models, by information criteria methods. Hence, if we wish to comparatively assess these types of models there is a need for the development of a causal test.

Data aggregation did not provide causal information as had been hoped, this is largely due to the fact that models were highly saturated over a short lag. The seasonal models were interesting because it appeared that the model for Spring and the full season data were "closer" than they were to the Autumn model. This analysis is supported by the plots in figure 23. Analysis of the difference between models is very important in environmental and ecological applications and hence the development of a metric to measure the "distance" between alternative models in a way which provides insight into what these differences are would be a very worthwhile study.

It does appear that GMTS satisfies the requirement of G-Causality, that is, it takes a sensibly selected subset, a CIG, and assesses it with information criteria, with a methodical way of removing edges (by moralisation). Further, in cases where the existence of contemporaneous causal links is known one way or the other GMTS does appear causally appropriate. However, when contemporaneous causal models are being compared with non-contemporaneous causal models this modelling strategy appears only to be causally sensitive for reasons outlined above, perhaps Granger was right.

Finally, in the case of the 30 minute data the SINful approach was adapted to time series. In this case it created a model which was not favourable compared with the GMTS model and the traditional modelling approach. It appears that, despite being a procedure which reduces selection errors, it has provided too strict a criteria and important explanatory links have been removed. It is likely that this criteria could work with other datasets, however, given that it has removed a number of important explanatory variables in this case, it is unlikely the resulting model could be considered causal.

6 Case Study III: Ecological Data

> Statistical inference is particularly difficult in ecology. Statistical populations are not biological populations ... biological populations and communities change in space and time in such a complex manner that if we specify a statistical population very broadly, we cannot sample it in a random manner. This fundamental problem undercuts the very foundation of normal statistical inference in ecology.
>
> -Charles J Krebs, Ecological Statistician

The data for this case study was kindly provided by Mike Fitzgerald who, along with Brian Karl, was responsible for collecting these data for nearly 23 years, while working for DSIR/Landcare Research.

6.1 The Causal Question

These data poses an interesting causal question, does beech seed masting cause substantial increase in mouse populations? This view was proposed following similar studies in the Northern Hemisphere starting in the 1970's. More recently, this "traditional view" has been challenged by the view that the observed relationship between seed masting and mouse numbers is not the causal relationship, that in fact, flowerfall and seedfall may provide a food source for invertebrates, and it is the rise in invertebrate numbers which cause mouse population growth (Fitzgerald, M., *et al.*, 2004, and Choquenot, D., *et al.*,2000). Hence, this dataset appeared to be an interesting case to study using GMTS.

6.2 About the Data

These data were collected in the Orongorongo valley near Wellington between August 1971 and November 1996. From the data provided, three variables, suitable to test the ecological research question, were chosen. These were: mice numbers, beech seedfall and mouse breeding numbers.

Mice Numbers

Mice were trapped in snap-traps which had metal-covers to exclude possums. In total, there were 116 trap sites set up at 50m intervals, throughout hard beech forest. Trapping took place over 3 consecutive nights, four times a year (February, May, August and November). Between 1971 and 1993 this was carried out by Fitzgerald and Karl. From 1994 to 1996 data were collected by Alley and Berben. However, when the data were analysed it was noticed that the data had changed significantly from one pair of collectors to the next, and it is unlikely this is due to differing conditions, but simply different styles of collection. Hence, analysis excluded the final 2 years, and only analysed the data up until the end of 1993 (Fitzgerald, M., 2004, pers comm).

Once trapped each mouse was counted and autopsied. The number of mice captured was then transformed into a density index, and it is these values used in this analysis. The formula for density is given as:

$$N_t = \frac{-100a_2}{(a_1 + a_2)} \log \frac{a_0}{T}, \tag{34}$$

(Fitzgerald, *et al.*,2004,170) where, N_t is the population density at time t, a_0 is the number of traps not sprung, a_1 is the number of traps sprung but empty, or containing another species and a_2 is the number of traps containing a mouse. $T = a_0 + a_1 + a_2$. This estimate had a standard error given as:

$$SE(N_t) = 100\sqrt{\frac{a_1 a_2[\log \frac{a_0}{T}]^2}{(T - a_0)^3} + \frac{a_2^2}{a_0 T(T - a_0)}} \tag{35}$$

(ibid). This index was found assuming a Poisson distribution of trap rates and that trappability was constant for all ages and sexes. The authors acknowledge that these assumptions may have caused some bias. In total there are 90 mouse density measurements.

Beech Seed

Beech seedfall, along with other flower and leaf litter, were measured using traps left under mature trees with an area of 0.28m^2. Traps were cleared monthly, but data were reported annually. From 1971 three traps were used, but this was increased to 21 in 1974. The mean of the initial three traps appeared representative of subsequent data from all 21, and data from these three traps were used to provide a density estimate seedsm^{-2}. Seedfall occurred from February through until May. In total there are 24 seedfall observations.

Mouse breeding

Breeding mouse data were given both as a number of pregnant females and as a percentage of adult females. Adulthood was assessed based on Lidicker's toothwear classes. A mouse with toothwear class above 3 i.e. greater than 4 months old was considered an adult. A mouse was considered pregnant if the autopsy found live or resorbing embryos in utero. There are 23 observations for breeding, although the last was discarded because it recorded an incomplete breeding cycle. Breeding took place in the Orongorongo Valley in both Spring and Summer but only one value was reported annually. The data used for the subsequent analysis was the percentage of breeding mice because it gave a very different, but I believe more accurate, description of the process. For example, in the November 1971-August 1972 observation 50 breeding mice were recorded, this corresponded to 4.0% of the adult female population. In 1974-1975 year 16 breeding mice were recorded corresponding to 56.2% of the adult female population.

6.3 The Problem of Ecological Data

The dataset presented in this case study is not atypical of ecological datasets. Ecological data can be expensive to collect, and hence there can be fewer observations than might be ideal. The processes under study can be slow to develop and processes may have different periodicity which is an important part of the data and needs to be preserved. In this case while one series, X_t (mouse numbers), persist over each season, other dependent processes such as mouse breeding (Z_t) occurs over only two seasons, with breeding being 0 (or not significantly different from 0) for the other two seasons. The two variables (Y_t and Z_t) are approximated by Poisson process and hence is not conducive to finding a set of equations of the form $[X_t, Y_t, Z_t]^T$, where each variable is modelled at the same time period. Beech seed fall, the main causal interest of this study only occurs once per year AND mast years, years of excessive beech flowering and seedfall (which ideally we would like to be able to model separately and compare with non-mast years) only occur every 4-7 years. This means that over the 23 years of study, a maximum of about 6 mast year observations are possible. In cases like these, all the research funding in the world cannot provide more data in any hurry. In addition, over such a long period of time theories change. In this case the "seed mast" view was challenged by the "invertebrate" view but by then it was too late to get a corresponding invertebrate series of sufficient length to test.

Furthermore, in this dataset (as with many others) it is not possible to go out into the field, and find out exactly how many mice there are, especially if each mouse needs to be autopsied first. Hence, ways of estimating the population must be devised, and this process is susceptible to bias.

6.4 So What Can We Do?

The easy answer would be to walk away from this dataset in disgust due to the fact that it does not really display the properties that time series analysis requires. However, one cannot always choose the data one gets to analyse, and, one way or another the scientist who spent their time and effort collecting the data will want some answers.

Before presenting the method used in this case it pays to mention the initial analysis carried out. To begin, it was intended that ordinary GMTS analysis take place using the three series mouse numbers X_t, seedfall Y_t and mouse breeding Z_t. It quickly became apparent that this was highly problematic. In order to carry out the ordinary GMTS method it was proposed that, for the two variables with only one value per year, i.e. the breeding and seedfall data, we generate data from the observation according to an appropriate distribution to simulate values for seasons within each year, and hence remove the zeros. However, after various attempts, results did not provide any useful output. At the same time I became more convinced that the 0 entries were justifiably 0. For example, 0 seedfall in July is a real value because seed does not fall in July. In addition I came to realisation that the loss of seasonal information which occurred when applying the original GMTS approach was a large drawback in this case. That is, a model where a present value is given in terms of the past value for all time periods looks to find the unchanging relationship between past and present variables. In ecological data, we are often interested in looking at the changing relationships, in this case the seasonal change, and in particular the comparison with extreme events such as seed masting.

