26/2003

Raport Badawczy

Research Report

RB/31/2003

A method for modelling a non-stationary system. Greenhouse gas emission case

Z. Nahorski, W. Jęda, M. Jonas

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 8373578

fax: (+48) (22) 8372772

Kierownik Pracowni zgłaszający pracę:
Prof. dr hab. inż. Zbigniew Nahorski

Warszawa 2003

A method for modelling of a non-stationary system. Greenhouse gas emission case

Zbigniew Nahorski[1], Waldemar Jęda[1], Matthias Jonas[2]
[1]Systems Research Institute, Polish Academy of Sciences
[2]International Institute of Applied System Analysis, Laxenburg, Austria

Abstract

The parties that signed the Annex I to the Kyoto Protocol must report their emissions of the greenhouse gases each year. Uncertainty of reported values obtained by aggregation of the partial emissions from all processes and provided so far for several countries is very high. Independent calculation of the estimates could confirm or question the values obtained up to now. The aim of this paper is to investigate procedures for doing it. They are statistical procedures, which benefit from the yearly reported observations while assuming temporal smoothness of the emission curve. Another goal is to provide estimates of the real emissions, which are more accurate than the observed values and could be used in testing fulfilment of the Kyoto obligations.
We consider two methods for empirical estimation of the standard deviation of the accounting errors. For this we use signal processing methods: a smoothing procedure based on the spline functions and a parametric model with a time-varying parameter. They are verified on historical observations of the greenhouse gas emissions from combustion of the fossil fuels.
The results obtained are promising and give estimates of variances that agree well with those calculated by the aggregation of variances from the partial processes. A simple piecewise exponential model was found to fit well the data in the periods of steady growth or decay.

Keywords: Kyoto Protocol, uncertainty, smoothing splines, variable parameter model

1. Introduction

The Kyoto Protocol contains obligations to decrease the emission of the greenhouse gases of 5.2% below 1990 level by the first period up to 2008-2012. However, the uncertainty assessments performed are big and in most times exceed, sometimes very considerably, the reductions agreed upon in the Annex I to the Protocol, see the comparison of the uncertainty assessments in Nahorski at al. (2003).

The parties who signed the Annex I have to monitor their emissions starting from the base year, which is mainly 1990. This redundancy in observations can be favourably used to reduce the uncertainty of individual observations in the commitment period 2008-2012, using statistical inference. In this paper we propose a method for estimating emissions and their variances using the smoothing splines (Wahba, 1990; Gu, 2002). Further analysis of obtained results extends the method to the parametric model with a time variable coefficient, which finally turns out to be well fitted as a stepwise one.

The spline functions are the piecewise polynomial functions, which are smooth enough at the joint points. They were invented to interpolate a function known only in a set of finite arguments. To approximate a function known in errors, the smoothing splines are applied. They need not to go through the observed points, better estimating the real function. Within this methodology it is also possible to estimate the variance of errors.

Our application of the smoothing splines aims in getting better estimates of the greenhouse gas emissions reported by countries and to reduce their uncertainties. For this, we assume that the emission, as a function of time, is regular enough, permitting good approximation by the smoothing splines.

Other methods could be tried to solve the problem. A simple alternative would be to use another smoothing method. Those based on the wavelets might be promising ones (Debnath, 2002; Walter, 1994). Popular methods in the automatic control literature use the parametric models with calculation of the state errors following an earlier phase of the parameter estimation. In some of them, like in the extended Kalman filter, the parameters and the states are estimated simultaneously. Similar results can be obtained using the method of Cox & Bryson (1980) where the control theory approach is explicitly applied. To use this kind of methods, the parametric model is needed. Apart of that, at least some of them require quite long data samples to converge.

