Econometric Methods and Applications in Modelling Non-Stationary Climate Data

Felix Pretis

Nuffield College & Programme for Economic Modelling
Institute for New Economic Thinking
at the Oxford Martin School
University of Oxford

Thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in Economics

Submitted in Michaelmas 2014
Awarded in Trinity 2015

CONTENTS

PREFACE

The DPhil has been a highly non-stationary process, I have encountered many shifts in new and fascinating directions not anticipated at the start. I am thankful to all those who provided the necessary intercept-corrections to steer this thesis towards completion.

My deepest thanks go to my supervisor, David Hendry, who has been a constant source of inspiration and support throughout. He has made me strive to be a better academic, researcher, and teacher. Without his never-failing trust, encouragement, optimism, and eagerness to explore new ideas this thesis would not have been realized.

It has been an honour to be part of the academic family in Oxford and the wider world – from my first conference presentation in Washington, to seminars, and teaching courses from Cuba to Japan. I am very grateful to Jennifer Castle, Bent Nielsen, Jurgen Doornik, Vanessa Berenguer-Rico, Tony Atkinson, Janine Aron, John Muellbauer, Steve Bond, Cameron Hepburn, and Myles Allen, for many fruitful discussions, collaborations, and support.

In the wider world of academia I am indebted to Robert Kaufmann who has been extremely generous with his time and always enthusiastic about new projects, Michael L. Mann, David I. Stern, Jason Smerdon (for initiating the process that has lead to work on volcanic eruptions), Lea Schneider, Katarina Juselius, Søren Johansen, Niels Haldrup, Eric Hillebrand, Timo Teräsvirta, Anders Rahbek, Sven Crone, Roger Fouquet (for being excited about what turned out to be the first publication in this thesis), Neil Ericsson (for extremely detailed comments and very enjoyable conversations on scuba diving), Gunnar Bårdsen, Roger Hamersland, Andre Kallak Anundsen, Heike Langenberg (for being a very supportive editor at *Nature Geoscience*), Peter Thorne, Guillaume Chevillon, and Neil Ferguson (for his invitation to Imperial College, exploring the world of bio-statistics). It has been extremely enjoyable working on the *isat* function with Genaro Sucarrat and James Reade. I also thank James for his teaching material.

Coming into Eagle House on a Monday morning has been some of the best time of the DPhil thanks to Angela Wenham, Max Roser, Salvatore Morelli, Sebastian Königs, Andrew Martinez, Christine Francis, and Oleg Kitov. The meetings and coffee breaks have lead to many great (and sometimes even workable) ideas.

I would also like to thank the participants of the Gorman workshops, econometric seminars, and conferences including: OxMetrics, the International Meeting on Statistical Climatology, the first International Workshop on Econometric Applications in Climatology, the EMoD Econometrics conference, the International Surface Temperature Initiative at the National Center for Atmospheric Research, and the NCAS – NERC ES4 programme.

Teaching has deepened my understanding of econometrics and statistics – and I am grateful to my students from Oxford, Havana, and Akita, for stimulating questions – ranging from modelling Cuba's economy following the collapse of the Soviet Union, to modelling insect acoustics (with a special note to Olga Tserej Vasquez). In particular I am thankful to Hiroya Ichikawa for inviting me to teach at Akita International University, and Lazaro Peña Castellanos for bringing me to Cuba.

Special thanks have to be given to Kohei Kawamura who was the first to encourage me to go to Oxford, and Margaret Meyer for switching supervisees with David Hendry at the beginning of my time at Oxford – without both of them, the path of my academic career may have been a different one.

Without funding from the Oxford Martin School, the Open Society Foundations, and Nuffield College, this DPhil would have been impossible - it is a pleasure to thank these funding bodies, and James Martin himself for interest in this project.

I would like to thank Nicholas Chesterley, Claudia Herresthal, David Millican, Sarah-Wilkins Laflamme, Christine Matthey, Stephanie Schmölzer, Dieter Turk, Tania Thiel, Julius Pirklbauer, and Maggie Ziriax for their advice, support, and friendship.

The greatest appreciation goes to my family, Ulrike Kern, Manfred Pretis, Aleksandra Dimova, Herwig Platzer, Elfriede and David Pretis, Gert Kern, and Toffl, without whom I would not have reached this point.

Felix Pretis
Nuffield College
December 2014

ABSTRACT

Understanding of climate change and policy responses thereto rely on accurate measurements as well as models of both socio-economic and physical processes. However, data to assess impacts and establish historical climate records are non-stationary: distributions shift over time due to shocks, measurement changes, and stochastic trends - all of which invalidate standard statistical inference. This thesis establishes econometric methods to model non-stationary climate data consistent with known physical laws, enabling joint estimation and testing, develops techniques for the automatic detection of structural breaks, and evaluates socio-economic scenarios used in long-run climate projections.

Econometric cointegration analysis can be used to overcome inferential difficulties stemming from stochastic trends in time series, however, cointegration has been criticised in climate research for lacking a physical justification for its use. I show that physical two-component energy balance models of global mean climate can be mapped to a cointegrated system, making them directly testable, and thereby provide a physical justification for econometric methods in climate research.

Automatic model selection with more variables than observations is introduced in modelling concentrations of atmospheric CO_2, while controlling for outliers and breaks at any point in the sample using impulse indicator saturation. Without imposing the inclusion of variables *a-priori*, model selection results find that vegetation, temperature and other natural factors alone cannot explain the trend or the variation in CO_2 growth. Industrial production components, driven by business cycles and economic shocks, are highly significant contributors.

Generalizing the principle of indicator saturation, I present a methodology to detect structural breaks at any point in a time series using designed functions. Selecting over these break functions at every point in time using a general-to-specific algorithm, yields unbiased estimates of the break date and magnitude. Analytical derivations for the split-sample approach are provided under the null of no breaks and the alternative of one or more breaks. The methodology is demonstrated by detecting volcanic eruptions in a time series of Northern Hemisphere mean temperature derived from a coupled climate simulation spanning close to 1200 years.

All climate models require socio-economic projections to make statements about future climate change. The large span of projected temperature changes then originates predominantly from the wide range of scenarios, rather than uncertainty in climate models themselves. For the first time, observations over two decades are available against which the first sets of socio-economic scenarios used in the Intergovernmental Panel on Climate Change reports can be assessed. The results show that the growth rate in fossil fuel CO_2 emission intensity (fossil fuel CO_2 emissions per GDP) over the 2000s exceeds all main scenario values, with the discrepancy being driven by underprediction of high growth rates in Asia. This underestimation of emission intensity raises concerns about achieving a world of economic prosperity in an environmentally sustainable fashion.

STATEMENT OF PUBLICATION, CO-AUTHORSHIP, AND WORDCOUNT

Some chapters of this thesis have been published or are currently under review at peer-reviewed journals. Most of the work has also been presented at seminars and conferences. Parts of this thesis were written as co-authored papers. The following section provides an overview of publication and co-authorship.

CHAPTER 1

Chapter 1 merges two published papers to highlight hazards in econometric models of climate change. The papers are:

- Pretis, F. and Hendry, D.F. 2013. "Some Hazards in Econometric Models of Climate Change". *Earth System Dynamics*. 4, 375-484.
- Pretis, F. and Allen, M.R. 2013. "Climate Science: Breaks in Trends". *Nature Geoscience*. 6, 992-993.

The first paper, Pretis & Hendry (2013), is co-authored with David F. Hendry (University of Oxford, Department of Economics). I conducted the analysis of the climate data, D.F. Hendry provided the overview of hazards applied to an example of car deaths. The second paper, Pretis & Allen (2013), is co-authored with Myles R. Allen (University of Oxford, Environmental Change Institute and Department of Physics). M. R. Allen provided the calibrated climate model. The paper, together with the work by Estrada et al. (2013), was covered by *BBC News*, *National Geographic*, *Science News*, *The Australian*, *ORF*, *Der Focus*, *Der Spiegel*, *Der Tagesspiegel*, and the radio station *Deutschlandfunk*.

I am grateful to David F. Hendry, Heike Langenberg, Stephen Smith, Vanessa Berenguer-Rico, Bent Nielsen, Margaret Ziriax, and four anonymous referees for comments on the papers.

CHAPTER 2

Chapter 2 is single-authored and has been presented at the EMoD Econometrics Conference at the Oxford Martin School in September 2014, and at the Econometrics Seminar at the University of Copenhagen in November 2014.

I thank David Hendry, Robert Kaufmann, Janine Aron, Niels Haldrup, Søren Johansen, Katarina Juselius, Eric Hillebrand, Anders Rahbek, Bent Nielsen, Heine Bohn Nielsen, Andrew Martinez, and Max Roser for helpful comments and suggestions.

CHAPTER 3

Chapter 3 has been published and originated out of my MPhil Thesis for the Degree of Master of Philosophy in Economics at the University of Oxford (awarded in Trinity Term 2012). The published paper is co-authored with David F. Hendry and has appeared as:

- Hendry, D.F. and Pretis, F. 2013. Anthropogenic Influences on Atmospheric CO_2. in *Handbook on Energy and Climate Change*. R. Fouquet ed. Edward Elgar.

An earlier version of the chapter is available as the Oxford Economics Discussion Paper:

– Hendry, D.F. and Pretis, F. 2011. Anthropogenic Influences on Atmospheric CO_2. *Department of Economics Discussion Paper 584*.

The primary parts originally contributed by David F. Hendry to the chapter (sections 3.3.2 and 3.3.3) have been re-written for this thesis. The research has also been presented at the 11^{th} *OxMetrics* Users conference Washington DC 2012, as well as at the 1^{st} *International Workshop on Econometric Applications in Climatology* (University of Guelph), the Oxford Econometrics Workshop, and Gorman Research Workshop.

I am thankful to Jennifer Castle, Bent Nielsen, Jurgen Doornik, Steve Bond, participants at the 11^{th} *OxMetrics* Users Conference (George Washington Unversity, March 2012), and Oxford Gorman and Econometrics Workshops for helpful comments and suggestions. I thank Roger Fouquet for useful comments and Duo Quin for helpful notes on Chinese production.

CHAPTER 4

Chapter 4 is co-authored with Lea Schneider (Johannes Gutenberg University, Department of Geography) and Jason E. Smerdon (Columbia University, Lamont-Doherty Earth Observatory). I developed the methodology, conducted the analysis, derived the energy balance model formulation of the break function, and drafted the paper. The idea for the application, the volcanic data, and overview of volcanic impacts were contributed by Lea Schneider and Jason E. Smerdon. All authors were involved in editing the final paper. I presented the paper at the CREATES seminar at Aarhus University, the 14^{th} OxMetrics Conference at George Washington University, and at the National Center for Atmospheric Research in Boulder Colorado, as part of the SAMSI (Statistical and Applied Mathematical Sciences Institute) and IMAGe (Institute for Mathematics to Applied Geosciences) workshop on the International Surface Temperature Initiative (ISTI).

I thank David Hendry, Max Roser, Oleg Kitov, Bent Nielsen, Søren Johansen, Niels Haldrup, Eric Hillebrand, John Muellbauer, Timo Teräsvirta, Guillaume Chevillon, Bent Nielsen, participants of the CREATES Aarhus University seminar, participants of the 14^{th} OxMetrics Conference, and participants at the NCAR/SAMSI/IMAGe workshop for helpful comments and suggestions.

CHAPTER 5

Chapter 5 is co-authored with Max Roser (University of Oxford, Institute for New Economic Thinking). The paper was written jointly. M. Roser produced the downscaling data and I provided the growth rate decompositions and draft versions of non-spatial graphs.

I am grateful to Tony Atkinson, Sven Crone, Alexis Grigorieff, Neil Ericsson, Cameron Hepburn, David F. Hendry, Oleg Kitov, Andrew Martinez, Peter Thorne, and Steven Smith for helpful comments and suggestions.

EPILOGUE (CHAPTER 6)

The epilogue (chapter 6) is single-authored.

ADDITIONAL MATERIAL

Not all the work conducted during my DPhil has been included in the final thesis. These projects, however, are closely linked to the work presented here. They include the work on break detection through the use of step-indicator saturation (SIS), which resulted in two papers:

- Doornik J.A, D.F Hendry, and F. Pretis. 2013. "Step Indicator Saturation". *Department of Economics Discussion Paper 658*
- Castle, J.L., Doornik J.A, D.F Hendry, and F. Pretis. 2015. "Detecting Location Shifts by Step-Indicator Saturation during Model Selection". *Econometrics.* 3:2.

This work is also reflected in the ongoing development of implementing indicator saturation based break detection into the statistical software environment *R* through the function *isat* in the *gets* package, jointly with Genaro Sucarrat (Norwegian Business School) and James Reade (University of Reading). This will allow the wider community to use the proposed methods, ranging from impulses, to step indicators, and designed break functions.

Together with the work on designed break functions (Chapter 4), the work on SIS has also led to an additional (single-authored) working paper on non-linear estimation and the detection of breaks:

- Pretis, F. 2014. "Semiparametric Indicator Saturation".

The detection of breaks can also be a valuable tool in assessing relative model performance. Together with Robert Kaufmann (Boston University, Department of Earth & Environment) and Michael L. Mann (George Washington University, Department of Geography) I investigate this for competing climate models of the recent slowdown in observed warming in:

- Pretis, F., Mann, M.L. and Kaufmann, R. K. 2015. "Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection". forthcoming in *Climatic Change.*

WORD COUNT

This thesis comprises 172 pages (including references) with an average page containing around 430 words, resulting in an approximate total document length of 73,500 words.

1

INTRODUCTION: HAZARDS IN ECONOMETRIC MODELS OF CLIMATE CHANGE

Abstract

Econometric tools based on cointegration and model selection have the potential to contribute to existing climate research through rigorous observation-based analyses, however, these have to be conducted carefully to avoid statistical hazards. The importance of this is highlighted in the first chapter by assessing two studies from the recent literature using econometric techniques in climatology after a brief description of a basic model of the Earth's climate and changes therein. The chapter concludes with an overview of the thesis.

This chapter has been published as:

Pretis, F. and Hendry, D.F. 2013. "Some Hazards in Econometric Models of Climate Change". *Earth System Dynamics*. 4, 375-484.

and

Pretis, F. and Allen, M.R. 2013. "Climate Science: Breaks in Trends". *Nature Geoscience*. 6, 992-993.

1.1 INTRODUCTION

Human influences on climate are clear (IPCC, 2013), yet it has proved difficult to establish robust links between specific human actions and climate outcomes given natural variability and the highly varying observed climate time series. Understanding of climate change and responses thereto rely on accurate measurements as well as models of both socio-economic and physical processes. However, data to assess impacts and establish historical climate records are non-stationary: distributions shift over time due to shocks, measurement changes, and stochastic trends - all of which invalidate standard statistical inference and require careful data analyses to obtain viable inferences.

In this thesis I develop and apply econometric methods to augment climate research by disentangling complex links between human actions and natural variability and their climate responses masked by stochastic trends, omitted variables and breaks. To do so, I establish econometric methods to model the non-stationary nature of the data consistent with known physical laws and socio-economic processes, enabling joint estimation and testing. I propose procedures to detect structural change - both to generate robust models, as well as to use break detection as an aim in itself - by detecting previously unknown events and quantifying their magnitudes.

Econometric tools based on cointegration and model selection have the potential to contribute to existing climate research through rigorous observation-based analyses, however, these have to be conducted carefully to avoid statistical hazards. The importance of this is highlighted in the first chapter by assessing two studies from the recent literature using econometric techniques in climatology after a brief description of a basic model of the Earth's climate and changes therein.

1.2 ECONOMETRIC MODELS OF CLIMATE CHANGE

Earth's climate is driven by natural variability and anthropogenic (human) influences - see IPCC (2013) for a comprehensive overview. Natural variability broadly stems from solar radiation, Earth's orbit and tilt, the composition of its atmosphere and ocean circulation. Anthropogenic factors include emission of greenhouse gases (GHGs), aerosols, and land use changes. The concentration of GHGs has direct effects on Earth's climate – in highly simplified terms (i.e., using a model with an atmosphere made up of a single isothermal slab), heat enters the Earth's atmosphere as short-wave radiation from the Sun, and is radiated in the form of long-wave radiation from the warmed surface to the atmosphere, where greenhouse gases absorb some of that heat. This heat is re-radiated, so some radiation is directed back towards the Earth's surface. Thus, greater concentrations of greenhouse gases increase the amount of absorption and hence re-radiation. The most prominent greenhouse gases include water vapour, carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), ozone (O_3), and chlorofluorocarbons (CFCs). The physics of greenhouse gases are reasonably well understood, and date from insights in the late 19th century by Arrhenius (1896), who showed that atmospheric temperature change was proportional to the logarithmic change in CO_2. The longevity of effects of different gases varies drastically across the gases, for example N_2O has

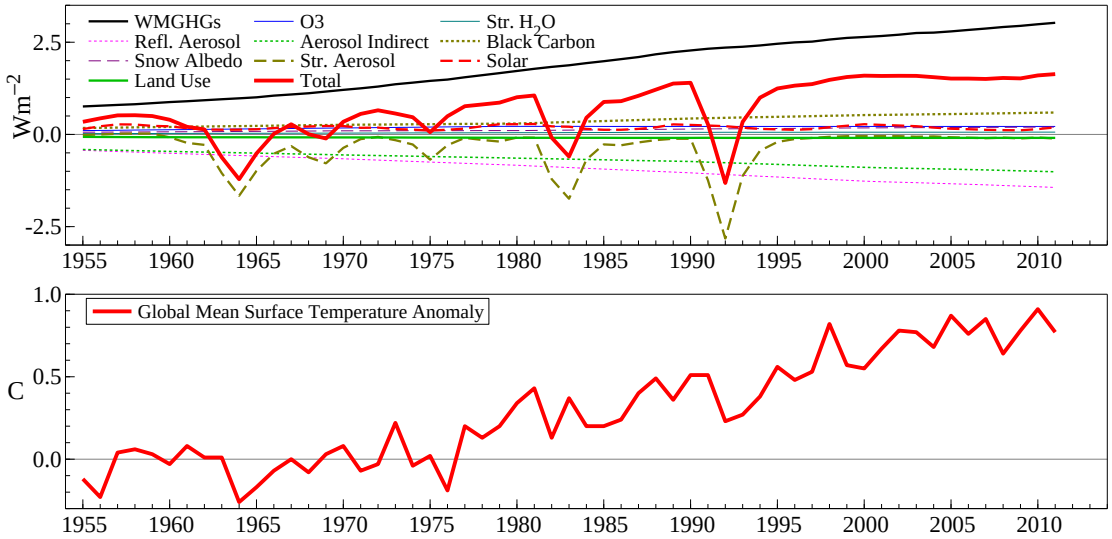


Figure 1: Time series of Radiative Forcing (Top) and Global Mean Surface Temperature Anomaly (bottom). Total forcing is shown in red as the sum of the individual factors including well-mixed greenhouse gases (WMGHGs, black), stratospheric aerosols (volcanic eruptions, dashed green) etc.

a lifetime[1] around 121 years, while CO_2 is extremely long-lived with no lifetime that can be calculated (Myhre et al. 2013). Thus, any shock in a source of emissions in CO_2 will remain present unless a sink actively removes it.

Together with many other factors (e.g. solar irradiance, reflective aerosols from volcanic eruptions, soot, etc.) the effect of GHGs is quantified through the concept of radiative forcing. In simple terms, this captures the change in energy in the atmosphere due to a change in concentration of a GHG or other factor. The radiative forcing effect, measured in watts per square meter (wm^{-2}), is calculated for the different forcing agents (e.g. GHGs, aerosols) and links the effect of the concentration of a gas on to temperatures. Myhre et al. (2001) provide an extensive overview of the different radiative forcing of various greenhouse gases: a well known equation describes the radiative forcing effect of atmospheric CO_2 as $F_{CO_2} = 5.35\ln(CO_2/CO_{2,1860})$ where $CO_{2,1860}$ denotes the pre-industrial concentration, generally assumed to be 280ppm. Figure 1 (top) shows the radiative forcing of different factors over the latter half of the 20th century (Meinshausen et al. 2011). The bottom panel shows the global mean surface temperature anomaly[2] over the same time period (Hansen et al. 2010). Greenhouse gases are grouped together[3] as well-mixed GHGs (WMGHGs) due to their long lifetime and property that they are "well-mixed" – concentrations are at approximately the same level across the globe (Myhre et al., 2013).

The effect of GHGs and other factors onto temperatures can be modelled using energy balance models (EBMs) equating incoming and outgoing energy. Equation (1) shows a simple

[1] Lifetime is defined as the time it takes to turn over the global atmospheric burden, the total burden or concentration divided by mean global sink (i.e. how much is removed per year) - see Myhre et al. (2013).

[2] Global mean surface temperatures are traditionally reported as anomalies relative to a reference period. The reference period here is 1951-1980.

[3] The major gases grouped together as WMGHGs include CO_2, CH_4, N_2O, and SF_6 (Sulfur hexafluoride).

zero-dimensional energy balance model (see e.g. Kaufmann et al. 2013) describing the change in atmospheric temperatures (T) over time t:

$$C\frac{dT}{dt} = F - \lambda T \tag{1}$$

where C is the heat capacity, F denotes the radiative forcing in wm^{-2}(e.g. from GHGs) and λT captures the feedback of increased outgoing long-wave radiation from higher temperatures. The forcing F consists of the aggregate of all forcing effects. An increase in the forcing F (e.g. due to increased emissions of CO_2) leads to an increase in temperatures $dT/dt > 0$ subject to a feedback of $-\lambda T$.

The simple EBM (1) provides the motivation and starting point for the use of econometric analyses of climate dynamics. In simple zero-dimensional energy balance models the time-series properties of radiative forcing should be transferred onto temperature, in other words if both GHGs and temperatures are driven by the same stochastic trends or shocks, they should cointegrate or co-break. This provides the basis for using statistical methods to link effects of greenhouse gases to temperature responses. The transfer of time series properties may not hold for a limited number of observations, sufficiently noisy time series, or if the model is inappropriate. However, Kaufmann et al. (2013) provide evidence that this result generalizes from simple to more complex climate models: the stochastic trend in model temperature is driven by the stochastic trends in the anthropogenic forcing series. Equally, if there are breaks in anthropogenic forcing, one could expect these breaks to be present in observed temperature series as well.

There are various reasons why the observed record of temperature and other climate variables may not exhibit the warming patterns suggested by theory or large-scale simulated coupled climate models.[4] Empirical modelling can play a role in investigating these underlying issues and test whether observations exhibit the relationships implied by theory. Econometric methods specialised in the analysis of non-stationary time-series, model-selection, and break detection can augment climate research by investigating these empirical links to test theories, estimate parameters, identify policy as well as other previously unknown effects, and assess model robustness and stability.

In this thesis, I first provide an overview of hazards that arise in time-series econometric models of climate (chapter 1), I then show how a physical system can be mapped to a cointegrated vector autoregression (CVAR), providing an econometric system approach to the estimation of relationships between temperatures, ocean heat, and radiative forcing (chapter 2). Chapter 3 then explores how one of the forcing series, atmospheric CO_2, can be modelled endogenously without prior restrictions on which contributing variables are relevant, using model selection methodology for more variables than observations. This approach introduces indicator saturation methods to detect structural breaks and outlying observations in the form of impulses to the modelling of climate data. I then generalize the principle of indicator saturation to designed break functions - treating the detection of breaks as an aim in itself (chapter 4). The resulting methodology is applied to detect volcanic eruptions in temperature time series. All climate models, ranging from small EBMs to large-scale simulated models rely on socio-economic projections to make statements about future climate change. Chapter

[4] Coupled climate models refer to large-scale computer simulations coupling ocean and atmosphere dynamics.

5 assesses the performance of the initial socio-economic scenarios used in IPCC reports to provide evidence on the course of global development.

1.3 HAZARDS IN ECONOMETRIC MODELS OF CLIMATE CHANGE

There are at least six key reasons why econometric models may lead to spurious conclusions: data measurement errors (e.g. inaccurate interpolation); unmodelled shifts (when a change in policy or measurement shifts a relationship); mistaken inference; incorrectly modelled relations (when the residuals from the estimated relationship do not satisfy the statistical properties of the assumed error processes, so claimed inferences are invalid); omitted variables' bias (omitting relevant explanatory variables); and aggregation bias (mixing data from very different populations). All six can powerfully distort any empirical statistical study, and need to be controlled for to avoid coming to fallacious conclusions.

In this chapter I highlight some of these hazards[5] based on two recent papers: Estrada et al. (2013), on detecting and attributing breaks in linear trends, and Beenstock et al. (2012) on cointegration of temperatures and greenhouse gases. In particular the focus here lies on unmodelled shifts (through policy intervention and measurement changes), incorrectly modelled relations (through stochastic trends and mistaken physical relationships), and omitted variables (primarily ocean effects). While the individual hazards are specific in their nature, they illustrate technical difficulties which may lead to spurious conclusions, and are representative of the broad range of hazards encountered in time series analysis of climate data. The individual hazards span three common themes that have to be tackled: structural breaks, stochastic trends, and consistency of statistical models with underlying physical laws.

Global temperature records and radiative forcing of greenhouse gases (GHGs) are non-stationary time series, the statistical properties of which invalidate standard inference procedures that seek to detect relationships between them. Non-stationarity comes in many forms, two of which are particularly prominent: stochastic trends from economic activity and sudden shifts through policy interventions, natural shocks, or changes in measurement.

Cointegration analysis can be used to overcome the inferential difficulties resulting from stochastic trends when that is the only source of non-stationarity, and is applied to test whether there exist combinations of non-stationary variables that are themselves stationary (see Hendry & Juselius 2001). Cointegration analysis crucially relies on the time-series properties of the available data, and while tests can be performed on an estimated equation's residual in bi-variate models (see Engle & Granger 1987), cases with more than two variables require testing in a system setting. The literature[6] on stochastic trends and cointegration differentiates between series being stationary, trend stationary, first difference stationary, denoted I(1), and stationary in second differences, I(2). Stationary variables are drawn from distributions that are invariant over time. Trend stationarity implies that a series is stationary once a linear trend component is removed. Integrability refers to one aspect of the stationarity properties of a time series. Series that are integrated of order one, I(1), and two, I(2), are stationary if differenced once or twice respectively.[7] An I(1) process contains a unit-root, that is, the characteristic polynomial of the process has a root at 1. The presence of more than one unit

[5] Pretis & Hendry (2013) provides a complete treatment of all six hazards relative to the work of Beenstock et al.

[6] The focus here lies on linear processes, cointegration in non-linear processes can also be considered – see Berenguer-Rico & Gonzalo (2014).

[7] In general a series that is I(k) needs to be differenced k times to be stationary.

root is indicative of higher order integration (e.g. I(2)). The level of integration is traditionally determined using unit-root tests.[8]

Unmodelled shifts are unaccounted changes in the distributions of the variables in the model over time. These unmodelled shifts (due to many potential causes, including technological innovations, changes in legislation, wars, major geophysical disturbances, changes in measurement etc.) can play havoc with statistical inference: they add an additional non-stationarity to that induced by integrating forces (such as unit roots). For example, standard inference made under the assumption of stationarity will be invalid if there is a shift in the mean—the underlying distribution of that variable is then not invariant over time. Further, unmodelled shifts distort the relationships between the variables that have been included; lead to residuals with properties that differ from the assumed error processes and thus invalidate inference; and can induce forecast failure out of sample. These shifts or structural breaks require explicit modelling, however, their timing and magnitudes are often unknown *a-priori*. In any data series, there might be an unknown number of location shifts of unknown durations and magnitudes, possibly from unknowingly omitted variables: data-based, sample-wide procedures are required to detect changes.

1.3.1 *Structural Breaks in Climate Data*

There are uncountable possible reasons for breaks in observed data. Common candidates include policy interventions and changes in measurement. These shifts or breaks have to be modelled for valid inference and identified using sample wide searches if they are previously unknown. Breaks themselves can be informative - they can provide insight into whether policy is effective or identify previously unknown events such as volcanic eruptions (see Chapter 4), and be used to evaluate models (see Pretis et al. 2015).

Policy Interventions

Estrada and colleagues (2013) investigate the relationships between the time series of global mean temperatures and atmospheric greenhouse gas concentrations using tests for breaks in trends, arguing that both time series of temperature and atmospheric greenhouse gas concentrations can be approximated by deterministic, linear trends with breaks. Using tests to identify these break points in each time series, they find that the trends and breaks in the evolution of temperature and of atmospheric greenhouse gas concentrations match up. Based on these matches, Estrada et al. conclude that two changes in human-made emissions may have contributed to the relatively slow warming over the past fifteen years or so: reductions in the emissions of chlorofluorocarbon (CFC) greenhouse gases that were regulated under the Montreal Protocol from 1989 to protect the Earth's stratospheric ozone layer; and a cut-back in methane emissions as a result of changes in farming practices since the 1990s. They also identify the impacts of the two World Wars and the Great Depression on carbon dioxide

[8] System-based cointegration analysis (Johansen, 1988), as used in Chapter 2, does not require pre-testing for unit-roots. The term unit root stems from the case when unity is a solution to the characteristic polynomial of a process. As an example, let y_t be the first-order autoregressive process: $y_t = \alpha y_{t-1} + \epsilon_t$ where ϵ_t is a white-noise process. Note that y_t is a random walk, which is an I(1) process, and has a unit root if $\alpha = 1$. To see this, using the lag-operator L (see Hendry 1995, Ch4.), the process can be expressed as $y_t = \alpha L y_t + \epsilon_t$, so re-arranged to $(1 - \alpha L)y_t = \epsilon_t$ where $(1 - \alpha L)$ is a first order polynomial in L. This polynomial has a root of $1/\alpha$, and thus a unit-root if $\alpha = 1$.

emissions, on the rate of warming during the 20th century. Estrada et al. identify the breaks using the sample-wide Perron and Yabu (2009) procedure which tests for a single break in a linear trend.

There are concerns with both the break detection approach as well as with the underlying models. The Perron-Yabu (2009) test allows for at most one break at a time, thus, requiring a sequential testing procedure to detect multiple breaks. The sequential approach implies that the detection of possible first breaks will be affected by the presence of further breaks and may therefore lead to spurious results. A general to specific approach starting with a general model of a break at every point without an upper limit on the number of breaks can offer improvements – chapters three and four of this thesis propose and apply methods for multiple breaks. A further concern is Estrada et al.'s search for breaks in linear trends - if the linear trend model is incorrect (as section 1.3.2 suggests), it is likely that stochastic movement could be attributed to changes in deterministic trend functions resulting in spurious detection.

Changes in Measurement

Changes in measurement equally can manifest themselves as breaks in observations and wreak havoc with conventional inference. In their empirical statistical study of temperature and radiative forcing of greenhouse gases, Beenstock et al. (2012) present statistical tests that purport to show that these variables have different integrability properties, and hence cannot be related unless they polynomially cointegrate. Beenstock et al. then show that their constructed measure of anthropogenic forcing (a component of F in equation (1)), an "anthropogenic anomaly", does not cointegrate with observed temperature, presenting this as evidence against anthropogenic global warming. Beenstock *et al.* address an interesting question, to do so they rely heavily on the time-series properties, in particular the order of integration of the data to reach their conclusions. They argue that temperatures are I(1) while individual anthropogenic GHGs are I(2). In their analysis, the GHGs cointegrate into an I(1) process, an "anthropogenic anomaly", which, however, they find is not cointegrated with global mean temperatures. Cointegration methods as applied in Beenstock *et al.* require that series first exhibit the same degree of integration before a meaningful relationship between them can be established. Therefore, unit-root tests to determine the level of integration often play a major part in single-equation cointegration analyses.

We obtained the data on concentration of greenhouse gases used by Beenstock *et al.* (2012, see their section 3.1) using the values provided by Myhre et al. (1998) to convert the series into their radiative forcing equivalents. The fact that the measured series of GHGs come from a variety of different sources is omitted from Beenstock *et al*. If the measurements were identical in all sources, this would not be an issue: however, our graphs reveal sharp differences in the data properties. Consider the CO_2 and N_2O series. Both are initially based on ice core data (up to the dates indicated by the vertical lines in Figure 2: 1850 until 1958 for CO_2 and 1850 until 1978 for N_2O) followed by flask and other measurements thereafter. Figure 2b shows that up until approximately the point when the switch from ice-core to non-ice core data was made, many of the changes have precisely the same magnitude, revealing an artificial pattern different to the latter half of the sample. Moreover, there are large changes in the variances of the second differences of both series at the measurement system switch. Despite the well-established problems for unit-root tests, the data are analyzed in Beenstock *et al.* as if they come from the same populations. Interacting with unmodelled shifts, measurement errors can

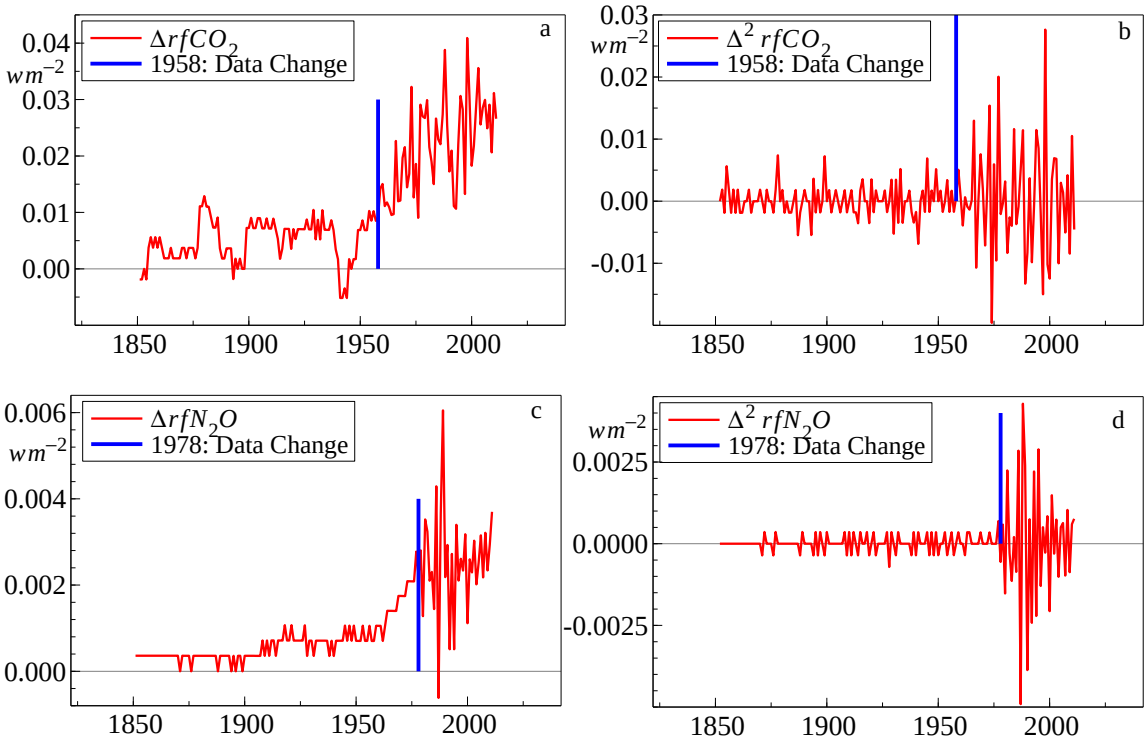


Figure 2: Time series of the first and second differences of $rfCO_2$ and rfN_2O

can lead to false interpretations of the stationarity properties of data. In the presence of these different measurements exhibiting structural changes, a unit-root test[9] on the entire sample could easily not reject the null hypothesis of I(2) even when the data are in fact I(1). Indeed, once we control for these changes, our results contradict the findings in Beenstock *et al.* (see their section 3.1 and Table 1). Our results are presented here in Table 1 and Table 2 below.

As is to be expected from the data in Figure 2, but based on augmented Dickey–Fuller (ADF) tests (see Dickey & Fuller 1981), the first difference of annual radiative forcing of CO_2 is stationary initially around a constant (over 1850–1957), then unit-root non-stationary (over 1958–2011). Although these tests are based on sub-samples corresponding to the shift in the measurement system, there is sufficient power to reject the null hypothesis of a unit root. In a similar manner, unit-root tests reject non-stationarity of the first difference of N_2O for the second set of observations (1978–2011). Unit-root non-stationarity cannot be rejected for 1850–1978 for N_2O: however, given the manifestly artificial appearance of the data such a result should be interpreted with extreme caution. Particularly, the ΔN_2O series appears to exhibit a step shift in the early 1900s, which also leads to spurious results in unit-root tests. The split points for our sub-sample unit-root tests are given by the extraneous dates of change in the measurement system, so the tests do not need to allow for that choice.

Given these time-series properties including apparent shifts, interpolation, and changes in variance, assuming that all annual anthropogenic GHGs are I(2) over the entire sample is an incorrect starting point, and the findings in Beenstock et al. appear to be an artefact of

[9] Unit-root tests are used to determine the level of integration of time series. For the ADF (see Dickey & Fuller 1981) test used here, rejection of the null hypothesis provides evidence against the presence of a unit-root and suggests that the series is I(0) (stationary) rather than I(1) (integrated).

Table 1: ADF Unit-root Tests on $\Delta rfCO_2$

1850–1957 constant			1958–2011 constant		
D-lag (AIC)	**t-adf**	**Reject H_0**	**D-lag (AIC)**	**t-adf**	**Reject H_0**
5 (-12.11)	-3.737	**	5 (-9.68)	-2.055	
4 (-12.05)	-2.910	*	4 (-9.72)	-2.132	
3 (-12.06)	-2.948	*	3 (-9.76)	-2.124	
2 (-12.08)	-3.146	*	2 (-9.74)	-2.645	
1 (-12.07)	-2.706		1 (-9.72)	-3.529	*
0 (-12.05)	-3.544	**	0 (-9.73)	-4.815	**

ADF unit-root tests: the null hypothesis H_0 is that the series has a unit root so is non-stationary. Rejecting the null hypothesis suggests no unit-root non-stationarity. D-lag specifies the number of lags included in the ADF unit root test, where the AIC is given in parentheses. Unit root test outcome: ** indicates rejection of the null hypothesis at 1% and * at 5%.

Table 2: ADF Unit-root Tests on ΔrfN_2O

1850–1978 constant			1979–2011 constant		
D-lag (AIC)	**t-adf**	**Reject H_0**	**D-lag (AIC)**	**t-adf**	**Reject H_0**
5 (-17.20)	3.621		5 (-13.35)	-3.856	**
4 (-17.20)	3.461		4 (-13.29)	-3.318	*
3 (-17.20)	3.135		3 (-13.33)	-3.594	*
2 (-17.18)	2.477		2 (-13.37)	-3.993	**
1 (-17.02)	0.861		1 (-13.41)	-4.721	**
0 (-16.64)	-1.900		0 (-13.47)	-7.711	**

ADF unit-root tests: the null hypothesis H_0 is that the series has a unit root and is non-stationary. Rejecting the null hypothesis suggests no unit-root non-stationarity. D-lag specifies the number of lags included in the ADF unit root test, where the AIC is given in parentheses. Unit root test outcome: ** indicates rejection of the null hypothesis at 1% and * at 5%.

pooling data with very different measurement systems and behaviour in the two sub-samples. Measurement problems also extend to other radiative forcing series such as black carbon or reflective aerosols not plotted here.

In line with our analysis suggesting that the presence of measurement changes affects the unit root test outcomes, Stern & Kaufmann (2000) show that when using univariate unit root tests, the results can vary considerably depending on which type of test is used and also vary across different anthropogenic gases. For example, CO_2 appears to be I(1) in two out of their four tests, and I(2) in the others (Table 1 in their paper). Similarly, N_2O appears I(1) in three out of the four test types. Second, Kaufmann & Stern (2002) test the time-series properties of the aggregate of the radiative forcing of all the major greenhouse gases (CO_2, CH_4, N_2O, CFC-11, CFC-12) and find them to be I(1). This result on the aggregate is important - as the following section (and Chapter 2) argue - it is the time series properties of the aggregate rather than those of the individual series that determine the corresponding properties of temperatures.

Crucially, the level of integration, and thus stationarity, of data is not intrinsic to its process and can change over time. There is nothing inherent in the physical data generating process that makes anthropogenic forcings, or other variables, I(1) or I(2). The observational data may, over some period, be consistent with a process that is I(1) or I(2), but this is not an intrinsic property that cannot change. There are many examples of changes, two being that the level of CO_2 emissions is related to economic activity, which has varied over time, and emissions of CFCs which only arose in the latter part of the 20th century. Both of these may be stationary in second differences from the 1950s onwards, but because of the underlying processes, they may well have been stationary in first differences or levels before then, or in the case of CFCs non-existent before their discovery and declining after the Montreal Protocol of 1989 as the example of Estrada et al. (2013) shows. The claim that all greenhouse gases are always I(2) is incorrect. Such a result is also inconsistent with the tests conducted in the previous literature and with our analysis.

As shown, data measurement errors induced by changes in measurement, can mislead any form of inference. Empirical relationships then represent correlations between what was incorrectly measured, not what actually happened. Interpolation, unless perfect, creates measurement errors and measurement errors in explanatory variables which induce downwards biases in parameter estimation. Further, interpolation leads to negative error autocorrelation in dynamic relations which invalidates standard statistical inference unless this auto-correlation is handled appropriately. To see this, let $\{Y_t\}$ be a stationary autoregressive process:

$$Y_t = \gamma Y_{t-1} + e_t$$

where $e_t \sim \mathsf{ID}[0, \sigma_e^2]$ (denoting independent sampling from a constant distribution with mean 0 and variance σ_e^2). Suppose $\{Y_t\}$ is only observed with error as $\widetilde{Y}_t = Y_t + v_t$ where $v_t \sim \mathsf{ID}[0, \sigma_v^2]$ is a random error of measurement or interpolation, then the observed variable is:

$$\widetilde{Y}_t = \gamma \widetilde{Y}_{t-1} + e_t + v_t - \gamma v_{t-1} = \gamma \widetilde{Y}_{t-1} + u_t - \rho u_{t-1}$$

where ρ depends on σ_e^2 and σ_v^2. The presence of u_t and u_{t-1} in the above equation implies that the model will automatically exhibit negative error autocorrelation. In turn, this negative error autocorrelation strongly affects the outcomes of integration and cointegration tests (see

Schwert 1987, and Hendry 1995, Ch. 12). An additional major concern is that periods of interpolation and regular measurement are combined and treated as if they stemmed from the same measurement process despite changes in error autocorrelation and variance.

1.3.2 *Stochastic Trends and (In)Correctly Modelled Relations*

There is an ongoing debate in the literature on how to model non-stationarity in climate data. Some studies (Kaufmann et al. 2013, Kaufmann & Stern 2002) argue that temperature and GHGs are best approximated by stochastic trends that are in their nature random and driven by underlying processes (e.g. economic activity which are stochastic trends themselves). On the other hand, Estrada et al. argue that temperature and also GHGs can be reasonably approximated by linear, deterministic trends with breaks in the trend function. There are some concerns over this, in particular that GHGs are closely linked to economic activity and the long lived nature of some GHGs means that shocks (e.g. increases in CO_2 emissions) will affect atmospheric concentrations for a long time - the concentration of that gas will not quickly revert to an underlying trend. This is more consistent with a stochastic trend model, rather than a linear trend model. Such model mis-specification can lead to spurious detection of breaks.

Beenstock et al., in turn, rely on a stochastic trend model and use cointegration techniques to tame the apparent non-stationarity. However, hazards equally arise in this world. The main message of Hendry (1995) is that before any statistical inferences can be conducted, a model must be congruent, or well-specified in that it satisfies the assumptions on which the statistical analysis relies. The unit root tests in Table 1 in Beenstock *et al.* make many untested assumptions, including accurate data, that there is a single measurement regime, and that no location shifts occurred.

Beenstock *et al.*'s conclusion that anthropogenic forcings do not cointegrate with the observed temperature anomaly relies on their constructed measures of anthropogenic forcing which approximates parts of the F term in (1), the "anthropogenic anomaly" (see their part 3.2). In light of model specification, we assess this method of constructing the measure of anthropogenic forcing. The measures of anthropogenic forcing, the "anthropogenic anomalies" in in Beenstock *et al.* (given by equations 9 and 10 in Beenstock *et al.*, reproduced here as equations (2) and (3)) are the residuals g_1 and g_2 of a single regression of radiative forcing of CO_2 (denoted here as "rf") on the forcing of other greenhouse gases.

$$\mathrm{rfCO_2} = 10.972 + 0.046\mathrm{rfCH_4} + 10.134\mathrm{rfN_2O} + g_1 \tag{2}$$

$$\mathrm{rfCO_2} = 12.554 + 0.345\mathrm{rfCH_4} + 9.137\mathrm{rfN_2O} + 1.029\mathrm{BC} + 0.441\mathrm{ReflAer} + g_2 \tag{3}$$

where BC is their radiative forcing of black-carbon concentration, and ReflAer is their radiative forcing of all reflective aerosols. Such regressions are a variant for possibly I(2) variables of the approach in Engle & Granger (1987). Banerjee et al. (1986) demonstrated that this type of test imposed 'common-factor' restrictions of the form criticized by Hendry & Mizon (1978) and Mizon (1995), as a consequence of which the test often lacks power and is substantively inferior to the systems approach in Johansen (1988). We now consider their 'anthropogenic anomaly'.

Despite the above ADF unit root test outcomes, suppose one accepted the starting point of Beenstock *et al.* that all anthropogenic variables are I(2). They state that equations (2) and (3) are to test for cointegration between the anthropogenic series. However, cointegration is a system property, so the variables need to be treated as such. To establish cointegration between the variables in Equation (2) ($rfCO_2$ regressed on $rfCH_4$ and rfN_2O), the full system of three variables needs to be considered (see Hendry & Juselius 2001). This is further complicated here as the variables are assumed to be I(2), so an I(2) cointegration analysis is required (see Juselius 2006): the system has at most full rank (=3), or if there is (polynomial) cointegration the system may exhibit reduced rank of one or two, and if no (polynomial) cointegration, rank zero. If there is reduced rank (which has to be tested in an I(2) procedure), the system needs to be decomposed into the cointegrating relations (which are I(0) and therefore stationary) and the common underlying stochastic trends (of which some may be I(1) and some I(2) trends). Given the assumed I(2) property, if the three variables cointegrate, there may be up to two I(1) cointegrating relations between the three series, and thus two potential anthropogenic anomaly measures. The single anthropogenic anomaly given in their equation (2) is then a linear combination of these measures of anthropogenic forcing. The same problem generalizes to their equation (3), with there being five variables in the system and a much larger set of potential cointegrating relations. The system of five variables may have full rank (=5), rank zero, or reduced rank between one and four if there is cointegration, implying up to four I(1) cointegrating relations and up to four measures of the anthropogenic anomaly. Thus, even if their starting point that all anthropogenic are I(2) is accepted, then their measure of the anthropogenic anomaly is likely only one of many, given the large number of potential cointegrating relations. There could well be a residual (anthropogenic anomaly) that does cointegrate with temperature and solar irradiance. A further problem arises in the physical implications of the constructed "anthropogenic anomaly" as the next section shows.

1.3.3 *Consistency to Physical Laws and Omitted Variables*

It is important to recognize that any attribution analysis requires some form of model to establish cause and effect. Every estimated statistical model implies a physical model, even if not explicitly stated or assumed. Cointegration and break detection can be useful tools, but, econometric models in themselves are often (more often than not validly) criticised for lacking a physical basis. To draw conclusions and quantify the magnitude of effects it is important to have a strong physical basis for the use of statistical models.

The challenges of underlying physical models show themselves in Beenstock et al. in two ways. First, how radiative forcing of anthropogenic greenhouse gases is used and modelled. Given that the starting point of assuming all anthropogenic variables are I(2) is incorrect, the measures of the anthropogenic anomaly (g_1 and g_2) are inappropriate. In Beenstock et al. the measures of anthropogenic forcing are the residuals of regressions of $rfCO_2$ on radiative forcing of the other greenhouse gases. This means that the measure of anthropogenic forcing used in Beenstock *et al.* is really the variation in radiative forcing of CO_2 that is unexplained by the variation in other greenhouse gases. In a basic energy balance model, radiative forcings are measured in a single unit (wm^{-2}) and summable (see equation (1) and Figure 1). The total effect of all forcings together (while taking feedbacks into account) is what is important. Taking the unexplained variation in radiative forcing of CO_2 as a measure of anthropogenic

forcing (i.e. the "anomaly") is then incorrect and does not measure what Beenstock *et al.* state it does. Their main test of anthropogenic global warming (the regression of temperature on solar irradiance and the anthropogenic anomaly in their Table 3) is then a regression of temperature on solar irradiance and a residual, which need not capture any anthropogenic component at all, and does not capture the main anthropogenic forcing component. This leads to the main point (alluded at in the section 1.3.1): the time series properties of individual GHGs are not what is directly relevant, total forcing F is what matters. When aggregating the effect across GHGs and other factors (e.g. solar irradiance, volcanic forcing etc.) the weights in the sum are given by the radiative forcing equivalent, not by a single equation cointegrating relationship as used in Beenstock et al. Thus, even if some of the series are I(2), the physical cointegrating relationship is given by the sum of the forcing. This property is also highlighted in Chapter 2, and resolves much of the debate around the different orders of integration of gases and temperatures.

The second concern is the tendency to "leave the ocean out", prevalent in many econometric studies of climate change. The simple EBM in equation (1) equally does not model any ocean effect. Ocean-atmosphere dynamics should not be ignored in empirical models of climate, whether it is heat exchange (see Chapter 2), or cyclical fluctuations such as the El Niño Southern Oscillation (see Chapter 3). In particular the ocean's role as a large heat sink is crucial. Even though ocean heat content has steadily been increasing (see Chapter 2), Beenstock et al. omit the ocean entirely from their analysis based on arguments of different order of integration, stating that ocean heat is not I(2) and should therefore be omitted.

Omitted variables induce biases in general (unless orthogonal to all included variables), and in Beenstock *et al.*'s analysis based on cointegration, the negative effects of having omitted important factors may be even more pronounced. Suppose one accepts their analysis of integration properties (see above in section 1.3.1), and the anthropogenic anomaly, (see above in section 1.3.2), the main equation to test for cointegration between the anthropogenic component and temperature in Beenstock *et al.* is a regression of temperature on the anthropogenic anomaly and solar irradiance (reproduced here as equation (4) to approximate equation (1), see section 3.3 and Table 3 of cointegration tests in Beenstock *et al.*):

$$\text{Temperature}_t = \beta_0 + \beta_1 \text{Solar Irradiance}_t + \beta_2 \text{Anthropogenic Anomaly}_t + \epsilon_t \qquad (4)$$

Their main conclusion stems from the fact that, using unit-root tests, the error term ϵ_t of this regression is non-stationary, I(1), suggesting that there is no cointegration between the three series. Any omitted I(1) variable in this equation will induce the error term to appear I(1), and lead to spurious rejection of cointegration. Ocean heat uptake is one of many factors missing in this equation, and, according to Beenstock *et al.*'s analysis is I(1). Beenstock *et al.* state that the omission of ocean heat is not a concern as ocean heat content and temperature do not cointegrate (Table 4 in Beenstock *et al.*), however, this cointegration test is undertaken using temperature, ocean heat and water vapour alone, rather than considering the full system where the anthropogenic measure and solar radiation are also included. These individual regressions do not capture the important system property of cointegration (emphasized in section 1.3.2). Further, any other omitted forcing series (CFCs, ozone) may also be I(1). Any omitted I(1) variable in their main test of cointegration between temperature, the anthropogenic anomaly and solar radiation will induce an I(1) stochastic trend in the residual. Given that the regression only consists of a constructed measure of anthropogenic

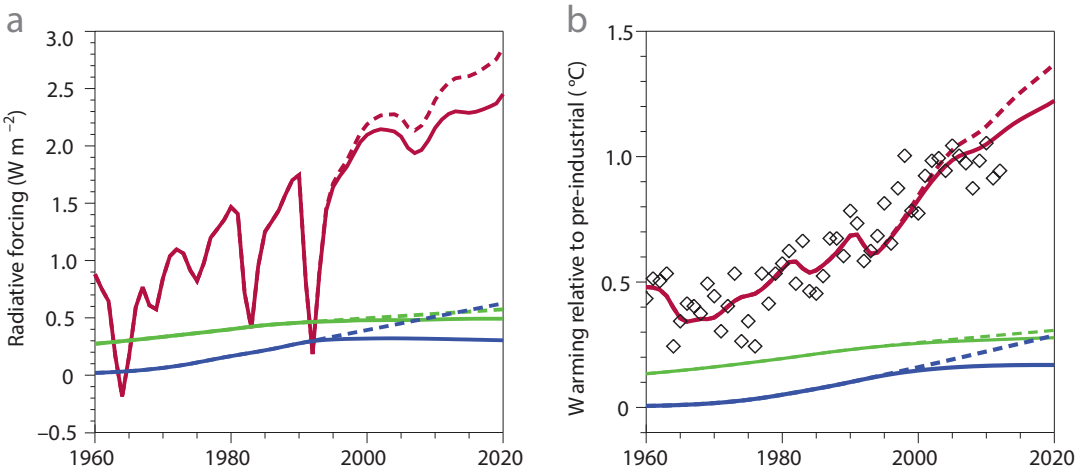


Figure 3: Radiative forcing and global mean temperature since 1960. (**a**), Between 1960 and 2020, various factors are known to have affected the balance between incoming and outgoing energy on Earth. These factors include atmospheric methane concentrations (green) and levels of CFCs and other gases whose emissions are regulated under the Montreal Protocol (blue). Total radiative forcing (from Meinshausen et al. (2011)) is depicted in red. Without implementation of the Montreal Protocol, it is assumed that CFC levels would have risen beyond 1990 in a linear extrapolation (blue dashed line). (**b**), An energy balance model allowing for faster shallow ocean and slow deep ocean warming (Held et al. 2010) highlights the temperature response from each factor (same colours as in a), and suggests that as a result of the implementation of the Montreal Protocol, temperatures today are almost 0.1 C cooler than they would have been without it (dashed versus solid red lines). Observed annual mean temperatures, offset to have the same mean as model temperatures over the period 1900 to 2012, are included as diamonds.

forcing and solar irradiance, there are many factors that will lead to the error term appearing I(1) and thus, to spurious rejection of cointegration.

Statistical models of climate require a physical foundation. Estrada et al. detect a break around the time of the Montreal Protocol and thus attribute (parts of) the recent temperature slowdown (the warming "hiatus") to this policy intervention as well as hypothesized changes in agricultural practices affecting methane emissions around the same time. However, while suggesting this link between policy and the observed slowdown in warming, they do not directly evaluate the magnitude of their break effect and whether it is sufficient in explaining the entirety of the observed phenomenon.

Estrada et al.'s conclusion regarding the contribution of the Montreal Protocol and reductions in methane emissions from agriculture can be tested with a more conventional approach (Boucher & Reddy 2008) that uses an analytically tractable calibrated model which accounts for faster warming in the shallow ocean and the slow component of deep ocean warming (Held et al. 2010). Using such a model to construct a simple null-hypothesis (Fig. 3), and linearly extrapolating the trend in radiative forcing from the late 1980s, there is indeed a break around 1990 in response to the Montreal Protocol. The impact of this change is small but non-negligible: temperatures could be almost 0.1C warmer than they are today without the reductions in CFC emissions. A trend break in methane forcing can also be constructed through extrapolation in this way, but its effect is much smaller.

This simple analysis supports the idea that the CFC emissions reductions, that were enforced by the Montreal Protocol, contributed to the warming hiatus of the past 15 years.

However, it also suggests that this is unlikely to be the whole story. To fully compensate for the warming that would have been expected between 1998 and 2012, the reduction in radiative forcing from CFCs would have had to be much stronger. The untidy conclusion is that the explanation of the much-discussed warming hiatus over the past decade and a half is unlikely to be a simple single issue: various neglected effects that are difficult to capture adequately through deterministic linear trend models, such as stratospheric volcanic aerosols, deep ocean heat uptake, the unusually strong solar minimum and internal climate variability (Kosaka & Xie 2013) are all likely to have played a role.

1.4 STRUCTURE OF THE THESIS

Econometric time series models of climate data have to account for the underlying non-stationarity of the data and be consistent with physical laws. Stochastic trends and structural breaks have to be modelled for inference to be valid. Equally, breaks in themselves can provide insight into previously unknown events, and assess models and policy alike. Through the work of Beenstock et al. and Estrada et al. I highlighted some of the challenges encountered in the use of time-series econometric methods in climate research. The following chapters are not a response to the work of Beenstock et al. and Estrada et al., but, in encountering some of these challenges, show directions through which they can be addressed. Without making the claim of providing a complete solution approach, this opens up a new line of research and provides a strong case of support for the use of econometric methods using energy balance models based on cointegration, and model selection based break detection. The remainder of the thesis is composed of chapters two to six, here I provide a short overview of the topics covered and contributions to econometric methodology and the wider literature.

Chapter 2: Linking Cointegration and Energy Balance Models

While econometric cointegration methods are powerful tools to handle stochastic trends, they are often criticised for lacking a physical justification for their use. In turn, physical energy balance models such as the one used to evaluate the claims made by Estrada et al. above, are pre-dominantly calibrated rather than estimated, and not tested as a system. The second chapter of this thesis shows that econometric models - restricted cointegrated vector auto-regressions (CVARs) - are equivalent to two-component energy balance models taking ocean interactions into account.

This chapter contributes to econometric methodology by providing a physical basis for the use of CVARs, and allows energy balance models to be estimated jointly as a system and tested as such. The application of the modelling approach in this chapter provides empirical links between radiative forcing, observed global mean temperatures and ocean heat content. Using a restricted CVAR I establish a complete econometric system model of global mean climate consistent with physical processes modelling the interaction between atmosphere, ocean, and radiative forcing.[10]

[10]This indirectly addresses many of the issues encountered in the work by Beenstock et al.

Chapter 3: Automatic Model Selection and Break Detection

The CVAR estimated energy balance model in Chapter 2 takes radiative forcing as exogenous and does not account for potential breaks in the observed time series. The third chapters explores how the most prominent of the greenhouse gas forcing series, atmospheric CO_2, can be modelled without imposing *a-priori* importance of variables while being robust to structural breaks and outliers. Using automatic model selection starting with a general unrestricted model incorporating impulse indicator saturation (IIS, Hendry et al. 2008), individual contributions to atmospheric CO_2 are determined.

On the methodology side, the chapter introduces econometric automatic model selection with more variables than observations to climate research. Impulse indicator saturation is applied to allow for outlying observations and structural breaks at any point in the sample, as well as a tool to indicate model misspecification. The application establishes the strong anthropogenic contribution to atmospheric CO_2, natural factors are necessary but not sufficient to account for the observed trend.

Chapter 4: Detection of Breaks using Designed Functions

Model selection-based break detection is not limited to impulses (IIS) or step shifts (SIS, Castle et al. 2015, Doornik et al. 2013) but rather can be generalized to any break form. The fourth chapter expands on the indicator saturation based break detection methodology and generalizes it to any form of designed break function. This model selection approach takes break detection as the aim in itself and demonstrates its use to improve historical measurements: to detect volcanic eruptions in the historical record without prior knowledge of their occurrence. This is done while remaining consistent with the underlying physical processes - the break form is determined by a simple energy balance model.

The work contributes to econometric methodology by generalising results from the indicator saturation literature to break detection using designed break functions, demonstrating the feasibility by providing analytical derivations for the split-sample approach under the null of no breaks and the alternative of one or more breaks. Monte Carlo simulations confirm the approximate size and power. The application demonstrates that break functions designed on theory models can be used to detect previously unknown breaks – in particular this holds promise for future volcanic detection efforts using real-world reconstructions of temperature variability over the last several millennia, and forecasting through breaks ranging from eruptions to recessions.

Chapter 5: Long-Run Socio-Economic Scenarios

Any prediction (using a simulated or empirical model) of climate change requires projections of how future emissions of greenhouse gases will evolve. The wide range of different global socio-economic scenarios in the IPCC reports results in high uncertainty about future climate change. It is therefore crucial to understand on which trajectory the world is developing. For the first time, observations over two decades are available against which scenarios can be compared. The fifth chapter evaluates the socio-economic scenarios on which IPCC long-run climate predictions are based, and shows that the growth rate in fossil fuel CO_2 emission intensity - fossil fuel CO2 emissions per GDP - over the 2000s was rising while all main emission scenarios envisaged a decline. The underprediction is explored by decomposing

global model errors of growth rates into individual country contributions using down-scaling and country-level decompositions. Emphasizing the importance of observation-based models and consistency of initial values across different scenarios, the chapter provides an outlook into expected future climate developments.

Chapter 6: Epilogue

The epilogue in chapter six summarizes the findings and provides an outlook into future research building on the work of this thesis.

Summary

On a purely methodological side, the thesis establishes a physical-science basis for the use of econometric cointegration methods in climate research. It introduces automatic model selection with more variables than observations together with the detection of breaks using indicator saturation to the climate literature. Break detection is formulated as a model selection problem and addressed using designed functions. Global model errors are decomposed to individual contributions using a proposed growth rate decomposition.

The applications in the thesis first establish a strong empirical link between observed global mean temperature, radiative forcing, and ocean heat content, by estimating a complete model of global mean climate. Results suggest a possible explanation for the low estimates of the temperature response to the doubling of CO_2 in other observation-based models. Anthropogenic contributions to atmospheric CO_2 are established, showing that natural factors are necessary but not sufficient to explain the observed trend in atmospheric CO_2. Using econometrics, volcanic eruptions can be detected without prior knowledge of their occurrence, holding promise to augment historical records and the use of designed functions in robust forecasting. Turning from historical data to future projections, growth in CO_2 emission intensity has been underpredicted in all main initial socio-economic scenarios used for climate predictions, this underestimation of emission intensity raises some concerns about achieving a world of economic prosperity in an environmentally sustainable fashion.

2

ESTIMATION OF NON-STATIONARY SYSTEMS IN CLIMATOLOGY: THE EQUIVALENCE OF TWO-COMPONENT ENERGY BALANCE MODELS AND COINTEGRATED VARS

Abstract

Emissions of greenhouse gases, global mean temperatures, and other climate observations are non-stationary time series invalidating the use of standard statistical inference procedures. Econometric cointegration analysis can be used to overcome some of these inferential difficulties, however, cointegration has been criticised in climate research for lacking a physical justification for its use. Here I show that a physical two-component energy balance model of global mean climate is equivalent to a cointegrated system that can be mapped to a cointegrated vector autoregression, making it directly testable, and providing a physical justification for econometric methods in climate research. Doing so opens the door to investigating the empirical impacts of shifts from both natural and human sources. My approach finds statistical support of the model using global mean surface temperatures, $0-700$m ocean heat content and radiative forcing. The model results show that previous empirical estimates of the temperature response to the doubling of CO_2 may be misleadingly low due to model mis-specification.

2.1 INTRODUCTION

Emissions of greenhouse gases, global mean temperatures, and other climate observations are non-stationary time series invalidating the use of standard statistical inference procedures (Pretis & Hendry 2013, Kaufmann et al. 2013, Kaufmann & Stern 2002, Stern & Kaufmann 2000). Econometric cointegration analysis can be used to overcome the inferential difficulties resulting from stochastic trends when that is the only source of non-stationarity, and is applied to test whether there exist combinations of non-stationary variables that are themselves stationary (see Hendry and Juselius 2001 for an overview, Juselius 2011 for an application to climate data, and Kaufmann & Juselius 2013 for a paleo-climate analysis). Cointegration, however, has been criticised in climate research for lacking a physical justification. In turn, multiple-component physical energy-balance models (see Gregory et al. 2002, Held et al. 2010, Schwartz 2012, Pretis & Allen 2013) that model the ocean and atmosphere interactions have tracked global mean climate well while remaining analytically tractable, but are rarely formally tested as estimation is difficult due to the non-stationarity of observations. Here I show that two-component energy-balance models of global mean climate can be mapped to a cointegrated system, both making them directly testable, and providing a physical justification for econometric methods that account for data non-stationarity. This makes it possible to estimate a complete system model of global mean-climate model using econometric methods. Doing so opens the door to investigating the empirical impacts of shifts from both natural and human sources, and enables a close linking of data-based macroeconomic models with climate systems.

