A multicointegration model of global climate change

BRUNS, Stephan; Csereklyei, Z & Stern, DI (2020) A multicointegration model of global climate change. In: JOURNAL OF ECONOMETRICS, 214 (1) , p. 175 -197.

DOI: 10.1016/j.jeconom.2019.05.010
Handle: http://hdl.handle.net/1942/30323

Bruns, Stephan B.; Csereklyei, Zsuzsanna; Stern, David I.

Working Paper

A multicointegration model of global climate change

cege Discussion Papers, No. 336

Provided in Cooperation with:
Georg August University of Göttingen, cege - Center for European, Governance and Economic Development Research

Suggested Citation: Bruns, Stephan B.; Csereklyei, Zsuzsanna; Stern, David I. (2018) : A multicointegration model of global climate change, cege Discussion Papers, No. 336, University of Göttingen, Center for European, Governance and Economic Development Research (cege), Göttingen

This Version is available at:
http://hdl.handle.net/10419/173787

Center for European, Governance and Economic Development Research
cege

Discussion Papers

Number 336 – January 2018

A MULTICOINTEGRATION MODEL OF GLOBAL CLIMATE CHANGE

Stephan B. Bruns, Zsuzsanna Csereklyei and David I. Stern

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

ISSN: 1439-2305

A Multicointegration Model of Global Climate Change

Stephan B. Bruns

Department of Economics, University of Göttingen, Humboldtallee 3, 37073 Göttingen, Germany. stephan.bruns@uni-goettingen.de. Phone: +49-551-39-21391.

Zsuzsanna Csereklyei

Crawford School of Public Policy, The Australian National University, 132 Lennox Crossing, Acton, ACT 2601, Australia. E-mail: zsuzsanna.csereklyei@anu.edu.au. Phone: +61-2-6125-2295.

David I. Stern

Crawford School of Public Policy, The Australian National University, 132 Lennox Crossing, Acton, ACT 2601, Australia. E-mail: david.stern@anu.edu.au. Phone: +61-2-6125-0176.

24 January 2018

Abstract

We model the role of the ocean in climate change, using the concept of multicointegration. Surface temperature and radiative forcing cointegrate and the accumulated cointegration disequilibria represent the change in Earth system heat content, which is predominantly stored in the ocean. System heat content in turn cointegrates with surface temperature. Using a multicointegrating I(2) model, we find that the climate sensitivity is 2.8ºC and the rate of adjustment to equilibrium is realistically slow. These results contrast strongly with those from I(1) cointegration models and are more consistent with global circulation models. We also estimate Earth system heat content as a latent variable for the full period, 1850-2014, and this predicted heat content cointegrates with available ocean heat content observations for 1940-2014.

JEL Codes: C32, Q54

Acknowledgements: We thank Steven Smith for sending us the data on emissions of sulfur dioxide and black and organic carbon and Lijing Cheng for sending us ocean heat content data. We thank Myles Allen, Tom Engsted, Niels Haldrup, David Hendry, Soren Johansen, Katarina Juselius, and Felix Pretis for very useful comments on our presentation at the conference "Econometric Models of Climate Change" in Aarhus, Denmark. We thank Boriss Siliverstovs for sharing his data so that we could replicate his paper as a check on our methods. We also thank the journal's three anonymous referees for their detailed comments. All remaining errors are ours.

1. Introduction

Though most research on the historical effects of anthropogenic greenhouse gases on the global climate compares the output of simulation models such as general circulation models (GCMs) and energy balance models (EBMs) with observed temperatures and other variables (Barnett *et al.*, 2005), applying econometric time-series techniques to modeling the global climate system and the effects of human activity on climate change is increasingly popular (e.g. Bindoff *et al.*, 2013; Beenstock *et al.*, 2012; Schmith *et al.* 2012; Pretis *et al.*, 2015; Estrada and Perron, 2016; Storelvmo *et al.*, 2016, Phillips *et al*., 2017). But most such research – apart from Stern (2006), Stern and Kaufmann (2014), and Pretis (2015) – focuses on the atmosphere alone, which, in the absence of an ocean, would adjust quickly to changes in forcing. Climate change is a long-term problem because the oceans and other relevant Earth systems adjust very slowly – over centuries or millennia. Atmospheric energy imbalances are stored as increases in ocean heat content (OHC), and in the long run, the temperature of the atmosphere must be in equilibrium with that of the ocean. Estimating the climate sensitivity using only relatively short atmospheric temperature time series without taking heat storage into account might, therefore, underestimate the climate sensitivity to increasing greenhouse gases and overestimate the rate of adjustment of temperature to forcing (Stern, 2006). In this paper, we model energy imbalances and heat storage using the concept of multicointegration. We find that the equilibrium climate sensitivity (ECS) is 2.8ºC with a 66% confidence interval of 2.2 to 3.7 ºC, and the rate of adjustment to equilibrium occurs over two centuries, mimicking, to some degree, that found in global circulation models.[1] These results contrast strongly with those from I(1) cointegration models, which have much more rapid adjustment to a lower estimated ECS. Our prediction of system heat content also cointegrates with observed ocean heat content.

Multicointegration models, first introduced by Granger and Lee (1989), are designed to handle long-run cointegrating relationships between variables where the accumulated cointegration

[1] "ECS is defined as the equilibrium change in annual mean global surface temperature following a doubling of the atmospheric CO_2 concentration… while TCR is defined as the annual mean global surface temperature change at the time of CO_2 doubling following a linear increase in CO_2 forcing over a period of 70 years" (Collins *et al*., 2013, 1110). The Earth System Sensitivity that includes very slow feedbacks may be 30-50% higher than typical estimates of the ECS (Lunt *et al*., 2010). Given that we use data for the last 165 years our estimates only incorporate feedbacks operating within this time scale and so we estimate the TCR and ECS in this paper.

error is itself cointegrated with the original variables.[2] Such a relationship is typically found for flow and stock variables where cointegration is found among the flows and between the flows and the accumulated cointegration disequilibria, which form a stock variable. For example, Granger and Lee (1989) and Engsted and Haldrup (1999) examine production, sales, and inventory in manufacturing, Engsted and Haldrup (1999) and Siliverstovs (2006) housing starts and unfinished stock, and Berenguer-Rico and Carrion-i-Silvestre (2011) government deficits and debt. In the case of the climate system, there is a long-run equilibrium between surface temperature and radiative forcing which represents the energy balance of the earth system.[3] A further long-run equilibrium exists between surface temperature and the accumulated deviations from this energy balance, which represents the stock of heat of the earth system. This stock of heat is predominantly stored in the ocean (Levitus *et al.*, 2012). Therefore, a multicointegration model can incorporate an ocean that heats or cools when the atmosphere is out of long-run equilibrium, and in which the ocean also reciprocally governs atmospheric temperature. This mechanism is identical to that in simple two-layer physical EBMs.

Much research in recent years has been devoted to explaining the so-called "hiatus" in surface warming that extended from 1998 to 2014 (e.g. Kaufmann *et al.*, 2011; Estrada *et al.*, 2013; Pretis *et al.*, 2015). Researchers have emphasized various factors that may have played a role in the hiatus including aerosols (Andersson *et al.*, 2015; Kaufmann *et al.*, 2011, Smith *et al.*, 2016), solar activity (Kaufmann *et al.* 2011), or even the measurement of surface temperature itself (Gramling, 2015; Karl *et al.*, 2015; Curry, 2014; Medhaug *et al.*, 2017). However, the recent consensus is that interchange of heat between the atmosphere and ocean (Fyfe *et al.* 2016; Tollefson, 2014; Trenberth and Fasullo, 2010, 2012; Balmaseda *et al.*, 2013; Glecker *et al.*, 2016; Kouketsu *et al.*, 2011; Meehl *et al.*, 2011; Palmer *et al.*, 2011, Watanabe *et al.*, 2013, Medhaug *et al*., 2017), which may be mediated by the decadal variability of oceans, in particular the Pacific Decadal Oscillation (Fyfe *et al.*, 2016; Smith *et al.*, 2016; Trenberth, 2015), largely

[2] Though previous research did not use the concept of multicointegration, this type of model has a long history in the control theory literature (e.g. Philips, 1954; Hendry and von Ungern-Sternberg, 1981).

[3] Both radiative forcing and surface temperature can also be interpreted as flow variables. Radiative forcing is expressed as watts per square meter and, therefore, reflects a rate per unit of time. It reflects the additional flow of radiation from the atmosphere to the surface due to the addition of greenhouse gases and other factors. Temperature is a property that governs the transfer of thermal energy, or heat, between one system and another. Therefore, it also implicitly reflects a flow rate, in this case the flow of long-wave radiation from the surface to the atmosphere. See Section 3 for a more detailed description.

explain the hiatus. Medhaug *et al*. (2017) argue that a combination of all the above factors is needed to explain the global warming hiatus. Therefore, it is timely to put forward an econometric approach that explicitly models the interaction of the atmosphere and ocean. In addition to estimating the climate sensitivity and the rate of adjustment of surface temperature to equilibrium from an increase in radiative forcing, we also use our model to attempt to attribute the hiatus in surface temperature increase.

A few existing papers using econometric methods consider the role of OHC in the climate change process. Stern and Kaufmann (2014) control for OHC when testing for Granger causality between radiative forcing and surface temperature. Pretis (2015) uses both surface temperature and OHC data and notes that disequilibria contribute to the increase in heat stored in the ocean. His focus is on showing the equivalence of EBM and vector autoregressive (VAR) models and the analysis is restricted to 1955-2011 due to limited availability of OHC data. Stern (2006) estimated the first time-series models that included both surface temperature and OHC using the concept of multicointegration to impose energy balance on the model. His model was based on the bivariate, two stage approach of Granger and Lee (1989) and do not exploit the fact that multicointegration implies that the data can be modeled by an I(2) VAR (Engsted and Johansen, 1999; Engsted and Haldrup, 1999). Because of a relatively short OHC series, Stern modeled OHC as a latent variable constrained by available observations using a state space model.

In this paper, we exploit the I(2) representation of the multicointegration model, which effectively allows us to estimate OHC as a latent variable. Compared to the length of the surface temperature time series (1850-2014) there are still only limited observations on OHC (1940-2014). Using the full 165 years of available data on surface temperature and radiative forcing we estimate a model that is not constrained by observations on OHC. Despite not using any data on OHC, the predicted heat content from this model matches simulated results from an energy balance model for 1850-2005 (Marvel *et al*., 2016) quite well and cointegrates with observed OHC (Cheng *et al*., 2017) for the available period since 1940. This empirical estimate of OHC may help future research to better understand the role of the ocean in climate change.

Even though we only use 165 years of data in our estimation, equilibrium in response to a doubling of carbon dioxide is not reached for more than two centuries. Our estimates of the ECS and the Transient Climate Response (TCR) are close to the consensus in the scientific literature. These results contrast strongly with those produced using I(1) cointegration models,

which have much more rapid adjustment to a lower estimated ECS. An I(1) model that includes observed OHC did not perform any better, which is likely due to the very short sample available. The primary advantages of our approach over such an I(1) model is that we can use a much longer time series and we can exploit the super-super consistency property of the I(2) estimator of the climate sensitivity.

The paper is structured as follows. The next section reviews recent relevant research. The third section introduces the model and methods. The fourth section of the paper examines the data and the fifth presents results. The sixth section examines the potential effect of measurement error on the results. The final section provides some conclusions.

2. Prior Research

2.1. The Role of the Oceans in Global Climate Change

At the time of writing this paper, the surface warming slowdown has ended, with record high global temperatures from 2015 to 2017. Nevertheless, the period between 1998 and 2014, when surface temperatures increased much more slowly than in the previous quarter century, has been the subject of intense scrutiny. As the search for the missing pieces of the puzzle began, a number of potential culprits surfaced.

Among the suggested candidates were an increase in anthropogenic sulfur emissions (Kaufmann *et al.*, 2011), declining solar irradiance (Tollefson, 2014, Trenberth 2015; Kaufmann *et al.*, 2011), and an increase in volcanic aerosols (Andersson *et al.*, 2016) over the examined period, which also coincided with a negative phase of the Pacific Decadal Oscillation (PDO). Similarly, anthropogenic sulfate aerosols contributed to the earlier hiatus period from the 1950s to the 1970s (Kaufmann *et al*., 2006). Smith *et al.* 2016 recently suggested that anthropogenic aerosol emissions might be a driver of the negative PDO. This is, however, in contrast with the findings of Kosaka and Xie (2013) who attribute the hiatus with high probability to internal variability, instead of forcing.

Karl *et al.* (2015) argue that the apparent hiatus was due to mismeasurement of surface temperature data. They correct the temperature data for several biases finding the resulting warming trends between 1950-1999 and 2000-2014 to be "virtually indistinguishable".[4] However, their approach was critiqued, by among others Fyfe *et al.* (2016) who argue that the

[4] See also Hausfather *et al*. (2017) who find cooling biases in recent years in several existing sea surface temperature series.

starting and ending dates of the observation period matter, as the 1950-1970 period also included a long hiatus.

The majority of recent studies do agree that exchange of heat between the atmosphere and the oceans is a key player in explaining the surface warming slowdown. Nonetheless, the mechanisms by which oceans absorb and then again release heat were not well understood until recently, when this process was found to be closely linked to the decadal oscillation of the oceans, in particular the Pacific Decadal Oscillation (PDO) (Kosaka and Xie, 2013; Meehl *et al.* 2011). Hiatuses might be relatively common climate occurrences, where enhanced heat uptake by the ocean is linked to La Nina-like conditions associated with a negative PDO (Meehl *et al.*, 2011). By contrast, the positive phase of the PDO favors El Nino conditions and injects heat into the atmosphere (Tollefson, 2014). Stronger trade winds during La Nina episodes drive warm surface water westwards across the Pacific, then down into the lower layers of the ocean. Simultaneously, cold water upwells in the eastern Pacific (Trenberth and Fasullo, 2012). It is also possible that extreme La Nina events, such as that in late 1998, may tip the ocean into a cool phase of the PDO (Tollefson, 2014). On the other hand, Hedemann *et al*. (2017) argue that internal variability of the top of the atmosphere energy imbalance may play just as large a role in hiatuses as internal variability of the ocean.

Observations of OHC are sparse prior to systematic observation through the Argo float system in recent decades. Therefore, the further we go back in time, the larger the uncertainty surrounding ocean heat uptake and the larger potential biases might be. The vast majority of warming is concentrated in the top 2000m of the ocean (Purkey and Johnson, 2010). Levitus *et al.* (2012) provide pentadal OHC estimates for the upper 2000 meters since the late 1950s (annual since 2005) and annual observations for the upper 700m from 1955. Cheng *et al*. (2017) estimate annual changes in ocean heat content for the top 700m and top 2000m as far back as 1940, but uncertainties are again high for the 20th Century.

Estimates for 1955-2010 (Levitus *et al.*, 2012) show a rate of heat uptake of 0.27Wm^{-2} for the upper 2000m of the world ocean but the uptake has varied over time. Half of the heat accumulated since 1865 accumulated after 1997 (Gleckler *et al.*, 2016). Balmaseda *et al.* (2013) estimate that the rate of heat uptake in the 2000s was 0.84 Wm^{-2} for the entire ocean with 0.21Wm^{-2} of that being stored below 700m, but in the 1990s uptake was negative (-0.18Wm^{-2})

though other sources find a lower but positive rate of uptake in that period.[5] However, during the recent hiatus period, the upper layers of the ocean did not show enough warming to account for the imbalance in the energy system (Balsameda *et al.* 2013). This "missing energy" was actually stored in the deep oceans (Trenberth and Fasullo, 2012). Cheng *et al*. (2017) estimate the annual rate of increase in OHC in the upper 700 meters of the ocean was 0.09Wm^{-2} from 1960 to 1991 and 0.61Wm^{-2} from 1992 to 2015. For the 700-2000m layer they estimate the rates of change at 0.04Wm^{-2} and 0.37Wm^{-2}, respectively.