Although our approach starts with a purely nonparametric model, Sec. 3, in the consequent steps we move to the analysis of parametric models with a time varying parameter, Sec. 4. Our final conclusion, presented in Sec. 4.4, is that this parameter may be actually taken as a constant one in shorter time periods, at least for the data from the fossil fuel combustion, used in the paper. This allowed us to use the simple regression method to estimate the parameter in each period. This approach has several advantages for short sequences. The difficulty lies, however, in predicting the points where the parameter jumps from one value to another.

2. Notation used

Let us introduce some basic notation. By $x(t)$, as a function of time, we denote the integral of the real emission calculated on the interval $(t-1, t]$ where t is expressed in years. Thus, the integral is calculated over the one-year-back period. So, if the instantaneous emission is called $em(t)$, then $x(t) = \int_{t-1}^{t} em(\tau)d\tau$. In the sequel we call $x(t)$ the emission. The integer value of t is assigned to the end of the year. The emission in the basic year t_0 will be denoted $x(t_0) = x_0$. The important commitment period years, 2008–2010, will be denoted T_i, where i = 8–10. The emission reports provided by the Annex I Parties are prepared by summing up partial emissions of all involved activities during a year. Due to uncertainties in assessing the exact quantities of all emissions, they are in errors. We denote the observed (reported) values $y(t_i)$ or shortly y_i. The index i begins here at 0 and takes the consecutive integer values. Therefore $T_8 \neq t_8$, and similar for other T_i's. The real emissions are unknown and can be only estimated. Hats will mark the estimated values, so $\hat{x}(t)$ is the estimated emission.

By δ we denote the fraction of the emission to be reduced within the Kyoto obligations until the commitment period. Thus at the commitment period the emission should be not greater than $(1-\delta)x_0$. Obviously, the percentage reduction required by the Kyoto protocol is 100δ but we often refer directly to δ in percents. The value of δ is not greater than few percents.

As it is common to express obligations in percents, it is useful to work not with the straight observations but with their logarithms. Let us denote $\hat{X}(T_i) = \ln \hat{x}(T_i)/\hat{x}_0$, thus $\hat{X}(T_i)$ is the logarithm of the normalized emission. As in our case $\hat{x}(T_i)/\hat{x}_0$ is close to 1, then it approximately holds

$$\hat{X}(T_i) = \ln\frac{\hat{x}(T_i)}{\hat{x}_0} \cong \frac{\hat{x}(T_i)}{\hat{x}_0} - 1 = \frac{\hat{x}(T_i) - \hat{x}_0}{\hat{x}_0} \qquad (1)$$

Thus, $\hat{X}(T_i)$ may be interpreted as the relative change of $\hat{x}(T_i)$ with respect to $\hat{x}_0$ and may be expressed in percents.

The logarithmic values will additionally prove useful in the sequel when we discuss the evolution of the emission curves in time. Our calculations suggest that a first approximation to them is a piecewise exponential curve. Working with the logarithms we get this way a piecewise linear functions, easy for any algebraic manipulations.

3. Estimating the uncertainty parameters by a nonparametric method

3.1 Basic assumptions and simplifications

3.1.1 Data treatment

The function $x(t)$, as an integral of a positive function, is continuous and positive. We have $dx(t)/dt = em(t) - em(t-1)$. If we take a reasonable assumption that *em* is a continuous function of t, then $x(t)$ is even continuously differentiable. We assume more, namely, that $x(t)$ is a twice continuously differentiable function.

We assume that the emission process can be only observed with errors in equally spaced time intervals $\Delta t = 1$ year. We introduce a simplified notation $x(t_i = i\Delta t) \equiv x_i$, $x_i > 0$.

4.1.2 Uncertainty treatment

We assume that the real process x_i is observed with multiplicative errors $\varepsilon_i = u_i x_i$, where

$$E(u_i) = m,$$
$$E[(u_i - m)^2] = \sigma_i^2,$$
$$\mathrm{cov}(u_i, u_j) = \rho_{ij}$$

Thus, the observations can be presented in the following way

$$y_i = x_i + u_i x_i = (1 + u_i)x_i \ , \ i = 0,1,...,N \ .$$

where y_i are the observed emissions, x_i the (unknown) real emissions, and u_i their relative uncertainties.

