Modelling the Interactions in Paleoclimate Data from Ice Cores*

James Davidson
University of Exeter

Alemtsehai Turasie
CPTEC/INPE, São Paulo

August 21st 2013

Abstract

This paper considers methods for testing directions of causation in the paleoclimate series for temperature and CO2 concentration derived from Antarctic ice cores from 800,000 years BP. These series are well-known to move together in the transitions between glacial and interglacial periods, but the dynamic relationship between the series is open to dispute. Bivariate models are constructed, in the context of which we perform tests for Granger causality, or in other words for asymmetry in the pattern of dynamic interactions. An important question is the stationarity of the series, because different statistical techniques are called for in stationary and integrated models. Previous work with climate data has focused on a cointegration approach appropriate to integrated series, but a range of tests show no evidence of integrated behaviour in these seres. A second important question is linearity, whether the dynamics in mean can be adequately represented by a high-order vector autoregressive process. This modelling approach is compared with a nonlinear Markov switching mechanism, in which the glacial/interglacial switches are controlled by a common hidden discrete process with fixed conditional probabilities of changes of state. A further characteristic of the data that comes to light is pronounced conditional heteroscedasticity, with much larger disturbances in evidence around the glacial/interglacial transitions.This data feature is effectively accounted for by fitting a GARCH component.

1 Introduction

Climatic data gathered from cores drilled in the Antarctic ice cap have played a prominent role in the debate on the causes of climate change. Measurements include the concentration of carbon dioxide and other gases in air trapped in bubbles in the ice, and the proportions in the water content of the heavy hydrogen isotope deuterium. It is argued that because water molecules containing deuterium are heaver than the commoner variety, the water vapour resulting from ocean evaporation has a lower deuterium content than the ocean, and the effect is more pronounced in colder climatic periods. The deuterium content of water vapour subsequently falling as snow is accordingly held to reflect the surface temperature of the ocean. A temperature proxy, in degrees Celsius, is generated as a linear transformation of the deuterium concentration. The samples are dated by counting the seasonal snowfall variations, subject to corrections to allow for the fact that gases are only locked into bubbles and isolated from the atmosphere at a date (the gas age) some time after the snowfall (the ice age).

Time series for temperature and CO_2 concentration, plotted in Figure 1. The ordering of observations is chronological, with the oldest to the left and most recent to the right, with the

*We thank David Stephenson, David Stern and participants at the Guelph 2013 International Workshop on. Econometric Applications in Climatology for their comments. Errors are ours alone.

horizontal axis showing age in thousands of years BP. Temperatures are measured as the difference in degrees Celsius from the average of the last 1000 years, and CO_2 concentration is measured in units of parts per ten thousand by volume.[1] The series show substantial periodic variations dividing the historical record into the so-called glacial and interglacial periods. There is clearly a strong association between the temperature and CO_2 records over these cycles, which are assumed to reflect variations in the precession, obliquity and eccentricity of the Earth's orbit (see Imbrie et al. (1992), Berger et al. (1978), Lüthi et al. (2008)). However, the precise mechanism by which these geophysical phenomena affect temperature and CO_2 concentration in the way that they do remains unknown.

There has however been speculation that greenhouse gas concentrations may have driven temperature over the glacial cycle. It has in fact been observed (Caillon et al. 2003) that at least in one specific episode (Termination III) the changes in the temperature record appear to lead those in CO_2 concentration by an order of 800±200 years. Fischer et al. (1999) estimated that increases in CO_2 lagged temperatures in the Vostok record by 600±400 years at the start of the last three Terminations. Mudelsee (2001) undertakes a parabolic regression analysis to determine phase relations in the Vostok data, and finds a lag of 1300 years of CO_2 behind the deuterium proxy. However, a formal statistical test of causality applied to the whole 800,000 years span of available data, has not so far been attempted. This paper attempts to bring new methods to bear on this problem. The econometric testing technique to be applied involves constructing a dynamic parametric model to capture the characteristics of the data, within which the restrictions of interest can be defined.

The fundamental idea of testing directions of causation in economic data is due to Granger (1969). So-called Granger causation of a time-series variable Y by another variable X is the phenomenon that the information contained in X can be used to improve forecasts of future values of Y. Two-way Granger causation, in this sense, would be likely to imply that X and Y are both driven by some, possibly unobserved, third force. One-way Granger causation, on the other hand, where the property does not obtain when the roles of the variables are interchanged, is evidence in favour of a direct causal mechanism.

Given the close relationship over time between the CO_2 and temperature series, forecasting potential in at least one direction is to be expected. The hypothesis that variations in CO_2 concentration drive variations in temperature would be supported by the finding that CO_2 Granger-causes temperature, whereas temperature does not Granger-cause CO_2. Such a finding would be a key piece of evidence for the hypothesis of global warming due to anthropogenic increases in CO_2. In any other case, however, we should have to conclude that the paleoclimate record can tell us little, one way or the other, about anthropogenic global warming.

2 The Data Set

The data used in this study are taken from the website of the US National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center.[2] The source of the data is the European Project for Ice Coring in Antarctica (EPICA) Dome C. The raw data on temperature, described in Jouzel et al. (2007), take the form of 5,800 core samples, for each of which are reported the depth of the sample, the imputed date, the deuterium content of the ice, and the imputed temperature. The formula used to construct the temperatures, discussed in Jouzel et al. (2003), is approximately linear, but varies depending on the age of the samples, as can be

[1] The published data for CO_2 concentration are expressed in parts per million by volume (ppmv). We divided this series by 100 for the purposes of plotting and calculations, to minimise the differences of scale.

[2] http://www.ncdc.noaa.gov/paleo/pubs/ipcc2007/fig63.html

seen in the scatter plot of the two series (See Figure 2). This plot distinguishes the age of the observations by the shading of the points, with the darkest points being the most recent and the lightest, the oldest. While the mean interval between observations is 138 years, the older observations are considerably sparser than the recent ones, as can be seen in the time plot of these intervals in Figure 3(a). However, the intervals exceed 1000 years on only about 25 occasions, at the earliest dates.

The CO_2 measurements, on the other hand, number only 1098 to cover the same 800,000 year period. This is the concentration of the gas in trapped air bubbles, representing a much smaller proportion of each sample than water. Also, this is a composite series combining data from the EPICA and Vostok sites. The intervals between the observations are plotted in Figure 3(b). Don't overlook the large difference in both the vertical and horizontal scales in these plots. The mean interval between CO_2 measurements is 729 years, but over 40 of the intervals exceed 2000 years. These gaps presumably correspond to periods of low precipitation, so that a shorter time span yields too small a sample for analysis.

A key issue in the interpretation of these series is the dating of the CO_2 observations. Following an initial deposit of fallen snow, atmospheric air is capable of diffusing into the partially consolidated snow (so-called 'firn') for some time before it becomes compressed into ice and the air therefore trapped. It is important to distinguish the ice age (the age assigned to the deuterium record) from the gas age (the age assigned to the trapped air). This issue is analysed in Barnola et al. (1991). Since matters of timing are critical to the question investigated here, we acknowledge that our results depend critically on the correctness of these assigned ages.

The main issue needing to be treated here is that of transforming the data into a form amenable to parametric modelling. The essential requirement is for a set of observations equally spaced in time. The usual modelling framework cannot accommodate irregular observation dates. Time series with irregularly spaced observations have been studied in a univariate context, for example by Robinson (1977) using a simple diffusion model, but a multivariate analysis of dated interactions between variables with different irregular observation frequencies lies beyond the reach of standard econometric methods. However, a feasible approach is to choose a suitable set of equally spaced dates, and then attempt to impute values to the variables on these dates by interpolation.

Because of the lower frequency of the CO_2 observations, it will be necessary to discard the greater part of the temperature measurements to define matching data frequencies. Simple linear interpolation corresponds to joining up the (recorded date, recorded value) pairs by straight lines and then reading off the values of this continuous record at the chosen intervals. The obvious alternative method is to fit a cubic spline, which would be appropriate when the data points are sparser on average than the desired intervals. In these data, the two methods must deliver very similar results, and linear interpolation is much simpler to implement and to interpret.

Given the thousand-odd CO_2 observations, it will be appropriate to extract from these somewhat fewer than 1000 data points. Choosing intervals of 1000 years results in 798 imputed data points. Choosing 500 year intervals yields twice as many points, but then the majority of the imputed CO_2 values would depend on the same observations as their neighbours, with a consequent spurious dependence in the series. On the other hand, choosing 1500 year intervals would yield 534 data points, so that with this frequency almost half the CO_2 observations would effectively be discarded. There is, inevitably, a trade-off of disadvantages here. Higher frequencies risk distorting the relative timing of CO_2 and temperature observations, such that a nominal 'lagged value' in one variable may often depend on a contemporaneous shock. Lower frequencies, on the other hand, remove key information through aliasing. In either case, there are sections of the sample where the CO_2 data are very sparse and several successive time periods must depend on the same data points. Figure 3 indicates the locations of these problematic points. It seems that

Lag(+)/Lead(−) of CO_2	2	1	0	−1	−2
Levels	0.790	0.845	0.884	0.891	0.876
Differences	0.013	0.198	0.430	0.346	0.170

Table 1: Correlations beween temperature and lags and leads of CO_2.

this problem is unavoidable, but there are hopefully too few such points to significantly distort the outcome of the analysis. Given these considerations, we elect to use the 1000 year interval data in this study.

Valid tests of Granger causality depend critically on the timing of the observations, and given the element of uncertainty regarding the gas age-ice age difference, there is a temptation to experiment with alternative dating schemes. However, it is not clear that any simple translation would be appropriate, and to second-guess the authors of the data sets without the requisite expertise appears in any case unwise. By construction, our interpolated data set will not be able to discriminate timing differences of less than 1000 years, and our tests will require lagged effects of at least this order to have power. It is encouraging to observe that in our interpolated series plotted in Figure 1, the main glacial-interglacial shifts appear closely coordinated. More formally, a natural check to perform is to compare the contemporaneous correlations of the series with those of low-order leads and lags. This is done in Table 1, for the levels and differences of the series respectively.[3] Interestingly there is a slightly higher correlation of temperature with future CO_2 than with contemporaneous CO_2 although this effect is not apparent in the differenced series, which is certainly the more relevant case in this exercise.

3 Initial Examination of the Data

Previous econometric work on climatological series from the more recent past has tended to focus on nonstationary time series modelling using a cointegration approach; see, among other studies of this type, Kaufman and Stern (2002), Kaufman et al. (2007, 2010) and Mills (2007). However, there is little visual evidence for unit root or other nonstationary behaviour in the series plotted in Figure 1, Notwithstanding the pronounced alternation of glacial and interglacial episodes, we also note that the series are confined within fairly well-defined limits, and even exhibit a type of mean reversion.

The patterns of variation also show quite a pronounced asymmetry, with more extreme variations evident in the inter-glacial periods than in the glacial periods. This suggests that a monotone transformation may be appropriate for modelling the series subject to a Gaussian paradigm. Before doing any further analysis, we take logarithms of the series after first adding a constant to the temperature series sufficiently large to ensure positivity and an approximately symmetric distribution of data points around the central tendency. The chosen shift is 16 units. The CO_2 series, being positive by construction, is not shifted. Let it be emphasized that this transformation would not allow estimated coefficients to be interpreted as elasticities.[4] The parameterization is from this point of view arbitrary, but is chosen in conjunction with the functional forms to be adopted to yield representations of the data generation process that are hopefully adequate for our purpose. A more detailed justification of these choices is given in the sequel.. These transformed series are plotted in Figure 4.

