Referee's Report:

"Polynomial cointegration tests of anthropogenic impact on global warming"

by M Beenstock, Y Reingewertz and N. Paldor

This paper reports the results of a data analysis seeking a statistical link between global temperature and supposedly anthropogenic effects on the atmosphere, specifically concentrations of CO_2, methane and nitrous oxide measured over the last 120+ years.

The context of this analysis is the commonly asserted claim that because both CO_2 concentration and global temperature have increased over the past century, so there must be a causal association between them, the so-called anthropogenic global warming (AGW) hypothesis. Of course, there is a physical theory (the "greenhouse effect") that predicts this association, in principle, but climate dynamics appear to be so complex in practice that whether the effect is trivial or catastrophic appears difficult to determine.

GCMs in common use by climate scientists are said to be constructed and calibrated on the assumption that the observed common trend is due directly to the greenhouse effect. When extrapolated, this assumption easily allows such models to generate catastrophic future scenarios. However, the observed upward drift in global temperature can evidently have many possible causes. There is now strong evidence that errors of measurement due to the urban heat island effect are an important factor.[1] There is also the simple fact that the natural cycles that produced the medieval warm period and little ice age must still be in operation. In short, common upward (or for that matter, downward) drifts cannot of themselves anything about the relationships between time series, causal or otherwise. This fact has been know at least since Yule reported the correlation (0.95) between standardized mortality and the proportion of marriages solemnised by the Church of England, 1866-1911.[2] Another commonly cited reference on the spurious regression phenomenon is Hendry (1980)[3].

However, as first pointed out by Clive Granger, there is one situation where it can be possible to extract more informative conclusions from time series evidence, by invoking the phenomenon of a *stochastic* trend – the type of pattern represented mathematically by a random walk, or Brownian motion process. If the time series in question have Brownian characteristics, then observing cointegration (in effect, being able to fit the random changes in drift of two or more series together) would provide positive evidence of association – though not of causation, note, without at the least an examination of short-run dynamics.

The form of time series process appropriate to this model is referred to as I(1), having the property that the sequence of first differences is a stationary process and (at worst) weakly dependent. There is a range of statistical tests (the Dickey-Fuller and Phillips-Perron procedures being the most popular) having the I(1) property as the null hypothesis. These tests are typically implemented by considering

[1] See for example R. McKitrick and L. Tole, "Evaluating explanatory models of the spatial pattern of surface climate trends using model selection and Bayesian averaging methods" *Climate Dynamics* (in press) 2012.

[2] G. Udny Yule, "Why do we sometimes get nonsense correlations between time series?" *Jnl. Royal Statist. Soc.* 89 (1926), 1-68

[3] D. F. Hendry, "Econometrics: Alchemy or Science?" *Economica* 47, (1980), 387-406.

the one-sided alternative hypothesis under which the series are closer to being stationary and trend-free.

In the event these authors, who have gone to a great deal of trouble to obtain results that are comprehensive and robust, find that while the temperature and solar irradiance series are indeed I(1) on the standard tests, the so-called anthropogenic greenhouse gas (GHG) series are I(2), requiring differencing twice to yield a stationary series.

This fact immediately rules out the possibility of extracting evidence for AGW from an analysis of the time series. The prerequisites for a cointegrated relationship do not exist. This is not to say that the common upward drifts *cannot* be causally connected. It's simply that this avenue to showing a statistically significant relation is ruled out. Previous attempts at cointegration tests that overlooked this fact, cited in the paper, must accordingly be discounted.

The authors also show evidence of a cointegrating relationship between the temperature (corrected for solar irradiance) and changes in the anthropogenic variables. This is odd since the physical theories do not predict such a relation. This could of course be a case of a false positive, but is nonetheless a finding worth reporting.

In summary, this paper reports some important evidence regarding the AGW controversy. The results show that simply examining the historical record cannot throw any light on AGW, one way or the other. This is not the same as showing evidence against AGW, but it does debunk effectively, using an impeccable statistical methodology, the naïve argument from upward drift. This paper is an important contribution to the literature on AGW although it does require some revision, as detailed in the comments, before being suitable for publication.

Text comments (page order)

1. (Page 563, line 22) "For the correlation to be genuine … " should read "for the relation to be genuine…". (We agree, that cointegration is not correlation.)