Estimates of deep ocean heat fluxes are limited. Johnson *et al.* (2016) estimate net ocean heat uptake in the top 1800m of the ocean of 0.71Wm^{-2} from 2005 to 2015, and 0.07Wm^{-2} below 1800m. Kouketsu *et al.* (2011) calculate world ocean temperature changes for the 1990s and 2000s for waters below 3000m, estimating the heat uptake to be around 0.056Wm^{-2}. Purkey and Johnson (2010) estimate the heat uptake below 4000m to be 0.027 Wm^{-2}. Cheng *et al*. (2017) estimate that 12.8% of the change in ocean heat content is accumulating below 2000m. Cheng *et al*. (2017) also estimate that non-ocean heat uptake at 7% of the total. These estimates guide our comparison of our estimated total Earth system heat content and observations on the top 2000m of the ocean. We assume that heat uptake by the top 2000m of the ocean constitutes around 81% of the total increase in Earth system heat content and about 88% of total ocean heat uptake is in the top 2000m. Between 50% and 56% of total system heat content accumulates in the top 700m of the ocean.

2.2. Recent Estimates of the Climate Sensitivity

Estimates of the climate sensitivity have been the focus of ongoing debate. Estimates vary widely (Armour, 2016, Knutti *et al.* 2017) and there are notable differences between observation- and model-based sensitivity estimates (Richardson *et al*., 2016; Knutti *et al*., 2017). The consensus in the *IPCC 5th Assessment Report* (Bindoff *et al*., 2013) is that the equilibrium climate sensitivity (ECS) falls in the range of 1.5-4.5ºC with more than 66% probability,[6] while the transient climate response (TCR) falls in the range 1-2.5ºC with more than 66% probability. Armour (2016) notes that the range of ECS supported by recent observations is 1-4ºC with a best estimate of around 2ºC and the TCR is estimated at 0.9-2.0ºC. This suggests that climate

[5] Heat uptake is expressed either in joules of energy stored or as a flow of Wm^{-2}. But the latter is usually expressed on terms of the equivalent heat flow over the entire surface of the Earth that would equate to the rate of heat uptake and not the heat flow per square meter of ocean at the depth quoted. We have normalized all measurements quoted in this Section to Wm^{-2} over the entire surface of the Earth.

[6] This is the official translation of the term "likely" as used by the IPCC.

model based estimates are too sensitive. Richardson *et al*. (2016) note that sea surface temperature measurements measure water rather than air temperature, and the water has warmed more slowly than the air. Additionally, the most poorly measured regions on Earth, such as the Arctic, have also warmed the most. Richardson *et al*. (2016) process the CMIP5 model output in the same way as the HADCRUT4 temperature series is constructed – using seawater temperatures and under-sampling some regions. They infer an observation-based best estimate for the TCR of 1.66ºC, with a 5–95% range of 1.0–3.3ºC, consistent with the climate models considered in the *IPCC 5th Assessment Report*.[7] Our primary estimates use the Berkeley Earth surface temperature series, which provides much better coverage of polar regions than HADCRUT4. We compare these results to a model using the HADCRUT4 data as a robustness check.

Marvel *et al*. (2016) argue that the efficacy of some forcings is less than that of greenhouse gases and anthropogenic aerosols. They use single-forcing experiments to estimate these efficacies and revise TCR and ECS estimates upward to 1.7K and to 2.6-3.0ºC, depending on the feedbacks included. Armour (2016) highlights the joint (multiplicative) importance of the Richardson *et al.* (2016) and the Marvel *et al.* (2016) studies, which together should raise observational ECS by 60%, reconciling the discrepancy between observation and model based estimates. We test this idea by estimating models with both total aggregated radiative forcing using a uniform efficacy and radiative forcing adjusted for the lower efficacy of some forcings. Knutti *et al*. (2017) carry out an extensive survey of the literature, concluding that based on estimates constrained by different lines of evidence an ECS value of 3ºC is most likely. On the other hand, Brown and Caldeira (2017) argue that models that better simulate the current energy budget predict greater future warming and that the mean observationally informed ECS is 3.7ºC with a 25-75% interval of 3-4.2ºC. But Cox *et a*l. (2018) find that based on models that estimate the observed climate variability better, the ECS is 2.8 ºC with a 66% confidence interval of 2.2-3.4 ºC.

2.3. Issues in the Climate Econometrics Literature

Most studies of global climate change using econometric methods have ignored the role of the ocean. These studies produce implausible estimates of the rate of adjustment of surface temperature to long-run equilibrium. For example, Kaufmann and Stern (2002) find that the

[7] Using CRU land surface measurements only for the period 1964-2010 and an econometric dynamic panel data method that allows for potential long-run cointegration, Storelvmo *et al.* (2016) estimate the TCR at 2.0 +/- 0.8ºC. But this TCR only applies to land surface temperatures.

rate of adjustment of temperature to changes in radiative forcing is around 50% per annum even though they estimate that the climate sensitivity is 2.03ºC. Similarly, Kaufmann *et al.* (2006) estimate a climate sensitivity of 1.8ºC, while the adjustment coefficient implies that more than 50% of the disequilibrium between forcing and temperature is eliminated each year.

Simple AR(1) I(1) autoregressive models of this type assume that temperature adjusts in an exponential fashion towards the long run equilibrium. The estimate of that adjustment rate tends to go towards that of the fastest adjusting process in the system. Schlesinger *et al.* (no date) illustrate these points with a very simple first order autoregressive model of global temperature and radiative forcing. They show that such a time-series model approximates a model with a simple mixed layer ocean. Parameter estimates can be used to infer the depth of such an ocean. The models that they estimate have inferred ocean depths of 38.7-185.7 meters. Clearly, an improved time-series model needs to simulate a deeper ocean component.

As described in the Introduction, Stern (2006) used a state-space model inspired by multicointegration. The estimated climate sensitivity for the preferred model is 4.4ºC, which is much higher than previous time-series estimates and temperature responds much more slowly to increased forcing. However, this model only used data on the top 300m of the ocean. Pretis (2015) estimates an I(1) VAR for surface temperature and the heat content of the top 700m of the ocean for observed data for 1955-2011. The climate sensitivity is 1.67ºC for the preferred model but 2.16ºC for a model that excludes the level of volcanic forcing from the cointegrating relationship. The adjustment rate between surface temperature and the disequilibrium between surface temperature and radiative forcing is 0.11. As we describe in the Model section, below, we estimate models that explicitly model ocean heat storage using the concept of multicointegration and exploit the implied I(2) properties of multicointegrating models which allows us to obtain parameter estimates for 1850 to 2014. We compare our results to I(1) models with and without observed OHC. This comparison shows the added value of the I(2) multicointegration approach.

A second major issue relates to the time-series nature of the observed climate variables. Francisco Estrada, Pierre Perron, and others have argued "that both temperature and radiative forcing variables can be best represented as trend stationary processes with structural changes occurring in the slope of their trend functions and that they share a common secular trend and common breaks, largely determined by the anthropogenic radiative forcing" (Estrada and Perron, 2016, 1). Previously, Gay-Garcia *et al.* (2009) argued that the climate system cannot be

characterized by stochastic trends because: “assuming a unit root implies that global and hemispheric temperatures are highly unstable processes and therefore single events such as isolated solar flares, the 1974 La Nina (as well as other internal variation) or the 1883 Krakatau eruption would have changed the long-run path of global temperatures and their effect would be present even today” (336). As Kaufmann *et al.* (2010) explain in their reply, we would only expect global temperature to contain a stochastic trend because radiative forcing does. In the absence of forcing, the system would display levels stationary behavior.

The stochastic trend model assumes that all shocks are permanent. But the deterministic trend with breaks model assumes that only some shocks are permanent – these change the level or slope of the trend of the series. In particular, a shift in the level of the series is simply a very large shock. With enough breakpoints, any random walk could be approximated by a piecewise linear trend stationary process (Rappoport and Reichlin, 1989). As Kejriwal and Lopez (2013, 899) state: “a unit root process can be viewed as a limiting case of a stationary process with multiple breaks, one that has a break (permanent shock) every period”.

Both these models are approximations to reality. Kaufmann *et al.* (2013) find that they cannot reject the null of a unit root against the alternative of a deterministic trend with up to two breakpoints for radiative forcing series and, furthermore, show that “Monte Carlo simulations of the cointegration/error-correction model can generate time series for temperature that appear to be trend-stationary-with-a-break when analyzed with the statistical methodology used by Gay *et al.* (2009).” (Dergiades *et al.*, 2016, 68).

Similarly, Perron and Wada (2009) argue that U.S. GDP is best characterized by a trend stationary process with a slope break in 1973. But others find that the series is best represented by a random walk (Cushman, 2016). An improved testing procedure developed by Kejriwal and Lopez (2013) cannot reject the null hypothesis of a unit root in the GDP of 19 developed countries, though there is extensive evidence of breaks in the series. Berenguer-Rico and Carrion-i-Silvestre (2011) develop a single equation approach to testing for multicointegration in the presence of a potential structural break. However, this is not really relevant, because while there may be a structural break in the variables, we do not expect there to be a structural break in the cointegrating parameters. Instead, we test our estimated models for residual instability to see whether we may still be missing breaks in the relationships.

3. Methods

3.1. The Long-Run Relations

Long wave radiation from the Earth to space is proportional to the temperature of the surface and atmosphere. If there is an increase (decrease) in the incoming radiation from the sun that is absorbed – due to a change in solar irradiance or a change in the Earth's albedo – the equilibrium temperature must increase (decrease) to balance incoming and outgoing radiation. The atmosphere absorbs radiation from the surface and reradiates some of the energy back towards the surface.[8] The level of this back radiation is a function of the proportion of greenhouse gases in the atmosphere, mainly water and carbon dioxide. An increase in the share of greenhouse gases in the atmosphere increases the back radiation and, therefore, increases the necessary radiation from the surface that is needed to balance the back radiation. This requires an increase in the temperature of the surface. As the Earth system adjusts to a new equilibrium following an increase in greenhouse gases it will store heat as its temperature increases. This is the phenomenon of global warming. Though greenhouse gases only directly warm the atmosphere and surface, the temperature of the surface must be in equilibrium with that of the ocean. The heat capacity of the ocean is much larger than that of the atmosphere and so the storage of heat in the ocean and its warming dominates the global warming process, though some heat is stored in the atmosphere or used to melt ice. For simplicity, we model the overall increase in heat stored in the Earth system as a single variable and assume that surface temperature has a long-run equilibrium with that quantity. The global energy balance is given by

$$f_t - \lambda s_t = q_t \tag{1}$$

where *f* is radiative forcing in watts per square meter, *s* is surface temperature in Celsius, q is the uptake of heat by the Earth system in watt years per square meter,[9] which is assumed to be a stationary process, and *t* indicates the year of observation. The climate feedback parameter, λ, determines the equilibrium climate sensitivity to a doubling of carbon dioxide concentrations given by $ECS = 5.35\ln(2)/\lambda$. Equation (1) states that the disequilibrium between current radiative forcing and current temperature is stored in the Earth system. Assuming that *f* and *s*

[8] This back radiation is actually about double the radiation absorbed by the surface directly from the sun.

[9] We can convert this to joules by multiplying by the number of seconds in a year and the surface of the Earth in square meters. This constant is approximately $1.609*10^{22}$. Empirical implementation of (1) and (2) will each require a constant, as discussed below.

are I(1) time series, as has been shown by previous research (e.g. Kaufmann *et al*., 2013; Pretis, 2015), there should be a cointegrating relationship between them so that q_t is a stationary stochastic process. In the long run, surface temperature must also be in equilibrium with system heat content, $Q_t = \sum_{i=1}^{t} q_i$, so that

$$Q_t - \phi s_t = u_t \tag{2}$$

where u_t is another stationary stochastic process. Therefore, there is multicointegration between *f* and *s* as surface temperature, *s*, cointegrates with Q, the accumulated errors from the first cointegrating relation.

3.2. I(2) Models

Although multicointegration was initially developed for I(1) Vector Error Correction Models (VECM) with two instead of the usual single error correction mechanism (Granger and Lee (1989,1990), it was later shown that multicointegration implies an I(2) representation of the system (Engsted and Johansen, 1999). In the case of our model, the accumulated cointegration error Q_t can be obtained by cumulating the cointegrating relationship in (1): $F_t - \lambda S_t = Q_t$ where $F_t = \sum_{i=1}^{t} f_i$ and $S_t = \sum_{i=1}^{t} s_i$ are the partial sums of the original variables, where 1 indicates the first year of the sample, which is 1850 for the time series introduced in the next section. The accumulated I(1) variables F_t and S_t are by definition I(2).

Johansen (1992) shows that the p-dimensional I(2) VECM of order k is given by

$$\Delta^2 \boldsymbol{X}_t = \boldsymbol{\Pi X}_{t-1} - \boldsymbol{\Gamma} \Delta \boldsymbol{X}_{t-1} + \sum_{i=1}^{k-2} \boldsymbol{\Psi}_i \Delta^2 \boldsymbol{X}_{t-i} + \boldsymbol{\Phi D}_t + \boldsymbol{\epsilon}_t \tag{3}$$

with reduced rank conditions

$$\boldsymbol{\Pi} = \boldsymbol{\alpha \beta}' \tag{4}$$

and

$$\boldsymbol{\alpha}_\perp' \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp = \boldsymbol{\xi \eta}' \tag{5}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $p \times r$, $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are $(p-r) \times s_1$, $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ are $p \times (p-r)$ with $r < p$

and for which $\boldsymbol{\alpha}_{\perp}'\boldsymbol{\alpha} = 0$ and $\boldsymbol{\beta}_{\perp}'\boldsymbol{\beta} = 0$, $\boldsymbol{D}_t$ collects the deterministic components[10], $\boldsymbol{\epsilon}_t$ are iid $N(\mathbf{0}, \boldsymbol{\Omega})$, initial values are assumed to be fixed, and $t = 1, \dots, T$. We use the notation $\overline{\boldsymbol{\alpha}} = \boldsymbol{\alpha}(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}$ and define $\boldsymbol{\beta}_1 = \overline{\boldsymbol{\beta}}_{\perp}\boldsymbol{\eta}$ and $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{\perp}\boldsymbol{\eta}_{\perp}$ which are $p \times s_1$ and $p \times (p - r - s_1)$. The number of relations that cointegrate to I(0) is denoted by r while s_1 denotes the common I(1) trends and $s_2 = p - r - s_1$ the common I(2) trends. Hence the I(2) model is characterized by

$$r : \boldsymbol{\beta}'\boldsymbol{X}_t - \boldsymbol{\delta}\boldsymbol{\beta}_2'\Delta\boldsymbol{X}_t \sim I(0) \tag{6}$$

$$s_1 : \boldsymbol{\beta}_1'\boldsymbol{X}_t \sim I(1) \tag{7}$$

$$s_2 : \boldsymbol{\beta}_2'\boldsymbol{X}_t \sim I(2) \tag{8}$$

where the r polynomially cointegrating relations usually need $\Delta\boldsymbol{X}_t$ to cointegrate to I(0) with multicointegration parameter $\boldsymbol{\delta} = \overline{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}\overline{\boldsymbol{\beta}}_2$.