Here I first show the mathematical equivalence of a two-component energy balance model to a restricted cointegrated system that can be mapped to a cointegrated vector autoregression (CVAR). This provides a system-based statistical framework in which an EBM can be estimated jointly (rather than one equation at a time) and therefore can be tested against observed data. Equally, this mathematical equivalence provides a strong physical basis for the application of CVARs in climatology, and places the entire CVAR toolkit at the disposal of energy balance models.

Climate observations are predominantly non-stationary time series invalidating the use of standard statistical inference procedures such as ordinary least squares regression and simple correlations. This makes the estimation of parameters in EBMs from observations a challenge. Additionally, most statistical approaches which estimate model parameters from observations do not formally test whether the EBM is an appropriate model for the system but rather estimate one equation at a time. It is, however, exactly the system nature of two-component EBMs and the non-stationarity of the data that provides a useful property that allows EBMs to be tested against observations. As I will show, a two-component EBM is equivalent to a restricted CVAR and this equivalence implies directly testable assumptions: if the model holds then the modelled series together form stationary relations (cointegrate), restrictions across the estimated equations are not rejected, and individual parameters are statistically significant.

As an application, I estimate a two component EBM through the use of a CVAR and test it against the observational record. The approach here is highly simplified climatologically and takes the fast, mixed component as the atmosphere at the surface, while the slower

component is proxied by the ocean heat content[1] from 0-700m. I find that tests against the statistical properties implied by the EBM do not reject the model. The time series form stationary relations (cointegrate) consistent with theory. Previous empirical estimates of climate sensitivity (ECS) may be misleadingly low due to model mis-specification (e.g apparent through residual autocorrelation), partly induced by volcanic aerosol forcing. Recursive estimates suggest that the recent slowdown in temperature increase, the "hiatus" (Kaufmann et al. 2011, Meehl et al. 2011, Watanabe et al. 2013), does not strongly affect the estimate of ECS. A simple forecasting exercise shows that the hiatus can be approximated through changes in radiative forcing and ocean heat uptake.

2.2 TWO-COMPONENT ENERGY BALANCE MODELS

Two component energy balance models are characterised by a mixed upper layer (the shallow ocean/atmosphere) with low heat capacity and a deeper ocean layer with higher heat capacity (Gregory et al. 2002, Held et al. 2010, Schwartz 2012). The upper, mixed (denoted by subscript m) component responds more quickly to perturbations (radiative forcing such as an increase in greenhouse gases) and refers to the mixed layers of the atmosphere and shallow ocean. The deeper (denoted by subscript d) component responds slower and allows for delayed, recalcitrant warming (Held et al. 2010). In other words, if external forcing stopped and the system was in disequilibrium, high temperatures in the deep component would lead to a slow warming of the mixed atmosphere component (and vice versa). This energy balance model explicitly takes ocean-atmosphere interactions into account. A simple two-component EBM is given here as differential equations in (5)-(6). Let T_m be the temperature anomaly (deviations from steady state) for the mixed layer, C_m is the component heat capacity and F denotes radiative forcing. Heat capacity denotes the amount of energy needed to change the temperature of the component. Net heat flux denoting the net incoming energy is given by $Q = F - \lambda T_m$ where climate feedback is defined as λ, the change in downward net flux with temperatures T_m. The term λT captures the increasing outgoing long-wave radiation with increased temperatures.

The upper component of the model describes the changes in heat content relative to the steady state $C_m \frac{dT_m}{dt}$ as a function of net flux $F - \lambda T_m$ and the heat exchange H with the lower component:

$$C_m \frac{dT_m}{dt} = -\lambda T_m - H + F \tag{5}$$

The change in the heat content of the lower component $C_d \frac{dT_d}{dt}$ is given by:

$$C_d \frac{dT_d}{dt} = H \tag{6}$$

where C_d denotes the effective heat capacity of the deep compartment and T_d is the associated temperature anomaly. The heat exchange between the bottom and upper component is assumed to be proportional to the difference in temperatures:

$$H = \gamma(T_m - T_d) \tag{7}$$

[1] The shallow depth implies the lower "slow" component will not be as slow as suggested in other studies.

Substituting this expression into the above equations yields the two component model as a system of two differential equations:

$$C_m \frac{dT_m}{dt} = -\lambda T_m - \gamma(T_m - T_d) + F \tag{8}$$

$$C_d \frac{dT_d}{dt} = \gamma(T_m - T_d) \tag{9}$$

Parameters are traditionally estimated using individual equations, or calibrated based on theory or general circulation model (GCM) simulations. I show in section 2.4 that this system of differential equations in discrete time is mathematically equivalent to a cointegrated vector auto-regression which can be estimated using econometric techniques.

2.3 COINTEGRATED VECTOR AUTOREGRESSIONS

Cointegration analysis can be used to overcome difficulties associated with stochastic trends in time series and applied to test whether there exist combinations of non-stationary series that are themselves stationary. To model both the short-run and long-run relations between non-stationary time series in energy balance models, I outline the structure of cointegrated vector auto-regressions in econometrics[2] (CVAR) and in the following section show the equivalence to the two-component energy balance model.

Let $\mathbf{Y}_t = (Y_{1,t}, Y_{2,t}, ..., Y_{p,t})'$ denote a p-dimensional vector of p non-stationary (specifically I(1) or integrated of order one[3]) variables to be modelled for which observations are available over time. In equilibrium-correction form the CVAR model is given by:

$$\Delta\mathbf{Y}_t = \Gamma_1 \Delta\mathbf{Y}_{t-1} + ... + \Gamma_k \Delta\mathbf{Y}_{t-k} + \Pi\mathbf{Y}_{t-1} + \Phi + \boldsymbol{\epsilon}_t \tag{10}$$

where Δ is the first difference operator, such that $\Delta Y_t = Y_t - Y_{t-1}$. The CVAR describes changes in the endogenous $\mathbf{Y}_t$ variables as a function of past changes up to k lags through the cofficient matrices Γ_1, Γ_k and changes in the long run-dynamics as captured through the coefficient matrix Π. The rank of Π determines the presence of cointegrating, or long-run, stationary relations. If the rank of $\Pi = 0$, there are no cointegrating relations, and if Π is full rank, the system must be stationary. The matrix Π will have full rank if the series tested are stationary ($\sim$ I(0)) or if all internal stochastic trends can be explained by exogenous variables entering the cointegrating relation. The matrix has a rank of zero if the series do not cointegrate, and reduced rank (p $>$ rank $>$ 0) if the series cointegrate. The number of cointegrating relations is chosen using the trace statistic which tests the null hypothesis that there are $r - 1$ cointegrating relations against the null hypothesis that there are r relations (Johansen 1988, Johansen 1995). To alleviate concerns associated with small samples, a bootstrapping procedure to test the cointegrating rank (see Cavaliere et al. 2012) could be used. Using a system approach (Johansen 1988) avoids many problems associated with single equation cointegration procedures (e.g. Engle & Granger 1987) and does not require pre-testing of the individual series for stationarity (for a discssion of potential hazards see Pretis & Hendry 2013).

[2] See Hendry & Juselius (2001), and Juselius (2006) for a comprehensive overview of cointegration methods.
[3] A series that is $I(s)$, or integrated of order s needs to be differenced s times to be stationary.

In cointegration the long-run dynamics are modelled explicitly by decomposing the estimated matrix Π under reduced rank into a cointegrating vector β and the adjustment coefficients α to this cointegrating vector:

$$\alpha\beta' = \Pi \tag{11}$$

It then holds that $\beta'\mathbf{Y_t}$ forms a stationary combination that describes the long-run dynamics between the endogenous variables. The elements of α describe how the individual variables adjust to this long-run relation.

Consider the case of three endogenous variables: $Y_{1,t}, Y_{2,t}, Y_{3,t}$ with a single lag in the CVAR specification where I have assumed for illustration purposes the presence of a single cointegrating relation (rank of $\Pi = 1$):

$$\begin{bmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \\ \Delta Y_{3,t} \end{bmatrix} = \mathbf{\Gamma} \cdot \begin{bmatrix} \Delta Y_{1,t-1} \\ \Delta Y_{2,t-1} \\ \Delta Y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \cdot \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \end{bmatrix} \cdot \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \end{bmatrix} + \epsilon_t \tag{12}$$

The cointegrating relation is given by $\beta'\mathbf{Y}_{t-1}$ which is stationary. In equilibrium it holds that:

$$\beta'\mathbf{Y}_{t-1} = (\beta_1 Y_{1,t-1} + \beta_2 Y_{2,t-1} + \beta_3 Y_{3,t-1}) = 0 \tag{13}$$

If the expression is not in equilibrium, $\beta'\mathbf{Y}_{t-1} \neq 0$, then this disequilibrium moves the variables towards the new equilibrium through the adjustment coefficients $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$.

2.4 LINKING EBMS TO COINTEGRATION AND MAPPING TO CVARS

I now show that the system of differential equations of the two-component EBM is equivalent to a restricted cointegrated system and can be mapped to a restricted CVAR. The energy balance model in (8) and (9) can be expressed as

$$\frac{dT_m}{dt} = \frac{1}{C_m}\left[-\lambda T_m + F\right] + \frac{\gamma}{C_m}\left[T_m - T_d\right] \tag{14}$$

$$\frac{dT_d}{dt} = \frac{\gamma}{C_d}\left[T_m - T_d\right] \tag{15}$$

with no explicit formulation given for $\frac{dF}{dt}$ in the EBM. Let α and β denote $(p \times r)$ matrices (where $p = 3, r = 2$) corresponding to:

$$\alpha = \begin{bmatrix} \frac{1}{C_m} & \frac{\gamma}{C_m} \\ 0 & \gamma \\ 0 & 0 \end{bmatrix} \text{ and } \beta' = (\beta_1, \beta_2)' = \begin{bmatrix} -\lambda & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \tag{16}$$

Then the two-component EBM can be written as the following system:

$$d\mathbf{Y} = \alpha\beta'\mathbf{Y}dt \tag{17}$$

where $\mathbf{Y} = (T_m, T_d, F)'$, a $(3 \times 1$ vector).

Assuming stochastic processes (e.g. due to imperfect measurement and omitted variables) the two component EBM can be written with a noise term $\mathbf{v}$ as an Ornstein-Uhlenbeck process given by the stochastic differential equation:

$$d\mathbf{Y} = \Pi\mathbf{Y}dt + \mathbf{v} \tag{18}$$

$$= \alpha\beta'\mathbf{Y}dt + \mathbf{v} \tag{19}$$

where Π is a $(p \times p)$ matrix of reduced rank $r = 2$ such that $\Pi = \alpha\beta'$. The noise term $\mathbf{v} = DdW$ captures omitted effects from the model, with D being a $(p \times p)$ matrix and W a p-dimensional Brownian motion. The following analysis relies on the results in Kessler & Rahbek (2004) and Kessler & Rahbek (2001).

The formulation of the stochastic two-component energy balance model then implies that, if the continuous Ornstein-Uhlenbeck process is integrated of order one, I(1), the series in the EBM cointegrate into two stationary relations given by $\beta'\mathbf{Y}$ with adjustment coefficients of α. The physical stochastic two-component energy balance model is equivalent to a restricted cointegrated system in continuous time in (20) and (21) where the expressions in brackets correspond to the cointegrating vectors, and the coefficients outside the brackets are the α adjustment coefficients in the cointegration model:

$$\frac{dT_m}{dt} = \underbrace{\frac{1}{C_m}}_{\alpha_{1,1}} \underbrace{[-\lambda T_m + F]}_{\beta_1'\mathbf{Y}} + \underbrace{\frac{\gamma}{C_m}}_{\alpha_{1,2}} \underbrace{[T_m - T_d]}_{\beta_2'\mathbf{Y}} + v_{1,t} \tag{20}$$

$$\frac{dT_d}{dt} = \underbrace{\frac{\gamma}{C_d}}_{\alpha_{2,2}} \underbrace{[T_m - T_d]}_{\beta_2'\mathbf{Y}} + v_{2,t} \tag{21}$$

The energy balance model implies that the three variables of upper component temperatures, lower component temperatures, and radiative forcing cointegrate into two stationary relations. The first $\left(\beta_1'\mathbf{Y} = [-\lambda T_m + F]\right)$, an equation linking upper component temperatures to radiative forcing describing the net heat flux (Q). The second $\left(\beta_2'\mathbf{Y} = [T_m - T_d]\right)$, an equation describing the heat transfer between the upper and lower component. Upper component temperatures adjust to both cointegrating vectors, lower component temperatures adjust to the second cointegrating vector ($\alpha_{2,1} = 0$), and radiative forcing is weakly exogenous ($\alpha_{3,1} = \alpha_{3,2} = 0$). Cointegration between these variables is a testable property, and cointegration analysis can further be used to estimate and conduct inference on the parameters α and β.

The differential equations describe the two-component energy balance model in continuous time, while sampling happens at discrete intervals. Inference on cointegration and parameter estimates in the continuous-time EBM can be conducted based on discrete observations using the results in Kessler & Rahbek (2004).[4] It is important to note though that the discretization method moving from discrete to continuous time can have effects on the auto-correlation structure of the error term, and thus affect the statistical properties of the estimates (Sargan

[4] In particular, Theorem 1 in Kessler & Rahbek (2004).

1974). Given the specification in equation (19) as an Ornstein-Uhlenbeck process, discrete observations $\mathbf{Y}_t$ of $\mathbf{Y}$ follow a VAR process as in equation (22):

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \boldsymbol{\epsilon_t} \tag{22}$$

where $\boldsymbol{\epsilon}_t \sim N(0, \Sigma)$. The coefficient matrix $\mathbf{A}$ in the VAR in (22) equals the matrix exponential of the coefficient matrix Π in equation (19), $\mathbf{A} = \exp(\Pi)$. This result derives from solving the Ornstein-Uhlenbeck process[5] in equation (19) – see appendix 2.7.1. Let $\mathbf{P} = \mathbf{A} - I$, then (22) can be written in equilibrium correction form as:

$$\Delta\mathbf{Y}_t = \mathbf{P}\mathbf{Y}_t + \boldsymbol{\epsilon}_t \tag{23}$$

There is then a direct mapping of the EBM parameters from equation (19) to the parameters in the discrete-time model (23). In particular, the coefficient matrix $\mathbf{P}$ in the CVAR relates to the EBM parameters α, β as:

$$\mathbf{P} = \mathbf{A} - I = \exp(\Pi) - I = \exp\left(\alpha\beta'\right) - I = \alpha\kappa\beta' = \tilde{\alpha}\beta' \tag{24}$$

where $\tilde{\alpha} = \alpha\kappa$ and the $(r \times r)$ matrix $\kappa = \left(\beta'\alpha\right)^{-1}\left[\exp\left(\beta'\alpha\right) - I\right]$, see appendix 2.7.1 for a proof. This mapping implies that the two-component continuous-time EBM can be tested and estimated using a CVAR estimated on discrete observations: the rank of the coefficient matrix of the discrete-time VAR model equals that of the EBM, $\text{rank}(\mathbf{P}) = \text{rank}(\Pi) = \text{rank}(\alpha\beta')$, and the matrix $\widehat{\mathbf{P}} = \hat{\tilde{\alpha}}\hat{\beta}'$ yields estimates of the EBM parameters given by β'. The adjustment coefficients α are not identified given that the CVAR parameters $\tilde{\alpha}$ correspond to $\alpha\kappa$. Thus, restrictions in α are not directly preserved in the CVAR formulation. In other words, while in the continuous time EBM the lower component temperatures T_d do not adjust to the first cointegrating vector, $\alpha_{2,1} = 0$, this does not directly imply that the same holds for the discrete-time CVAR, $\tilde{\alpha}_{2,1} \neq 0$. An intuitive explanation for this comes from the fact that at the discrete time interval the lower component may adjust to the upper component given the feedback between T_m and T_d. The extent of this can be tested by assessing whether $\tilde{\alpha}_{2,1}$ equals zero or not using CVAR estimates.[6] One exception on restrictions linking α and $\tilde{\alpha}$ occurs under weak-exogeneity. The forcing series is assumed to be weakly-exogenous (not adjusting to the cointegrating relations) in the EBM, thus $\alpha_{3,1} = \alpha_{3,2} = 0$. Given $\tilde{\alpha} = \alpha\kappa$ this implies that $\tilde{\alpha}_{3,1} = \tilde{\alpha}_{3,2} = 0$, forcing should also be weakly exogenous in the CVAR formulation. Writing the model (23) in matrix form we can formulate equations (20) and (21) as a discrete CVAR where the CVAR parameters (given by $\tilde{\alpha}$ and β) correspond to (combinations of) parameters of the EBM:

$$\begin{bmatrix} \Delta T_{m,t} \\ \Delta T_{d,t} \\ \Delta F_t \end{bmatrix} = \begin{bmatrix} \tilde{\alpha}_{1,1} \\ \tilde{\alpha}_{2,1} \\ \tilde{\alpha}_{3,1} \end{bmatrix} \cdot \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \end{bmatrix} \cdot \begin{bmatrix} T_{m,t-1} \\ T_{d,t-1} \\ F_{t-1} \end{bmatrix} \tag{25}$$

$$+ \begin{bmatrix} \tilde{\alpha}_{1,2} \\ \tilde{\alpha}_{2,2} \\ \tilde{\alpha}_{3,2} \end{bmatrix} \cdot \begin{bmatrix} \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \end{bmatrix} \cdot \begin{bmatrix} T_{m,t-1} \\ T_{d,t-1} \\ F_{t-1} \end{bmatrix} + \boldsymbol{\epsilon}_t \tag{26}$$

[5] Where for simplicity the frequency of observations is assumed to be equal to one.
[6] Sections 2.5 and 2.7.5 investigate this further.

The cointegrating vectors and adjustment coefficients are linked to the continuous-time two-component EBM such that: $\beta_{1,1} = -\lambda, \beta_{1,2} = 0, \beta_{1,3} = 1$, linking upper component temperatures and radiative forcing in the first cointegrating relation; $\beta_{2,1} = -\beta_{2,2} = 1$, and $\beta_{2,3} = 0$, linking upper component temperatures and lower component temperatures in the second cointegrating relation; and $\tilde{\alpha}_{3,1} = \tilde{\alpha}_{3,2} = 0$, as forcing is assumed to adjust to neither cointegrating relation. In matrix notation this can be expressed as:

$$\begin{aligned} \begin{bmatrix} \Delta T_{m,t} \\ \Delta T_{d,t} \\ \Delta F_t \end{bmatrix} &= \begin{bmatrix} \tilde{\alpha}_{1,1} \\ \tilde{\alpha}_{2,1} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} -\lambda & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} T_{m,t-1} \\ T_{d,t-1} \\ F_{t-1} \end{bmatrix} \\ &+ \begin{bmatrix} \tilde{\alpha}_{1,2} \\ \tilde{\alpha}_{2,2} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} T_{m,t-1} \\ T_{d,t-1} \\ F_{t-1} \end{bmatrix} + \boldsymbol{\epsilon}_t \end{aligned} \tag{27}$$

This yields the discrete-time CVAR representation of the stochastic energy balance model and can be estimated using standard CVAR methods.

2.4.1 *Testable Properties implied by the EBM-CVAR mapping*

The equivalence of a two component EBM and a CVAR provides direct restrictions which can be tested against observations. Formally the EBM implies the following properties of the estimated CVAR (10) which can then be tested:

1. rank(Π)=2 (Trace test): the time series cointegrate to two stationary relations, linking upper component temperatures with radiative forcing, and lower component temperatures with upper component temperatures.
2. Restrictions in (27) are not rejected (Likelihood ratio test).
3. Estimates for β (and $\tilde{\alpha}$) are statistically different from zero, (Likelihood ratio tests) and theory-consistent.

In particular, for the EBM to hold we require that the three series – upper component temperatures, lower component temperatures, and radiative forcing – cointegrate into two stationary relations: The first, an equation linking upper component temperatures to radiative forcing, describing the net heat flux (Q):

$$\mathbf{h}_{1,t} = \left(\beta_{1,1} T_{m,t-1} + \beta_{1,3} F_{t-1} \right) \tag{28}$$

where $\beta_{1,1} = -\lambda$, and $\beta_{1,3} = 1$. The second, an equation describing the heat transfer between the upper and lower component:

$$\mathbf{h}_{2,t} = \left(\beta_{2,1} T_{m,t-1} + \beta_{2,2} T_{d,t-1} \right) \tag{29}$$

where $\beta_{2,1} = 1 = -\beta_{2,2}$. The order of cointegration, can be tested using the Johansen trace test (Johansen 1988).

With respect to the physical structure between variables, this implies that any stochastic trends present in the radiative forcing (where the stochastic component is driven by anthro-

pogenic emission of greenhouse gases from economic activity) will be imparted onto the temperatures of the mixed component, through which it transfers onto the lower component. This generalizes the results in Kaufmann et al. (2013) to the two-component EBM.

Notably, estimating a two-component EBM through a CVAR avoids potential hazards associated with statistical and econometric climate models. The Johansen cointegration procedure requires no pre-testing of individual series for the order integration (up to I(1)) and takes the system nature of the model into account. The model also relies on the aggregate of radiative forcing, rather than single series which, individually, may suffer from data measurement problems.

2.4.2 *Estimation using ocean heat content observations*

The EBM characterises changes in temperatures while the data available for applications often is a mixture of heat content and temperature observations. Given the availability of data, the application in section 2.5 uses ocean heat content (H_t) for the lower component and temperatures for the upper component. This transformation yields a re-parametrised, but otherwise identical EBM for estimation where I use $H_t \approx C_d T_{d,t}$, the re-parametrised EBM is given as:

$$\begin{aligned}\begin{bmatrix}\Delta T_{m,t}\\ \Delta H_{,t}\\ \Delta F_t\end{bmatrix} &= \begin{bmatrix}\tilde{\alpha}_{1,1}\\ \tilde{\alpha}_{2,1}\\ 0\end{bmatrix}\cdot\begin{bmatrix}-\lambda & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}T_{m,t-1}\\ H_{t-1}\\ F_{t-1}\end{bmatrix} \\ &+ \begin{bmatrix}\tilde{\alpha}_{1,2}\\ \tilde{\alpha}_{2,2}\\ 0\end{bmatrix}\cdot\begin{bmatrix}1 & -1/C_d & 0\end{bmatrix}\cdot\begin{bmatrix}T_{m,t-1}\\ H_{t-1}\\ F_{t-1}\end{bmatrix}+\boldsymbol{\epsilon}_t\end{aligned} \tag{30}$$

To determine the cointegrating rank of the model, estimation includes an unrestricted constant. If the rank equals two, then the first corresponding cointegrating vector is restricted and given by:

$$\mathbf{h}_{1,t} = \left(\beta_{1,1} T_{m,t-1} + \beta_{1,3} F_{t-1}\right) \tag{31}$$

The second, a restricted equation for temperature changes in the lower component:

$$\mathbf{h}_{2,t} = \left(\beta_{2,1} T_{m,t-1} + \beta_{2,2} H_{t-1}\right) \tag{32}$$

For identification both cointegrating vectors are normalized such that $\beta_{1,3} = 1$ and $\beta_{2,1} = 1$. The additional restrictions given by the EBM are: lower component temperatures do not enter the first cointegrating vector, $\beta_{1,2} = 0$; radiative forcing does not enter the second cointegrating vector, $\beta_{2,3} = 0$, and is weakly exogenous, $\alpha_{3,1} = \alpha_{3,2} = 0$ which implies that $\tilde{\alpha}_{3,1} = \tilde{\alpha}_{3,2} = 0$. This is a simplification since some of the radiative forcing series are determined endogenously, CO_2 concentrations generally vary with temperatures as does ice-albedo. The weak exogeneity assumption is relaxed in a simple extension of the model (see model C in Table 4) where an aggregate of radiative forcing is modelled endogenously.

Every parameter in the cointegrating vector of the CVAR has a structural interpretation. To estimate the model parameters of interest, it then holds that in the second cointegrating vector

(32) $\beta_{2,2} = -1/C_d$, providing an estimate of the heat capacity of the deep component. The climate feedback parameter λ can be determined using $\beta_{1,1} = -\lambda$. Approximate standard errors and confidence intervals can be derived using the asymptotic normality of the cointegrating parameters. In the case of parameters of interest being non-linear transformations of the cointegrating estimates (such as $-1/C_d = \beta_{1,1}$), I rely on the δ-method to provide approximate standard errors.[7] Based on these restrictions all EBM parameters in the approximation can be identified and the over-identifying restrictions can be tested.

Appendix 2.7.5 considers an alternate approach mapping the EBM to a CVAR using a simple discrete-time approximation. In this setup further restrictions on the adjustment coefficients are imposed for identification of additional EBM parameters.

2.5 APPLICATION

Here I estimate a CVAR as a simple test of the two-component EBM, and to obtain estimates of the model parameters, notably estimated climate sensitivity, while accounting for the non-stationary nature of the climate time series.

2.5.1 *Data*

The upper/mixed component is specified as the atmosphere at the surface and temperatures are taken as global mean surface temperature anomalies from the GISS dataset (Hansen et al. 2010). The lower component is proxied by the ocean heat content anomalies from 0-700m (Levitus et al. 2009). This is a crude approximation as it omits the deep ocean and will therefore provide a high estimate of the adjustment speed of this component.[8] The series are graphed in Figure 4. Radiative forcing time series (Figure 5) are taken from Meinshausen et al. (2011). The aggregate of the forcing used for the model includes well-mixed greenhouse gases (WMGHGs), stratospheric water vapour (Str. H_2O), stratospheric ozone (O_3), black carbon, stratospheric aerosols, snow albedo changes, land use changes, indirect aerosol effects and reflective aerosols. The short time period from 1955-2010 masks the dramatic increase in radiative forcing stemming from WMGHGs since the late 19th century.

To directly estimate the parameters of the EBM here I rely on the aggregate of radiative forcing. An extension could consider disaggregates by separating out volcanic forcing in the form of stratospheric aerosols. Stern & Kaufmann (2014) test for links between individual forcing series and surface temperatures, further work could focus on more detailed models using individual forcing series found to be significant in Stern & Kaufmann (2014), while series with little impact (e.g. black carbon) could be omitted. Few econometric models accounting for non-stationarity in the form of stochastic trends also consider ocean heat content and thus the heat transfer between the atmosphere/top component and the ocean. Stern (2006) is a notable exception. Relative to Stern (2006) here I provide a direct link to the two component EBM and estimate the system jointly as a CVAR.

[7] Given a sequence of random variables X_n satisfying $\sqrt{n}(X_n - \mu) \xrightarrow{D} N\left(0, \sigma^2\right)$ and a continuous function $g(\mu)$, using a Taylor-expansion we can approximate the distribution of $g(X_n)$ by $\sqrt{n}\left(g(X_n) - g(\mu)\right) \xrightarrow{D} N\left(0, \sigma^2[g'(\mu)]^2\right)$.

[8] Data for deeper ocean heat content for 0-2000m is available but only through pentadal averages (Levitus et al. 2012).

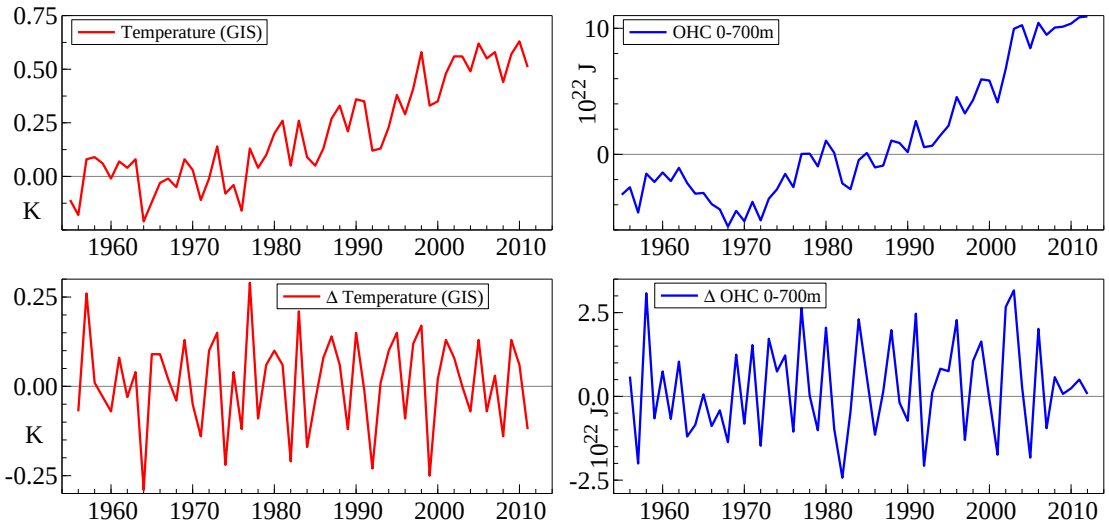


Figure 4: Left: Global mean surface temperatures in levels (top) and first differences (bottom). Right: Ocean heat content from 0-700m in levels (top) and first differences (bottom).

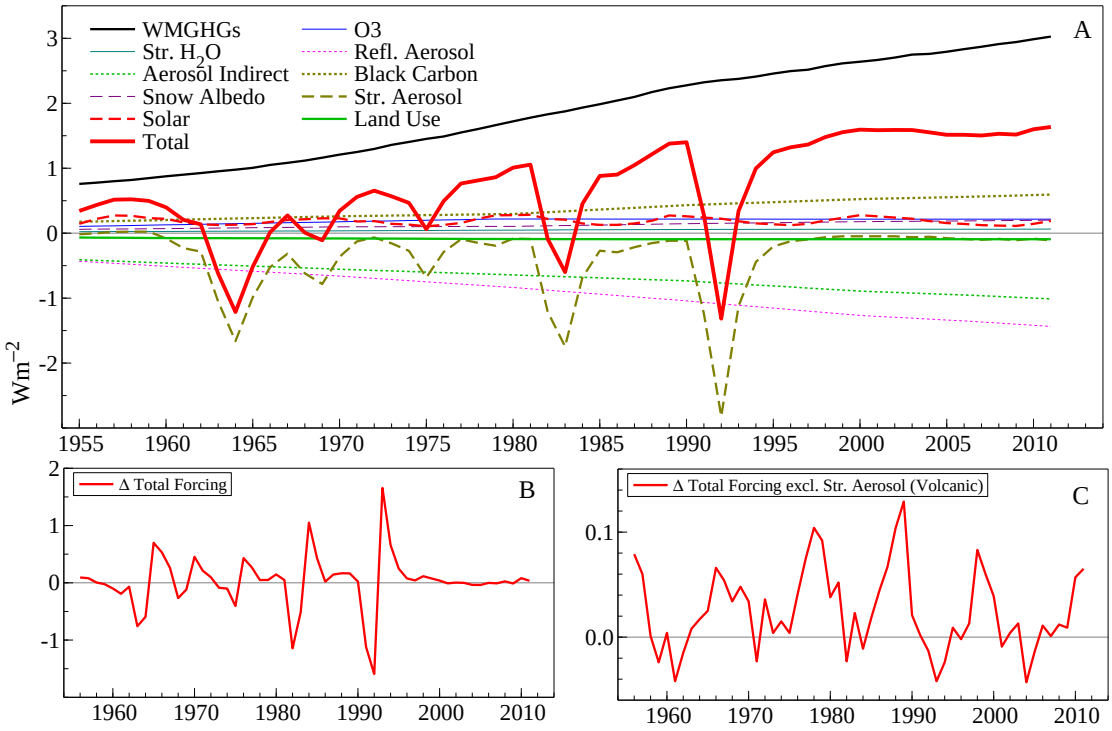


Figure 5: Radiative forcing from 1955-2011 in Wm^{-2}. Panel A shows disaggregated forcing, panel B shows the first difference of total forcing, panel C graphs first difference of total forcing excluding stratospheric aerosol forcing from volcanic eruptions. Data obtained from Meinshausen et al. (2011).

Data availability for ocean heat content constrains the sample to a period from 1955 onwards. The estimation period is then 1955-2011 at an annual frequency, resulting in T=56 observations. This time period, together with the use of an aggregate of radiative forcing rather than individual series, alleviates some concerns over measurement changes and the data quality of the radiative forcing series (see Pretis and Hendry 2013).

While the time series properties of radiative forcing series are heavily debated in the econometric literature (Kaufmann & Stern 2002, Stern & Kaufmann 2000, Beenstock et al. 2012, Pretis and Hendry 2013), the individual order of integration of these series is not directly relevant. Within the EBM formulation it is aggregate forcing that drives upper component temperatures. On initial inspection it may appear inconsistent that global mean surface temperatures are well approximated by an I(1) process while well-mixed greenhouse gases (in particular) atmospheric CO_2 concentrations can be argued to be I(2) (see appendix section 2.7.6 and above sources). However, jointly the individual forcing series sum (or cointegrate) to an I(1) aggregate of forcing, where the I(1) cointegrating relation is given simply by the sum of the individual forcing series. Given the same unit of measurement and the aggregate effect, forcings are directly summable – this is a crucial oversight in the analysis of Beenstock et al. (2012) who argue that, due to the different time series properties, temperatures and WMGHGs cannot be related. While individual forcing series may be I(2) or I(0), aggregate forcing appears to be consistent with an I(1) process, and thus exhibits approximately the same order of integration as global mean surface temperatures.[9]

2.5.2 *Model Specification*

Determining the number of cointegrating relations and conducting inference on the estimated parameters relies on well-specified models (Juselius 2006). While the theoretical EBM model does not imply additional lags, the time series properties of the data may support a longer lag length. To formally determine the lag structure I estimate a general unrestricted VAR starting with three lags. Removing the third lag is not rejected ($p = 0.08$), while dropping both the second and third lag is rejected ($p = 0.003$). The model with the lowest Schwarz criterion (Schwarz 1978) includes just one lag (SC=3.28, relative to 3.45 for two, and 3.76 for three lags). Assessing the diagnostic tests of the unrestricted models, a VAR(1) model rejects no-residual autocorrelation (p=0.003), while a VAR(2) passes the residual vector autocorrelation test (p=0.59) and dropping the second lag in the VAR(2) is rejected (p=0.005). Many discrete approximations use the average of two periods which further provides a justification for the use of two lags. Additional results on the dynamic stability and unit-root properties of the estimated VAR models are provided in appendix 2.7.2.

I proceed by estimating four variations of the CVAR EBM model. First, model A is estimated with a single lag corresponding to the theoretical two-component EBM. While this model is likely mis-specified and thus inference should be interpreted with care, it is included as a comparison to show the estimates if just the theory model is estimated. Second, model B includes two lags as suggested by the tests determining the lag structure. Once the restrictions implied by the EBM are imposed, both models A and B do not model forcing explicitly. Nevertheless, some of the forcing series are not exogenous but rather may vary with observed

[9] While not required by the Johansen methodology, for completeness univariate unit root tests are provided in the appendix in section 2.7.6

temperatures. Model C (estimated with 2 lags) relaxes the strong restrictions from model B by allowing aggregate forcing to adjust to the first cointegrating relation. This changes the system to three variables, however, no additional variables are included to explain forcing. In other words, radiative forcing is modelled endogenously and is allowed to adjust to the first cointegrating vector ($\tilde{\alpha}_{3,1} \neq 0$). Further research could expand the EBM to disaggregate the forcing series and also incorporate a macroeconomic VAR to model the anthropogenic emissions, as well as additional equations modelling natural variability (notably snow albedo changes).

An additional concern for residual autocorrelation are the time series properties of volcanic forcing (stratospheric aerosols) which act more as structural breaks than continuous series and may induce auto-correlation in the error terms. Therefore, a preliminary fourth specification, reported as Model V (in appendix 2.7.3) provides an initial attempt at modelling a simple disaggregate of the forcing series where volcanic aerosol forcing are modelled as breaks to avoid concerns of estimating climate sensitivities based on volcanic forcing in general (see e.g Lindzen & Giannitsis 1998). Only limited results are reported on model V as the precise modelling of volcanic eruptions requires a detailed treatment exceeding the scope of this paper in which the application to real data serves primarily as an illustration of estimating EBMs using CVARs.

In summary, model A corresponds directly to the EBM theory with a single lag, model B is estimated with two lags to account for additional dynamics and residual autocorrelation, model C is equivalent to model B with weak-exogeneity of the forcing series relaxed. Model V in appendix 2.7.3 provides an initial attempt at accounting for break-like volcanic forcing. Full results are reported for models A-C.

2.5.3 *Results*

The time series of global mean surface temperature anomalies, 0-700m ocean heat content anomalies, and total radiative forcing cointegrate to two stationary relations consistent with the theory of a two-component EBM. Likelihood ratio tests cannot reject the presence of two cointegrating vectors (rank = 2) in all the models estimated here (p_A=0.58, $p_B = p_C = 0.76$ for models A, B, C respectively), consistent with the theory provided by the two-component EBM (3). A single cointegrating relation (rank=1) is rejected in all models (see Table 3 for full results).[10]

Table 4 provides the estimated parameters of the CVAR. Fitted values are given in levels in Figure 6 and first differences in Figure 7. Residual plots are provided in appendix 2.7.4. The imposed restrictions on the CVAR are not rejected at the 1% level (p=0.02) in model A (see Table 4), though this should be interpreted with care given the failure of the residual autocorrelation mis-specification test of the model. The restrictions are, however, rejected in model B which controls for residual autocorrelation. To explore the rejection of the restrictions, model C relaxes weak-exogeneity of the forcing series. Taking into account the error serial correlation through the extension of the model to two lags, the model restrictions are not rejected in model C once forcing is allowed to adjust to upper component temperatures (p=0.32). This is a notably strong result – it is not unusual for restrictions of this complexity

[10] Future work will explore the effects of the small sample sizes using the bootstrap rank testing approach in Cavaliere et al. 2012.

to be rejected.[11] For completeness, despite the rejection of the restrictions in model B and mis-specification in A, the proceeding analysis reports results for all models A, B, and C.

Normality of the residuals cannot be rejected in diagnostic tests of the models A and B, it is, however, rejected for model C. The rejection of normality in C likely stems from the forcing series as indicated by the single-equation diagnostic tests – normality is not rejected in all but the forcing series. This can be driven by the volcanic impacts resembling structural breaks. Model V in (appendix 2.7.3), modelling volcanic forcing as breaks, shows improvements in diagnostic test results (see also the residual plots in appendix 2.7.4) and the EBM restrictions (including weak exogeneity of forcing) are then not rejected (p=0.17).

The estimated parameters of the EBM in the cointegrating relations are highly significant and estimated with theory-consistent signs. There is little evidence that the ocean heat content series adjusts to the first cointegrating vector ($\tilde{\alpha}_{2,1}$ is insignificant in Table 4). This is consistent with the continuous-time EBM where the lower component only adjusts to the second cointegrating relation in (21).[12] The two cointegrating relations representing the equations of the EBM are reported here as an example for the EBM theory model A:

$$\widehat{\mathbf{h}}_{1,t} = \left(\underset{(0.51)**}{-\,2.71}\; T_{m,t-1} + F_{t-1} \right) \tag{33}$$

$$\widehat{\mathbf{h}}_{2,t} = \left(T_{m,t-1} - \underset{(0.003)**}{0.044}\; H_{t-1} \right) \tag{34}$$

where standard errors are given in parentheses and the coefficients on F_{t-1} in the first equation, and the coefficient on $T_{m,t-1}$ in the second equation are normalized to 1 for identification.

Table 3: Cointegration Tests of the Two-Component EBM

	A: Base, 1-lag	**B & C: 2-lag**
Rank=0	61.52 [p<0.00]**	57.90 [p<0.00]**
Rank=1	20.35 [p<0.01]**	22.46 [p<0.01]**
Rank=2	0.31 [p=0.58]	0.09 [p=0.76]
Obs.	56	55

Cointegration rank tests using the Johansen (1988) trace test. p-values are reported in brackets [ ], * indicates rejection at 5%, ** indicates rejection at 1%.

[11] Equally future work will explore the effects of the small sample sizes using the bootstrap hypotheses testing approach in Cavaliere et al. 2014.

[12] Results for models where an additional restriction of $\tilde{\alpha}_{2,1} = 0$ is imposed are reported in appendix 2.7.5

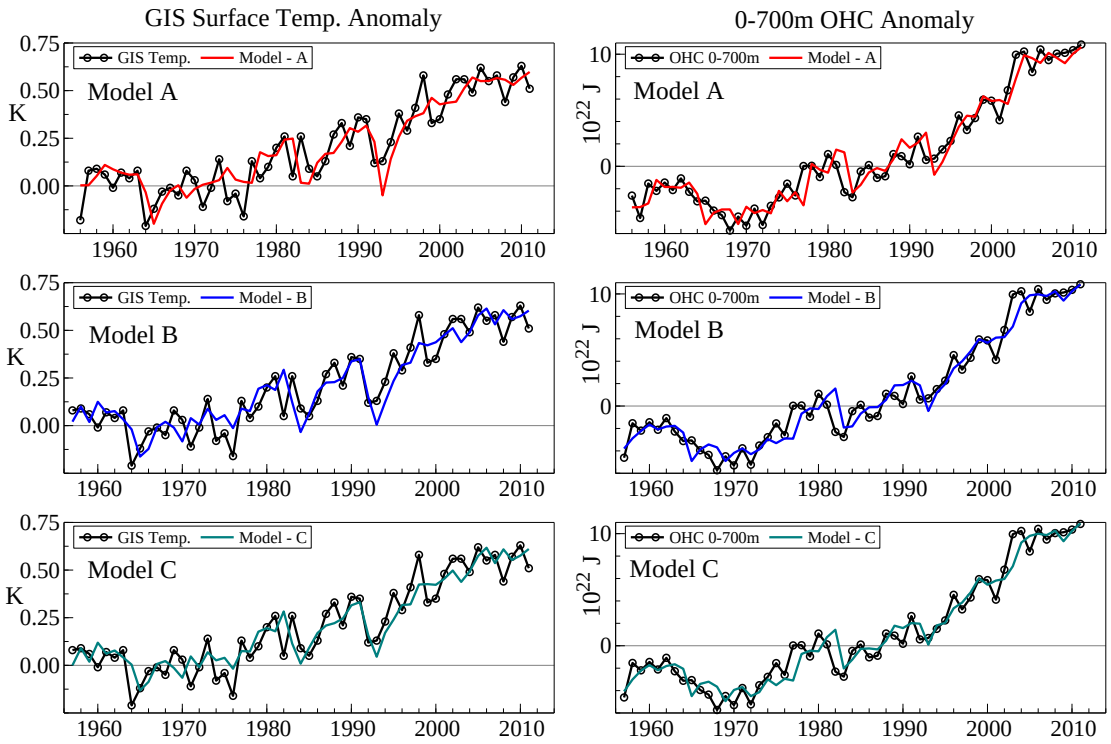


Figure 6: EBM Model Fit (estimated as a CVAR). Left panels shows model fit and observed (GIS) upper-component (global mean-surface) temperature anomalies. Right panels shows model fit and observed 0-700m ocean heat content (OHC) anomalies (Levitus et al. 2009) which proxy lower component heat content in the EBM.

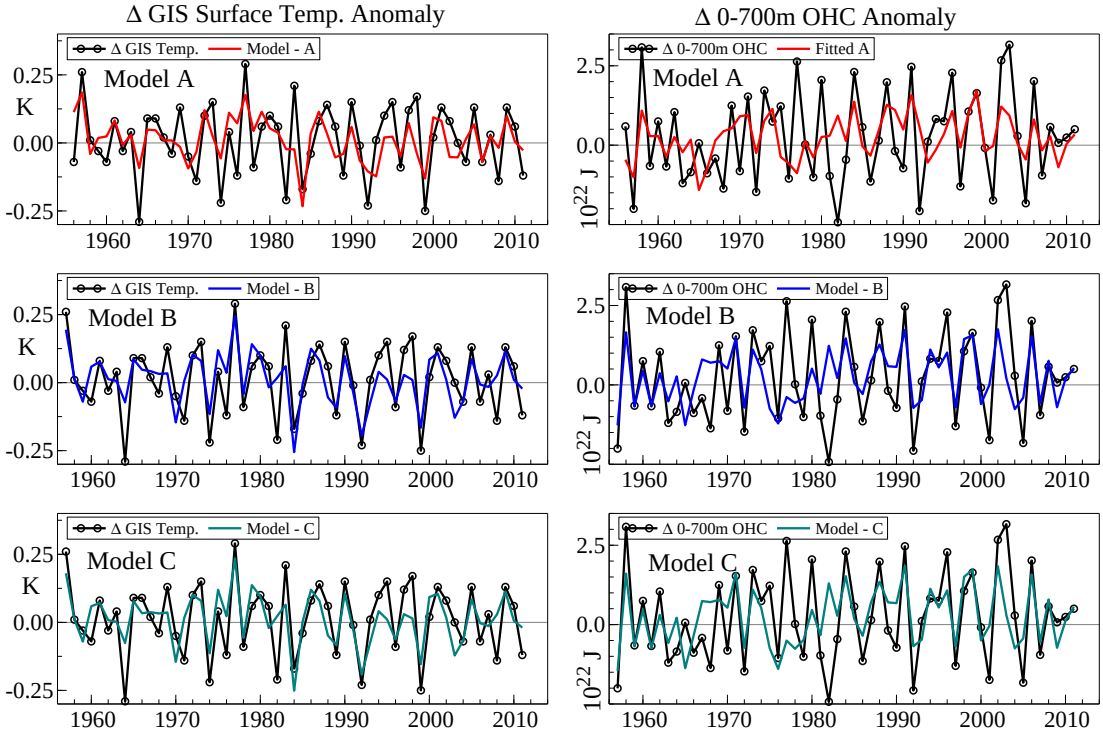


Figure 7: EBM Model Fit of First Differences (estimated as a CVAR). Left panels shows model fit and observed first differences of (GIS) upper-component (global mean-surface) temperature anomalies. Right panels shows model fit and first differences of observed 0-700m ocean heat content anomalies (Levitus et al. 2009) which proxy lower component heat content in the EBM.

Table 4: EBM Cointegration Model Parameter Estimates where $\tilde{\alpha} = \alpha\kappa$

	EBM/CVAR Model		
Coint. Relations	**A: Base**	**B: 2-lag**	**C: Endo. Forc.**
Vector 1 (Mixed)			
$\beta_{1,1}$	-2.71 (0.51)**	-2.29 (0.41)**	-2.21 (0.32)**
$\beta_{1,2}$	-	-	-
$\beta_{1,3}$	1	1	1
Adjustment			
$\tilde{\alpha}_{1,1}$	0.11 (0.03)**	0.15 (0.03)**	0.11 (0.03)**
$\tilde{\alpha}_{2,1}$	0.60 (0.33)	0.53 (0.38)	-0.03 (0.41)
$\tilde{\alpha}_{3,1}$	-	-	-0.55 (0.15)**
Vector 2 (Deep)			
$\beta_{2,1}$	1	1	1
$\beta_{2,2}$	-0.044 (0.003)**	-0.041 (0.004)**	-0.041 (0.004)**
Adjustment			
$\tilde{\alpha}_{1,2}$	-0.41 (0.12)**	-0.46 (0.15)**	-0.47 (0.15)**
$\tilde{\alpha}_{2,2}$	7.82 (1.48)**	7.43 (1.99)**	7.35 (1.99)**
$\tilde{\alpha}_{3,2}$	-	-	-
LR Test of Restric.			
$\chi^2(2)$, C: $\chi^2(1)$	12.87 [p=0.012]*	16.32 [p=0.003]**	2.47 [p=0.12]
Diagnostic Tests			
AR (1-2) F-Test	2.96 [p=0.005]**	1.32 [p=0.24]	1.14 [p=0.32]
Normality $\chi^2(4)$, C: $\chi^2(6)$	0.76 [p=0.94]	7.19 [p=0.13]	20.36 [p=0.002]**
Observations T	56	55	55
log-likelihood	-72.48	-59.85	-51.84
EBM Estimates			
λ (W$m^{-2}K^{-1}$)	2.71 (0.51)	2.29 (0.41)	2.21 (0.32)
ECS (K)	1.37 (0.25)†	1.62 (0.29)†	1.67 (0.24)†
C_m (W year $m^{-2}K^{-1}$)	-	-	-
C_d (W year $m^{-2}K^{-1}$)	22.72 (1.54)†	24.39 (2.38)†	24.39 (2.38)†
γ	-	-	-

Model estimates based on CVAR estimation. Standard errors are given in parentheses ( ) while p-values are reported in brackets [ ]. Standard errors are provided where available. If no standard errors are reported, then parameter is restricted or derived from estimated model parameters. * indicates significance at 5%, ** indicates significance at 1%. Standard errors derived using δ-method are marked using †. Left column specifies parameters, right columns shows estimation results. Dash - marks imposed restriction and no identification in the case of the structural EBM parameters given the theoretical result of $\tilde{\alpha} = \alpha\kappa$.

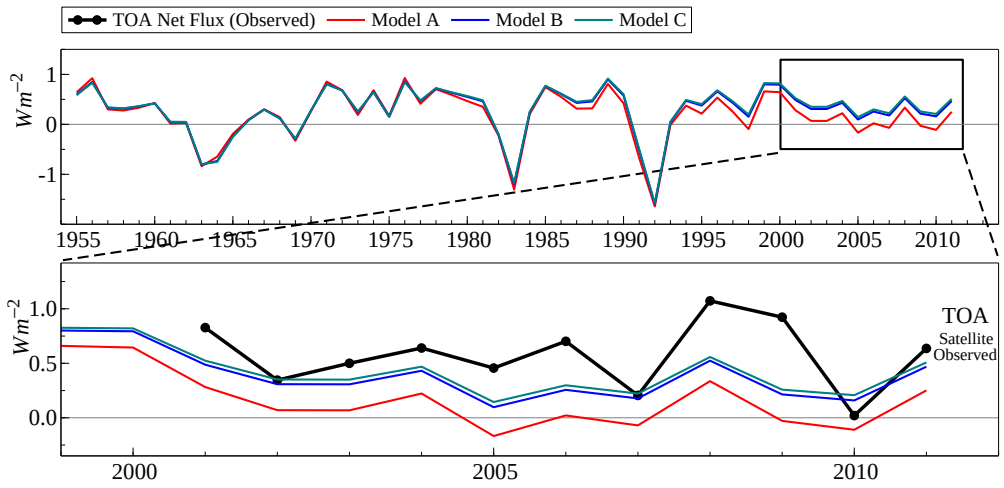


Figure 8: Net heat flux based on the first cointegrating vector and observed top of the atmosphere (TOA) net flux. Top panel graphs the net heat flux as described by the estimated first cointegrating vector for models A-C from 1955-2011. Bottom panel graphs the net flux from 2000-2011 together with annually averaged observed TOA net flux from the CERES Energy Balanced and Filled (EBAF) all skies data (black line with points) (Loeb et al. 2009) which have not been used to estimate the model and serve as an informal test of the model against a different set of independently measured observations.

2.5.4 *Discussion*

Physical Interpretation of First Cointegrating Vector

Based on the EBM in (5) and the equivalent cointegration formulation (28), the first cointegrating relation used to estimate climate feedback has a direct physical interpretation and describes the net heat flux or net energy coming into the system: $Q = F - \lambda T_m$. As an informal test of the model I compare the estimated cointegrating vector which corresponds to estimated net heat flux Q against the independently observed net heat flux from satellite-based measurements of annually averaged net flux at the top of the atmosphere (TOA) from the CERES Energy Balanced and Filled (EBAF) All Skies data (Loeb et al. 2009). It is important to emphasize that this data series is not used in the estimation of the model but is a different set of observations from satellite measurements of the quantity captured by the first cointegrating relation. Figure 8 plots the first cointegrating relation together with the satellite measurements. While the models under-estimate the volatility and level slightly, the first cointegrating vectors from models B and C remain close to the satellite-observed record. Model A, which does not control for residual autocorrelation (and results in high estimates of climate feedback – see section 2.5.4), consistently lies below satellite-observed TOA net flux.

Climate Feedback and Sensitivity

The estimates of the climate response to increasing emissions is characterised by climate feedback λ and the equilibrium climate sensitivity.[13] Estimates of climate feedback and

[13] ECS for a doubling of CO_2 is derived from the steady-state equilibrium of the EBM. Using (8) and (9) in equilibrium it holds that $dT_m/dt = dT_d/dt = 0$ and thus $T_m = F/\lambda$. ECS is then defined as the equilibrium temperature response T_m given a radiative forcing of doubling of CO_2: ECS = $F_{2\times CO_2}/\lambda$, where $F_{2\times CO_2} \approx 3.7 Wm^{-2}$.

equilbrium climate sensitivity (see Table 4) depend on the model specification, where model mis-specification likely leads to an over-estimation of climate feedback (and thus an under-estimation of climate sensitivity). Figure 9 plots the densities of estimated climate feedback (λ) and sensitivity (ECS) relying on approximate asymptotic normality.[14]

Using model A which does not control for residual-autocorrelation, λ is estimated to be 2.71 (0.51) $Wm^{-2}K^{-1}$, equivalent to an equilibrium climate sensitivity of 1.37 (0.25) K for a radiative forcing of $3.7Wm^{-2}$ for a step-shift doubling of CO_2. This is lower than IPCC (IPCC 2013) best estimates of around 3K. A possible reason for why climate feedback may have been over-estimated (and ECS under-estimated) when using observation-based approaches (for example Schwartz (2007) finds ECS for a doubling of CO_2 to be approximately 1.1K) could be model mis-specification (e.g apparent through residual autocorrelation) and stochastic trends in the data which are not accounted for when relying on correlations and not estimating the system as a whole. Once residual autocorrelation is corrected for through the use of additional lags (model B) the estimates of climate feedback are lower (and those of ECS higher). Model B (two-lags) and model C (endogenous forcing) estimate λ at 2.29 (0.41) and 2.21 (0.32) $Wm^{-2}K^{-1}$ respectively, with associated equilibrium climate sensitivities of 1.62 (0.29) K and 1.67 (0.24) K. However, this should be interpreted with care as the now endogenous forcing is modelled as a simple aggregate. Appendix 2.7.5 reports the results when additional restrictions are placed on $\tilde{\alpha}$, generally resulting in higher estimates of ECS.

There are concerns on estimating λ and ECS using volcanic forcing data (Lindzen & Giannitsis 1998). Model V makes an initial attempt in correcting for this by treating volcanic forcing as transitory breaks, entering the CVAR unrestrictedly in first differences. The climate feedback estimate in model V is notably lower than in models A-C, resulting in a higher estimate of ECS of 2.16 (0.56) K, however, associated with higher uncertainty (see Figure 9, model V, purple).

Overall, EBM estimates based on CVAR estimation show that the IPCC estimated range of equilibrium climate sensitivities is robust to the non-stationary nature of the data at hand, though at the higher end of the CVAR estimated distributions.

Component Heat Capacity

While the upper-component heat capacity C_m (the amount of energy needed to change the temperature) is not directly identified in the discrete CVAR representation of the continuous-time EBM, the heat capacity of the lower component can be estimated using the parameters in the second cointegrating vector. The lower component heat capacity C_d is given by the CVAR mapping as $\beta_{2,2} = -1/C_d$. Raw estimates of the heat capacity of the lower component using observational data and a two-component EBM estimated as a CVAR from model A suggests an effective ocean 0-700m heat capacity of 22.72 (2.70) W year $m^{-2}K^{-1}$. This increases to 24.39 (2.38) when two lags are considered (models B) and when we allow forcing to be endogenous (Model C). Thus, when estimating the EBM as a system and taking the time series properties into account, the raw estimate is higher than found by Schwartz (2012), who finds the heat capacity of the lower component up to 700m to be 14.1 W year $m^{-2}K^{-1}$ using OLS regression and suggests that his finding is a strong under-estimation. We may still under-estimate the total effective heat capacity of the system, mostly due to omission of the deeper ocean (>700m). Equally, other heat sinks are omitted in this simple model. If we

[14]The results should be interpreted with care given the presence of residual autocorrelation in model A.

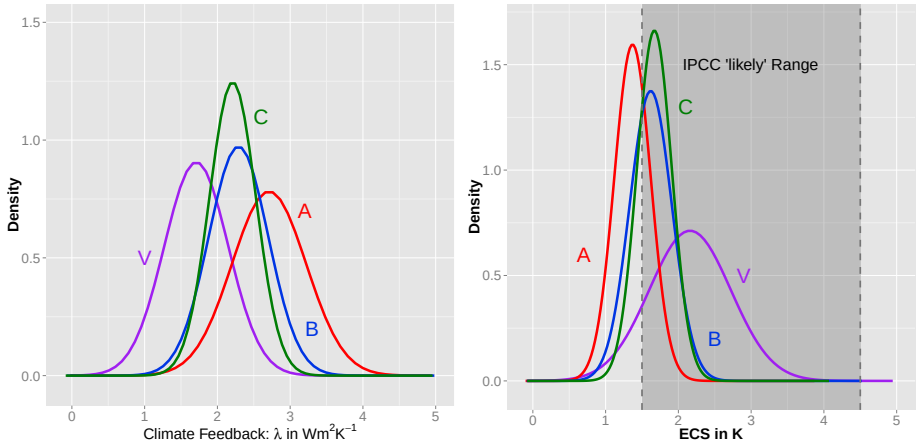


Figure 9: Estimated densities of climate feedback (λ, left) and equilibrium climate sensitivity (ECS, right), the equilibrium temperature response for a doubling of CO_2 for models A (theory, red), B (2-lags, blue), and C (2-lags & endogenous forcing, green). Model V (purple) models volcanic forcing as breaks (see appendix 2.7.5). Densities are plotted based on the asymptotic normality of the CVAR parameter estimates. The IPCC "likely" range is denoted by vertical dashed lines in the right panel.

follow Schwartz's approach of correcting the 0-700m estimate upwards (by 30% for deeper ocean, and 19% for other heat sinks), this yields an estimate of 36.34 W year $m^{-2}K^{-1}$ (using models B, C).

If we rely on a simple discrete-time approximation to map the EBM parameters to a CVAR instead, the effective heat capacity of the upper component can be estimated using estimates of the adjustment coefficients (see appendix 2.7.5).

Stability and Time Dependence

RECURSIVE ESTIMATION The CVAR specification of the EBM enables the stability of the model to be assessed over time using recursive estimation - I recursively estimate the model starting with a base sample of 20 observations from 1955-1976 onwards, expanding the sample up until 2011. This permits an assessment of how the estimates would have changed over time and can provide insight whether the recent slowdown in increase in surface temperatures (the "warming hiatus") affects the estimates of climate sensitivity. Figure 10 graphs the recursively estimated climate feedback and ECS over the sample period for the different model specifications. Consistent with the findings of Otto et al. (2013) (who use a single component EBM), the ECS and climate feedback are relatively stable over the sample prior to 2003. Notably, there is a dip in the recursively estimated system following 2003, where estimates including the time period thereafter yield a lower climate feedback and thus higher ECS with overall higher uncertainty as indicated by the $\pm$ 2 standard error interval.[15] This sudden change in 2003 is likely driven by unexpected increased ocean heat uptake during this time period (see Figure 4).

[15]The wide range of the $\pm$ 2 standard error interval for ECS in Figure 10 primarily stems from the non-linear transformation of the climate feedback estimate $\hat{\lambda}$. The estimate for λ in model B falls close to zero in 2001-2003 leading to very high values (and uncertainty) of ECS which is a non-linear transformation of λ.

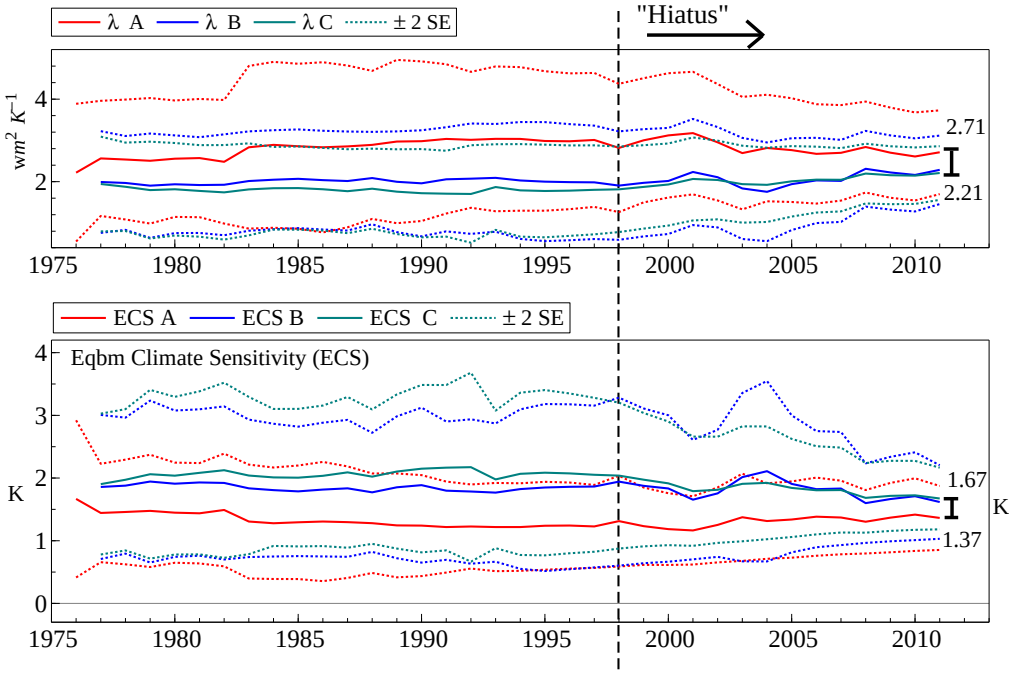


Figure 10: Recursive Estimates of Climate Feedback λ and Equilibrium Climate Sensitivity (ECS). Top panel shows the recursively estimated climate feedback λ starting with a base sample from 1956-1976 extended up to 1956-2011 for models A, B, and C. Bottom panel shows the corresponding ECS for a radiative forcing of 3.7 Wm^{-2} for a doubling of CO_2. Dashed lines graph the interval $\pm$ 2 standard errors, where standard errors for ECS are derived using the δ-method. While the $\pm$ 2 SE range is wide, the full sample 95% confidence interval is strictly positive for both ECS and λ. The range of full-sample point-estimates is denoted by black bars at the end of the sample.

The inclusion of the hiatus period (approximately from 1998 onwards), despite little increase in surface temperatures, suggests that estimates of ECS are relatively stable over time. This is likely driven by increased ocean heat uptake towards the end of the sample.

FORECASTING THROUGH THE HIATUS To further investigate the performance and the stability of the EBM I re-estimate the models up until 1997 and forecast the remaining period (2011) 5-steps (years) ahead using the model estimates. While forecast performance is not a directly useful model evaluation tool (Clements & Hendry 2005), it can provide insights into model stability. The 5-step forecasts used here rely on dynamic forecasting in a closed system – next period's forecast are used as starting points for succeeding forecasts up to five steps ahead (see e.g. Clements & Hendry 1999). Figure 11 graphs the 5-step forecasts of the temperature and ocean heat content anomalies. While the forecast confidence intervals are wide, the mean-level forecast tracks observed temperatures and ocean heat content closely. This is consistent with Kaufmann et al. (2011): changes in radiative forcing over this time period together with ocean heat uptake can (in part) account for the slow down in warming. However, it appears that ocean heat uptake is generally slightly under-predicted relative to a model estimated prior to 1997. This may provide some evidence that increased ocean heat uptake accounts for some of the "missing energy" recently debated (Trenberth & Fasullo 2010, Trenberth & Fasullo 2012).

