Greenhouse statistics

Tol, R.S.J.; de Vos, A.F.

1993

Faculteit der Economische Wetenschappen en Econometrie
Vakgroep Econometrie

Serie Research Memoranda

Greenhouse Statistics
a different look at climate research

Richard S.J. Tol and
Aart F. de Vos

Research Memorandum 1993-77

December 1993

vrije Universiteit *amsterdam*

GREENHOUSE STATISTICS
a different look at climate research

Richard S.J. Tol[†] *and Aart F. de Vos*[‡]

23 December 1993

† Institute for Environmental Studies, Vrije Universiteit, De Boelelaan 1115, NL-1081 HV Amsterdam, The Netherlands

‡ Department of Econometrics, Vrije Universiteit, De Boelelaan 1105, NL-1081 HV Amsterdam, The Netherlands

Greenhouse Statistics — A Different Look at Climate Research

The greenhouse effect is a hot topic. Although there is some consensus on the enhanced greenhouse effect being real[1], the unequivocal detection of human-induced climatic change is not expected before the turn of the century[2], though some expect it to take much longer[3]. Until then, uncertainty is bound to prevail and this may easily confuse decision-makers. This is one of the causes of the delay in the policy-making on greenhouse gas abatement[4]. But delay could be harmful. Therefore, we think it is worthwhile to map these uncertainties by aiming at a statistical statement on the size and the significance of the impact of the rising atmospheric concentrations of carbon dioxide (CO_2) — as a proxy to all greenhouse gases — on the global mean surface air temperature (*GMT*). We analyse the relationship between *GMT* and atmospheric CO_2 for the period 1870-1991 with a variety of models. Our conclusion is that the expected effect corresponds with the climatological outcomes[2] and that, even when ample room is left for the long-term natural variability, the hypothesis that the observed temperature rise is not related to the enhanced greenhouse effect is rejected at the 1% significance level. This research article is an update and extension of earlier work[5].

The important question whether and by how much greenhouse gases influence the climate cannot (yet) accurately be answered by the global circulation models; this is due to the insufficient climatological knowledge, the limited computer-capacity and the large natural variability and chaotic nature of the system[2]. Our alternative are sophisticatedly simple statistical models, inspired by econometrics where the failure of large models has led to a return to simplicity[6,7].

Earlier statistical models are in our view less efficient. They use pre-whitened data[8,9,10], detrended data[11], calculate signal-to-noise ratios in an improper way[7,8] or do not consider the influence of El Niño[10,12]. And above all: they ignore the long-term natural variability.

The most relevant facts are fairly simple. The *GMT* (Climate Research Unit, East Anglia) shows nonstationary behaviour, irregular but with an overall rise of 0.5°C, with a slight decrease between 1940 and 1975 (cf. Fig. 2). The carbon dioxide[13] record is explosive; its natural logarithm increases linearly up to 1960 and accelerates afterwards. In time series jargon: the $\ln CO_2$ series is nonstationary, i.e., without fixed equilibrium level or even trending, the temperature series might be so.

Relations between nonstationary series have received ample attention in the recent econometric literature[14,15]. On the one hand there is the risk of *spurious correlation*: independent nonstationary series always seem to correlate. On the other hand, if the nonstationary character of a series (*GMT*) is explained by another nonstationary series (CO_2), this *cointegration*[14] leads to robust (i.e. independent of modelling details) estimates of the effect. Some authors have applied this theory to the link CO_2-GMT[12,16].

We share the opinion[17] that this simple procedure does not answer the most relevant question. The results of two simple models will show what is going on and point the route to manoeuvre between the Scylla of spurious correlation and the Charybdis of throwing aside too much information.

The first model is

$$GMT_t = \underset{(2.6527)}{-28.9157} + \underset{(0.4649)}{5.0460} \ln CO_{2;t-20} + \varepsilon_t \tag{1a}$$

with

$$(1 - \underset{(0.0828)}{0.4156}\, L)\varepsilon_t = u_t; \qquad \hat{\sigma} = 0.1038 \tag{1b}$$

an ARX(1)-model: temperature is explained from a constant, the natural logarithm of atmospheric carbon dioxide, with a lag of twenty year, and from an Gaussian AR(1)-process ε_t, representing unexplained stationary deviations. The main criterion to judge the model quality is σ: the estimated standard deviation of the one-step ahead forecast error.