The function $x(t)$ is generally not known. Therefore, we can also introduce the errors proportional to the observed value $\varepsilon_i = u_i y_i$. This yields

$$y_i = x_i + u_i y_i$$
$$y_i = \frac{1}{1 - u_i} x_i$$

For small u_i we have $1/(1-u_i) \cong 1 + u_i$ and both models are approximately the same. However, for higher u_i it is not so.

There is an important difference between two above models. In the former y_i is linear in u_i, which is the only stochastic variable at the right hand side. This is not the case in the latter.

The above dependencies are also true for i = 0. Dividing sides and taking the logarithms we get

$$Y_i = X_i + \ln \frac{1 + u_i}{1 + u_0}$$

or, respectively,

$$Y_i = X_i + \ln\frac{1-u_0}{1-u_i}$$

where $Y_i = \ln y_i / y_0$ and $X_i = \ln x_i / x_0$. For small u_0 and u_i approximately it holds

$$\ln\frac{1+u_i}{1+u_0} \approx u_i - u_0 \approx \ln\frac{1-u_0}{1-u_i}$$

resulting in the identical expression

$$Y_i = X_i + u_i - u_0$$

The errors $v_i = u_i - u_0$ have the zero mean, $E(v_i) = 0$, and the variance $\sigma_{v_i}^2 = \sigma_i^2 + \sigma_0^2 - 2\rho_{i0}\sigma_i\sigma_0$. Their covariance is equal to

$$\operatorname{cov}(v_i, v_j) = E[(u_i - u_0)(u_j - u_0)] = \rho_{ij} - \rho_{i0} - \rho_{j0} + \sigma_0^2$$

It is equal zero, if all summands are equal. But generally the sequence is correlated, even if the original errors u_i are not. We assume, however, that the correlation is negligibly small. As noticed by Wahba (1990, s. 4.9), correlation of errors may considerably worsen the smoothing results, as far as reconstruction of the original function is considered.

3.2. Smoothing and uncertainty analysis

3.2.1 Smoothing splines

Let us consider some abstract data z_i generated by the following system

$$z_i = f(t_i) + e_i, \qquad i = 0, 1, 2, \dots, N$$

The vector

$$\mathbf{e} = (e_0, \dots, e_N) \propto \mathrm{N}(0, \sigma^2\mathbf{I})$$

contains the set of observation errors. We want to recover the function $f(t)$, assumed to be smooth enough, knowing only the erroneous observations z_i, $i = 0, 1, \dots, N$. For this we use splines.

In the interpolating splines an approximation $\hat{z}(t)$ to $f(t)$ is obtained assuming that $\hat{z}(t)$ is a polynomial of an order m (we use $m = 3$) on each segment $[t_i, t_{i+1})$, $i = 0,1,2,\dots,N-1$, satisfying $\hat{z}(t) = z_i$ and having the continuous derivatives up to the order $m-1$ on the whole interval (t_1, t_N). In the presence of noise the interpolating spline generally quickly varies in time, overshooting and undershooting considerably the function $f(t)$.

Much better approximation can be then achieved using the smoothing splines. Their idea is to find the function $\hat{z}(t)$ that does not need to go directly through the observed points z_i, in order to get a function with smaller $(m-1)$th derivative, see Fig. 1.

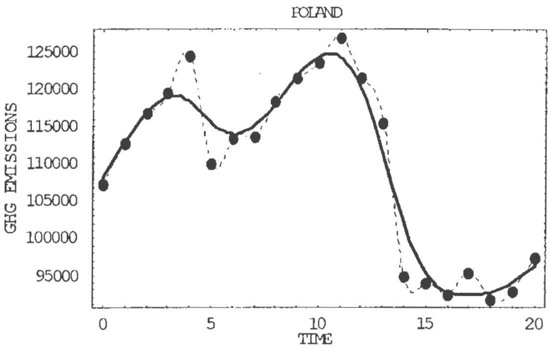


Figure 1. The interpolating spline (dashed curve) and the smoothing spline (solid curve).

