ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

SOME FALLACIES IN ECONOMETRIC MODELLING OF CLIMATE CHANGE

David F. Hendry and Felix Pretis

Number 643
February 2013

Manor Road Building, Manor Road, Oxford OX1 3UQ

Some Fallacies in Econometric Modelling of Climate Change

David F. Hendry and Felix Pretis
Department of Economics, and Institute for Economic Modelling,
Oxford Martin School, University of Oxford, UK. *

Abstract

We demonstrate major flaws in the statistical analysis of Beenstock, Reingewertz and Paldor (2012), discrediting their initial claims as to the different degrees of integrability of CO_2 and temperature.

JEL classifications: C1, Q5.
KEYWORDS: Econometric Modelling; Location Shifts; Data Measurements; Climate Change.

1 Introduction

In their analysis of temperature and greenhouse gases, Beenstock *et al.* (2012) present statistical tests that purport to show that those two variables have different integrability properties, and hence cannot be related. The physics of greenhouse gases are well understood, and date from insights in the late 19th century by Arrhenius (1896). He showed that atmospheric temperature change was proportional to the logarithmic change in CO_2). Heat enters the Earth's atmosphere as radiation from the sun, and is re-radiated from the warmed surface to the atmosphere, where greenhouse gases absorb some of that heat. This heat is re-radiated, so some radiation is directed back towards the Earth's surface. Thus, greater concentrations of greenhouse gases increase the amount of absorption and hence re-radiation. To 'establish' otherwise merely prompts the question 'where are the errors in the Beenstock *et al.* analysis?'. We will demonstrate several major flaws in their approach, such that none of their claimed conclusions has any evidential basis.

Section 2 uses an uncontroversial example to highlight the dangers of approaches that fail to address all the complications inherent in statistical analyses of observational data. Section 3 applies the reasoning to the apparently more controversial case of the relationship between greenhouse gases and temperature.

2 A case study

To highlight the problems with the analysis of temperature and greenhouse gases in Beenstock *et al.* (2012), we first use an example where the analysis is completely uncontroversial: road fatalities are due to people killed by or in moving vehicles.

Consider Figure 1 that records total vehicle distances driven in billions of km p.a. (denoted X_t) and road fatalities (Y_t), both for the UK.[1] The four panels labeled a, b, c, d respectively show X_t, Y_t, $X_t - X_{t-1} = \Delta X_t$ and ΔY_t. It is manifest from the graph that X_t and Y_t are highly non-stationary (do not have constant means and variances), and have strong opposite trends. Thus the further vehicles drive, the fewer the number of deaths. We can establish that finding 'rigourously' by a statistical analysis, but however sophisticated that may be, the implications that road fatalities are not due to moving vehicles, or that we can even reduce road fatalities by more driving, both remain absurd.

*This research was supported in part by grants from the Open Society Foundations and the Oxford Martin School. We are grateful to Myles Allen for comments on an earlier draft. Contact details: david.hendry@nuffield.ox.ac.uk and felix.pretis@economics.ox.ac.uk.

[1]Fatalities are only available continuously from 1979 onwards and interpolated from intermittent data for 1930–1979, an issue of some importance below.