[3] The series used in these calculations are the logarithmically-transformed cases as described in Section 3 of the paper.

[4] For this purpose we would need to express temperature in degrees Kelvin.

	log(Temp+16)		**log(CO2)**	
	Statistic	p-value	Statistic	p-value
Tests of I(1):				
ADF Test*	-4.73 [3]	< 0.01	-4.23 [3]	< 0.01
PP Test**	-4.94 [5]	< 0.01	-4.10 [5]	< 0.01
ERS, DF-GLS*	-4.56 [1]	< 0.01	-3.93 [3]	< 0.01
ERS, P Test**	0.530 [4]	< 0.01	0.757 [4]	< 0.01
Tests of I(0)				
KPSS Test**	0.129 [13]	< 1	0.228 [13]	< 1
V/S Test**	0.063 [13]	0.563	0.202 [13]	0.037
Modified RS Test**	1.072 [13]	< 0.8	1.65 [13]	< 0.1
RL Test†	-1.23 [12]	0.891	-0.499 [12]	0.691
HML test ‡	7.631	0	7.60	0

* Lags (in square brackets) chosen by Akaike's criterion.
** HAC variance computed with Parzen kernel and bandwidth (in square brackets) chosen by the Newey-West (1994) plug-in method.
† Bandwidth (in square brackets) from Lobato and Robinson (1998) formula.
‡ Setting $c = 1, L = 0.66$,.see Harris et. al. (2008) for details.

Table 2: Tests for unit roots (I(1))and stationarity (I(0).

Table 2 shows the results of a range of classic and more recently derived tests for unit roots and for the null hypothesis of stationarity and weak dependence, applied to the transformed series.[5] These null hypotheses are referred to respectively as "I(1)" and "I(0)". ADF denotes the augmented Dickey-Fuller test (Said and Dickey 1984), the PP test is from Phillips and Perron (1988), and the ERS tests are from Elliott, Rothenberg and Stock (1996). The KPSS test is from Kwiatkowski et al. (1992), the V/S test from Giraitis et. al. (2003), the modified RS test from Lo (1991), the RL test from Lobato and Robinson (1998), and the HML test from Harris et al. (2008). The p-values shown are computed exactly from asymptotic tables where available. In the case of p-values computed from Monte Carlo simulations, upper bounds according to the tabulated points are indicated by "<".

Clearly, the I(1) hypothesis is decisively rejected by all the tests. The results on the I(0) hypothesis are more ambivalent, in general failing to reject but with some disagreement between the criteria. The KPSS, V/S and RS tests represent different ways of comparing the cumulated process to a Brownian motion, while the RL test is based in the frequency domain and tests essentially that the spectral density is finite at he origin. The HML test, which notably rejects the null hypothesis decisively, is based specifically on the distribution of the high-order sample autocovariances.

It may be argued that I(1) is an improper null hypothesis in the present case due to the evidence that the series move within fixed upper and lower bounds, in spite of the appearance of stochastic trending behaviour within those bounds. A bounded random walk model with these features has been studied by Cavaliere (2005) and Granger (2010). Cavaliere considers the case

$$\begin{aligned} X_t &= \theta + Y_t \\ Y_t &= Y_{t-1} + \varepsilon_t + \underline{\xi}_t - \overline{\xi}_t \end{aligned}$$

where ε_t is a weakly dependent process and $\underline{\xi}_t$ and $\overline{\xi}_t$ are non-negative 'regulator processes' such

[5] All the calculations in this paper were carried out using the Time Series Modelling 4 econometrics package (Davidson 2013) which runs under the Ox programming system (Doornik 2009).

	log(Temp+16)	**log(CO2)**
s_{AR}	0.0837	0.0253
Sample initial value X_0	1.998	0.642
$\overline{b}$ = sample maximum	2.984	1.082
$\underline{b}$ = sample minimum	1.748	0.546
$\widehat{\overline{c}}$	0.417	0.613
$\widehat{\underline{c}}$	-0.106	-0.135
5% Critical Value from Simulations	-3.97	-3.61
1% Critical Value from Simulations	-4.58	-4.31

Table 3: Unit root tests subject to barriers: data features

that $\underline{b} - \theta \leq Y_t \leq \overline{b} - \theta$. and hence X_t is confined within the interval $[\underline{b},\overline{b}]$. If the bounds $\underline{b}$ and $\overline{b}$ are known, it is possible to consider a test in which the bounded random walk forms the null hypothesis. Cavaliere and Xu (2012) derive a test of this hypothesis based on a Monte Carlo simulation of the limit distribution of regulated Brownian motion. This simulation can be performed for the sample size appropriate to the observed data, but note that the bounds defining the approximating distribution must be linked to the long-run variance of the shock process ε_t. The limit process is a regulated Brownian motion $B_{\underline{c}}^{\overline{c}}$ where $\overline{c}$ and $\underline{c}$ are nuisance parameters on which the distribution depends. Cavaliere and Xu (2012) show that $\overline{c}$ and $\underline{c}$ can be consistently estimated by the respective formulae

$$\begin{aligned} \widehat{\overline{c}} &= T^{-1/2}(\overline{b} - X_0)/s_{AR} \\ \widehat{\underline{c}} &= T^{-1/2}(\underline{b} - X_0)/s_{AR} \end{aligned}$$

where T is sample size, X_0 is the initial value of the sequence and s_{AR} denotes the the autoregressive estimator of the long-run variance of the differences. Note the implication, that to validate the asymptotic approximation the bounds $\underline{b}$ and $\overline{b}$ must be assumed to diverge at the rate $T^{1/2}$; clearly, the Brownian approximation cannot operate if the bounds are too close together, relative to the range of variation of the difference process. The Cavaliere-Xu test is implemented by simulating the standard Dickey-Fuller statistic in a Monte Carlo exercise with artificial samples of, in the present case, length $T = 798$ to match the observed series. These are generated by a bounded random walk with bounds $\widehat{\overline{c}}$ and $\widehat{\underline{c}}$, as appropriate for each series, and independent Gaussian shocks ε_t^* having variance $1/T$. The bounds were imposed by the data generating equation

$$X_t^* = \widehat{\overline{c}} + ((X_{t-1}^* + \varepsilon_t^* - \widehat{\underline{c}})^+ + \widehat{\underline{c}} - \widehat{\overline{c}})^-$$

where $(.)^+$ and $(.)^-$ denote respectively the positive and negative parts of their arguments. The problem is to choose values for $\underline{b}$ and $\overline{b}$. The tighter these are chosen, the further the test distribution will be shifted to the left and the smaller the critical values. It makes sense to verify the test outcomes under the extreme case, setting the bounds to the actual recorded maxima and minima of the series, since a rejection in this case cannot be contradicted by different choices.

The relevant statistics for the construction of the test tabulations are shown in Table 3 together with the critical values estimated from 10,000 Monte Carlo replications, in which the augmented Dickey-Fuller statistics are computed from samples of size 798 with three lags, to match the selection for the results reported in Table 2. It is clear, by comparing the critical values in Table 3 with the ADF statistics in Table 2, that the bounded unit root hypothesis is also rejected by these data, at or near the 1% level.

Another view of these data properties is provided by Table 4, which shows the results of Johansen tests for reduced rank in the context of a vector autoregression of order 10, selected

	Maximum Eigenvalue Test		Trace test	
Rank	Statistic	p-value	Statistic	p-value
0	59.43	< 0.01	74.38	< 0.01
1	14.94	< 0.01	14.94	< 0.01

Table 4: Johansen tests for cointegrating rank

	$\hat{d}$ (Local Whittle)			$\hat{d}$ (GPH)			Bias test	
Bandwidth	$[T^{0.5}]$	$[T^{0.6}]$	$[T^{0.7}]$	$[T^{0.5}]$	$[T^{0.6}]$	$[T^{0.7}]$	Statistic	p-value
log(Temp+16)	0.271	0.741	0.949	0.0213	0.475	0.639	2.89	0.002
log(CO2)	0.586	0.762	0.913	0.447	0.629	0.800	5.45	0

Table 5: Semiparametric estimates of the long memory parameter

by optimizing the Akaike model selection criterion. Of course, the null hypothesis in these tests entails the condition that both series are I(1). Rejection of both the rank hypotheses implies that the VAR has full rank, and hence that the variables exhibit reversion to a finite long-run mean. The decisive rejection reinforces the conclusion that these variables ought to be treated as stationary. The result of these preliminary looks at the data is to decisively reject a 'nonstationary with cointegration' interpretation of the evidence.

If both the I(1) and I(0) hypotheses are rejected, which is one way to interpret the results of Table 2, the data might possibly be characterized as I(d) for $0 < d < 1$, where d is a parameter measuring the rate of divergence of the spectral density at the origin. We might, under this hypothesis, attempt to estimate d using a semiparametric estimator such as the Geweke-Porter Hudak (1982) log-periodogram regression (GPH), or the 'local Whittle' maximum likelihood procedure (Kunsch 1987, Robinson 1995). Estimates using these methods with alternative bandwidths are shown in Table 5. A conservative bandwidth of $O(T^{1/2})$ was proposed by Geweke and Porter-Hudak, whereas broader bandwidth choices have been shown to be optimal on an MSE criterion by Hurvich et al. (1998). The 'bias test' is from Davidson and Sibbertsen (2009) and is a variant on the Hausman test, comparing broad and narrow bandwith estimates of d. This statistic is asymptotically standard normal under the null hypothesis of a pure fractionally integrated process.

What is notable in the present case is how very sensitive these estimates are to both the estimator and the bandwidth choice. This is a result that is predictable given the form of the autocovariance functions on which they are based, which are plotted in Figure 5. A linear representation of long-range dependence incorporating the fractional differencing operator would, on this evidence, have great difficulty in providing a complete description of these data series.

4 Linear autoregressive analysis

Econometric testing methodologies are normally embodied in a parametric time series framework and this almost always involves linear modelling. Granger non-causation hypotheses in stationary data are typically tested as restrictions on the coefficients of a vector autoregression (VAR). In nonstationary data, the nonstationarity is typically attributed to a unit autoregressive root, and in this case the usual approach is to construct a reduced rank VAR, otherwise known as a vector error correction model (VECM).

When applied to economic or financial data, the nature of the data is such that one of these two approaches is found to cover most eventualities. In particular, economic data sets typically appear quite compatible with the linear model paradigm. After allowing for factors such as

seasonal patterns, vector autoregressions, with or without unit roots, are found to model economic time series quite successfully and parsimoniously. From a theoretical point of view, appeal may be made to the Wold decomposition theorem (see e.g. Davidson 2000) which states that every stationary process has a linear moving average representation with respect to a sequence of uncorrelated disturbances. An autoregressive process of sufficiently high order can approximate this representation arbitrarily closely. This result does not exclude the possibility of unexplained dependence involving second or higher moments of the disturbances, but it does offer powerful support for a linear representation of the conditional mean.