In our bivariate case with $\boldsymbol{X}_t = [F_t, S_t]'$, the I(2) model simplifies. If multicointegration is present, then $r = 1, s_1 = 0$, and $s_2 = 1$. Thus, the two cointegrating relationships outlined in (1) and (2) are represented by the multicointegrating relationship in (6) while the system is driven by a common I(2) trend.

However, the multicointegrating relationship (6) includes both f_t and s_t in $\Delta\boldsymbol{X}_t$ while from a physical perspective only s_t is needed in the multicointegrating relationship. The corresponding theoretical equation, (2), models the surface temperature and the heat stored in the Earth system reaching thermal equilibrium, which does not depend on the level of radiative forcing in the atmosphere but only on the relative temperature of the atmosphere and ocean. From a statistical perspective, only one of s_t or f_t is needed to achieve multicointegration, as these two variables cointegrate (Engsted and Haldrup, 1999; Siliverstovs, 2006). Therefore, we re-parameterize the multicointegrating relation (Rahbek *et al*., 1999; Kongsted and Nielsen, 2004) to exclude forcing, in order to achieve physical interpretability of this relationship. We estimate the model by using the I(2) to I(1) transformation that also allows us to use standard inference on the multicointegration parameter (Kongsted and Nielsen, 2004; Kongsted, 2005; Siliverstovs, 2006). The I(2) to I(1) transformation is given by

[10] For ease of presentation, we introduce the I(2) model without discussing the deterministic components in detail. Restrictions on the deterministic components follow the widely used formulation of Rahbek *et al*. (1999) and are discussed in detail for the final model that we estimate towards the end of this section.

$$\widetilde{\boldsymbol{X}}_t = \begin{pmatrix} \boldsymbol{B}'\boldsymbol{X}_t \\ \boldsymbol{v}'\Delta\boldsymbol{X}_t \end{pmatrix} \tag{9}$$

where $\boldsymbol{B} = \boldsymbol{\beta}$ as estimated in (3) by the procedure outlined in Johansen (1997), and $\boldsymbol{v} = [0,1]'$, so that only s_t is selected. Only I(1) variables are contained in $\widetilde{\boldsymbol{X}}_t$ and the transformed I(1) VECM can be estimated as

$$\Delta\widetilde{\boldsymbol{X}}_t = \tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\beta}}'\widetilde{\boldsymbol{X}}_{t-1} + \sum_{i=1}^{k-1} \tilde{\Gamma}_i \Delta\widetilde{\boldsymbol{X}}_{t-i} + \tilde{\Phi}\boldsymbol{D}_t + \tilde{\boldsymbol{\epsilon}}_t \tag{10}$$

where $\tilde{\boldsymbol{\beta}}' = \left(1, -\tilde{\delta}\right)$,[11] so that the reformulated multicointegrating relationship that guarantees physical interpretability is given by

$$\boldsymbol{\beta}'\boldsymbol{X}_t - \tilde{\delta}\boldsymbol{v}'\Delta\boldsymbol{X}_t \sim I(0). \tag{11}$$

Engsted and Haldrup (1999) derive both stock and flow VECM representations for multicointegrating variables. We use the bivariate flow VECM representation and substitute in the re-parameterized multicointegrating relation in (11) (Siliverstovs, 2006). This flow VECM representation of the I(2) model is the basis for the results in the next section and, thus, we now discuss restrictions on the deterministic components, which were not discussed up till now for ease of presentation. We follow the specification proposed by Rahbek *et al*. (1999), who restrict the linear trend and the constant to the cointegration spaces.[12] A time trend is needed in the CI(2,1) cointegrating relationship between the two accumulated I(2) variables, S_t, F_t, in order to account for initial conditions (Engsted and Haldrup, 1999). Similarly, a constant is needed in the multicointegrating relationship between $\boldsymbol{\beta}'\boldsymbol{X}_t$ and $\boldsymbol{v}'\Delta\boldsymbol{X}_t$. For the I(1) VECM presented in (10) that results from the I(2) to I(1) transformation we use a restricted constant to be consistent with the I(2) model. Hence, the bivariate flow VECM representation for multicointegrating variables becomes

$$\Delta^2\boldsymbol{X}_t = \boldsymbol{\alpha}\left[\breve{\boldsymbol{\beta}}'\breve{\boldsymbol{X}}_{t-1} - \breve{\boldsymbol{\delta}}\Delta\breve{\boldsymbol{X}}_{t-1}\right] - \zeta\Delta\breve{\boldsymbol{\beta}}'\breve{\boldsymbol{X}}_{t-1} + \sum_{i=1}^{k-2} \boldsymbol{\Psi}_i\Delta^2\boldsymbol{X}_{t-i} + \boldsymbol{u}_t \tag{12}$$

where $\breve{\boldsymbol{X}}_t = [F_t, S_t, t]'$ and so $\Delta\breve{\boldsymbol{X}}_t = [f_t, s_t, 1]'$. Then $\breve{\boldsymbol{\beta}} = [\boldsymbol{\beta}', \beta_0]' = [1, -\lambda, \beta_0]'$ and $\breve{\boldsymbol{\delta}} =$

[11] See Kongsted and Nielsen (2004) for the remaining transformed coefficients.
[12] The means of Δf and Δs in our sample are not significantly different from zero and so, unlike Siliverstovs (2006) we do not need an unrestricted constant in our specification.

$[\tilde{\delta}\boldsymbol{v}', \gamma]' = [0, \tilde{\delta}, \gamma]'$.[13] The coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ correspond to the theoretical equations (1) and (2) where β_0 and γ account for initial conditions. Furthermore, $\breve{\boldsymbol{\beta}}'\breve{\boldsymbol{X}}_t = \hat{Q}_t$ is the predicted system heat content that is generated as a latent variable in the multicointegrating framework and $\Delta\breve{\boldsymbol{\beta}}'\breve{\boldsymbol{X}}_t$ is the radiative imbalance in the atmosphere between radiative forcing and temperature given in (1), which is also equal to the system heat uptake. The adjustment parameters $\boldsymbol{\alpha}$ estimate the speed with which $\Delta^2\boldsymbol{X}_t = [\Delta f_t, \Delta s_t]'$ adjusts to deviations from the multicointegrating relation. Hence, $\boldsymbol{\alpha}$ estimates the adjustment speed to both cointegrating relations in (1) and (2). The parameter $\boldsymbol{\zeta}$ estimates the adjustment to the heat uptake of the previous period.

3.3. I(1) Models

For comparison with the I(2) models, we also estimate several I(1) models. Given revisions and updates to the data, this provides a more comparable benchmark than simply referring to previous studies. Most generally, we estimate the 3-dimensional I(1) VECM for observed OHC, surface temperature, and radiative forcing that is given by

$$\Delta\boldsymbol{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\breve{\boldsymbol{y}}_{t-1} + \sum_{i=1}^{k-1}\boldsymbol{\Psi}_i\Delta\boldsymbol{y}_{t-i} + \boldsymbol{\varepsilon}_t \tag{13}$$

where $\boldsymbol{y}_t = [f_t, s_t, h_t]$ and h is observed OHC in watt-years per square meter and $\breve{\boldsymbol{y}}_t = [f_t, s_t, h_t, 1]$. We again use a restricted constant to be consistent with the models above. For this model, the sample is restricted by the availability of observations for OHC to the period 1940 to 2014. Despite using actual observations of OHC, we expect this model to perform worse than our I(2) VECM because the I(2) estimator is super-super consistent and there is a much shorter time series available for this I(1) model. We also estimate the 2-dimensional I(1) VECM given by

$$\Delta\boldsymbol{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\breve{\boldsymbol{x}}_{t-1} + \sum_{i=1}^{k-1}\boldsymbol{\Psi}_i\Delta\boldsymbol{x}_{t-i} + \boldsymbol{\varepsilon}_t \tag{14}$$

where $\boldsymbol{x}_t = [f_t, s_t]$ and $\breve{\boldsymbol{x}}_t = [f_t, s_t, 1]$. We estimate this model for the full sample period using both the optimal number of lags and a simple AR(1) model too. These models are equivalent

[13] See Rahbek *et al.* (1999) for the restrictions imposed on β_0 and γ in estimating the model.

to a zero-dimensional single-layer energy balance model (Kaufmann *et al*., 2013), whereas the I(2) VECM is equivalent to a two-layer zero-dimensional energy balance model.

All estimation and testing was carried out using RATS 9.1 (Estima, 2014) and CATS 2.0 (Dennis *et al*., 2006).

4. Data

Figures 1 to 3 present the key time series used in this study. The sources of the data are documented in Appendix I. Figure 1 presents the two versions of the radiative forcing aggregate that we use. Forcing trends upwards over time with large negative departures following major volcanic eruptions, which cluster in the late 19th Century – starting with the 1883 eruption of Krakatoa – and late 20th Century – the largest of which was Pinatubo in 1992.[14] The period from the late 1990s represents a step change in the level of forcing compared to the period up till then. The level of forcing in 2014 was between 2.65 (partial efficacy) and 2.98 (full efficacy) Wm^{-2} and is about three quarters of what would be the effect of doubling CO_2 alone relative to the preindustrial era, which would increase forcing by 3.7 Wm^{-2}. By multiplying the radiative forcing of ozone, volcanic aerosols, and solar irradiance by 0.5, the series with modified efficacies approximately account for Marvel *et al*.'s (2016) findings that these forcings have less effect than greenhouse gases and anthropogenic aerosols.[15] This series generally shows a smaller increase and variance.

[14] Though volcanic eruptions only affect radiative forcing for two to three years, they have much more persistent effects on ocean heat content (Marvel *et al*., 2016) and thus also on surface temperature. Omitting volcanic forcing entirely or excluding it from the cointegration relations would result in system heat content following a very different path than it did historically.

[15] We do not use the precise efficacies that they estimate because these differ depending on the timeframe.

Figure 1. Radiative Forcing

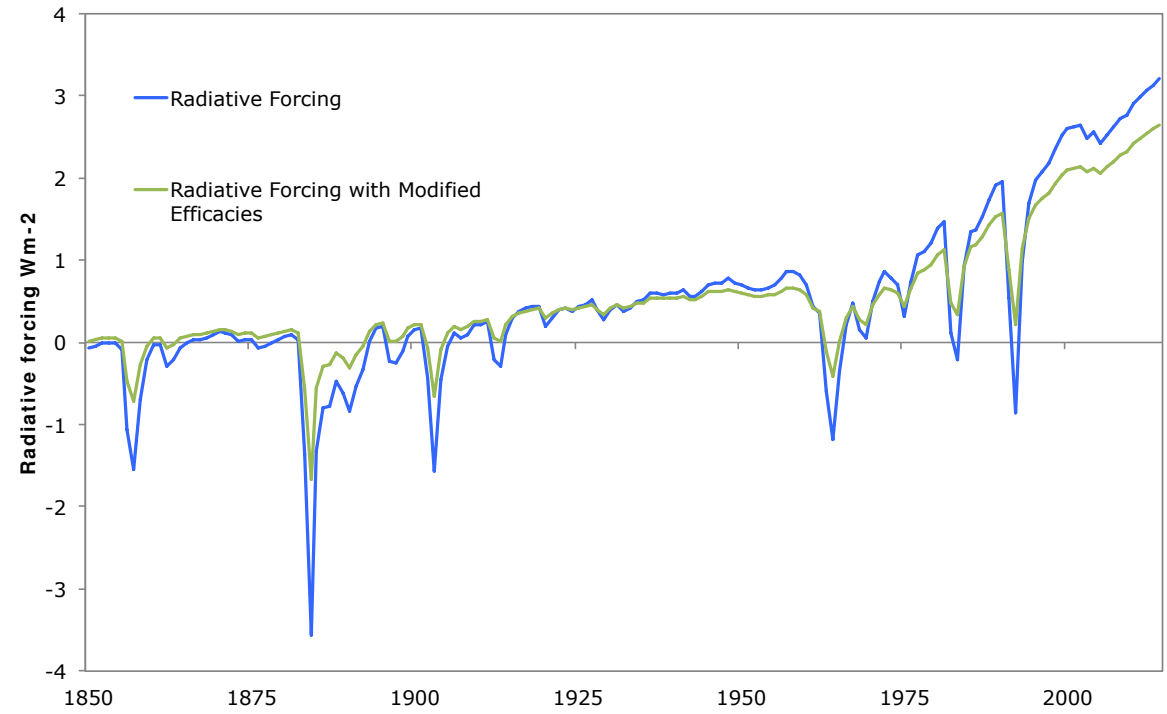


Figure 2. Surface Temperature

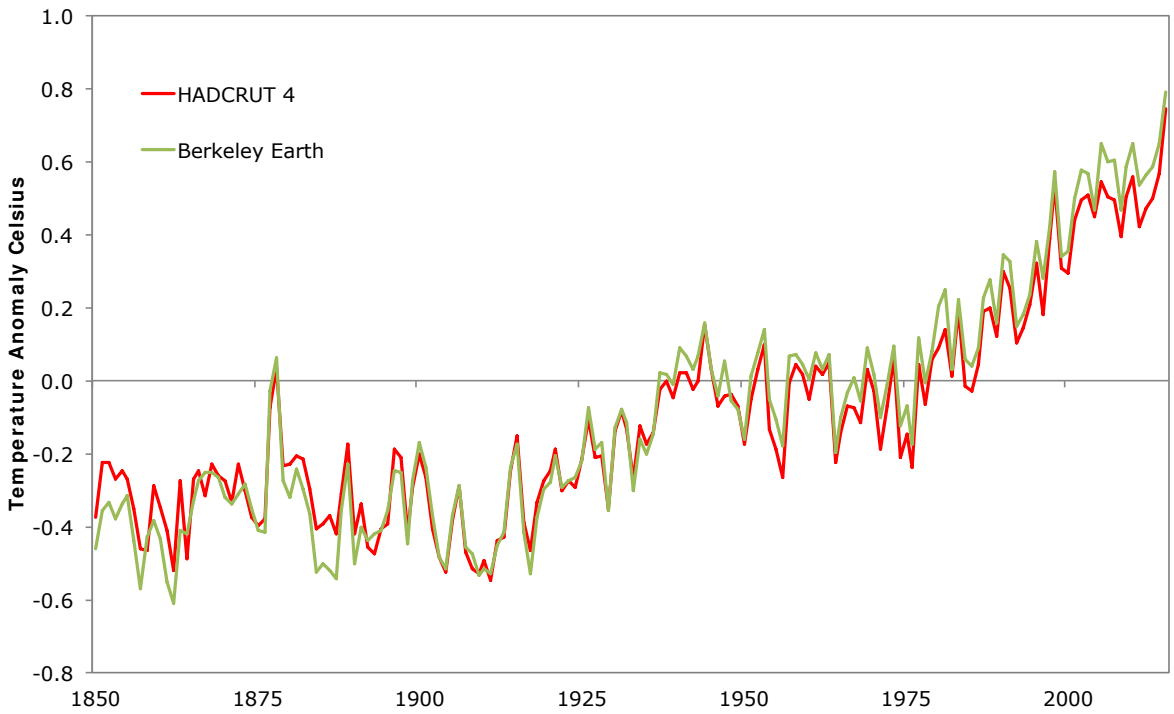
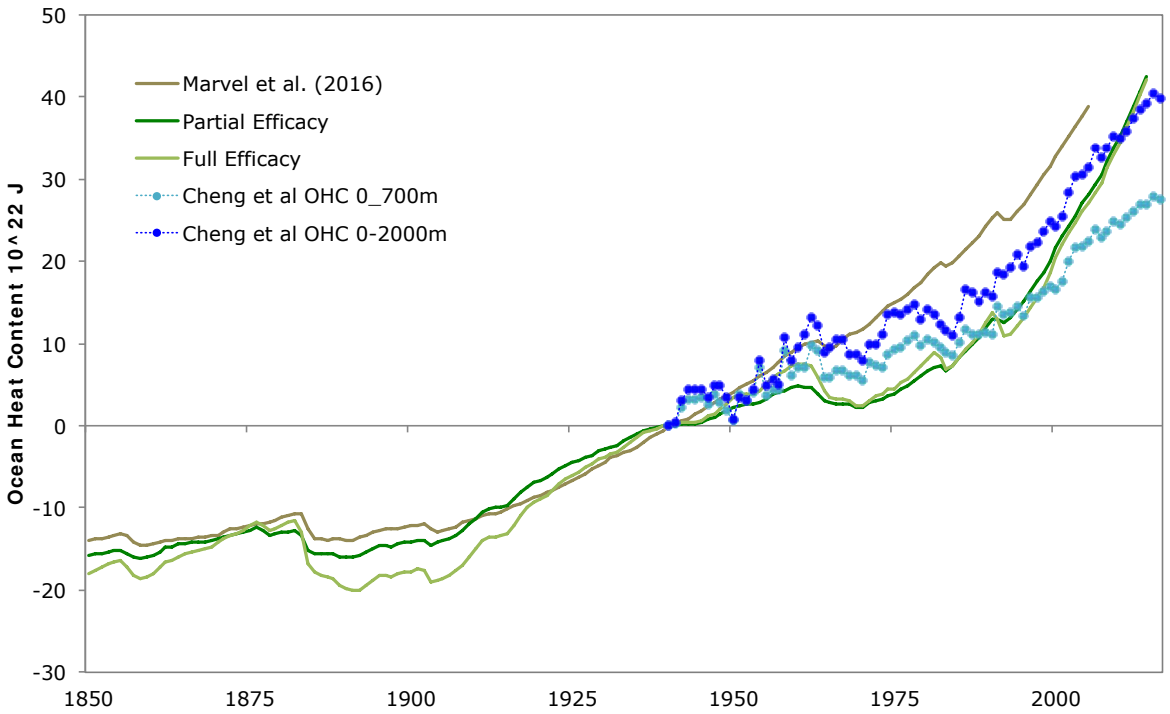