6.5 Modelling with GMTS

There are a number of reasons GMTS modelling will be problematic in this application.

- Due to the staggering of activity periods each series has, e.g. mouse breeding occurs from November to February and seedfall occurs in February, the model will have to investigate, and deal with, the relationship among lagged variables. This is because whereas in ordinary time series modelling the present is modelled in terms of the past, in this case there are relationships between different past variables which are essential to preserve. For example, in figure 36 seedfall (Y_t) and mouse breeding (Z_t), display a relationship between lagged variables Y_{t-1} and Z_{t-2} which is of equal importance to the relationship mouse numbers have with themselves X_t and X_{t-1}.

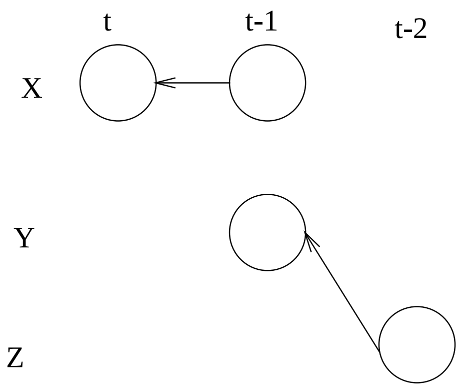


Figure 36: This figure represents 3 processes which occur at different periods. The edge between X_t and X_{t-1} is the sort of link GMTS is concerned with. However in this case the link between Y_{t-1} and Z_{t-2} is also important.

- Although aggregating yearly would allow a conventional time series approach to be applied, the results of this would be uninformative because (a) seasonal effects would be lost and (b) what was a small number of observations would become even smaller, for not much gain in informativeness.

Nevertheless, there are other reasons for wanting to carry out this procedure.

- Graphical modelling of this process, and in particular, a causally sensitive GM approach could potentially be very useful given the causal question being discussed.

- Bad time series or not, this is still a time series and so a model which preserves the "arrow of time" is still preferable. However, seasonality is a VERY important property to preserve. The model would be deeply flawed if it gave a model of $X_t = \alpha_1 X_{t-1} ... \alpha_k X_{t-k}$, for example where the difference between time t and $t-1$ could not be given a seasonal interpretation.

6.6 A Possible Modification to the GMTS approach

It was clear that GMTS would require some manipulation. First, to preserve seasonality, the matrix G of lagged variables was created, so that each lag denoted a season. For example, the vector $X = [x_t, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4} ... x_{t-k}]^T$, where x_t was an observation of

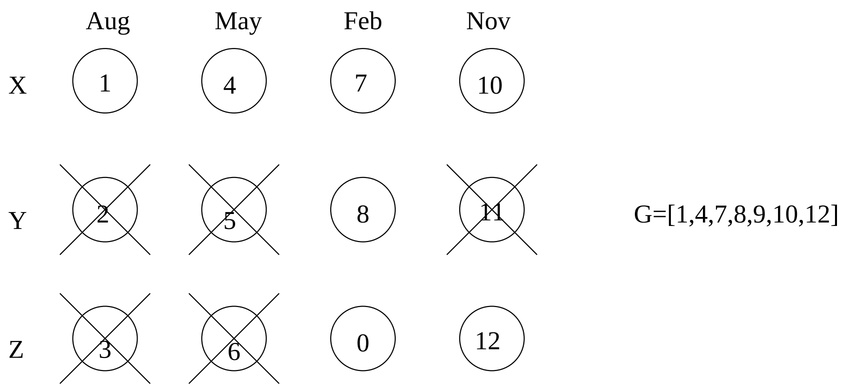


Figure 37: Schematic diagram of the construction of the matrix G to remove 0 but preserve the right direction of time

each of the three variables from August 1993 and x_{t-k} was an observation from November 1970, was split into 4 vectors $Xaug = [x_t, x_{t-4}...]^T, Xmay = [x_{t-1}, x_{t-5}...]^T$, $Xfeb = [x_{t-2}, x_{t-6}...]^T, Xnov = [x_{t-3}, x_{t-7}...]^T$. These were compiled so that $G = [Xaug, Xmay, Xfeb, Xnov]$ and the CIG found as usual. Wherever there was a row/column of 0's, the partial autocorrelation matrix would output complete rows and columns of `NaN` (Not a Number) as a result of division by 0. This was unavoidable, however it is straightforward to assume that when beech seed does not fall it has no effect on other variables, and similarly when mice do not breed mouse breeding has no effect on other variables, hence all 0 rows/columns were removed and the process proceeds. Figure 37 shows schematically how the matrix `G` is modified to create a new `G` with 0 rows and columns removed.

Second, this process was repeated with different *leading season's* until four graphical models are produced. This is done ensuring that the model is temporally faithful. For example, when the leading season changes from August to May, in our example x_t is dropped from the top of the vector and x_{t-k-1} is added at the bottom. Doing this, as opposed to aggregating the whole season's values and reverting to an ordinary regression, loses a couple of degrees of freedom, but this is unavoidable if a temporal interpretation is to be preserved.

In the analysis which follows the only data points imputed for this modelling were that of the breeding percentage. This is because only one value was given to cover two time periods, even though breeding were possible in both. This differs from the case of seedfall where one yearly value referred to one seasons activity. The breeding data needed to be “split” between the two seasons because there was no information to more accurately model breeding differences between seasons. Initially, this was done simply by dividing each observation equally between the two seasons. However, this meant that there were linearly dependent rows and hence the CIG could not be calculated. To overcome this problem normal random variables were generated by `MATLAB`. This was done by sampling repeatedly and taking means from a uniform distribution with mean 0.5 and range (0,1). Hence, for each observation pN the spring value was defined as $pN * rand$ and the summer value defined as $pN * (1 - rand)$.

The remainder of the process does not differ from the ordinary GMTS method and the resulting model from the above analysis is given in the top left of figure 38.

A model was generated for each permutation of seasons and these the remaining models as well as the combined model are given in figure 38.

6.7 Results

Regression coefficients were fitted using the `R` procedure `lm`. Care must be taken in using this procedure for time series because `lm` does not always maintain the order of variables, especially in the residual plots. The subsequent graphs were checked thoroughly to make sure this had not occurred.

The models presented in figure 38 are interesting because there is very little interaction between variables. With the omission of the link from February to May the mouse

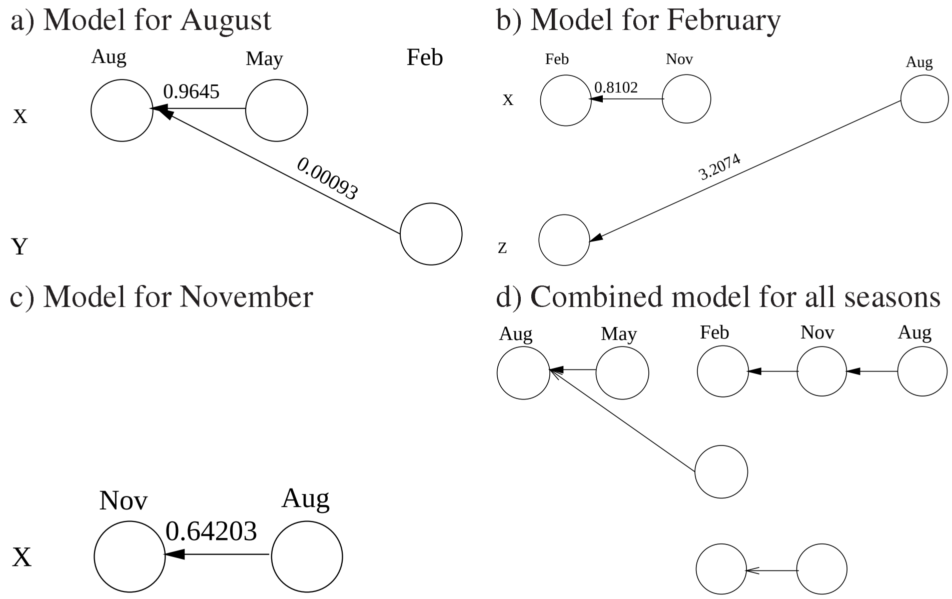


Figure 38: Graphical models for each different permutation of seasons. Note that there is no model for May