The logarithmic specification is in line with the theory; the data are little informative in this respect but if one adds the 'fact' that without the greenhouse effect the earth would be about 33°C colder, one has to choose for the natural logarithm. The twenty-year lag also mainly stems from the theoretically expected slow response of the *GMT* to changes in atmospheric CO_2; in subsequent models we will use 'distributed lags' which is more elegant but does not change the conclusions. The data mildly support the idea of a slow response.

Replacing the AR(1) process by more elaborate stationary processes (or even first-order nonstationary processes) does neither improve the fit nor change the conclusions. So, in any model similar to (1), we find a strongly significant effect of CO_2 — in (1) a t-value of 10.85. However, adding a deterministic trend to account for a possible 'spontaneous' long term rise in the temperature, we get the following result (t scaled such that the coefficient represents the rise in a century):

$$GMT_t = -\underset{(12.2714)}{35.0133} + \underset{(2.1990)}{6.1398} \ln CO_{2;t-20} - \underset{(0.2161)}{0.1098}\, t + \varepsilon_t \tag{2a}$$

with

$$(1 - \underset{(0.0836)}{0.4101}\, L)\varepsilon_t = u_t; \qquad \hat{\sigma} = 0.1042. \tag{2b}$$

Model (2) is almost identical to model (1) but for the t-value of the CO_2 coefficient which drops to 2.79, still significant at the 95% level but considerably less convincing. The message is clear: lnCO_2 explains the rise in *GMT* much better than a linear trend (standard model selection would simply discard the extremely insignificant term trend) but if one believes that a spontaneous long term rise is possible, model (2) provides the proper picture.

Prior belief in the possibility of a century-long spontaneous rise in temperature is thus crucial in judging the outcomes. In Bayesian statistics, prior beliefs are essential in making probability statements; often subjective priors are used and there is much debate on the value of the resulting inference. In this case however, reasonably objective prior knowledge is available in the form of a *GMT* record for the 10,000 years preceding our sample (ref. 2, page 202, middle figure). This 'prior *GMT*' record shows that long periods with considerable changes in temperature have occurred, but also that these changes develop gradually (the recently reported large and abrupt changes in temperature[18,19] are, sofar, local events and the presumed trigger to these events, a collapse of Greenland's ice-sheet[20] would not have passed unnoticed). A time series analysis of the prior *GMT* record predicts a rise of 0.01°C in the 20th century, with a standard error of 0.12°C (thus, a rise of more than 0.25°C per century is *a priori* implausible). Re-estimating (2) with the corresponding prior (in classical statistics known as mixed estimation[21]) results in virtually the same estimates for the CO_2 parameter, but its t-statistic rises to 4.58. For a prior standard deviation of 0.24°C/century — rather conservative — a t-value of 3.37 results, still considerably more significant than the 2.79 of (2) which corresponds to an infinitely large prior standard error.

If these simple models are replaced by a more sophisticated one, the message that the influence of the enhanced greenhouse effect on the *GMT* is real appears to be further confirmed. We replace the twenty-year lag of lnCO_2 by a 'distributed lag' (a second order Almon[22] lag with 40 lags and zero restrictions at both sides), representing a gradual effect on *GMT*. A couple of popular other explanatory variables of the global mean temperature are also taken up. These are the dust veil index (*DVI*) for the volcanic activity[23], the sunspot numbers (*SSN*) for the solar activity[24] (both are divided by thousand to get the parameters in a proper range) and the southern oscillation index[23] (*ENSO*); these indices do not raise problems with respect to multicollinearity. Lagged *GMT* captures the first-order auto-correlation. This change in presentation requires some re-scaling of the direct regression results, which is done in such a way that the coefficients of CO_2 and t (with an *a priori* standard error of 0.12°C/century) may directly be compared to the previous results. The regression results are:

$$\begin{aligned} GMT_t = & \underset{(5.1794)}{-17.9785} + \underset{(0.0780)}{0.4309}\, GMT_{t-1} + \underset{(1.5179)}{5.5317}\, (1-0.4309)\ln CO_{2;ALM(40,2)} + \\ & + \underset{(0.1858)}{0.3792}\, SSN_{t-1} - \underset{(0.0329)}{0.0407}\, DVI_t - \underset{(0.0372)}{0.1182}\, DVI_{t-1} - \underset{(0.0372)}{0.0998}\, DVI_{t-2} - \\ & - \underset{(0.0352)}{0.0619}\, ENSO_t - \underset{(0.0116)}{0.0332}\, ENSO_{t-1} - \underset{0.0841)}{0.0386}\, t + u_t \end{aligned} \qquad (3)$$

$$\hat{\sigma} = 0.0879.$$

The residual standard error is considerably lower due to the significant and plausible (in sign) contributions of the added explanatory variables.

Figure 1 depicts the influence of the short-term forcings. El Niño contributes 78% to the explained short-term variability, volcanic dust 14% and solar activity only 8%. The observed stabilisation of the *GMT* from 1940 to 1975 is, by our model, explained by a little overshooting of the *GMT* at the end of the 30's plus a restoring to the equilibrium value in the early 40's, followed by a period of a strong negative *ENSO* forcing, in turn followed by another period of strong negative *ENSO* accompanied by renewed volcanic activity. Our model attaches more weight to El Niño than Jones[25]; this may be explained by our inclusion of other variables.