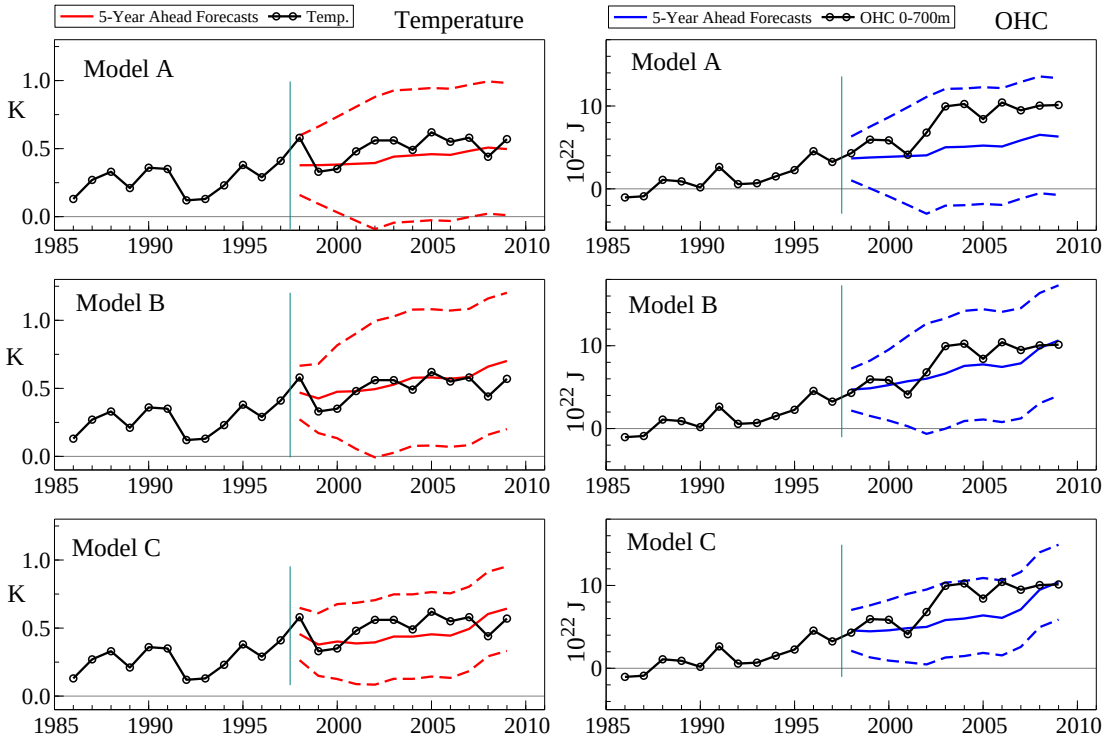


Figure 11: EBM/CVAR Model 5-Year Ahead Forecasts of Surface Temperature (left) and Ocean Heat Content (right) during the Hiatus for models A, B, and C.

2.6 CONCLUSION

I demonstrate the mathematical equivalence of a two-component energy balance model to a cointegrated system that can be mapped to a cointegrated vector autoregression. This result provides directly testable properties of an EBM when estimated on the observational record, while accounting for non-stationarity in the form of stochastic trends in the data. The equivalence result further provides a physical basis for the use of econometric CVARs in climate research, and places the entire tool-kit of CVARs at the disposal of energy balance models, ranging from tests of stability, to forecasts. A simple application of econometric estimation of a two-component EBM using global mean surface temperature anomalies, 0-700m ocean heat content anomalies, and aggregates of radiative forcing suggests that the model is not rejected. The time series form two stationary relations (cointegrate), consistent with the theoretical model and individual parameters are statistically significant with theory-consistent signs and magnitudes. CVAR estimation of a two-component EBM shows that estimates of ECS are relatively stable over the sample of observations, where the recent slowdown in warming has little effect on the estimated climate sensitivity. However, careful thought has to be given how the radiative forcing series is modelled. Using the model parameter estimates, I find that model mis-specification (e.g. apparent through residual autocorrelation) results in high values of the observationally determined climate feedback, and in turn, to low estimates of equilibrium climate sensitivity. Future work will explicitly model the socio-economic aspects determining disaggregates of the forcing series, and explore un-modelled shifts in the data.

2.7 APPENDIX

2.7.1 *Further Results on the Continuous to Discrete Mapping*

The energy balance model is given as:

$$d\mathbf{Y} = \Pi\mathbf{Y}dt + DdW \tag{35}$$

The solution to the Ornstein-Uhlenbeck process can be found using a change-of-variable approach. Define $\mathbf{Z}$ such that:

$$\mathbf{Z} = \exp(-\Pi t)\mathbf{Y} \tag{36}$$

where $\exp(-\Pi t)$ denotes the matrix exponential of $-\Pi t$, where for a $n \times n$ matrix $\mathbf{A}$ $\exp(\mathbf{A})$ is defined as:

$$\exp(\mathbf{A}) = \sum_{k=0}^{\infty} \frac{1}{k!}\mathbf{A}^k = I + \mathbf{A} + \frac{1}{2!}\mathbf{A}^2 + \dots \tag{37}$$

with the inverse of $\exp(\mathbf{A})$ given by $\exp(\mathbf{A})^{-1} = \exp(-\mathbf{A})$. Pre-multiplying (36) by $\exp(\Pi t)^{-1}$ yields:

$$\mathbf{Y} = \exp(\Pi t)\mathbf{Z} \tag{38}$$

Using (36) we can write $d\mathbf{Z}$ as:

$$\begin{aligned} d\mathbf{Z} &= -\Pi\exp(-\Pi t)\mathbf{Y}dt + \exp(-\Pi t)d\mathbf{Y} && (39)\\ &= -\Pi\exp(-\Pi t)\mathbf{Y}dt + \exp(-\Pi t)(\Pi\mathbf{Y}dt + DdW) && (40)\\ &= \exp(-\Pi t)DdW && (41) \end{aligned}$$

Integrating from 0 to t yields:

$$\mathbf{Z} = \mathbf{Z}_0 + \int_0^t \exp(-\Pi u)DdW_u \tag{42}$$

Substituting for $\mathbf{Z}$ in (38) provides the solution to the Ornstein-Uhlenbeck process as:

$$\begin{aligned} \mathbf{Y} &= \exp(\Pi t)\mathbf{Y}_0 + \exp(\Pi t)\int_0^t \exp(-\Pi u)DdW_u && (43)\\ &= \exp(\Pi t)\left(\mathbf{Y}_0 + \int_0^t \exp(-\Pi u)DdW_u\right) && (44) \end{aligned}$$

Discrete observations $\mathbf{Y}_t$ of $\mathbf{Y}$ at a frequency of one then follow a VAR process as:

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_\mathbf{t} \tag{45}$$

where $\mathbf{A} = \exp(\Pi)$, and $\boldsymbol{\epsilon}_t \sim N(0, \Sigma)$. In equilibrium correction form the VAR is written as:

$$\Delta\mathbf{Y}_t = \mathbf{P}\mathbf{Y}_t + \boldsymbol{\epsilon}_t \tag{46}$$

with

$$\mathbf{P} = \mathbf{A} - I = \exp(\Pi) - I = \exp\left(\alpha\beta'\right) - I = \alpha\kappa\beta' \tag{47}$$

where κ is a $(r \times r)$ matrix $\kappa = \left(\beta'\alpha\right)^{-1}\left[\exp\left(\beta'\alpha\right) - I\right]$. Proof:

$$\begin{aligned}
\mathbf{P} &= \exp\left(\alpha\beta'\right) - I \\
&= I + \alpha\beta' + \alpha\beta'\alpha\beta'\frac{1}{2!} + \cdots - I \\
&= \alpha\left(I + \beta'\alpha\frac{1}{2!} + \ldots\right)\beta' \\
&= \alpha\left(\left(\beta'\alpha\right)^{-1}\left[\beta'\alpha + \beta'\alpha\beta'\alpha\frac{1}{2!} + \ldots\right]\right)\beta' \\
&= \alpha\left(\left(\beta'\alpha\right)^{-1}\left[I + \beta'\alpha + \beta'\alpha\beta'\alpha\frac{1}{2!} + \cdots - I\right]\right)\beta' \\
&= \alpha\left(\left(\beta'\alpha\right)^{-1}\left[\exp\left(\beta'\alpha\right) - I\right]\right)\beta'
\end{aligned} \tag{48}$$

For additional results see Kessler & Rahbek (2004).

2.7.2 *Eigenvalues of the Companion Matrix*

To further assess dynamic stability, Table 5 reports the moduli of the eigenvalues of the companion matrix of models A and B & C for the unrestricted VAR and reduced rank estimates (r=2) without the EBM theory restrictions imposed. The indicative results of the eigenvalue analysis suggest a single unit root: one eigenvalue lies on the unit circle, all other eigenvalues are within the unit circle. There is no evidence of an explosive process and the system does not appear to be I(2) – once a reduced rank of two is imposed, all other eigenvalues are well within the unit circle.

Table 5: Eigenvalues of the companion matrix of models A and B & C for the unrestricted VAR and reduced rank (r=2) estimates (without EBM theory restrictions).

Model A (1×3 roots)						
VAR(1)	0.98	0.49	0.17			
r=2	**1.00**	0.48	0.17			
Models B & C (2×3 roots)						
VAR(2)	1.00	0.65	0.65	0.40	0.40	0.07
r=2	**1.00**	0.64	0.64	0.40	0.40	0.07

2.7.3 *Modelling Volcanic Forcing as Breaks*

Volcanic forcing $(F_{t,\text{Volc.}})$ closely resembles an impulse, where the first difference of volcanic forcing $(\Delta F_{t,\text{Volc.}})$ appears similar to a transitory shock dummy e.g. $(\ldots, 0, 0, 1, -1, 0, 0, \ldots)$ – see Figure 12 which plots the level and first difference of stratospheric aerosol (volcanic)

forcing. This break-like behaviour likely drives the rejection of normality in Model C in

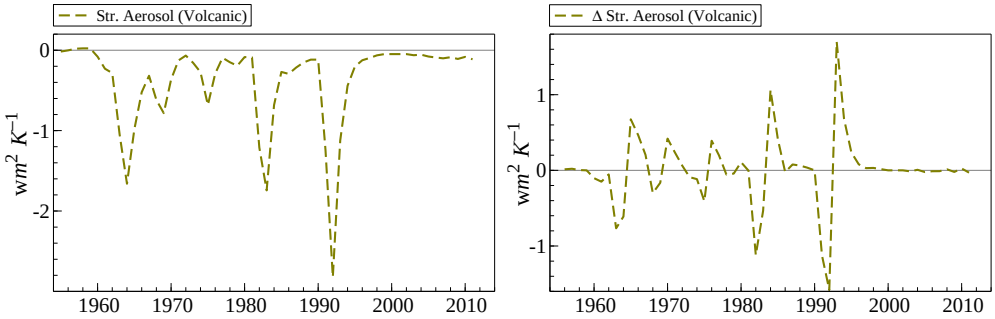


Figure 12: Radiative forcing of stratospheric aerosols from volcanic eruptions in levels, $F_{t,\text{Volc.}}$ (left) and first differences $\Delta F_{t,\text{Volc.}}$ (right)

section 2.5. There are multiple different ways in which volcanic forcing can be modelled within the CVAR specification. The aggregate forcing series F_t could be disaggregated and separated into exogenous and endogenous forcing series, alternatively eruptions could be treated as transitory shocks. Here I follow the latter approach, future work will explore the modelling of eruptions within the CVAR in more detail.

Treating volcanic eruptions as closely resembling transitory breaks, I re-estimate the CVAR EBM from section 2.5 with 2-lags, where the forcing series $\tilde{F}_t = F_t - F_{t,\text{Volc.}}$ enters the first cointegrating relation, and is itself restricted to be weakly exogenous. The first difference of volcanic forcing, $\Delta F_{t,\text{Volc.}}$ enters the model unrestrictedly, similar to a transitory shock dummy.

Table 6 provides the estimation results where model V refers to the model including the first difference of volcanic forcing unrestrictedly. The model now passes the diagnostic tests for residual autocorrelation (p=0.21) and normality (p=0.22). The EBM restrictions, including weak-exogeneity of the remaining forcing series $\tilde{F}_t$, are not rejected (p=0.17). The estimate of ECS is notably higher relative to models A-C. Figure 9 in section 2.5 plots the approximate density of ECS and λ when estimated using model V.

Table 6: EBM Cointegration Model Parameter Estimates - Unrestricted Volcanic Forcing

	EBM/CVAR Model
Coint. Relations	**V: Volc. Breaks**
Vector 1 (Mixed)	
$\beta_{1,1}$	-1.71 (0.44)**
$\beta_{1,2}$	-
$\beta_{1,3}$	1
Adjustment	
$\tilde{\alpha}_{1,1}$	0.21 (0.06)**
$\tilde{\alpha}_{2,1}$	0.09 (0.81)
$\tilde{\alpha}_{3,1}$	-
Vector 2 (Deep)	
$\beta_{2,1}$	1
$\beta_{2,2}$	-0.042 (0.004)**
Adjustment	
$\tilde{\alpha}_{1,2}$	-0.65 (0.17)**
$\tilde{\alpha}_{2,2}$	6.43 (2.04)**
$\tilde{\alpha}_{3,2}$	-
LR Test of Restric.	
$\chi^2(2)$, C: $\chi^2(1)$	6.49 [p=0.17]
Diagnostic Tests	
AR (1-2) F-Test	1.37 [p=0.21]
Normality $\chi^2(4)$, C: $\chi^2(6)$	5.75 [p=0.22]
Observations T	55
log-likelihood	84.11
EBM Estimates	
λ (W$m^{-2}K^{-1}$)	1.71 (0.44)
ECS (K)	2.16 (0.56)†
C_m (W year $m^{-2}K^{-1}$)	-
C_d (W year $m^{-2}K^{-1}$)	23.8 (2.26)†
γ	-

Model estimates based on CVAR estimation. Standard errors are given in parentheses ( ) while p-values are reported in brackets [ ]. Standard errors are provided where available. If no standard errors are reported, then parameter is restricted or derived from estimated model parameters. * indicates significance at 5%, ** indicates significance at 1%. Standard errors derived using δ-method are marked using †. Left column specifies parameters, right columns shows estimation results. Dash - marks imposed restriction and no identification in the case of the structural EBM parameters given the theoretical result of $\tilde{\alpha} = \alpha\kappa$.

2.7.4 *Residual Plots*

Figure 13 plots the scaled model residuals from models A-C (see section 2.5) and model V (see section 2.7.3). The outlying observations in 1981/1982 are associated with strong volcanic forcing from the "El Chichón" eruption. Model V where volcanic forcing is modelled as transitory breaks alleviates some of these problems.

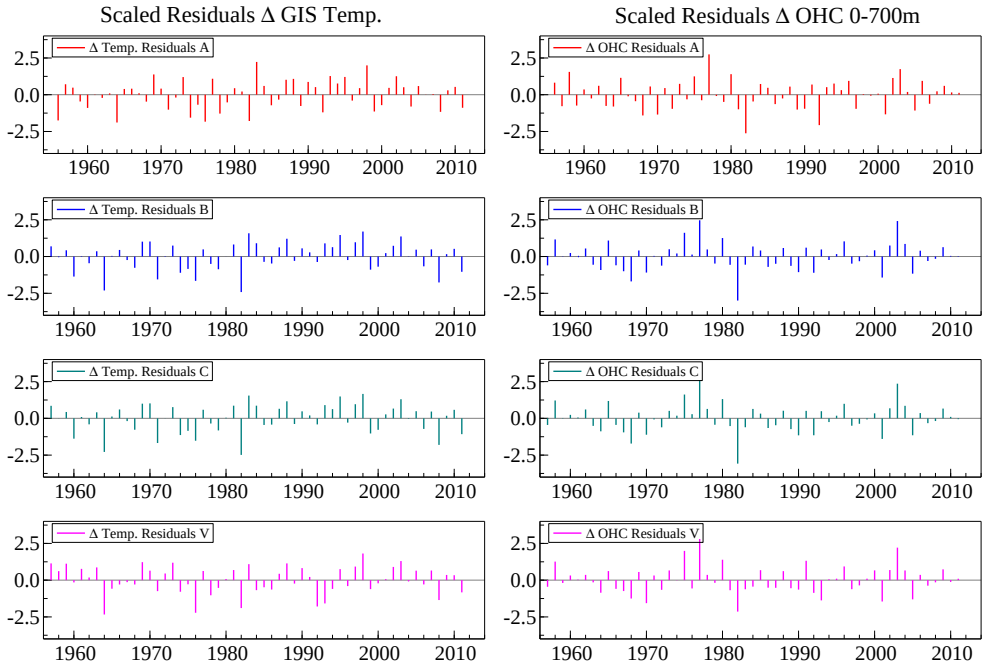


Figure 13: Scaled model residuals from models A-C (section 2.5) and model V (section 2.7.3)

2.7.5 *Discrete Time Approximation of the two-component EBM*

Section 2.4 relies on the discrete-time VAR formulation of the continuous Ornstein-Uhlenbeck process. An alternative linking of EBMs to CVARs is to rely on a simple first-order Euler discrete time approximation where $\frac{dT_m}{dt} \approx \Delta T_{m,t} = T_{m,t} - T_{m,t-1}$ and $\frac{dT_d}{dt} \approx \Delta T_{d,t} = T_{d,t} - T_{d,t-1}$. The system is then approximated as:

$$\Delta T_{m,t} = \frac{1}{C_m}\left[-\lambda T_{m,t-1} + F_{t-1}\right] - \frac{\gamma}{C_m}\left[T_{m,t-1} - T_{d,t-1}\right] + u_{1,t} \tag{49}$$

and

$$\Delta T_{d,t} = \frac{\gamma}{C_d}\left[T_{m,t-1} - T_{d,t-1}\right] + u_{2,t} \tag{50}$$

This system of equations is equivalent to a restricted cointegrated vector autoregression (CVAR) where the expressions in brackets correspond to the cointegrating vectors, and the coefficients outside the brackets are the α adjustment coefficients in the cointegration model. The error terms (here $u_{1,t}$ and $u_{2,t}$) will capture omitted effects from the model. In matrix notation this approximation is given by:

$$\begin{aligned}\begin{bmatrix}\Delta T_{m,t}\\ \Delta T_{d,t}\\ \Delta F_t\end{bmatrix} &= \begin{bmatrix}\frac{1}{C_m}\\ 0\\ 0\end{bmatrix}\cdot\begin{bmatrix}-\lambda & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}T_{m,t-1}\\ T_{d,t-1}\\ F_{t-1}\end{bmatrix}\\ &+ \begin{bmatrix}-\frac{\gamma}{C_m}\\ \frac{\gamma}{C_d}\\ 0\end{bmatrix}\cdot\begin{bmatrix}1 & -1 & 0\end{bmatrix}\cdot\begin{bmatrix}T_{m,t-1}\\ T_{d,t-1}\\ F_{t-1}\end{bmatrix} + \mathbf{u}_t\end{aligned} \tag{51}$$

where $\mathbf{u}_t = (u_{1,t}, u_{2,t}, u_{3,t})'$ and the cointegrating vectors and adjustment coefficients are linked through the approximation to the EBM such that: $\tilde{\alpha}_{1,1} = 1/C_m$, $\tilde{\alpha}_{2,1} = \tilde{\alpha}_{3,1} = 0$, $\beta_{1,1} = -\lambda, \beta_{1,2} = 0, \beta_{1,3} = 1$ further $\tilde{\alpha}_{2,2} = \gamma/C_d$, $\tilde{\alpha}_{1,2} = -\gamma/C_m$, $\tilde{\alpha}_{3,2} = 0$, $\beta_{2,1} = 1 = -\beta_{2,2}$ and $\beta_{2,3} = 0$.

Note that while the motivation linking EBMs to CVARs in this approximation is slightly different to the one provided in section 2.7.5 (a simple approximation using Euler-discrete time approach compared to the solution of the Ornstein-Uhlenbeck process), in practice the estimation procedure is near-identical. Cointegration is preserved in this CVAR formulation, testing for the rank and estimating the long-run relations given by β is exactly identical to the results in 2.7.5, therefore test results in Table 3 still apply. The main difference lies in the interpretation of the adjustment coefficients α. Given the simplifying approximation, the CVAR adjustment coefficients $\tilde{\alpha}$ have a structural interpretation where $\tilde{\alpha} \approx \alpha$.

If this discrete time approximation is considered, then the EBM implies additional restrictions on the adjustment coefficients and provides further estimates of the structural parameters. These restrictions are: lower component temperatures only adjust to the second cointegrating vector, $\tilde{\alpha}_{2,1} = 0$; the adjustment of the lower component to the second cointegrating vector $\tilde{\alpha}_{2,2}$ equals the coefficient γ determining the rate of heat transfer. This provides a cross-equation over-identifying restriction since γ is also identified in the first cointegrating relation, therefore $\tilde{\alpha}_{2,1} = -\tilde{\alpha}_{2,2}\tilde{\alpha}_{1,1} = \gamma/C_m$. Upper component heat capacity is given through $\tilde{\alpha}_{1,1} = 1/C_m$.

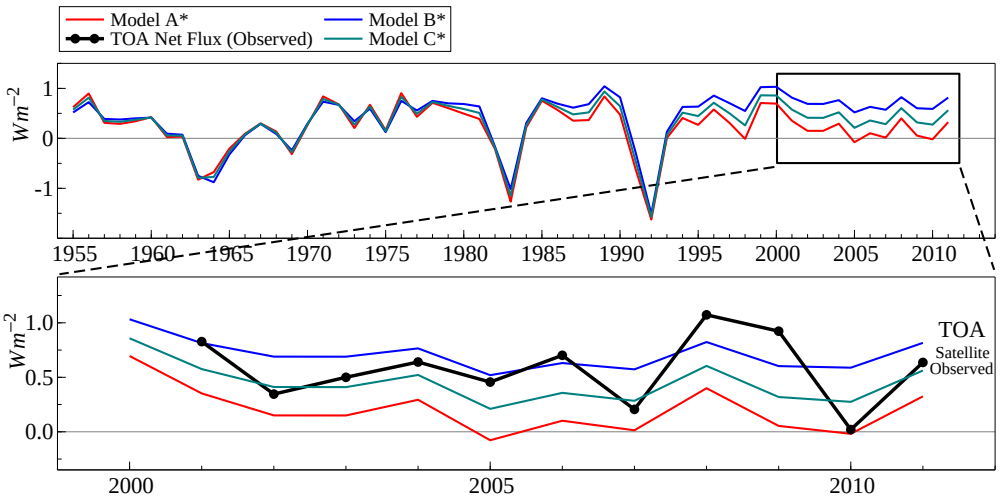


Figure 14: Appendix Model – Net heat flux based on the first cointegrating vector and observed top of the atmosphere (TOA) net flux. Top panel graphs the net heat flux as described by the estimated first cointegrating vector for models A*-C* from 1955-2011. Bottom panel graphs the net flux from 2000-2011 together with annually averaged observed TOA net flux from the CERES Energy Balanced and Filled (EBAF) all skies data (black line with points) (Loeb et al. 2009) which have not been used to estimate the model and serve as an informal test of the model against a different set of observations.

I re-estimate models A-C, and V, using the restrictions in (51), the models with additional restrictions are denoted as A*-C* and V*. Estimation results when the additional restrictions on α are imposed are provided here in Table 7 for models A* (1-lag, EBM theory), B* (2-lags), C* (2-lags, endogenous forcing), and model V* (2-lags unrestricted volcanic forcing). Figures 14 to 16 repeat the analysis from section 2.5 for the present models. The additional restrictions on α here can also be motivated by the fact that the adjustment coefficients $\tilde{\alpha}_{2,1}$ of ocean heat to the first cointegrating vector in Table 4 in section 2.5 are insignificant.

Comparing Table 4 to 7 there is little difference in estimation results of the models with additional α restrictions relative to those used in section 2.5. Estimates of ECS are generally higher with additional α restrictions imposed, with Models C* and V* estimates for ECS of 1.76 (0.29)K and 3.10 (1.72)K respectively.

Given the structural interpretation of the adjustment coefficients here, estimates for C_m can be obtained using $\tilde{\alpha}_{1,1} = 1/C_m$. Theory consistent, the effective heat capacity of the mixed component (here proxied through the use of global mean surface temperatures) is lower than that of the deeper component. The estimates of 12.19 (2.68) W year $m^{-2}K^{-1}$ for model A*, (8.33 for models B* and 11.11 for model C*) are relatively close in orders of magnitude to the lower component heat capacity. This result likely stems from the fact that the deep component in the present application does not directly represent the deep ocean, but rather only the first 700m. Temperatures for this range are therefore likely to be close to the mean surface temperature.

Table 7: Appendix Model – EBM Cointegration Model Parameter Estimates with additional α restrictions

	EBM/CVAR Model			
Coint. Relations	**A*: Base**	**B*: 2-lag**	**C*: Endo. Forc.**	**V*: Volc. Breaks**
Vector 1 (Mixed)				
$\beta_{1,1}$	-2.57 (0.94)**	-1.61 (0.77)*	-2.10 (0.35)**	-1.20 (0.67)
$\beta_{1,2}$	-	-	-	-
$\beta_{1,3}$	1	1	1	1
Adjustment				
$\alpha_{1,1}$	0.082 (0.018)**	0.12 (0.03)**	0.09 (0.02)**	0.18 (0.05)**
$\alpha_{2,1}$	-	-	-	-
$\alpha_{3,1}$	-	-	-0.55 (0.13)**	-
Vector 2 (Deep)				
$\beta_{2,1}$	1	1	1	1
$\beta_{2,2}$	-0.043 (0.005)**	-0.037 (0.005)**	-0.041 (0.004)**	-0.038 (0.005)**
Adjustment				
$\alpha_{1,2}$	-0.52	-0.65	-0.57	-0.83
$\alpha_{2,2}$	6.28 (1.24)**	5.31 (1.14)**	6.20 (1.53)**	4.67 (1.07)**
$\alpha_{3,2}$	-	-	-	-
LR Test of Restric.				
$\chi^2(4)$, C: $\chi^2(3)$	12.87 [p=0.012]*	19.45 [p=0.001]**	3.36 [p=0.34]	6.49 [p=0.17]
Diagnostic Tests				
AR (1-2) F-Test	2.82 [p=0.01]**	1.03 [p=0.42]	0.89 [p=0.59]	1.42 [p=0.20]
Normality $\chi^2(4)$, C: $\chi^2(6)$	1.25 [p=0.87]	5.57 [p=0.23]	19.34 [p=0.004]**	5.67 [p=0.22]
Observations T	56	55	55	55
log-likelihood	-74.46	-61.42	-53.36	83.43
EBM Estimates				
λ (W$m^{-2}K^{-1}$)	2.57 (0.94)	1.61 (0.77)	2.10 (0.35)	1.20 (0.67)
ECS (K)	1.44 (0.53)†	2.30 (1.10)†	1.76 (0.29)†	3.10 (1.72)†
C_m (W year $m^{-2}K^{-1}$)	12.19 (2.68)†	8.33 (2.08)†	11.11 (2.47)†	5.55 (1.54)†
C_d (W year $m^{-2}K^{-1}$)	23.25 (2.70)†	27.02 (3.65)†	24.39 (2.38)†	26.31 (3.46)†
γ	6.28 (1.24)	5.31 (1.14)	6.20 (1.53)	4.67 (1.07)

Model estimates based on CVAR estimation. Standard errors are given in parentheses ( ) while p-values are reported in brackets [ ]. Standard errors are provided where available. If no standard errors are reported, then parameter is restricted or derived from estimated model parameters. * indicates significance at 5%, ** indicates significance at 1%. Standard errors derived using δ-method are marked using †. Left column specifies parameters, right columns shows estimation results. Dash - marks imposed restriction, specifically the restrictions based on the EBM are: $\beta_{1,2} = 0$, $\beta_{1,3} = 1$, $\alpha_{2,1} = 0$, $\alpha_{3,1} = 0$, $\beta_{2,1} = 1$, $\alpha_{1,2} = -\alpha_{2,2}\alpha_{1,1}$, $\alpha_{3,2} = 0$.

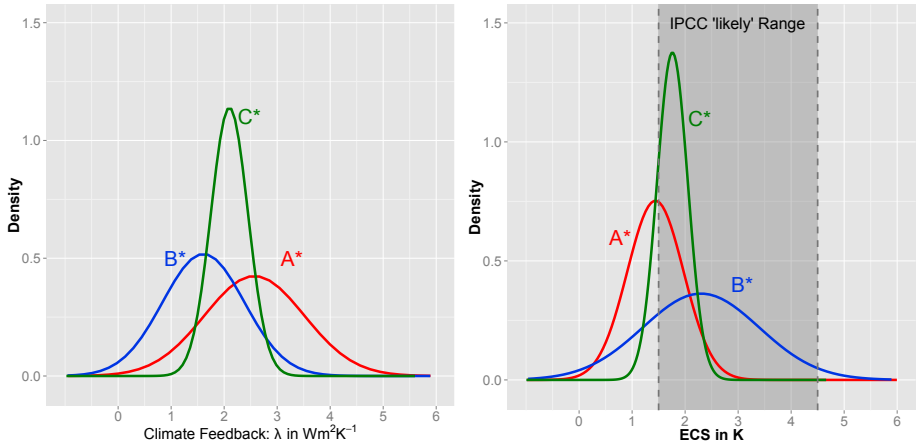


Figure 15: Appendix Model – Estimated densities of climate feedback (λ, left) and equilibrium climate sensitivity (ECS, right), the equilibrium temperature response for a doubling of CO_2 for models A* (theory, red), B* (2-lags, blue), and C* (2-lags & endogenous forcing, green). Densities are plotted based on the asymptotic normality of the CVAR parameter estimates. The IPCC "likely" range is denoted by vertical dashed lines in the right panel.

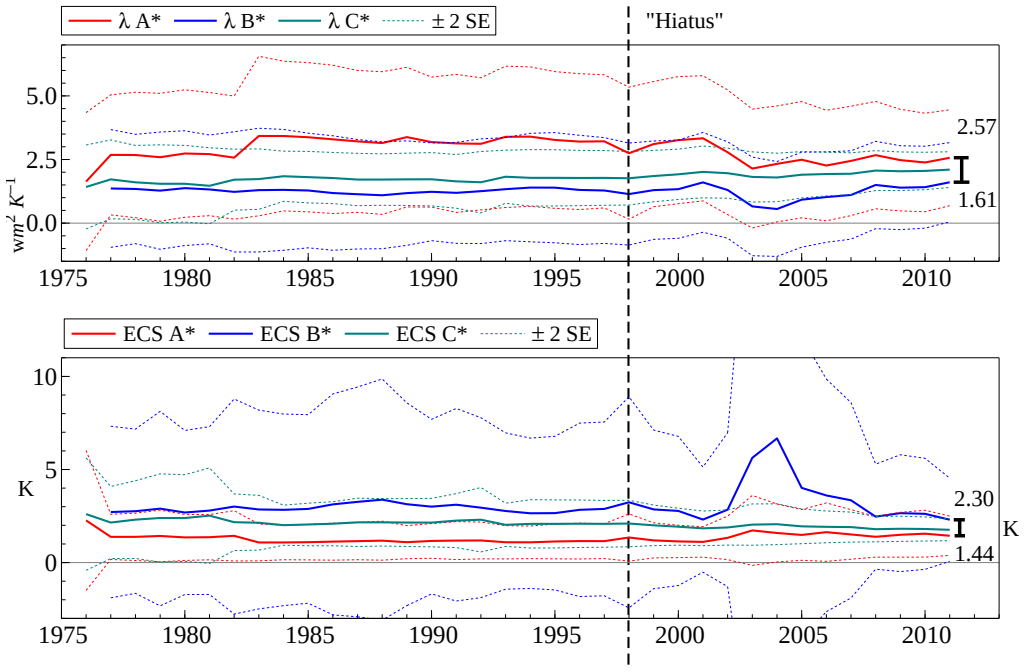


Figure 16: Appendix Model – Recursive Estimates of Climate Feedback λ and Equilibrium Climate Sensitivity (ECS). Top panel shows the recursively estimated climate feedback λ starting with a base sample from 1956-1976 extended up to 1956-2011 for models A*, B*, and C*. Bottom panel shows the corresponding ECS for a radiative forcing of 3.7 Wm^{-2} for a doubling of CO_2. Dashed lines graph the interval $\pm$ 2 standard errors, where standard errors for ECS are derived using the δ-method. While the $\pm$ 2 SE range is wide, the full sample 95% confidence interval is strictly positive for both ECS and λ. The range of full-sample point-estimates is denoted by black bars at the end of the sample.

2.7.6 *Univariate Unit Root Tests*

While the Johansen cointegration procedure does not require pre-testing for unit-roots, for completeness I provide the results of uni-variate augmented Dickey-Fuller (ADF) (Dickey & Fuller 1981) unit root tests on both levels and first differences for global mean surface temperatures, 0-700m ocean heat content, total radiative forcing, total radiative forcing excluding stratospheric aerosols (volcanic forcing) and well-mixed greenhouse gases (WMGHGs). The null hypothesis H_0 is that the series has a unit root, rejecting H_0 suggests no unit-root non-stationarity. D-lag specifies the number of lags included in the ADF test where choice of lag length is based on the lowest AIC. A constant is included in the ADF test specification. Test outcomes: ** indicates rejection of H_0 at 1% and * at 5%.

Table 8: GIS Temperature ADF Unit Root Tests

	GIS Temp.			Δ GIS Temp.	
D-Lag	**t-adf**	**AIC**	**D-Lag**	**t-adf**	**AIC**
3	-0.1707	-4.252	3	-4.329**	-4.263
2	-0.5416	-4.229	2	-6.850**	-4.291
1	-1.168	-4.161	1	-8.252**	-4.262
0	-1.949	-4.1	0	-10.26**	-4.172

Table 9: 0-700m OHC ADF Unit Root Tests

	0-700m OHC			Δ 0-700m OHC	
D-Lag	**t-adf**	**AIC**	**D-Lag**	**t-adf**	**AIC**
3	1.02	0.6989	3	-3.065*	0.6295
2	0.8619	0.6704	2	-5.301**	0.6821
1	0.3142	0.7122	1	-7.107**	0.6468
0	-0.1528	0.7377	0	-9.019**	0.675

Table 10: Total Radiative Forcing ADF Unit Root Tests

	Total Forcing			Δ Total Forcing	
D-Lag	**t-adf**	**AIC**	**D-Lag**	**t-adf**	**AIC**
3	-1.655	-1.476	3	-4.969**	-1.446
2	-1.789	-1.515	2	-5.400**	-1.457
1	-2.867	-1.454	1	-7.166**	-1.488
0	-2.299	-1.422	0	-6.182**	-1.335

Table 11: Total Forcing (excluding Stratospheric Aerosols) ADF Unit Root Tests

Total Forcing (excl. Str. Aerosols)			Δ Total Forcing (excl. Str. Aerosols)		
D-Lag	**t-adf**	**AIC**	**D-Lag**	**t-adf**	**AIC**
3	-0.5852	-6.789	3	-4.764**	-6.917
2	-0.7223	-6.763	2	-3.794**	-6.821
1	-0.7475	-6.801	1	-3.257*	-6.791
0	-0.7114	-6.43	0	-3.543*	-6.829

Table 12: Well-Mixed Greenhouse Gases (WMGHGs) ADF Unit Root Tests

WMGHGs			Δ WMGHGs		
D-Lag	**t-adf**	**AIC**	**D-Lag**	**t-adf**	**AIC**
3	-0.9117	-8.967	3	-1.793	-8.997
2	-0.6269	-8.902	2	-2.114	-8.989
1	-0.5831	-8.903	1	-2.898	-8.933
0	-0.6148	-8.619	0	-3.980**	-8.935

3

ANTHROPOGENIC INFLUENCES ON ATMOSPHERIC CO_2

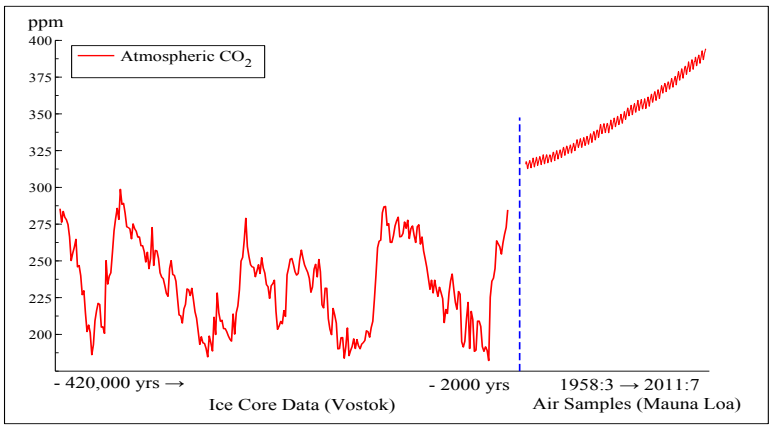
Abstract

We identify anthropogenic contributions to atmospheric CO_2 measured at Mauna Loa using the econometric automatic model selection algorithm *Autometrics* and introduce impulse indicator saturation to econometric models of climate data. We find that vegetation, temperature and other natural factors alone cannot explain the trend or the variation in CO_2 growth. Industrial production components, driven by business cycles and economic shocks, are highly significant contributors.

This chapter has been published as:

Hendry, D.F. and Pretis, F. 2013. "Anthropogenic Influences on Atmospheric CO_2" in *Handbook on Energy and Climate Change*. R. Fouquet ed. Edward Elgar.

Figure 17: Historic and present levels of CO_2



3.1 INTRODUCTION

We identify anthropogenic (human) contributions to atmospheric carbon dioxide (CO_2) measured at Mauna Loa using the econometric automatic model selection algorithm *Autometrics* and introduce impulse indicator saturation to econometric models of climate data. The present work provides an application of a relatively recent development in econometrics - selecting and then estimating models with more variables than observations. Estimating the determinants of atmospheric CO_2 is traditionally a challenge due to the complex systems of data involved. The atmospheric concentration of Carbon dioxide is a highly autocorrelated, non-stationary time series, and globally there exist a large number of potential carbon sources and sinks. There is mixed evidence in the literature on anthropogenic contributions to atmospheric CO_2: the long-term trend is widely attributed to human factors, while the main seasonal fluctuations are thought to be driven by the biosphere. However, the statistical measures applied are often somewhat unsatisfactory due to the complexities of dealing with large numbers of variables. Econometric model selection can augment the existing approaches to this problem.

Atmospheric carbon dioxide has varied over time, covering very low levels and high levels over 1000 parts per million (ppm)[1] (see Hendry 2011 for an overview). However, over the past 400,000 years, CO_2 has been drastically below the current levels (Figure 17, Petit et al., 1999, Tans & Keeling, 2011). The current strong increasing trend and the radiative forcing effect contributing to an increase in global mean temperatures provides a motivation to establish and estimate an underlying anthropogenic signature. This is of particular relevance when considering target levels of CO_2 for policy. Initial proposed target levels of 350ppm (Hansen et al. 2008) have already been exceeded. New targets aim for a stabilization around 500ppm, however, to achieve this, given the significant anthropogenic contribution (confirmed in the present study), drastic reductions in emissions will be required (see Hepburn & Stern 2008 for a discussion).

We introduce a new approach to modelling changes in atmospheric CO_2 using a model selection algorithm which allows for a larger number of potential variables than observations

[1] 1 ppm by volume of CO_2 in the atmosphere is approximately equal to 2.13 gigatons of carbon: see Clark (1982) and http://cdiac.ornl.gov/pns/convert.html.

without *a priori* forcing any to be significant or to be excluded. Using this method, the main relevant explanatory variables are determined and their magnitudes estimated while irrelevant factors are dropped from the model.

The model controls for a number of natural carbon sources and sinks: vegetation measured by the Normalized Difference Vegetation Index (NDVI); temperature (measured as anomalies in the Northern hemisphere); weather phenomena (measured through the Southern Oscillation Index); as well as accounting for dynamic transport by including seasonal interaction terms. This allows an estimate of the anthropogenic contribution to CO_2 as measured by industrial output indices and fossil fuel use for different geographical areas. Due to the high dimensional properties of this system, this initial study focuses on identification of contributing factors using a single equation approach. Endogeneity of surface temperatures is controlled using instrumental variables in line with the analysis by Kaufmann et al. (2006).

Based on this initial study, the resulting estimates describe the direct effects on CO_2 growth within the estimated conditional model and the proportional contribution of each factor. Without imposing individual variables to be included in the final model, we find that vegetation, temperature and other natural factors alone cannot explain either the trend or all the variation in CO_2 growth. Industrial production components, driven by business cycles and shocks, are highly significant contributors.

Section 3.2 provides an overview of related literature as an introduction of climate change modelling to econometrics. Section 3.3 discusses model selection, impulse indicator saturation (IIS)–which we use to detect multiple breaks in the models–and the *Autometrics* algorithm. Section 3.4 describes the data used, while section 3.5 outlines the estimation procedures, section 3.6 reports the main results and section 3.7 concludes.

3.2 LITERATURE REVIEW

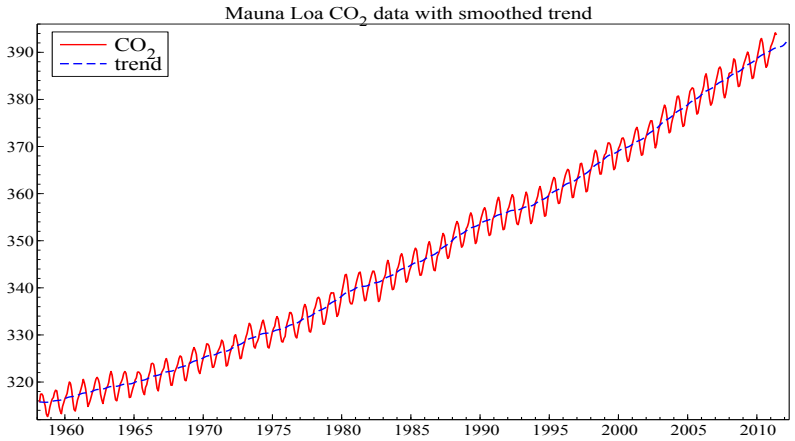
There is a plethora of literature on atmospheric CO_2 and its link to anthropogenic factors. Key aspects in the literature are finding an appropriate measure for anthropogenic activity and sufficient controls for natural effects such as vegetation and oceanic absorption. Atmospheric CO_2 has been measured consistently and regularly since 1958, mostly due to the effort of Charles Keeling who initiated and supported the measurement at Mauna Loa, Hawaii (and later other measurement stations: see Sundquist & Keeling 2009). This led to the now well established and often cited "Keeling Curve", showing the increasing trend and highly seasonal pattern in atmospheric CO_2 (Figure 18). We are primarily interested in identifying the anthropogenic contribution to atmospheric CO_2, so the following section reviews existing evidence and factors that need to be included in models.

3.2.1 *Sources and Sinks*

Economic sources

While atmospheric CO_2 has been consistently measured at multiple sites for a long time (see C. D. Keeling et al. 1976, and Sundquist & Keeling 2009), the choice of economic/anthropogenic variables is less straight forward. Three variables are regularly used: estimates of fossil fuel

Figure 18: Atmospheric CO_2 measured at Mauna Loa—the "Keeling Curve"



emissions, population, and cement production. In most cases, variables are measured on an annual basis and interpolated to monthly frequencies.

A standard measure for estimating fossil fuel emissions is the product of the amount of fuel produced, the proportion of the fuel that is oxidized, and the fuel carbon content (Marland & Rotty 1984). Variations of this data are used in Erickson et al. (2008), Jones & Cox (2005), Randerson et al. (1997), and Nevison et al. (2008). Recent data using this methodology are available at an annual frequency by Marland et al. (2011). In contrast, Hofman et al. (2009) as well as Newell & Marcus (1987) focus on population as a measure of human industrial output. Granados et al. (2011) extend the model of population by including a measure of global GDP. Cement production is a further major component of anthropogenic emissions. CO_2 emissions in production through limestone calcination, kiln operation and power generation are estimated to make up approximately five percent of anthropogenic emissions (Worrell et al. 2001).

These measures provide a good starting point to capture anthropogenic emissions in the long run, but due to the annual measurement frequency do not capture short-run fluctuations and seasonality. Population and GDP are too broad as measures to capture variability other than a trend. Measurement could be improved through supplementing overall fuel emissions by disaggregate individual fuel consumption. However, most importantly, a higher-frequency (monthly) anthropogenic measure is required to capture seasonal and short-term effects.

Vegetation and Transport

Aside from anthropogenic emissions, the terrestrial biosphere (vegetation) is a major factor in the carbon cycle's sources and sinks. Atmospheric carbon dioxide falls and rises seasonally each year due to photosynthetic activity (during Summer) and respiratory release (during Winter) of CO_2 in the terrestrial biosphere (C. D. Keeling et al. 1996, Buermann et al. 2007). The intensity of these effects depends on the length of the growing season, a fertilization effect (feedback to plant growth from increased CO_2), and shifts in seasonal patterns (Kaufmann et al. 2008).

The Normalized Difference Vegetation Index (NDVI) provides a direct measure of photosynthetic activity. Chlorophyll found in plants absorbs visible light for photosynthesis and reflects near infra-red light (Myeni et al. 1995). The more active a plant is (indicated by higher

density of green leaves), the more visible radiation is absorbed and the more near infra-red is reflected. Thus, the difference between the two measures increases with higher leaf density. Using satellite-based remote sensing the intensity of reflected visible and infra-red light can be measured. Using the ratio of the difference between the two measures, the NDVI is defined by:

$$\text{NDVI} = \frac{\rho_{NearIR} - \rho_{Visible}}{\rho_{Visible} + \rho_{NearIR}} \tag{52}$$

where $\rho_{Visible}$, and ρ_{NearIR} measure radiation of visible and near infra-red light respectively. NDVI is therefore an index ranging from -1 to $+1$, with values around 0 denoting non-vegetation objects, and values near 0.7 and above indicating dense vegetation (Tucker et al. 2010). The NDVI provides a direct measure for vegetation activity with high values close to 1 during the growing season (Summer in the Northern Hemisphere), and lower values closer to 0 during the less active season (Winter in the Northern Hemisphere). Kaufmann et al. (2008) investigate the link between NDVI and atmospheric carbon dioxide using econometric methods and find evidence that NDVI values "Granger cause" CO_2. There is also evidence of a feedback effect of increased CO_2, leading to enhanced vegetation activity.

Due to transport airflows, the primary influence of vegetation on measured atmospheric CO_2 depends on the location of the measurement station. In the case of Mauna Loa, Hawaii, the seasonal variation due to the biosphere is driven by long-range transport from Eurasia during Winter and short-range transport from North America during Summer (see Buermann et al. 2007, Taguchi et al. 2003). Narrowing down the time frame, Levin et al. (1995) suggests that Eurasian airflows dominate from October to June while North American airflows are dominant from July to September.

In terms of long-term development of terrestrial vegetation, there has been a greening trend—an increase of the growing season in the Northern Hemisphere (Lucht et al. 2002). This trend, however, was interrupted by the eruption of Mount Pinatubo in June 1991 which led to a decline in vegetation during 1992–1993. This poses the question whether volcanic influence on atmospheric CO_2 needs to be controlled for separately from the biosphere. Lucht et al. (2002) state that the main channel through which volcanoes affect atmospheric CO_2 is indirect through temperature while Hofman et al. (2009) propose that the Pinatubo eruption enhanced photosynthesis through scattered sunlight. Overall, Gerlach (2011) finds the direct effect of volcanic activity on measured atmospheric carbon dioxide in the short run over the recent observational record to be small, thus accounting for vegetation and temperature is expected to be sufficient without considering volcanoes separately. However, sudden breaks due to large eruptions will be detected through our IIS (3.3.3) procedure.

Oceanic absorption and El Niño

A large amount of carbon dioxide is transferred between the ocean and the atmosphere, where the ocean acts as both a source and a sink of atmospheric CO_2. Temperature is a key factor in oceanic CO_2 transfer, as higher sea surface temperature reduces uptake and increases outgassing (Watson et al. 1995). Increased wind speed in turn leads to higher oceanic CO_2 uptake. An additional effect is upwelling—dense cold water driven to the surface releases carbon dioxide stored in the ocean. Absorption and upwelling play opposite roles; which dominates is debated and depends on the geographical region.

The atmospheric fluctuations of air pressure differences known as Southern Oscillation affect oceanic absorption through the two channels described above. Southern Oscillation describes the change of air pressure differences between Tahiti and Darwin, Australia (see Troup 1965, Bacastow 1976, C. D. Keeling & Revelle 1985). It is measured as an index (SOI) from the Australian Bureau of Meteorology 2011, and is defined as:

$$\text{SOI} = 10 \cdot \frac{\Delta P_t - \overline{\Delta P}_t}{\sigma_{\Delta P_t}} \tag{53}$$

where ΔP_t is the difference in the average of mean sea level pressure between Tahiti and Darwin for month t. $\overline{\Delta P}_t$ is the long-run monthly average of ΔP_t and $\sigma_{\Delta P_t}$ is the long-run standard deviation of ΔP_t for the given month. Negative values of the SOI are generally referred to as El Niño years, while positive values correspond to episodes of La Niña. However, the effect on oceanic absorption is not so clear cut. Episodes of La Niña (SOI > 0) are associated with increased wind speeds thus making uptake easier. Nevertheless, increased wind also increases upwelling which leads to a release of oceanic CO_2. Bacastow et al. (1985) suggests that easier absorption outweighs upwelling during episodes of La Niña (SOI > 0) resulting in higher absorption of carbon dioxide by the ocean when the SOI is positive. In turn, this implies less absorption during El Niño (SOI < 0) years.

On the contrary, Francey et al. (1995) find the opposite—during La Niña years (SOI > 0) oceanic absorption is relatively lower because of the large upwelling effect. C. D. Keeling & Revelle (1985) side with Francey on the theoretical model that upwelling should outweigh increased absorption, but empirically find that less atmospheric CO_2 is absorbed during El Niño episodes (SOI < 0) which agrees with Bacastow's (1976) findings.

Another factor that is not often considered in the literature is carbon dioxide use by oceanic algae, as Ritschard (1992) mentions. Nevertheless, data on algae is limited as they are not covered by the NDVI satellite measures, and consequently are not considered in our study.

Looking at the bigger picture, most evidence suggests that the ocean has become a carbon sink for anthropogenic emissions (Christopher et al. 2007). However, the absolute magnitude of this effect is not known, and they estimate approximately 48 percent of fossil fuel emissions are absorbed in the ocean; and atmospheric CO_2 would be approximately 55 ppm (parts per million) higher if there were no oceanic uptake. Orr et al. (2001) similarly find the ocean to be a net carbon sink but, as well as Nevison et al. (2008), suggest that most models over-estimate the proportion of carbon dioxide emissions absorbed by the ocean.

3.2.2 *Modelling Methodology*

The models for atmospheric CO_2 can broadly be classed into two categories: global carbon cycle atmospheric transport and econometric/statistical models. Data for both is often decomposed into a long-run trend, cycle, and noise using non-parametric Fourier series or the Hodrick–Prescott (HP) filter.

Global carbon cycle models describe spatial three dimensional models with vertical levels based on solving the fundamental equations for conservation of mass, momentum, and energy. For different latitude and longitude grid resolutions, these models simulate carbon emissions and global transport (for the set-up and methodology of these models see Hansen et al. 1983,

and Kawa et al. 2004). Within this group of models, Erickson et al. (2008), Nevison et al. (2008), Randerson et al. (1997) and C. D. Keeling et al. (1995) use annual emissions data to analyse anthropogenic effects.

Statistical approaches also vary in methodology. Thoning & Tans (1989), C. D. Keeling et al. (1976) and Enting (1987) use Fourier series decompositions to study trend and seasonal cycle. Granados et al. (2011) and Granados et al. (2009) use cointegration and time-series regression to study links between population, GDP and HP-filtered CO_2 growth. However, there is no actual measure of anthropogenic emissions used in these studies. Hofman et al. (2009) use regression and graphical comparisons of carbon and population, while Jones & Cox (2005) regress growth rates of CO_2 on global emissions and cement production. Newell & Marcus (1987) look at the simple correlation between levels of carbon dioxide and global population.

The analysis of model CO_2 and observational data often suffer from similar problems based on the time series properties of the data and the potential for sudden changes or structural breaks. Correlation of time series alone is not an appropriate measure of dependence between them. The low frequency of measurement of emissions data is problematic and models are restricted by an initial choice of a small number of independent variables. Original data are rarely used: instead series are decomposed. This step is not necessary *a priori*, especially when explanatory variables that are seasonal themselves are available. Additionally, the regression analyses applied in many papers are not robust to outliers or structural breaks, do not always handle non-stationarity, and present few tests for mis-specification.

3.2.3 *Summary of the Main Findings*

Long-term trend

The long-term trend in atmospheric CO_2 is fossil fuel induced. Pre-industrial levels of CO_2 are estimated to be around 260-280 ppm (see Wigley 1983, and Hofman et al. 2009) based on ice core, tree ring, and oceanic data. Consistent and repeated measurement, starting with Keeling's work in 1958, have documented the rise in CO_2 to a current level of approximately 400 ppm measured at Mauna Loa. The rate of increase of CO_2 is proportional to combustion of fossil fuels (see C. D. Keeling 1973, C. D. Keeling et al. 1976, C. D. Keeling et al. 1995, Thoning & Tans 1989). Using population as a proxy measure for emissions yields similar results (Hofman et al. 2009, Newell & Marcus 1987, Granados et al. 2009, and Granados et al. 2011).

Seasonal variation and amplitude

Seasonal fluctuations and changes in amplitude are mainly attributed to factors in the biosphere rather than industrial emissions. There are two effects described in the literature, one is the general pattern of seasonality, the second is an increase in the amplitude of this seasonality. In particular a perceived increase in the growing season is alleged to be the driving force behind increases in amplitude.

Many studies propose that the seasonal component of atmospheric CO_2 reflects the inter-annual uptake by plants. This is supported by the fact that the amplitude of this seasonality for a given season decreases towards the equator (C. D. Keeling et al. 1976). In particular,

Enting (1987) argues that vegetation is sufficient to account for most of the inter-annual variation and that economic data does not show the required seasonality. While the peak to trough ratio measured at Mauna Loa was approximately 0.8 for the time period Enting investigates, he suggests that industrial emissions are not sufficient to cause this seasonal change. However, as is obvious from many economic time series, there is high seasonality in production and therefore in emissions.

The amplitude of the seasonal effect has been increasing over time. C. D. Keeling et al. (1996), Randerson et al. (1997), Kohlmaier et al. (1989), and Bacastow et al. (1985) characterise the increase as a result of a lengthening growing season with only a very small effect directly from fossil fuel emissions. The effect from anthropogenic emissions in these studies ranges from 0.01 to 0.2 percent on the change of amplitude. Additionally, C. D. Keeling et al. (1995) find that changes in the overall growth rate of CO_2 are driven by changes in vegetation and temperature rather than changes in industrial emissions.

A major issue with many of the above mentioned studies is that anthropogenic emissions and production data are measured annually and therefore do not have the required frequency to be able to account for seasonal fluctuations. In a recent paper, Erickson et al. (2008) investigate this issue and find that economic data would suggest the highest anthropogenic fluxes occur at the same time as the respiration phase of plants (Winter in the Northern Hemisphere). Once models account for seasonality in fuel consumption this will then lead to a diminished effect of seasonality from the biosphere.

3.2.4 *Contribution of this Study*

There are recurring challenges in existing studies of anthropogenic contributions to CO_2. Climate and atmospheric carbon fluxes are complex systems, nevertheless, many of the models are restricted by *a priori* selections of explanatory variables. The data used to account for anthropogenic emissions are often measured at too low a frequency to capture any seasonality. The main series of carbon dioxide is often decomposed into cycles and trends, something that is not necessary if the explanatory data is measured at a reasonable frequency. On the one hand, a significant number of the papers that approach the problem from a statistics or economic point of view do not sufficiently control for the biosphere or other natural factors. On the other hand, many models coming from a natural science background use statistical methods that are ill fitted given the time-series characteristics of the data. Modern econometric methods can make an interdisciplinary contribution to these problems.

To address these issues, we introduce an extended General to Specific (GETS) modelling approach based on automatic model selection, encompassing and the theory of reduction. This allows for a large number of candidate explanatory variables, in particular, models can be selected with more explanatory variables than observations. It is therefore possible to include many lags to capture time dynamics as well as a wide range of controls for natural factors and industrial output measures. As our main measure of anthropogenic productivity and emissions is industrial production (measured at monthly intervals), the data is analysed as a whole without requiring prior decompositions into trends and cycles. Models are also not restricted to a tight *a priori* selection of variables. Using IIS, the methods are robust to outliers and structural breaks, handle unit roots reasonably well, and provide a straight-forward method of testing for mis-specification.

Overall, the literature indicates a clear necessity to control for the biosphere, temperature, El Niño effects, and long (as well as short-run) anthropogenic measures. Following the idea of encompassing, we try to include variables normally included in econometric analyses of this topic as well as natural components. Intuitively, our approach is to utilise a large number of potential determinants controlling for the above mentioned factors, and then use automatic model selection techniques to determine which forces are significant. Starting with a theory-based, but very broad General Unrestricted Model (GUM), the initial system is reduced to a specific model. This is a comparatively agnostic and data-driven approach that imposes few restrictions on explanatory variables while being robust to sudden shifts (structural breaks). Naturally this approach is a simplification by relying on global averages and using a single equation approach can only mimic a basic model of atmospheric carbon uptake.

3.3 METHODOLOGY

The carbon cycle, with its many potential sinks and sources is a complex system which makes it near-impossible to correctly specify an appropriate model *a priori*. We thus require a methodology that can handle more variables (N) than observations (T) and is based on model selection. The literature for $N > T$ is relatively new and can be split into two categories:[2] general to specific (GETS) modelling, and models based on penalised shrinkage estimators. GETS (which is outlined in more detail below) relies on the theory of reduction in which the basic principle is to reduce a very general model to a specific one (see Campos et al. 2005).

The alternative are models based on penalised shrinkage estimators in the sparse modelling literature (see Tibshirani 1996, and Stodden 2006). A model is said to be sparse, if for a large number of variables, the associated coefficient vector only has few non-zero rows. Penalised shrinkage estimators such as the LASSO (Least absolute shrinkage and selection operator, Tibshirani 1996), least-angle-regression (LARS, Stodden 2006) or Stagewise Orthogonal Matching Pursuit (StOMP, Donoho et al. 2006) therefore rely on a penalisation parameter on coefficients. The models are estimated subject to a constraint that some of the entries in the coefficient vector are smaller than a penalisation parameter. While many theoretical results exist for this literature, in practice LASSO and LARS are closely related to forward stepwise selection (Stodden 2006) in which variables are selected after each other, starting with the highest correlated ones. This approach suffers from well known problems associated with forward stepwise regression, in particular these methods often fail in selection when negative correlation between regressors is present (see Castle et al. 2011a). It is for these reasons that this thesis will rely on general to specific modelling to model complex equations.

The unknown data generating process (DGP) is the underlying structure that creates the data. Empirical modelling will always deal with a subset of variables of the DGP, thus an important factor is the local data generating process (LDGP)—the generating process in the space of the variables under analysis: see Hendry (2009). The approach, therefore, is to construct a set of data based on broad theoretical assumptions, which nests the LDGP, then within this set, reduce the model from its general form down to a specific representation. This is a two step procedure. Step one: define a set of N variables that include the LDGP

[2] In the field of machine learning there also exist other approaches dealing with $N > T$, such as Random Forests (Siroky 2009), though these are not directly applicable to regression analysis which is the focus here.

as a sub-model. Step two: starting with that general model as a good approximation of the overall properties of the data, reduce its complexity by removing insignificant variables, while checking that at each reduction the validity of the model is preserved. This is the basic framework of GETS modelling.

This section introduces theoretical concepts of model selection, their use in mis-specification testing, followed by the introduction of impulse indicator saturation (IIS) and its generalized version. All these concepts are then united and applied through the automatic search algorithm *Autometrics* (see Doornik 2009a). The algorithm combines these features through automated selection based on GETS while handling more variables than observations with IIS for detecting breaks and outliers, and mis-specification testing.

3.3.1 *Model Selection in General to Specific Modelling*

GETS modelling is based on the theory of reduction which attempts to explain the discovery of econometric models based on the unobserved underlying DGP (see Hendry 1995, Ch. 9). The theory describes the process of moving from the unobserved DGP to the local DGP (LDGP), which is a simplified admissible version of the underlying DGP that can be modelled using the observed variables $\mathbf{y}_t, \mathbf{x}_t$. The choice of variables $\mathbf{y}_t, \mathbf{x}_t$ will define the properties of the LDGP to be modelled. The primary aim is to achieve a final model that is congruent – defined as matching the empirical properties of the LDGP. From the theory of reduction (Hendry 1995, p.365), there are five criteria that a final model should fulfil to be congruent: 1) Homoskedastic innovation errors – for the purpose of the present analysis this is tested post-selection. 2) Weakly exogenous conditioning variables for the estimated parameters of interest – this is currently taken as an assumption here, discussed in section 3.5 and indirectly tested for in section 3.6.3. 3) Constant invariant parameters – tested here for sample constancy using the Chow (1960) test. 4) Theory consistent estimates – these are evaluated post-selection. 5) Data admissible formulations on accurate observations – impulse indicator saturation (see section 3.3.3) is used to check for outliers. Additional to the steps mentioned here, congruence is evaluated by a wide range of further diagnostic and misspecification tests and through the interpretation of selected indicators. Further, to be un-dominated, a model should encompass all other valid competing sub-models (see e.g., Mizon & Richard 1986, Hendry & Richard 1989, and Bontemps & Mizon 2008). The final aim for a congruent model is to be encompassing of other models, that is to say able to explain outcomes of other valid models within its own framework. Here we attempt to encompass sub-models by covering a large system of potential variables that nest other models in the literature.

3.3.2 *Misspecification Testing*

Using more variables than observations, in particular IIS (further outlined in the next section) allows for a new approach to control for model misspecification. By not being restricted in the number of variables to include initially, the idea is to be agnostic about the first formulation and use model reduction steps to achieve a well specified model. Impulse indicators can be an especially useful source of information on misspecification – a large number of outliers (selected through indicators) provide evidence for potential functional form misspecification. Nevertheless, familiar diagnostic tests are still required to achieve a valid statistical model,

where valid refers to a model corresponding to the statistical assumptions that are being made (see Castle & Hendry 2011). These diagnostics include tests for heteroskedasticity (non-constant variance) of the error terms, non-normality, non-constancy of parameters, non-independence of errors, conditioning variables being independent of the errors and non-linearities in functional form. All these can easily be tested during selection when $N < T$, but when the number of variables exceeds the number of observations these tests will primarily be applied on the terminal models. These models are the relevant models of interest on which these properties should be tested, rather than models on sub-sets of variables that are encountered during selection. These tests fall into the category of tests of congruence; which exact tests are used in the present application is outlined in section 3.5.

3.3.3 *More Variables than Observations: $N > T$*

The model selection approach introduced here allows for more variables than observations to be used in modelling ($N > T$). For *Autometrics* this was first introduced through impulse indicator saturation, and has recently been extended to the general case.

Impulse indicator saturation

Empirical modelling often faces a large number of potential unknown unknowns. For a given series of data, there might be an unknown number of location shifts for unknown durations for unknowingly omitted variables. Therefore, an agnostic approach to detecting these breaks and outliers should be applied. Impulse indicator saturation (IIS) was developed to solve this problem. IIS creates a zero/one indicator variable for every observation in the sample, such that for T observations there are T variables that correspond to $1_{j=t}$ indicators. The variables are equal to one when $j = t$ for each $j = 1...T$ and zero otherwise. By definition (and since an intercept is always included), this leads to the situation of facing more variables than observations, something that could not be handled in the past. The postulated solution to this problem is to split the sample of indicators into blocks and estimate the model on these partitions of variables. Consider the basic case of estimating the mean of a sample using a split-half approach proposed in Hendry et al. (2008). In step one, only the first $T/2$ indicators are included, this is equivalent to dummying out the first part of the sample and estimating the mean on half of the sample. Indicators that deviate significantly from this estimated mean will be retained. This step is repeated for the other half of indicators, similarly significant indicators are retained. In the final step all previously retained indicators are included (which will generally lead to $N < T$), and the mean is estimated on the full sample while simultaneously selecting significant indicators. Figure 19 (adapted from Hendry & Santos 2010) shows these steps in the general case of estimating the mean under the null hypothesis of no breaks for T=100 for a process $y_t \sim$ iid $\mathrm{N}(10, 1)$. Under the null hypothesis, for a significance level of selection α, the expected rate of retention for irrelevant indicators is αT (Hendry & Mizon 2011). Here just one indicator is retained, as is to be expected for $\alpha = 0.01$.