With this in mind, consider modelling the data in Figure 4. Under a stationarity assumption, we should be able to represent this pattern by an autoregression of sufficiently high order. An ideal situation would be able to model the climatic interactions by a simple bivariate system in which one variable drives, or is driven by, the other. A finding of Granger non-causality would.provide unambiguous evidence about these interactions.

However, the rejection of Granger noncausality would have a number of possible interpretations. The usually accepted scenario is that the glacial cycle is the response of the observed processes to some geophysical factors relating to changes in solar output and in the Earth's orbit about the sun. The so-called Milankovitch cycles, regular variations in the eccentricity, obliquity and precession of the Earth's orbit can be reconstructed using a public domain computer program (see Paillard et al. 1996), and these can be included in the model as non-stochastic explanatory variables. The series in question are plotted in Figure 6.[6] Unobserved drivers, such as variations in solar output, are modelled as nominally stochastic, although fixed cycles might be captured in practice by nonstochastic difference equations, since it is not necssary to specify the distribution of the associated shock processes. In practice, these these unobserved factors need to be proxied by the observed data themselves, in the context of a bivariate VAR.

To show how this works, start with the complete system. Let $\boldsymbol{A}(L)$ $(m \times m)$ denote a finite-order matrix polynomial in the lag operator, $\boldsymbol{\Lambda}(L)$ a $m \times p$ matrix polynomial, and $\boldsymbol{u}_t$ a m-vector of zero-mean random shocks. Letting

$$\boldsymbol{A}(L)\boldsymbol{y}_t = \boldsymbol{\Lambda}(L)\boldsymbol{d}_t + \boldsymbol{u}_t$$

describe the evolution of the m-vector of random processes $\boldsymbol{y}_t$ with $\boldsymbol{A}_0 = \boldsymbol{I}_m$, where $\boldsymbol{d}_t$ $(p \times 1)$ denotes known cycles or trends in addition to a fixed intercept, partition this system as

$$\begin{bmatrix} \boldsymbol{A}_{xx}(L) & \boldsymbol{A}_{xz}(L) \\ \boldsymbol{A}_{zx}(L) & \boldsymbol{A}_{zz}(L) \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{z}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_x(L)\boldsymbol{d}_t \\ \boldsymbol{\Lambda}_z(L)\boldsymbol{d}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_{xt} \\ \boldsymbol{u}_{zt} \end{bmatrix} \quad \begin{matrix} (2 \times 1) \\ (m_z \times 1) \end{matrix} \tag{4.1}$$

where $m = 2 + m_z$, $\boldsymbol{y}_t = (\boldsymbol{x}_t', \boldsymbol{z}_t')'$, $\boldsymbol{x}_t = (T_t, C_t)'$ (2×1) where (according to our chosen functional form) $T_t = \log(\text{temperature } +16)$ and $C_t = \log(\text{CO}_2)$, and $\boldsymbol{z}_t$ $(m_z \times 1)$ is the vector of unobserved processes. Solving out $\boldsymbol{z}_t$ yields the reduced bivariate system

$$\boldsymbol{B}(L)\boldsymbol{x}_t = \boldsymbol{\Upsilon}(L)\boldsymbol{d}_t + \boldsymbol{v}_t \tag{4.2}$$

where

$$\boldsymbol{B}(L) = |\boldsymbol{A}_{zz}(L)|\boldsymbol{A}_{xx}(L) - \boldsymbol{A}_{xz}(L)\text{adj}\boldsymbol{A}_{zz}(L)\boldsymbol{A}_{zx}(L) \tag{4.3a}$$

$$\boldsymbol{\Upsilon}(L) = |\boldsymbol{A}_{zz}(L)|\boldsymbol{\Lambda}_x(L) - \boldsymbol{A}_{xz}(L)\text{adj}\boldsymbol{A}_{zz}(L)\boldsymbol{\Lambda}_z(L) \tag{4.3b}$$

[6] It is possible using the same software to compute series for imputed insolation at different latitudes. In this version of the model we have not included variables of this type since they introduced excessive collinearity. It may in any case be better to let the model determine how the orbital variations influence the target variables than to introduce an extra layer of imputation.

$$\boldsymbol{v}_t = |\boldsymbol{A}_{zz}(L)|\boldsymbol{u}_{xt} - \boldsymbol{A}_{xz}(L)\text{adj}\boldsymbol{A}_{zz}(L)\boldsymbol{u}_{zt} \tag{4.3c}$$

Consider first the case where $\boldsymbol{A}_{zx}(L) = \boldsymbol{0}$, such that $\boldsymbol{x}_t$ does not Granger-cause $\boldsymbol{z}_t$, and there are no feedbacks from observables to unobservables. In this case, notice that $\boldsymbol{B}(L) = |\boldsymbol{A}_{zz}(L)|\boldsymbol{A}_{xx}(L)$, and since $|\boldsymbol{A}_{zz}(L)|$ is a scalar polynomial, any zero restrictions on $\boldsymbol{A}_{xx}(L)$, such as a triangular or diagonal structure, are shared by $\boldsymbol{B}(L)$. It should be diagonal if temperature and CO_2 levels are unconnected, but we would expect at least an upper triangular structure if C_t drives T_t through the greenhouse effect.

System (4.2) has a vector ARMA structure, noting that the right-hand side term in (4.3c) is a sum of finite-order moving averages. As is well-known, a process so formed has a terminating autocovariance sequence and hence a representation as a finite-order moving average of white noise elements. In other words, there exists a vector of white noise elements $\boldsymbol{\varepsilon}_t = (\varepsilon_{Tt}, \varepsilon_{Ct})'$ such that

$$\boldsymbol{v}_t = \boldsymbol{\theta}(L)\boldsymbol{\varepsilon}_t. \tag{4.4}$$

To tie down the causality question more fully, we have to consider the role of the Milankovitch variables and the unobservables, represented here by the shocks $\boldsymbol{u}_{zt}$. If these drive both temperature and CO_2, there is no need for a greenhouse effect hypothesis to account for the observed co-movements of the variables. An alternative possibility is that these drive CO_2 only, so that the greenhouse hypothesis is needed to account for the subsequent changes in temperature. Consider this latter scenario. It arises if the first rows of, respectively, $\boldsymbol{A}_x(L)$ $(2 \times p)$ and $\boldsymbol{A}_{xz}(L)$ $(2 \times m_z)$ are zero. Writing $\boldsymbol{u}_{xt} = (u_{Tt}, u_{Ct})'$, the first element of $\boldsymbol{v}_t$ is in this case $|\boldsymbol{A}_{zz}(L)|u_{Tt}$, whereas u_{Tt} does not enter into the second element of $\boldsymbol{v}_t$. It follows that (4.4) could be partitioned as

$$\begin{bmatrix} v_{Tt} \\ v_{Ct} \end{bmatrix} = \begin{bmatrix} |\boldsymbol{A}_{zz}(L)| & 0 \\ 0 & \theta_{CC}(L) \end{bmatrix} \begin{bmatrix} u_{Tt} \\ \varepsilon_{Ct} \end{bmatrix}$$

where ε_{Ct} is a white noise disturbance derived from $\boldsymbol{u}_{zt}$ and u_{Ct}, and $\theta_{CC}(L)$ is the scalar lag polynomial defined implicitly by the autocovariance structure of v_{Ct}. Thus, we would be able to write system (4.2) in VAR form as

$$\boldsymbol{B}^*(L)\boldsymbol{x}_t = \boldsymbol{\Upsilon}^*(L)\boldsymbol{d}_t + \boldsymbol{\varepsilon}_t \tag{4.5}$$

where $\boldsymbol{B}^*(L) = \boldsymbol{\Theta}(L)^{-1}\boldsymbol{B}(L)$ and $\boldsymbol{\Upsilon}^*(L) = \boldsymbol{\Theta}(L)^{-1}\boldsymbol{\Upsilon}(L)$, and since $\boldsymbol{\Theta}(L)$ is a diagonal matrix, Granger non-causality restrictions on $\boldsymbol{B}(L)$ are preserved in $\boldsymbol{B}^*(L)$.

It follows that if the chain of causation runs in one direction, from $\boldsymbol{dt}$ and $\boldsymbol{z}_t$ to C_t and thus to T_t, this fact should be detectable by tests on the reduced VAR system. Temperature would not Granger-cause CO_2. In the alternative case where both C_t and T_t are driven by the unobservables, both rows of $\boldsymbol{A}_{xz}(L)$ contain non-zero terms, hence v_{Tt} depends on $\boldsymbol{u}_{zt}$ as well as u_{Tt}, and $\boldsymbol{\Theta}(L)$ is in general not diagonal. In this case Granger non-causality restrictions on the VAR system won't hold in general, regardless of the form of $\boldsymbol{A}_{xx}(L)$. There is of course the further possibility that $\boldsymbol{A}_{zx}(L) \neq \boldsymbol{0}$, because temperature and CO_2 themselves influence some relevant unobservables, and in this worst case the restrictions of interest may well be unidentified.

We conclude that analysis of the reduced system has the potential to be informative, but there is an asymmetry in respect of the conclusions obtainable from test outcomes. Granger non-causality of CO_2 by temperature in the VAR system (4.5) would be a strong finding, indicating that the cycles in temperature could be driven by the greenhouse effect, following movements in CO_2. Rejection of the hypothesis, on the other hand, while not contradicting the existence of a greenhouse effect, could be due at least in part to the unobservables driving the temperature cycle directly. We should need more information about the geophysical environment to be distinguish between these alternatives. There is of course the possibility of modelling the moving average

structure of (4.4) and so estimate $\boldsymbol{B}(L)$ directly. However, since the orders of lags involved are unknown and long, and the estimation of over-parameterised ARMA systems is tricky and subject to identification problems, this would be an ambitious exercise. It is not attempted here.

5 The VAR Model

Let r denote the maximum lag length for the system, to be chosen. Since the cyclical variables $\boldsymbol{d_t}$ vary smoothly over a relatively long time scale, we simplify the setup by including only current variables, replacing $\boldsymbol{\Upsilon}^*(L)$ by a constant matrix $\boldsymbol{\Upsilon}_0^*$. Next, rather than adopting the standard VAR format we choose the VECM form which is written as

$$\boldsymbol{\Gamma}(L)\Delta\boldsymbol{x}_{1t} = \boldsymbol{\Upsilon}_0^*\boldsymbol{d}_t + \boldsymbol{\Pi}\boldsymbol{x}_{1,t-1} + \boldsymbol{\varepsilon}_t. \tag{5.1}$$

where $\boldsymbol{\Gamma}_0 = \boldsymbol{I}_2$, $\boldsymbol{\Gamma}_j = -\sum_{k=j+1}^{r}\boldsymbol{B}_j^*$ for $j = 1, \ldots, r-1$ and

$$\boldsymbol{\Pi} = \begin{bmatrix} \pi_{TT} & \pi_{TC} \\ \pi_{CT} & \pi_{CC} \end{bmatrix} = -\boldsymbol{B}^*(1). \tag{5.2}$$

Note that the two parameterizations are equivalent, but the VECM setup has the benefit that it allows straightforward testing of the Granger non-causality hypotheses which are embodied in the restrictions $\pi_{CT} = 0$ and $\pi_{TC} = 0$ respectively. Under the restriction $\pi_{CT} = 0$, for example, the level of T_t, cannot affect the subsequent path of C_t. This single restriction does not rule out correlations between the changes of the two series, but such correlations cannot induce a systematic relationship between the levels of the variables. If, equivalently, the sum of the off-diagonal lag coefficients in the VAR form is zero, the response of the left-hand side variable to a step change in the other is likewise zero in the long run. Hence, we can base our tests on these coefficients alone. In the cases where the lag order exceeds 1 this is a weaker restriction than that all the corresponding lag coefficients are jointly zero, but it is the relevant restriction for the present purpose. If it is rejected, we may be sure, in a sufficiently large sample, that the stronger restriction will likewise be rejected.