population displays a Markov property, i.e. each node in the X direction is linked with the node of its previous variable. For example, X_{aug} is linked to X_{may} and X_{feb} is linked to X_{nov}. The regression coefficients for the X series does not appear to imply growth, which is unusual. There is a surprising omission of a link between breeding and mouse numbers. The relationship breeding has to itself, is probably an artefact of the way in which the data were split. On discovering this, I reverted to modelling this as one event but this did not create any interaction, and hence the above model was retained, without the inclusion of the regression coefficient (which was about 1). Under this model, there does appear to be a causal link between beech seedfall and mouse numbers. In addition there is a distant effect by mouse numbers on breeding. It maybe more appropriate (in light of the way breeding data was imputed), to shift the link from joining mouse numbers in August and mouse breeding the following February, so that it joins mouse breeding in November instead.

The errors in figures 39, 40,41 and 42 are not particularly stable and display a divergence in errors. The Cook's distance plot indicates that mast years have very high influence on

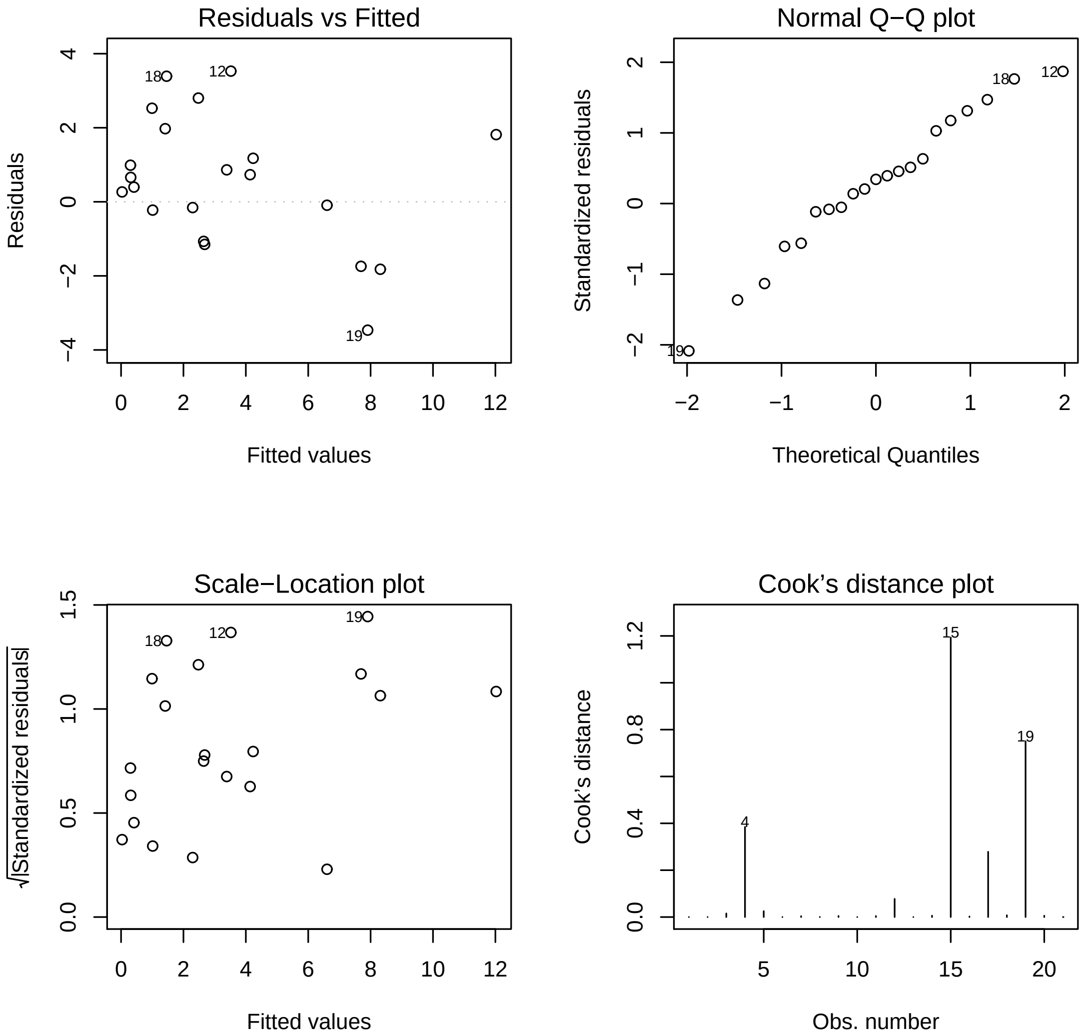


Figure 39: Errors for August model

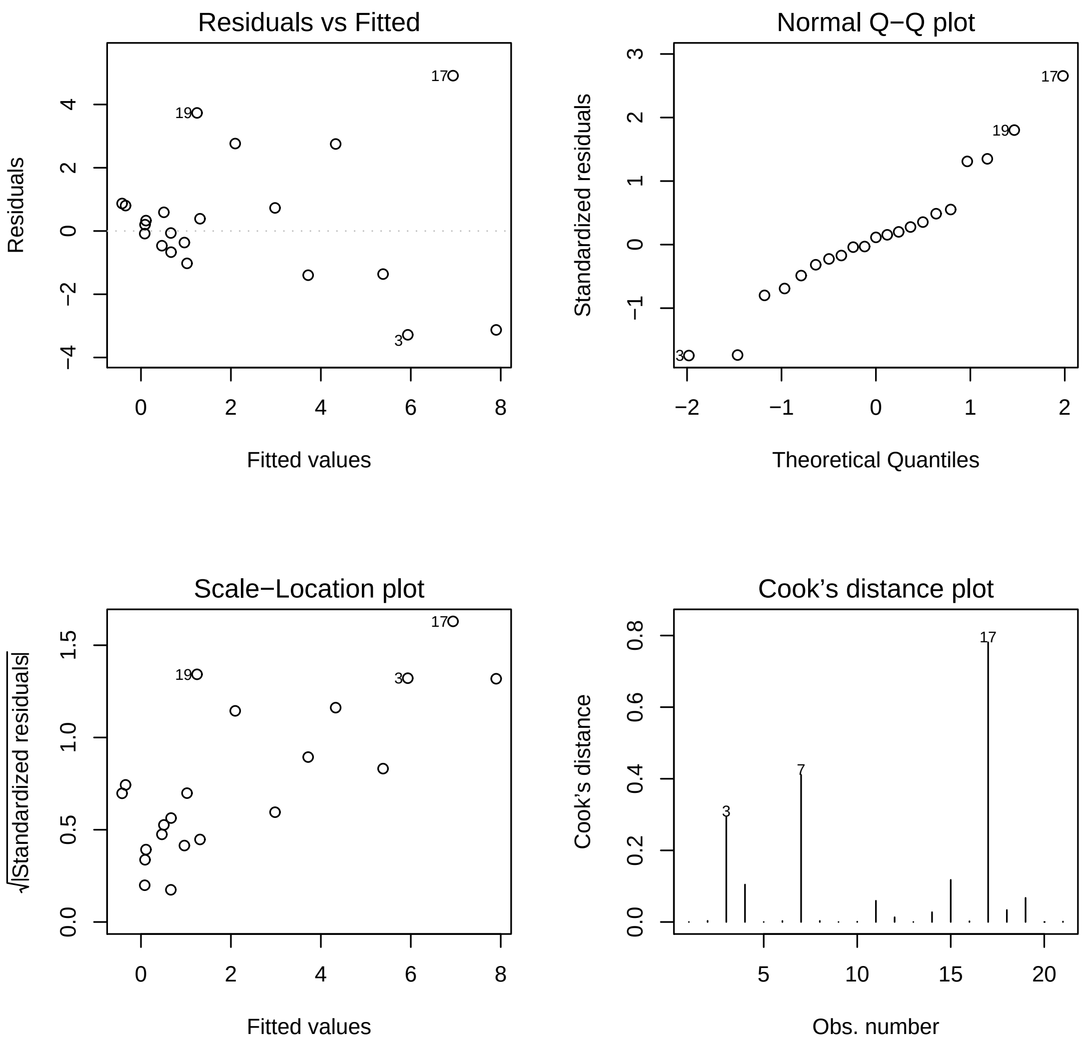


Figure 40: Errors for February,X

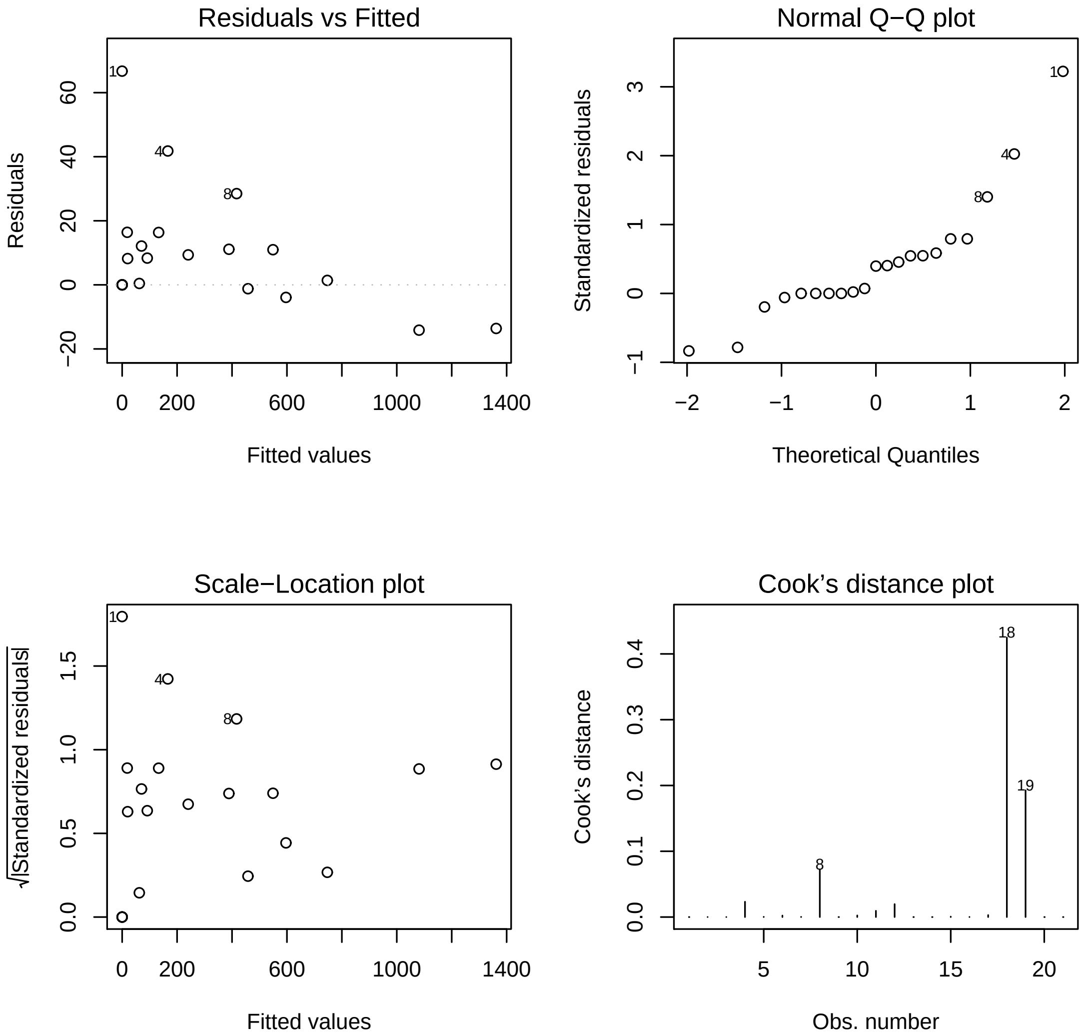


Figure 41: Errors for Feb,Y

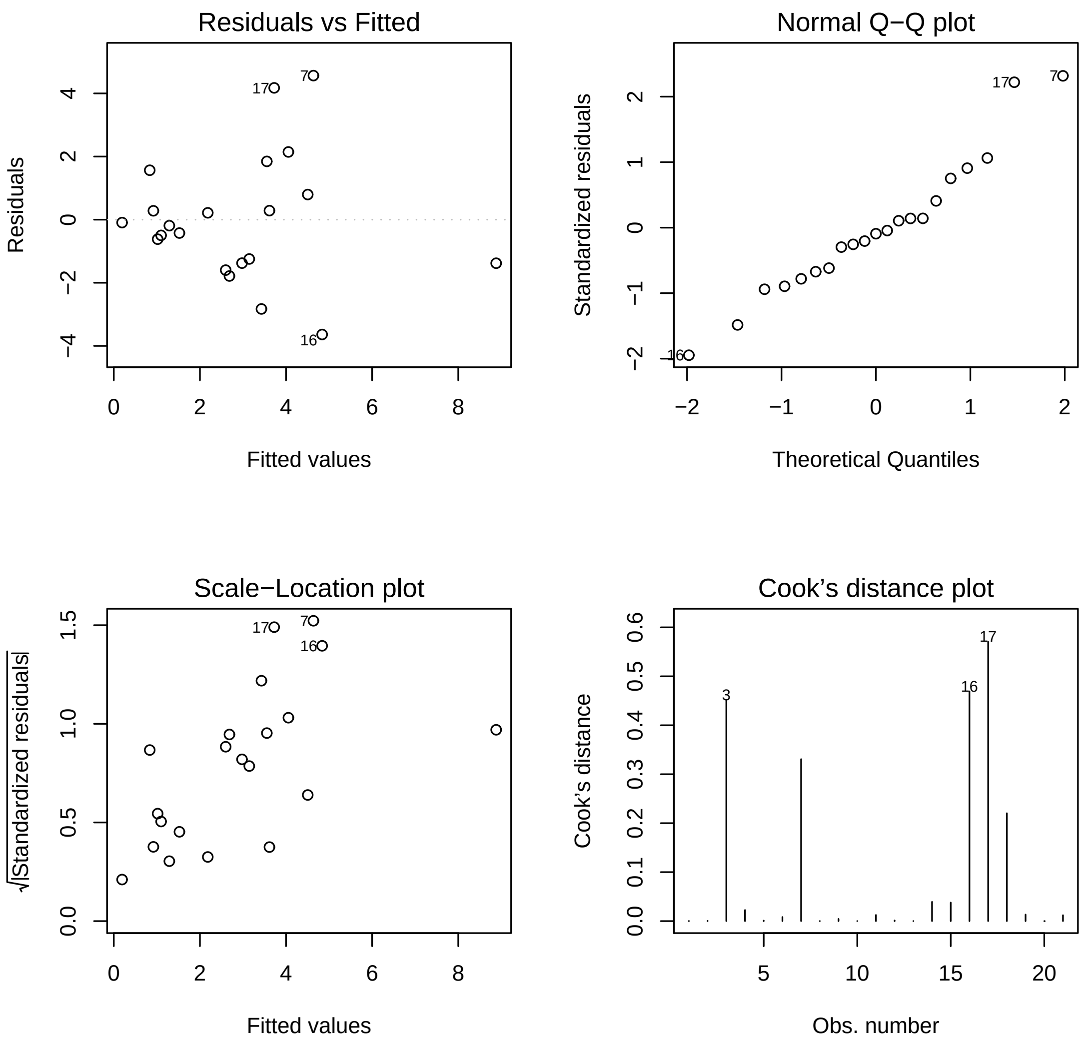


Figure 42: Errors for February November

the model. Unfortunately, to remove these from the model would leave a dataset which is already too small, unuseable. It would be desirable to model the mast years separately, however this really is impossible with the dataset as it is.