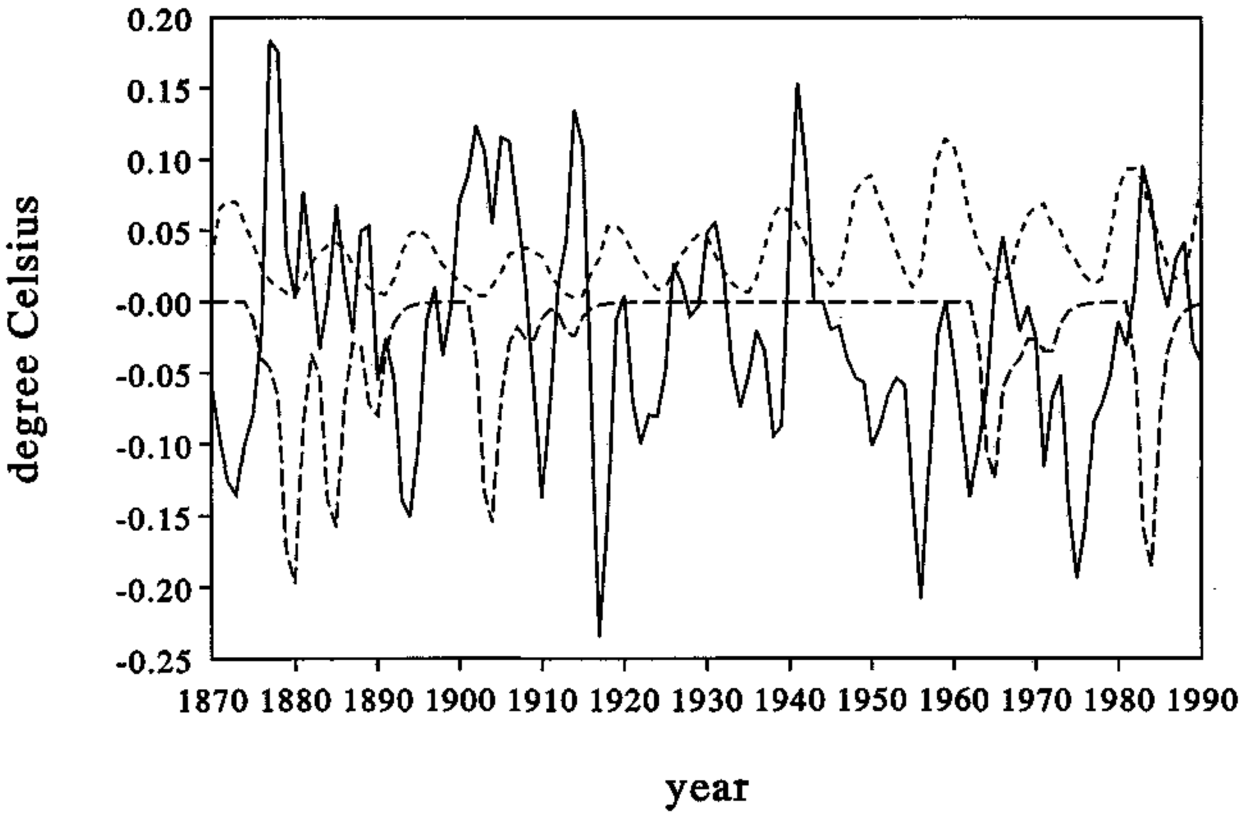

Figure 1. The influence, expressed in degrees Celsius, of *SSN* (dotted line), *ENSO* (solid line) and *DVI* (dashed line) on *GMT* according to model (3).

The model is extensively tested for normality[26], serial correlation[27], heteroscedasticity[28] and non-linearity[29] and found to perform well. The RESET test[30] is not passed, however, which may point at some misspecification. This little flaw is more than compensated by another result. Re-estimating the model for the period up to 1940 and forecasting the remaining 51 observations (conditional on the exogenous variables, but using the forecast of *GMT* as lagged variable) provides Figure 2. The overall quality of the forecast is quite remarkable for a time-series model like this, and restores the confidence in the parameter stability and other aspects.