Figure 2 shows the two surface temperature series. Though most research has been dedicated to correcting bias in sea surface temperatures (e.g. Hausfather *et al*., 2017), the Berkeley Earth dataset focuses on correcting bias in land surface temperatures (Rohde *et al*., 2013) and employs a modified version of the Hadley sea surface temperature series that extrapolates to areas without observations. The two series are very similar to each other in both trend and short-run fluctuations. HADCRUT 4 shows less warming due to its low coverage of Arctic regions (Karl *et al.*, 2015). Temperature does seem to track forcing quite well, starting a strong upward trend in the late 1970s just as forcing does.

Finally, Figure 3 compares simulated and observed OHC series, where all series have been normalized to zero in 1940. The Marvel *et al.* (2016) series is for their model simulation with all forcings. We multiply their original data by 0.88 to scale it to the heat content of the top 2000m of the ocean. This shows a stronger increase in heat content than Cheng *et al*.'s (2017) observations and surprisingly strong heat uptake before 1960, which actually matches Stern's (2006) results quite closely. Our own estimates will be discussed in the next section.

Figure 3. Predicted and Observed Ocean Heat Content



5. Results and Discussion

5.1. I(2) Models

Our main estimates, Models I and II, use the Berkeley Earth temperature series and we then test the effect of using the HADCRUT series (Model III). Model I uses unadjusted radiative forcing while the other models use the series adjusted for partial efficacy.

We considered a maximum of four lags and select the most parsimonious model that has adequate serial correlation properties. It is important not to select too long a lag length because this can induce artificially large roots to be found in the model (Granger and Jeon, 2006; Onatski and Uhlig, 2012), which could result in not being able to reject the null of non-cointegration (Onatski and Uhlig, 2012). Table 1 reports information criteria and residual autocorrelation tests for all the I(2) models. The Bayesian Information Criterion (Schwarz, 1978) would pick two lags in each case. The Hannan-Quinn Criterion (Hannan and Quinn, 1979) would pick three lags for Model I and two lags for the other models. The Akaike Information Criterion (Akaike, 1973) picks three lags for Model I and very marginally picks 4 lags over 3 for Models II and III. Setting a 0.05 significance threshold for the LM(1) and LM(k) tests, the minimum adequate lag length for Models I to III is three lags. The Ljung-Box portmanteau test statistic also points to three lags with the exception of Model II where the p-value is 0.04. All things considered, we picked three lags for Models I to III.

Table 1. Lag Length Selection of I(2) Models

		Information Criteria			Autocorrelation Tests		
	k	SC	H-Q	AIC	LM(1)	LM(k)	Ljung-Box(40)
Model I	2	-6.092	-6.228	-6.344	0.012	0.001	0.004
Berkeley, full efficacy	3	-6.051	-6.233	-6.368	0.110	0.387	0.120
	4	-5.976	-6.203	-6.359	0.565	0.070	0.298
Model II	2	-7.402	-7.538	-7.655	0.029	0.001	0.002
Berkeley, partial efficacy	3	-7.350	-7.531	-7.668	0.057	0.523	0.040
	4	-7.286	-7.513	-7.669	0.279	0.041	0.234
Model III	2	-7.436	-7.573	-7.689	0.029	0.001	0.024
HADCRUT, partial efficacy	3	-7.382	-7.564	-7.700	0.051	0.600	0.151
	4	-7.319	-7.547	-7.703	0.322	0.064	0.360

Notes: p-values for autocorrelation test statistics, k is number of lags. SC: Schwarz Criterion, H-Q: Hannan-Quinn Criterion, AIC: Akaike Information Criterion. LM(k): LM-Test for autocorrelation of order k.

Table 2a reports rank test statistics and Table 2b the roots of the companion matrix for full rank and relevant reduced rank specifications. Using the rank test statistics and the *p*-values from the asymptotic distribution, we can reject non-cointegration for all three models, but we can also reject the restriction $r = 1, s_2 = 1, s_1 = 0$ at the 0.05 level and for Models II and III at the 0.01 level also. On the other hand, we cannot reject the hypothesis $r = 1, s_2 = 0, s_1 = 1$. However, when we impose $r = 1, s_2 = 1, s_1 = 0$ there appear to be two unit roots (Table 2b), though for Models II and III the third root is quite large. When we impose $r = 1, s_2 = 0, s_1 = 1$ there is one explosive and one unit root, which is not compatible with the hypothesis that there is a single I(1) trend in the data. In Section 6, we simulate critical values for the null of $r = 1, s_2 = 1, s_1 = 0$ for a sample size of 165 with and without measurement error. Even in the absence of measurement error, none of the test statistics for the test of the null of $r = 1, s_2 = 1, s_1 = 0$ are significant at the 0.01 level for a sample of this size. With a larger level of measurement error, they are not significant even at the 0.05 level. Therefore, we accept the $r = 1, s_2 = 1, s_1 = 0$ restrictions.

Table 2a. Rank Tests for I(2) Models

	$p-r$	r	$s_2 = 2$	$s_2 = 1$	$s_2 = 0$
Model I	2	0	94.208	47.176	44.910
Berkeley, full efficacy			(0.000)	(0.001)	(0.000)
	1	1		22.968	2.345
				(0.017)	(0.929)
Model II	2	0	81.600	45.269	43.679
Berkeley, partial efficacy			(0.000)	(0.002)	(0.000)
	1	1		27.643	2.723
				(0.003)	(0.895)
Model III	2	0	81.175	43.983	42.476
HADCRUT, partial efficacy			(0.000)	(0.003)	(0.000)
	1	1		25.340	1.893
				(0.007)	(0.961)

Notes: p-values based on the asymptotic distribution in parentheses. See Table 10 for critical values with a sample size of 165 and potential measurement error.

Table 2b. Moduli of Estimated Characteristic Roots

	r	s_2	s_1	1	2	3	4	5	6
Model I	2	0	0	1.023	0.970	0.543	0.543	0.309	0.309
Berkeley, full efficacy	1	1	0	1.000	1.000	0.621	0.326	0.302	0.302
	1	0	1	1.023	1.000	0.549	0.549	0.313	0.313
Model II	2	0	0	1.032	0.987	0.529	0.529	0.244	0.244
Berkeley, partial efficacy	1	1	0	1.000	1.000	0.886	0.395	0.395	0.001
	1	0	1	1.027	1.000	0.515	0.515	0.249	0.249
Model III	2	0	0	1.029	0.975	0.507	0.507	0.459	0.013
HADCRUT, partial efficacy	1	1	0	1.000	1.000	0.856	0.410	0.410	0.078
	1	0	1	1.030	1.000	0.519	0.519	0.407	0.035

Table 3 reports parameter estimates for the I(2) models. Residual autocorrelation is adequate for all models and the residuals of the temperature equation are normally distributed. This is not the case for the radiative forcing equation, but Gonzalo (1994) showed that the FIML Johansen procedure is rather robust to minor departures from the model assumptions due to non-normality.

Table 3. Main I(2) Results

		Model I Berkeley, full efficacy	**Model II** Berkeley, partial efficacy	**Model III** HADCRUT, partial efficacy
Multicoint. Vector (mci)				
$\breve{\boldsymbol{\beta}}$	F_{t-1}	1.000	1.000	1.000
	$S_{t-1}, -\lambda$	-1.709 (-5.041)	-1.326 (-3.940)	-1.567 (-4.639)
	t, β_0	-0.401 (-3.941)	-0.399 (-3.379)	-0.492 (-4.550)
$\breve{\boldsymbol{\delta}}$	$\hat{Q}_{t-1}$	1.000	1.000	1.000
	$s_{t-1}, -\phi$	-41.482 (-17.653)	-33.355 (-15.646)	-30.695 (-15.347)
	$1, \gamma$	-17.577 (-21.780)	-13.821 (-17.747)	-9.269 (-14.738)
VECM				
Δf_t equation				
$\boldsymbol{\alpha}$	mci_{t-1}	0.007 (0.839)	0.020 (3.648)	0.026 (4.054)
$\boldsymbol{\zeta}$	$\Delta\hat{Q}_{t-1}$	-0.216 (-3.77)	-0.119 (-2.867)	-0.182 (-3.711)
Δs_t equation				
$\boldsymbol{\alpha}$	mci_{t-1}	0.013 (7.083)	0.016 (6.573)	0.017 (6.046)
$\boldsymbol{\zeta}$	$\Delta\hat{Q}_{t-1}$	0.046 (3.55)	-0.009 (-0.461)	-0.024 (-1.105)
Diagnostic Tests				
Multivariate q-test (Hosking)		0.2310	0.1142	0.2457
Jarque-Bera:				
Δf_t		0.0000	0.0000	0.0000
Δs_t		0.6245	0.6246	0.7133
Obs.		162	162	162

Notes: t-statistics for coefficients in parentheses. p-values for diagnostic tests. Coefficients for lagged second differences not reported.

For Model I, the ECS is 2.17ºC and is highly significantly different to zero (t = 5.04). Radiative forcing is weakly exogenous (Engle *et al*., 1983) – it does not respond to deviations from the multicointegrating relationship which reflect disequilibria in both the long-run relation between OHC and surface temperature and the long-run relation between surface temperature and forcing. The adjustment rate of temperature is slow, as only 1.3% of the deviation from equilibrium is removed each year, but this parameter is highly significant (t = 7.08). Both dependent variables do respond to the first difference of OHC. The coefficient of s_t in $\boldsymbol{\delta}$ in Table 3 corresponds to $-\phi$ in (2) so that in equilibrium the increase in system heat content in watt years per square meter is 41 times the increase in surface temperature in Celsius. Using data on the mass of the atmosphere and its heat capacity, increasing the temperature of the atmosphere by 1ºC requires an increase in heat content of $5*10^{21}$ joules, which translates to 0.31 watt years per square meter. Therefore, there is a 41 watt years per square meter increase in system heat content for a 0.31 watt years per square meter increase in atmospheric heat content, implying that less than 1% of heat content is directed to warming the atmosphere. The literature reviewed above, found that 3% of the heat content resided in the atmosphere, but this 3% is not a long-run equilibrium proportion as the oceans are far from equilibrium. For Model II the climate sensitivity is 2.8ºC, which is close to the IPCC consensus. The 66% confidence interval – equivalent to the IPCC's "likely" range – is from 2.25 to 3.69 ºC. This is almost identical to Cox et al.'s (2018) estimate and confidence interval. The 95% confidence interval is from 1.87 to 5.57 ºC. The rate of adjustment of temperature is 1.6% per year, which is close to that in Model I. Radiative forcing is not weakly exogenous in this specification. The estimate of ϕ again implies that roughly 1% of the heat content increase occurs in the atmosphere in the long run.

We re-estimate the temperature equation using recursive least squares, taking the estimated cointegrating vectors as given, and compute the CUSUM and CUSUMSQ tests (Brown *et al*., 1975) using critical values from Edgerton and Wells (1994) and Zeileis (2004). In neither case, is there any sign of structural instability in the surface temperature equation. Of course, these tests do not test for instability in the cointegrating relationship but only in $\boldsymbol{\alpha}$, $\boldsymbol{\zeta}$, and the short-run effects given the cointegrating vectors. Unsurprisingly, the tests suggest breaks in the radiative forcing equation associated with the Krakatoa eruption and other large volcanic eruptions, but these breaks only have transitory effects on radiative forcing.

Figure 3 compares the predicted OHC for the top 2000m of the ocean for Models I and II. We convert the predicted Earth system heat content to joules and scale it to the heat content of the top 2000m of the ocean by multiplying by 0.81 as explained in Section 2.1. We multiply the Marvel *et al.* (2016) series by 0.88 to scale it to the top 2000m of the ocean. Both our predicted series are quite similar to Marvel *et al.*'s (2016) simulated series in terms of short run fluctuations and total heat uptake. As we would expect, the full efficacy model shows much larger declines in OHC in the wake of volcanic eruptions than the partial efficacy model does. Our reconstructed series track observed heat content in the top 2000m of the ocean fairly well but show less heat uptake in the 1940s and 1950s and more heat uptake recently. Total heat uptake since 1940 is similar.

In order to assess whether measurement errors affect our estimates of system heat content, we regress observed ocean heat content (Cheng *et al.*, 2017) on the predicted system heat content from Model II (Table 4).[16] The estimated coefficient for the top 2000m of the ocean is 0.71, which is close to, but significantly less than, the estimate of 0.81 mentioned in Section 2.1. The Phillips and Ouliaris (1990) cointegration test statistic is -2.95, which is statistically significant at the 0.05 level (MacKinnon, 2010). We also test for cointegration with heat content in the top 700m of the ocean. The estimated coefficient of 0.49 is close to expectations (Section 2.1) and we can strongly reject the null of non-cointegration.