6.8 Conclusion

Modelling these data using GMTS was not an overwhelming success. However, it would be interesting to further investigate this type of method in further non-time series applications.

While the graphical model did not give too much new insight and omitted some unusual links, like a link from mouse breeding to mouse numbers, it was basically sensible, i.e. it reflected what we know to be the true system and this is encouraging.

In regards to this particular case it would be interesting to retry the approach stated after some careful simulation of further data. This would probably require access to the raw data.

7 Summary

In this thesis I have had two main tasks, to discuss causation in statistics and in particular GMTS as a causal model as well as the practical application of GMTS methodology in environmental and ecological datasets.

Causality

- The development of statistical methods for finding causal links is different from a philosophical approach to causality and it is important that this distinction is understood if progress in this discipline is to be made and statistical methods adopted. In short, statistical treatments of causality need to focus on *finding* causal links rather than trying to *define* what causation is.

- The definition of G-causality provides a causally sensitive approach to modelling, so long as the set of input parameters are selected in a manner which ensures causal links are not excluded, such as is the case with conditional independence.

- GMTS provides a practical and causally sensitive approach to modelling and is consistent with the notion of Granger Causality. However, as was discovered in case study II it appears that there may be some bias in the selection of structural VAR models.

- Further research is required in order to discover if moralisation is really as important in practise as the theory suggests, i.e. in what type of models will the removal of moral edges produce significant improvement in a model with respect to information criteria. In addition, theory needs to be developed to govern the principle of moralisation, when in conjunction with a certain causal direction it entails the removal of a particular type of edge, for example the removal of serial links.

- Ultimately, there is a need for the development of a causal test, similar to the information criteria tests, which will not be biased towards contemporaneous models in the way that the information criteria are.

Application to Ecology and the Environment

- In practice, GMTS improved over traditional modelling in case study I and all cases in case study II.

- Much of the potential hassle, which has occurred when manually specifying alternative DAGs from the CIG, is removed as it has been shown that this process can be automated.

- Even in the most unlikely scenario, as with case study III, GMTS was still able to generate a model which, although flawed in some ways looked basically sensible. Further research into GMTS type strategies for linear models and generalized linear models would be a worthwhile avenue to explore, especially in this case if careful data generation can be carried out to compare mast years with non mast years.

- In environmental and ecological applications change can be more important than similarity, and, a measure of change in processes as seasons progress etc. would be of benefit to scientists in these disciplines. Hence, the development of a metric which can provide a measure of change between seasons is a worthwhile task for future study.

- Although it appears that CIGs do not provide an optimal set of causal variables, i.e. they contain some links which may not be causal, it does seem that it is nearly optimal. As already mentioned CIG based models performed better than the saturated models, but also, in case study II where the SINful approach was adapted a CIG based model prevailed as the best model.

A List of MATLAB Code Used

A.1 cigts

```
function [R,crit,Rsig,IRsig]=cigts(X,lags,tv)
%CIGTS Conditional independence graph for time series models
%  Ver 6.1 (April, 2001)
%
%  [R,crit,Rsig,IRsig]=cigts(X,lags,tv) computes the conditional
%    independence graph (CIG) for vector autoregressive
%    time series models (VAR).
%
%  You need to specify:
%    X - the matrix containing the time series as column vectors;
%    Example: if you wish to analyze the CIG between the time
%    series Y, W, Z and their lagged series, then your X is
%    X=[Y W Z];
%
%    lags is the number of lags you want to consider in the model;
%    Example: if you set lags equal to 1, cigts will analyse the
%      conditional independence or partial correlation between
%      Y(t), W(t), Z(t), Y(t-1), W(t-1) and Z(t-1);
%
%    tv is the t-value corresponding to the alpha level
%    of probability;
%    Example: if you wish alpha=0.05 then your t-value (for
%    large samples) is 1.96.
%
%  CIGTS returns:
```

```
%    R - the "conditional independence/partial correlation" matrix
%      (R) between the current and lagged time series;
%      Example: if X=[Y W Z] and lags=1, R is the
%      "conditional independence/partial correlation"
%      matrix between the variables Y(t), W(t), Z(t), Y(t-1),
%      W(t-1) and Z(t-1);
%
%    crit - the critical value to reject the null hypothesis of
%      independence/incorrelation, calculated according
%      the statistics exposed by Reale and Tunnicliffe Wilson.
%
%    Rsig - is a "conditional independence/partial correlation"
%       matrix where the significant partial correlations are
%      indicated as ones and the non signicant ones as zeroes.
%
%    IRsig - is equivalent to Rsig but with the variables
%      and the lags indicated on the left of the partial
%      correlations to facilitate the reading.
%
%    With this procedure only the relations with current variables
%    are tested.
%
%Reference: Reale and Tunnicliffe Wilson, Journal of the Italian
%           Statistical Society (2001)
[rows cols]=size(X);


for i=1:lags+1
  G(1:rows-lags,i+(cols-1)*(i-1):i*cols)=X(lags+2-i:rows+1-i,:);
```

```
end

[nv m]=size(G);

if nv-(m-1)<30
    warning('You have less than 30 degrees of freedom')
end

if nv-(m-1)<0
    warning('You have not enough observations: ...
    reduce the number of lags')
    break
end

V=corrcoef(G); W=inv(V); S=sqrt(diag(W)); R1=W./(S*S');
R=2*eye(m)-R1; crit=tv/sqrt(tv^2+nv-(m-1)); Asig=abs(R)>crit;
P1=ones(length(V)); P2=triu(P1); P3=P2-tril(P1).*triu(P1);
nodisplay=sym('x'); P4=P3*nodisplay; Bsig=tril(Asig)+P4;
Rsig=Bsig(:,1:cols); Vars=(1:cols)'*ones(1,lags+1);
FVars=reshape(Vars,(lags+1)*cols,1); Ind=ones(cols,1)*(0:lags);
FInd=reshape(Ind,(lags+1)*cols,1); variables=sym('variables');
index=sym('lags'); pcs(1:cols)=sym('pc');
IRsig=[variables index
pcs(1:cols);FVars FInd Rsig];
```

A.2 DAGfit

```
%Code DAGfit by M Reale.  Modified for presentation by C Meurk
%
%Input:
%
%
%G - Matrix of Lagged Variables
%edges - equations for regression
%k-number of parameters
%
%fit - function by M Reale (omitted), performs OLS

n=length(G); hq=2*log(log(n)); sz=log(n);

'model a' %All tentative links included
k=13; [s1,b1,t1,e1]=fit(G(:,[1,edges]));
[s2,b2,t2,e2]=fit(G(:,[2,edges]));
[s3,b3,t3,e3]=fit(G(:,[3,edges]));
[s4,b4,t4,e4]=fit(G(:,[4,edges]));
%corrcoef([e1,e2,e3])
dev=n*(log(s1)+log(s2)+log(s3)+log(s4)); aic=dev+2*k;
caic=aic+(2*k*(k+1))/(n-k-1); hic=dev+hq*k; sic=dev+sz*k;
[k,dev,aic,caic,sic] tv=[t1;0;t2;0;t3;0;t4];
bv=[b1;0;b2;0;b3;0;b4]; [bv,tv]
```