Another useful feature of this parameterization is that it is equally appropriate, whether the data are stationary or feature stochastic trends. In the former case the matrix $\boldsymbol{\Pi}$ must have full rank while in the latter case it has reduced rank, say $s < 2$. In this case it is customary to write $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $2 \times s$ matrices such that $\boldsymbol{\beta}'\boldsymbol{x}_t$ is a stationary process and $\boldsymbol{\beta}$ is the matrix whose columns are the so-called cointegrating vectors. However, provided $s = 2$, which is the conclusion from the Johansen (1981) tests for cointegrating rank described in Section 3, we are able to assume that the elements of $\boldsymbol{\Pi}$ are asymptotically normally distributed. Standard asymptotic inference can be applied to tests of significance of these elements, provided the rank of the matrix is invariant under null and alternative hypotheses.

A further feature of these data, which becomes clear on attempting to specify a system of the form (5.1) for $\boldsymbol{x}_t = (T_t, C_t)$, is that a linear structure cannot account for all the dependence in the series. In particular, much larger shocks than usual appear to attend the periods of glacial/interglacial transition. Since these episodes of high disturbance volatility occur in clusters, a natural model to explain this pattern is generalized autoregressive conditional heteroscedasticity or GARCH (Bollerslev 1987).

Letting $E_{t-1}\boldsymbol{\varepsilon}_t = \boldsymbol{h}_t$ denote the conditional variance of the process $\boldsymbol{\varepsilon}_t$, the first-order multivariate GARCH equation takes the form

$$\boldsymbol{h}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}\boldsymbol{\varepsilon}_{t-1}^2 + \boldsymbol{\beta}\boldsymbol{h}_{t-1} \qquad (2 \times 1) \tag{5.3}$$

	ΔT_t	ΔC_t
Intercept	-0.683 (0.161)**	-0.009 (0.0054)
Eccentricity	0.374 (0.227)	0.059 (0.080)
Obliquity	0.035 (0.007)**	0.0005 (0.002)
Precession	-0.0233 (0.128)	0.047 (0.037)
T_{t-1}	-0.087 (0.032)**	0.021 (0.009)*
C_{t-1}	0.107 (0.065)	-0.066 (0.019)**
See Table 7 for dynamic terms		
$\sqrt{\text{GARCH } \omega}$	0.028 (0.0052)$^-$	0.015 (0.001)$^-$
GARCH δ	0.998 (0.036)**	0.733 (0.093)**
GARCH β	0.705 (0.061)**	0.421 (0.118)**
R^2 (levels)	0.912	0.962
Jarque-Bera Stat.	30.24	195.6
Box-Pierce (17)	18.15	18.58
Box-Pierce,Sq (25)	18.56	25.15

Table 6: VAR(8) + Garch(1,1) model with Gaussian likelihood, equations

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are 2×2 matrices. Letting $\hat{\boldsymbol{h}}_t$ denote the diagonal matrix in the elements $\boldsymbol{h}_t$, the shock process is represented as $\boldsymbol{\varepsilon}_t = \hat{\boldsymbol{h}}_t^{1/2} \boldsymbol{e}_t$ where by hypothesis $\boldsymbol{e}_t \sim$ i.i.d.$(\boldsymbol{0}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega}$ is a fixed covariance matrix. Since the GARCH model requires the distribution of the shocks to be symmetric, our decision to work with the logarithmically transformed variables is made on the basis of ensuring symmetry to the best approximation. Note the so-called "ARMA in squares" form of the GARCH model, having the representation

$$\boldsymbol{\varepsilon}_t^2 = \omega + \delta \boldsymbol{v}_{t-1}^2 + \boldsymbol{w}_t - \boldsymbol{\beta} \boldsymbol{w}_{t-1} \tag{5.4}$$

where $\boldsymbol{\delta} = \boldsymbol{\alpha} + \boldsymbol{\beta}$ and $\boldsymbol{w}_t = \boldsymbol{\varepsilon}_t^2 - \boldsymbol{h}_t$. This is a convenient form for estimation, and has the benefit that the eigenvalues of $\boldsymbol{\delta}$ define the covariance stationarity conditions for the system which can accordingly be monitored directly. $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are the parameters we report.

The results of estimating this model by Gaussian quasi-maximum likelihood are shown in Tables 6 and 8. Figures 7 and 8 show the residuals from the mean equation with two-standard error bands, before and after normalization by $1/\sqrt{\boldsymbol{h}_t}$. In Table 6, standard errors are shown in parentheses following the point estimates, with one and two stars indicating significance at the 5% and 1% levels respectively. These standard errors are computed using the robust formula for the covariance matrix, that is to say, from the diagonal elements of the matrix $\boldsymbol{V} = \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}^{-1}$ where $\boldsymbol{A}$ denotes the covariance matrix of the scores and $\boldsymbol{Q}$ the Hessian matrix of the criterion function. In other words, the information matrix equality $\boldsymbol{A} = -\boldsymbol{Q}$ is not invoked, as befits the case when Gaussianity of the disturbances is not being assumed.

The maximum lag length in the mean model, $r = 8$, has been chosen to maximize the Akaike (1969) selection criterion.[7] Notice that the lag coefficients are estimated unrestrictedly even if nominally insignificant. While truncating the maximum lag is a natural restriction, we take the view that suppressing intermediate lag coefficients is best avoided unless there is an explicit modelling justification, for example seasonal effects, otherwise such restrictions could distort the lag structure more seriously that allowing it to be estimated inefficiently. However,the off-diagonal

[7]The Akaike and Schwarz selection criteria are here defined in the form "log-likelihood less penalty", hence larger values point to preferred models. Note that some textbooks give a definition reversing the sign, as well as normalizing by sample size.

	ΔT_t	ΔC_t
ΔT_{t-1}	-0.008 (0.059)	0.045 (0.016)**
ΔC_{t-1}	0.875 (0.164)**	0.356 (0.060)**
ΔT_{t-2}	-0.117 (0.054)	0.035 (0.014)**
ΔC_{t-2}	0.095 (0.141)	-0.140 (0.049)**
ΔT_{t-3}	-0.026 (0.054)	0.034 (0.015)*
ΔC_{t-3}	-0.117 (0.136)	-0.108 (0.046)*
ΔT_{t-4}	-0.050 (0.046)	0.027 (0.013)
ΔC_{t-4}	0.060 (0.121)	-0.086 (0.041)*
ΔT_{t-5}	-0.055 (0.048)	0.010 (0.011)*
ΔC_{t-5}	0.104 (0.131)	0.032 (0.037)
ΔT_{t-6}	0.065 (0.052)	0.007 (0.013)
ΔC_{t-6}	-0.059 (0.128)	0.002 (0.037)
ΔT_{t-7}	-0.026 (0.052)	-0.0018 (0.013)
ΔC_{t-7}	0.031 (0.114)	0.007 (0.041)
ΔT_{t-8}	-0.014 (0.047)	-0.006 (0.014)
ΔC_{t-8}	0.110 (0.101)	0.009 (0.040)

Table 7: VAR(8) + Garch(1,1) model with Gaussian likelihood, dynamic terms

Log-likelihood:	3747.3
Akaike Criterion:	3696.7
Schwarz Criterion	3577.27
Contemporaneous Correlation	0.464
CM Autocorrelation Test (24):	13.64
CM Neglected ARCH Test (24):	28.76
CM Functional Form Test (4):	9.75*

Table 8: Var(8) + GARCH(1,1) model with Gaussian likelihood, system statistics

	ΔT_t	ΔC_t
Intercept	-0.663 (0.139)**	0.021 (0.04)
Eccentricity	0.291 (0.220)	0.0 (0.071)
Obliquity	0.039(0.006)**	-0.001 (0.002)
Precession	-0.266 (0.118)*	0.0402 (0.032)
T_{t-1}	-0.119 (0.029)**	0.024 (0.007)**
C_{t-1}	0.156 (0.060)**	-0.067 (0.015)**
See Table 10 for dynamic terms		
$\sqrt{\text{GARCH } \omega}$	0.027 (0.0041)	0.016 (0.001)
GARCH δ	0.955 (0.030)**	0.758 (0.071)**
GARCH β	0.698 (0.056)**	0.475 (0.094)**
R^2 (levels)	0.912	0.962
Jarque-Bera Stat.	36.01	473.4
Box-Pierce (17)	18.81	18.27
Box-Pierce,Sq (25)	19.73	25.19

Table 9: VAR(8) + Garch(1,1) Model with Student t likelihood, equations

elements of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ in (5.4) have been fixed at zero, due to strong indications of an identification problem.[8].

Table 6 included diagnostics for the individual equations, including R^2s (computed for the variables in levels, not differences), the Jarque-Bera (1980) test for residual Gaussianity (asymptotically chi-squared with two degrees of freedom under the null hypothesis) and also Box-Pierce (1970) tests for residual autocorrelation in levels and squares. Table 8 contains statistics computed for the model as a whole, including the contemporaneous correlation between the normalized residuals and diagnostic conditional moment tests, which on the null hypothesis of correct specification are asymptotically chi-squared with the degrees of freedom shown. The autocorrelation and neglected ARCH tests are based on the covariances between residuals and squared residuals, respectively, and their respective lags from both equations up to sixth order. The functional form test, which leads to a rejection at the 5% level, is based on the correlation between residuals and the squared fitted values for each equation. (See Newey (1985) and Tauchen (1985), and also Davidson (2000) for background on conditional moment tests.)

A noteworthy feature of these estimates is that the residuals appear distinctly heavy-tailed, as indicated by both the residual plots. This fact is also reflected by magnitudes of the Jarque-Bera test statistics, particularly in the case of CO_2. This finding suggests using the Student-t distribution to construct the likelihood function, with the degrees of freedom to be estimated as an additional parameter. The distribution is symmetric and the Gaussian enters as a special case in this set-up, but otherwise the form of the density function limits the influence of outlying data points in determining the fit. The available algorithm does not allow the degrees of freedom to differ between variables, so both of these estimation options represent a compromise. The Gaussian estimator is known to be consistent and asymptotically normally distributed under the usual regularity conditions, subject only to existence of at least second moments. The Student t estimator does not share this robustness property, and Newey and Steigerwald (1997) show that bias is possible in GARCH models in cases where the true shock distribution is asymmetric. In the event, there is no evidence of asymmetry and this second set of estimates, shown in Tables

[8]Fitting the off-diagonal elements of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ one at a time yielded small and insignificant values. Attempting to fit both at once resulted in a failure of the search algorithm, with the positivity restrictions on the $\boldsymbol{h}_t$ sequence apparently violated so that the likelihood could not be computed.