2. (Page 565, line 28-566, line 1) No! The null hypothesis of the KPSS test is *not* stationarity. It is "I(0)", which means, essentially, stationary and weakly dependent. More technically, I(0) means that the process has summable autocovariances (but not summing to zero) and hence that its spectral density is both finite and strictly positive at the origin. You are quite correct to be using the KPSS test to provide reinforcing evidence on integration order, since if you cannot reject I(0), this reinforces the conclusion that the process is not I(1). However, it is perfectly possible for the process to be in the class of fractionally integrated processes, I(d) for $0 < d < 1$. Such processes are stationary for the cases with $0 < d < ½$, but are not weakly dependent, and certainly not in the null set of the KPSS test. Rejection on both tests is definitely not an anomalous result, but it does require one to entertain a wider class of stochastic processes that I(0) and I(1).

3. (Page 566, lines 5-8) This sentence is confused. Please see Point 2 above for clarification.

4. (Page 566, line 10) "Trend stationary" variables are not I(1). Remember that I(0) includes the condition that that the autocovariance sequence has a non-zero sum (i.e. positive spectral density at the origin). According to this rule, which excludes the "over-differenced" (I(−1)) cases in particular, trend-stationary series are I(0) after subtracting the deterministic component.

5. (Page 566, line 27) The PP statistic in fact corrects both the mean and the standard deviation of the DF statistic, to allow for autocorrelation. However, maybe this sort of detail is unnecessary for this paper?

6. (Page 566, line 28) The KPSS uses an HAC estimate of the variance in common with the PP correction, but otherwise the corrections are different. Again, is this kind of detail needed, for your target readership?

7. (Page 567, line 21) Of course, directions of causality cannot be identified from cointegrating relations, which are long-run and (on that time scale) inherently simultaneous. They could be due to causal factors running in either direction or, indeed, neither direction if the observed relation is due to an unobserved common cause.

8. (Page 570, line 7) Under-powered means high probability of false negatives (failure to reject false null), *not* of false positives. “Accept false positive results” is a confused form of words, because “false positive” implies a *rejection* of the null.

9. (Page 570, lines 17,18) The powers of T appear to be misprints. Should these read “T-consistent” and “T^2-consistent” respectively?

10. (Page 571, line 7) The “SEBM” requires, at least, a literature reference.

11. (Page 571, line 15) In equation 2, the upper limit of the sum should be t–1 not ∞. $\kappa\rho^t$ must evidently represent the sum of terms from t onwards.

12. (Page 573, line 8) Note that equation (4) (with $F = F_A + F_B$), together with equation (6), represent a system with cointegrating rank 2. In the absence of restrictions (given that F_A and F_B are actually sums of terms with unknown coefficients) this system is not identified. In other words, running regressions with the form of (5) and the counterpart of (6) with appropriate substitutions will give the same result in the limit, apart from normalization, and the estimated coefficients will be an unknown linear combination of the two cointegrating relations. However, the “reduced form” equation (7) is unique (the reduced system excluding F_B has cointegrating rank 1) and its coefficients are identified. While β is not identified, it is true that $\psi_1 = 0$ if and only if $\beta = 0$. The exposition here is not incorrect, but I suggest it might be helpful to convey the relevant ideas in this framework, by invoking the system of cointegrating relations.

13. (Page 574, line 13). The paper by Clemente et al. (1998) appears to describe a modification, for allowing a double break, of the test of Perron and Vogelsang (1992). There is no mention of a double break here. Should not the reference be to the 1992 paper?

14. (Page 574, line 15) “… break occurs in 1964, …”. This sentence is ambiguous, please clarify! What was the null hypothesis under test, here? If this was “(I(2)” with alternative “I(1) with a break in trend” , and the result was non-rejection, as the quoted statistic suggests, then the date 1964 is irrelevant. The test might report 1964 as the point in the sample where the statistic was minimized, but under the null hypothesis this is nothing but a random drawing. Quoting it must surely confuse the reader.

15. (Page 574, line 17) Table 2 should give the values of the test statistics, and indicate critical values or p-values. Readers must have the opportunity to see directly the evidence for the

integration orders. It's not enough to just give the outcome of tests which have error probabilities attached.

16. (Page 575, line 9) See the footnote: Granger and Lee (1989) coined the term "multicointegration" to describe a situation where there is a cointegrating vector for I(1) variables whose cumulated residuals are also I(1) and cointegrated with the other variables of the model. I guess that this describes the present situation if we think in terms of the I(1) series of differenced GHG concentrations. It might be good to explain all this a bit more explicitly.

17. (Page 575, line 12) Using the term "anthropogenic trend" to refer to the cointegrating residual g_1 is extremely odd. This is the component of the trend in CO_2 concentration *not* explained by the trends in the other GHG concentrations. Since we have the three GHG variables here, the cointegrating rank of this sub-system is either 1 or 2. If it is 1, then equation (9) represents the unique cointegrating vector, and the subsystem has two common trends, that might be called anthropogenic trends. If the cointegrating rank is 2, on the other hand, there is a single common trend. This would be the unique anthropogenic trend, in other words, the trend shared by all three series). But g_1 is none of these! It might be worth investigating this issue further, using Johansen's rank test?