A.3 Case Study I, Code For Generating Contemporaneous Models

```
%Creation of Matrix

d=[]; for n=1:10

    a=[ones(1,2^(n-1)),zeros(1,2^(n-1))];

    b=a'*ones(1,2^(10-n));

    c=reshape(b,1,2^10);

    d=[c;d];

end

    UD=[ones(1,2^10);d];

    [r c]=size(d);

%Removal of Cycles and DAGs where no moral link in CIG

xx=[]; for n=1:2^10; if  ~((UD(2,n)==1 & UD(3,n)==0 &
UD(4,n)==1)|(UD(2,n)==0 & UD(3,n)==1 & UD(4,n)==0))

        xx=[xx,n];

end end


UD=UD(:,xx);


[r c]=size(UD);


xx=[]; for n=1:c;

    if ~((UD(6,n)==0 & UD(7,n)==1 & UD(9,n)==1) | ...

    (UD(6,n)==1 & UD(7,n)==0 & UD(9,n)==1))

  xx=[xx,n];

    end

end UD=UD(:,xx); [r c]=size(UD); xx=[]; for n=1:c;

  if ~((UD(8,n)==1 & UD(9,n)==0 & UD(10,n)==0 & UD(11,n)==0)|...

    (UD(8,n)==0 & UD(9,n)==1 & UD(10,n)==1 & UD(11,n)==1))
```

```
 xx=[xx,n];
    end
end UD=UD(:,xx) [r c]=size(UD); xx=[]; for n=1:c;
    if ~((UD(1,n)==1 & UD(2,n)==1))
 xx=[xx,n];
    end
end UD=UD(:,xx) [r c]=size(UD); xx=[]; for n=1:c;
    if ~(UD(3,n)==0 & UD(5,n)==1)
xx=[xx,n];
    end
end UD=UD(:,xx) [r c]=size(UD) xx=[]; for n=1:c;
    if ~(UD(8,n)==1 & UD(9,n)==1)
xx=[xx,n];
    end
end UD=UD(:,xx) [r c]=size(UD) xx=[]; for n=1:c;
    if ~(UD(10,n)==0 & UD(11,n)==1)
xx=[xx,n];
    end
end UD=UD(:,xx) [r c]=size(UD) xx=[]; for n=1:c;
    if ~(UD(6,n)==0 & UD(10,n)==1)
xx=[xx,n];
    end
end UD=UD(:,xx) [r c]=size(UD) xx=[]; for n=1:c;
    if ~(UD(5,n)==0 & UD(9,n)==0)
xx=[xx,n];
    end
end UD=UD(:,xx) [r c]=size(UD) xx=[]; for n=1:c;
    if ~((UD(1,n)==1 & UD(3,n)==1))
```

```
 xx=[xx,n];
    end
end UD=UD(:,xx) [r c]=size(UD); xx=[]; z=[1,2]; y=[1,3]; x=[1,4];
w=[3,4]; v=[4,8]; u=[5,7]; t=[5,6]; s=[6,9]; r=[6,7]; q=[7,8];
l=[8,9];

zz=[z;y;x;w;v;u;t;s;r;q;l]; zzz=[zz,zz,zz,zz,zz,zz];
for m=1:c
    for n=1:11
  if UD(n,m)==0
        zzz(n,[(2*m-1),2*m])=fliplr(zzz(n,[(2*m-1),2*m]));
  end
    end
end aa=zzz; [r c]=size(aa); bb=zeros(r,r,c);
for m=1:2:c
    for n=1:r
       bb(aa(n,m),n+1,m)=aa(n,m+1);
    end
end
bb;
bb=bb(:,:,[1:2:c])
```

B List of R Code Used

B.1 spacedata

```
 #spacedata written by C Meurk
rm(list=ls()) data<-read.table("H:rainfall_data.txt",h=T)
library(survival)

datetime <- as.POSIXct(strptime(paste(data$V1, data$V2), ...
                              "%d/%m/%Y %H:%M"), tz="GMT")

mindata<-seq(from=min(datetime),to=max(datetime),by="mins")
newdata<-matrix(0,length(mindata),1)

newdata[match(format.POSIXct(datetime,"%Y %m %d %H %M",...
tz="GMT"),format.POSIXct(mindata,"%Y %m %d %H %M",tz=...
                                          "GMT"))]<-data$V3
nr<-as.numeric(newdata) lnr<-length(nr)
nr<-matrix(nr[1:(lnr)],nrow=30) clump<-colSums(nr)
write.table(clump,file="H:x30.txt",sep=" ",row.names=FALSE)
```

As a brief aside, it is important to note some of R's foibles when it comes to time. The rainfall data were all recorded in one timezone (New Zealand Standard Time, NZST). However R, in conjunction with my computer's operating system, was determined to convert the times into New Zealand Daylight Savings Time (NZDT) where appropriate. This was problematic as when trying to align values by the minute, the command `match` could not match the data correctly surrounding time changes causing either a resulting vector 2 observations (i.e. an hour) shorter than my other series, or error messages. After extensive search and helpful comments from the `[R] help` mailing list (`https://stat.ethz.ch/mailman/listinfo/r-help`), I was able to solve the problem to the extent that I had a correct vector of values. Nevertheless (and this is acknowledged as a general problem for R especially when run from a Windows operating system), R still outputs values in NZST and NZDT but also is 12 hours behind the actual observed time. Although a small annoyance when storing time, R actually stores the date as seconds past 1/1/70 and hence it did not cause errors in my results.

B.2 CREATEX

```
#CREATEX, creates matrix of column vectors for analysis with
#MATLAB function cigts, eg for 60 minute aggregation takes x60
#from modified code spacedata

x<-read.table("H:x60.txt",h=T)
y<-read.csv("H:plot_1_theta1.txt",h=T)
z<-read.table("H:plot_6_theta1.txt",h=T) y<-y[,4] ly<-length(y)
y<-y[1:(ly-1)] ny<-matrix(y,nrow=2) ny<-colMeans(ny) z<-z[,4]
lz<-length(z) z<-z[1:(lz-1)] nz<-matrix(z,nrow=2) nz<-colMeans(nz)
X<-cbind(x,ny,nz)
write.table(X,file="H:X60.txt",sep="",row.names=FALSE)

#This code is easily adapted for any time aggregation
```

C Full CIG for Chapter 3

References

[1] Akaike, H., Information theory as an extension of the maximum likelihood principle. Second International Symposium on Information Theory, Budapest (1973).

[2] Burnham, K., and Anderson, D., Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach 2nd edn, Springer-Verlag (1998).

[3] Chatfield C., The Analysis of Time Series: An Introduction, 4th edn. Chapman & Hall (1989)

[4] Choquenot, D., and Ruscoe, W., Mouse Population Eruptions in New Zealand Forests: The Role of Population Density and Seedfall, Journal of Animal Ecology, 69(6)1058-1070 (2000).

[5] Davies, T., Working title: Surface run-off mechanisms in the Pukemanga catchment, Whatawhata, Waikato, NI, NZ, unpublished paper, (2004).

[6] Dawid, A., Conditional Independence in Statistical Theory, Journal of the Royal Statistical Society, Series B (Methodological) 41(1)1-31 (1979).

[7] Dickey, D., and Fuller, W., Distribution of the Estimators for Autoregressive Time Series with Unit Root, Journal of the American Statistical Association, 74(366)427-431, (1979).

[8] Doob,J., Stochastic Processes, John Wiley & Sons, Inc. (1953)

[9] Drton, M., and Perlman, M., A SINful Approach to Gaussian Graphical Model Selection, online publication at http://www.stat.washington.edu/drton/papers.html, (2004).

[10] Drton, M (2004). SIN: A SINful Approach to Selection of Gaussian Graphical Markov Models. R package version 0.2. http://www.r-project.org, http://www.stat.washington.edu/drton

[11] Edwards, D., Introduction to Graphical Modelling, 2nd edn. Springer Texts in Statistics (2000).

[12] Fitzgerald, B., Efford, M. and Karl, B., Breeding of house mice and the mast seeding of southern beeches in the Orongorongo Valley, New Zealand, New Zealand Journal of Zoology, 31(2)167-184 (2004).

[13] Granger, C., Some Recent Developments in a Concept of Causality, Journal of Econometrics 39, 199-211 (1988).

[14] Hamilton, J., Time Series Analysis, Princeton University Press, (1994).

[15] Hannan, E. and Quinn, B., The determination of the order of an autoregression. Journal of the Royal Statistical Society, Series B, 41(2)190-195 (1979).