	ΔT_t	ΔC_t
ΔT_{t-1}	-0.171 (0.058)	0.046 (0.014)**
ΔC_{t-1}	0.957 (0.158)**	0.366 (0.052)**
ΔT_{t-2}	-0.1218 (0.051)	0.031 (0.013)**
ΔC_{t-2}	0.134 (0.140)	-0.117 (0.046)**
ΔT_{t-3}	-0.050 (0.050)	0.030 (0.014)
ΔC_{t-3}	-0.066 (0.125	-0.080 (0.043)*
ΔT_{t-4}	-0.041 (0.044)	0.023 (0.012)
ΔC_{t-4}	0.045 (0.119)	-0.080 (0.038)*
ΔT_{t-5}	-0.058 (0.046)	0.011 (0.010)
ΔC_{t-5}	0.151 (0.121)	0.035 (0.036)
ΔT_{t-6}	0.048 (0.048)	0.007 (0.011)
ΔC_{t-6}	-0.027 (0.120)	0.011 (0.036)
ΔT_{t-7}	-0.045 (0.119)	-0.012 (0.010)
ΔC_{t-7}	0.073 (0.116)	-0.011 (0.035)
ΔT_{t-8}	-0.017 (0.044)	-0.004 (0.010)
ΔC_{t-8}	0.097 (0.095)	0.0016 (0.032)

Table 10: VAR(8) + Garch(1,1) model with Student t likelihood, dynamic terms

Log-likelihood:	3769.2
Akaike Criterion:	3717.1
Schwarz Criterion	3595.75
Contemporaneous Correlation	0.468
$\sqrt{\text{Student } t \text{ degrees of freedom}}$	3.556 (0.313)
CM Autocorrelation Test (24):	20.58
CM Neglected ARCH Test (24):	22.68
CM Functional Form Test (4):	9.032

Table 11: Var(8) + GARCH(1,1) model with Student t likelihood, system statistics

9 and 11 (the standard errors are again computed by the robust formula) show at most minor differences from the Gaussian case.

6 The Markov-Switching Model

In the last section we modelled the data using a VAR structure to represent the conditional mean. This is the conventional modelling approach, but the autoregressive structure has to work very hard here to reproduce the cyclical pattern of the data. By hypothesis we have assumed that the unobserved process $\boldsymbol{z}_t$ in (4.1) is capable of generating such a pattern through purely linear interactions of uncorrelated shocks. By the Wold theorem this should be possible, accepting the hypothesis of stationarity, but the autocorrelation pattern of the data suggests a high order process, which is difficult to estimate in all but the longest realizations. Although optimizing the Akaike criterion in the available sample, the chosen lag length is almost certainly too short to capture such a pattern with accuracy.

It is therefore natural to consider a nonlinear modelling framework. Suppose that the glacial and interglacial periods can be thought of as the result of changes in geophysical conditions under which different dynamic interactions prevail. The stochastic mechanism that induces these shifts is unobserved, but we can deduce its key properties from the observed processes. The simplest mechanism that we can propose is a fixed probability of switching states conditional on the current state, the so-called Markov switching model. Since the original proposal by Hamilton (1989), Markov-switching has become a popular device in econometrics to capture the transitions between (for example) boom to recession episodes in the macro-economy.

Our second model therefore proposes different VAR-GARCH structures for the glacial and interglacial periods, although the VAR component is now much simplified, entailing dependence on just two lag terms. Otherwise, the functional forms in each regime follow equations (5.1) and (5.4), with the parameters allowed to differ in the glacial and interglacial periods as these are distinguished by the switching algorithm. It is not possible to infer with certainly that one or the other 'regime' prevails at a given date. In principle, the contributions to the likelihood function for each observation are the alternative probability densities weighted by unobserved indicators, taking the values one or zero depending which 'regime' currently prevails. In an application of the EM algorithm (Dempster, Laird and Rubin 1977), the indicators are replaced at each optimization step by their conditional means, the so-called filter probabilities, so that the likelihood contributions are constructed as a probability-weighted average of regime densities. The sum of (the logarithms of) these terms is maximized with respect to the model parameters. The filter probabilities are constrained to evolve according to the Markov updating rule, and the fixed transition probabilities are additional parameters that need to be estimated.

The Markov switching model is clearly a simplified version of reality just as is the linear autoregression, but the simplification is of a different form because the hidden mechanism takes the form of a discrete switch rather than a continuous, linear adjustment to a succession of shocks. The estimates of the Markov-switching model, which is fitted with Student t conditional densities (whose degress of freedom parameter does not switch), are shown in Tables 12, 13 and 14. Figures 9 and 10 show, respectively, the estimated filter probabilities of the interglacial regime, and their smoothed counterparts, the latter being computed using the full sample rather than just the data preceding the dates in question; see Kim andNelson (1999) for details.

The temperature equation appears to exhibit some residual autocorrelation, although it is important to bear in mind that the 'residuals' in these models are computed as the filter-probability weighted averages of the residuals computed for each regime. It is not clear what the properties of the usual tests will be in these cases. Nonetheless, the potential problem is alleviated by computing standard errors and tests using a heteroscedasticity and autocorrelation consistent (HAC)

		ΔT_t	ΔC_t
Regime 1	Intercept	−0.525 (0.159)	-0.065 (0.017)
(Inter-glacial)	Eccentricity	−0.038 (0.326)	−0.066 (0.110)
	Obliquity	0.029 (0.006)**	0.0034 (0.001)**
	Precession	−0.065 (0.100)	0.038 (0.032)
	T_{t-1}	−0.132 (0.049)**	0.011 (0.013)
	C_{t-1}	0.188 (0.084)*	−0.042 (0.023)*
	ΔT_{t-1}	0.359 (0.175)*	0.166 (0.080)*
	ΔC_{t-1}	0.398 (0.195)*	0.152 (0.090)
	$\sqrt{\text{GARCH } \omega}$	0.026 (0.004)$^-$	0.0085 (0.002)$^-$
	GARCH δ	0.848 (0.112)**	0.895 (0.084)**
	GARCH β	0.748 (0.244)*	0.564 (0.163)**
Regime 2	Intercept	0.464 (0.227)*	0.042 (0.037)
(Glacial)	Eccentricity	1.500 (0.458)**	0.042 (0.135)
	Obliquity	0.040 (0.009)**	−0.0017 (0.0016)
	Precession	−0.902 (0.230)**	−0.036 (0.071)
	T_{t-1}	−0.421 (0.057)**	0.039 (0.014)**
	C_{t-1}	0.552 (0.100)**	−0.117 (0.030)**
	ΔT_{t-1}	−0.054 (0.054)	−0.0014 (0.011)
	ΔC_{t-1}	1.020 (0.259)**	0.443 (0.064)**
	$\sqrt{\text{GARCH } \omega}$	0.035 (0.025)$^-$	0.017 (0.002)$^-$
	GARCH δ	0.975 (0.045)**	0.511 (0.154)**
	GARCH β	0.851 (0.065)**	0.302 (0.150)*
R^2 (levels)		0.924	0.963
Jarque-Bera.		11.26**	170.7**
Box-Pierce (24)		32.66	26.85
Box-Pierce, Sq (25)		21.11	22.34

Table 12: Markov Switching VAR(2) + Garch(1,1) model with Student t likelihood, equations

estimate of the covariance matrix. The Parzen kernel is used with a bandwidth of three, chosen by the Newey and West (1994) plug-in procedure.

7 Tests of Granger Causality

The focus of attention in these results is the coefficients of C_{t-1} in the ΔT_t equation and T_{t-1} in the ΔC_t equation. If these are significantly positive, this indicates influence of the levels of one variable on the future path of the other. Nonzero coefficients of the lagged differenced variables indicate the existence of temporary influences, but not permanent ones. Given the sympathetic behaviour of the two series, we should expect that at least one series "drives" the other, and

	P(·|Interglacial)	P(·|Glacial)
P(Interglacial|·)	0.9642	0.0341
P(Glacial|·)	0.038	0.9659

Table 13: Markov Switching VAR(2) + Garch(1,1) model with Student t likelihood, estimated transition probabilities

Log-likelihood:	3853.0
Akaike Criterion:	3804.0
Schwarz Criterion	3689.3
Contemporaneous Correlation	0.484
$\sqrt{\text{Student } t \text{ degrees of freedom}}$	4.416 (0.825)
CM Autocorrelation Test, (24):	42.6**
CM Neglected ARCH Test, (24):	13.81
CM Functional Form Test (4):	1.056

Table 14: Var(8) + GARCH(1,1) Model with Student t likelihood, system statistics

possibly each drives the other. The most interesting outcome would be the case where one series does not drive the other, in other words, Granger non-causality of one series with respect to the other.

Consider first the VAR(8) model, and in particular Table 9, although the Gaussian MLE yields essentially similar results. Using the usual asymptotic test criteria, it appears that non-causality is rejected in both directions, although at the 1% significance level in the case of CO_2 by temperature, while only at the 5% level in the case of temperature by CO_2. Indeed, the latter result is rather close to the borderline for non-rejection at 5%. However, such results are always problematic because of the uncertainty associated with asymptotic criteria in the face of possible model mis-specification. Therefore, we have attempted to provide supplementary evidence on these findings in the form of bootstrapped critical values. The method adopted was to use the fitted models to generate artificial series using randomly resampled residuals to provide the shock processes and the actual pre-sample observations to set initial conditions. The two equation residuals are resampled in pairs to preserve the contemporaneous correlations. Then, the model was estimated from the bootstrapped data set. This procedure was repeated 3000 times, allowing the distributions of the centred t statistics to be tabulated. These are the distributions of the statistics having the form $(\hat{\pi}^*_{ij} - \hat{\pi}_{ij})/\text{s.e.}(\hat{\pi}^*_{ij})$ where $\hat{\pi}^*_{ij}$ denotes the estimate from the bootstrap sample of an off-diagonal element of the $\boldsymbol{\Pi}$ matrix in (5.2), and $\hat{\pi}_{ij}$ is the 'pseudo-true' parameter represented by the estimate from the original sample. These provide the bootstrap estimates of the null distribution of the one-tailed t test, with rejection in the range of positive values. The relevant quantiles from these distributions are shown in Table 15 where the column headings indicate upper tail areas. The usual asymptotic quantiles from the standard normal distribution are also shown for comparison.

In addition to bootstrapping the VAR(8) model, a second set of experiments used the Markov-switching model to generate the bootstrap data. This option allows us to check on the effect on the test statistics of fitting a misspecified model, in this case, the VAR(8) model where the true data generation process (DGP) is assumed to be the Markov switching case. Both of these sets of results are reported in Table 15. It is apparent that, assuming the VAR model is correct, the bootstrap returns critical values that are close to the asymptotic ones. Assuming the true model is Markov-switching, the estimated quantiles turn out to be somewhat smaller, suggesting that asymptotic tests could tend to over-reject a true null hypothesis. However, none of these cases leads to different conclusions from the conventional test procedure. Arguably, this allows us to be reasonably confident that the findings with the VAR(8) model are robust.