Distributional properties of IIS are analysed by Hendry et al. (2008) for estimating a sample mean under the null-hypothesis of no breaks or outliers and generalised to more than two blocks. Johansen & Nielsen (2009) extend this analysis for the inclusion of regressors for both

Figure 19: Impulse indicator saturation under the null of no break

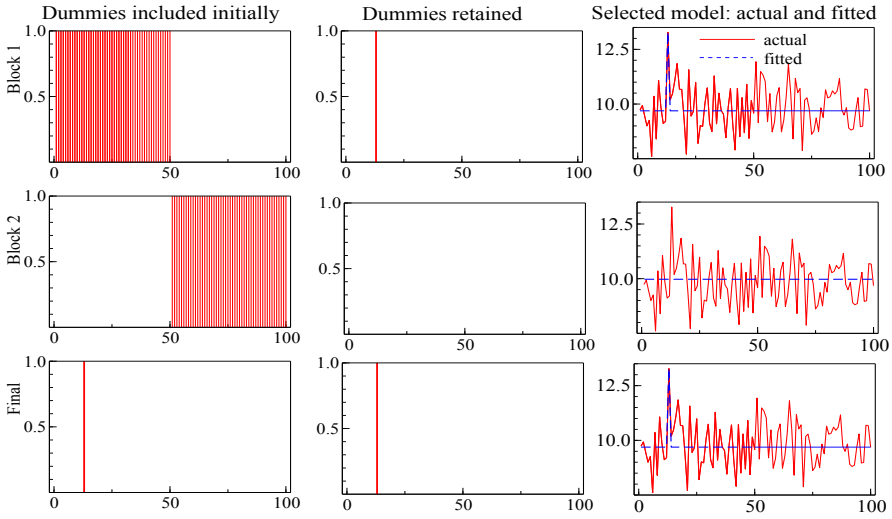
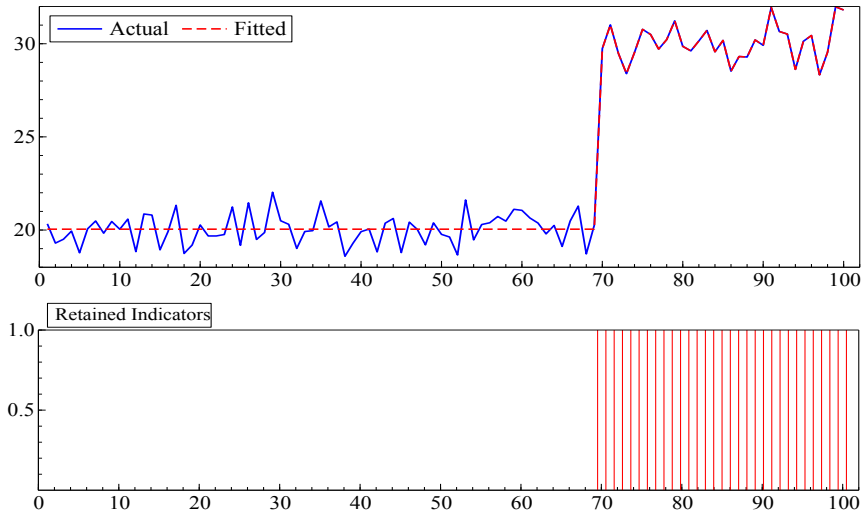


Figure 20: Impulse indicator saturation under a single location step-shift



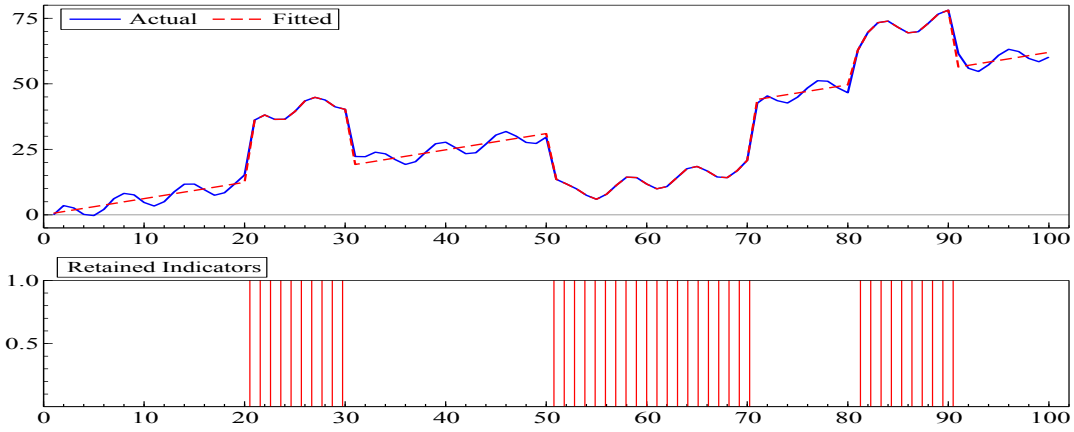
stationary and unit-root autoregressions, similarly under the null hypothesis of no breaks in the DGP.

To see IIS operating in practice when faced with a one time location shift, consider estimating the mean under a one time break for observations from $t = z...T$. Let the DGP be given as:

$$y_t = \mu + \lambda 1_{\{t>z\}} + \epsilon_t \tag{54}$$

Figure 20 shows a series generated according to (54) with $\mu = 20$, $z = 70$, $\epsilon \sim$ iid $\mathrm{N}(0, 1)$ and a large location shift for illustration purposes: $\lambda = 10\sigma_\epsilon$. At $\alpha = 0.01$ the break is correctly identified as the retained indicators show. Using IIS this break is identified without prior knowledge of the magnitude, timing or duration of it taking place. One might argue that the break in Figure 20 is still relatively easy to spot simply by data inspection alone, however;

Figure 21: IIS outcomes for six location shifts in data with seasonal cycles



IIS remains successful even in more complex scenarios. Consider an autoregressive, trending DGP (55) with a seasonal cycle not unlike the series of atmospheric CO_2:

$$y_t = \rho y_{t-1} + \beta_1 t + \beta_2 \sin(t) + \sum_{i=1}^{b} \lambda_i 1_{\{z_{i,1} > t > z_{i,2}\}} + \epsilon_t \tag{55}$$

where b is the number of location shifts and $\sin(t)$ simulates a simple seasonal cycle. This series is shown in Figure 21 for $\rho = 0.2$, $\beta_1 = 0.5$, $\beta_2 = 3$ and six breaks, $b = 6$, of varying durations at $t = 20, 30, 50, 70, 80$ and 90 for λ ranging from $10\sigma_\epsilon$ to $20\sigma_\epsilon$. As Figure 21 shows, IIS successfully identifies all six breaks even in the presence of a trend and seasonal component. These large breaks here are shown for illustration purposes, IIS has been shown to detect breaks of much smaller magnitude with high power (Castle et al. 2011c). IIS may appear surprising, though some commonly used econometric techniques are variants of the general IIS approach. Many techniques that "dummy out" observations are related. Recursive estimation of regression coefficients is identical to IIS over future samples with the number of indicators reduced in each recursion. Similarly, the Chow (1960) test for parameter constancy can be seen as IIS over sub-samples of the data for $T - k + 1$ to T (without selection) as Salkever (1976) showed. There are two clear purposes of IIS. First, as the examples above demonstrated, IIS is successful in identifying location shifts, breaks, and outliers. Addressing the success of IIS in identifying breaks, Hendry & Santos (2010) demonstrate that the key factors influencing success of IIS for a single location shift are the magnitude of the shift, the duration of the break, selection significance level α and error variance σ_ϵ. Castle et al. (2011c) extend this to multiple breaks and shifts within the sample and compensating for non-normality. These results are intuitive, as the larger a break and the higher the signal to noise ratio in the DGP is, the easier it will be to identify structural changes.

A second purpose of IIS is that, as mentioned in section 3.3.2, it can be a good indicator of model misspecification. A large number of retained dummies is indicative of model misspecification in the sense that a large fraction of the data cannot be explained by the estimated model, which could apply to non-linearities, omitted variables or other forms of misspecification.

So far IIS has focused on single impulses through indicator variables. This can be extended to a more general approach (see Chapter 4), so rather than just focusing on single breaks,

there could be step-shifts or breaks in coefficients where each independent variable $x_{i,t}$ is interacted with a time indicator dummy $1_{\{j=t\}}$ for each $j = 1...T$. (e.g. N. R. Ericsson 2011, Doornik et al. 2013, Castle et al. 2015).

General case of $N > T$

A natural extension of IIS is to move from allowing more variables than observations with indicators to the general case of all forms of independent variables (see Hendry & Krolzig 2005 and Hendry & Johansen 2011). Suppose there are N total regressors partitioned into J blocks of n_j, where $N = \sum_{j=1}^{J} n_j$ such that $N > T$ and $n_j < T$ for all j. Consequently the total number of variables N exceeds the number of observations T but total variables can be partitioned into J blocks n_j each smaller than T. Their approach suggests randomly partitioning the set of variables into blocks of n_j, applying GETS to each block retaining the selected variables and crossing the groups to mix variables. The next step is to use the union of selected variables from each block to form a new initial model and repeat the process until the final union of selected variables is sufficiently small. *Autometrics* implements a variant of this algorithm to handle the general case of $N > T$, with a key addition being an expansion step to search for omitted variables.

3.3.4 *Autometrics*

Autometrics (see Doornik 2009a) is the latest installment in the automated GETS methodology and is available in the *OxMetrics* (Doornik 2010b) software package. The algorithm is based on the following main components:

1. GUM: The general unrestricted model (GUM) is the starting point of the search. The GUM should be specified based on broad theoretical considerations to nest the LGDP.

2. Pre-Search: prior to specific selection, a pre-search lag reduction is implemented to remove insignificant lags, speeding up selection procedures and reducing the fraction of irrelevant variables selected (denoted the gauge of the selection process). Pre-search is only applied if the number of variables does not exceed the number of observations ($N < T$).

3. Search Paths: *Autometrics* uses a tree search to explore paths. Starting from the GUM, *Autometrics* removes the least significant variable as determined by the lowest absolute t-ratio. Each removal constitutes one branch of the tree. For every reduction, there is a unique sub-tree which is then followed; each removal is back-tested against the initial GUM using likelihood ratio tests (equivalent to an F-test). If back-testing fails, no sub-nodes of this branch are considered (though different variants of this removal exist). Branches are followed until no further variable can be removed at the pre-specified level of significance α. If no further variable can be removed, the model is considered to be terminal.

4. Diagnostic Testing: each terminal model is subjected to a range of diagnostic tests based on a separately chosen level of significance. These tests include tests for normality (based on skewness and kurtosis), heteroskedasticity (for constant variance using squares), the Chow test (for parameter constancy in different samples), and residual autocorrelation

and autoregressive conditional heteroskedasticity. Parsimonious encompassing of the feasible general model by sub-models both ensures no significant loss of information during reductions, and maintains the null retention frequency of *Autometrics* close to α: see Doornik (2008). Both congruence and encompassing are checked by *Autometrics* when each terminal model is reached after path searches, and it backtracks to find a valid less reduced earlier model on that path if any test fails. This repeated re-use of the original mis-specification tests as diagnostic checks on the validity of reductions does not affect their distributions (see Hendry & Krolzig 2003).

5. Tiebreaker: as a result of the tree search, multiple valid terminal models can be found. The union of these terminal models is referred to as the terminal GUM. As a tiebreaker to select a unique model, the likelihood-based Schwarz (1978) information criterion (SIC) is used, though other methods are also applicable, and terminal models should be considered individually.

In simulation experiments, models are primarily evaluated based on three concepts: gauge, potency and the magnitudes of the estimated parameters' unconditional and conditional (on selection) root mean-square errors (RMSEs) around the DGP values (see Doornik & Hendry 2009). Potency, the model selection equivalent of power, is defined as the proportion of relevant variables (featured in the DGP) selected in the final model. For a given (linear) DGP, let $N+1$ be the total number of potential explanatory variables, of which the first q are relevant ($\beta_1, ...\beta_q \neq 0$), and $N+1-q$ are irrelevant ($\beta_{q+1}, \beta_{q+2}, ...\beta_N = 0$). Let $\hat{\beta}_{k,i}$ denote the coefficient of the selected variable k in replication i. For M replications, $i = 1, 2, ..., M$, potency is thus defined as:

$$\text{Potency} = \frac{1}{q}\frac{1}{M}\sum_{k=1}^{q}\sum_{i=1}^{M} 1_{\{\hat{\beta}_{k,i}\neq 0\}} \tag{56}$$

Gauge, the model selection equivalent of size (also referred to as false-positive rate), measures the fraction of retained irrelevant variables:

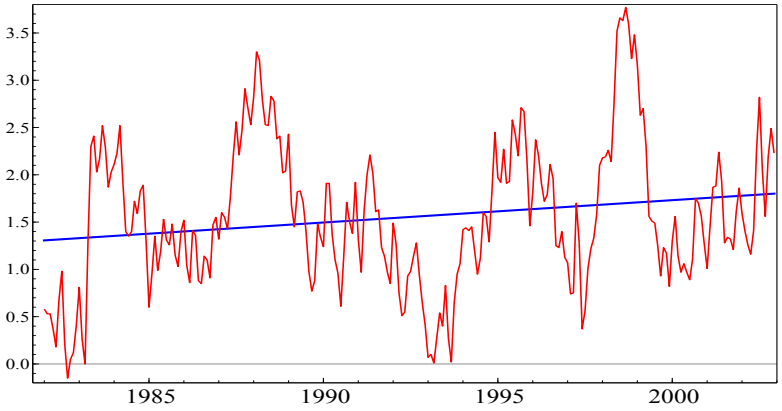
$$\text{Gauge} = \frac{1}{N+1-q}\frac{1}{M}\sum_{k=q+1}^{N}\sum_{i=1}^{M} 1_{\{\hat{\beta}_{k,i}\neq 0\}} \tag{57}$$

Low gauge (close to zero) and high potency (close to 1) are preferred, as are small RMSEs (see Castle et al. 2011a).

The main calibration decision in the search algorithm is the choice of significance level α at which selection occurs. Selection continues until retained variables are significant at α, though it can be the case that variables in the final model are also retained at a level above α if removal leads to diagnostic tests failing. α is approximately equal to the gauge of selection. Further, the choice of diagnostic tests and lag length selection for residual autocorrelation and autoregressive conditional heteroskedasticity need to be set.

In the general case of $N > T$ and IIS, *Autometrics* groups variables into two categories: selected and not selected (Doornik 2010a). Not currently selected variables are split into sub-blocks and the algorithm proceeds by alternating between two steps: first, the expansion step, selection is run over not-selected sub-blocks to detect omitted variables. Second, the reduction step, a new selected set is found by running selection on the system augmented with the omitted variables found in step one. This is repeated until the dimensions of the terminal

Figure 22: Annual changes in atmospheric CO_2



model are small enough and the algorithm converges, so the final model is unchanged by further searches for omitted variables.

Autometrics has been applied successfully in a range of fields: see, for example, Hendry & Mizon (2011) on US food expenditure, Bårdsen et al. (2010) on unemployment in Australia, and Castle et al. (2011a) for a comparison with other selection methods.

Nevertheless overall selections should be interpreted carefully. Successful identification of the underlying LGDP can be adversely affected by collinearity of the independent variables. Most simulations of *Autometrics* with large numbers of variables use orthogonal regressors, which makes selection easier. Furthermore, when $N > T$, in the block selection algorithm of *Autometrics*, adding or dropping a variable from the initial GUM may change the block partitioning of variables, so the selection is not invariant to the initial specification.

The next section covers the data used to construct the GUM in an attempt to nest the LGDP for atmospheric carbon dioxide. In the section following, *Autometrics* is then used to determine the anthropogenic contributions to CO_2.

3.4 DATA

3.4.1 CO_2

The atmospheric CO_2 data used here are taken from Keeling's measurements at Mauna Loa, available from Tans & Keeling (2011) (Scripps Institution of Oceanography). The time series of carbon dioxide in monthly averages runs from 1958:3 until 2011:7 at the time of writing. Simple inspection of the data shows that both the level (see Figure 18) and the annual change (see Figure 22) are increasing over time.[3]

The seasonal fluctuations are apparent in the data and as Buermann et al. (2007) poetically describe it, the regular seasonal cycle of CO_2 at Mauna Loa "records the breathing of the Northern Hemisphere biosphere". We empirically model atmospheric CO_2 based on a simplified version of the carbon cycle model used in Allen et al. (2009). While the Allen et al. (2009) model consists of three carbon pools with varying time scales, we abstract from this

[3] The reason annual changes of CO_2 and following graphs are shown from 1982–2002 is that due to data availability the estimation sample is limited to this time period, as section 3.4.2 describes.

and model atmospheric CO_2 as a single pool where the equilibrium level of CO_2 increases with cumulative emissions:

$$\frac{d\text{CO}_2}{dt} = \mathbf{E}_t'\boldsymbol{\beta} \tag{58}$$

where $\mathbf{E}_t$ is a vector of sources and sinks which, weighted by $\boldsymbol{\beta}$, denotes net emissions per year. The simplification of three to one carbon pool implies that all sources and sinks are modelled to adjust on the same time scale. While this is a highly simplified version of the Allen et al. (2009) model, it does account for the long-atmospheric lifetime of the gas (see Myhre et al. 2013) by modelling CO_2 as a stock variable driven by cumulative emissions. In contrast to Allen et al (2009) the model for this study is estimated, rather than calibrated. An alternative approach is followed by Kaufmann et al. (2006) who model CO_2 in levels rather than first-differences. Approximating the relationship in (58) in discrete time and assuming a stochastic process due to omitted variables and measurement errors yields:

$$\frac{d\text{CO}_2}{dt} \approx \Delta\text{CO}_{2,t} = \mathbf{E}_t'\boldsymbol{\beta} + \epsilon_t \tag{59}$$

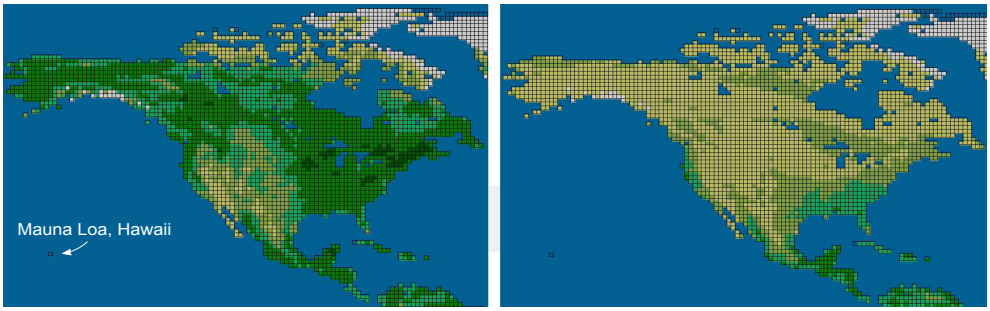
The dependent variable modeled is therefore $\Delta\text{CO}_{2,t}$. Equations (58)–(59) suggest a potential linear relationship between the change in atmospheric carbon dioxide and netflow. This is consistent with Scheffer et al. (2006) who argue that while the radiative forcing effect of CO_2 on temperatures is approximately logarithmic, the feedback effect of temperatures on CO_2 modelled here as one element in $\mathbf{E_t}$ has an approximate linear relationship. Tests for non-linearity are considered in the results section. The following section identifies variables that make up the netflow $\mathbf{E}_t$, both anthropogenic as well as natural sources and sinks.

3.4.2 *Vegetation*

We use NDVI data to account for vegetation effects on carbon dioxide. Data are available for the NDVI from the Oak Ridge National Laboratory Distributed Archive Center (see Tucker et al. 2010) ranging from 1981:7 until 2002:12 at spatial resolutions of 0.25, 0.5 and 1.0 degrees latitude and longitude. CO_2 measured at Mauna Loa is driven by North American airflow during the Summer and airflow from Eurasia during the Winter. Therefore, the NDVI data is split into two main regions: North America and Eurasia. Figure 23 shows the NDVI measure in detail for North America in August and January respectively. Using 1° spatial resolution an algorithm then takes the average of every 3×3 observation grid on land within the two regions (excluding water, permanent ice and missing observations). To capture the main variation of vegetation the North American region is defined by the rectangle ranging from 86°N/167°W to 14°N/48°W, and the Eurasian region by the rectangles ranging from 75°N/9°E to 36°N/51°E and 76°N/52°E to 7°N/358°E. This generates 198 time series variables for North America and 567 for Eurasia.

Due to the nature of a common growing season in the Northern Hemisphere, the generated time series are highly collinear. Principal components (PCs) are used to reduce the number of variables, while retaining most of the variation in the data. For p random variables $Y_1, ...Y_p$ define $\mathbf{Y} = (Y_1 - \overline{Y_1}, ...Y_p - \overline{Y_p})$ and let C_j denote the jth principal component, where $j = 1, ...p$. The component C_j will be a weighted sum of variables $Y_1 - \overline{Y_1}, ...Y_p - \overline{Y_p}$.

Figure 23: NDVI at 1° North America: August (left) and January (right) 1981



$$C_j = \mathbf{w}^{(j)'}\mathbf{Y} = w_1^{(j)}(Y_1 - \overline{Y_1}) + ... + w_p^{(j)}(Y_p - \overline{Y_p}) \tag{60}$$

The weights $\mathbf{w}^{(j)} = (w_1^{(j)}, ..., w_p^{(j)})'$ are chosen to maximize the variance of the associated component under the constraint that the component is orthogonal to the previous components and that $w^{(j)}$ has unit length:

$$\mathbf{w}^{(j)} = \arg\max \text{Var}(\mathbf{w}^{(j)'}\mathbf{Y}) \tag{61}$$

$$\text{s.t. } \mathbf{w}^{(j-1)'}\mathbf{w}^{(j)} = 0 \text{ for } j = 2, ...p$$

$$\text{s.t. } \mathbf{w}^{(j)'}\mathbf{w}^{(j)} = 1$$

By construction the first component has the highest variance and captures the largest fraction of variation in $\mathbf{Y}$. Although PCs are just linear transformations of the original times series, they have three potential advantages. First, PCs are mutually orthogonal, so adding or eliminating any one PC has little effect on the coefficient estimates of others, making selection more robust. Secondly, using components rather than individual variables reduces the number of total variables in the initial model, further aiding with selection. Thirdly, linear combinations of 'small' effects can be statistically significant (so retained during model selection) when individual time series would not be: see Castle et al. (2011b) for a more detailed discussion. The contributions of individuals variables can be disentangled if needed. Since this process captures the overall variation in vegetation, it should also reduce the problem of random noise due to cloud cover at the time of satellite measurement.

Naturally the issue of employing principal components on non-stationary time series arises. While vegetation follows a relatively seasonally stationary pattern, section 3.4.4 and 3.4.4 also employ principal components on industrial production and emissions which are clearly non-stationary.

There are two common concerns raised when using principal components on non-stationary series (see Machado et al. 2001, Lansangan & Barrios 2009). First, in general due to non-converging second moments of non-stationary series, if combined with stationary data, the non-stationary variables will receive relatively larger component weights. Second, simultaneous drifting of non-stationary variables could spuriously lead to high correlation between variables that are not directly related. However, Harris (1997) shows that while

principal components are not asymptotically efficient, they are consistent when applied to cointegrating series. Addressing these issues for the present application, all series in section 3.4.4 and 3.4.4 are non-stationary, the issue of relative weighting compared to stationary series should thus be minimal. We conduct a quick cointegration analysis (see 3.4.4 and 3.4.4) for the production and emission series suggesting that both cointegrate. Principal components should then be consistent and capture the main underlying features of production and emissions rather than spurious correlation. Overall, principal components here serve primarily as linear combinations of the variables of interest, if all components were included this is equivalent to including the full non-stationary system with rotated axes. Therefore, given that non-stationary series are not mixed with stationary ones, the series cointegrate, and are primarily used as linear combinations of variables, we proceed by using regular components. Possible further work could extend this to employ sparse components (Lansangan & Barrios 2009) which were developed for use on non-stationary series.

For the following analysis, the first three principal components of vegetation are entered for both North America and Eurasia. Cumulatively they explain 93.6 percent of the variation in North America and 88.5 percent of Eurasian variation. Table 13 summarizes the principal components that account for variation in the biosphere and Figure 24 shows the seasonal variation present in the biosphere as measured by PCs. The seasonal pattern is near constant, closely resembling a sine-wave. The first principal component shows a higher amplitude in Eurasia compared to North America but the seasonal pattern is nearly identical. NDVI implicitly covers changes in land-use since it is a measure of photosynthetic activity for a particular area. A forest that is cut down would result in a change of NDVI from around $+1$ to closer to zero for that particular region. However, once NDVI is calculated for large regions and reduced in dimensionality (by PCs), changes in land use would have to occur on a grand scale to be identified in the time series. The principal components of NDVI should, therefore, be interpreted primarily as the variation in plant activity of photosynthesis and respiratory release.

Table 13: Principal components for vegetation, 1981:7–2002:12

North America	Proportion of variance	Cumulative
Principal Component 1	0.831	
Principal Component 2	0.064	0.895
Principal Component 3	0.042	0.937
Eurasia		
Principal Component 1	0.735	
Principal Component 2	0.108	0.843
Principal Component 3	0.042	0.885

3.4.3 *Oceanic Effects: Temperature and Southern Oscillation*

The measure of temperature used here is the anomaly in land and sea surface temperature for the Northern Hemisphere. The temperature anomaly measure is expected to capture the main factors of ocean CO_2 absorption and is available from the NASA Goddard Institute for Space Studies (GISS) (2011) Surface Temperature Analysis from 1880:1–2011:9. The data is measured as an index in 0.01 degrees Celsius of deviations from the 1951–1980 base

Figure 24: First principal components for Eurasian and North American NDVI

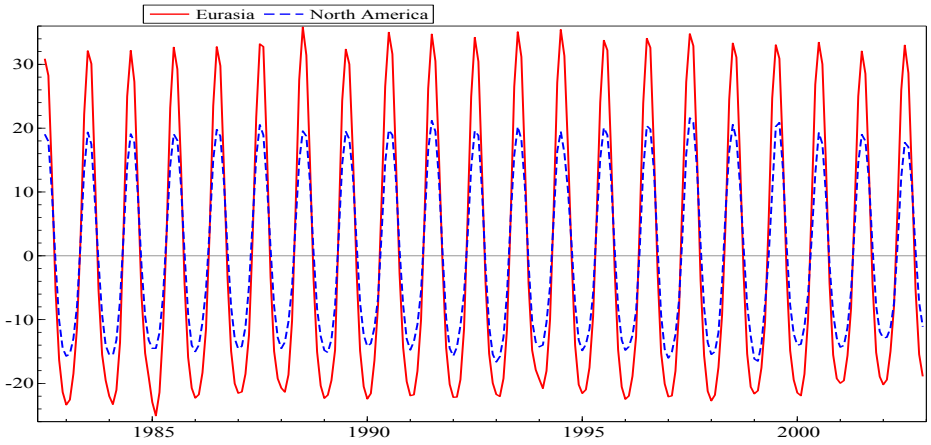
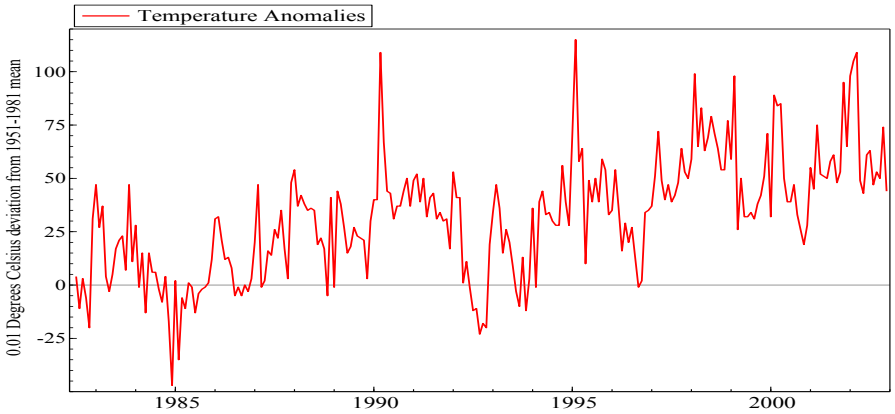


Figure 25: Land and Sea temperature anomalies for the Northern Hemisphere

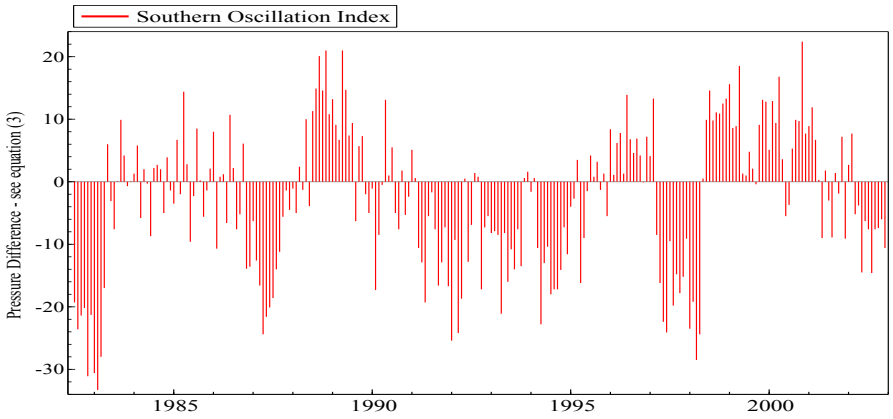


period (see Hansen & Lebedeff 1987, and Hansen et al. 2010 for the detailed measurement methodology). Land measures are taken from multiple stations and are combined and corrected for urban and other non-climatic factors. Sea surface temperature measures are restricted to ice-free regions. Figure 25 shows temperature anomalies from 1982:7–2002:12. Following a spike in temperature anomalies 1990 associated with a strong el Niño event, the low NH temperature anomalies in the early 1990s are partly driven by the Pinatubo eruption, returning to pre-Pinatubo levels around 1995 (Hansen et al. 1996).

The feedback effect of CO_2 is one of the main concerns in climate change and regularly estimated through statistical energy balance models and large-scale global climate models. The level of atmospheric CO_2 feeds to temperature which, in turn, affects the rate of growth of CO_2 particularly through oceanic absorption. If the feedback occurs at a lag, the single-equation approach is valid using the conditional model where temperatures are considered weakly exogenous (see 3.5). While lagged variables are exogenous, the selection algorithm may incorrectly select a lagged rather than contemporaneous variable, in this case the arising problem of endogeneity is addressed here post-selection following the approach of Kaufmann et al. (2006) using instrumental variables.

To take account of weather phenomena through the Southern Oscillation we include the Southern Oscillation index (SOI). Data on the SOI is available from the Australian Bureau of

Figure 26: Southern Oscillation index



Meteorology, (2011) from 1876:1 until 2011:9. Figure 26 shows the SOI (equation (53) specifies the units) for 1982 until 2002, with a noticeably strong episode of El Niño in 1997–1998 (SOI < 0).

3.4.4 *Economic Indicators*

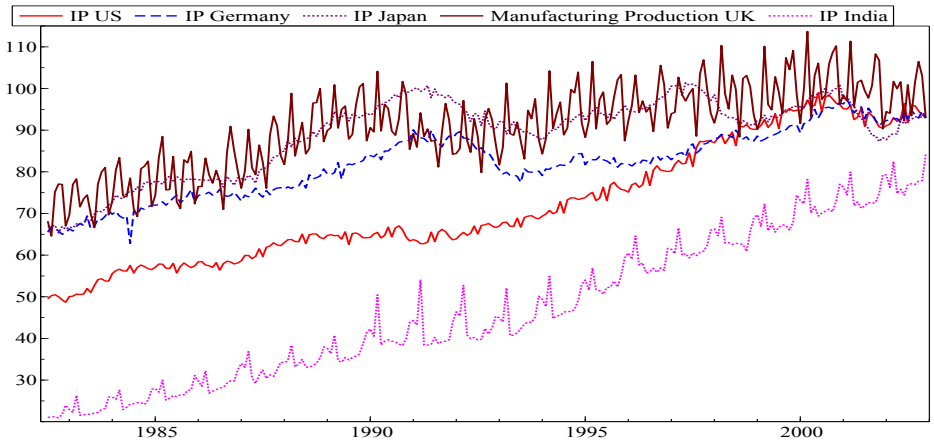
Anthropogenic contributions to atmospheric CO_2 are normally approximated by low-frequency economic indicators; for example annual GDP, population, an estimate of total CO_2 emissions or cement production (see section 3.2). This works reasonably well when trying to explain the long-run dynamics of carbon. However, using only low-frequency annual measures does not allow for estimation of any effect of anthropogenic emissions on the seasonal variation of CO_2. High-frequency (monthly) measures of anthropogenic output permit a richer analysis. Here we use a combination of multiple low-frequency (annual) and high-frequency (monthly) indicators. The annual data is included to provide a robustness check and potentially account for long-run growth. Monthly measures are included for short-run dynamics, which could explain the seasonal fluctuations as well as long-run growth. To capture atmospheric transport, variables are chosen to reflect North America as well as Europe/Asia.

High-frequency (monthly) measures

The main high-frequency indicators for anthropogenic contribution to CO_2 used here are monthly industrial production indices for multiple regions. Industrial production for North America is given by the US Industrial Production Index (2005=100) available from the Federal Reserve, (2011a). The data are not seasonally adjusted and range from 1919:1 until 2011:5. The index measures real output in the sectors covering manufacturing, electric and gas utilities and mining and thus accounts for a large share in business-cycle fluctuations. Activities associated with industrial production account for approximately 52% of emissions for the United States (Environmental Protection Agency - EPA 2011).

To cover Europe and Asia, industrial production indices for the United Kingdom (UK), Germany, India and Japan are included. These measures function as a proxy for business-cycle fluctuations in the Eurasian region. Ideally Chinese and Russian production should also be included, however there are no data available on industrial production for both countries at

Figure 27: Industrial Production indices, 2005=100



the required frequency and time span.[4] UK industrial production is measured as an index (2005=100) of non-seasonally adjusted manufacturing activities. The data are obtained from the Office of National Statistics, 2011 and ranges from 1968:1–2011:6. German and Japanese industrial production is measured by the industrial production index (2005=100) covering manufacturing, mining and electricity, gas and water supply. The series are available only in seasonally-adjusted format from the OECD (2011) from 1960:1 until 2011:2, where for Germany after October 1990 the data account for the accession of the German Democratic Republic to West Germany. The Indian industrial production index (2005=100) covers manufacturing, mining and electricity (Government of India, Ministry of Statistics, 2011) and ranges from 1981:4–2011:5. Table 14 and Figure 27 summarize the high-frequency measures.

Table 14: High-frequency (monthly) variables

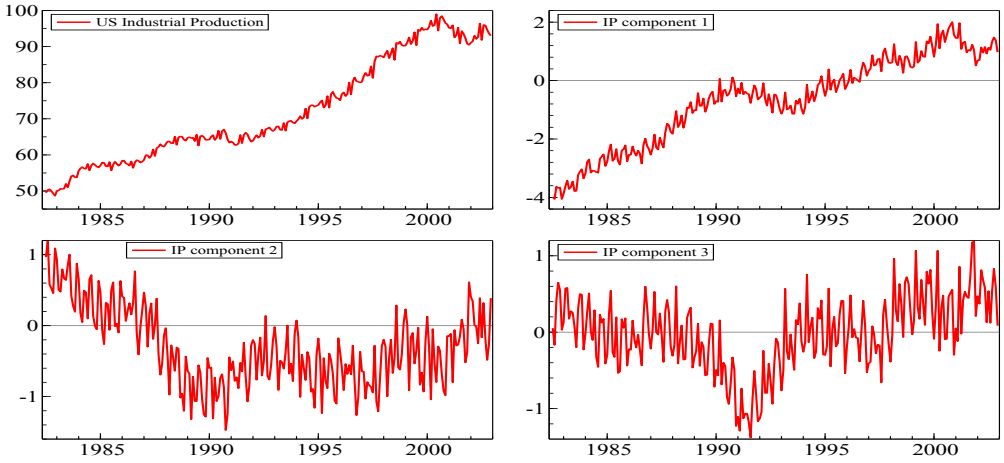
Measure	Description	Range	Source
US Industrial Production	Index 2005=100	1958:3–2011:5	US Federal Reserve
UK Industrial Production	Index 2005=100	1968:1–2011:6	ONS
Germany Industrial Production	Index 2005=100	1960:1–2011:2	OECD
India Industrial Production	Index 2005=100	1981:4–2011:5	Govt. of India
Japan Industrial Production	Index 2005=100	1960:1–2011:2	OECD

The seasonal adjustment of German and Japanese industrial production is visible in their dampened seasonal cycles. The higher seasonal variation in UK industrial production likely stems from it covering primarily manufacturing, which is more volatile to business cycles and seasonal factors than mining and energy production included in the other indices.

These industrial production series are highly collinear, so we again employ principal components to reduce the dimensionality and work with orthogonal variables. This improves robustness in selection as discussed above, as well as it reduces the number of variables and lags to select over (for a reduction from 5 to 3 variables at 12 lags, 24 variables fewer are initially included for selection). As mentioned in section 3.4.2, to support the use of principal components on these non-stationary variables, we provide a quick analysis of

[4] Chinese gross industrial output is only available from 1979–1999 with a change in measure thereafter. Inclusion of this series did not improve selection at the cost of a reduction in observations.

Figure 28: Industrial production principal components



cointegration between the series. We estimate a preliminary VAR covering all 5 industrial production series at 12 lags with a restricted linear trend and unrestricted constant. Testing for cointegrating rank (Hendry & Juselius 2001), reduced rank of 3 cannot be rejected at the 1% level in the VAR of 5 variables, providing evidence for cointegration. Table 15 summarizes the first three industrial production components, which cumulatively explain approximately 97% of variation in production. An alternative specification for future analysis could be to separate the seasonally adjusted from the non-seasonally adjusted variables before taking principal component transformations. Figure 28 compares US industrial production to the three components used in selection.

Table 15: Principal components for Industrial Production, 1981:4–2011:2

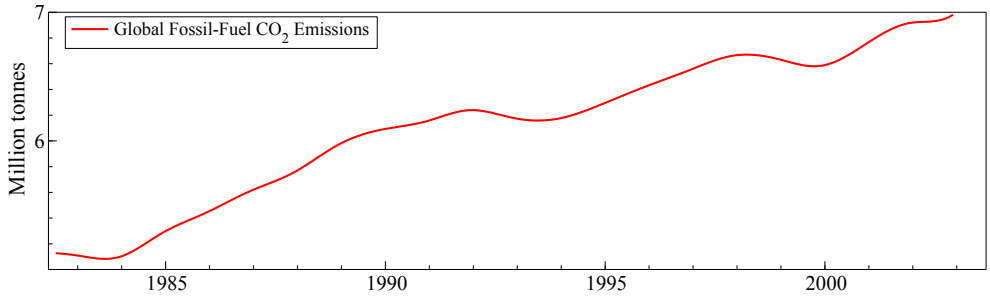
IP	Proportion of variance	Cumulative
Principal Component 1	0.809	
Principal Component 2	0.119	0.928
Principal Component 3	0.047	0.975

These are the first 3 anthropogenic high-frequency components included in the GUM. While industrial production reflects the intensity of economic activity associated with higher emissions, it does not account for changing emission intensity. More efficient processes could lead to an increase in industrial production without an equivalent increase in emissions – this effect is visible for the US, the only country for which high frequency emission data is available (see appendix 3.8.1). This is a crucial missing measure and difficult to control for: an attempt is made by including overall long-run emissions estimates in addition to production.

Low-frequency (annual) measures

As a robustness check we also include the conventional measure of estimated global total CO_2 emissions available only at annual frequency. Total estimated fossil-fuel CO_2 emissions in million metric tonnes cover CO_2 emissions for North America, Western Europe, Eastern

Figure 29: Global (cubic-spline interpolated) CO_2 emissions



Europe, Central Asia and the Far East.[5] Emissions are estimated based on the burning of fossil fuels, cement manufacture and gas flaring by the US Department of Energy (see Marland et al. 2011) from 1950–2007.

In order to usefully combine the annual measure with the monthly variables listed above, global fossil fuel CO_2 emissions are interpolated to monthly observations using cubic splines (see Figure 29). While this is restrictive since any seasonal variation is not observed, seasonal dummy variables and IIS can 'pick up' any systematic or large deviations induced by spline-interpolation. The annual measure is included to account for potential low-frequency movements, but the industrial production short-term indicators are expected to be sufficient to capture inter-annual dynamics.

3.5 ESTIMATION

3.5.1 *Overview*

The joint model of the system outlined above can be described as a vector autoregression (VAR) (see e.g., Hendry 1995). There are two main groups of variables: the main variable of interest $\Delta CO_{2,t}$, and the economic and natural factors which are grouped in the (20×1) vector $\mathbf{X}_t = (\text{IP}_{1-3,t}, \text{Emission}_{1-3,t}, \text{NDVI}_{1-3\text{ Eur, NA},t}, \text{Seasonal NDVI}_{1-3\text{ Eur, NA},t}, \text{Temp}_t, \text{SOI}_t)'$. Let $\mathbf{Y}_t$ denote the full system of these variables, $\mathbf{Y}_t = (\Delta CO_{2,t}, \mathbf{X}_t')'$. The joint model is then given by equation (62), where the vector $\mathbf{Z}_t$ contains deterministic terms and impulse indicators.

$$\mathbf{Y}_t = \left(\begin{array}{c} \Delta CO_{2,t} \\ \hdashline \text{IP}_{1-3,t} \\ \text{Emission}_t \\ \hdashline \text{NDVI}_{1-3\text{ Eur, NA},t} \\ \text{Seasonal NDVI}_{1-3\text{ Eur, NA},t} \\ \text{Temp}_t \\ \text{SOI}_t \end{array} \right) = \sum_{j=1}^{s} \mathbf{A}_j \mathbf{Y}_{t-j} + \Phi \mathbf{Z}_t + \mathbf{e}_t \tag{62}$$

[5] See Marland et al. 2011 for a detailed listing of which countries are included. Eastern Europe includes Russia, Central Asia includes China, and India is covered by the Far East category.

The joint model (62) can be written as a marginal model (63), and a conditional model (64) for $\Delta CO_{2,t}$.

$$\mathbf{X}_t = \sum_{j=1}^{s} \mathbf{A}_{x,j}\mathbf{Y}_{t-j} + \Phi_x \mathbf{Z}_t + \mathbf{e}_{x,t} \tag{63}$$

$$\Delta CO_{2,t} = a_j \sum_{j=1}^{s} \Delta CO_{2,t-j} + \sum_{j=0}^{s} \mathbf{X}'_{t-j}\mathbf{B}_j + \mathbf{Z}_t{}'\boldsymbol{\phi} + \epsilon_{\Delta CO_2 \cdot x,t} \tag{64}$$

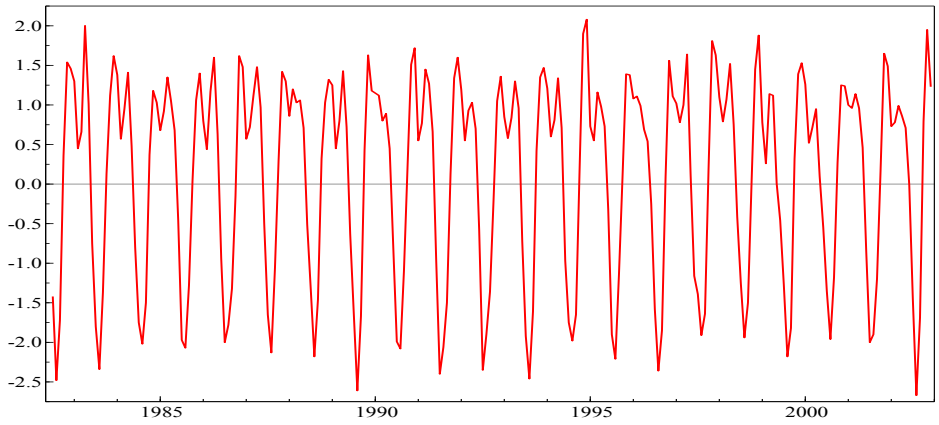
The vector $\mathbf{Z}_t$ consists of non-lagged deterministic terms such as a linear time trend, centred seasonal variables and impulse indicators. Let q denote the total number of explanatory variables that appear on the right-hand side of equation (64). Then the change in atmospheric CO_2 in the conditional model is modelled as a function of past values of the change in CO_2, current and past values of selected independent variables $\mathbf{X}_{t-j}$, and deterministic components. The VAR model specified in (62) contains a large number of variables and it is thus infeasible to select models from a general unrestricted system jointly.[6] Under the assumption of weak-exogeneity we proceed by focusing on the conditional model (64) alone, without modelling the marginal model (63). This is restrictive, as a single equation model implies the existence of only a single cointegrating vector and rules out potential feedback effects. However, when concerned with selection, due to the large number of variables this is a necessary initial step whose primary purpose is identification of relevant variables. This could be followed by a VAR analysis of the selected system. When no selection is conducted, an energy balance model coupled with a model for carbon pools could be used (e.g. Allen et al. 2009). In particular, the cointegrated vector autoregression energy balance model in Chapter 2 could be extended to model CO_2 and other anthropogenic factors endogenously. Equally, an approach following Kaufmann et al. (2006) could be taken, where a complete system model is constructed through estimation of a series of single equation models. Their approach does not make use of model selection techniques and defines all relevant variables *a-priori*. In the present model it is probably safe to assume that in the short time period considered here, the causal effect (if it exists) between production and emissions to CO_2 runs in only a single direction. This may not strictly hold for other variables included (in particular the lowest block in equation 62). However, weak-exogeneity does not rule out two-way causality. For example, if the feedback of CO_2 on temperatures is not directly contemporaneous, the conditional model remains valid. Section 3.6.3 investigates the issue of exogeneity further by using IV estimation to account for endogeneity of temperature.

For the current analysis, the dependent variable ΔCO_2 is modelled as a finite autoregressive distributed lag model (ADL) as specified in equation (64) (see e.g., Hendry 1995, Ch.6). Figure 30 graphs $\Delta CO_{2,t}$ for 1982:7–2002:12. The seasonal variation is so large that it is difficult to visually discern the slow but persistent growth.

There is a large number of potential explanatory variables $\mathbf{X}_{t-j}$ in modelling atmospheric CO_2. The model needs to account for all the above mentioned anthropogenic and natural factors as well as their lag reactions, as CO_2 is a highly autocorrelated time series. Adding IIS quickly moves the general unrestricted model to a situation with more explanatory variables than observations. This used to be a major difficulty in modelling, however, as outlined in

[6] Recent research in this field is being conducted, see Doornik & Hendry (2012).

Figure 30: Dependent variable $\Delta CO_{2,t}$ over 1982:7–2002:12



section 3.3, can now be handled by estimation in blocks using *Autometrics* to select variables to retain in the final model in the form of (64). The estimation procedure operates as follows: first the theory-motivated GUM is specified, then estimation in blocks following the *Autometrics* algorithm selects down to individual terminal models. The union of terminal models is captured by the final GUM. Formally, the selection for the final model can use the likelihood-based SIC, though as each terminal model represents a valid representation, final model selection can be based on other theoretical considerations.

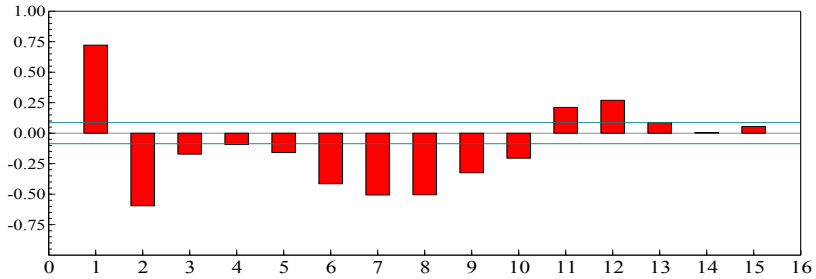
Formulation of the GUM

The dependent variable that is being modelled is ΔCO_2 (the change in atmospheric carbon dioxide measured at Mauna Loa). While *Autometrics* has been shown to be effective at recovering the data generating process in large models, the algorithm is not perfect and is sensitive to initial specification. As a robustness check, therefore, we estimate multiple variations of initial specifications. Further robustness checks and theoretical considerations of robustness are investigated in section 3.6.4. The GUMs 1 and 2 are selected for different initial variables, though they always include the main variables of interest: short-term dynamics captured by monthly industrial production and the control variables for natural factors—temperature, SOI and vegetation. Non-anthropogenic emissions are captured through potential oceanic release of CO_2 (SOI and temperature) and the respiratory release phase in vegetation (NDVI close to zero). The models include general control variables of a constant, a linear time trend and 11 centred seasonal dummy variables (annual frequency -1 dummies with mean zero in the long run), which are subject to selection so not automatically included in terminal models.

To capture inter-seasonal transport dynamics, interaction terms for Winter/Summer are included with each vegetation measure. Summer is defined as May–October and Winter is defined as November–April. Thus, binary variables (Winter and Summer Weights) are interacted with all region-specific NDVI variables and included in the GUM. This could be extended to smoothed weights following a sine/cosine pattern, but binary weights are expected to capture the main seasonal effect of different atmospheric transport.

Specification 1 includes all variables measured at a monthly frequency: natural control variables as well as the first three components of industrial production. Specification 2 includes

Figure 31: Partial-autocorrelation function for ΔCO_2



all short-term dynamics as in specification 1 and adds the first low-frequency interpolated fossil fuel emission variable. The GUM does not initially include any non-linear terms (apart from seasonal NDVI interaction), while this is restrictive, we test for non-linearities post-model selection. Further, a large number of retained impulse indicators could provide evidence for functional form misspecification and potential non-linearities.

Since ΔCO_2 is a highly autocorrelated series when measured monthly, a long lag length is allowed. Based on the partial-autocorrelation function (PACF) of ΔCO_2 and the monthly frequency of the data, the longest lag length is selected to be 12, as longer lags fall below the 95% critical level. Similar results are obtained when likelihood-based lag selection criteria are used. Figure 31 shows the PACF for ΔCO_2. Additionally, 12 lags of each independent variable are added to GUM 1. Lag lengths for GUM 2 are specified based on selection from GUM 1 to compensate for a larger number of independent variables (see section 3.8.2).

Table 16: GUM: Specification 1

Variables included	lag length
Temperature	12
NDVI PC1 Eurasia (Eur) + Winter Interaction	12
NDVI PC2 Eurasia (Eur) + Winter Interaction	12
NDVI PC3 Eurasia (Eur) + Winter Interaction	12
NDVI PC1 North America (NA) + Summer Interaction	12
NDVI PC2 North America (NA) + Summer Interaction	12
NDVI PC3 North America (NA) + Summer Interaction	12
SOI	12
Industrial Production Comp. 1	12
Industrial Production Comp. 2	12
Industrial Production Comp. 3	12
Constant	yes
Trend	yes
Centred Seasonal Variables	yes
Impulse Indicators	yes
Total variables	492

As controls for the terrestrial biosphere are included in every GUM, it is the NDVI measure that defines the maximum number of available observations. Carbon dioxide is measured at Mauna Loa from 1958:3 until present, but NDVI data is only available from 1981:7 until 2002:12, so the vegetation control limits the maximum number of observations to 258. A

maximum lag length of 12 then leads to the estimated sample size being 246 observations ($T = 246$).

Impulse indicator saturation includes one binary variable for each observation. This means that 246 individual binary variables are added to the GUM. Tables 16 and 19 provide an overview of the variables making up the specific GUMs.

3.5.2 *Non-stationarity*

A non-stationary process is a process whose distribution changes over time, for example the mean or variance of the process are non-constant (see Hendry & Juselius 2000, for a detailed discussion). The trending level of CO_2 measured at Mauna Loa is non-stationary as its mean is increasing with time. There are numerous forms of non-stationarity. The time series could be an integrated process (a time series is said to be integrated of order r, or I(r), if differencing the series r times yields a stationary process), or alternatively there could be structural breaks/shifts in coefficients, levels or variances. Non-stationarity through structural breaks need not be removed by differencing. IIS is used to detect structural breaks (as well as mis-specification): if a large number of impulse indicators are selected, the model may be mis-specified, data badly mis-measured; or there are breaks in the data.

The time series properties of atmospheric greenhouse gas concentrations, other radiative forcing series, and global mean temperatures are heavily debated in the literature (Stern & Kaufmann 2000, Kaufmann & Stern 2002, Beenstock et al. 2012, Pretis & Hendry 2013). Determination of an appropriate order of integration and the presence of breaks is further complicated due to concerns about data measurements, however, due to the restricted sample period here, concerns about the change of measurement (Pretis & Hendry 2013)[7] of atmospheric CO_2 are mostly alleviated. While there is convincing evidence that global surface temperatures are I(1) (Kaufmann et al. 2013), there is mixed evidence for the properties of atmospheric CO_2. Over the sample period used in this study, the annual change in CO_2 is increasing (see Figure 22). ΔCO_2 for the present analysis could therefore be I(1), or I(0) around a trend. However, unit-root non-stationary data is not a problem in automatic model selection so long as it is approached correctly. Selection in *Autometrics* is primarily based on likelihood ratios (which can be interpreted as t- and F-statistics) that under non-stationarity can follow non-standard distributions. Sims et al. (1990) show that the limit distributions of these test statistics are standard if the original equation can be re-written in terms of coefficients on mean-zero stationary variables. That is to say, the actual re-parametrization is not required, merely the existence of a linear re-parametrization in mean-zero variables is sufficient for the test statistics in the original equation to have standard distributions. So long as the equation can be written that way, selection based on t- and F-statistics will be valid. For example, suppose ΔCO_2 is I(1) and was modelled as an ADL(2,2) model with a single I(1) independent variable $x_{1,t}$:

$$\Delta CO_{2,t} = \theta_1 \Delta CO_{2,t-1} + \theta_2 \Delta CO_{2,t-2} + \beta_0 x_{1,t} + \beta_1 x_{1,t-1} + \beta_2 x_{1,t-2} + \epsilon_t \tag{65}$$

[7] Pre 1958 measurements of CO_2 rely on approximations based on ice-core reconstructions, while data from 1958 onwards stems from in-situ measurements.

This can be re-written in equivalent equilibrium correction form:

$$\Delta^2 \mathrm{CO}_{2,t} = \lambda_1 \underbrace{\Delta^2 \mathrm{CO}_{2,t-1}}_{I(0)} + \kappa_0 \underbrace{\Delta x_{1,t}}_{I(0)} + \kappa_1 \underbrace{\Delta x_{1,t-1}}_{I(0)} + \alpha(\Delta \mathrm{CO}_{2,t-1} - \delta x_{1,t-1}) + \epsilon_t \tag{66}$$

where the coefficients in (66) are functions of the coefficients in (65). The variables $\Delta^2 \mathrm{CO}_{2,t-1}$, $\Delta x_{1,t}$, $\Delta x_{1,t-1}$ are mean zero and stationary. Under cointegration $(\Delta \mathrm{CO}_{2,t-1} - \delta x_{1,t-1})$ will also be stationary. The existence of this re-parametrization is sufficient – the actual re-parametrization is not necessary for the required test statistics to have standard distributions and it is therefore acceptable to work with the model given in equation (65) rather than (66). The estimation process here proceeds with the GUM in form of the ADL model (64) rather than in equilibrium correction form (66). While equilibrium correction models can be beneficial when dealing with integrated data, the re-parametrisation is not necessary as the previous section describes. Further, in model selection with a large number of potentially relevant variables, it is not straightforward to specify the equilibrium correcting term *a priori* in the GUM.[8]

There are potential problems during selection when a path is considered in which this re-parametrization is not possible. Such a difficulty is hard to avoid, so is handled here by using tight significance levels that allow for possible non-standard distributions. As Figure 22 showed, $\Delta_{12}\mathrm{CO}_2$ manifests a strong trend so is clearly not stationary: in monthly data, that trend is 'swamped' by the seasonal variation. Due to this strong seasonality, augmented Dickey–Fuller (Dickey & Fuller 1981) type tests would not successfully identify a unit root. Seasonal unit-root tests are somewhat unreliable as other determinants, such as breaks, are not included, but based on the apparent stochastic trend in $\Delta_{12}\mathrm{CO}_2$, and the long atmospheric lifetime of CO_2 it is likely that $\Delta\mathrm{CO}_2$ is well approximated by a process integrated of order one I(1) over the sample period.

If the system is unbalanced, e.g. the order of integration of the dependent variable (ΔCO_2) is lower than the independent variables (Banerjee et al. 1993, Ch. 6) estimation is still feasible. This topic has received less attention in the literature, though Baffes (1997) proposes a post-estimation test for unbalanced or near to unbalanced regressions. Standard single equation cointegration models rely on tests of stationarity on the residuals of the model. In the I(1) on I(1) case, if the residual itself is a stationary process the system is said to cointegrate. In an unbalanced system, stationarity of the residual is necessary but not sufficient to establish a sensible relationship. Stationarity of the residuals in I(0) on I(1) models can be the result of either a stationary combination of I(1) variables (cointegration), or if the model fails, the residuals could simply pick up the stationarity of the dependent variable. If the residuals are I(0) we require an additional test whether the model is appropriate or the stationarity is simply the result of a failing model. For this matter Baffes (1997) proposes to test the stationarity of the predicted values. Therefore an appropriate post-estimation testing procedure in unbalanced models could be to first test for residual stationarity, followed by a test for stationarity of the fitted values.

However, given the trend present in $\Delta_{12}\mathrm{CO}_2$ and thus also in $\Delta\mathrm{CO}_2$, we proceed by estimating the model as specified in equation (64), relying on the fact that it can be re-

[8] Selection in an equilibrium correction framework could be conducted jointly over a complete set of first differences and levels while modelling the endogenous variable in first differences and forcing the inclusion of a one-period lagged dependent variable in levels.

written as coefficients on stationary mean zero variables (see Banerjee et al. 1993), with tight significance levels to account for selection effects where an I(0) re-parametrization is not possible. To cover the case of unbalanced models, the model residuals are tested for unit-roots post-selection, however, this test will only have power if the algorithm fails, since *Autometrics* selects congruent models.

3.5.3 *Algorithm specification*

The *Autometrics* algorithm is used to estimate and select within the GUMs described in tables 16 and 19. Selection is done at a 0.1% significance level, considerably tighter than the conventional 5% or 1% used in the literature. In models starting with K irrelevant variables, this implies that on average $0.001K$ irrelevant variables are retained in the terminal model, with K=500 this implies 0.5 of a variable will be retained by chance. The current model is reduced by removing the least significant variable until variables cannot be dropped at the 0.1% level. At the termination of each path, models are backtested to the initially specified GUM when feasible. Diagnostic tests, defined formally below (see Doornik & Hendry 2009), are conducted at a 1% level, for normality, heteroskedasticity, coefficient constancy (set to a 70% sample split), residual autocorrelation and autoregressive conditional heteroskedasticity, both based on 8 lags. The specified GUM includes more variables than observations with IIS, so diagnostic tests are only applied to terminal models, and if satisfactory, conventional standard errors are used.

3.6 RESULTS

GUM Specification 1 includes all natural controls and monthly anthropogenic components; specification 2 includes all natural controls as well as monthly and annual anthropogenic components. The crucial feature of *Autometrics* is determining the selection of variables, rather than their estimated coefficients, although bias corrections are feasible in specific cases. Below, we also note the relative importance of individual variables through decompositions of explained variance.

The selection algorithm results in 14 terminal models for specification 1, and 18 terminal models for specification 2. Detailed results for GUM specification 2 are reported in the appendix. It may surprise that so many congruent undominated different representations can be found at such a tight significance level as 0.1%, but this merely reflects the high collinearity both between the different series and over time as represented by their lagged values. Most of the terminal models are minor variations on others as the final GUMs had 21 and 25 variables respectively.[9] The final models are selected from the set of terminal models by the smallest Schwarz (1978) information criterion value. Equation (67) shows the selected final models for specification 1, the model for specification 2 is reported in the appendix.

[9] Detailed results are available on request.

$$
\begin{aligned}
\widehat{\Delta \mathrm{CO}}_{2,t} = & \underset{(0.053)}{0.24}\ \Delta \mathrm{CO}_{2,t-1} - \underset{(0.049)}{0.67}\ \Delta \mathrm{CO}_{2,t-2} + \underset{(0.05)}{0.20}\ \Delta \mathrm{CO}_{2,t-3} \\
& - \underset{(0.037)}{0.32}\ \Delta \mathrm{CO}_{2,t-4} + \underset{(0.057)}{0.24}\ \mathrm{IP}_{1_{t-1}} - \underset{(0.056)}{0.30}\ \mathrm{IP}_{1_{t-4}} \\
& - \underset{(0.034)}{0.20}\ \mathrm{IP}_{2_{t-4}} + \underset{(0.035)}{0.15}\ \mathrm{IP}_{3_t} + \underset{(0.0005)}{0.003}\ \mathrm{Temp}_{t-4} \\
& - \underset{(0.0013)}{0.007}\ \mathrm{SOI}_{t-5} - \underset{(0.007)}{0.042}\ \mathrm{NDVI}_{1,\,\mathrm{Eur}_{t-1}} + \underset{(0.003)}{0.019}\ \mathrm{NDVI}_{1,\,\mathrm{Eur}_{t-10}} \\
& - \underset{(0.006)}{0.020}\ \mathrm{NDVI}_{1,\,\mathrm{Eur}_{t-12}} + \underset{(0.007)}{0.026}\ \mathrm{w\_NDVI}_{3,\,\mathrm{Eur}_{t-8}} \qquad (67) \\
\widehat{\sigma} = 0.212 \quad & T = 246 \quad n = 14 \quad \mathsf{SIC} = -0.006 \quad \mathsf{F}_{\mathsf{ar}}(8, 224) = 1.55 \\
\chi^2_{\mathsf{nd}}(2) = 1.47 \quad & \mathsf{F}_{\mathsf{reset}}(2, 230) = 1.42 \quad \mathsf{F}_{\mathsf{arch}}(8, 230) = 1.07 \quad \mathsf{F}_{\mathsf{het}}(28, 217) = 1.42
\end{aligned}
$$

Let $\mathsf{F}_{\mathsf{name}}$ denote an approximate Lagrange-multiplier F-test, then F_{ar} tests for autocorrelation of order k (see Godfrey 1978, for original test under stationarity), $\mathsf{F}_{\mathsf{het}}$ tests for heteroscedasticity (see White 1980, for original test under stationarity[10]), $\mathsf{F}_{\mathsf{arch}}$ tests for k^{th}-order autoregressive conditional heteroskedasticity (ARCH: see Engle 1982, for original test under stationarity), $\mathsf{F}_{\mathsf{reset}}$ tests for functional-form mis-specification (see Ramsey 1969), and $\chi^2_{\mathsf{nd}}(2)$ tests for non-normality (Doornik & Hansen 2008).

First, consistent with theory, controls for natural factors are selected in both final models. Temperature anomalies enter the model with a positive coefficient, likely capturing the effect of oceanic uptake such that atmospheric CO_2 increases with higher temperatures. Vegetation controls through the principal components of NDVI are selected in both models, as is the control for Southern Oscillation. However, a key finding is that in both terminal models, natural controls are insufficient to account for the variation in the change of atmospheric CO_2 in this conditional model. Anthropogenic factors captured through components of industrial output indices are consistently selected in both models. Selection of these components is highly stable over the two models as the selected production components are near-identical in models 1 and 2. Selection of these is robust to the addition of emissions components which are not selected in specification 2, suggesting that the high frequency measures provide a better approximation for anthropogenic emissions measured at Mauna Loa.

Second, most selected variables enter the model in lagged form. Only the third principal component of production (in specification 1 and 2) and the first component of Eurasian NDVI (in specification 2) have an estimated immediate effect on the growth of CO_2. Most anthropogenic emissions and vegetation growth "lead" measured atmospheric CO_2 by suggested time periods of 1 to 5 months (with the exception of the NDVI series). Relative to the initial sizes of the GUMs, few variables are retained, yet relative to the tight significance levels, many more are retained than could be attributed to chance (less than 1 on average).

Third, the final models are well specified. As a result of the algorithm, all models pass tests for normality, heteroskedasticity, residual autocorrelation and autoregressive conditional heteroskedasticity. There are no indicators from IIS retained in the final models suggesting

[10] Caceres & Nielsen (2005) show that the White test is unaffected by non-stationarity (for non-explosive processes) under the null of no misspecification.

that the model is correctly specified and there appear to be no real structural breaks or shifts in the change of atmospheric CO_2.