[16] Hicks, J., Causality in Economics, Basil Blackwell Publisher, (1979).

[17] Krebs, C., Ecological Methodology 2nd edn, Addison-Welsey Educational Publishers Inc., (1999).

[18] Ovid Technologies, The Philosopher's Index, WebSPIRS 5 (2005).

[19] Pearl, J., Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann (1988)

[20] Pearl, J., Causality: Models, Reasoning and Inference, Cambridge University Press (2000)

[21] Phillips, P and Perron, P., Testing for a Unit Root in Time Series Regression, Biometrika, 75(2)335-346 (1988).

[22] Plantinga, A., Warrant: The Current Debate, Oxford University Press, (1993).

[23] Plantinga, A., Warrant and Proper Function, Oxford University Press, (1993).

[24] R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[25] Reale, M., A Graphical Modelling Approach to Time Series, PhD thesis - Lancaster University (1998)

[26] Reale, M., Tunnicliffe-Wilson, G., Causal Diagrams for I(1) structural VAR models (2002)

[27] Schwarz,G., Estimating the dimension of a mode, Annals of Statistics, (6)461-464.

[28] Shaw, J.E.H., Some Quotable Quotes for Statistics, www.ewartshaw.co.uk.

[29] Sidak, Z. Rectangular confidence regions for the means of multivariate normal distributions, Journal of the American Statistical Association 62(318)626-633 (1967).

[30] Stephens, D., The equilibrium exchange rate according to PPP and UIP, Reserve Bank of New Zealand DP2004/03 (2004).

[31] Yue, S., Pilon, P., and Phinney, B., Canadian Streamflow trend detection; impacts of serial and cross-correlation, Hydrological Sciences Journal 48(1)51-63 (2003).

Model A	Model B	Model C	Model D	Model E	Model F	Model G	Model H	Model I	Model J	Model K	Model L	Model M
$[1, 2]$	$[1, 2]$	$[1, 2]$	$[1, 2]$	$[1, 2]$	$[1, 2]$	$[1, 2]$	$[1, 2]$	$[1, 2]$	$[1, 2]$	$[1, 2]$	$[1, 2]$	0
$[2, 3, 7]$	$[2, 3, 7]$	$[2, 3, 7]$	$[2, 3, 7]$	$[2, 3, 7]$	$[2, 3, 7]$	$[2, 7]$	$[2, 7]$	$[2, 3]$	$[2, 3]$	$[2, 3]$	0	0
$[3, 4]$	$[3, 4]$	$[3, 4]$	$[3, 4]$	0	0	$[3, 2]$	$[3, 2]$	$[3, 4]$	$[3, 4]$	0	$[3, 2]$	$[3, 2]$
$[4, 5]$	$[4, 5]$	0	0	$[4, 3]$	$[4, 3]$	$[4, 3]$	$[4, 3]$	$[4, 5]$	0	$[4, 3]$	$[4, 3]$	$[4, 3]$
0	0	$[5, 4]$	$[5, 4]$	$[5, 4]$	$[5, 4]$	$[5, 4]$	$[5, 4]$	0	$[5, 4]$	$[5, 4]$	$[5, 4]$	$[5, 4]$
$[6, 7]$	0	$[6, 7]$	0	$[6, 7]$	0	$[6, 7]$	0	$[6, 7]$	$[6, 7]$	$[6, 7]$	$[6, 7]$	$[6, 7]$
0	$[7, 6]$	0	$[7, 6]$	0	$[7, 6]$	0	$[7, 6]$	$[7, 2]$	$[7, 2]$	$[7, 2]$	$[7, 2]$	$[7, 2]$
0	0	0	0	0	0	0	0	0	0	0	0	0
$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$	$[9, 8]$

Table 2: Table of regression equations corresponding the alternate DAGs in figure 15.

Orientation down	[1, (10), (11), (20), (38)]	[2, (11), (12), (21)]	[3, (12), (13), (22)]	[4, (13), (14), (23)]	[6, (42), (49)]	[8, (17)]
	[2, 1, 10, 11, 20, 38]	[3, 2, 11, 12, 21]	[4, 3, 12, 13, 22]	[5, 4, 13, 14, 23]	[7, 6, 42, 49]	[9, 8, 17]
Orientation Up	[1, 2, 10, 11, 20, 38]	[2, 3, 11, 12, 21]	...			
	[2, (10), (11), (20), (38)]	[3, (11), (12), (21)]	...			

Table 3: Table showing the regularity of the contemporaneous relationship. Given the regularity, it can be seen that these values are generated automatically.

Model A	Model B	Model C
[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 3, 7, (10), 11, (20), (38), 12, 21] [3, 4, 28, 51, (11), (21), 12, 13, 22] [4, 5, 10, 11, 31, 40, (12), (22), 13, 14, 23] [5, 26, (13), (14), (23)] [6, 7, 15, 27, 29, 42, 49] [7, 16, 28, 37, 44, (42), (49)] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]	[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 3, 7, (10), 11, (20), (38), 12, 21] [3, 4, 28, 51, (11), (21), 12, 13, 22] [4, 5, 10, 11, 31, 40, (12), (22), 13, 14, 23] [5, 26, (13), (14), (23)] [6, 15, 27, 29, (42), (49)] [7, 6, 16, 28, 37, 44, 42, 49] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]	[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 3, 7, (10), 11, (20), (38), 12, 21] [3, 4, 28, 51, (11), (21), 12, 13, 22] [4, 10, 11, 31, 40, (12), (13), (22), (14), (23)] [5, 4, 26, 13, 14, 23] [6, 7, 15, 27, 29, 42, 49] [7, 16, 28, 37, 44, (42), (49)] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]
Model D	Model E	Model F
[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 3, 7, (10), 11, (20), (38), 12, 21] [3, 4, 28, 51, (11), (21), 12, 13, 22] [4, 10, 11, 31, 40, (12), (13), (22), (14), (23)] [5, 4, 26, 13, 14, 23] [6, 15, 27, 29, (42), (49)] [7, 6, 16, 28, 37, 44, 42, 49] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]	[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 3, 7, (10), 11, (20), (38), 12, 21] [3, 28, 51, (11), (12), (21), (13), (22)] [4, 3, 10, 11, 31, 40, 12, 13, 22, (14), (23)] [5, 4, 26, 13, 14, 23] [6, 7, 15, 27, 29, 42, 49] [7, 16, 28, 37, 44, (42), (49)] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]	[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 3, 7, (10), 11, (20), (38), 12, 21] [3, 28, 51, (11), (12), (21), (13), (22)] [4, 3, 10, 11, 31, 40, 12, 13, 22, (14), (23)] [5, 4, 26, 13, 14, 23] [6, 15, 27, 29, (42), (49)] [7, 6, 16, 28, 37, 44, 42, 49] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]
Model G	Model H	Model I
[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 7, (10), (11), (20), (38), (12), (21)] [3, 2, 28, 51, 11, 12, 21, (13), (22)] [4, 3, 10, 11, 31, 40, 12, 13, 22, (14), (23)] [5, 4, 26, 13, 14, 23] [6, 7, 15, 27, 29, 42, 49] [7, 16, 28, 37, 44, (42), (49)] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]	[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 7, (10), (11), (20), (38), (12), (21)] [3, 2, 28, 51, 11, 12, 21, (13), (22)] [4, 3, 10, 11, 31, 40, 12, 13, 22, (14), (23)] [5, 4, 26, 13, 14, 23] [6, 15, 27, 29, (42), (49)] [7, 6, 16, 28, 37, 44, 42, 49] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]	[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 3, (10), 11, (20), (38), 12, 21] [3, 4, 28, 51, (11), (21), 12, 13, 22] [4, 5, 10, 11, 31, 40, (12), (22), 13, 14, 23] [5, 26, (13), (14), (23)] [6, 7, 15, 27, 29, 42, 49] [7, 2, 16, 28, 37, 44, (42), (49)] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]
Model J	Model K	Model L
[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 3, (10), 11, (20), (38), 12, 21] [3, 4, 28, 51, (11), (21), 12, 13, 22] [4, 10, 11, 31, 40, (12), (22), (13), (14), (23)] [5, 4, 26, 13, 14, 23] [6, 7, 15, 27, 29, 42, 49] [7, 2, 16, 28, 37, 44, (42), (49)] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]	[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, 3, (10), 11, (20), (38), 12, 21] [3, 28, 51, (11), (21), (12), (13), (22)] [4, 3, 10, 11, 31, 40, 12, 13, 22, (14), (23)] [5, 4, 26, 13, 14, 23] [6, 7, 15, 27, 29, 42, 49] [7, 2, 16, 28, 37, 44, (42), (49)] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]	[1, 2, 15, 35, 37, 47, 10, 11, 20, 38] [2, (10), (11), (20), (38), (12), (21)] [3, 2, 28, 51, 11, 12, 21, (13), (22)] [4, 3, 10, 11, 31, 40, 12, 13, 22, (14), (23)] [5, 4, 26, 13, 14, 23] [6, 7, 15, 27, 29, 42, 49] [7, 2, 16, 28, 37, 44, (42), (49)] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]
Model M		
[1, 15, 35, 37, 47, (10), (11), (20), (38)] [2, (10), (11), (20), (38), (12), (21)] [3, 2, 28, 51, 11, 12, 21, (13), (22)] [4, 3, 10, 11, 31, 40, 12, 13, 22, (14), (23)] [5, 4, 26, 13, 14, 23] [6, 7, 15, 27, 29, 42, 49] [7, 2, 16, 28, 37, 44, (42), (49)] [8, 53, (17)] [9, 8, 16, 18, 27, 29, 36, 17]		