Next, consider the Markov switching model. Here, we have the opportunity to test causality in each of the regimes, and so determine whether behaviour of the series is different in the glacial and interglacial periods. Indeed, this proves to be the case. In the interglacial regime, both effects are insignificant and of a similar magnitude. In the glacial regime, on the other hand,

	t statistic	DGP	10%	5%	1%
CO2(−1) in Temp equation	2.60	VAR(8)	1.23	1.59	2.25*
		Markov Sw.	2.16	2.60*	3.45
Temp(−1) in CO2 equation	3.342	VAR(8)	1.54	1.84	2.51*
		Markov Sw.	1.53	1.83	2.60*
N(0,1)			1.28	1.64	2.32

Table 15: Bootstrap critical values for causality tests in the VAR(8) model

	Regime	t statistic	DGP	10%	5%	1%
CO2(−1) in	Interglacial	2.23	Markov Sw.	1.00	1.36	2.13
Temp equation			VAR(8)	0.07	0.68	1.61
	Glacial	5.52	Markov Sw.	0.36	0.67	1.42*
			VAR(8)	−0.34	0.96	1.90*
Temp(−1) in	Interglacial	0.846	Markov Sw.	1.06	1.57	2.33
CO2 equation			VAR(8)	2.43	3.45	5.38
	Glacial	2.785	Markov Sw.	1.09	1.49	2.49*
			VAR(8)	1.62	2.77*	5.11
N(0,1)				1.28	1.64	2.32

Table 16: Bootstrap critical values for causality tests in the Markov switching model

both noncausality hypotheses are rejected at the 5% level, although in this case the effect of CO_2 on temperature appears larger. Refer to Figure 9 to see which regime is estimated to predominate in each historical period.

Bootstrap critical values were also computed for this model and reported in Table 16, although in this case a modified version of the fitted model has been used in which the transition probabilities are fixed at their sample values, rather than being fitted freely along with the other parameters. Without such a strategy, there is quite a large risk of convergence failure with these models, due to the often poorly conditioned form of the likelihood contours. Even with this option in place, 600 out of the 3000 replications resulted in convergence failure, so that the reported results are based on the 2400 successful cases.

Using the Markov-switching model as DGP, the simulation results prove to depend rather critically on the regime. The interglacial estimates are distributed reasonably close to the asymptotic case, whereas the distributions for the temperature equation in the glacial period are notably different, and point to the possibility of under-rejection using asymptotic criteria. Evidently, the data are somewhat less informative about temperature in the glacial periods. Nonetheless, the test outcomes are the same as from the asymptotic criteria, suggesting that these findings are reasonably robust.

The table also shows the results of estimating the misspecified Markov-switching model, where the generated data are obtained from the VAR(8). Here there is a more serious problem, due to the underidentification of regime-dependent parameters when there is in fact just a single regime. The extreme distortion of the test distributions is notable, interestingly affecting the two equations in rather different fashion. These results might suggest that the Markov switching model, in spite of giving a better and more parsimonious fit to the data on conventional criteria, may offer a greater hazard to valid inference if it is misspecified, and the test results should be treated with relatively more caution than in the linear case.

8 Concluding Remarks

This paper has applied econometric techniques to the problem of testing directions of causation from historical data on temperature and atmospheric CO_2 derived from ice-core measurements. There are a number of issues with the study of data of this type, not least the problem of irregularly spaced and non-coincident observations. We have concluded that 1000 years is the shortest acceptable interval over which to create interpolated observations, while recognising that there are a number of episodes over which the CO_2 measurements enter two or more consecutive data points. It is inevitable that this data feature will induce some artefactual autocorrelation in the series, and may distort the short-run dynamics as estimated in these models. In Tables 7 and 10 we note that the change in CO_2 lagged one period takes a relatively large coefficient.in both equations. This could represent an important short-run effect, but might also be simply an artefact due to sparse data. On the other hand, going to 1500 or 2000 year intervals would run the risk of aliasing the interactions in which we are primarily interested. Hopefully, the modelled dynamics are able to account for such data artefacts as well as the historical interactions.

We have constructed two alternative models that both fit the data reasonably well while offering different approximations to the actual dynamic structure, each with its own limitations. The general conclusion we draw, taking the results together, is that there exists a two-way interaction, and the glacial-interglacial cycles exhibited by both series are clearly driven by a common underlying geophysical cause. Contrasting the two sets of estimates obtained, we should note that Markov-switching systems contain parameters whose identification depends on the existence of clearly defined regime differences. This can make them numerically challenging to estimate and there is typically a good deal of specification uncertainty. The linear-in-mean specification is cruder and more obviously descriptive, but it is also more numerically robust and our bootstrap calculations indicate that the tests of causality could be more or less correctly sized even when the model is wrong. In other words, the linear form can approximate the switching process adequately.from the point of view of inference on the restrictions of interest. This finding increases our confidence in the conclusions.

However, we should conclude with a cautionary remark regarding interpretation of the findings. For the reasons discussed in Section 4, the rejection of Granger non-causality means that without additional information about the climate system, we are not able to conclude very much about the direct interactions of our observable variables. Consider, for example, using the equations to simulate the effects of a shock to CO_2 levels. This would certainly show a response by temperature, just as a shock to temperature would show a response by CO_2. However, our inability to impose ceteris paribus conditions means that such an exercise could tell us little about the outcome of a controlled experiment, such as the injection of anthopogenic CO_2 into the atmosphere.

References

[1] Akaike, H (1969), Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 21, 243-247.

[2] Barnola , J-M, P. Pimienta, D. Raynaud and Y. S. Korotkevich (1991) CO_2-climate relationship as deduced from th Vostok ice core: a re-examination based on new measurements and on a re-evaliation of the air dating. *Tellus* 43B, 83-90

[3] Berger, A. L. (1978) Long-term variations of daily insolation and Quaternary climatic change. *J. Atmos. Sci.* 35, 2362–2367.

[4] Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics,* 31, 307-327.

[5] Box, G. E. P., and D. A. Pierce (1970). The distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American Statistical Association* 5, 1509 26.

[6] Caillon, N et al. (2003) Timing of Atmospheric CO2 and Antarctic temperature changes Across Termination III. *Science* 299, 1728-1731

[7] Cavaliere, G. (2005) Limited time series wuth a unit root. *Econometric Theory* 21(5), 907-945.

[8] Cavaliere, G. and F. Xu (2011) Testing for unit roots in bounded time series. Working Paper, University of Bologna.

[9] Davidson, J. (2000) *Econometric Theory.* Blackwell Publishers.

[10] Davidson, J. and P. Sibbertsen (2009) Tests of bias in log-periodogram regression. *Economics Letters* 102, 83-86.

[11] Davidson, J. (2013) *Time Series Modelling 4.37* at http://www.timeseriesmodelling.com.

[12] Doornik, J. (2009) *An Object-oriented Matrix Programming Language - Ox 6.* London, Timberlake Consultants Ltd.

[13] Elliott, G., T. J. Rothenberg and J. H. Stock (1996) Efficient tests for an autoregressove unit root, *Econometrica* 64 (4), 813-836

[14] Fischer, H., M Wahlen, J. Smith, D. Mastroianni and B. Deck (1999) Ice core records of atmospheric CO2 around the last three glacial terminations. *Science* 283, 1712-1714.

[15] Geweke, J. and S. Porter-Hudak (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221-237.

[16] Giraitis, L., P. Kokoskka, R. Leipis and G. Teyssiere (2003) Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* 112, 265-294.

[17] Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37, 424-38.

[18] Granger, C. W. J. (2010) Some thoughts on the development of cointegration. *Journal of Econometrics* 158, 3-6.

[19] Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57(2) 357-384.

[20] Hamilton, J. D. and R. Susmel (1994) Autoregressive conditional heteroscedasticity and changes in regime. *Journal of Econometrics* 64, 307-333

[21] Harris, D, B. McCabe and S. Leybourne (2008) Testing for long memory. *Econometric Theory* 24(1), 143-175.

[22] Hurvich, C. M., R. Deo and J. Brodsky (1998) The mean squared error of Geweke and Porter-Hudak's estimator of a long memory time series, *Journal of Time Series Analysis* 19, 19-46.

[23] Imbrie, J. et al. (1992) On the structure and origin of major glaciation cycles. 1. Linear responses to Milankovich forcing. *Paleoceanography* 7, 701–738 (1992).

[24] Jarque, C. M., and A. K. Bera (1980). Efficient tests for normality, heteroskedasticity and serial independence of regression residuals, *Economics Letters*, 6, 255-59.

[25] Johansen, S. (1991). Estimation and hypothesis testing of cointegration in Gaussian vector autoregressive models, *Econometrica* 59, 1551-80.

[26] Jouzel, J. et al, (2003) Magnitude of isotope/temperature scaling for interpretation of central Antarctic ice cores. *Journal of Geophysical Research*,108, D12, 4631 doi:10 1029/2002JD002677

[27] Jouzel, J. et al. (2007) Orbital and Millennial Antarctic Climate Variability over the Past 800,000 Years, *Science* 317, 793-796.

[28] Kaufmann, A. and D. I. Stern, (2002) Cointegration analysis of hemispheric temperature relations, *J. Geophys. Res.*, 107, ACL8.1–8.10.

[29] Kaufmann, A., H. Kauppi,. and J. H. Stock (2006) Emissions, concentrations and temperature: a time series analysis, *Climatic Change* 77, 248–278.

[30] Kaufmann, A., H. Kauppi and J. H. Stock (2010) Does temperature contain a stochastic trend? *Climatic Change* 101, 395–405.

[31] Kim, C.-J. and C. R. Nelson (1999) *State-space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications* MIT Press.

[32] Kunsch, H. R. (1987) Statistical aspects of self-similar processes, *Proceedings of 1st World Congress of the Bernoulli Soc.* (Eds. Yu Prohorov and V. V. Sazanov) VNU Science Press, Utrecht 1, 67-74.

[33] Kwiatkowski, D.., P. C. B. Phillips, P. Schmidt and Y. Shin (1992) Testing the null of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159-178

[34] Lo, A W. (1991) Long-term memory in stock market prices, *Econometrica* 59, 5, 1279-1313

[35] Lobato, I. N. and P. M.. Robinson (1998) A Nonparametric Test for I(0), *Review of Economic Studies* 65 (3), 475-495.

[36] L. Loulergue et al. (2007) New constraints on the gas age-ice age difference along the EPICA , ice cores, 0–50 kyr *Climate of the Past* 3, 527-540.

[37] Lüthi, D. et al. (2008) High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* 453, 379-382.

[38] Mills, T. C.(2007) Time series modelling of two millenia of northern hemisphere temperatures: long memory or level shifts, *J. Roy. Stat. Soc.* A 170, 83–94.

[39] Mudelsee, M. (2001) The phase relations among atmospheric CO2 content, temperature and global ice volume over the past 420 ka. *Quaternary Science Reviews* 20, 583-589.

[40] Newey, W. K. (1985) Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53, 1047-70.

[41] Newey, W. K. and D. G. Steigerwald (1997) Asymptotic bias for quasi-maximum likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65(3), 587-599.

[42] Newey, W. K., and K. D. West (1994). Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* 61, 631-653

[43] Paillard, D., Labeyrie, L., and Yiou, P. (1996), Macintosh program performs time-series analysis, Eos Trans. AGU, 77, 379.

[44] Parrenin F.„ T. D. Van Ommen, and E.W. Wolff (2007) Preface, The EPICA (EDC and EDML) ice cores age scales *Climate of the Past* Special Issue, pp 1-3.

[45] Phillips, P.C. B. and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika* 75, 335-346.

[46] Robinson, P. M. (1977) Estimation of a time series model from unequally spaced data. *Stochastic Processes and their Applications* 6, 9-24.