18. (Page 576, line 3) The presentation of test outcomes in graphical form does have a point in terms of exposition, but I think that telling readers the actual test outcomes should take precedence. The figures are also somewhat misleading, since readers will think that the areas of the shaded boxes have some significance, when in fact they are arbitrary. I urge the authors to replace Figures 2 and 3 with tables of test statistics and critical values or *p*-values.

19. (Page 576, line 12) "… slightly submarginal."? Odd choice of terminology! Please be more explicit.

20. (Page 577, line 7) In equation 12, g_2 appears to have a negative coefficient. Why is this called positive here?

21. (Page 578, line 7) Evidently a sign is missing in equation (13).

22. (Page 578, line 21). The words "KPSS statistic" do not refer to a test. Of course the KPSS test is not a test for cointegration, polynomial or otherwise. I guess what is meant here is the Shin (1994) test, which applies the KPSS statistic to regression residuals. Better to say this explicitly.

23. (Page 579, line 23) Words "with global temperature" are redundant, please delete. Why not give the results of this cointegration test? Bound to be of interest to readers.

24. (Page 579, line 26) Note that equation (15) has not yet been given, and it contains no I(2) variables. Should this refer to something else?

25. (Page 580, line 5) I recommend handling the treatment of nonlinear transformation with caution! Integration and differencing are linear operators, and their properties are not preserved under nonlinear transformations. The logarithm of an I(*d*) process is not I-anything! There is certainly no possibility of an I(2) process being transformed to I(1), except by the linear operations of differencing and cointegration.

26. (page 580, line 7) The Banerjee et al. (1993) reference is missing. Does it really say that? I'm sceptical!

27. (Page 580, line 18) Similarly to the last comment, I cannot conceive of the product of an I(2) and I(0) being I(1). The whole discussion of this paragraph is misguided, and should be deleted.

28. (Page 581, line 14) I think that in the equation, CH_2 is a misprint for CH_4.

29. (Page 581, lines 19-20) It is perfectly possible to have the KPSS and ADF/PP test both reject, even asymptotically. Please see Point 2 above. Non-rejection by the KPSS (i.e.. Shin 1994) test is certainly supporting evidence for cointegration, but if (as I think we should) we discriminate between I(1) residuals and I(d) for $d < 1$, then cointegration does not imply I(0) residuals, Indeed, the residuals could be nonstationary (the fractional cases with $½ \leq d <1$) but still technically cointegrating. I think that most people would accept this designation. I suggest that it would be worth adding a paragraph somewhere to explain these distinctions.

30. (Page 583, line 9) The reference given to the Engle-Yoo paper is the wrong one. There are two papers by these authors, both reprinted in the Engle and Granger volume "Long-run Economic Relationships" (OUP 1991) . (This volume is the correct reference to give, since the paper cited did not appear anywhere else.)

31. (Page 583, second paragraph) There are a number of references to alternative estimates of equations previously quoted, but this is a bit sloppy. The equations referred to are particular estimated cases, not generic forms, so the discussion is confusing. I suggest that a solution might be to give the equations in generic form with coefficients α_1, α_2 and so on, and then give the estimated coefficients in a table. This would also have the advantage of letting the alternative estimates (Engle-Yoo 3-stage, DOLS) be easily reported, in the same table. These would be of interest to readers.

32. (Page 583, line 24) Please note that the usual DF cointegration test tabulations are not appropriate for residuals estimated by DOLS, since the limit distributions are different. Best to avoid comparisons unless you have suitable tables for the latter case.

33. (Page 586, line 19) A Type 1 error is surely a false *positive* (rejecting a true hypothesis) A Type II error is a false negative. See also Point 8 above. While usage is not completely clear-cut, it seems to me that the majority of people would think of rejecting the null as the "positive" outcome. If we cannot agree about this, then it would save confusion to omit these designations entirely.

34. (Page 591) As remarked before, it's essential to give the actual test outcomes here. The"*d*" designation is a sample estimate, not a datum, so we should see the evidence.

35. (Page 593). It would be helpful to plot the differences of the GHG series. Larger, more detailed plots of all the series would also be desirable

36. (Pages 594/595) I strongly recommend, as mentioned before, replacing these plots with tables of the test statistics, and critical values and/or *p*-values. While the available tables are selective so exact *p*-values cannot be given, one can report *p*-value bounds from the tabulated points, for example, "$0.01 < p < 0.05$". Other authors use a star system - 1 star for rejection at 10%, 2 stars for 5%, 3 stars for 1%, and so forth.