Table 4: Alternative model specification. In cases where an edge was represented twice - once as real, once as moral, the edge was deemed to be real. Edges are not necessarily in order, however, ordering is carried out prior to regression.

Model Type	AIC	HIC	SIC
Sat cVAR	-3.3594e+003	-2.8990e+003	-2.2244e+003
Sat sVAR	-5.2648e+003	-4.7119e+003	-3.9015e+003
Model A	-4.5644e+003	-4.4873e+003	-4.3742e+003
Model B	-4.5147e+003	-4.4375e+003	-4.3244e+003
Model C	-4.5711e+003	-4.4940e+003	-4.3809e+003
Model D	-4.5118e+003	-4.4346e+003	-4.3216e+003
Model E	-4.5447e+003	-4.4676e+003	-4.3545e+003
Model F	-4.4854e+003	-4.4082e+003	-4.2952e+003
Model G	-4.5474e+003	-4.4703e+003	-4.3572e+003
Model H	-4.4881e+003	-4.4109e+003	-4.2979e+003
Model I	-4.6927e+003	-4.6155e+003	-4.5024e+003
Model J	-4.6898e+003	-4.6126e+003	-4.4996e+003
Model K	-4.6634e+003	-4.5862e+003	-4.4732e+003
Model L	-4.6595e+003	-4.5824e+003	-4.4693e+003
Model M	-4.5127e+003	-4.4355e+003	-4.3224e+003

Table 5: table of AIC,HIC and SIC values for candidate models.

Series	Test-Statistic	Outcome	Deduced status ($\alpha = 0.05$)
Rainfall	-7.5191	reject H	$I(0)$
Soil Moisture 10cm, site 1	-0.3862	$-$(reject H)	$I(1)$
Soil Moisture at 10cm depth site 6	-0.2525	$-$(reject H)	$I(1)$

Model	AIC	SIC	HIC
TradSat	-238790	-238470	-237810
CCM	-272840	-272720	-272460
CCnoml	-269880	-269770	-269550
noCC	-262230	-262140	-261760

Table 6: TradSat refers to a saturated sVAR, CCM is a contemporaneous causal model with all moral links included. CCnoml is a contemporaneous causal model with no moral links included and noCC is the non-contemporaneous model. This table indicates that CCM was favoured by all criteria.

Model	AIC	SIC	HIC
TradSat	-64095	-64005	-63826
CCM	-96117	-96032	-95863
CCnoml	-44139	-41187	-41367
noCC	-89066	-88988	-88834

Table 7: TradSat refers to a saturated sVAR, CCM is a contemporaneous causal model with all moral links included. CCnoml is a contemporaneous causal model with no moral links included and noCC is the non-contemporaneous model. This table indicates that CCM was favoured by all criteria.

Model	AIC	SIC	HIC
TradSat	-10303	-10092	-9692
CCM	-18053	-18187	-18440
CCnoml	-13336	-13271	-13146
noCC	-22906	-22826	-22675

Table 8: TradSat refers to a saturated sVAR, CCM is a contemporaneous causal model with all moral links included. CCnoml is a contemporaneous causal model with no moral links included and noCC is the non-contemporaneous model. This table indicates that noCC was favoured by all criteria.

variables	lags	pc	pc	pc	pc	pc	pc	pc	pc	pc
1	0	1	x	x	x	x	x	x	x	x
2	0	1	1	x	x	x	x	x	x	x
3	0	0	1	1	x	x	x	x	x	x
4	0	0	0	1	1	x	x	x	x	x
5	0	0	0	0	1	1	x	x	x	x
6	0	0	0	0	0	0	1	x	x	x
7	0	0	1	0	0	0	1	1	x	x
8	0	0	0	0	0	0	0	0	1	x
9	0	0	0	0	0	0	0	0	1	1
1	1	1	1	0	1	0	0	0	0	0
2	1	1	1	1	1	0	0	0	0	0
3	1	0	1	1	1	0	0	0	0	0
4	1	0	0	1	1	1	0	0	0	0
5	1	0	0	0	1	1	0	0	0	0
6	1	1	0	0	0	0	1	0	0	0
7	1	0	0	0	0	0	0	1	1	1
8	1	0	0	0	0	0	0	0	1	1
9	1	0	0	0	0	0	0	0	1	1
1	2	0	0	0	0	0	0	0	0	0
2	2	1	1	0	0	0	0	0	0	0
3	2	0	1	1	0	0	0	0	0	0
4	2	0	0	1	1	0	0	0	0	0
5	2	0	0	0	1	1	0	0	0	0
6	2	0	0	0	0	0	0	0	0	0
7	2	0	1	0	0	0	0	0	0	0
8	2	0	0	0	0	1	0	0	0	0
9	2	0	0	0	0	0	1	0	1	1
1	3	0	0	1	0	0	0	1	1	0
2	3	0	0	0	0	0	1	0	1	1
3	3	0	0	0	0	0	0	0	0	0
4	3	0	0	0	1	0	0	0	0	0
5	3	0	0	0	0	0	0	0	0	0
6	3	0	0	0	0	0	0	0	0	0
7	3	1	1	0	0	0	0	0	0	0
8	3	1	0	0	0	0	0	0	0	0
9	3	0	0	0	0	0	0	0	0	1
1	4	1	0	0	0	0	0	1	0	0
2	4	1	1	0	0	0	0	0	0	0
3	4	0	0	0	0	0	0	0	0	0
4	4	0	0	0	1	0	0	0	1	0
5	4	0	0	0	0	0	0	0	1	0
6	4	0	0	0	0	0	0	0	0	0
7	4	0	0	0	0	0	1	0	0	0
8	4	0	0	0	0	0	0	0	0	0
9	4	0	0	0	0	0	0	0	0	0
1	5	0	0	0	0	0	0	0	0	0
2	5	1	0	0	0	0	0	0	0	0
3	5	0	0	0	0	0	0	0	0	0
4	5	0	0	0	0	0	1	1	1	0
5	5	0	0	0	0	0	0	0	1	0
6	5	0	0	1	0	0	0	0	0	0
7	5	0	0	0	0	0	0	0	0	0
8	5	0	0	0	0	0	0	0	1	0
9	5	0	0	0	0	0	0	0	1	0

Table 9: `MATLAB` output of CIG, an entry of 0 indicates conditional independence and an entry of 1 indicates conditional dependence i.e. that an edge can be drawn between two nodes