[47] Robinson, P. M. (1995) Gaussian semiparametric estimation of long-range dependence, *Annals of Statistics* 23, 1630–1661

[48] Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika*, 71, 599-607.

[49] Tauchen, G. (1985) Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30, 415-43

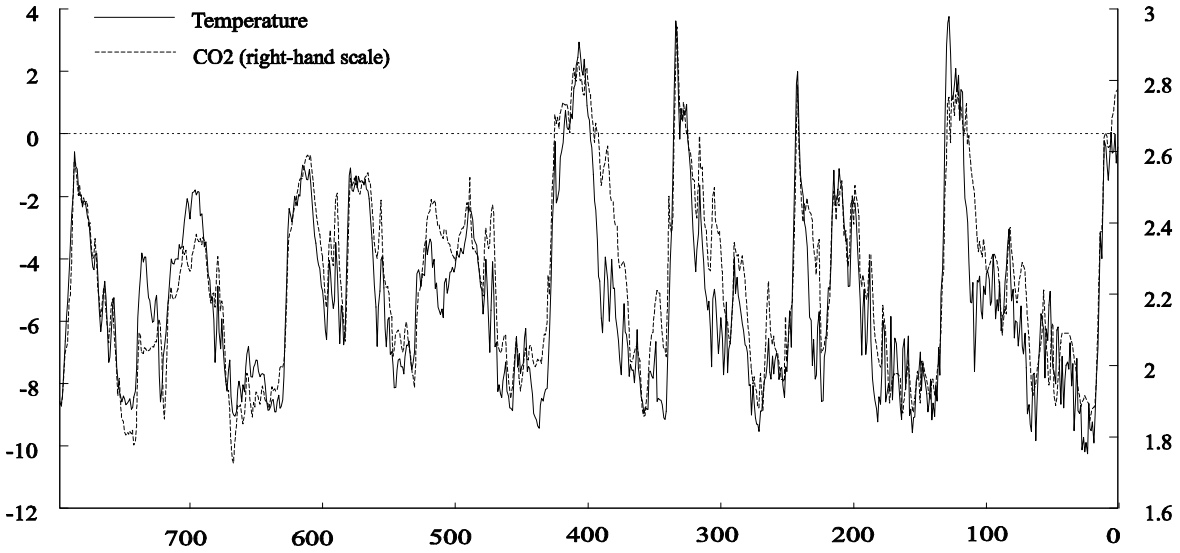

Figure 1: Temperature and CO2 series plotted at 1000 year intervals by linear interpolation.

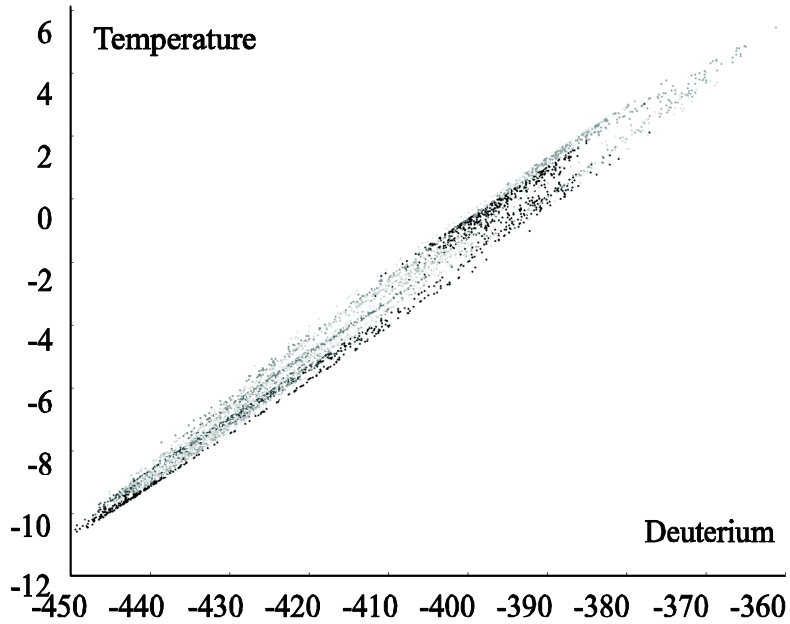

Figure 2: Scatter plot of deuterium measurements vs. imputed temperatures. (Darkest points are the most recent, lightest are earliest.)