If we restrict our attention to the third order polynomials, then the task is to find a smooth function $\hat{z}(t)$, which minimises the sum

$$\frac{1}{N+1}\sum_{i=0}^{N}(z_i - \hat{z}(t_i))^2 + \lambda\int_{t_0}^{t_N}\left(\hat{z}^{(2)}(t)\right)^2 dt \tag{2}$$

where

$$\hat{z}(t) = a_i + b_i(t-t_i) + c_i(t-t_i)^2 + d_i(t-t_i)^3$$
$$t \in [t_i, t_{i+1}), \qquad i = 0,1,\ldots,N-1$$

The solution of the problem, for a given λ, is delivered e. g. in Wahba (1990) and can be written in a general form

$$\hat{z}(t_i) = a_i = \sum_j A_{ij}(N,\lambda)z_j \qquad \frac{d\hat{z}(t_i)}{dt} = b_i = \sum_k B_{ik}(N,\lambda)z_k \tag{3}$$

see also Gu (2002), where A_{ij} and B_{ik} are coefficients which do not depend on the data z_i and can be precomputed. Thus, both $\hat{z}(t_i)$ and $d\hat{z}(t_i)/dt$ are kernel estimators.

3.2.2 Uncertainty analysis

The solution depends on the value of λ. This value is estimated by the generalized cross validation method (Wahba, 1990) by minimising over λ the criterion

$$V(N,\lambda) = \frac{\sum_{i=0}^{N}[z_i - \hat{z}_i(N,\lambda)]^2}{N+1-\sum_{i=0}^{N}A_{ii}(N,\lambda)} \tag{4}$$

where $\hat{z}_i(N,\lambda)$ is the solution of the problem (2), in which the $\hat{z}_i$ observation is dropped. The optimal value will be denoted $\hat{\lambda}$. The optimal value of the criterion can be used as an estimate of σ^2, i.e.

$$\hat{\sigma}^2(N) = V(N,\hat{\lambda}) \tag{5}$$

The expression in the denominator of (4) can be interpreted as for the degrees of freedom of the noise, in analogy to the degrees of freedom in the regression analysis. However, in contrast to the regression analysis, only consistency of the estimate $\hat{\sigma}^2(N)$ for the smoothing splines has been proved theoretically (Gu, 2002, th.3.4), while other good statistical properties have been checked on numerical simulations.

The estimated variance of $\hat{z}(t_i)$ is now

$$\hat{\sigma}_{\hat{z}}^2(N,t_i) = \hat{\sigma}^2(N)A_{ii}(N,\hat{\lambda}) \tag{6}$$

Table 1. Optimal values of λ and estimated standard deviations of observation errors for different countries and two time periods

Years 1950-1998

Country	$\hat{\lambda}$	Std. Dev. "edge" [%]	Std. Dev. "middle" [%]
ARGENTINA	0,06	2,7	2,3
AUSTRALIA	0,06	2,1	1,8
AUSTRIA	0,15	3,4	2,7
BELGIUM	0,07	2,8	2,3
BRAZIL	0,31	2,6	1,9
CANADA	0,03	2,1	1,9
CHINA	0,03	5,2	4,7
CUBA	0,16	8,4	6,6
EGYPT	1,16	5,0	3,4
FINLAND	0,03	5,3	4,8
FRANCE	0,14	2,9	2,3
GREECE	0,14	3,5	2,8
ICELAND	1,64	5,3	3,5
IRELAND	0,11	5,4	4,3
ISRAEL	0,03	3,8	3,4
ITALY	0,10	2,0	1,6
JAPAN	0,01	2,8	2,7
LUXEMBOURG	0,05	3,3	2,9
MEXICO	0,76	2,5	1,7
NETHERLANDS	0,08	3,4	2,8
NEW ZEALAND	5,11	2,8	1,8
NORWAY	3,44	6,6	4,2
POLAND	0,71	2,1	1,5
PORTUGAL	3,34	3,0	1,9
ROMANIA	0,19	2,5	1,9
SPAIN	0,03	3,4	3,0
SWEDEN	3,69	4,0	2,5
SWITZERLAND	0,11	4,1	3,3
TURKEY	0,11	3,9	3,1
UNITED KINGDOM	0,15	2,0	1,6
USA	0,02	1,8	1,6