Table 4. Cointegration between Predicted System Heat Content and Observed Ocean Heat Content

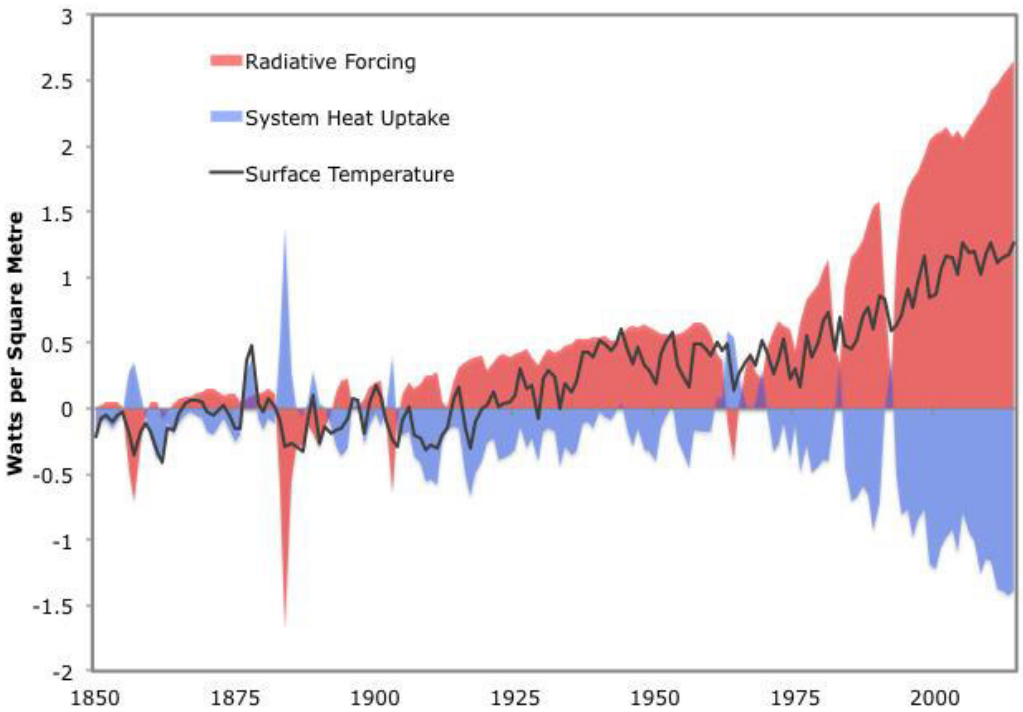
	Top 700m	Top 2000m
Constant	4.256 (0.701)	5.668 (0.421)
Predicted System Heat Content	0.494 (0.031)	0.712 (0.021)
R-squared	0.934	0.938
Phillips-Ouliaris Cointegration Test	-3.63	-2.95
T	75	75

Notes: Means of series adjusted as in Figure 3. Robust standard errors in parentheses. Critical values for Phillips-Ouliaris cointegration test (McKinnon, 2010): 10%: -2.59, 5%: -2.90, 1%: -3.52.

[16] In Section 6 we examine the measurement error issue in detail.

Figure 4 illustrates the surface radiative disequilibrium and ocean (system) heat uptake for Model II. The gap between radiative forcing and transformed temperature is the disequilibrium in watts per square meter. In 2014, it was 1.38Wm^{-2}, which, given the climate sensitivity, implies that temperature needs to rise by 1.0ºC to reach long-run equilibrium. Due to multicointegration, this disequilibrium is exactly equal to the annual predicted take up of heat by the Earth system – mostly the ocean - in watt years per square meter. As we can see, following volcanic eruptions, heat is released from the ocean, mitigating the cooling effects. In the first sixty or so years of the sample, except following large eruptions, surface temperature tracks heat uptake by the ocean quite closely. But then temperature increasingly follows the rapidly ramping radiative forcing, but mostly lags forcing resulting in ocean heat uptake. According to our model, the hiatus in temperature after 1998 can largely be accounted for by an increase in ocean heat uptake, though there is a plateau in radiative forcing from 1999 to 2005 as noted by Kaufmann *et al*. (2011). By contrast, during the slow-down in warming from the 1940s to 1970s radiative forcing grew very slowly but heat uptake was also very low.

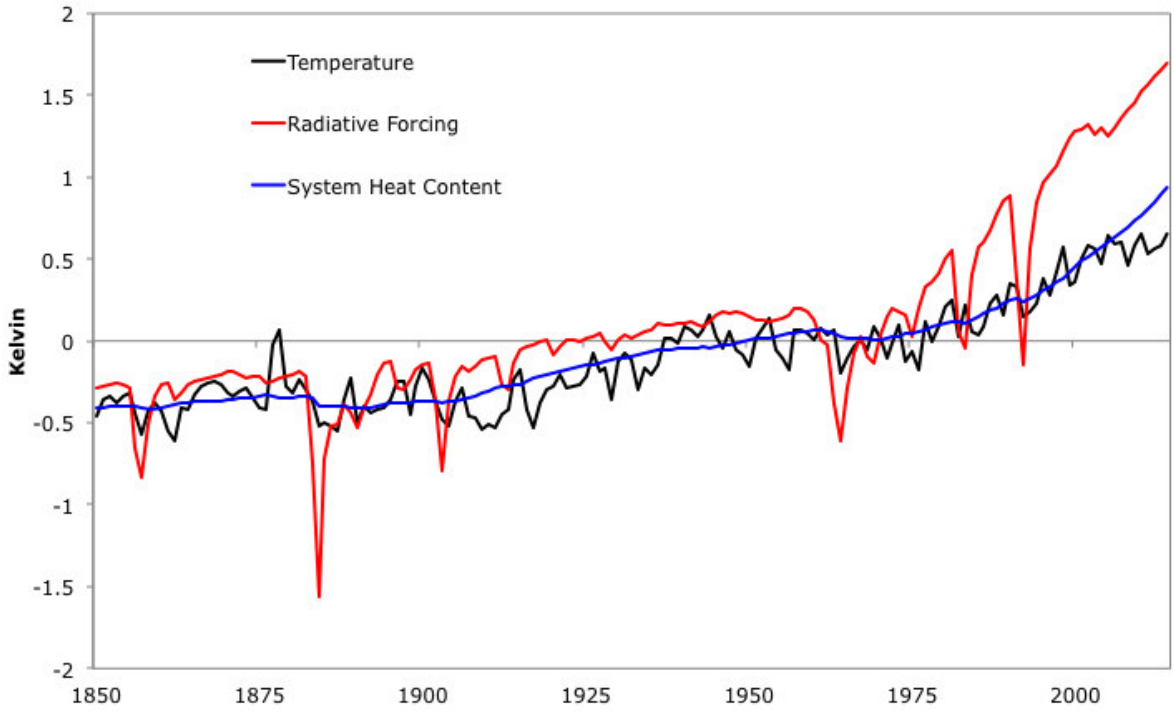
Figure 4. Surface Radiative Disequilibrium and System Heat Uptake



Notes: Data is from Model II, which uses Berkeley Earth temperature and partial efficacy of radiative forcing. Temperature has been transformed to units of radiative forcing according to Equation (1). The sign of system heat uptake has been reversed so that negative numbers are storage of heat in the system and positive are release of heat (mostly from the ocean).

An alternative way of visualizing this relationship is shown in Figure 5 where instead of heat uptake we plot heat content and scale the variables to the temperature variable. Here the gap between observed temperature and radiative forcing is the "committed warming" – the increase in temperature that would occur if radiative forcing was held constant from that year onwards.[17] The Figure shows that surface temperature and system heat content are quite tightly coupled, while radiative forcing has in recent decades diverged upwards. We also see that the "hiatus" in recent years resulted in quite a wide deviation between surface temperature and ocean heat content.

Figure 5. Temperature, Radiative Forcing, and System Heat



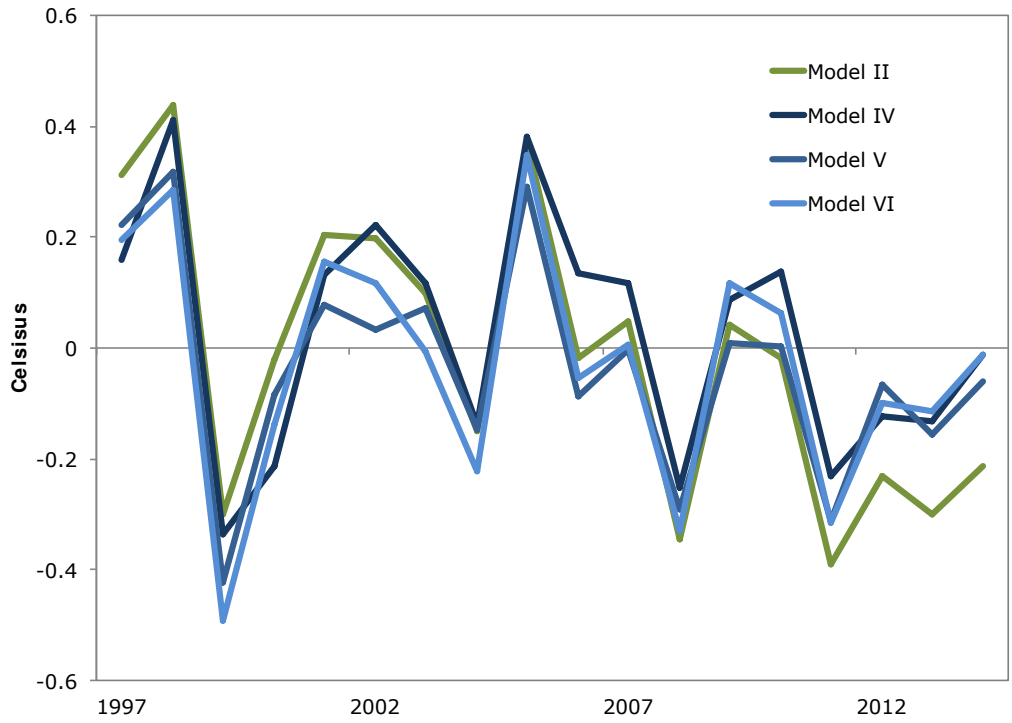
Note: Radiative forcing has been transformed to temperature units using Equation (1), and system heat content using Equation (2).

The role of "internal variability" in driving this disequilibrium during the hiatus is shown in Figure 6 for Model II and the three I(1) models discussed in the next subsection. We orthogonalize the reduced form residuals using the Cholesky decomposition with radiative

[17] If greenhouse gas emissions were cut to zero, temperature is, in fact, expected to stay constant rather than rise to the long-run equilibrium. This is because the effect of removal of carbon dioxide from the atmosphere is expected to roughly balance the adjustment of ocean heat content to equilibrium (Solomon *et al*., 2009; Mauritsen and Pincus, 2017). This is why the TCR is actually more relevant for climate policy than the ECS.

forcing ordered first and surface temperature second. Therefore, the reduced-form residual of the surface temperature equation is decomposed into a component associated with shocks to radiative forcing and an orthogonal component that does not instantaneously affect radiative forcing. The graph shows that this component is positive during El Nino episodes such as in 1997-98, 2005, and 2009, and negative during La Nina episodes such as in 1999-2000, 2008, and 2011. There is also a downward trend in the mean shock over the course of the hiatus, which appears to be associated with the shift to negative PDO values. This is most pronounced for the I(2) model, Model II, and least pronounced for Model IV. This shows that the I(2) multicointegration model best models known components of internal variability.

Figure 6. Internal Variability Shocks During the Hiatus



To test the sensitivity of the estimates to data sources, we also estimated the partial efficacy model using HADCRUT surface temperature data (Model III). Results for Model III are similar to those for the Berkeley series (Model II). As expected, due to the lesser coverage of the Arctic area in the HADCRUT data, the climate sensitivity is lower (2.37ºC).

5.2. I(1) Models

All the models in this subsection use the Berkeley Earth temperature series and the partial efficacy radiative forcing series. Model IV also adds observed heat content in the top 2000m of the ocean and so its sample period is restricted to 1940-2014. Models V and VI only include the first two variables – Model V uses the optimal lag length, while Model VI is an AR(1) model. We consider a maximum of three lags for Model IV and four lags for Model V. Table 5 reports diagnostics for lag length selection. Unusually, all three information criteria select a single lag for the model with observed OHC (Model IV). For Model V, the best-fit model has four lags. Table 6 reports cointegration rank tests. All the models cointegrate. Model IV has two cointegrating vectors and Models V and VI, one each.

Table 5. Lag Length Selection of I(1) Models

	k	SC	H-Q	AIC	LM(1)	LM(k)	Ljung-Box(n)
Model IV Berkeley, with observed OHC, partial efficacy	1	-7.305	-7.534	-7.574	0.045	0.045	0.277
	2	-7.034	-7.434	-7.489	0.289	0.113	0.084
	3	-6.699	-7.27	-7.327	0.86	0.754	0.144
Model V and VI Berkeley, partial efficacy	1	-7.350	-7.418	-7.499	0.121	0.121	0.000
	2	-7.271	-7.385	-7.486	0.000	0.000	0.000
	3	-7.312	-7.471	-7.592	0.029	0.009	0.078
	4	-7.262	-7.467	-7.605	0.444	0.122	0.269

Notes: n = 18 for Model IV and n= 40 for Models V and VI. For other information see Table 1.

Table 6. Rank Tests for I(1) Models

	$p-r$	r	Trace	5% Critical Value	p-value
Model IV	3	0	64.078	35.070	0.000
	2	1	23.136	20.164	0.018
	1	2	5.350	9.142	0.256
Model V	2	0	18.578	20.164	0.083
	1	1	2.967	9.142	0.595
Model VI	2	0	41.999	20.164	0.000
	1	1	1.636	9.142	0.839

Notes: Trace test statistics and critical values use a correction for small samples (Doornik, 1998)

Table 7. Main I(1) Results

	Model IV Berkeley with observed OHC, full efficacy	**Model V** Berkeley, partial efficacy, 4 lags	**Model VI** Berkeley, partial efficacy, 1 lag
Coint. Relation 1			
f_{t-1}	1	1	1
$s_{t-1}, -\lambda$	-3.054 (-13.087)	-2.259 (-10.829)	-2.396 (-14.290)
h_{t-1}	0		
1	-0.446 (-6.628)	-0.723 (-10.948)	-0.757 (-13.764)
Coint. Relation 2			
f_{t-1}	0		
$s_{t-1}, -\phi$	-31.232 (-10.638)		
h_{t-1}	1		
1	16.973 (20.040)		
VECM			
Δf_t equation			
α_{11}	-0.049 (-0.604)	-0.087 (-1.307)	-0.075 (-1.495)
α_{12}	0.009 (1.243)		
Δs_t equation			
α_{21}	0.175 (5.411)	0.108 (3.461)	0.136 (5.837)
α_{22}	0.008 (2.608)		
Δh_t equation			
α_{31}	0.775 (2.293)		
α_{32}	-0.120 (-3.995)		
Diagnostic Tests			
Multivariate q-test (Hosking)	0.2693	0.6559	0.0023
Jarque-Bera:			
Δf_t	0.0000	0.0000	0.0000
Δs_t	0.3507	0.3625	0.8526
Δh_t	0.8919		
Obs.	74	161	164

Notes: t-statistics for coefficients in parentheses. p-values for diagnostic tests. Coefficients for lagged first differences not reported.