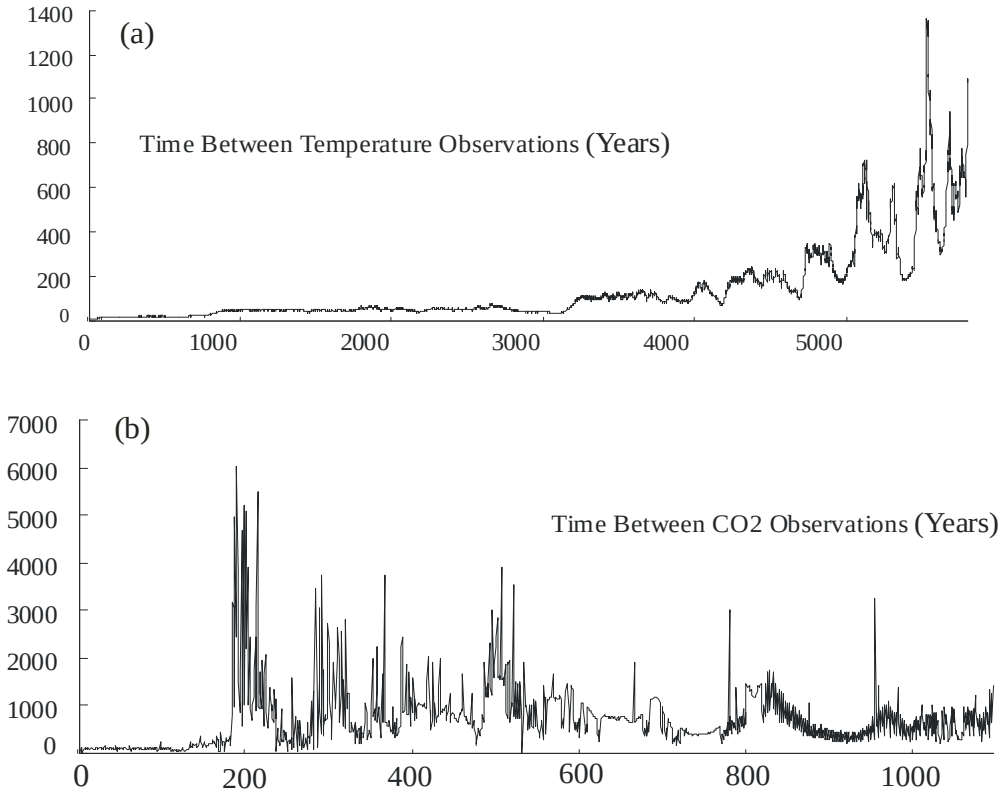

Figure 3: Time between observation dates

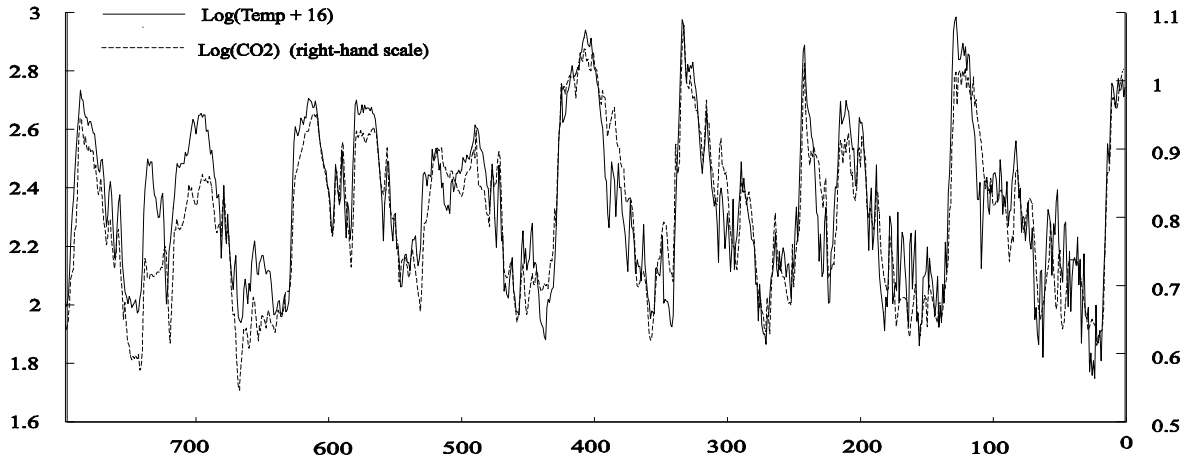

Figure 4: Logarithmically transformed series

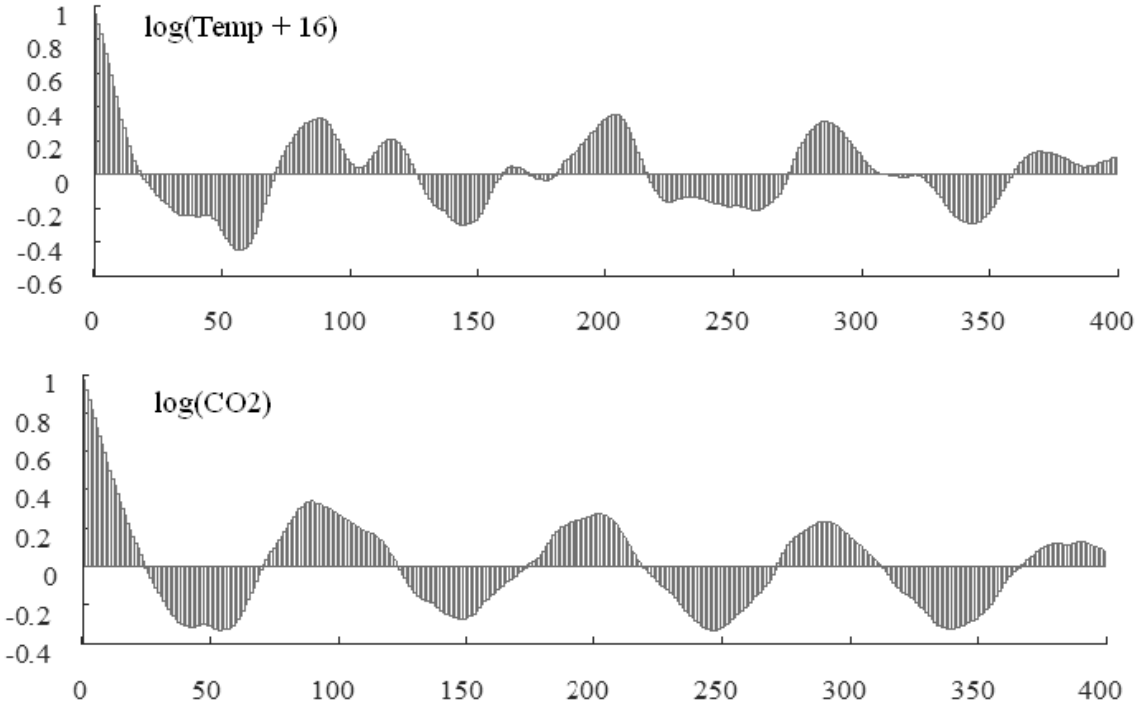

Figure 5: Autocorrelation Functions

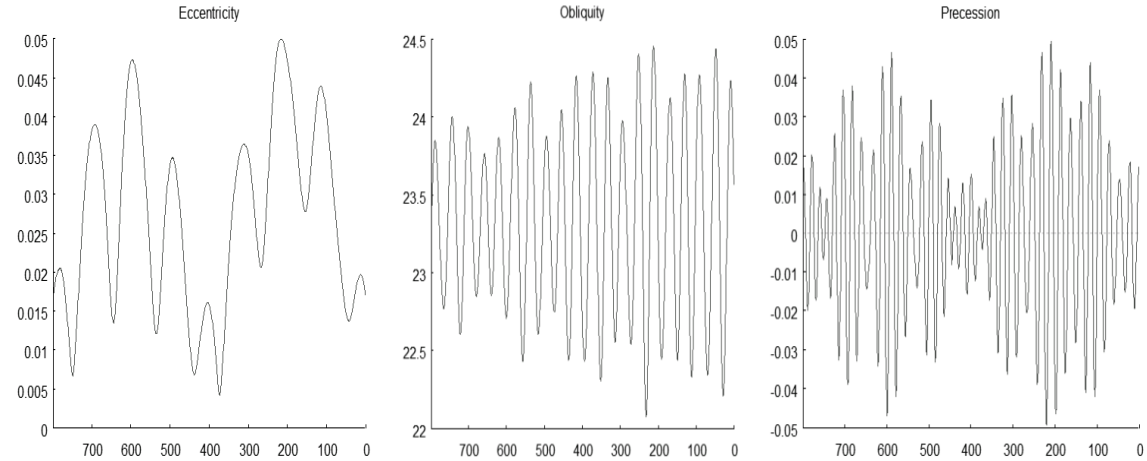

Figure 6: Orbital variables

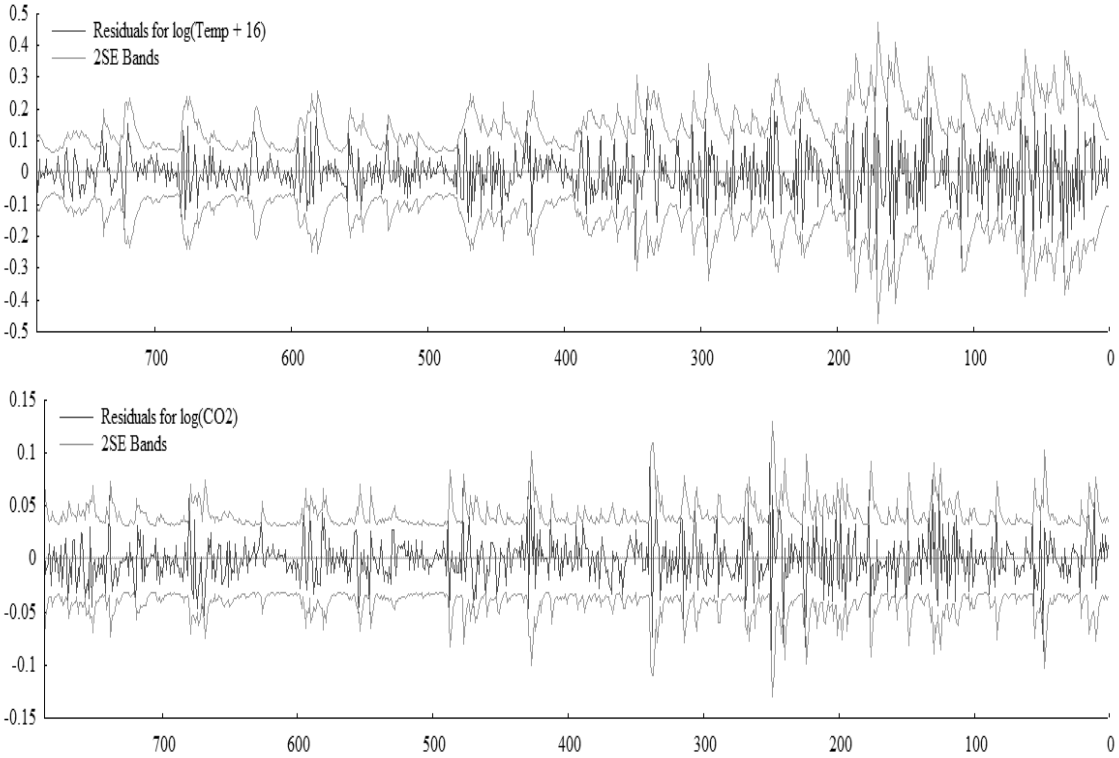

Figure 7: Residuals from the VAR(8) Model

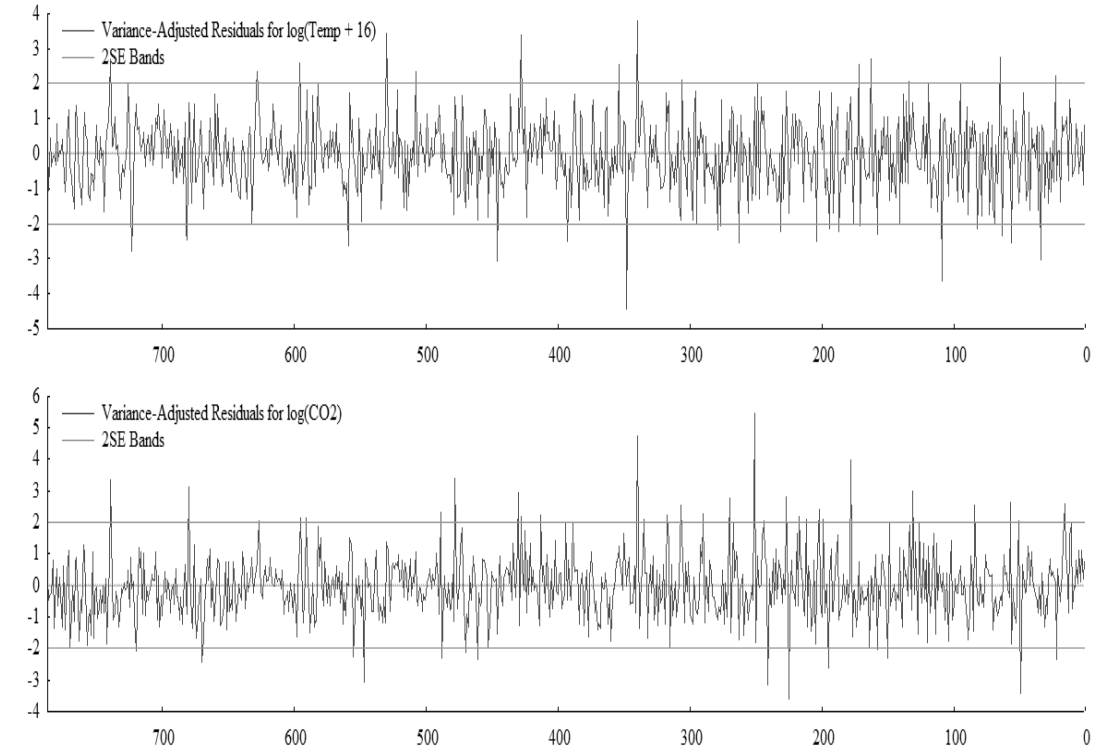

Figure 8: Variance-adjusted residuals from the VAR(8) model, Table 6

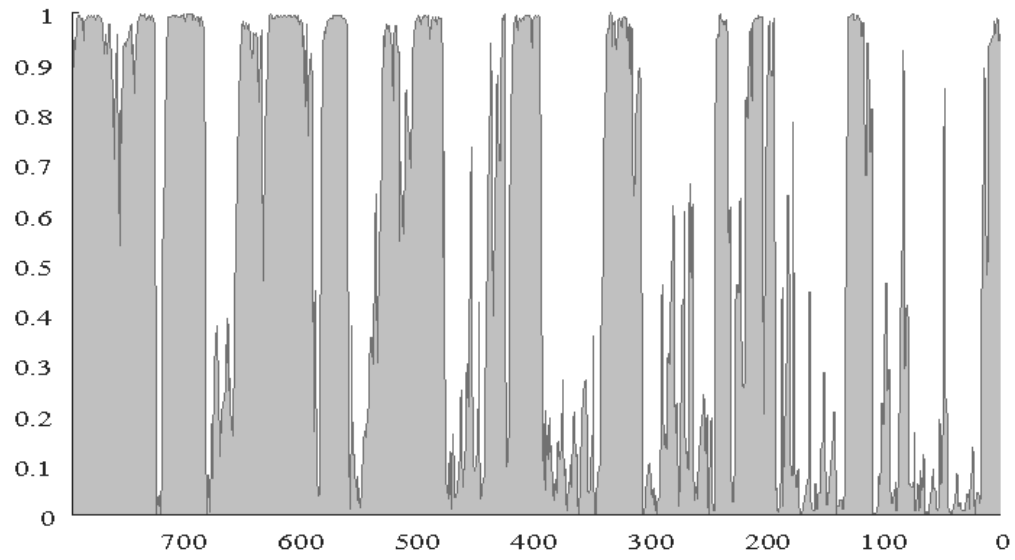

Figure 9: The filter probabilities of Regime 1 (inter-glacial).

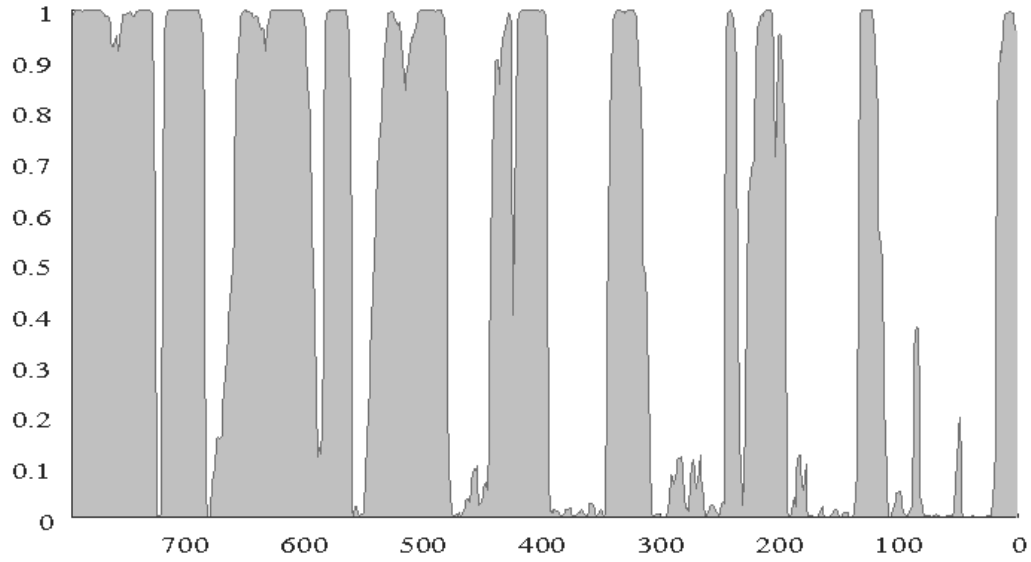

Figure 10: Smoothed probabilities of Regime 1 (inter-glacial)