Fourth, no deterministic variables are selected: no constant, time trend or centred seasonals appear in the final models. Consequently, although a possibility from the specification of the GUM, there is no autonomous growth despite the growth in $\Delta_{12}CO_2$ and no 'exogenous' seasonality even though Figure 30 showed such constant seasonal patterns. This suggests that changes in CO_2 are well approximated by the selected final variables covering anthropogenic and natural factors. Linearity of both models is not rejected using a non-linear factor based test (Castle & Hendry 2010), (Model 1: p=0.458, Model 2: p=0.23). If constants are included post-selection (both not statistically different from zero in models 1 and 2), R^2 can be used as a rough measure of goodness of fit. Both final models exhibit a high goodness of fit (R^2 = 0.974 and = 0.973 respectively), in comparison a random walk model with deterministic seasonal indicators results in a much lower goodness of fit, R^2 = 0.753. High goodness of fit of the models is not a straight result of selection as *Autometrics* does not directly maximize the goodness of fit. Moreover, R^2 measures should not be attributed much weight when assessing models, as there are preferred likelihood-based measures that also account for the number of parameters included.

3.6.1 *GUM: Specification 1*

Specification 1 in equation (67) covers all variables measured at a monthly frequency. *Autometrics* in specification 1 with 246 observations estimated 345 models reducing the number of explanatory variables from an initial 492 (split into initial 6 blocks) down to 14 in the final model. The final model passes the stationarity test on the residuals, the presence of a unit root is rejected at the 1% level when using ADF tests with lag specifications from 1 to 12. To alleviate concerns of an unbalanced regression we follow Baffes (1997) approach and also conduct conventional unit root tests on the predicted values. A unit root is rejected for the fitted values at the 1% level for lags 1 to 12. There are no impulse indicator variables selected in the final GUM. Together this provides evidence for a well specified model that forms a stationary relationship.

Selection: neither the constant, the linear time trend nor centred seasonal variables feature in the final conditional model, so that the seasonality and increase in the growth of CO_2 are explained by the anthropogenic and natural factors. All selected variables (apart from IP_{3_t}) enter the model as lags, suggesting a delay in the effect of CO_2 emissions/absorption and measurement at Mauna Loa. The longest lag on an anthropogenic component is 4 months.

The anthropogenic sources that are selected are all three principal components for industrial production at lag lengths ranging from immediate t to $t-4$. Component 1 is selected at $t-1$ and $t-4$ with opposite signs. This may suggest that it enters the model as a difference, however, a likelihood ratio test rejects the model with IP_1 entering as a difference (LR = 26.856, $p < 0.0001$). As these variables are principal components of production indices, the coefficients are not straightforward to interpret. The key result is the consistent selection, relative importance and lag length of these, rather than the signs of individual estimated coefficients.

The non-seasonally weighted principal components for NDVI are mostly selected, but only the Eurasian region is included. Given that the growth cycle is relatively similar in North

America and Eurasia this should not be over-interpreted. It is likely that the PCs for North American and Eurasian vegetation capture a very similar pattern and are to a considerable extent interchangeable. The negative coefficient on the first PC of vegetation (at $t-1$, as well as in the long-run solution below) supports the theory that increased vegetation activity slows down CO_2 growth. The near equal magnitude, opposite signs on $t-10$ and $t-12$ suggest a difference, a pattern also seen in (77). The coefficient on temperature is positive at a lag of $(t-4)$, likely capturing reduced oceanic absorption under higher temperature. Southern Oscillation enters the model at a lag of five months with a negative coefficient. Thus, during El Niño episodes ($SOI < 0$) growth in CO_2 appears to increase, in line with findings of other papers (see Bacastow 1976, C. D. Keeling & Revelle 1985).

To quantify and assess the relative importance of each regressor, we decompose the total explained variation in ΔCO_2. Decomposition is not straightforward when independent variables are correlated. The partial R^2 provides a measure of the marginal contribution to explained variation for a given variable, while holding other factors constant.[11] Table 17 ranks the selected variables by the partial R^2.

Table 17: GUM: Specification 1 relative importance

Variable	Partial R^2
$\Delta CO_{2,t-2}$	0.4472
$\Delta CO_{2,t-4}$	0.2461
$NDVI_{1,\,Eur_{t-10}}$	0.1583
$IP_{2_{t-4}}$	0.1259
$NDVI_{1,\,Eur_{t-1}}$	0.1259
$Temp_{t-4}$	0.1225
$IP_{1_{t-4}}$	0.1105
SOI_{t-5}	0.0836
$\Delta CO_{2,t-1}$	0.0796
IP_{3_t}	0.0756
$IP_{1_{t-1}}$	0.0727
$\Delta CO_{2,t-3}$	0.0619
$w\_NDVI_{3,\,Eur_{t-8}}$	0.0578
$NDVI_{1,\,Eur_{t-12}}$	0.0483

Based on this decomposition, the single largest (non-autoregressive) contribution comes from the Eurasian NDVI principal component of vegetation followed by industrial production. The anthropogenic components explain a large fraction of the variation in atmospheric CO_2. The result of the vegetation cycle accounting for a large fraction of the variation in CO_2 is robust to the model selection approach employed here.

3.6.2 *Addressing Endogeneity - Instrumental Variable Estimation*

While the temperature variable is selected at a lag and is thus not endogenous by construction, it can be the case that the selected lag specification is incorrect and temperatures should enter the model contemporaneously, in which case the radiative forcing effect of CO_2 concentrations

[11] An intercept is forced to be retained in selection of these models, though it is not statistically different from zero for model 1 and 2.

directly leads to an endogeneity problem.[12] To address this issue we use an instrumental variable approach post-selection. Instrument choice is based on Kaufmann et al. (2006) where the endogenous temperature anomaly variable is instrumented using stratospheric aerosol (volcanic activity) time series (Sato et al. 1993). Large volcanic eruptions inject sulphate aerosols into the stratosphere and induce surface cooling (Kelly & Sear 1984). Thus, while not affecting CO_2 concentrations in the short-run, these eruptions affect global mean temperature anomalies. The selected model is therefore re-estimated where the lagged temperature variable is replaced by its contemporaneous counterpart, which is then instrumented using the aerosol observations up to a lag of 3 months based on approximate durations of atmospheric transport of large scale eruptions (McCormick et al. 1995). The estimated models for both GUM specification 1 and 2 are reported in the appendix. While not directly comparable due to the difference between contemporaneous and lagged variables, the IV estimates of the temperature effect on CO_2 are slightly lower than the standard OLS estimates. The static long-run estimates of temperature on CO_2 drops from 0.0018 (Model 1) to 0.0015 (Model 1 IV)[13]. The instruments are jointly significant in the first stage (F(4,229) = 10.2 for model 1, and F(4, 229) = 10.45 for model 2). However, there is some concern of weak instruments, while jointly significant, individual significance of the instrumental variables in the first stage regressions is low.

An alternative to replacing the lagged temperature variable with its contemporaneous version post-selection is to follow Hendry & Johansen (2014) and force the retention of the contemporaneous temperature variable in the GUM. When forcing contemporaneous temperatures, selection remains relatively stable compared to the previous model – industrial production, SOI and NDVI are selected with minor variations in the lag structure.[14] The coefficient on temperature is positive and significant. Thus, in this particular application there appears little difference in forcing temperatures pre-selection compared to adjusting the variable post-selection.

3.6.3 *Test of Super Exogeneity*

Given concerns on weak exogeneity and parameter stability of some of the included variables in section 3.5, here we address some of these issues by providing a test for super-exogeneity. In particular, failure of weak exogeneity of temperature to the parameters in the conditional model (64) is a possibility. However, weak-exogeneity of temperature does not rule out two-way granger causality – temperatures can be weakly exogenous to CO_2 if feedback is not directly contemporaneous, as the selected model suggests. Modelling CO_2 in the conditional model then remains valid under weak-exogeneity. A thorough test of weak exogeneity requires the variables to be modelled as a system of equations, which, given the model selection approach here, is not straightforward. A VAR on a subset of variables, or analysis of partial systems (see Harbo et al. 1998) would also be potential approaches. An alternative to this is to test the variables for super exogeneity instead - a condition which relies on weak exogeneity. For a conditioning variable to be super exogenous to the

[12] On a monthly time-scale weak exogeneity of temperatures is feasible since the feedback may not be contemporaneous but occur at a lag. On an annual time scale, Chapter 2 provides a complete system model tackling feedback consistent with an energy balance model.

[13] For model 2 the estimated long-run effect drops from 0.0019 to 0.0015.

[14] Full results are omitted here, available upon request.

parameters in the conditional model, we require weak exogeneity of the conditioning variable and parameter invariance to changes in parameters in the marginal model (see Hendry 1995, Ch.5). Parameters in the conditional model should be invariant to interventions affecting parameters in the marginal model. A test for super exogeneity is therefore not a direct test of weak exogeneity, super exogeneity could be rejected due to failure of parameter constancy, failure of invariance or failure of weak exogeneity. However, if super exogeneity cannot be rejected this provides evidence that weak exogeneity of the conditioning variables cannot be rejected either.

Table 18: Retained Indicators from Marginal Models

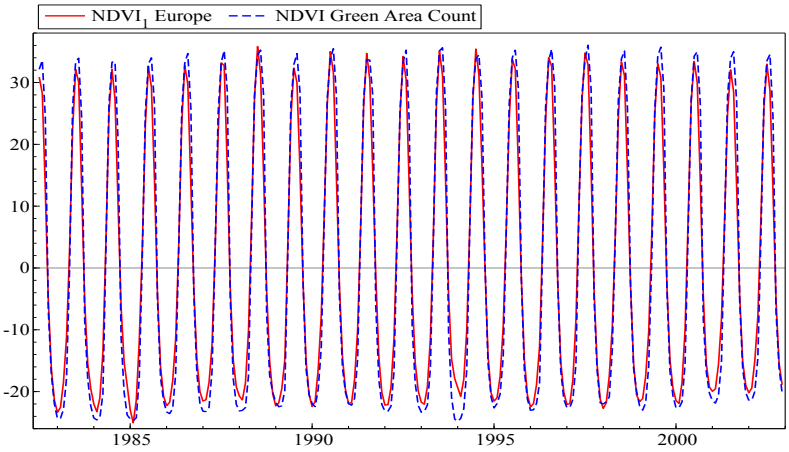
Marginal Model	**Indicators for Model 1**
IP_1	None
IP_2	2002(5), 2002(6)
IP_3	1985(2), 2001(1)
Temp.	1984(12), 1995(2), 2001(11)
SOI	None
$NDVI_{1,Eur}$	1983(5)
$w\_NDVI_{3,Eur}$	1983(1), 1984(3), 1984(4), 1985(1), 1985(2), 1987(3), 1989(3), 1989(4), 1989(5)

Here we apply an automatic test for super exogeneity based on impulse indicators (Hendry & Santos 2010). The idea is to estimate congruent marginal models for the conditioning variables and detect interventions/shifts using impulse indicators. If super exogeneity holds, these interventions should not affect parameters in the conditional model. These indicators are then included in the conditional model and tested for joint significance using an F-test which under the null of super exogeneity will follow an approximate F-distribution. Using *Autometrics* we estimate marginal models (63) for each of the conditioning variables present in the final model 1 (67), and retain significant indicators. The models are selected based on a 0.1% level of significance and by construction satisfy the standard diagnostic tests. Table 18 lists the retained indicators for each marginal model of the conditioning variables. The retained indicators are then included in the conditional model (67) and tested for joint significance. The null hypothesis of super exogeneity of the conditioning variables to the parameters cannot be rejected $F = 0.457$ $(p = 0.96)$ for model 1. This suggests that the parameters in the conditional model are invariant to changes in the marginal models and provides potential evidence that weak exogeneity of the variables cannot be rejected either for the short time series considered here. However, this result can stem from a relatively low power to reject super-exogeneity of the test and the short sample period considered. The test for super-exogeneity here should then primarily be interpreted as a evidence for parameter constancy.

3.6.4 *Model Robustness Checks by Extra Variables*

The series of atmospheric CO_2 is clearly trending and it might be the case that industrial production is only consistently selected as it is one of the few similarly trending series in the

Figure 32: First NDVI PC for Europe and NDVI measure of green area



specified GUMs.[15] This is highly unlikely as the model already accounts for a wide range of (non-stationary) control variables. Industrial production is selected over a constant (trend in levels), a linear trend (quadratic trend in levels) and the interpolated emission measure as well as seasonal effects. However, to provide additional checks that the selection of production is not the result of simply spurious relations we consider the following additional robustness checks: deforestation and oil price are considered as additional explanatory factors, and further, 4 irrelevant (trending) simulated random walk variables are included in the GUMs.

DEFORESTATION AND THE OIL PRICE The current NDVI variables used here (see section 3.4.2) measure the annual cycle, but not the level of vegetation. One might argue that the level of Northern Hemisphere vegetation, in other words deforestation or forest coverage, should also play a role. To investigate this, rather than taking the average NDVI value, we modify the algorithm used in the previous section to provide a count of grid points where NDVI exceeds the threshold level of 0.4 for the entire Northern Hemisphere.[16] This provides a measure of total green area. As Figure 32 shows, the values of total green area are nearly identical to the simple annual cycle measure included in the model.

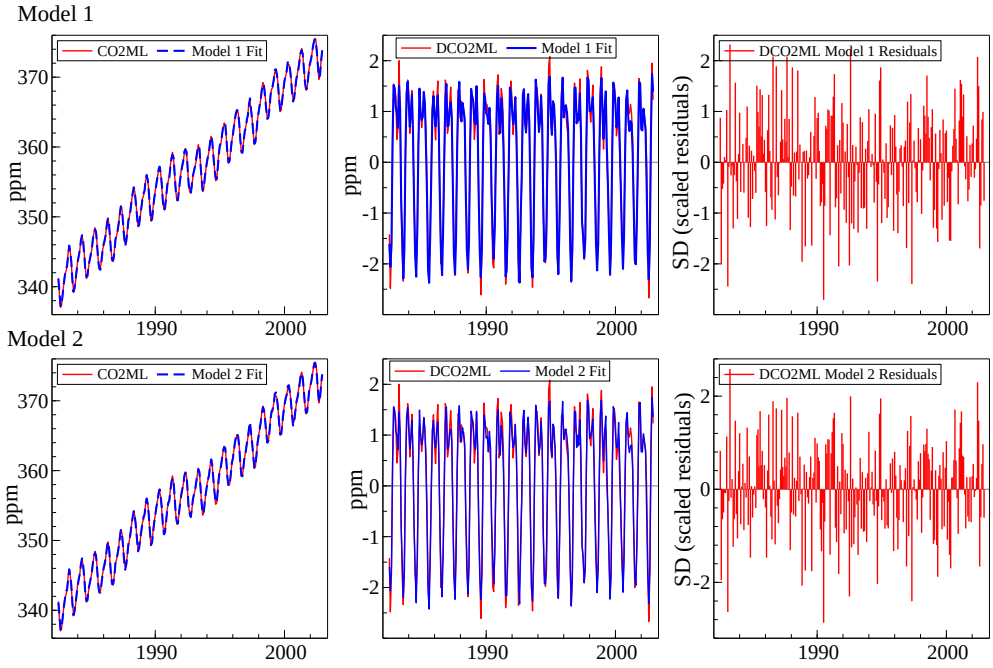
A regression of total green area on the first Eurasian and North American components results in a near perfect fit, $R^2 = 0.975$ (correlations with the components are: $\rho_{\text{Green,NDVI:Eur}} = 0.985$ and $\rho_{\text{Green,NDVI:NA}} = 0.984$ respectively). A linear trend in total green area is not significant, $p = 0.673$. This suggests that the currently included vegetation measure is sufficient and the annual vegetation cycle captures the same information as a measure of deforestation, which appears to remain stable in the Northern Hemisphere.

In the previous analysis it was assumed that production variables sufficiently capture emissions and already contain any relevant information captured in the oil price. As a precaution against this failing, the actual annual emissions measures were included to capture all emissions related information missed by production. Nevertheless, as a robustness check GUM1 is re-estimated by including 12 lags of the monthly US dollar spot price of WTI crude

[15] I would like to thank participants at the 11^{th} *OxMetrics* Users Conference at George Washington University, and participants at the Oxford Gorman and Econometrics Workshops for comments on this.

[16] The choice of 0.4 is slightly arbitrary, though changing the threshold would generally only affect the magnitude of the cycle, not the trend. If there is a declining/increasing trend in the number of "green" grid points, any reasonably chosen value between 0.2 and 0.7 should capture this.

Figure 33: Fitted and actual values in differences and levels, and residuals



oil (Federal Reserve 2011b), inflation adjusted using US CPI data (US Department of Bureau of Labour Statistics 2011). While there was some variation in the lags of selected terminal variables, the oil price was not selected in any of the specifications.

SIMULATED IRRELEVANT VARIABLES There is large scale simulation evidence for the success of *Autometrics* being able to identify irrelevant variables (see Castle et al. 2011a). Nevertheless, to provide extra robustness checks, here four artificial irrelevant sun-spot variables are added to GUM1. This is to provide an example rather than a full scale simulation. The artificial DGP for the variables $v_1, ..., v_4$ is given by equation (68).

$$v_{i,t} = v_{i,t-1} + u_t \text{ where } u_t \sim \text{iid N}(0,1) \text{ for } i = 1,2 \tag{68}$$

$$v_{i,t} = v_{i,t-1} + \delta t + u_t \text{ where } u_t \sim \text{iid N}(0,1) \text{ for } i = 3,4$$

The trend coefficient $\alpha = 0.0001$ is chosen to be of similar order of magnitude as the linear trend in ΔCO_2. While there was slight variation in lag selection, none of the artificial variables are selected, while industrial production, temperature, SOI and vegetation remain selected consistently. One impulse is selected at 1983:4. Overall, none of the simulated irrelevant variables, or oil price are retained, this provides evidence for the robust selection of Industrial production which is unlikely solely due to the similar trending properties.

3.6.5 *From Changes to Levels*

The models estimated as GUM specification 1 and GUM specification 2 describe the change in atmospheric carbon dioxide, ΔCO_2. While most of the analysis has been focused on the change in CO_2, it is possible to recover the estimated level from the models 1 and 2. Level

estimates for model 1 and model 2, based on $\widehat{\mathrm{CO}}_{2,t} = \widehat{\Delta \mathrm{CO}}_{2,t} + \mathrm{CO}_{2,t-1}$ are given in Figure 33 and show the close fit. The next step is to attribute the components of the long-run explanation. To do so, we derive the relation after all dynamics from lagged variables have been solved out (the 'long-run solution': see Hendry 1995, p.212). In a simple ADL(1) of the form:

$$y_t = \lambda_1 x_{1,t} + \lambda_2 y_{t-1} + e_t \tag{69}$$

where $|\lambda_2| < 1$, the long-run conditional expected value is:

$$\mathsf{E}[y \mid x] = \lambda_1 x_1 / (1 - \lambda_2) = \beta_1 x_1 \tag{70}$$

Then, based on the theoretical specifications given in equations (58)–(59), the solved estimated model is in the approximate form given in equation (71):

$$\Delta \mathrm{CO}_{2,t} = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_q x_{q,t} + \epsilon_t \tag{71}$$

This can be re-written as:

$$\mathrm{CO}_{2,t} = \mathrm{CO}_{2,t-1} + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_q x_{qt} + \epsilon_t \tag{72}$$

Recursive substitution for $\mathrm{CO}_{2,t-1}, \mathrm{CO}_{2,t-2}, \ldots$ in equation (72) yields:

$$\mathrm{CO}_{2,t} = \mathrm{CO}_{2,0} + \beta_1 \sum_{j=1}^{t} x_{1,j} + \beta_2 \sum_{j=1}^{t} x_{2,j} + \cdots + \beta_q \sum_{j=1}^{t} x_{q,j} + \tilde{e}_t \tag{73}$$

where $\tilde{e}_t = \sum_{j=1}^{t} \epsilon_j$, CO_2 cumulates all past shocks to emissions.

We divide the variables into two different groups: let s be the number of variables $x_{i,t}$ that have a near-stationary cumulative sum (non-strongly-trending) $\sum_{j=1}^{t} x_{i,j} \sim \mathsf{I}(0)$, so that $q - s$ variables have non-stationary cumulative sums (trending) $\sum_{j=1}^{t} x_{i,j} \sim \mathsf{I}(r)$ where $r > 0$. Equation 73 can then be expressed as:

$$\mathrm{CO}_{2,t} = \mathrm{CO}_{2,0} + \mathbf{x}'_{s,t} \boldsymbol{\beta}_s + \mathbf{x}'_{q-s,t} \boldsymbol{\beta}_{q-s} + \tilde{e}_t \tag{74}$$

where $\mathbf{x}_{s,t}$ and $\mathbf{x}_{q-s,t}$ are $s \times 1$ and $(q-s) \times 1$ column vectors respectively with $\sum_{j=1}^{t} x_{i,j}$ as their row elements. Equation (74) implies that the trending level of CO_2 is a function of the near-stationary and non-stationary cumulative sums of variables in our model. Variables $\mathbf{x}_{s,t}$ whose cumulative sums are near-stationary by nature cannot drive the trend. Only explanatory variables with non-stationary cumulative sums, $\mathbf{x}_{q-s,t}$, determine the trend. This provides a straight-forward method of evaluating the underlying factors of the long-run trend. Quantifying the trend in the present application, atmospheric CO_2 concentrations rise by approximately 30ppm over the sample period (see Figure 34).

Out of the selected variables in models 1 and 2, only a sub-set exhibit strongly trending cumulative sums, which are all the anthropogenic factors and the temperature anomaly. The natural controls of NDVI while fluctuating seasonally only account for a drop of approximately

2.5ppm CO_2 over the sample period. SOI appears to be slightly increasing over the time frame considered, but its effect is small in magnitude when weighted by its long run coefficient, accounting for an approximate increase of 2ppm (see Figure 34). Importantly, neither final model includes a deterministic intercept or trend, which on summation would become a linear or a quadratic time trend. However, summed variables do not have a straightforward interpretation in the case of PCs of industrial output and temperature. The trending temperature anomaly is likely a mutually supporting feedback effect driven by the increase in anthropogenic activity, as mentioned in section 3.4. Overall, the trend in the levels of CO_2 in these conditional models is derived from the trends in the independent variables in both estimated models, so is attributed primarily to the PCs of production and partly by temperature. Specifically, the empirical equivalents of (74) for model 1 are:

$$\mathbf{x}'_{s,t}\boldsymbol{\beta}_s = -0.0038 \sum_{j=1}^{t} SOI_j + (-0.0275 \sum_{j=1}^{t} \text{NDVI}_{1,\,\text{Eur}_j} + 0.0169 \sum_{j=1}^{t} w\_\text{NDVI}_{3,\,\text{Eur}_j}) \tag{75}$$

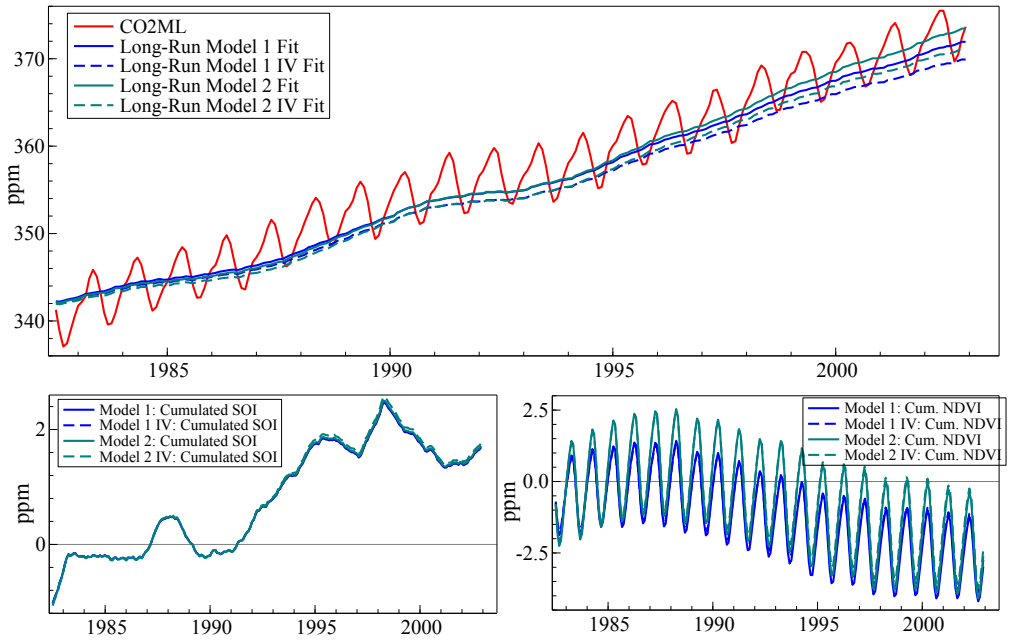
$$\mathbf{x}'_{q-s,t}\boldsymbol{\beta}_{q-s} = -0.037 \sum_{j=1}^{t} \text{IP}_{1_j} - 0.127 \sum_{j=1}^{t} \text{IP}_{2_j} + 0.097 \sum_{j=1}^{t} \text{IP}_{3_j} + 0.0018 \sum_{j=1}^{t} \text{Temp}_j \tag{76}$$

Figure 34 shows the resulting coefficient-weighted cumulative sums of vegetation (NDVI_{EUR} principal components), SOI, and the combined anthropogenic components (IP_1 to IP_3) and temperature trend for both models 1 and 2, as well as their IV estimates (dashed lines), together with the recorded level of CO_2. The industrial production components and temperature approximate the level of CO_2 well (top panel in Figure 34), marking a slight slow down in the trend around 1991–1993. The long-run trend slightly under-estimates the level of CO_2 towards the end of the sample, this likely stems from the (small) positive contribution of cumulative southern oscillation effects not included in the sum (Figure 34 bottom panels), and could also stem from the omission of Chinese production for which no data was available. A further concern, which likely acts in the opposite direction however, is the decreasing emission intensity of production (see Figure 35 in Appendix 3.8.1), a feature that is not captured in the model here due to monthly emission data not being available for countries other than the US.

Both cumulative near-stationary components vary over a small range so contribute little to the long-run changes. In line with previous studies the seasonality is predominantly driven by changes in vegetation as modelled using the NDVI series (bottom right panel in Figure 34). Using instrumental variables to control for some of the endogeneity results in slightly lower long-term trend estimates. Even though the model is estimated in net inflows to atmospheric carbon dioxide, in re-parametrized form it can explain the long-run trend—and, within the conditional model, attributes it primarily to anthropogenic emissions. This is an outcome of the data analysis alone and is not enforced.

Overall, based on *Autometrics* selection, natural factors such as vegetation, temperature, Southern Oscillation are necessary but not sufficient in explaining changes in the conditional model for atmospheric CO_2 measured at Mauna Loa. Industrial production variables are consistently selected. Most estimated effects from selected variables affect ΔCO_2 with a lag, and there seem to be few or no structural breaks in the relationships being modelled. As a further robustness check, future work could involve applying the estimation method to other measurement stations such as Barrow, Alaska (R. F. Keeling et al. 2008). Atmospheric carbon dioxide at Barrow has been measured from 1974:2 until 2007:12 and displays higher amplitude and higher autocorrelation due to its location relative to Mauna Loa.

Figure 34: Level CO_2 and cumulative sums of anthropogenic and natural factors



Using relatively few assumptions on variable relevance *a-priori*, automatic model selection with GETS modelling can provide a tool to successfully model complex relationships. Starting from a broad GUM that nests the LDGP theoretically (accounting for natural and anthropogenic sinks and sources), the analysis proceeds with an agnostic approach to determine the key factors in changes in atmospheric CO_2. Shortcomings of automatic model selection are computational issues in *Autometrics* selection because of short cuts and block partitioning. However, the terminal models appear congruent (acceptable diagnostic tests, few to no indicators or seasonal dummies selected) and are also supported by theoretical conclusions from the broader literature—natural effects are selected with the expected signs on coefficients. Key in the results is that additionally to the natural determinants selected, all terminal models include a large number of anthropogenic factors.

3.7 CONCLUSION

This study provides an initial application of modelling more variables than observations in a complex system. We identified anthropogenic contributions to atmospheric CO_2 measured at Mauna Loa using an automatic model selection algorithm. Traditionally, estimation of anthropogenic effects on carbon dioxide relied on *a priori* selection of variables (which may not be appropriate in complex relations), low-frequency measures of anthropogenic emissions, and decompositions of time series. Using *Autometrics* in a general to specific modelling approach allows for model selection with more variables than observations, stringent mis-specification testing and a more agnostic way of modelling complicated interactions. The algorithm is applied to model changes in atmospheric CO_2, encompassing other approaches by controlling for natural as well as anthropogenic sinks and sources without *a priori* restrictions of the determinants. While not completely robust to initial specification, we find that natural factors

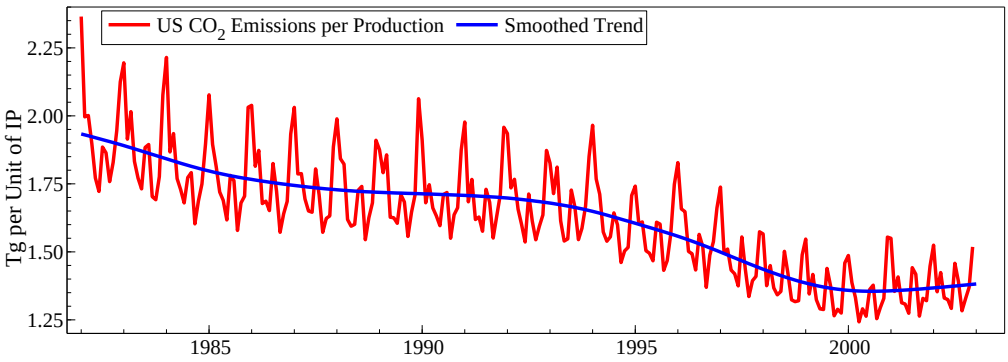
such as vegetation, temperature and Southern Oscillation are necessary, but not sufficient in explaining variation of atmospheric CO_2 in the estimated conditional model. Industrial production components measured monthly are highly significant and consistently selected in the estimated models and primarily account for the long-run trend when cumulated, whereas cumulated natural factors (except temperatures) to not trend strongly when cumulated. Producing congruent models, our methodology introduces GETS modelling through *Autometrics* as a useful tool in modelling complicated climate relationships.

3.8 APPENDIX

3.8.1 *US Emission Intensity*

Figure 35 shows the declining emission intensity of industrial production over the sample period of the estimated models.

Figure 35: US CO_2 Emission Intensity: Emissions per unit of Industrial Production



3.8.2 *GUM Specification 2: Including Annual Variables*

GUM specification 2 is estimated as a robustness check based on the inclusion of annually reported global fossil fuel CO_2 emissions. Table 19 outlines the variables included in GUM 2.

Table 19: GUM: Specification 2

Variables included	lag length
Temperature	6
NDVI PC1 Eurasia (Eur) + Winter Interaction	12
NDVI PC2 Eurasia (Eur) + Winter Interaction	12
NDVI PC3 Eurasia (Eur) + Winter Interaction	12
NDVI PC1 North America (NA) + Summer Interaction	12
NDVI PC2 North America (NA) + Summer Interaction	12
NDVI PC3 North America (NA) + Summer Interaction	12
SOI	6
Industrial Production Comp. 1	6
Industrial Production Comp. 2	6
Industrial Production Comp. 3	6
CO_2 Emissions	12
Constant	yes
Trend	yes
Centred Seasonal Variables	yes
Impulse Indicators	yes
Total variables	475

$$
\begin{aligned}
\widehat{\Delta \mathrm{CO}_2} \;=\;& \underset{(0.0525)}{0.205}\ \Delta \mathrm{CO}_{2,t-1} - \underset{(0.0487)}{0.639}\ \Delta \mathrm{CO}_{2,t-2} + \underset{(0.0479)}{0.177}\ \Delta \mathrm{CO}_{2,t-3} \\
& - \underset{(0.0321)}{0.249}\ \Delta \mathrm{CO}_{2,t-4} + \underset{(0.00998)}{0.0551}\ \mathrm{IP}_{1,t-4} - \underset{(0.0327)}{0.205}\ \mathrm{IP}_{2,t-1} \\
& + \underset{(0.0353)}{0.16}\ \mathrm{IP}_{3,t} + \underset{(0.000492)}{0.00295}\ \mathrm{Temp}_{t-4} - \underset{(0.00127)}{0.00575}\ \mathrm{SOI}_{t-5} \\
& - \underset{(0.00712)}{0.0394}\ \mathrm{NDVI}_{1,\mathrm{Eur}_{t-1}} + \underset{(0.00455)}{0.0308}\ \mathrm{NDVI}_{1,\mathrm{Eur}_{t-11}} - \underset{(0.0085)}{0.0432}\ \mathrm{NDVI}_{1,\mathrm{Eur}_{t-12}} \\
& + \underset{(0.0067)}{0.0224}\ \mathrm{w\_NDVI}_{3,\mathrm{Eur}_{t-8}}
\end{aligned} \tag{77}
$$

$$
\widehat{\sigma} = 0.212 \;\; T = 246 \;\; n = 19 \;\; \mathsf{SIC} = -2.86 \;\; \mathsf{F}_{\mathsf{ar}}(7, 226) = 1.32
$$

$$
\chi^2_{\mathsf{nd}}(2) = 1.25 \;\; \mathsf{F}_{\mathsf{reset}}(2, 231) = 0.62 \;\; \mathsf{F}_{\mathsf{arch}}(7, 232) = 0.62 \;\; \mathsf{F}_{\mathsf{het}}(26, 219) = 1.56
$$

Specification 2 in (77) covers all variables measured at a monthly frequency as well as interpolated annual components for long-term anthropogenic emissions. Lag selection for specification 2 is based on selection in specification 1. Selection in model 1 results in a maximum lag of 4 on industrial production components and 5 for temperature and SOI. The longest lag for vegetation was selected at $t - 12$. Therefore, GUM 2 starts with a maximum lag of 6 for industrial production components, and a maximum lag of 12 for vegetation. To allow for any potential lag length, annual cubic spline interpolated CO_2 emissions are included for up to 12 lags. *Autometrics* in specification 2 with 246 observations estimated 498 models reducing the number of explanatory variables from an initial 475 down to 13 in the final model. Unit roots are rejected for the model residuals and fitted values at the 1% level using ADF tests covering up to 12 lags. There is no impulse retained suggesting no major breaks or mis-specification.

Selection: the selection is highly stable relative to specification 1. Anthropogenic components, as well as temperature and Southern Oscillation are selected near-identically to model 1. No deterministic terms are selected.

Anthropogenic emissions are captured solely by the principal components for industrial production: the low-frequency fossil fuel measure is not selected. These findings suggest that high-frequency industrial production provides a better measure for anthropogenic factors than interpolated fuel emissions when working with monthly CO_2 observations. This may appear surprising given that anthropogenic emissions directly measure emitted carbon dioxide: however, this result likely stems from the annual frequency of carbon dioxide emissions that miss any seasonal component.

In terms of natural controls, the estimated coefficient on temperature is positive and that on Southern Oscillation is negative, both as in the previous model.

3.8.3 *IV Model Estimates*

Instrumental variable estimation where the temperature anomaly is instrumented using contemporaneous and lagged stratospheric aerosol time series (Sato et al. 1993).

IV MODEL ESTIMATES FOR GUM SPECIFICATION 1

$$\begin{aligned}\widehat{\Delta CO_2} = {} & \underset{(0.00138)}{0.00235}\ \text{Temp}_t + \underset{(0.0648)}{0.263}\ \Delta CO_{2,t-1} - \underset{(0.0551)}{0.697}\ \Delta CO_{2,t-2} \\ & + \underset{(0.0512)}{0.215}\ \Delta CO_{2,t-3} - \underset{(0.0396)}{0.34}\ \Delta CO_{2,t-4} - \underset{(0.0716)}{0.282}\ IP_{1,t-1} \\ & + \underset{(0.0714)}{0.34}\ IP_{1,t-4} - \underset{(0.0628)}{0.212}\ IP_{2,t-4} + \underset{(0.0523)}{0.158}\ IP_{3,t} \\ & - \underset{(0.00138)}{0.0059}\ SOI_{t-5} - \underset{(0.00899)}{0.0415}\ NDVI_{1,Eur_{t-1}} + \underset{(0.00295)}{0.0196}\ NDVI_{1,Eur_{t-10}} \\ & - \underset{(0.00684)}{0.0178}\ NDVI_{1,Eur_{t-12}} + \underset{(0.00726)}{0.0243}\ w\_NDVI_{3,Eur_{t-8}}\end{aligned}$$

IV MODEL ESTIMATES FOR GUM SPECIFICATION 2

$$\begin{aligned}\widehat{\Delta CO_2} = {} & \underset{(0.00137)}{0.00221}\ \text{Temp}_t + \underset{(0.0653)}{0.241}\ \Delta CO_{2,t-1} - \underset{(0.0577)}{0.665}\ \Delta CO_{2,t-2} \\ & + \underset{(0.0504)}{0.195}\ \Delta CO_{2,t-3} - \underset{(0.0352)}{0.26}\ \Delta CO_{2,t-4} + \underset{(0.0102)}{0.0556}\ IP_{1,t-4} \\ & - \underset{(0.0608)}{0.232}\ IP_{2,t-1} + \underset{(0.054)}{0.175}\ IP_{3,t} - \underset{(0.00139)}{0.00584}\ SOI_{t-5} \\ & - \underset{(0.00933)}{0.0377}\ NDVI_{1,Eur_{t-1}} + \underset{(0.00473)}{0.0321}\ NDVI_{1,Eur_{t-11}} - \underset{(0.00932)}{0.0428}\ NDVI_{1,Eur_{t-12}} \\ & + \underset{(0.00737)}{0.0199}\ w\_NDVI_{3,Eur_{t-8}}\end{aligned}$$

4

DETECTION OF BREAKS BY DESIGNED FUNCTIONS APPLIED TO VOLCANIC IMPACTS ON HEMISPHERIC SURFACE TEMPERATURES

Abstract

We present a methodology to detect structural breaks at any point in a time series using designed functions based on indicator saturation. Building on recent developments in econometric model selection for more variables than observations, we saturate a regression model with a full set of designed break functions. By selecting over these break functions at every point in time using a general-to-specific algorithm, we obtain unbiased estimates of the break date and magnitude. Analytical derivations for the split-sample approach are provided under the null of no breaks and the alternative of one or more breaks. Monte Carlo simulations confirm the approximate size and potency of the approach. We demonstrate the methodology by detecting volcanic eruptions in a time series of Northern Hemisphere mean temperature derived from a coupled climate simulation spanning close to 1200 years.

Breaks in time series come in many shapes, durations, and may occur at any point in time. Often an approximate shape of a break can be postulated *a-priori*, either from previous observations or theory. For example, smooth transitions are common in economic time series, while sudden drops followed by smooth reversions to the mean are typical in temperature records following a large volcanic eruption (e.g. Kelly & Sear 1984). While the approximate form of a break may be known, the timings and magnitudes of breaks are often unknown. Here we propose an approach for detecting breaks of any specified shape using an indicator saturation procedure. Our approach is based on recent developments in variable selection within regression models that involve more variables than observations (Castle et al. 2011a). By selecting over a complete set of designed break functions, our approach produces unbiased estimates of the break magnitude and timing.

A structural break is defined as a time-dependent change in a model parameter from the underlying data generating process. For example, a volcanic eruption leading to a rapid cooling corresponds to a temporary shift in the mean of the surface temperature process. The detection of structural breaks in time series has received significant attention in the recent literature. The focus has primarily remained on breaks in the mean through the form of step functions (Doornik et al. 2013, Castle et al. 2015), smooth transition functions (González & Teräsvirta 2008), breaks in regression coefficients (see e.g. Bai & Perron 1998, Bai & Perron 2003, Perron & Zhu 2005, Perron & Yabu 2009), or individual outliers or groups of outliers that can be indicative of different forms of breaks (see e.g. Hendry et al. 2008). Many of these methods, however, involve limiting prior specifications.

For instance, specific-to-general methods (e.g. Bai & Perron 2003 and the subsequent literature) are subject to an upper limit of breaks and require a minimum break length.[1] Equally, if breaks in multiple coefficients are considered, the methods impose a co-breaking restriction - different variables cannot break at different times.

Recent developments in impulse and step-indicator saturation (IIS, SIS) provide an alternative approach using a general-to-specific methodology (Doornik et al. 2013, Castle et al. 2015). By starting with a full set of step indicators and removing all but significant ones, structural breaks can be detected without having to specify a minimum break length, maximum break number or imposed co-breaking. Step-functions and impulses are nevertheless only the simplest of many potential break specifications and may not provide the closest approximation to the true underlying break. N. R. Ericsson (2012) proposes a wide range of extensions to impulse and step-shifts. Here we show that the principle of step-indicator saturation can be generalized to any form of deterministic break function. The advantage over existing methods is an expected higher frequency of detection (while the false-positive rate is controlled) when a break function approximates the true break,[2] high flexibility as multiple types of break functions can be selected over jointly, and improvements in robust forecasting where designed functions act as continuous intercept corrections.

The method is subsequently demonstrated as a way of detecting volcanic eruptions in a time series of Northern Hemisphere mean temperature from a coupled climate simulation

[1] The least-squares approach in Bai & Perron 2003 is also infeasible for outliers or volcanic breaks as considered here.

[2] For example step-indicator saturation exhibits higher power in detecting step shifts than using impulses alone – see Castle et al. (2015). This result generalizes to other forms of breaks.

spanning close to 1200 years. While a climate model with predefined forcing time series represents an ideal test bed for this approach, there is considerable uncertainty in the timing and magnitude of the climatic impact of historic volcanic eruptions over the last millennium (Schmidt et al. 2011). Our technique demonstrates that eruptions can be statistically detected without prior knowledge of their occurrence or magnitude- and hence may prove useful in estimating the past impact of volcanic events using proxy-reconstructions of hemispheric or global mean temperatures.

4.2 BREAK DETECTION USING DESIGNED FUNCTIONS

While intrinsically stochastic, without prior knowledge of the timings and magnitudes of breaks, a deterministic approach using a full set of break functions allows us to account for the underlying breaks and to model the responses deterministically. The detection of structural breaks can be formulated as a model selection problem where we select over a full set of break functions, a subset of which accurately describes the underlying true break. Consider a simple model as:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{78}$$

where $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are $(T \times 1)$ vectors, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector, and $\mathbf{Z}$ is a $(T \times k)$ matrix $\mathbf{Z} = (\mathbf{z}_1, ..., \mathbf{z}_k)$. We investigate the presence of structural breaks in any of the $\boldsymbol{\beta}$, where $\mathbf{z}$ may be a constant, trend or variable. For each break type at any point in time for each variable considered we augment the above model by a $(T \times T)$ break function matrix $\mathbf{D}$:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{79}$$

where $\boldsymbol{\gamma}$ is a $(T \times 1)$ vector. The specification of $\mathbf{D}$ is such that the first column $\mathbf{d}_1$ $(T \times 1)$ is set to denote some specified break function $d(t)$ of length L, where $d_{1,t} = d(t)$ for $t \leq L$ and 0 otherwise, $d_{1,t} = 0$ for $t > L$. All further columns $\mathbf{d}_j$ (for $j = 2, ..., T$) in $\mathbf{D}$ are set such that $d_{j,t} = d_{j-1,t-1}$ for $t \geq j$ and o otherwise. The break matrix $\mathbf{D}$ is then defined as $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_T)$, where $\mathbf{d}_j$ denotes a vector with break at time $t = j$:

$$\begin{aligned}
\mathbf{D} &= (\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_T) \\
\mathbf{d}_1 &= (d_1, d_2, ..., d_{L-1}, d_L, 0, ..., 0)' \\
\mathbf{d}_2 &= (0, d_1, d_2, ..., d_{L-1}, d_L, 0, ..., 0)' \\
\mathbf{d}_3 &= (0, 0, d_1, d_2, ..., d_{L-1}, d_L, 0, ..., 0)' \\
&\vdots
\end{aligned} \tag{80}$$

This specification provides a general framework within which multiple break types can be analysed - Table 20 provides a non-exhaustive overview.[3] The form of the break function $d(t)$ has to be designed *a-priori*, however, this is implicitly done in most structural break detection methods: for example, outlier detection through finding impulses (Impulse Indicator

[3] While the framework presented here provides an encompassing specification for many break types, the construction of $\mathbf{D}$ is not limited to this particular case. Additional sets of specifications for step shifts are considered in Doornik et al. (2013) and Castle et al. (2015). The appeal of the specification here is that the definition of $\mathbf{D}$ allows for a general framework under which properties can be analyzed where many of the previously proposed cases are a special case of $\mathbf{D}$.

Table 20: Break Function Specifications

	Break value: $d(t)$	**Length:** L
Deterministic Breaks		
General Case	$d(t)$	L
Impulses/Outliers (IIS)	1	1
Step Shifts (SIS)	1	T
Broken Trends	t	T
Volcanic Functions	see eq. (113)	3
Random Variables		
Coeff. on z_t (MIS)	$z_t \cdot d_{t,SIS}$	T

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & \dots & \dots & 0 \\ d_2 & d_1 & 0 & \dots & \dots & \vdots \\ \vdots & d_2 & d_1 & 0 & \dots & \vdots \\ d_L & \vdots & d_2 & d_1 & 0 & \vdots \\ \vdots & d_L & \vdots & d_2 & d_1 & 0 \\ 0 & 0 & d_L & d_3 & d_2 & d_1 \end{pmatrix}$$

Saturation, IIS, in Hendry et al. 2008) sets the break vector in $\mathbf{d}_1$ such that $d(t) = 1$ and $L = 1$, while a search for step-shifts (SIS) sets $d(t) = 1$ and $L = T$. Breaks in linear trends (see e.g. Perron & Yabu 2009, Perron & Zhu 2005, Estrada et al. 2013) can be constructed by setting $d(t) = t$ and $L = T$. Breaks in coefficients on random variables z_t (see e.g. N. R. Ericsson 2012, Bai & Perron 2003) can be constructed by interacting z_t with a full set of step shifts. Sudden declines followed by a smooth recovery to the mean in hemispheric temperature responses are introduced here as volcanic functions and considered in detail in sections 4.2.1 and 4.3. Linear combinations of multiple break-functions can allow for varying lengths of breaks without pre-specification.

Searching for breaks in k variables implies that the complete break matrix $\mathbf{D}$ is of dimension $(T \times kT)$. The inclusion of kT additional variables leads to the total number of variables N exceeding the number of observations, $N > T$, even for $k = 1$. Thus, a methodology allowing for more variables than observations is required.

Estimation of models with more variables than observations has relied on either shrinkage-based penalised likelihood methods (Tibshirani 1996, Zou & Hastie 2005, Tibshirani 2011) or general-to-specific methodology in the econometrics literature (see e.g. Castle et al. 2011a). Here we rely on the latter due to methods based on the forward-step-wise searches not performing as well in break detection contexts (see section 4.2.1 for a simple comparison). Cox and Snell (1974) discuss some of the challenges of the general variable selection problem and Hoover & Perez (1999) show the feasibility of general-to-specific model selection for $N << T$. When facing more variables than observations the general-to-specific approach is closely linked to robust statistics. Saturating a model with a full set of 0/1 indicator functions which are then selected over is equivalent to a robust one-step M-estimator using Huber's skip function (see Johansen & Nielsen 2009, Johansen & Nielsen 2013 for the iterated case, and Johansen & Nielsen 2014 for an overview). Here we generalize this allowing for any form of designed break function in place of impulses and formulate break detection as a model selection problem.

To estimate model (79) saturated with a full set of break functions $\mathbf{D}$ (where $N > T$) we rely on a block-partitioning estimation procedure (Doornik 2010a, Hendry & Johansen 2014). For this we partition $\mathbf{D}$ into b blocks of n_i variables such that $n_i << T$ and $\sum_{i=1}^{b} n_i = N$. We then select over break indicators in each block using a selection algorithm. The selection is repeated over the union of retained break functions. In the simplest case of testing for

a break in a single variable (e.g. the intercept), a split-half approach (see Figure 36 and *Algorithm 1* in section 4.5.4) is feasible: initially we include the first half of break indicators $\mathbf{D}_1 = (\mathbf{d}_1, ..., \mathbf{d}_{T/2})$ and retain all significant break indicators. We repeat the step for the second half of break functions d_j (for $j = T/2 + 1, ...T$) and finally combine the retained sets and only keep significant indicators. This split-half approach is considered here for analytical tractability in section 4.2.1.

In practice, however, we rely on a multi-split and multi-path search to lower the variance of the estimators, allow for any number of variables for a given set of observations, and to avoid a breakdown of the procedure if the breaks cannot be adequately modelled through split-half indicators. This is implemented here through the general-to-specific model selection algorithm *Autometrics* (*Algorithm 2* in section 4.5.4, described in Doornik 2009a, see Hendry & Pretis 2013 for an application to atmospheric CO_2 – in chapter 3). The algorithm (referred to as multi-path throughout the paper) uses a parallel step-wise backwards search, avoiding path dependence through a tree-structure. A simulation-based comparison to shrinkage methods is provided in section 4.2.1.

4.2.1 *Properties of Designed Break Functions in the Presence of Breaks*

To assess the theoretical power of the proposed methodology we first investigate the properties in the benchmark case of a single break matched by a correctly timed break indicator in section 4.2.1. Section 4.2.1 then assesses the properties of break indicator saturation when the break date and magnitude are unknown. Section 4.2.1 investigates uncertainty around the break date and 4.2.2 describes the properties in the presence of no breaks. Theory results are derived for general designed functions, simulation examples are based on a volcanic break as characterised by equation (113).

Power for Known Break Date

We investigate the theoretical power of detecting a break in a time series given a known break date. Consider a data generating process (DGP) and model for a single known break with an intercept:

$$y_t = \mu + \lambda_1 d_t + \epsilon_t \tag{81}$$

where $\epsilon_t \sim \text{iid } \mathrm{N}(0, \sigma_\epsilon^2)$. The break shifts μ to $\mu + \lambda_1 d_t$ where d_t is a break function of length L beginning at time $t = T_1$ where $T_1 + L \leq T$ such that $d_t \neq 0$ for $T_1 \leq t < T_1 + L$ and 0 otherwise. The estimators $\widehat{\mu}$ and $\widehat{\gamma}$ (where $\widehat{\gamma}$ is the estimator for λ_1) on a correctly specified model for a known break are given by:

$$\begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\gamma} - \lambda_1 \end{pmatrix} = \begin{pmatrix} T_d^{-1} \left(\sum_{t=T_1}^{T_1+L-1} d_t^2 \sum_{t=1}^{T} \epsilon_t - \sum_{t=T_1}^{T_1+L-1} d_t \sum_{t=T_1}^{T_1+L-1} d_t \epsilon_t \right) \\ T_d^{-1} \left(\sum_{t=T1}^{T_1+L-1} d_t \epsilon_t - \sum_{t=T1}^{T_1+L-1} d_t \sum_{t=0}^{T} \epsilon_t \right) \end{pmatrix} \tag{82}$$

where $T_d = T \left[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \frac{1}{T} \left(\sum_{t=T_1}^{T_1+L-1} d_t \right)^2 \right]$. The estimators are unbiased for the break and intercept: $E[\widehat{\mu} - \mu] = 0$ and $E[\widehat{\gamma} - \lambda_1] = 0$. For the special case when step-indicators are

chosen as the functional form of d_t and the single break lasts from $t = 0$ to $t = T_1$, equation (82) simplifies[4] to $\widehat{\mu} - \mu = \bar{\epsilon}_2$ and $\widehat{\gamma} - \lambda_1 = \bar{\epsilon}_1 - \bar{\epsilon}_2$.

The variance of the estimators is given by:

$$V\begin{pmatrix}\widehat{\mu}-\mu\\ \widehat{\gamma}-\lambda_1\end{pmatrix} = \sigma_\epsilon^2 T_d^{-1}\begin{pmatrix}\sum_{t=T_1}^{T_1+L-1} d_t^2 & -\sum_{t=T_1}^{T_1+L-1} d_t\\ -\sum_{t=T_1}^{T_1+L-1} d_t & T\end{pmatrix} \tag{83}$$

The distribution of the break estimator is then:

$$(\widehat{\gamma}-\lambda_1) \sim N\left(0, \sigma_\epsilon^2\left[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \sum_{t=T_1}^{T_1+L-1} d_t\bar{d}\right]^{-1}\right) \tag{84}$$

where $\bar{d} = 1/T\sum_{t=1}^{T} d_t$. The associated t-statistic is:

$$t_{\widehat{\gamma}} = \frac{\widehat{\gamma}-\lambda_1}{\sqrt{\widehat{V}(\widehat{\gamma})}} = \frac{(\widehat{\gamma}-\lambda_1)\sqrt{\left(\sum_{t=T_1}^{T_1+L-1} d_t^2 - \sum_{t=T_1}^{T_1+L-1} d_t\bar{d}\right)}}{\widehat{\sigma_\epsilon}} \tag{85}$$

Ignoring the estimation uncertainty of $\widehat{\sigma}_\epsilon$, the t-statistic for a test of the hypothesis: $\lambda_1 = 0$, therefore has the approximate normal distribution:

$$t_{\widehat{\gamma}} \approx \frac{\widehat{\gamma}\sqrt{\left(\sum_{t=T_1}^{T_1+L-1} d_t^2 - \sum_{t=T_1}^{T_1+L-1} d_t\bar{d}\right)}}{\sigma_\epsilon} \sim N(\psi, 1) \tag{86}$$

where ψ is the non-centrality of the t-statistic:

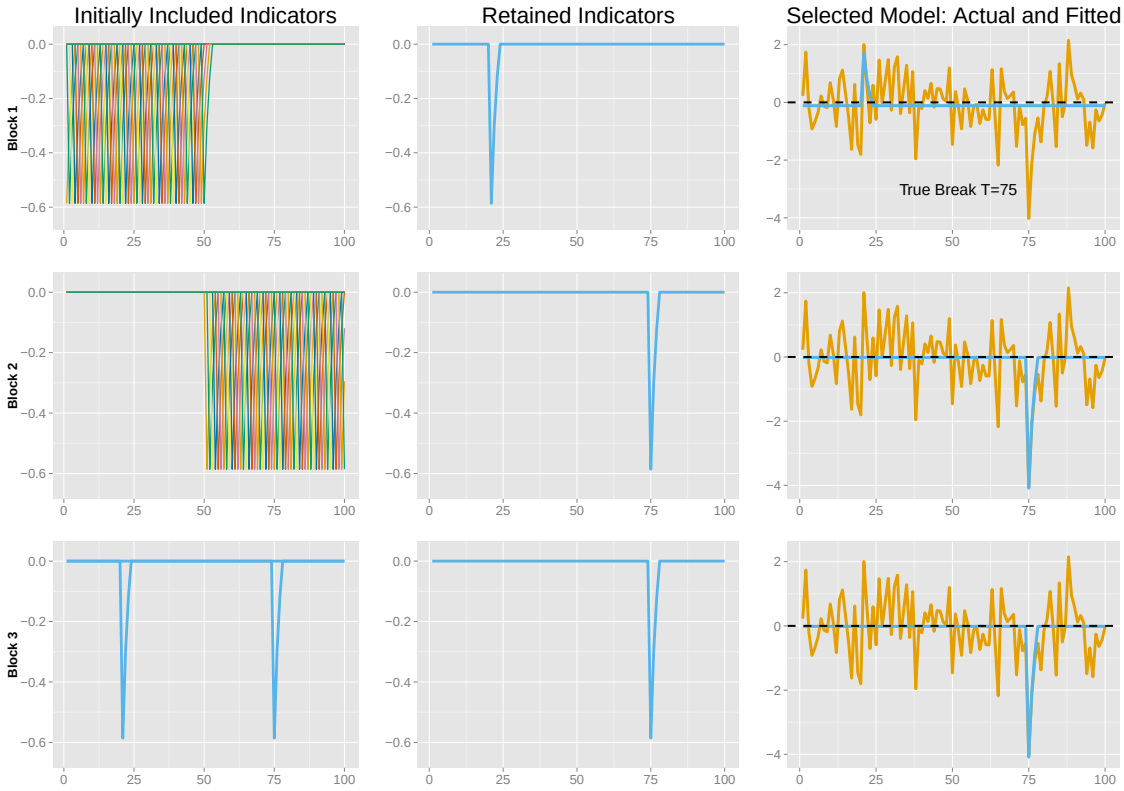
$$\psi = E\left[t_{\widehat{\gamma}}\right] = \frac{\lambda_1\sqrt{\left(\sum_{t=T_1}^{T_1+L-1} d_t^2 - \left(\sum_{t=T_1}^{T_1+L-1} d_t\right)^2 \frac{1}{T}\right)}}{\sigma_\epsilon} \tag{87}$$

Using the Cauchy-Schwarz inequality, the multiplying term $\sum_{t=T_1}^{T_1+L-1} d_t^2 - \left(\sum_{t=T_1}^{T_1+L-1} d_t\right)^2 \frac{1}{T}$, a function of the break form d_t and length L, will be positive. The non-centrality ψ increases in the break magnitude λ_1, varies with break length L and will depend on the underlying break function given by d_t. However, given the pre-specification of the break function, the theoretical power can be computed prior to estimation. The asymptotic distribution of the break estimator will be subject to further research to determine necessary conditions on the designed break functions and appropriate scaling to yield non-degenerate limit distributions (e.g. for linear trends or more complex specifications – see Phillips 2007).[5]

[4] See supplementary material 4.5.3 for proof. Where $\bar{\epsilon}_1 = 1/T_1\sum_{t=1}^{T_1}\epsilon_t$ and $\bar{\epsilon}_2 = 1/(T-T_1)\sum_{t=T_1+1}^{T}\epsilon_t$.

[5] The asymptotic rates of convergence will generally depend on the specification of the break function – varying scaling to obtain non-degenerate limit distributions may therefore be required. In the case of step functions ($d_t = 1, L = T$) and the simple no-intercept case, pre-multiplying the estimator by $\sqrt{T}$ yields asymptotic normality for the break estimator when $T^{-1}\sum_{t=T_1}^{T_1+L-1} d_t^2 = T^{-1}L \to \tau$ as $T \to \infty$. In other words, the ratio of break length to the sample size remains constant as the sample size increases – this can be interpreted as obtaining more information on the break period or sampling at higher frequencies as $T \to \infty$. A similar analysis can be applied to the volcanic functions considered here, where either the break length scales with the sample size, or alternatively the magnitude increases similar to the asymptotic analysis for a single impulse in Doornik et al. (1998).

Figure 36: Split-half approach for a single unknown break of the shape of a volcanic function at T=75. Left column shows included indicators in each step, middle column shows the retained indicators, right column graphs the selected model with actual and fitted data. Block 1 (top panel) includes the first half of break functions, block 2 (middle panel) includes the second half, while block 3 uses the union of retained indicators from blocks 1 and 2. Using a saturating set of break functions the break at T=75 is detected without prior knowledge and is the only break function retained.



Power for an Unknown Break Date

When the break date is unknown, we saturate the regression using a full set of specified break indicators and select significant breaks through a general-to-specific algorithm. Here we assess the feasibility of the proposed method by deriving the analytical power associated with the split-half approach for an unknown break. Figure 36 illustrates the split-half method for a single unknown break. In practice we rely on a multi-path, multi-block search algorithm (such as *Autometrics*, see *Algorithm 2* in section 4.5.4) to reduce the variance of the estimators.

Consider a single break falling into the first half of the sample beginning at time T_1 for L periods such that $0 < T_1 < T_1 + L < T/2$. In matrix form the DGP is given as:

$$\mathbf{y} = \lambda_1 \mathbf{d}_{T_1} + \boldsymbol{\epsilon} \tag{88}$$

where $\boldsymbol{\epsilon} \sim \text{iid } \mathsf{N}(\mathbf{0}, \sigma_\epsilon^2)$ for simplicity and the $(T \times 1)$ vector $\mathbf{d}_{T_1}$ denotes a break at $t = T_1$ for L periods. We specify the complete $(T \times T)$ matrix of break indicators as $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_T)$. Using a split-half approach, we assess the power and properties of detecting the single break when the break date is unknown. The split-half model for the first half of break functions is given by:

$$\mathbf{y} = \mathbf{D}_1 \gamma_{(1)} + \mathbf{v} \tag{89}$$

where $\gamma_{(1)} = (\gamma_1, \gamma_2, ...\gamma_{T/2})'$ and $\mathbf{D}_1 = (\mathbf{d}_1, ..., \mathbf{d}_{T/2})$. The estimator $\widehat{\gamma}_{(1)}$ equals:[6]

$$\begin{aligned} \widehat{\gamma}_{(1)} &= \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\mathbf{y} = \lambda_1\left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\mathbf{d}_{T_1} + \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\boldsymbol{\epsilon} \\ &= \lambda_1\mathbf{r} + \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\boldsymbol{\epsilon} \end{aligned} \tag{90}$$

where the $(T/2 \times 1)$ vector $\mathbf{r}$ is equal to one at $t = T_1$ and zero otherwise, $r_t = 1_{\{t=T_1\}}$. It follows that $E[\widehat{\gamma}_{(1)}] = \lambda_1\mathbf{r}$ and $V[\widehat{\gamma}_{(1)}] = \sigma_\epsilon^2\left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}$. We find for the first half:

$$\left(\widehat{\gamma}_{(1)} - \lambda_1\mathbf{r}\right) \sim N\left(\mathbf{0}, \sigma_\epsilon^2\left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\right) \tag{91}$$

The estimator $\widehat{\gamma}_{(2)}$ on the second half of indicators will miss the break in the first half and equals:

$$\widehat{\gamma}_{(2)} = \lambda_1\left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1}\mathbf{D}_2'\mathbf{d}_{T_1} + \left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1}\mathbf{D}_2'\boldsymbol{\epsilon} \tag{92}$$

For step-shifts, Castle et al. (2015) show that the indicator in $\mathbf{D}_2$ closest to the sample split will be retained in the second set of indicators. For the general form of break functions, retention in $\mathbf{D}_2$, when there is a break in the first half, will depend on the specific functional form. However, conditional on the break indicator being correctly retained in the first set $\mathbf{D}_1$, retention of irrelevant indicators in $\mathbf{D}_2$ does not affect the correct identification of the break overall: let $\mathbf{D}_{1*}$ and $\mathbf{D}_{2*}$ denote the set of retained break functions in the first and second set respectively, where retention is based on a retention rule such as $\mathbf{d}_j$ is retained if $|t_{\widehat{\gamma}_j}| \geq c_\alpha$. The final step in the split-half procedure is then to combine the retained indicators using $\mathbf{D}_U = [\mathbf{D}_{1*}\mathbf{D}_{2*}]$ and estimate the model:

$$\mathbf{y} = \mathbf{D}_U\gamma_{(U)} + \mathbf{v} \tag{93}$$

This yields the estimator $\widehat{\gamma}_{(U)}$ unbiased for the true break:

$$\widehat{\gamma}_{(U)} = \lambda_1\mathbf{r} + \left(\mathbf{D}_U'\mathbf{D}_U\right)^{-1}\mathbf{D}_U'\boldsymbol{\epsilon} \tag{94}$$

The carried-forward break function in $\mathbf{D}_{1*}$ correctly identifies the true break, and coefficients on all other break functions will thus be zero in expectation. The proof is identical to that given for the first half of indicators in appendix 4.5.2. This shows that, conditional on retaining the correct break indicator in $\mathbf{D}_1$, the retention of indicators in $\mathbf{D}_2$ does not affect the correct identification of the break, when the first and second set are combined and re-selected over. The distribution of the final split-half estimator is then given by:

$$\left(\widehat{\gamma}_{(U)} - \lambda_1\mathbf{r}\right) \sim \mathrm{N}\left(\mathbf{0}, \sigma_\epsilon^2\left(\mathbf{D}_U'\mathbf{D}_U\right)^{-1}\right) \tag{95}$$

Re-selection then results in only the true break indicator being retained in expectation.[7]

This result generalizes the specific case of step indicators presented in Doornik et al. (2013) and Castle et al. (2015). Even though the break date and magnitude are unknown, the use of a fully saturated set of break indicators allows us to obtain an unbiased estimate of the break

[6] Proof given in appendix 4.5.2.