For Model IV, the ECS is only 1.21ºC (Table 7). The estimate of ϕ is 31.2 (Table 7), which implies that the top 2000m of the ocean in equilibrium stores 100 times more of the increased heat than does the atmosphere. Using a similar model,[18] Pretis (2015) estimates the ECS at 1.67ºC and ϕ as 24.4. The rates of adjustment to equilibrium are all very fast compared to those in the I(2) multicointegrating model. The errors of the radiative forcing equation are non-normal, but the diagnostics are otherwise good. Model V has a somewhat higher ECS and slower adjustment to equilibrium. The residuals have good autocorrelation properties. As expected, the latter is not the case for Model VI as it uses only one lag and the adjustment of surface temperature to equilibrium is a little faster for this final model. Interestingly, α is not as large as past research found (e.g. Kaufmann and Stern, 2002; Kaufmann *et al.*, 2006). We think this is most likely due to improved measures of radiative forcing. We ran Model V on a sample restricted to 1865-1990, which is the period Kaufmann and Stern (2002) use and obtained similar estimates of α as we do here.

These results show the advantages of the I(2) multicointegration model. Adding observed OHC to our I(1) model in fact results in a lower ECS than that estimated by the models that do not include OHC data, This appears to be because the sample is much shorter for Model IV than for Models V and VI. When we ran the latter models using a 1940-2014 sample, we obtained very similar ECS and estimates of α as we did for Model IV. If we had a longer time series for observed OHC, we might obtain better estimates.

5.4. Simulations

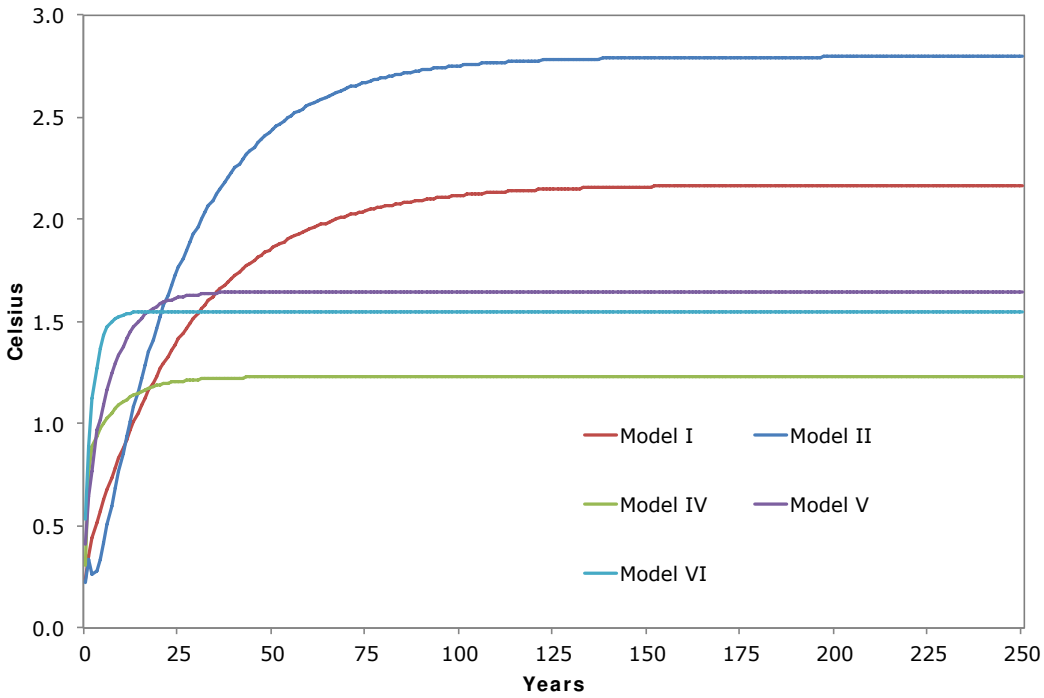
We carry out two simulations for each of the six models. One doubles radiative forcing in one year and then lets the model reach equilibrium holding radiative forcing constant. The point of this experiment is to see the adjustment path. The other experiment is a transient experiment where CO_2 is increased by 1% a year until it doubles. We then calculate the TCR at the doubling point. As we only investigate the effect of shocks to radiative forcing on temperature, we transform the estimated reduced form VECM into a structural VECM using the Cholesky decomposition with radiative forcing ordered first. We also decouple the radiative forcing equation as our experiments require us to control radiative forcing rather than allow for feedback.

[18] Pretis (2015) uses NOAA data on OHC in the top 700m of the ocean, in units of 10^{22} joules rather than watt years per square meter, the GISS surface temperature series, and a sample ending in 2011.

For the instant doubling experiment, we apply a shock of 3.708Wm^{-2} to radiative forcing in year 1 of the experiment and hold radiative forcing constant after that.[19] For the TCR experiment we increase radiative forcing by 0.053Wm^{-2} in year 1 and the following 69 years.

Figure 7 shows the response to instantaneous doubling of CO_2 for the various models. The I(2) models show a slow response. For Model II, temperature only increases 0.37ºC in the first year and takes 12 years to exceed a 1ºC increase, reaches 2ºC after 32 years and after 100 years 98% of the adjustment has occurred. By 200 years, adjustment is completed. So, adjustment to equilibrium does happen faster than in physical simulation models such as Li *et al*. (2013) where the surface equilibrates only after 1200 years, which is not surprising given that we estimated the model with only 165 years of data. On the other hand, our results show the strength of the multicointegration approach in modeling a slow adjustment process with a relatively short sample. Model I adjusts somewhat slower. The I(1) models – Models IV to VI – all reach equilibrium very fast as we would expect from the econometric results - the AR(1) model (VI) is fastest.

Figure 7. Response to Doubling Carbon Dioxide



[19] In both cases, the model is in long-run equilibrium prior to the perturbation.

Table 8 presents the ECS and TCR for each of our models. For the partial efficacy model with Berkeley temperature data the TCR is 1.85ºC. This is a bit higher than Richardson *et al.*'s (2016) estimate and is 66% of the ECS. The TCR has a smaller range than the ECS – there is a tendency for the TCR to be a larger proportion of the ECS for models with lower ECS's. For the I(1) models (IV to VI) TCR is always more than 90% of ECS, which is a consequence of their fast adjustment to equilibrium.

Table 8. Climate Sensitivities

Model	Description		TCR	ECS	Ratio
Model I	I(2)	Berkeley, full efficacy	1.49ºC	2.17 ºC	0.68
Model II	I(2)	Berkeley, partial efficacy	1.85 ºC	2.80 ºC	0.66
Model III	I(2)	HADCRUT, partial efficacy	1.72 ºC	2.37 ºC	0.73
Model IV	I(1)	Berkeley, with observed OHC, partial efficacy	1.17 ºC	1.21 ºC	0.97
Model V	I(1)	Berkeley, partial efficacy, 4 lags	1.53 ºC	1.65 ºC	0.93
Model VI	I(1)	Berkeley, partial efficacy, 1 lag	1.51 ºC	1.55 ºC	0.98

6. Measurement Error

If the original I(1) variables are observed subject to I(0) measurement error, then accumulating these variables into I(2) variables results in I(1) measurement error. We assess how I(1) measurement errors may affect tests for non-cointegration and multicointegration, the cointegrating parameter estimates, and predicted system heat content. In this section, we explicitly model measurement error in forcing and surface temperature by $f_t^* = f_t + v_t$ and $s_t^* = s_t + w_t$, where f_t and s_t denote the true unobserved values and the I(0) measurement errors are given by v_t and w_t. Both $\sum_{i=1}^{t} v_i$ and $\sum_{i=1}^{t} w_i$ are then I(1) measurement errors in observing the I(2) variables, which with measurement error are indicated by F_t^* and S_t^*. We use Monte Carlo simulations to assess the relevance of these I(1) measurement errors for our analysis.

In the case of cointegration between I(1) variables with I(0) measurement error, it is well known that estimates of cointegration parameters remain consistent (Phillips and Durlauf, 1986). Duffy and Hendry (2017) show that if the I(1) variables trend sufficiently, the consequences of even I(1) measurement error may be negligible. Hassler (2007) investigated the effect of I(1) measurement errors on the power of multicointegration tests in a VAR model using a Monte Carlo analysis. He concludes that the power of finding multicointegration is strongly affected by measurement error, but his simulation design relies on Granger and Lee's (1989) original approach, which does not take the I(2) representation of multicointegration into account.

Because we exploit the I(2) nature of the system and use maximum likelihood estimation, we expect the effect of I(1) measurement error on the inference on multicointegration and estimated cointegrating parameters to be less severe in our analysis. However, predicted system heat content may be contaminated by I(1) measurement error. Predicted system heat content with measurement error is given by

$$\hat{Q}_t^* = F_t^* - \hat{\lambda} S_t^* + \hat{\beta}_0 t$$

$$= Q_t + \left(\lambda - \hat{\lambda}\right) S_t^* + \left(\hat{\beta}_0 - \beta_0\right) t + \sum_{i=1}^{t} v_i - \lambda \sum_{i=1}^{t} w_i \tag{15}$$

where $\hat{\lambda}$ and $\hat{\beta}_0$ are estimates of λ and β_0, respectively. If $\hat{\lambda}$ is still super-super consistent for λ despite the accumulated measurement error, then $\left(\lambda - \hat{\lambda}\right)$ is $O_p(T^{-2})$ and hence the effect of $\left(\lambda - \hat{\lambda}\right) S_t^*$ could be relatively negligible, despite S_t^* being an I(2) variable. The effect of $\left(\hat{\beta}_0 - \beta_0\right) t$ could be similarly negligible. However, the I(1) measurement errors $\sum_{i=1}^{t} v_i$ and $\lambda \sum_{i=1}^{t} w_i$ could have stronger effects on estimated system heat content, $\hat{Q}_t$, which is also an I(1) variable. Duffy and Hendry (2017) suggest that such I(1) measurement errors might be cointegrated, which is likely to occur for cognate time series such as the main series of the National Income Accounts. However, in our case it is unlikely that these measurement errors cointegrate. Error in measuring temperature is unlikely to be related to error in measuring greenhouse gas concentrations, in formulating radiative forcing functions, or in estimating emissions of anthropogenic aerosols. Error in measuring temperature is small, especially in recent decades (Rohde *et al*., 2013), but error in estimating radiative forcing could be quite large (Myhre *et al*., 2013). We thus analyze the consequences of these I(1) measurement errors using a data-generating process that is based on our estimated models and empirically motivated strengths of measurement errors.

The data generating process for temperature is based on the estimates of our preferred I(2) VECM (Model II) while radiative forcing is modeled as an exogenous random walk:

$$\Delta f_t^{(i)} = e_{1t}^{(i)} \tag{16}$$

$$\Delta s_t^{(i)} = 0.015\left[\breve{\boldsymbol{\beta}}' \breve{\boldsymbol{X}}_{t-1}^{(i)} - \breve{\boldsymbol{\delta}} \Delta \breve{\boldsymbol{X}}_{t-1}^{(i)}\right] - 0.002 \Delta \breve{\boldsymbol{\beta}}' \breve{\boldsymbol{X}}_{t-1}^{(i)} + 0.060 \Delta f_t^{(i)} + 0.041 \Delta f_{t-1}^{(i)} + 0.045 \Delta s_{t-1}^{(\mathrm{i})} + e_{2t}^{(i)} \tag{17}$$

$$\begin{bmatrix} e_{1t}^{(i)} \\ e_{2t}^{(i)} \end{bmatrix} \sim N\left(0, \begin{bmatrix} 0.228^2 & 0 \\ 0 & 0.099^2 \end{bmatrix}\right) \tag{18}$$

where $\breve{\boldsymbol{X}}_t^{(i)} = \left[F_t^{(i)}, S_t^{(i)}, t\right]'$, $\Delta\breve{\boldsymbol{X}}_t^{(i)} = \left[f_t^{(i)}, s_t^{(i)}, 1\right]'$, $\breve{\boldsymbol{\beta}} = [1, -\lambda, \beta_0]'$, $\breve{\boldsymbol{\delta}} = \left[0, \tilde{\delta}, \gamma\right]'$ and i denotes the Monte Carlo iterations. The values of both $\breve{\boldsymbol{\beta}}$ and $\breve{\boldsymbol{\delta}}$ are taken from Model II in Table 3 and the accumulated variables are obtained by $\boldsymbol{X}_t^{(i)} = \boldsymbol{X}_{t-1}^{(i)} + \boldsymbol{x}_t^{(i)}$ with $\boldsymbol{X}_t^{(i)} = \left[F_t^{(i)}, S_t^{(i)}\right]'$ and $\boldsymbol{x}_t^{(i)} = \boldsymbol{x}_{t-1}^{(i)} + \Delta\boldsymbol{x}_t^{(i)}$ with $\boldsymbol{x}_t^{(i)} = \left[f_t^{(i)}, s_t^{(i)}\right]'$. The remaining parameters and error variances were fitted by estimating (17) by OLS using the data for Model II and taking the estimates of $\breve{\boldsymbol{\beta}}$ and $\breve{\boldsymbol{\delta}}$ as given. The variance of e_{1t} is set to the variance of the first difference of observed radiative forcing. This ensures that the signal to noise ratio in the models with measurement error are empirically relevant.

For each iteration, i, we also consider two cases of measurement errors where $c = 1$ denotes Case 1 and $c = 2$ denotes Case 2, respectively. The cumulative series with measurement error are computed using

$$\boldsymbol{X}_{tc}^{*(i)} = \sum_{i=1}^{t}\left(\boldsymbol{x}_t^{(i)} + \boldsymbol{m}_{tc}^{(i)}\right) \tag{19}$$

where $\boldsymbol{m}_{tc}^{(i)} = \left[v_{tc}^{(i)}, w_{tc}^{(i)}\right]'$ and $v_{tc}^{(i)} \sim N\left(0, \sigma_{fc}^2\right)$ and $w_{tc}^{(i)} \sim N(0, \sigma_{sc}^2)$. As described in Appendix II, we use the available sources to estimate the likely standard deviations of the measurement errors. For temperature, the mean standard deviation for the full period is 0.038°C. For recent years the standard deviation is 0.022°C. For radiative forcing, the mean standard deviation for the full period is 0.20 Wm^{-2}. For 2014 the standard deviation is 0.5 Wm^{-2}. We use two cases of measurement error in the Monte Carlo simulation to reflect these empirical standard deviations. For Case 1 we set $\sigma_f = 0.2$ and $\sigma_s = 0.04$ and for Case 2 we double the measurement errors used in Case 1.

For the generated data without measurement error and for the two cases with measurement error, we estimate I(2) VECM models using the Johansen maximum likelihood procedure with the correct lag lengths ($p = 3$). We compute the rank test statistics to assess the performance of identifying multicointegration and we then impose the rank restrictions $r = 1, s_1 = 0, s_2 = 1$ to estimate the cointegrating parameters and the corresponding predicted system heat content. We use a sample size of $T = 165$ as is the case in our empirical analysis and we run the Monte Carlo simulation with 2000 iterations. Finally, we also run the simulation again for $T = 1000$

without considering the two cases of measurement errors to study the asymptotic properties. Results are presented in Tables 9 to 11.

Table 9. Rejection Rates for Rank Tests

	$r = 0, s_2 = 2,$ $s1 = 0$	$r = 0, s_2 = 1,$ $s_1 = 1$	$r = 0, s_2 = 0,$ $s_1 = 2$	$r = 1, s_2 = 1,$ $s_1 = 0$
DGP (nominal)	1	1	1	0.05
No ME ($T = 165$)	0.9985	0.9985	1	0.1075
ME Case 1 ($T = 165$)	1	0.997	1	0.134
ME Case 2 ($T = 165$)	1	0.996	0.9995	0.292
No ME ($T = 1000$)	1	1	1	0.0685

Notes: Rejection rates based on critical values reported in Rahbek et al. (1999) using the 0.05 level of statistical significance. ME denotes measurement error.