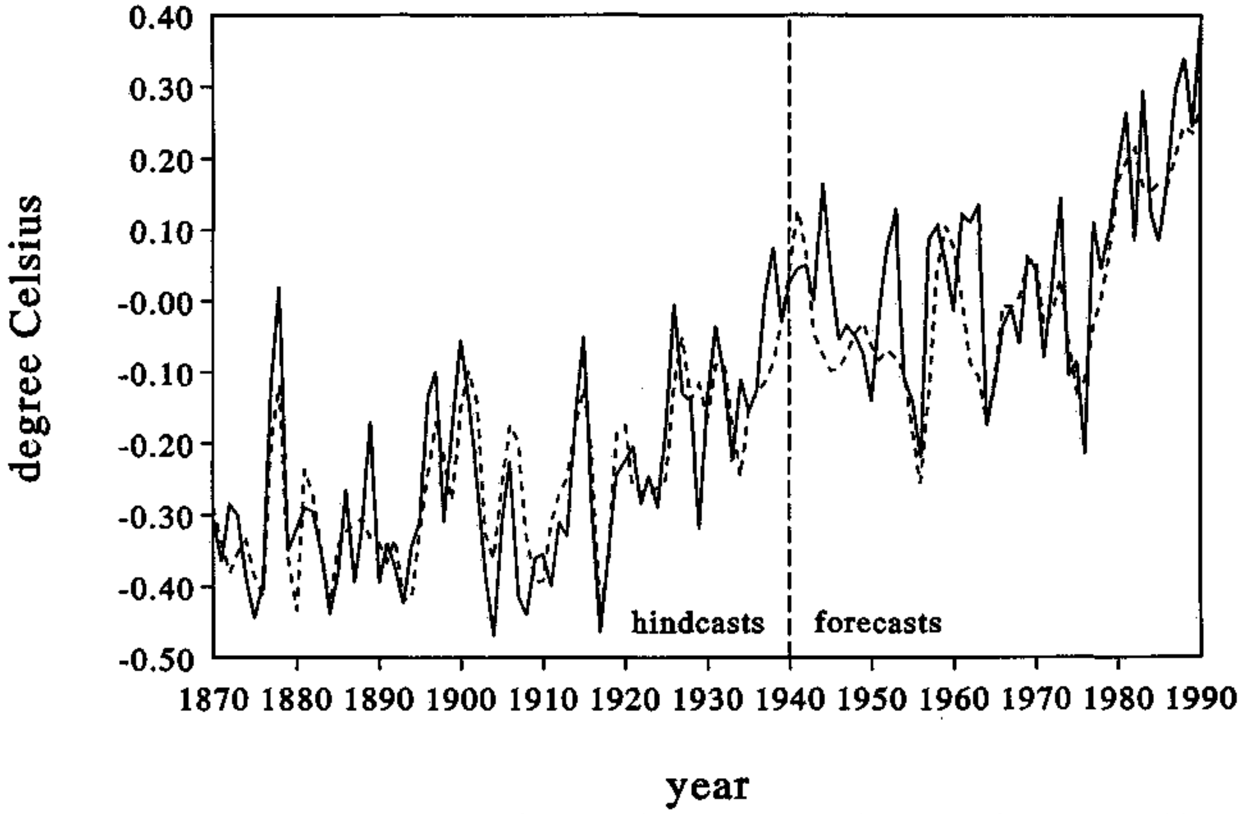

Figure 2. *GMT* as observed (solid line) and modelled (dotted line) by model (3), without trend. The period 1870-1940 (hindcasts) is used to estimate the parameters, the period 1941-1991 (forecasts) is used to validate the model.

The results on the CO_2 coefficient depend, like before, on the prior for the trend coefficient. Table 1 contains the outcomes for various priors, together with those for the ARX(1) model. All coefficients and standard errors are multiplied by ln2 in order to obtain the effect of a doubling of CO_2. The conclusions of model (2) are stronger due to the better description of the temperature. Noteworthy is that in the general model the greenhouse effect is also significant at the 99% level for $\sigma_\beta=\infty$ (i.e. unrestricted inclusion of the trend). Furthermore, it strikes that the size of the estimates hardly differs between the ARX(1) model and model (3). This is in line with the literature on cointegration: for inference on relations between nonstationary series, extension of a model with stationary explanatory variables hardly affects the conclusions.

Table 1. Equilibrium Temperature Change (°C) 2×CO_2

model	$\sigma_\beta=\infty$	$\sigma_\beta=0.24$	$\sigma_\beta=0.12$	$\beta=0$
ARX(1)	4.2558	3.8860	3.6245	3.4976
(1)-(2)	(1.5243)	(1.1528)	(0.7912)	(0.3222)
General	4.3952	4.1730	3.8343	3.3701
(3)	(1.4505)	(1.3073)	(1.0521)	(0.2855)

From this point of view, Table 1 may be considered to represent the extreme cases of a wide range of models. The ARX(1) model represents abstinence of modelling details, model (3) belief in detailed explanation. A balanced statement on the impact of the enhanced greenhouse effect, discounting for 'data mining' in model (3), would end up somewhere between these two. Likewise, moving from the left to the right of the table represents decreasing belief in the long-term natural variability, with a balanced judgement somewhere in the middle. So, we feel confident in saying that a doubling of the concentration of atmospheric greenhouse gases will imply a rise of the global mean surface air temperature between 2.0 and 5.6°C (3.8±2·0.9). Only new surprising data or alternative theories of the observed warming will be able to fundamentally alter our conclusions.

We are well aware of the many wrong and misleading results obtained by this type of analysis. But there are good examples as well. For a number of reasons we have faith that we belong to the latter category: (*i*) the effect of the global mean temperature on the atmospheric concentration of carbon dioxide is dominated by human intercourse; (*ii*) the influence of changing concentrations of greenhouse gases on the climate is theoretically formulated before the main empirical evidence occurred; (*iii*) we tried many more models than are reported; none of them gave conflicting answers; (*iv*) we applied a battery of tests and the model survived them.

A remaining problem is the interpretation of our coefficient. It would be incorrect to identify it with the 'causal' effect of CO_2 on *GMT*. Other greenhouse gases and other anthropogenic influences on the climate have increased together with CO_2. It is difficult to separate these effects statistically, as the increases largely run parallel. As long as this remains so, the conclusion remains valid that when carbon dioxide and other greenhouse gases double, temperature rises as predicted. The causal effect of CO_2 — the expected rise in temperature when atmospheric CO_2 doubles while other greenhouse gases remain stable — is smaller. About 2/3 of the enhanced greenhouse effect is considered to be due to CO_2, the causal coefficient thus becomes 2.6°C, about the same conclusion as that of the climatologists. However, our standard error (0.9°C for the total effect) is considerably smaller than theirs (about 1.5°C for the causal effect only). We think this is the result of proper modelling.