[7] The split-half approach is not the only way of analysing the theory of indicator saturation: rather than splitting the functions into a first and second half, alternatively one could consider including every other break function in two sets such that $\mathbf{D}_1$ covers breaks at $t = 1, 3, 5...$ and $\mathbf{D}_2$ covers breaks at $t = 2, 4, 6....$ Retention frequencies in this setup can be derived using the results in section 4.2.1.

magnitude and timing. The estimator then follows an approximate normal distribution subject to correct specification of the break function. Thus the estimated coefficient at the precise break time, $\widehat{\gamma}_{T_1}$, is in expectation equal to the break magnitude, while all other estimated coefficients are zero in expectation. This result generalizes to multiple breaks falling in a single split. As in the case of the known break timing, the variance of the estimator depends on the specified break function. Let $\delta_{k,j}$ denote the (k, j) element of the matrix $(\mathbf{D}_1'\mathbf{D}_1)^{-1}$. The variance of the coefficient at the breakpoint in the first half is therefore:

$$V[\widehat{\gamma}_{T_1}] = \sigma_\epsilon^2 \delta_{T_1,T_1} \tag{96}$$

For iid error terms $\boldsymbol{\epsilon}$ and $\mathbf{D}$ specified as a full set of step functions then $\delta_{j,j} = 2$. The break coefficient has twice the error variance. For the proposed volcanic function (derived and assessed in detail in section 4.3) modelling a single drop followed by a reversion to the mean we find that $\delta_{j,j} = 3.7$, thus $V[\widehat{\gamma}_{T_1}] = 3.7\sigma^2$. This can be compared to the known-break/single-indicator case where the variance is given by equation (83) and for the volcanic function equals $2.3\sigma^2$ (for T=100). Due to collinearity of break functions, the variance of the estimator is higher in a fully saturated model. In the more general case, δ_{T_1,T_1} depends on the specification of the break function but can be computed *a-priori*. The *t*-statistic is then given as:

$$t_{\widehat{\gamma}_{T_1}} = \frac{\widehat{\gamma}_{T_1}}{\widehat{\sigma}_\epsilon \sqrt{\delta_{T_1,T_1}}} \approx \frac{(\widehat{\gamma}_{T_1} - \lambda_1)}{\sigma_\epsilon \sqrt{\delta_{T_1,T_1}}} + \frac{\lambda_1}{\sigma_\epsilon \sqrt{\delta_{T_1,T_1}}} \sim \mathrm{N}\left(\frac{\lambda_1}{\sigma_\epsilon \sqrt{\delta_{T_1,T_1}}}, 1\right) \tag{97}$$

In practice we use sequential elimination of the break indicators or a multi-path search to eliminate insignificant indicators to reduce the variance of the estimators from a saturated model (96) closer to the single break (83) and increase the power of detection.

For dynamic time series models the above approach can be simply extended by including time-dependent covariates. Valid conditioning (e.g. through the inclusion of auto-regressive terms in the case of non-iid errors) can be ensured by always including the covariates in each block estimation step and only selecting over the break functions. Johansen & Nielsen (2009) provide the asymptotics under the null of no break for the special case of impulses as break functions for stationary and unit-root non-stationary autoregressive processes (see Johansen & Nielsen 2013 for the iterated version). The case for general break functions is discussed in section 4.2.2 and section 4.5.5 provides simulation results for an AR(1) model and DGP.

SIMULATION PERFORMANCE BASED ON VOLCANIC BREAK FUNCTION: Table 21 reports simulation results (T=100) for a DGP with a single unknown shift (a volcanic break) at $t = T_1$ of magnitude λ (over the entire break length in standard deviations (sd) of the error term, this implies a trough of 0.58λ) followed by a smooth reversion to the mean over a total length $L = 3$ (equation (113) provides the exact functional form). Simulations are assessed based on the retention frequency (potency) for a single break and average retention of spurious breaks

(gauge).[8] Let $\widehat{\gamma}_j$ denote the coefficient of the selected break function j in replication i. For M replications, $i = 1, 2, ..., M$, the retention frequency is defined as:

$$\text{Retention: } p_{T_1} = \frac{1}{M} \sum_{i=1}^{M} 1_{\{\widehat{\gamma}_{i,T_1} \neq 0\}} \tag{98}$$

The false-positive rate or gauge, the model selection equivalent of size, measures the fraction of retained irrelevant variables (i.e. in the case of impulses gauge measures the expected frequency of falsely detected outliers, see Johansen & Nielsen 2014 for distributional results). For a single break at $t = T_1$ with the correct break function of $d_j = T_1$ there are $T - 1$ irrelevant break indicators:

$$\text{Gauge} = \frac{1}{T-1} \frac{1}{M} \sum_{j=1, j \neq T_1}^{T} \sum_{i=1}^{M} 1_{\{\widehat{\gamma}_{i,j} \neq 0\}} \text{ for j} \neq T_1 \tag{99}$$

Low gauge (close to zero) and high retention (close to 1) are preferred.

The trade-off between potency and level of significance of selection α is shown in Figure 37 for a single volcanic break.[9] A multi-path search generally increases the power of detection relative to the split-half approach. Figure 38 shows the results for split-half (dashed) and multi-path (solid) selection when using volcanic functions for a break of $\lambda = 6$. Consistent with derived theory (96), the estimator has 3.7 times the variance of the error term when using split-half estimation for the given function. Using a multi-path search reduces the variance drastically. Any selection bias of the multi-path search estimates can be controlled through bias correction after selection (see Castle et al. 2011a). Appendix 4.5.5 provides simulation results for a simple autoregressive DGP and model.

Table 21: Potency of detecting an unknown break when using split-half and multi-path searches. Statistics were generated from 1000 simulations and detection significance was set to $\alpha = 0.01$, with a length of $L = 3$. Break magnitude λ corresponds to the full response in standard deviations of the error term ($\sigma = 1$) over the entire break, the trough is 0.58λ.

	Split-Half		Multi-Path	
	Potency	Gauge $\mathbf{D}_1$	Potency	Gauge
$\lambda = 6$, trough=3.48	0.69	0.013	0.88	0.015
$\lambda = 4$, trough=2.23	0.30	0.013	0.50	0.014
$\lambda = 2$, trough=1.16	0.06	0.013	0.11	0.015

COMPARISON TO SHRINKAGE-BASED METHODS Shrinkage-based methods using penalised likelihood estimation (Tibshirani 2011, Zou & Hastie 2005) provide an alternative to the general-to-specific algorithm used here in selecting models with more variables than observations. Figure 39 shows the simulation outcomes comparing multi-path indicator saturation (for $\alpha = 0.01$), the LASSO (Tibshirani 1996, estimated using LARS, see Efron et al. 2004) where cross-validation is used to determine the penalisation parameter, and the LASSO where the penalisation parameter is set such to approximate the false-positive rate of the IS procedure

[8] All simulations and applications using the multi-path search *Autometrics* are coded using the *Ox* programming language (Doornik 2009b). Simulations using the LASSO are coded using *R*.

[9] For a volcanic break, λ denotes the entire temperature response over the specified length L, thus the trough will be less than λ. For the present specification of $L = 3$, the initial trough of the function equals 0.58λ.

Figure 37: Left: Potency of detecting a volcanic break of magnitude λ for level of significance α using split-half and multi-path selection. Right: Proportion of spuriously retained break indicators (gauge). Break magnitude λ corresponds to the full response in standard deviations of the error term ($\sigma = 1$) over the entire break, the peak is 0.58λ.

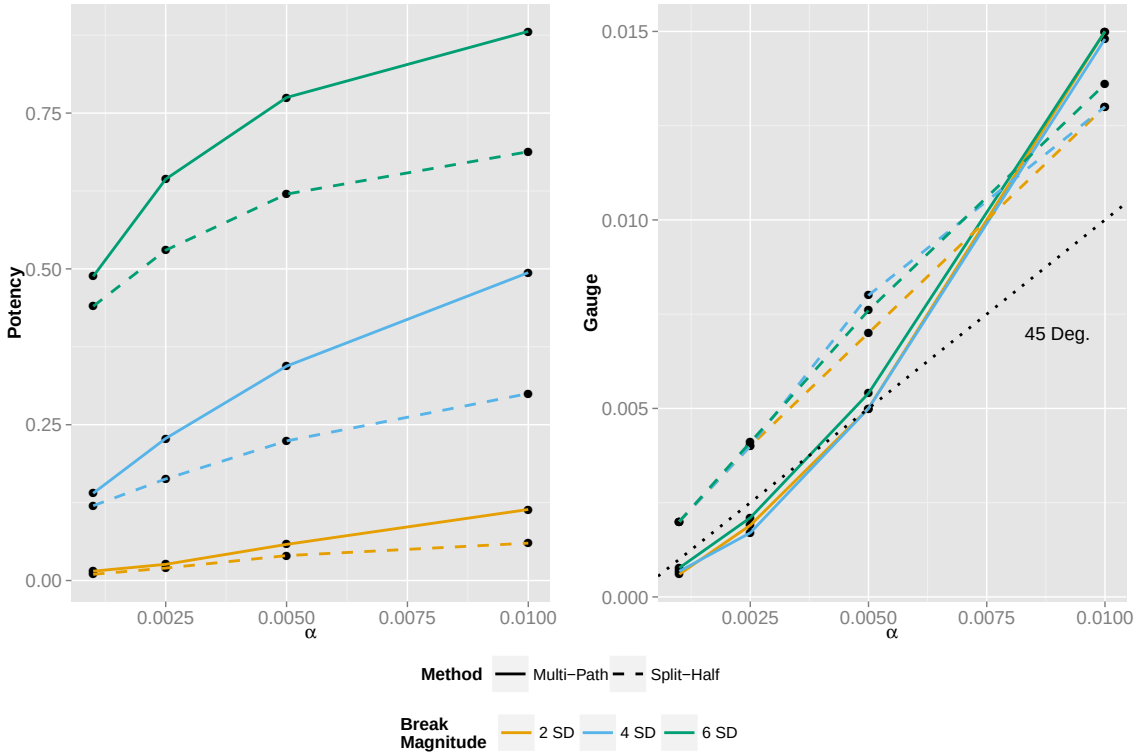


Figure 38: Estimated break indicator and variance for unknown break using split-half (blue) and multi-path (yellow) selection. Left panel shows a simulated time series, middle panel the distribution of the estimated coefficient, right panel shows the variance of the coefficient. Dashed lines show the true break magnitude and analytical variance of the split-half coefficient.

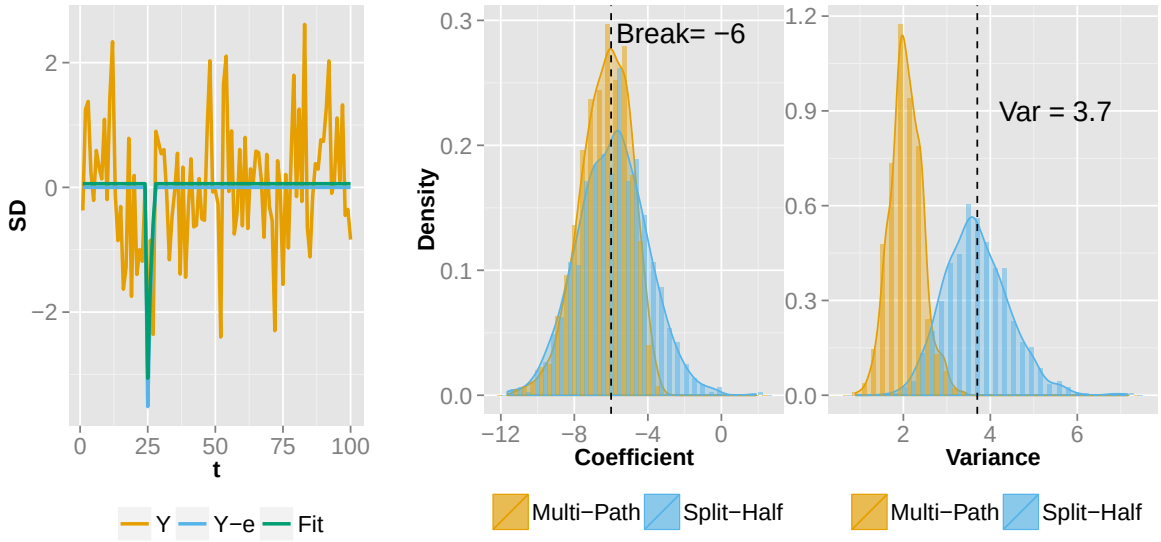
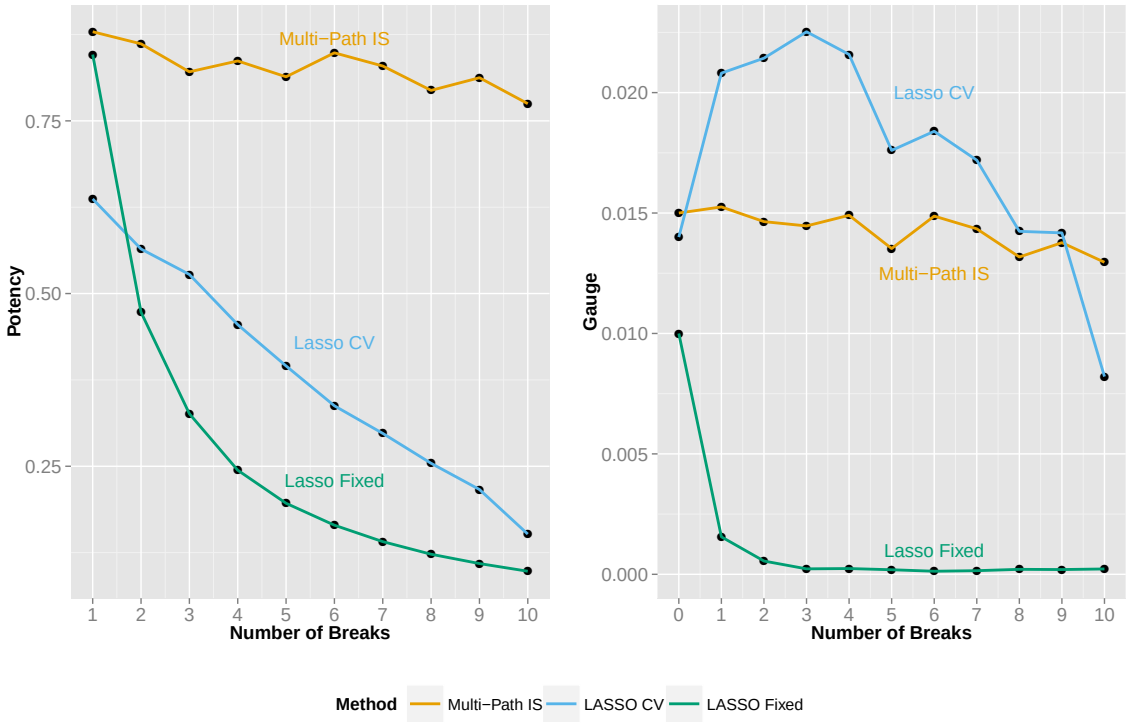


Figure 39: Left: Average potency of detecting increasing number of volcanic breaks using Multi-Path Indicator Saturation (IS, orange), Cross-Validated Lasso (blue), and Lasso with fixed penalisation parameter (green) where the penalisation parameter is set such that the false-positive rate approximates that of the indicator saturation procedure under the null of no break. Right: Corresponding false-positive rate (gauge). $M = 1000$ replications.



under the null of no breaks (≈ 0.01). The simulation uses a total break magnitude of 6 standard deviations (implying a peak of 3.48 sd) for an increasing number of evenly spaced breaks from 0 up to 10 in a sample of $T = 100$. The general-to-specific multi-path algorithm exhibits stable power exceeding that of the penalised likelihood methods across any number of breaks. The false-positive rate remains stable and close to the theory level of 0.01. The shrinkage-based procedures, due to their similarity to forward-selection, show decreasing potency (which lies consistently below that of the multi-path algorithm) as the number of breaks increases, and the false-positive rate is difficult to control.

Uncertainty on the break date

An estimated uncertainty on the break magnitude is simple to compute given the approximate distribution of the coefficient on the break functions. While of considerable interest, it is non-trivial, however, to quantify the uncertainty around the timing of the break. As Elliott & Müller (2007) point out, it is easier to differentiate between models with and without breaks than it is to estimate the uncertainty around the estimate of break dates. This is particularly true for the literature focusing on break detection using general to specific methodology. Here we provide an initial attempt at investigating the uncertainty around the timing of estimated break points when using break indicator saturation. We compute the analytical power of break indicators when the break function is correctly specified but the break time is not.

Consider a simple DGP without an intercept and a single structural break in the mean:

$$y_t = \lambda_1 d_{T_1,t} + \epsilon_t \tag{100}$$

The break shifts $E[y_t]$ from 0 to $\lambda_1 d_{T,1}$ at $t = T_1$ where d_{T_1} is a break function of length L beginning at time $t = T_1$ such that $T_1 + L < T$ and $d_{T_1} = (0, ..., d_1, d_2, ..., d_L, 0, ..., 0)$. The corresponding model is then:

$$y_t = \widehat{\gamma} d_{j,t} + v_t \tag{101}$$

As a benchmark we consider the case when the break date is correctly specified, this means that $d_{j,t} = d_{T_1,t}$. When the break is correctly specified the estimator for λ_1 is given by:

$$\widehat{\gamma}_{t=T_1} - \lambda_1 = \left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2 \right)^{-1} \left(\sum_{t=T_1}^{T_1+L} d_{T_1,t} \epsilon_t \right) \tag{102}$$

Similarly for a test of the hypothesis: $\lambda_1 = 0$, the t-statistic has an approximate non centrality of $E\left[t_{\widehat{\gamma},t=T_1}\right] = \psi = \frac{\lambda_1 \sqrt{\left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2\right)}}{\sigma_\epsilon}$ and the approximate normal distribution:

$$t_{\widehat{\gamma}} \approx \frac{\widehat{\gamma}_{t=T_1} \sqrt{\left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2\right)}}{\sigma_\epsilon} \sim \mathrm{N}(\psi, 1) \tag{103}$$

The non-centrality ψ increases in the break magnitude λ_1, varies with the break length L and will depend on the underlying break function given by d_t.

Now consider the model being incorrectly specified for the break date, such that $d_{j,t} \neq d_{T_1,t}$ and is shifted by K periods $d_{j,t} = d_{T_1 \pm K,t}$. The estimator for λ_1 is then:

$$\widehat{\gamma}_{t=T_1 \pm K} - \lambda_1 = \lambda_1 \left[\left(\sum_{t=T_1}^{T_1+L} d_{j,t}^2 \right)^{-1} \left(\sum_{t=T_1}^{T_1+L} d_{j,t} d_{T_1,t} \right) - 1 \right] + \left(\sum_{t=T_1}^{T_1+L} d_{j,t}^2 \right)^{-1} \left(\sum_{t=T_1}^{T_1+L} d_{j,t} \epsilon_t \right) \tag{104}$$

It follows that $\widehat{\gamma}_{t \neq T_1}$ is not an unbiased estimator for λ_1. Note that if d_j is functionally specified correctly such that the only difference to the true break function is through K lags: $\mathbf{d}_j = \mathbf{d}_{T_1 \pm K}$, then it holds that $\left(\sum_{t=T_1}^{T_1+L} d_{j,t}^2\right) = \left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2\right)$. Equally it holds that $\left(\sum_{t=T_1}^{T_1+L} d_{j,t} d_{T_1,t}\right) = \left(\sum_{t=T_1}^{T_1+L} d_{T_1 \pm K,t} d_{T_1,t}\right)$ for $K \leq L$ and 0 for $K > L$. Using this we derive an expression for the approximate t-statistic associated with the estimator given a break function time mis-specified by K lags:

$$E\left[t_{\widehat{\gamma},t=T_1 \pm K}\right] \approx \frac{E\left[\widehat{\gamma}_{t=T_1 \pm K}\right]}{\sigma_\epsilon \left(\sum_{t=T_1}^{T_1+L} d_{T_1}^2\right)^{-1/2}} = \frac{\lambda_1 \left(\sum_{t=T_1}^{T_1+L} d_{T_1 \pm K} d_{T_1}\right)}{\sigma_\epsilon \left(\sum_{t=T_1}^{T_1+L} d_{T_1}^2\right)^{1/2}} \tag{105}$$

This is equal to the non-centrality of the correct break date ψ scaled by a factor less than one, decreasing with the distance K from the correct date:

$$E\left[t_{\widehat{\gamma},t=T_1 \pm K}\right] \approx \psi \left(\frac{\sum_{t=T_1}^{T_1+L} d_{T_1 \pm K,t} d_{T_1,t}}{\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2} \right) \leq \psi \tag{106}$$

For a given break specification d_t and break length L, the corresponding power function can be computed to provide an approximate measure of power for detection of a break at $t = T_1$ in the neighbourhood of T_1. Note that $E\left[t_{\widehat{\gamma},t=T_1\pm K}\right]$ is zero outside a neighbourhood of L. The associated t-statistic of a break indicator further away from the true break date T_1 than the break length L is zero in expectation, since $\left(\sum_{t=T_1}^{T_1+L} d_{j,t}d_{T_1,t}\right) = 0$ for $K > L$. Intuitively, longer breaks increase the likelihood that a break indicator that is not perfectly coincident with the break date will appear significant, and we can expect the retention to be equal to the nominal significance level outside a $t = T_1 \pm L$ interval.

As before we consider the special case of volcanic functions and also provide results from step shifts for comparison. Figure 40 shows the analytical as well as simulated non-centrality and power around a true break date at t=26 of length $L = 3$ for $\alpha = 0.05$. The Monte Carlo simulations match the theoretical power and non-centralities closely.

For no break, the analytical power is uniform and equal to the nominal significance level. When there is a break, outside of the interval $T_1 \pm L$ the expected retention of the break indicator equals the nominal significance level. For a step shift, given (106), the non-centrality decreases linearly as the numerator falls by $1/L$ per shifted period relative to the correct break date. For longer breaks this implies that the power around the true break date is close to uniform. In the case of volcanic functions, due to the particular functional form, the power and retention probability drop more rapidly and peak clearly around the true break date. The special case presented here only considers the properties of a single time-mis-specified indicator in the model. When selecting from a full set of break functions (see section 4.2.1) it is less likely that a break function at $T_{1\pm K}$ appears significant if the correct T_1 indicator is included in the same model, a mis-timed indicator in a fully saturated model would likely appear significant only if a chance draw of the error offsets the shift.

4.2.2 *Properties under the Null of No Break*

Under the null hypothesis when there are no breaks in the DGP, there are two primary concerns regarding the inclusion of a full set of break functions in the statistical model. First, break indicators may be retained spuriously, and second, there may be concerns about the effect on the distribution of coefficients on variables that are known to be relevant when including a full set of break functions - in other words, whether saturating a model with irrelevant variables affects relevant ones.

First, we consider the spurious retention of break indicators. Under the null of no breaks, $\lambda_1 = 0$, the DGP from before (88) is given by:

$$\mathbf{y} = \boldsymbol{\epsilon} \tag{107}$$

Based on the above results, when using a split-half approach with a full set of break indicators, the expectation of the estimated coefficients in the first half is given by:

$$E[\widehat{\gamma}_{(1)}] = E\left[\left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\boldsymbol{\epsilon}\right] = \mathbf{0} \tag{108}$$

The same result generalizes to the union of retained indicators $\mathbf{D}_U$. Thus, the t-statistics of the included break functions will be centred around zero in expectation when there is no break. Using the selection rule that retains the break function $\mathbf{d}_j$ if $|t_{d_j}| > c_\alpha$, then $\alpha T/2$ indicators

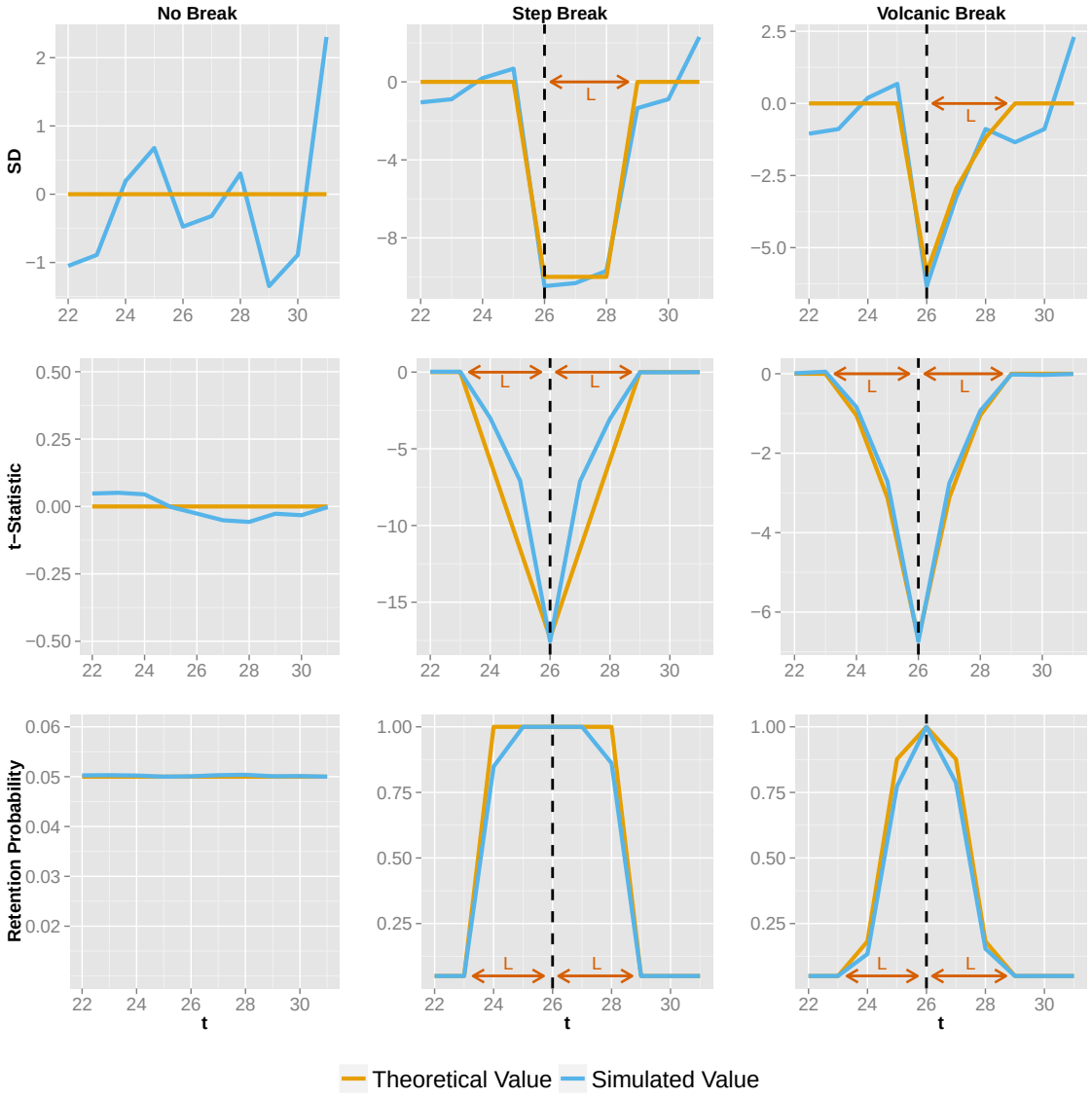


Figure 40: Power and Retention Frequency around the Break Date where the timing of the break functions is imposed without selection: Simulated data with and without shifts (top), associated non-centrality and simulated t-statistics (middle), analytical and simulated power (bottom) around break $\lambda = -10$ at $T = 26$ of length $L = 3$ and interval $T \pm K$ for $\alpha = 0.05$. Left shows no break, middle a step-break and right panel a volcanic function break. Analytical non-centralities and power are shown in yellow, simulated t-statistics and retention are shown in blue. Dashed lines mark the break occurrence. Outside of an interval $T = 26 \pm L$ the retention probability and analytical power are equal to the nominal significance level $= \alpha = 0.05$.

will be retained on average in each half. Combining the retained indicators in the final set, αT indicators are retained in expectation. The proportion of spurious indicators can thus be easily controlled through the nominal significance level of selection. The properties under the null are confirmed below using Monte-Carlo simulations.

Table 22 and Figure 41 report the simulation results when there are no breaks in the DGP but a full set of break functions (of the form of volcanic functions) is included. When using a split-half approach with a 1-cut variable selection decision based on the absolute t-statistic the proportion of irrelevant retained indicators is close to the nominal significance level. In practice when using a multi-path, multi-split procedure (here implemented through *Autometrics* in Doornik & Hendry 2009 and Doornik 2006) the false-positive rate is close to the nominal significance level for low levels of α. A conservative approach (low $\alpha \leq 1\%$) is recommended in practice.[10] When compared to results in Doornik et al. 2013 and Castle et al. 2015, there is little notable difference between different specifications of break functions, consistent with the analytical properties of irrelevant indicators.

Table 22: Retention of spurious volcanic break functions when there is no break

Significance Level	Split-Half One-Cut		Multi-Path Search
	Gauge $\mathbf{D}_1$	Gauge $\mathbf{D}_2$	Gauge $\mathbf{D}$
$\alpha = 0.05$	0.056	0.054	0.30
$\alpha = 0.01$	0.013	0.012	0.015
$\alpha = 0.005$	0.007	0.007	0.005
$\alpha = 0.0025$	0.004	0.004	0.002
$\alpha = 0.001$	0.002	0.002	0.001

We now assess the second consideration, which is the effect of including a full set of break indicators when theory variables $\mathbf{X}$ are included in the model and not selected over. These could include contemporaneous covariates or autoregressive dynamic variables. For the specific case when the elements of $\mathbf{D}$ are specified to be impulse indicators, Johansen & Nielsen (2009) derive the asymptotic distribution of β in the full split-half approach in stationary and unit-root non-stationary regressions using the equivalence of IIS and one step Huber-skip M estimators. For an iterated procedure (e.g. resembling the multi-block approach in *Autometrics*) the distributional results under the null for IIS are derived in Johansen & Nielsen (2013). For the general form of designed functions we follow theory for the sub-steps of split-half estimation where N<<T in each step, and appeal to simulation results for the overall algorithm. Consider a simple DGP:

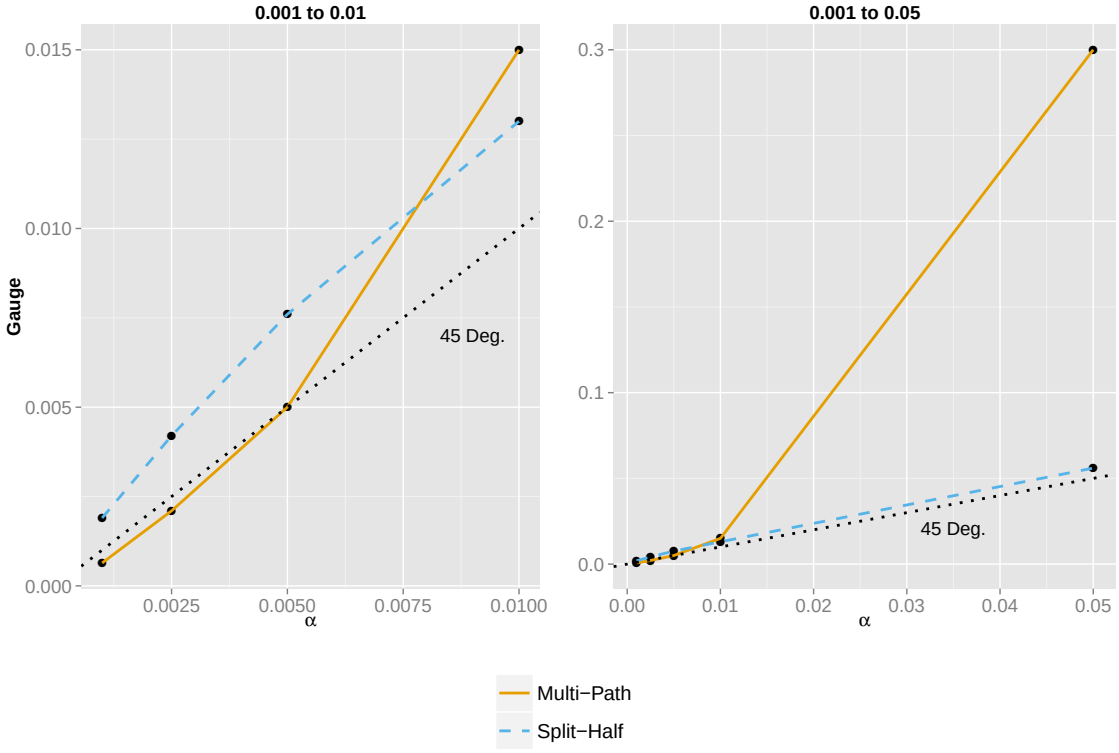
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{109}$$

where $\epsilon \sim \text{iid}(0, \sigma_\epsilon^2)$ and the elements of $\mathbf{X}$ (dynamic or static) are assumed to be relevant and not selected over. The model relying on the split-half approach saturated with the first half of the break functions is then:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}_1\boldsymbol{\gamma}_{(1)} + \mathbf{v} \tag{110}$$

[10] Results of high gauge for high significance levels (e.g. $\alpha \geq 0.05$) are consistent with previous results found by Bergamelli & Urga (2013) for step-functions. Once a large number of spurious breaks is retained, it becomes more likely to keep additional spurious breaks.

Figure 41: Proportion of irrelevant retained break functions (gauge) using split-half and multi-path selection for varying α when there is no break.



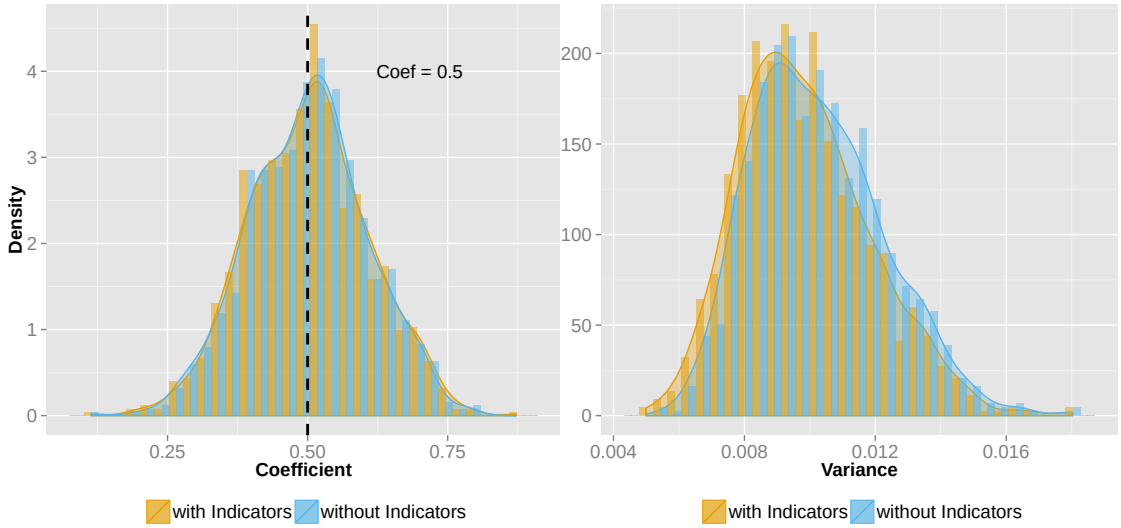
where the true $\gamma_{(1)} = \mathbf{0}$. Following Hendry & Johansen (2014), given that there is no break in the DGP, the inclusion of a full set of irrelevant additional variables $\mathbf{D}_1$ does not affect the distribution of the included relevant parameters β. Orthogonalizing $\mathbf{X}$ and $\mathbf{D_1}$ by regressing each column of $\mathbf{D}_1$ on $\mathbf{X}$ yields the estimator $\widehat{\beta}^*$ with asymptotic distribution:[11]

$$\sqrt{T}\begin{pmatrix}\widehat{\beta^*}-\beta\\ \widehat{\gamma}_{(1)}-0\end{pmatrix} \xrightarrow{D} \mathrm{N}\left[\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \sigma_\epsilon^2\begin{pmatrix}\Sigma_{\mathbf{XX}}^{-1} & \mathbf{0}\\ \mathbf{0} & \Sigma_{\mathbf{D}_1\mathbf{D}_1|\mathbf{X}}^{-1}\end{pmatrix}\right] \tag{111}$$

The distribution of the parameters $\widehat{\beta^*}$ on the correct variables $\mathbf{X}$ is unaffected by the inclusion of the orthogonalized break indicators $\mathbf{D}_1$ when there is no break. The equivalent result holds when the second half of break indicators $\mathbf{D}_2$ is included and the resulting union of retained indicators from $\mathbf{D}_1$ and $\mathbf{D}_2$ given that N<<T. Orthogonalization relative to shifts, however, is not necessary for estimation in practice. Figure 42 shows the simulated distribution of $\widehat{\beta}$ for a single x_t when a full set of break functions is included and selected at $\alpha = 0.005$ (yellow) and when break functions are not included (blue). The distribution of $\widehat{\beta}$ is unaffected by the saturation of a full set of break functions. In practice the main risk is the spurious retention of break indicators, but this can be controlled through a conservative selection mechanism (low α).

[11] Where for a matrix $\mathbf{Z}$ the symmetric positive-definite matrix $\Sigma_{\mathbf{ZZ}}$ is defined as $T^{-1}\mathbf{Z}'\mathbf{Z} \xrightarrow{P} \Sigma_{\mathbf{ZZ}}$. See supplementary material 4.5.6 for proof of the distributional result based on Hendry & Johansen (2014).

Figure 42: Simulated distribution and densities of coefficient $\widehat{\beta}$ (true $\beta = 0.5$) on forced parameter x_t: with (yellow), and without (blue), a full set of break functions.



4.3 Empirical illustration: detection of volcanic eruptions from model surface air temperature data

Large volcanic eruptions that inject significant amounts of sulphate aerosols into the stratosphere cause short-lived (multi-year) radiative imbalances that induce surface cooling. Over the course of the last several millennia there have been numerous eruptions that have had significant impacts on global mean temperatures. Identifying their climatic fingerprint is an important scientific endeavor that relies critically on the robust characterization of the timing and magnitude of past volcanism. Records of climatically relevant events primarily rely on sulphur deposits in ice cores (see e.g. Gao et al. 2008, and Crowley & Unterman 2012), however, there remains uncertainty on the precise timing, magnitude, and climatic impact of past volcanic activity (Brohan et al. 2012, Mann et al. 2012, Anchukaitis et al. 2012, G. Schmidt et al. 2011). Statistical methods such as the break detection methodology presented herein therefore can augment previous volcanic reconstruction estimates by providing additional characterizations of the timing and magnitude of temperature responses to volcanic eruptions when coupled with large-scale proxy estimates of past temperature variability. As a synthetic evaluation of the performance of the break indicator saturation method, we search for volcanic eruptions in surface air temperature output from model simulations. While there is some disagreement on the timing, magnitude, and climatic impact of real eruptions over the past several millennia, the present simulation is forced with deterministic (known, imposed) eruptions. It therefore can function as a useful tool for assessing the detection efficacy of the proposed statistical methodology in real-world scenarios when the timing and exact DGP of volcanic eruptions are uncertain.

For our empirical illustration we use the Northern Hemisphere (NH) mean surface air temperature from the combined Last Millennium (LM) and historical simulation of the National Center for Atmospheric Research (NCAR) Community Climate System Model 4

(CCSM4) (Landrum et al. 2013). These simulations were made available as part of the Coupled and Paleoclimate Model Intercomparison Projects Phases 5 and 3 (CMIP5/PMIP3), respectively (Taylor et al. 2012). Collectively, the two simulations span the period 850-2005 C.E. To imitate potential proxy reconstructions (e.g. tree-ring based), temperatures for extratropical land areas (30 deg −90 deg N) were extracted from the model and only summer months (June-August) were used to build annual averages. This time period is expected to show the strongest cooling in response to an eruption (e.g. Zanchettin et al. 2013 argue for a winter-warming effect). Temperatures are reported as anomalies relative to the 1850-1999 mean. The model is forced with the volcanic reconstruction by Gao et al. (2008) that reports volcanic activity as stratospheric sulphate loadings in teragrams (Tg). While the model is forced with multiple radiative forcing conditions (e.g. solar irriadiance, greenhouse gases, volcanoes, land cover changes, and anthropogenic aerosol changes), for the present experiments we treat these as unknown and work with the univariate NH mean temperature series, although multivariate models with more forcing variables could improve the detection algorithm. For a real-world scenario, however, estimates of climate-forcing and -sensitivity are uncertain (IPCC 2013) and may prove to be of limited use in explaining non-volcanic temperature variation in proxy reconstructions.

4.3.1 *Simulation Setup*

The break functions to capture the temperature response to a large-scale volcanic eruption is approximated through a simple zero-dimensional energy balance model with one time constant (see e.g. section 1 in Rypdal 2012, or section 1 in Schwartz 2012):

$$C\frac{dT'}{dt} = F - \alpha T' \tag{112}$$

where α is the climate feedback, C is the heat capacity, T' the temperature deviation from steady state and F denotes radiative forcing. The feedback response time of the model is given by $\tau = \frac{C}{\alpha}$. Assuming a volcanic forcing effect of an impulse injection of stratospheric aerosols decaying exponentially at rate $-1/\gamma$ yields the following functional form of a volcanic function for the associated temperature response:[12]

$$T'_t = d_t = \begin{cases} \frac{1}{C} e^{\frac{-\alpha}{C}t} F_1 \left(\frac{\alpha}{C} - \frac{1}{\gamma}\right)^{-1} \left[e^{t\left(\frac{\alpha}{C} - \frac{1}{\gamma}\right)} - 1\right] & t \leq L \\ 0 & t > L \end{cases} \tag{113}$$

Different parameter calibrations are explored in the simulation section below. The main results are reported for a normalized temperature response where the feedback response time is set to 1, length of the volcanic impact is set to $L = 3$ to approximate the theory. The decay of stratospheric aerosols is modelled as $\gamma = 0.5$ (function a) and $\gamma = 3$ (function b) to capture one-period and two-period cooling respectively. On visual inspection these calibrations closely match the average-model response based on a superposed epoch analysis (Mass & Portman 1989) of all large-scale volcanic eruptions in the climate model (see Figure 43). The average model response in temperature is a drop by approximately 1–1.5 degrees C, followed by a

[12] See supplementary material 4.5.7 for a derivation. For break detection the function is normalized to sum to 1 over L.

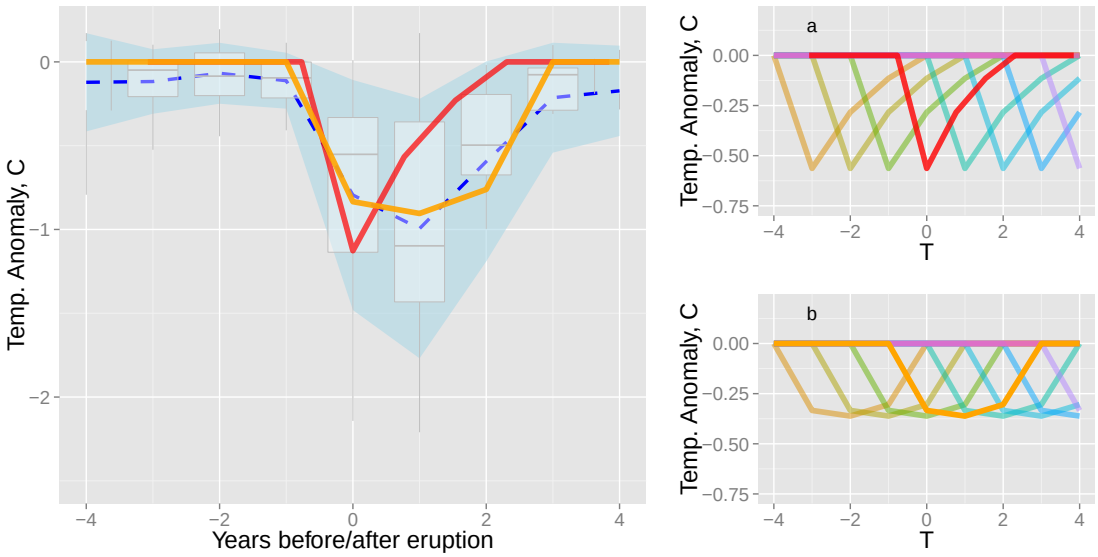


Figure 43: Superposed Epoch Analysis of the Model temperature response to simulated volcanic eruptions and sets of volcanic functions. Left: Superposed epoch analysis (Mass and Portman, 1989) of NH mean model temperature response to volcanoes with sulphate emissions $> 20Tg$ (42 events, dashed blue) with 1 sample standard deviation bands (shaded light blue) and distribution over volcanoes (box-plots). Approximate temperature response using a zero-dimensional energy balance model used as volcanic function (a) is given in red and function (b) in orange. Right: Sets of EBM based volcanic break functions for the two different specifications (a) (top) and (b) (bottom) to approximate the temperature response in years T relative to an eruption at $t = 0$.

smooth reversion to the previous mean over a 3-4 year period. While Gao et al. (2008) estimate the retention time for sulphate aerosols to be 2-3 years, a climatic perturbation of 4 years is in line with findings by Landrum et al. (2013). It is important to emphasize that the in-sample response to a volcanic eruption is not used to design the break function - the method is not trained and evaluated on the same set of observations.

In a more theoretical approach, which avoids particular shape parameters, a single peak (impulse) could be followed by auto-regressive reversion to the mean where we search over a full set of impulses and full set of breaking autoregressive coefficients. The DGP for the response variable NH temperature (T_t) is:

$$T_t = f(X_t, V_t) + \epsilon_t \tag{114}$$

To simulate sampling uncertainty of a proxy-based reconstruction we generate 100 replications of the outcome by adding $\epsilon_t \sim N(0, \sigma^2)$ to the NH mean temperature. The main results here are presented for simulations setting $\sigma = 0.2$ which is half the sample standard deviation of the NH time-series of 0.4, the effect of the magnitude of noise is explored in Figure 45. The function $f(X_t, V_t)$ mapping volcanic V_t and other forcing X_t on to temperature is unknown, and the observed forcing variables V_t and X_t are equally treated as unknown. As a proof

of concept we consider two models (intercept-only, and AR(1) with intercept)[13] to detect eruptions:

$$y_t = \mu + \gamma' \mathbf{d}_t + v_t \quad (115)$$

$$y_t = \rho y_{t-1} + \mu + \gamma' \mathbf{d}_t + v_t \quad (116)$$

where $\mathbf{d}_t$ is a full set of volcanic break functions (113) to be selected over.[14] To reduce computational requirements due to the varying simulation setup, the full sample is split into 10 sub-samples of $T = 115$ observations each.[15] There is little difference between full sample and sub-sampling performance (appendix 4.5.8 provides the results for a full sample simulation). Selection is conducted at $\alpha = 1/T$. Selection over T break functions in a time series of T observations in a correctly specified model implies that on average we expect to retain $\alpha T = T/T = 1$ break functions spuriously. Higher retention of break functions can be an indicator of model misspecification. Simulations are evaluated based on the retention frequency of known individual volcanic events (potency), the average potency over all volcanoes and the proportion of spurious eruptions detected (gauge).

4.3.2 *Simulation Results*

Figure 44 and Tables 23 and 24 show the results of detected volcanic events in 100 replications of the modelled NH mean temperature[16] using the model (a) volcanic function. The retained volcanic breaks (in red, magnitude indicates retention frequency) coincide predominantly with the simulated volcanic eruptions (sulphate graphed in blue and orange in Tg). Few spurious volcanoes are detected, and those that are spurious exhibit retention frequencies drastically lower than those of volcanoes used to force the model.

Most large-scale simulated volcanic eruptions are detected consistently: 74% of all larger ($> 20Tg$) NH eruptions are detected on average within an interval of ± 1 year (57% of all large global eruptions, many of which appear to have had little impact on NH temperatures).

Consistent with the basic analytical results presented in the previous section, the intervals of selection around the true break dates are small. While increasing the band from 0 to 1 generally yields an increase in potency, outside of ± 1 year there is little difference – the potency for $\pm$ 3 years is close to identical to that of ± 1 year (see Table 23). An uncertainty in break dates of $+/-$ 1 year can be the result of a monthly dated volcanic forcing record coupled with an annually dated temperature record. The season of sulphure injection - before or after summer - can cause offsets in the timing of the temperature response. Equally there

[13] Unless otherwise stated, results refer to the intercept-only case.

[14] Given the specification of the volcanic break function and if σ was the only noise added to the DGP then the approximate expected non-centrality for a single unknown break using (97) is $\lambda \left(0.2\sqrt{3.7)}\right)^{-1} \approx 0.4^{-1}\lambda$ where λ is the full temperature response following a volcanic eruption. Since the specified volcanic function has an approximate trough of 0.58 λ, a temperature drop of 1 degree after a volcanic eruption implies that overall $\lambda \approx 1.7$. Thus in absence of additional noise and for a single volcanic break with an immediate temperature response of 1 degree, the expected t-statistic is approximately ≈ 4.3. The analytical probability of detecting this eruption is roughly: $P(|t| > c_\alpha) \approx 0.96$ for $\alpha = 0.01$. Large eruptions should be consistently detected if the break function is correctly specified and if σ was the only source of noise.

[15] The total sample size is $T = 1155$, resulting in 9 sub-samples of $T = 115$ observations and one sub-sample of $T = 120$ observations. Significance levels are scaled accordingly.

[16] Retained volcanic functions with positive coefficients are dropped since these likely constitute positive outliers. The focus here lies on the detection of volcanic events which have a negative temperature response.

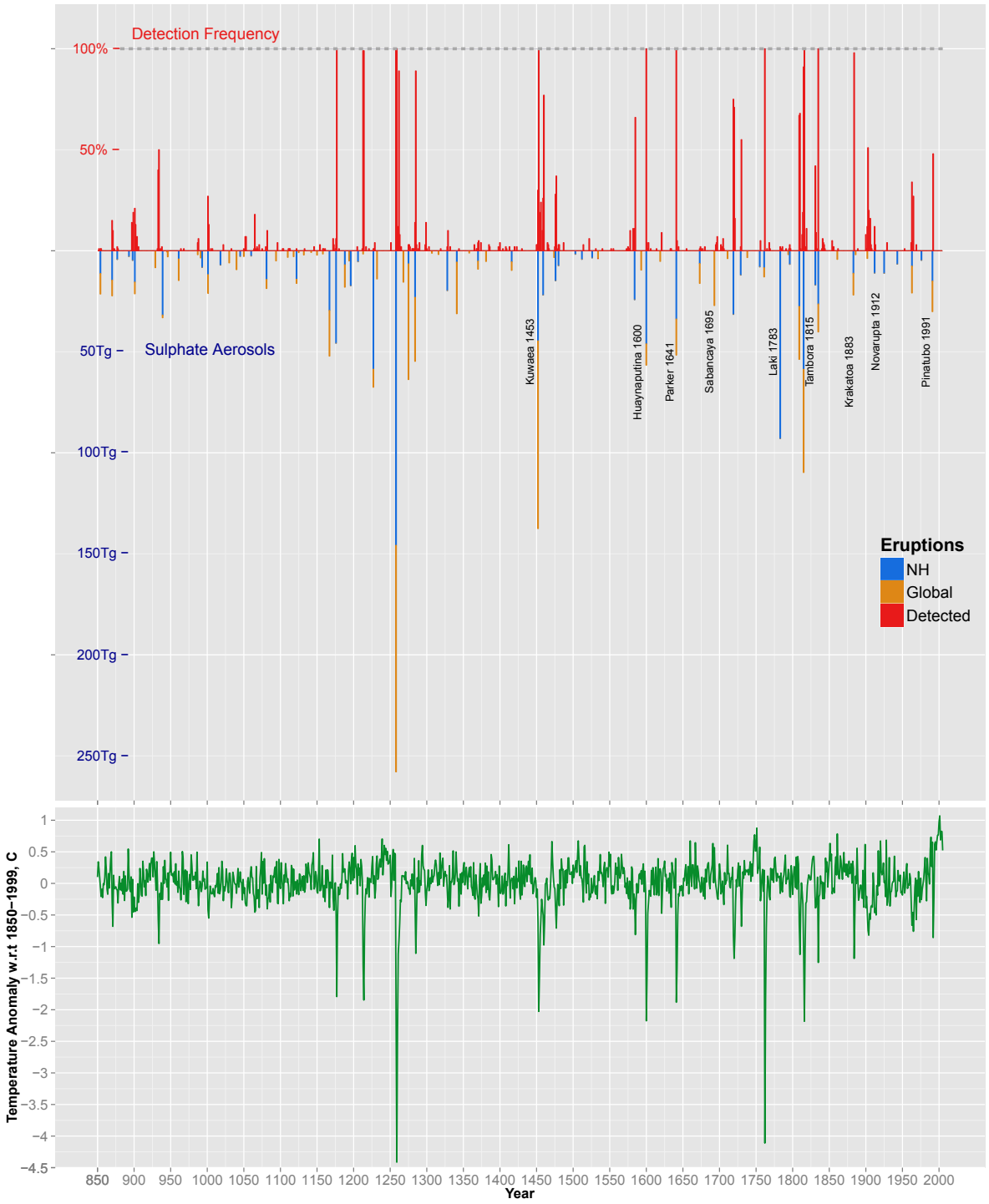


Figure 44: Detected Model Volcanic Eruptions from 850-2005. Top: Detected (red) volcanic eruptions in the model temperature series from 850-2005 using function (a) modelling a single-period drop followed by a reversion to the mean together with an intercept. Bar height indicates detection frequency $[0, 1]$ across 100 simulations. Stacked sulphure deposition record used to force model temperatures are shown for Northern hemisphere (blue) and global measurements (orange) in Tg. Bottom: simulated model mean temperature anomalies (green) used to detect the above volcanic eruptions. Mean NH surface temperature data is taken from the Last Millenium and historical simulation of the NCAR CCSM4 model as part of the CMIP5/PMIP3 data archive.

may be regional sampling biases based on the construction of the NH mean surface air temperature.

Augmenting the designed break functions (a) by an autoregressive model results in nearly similar potency and gauge relative to the baseline model using just a constant (see Table 23 and Figure 45).

The retention frequency of volcanic functions increases with the magnitude of sulphate emissions of the volcanic eruption (Figure 45). While the overall potency for all volcanoes in the Northern Hemisphere within a 1-year interval is 33%, this increases to 74% when larger volcanic eruptions over 20Tg are considered. Given that potency covers all of the volcanic forcing, much of which is small in magnitude, the result is unsurprising. In particular, the lower potency for small eruptions is not driven by an inconsistency in selection of the same volcano over multiple experiments, but rather in the variation in temperature response between volcanoes. Eruptions in 1641 (Parker) and 1600 (Huaynaputina) are detected 100% of the time while the eruption of 1783 (Laki) is not detected in any of the outcomes. In contrast to most of the other volcanoes Laki is a high-latitude volcano. Since the model uses spatially resolved sulphate estimates this eruption only affects the northernmost areas and causes only a minor cooling of -0.15 degrees which is much lower in magnitude than that of any of the other major volcanic events[17] (see Figure 43).

Equally, the potency is affected by the chosen standard deviation of the noise process added to the model mean. The main results here are reported for added noise with a standard deviation of half the sample standard deviation. Figure 45 shows the potency for varying levels of noise.

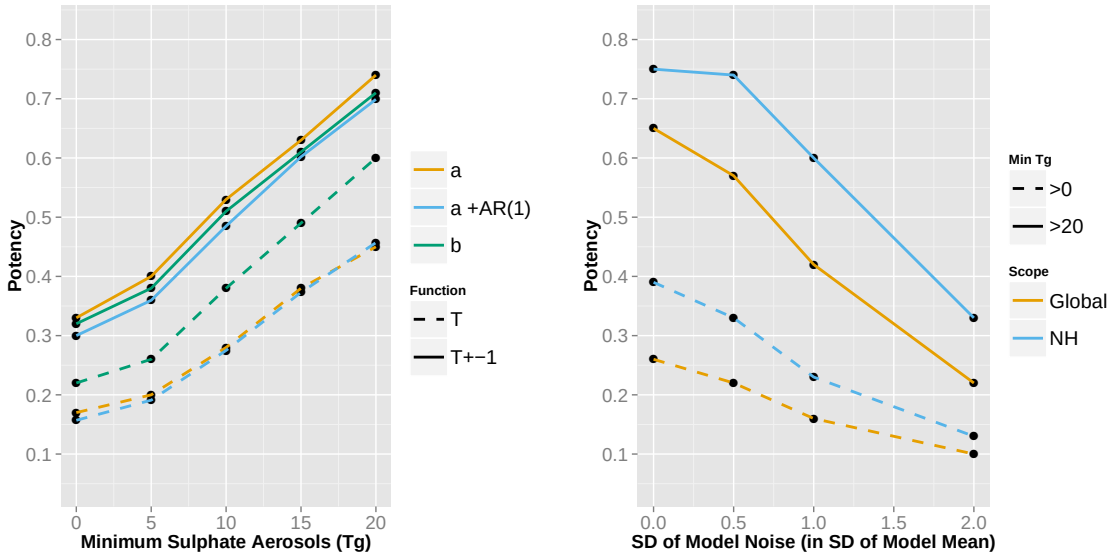
The proportion of spuriously detected volcanoes (gauge) at around 0.02 is close to the nominal significance level ($1/T \approx 0.01$). The fact that it is slightly higher is likely due to the misspecification of the model, which is only run on a constant (including an auto-regressive term in the alternate specification) and set of break functions. Any variability in temperature other than volcanic eruptions may be spuriously attributed to the shape of the volcanic functions. This could be controlled by augmenting the model with additional dynamics (e.g. further autoregressive terms, long-term fluctuations through sine-cosine processes) or known forcing series.

Results for volcanic functions (b) are reported in appendix 4.5.9 (see Tables 28 and 27). Volcanic functions (b) that capture the slower initial decline in temperature yield a slightly higher potency when measured at the precise timing (see Figure 45). Potency for $t = T$ for all NH volcanoes using (b) is 0.32 vs 0.17 for (a) (0.23 vs 0.11 for Global). This result stems from the single drop in function (a) often being most significant in the second period after an eruption if the cooling lasts for two periods. Once we consider the interval of $T \pm 1$ years or volcanoes of larger scale the results are nearly identical for functions (a) and (b). Differentiation between one or two-period cooling following an eruption and thereby further improvements in detection could be implemented by searching over functions of type (a) and (b) simultaneously.

In summary, large-scale volcanic eruptions can consistently be detected with high power within a ± 1 year interval. Even though the model is likely mis-specified when using only a constant, few spurious volcanic eruptions are retained. The signal to noise ratio remains,

[17]There is considerable uncertainty on the impact of the Laki eruption, for example A. Schmidt et al. (2012) find the NH peak temperature response to Laki to be around -1 degree, suggesting that the LM simulation used here may not reflect the entire impact of the eruption.

Figure 45: Left: detection potency of NH eruptions for given minimum sulphate emissions and timing for functions (a), (a) + AR(1), and (b) at the precise timing T (dashed) and in the interval of $T \pm 1$ (solid). Right: Detection for varying levels of noise added in the simulation for function (a) for all eruptions (dashed) and large eruptions over 20Tg (solid).



however, crucial in detection. When the method is applied to real world proxy reconstructions where lower temperature spikes and higher noise levels can be expected, a well specified baseline model for the temperature process will be required against which volcanic events can be detected to ensure a high power of detection.

Forecasting through Breaks

While structural breaks (such as volcanic eruptions) are by their nature stochastic, using a deterministic approach through a full set of break functions allows us to account for the underlying break and model the response deterministically. This can improve forecasts during breaks if the break function is well specified. Once the break is observed (in this case a volcanic eruption), a forecasting model can be augmented with a break indicator where the magnitude is determined through estimation in the first break period. This indicator then acts as a continuous intercept correction, thereby improving the forecast performance during the break. To illustrate this concept, Figure 46 shows a 1-step forecast for NH model mean temperatures following the simulated 1641 eruption, together with the root-mean-squared (RMSE) forecast errors for all NH ($> 20Tg$) model eruptions based on volcanic function (a). Using volcanic indicators to forecast through the breaks yields on average a lower forecast RMSE (RMSE = 0.51) when compared to a simple AR(1) model (RMSE = 0.71) or even a robust forecasting device (RMSE = 0.66) (Clements & Hendry 1999).[18] Crucially this depends on the correct specification of the break function - for volcanic eruptions further improvements could be achieved by switching to volcanic function (b) if the initial cooling lasts for two periods.

[18]The robust forecasting device is based on first differences using the forecasting model for $T+1|T$ given by: $y_{T+1|T} = y_T + \widehat{\rho}\Delta y_T$ where ρ is estimated using an AR(1) model. No error bars are shown on the green robust forecast in Figure 46 due to the non-standard distribution of the forecast.

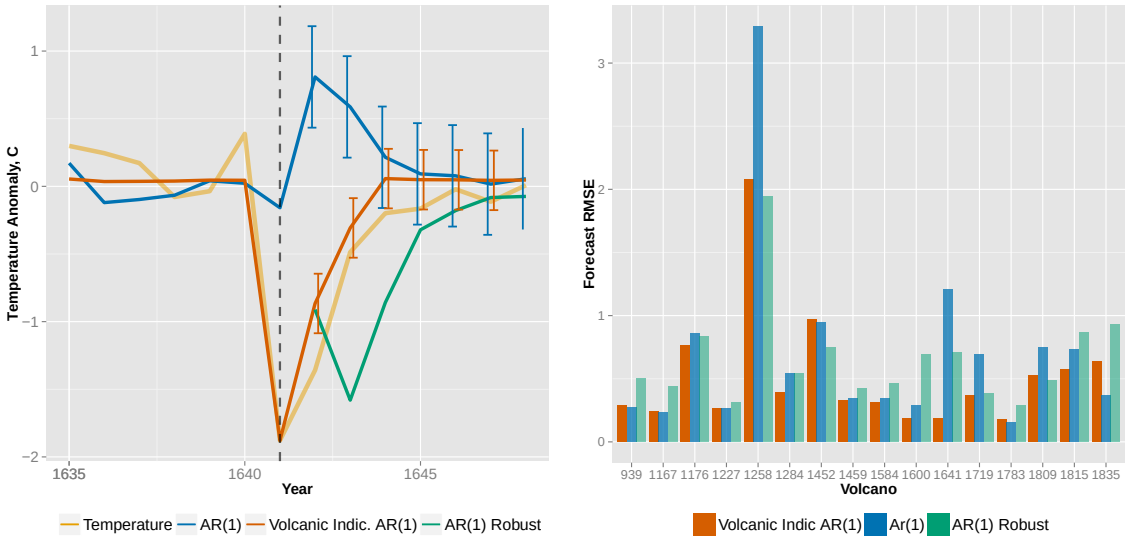


Figure 46: 1-Step forecasts through volcanic eruptions using break indicators. Left: Forecast performance across different methods: model mean temperature during the simulated 1641 eruption is shown in yellow, 1-step forecasts from 1641 onwards are shown for using an AR(1) model with volcanic indicator (red), an AR(1) model without a volcanic indicator (blue) and a robust AR(1) forecast (Clements & Hendry 1999). Models are estimated from 1605 until 1641. Right: 1-step forecast root-mean-squared-error over all NH model volcanic eruptions ($> 20Tg$) for an AR(1) model with volcanic indicator (red), without (blue), and robust AR(1) forecast (green). Using volcanic indicators, on average, improves the forecast performance during the break period. However, when no break occurs (little to no temperature response), using a break indicator can result in higher RMSE as seen for example for the 1783 model Laki eruption.

Saturating a statistical model with a full set of designed break functions and removing all but significant ones through a general-to-specific algorithm yields unbiased estimates of the break magnitude and time. By initializing the model with a full set of break functions many of the shortcomings associated with a forward selection or specific-to-general approach in break detection can be avoided. Analytical power and non-centralities can be derived for any deterministic break function and can be extended to breaks in random variables when interacted with the deterministic break specifications. The break detection procedure exhibits desirable properties both in the presence of breaks (stable potency across multiple breaks) and under the null hypothesis of no breaks where the spurious retention of break functions can be easily controlled through a chosen significance level of selection. The multi-path algorithm outperforms shrinkage-based estimators, especially when facing multiple breaks. We provide some initial insight into uncertainty on the break date by assessing the retention probability of mis-timed break estimators.

Break indicator saturation appears to be effective for detecting large-scale temperature responses to volcanic eruptions. This was shown using surface air temperature output from a last millennium and historical simulation with the NCAR CCSM4 model. Statistically searching over a set of break functions consistently detects large eruptions without prior knowledge of their occurrence. This holds promise for future volcanic detection efforts using real-world proxy reconstructions of temperature variability over the last several millennia. More broadly, break detection using designed functions and indicator saturation provide a framework to analyse the detection of breaks of any specified shape at any point in time.