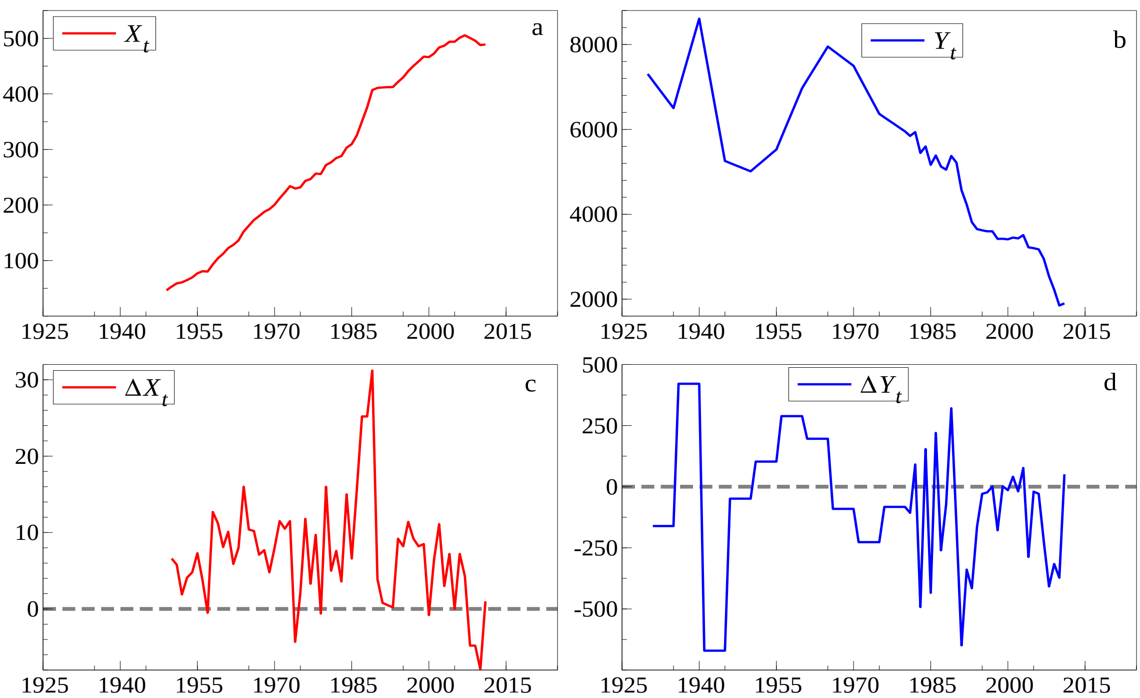

Figure 1: Vehicle kilometers driven and road fatalities p.a. in the UK

How can 'statistical evidence' fly in the face of the obvious? There are at least five key reasons why such a result occurs: omitted variables' bias (omitting relevant explanatory variables); aggregation bias (mixing data from very different populations); unmodelled shifts (when a change in legislation or technology shifts a relationship); incorrectly modelled relations (when the residuals from the estimated relationship do not satisfy the statistical properties of the assumed error processes, so claimed inferences are invalid); and data measurement errors. All five are powerful distorters as we now briefly discuss.

2.1 Omitting relevant explanatory variables

There are many obvious omitted potentially relevant explanatory variables: a partial list would include improved driving standards from more stringent tests; better road safety training; safer cars with improved impact designs and better brakes (abs); seat belts and air bags (see the analysis of the impact of the former in Harvey and Durbin, 1986); separation of traffic flows on motorways; reductions in drunk driving; and so on. Converse effects come from faster driving, driver overconfidence, driving after taking drugs, etc.

2.2 Aggregation bias

Aggregation bias is due to the total data including very distinct sub-populations with different characteristics, across age and sex; geographical location; urban and rural; and road types. For example, replacing a 'one lane each way' road with a motorway (still desperately needed in the North East of both England and Scotland) would increase kilometers driven and probably reduce deaths.

2.3 Unmodelled shifts

Unmodelled shifts (due to many potential causes, including wars) can play havoc with statistical inference: they add an additional non-stationarity to that induced by integrating forces (such as unit roots); distort the relationships between the variables that have been included; lead to residuals with properties that differ from the assumed error processes; and can induce forecast failure out of sample. The impact of the Second World War is visible in the fatalities graph (Figure 1 b).

2.4 Incorrectly modelled relations

Incorrectly modelled relations arise when the wrong functional form is imposed, say linear rather than non-linear, inadequate dynamics are allowed for, inducing residual autocorrelation, or heteroskedastic errors are not handled, all of which entail that estimated standard errors (on which tests are based) are far from the correct sampling uncertainty standard deviation.

2.5 Data measurement errors

Finally, data measurement errors can mislead any form of inference. Figure 1d showing the annual changes in road fatalities illustrates the interpolation over the early sample with constant periods followed by jumps, quite unlike any real data.

2.6 Mistaken inferences

Even when estimated standard errors correctly reflect the sampling standard deviations, there are two possible mistaken inferences arising from: (a) failing to reject a null hypothesis; and (b) rejecting a null hypothesis using a test with power against a specific alternative. We take theses in turn.

(a) Consider a sample of 100 observations on an accurately measured variable Z_t. The sample mean, $\widehat{\mu}$ is 0.005 and the estimated standard error $\widehat{\sigma}$ is 0.05. Then, under the hypothesis that $Z_t \sim \mathsf{ID}[\mu, \sigma^2]$, denoting independent sampling from a constant distribution with mean μ and variance σ^2, a Student's $\mathtt{t}$ test of the null hypothesis that $\mu = 0$ has the value of approximately 1.0. The null is not rejected at any reasonable significance level. But neither is the null that $\mu = 0.0025$ or even $\mu = -0.0025$. When these are quarterly growth rates of real income per capita, there is a dramatic difference between the substantive outcomes not reflected in the statistics, namely no growth; growth of approximately 1% p.a. and real incomes falling at 1% p.a. Not rejecting the null does not entail it is true, merely that evidence is inconclusive.