Table 10. Quantiles of the Trace Test Statistic for the Null of $r = 1, s_2 = 1, s_1 = 0$

	0.01	0.05	0.1
No ME ($T = 165$)	28.26	23.16	20.34
ME Case 1 ($T = 165$)	29.05	23.83	21.23
ME Case 2 ($T = 165$)	36.82	29.26	26.38
No ME ($T = 1000$)	25.03	21.03	18.41
Rahbek et al. (1999)	24.40	20.02	17.91

Notes: ME denotes measurement error.

Table 11. Cointegrating Parameter Estimates

	$-\hat{\lambda}$		$\hat{\beta}_0$	
	Mean	SD	Mean	SD
DGP	-1.326		-0.399	
No ME ($T = 165$)	-1.343	0.132	-0.409	0.114
ME Case 1 ($T = 165$)	-1.341	0.165	-0.407	0.136
ME Case 2 ($T = 165$)	-1.341	0.356	-0.409	0.201
No ME ($T = 1000$)	-1.326	0.002	-0.399	0.005

Notes: ME denotes measurement error. Mean and SD denote the mean and standard deviation of the estimated parameters obtained in the Monte Carlo simulation.

First, we analyze the rejection rates for the rank test to assess to what extent testing for multicointegration is affected by measurement errors. The first three columns of Table 9 present the frequency of rejecting the incorrect null hypothesis of non-cointegration. Despite measurement error, the rejection rate is always close to 1. Therefore, empirically relevant I(1) measurement errors do not affect the power of tests of the null of non-cointegration for a DGP that mimics our empirical model. For the fourth column of Table 9, the null hypothesis is true

and so the rejection rate is the size of the test. For $T = 165$, the rejection rate is 0.11 in the absence of measurement error and only slightly increases with measurement error (Case 1). For larger measurement errors (Case 2) the rejection rate increases to 0.29. For $T = 1000$, the rejection rate approaches the nominal level.

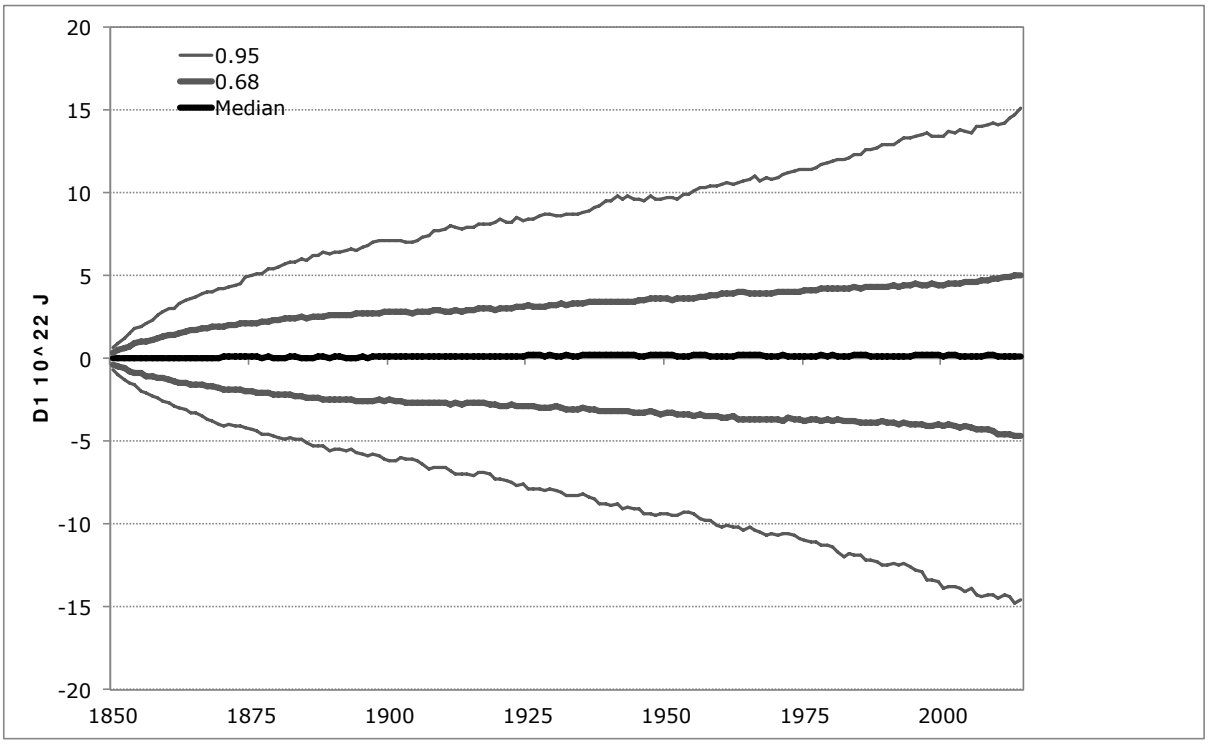
Moreover, Table 10 gives the quantiles of the trace test statistics as found in our simulation and compares them to the asymptotic results given by Rahbek *et al*. (1999). Our simulation with $T = 1000$ without measurement error shows that the distribution of trace test statistics does approach the asymptotic distribution, but in our smaller samples, especially in the presence of measurement error, the critical values can become substantially larger. These simulated values were used in combination with the roots of the companion matrix for inference on multicointegration in Section 5.

Second, we assess how measurement errors affect the cointegrating parameter estimates. The means of the estimated cointegrating parameters (Table 11) are slightly biased upwards in absolute value for $T = 165$, with and without measurement error, but are estimated consistently as indicated by the results for $T = 1000$. The standard deviations of the parameters obtained in the simulation increase with the standard deviation of the measurement error.

Finally, we assess how predicted system heat content is affected by measurement error. For this we calculate $D_{tc}^{(i)} = \hat{Q}_t^{(i)} - \hat{Q}_{tc}^{*(i)}$ for all i and t. Figure 8 plots quantiles of D_{t1} for each year t. As expected from the accuracy of the simulated cointegrating parameters, the median difference is close to zero. In 2014, the 0.68 interval extends to roughly +/-5*10^{22} joules, while the 0.95 interval extends to about +/-15*10^{22}. [20] For comparison, we estimate that system heat content increased by 72*10^{22} joules over the period. Obviously, measurement error can affect estimated heat content but would not greatly change conclusions about the direction or rate of increase. In Section 5.1 we also tested for cointegration between estimated system heat content, $\hat{Q}_t$, and observed ocean heat content. We found there (Table 4) that we can reject the null of non-cointegration suggesting that contamination of our estimate by spurious I(2) and I(1) trends is not very important. Therefore, in practice measurement error does not seem to strongly affect our findings.

[20] For D_{t2}, the 0.68 (0.95) interval extends a little further by +/- 9 (-25 to 29) in 2014.

Figure 8. Effect of Measurement Error (Case 1) on Estimated System Heat Content



7. Conclusions and Implications

We have shown how the concept of multicointegration can be used to impose energy balance on a simple time-series model of global climate change. Despite only using 165 years of data in our estimation, equilibrium in response to a doubling of carbon dioxide is not reached for two centuries. Despite not using any data on OHC our predictions of OHC match well those from an energy balance model and the available observations. Our estimates of the ECS and TCR are close to the consensus in the scientific literature. Our 66% confidence interval for the ECS of 2.25 ºC to 3.69ºC is narrower than the IPCC's estimate of 1.5 ºC to 4.5 ºC. These results contrast strongly with those produced using I(1) cointegration models that ignore heat storage in the ocean. An I(1) model that includes observed OHC did not perform any better than the other I(1) models, but this may be due to the very short sample available. Our results empirically verify the consensus in the climate science community on the likely impact of anthropogenic forcing on the climate. We think that we strengthen that consensus by using different methods than have been applied before to come to the same conclusion.

According to our model, the hiatus in surface warming that occurred between 1998 and 2014 can largely, but not entirely, be attributed to increased heat uptake. This contrasts with the period of flat global temperatures in the 1950s and 1960s, where radiative forcing grew slowly due to increasing anthropogenic sulfur emissions and several volcanic eruptions but heat uptake was low. The residuals in the surface temperature equation for Model II were negative in the period 2010-2014, averaging just under minus one standard deviation. Therefore, reversion to the mean would result in a jump in temperatures in 2015-2017 as actually occurred. The record high temperature of 0.80ºC in 2015 can be explained by a half a standard deviation positive residual in the temperature equation. On the other hand, 2016's 0.95 ºC would have been an almost two standard deviation positive residual. Therefore, slightly lower temperatures are likely in the next few years.

Appendix I: Data Sources and Construction

Raw data

All data series were collected as annual data for 1850 to 2014 unless otherwise stated.

Temperature: We used two different temperature data series:

Berkeley Earth (Rohde *et al*., 2013):

http://berkeleyearth.org/data/

HadCrut 4.0, using combined land (CRUTEM4) and marine (SST anomalies from HadSST3) temperature anomalies (Morice *et al.*, 2012):

http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/download.html

Ocean heat: We use data provided by Lijing Cheng for the top 700m and top 2000m of the ocean based on the data presented in Cheng *et al*. (2017). We also use Marvel *et al.*'s (2016) simulated OHC data for 1850 to 2005 (All forcings ensemble mean).

GHG concentrations: We use data for atmospheric greenhouse gas concentrations collated by Meinshausen *et al.* (2016):

http://www.climate-energy-college.net/cmip6

The computation of radiative forcing is described further below in this Appendix.

Aerosols: Data on emissions of sulfur dioxide and black and organic carbon 2014 were produced by the Community Emissions Data System (CEDS) Project (http://www.globalchange.umd.edu/ceds/) and were provided to us prior to publication by Stephen Smith. Ozone and stratospheric aerosols data are from Miller *et al.* (2014) (https://data.giss.nasa.gov/modelforce/Fi_Miller_et_al14_upd.txt). We assume that values for 2013-2014 are the same as for 2012.

Solar Irradiance:

We use the NOAA Climate Data Record (CDR) of Total Solar Irradiance (TSI), NRLTSI Version 2 (Coddington *et al.*, 2015) ftp://data.ncdc.noaa.gov/cdr/solar-irradiance/tsi/

Computing Radiative Forcing

We obtain radiative forcing as follows:

Well-mixed greenhouse gases: We apply the formulae from Table 6.2 in Ramaswamy *et al.* (2001) with the base year of 1850.

Solar irradiance: We multiply the irradiance by 0.175 as in Figure 2.5 in Shine *et al.* (1990) after subtracting the irradiance in 1850.

Tropospheric sulfate aerosol: Direct forcing is given by:

$$R_{Dt} = -0.4 h_t S_t / S_{2011}$$

where *S* is annual anthropogenic sulfur emissions in Tg S and *h* is the stack height term (Wigley and Raper, 1992). The estimated radiative forcing in 2011 is taken from Table 8.4 in Myhre *et al.* (2013). For indirect forcing, we apply a modification of the formula in Wigley and Raper (1992):

$$R_{It} = -0.45 ln\left(1 + \frac{h_t S_t}{19}\right) / ln\left(1 + \frac{h_t S_{2011}}{19}\right)$$

The natural burden is assumed to be 19Tg S (Boucher and Pham, 2002). The assumed forcing in 2011 is taken from Table 8.6 in Myhre *et al.* (2013).

Black carbon: We assume the forcing is linear: $R_t = 0.44 B_t / B_{2011}$, where *B* is emissions of black carbon and set the assumed forcing in 2011 based on Tables 8.4 and 8.6 in Myhre *et al.* (2013). This forcing includes both the absorption of heat by airborne black carbon (0.4Wm^{-2} in 2011) and the forcing due to reducing the albedo of snow (0.04Wm^{-2} in 2011).

Organic carbon: We assume the forcing is linear: $R_t = -0.09 O_t / O_{2011}$, where *O* is emissions of organic carbon and set the assumed forcing in 2011 based on Table 8.4 in Myhre *et al.* (2013).

Ozone: We use data from Miller *et al.* (2014).

Stratospheric sulfate aerosol: We use data from Miller *et al.* (2014).

Appendix II: Measurement Error

The Berkeley temperature data set provides estimates of the uncertainty (Rohde *et al*., 2013) that can be translated into the standard deviation of the measurement error. For the full period the mean standard deviation is 0.038°C. For recent years the standard deviation is 0.022°C. For radiative forcing, we use our radiative forcing data and the information on uncertainties in Table 8.6 and Figure 8.17 of Chapter 8 of the IPCC *5th Assessment Report* (Myhre *et al*., 2013). These give a +/- 45% confidence interval in percentage terms. We convert this to a standard deviation by dividing the percentage given by 1.65. We then multiply this percentage by the mean of each forcing in our sample to obtain a standard deviation in radiative forcing units and compute the standard deviation for total forcing making the simplifying assumption of zero covariance between each uncertainty. For the sample mean radiative forcing of the 1850 to 2014 period, the standard deviation is 0.20 Wm^{-2}. For 2014 the standard deviation is 0.5 Wm^{-2}.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N., Csáki, F. (Eds.) *2nd International Symposium on Information Theory*, *Tsahkadsor, Armenia, USSR, September 2-8, 1971*. Budapest: Akadémiai Kiadó. 267–281.

Andersson, S. M., Martinsson, B. G., Vernier, J. P., Friberg, J., Brenninkmeijer, C. A. M., Hermann, M., van Velthoven, P. F. J., Zahn, A., 2015. Significant radiative impact of volcanic aerosol in the lowermost stratosphere. *Nature Communications* 6, 7692.

Armour, K. C., 2016. Projection and prediction: Climate sensitivity on the rise**.** *Nature Climate Change* 6, 896–897.

Armour, K. C., 2017. Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks, *Nature Climate Change* 7, 331–335.

Ballantyne, A., Smith, W., Anderegg, W., Kauppi, P., Sarmiento, J., Tans, P., Shevliakova, E., Pan, Y., Poulter, B., Anav, A., Friedlingstein, P., Houghton, R., Running, S., 2017. *Nature Climate Change* 7, 148–152.

Balmaseda, M. A., Trenberth, K. E., E. Källén, E., 2013. Distinctive climate signals in reanalysis of global ocean heat content. *Geophysical Research Letters* 40, 1754–1759.

Barnett, T., Zwiers, F., Hegerl, G., Allen, M., Crowley, T., Gillett, N., Hasselmann, K., Jones, P., Santer, B., Schnur, R., Stott, P., Taylor, K., Tett, S., 2005. Detecting and attributing external influences on the climate system: A review of recent advances. *Journal of Climate* 18, 1291—1314.

Beenstock, M., Reingewertz, Y., Paldor, N., 2012. Polynomial cointegration tests of anthropogenic impact on global warming. *Earth System Dynamics* 3, 173–188.

Berenguer-Rico, V., Carrion-i-Silvestre, J. L., 2011. Regime shifts in stock-flow I(2)-I(1) systems: The case of US fiscal sustainability. *Journal of Applied Econometrics* 26, 298—321.

Bindoff, N. L., Stott, P. A., K. AchutaRao, M., Allen, M. R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I. I., Overland, J., Perlwitz, J., Sebbari, R., Zhang, X., 2013: Detection and attribution of climate change: from global to regional. In: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), *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 University Press, Cambridge and New York.

Boswijk, H. P., 2010. Mixed normal inference on multicointegration. *Econometric Theory* 26, 1565—1576.