In conclusion, we have tested the relationship between the increase in atmospheric greenhouse gases and global warming with a sophisticatedly simple and well-validated model which enables efficient statistical testing, incorporating the influence of the long-term natural variability. The hypothesis that there is no influence is rejected. We have not found a proof but can describe the phenomenon. We confirmed the climatological theory in sign and in size; the significance we obtained is higher. Thus, we have enlarged the empirical content of this hypothesis and reduced the uncertainties.

References

[1] Kerr, R.A., *Science*, **256**, 1138-1140 (1992).

[2] Houghton, J.T., Jenkins, G.J. & Ephraums, J.J. (eds.), *Climate Change — The IPCC Scientific Assessment*, Cambridge: Cambridge University Press (1990).

[3] Tennekes, H., *Weather*, **45**, 67-68 (1990).

[4] Freitas, C.R. de, *Area*, **23**, 11-18 (1991).

[5] Tol, R.S.J. & Vos, A.F. de, *Theo. Appl. Clim.* (forthcoming 1994).

[6] Sims, C.A., *Ectra*, **48**, 1-48 (1980).

[7] Zellner, A., *J.Ectrics*, **37**, 27-50 (1988).

[8] Schönwiese, C.-D., *Theo.Appl.Clim.*, **44**, 243-245 (1991).

[9] Schönwiese, C.-D. & Stähler, U., *Clim. Dyn.*, **6**, 23-33 (1991).

[10] Gilliland, R.L., *Clim. Change*, **4**, 111-131 (1982).

[11] Kuo, C., Lindberg, C. & Thompson, D.J., *Nature*, **343**, 709-714 (1990).

[12] Richards, G.R., *J.Climate*, **6**, 546-559 (1993).

[13] UNEP, *Environmental Data Report 1989/90 (2nd ed.)*, Oxford: Blackwell (1990).

[14] Engle, R.F. & Granger, C.W.J., *Ectra*, **55**, 661-692 (1987).

[15] Hendry, D.F., *Econometrics — Alchemy or Science?*, Oxford: Blackwell (1993).

[16] Galbraith, J.W. & Green, C., *Clim. Change*, **22**, 209-221 (1992).

[17] Sims, C.S. & Uhlig, H., *Ectra*, **59**, 1591-1601 (1991).

[18] Dansgaard, W. *et al.*, *Nature*, **364**, 218-220 (1993).

[19] GRIP Members, *Nature*, **364**, 203-207 (1993).

[20] Lehman, S., *Nature*, **365**, 109-110 (1993).

[21] Theil, H. & Goldberger, A.G., *Int. Econ. Rev.*, **2**, 65-78 (1961).

[22] Almon, S., *Ectra*, **30**, 407-423 (1962).

[23] Lamb, H.H., *Climate: Present, Past and Future, vol. 2*, London: Methuen (1977).

[24] Waldmeier, M. *The Sunspot Activity in the Years 1610-1960*, Zürich: Schulthess (1961).

[25] Jones, P.D., *Clim. Mon.*, **17**, 80-89 (1988).
[26] Jarque, C.M. & Bera, A.K., *Econ. Lett.*, **6**, 225-259 (1980).
[27] Ljung, G.M. & Box, G.E.P., *Biometrika*, **65**, 297-303 (1978).
[28] Harvey, A.C. *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge: Cambridge University Press (1989).
[29] McLeod, A.I. & Li, W.K., *J. Time Series Anal.*, **4**, 269-273 (1983).
[30] Ramsey, G.R., *J.Roy.Stat.Soc*, **B31**, 350-371 (1969).

Acknowledgements

The authors hereby express their gratitude to numerous collegues and attendants at seminars for discussions and comments, particularly Hans Coops, Richard Klein, Günther Können and Pier Vellinga, and to dr P.D. Jones (CRU, East Anglia) for kindly providing the data.