(b) Conversely, in the example just discussed, an investigator decides to test the assumption that the $\{Z_t\}$ are independent draws against the alternative that the series is a first-order autoregression, and strongly rejects the null hypothesis of independence. While that vitiates the analysis in (a), it does not imply that $\{Z_t\}$ is a first-order autoregression. Indeed it does not even imply that they are not drawn independently (see Mizon, 1995): the cause of residual autocorrelation could be due to a non-linear approximation or an unmodelled location shift (so μ has different values at different times, as has happened historically).

The occurrence of any or all of these problems precludes establishing the presence or absence of any meaningful bivariate relationship. Despite fatalities and distance driven having very different statistical properties, and even opposite trends, moving vehicles cause road deaths.

3 Statistical fallacies in 'polynomial cointegration tests'

All six problems just discussed potentially apply to the analysis in Beenstock *et al.* (2012), but three in particular stand out on just viewing the data series.

3.1 Incorrectly modelled relations

Although those authors reference Hendry (1995), they evidently fail to learn its main message: before any ststistical inference can be conducted, a model must be congruent, or well-specified in that it satisfies the assumptions on which the statistical analysis relies. All of the tests in their Table 1 make many untested assumptions, including accurate data and that there is a single regime.

3.2 Data measurement errors

We obtained the data on greenhouse gases used by Beenstock *et al.* (2012) in a similar fashion, using the values provided by Myhre, Highwood, Shine and Stordal (1998) to convert the series into their radiative forcing equivalents. Beenstock *et al.* omit the fact that the measured series of GHGs come from a variety of different sources. If the measurements were identical in all sources, this would not be an issue: however, our graphs reveal sharp differences in the data properties. Consider the CO_2 and N_2O series. Both are initially based on ice core data (indicated by the vertical line in Figure 3.2: 1850 until 1958 for CO_2 and 1850 until 1978 for N_2O) followed by flask and other measurements there-after. Figure 2b shows that up until approximately the point when the switch from ice-core to non-ice core data was made, almost all the changes have precisely the same magnitude, clearly revealing an artificial pattern different to the latter half of the sample. Nevertheless, the data are analyzed as if they come from the same populations.