Boucher, O., Pham, M., 2002. History of sulfate aerosol radiative forcings. *Geophysical Research Letters* 29(9), 22-1–22-4.

Brown, P. T., Caldeira, K. 2017. Greater future global warming inferred from Earth's recent energy budget. *Nature* 552, 45–50.

Brown, R. L., Durbin, J., Evans, J. M., 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society-B* 37, 149—192.

Cheng L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., Zhu, J., 2017. Improved estimates of ocean heat content from 1960 to 2015. *Science Advances* 3(3), e1601545.

Coddington, O., Lean, J. L., Lindholm, D., Pilewskie, P., Snow, M., NOAA CDR Program (2015): NOAA climate data record (CDR) of total solar irradiance (TSI), NRLTSI Version 2. NOAA National Centers for Environmental Information. doi:10.7289/V55B00C1

Cox, P. M., Huntingford, C., Williamson, M. S., 2018. Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature* 553, 319-322.

Curry, J., 2014. Uncertain temperature trend. *Nature Geoscience* 7, 83—84.

Cushman, D. O., 2016. A unit root in postwar U.S. real GDP still cannot be rejected, and yes, it matters. *Econ Journal Watch* 13(1), 5–45.

Dennis, J. G., Hansen, H., Johansen, S., Juselius, K., 2006. *CATS in RATS: Cointegration Analysis of Time Series, Version 2*. Estima, Evanston IL.

Dergiades, T., Kaufmann, R. K., Panagiotidis, T., 2016. Long-run changes in radiative forcing and surface temperature: The effect of human activity over the last five centuries. *Journal of Environmental Economics and Management* 76, 67—85.

Doornik, J. A., 1998. Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys* 12, 573–593.

Edgerton, D., Wells, C., 1994. Critical values for the CUSUMQ statistic in medium and large sized samples. *Oxford Bulletin of Economics and Statistics* 56(3), 355—365.

Engle, R. F., Hendry, D. F., Richard, J. F., 1983. Exogeneity. *Econometrica* 51, 277—304.

Engsted, T. Haldrup, N., 1999. Multicointegration in stock-flow models. *Oxford Bulletin of Economics and Statistics* 61, 237—254.

Engsted, T., Johansen, S., 1999. Granger's representation theorem and multicointegration. In: Engle, R. F., White, H. (Eds.), *Cointegration, Casuality, and Forecasting Festschrift in Honour of Clive Granger*, Oxford University Press, Oxford.

Estima, 2014. *RATS, Version 9.1*. Estima, Evanston IL.

Estrada F, Perron P, Martinez-Lopez B., 2013. Statistically derived contributions of diverse human influences to twentieth-century temperature changes. *Nature Geoscience* 6, 1050—1055.

Estrada, F., Perron, P., 2016. Extracting and analyzing the warming trend in global and hemispheric temperatures. *Working Paper*, Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México.

Fyfe J. C., Meehl, G. A., England, M. H., Mann, M. E., Santer, B. D., Flato, G. M., Hawkins, E., Gillett, N. P., Xie, S. P., Kosaka, Y., Swart, N. C., 2016. Making sense of the early-2000s warming slowdown. *Nature Climate Change* 6, 224—228.

Gay-Garcia, C., Estrada, F., Sanchez, A. 2009. Global and hemispheric temperature revisited. *Climatic Change* 94, 333—349.

Gleckler, P. J., Durack, P. J., Stouffer, R. J., Johnson, G. C., Forest, C. E., 2016. Industrial-era global ocean heat uptake doubles in recent decades. *Nature Climate Change* 6, 394—398.

Gonzalo, J., 1994. Five alternative methods of estimating long-run equilibrium relationships, *Journal of Econometrics* 60, 203–233.

Gramling, C., 2015. Lost and found: Earth's missing heat. *Science* 348 (6239), 1066—1067.

Granger, C. W. J., Jeon, Y., 2006. Dynamics of model overfitting measured in terms
Of autoregressive roots. *Journal of Time Series Analysis* 27(3), 347—365.

Granger, C. W. J., Lee, T. H., 1989. Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error correction models. *Journal of Applied Econometrics* 4, S145—S159.

Haldrup, N., 1998. An econometric analysis of I(2) variables. *Journal of Economic Surveys* 12(5), 595—650.

Hannan, E. J., Quinn, B. G., 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B* 41, 190–195.

Hansen, J., Ruedy, R., Sato, M., Lo, K., 2010. Global surface temperature change. *Reviews of Geophysics* 48, RG4004.

Hausfather, Z., Cowtan, K., Clarke, D. C., Jacobs, P., Richardson, M., Rohde, R., 2017. Assessing recent warming using instrumentally homogenous sea surface temperature records. *Science Advances* 3, e1601207.

Hedemann, C., Mauritsen, T., Jungclaus, J., Marotzke, J., 2017. The subtle origins of surface-warming hiatuses. *Nature Climate Change* 7, 336–339.

Hendry, D. F., von-Ungern Sternberg, T., 1981. Liquidity and inflation effects of consumer expenditure. In Deaton, A. S. (Ed.), *Essays in the Theory and Measurement of Consumers' Behaviour*. Cambridge University Press, Cambridge.

Johansen, S., 1995. A statistical analysis of cointegration for I(2) variables. *Econometric Theory* 11, 25—59.

Johansen, S., 1997. Likelihood analysis of the I(2) model. *Scandinavian Journal of Statistics* 24(4), 433–462.

Johnson, G. C., Lyman, J. M., Loeb, N. G., 2016. Improving estimates of Earth's energy imbalance. *Nature Climate Change* 6(7), 639—640.

Karl, T. R., Arguez, A., Huang, B., Lawrimore, J. H., McMahon, J. R., Menne, M. J., Peterson, T. C., Vose, R. S., Zhang, H.-M., 2015. Possible artifacts of data biases in the recent global surface warming hiatus. *Science* 348 (6242), 1469—1472.

Knutti, R., Rugenstein, M. A. A., Hegerl, G. C., 2017. Beyond equilibrium climate sensitivity. *Nature Geoscience* 10, 727–736.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Jones, P. D., 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset. *Journal of Geophysical Research* 117, D08101.

Kaufmann, R. K., Kauppi, H., Stock, J. H., 2006. Emissions, concentrations, and temperature: a time series analysis. *Climatic Change* 77(3—4), 249–278.

Kaufmann, R. K, Kauppi, H., Stock, J. H., 2010. Does temperature contain a stochastic trend? Evaluating conflicting statistical results. *Climatic Change* 101, 395–405.

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., Kauppi, H., Mann, M., Stock, J., 2013. Does temperature contain a stochastic trend: linking statistical results to physical mechanisms. *Climatic Change* 118(3), 729–743.

Kaufmann, R. K., Stern, D. I., 2002. Cointegration analysis of hemispheric temperature relations. *Journal of Geophysical Research* 107(D2), Article No. 4012.

Kejriwal, M., Lopez, C., 2013. Unit roots, level shifts, and trend breaks in per capita output: a robust evaluation, *Econometric Reviews* 32(8), 892–927.

Kongsted, H. C., 2005. Testing the nominal-to-real transformation. *Journal of Econometrics* 124(2), 205–225.

Kongsted, H. C., Nielsen, H. B., 2004. Analysing I(2) Systems by transformed vector autoregressions. *Oxford Bulletin of Economics and Statistics* 66(3), 379—397.

Kosaka, Y., Xie, S.-P., 2013. Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* 501, 403–408.

Kouketsu, S., *et al.*, 2011. Deep ocean heat content changes estimated from observation and reanalysis product and their influence on sea level change. *Journal of Geophysical Research*

116, C03012.

Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., Mishonov, A. V., Reagan, J.R., Seidov, D., Yarosh, E.S., Zweng, M.M., 2012. World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophysical Research Letters* 39, L10603.

Levitus, S., Antonov, J. L., Boyer, T. P., Stephens, C., 2000. Warming of the world ocean. *Science* 287, 2225—2229.

Levitus, S., Antonov, J. L., Wang, J., Delworth, T. L., Dixon, K. W., Broccoli, A. J., 2001. Anthropogenic warming of Earths climate system. *Science* 292, 267—270.

Li, C., von Storch, J.-S., Marotzke, J., 2013. Deep-ocean uptake and equilibrium climate response. *Climate Dynamics* 40, 1071—1086.

Lunt, D. J., Haywood, A. M., Schmidt, G. A., Salzmann, U., Valdes, P. J., Dowsett, H. J., 2010. Earth system sensitivity inferred from Pliocene modelling and data. *Nature Geoscience* 3, 60—64.

Mauritsen, T., Pincus, R., 2017. Committed warming inferred from observations. *Nature Climate Change 7*(9), 652–655.

Marvel, K., Schmidt, G. A., Miller, R. L., Nazarenko, L., 2016. Implications for climate sensitivity from the response to individual forcings. *Nature Climate Change* 6(4), 386—389.

MacKinnon, J. G., 2010. Critical values for cointegration tests. *Queen's Economics Department Working Paper* 1227.

Medhaug, I, Stolpe, M. B., Fischer, E. M., Knutti, R., 2017. Reconciling controversies about the global warming hiatus. Nature 545, 41–47.

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, 360–364.

Meinshausen, M., Elisabeth, V., Alexander, N., Lorbacher, K., Nicolai, M., David, E., Paul, F., Montzka, S., Rayner, P., Trudinger, C., Krummel, P., Beyerle, U., Cannadell, P., Daniel, J., Law, R., ODoherty, S., Prinn, R., Reimann, S., Rubino, M., Velders, G., Vollmer, M., Weiss, R. (2016). Historical greenhouse gas concentrations. *Geoscientific Model Development Discussions*, gmd-2016-169.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A, Stephens, G., Takemura, T., Zhang, H., 2013. Anthropogenic and natural radiative forcing. In: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), *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 University Press, Cambridge and New York.

Miller, R. L., Schmidt, G. A., Nazarenko, L. S., Tausnev, N., Bauer, S. E., Del Genio, A. D., Kelley, M., Lo, K. K., Ruedy, R., Shindell, D. T., Aleinov, I., Bauer, M., Bleck, R., Canuto, V.,

Chen, Y.-H., Cheng, Y., Clune, T. L., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Menon, S., Oinas, V., Pérez García-Pando, C., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Russell, G. L., Sato, M., Sun, S., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., Zhang, J., 2014. CMIP5 historical simulations (1850-2012) with GISS ModelE2. *Journal of Advances in Modeling Earth Systems* 6(2), 441—477.

Myhre, G., D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura and H. Zhang, 2013: Anthropogenic and Natural Radiative Forc-ing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Onatski, A., Uhlig, H., 2012. Unit roots in white noise. *Econometric Theory* 28(3), 485–508.

Palmer, M. D., McNeall, D. J., Dunstone, N. J., 2011. Importance of the deep ocean for estimating decadal changes in Earth's radiation balance. *Geophysical Research Letters* 38, L13707.

Phillips, P. C. B., and Durlauf, S. N., 1986. Multiple time series regression with integrated processes. *Review of Economic Studies* 53(4), 473–495.

Perron, P., Wada, T., 2009. Let's take a break: Trends and cycles in US real GDP. *Journal of Monetary Economics* 56, 749–765.

Phillips, A. W., 1954. Stabilization policy in a closed economy, *Economic Journal* 64, 290—323.

Phillips, P. C., Ouliaris, S. 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.

Pretis, F., 2015. Econometric models of climate systems: The equivalence of two-component energy balance models and cointegrated VARs. *Oxford Department of Economics Discussion Paper* 750.

Pretis, F., Allen, M. 2013. Breaks in trends. *Nature Geoscience* 6, 992—993.

Pretis, F., Hendry, D. F., 2013. Comment on "Polynomial cointegration tests of anthropogenic impact on global warming" by Beenstock et al. (2012) – some hazards in econometric modelling of climate change. *Earth System Dynamics* 4, 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. *Climatic Change* 131, 705–718.

Proistosescu, C., Huybers, P. J., 2017. Slow climate mode reconciles historical and model based estimates of climate sensitivity. Science Advances 3(7), e1602821.

Purkey, S. G., Johnson, G. J., 2010. Warming of global abyssal and deep southern ocean waters between the 1990s and 2000s: contributions to global heat and sea level rise budgets. *Journal of Climate* 23, 6336–6351.

Rahbek, A., Kogsted, H. C., Jørgensen, C., 1999. Trend stationarity in the I(2) cointegration model. *Journal of Econometrics* 90, 265–289.

Richardson, M., Cowtan, K., Hawkins, E., Stolpe, M. B., 2016. Reconciled climate response estimates from climate models and the energy budget of Earth, *Nature Climate Change* 6, 931–935.

Ramaswamy, V. *et al*., 2001. Radiative forcing of climate change. In: Houghton, J. T. *et al.* (Eds.), *Climate Change 2001: The Scientific Basis*, Cambridge University Press, Cambridge.

Rappoport, P., Reichlin, L., 1989. Segmented trends and non-stationary time series. *Economic Journal* 99, 168–177.

Rohde, R., Muller, A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., Wickham, C., 2013. A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinformatics &. Geostatistics: An Overview* 1, 1–7.

Shine, K. P., Derwent, R. G., Wuebbles, D. J., Morcrette, J.-J., 1990. Radiative forcing of climate. In: Houghton, J. T., G. J. Jenkins, J. J. Ephraums (Eds.), *Climate Change: The IPCC Scientific Assessment*, Cambridge University Press, Cambridge and New York, 41–68.

Schlesinger, M. E., Andronova, N. G., Kolstad, C. D., Kelly, D. L., no date. On the use of autoregression models to estimate climate sensitivity. Mimeo. Climate Research Group, Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, IL.

Schmith, T., Johansen, S., Thejll, P., 2012. Statistical analysis of global surface temperature and sea level using cointegration methods. *Journal of Climate* 25, 7822–7833.

Schwarz, G. E., 1978. Estimating the dimension of a model. *Annals of Statistics* 6(2), 461–464.

Siliverstovs, B., 2006. Multicointegration in US consumption data. *Applied Economics* 38(7), 819–833.

Smith, D. M., *et al.*, 2016. Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change* 6, 936–940.

Solomon, S., Plattner, G. K., Knutti, R., Friedlingstein, P. 2009. Irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Sciences 106*(6), 1704–1709.

Stern, D. I., 2006. An atmosphere-ocean multicointegration model of global climate change. *Computational Statistics and Data Analysis* 51(2), 1330–1346.

Stern, D. I., Kaufmann, R. K., 2014. Anthropogenic and natural causes of climate change. *Climatic Change* 122, 257–269.

Stock, J. H., Watson, M., 1993. A simple estimator of cointegrating vectors in higher order integrating systems. *Econometrica* 61, 783–820.

Storelvmo, T., Leirvik, T., Lohmann, U., Phillips, P. C. B., Wild, M., 2016. Disentangling greenhouse warming and aerosol cooling to reveal Earths climate sensitivity. *Nature Geoscience* 9, 286–389.

Tollefson, J., 2014. Climate change: The case of the missing heat. *Nature* 505, 276–278.

Trenberth, K. E., 2015. Has there been a hiatus? *Science* 349, 691–692.

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 Nino to global warming. *Surveys in Geophysics,* 33(3-4), 413–426.

Wigley, T. M. L., Raper, S. C. B., 1992. Implications for climate and sea level of revised IPCC emissions scenarios. *Nature* 357, 293–300.

Zeileis, A., 2004. Alternative boundaries for CUSUM tests. *Statistical Papers* 45(1), 123–131.