Table 23: Potency and Gauge for Volcanic Functions (a)

Function (a)	T	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
Potency NH Tg> 20	0.45	0.74	0.74	0.74
Potency NH Tg> 0	0.17	0.33	0.34	0.35
Potency Global Tg> 20	0.32	0.57	0.59	0.59
Potency Global Tg> 0	0.11	0.22	0.25	0.26
Gauge NH	0.02			
Gauge Global	0.02			
Function (a) + AR(1)				
Potency NH Tg> 20	0.46	0.70	0.70	0.70
Potency NH Tg> 0	0.16	0.30	0.31	0.31
Potency Global Tg> 20	0.31	0.52	0.54	0.54
Potency Global Tg> 0	0.11	0.20	0.22	0.23
Gauge NH	0.02			
Gauge Global	0.02			

Table 24: Potency of detection of Volcanic Eruptions $> 20Tg$ using volcanic functions (a) for intervals $t = T \pm 1, 2, 3$

NH Volcano	Tg	Potency $t = T$	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
939	31.83	0	0.02	0.02	0.03
1167	29.535	0	0	0	0
1176	45.761	0.06	1	1	1
1227	58.644	0.01	0.02	0.06	0.06
1258	145.8	1	1	1	1
1284	23.053	0.14	0.97	0.97	0.97
1452	44.6	0.3	1	1	1
1459	21.925	0.26	0.98	0.98	0.98
1584	24.228	0.11	0.77	0.8	0.8
1600	46.077	1	1	1	1
1641	33.805	1	1	1	1
1719	31.483	0.75	1	1	1
1783	92.964	0.02	0.02	0.03	0.05
1809	27.558	0.67	0.99	0.99	0.99
1815	58.694	0.91	1	1	1
1835	26.356	1	1	1	1

Global Volcano	Tg	Potency $t = T$	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
854	21.387	0	0.02	0.03	0.03
870	22.276	0	0.25	0.25	0.25
901	21.283	0	0.34	0.5	0.54
939	33.128	0	0.02	0.02	0.03
1001	21.011	0	0.4	0.4	0.4
1167	52.114	0	0	0	0
1176	45.761	0.06	1	1	1
1227	67.522	0.01	0.02	0.06	0.06
1258	257.91	1	1	1	1
1275	63.723	0	0.06	0.08	0.08
1284	54.698	0.14	0.97	0.97	0.97
1341	31.136	0	0	0	0.01
1452	137.5	0.3	1	1	1
1459	21.925	0.26	0.98	0.98	0.98
1584	24.228	0.11	0.77	0.8	0.8
1600	56.591	1	1	1	1
1641	51.594	1	1	1	1
1693	27.098	0	0	0.03	0.07
1719	31.483	0.75	1	1	1
1783	92.964	0.02	0.02	0.03	0.05
1809	53.74	0.67	0.99	0.99	0.99
1815	109.72	0.91	1	1	1
1835	40.16	1	1	1	1
1883	21.864	0	0.98	0.98	0.98
1963	20.87	0	0.43	0.63	0.63
1991	30.094	0	0.48	0.48	0.48

4.5 APPENDIX

4.5.1 *Power for a known break*

The DGP is:

$$y_t = \mu + \lambda_1 d_t + \epsilon_t \tag{117}$$

The break shifts μ to $\mu + \lambda_1 d_t$ where d_t is a break function of length L beginning at time $t = T_1$ where $T_1 + L \leq T$ such that $d_t \neq 0$ for $T_1 \leq t < T_1 + L$ and 0 otherwise. The correctly specified model is:

$$y_t = \widehat{\mu} + \widehat{\gamma} d_t + \epsilon_t \tag{118}$$

Expressions for the estimators $\widehat{\mu}, \widehat{\gamma}$ are given by:

$$\begin{aligned}
\begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\gamma} - \lambda_1 \end{pmatrix} &= \begin{pmatrix} T & \sum_{t=T_1}^{T_1+L-1} d_t \\ \sum_{t=T_1}^{T_1+L-1} d_t & \sum_{t=T_1}^{T_1+L-1} d_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} \epsilon_t \\ \sum_{t=1}^{T} d_t \epsilon_t \end{pmatrix} \\
&= T^{-1} \left[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \left(\sum_{t=T_1}^{T_1+L-1} d_t \right)^2 \right]^{-1} \cdot \\
&\quad \begin{pmatrix} \sum_{t=T_1}^{T_1+L-1} d_t^2 \sum_{t=1}^{T} \epsilon_t - \sum_{t=T_1}^{T_1+L-1} d_t \sum_{t=T_1}^{T_1+L-1} d_t \epsilon_t \\ - \sum_{t=T_1}^{T_1+L-1} d_t \sum_{t=t=1}^{T} \epsilon_t + T \sum_{t=T_1}^{T_1+L-1} d_t \epsilon_t \end{pmatrix} \\
&= \begin{pmatrix} T_d^{-1} \left(\sum_{t=T_1}^{T_1+L-1} d_t^2 \sum_{t=1}^{T} \epsilon_t - \sum_{t=T_1}^{T_1+L-1} d_t \sum_{t=T_1}^{T_1+L-1} d_t \epsilon_t \right) \\ T_d^{-1} \left(\sum_{t=T1}^{T_1+L-1} d_t \epsilon_t - \sum_{t=T1}^{T_1+L-1} d_t \sum_{t=0}^{T} \epsilon_t \right) \end{pmatrix}
\end{aligned} \tag{119}$$

where $T_d = T \left[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \frac{1}{T} \left(\sum_{t=T_1}^{T_1+L-1} d_t \right)^2 \right]$.

4.5.2 *Proof of equation (90)*

To show that $\widehat{\gamma}_{(1)} = \lambda_1 \mathbf{r} + (\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{D}_1' \boldsymbol{\epsilon}$ we need to show that $\lambda_1 (\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{D}_1' \mathbf{d}_{T_1} = \lambda_1 \mathbf{r}$. This is equivalent to showing that:

$$(\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{D}_1' \mathbf{d}_{T_1} = \mathbf{r} \tag{120}$$

Pre-multiply the above expression by $(\mathbf{D}_1' \mathbf{D}_1)$:

$$\mathbf{D}_1' \boldsymbol{\iota}_{T_1} = (\mathbf{D}_1' \mathbf{D}_1) \mathbf{r} \tag{121}$$

$$\mathbf{D}_1' \boldsymbol{\iota}_{T_1} = \mathbf{D}_1' (\mathbf{D}_1 \mathbf{r}) = \mathbf{D}_1' ([\mathbf{d}_1, ..., \mathbf{d}_{T_1}, ..., \mathbf{d}_{T/2}] \mathbf{r}) \tag{122}$$

$$\mathbf{D}_1' \mathbf{d}_{T_1} = \mathbf{D}_1' \mathbf{d}_{T_1} \tag{123}$$

which follows from $\mathbf{D}_1 = [\mathbf{d}_1, ..., \mathbf{d}_{T_1}, ..., \mathbf{d}_{T/2}]$ and $\mathbf{D}_1\mathbf{r} = \mathbf{d}_{T_1}$, the T_1 column of $\mathbf{D}_1$, where

$$\mathbf{d}_{T_1} = \begin{pmatrix} 0 \\ \vdots \\ d_1 \\ \vdots \\ d_L \\ \vdots \\ 0 \end{pmatrix} \text{ and } \mathbf{r} = \begin{pmatrix} 0 \\ \vdots \\ 1_{T_1} \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} \tag{124}$$

The same result applies to the estimator $\widehat{\gamma}_{(U)}$ over the retained indicators $D_U = [\mathbf{D}_{1*}\mathbf{D}_{2*}]$ where $\widehat{\gamma}_{(U)} = \lambda_1\mathbf{r} + (\mathbf{D}_U'\mathbf{D}_U)^{-1}\mathbf{D}_U'\boldsymbol{\epsilon}$. Assuming that the retained indicators of the first half $\mathbf{D}_{1*}$ include the true break indicator, then it holds that

$$(\mathbf{D}_U'\mathbf{D}_U)^{-1}\mathbf{D}_U'\mathbf{d}_{T_1} = \mathbf{r} \tag{125}$$

for any retained indicators in $\mathbf{D}_{2*}$ by the same steps as above.

4.5.3 *Proof of generalization of step-functions for known break*

Proof that $\widehat{\mu} - \mu = \bar{\epsilon}_2$ and $\widehat{\gamma} - \lambda_1 = \bar{\epsilon}_1 - \bar{\epsilon}_2$ when d_t take the form of SIS-type step indicators and a single break from $t = 0$ to T_1 is considered. Note, under these specifications $\sum_{t=T_1}^{T1+L} d_t^2 = \sum_{t=T_1}^{T1+L} d_t = T_1$. Then

$$\widehat{\mu} - \mu = \frac{T_1 \left(\sum_{t=1}^{T} \epsilon_t - \sum_{t=1}^{T_1} \epsilon_t \right)}{T \left(T_1 - T_1^2 / T \right)} = \frac{\sum_{t=T_1+1}^{T} \epsilon_t}{T - T_1} = \bar{\epsilon}_2 \tag{126}$$

and

$$\begin{aligned}
\widehat{\gamma} - \lambda_1 &= \frac{T \sum_{t=1}^{T_1} \epsilon_t - T_1 \sum_{t=1}^{T} \epsilon_t}{T_1 (T - T_1)} && (127)\\
&= \frac{T \sum_{t=1}^{T_1} \epsilon_t - T_1 \left(\sum_{t=1}^{T_1} \epsilon_t + \sum_{t=T_1+1}^{T} \epsilon_t \right)}{T_1 (T - T_1)} && (128)\\
&= \frac{T_1 (T - T_1) \bar{\epsilon}_1 - T_1 (T - T_1) \bar{\epsilon}_2}{T_1 (T - T_1)} && (129)\\
&= \bar{\epsilon}_1 - \bar{\epsilon}_2 && (130)
\end{aligned}$$

using $\bar{\epsilon}_1 = 1/T_1 \sum_{t=1}^{T_1} \epsilon_t$ and $\bar{\epsilon}_2 = 1/(T - T_1) \sum_{t=T_1+1}^{T} \epsilon_t$.

4.5.4 *Algorithm Specification*

Here we outline the split-half algorithm (*Algorithm 1*, see Figure 36) used for theory-derivations and simulations, and sketch the multi-path algorithm (*Algorithm 2*) used in simulations and in practice:

***Algorithm 1*: Split-Half Indicator Saturation for Designed Break Functions.** Choose a target level of significance α.

1. Design a break function and construct the $(T \times T)$ break matrix $\mathbf{D}$ using equation (81).
2. Split the break matrix $\mathbf{D}$ into two sets: $\mathbf{D}_1 = (\mathbf{d}_1, ..., \mathbf{d}_{T/2})$ and $\mathbf{D}_2 = (\mathbf{d}_{T/2+1}, ..., \mathbf{d}_T)$.
3. Compute the least squares estimators $\widehat{\gamma}_{(1)}$ and $\widehat{\gamma}_{(2)}$ using (90) and (92) fixing an intercept (and additional covariates if required) in each regression.
4. Construct the union set of k retained break functions as $\mathbf{D}_U = (\mathbf{d}_j, ..., \mathbf{d}_{j+k})$ where $\mathbf{d}_j$ is included if $|t_{\widehat{\gamma}_{i,j}}| \geq c_\alpha$ in step 3, where $i = 1, 2$ for each set.
5. Compute the least squares estimator $\widehat{\gamma}_{(U)}$ using (94), fixing an intercept (and additional covariates if required) in each regression. Retain final break functions j if $|t_{\widehat{\gamma}_{U,j}}| \geq c_\alpha$.

***Algorithm 2*: Multi-Path Indicator Saturation for Designed Break Functions.**

1. Design a break function and construct the $(T \times T)$ break matrix $\mathbf{D}$ using equation (81).

2. Use a mulit-path and multi-block algorithm: for the present simulations and application we rely on the parallel step-wise backwards tree-search algorithm *Autometrics* (Doornik 2009a) fixing an intercept (and additional covariates if required) in the general unrestricted model (79).

3. Retain the break functions from the terminal model chosen by the specified tiebreaker criterion.

4.5.5 *Simulation results for a simple dynamic DGP*

Here we present the results on power and false-positive rate when using a multi-path search and the DGP includes an auto-regressive term of order one. The DGP is given by:

$$y_t = \rho y_{t-1} + \lambda_1 d_t + \epsilon_t \tag{131}$$

where $\epsilon \sim \text{iid}N(0, \sigma^2)$, $\sigma^2 = 1$ and $\rho = 0.3$. The model includes an AR(1) term and a full set of break functions. Figure 25 and Table 25 provide the results. The null-retention (false-positive rate) remains well calibrated to the nominal size, the power is slightly lower compared to the static case, likely due to the shape of the break mimicking an auto-regressive process and thereby being detected less frequently.

Table 25: Power of detecting an unknown break when multi-path searches in an auto-regressive DGP and model. Statistics were generated from 1000 simulations and detection significance was set to $\alpha = 0.01$, with a length of $L = 3$. Break magnitude λ corresponds to the full response over the entire break, the peak is 0.58λ.

	Multi-Path	
	Potency	Gauge
$\lambda = 6$, peak=3.48	0.71	0.018
$\lambda = 4$, peak=2.23	0.34	0.015
$\lambda = 2$, peak=1.16	0.09	0.016

4.5.6 *Proof of null result when forced variables are included*

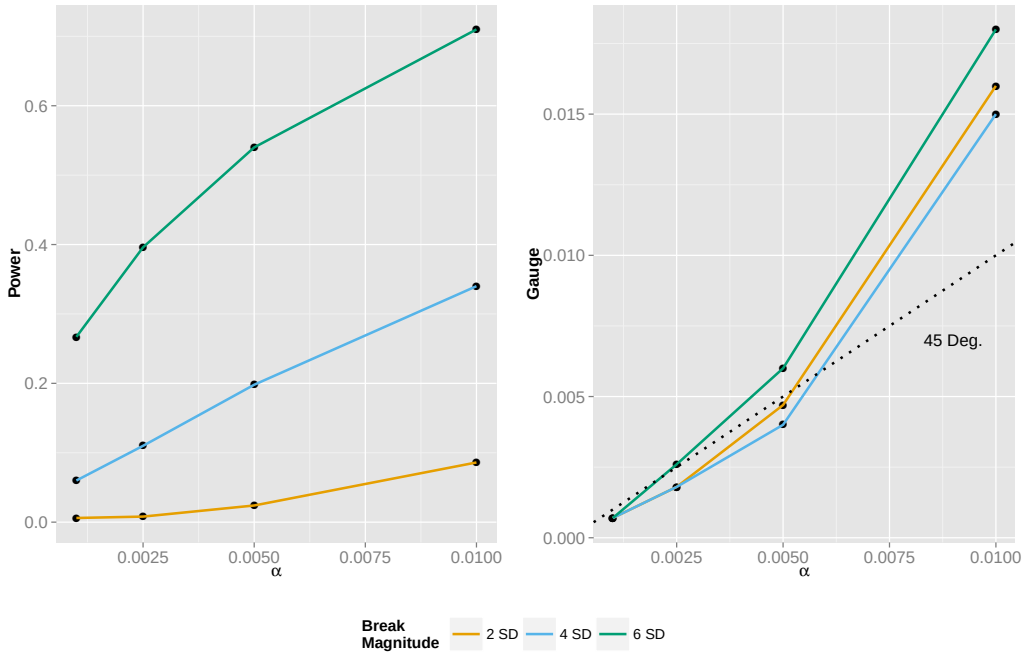
We investigate the approximate distributional consequences of including a full set of irrelevant break variables into a model with assumed K relevant variables $\mathbf{X}$ with associated coefficients β. This is based on the results of Hendry & Johansen (2014). The DGP is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{132}$$

where $\epsilon \sim \text{iid}(0, \sigma_\epsilon^2)$. The model is given as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}_1\boldsymbol{\gamma}_{(1)} + \mathbf{v} \tag{133}$$

Figure 47: Left: Potency of detecting a volcanic break of magnitude λ for level of significance α using multi-path selection for an autoregressive DGP and model. Right: Proportion of spuriously retained break indicators (gauge). Break magnitude λ corresponds to the full response over the entire break, the peak is 0.58λ.



where the true $\gamma_{(1)} = \mathbf{0}$. Consider first orthogonalizing $\mathbf{X}$ and $\mathbf{D_1}$ by regressing each column of $\mathbf{D}_1$ on $\mathbf{X}$, this yields the $(K \times T/2)$ matrix of coefficients $\widehat{\Gamma}$:

$$\widehat{\Gamma} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}_1 \tag{134}$$

The $(T \times T/2)$ residuals $\widehat{\mathbf{u}}$ are defined through:

$$\mathbf{D}_1 = \mathbf{X}\widehat{\Gamma} + \widehat{\mathbf{u}} \tag{135}$$

where $\widehat{\mathbf{u}}'X = \mathbf{0}$. Substituting for $\mathbf{D}_1$ in the model equation yields:

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \left(\mathbf{X}\widehat{\Gamma} + \widehat{\mathbf{u}}\right)\gamma_{(1)} + \mathbf{v} && (136)\\
&= \mathbf{X}\left(\boldsymbol{\beta} + \widehat{\Gamma}\gamma_{(1)}\right) + \widehat{\mathbf{u}}\gamma_{(1)} + \mathbf{v} && (137)\\
&= \mathbf{X}\beta^* + \widehat{\mathbf{u}}\gamma_{(1)} + \mathbf{v} && (138)
\end{aligned}$$

where $\beta^* = \left(\boldsymbol{\beta} + \widehat{\Gamma}\gamma_{(1)}\right)$. Then:

$$\begin{aligned}
\begin{pmatrix} \widehat{\beta^*} - \beta \\ \widehat{\gamma_{(1)}} - 0 \end{pmatrix} &= \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\widehat{\mathbf{u}} \\ \widehat{\mathbf{u}}'\mathbf{X} & \widehat{\mathbf{u}}'\widehat{\mathbf{u}} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'\epsilon \\ \widehat{\mathbf{u}}'\epsilon \end{pmatrix} && (139)\\
&= \begin{pmatrix} \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\epsilon \\ \left(\widehat{\mathbf{u}}'\widehat{\mathbf{u}}\right)^{-1}\widehat{\mathbf{u}}'\epsilon \end{pmatrix} && (140)
\end{aligned}$$

since $\widehat{\mathbf{u}}'X = \mathbf{0}$. Under the assumption that $T^{-1}X'X \xrightarrow{P} \Sigma_{XX}$ The asymptotic distribution is then:

$$\sqrt{T}\begin{pmatrix}\widehat{\beta^*} - \beta \\ \widehat{\gamma_{(1)}} - 0\end{pmatrix} \xrightarrow{D} \mathrm{N}\left[\begin{pmatrix}\mathbf{0} \\ \mathbf{0}\end{pmatrix}, \sigma_\epsilon^2 \begin{pmatrix}\Sigma_{\mathbf{XX}}^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{D_1 D_1|X}}^{-1}\end{pmatrix}\right] \tag{141}$$

The distribution of the parameters β on the correct variables $\mathbf{X}$ is unaffected by the inclusion of the break indicators $\mathbf{D}_1$ when there is no break. The equivalent result holds when the second half of break indicators $\mathbf{D}_2$ is included. In practice the main risk is the spurious retention of break indicators, however, this can be controlled through a conservative selection mechanism.

4.5.7 *Model Response To Large Volcanic Eruption*

The basic energy balance model in terms of deviations from the steady state is given by:

$$C\frac{dT'}{dt} = F(t) - \alpha T' \tag{142}$$

The differential equation can be solved using an integrating factor. Re-arranging the above model yields:

$$\frac{dT'}{dt} + \frac{\alpha}{C}T' = \frac{1}{C}F(t) \tag{143}$$

$$\frac{dT'}{dt} + \frac{1}{\tau}T' = \frac{1}{C}F(t) \tag{144}$$

Using the integrating factor $I(t) = e^{\frac{t}{\tau}}$ yields

$$\frac{d}{dt}T'e^{\frac{t}{\tau}} = \frac{F(t)}{C}e^{\frac{t}{\tau}} \tag{145}$$

$$T'e^{\frac{t}{\tau}} = \int_0^t \frac{F(u)}{C}e^{\frac{u}{\tau}}du \tag{146}$$

$$T' = \frac{1}{C}e^{\frac{-t}{\tau}}\int_0^t F(u)e^{\frac{u}{\tau}}du \tag{147}$$

This expression can be used to solve for the temperature response following a large volcanic eruption under the assumption that the volcanic eruption is approximated by a pulse forcing followed by exponential decay at rate $-1/\gamma$. Let $F(t)$ denote the volcanic forcing:

$$F(t) = 0 \text{ for } t \leq 0 \text{ and } F(t) = F_1 e^{\frac{-t}{\gamma}} \text{ for } t > 0 \tag{148}$$

Using this expression yields:

$$T' = \frac{1}{C}e^{\frac{-t}{\tau}}\int_0^t F(u)e^{\frac{u}{\tau}}du \tag{149}$$

$$T' = \frac{1}{C}e^{\frac{-\alpha}{C}t}F_1\left(\frac{\alpha}{C} - \frac{1}{\gamma}\right)^{-1}\left[e^{t\left(\frac{\alpha}{C} - \frac{1}{\gamma}\right)} - 1\right] \tag{150}$$

4.5.8 *Simulation Results using a Full Sample Search*

Table 26 provides simulation results when a full-sample rather than a sub-sample approach is used for function (a) with an intercept-only model. There is little difference relative to the sub-sampling method at the cost of increased computational time. Using a 3Ghz processor the sub-sample approach requires ≈ 5 seconds to cover the entire sample for one replication (across 10 subsamples), compared to ≈ 5 minutes for one replication using a full-sample approach.

Table 26: Potency and Gauge for Volcanic Functions (a) using a full sample search

	T	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
Potency NH Tg$>$ 20	0.32	0.67	0.67	0.67
Potency NH Tg$>$ 0	0.22	0.53	0.57	0.58
Potency Global Tg$>$ 20	0.23	0.52	0.57	0.58
Potency Global Tg$>$ 0	0.22	0.53	0.57	0.58
Gauge NH	0.02			
Gauge Global	0.02			

4.5.9 *Results for Volcanic Function(b)*

Table 27: Potency and Gauge for Volcanic Functions (b)

	T	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
Potency NH Tg$>$ 20	0.60	0.71	0.72	0.72
Potency NH Tg$>$ 0	0.22	0.32	0.33	0.34
Potency Global Tg$>$ 20	0.40	0.55	0.56	0.57
Potency Global Tg$>$ 0	0.15	0.22	0.25	0.26
Gauge NH	0.02			
Gauge Global	0.02			

Table 28: Detection of Volcanic Eruptions $> 20Tg$ using volcanic functions (b)

NH Volcano	Tg	Potency $t = T$	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
939	31.83	0	0	0.02	0.02
1167	29.535	0.01	0.02	0.04	0.04
1176	45.761	0.6	1	1	1
1227	58.644	0	0.01	0.03	0.03
1258	145.8	1	1	1	1
1284	23.053	0.76	0.82	0.82	0.82
1452	44.6	0.7	1	1	1
1459	21.925	0.81	1	1	1
1584	24.228	0.51	0.66	0.69	0.69
1600	46.077	0.99	1	1	1
1641	33.805	0.92	1	1	1
1719	31.483	0.94	0.98	0.98	0.98
1783	92.964	0	0.01	0.01	0.01
1809	27.558	0.9	0.98	0.98	0.98
1815	58.694	0.65	1	1	1
1835	26.356	0.77	0.9	0.98	0.98
Global Volcano	**Tg**	**Potency $t = T$**	**$t = T \pm 1$**	**$t = T \pm 2$**	**$t = T \pm 3$**
854	21.387	0	0.03	0.04	0.04
870	22.276	0	0.24	0.24	0.24
901	21.283	0	0.5	0.63	0.65
939	33.128	0	0	0.02	0.02
1001	21.011	0	0.34	0.34	0.34
1167	52.114	0.01	0.02	0.04	0.04
1176	45.761	0.6	1	1	1
1227	67.522	0	0.01	0.03	0.03
1258	257.91	1	1	1	1
1275	63.723	0	0.01	0.01	0.01
1284	54.698	0.76	0.82	0.82	0.82
1341	31.136	0	0.01	0.02	0.03
1452	137.5	0.7	1	1	1
1459	21.925	0.81	1	1	1
1584	24.228	0.51	0.66	0.69	0.69
1600	56.591	0.99	1	1	1
1641	51.594	0.92	1	1	1
1693	27.098	0	0.01	0.03	0.11
1719	31.483	0.94	0.98	0.98	0.98
1783	92.964	0	0.01	0.01	0.01
1809	53.74	0.9	0.98	0.98	0.98
1815	109.72	0.65	1	1	1
1835	40.16	0.77	0.9	0.98	0.98
1883	21.864	0	0.95	0.95	0.95
1963	20.87	0	0.47	0.48	0.48
1991	30.094	0	0.38	0.38	0.38

5

WORLD CO_2 EMISSION INTENSITY IS RISING FASTER THAN IPCC CLIMATE SCENARIOS ENVISAGED

Abstract

The large span of long-run projected temperature changes in IPCC (Intergovernmental Panel on Climate Change) reports does not predominately originate from uncertainty across climate models; instead it is the wide range of different global socio-economic scenarios that results in high uncertainty about future climate change. For the first time, observations over two decades are available against which the first two sets of socio-economic scenarios (IS92 and SRES) can be assessed. Here we compare these socio-economic scenarios created in both 1992 and 2000 against the recent observational record. We find that the growth rate in fossil fuel CO_2 emission intensity (fossil fuel CO_2 emissions per GDP) over the 2000s exceeds the projections of all six 1992 and all four 2000 marker emission scenarios. Studying the differences between projections and observations we find that the relative discrepancy is driven by underprediction of high growth rates in Asia, in particular in Russia and China. This underestimation of emission intensity raises concerns about achieving a world of economic prosperity in an environmentally sustainable fashion.

The large span of projected temperature changes in IPCC (Intergovernmental Panel on Climate Change, 2013) reports does not predominately originate from uncertainty across climate models; instead it is the wide range of different global socio-economic scenarios that results in high uncertainty about future climate change. While the physical-science basis of models is very much the focus of the debate in climate research (Rowlands et al. 2012), the underlying socio-economic scenarios that determine emissions of greenhouse gases have received comparably less attention. For the first time observations over two decades are available against which the first two sets of socio-economic scenarios underlying the IPCC reports can be assessed. Here we compare these socio-economic scenarios created in both 1992 (IS92 Leggett et al. 1992, Pepper 1992) and 2000 (SRES Nakicenovic & Swart 2000) against the recent observational record. We find that observed fossil fuel CO_2 emission intensity - fossil fuel CO_2 emissions per GDP - was rising over the 2000s while all six 1992 and all four 2000 marker emission scenarios envisaged a decline. Studying the differences between projections and observations we find that the relative discrepancy is driven by underprediction of high growth rates in Asia, in particular in Russia and China.

Three sets of socio-economic projections have been used in the IPCC reports - the 2nd IPCC report used the IS92 projections, the 3rd and 4th IPCC report used the SRES (Special Report on Emission Scenarios) and the 5th IPCC report relied on the Representative Concentration Pathways (RCPs). We assess all main socio-economic indicators in IS92 and SRES (real gross domestic product GDP, population and fossil fuel CO_2 emissions) with a focus on emission intensity as it is the crucial measure for the environmental impact of economic growth. This combines two projected socio-economic series, real gross domestic product (GDP) and fossil-fuel CO_2 emissions.[1]

There are six socio-economic paths in the 1992 scenarios (named A-F). The SRES projections are made up of multiple individual scenarios falling within four broad groups (named A1, A2, B1, and B2). As the use of sub-group averages is not appropriate (Manning et al. 2010), we instead study the four primary SRES marker scenarios[2] together with the six IS92 scenarios. Emission scenarios used in the early IPCC climate models are reported at a decadal interval (1990, 2000, 2010). Different SRES scenarios were allowed varying initial values at the start of the scenario projections (see Figure 1 panel a, and supplementary Figures S1-S2). This results in small differences for global population (up to 0.38%) but in large differences in world GDP and CO_2 emissions up to 4.3% between scenarios A2 and B1. This variation in starting values complicates any study of the accuracy in levels and is glossed over in many analyses of the IPCC scenarios. Discrepancies in initial values make it necessary to focus on growth rates, and since scenarios cannot be expected to capture short-term year-on-year fluctuations, a comparison on a decadal scale is appropriate.

Unlike temperature forecasts derived from climate model ensembles, the socio-economic scenarios are not probabilistic forecasts. Equally, uncertainties on observed CO_2 emissions

[1] We investigate population projections and fossil-fuel CO_2 emissions per GDP per capita which combines all three socio-economic scenarios in supplementary material 6.5.

[2] These are: A1 AIM, A2 ASF, B1 IMAGE, B2 MESSAGE.

and GDP are not available. Therefore we cannot employ a formal test of a scenario against the observed record.[3] Instead we assess the relative performance of scenarios against observations.

The adequacy of IPCC projections has been discussed previously (Höök et al. 2010, van Vuuren et al. 2010, Richardson et al. 2011), however, the focus has remained solely on levels, which is less meaningful given the aforementioned variation in initial values. Additionally the focus remained primarily on CO_2 emissions and not the wider socio-economic variables or crucially, emission intensity. Due to a lack of data, few studies investigated the performance of scenarios inclusive of 2010. Earlier studies using data up to 2009 find the level of CO_2 emissions to fall within the SRES range (van Vuuren et al., 2010), and using within model group averages, to lie on the upper scale of the 2000 scenarios (Richardson et al., 2011). However, wider socio-economic variables, emission intensity, or the earlier 1992 scenarios are not considered. On the physical side, a first assessment of an early probabilistic temperature forecast based on a single scenario finds that observed global mean temperatures fall within the predicted interval (Allen et al., 2013). We provide a similar analysis by assessing the accuracy of the underlying socio-economic scenarios. This equally permits an assessment of whether the single scenario chosen in the temperature forecast analysis (Allen et al., 2013) not only matches observed temperatures, but also the socio-economic evolution which determines the anthropogenic component in changes in climate.

5.2 DATA

IPCC scenario data are obtained from the IPCC Data Distribution Center (IPCC 2014). Observed fossil-fuel emissions are available at global and national level (Boden et al. 2013). Global and national population data are obtained from the UN Population Division (DESA, UN 2014). Gross domestic product (GDP) on global and national scale is measured in 1990 market exchange rate adjusted USD to be consistent with SRES measures (World Bank 2014).

5.3 ANALYSIS

The world has seen rapid growth in CO_2 emission intensity over the 2000s not envisaged by any of the main scenarios (Figure 48, panel b). While all main IPCC scenarios projected declining emission intensity, the observational record shows that emission intensity was in fact rising. The average decadal growth rate in emission intensity over the 2000s of 0.37% per year considerably exceeded even the closest marker scenario (A1) growth rate projecting a decline of -0.1%. The remaining SRES and IS92 scenarios project declines ranging from -0.3% to -1.75%. The plot of levels of emissions intensity (Figure 48, panel a) hides this discrepancy between scenarios and observations: the level projections only appear to match the observations closely due to the mismatch in starting values. These results are consistent for fossil-fuel CO_2 emissions per GDP per capita: observed growth rates in per capita terms exceed all projected IS92 and SRES marker scenarios despite the "Great Recession" (supplementary Figure 59).

To investigate the systematic discrepancy in global emission intensity growth rates between observations and scenarios, we downscale the regional scenario values and assess the accuracy

[3] Equally a test of the average annual growth rate against scenario growth rates provides little insight since the assumption that the observed growth rate in a particular year is equal to the scenario growth rate is not made in the scenarios.

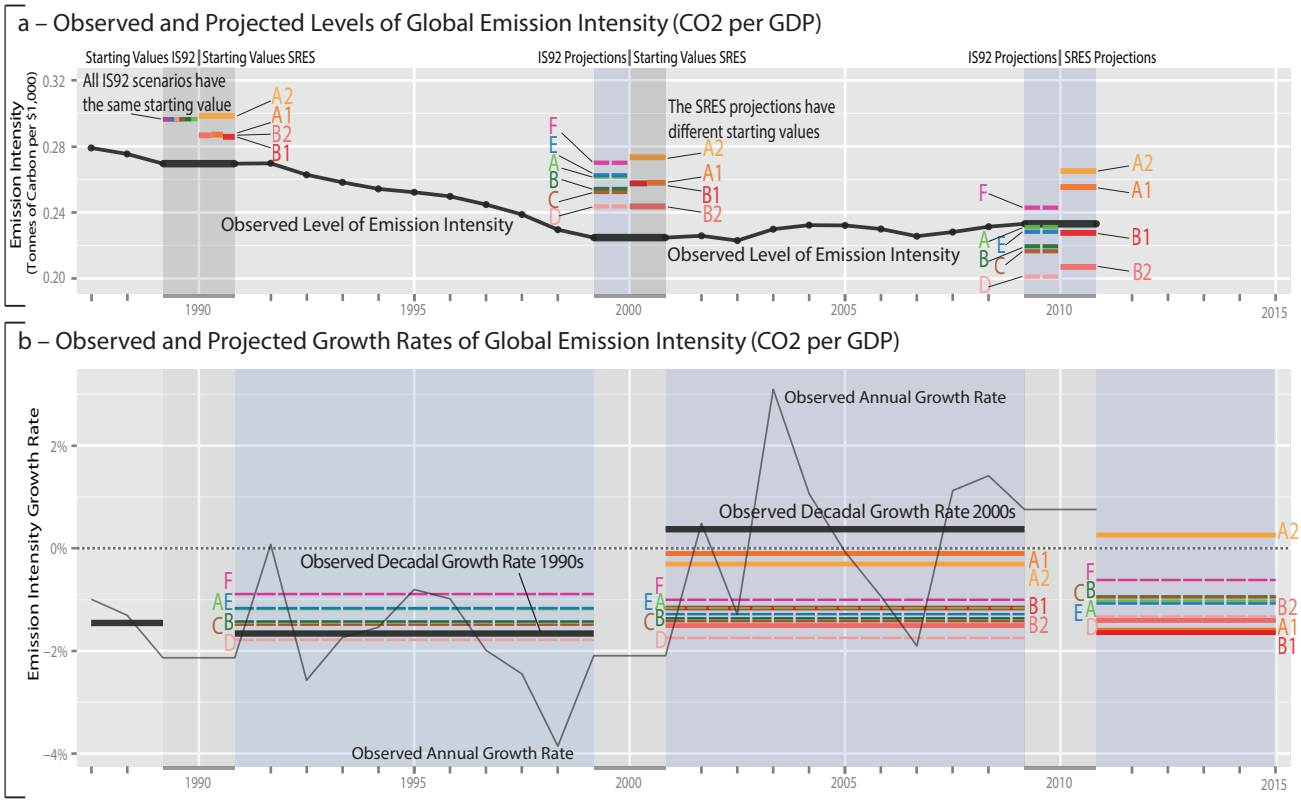


Figure 48: Observed and projected global emission intensity (fossil fuel CO_2 emissions per GDP) in levels (a) and growth rates (b). Panel a graphs global observed emission intensity (black) together with decadal IS92 (dashed colour) and SRES marker projections (solid colour). Starting values are shaded grey while projected values are shaded light blue. Note that initial values for SRES vary across scenarios. Panel b shows observed annual growth rates (continuous black) together with observed decadal growth rates (horizontal black) over both decades. Projected growth rates are shown in colour for IS92 (dashed) and SRES marker projections (solid). Observed decadal growth rates exceed all scenario projections over the 2000s.

based on a country-level dis-aggregation for the SRES projections. By downscaling the regional values of the scenarios to a country level, we then compare the observed against the projected country growth rates (Figure 49, panel a). The apparent regional differences are consistent across all four marker scenarios: the observed growth rates of fossil-fuel CO_2 emission intensity in Sub-Saharan African and South Asian countries greatly exceed those of the scenarios. Across all scenarios, growth rates observed in Latin American are predominantly below scenario projections.

To quantify the country-by-country contribution to the discrepancy between projected and observed global growth rates, the importance of the country in terms of the level of GDP and emissions intensity has to be considered. We therefore decompose the aggregate (global) difference in observed and scenario growth rates into individual country contributions (see Methodology 5.5.2). While Sub-Saharan African countries exhibit the highest deviations from scenario projections (Figure 49, panel a), their overall contribution to the global difference in growth rates is small (Figure 49, panel b). When weighted according to their importance with respect to global emission intensity, it becomes apparent that changes in emission intensity and GDP in China, Russia, and wider Asia account for the dominant share of the difference between global observed and SRES projected growth rates in all four SRES marker scenarios (Figure 49, panel b). It is important to emphasize that this is relative to the SRES projections and that the emission intensity in China and Russia actually declined over the 2000s (see supplementary Figure 58). The discrepancy is primarily driven by unanticipated growth in GDP in Asia. Rapid growth in Asia was unanticipated in most forecasts: even over much shorter forecast horizons of one quarter to one-and a half years, GDP forecasts produced by the Federal Reserve exhibit the highest forecast errors for China over the 2000s (R. Ericsson N et al. 2014). On a decadal scale, observed Chinese GDP growth exceeds the Consensus Economics (2013) forecasts by 7%, a similar magnitude as SRES projections, which are exceeded by 2-11%. IPCC projections do not appear systematically worse than alternative forecasts.

More broadly, we assess which scenario is closest to the observed record measured by the smallest proportional deviation from observed levels and lowest absolute difference to growth rates. The best fitting scenarios of each main socio-economic projection are listed in the supplementary material for both levels and growth rates. Population projections exhibit the lowest deviations from observations (supplementary Figure 52). No single scenario uniquely dominates other scenarios when assessed against the observed variables over both time intervals (see Figures 48, 50, supplementary Figures 51-56, and supplementary Table 29). Notably, earlier IS92 scenarios are not systematically worse than later SRES projections. The 1992 A scenario chosen for the assessment of the temperature forecast by Allen et al. is the closest IS92 scenario in levels of CO_2 emissions, supporting the choice of this scenario as the level of CO_2 emissions is most relevant for a global mean temperature forecast. There is no forecast failure as the levels of observed CO_2 emissions are closely matched. However, over the 2000s, nine out of the ten projections underestimated decadal growth in fossil-fuel CO_2 emissions (Figure 50, panel b). While forecasts are not rejected in the short-run, we may expect increasing divergence from scenario values in the long run based on the under-projection of both growth rates in CO_2 emission, and growth rates of emission intensity. Evidence for this can already be seen through the accelerating accumulation of concentrations of CO_2 in the atmosphere – the growth between 2012 and 2013 was the highest observed (World

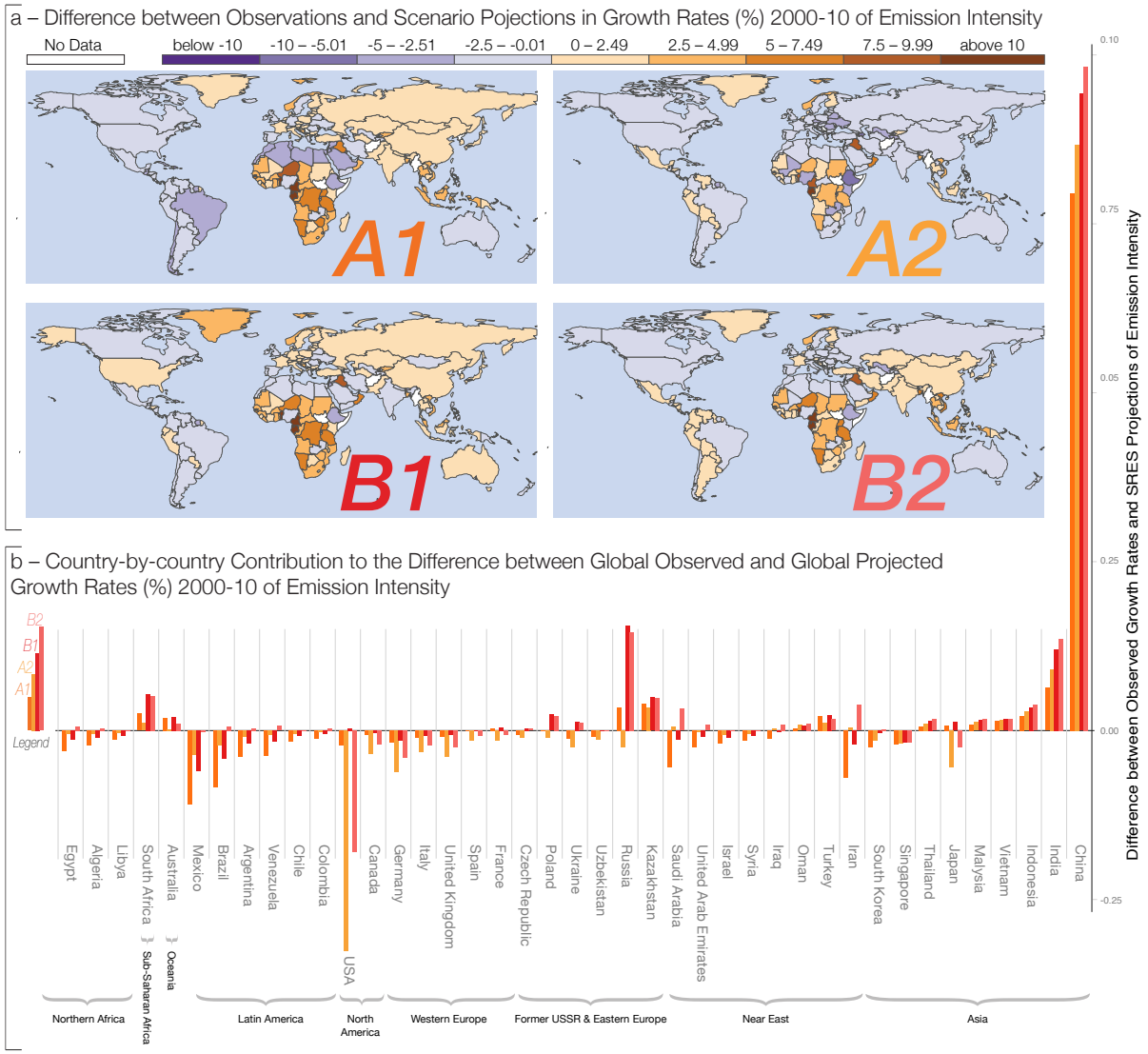


Figure 49: Country-by-country downscaled difference between observed and projected decadal growth rates in emission intensity for SRES marker projections over 2000-2010. The four world maps in panel a show the difference between observed and projected growth rates in fossil-fuel CO_2 emissions per GDP. Growth rates in emission intensity underestimated by the projections are shown in brown tones, overestimated growth rates in purple tones. Panel b graphs the country-by-country contribution to the difference between global observed and global projected growth rates in emission intensity. Only countries contributing more than 0.001 towards the difference are shown. While growth rates in Sub-Saharan Africa exceeded scenario projections, they contributed little to the overall magnitude of the difference in global observed and projected growth rates. The primary contribution to the differences in global growth rates relative to SRES projections stems from changes in Asia and China in particular.

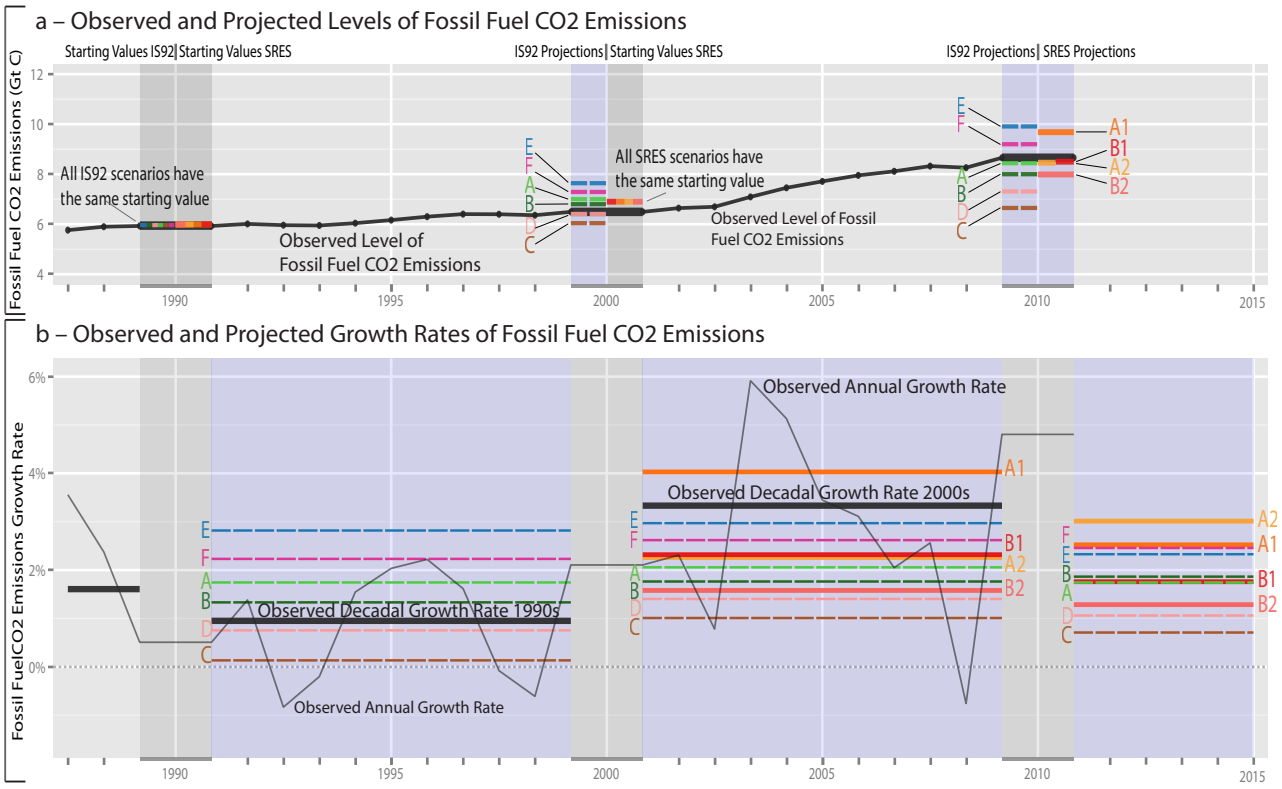


Figure 50: Observed and projected global fossil-fuel CO_2 emissions in levels (a) and growth rates (b). Panel a graphs global observed annual fossil-fuel CO_2 emissions (black) together with decadal IS92 (dashed colour) and SRES marker projections (solid colour). Starting values are shaded grey while projected values are shaded light blue. Panel b shows observed annual growth rates (continuous black) together with observed decadal growth rates (horizontal black) over both decades. Projected growth rates are shown in colour for IS92 (dashed) and SRES marker projections (solid). Observed decadal growth rates exceed nine out of ten scenario projections over the 2000s.

Meteorological Organization 2014) since 1984, and is also supported through the future outlook on CO_2 emissions by Friedlingstein et al. (2014).

5.4 CONCLUSION

The evaluation of the IPCC projections against the observed record over two decades requires careful analysis of the growth rates due to substantial differences in initial values in levels across the different projections. The analysis of the growth rates revealed that global emissions intensity growth exceeded all six 1992 and four 2000 marker scenarios over the decade 2000-2010. Similarly, growth rates of fossil-fuel CO_2 emissions exceeded projections in all but one scenario over the same period. This underestimation of emission intensity raises concerns about achieving a world of economic prosperity in an environmentally sustainable fashion.

5.5 APPENDIX A - METHODS

5.5.1 *Downscaling the SRES Projections*

Individual country growth rates are obtained using linear downscaling: countries are assigned their corresponding regional growth rate based on the four regions defined in the IPCC SRES projections. These regions are: REF – countries undergoing economic reform , OECD90, ASIA, and ALM - Africa and Latin America.[4] Each region in the SRES projections has a common growth rate across all countries. Figure 49 panel a plots the difference between observed and projected (regionally common) growth rates. Due to the variation in observed growth rates, there is variation in differences within each region. For level reconstructions the projected growth rates are applied to the individual country's[5] observed values in 2000 (Boden et al. 2013, DESA, UN 2014, World Bank 2014). More sophisticated downscaling methods are available - van Vuuren et al. (2007) released a downscaled SRES data set. In this data set the starting values are corrected, and while this adjustment of starting values is sensible in general, it renders the data impossible for an assessment of the original IPCC projections. Linear downscaling is chosen here as it does not require additional assumptions on regional convergence and preserves the differences in initial values which are crucial for the task at hand. Over long time scales we expect the values of different downscaling methods to diverge due to compounding, however, over the short time period considered here (2000-2010), the type of downscaling approach does have less of an effect on the final values.

5.5.2 *Decomposition of Aggregate Growth Rates*

We are interested in explaining the difference between observed global and projected global growth rates in emission intensity (visible in Figure 48 panel b). For this we investigate which countries are most important in explaining the aggregate difference. We use individual country level data from 2000-2010 to calculate observed country-level growth rates. For each scenario we use the downscaled projected country level growth rates described in the methodology

[4] While the IPCC regional list does not include the United Kingdom, we include it within the OECD90 region.
[5] Starting values in 2000 are chosen to be able to include a wide range of Eastern European countries for which no observations in 1990 are available.

section 5.1. To attribute the discrepancies in global growth rates to individual contributions in growth rates at country-level (plotted in Figure 49, panel b), we use the following approach to decompose aggregate growth rates: First, we show how to decompose aggregate growth rates into individual country contributions (the method is naturally equivalent for both observed and projected growth rates). Second, this allows us to decompose the aggregate difference into individual country-level contributions.

Decomposing Aggregate Growth Rates

Let $Y_t = \sum_j Y_t^j$ denote aggregate GDP over countries j, and let $Z_t = \frac{C_t}{Y_t}$ denote aggregate fossil fuel CO_2 emissions per GDP, where aggregate fossil fuel emissions $C_t = \sum_j C_t^j$ are summed over countries j. Each country j's fossil fuel CO_2 emissions per GDP is defined as $Z_t^j = \frac{C_t^j}{Y_t^j}$. The corresponding aggregate (G_t) and individual (G_t^j) growth rates are given by $G_t = \frac{Z_t - Z_{t-1}}{Z_{t-1}}$ and $G_t^j = \frac{Z_t^j - Z_{t-1}^j}{Z_{t-1}^j}$ respectively. The aggregate growth rate can then be re-expressed as:

$$G_t = \frac{Z_t - Z_{t-1}}{Z_{t-1}} = \frac{1}{Z_{t-1}} \left(\frac{\sum_j C_t^j}{Y_t} - \frac{\sum_j C_{t-1}^j}{Y_{t-1}} \right) \tag{151}$$

From above we use that the equation for the individual growth rate can then be re-arranged to yield:

$$G_t^j \frac{C_{t-1}^j}{Y_{t-1}^j} = \frac{C_t^j}{Y_t^j} - \frac{C_{t-1}^j}{Y_{t-1}^j} \tag{152}$$

This provides an expression for country j's fossil fuel emissions C_t^j:

$$C_t^j = (1 + G_t^j) C_{t-1}^j \frac{Y_t^j}{Y_{t-1}^j} \tag{153}$$

Therefore we can write the aggregate growth rate as a function of disaggregate growth rates by substituting for C_t^j in:

$$G_t = \left(\sum_j \frac{1}{Y_t} C_t^j - \frac{1}{Y_{t-1}} C_{t-1}^j \right) \frac{1}{Z_{t-1}} \tag{154}$$

This can be simplified to:

$$\begin{aligned} G_t &= \left(\sum_j \frac{C_{t-1}^j}{Y_{t-1}^j} \left[\frac{Y_t^j}{Y_t} (1 + G_t^j) - \frac{Y_{t-1}^j}{Y_{t-1}} \right] \right) \frac{1}{Z_{t-1}} && (155) \\ &= \left(\sum_j \frac{Z_{t-1}^j}{Z_{t-1}} \left[\frac{Y_t^j}{Y_t} - \frac{Y_{t-1}^j}{Y_{t-1}} + G_t^j \frac{Y_t^j}{Y_t} \right] \right) && (156) \\ &= \left(\sum_j \frac{Z_{t-1}^j}{Z_{t-1}} \left[\Delta \frac{Y_t^j}{Y_t} + G_t^j \frac{Y_t^j}{Y_t} \right] \right) = \left(\sum_j d_j \right) && (157) \end{aligned}$$

where $d_j = \frac{Z^j_{t-1}}{Z_{t-1}}\left[\Delta\frac{Y^j_t}{Y_t} + G^j_t\frac{Y^j_t}{Y_t}\right]$ and the term $\Delta\frac{Y^j_t}{Y_t} = \frac{Y^j_t}{Y_t} - \frac{Y^j_{t-1}}{Y_{t-1}}$ captures the change in the proportion of country j's GDP relative to total GDP. This yields the aggregate growth rates as the sum of the individual components j and allows for a decomposition of the aggregate growth rates into individual contributions d_j. The individual contributions can be decomposed further into an "emission intensity growth rate effect" and a "relative change in GDP effect":

$$d_j = \frac{Z^j_{t-1}}{Z_{t-1}}\left[\underbrace{\Delta\frac{Y^j_t}{Y_t}}_{\text{GDP Effect}} + \underbrace{G^j_t\frac{Y^j_t}{Y_t}}_{\text{Growth Rate Effect}}\right] \tag{158}$$

If the ratio of the particular countrys GDP to global GDP is unchanged $\left(\Delta\frac{(Y^j_t}{Y_t} = 0\right)$, then the only contribution to the overall growth rate is derived from the growth rate effect: $\frac{Z^j_{t-1}}{Z_{t-1}}G^j_t\frac{Y^j_t}{Y_t}$. Whether the contribution to the overall growth rate is positive or negative therefore depends on the change in the ratio of a country's GDP relative to the global GDP, and a country's growth rate in emissions per GDP, G^j_t scaled by the weight of the country's GDP relative to the global GDP.

Decomposing the Difference between Observed and Projected Aggregate Growth Rates

When comparing an observed growth rate (G_t) against a scenario predicted growth rate ($\hat{G}_t$), using the same decomposition procedure as above, the difference between observed and predicted growth rates can be attributed to disaggregated country contributions, and further into relative GDP change and emission intensity growth rate effects. Given the summability, simple regional aggregates can be considered as well as country-level disaggregation. These are plotted in Figure 49 (panel b) for individual contributions to the overall difference in growth rates exceeding 0.1%.

$$G_t - \hat{G}_t = \sum_j d_j - \sum_j \hat{d}_j \tag{159}$$

$$= \sum_j \frac{Z^j_{t-1}}{Z_{t-1}}\left[\Delta\frac{Y^j_t}{Y_t} + G^j_t\frac{Y^j_t}{Y_t}\right] - \sum_j \frac{\hat{Z}^j_{t-1}}{\hat{Z}_{t-1}}\left[\Delta\frac{\hat{Y}^j_t}{\hat{Y}_t} + \hat{G}^j_t\frac{\hat{Y}^j_t}{\hat{Y}_t}\right] \tag{160}$$

$$= \sum_j \underbrace{\left(\frac{Z^j_{t-1}}{Z_{t-1}}\left[\Delta\frac{Y^j_t}{Y_t}\right] - \frac{\hat{Z}^j_{t-1}}{\hat{Z}_{t-1}}\left[\Delta\frac{\hat{Y}^j_t}{\hat{Y}_t}\right]\right)}_{\text{GDP Effect}} + \sum_j \underbrace{\left(\frac{Z^j_{t-1}}{Z_{t-1}}\left[\hat{G}^j_t\frac{Y^j_t}{Y_t}\right] - \frac{\hat{Z}^j_{t-1}}{\hat{Z}_{t-1}}\left[\hat{G}^j_t\frac{\hat{Y}^j_t}{\hat{Y}_t}\right]\right)}_{\text{Growth Rate Effect}}$$

5.6 APPENDIX B

5.6.1 *Observed Socio-Economic Indicators and Scenario Projections*

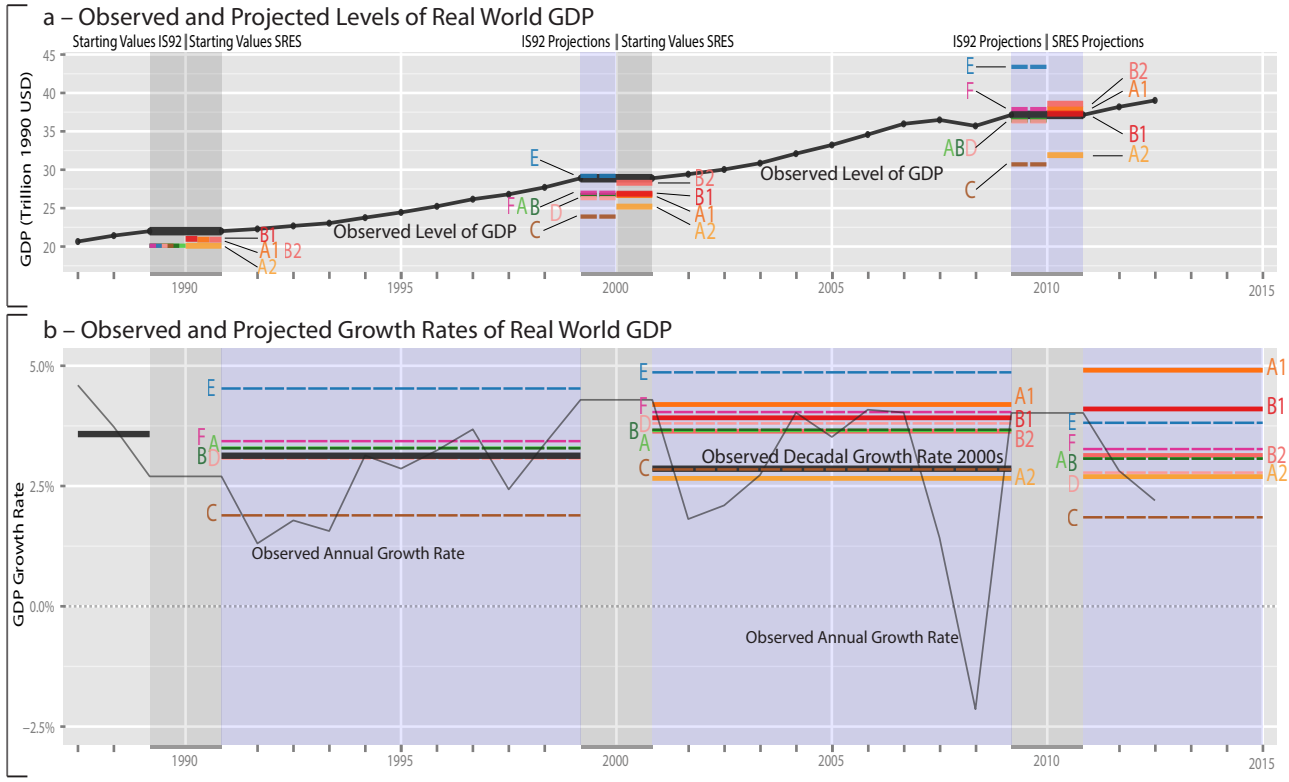


Figure 51: Global real GDP in levels in 1990 USD (top) and growth rates (bottom). Observed are shown in black, projections in colour.

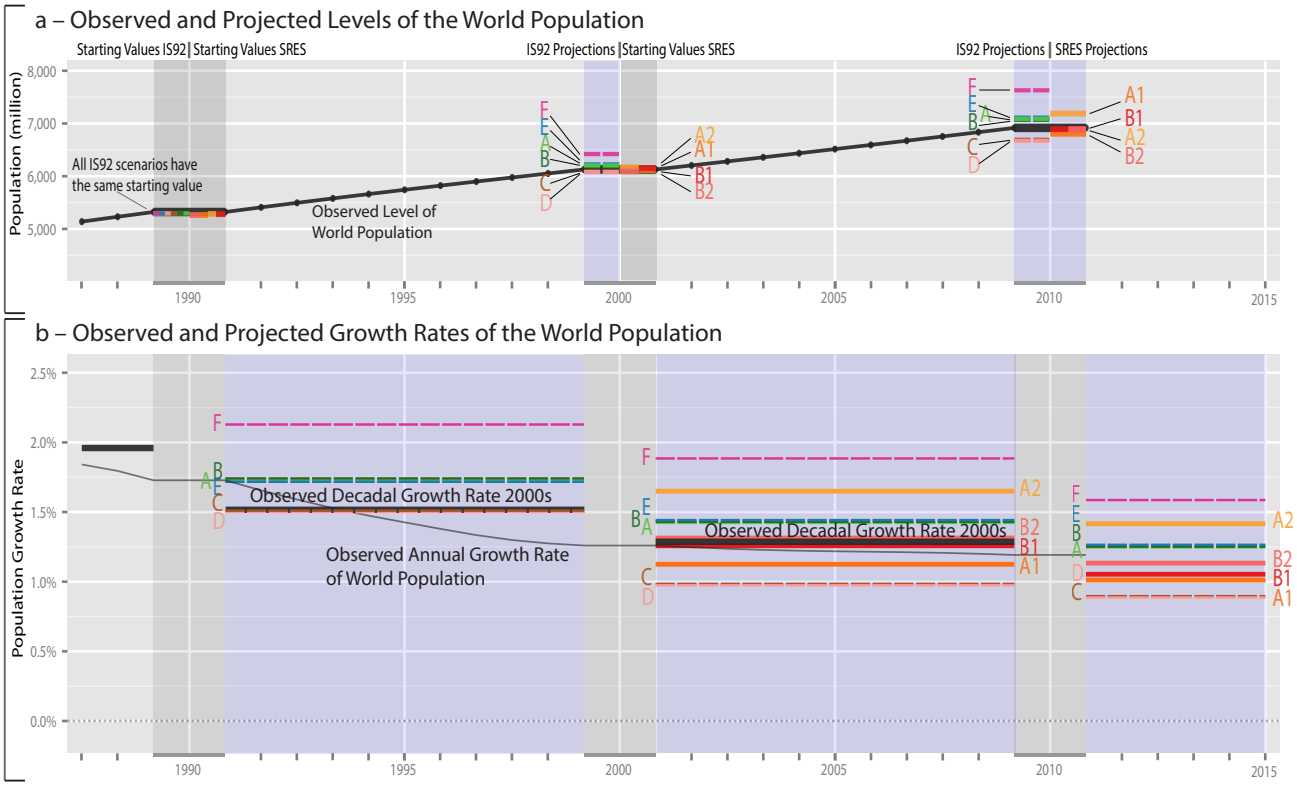


Figure 52: Global population in levels (top) and growth rates (bottom). Observed are shown in black, projections in colour.

5.6.2 *Relative Scenario Performances*

Table 29 lists the most accurate scenario based both on levels and growth rates for both 1990s and 2000s. Accuracy here is measured by the smallest absolute percentage deviation in levels, and the smallest absolute deviation in growth rates.

Table 29: Closest SRES and IS92 scenarios for 1990-2000 and 2000-2010 based on proportional deviation from observed levels and absolute difference from observed growth rates. Deviations are shown in parentheses.