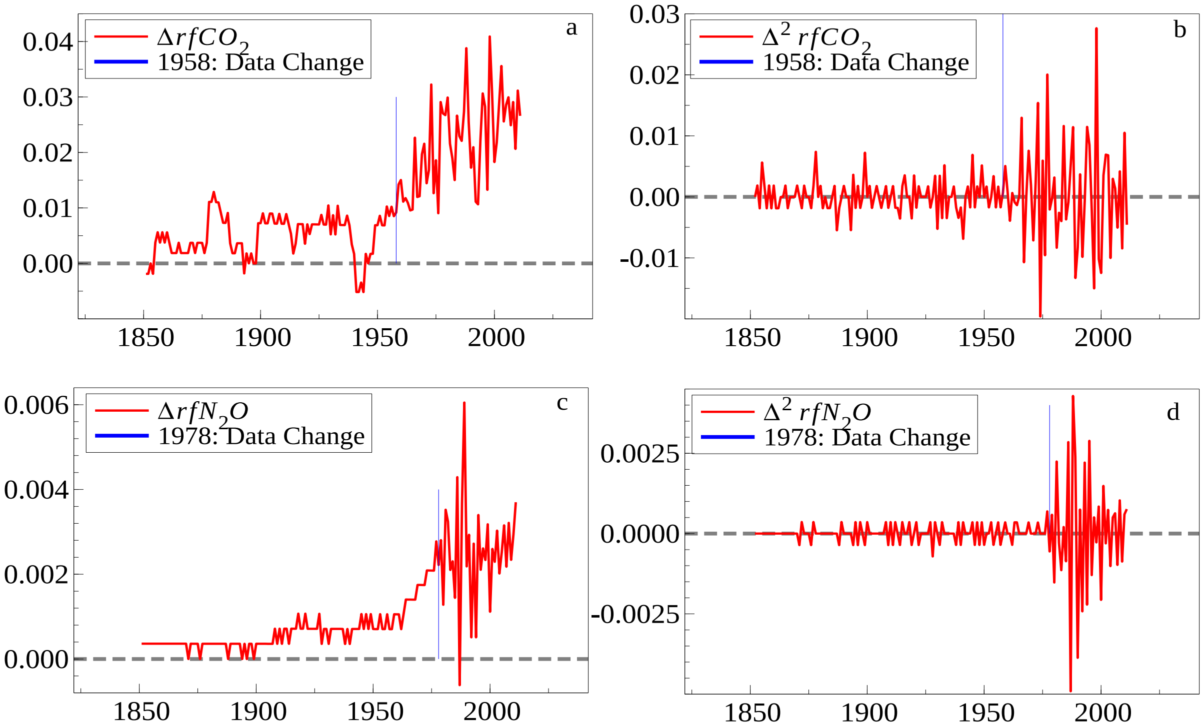

Figure 2: Time series of the first and second differences of rfCO2 and rfNO2

We worry that their Figure 2—reproduced here as Figure 3.2—was shown in one of the few formats that helps to camouflage the serious problem of regime shifts.

Interacting with unmodelled shifts, measurement errors can can lead to false interpretations of the stationarity properties of data. In the presence of these different measurements and structural changes, a unit-root test on the entire sample could easily not reject the null hypothesis of I(2) even when the data are clearly I(1). Indeed, once we control for these changes, our results (see Table 1 and Table 2 below) contradict the findings in Beenstock *et al.*

3.3 Sub-sample unit-root rests

As is to be expected following the inspection of the data, the first difference of radiative forcing of CO_2 is stationary initially around a constant (over 1850–1957) and then around a linear trend (over 1958–2011). Although these tests are based on sub-samples, there is clearly sufficient power to reject the null hypothesis of a unit root. In a similar manner, unit-root tests reject non-stationarity of the first difference

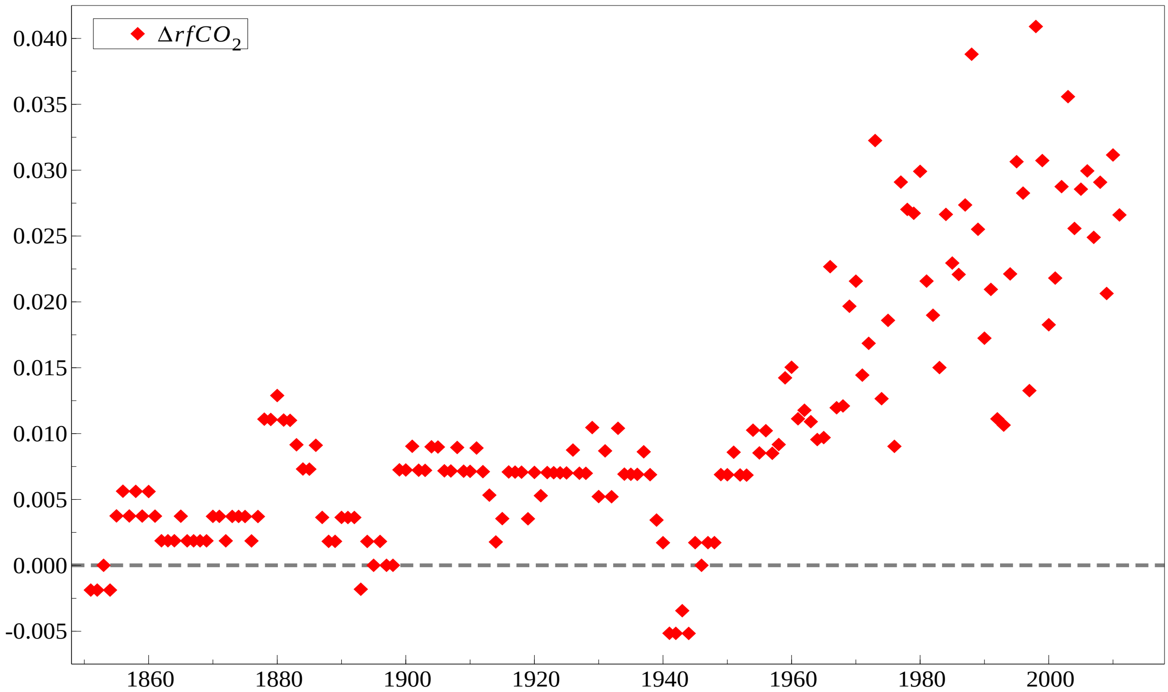

Figure 3: Time series of the first differences of rfCO2 from Beenstock *et al.*

of N_2O for the second set of observations (1978–2011). Unit-root non-stationarity cannot be rejected for 1850–1978 for N_2O: however, given the manifestly artificial appearance of the data such a result should be interpreted with extreme caution. Particularly, the ΔN_2O series appears to exhibit a step shift in the early 1900s, which also leads to spurious results in unit-root tests.

Given these time series properties, the starting point of the analysis in Beenstock *et al.* (2012) is incorrect, and their findings appear to be an artefact of pooling data with very different measurement systems and behaviour in the two sub-samples.

Table 1: ADF Unit Root Tests on ΔrfCO2, ** reject at 1%, * 5%

1850–1957 constant			1958–2011 const. & trend		
D-lag	**t-adf**	**Reject H0**	**D-lag**	**t-adf**	**Reject H0**
5	-3.737	**	5	-4.089	*
4	-2.91	*	4	-3.807	*
3	-2.948	*	3	-3.383	
2	-3.146	*	2	-4.197	**
1	-2.706		1	-5.365	**
0	-3.544	**	0	-6.563	**

4 Conclusion

An appropriate analysis of this data would require separate models for the pre and post ice-core measurements, taking account of the myriad influences impinging on the climate and temperature, its composition and distribution, as well as all the important sinks and sources. The aim of this brief note is merely to

Table 2: ADF Unit Root Tests on ΔrfN2O, ** reject at 1%, * at 5%

1850–1978 const. & trend			1978–2011 constant		
D-lag	**t-adf**	**Reject H0**	**D-lag**	**t-adf**	**Reject H0**
5	2.098		5	-3.832	**
4	1.864		4	-3.347	*
3	1.427		3	-3.636	**
2	0.8014		2	-4.048	**
1	-0.6185		1	-4.793	**
0	-3.87		0	-7.845	**

demonstrate that the conclusions claimed by Beenstock *at al.* about the different degrees of integrability of temperature and CO_2 are rejected once the regime-shift nature of the measurement system is taken into account. Indeed, a simple bivariate plot of temperature and log(CO_2ML) over the second period, matched by means and ranges, suggests the obvious: they are closely related. Such results match the most recent findings on Global Temperature from NASA:
see http://science.nasa.gov/science-news/science-at-nasa/2013/15jan_warming/

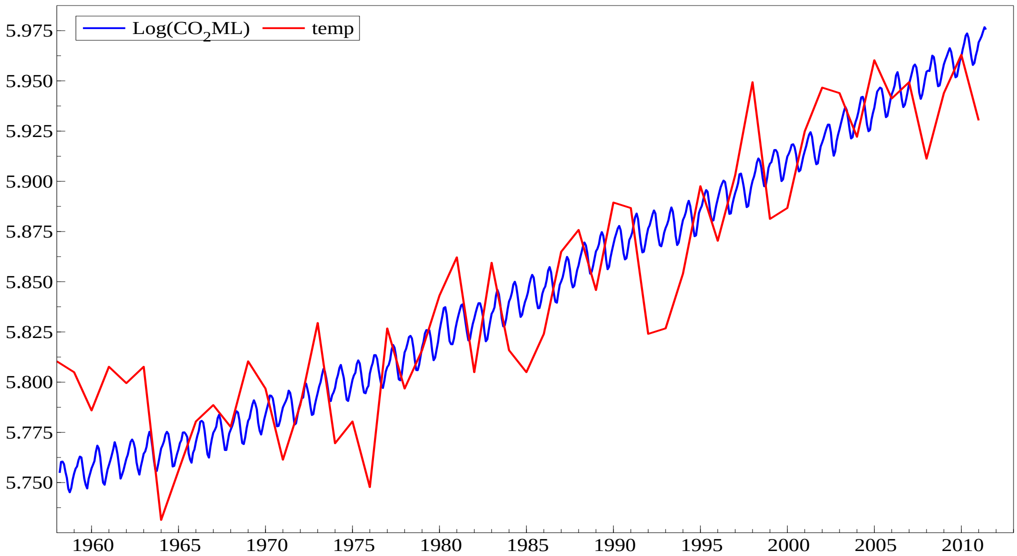

Figure 4: Time series graphs of temperature and log(CO_2ML), matched by means and ranges

References

Arrhenius, S. A. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science (fifth series)*, **41**, 237–275. http://www.globalwarmingart.com/images/1/18/Arrhenius.pdf.

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

Harvey, A. C., and Durbin, J. (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society, Series B*, **149**, 187–227.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Mizon, G. E. (1995). A simple message for autocorrelation correctors: Don't. *Journal of Econometrics*, **69**, 267–288.

Myhre, G., Highwood, E. J., Shine, K., and Stordal, F. (1998). New estimates of radiative forcing due to well mixed greenhouse gases. *Geophysical Research Letters*, **25**, 2715–2718.