	GDP		**Population**		**Foss. Fuel CO_2**		**CO_2 Intensity**	
Time	*Level*	*Growth*	*Level*	*Growth*	*Level*	*Growth*	*Level*	*Growth*
IS92 90s	E (0.011)	D (-0.045)	C/D (0.013)	C/D (0.206)	D (-0.012)	D (-0.196)	D (0.084)	D (-0.122)
IS92 00s	B (-0.017)	C (-0.015)	A/B/E (0.025)	A/B/E (0.137)	A (-0.025)	E (-0.361)	E (-0.008)	F (-1.375)
SRES 00s	B1 (0.004)	A2 (-0.201)	B1 (-0.003)	B2 (0.026)	B1 (-0.018)	A1 (0.697)	B1 (-0.023)	A1 (-0.474)

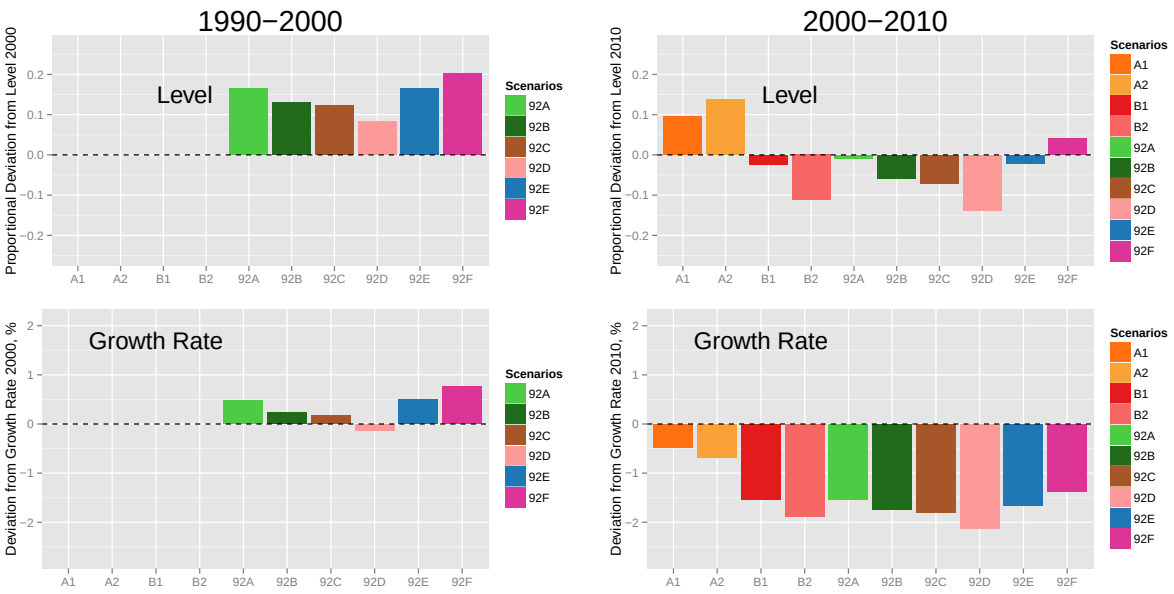


Figure 53: CO_2 Intensity Differences to Observed Levels (top) and Growth Rates (bottom) over 1990-2000 (left) and 2000-2010 (right)

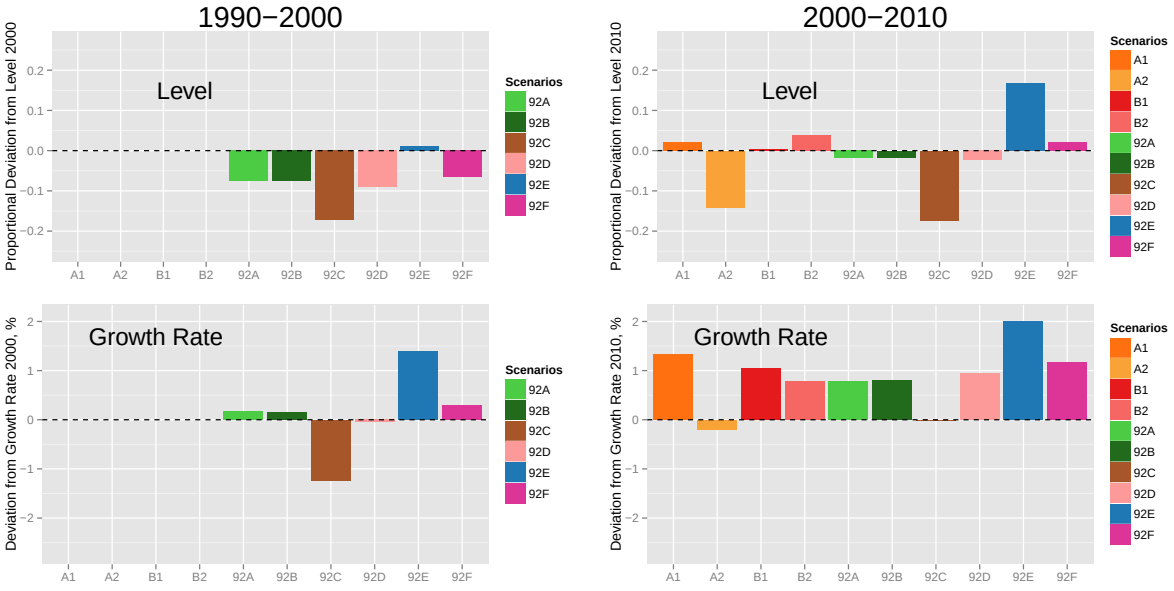


Figure 54: GDP Differences to Observed Levels (top) and Growth Rates (bottom) over 1990-2000 (left) and 2000-2010 (right)

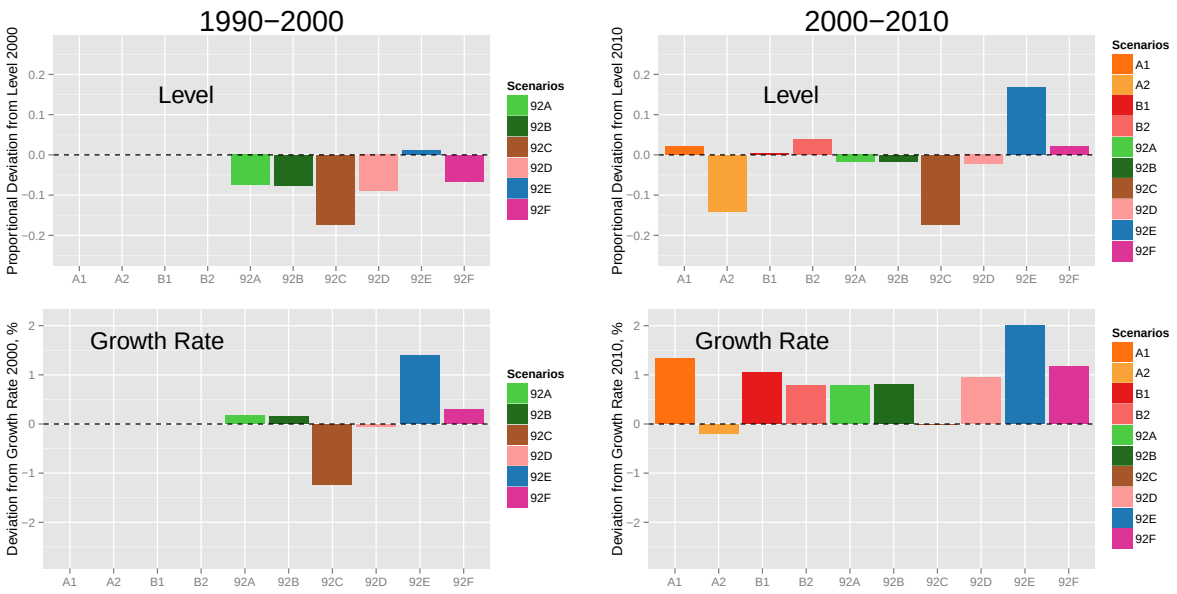


Figure 55: Population Differences to Observed Levels (top) and Growth Rates (bottom) over 1990-2000 (left) and 2000-2010 (right)

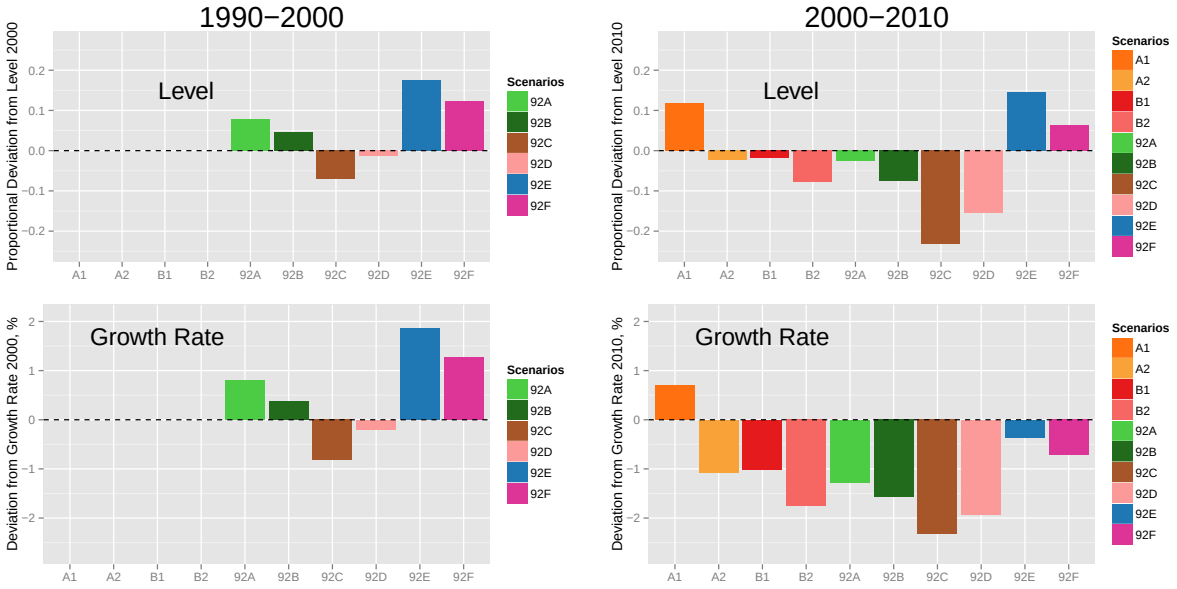


Figure 56: Fossil Fuel CO_2 Differences to Observed Levels (top) and Growth Rates (bottom) over 1990-2000 (left) and 2000-2010 (right)

5.6.3 *Total Growth Rate Decomposition into GDP and pure Emission Intensity Growth Rate Effect*

Country-by-country contributions to the difference in observed world growth rates and projected growth rates can be further decomposed into a GDP effect and a growth rate effect using the results from section 5.5.2:

$$G_t - \hat{G}_t = \underbrace{\sum_j \left(\frac{Z^j_{t-1}}{Z_{t-1}} \left[\Delta \frac{Y^j_t}{Y_t} \right] - \frac{\hat{Z}^j_{t-1}}{\hat{Z}_{t-1}} \left[\Delta \frac{\hat{Y}^j_t}{\hat{Y}_t} \right] \right)}_{\text{GDP Effect}} + \underbrace{\sum_j \left(\frac{Z^j_{t-1}}{Z_{t-1}} \left[\hat{G}^j_t \frac{Y^j_t}{Y_t} \right] - \frac{\hat{Z}^j_{t-1}}{\hat{Z}_{t-1}} \left[\hat{G}^j_t \frac{\hat{Y}^j_t}{\hat{Y}_t} \right] \right)}_{\text{Growth Rate Effect}}$$

Figure 57 graphs the difference in growth rate decomposition for the SRES projection A1 for countries with an overall effect exceeding 0.01 (1%). China contributes the largest absolute amount, this is primarily driven by changes in the ratio of China's GDP relative to aggregate GDP. The largest CO_2 intensity growth rate effects stem from Russia and South Africa, however, the total effect of these is comparatively minor due to their small weight in global emission intensity.

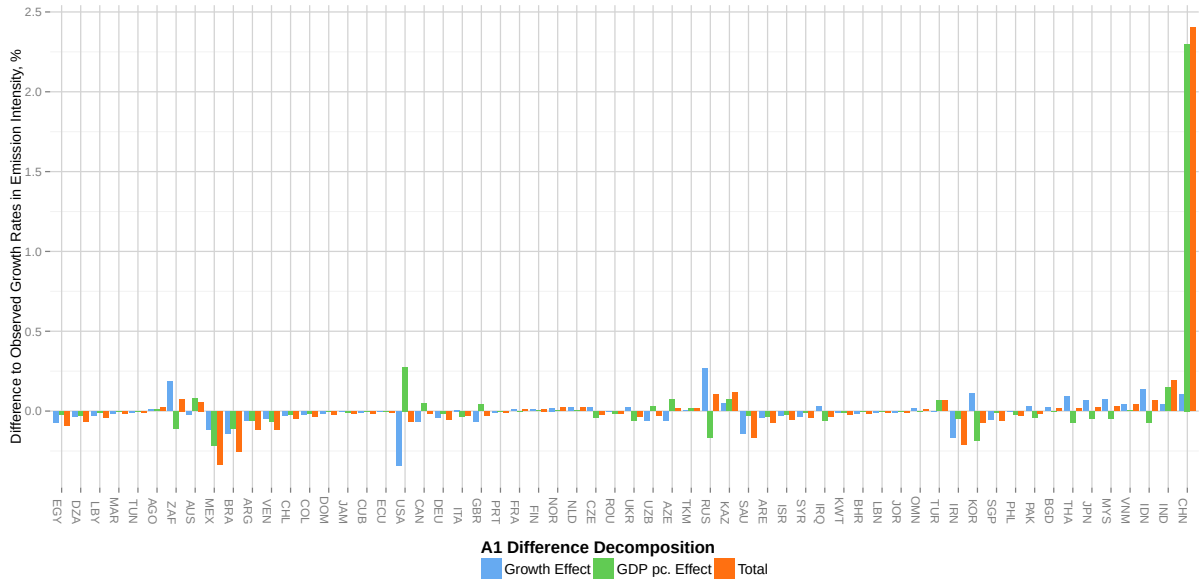


Figure 57: Country-by-country contribution to the Global Difference between Observations and Scenario Projections (A1) in Growth Rates 2000-10 of Emission Intensity Decomposed into CO_2 intensity Growth (blue), GDP (green) and Total Effects (orange).

5.6.4 *Observed Country-by-Country Growth Rates of Emission Intensity*

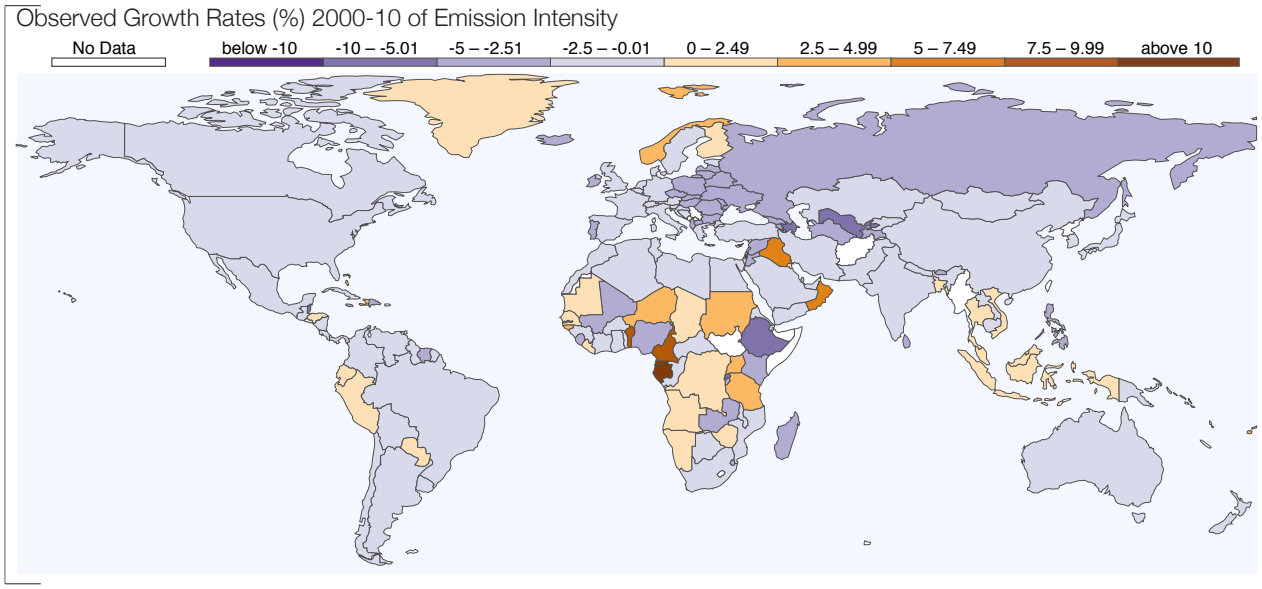


Figure 58: Observed average decadal growth rates of Emission Intensity (CO_2 per GDP) by country over 2000-2010.

5.6.5 *Fossil-fuel CO_2 Emissions per GDP per Capita*

Here we present results of emission intensity measured as fossil-fuel CO_2 emissions per GDP per capita. Consistent with the results in the main text, decadal emission intensity per capita growth over the 2000s exceeded all IS92 and SRES marker scenarios (see Figure 59).

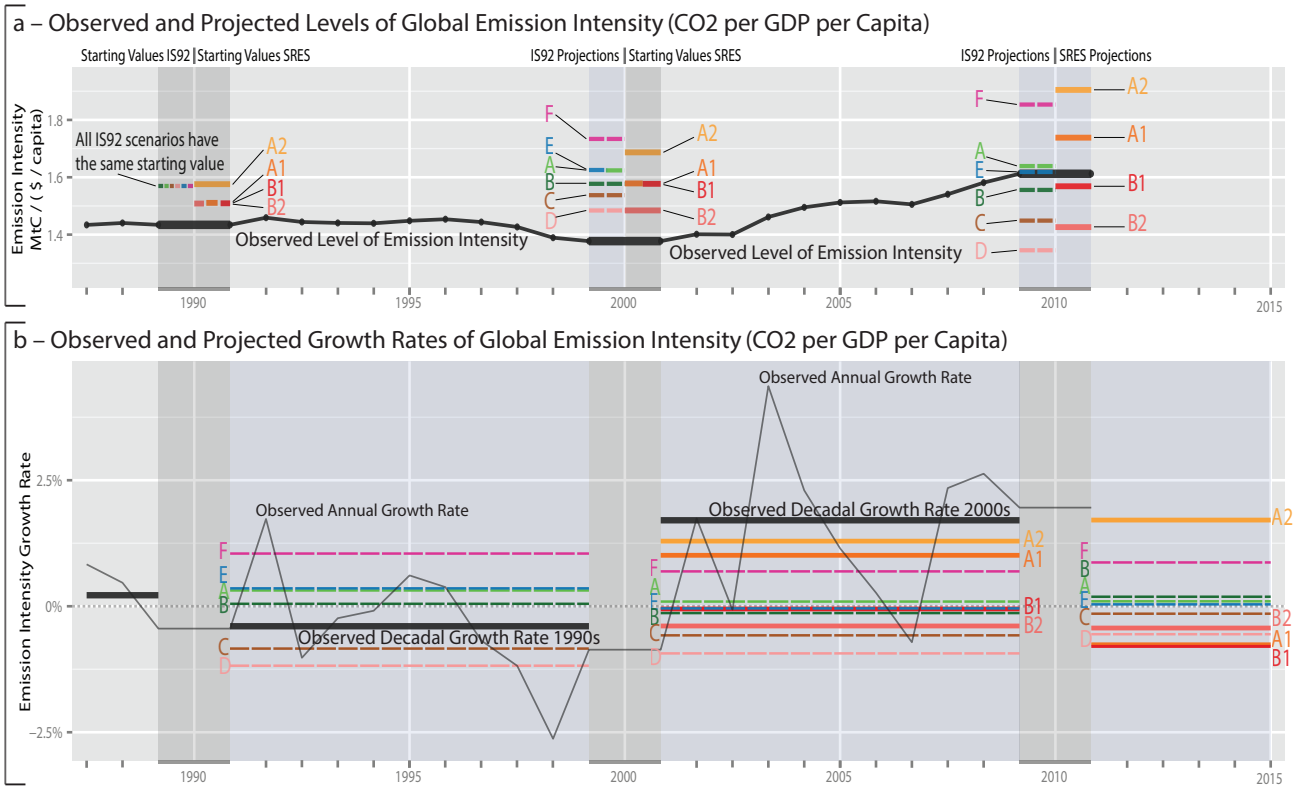


Figure 59: Observed and projected global emission intensity measured in fossil-fuel CO_2 emissions per GDP per capita in levels (a) and growth rates (b). Panel a graphs global observed annual fossil-fuel CO_2 emissions per GDP (black) together with decadal IS92 (dashed colour) and SRES marker projections (solid colour). Starting values are shaded grey while projected values are shaded light blue. Note that initial values for SRES vary across scenarios. Panel b shows observed annual growth rates (continuous black) together with observed decadal growth rates (horizontal black) over both decades. Projected growth rates are shown in colour for IS92 (dashed) and SRES marker projections (solid). Observed decadal growth rates exceed all scenario projections over the 2000s.

6

EPILOGUE

From the modelling of stochastic trends, to variable selection, the detection of breaks, and evaluation of scenarios – the proposed econometric tool-kit to model non-stationary climate data is vast. If we zoom out and look at the big picture, there is a natural progression in the sequence of this thesis.

Following an overview of hazards in econometric models of climate data, I show how a physical system can be mapped to a cointegrated vector autoregression (CVAR), providing a system approach to the estimation of relationships between temperatures, ocean heat, and radiative forcing. Exploring the restrictions in this CVAR estimated energy balance model (EBM), it becomes apparent that a congruent model of radiative forcing is crucial. I then explore how one of the radiative forcing factors, atmospheric CO_2, can be modelled endogenously without prior restrictions on which contributing variables are relevant. The model selection approach introduces indicator saturation methods in the form of impulses to the modelling of climate data. Based on the work on step indicator saturation developed alongside this thesis (see Doornik et al. 2013, Castle et al. 2015), I generalize the principle of indicator saturation to designed break functions - treating the detection of breaks as an aim in itself. All climate models, ranging from small EBMs (e.g. chapters 2 and 4) to large-scale simulated climate models require socio-economic projections to make statements about future climate change. The large span of projected temperature changes then originates predominantly from the wide range of scenarios, rather than uncertainty in climate models. The final chapter assesses how the initial scenarios have fared so far to establish on which trajectory the world is developing.

In exploring econometric methods and applications in modelling non-stationary climate data, the thesis makes contributions to both econometric methodology and its applications.

6.1 CONTRIBUTIONS TO ECONOMETRIC METHODOLOGY

Setting the stage by highlighting methodological hazards in econometric models of climate change, the first chapter emphasizes the importance of unmodelled shifts, stochastic trends, and consistency with known physical laws when working with climate data.

The second chapter shows that a two-component energy balance model can be mapped to a restricted cointegrated VAR providing a physical basis for the use of CVARs in modelling climate data. It offers testable restrictions and a system-based methodology for estimating physical models. This places the entire tool-kit of CVARs at the disposal of energy balance models, ranging from tests of fit or stability, to forecasts. It also sheds light on the dynamics of energy balance models themselves – they are implicitly equilibrium correction models, from which forecasts have to be made with caution (Hendry 2006).

The third chapter introduces automatic model selection with more variables than observations and impulse indicator saturation (IIS) to modelling of climate data. General-to-specific automatic model selection for monthly changes in atmospheric CO_2 concentrations yields

congruent models from which the level can be recovered, contributions decomposed, and the gauge of selection controlled.

Generalizing the principle of break detection (based on IIS and step indicator saturation - SIS) as a model selection problem, the fourth chapter demonstrates the feasibility of detecting breaks using designed functions. Relying on a split-half approach, theoretical results show that the method yields unbiased estimates of the break date and magnitude. The coefficient on the correct break function follows an approximate normal distribution where the variance scaling factor can be computed *a-priori*. Theory results are confirmed using Monte-Carlo simulations, which further show that using a multi-path algorithm (e.g. *Autometrics*) improves detection. A general-to-specific approach also outperforms shrinkage-based methods when multiple breaks are present. A first investigation into the uncertainty of the break timing computes the theoretical power (confirmed by simulations) for mis-timed break functions. The gauge in selecting over break functions is easily controlled through the nominal significance level of selection. The approach also holds promise for robust forecasting during breaks acting as a continuous intercept correction. The chapter provides the basis for the study of designed break functions (derived from theory or prior observations), as well as for breaks in more commonly used deterministic terms such as linear trends.

The fifth chapter demonstrates how aggregate differences in growth rates can be decomposed into individual contributions when down-scaling is necessary due to a coarser resolution of the projected values relative to observational data.

6.2 CONTRIBUTIONS OF APPLICATIONS

The argument against anthropogenic global warming based on an apparent lack of polynomial cointegration between observed global mean temperatures and radiative forcing of GHGs made by Beenstock et al. (2012) does not stand up to scrutiny, but is rather a modelling artefact, as the first chapter illustrates. The chapter further assesses the magnitude of the impact of the Montreal protocol on the recent observed slowdown in warming – while it is likely to have played a role, the magnitude of the effect cannot account for the entirety of the observed slowdown in warming.

The second chapter estimates a complete system model of global mean climate taking ocean-atmosphere interactions into account. Estimates of the model show that observations cointegrate consistently with theory. Likelihood ratio tests of the model's restrictions highlight one of the main challenges in estimating energy balance models – the modelling of the radiative forcing series. Endogenizing forcing leads to non-rejection of the restrictions in the CVAR, and reduces uncertainty in parameter estimates, however, break-like volcanic eruptions need to be controlled for. Model results also suggest that previous observation-based estimates of the temperature response to an increase in CO_2 may be artificially low due to model misspecification (e.g. apparent through serial correlation in the error terms).

Modelling a single forcing factor, atmospheric CO_2, the third chapter determines contributions using high-frequency observations without prior restrictions on the inclusion of variables. Natural factors are necessary but not sufficient in explaining the long-run underlying trend. Anthropogenic factors captured through industrial production (together with temperature effects) explain the long-run evolution.

The fourth chapter shows that the detection of historical volcanic eruptions without prior knowledge of their occurrence using theory-derived designed break functions is feasible. The false-positive rate is easily controlled and the method yields a high detection frequency. This holds promise for the creation of an updated record of eruptions relying on real-world proxy-based temperature reconstructions, and general applications of designed break functions such as robust forecasting through recessions based on continuous intercept corrections.

The fifth chapter shows that the growth rate in fossil fuel CO_2 emission intensity, fossil fuel CO_2 emissions per GDP, over the 2000s was rising while all main IPCC (Intergovernmental Panel on Climate Change) emission scenarios envisaged a decline. Equally, growth in observed emission intensity per capita and observed CO_2 emissions alone lie at the top end of scenarios. While no single scenario dominates others in performance, the trajectory of underestimating emission intensity raises concerns about achieving a world of economic prosperity in an environmentally sustainable fashion.

6.3 WHERE TO GO FROM HERE

Linking statistical and physical system models is a new development, allowing climate models to be estimated empirically, consistent with physical laws, and tested for stability to assess uncertainties. Pilot work in this thesis shows that a two-component energy-balance model of global mean climate can be mapped to a cointegrated system. This provides the basis for linking physical and econometric models and is generalizable to larger systems. Such systems allow the socio-economic side to be modelled directly and provide a data-based alternative to theory-based integrated assessment models to estimate feedbacks on the economy.

Statistical models of human effects on climate, such as the EBM in Chapter 2, mainly rely on global averages ignoring the spatial dimension. Spatially-resolved gridded data (e.g. using the EDGAR 1996 database now covering almost 25 years) can be used to estimate spatial models of human influence on climate using emissions, precipitation and temperature series. This permits estimation of local policy effects, as well as testing robustness of global models. This will also spawn new methodology since although spatial panel procedures exist, spatial panel cointegration handling stochastic trends and break detection have not been applied jointly.

By saturating models with full sets of break functions then eliminating insignificant ones, multiple breaks can be detected. However, many challenges remain to be solved. In particular what remains open are: the properties of indicator saturation using step shifts and general designed break functions when series exhibit stochastic trends; a complete quantification of the uncertainty on the break dates; an evaluation of the performance of indicator saturation in a cointegrated systems; and an assessment of how more variables than observations should be split into sub-blocks for estimation. Equally, the IS models so far have relied on linearity in parameters, in a working paper I have explored the joint modelling of non-linear functions and break detection using semi-parametric estimation. The results hold promise for a generalization of IS to non-linear models using series and other non-parametric estimators. Crucially these methods conduct variable selection jointly with break detection. Such methods can establish the impact of shocks on observed variables and assess whether policy changes are effective.

Historical data can be unreliable from measurement changes, and using these new break detection methods to establish changes in measurement will reduce the uncertainty about past climate change by correcting historical records. Research will continue following requests of collaboration with the UK MET office, the National Center for Atmospheric research, and building on the already established work with L. Schneider (Johannes Gutenberg University) and J. Smerdon (Columbia University) on the detection of eruptions. Volcanic eruptions have had devastating impacts on peoples lives, both locally and globally, yet there is uncertainty about the precise timing of eruptions. Chapter 4 established the feasibility of modelling temperature responses to volcanic eruptions as structural breaks. Using reconstructed observed temperatures, work with J. Smerdon and L. Schneider will lead to an updated record of historical eruptions. Historical surface temperature data are unreliable and plagued by measurement changes: e.g. a change from bucket to engine measurements introduced a step shift in sea surface temperature records. Measurement errors of this type are prevalent throughout temperature records – concerns about this have spawned the International Surface Temperature Initiative (ISTI) organised in part through the UK MET office. Indicator-based break detection methods hold promise in identifying changes in measurements, ultimately resulting in more accurate temperature records, including one of the main indicators of climate change, the global mean surface temperature anomaly. Break detection can also single out explanations for observed phenomena, including testing for effects of the recent economic recession on climate.

Chapter 5 shows that growth in CO_2 emissions per real GDP was under-projected in long-run climate scenarios. Further work will investigate how non-probabilistic scenarios should be assessed and study whether the scenarios exhibit systematic forecast biases. Break detection as forecast evaluation has been applied by Ericsson et al. (2014) for debt but not for climate scenarios: comparisons of IPCC estimates with other agencies (e.g. the IMF) will assess their relative performance and lead to recommendations for future scenarios. Crucially, only once EBMs are mapped to statistical systems can these additional features be tackled jointly.

Econometric tools can augment climate research: future research can build on the established physical science basis for their use which allows models to be estimated, tested and applied for policy evaluation using global and spatially resolved data. IS methods will evaluate the impacts of policy, and improve the historical record. Forecast evaluation of socio-economic projections can potentially reduce the uncertainty about future climate change.

Allen, M. R., Frame, D. J., Huntingford, C., Jones, C. D., Lowe, J. A., Meinshausen, M., & Meinshausen, N. (2009). Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, *458*(7242), 1163–1166.

Allen, M. R., Mitchell, J. F., & Stott, P. A. (2013). Test of a decadal climate forecast. *Nature Geoscience*, *6*(4), 243–244.

Anchukaitis, K. J., Breitenmoser, P., Briffa, K. R., Buchwal, A., Büntgen, U., Cook, E. R., ... others (2012). Tree rings and volcanic cooling. *Nature Geoscience*, *5*(12), 836–837.

Arrhenius, S. A. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science (fifth series)*, *41*, 237–275. (http://www.globalwarmingart.com/images/1/18/Arrhenius.pdf)

Australian Bureau of Meteorology. (2011). *Southern Oscillation Index Archive.* Available on-line [http://www.bom.gov.au/climate/current/soihtm1.shtml].

Bårdsen, G., Hurn, S., & McHugh, Z. (2010). *Asymmetric unemployment rate dynamics in Australia* (CREATES Research Paper 2010-2). Denmark: Aarhus University.

Bacastow, R. B. (1976). Modulation of atmospheric carbon dioxide by the southern oscillation. *Nature*, *261*, 116–118.

Bacastow, R. B., Keeling, C. D., & Whorf, T. P. (1985). Seasonal amplitude increase in atmospheric CO2 concentration at Mauna Loa, Hawaii 1959-1982. *Journal of Geophysical Research*, *90*, 10,529–10,540.

Baffes, J. (1997). Explaining stationary variables with non-stationary regressors. *Applied Economics Letters*, *4:1*, 69–75.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*, 47–78.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*, 1–22.

Banerjee, A., Dolado, J., Hendry, D. F., & Smith, G. (1986). Exploring equilibrium relationships in econometrics through statistical models: some monte carlo evidence. *Oxford Bulletin of Economics and Statistics*, *48*, 253-277.

Banerjee, A., Dolado, J. J., Galbraith, J. W., & Hendry, D. F. (1993). *Co-integration, error correction and the econometric analysis of non-stationary data*. Oxford: Oxford University Press.

Beenstock, M., Reingewertz, Y., & Paldor, N. (2012). Polynomial cointegration tests of anthropogenic impact on global warming. *Earth System Dynamics*, *3*(2), 173–188.

Berenguer-Rico, V., & Gonzalo, J. (2014). Summability of stochastic processes - a generalization of integration for non-linear processes. *Journal of Econometrics*, *178*, 331–341.

Bergamelli, M., & Urga, G. (2013). Detecting multiple structural breaks: a monte carlo study and application to the fisher equation for US. *Discussion Paper: Cass Business School, London*.

Boden, T., Marland, G., & Andres, R. (2013). *Global, regional, and national fossil-fuel CO_2 emissions.* Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. doi 10.3334/CDIAC/00001_V2013.

Bontemps, C., & Mizon, G. E. (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics*, *70*, 721–750.

Boucher, O., & Reddy, M. (2008). Climate trade-off between black carbon and carbon dioxide emissions. *Energy Policy*, *36*(1), 193–200.

Brohan, P., Allan, R., Freeman, E., Wheeler, D., Wilkinson, C., & Williamson, F. (2012). Constraining the temperature history of the past millennium using early instrumental observations. *Climate of the Past Discussions*, *8*(3), 1653–1685.

Buermann, W., Lintner, B. R., Koven, C. D., Angert, A., Pinzon, J. E., Tucker, C. J., & Fung, I. Y. (2007). The changing carbon cycle at Mauna Loa observatory. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 4249–4254.

Bureau of Labour Statistics. (2011). *US Consumer Price Index, CPI.* Available on-line [http://www.bls.gov/cpi/].

Caceres, C., & Nielsen, B. (2005). *Asymptotic properties of White's test for heteroskedasticity*. Unpublished doctoral dissertation, University of Oxford.

Campos, J., Ericsson, N. R., & Hendry, D. F. (2005). Editors' introduction. In J. Campos, N. R. Ericsson, & D. F. Hendry (Eds.), *Readings on general-to-specific modeling* (pp. 1–81). Cheltenham: Edward Elgar.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2011a). Evaluating automatic model selection. *Journal of Time Series Econometrics*, *3 (1)*, DOI: 10.2202/1941-1928.1097.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2011b). *Model selection in equations with many 'small' effects* (Discussion Paper 528). Oxford University: Economics Department.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2011c). Model selection when there are multiple breaks. *Journal of Econometrics*, *169:2*, 239-246.

Castle, J. L., Doornik, J. A., Hendry, D. F., & Pretis, F. (2015). Detecting locations shifts by step-indicator saturation during model selection. *forthcoming in Econometrics*.

Castle, J. L., & Hendry, D. F. (2010). A low-dimensional portmanteau test for non-linearity. *Journal of Econometrics*, *158*, 231-245.

Castle, J. L., & Hendry, D. F. (2011). *Checking the robustness and validity of model selection: An application to UK wages.* (Working Paper)

Castle, J. L., & Shephard, N. (Eds.). (2009). *The methodology and practice of econometrics*. Oxford: Oxford University Press.

Cavaliere, G., Nielsen, B., Heino, & Rahbek, A. (2014). Bootstrap testing of hypotheses on co-integration relations in var models. *Econometrica*, *forthcoming*.

Cavaliere, G., Rahbek, A., & Taylor, A. (2012). Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica*, *80*(4), 1721–1740.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*, 591–605.

Christopher, S. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., ... Rios, A. F. (2007). The oceanic sink for anthropogenic CO2. *Science*, *305*, 367–371.

Clark, W. C. (Ed.). (1982). *Carbon dioxide review*. New York: Oxford University Press.

Clements, M. P., & Hendry, D. F. (1999). *Forecasting non-stationary economic time series*. Cambridge, Mass.: MIT Press.

Clements, M. P., & Hendry, D. F. (2005). Evaluating a model by forecast performance. *Oxford*

Bulletin of Economics and Statistics, *67*(s1), 931–956.

Consensus Economics. (2013). Asia-pacific consensus forecasts.

Cox, D. R., & Snell, E. J. (1974). The choice of variables in observational studies. *Applied Statistics*, 51–59.

Crowley, T. J., & Unterman, M. B. (2012). Technical details concerning development of a 1200-yr proxy index for global volcanism. *Earth System Science Data Discussions*, *5*(1), 1–28.

DESA, UN. (2014). World population prospects, the 2012 revision. *New York: Department for Economic and Social Affairs*.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, *49*, 1057–1072.

Donoho, D. L., Tsaig, Y., Drori, I., & Starck, J. (2006). *Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursiut.* Available on-line [http://www.cs.tau.ac.il/ idrori/StOMP.pdf].

Doornik, J. A. (Ed.). (2006). *Ox*. London: Timberlake Consultants Ltd.

Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics*, *70*, 915–925.

Doornik, J. A. (2009a). Autometrics. In J. L. Castle & N. Shephard (Eds.), (pp. 88–121). Oxford: Oxford University Press.

Doornik, J. A. (2009b). An object-oriented matrix programming language Ox 6.

Doornik, J. A. (2010a). *Econometric model selection with more variables than observations* (Working paper). University of Oxford: Economics Department.

Doornik, J. A. (2010b). *Oxmetrics software package.* Timberlake Consultants [http://www.timberlake.co.uk/software/?id=64].

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, *70*, 927–939.

Doornik, J. A., & Hendry, D. F. (2009). *Empirical econometric modelling using PcGive: Volume I*. London: Timberlake Consultants Press.

Doornik, J. A., & Hendry, D. F. (2012). *Automatic model selection of multivariate dynamic econometric models.* 11^{th} OxMetrics Users Conference Presentation. Available on-line [http://www.gwu.edu/ forcpgm/OxMetrics2012-Papers/Doornik-OxMetricsUserConfererenc-March2012-Abstract.pdf].

Doornik, J. A., Hendry, D. F., & Nielsen, B. (1998). Inference in cointegrated models: UK M1 revisited. *Journal of Economic Surveys*, *12*, 533–572.

Doornik, J. A., Hendry, D. F., & Pretis, F. (2013). Step-indicator saturation. *Oxford Economics Deprtament Working Paper*, *658*.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of statistics*, *32*(2), 407–499.

Elliott, G., & Müller, U. K. (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics*, *141*(2), 1196–1218.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation,

estimation, and testing. *Econometrica*, *55:2*, 251-276.

Enting, I. G. (1987). The interannual variation in the seasonal cycle of carbon dioxide concentration at Mauna Loa. *Journal of Geophysical Research*, *92*, 5497–5504.

Environmental Protection Agency - EPA. (2011). *Inventory of U.S. Greenhouse Gas Emissions and Sinks 1990-2009.* Available on-line [http://www.epa.gov/climatechange/emissions/downloads11/US-GHG-Inventory-2011-Complete-Report.pdf].

Erickson, D. J., Mills, R. T., Gregg, J., J. Blasing, T. J., Hoffmann, F. M., Andres, R. J., ... Kawa, S. R. (2008). An estimate of monthly global emissions of anthropogenic CO2: Impact on the seasonal cycle of atmospheric CO2. *Journal of Geophysical Research*, *113*.

Ericsson, N. R. (2011). Justifying empirical macro-econometric evidence in practice. *Journal of Economic Surveys: Communication with Economists Conference*, –.

Ericsson, N. R. (2012). Detecting crises, jumps, and changes in regime. *Working Paper, Board of Governors of the Federal Reserve System, Washington, DC*.

Ericsson, R., N, Fiallos, J., E, & Seymour, J. E. (2014). Assessing greenbook forecasts of foreign GDP growth. *Federal Reserve Board Working Paper*.

Estrada, F., Perron, P., & Martínez-López, B. (2013). Statistically derived contributions of diverse human influences to twentieth-century temperature changes. *Nature Geoscience*, *6*, 1050-1055.

Federal Reserve. (2011a). *Industrial Production and Capacity Utilization.* Available on-line [http://www.federalreserve.gov/releases/g17].

Federal Reserve. (2011b). *West Texas Intermediate Crude Oil Spot Price.* Available on-line [http://research.stlouisfed.org/fred2/series/OILPRICE].

Francey, R. J., Tanis, P. P., Allison, C. E., Enting, I. G., White, J. W. C., & Trolier, M. (1995). Changes in oceanic and terrestrial carbon uptake since 1982. *Nature*, *373*, 326–330.

Friedlingstein, P., Andrew, R., Rogelj, J., Peters, G., Canadell, J., Knutti, R., ... others (2014). Persistent growth of co2 emissions and implications for reaching climate targets. *Nature Geoscience*, *7*(10), 709–715.

Gao, C., Robock, A., & Ammann, C. (2008). Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *113*(D23).

Gerlach, T. M. (2011). Changes in oceanic and terrestrial carbon uptake since 1982. *EOS, Transactions, American Geophysical Union*, *92*, 201–202.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, *46*, 1303–1313.

González, A., & Teräsvirta, T. (2008). Modelling autoregressive processes with a shifting mean. *Studies in Nonlinear Dynamics & Econometrics*, *12*(1).

Government of India, Ministry of Statistics. (2011). *Index of Industrial Production (IIP).* Available on-line [http://mospi.nic.in/].

Granados, J. A. T., Ionides, E. L., & Carpintero, O. (2009). *A threatening link between world economic growth and atmospheric CO2 concentrations.* (Working Paper)

Granados, J. A. T., Ionides, E. L., & Carpintero, O. (2011). *Climate change and the world economy: Short-run and long-run determinants of amtospheric CO2.* (Working Paper)

Gregory, J., Stouffer, R., Raper, S., Stott, P., & Rayner, N. (2002). An observationally based estimate of the climate sensitivity. *Journal of Climate*, *15*(22), 3117–3121.

Hansen, J., & Lebedeff, S. (1987). Global trends of measured surface air temperature. *Journal of Geophysical Research*, *92*, 13,345–13,372.

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, *48*(4), RG4004.

Hansen, J., Ruedy, R., Sato, M., & Reynolds, R. (1996). Global surface air temperature in 1995: Return to pre-pinatubo level. *Geophysical Research Letters*, *23*(13), 1665–1668.

Hansen, J., Russel, G., Rind, D., Stone, P., Lacis, A., Lebedeff, S., ... Travis, L. (1983). Efficient three-dimensional global models for climate studies: Models I and II. *Monthly Weather Review*, *111*, 609–662.

Hansen, J., Sato, M., Kharecha, P., Beerling, D., Berner, R., Masson-Delmotte, V., ... Zachos, J. C. (2008). Target atmospheric CO2: Where should humanity aim? *The Open Atmospheric Science Journal*, *2*, 217–231.

Harbo, I., Johansen, S., Nielsen, B., & Rahbek, A. (1998). Asymptotic inference on cointegrating rank in partial systems. *Journal of Business & Economic Statistics*, *16:4*, 388-399.

Harris, D. (1997). Principal component analysis of cointegrated time series. *Econometric Theory*, *13:4*, 529–557.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, *23*(9), 2418–2427.

Hendry, D. F. (1995). *Dynamic econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics*, *135*(1), 399–426.

Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In T. C. Mills & K. D. Patterson (Eds.), *Palgrave handbook of econometrics* (pp. 3–67). Basingstoke: Palgrave MacMillan.

Hendry, D. F. (2011). Climate change: Possible lessons for our future from the distant past. In S. Dietz, J. Michie, & C. Oughton (Eds.), *The political economy of the environment* (pp. 19–43). London: Routledge.

Hendry, D. F., & Johansen, S. (2011). *The properties of model selection when retaining theory variables* (Discussion Paper). University of Copenhagen: Economics Department.

Hendry, D. F., & Johansen, S. (2014). Model discovery and Trygve Haavelmo's legacy. *Econometric Theory*, *forthcoming*.

Hendry, D. F., Johansen, S., & Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, *23*, 337-339.

Hendry, D. F., & Juselius, K. (2000). Explaining cointegration analysis: Part I. *Energy Journal*, *21*, 1–42.

Hendry, D. F., & Juselius, K. (2001). Explaining cointegration analysis: Part II. *Energy Journal*, *22*, 75–120.

Hendry, D. F., & Krolzig, H.-M. (2003). New developments in automatic general-to-specific modelling. In B. P. Stigum (Ed.), *Econometrics and the philosophy of economics* (pp. 379–419). Princeton: Princeton University Press.

Hendry, D. F., & Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, *115*, C32–C61.

Hendry, D. F., & Mizon, G. E. (1978). Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *The Economic Journal*, *88*, 549-563.

Hendry, D. F., & Mizon, G. E. (2011). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, *3 (1)*, DOI: 10.2202/1941-1928.1100.

Hendry, D. F., & Pretis, F. (2013). Anthropogenic influences on atmospheric CO2. *Handbook on Energy and Climate Change*, 287.

Hendry, D. F., & Richard, J.-F. (1989). Recent developments in the theory of encompassing. In B. Cornet & H. Tulkens (Eds.), *Contributions to operations research and economics. the xxth anniversary of core* (pp. 393–440). Cambridge, MA: MIT Press.

Hendry, D. F., & Santos, C. (2010). An automatic test of super exogeneity. In M. W. Watson, T. Bollerslev, & J. Russell (Eds.), *Volatility and time series econometrics* (pp. 164–193). Oxford: Oxford University Press.

Hepburn, C., & Stern, N. (2008). A new global deal on climate change. *Oxford Review of Economic Policy*, *24*(2), 259–279.

Hofman, D. J., Butler, J. H., & Tans, P. P. (2009). A new look at atmospheric carbon dioxide. *Atmospheric Environment*, *43*, 2084–2086.

Höök, M., Sivertsson, A., & Aleklett, K. (2010). Validity of the fossil fuel production outlooks in the IPCC emission scenarios. *Natural Resources Research*, *19*(2), 63–81.

Hoover, K. D., & Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, *2*, 167–191.

IPCC. (2013). *Fith assessment report: Climate change 2013: Working Group I Report: The Physical Science Basis*. Geneva: IPCC. Retrieved from `https://www.ipcc.ch/report/ar5/wg1/`

IPCC. (2014). *Data distribution center.* Online [http://sedac.ipcc-data.org/ddc/sres/index.html and http://sedac.ipcc-data.org/ddc/is92/index.html].

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, *12*, 231-254.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.

Johansen, S., & Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In J. L. Castle & N. Shephard (Eds.), (pp. 1–36). Oxford: Oxford University Press.

Johansen, S., & Nielsen, B. (2013). Outlier detection in regression using an iterated one-step approximation to the huber-skip estimator. *Econometrics*, *1*(1), 53–70.

Johansen, S., & Nielsen, B. (2014). Outlier detection algorithms for least squares time series regression. *Available at SSRN 2510281*.

Jones, C. D., & Cox, P. M. (2005). On the significance of atmospheric CO2 growth rate anomalies in 2002-2003. *Journal of Geophysical Research*, *32*, –.

Juselius, K. (2006). *The cointegrated VAR model: Methodology and applications*. Oxford: Oxford University Press.

Juselius, K. (2011). Co-integration analysis of climate change: An exposition. *Working Paper*.

Kaufmann, R. K., & Juselius, K. (2013). Testing hypotheses about glacial cycles against the observational record. *Paleoceanography*, *28*(1), 175–184.

Kaufmann, R. K., Kauppi, H., Mann, M. L., & Stock, J. H. (2011). Reconciling anthropogenic climate change with observed temperature 1998–2008. *Proceedings of the National Academy of Sciences*, *108*(29), 11790–11793.

Kaufmann, R. K., Kauppi, H., Mann, M. L., & Stock, J. H. (2013). Does temperature contain a stochastic trend: linking statistical results to physical mechanisms. *Climatic change*, 1–15.

Kaufmann, R. K., Kauppi, H., & Stock, J. H. (2006). Emissions, concentrations, & temperature: a time series analysis. *Climatic Change*, *77*(3-4), 249–278.

Kaufmann, R. K., Paletta, L. F., Tian, H. Q., Myeni, R. B., & D'Arrigo, R. D. (2008). The power of monitoring stations and a CO2 fertilization effect: Evidence from causal relationships between NDVI and carbon dioxide. *Earth Interactions*, *12*, 1–23.

Kaufmann, R. K., & Stern, D. I. (2002). Cointegration analysis of hemispheric temperature relations. *Journal of Geophysical Research*, *107:2*.

Kawa, S. R., Erickson, D. J., Pawson, S., & Zhu, Z. (2004). Global CO2 transport simulations using meteorological data from NASA data assimilation system. *Journal of Geophysical Research*, *109*, D18312.

Keeling, C. D. (1973). Industrial production of carbon dioxide from fossil fuels and limestone. *Tellus*, *25:2*, 174–198.

Keeling, C. D., Bacastow, R. B., Brainbridge, A. E., Ekdahl, C. A., Guenther, J. P. R., & Waterman, L. S. (1976). Atmospheric carbon dioxide variations at Mauna Loa observatory, Hawaii. *Tellus*, *6*, 538–551.

Keeling, C. D., Chin, J. F., & Whorf, T. P. (1996). Increased activity of northern vegetation inferred from atmospheric CO2 measurements. *Nature*, *382*, 146–149.

Keeling, C. D., & Revelle, R. (1985). Effects of el niño/southern oscillation on the atmospheric content of carbon dioxide. *Meteoritics*, *20:2*, 437–450.

Keeling, C. D., Whorf, T. P., Whalen, M., & van der Plicht, J. (1995). Interannual extremes in rate rise of atmospheric carbon dioxide since 1980. *Nature*, *375*, 666–670.

Keeling, R. F., Piper, S. C., Bollenbacher, A. F., & Walker, S. J. (2008). *Atmospheric CO2 records from sites in the sio air sampling network.* In Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. Available on-line [http://cdiac.ornl.gov/trends/co2/sio-bar.html].

Kelly, P. M., & Sear, C. B. (1984). Climatic impact of explosive volcanic eruptions. *Nature*, *311*(5988), 740–743.

Kessler, M., & Rahbek, A. (2001). Asymptotic likelihood based inference for co-integrated homogenous gaussian diffusions. *Scandinavian Journal of Statistics*, *28*(3), 455–470.

Kessler, M., & Rahbek, A. (2004). Identification and inference for multivariate cointegrated and ergodic gaussian diffusions. *Statistical inference for stochastic processes*, *7*(2), 137–151.

Kohlmaier, G. H., Sire, E. O., Janecek, A., Keeling, C. D., Piper, S. C., & Revelle, R. (1989). Modelling the seasonal contribution of a CO2 fertilization effect of the terrestrial vegetation to the amplitude increase in atmospheric CO2 at Mauna Loa observatory. *Tellus*, *41B*, 487–510.

Kosaka, Y., & Xie, S.-P. (2013). Recent global-warming hiatus tied to equatorial pacific surface

cooling. *Nature*, *501*(7467), 403–407.

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., & Teng, H. (2013). Last millennium climate and its variability in ccsm4. *Journal of Climate*, *26*(4), 1085–1111.

Lansangan, J. R., & Barrios, E. B. (2009). Principal component analysis of nonstationary time series data. *Statistics and Computing*, *19:2*.

Leggett, J., Pepper, W. J., Swart, R. J., Edmonds, J., Meira Filho, L., Mintzer, I., & Wang, M. (1992). Emissions scenarios for the IPCC: an update. *Climate change*, 71–95.

Levin, I., Graul, R., & Trivett, N. B. A. (1995). Long-term observations of atmospheric CO2 and carbon isotopes at continental sites in Germany. *Tellus*, *47B*, 23–34.

Levitus, S., Antonov, J., Boyer, T., Baranova, O., Garcia, H., Locarnini, R., ... others (2012). World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophysical Research Letters*, *39*(10).

Levitus, S., Antonov, J., Boyer, T., Locarnini, R., Garcia, H., & Mishonov, A. (2009). Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophysical Research Letters*, *36*(7).

Lindzen, R. S., & Giannitsis, C. (1998). On the climatic implications of volcanic cooling. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *103*(D6), 5929–5941.

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., ... Wong, T. (2009). Toward optimal closure of the earth's top-of-atmosphere radiation budget. *Journal of Climate*, *22*(3).

Lucht, W., Prentice, I. C., Myeni, R. B., Sitch, S., Friedlingstein, P., Cramer, W., ... Smith, B. (2002). Climate control of the high-latitude vegetation greening trend and Pinatubo effect. *Science*, *296*, 1687–1689.

Machado, J. F., Marques, C. R., Neves, P. D., & da Silva, A. G. (2001). *Using the first principal component as a core inflation indicator.* (Banco de Portugal Working Paper 9-01)

Mann, M. E., Fuentes, J. D., & Rutherford, S. (2012). Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures. *Nature Geoscience*.

Manning, M., Edmonds, J., Emori, S., Grubler, A., Hibbard, K., Joos, F., ... others (2010). Misrepresentation of the ipcc co2 emission scenarios. *Nature Geoscience*, *3*(6), 376–377.

Marland, G., Boden, T. A., & Andres, R. (2011). *Global, regional, and national fossil fuel CO2 emissions.* In Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. (Available on-line [http://cdiac.ornl.gov/trends/emis/overview.html])

Marland, G., & Rotty, R. M. (1984). Carbon dioxide emissions from fossil fuels: a procedure for estimation and results for 1950-1982. *Tellus*, *36B*, 232–261.

Mass, F., Clifford, & Portman, A., David. (1989). Major volcanic eruptions and climate- a critical evaluation. *Journal of Climate*, *2*(6), 566–593.

McCormick, M. P., Thomason, L. W., Trepte, C. R., et al. (1995). Atmospheric effects of the Mt. Pinatubo eruption. *Nature*, *373*(6513), 399–404.

Meehl, G. A., Arblaster, J. M., Fasullo, J. T., Hu, A., & Trenberth, K. E. (2011). Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Climate Change*, *1*(7), 360–364.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M., Lamarque, J., ... others (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1-2), 213–241.

Mizon, G. E. (1995). A simple message for autocorrelation correctors: Don't. *Journal of Econometrics*, *69*(1), 267–288.

Mizon, G. E., & Richard, J. F. (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, *54*, 657–678.

Myeni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. *Transactions on Geoscience and Remote Sensing*, *33:2*, 481–486.

Myhre, G., Highwood, E. J., Shine, K. P., & Stordal, F. (1998). New estimates of radiative forcing due to well mixed greenhouse gases. *Geophysical research letters*, *25*(14), 2715–2718.

Myhre, G., Myhre, A., & Stordal, F. (2001). Historical evolution of radiative forcing of climate. *Atmospheric Environment*, *35*, 2361-2373.

Myhre, G., Shindell, D., Breon, F., Collins, W., Fuglestvedt, J., Huang, J., ... Zhang, H. (2013). *Anthropogenic and natural radiative forcing* (Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change). Cambridge, UK: Cambridge University Press.

Nakicenovic, N., & Swart, R. (2000). Special report on emissions scenarios. *Special Report on Emissions Scenarios, Edited by Nebojsa Nakicenovic and Robert Swart, pp. 612. ISBN 0521804930. Cambridge, UK: Cambridge University Press, July 2000.*, *1*.

NASA Goddard Institute for Space Studies (GISS). (2011). *GISS - Surface Temperature Analysis.* Available on-line [http://data.giss.nasa.gov/gistemp/].

Nevison, C. D., Mahowald, N. M., Doney, S. C., Lima, I. D., van der Werf, G. R., Randerson, J. T., ... McKinley, G. A. (2008). Contribution of ocean, fossil fuel, land biosphere, and biomass burning carbon fluxes to seasonal and interannual variability in atmospheric CO2. *Journal of Geophysical Research*, *113*.

Newell, N. D., & Marcus, L. (1987). Carbon dioxide and people. *Society for Sedimentary Geology*, *2:1*, 101–103.

OECD. (2011). *Composite Leading Indicators: MEI.* Available on-line [http://stats.oecd.org/Index.aspx].

Office of National Statistics. (2011). *Primary Production.* Available on-line [http://www.statistics.gov.uk/hub/business-energy/].

Olivier, J. G. J., Bouwman, A., Berdowski, J., Veldt, C., Bloos, J., Visschedijk, A., ... others (1996). Description of EDGAR version 2.0: A set of global emission inventories of greenhouse gases and ozone-depleting substances for all anthropogenic and most natural sources on a per country basis and on 1 degree x 1 degree grid.

Orr, C. J., Maier-Reimer, E., Mikolajewicz, U., Monfray, P., Sarmiento, J. L., Toggweiler, J. R., ... Boutin, J. (2001). Estimates of anthropogenic carbon uptake from four three-dimensional global ocean models. *Global Biogeochemical Cycles*, *15:1*, 43–60.

Otto, A., Otto, F. E., Boucher, O., Church, J., Hegerl, G., Forster, P. M., ... others (2013). Energy budget constraints on climate response. *Nature Geoscience*.

Pepper, W. (1992). *Emission scenarios for the IPCC: An update: Assumptions, methodology, and results*. US Environmental Protection Agency.

Perron, P., & Yabu, T. (2009). Testing for shifts in trend with an integrated or stationary noise

component. *Journal of Business & Economic Statistics*, *27*(3).

Perron, P., & Zhu, X. (2005). Structural breaks with deterministic and stochastic trends. *Journal of Econometrics*, *129*(1), 65–119.

Petit, J., Jouzel, J., Raynaud, D., Barkov, N., Barnola, J., Basile, I., ... Stievenard, M. (1999). *Climate and atmospheric history of the past 420,000 years from the vostok ice core, antarctica.* IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series 2001-076. NOAA/NGDC Paleoclimatology Program, Boulder CO, USA.

Phillips, P. C. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory*, *23*(04), 557–614.

Pretis, F., & Allen, M. (2013). Climate science: Breaks in trends. *Nature Geoscience*, *6*, 992-993.

Pretis, F., & Hendry, D. (2013). Some hazards in econometric modelling of climate change. comment on" polynomial cointegration tests of anthropogenic impact on global warming" by beenstock et al.(2012). *Earth System Dynamics*, *4*(2), 375–384.

Pretis, F., Mann, M. L., & Kaufmann, R. K. (2015). Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection. *forthcoming in Climatic Change*.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 350–371.

Randerson, T. J., Thompson, M. V., Conway, T. J., Fung, I. Y., & Field, C. B. (1997). The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide. *Global Biogeochemical Cycles*, *11:4*, 535–560.

Richardson, K., Steffen, W., & Liverman, D. (2011). *Climate change: Global risks, challenges and decisions*. Cambridge University Press.

Ritschard, R. L. (1992). Marine algae as a CO2 sink. *Water, Air and Soil Pollution*, *64*, 289–303.

Rowlands, D. J., Frame, D. J., Ackerley, D., Aina, T., Booth, B. B., Christensen, C., ... others (2012). Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nature Geoscience*, *5*(4), 256–260.

Rypdal, K. (2012). Global temperature response to radiative forcing: Solar cycle versus volcanic eruptions. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *117*(D6).

Salkever, D. S. (1976). The use of dummy variables to compute predictions, prediction errors and confidence intervals. *Journal of Econometrics*, *4*, 393–397.

Sargan, J. (1974). Some discrete approximations to continuous time stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 74–90.

Sato, M., Hansen, J. E., McCormick, M. P., & Pollack, J. B. (1993). Stratospheric aerosol optical depths, 1850–1990. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *98*(D12), 22987–22994.

Scheffer, M., Brovkin, V., & Cox, P. M. (2006). Positive feedback between global warming and atmospheric CO2 concentration inferred from past climate change. *Geophysical Research Letters*, *33*(10).

Schmidt, A., Thordarson, T., Oman, L. D., Robock, A., & Self, S. (2012). Climatic impact of the long-lasting 1783 laki eruption: Inapplicability of mass-independent sulfur isotopic composition measurements. *Journal of Geophysical Research: Atmospheres*, *117*(D23), n/a–n/a. Retrieved from `http://dx.doi.org/10.1029/2012JD018414` doi: 10.1029/2012JD018414

Schmidt, G., Jungclaus, J., Ammann, C., Bard, E., Braconnot, P., Crowley, T., ... others (2011). Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1. 0). *Geoscientific Model Development*, *4*(1).

Schwartz, S. E. (2007). Heat capacity, time constant, and sensitivity of earth's climate system. *Journal of Geophysical Research*, *112*(D24), D24S05.

Schwartz, S. E. (2012). Determination of earth's transient and equilibrium climate sensitivities from observations over the twentieth century: strong dependence on assumed forcing. *Surveys in geophysics*, *33*(3-4), 745–777.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Schwert, G. W. (1987). Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics*, *20*(1), 73–103.

Sims, C. A., Stock, J. H., & Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, *58*, 113–144.

Siroky, S. D. (2009). Navigating random forests and related advances in algorithmic modeling. *Statistics Surveys*, *3*, 147-163.

Stern, D. I. (2006). An atmosphere–ocean time series model of global climate change. *Computational statistics & data analysis*, *51*(2), 1330–1346.

Stern, D. I., & Kaufmann, R. K. (2000). Detecting a global warming signal in hemispheric temperature series: A structural time series analysis. *Climatic Change*, *47*, 411-438.

Stern, D. I., & Kaufmann, R. K. (2014). Anthropogenic and natural causes of climate change. *Climatic Change*, *122*(1-2), 257–269.

Stodden, V. (2006). *Model selection when the number of variables exceeds the number of observations.* Thesis. Available on-line [http://www.stanford.edu/ vcs/Papers.html].

Sundquist, E. T., & Keeling, R. F. (2009). The Mauna Loa carbon dioxide record: Lessons for long-term earth observations. , *183*, 27–35.

Taguchi, S., Murayama, S., & Higuchi, K. (2003). Sensitivity of inter-annual variation of CO2 seasonal cycle at Mauna Loa to atmospheric transport. *Tellus*, *55B*, 547–554.

Tans, P., & Keeling, R. (2011). *Mauna Loa, monthly mean carbon dioxide.* Scripps Institution of Oceanography. (scrippsco2.ucsd.edu/) and NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/) Available on-line [http://www.esrl.noaa.gov/gmd/ccgg/trends/].

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

Thoning, K. W., & Tans, P. P. (1989). Atmospheric carbon dioxide at Mauna Loa observatory 2. Analysis of the NOAA GMCC data, 1974-1985. *Journal of Geophysical Research*, *94*, 8549–8565.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58:1*, 267-288.

Tibshirani, R. (2011). Regression shrinkage and selection via the Lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273–282.

Trenberth, K. E., & Fasullo, J. T. (2010). Tracking earth's energy. *Science*, *328*(5976), 316–317.

Trenberth, K. E., & Fasullo, J. T. (2012). Tracking earth's energy: from El Niño to global warming. *Surveys in geophysics*, *33*(3-4), 413–426.

Troup, A. J. (1965). The 'Southern Oscillation'. *Quarterly Journal of the Royal Meteorological*

Society, *91*, 490–506.

Tucker, C. J., Pinzon, J., Brown, M., & GIMMS/GSFC/NASA. (2010). *ISLSCP II GIMMS Monthly NDVI, 1981-2002.* In Hall, Forrest G., G. Collatz, B. Meeson, S. Los, E. Brown de Colstoun, and D. Landis (eds.). ISLSCP Initiative II Collection. Data set. Available on-line [http://daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

van Vuuren, D. P., Edmonds, J., Smith, S. J., Calvin, K. V., Karas, J., Kainuma, M., ... others (2010). What do near-term observations tell us about long-term developments in greenhouse gas emissions? *Climatic change*, *103*(3-4), 635–642.

van Vuuren, D. P., Lucas, P. L., & Hilderink, H. (2007). Downscaling drivers of global environmental change: Enabling use of globalSRES scenarios at the national and grid levels. *Global Environmental Change*, *17*(1), 114–130.

Watanabe, M., Kamae, Y., Yoshimori, M., Oka, A., Sato, M., Ishii, M., ... Kimoto, M. (2013). Strengthening of ocean heat uptake efficiency associated with the recent climate hiatus. *Geophysical Research Letters*.

Watson, A. J., Nightingale, P. D., & Cooper, D. J. (1995). Modeling atmosphere ocean CO_2 transfer. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences*, *353*, 41–51.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817–838.

Wigley, T. M. L. (1983). The pre-industrial carbon dioxide level. *Climate Change*, *5*, 315–320.

World Bank. (2014). *World development indicators – economic growth and GDP*. World Bank Publications.

World Meteorological Organization. (2014). The state of greenhouse gases in the atmosphere based on global observations through 2013. *Greenhouse Gas Bulletin*, *10*.

Worrell, E., Price, L., Martin, N., Hendriks, C., & Meida, L. O. (2001). Carbon dioxide emissions from the global cement industry. *Annual Review of Energy and the Environment*, *26*, 303–329.

Zanchettin, D., Timmreck, C., Bothe, O., Lorenz, S. J., Hegerl, G., Graf, H.-F., ... Jungclaus, J. H. (2013). Delayed winter warming: A robust decadal response to strong tropical volcanic eruptions? *Geophysical Research Letters*, *40*(1), 204–209.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.