Years 1970-1998

Country	$\hat{\lambda}$	Std. Dev. "edge" [%]	Std. Dev. "middle" [%]
ARGENTINA	3,2E-10	0,4	0,4
AUSTRALÍA	7,4E+05	1,5	0,9
AUSTRIA	2,3E+05	2,0	1,1
BELGIUM	1,2E-01	2,8	2,3
BRAZIL	1,9E-01	1,7	1,3
CANADA	3,2E-10	0,5	0,5
CHINA	1,3E-01	1,7	1,4
CUBA	4,3E+05	3,3	1,9
EGYPT	3,2E-01	3,5	2,6
FINLAND	3,5E-02	4,3	3,8
FRANCE	5,5E-01	3,2	2,3
GREECE	1,6E-01	2,8	2,2
ICELAND	1,7E+05	4,1	2,7
IRELAND	3,0E-01	3,0	2,2
ISRAEL	2,2E-01	2,6	2,0
ITALY	8,6E-01	1,9	1,3
JAPAN	6,7E-02	2,1	1,8
LUXEMBOURG	1,3E-01	3,5	2,8
MEXICO	2,2E+05	2,6	1,7
NETHERLANDS	4,1E-02	4,2	3,7
NEW ZEALAND	5,3E-02	3,4	2,9
NORWAY	3,5E+05	8,3	5,2
POLAND	6,7E-01	2,5	1,8
PORTUGAL	4,0E+05	3,1	1,9
ROMANIA	3,2E-01	2,9	2,1
SPAIN	8,2E-01	2,5	1,7
SWEDEN	3,6E+05	3,7	2,3
SWITZERLAND	1,1E+05	2,8	1,9
TURKEY	1,6E-02	3,7	3,4
UNITED KINGDOM	1,1E+05	2,0	1,4
USA	8,7E-05	0,4	0,4

3.2.3 Application to real data

The above analysis was applied for smoothing the data $Y_i = \ln y_i / y_0$ to obtain the smoothed values. Equation (6) has been used to calculate the estimates of the standard deviations $\hat{\sigma}_{\xi}(N,t_i)$ for the emission from the fossil fuels provided by Marland et al. (1999), in the periods 1950-1998 and 1970-1998. As the values $A_{ii}(N,\hat{\lambda})$ do not change significantly in i except for time points at the beginning and the end of the data, see Table 1, where the values at the "edge" and in the "middle" of the data period are compared. The value $\hat{\sigma}_{\xi}(N,t_i)$ depends also on the number of data used. This dependence is visible, although mostly not crucial, in the results presented in Table 1 for different time periods. For few cases, like e.g. Argentine, Canada, USA, reduction of the number of data caused big drop of the standard deviation value.

Wahba (1990, sec. 4.9) recommends using at least 25-30 observations when applied the smoothing splines. The data used in calculating the values in the left side of Table 1 contained 29 points, just satisfying the recommendations. However, for many countries, the corresponding standard deviations differ for different length of data. At least in some cases this is correlated with extreme values of λ, either very close to zero, like for Argentina, Canada and USA, or very high, like for Austria and Cuba. This phenomenon is also mentioned by Wahba (1990, a. 4.9). This may suggest that the data in the shorter sequence are too short.

The estimated values agree quite well in magnitudes with the common idea of the errors made in calculation of the fossil fuel emission, believed to be of few percents. A little bigger figures obtained in some of our calculations may be connected with some additional factors that might have influenced the calculated estimates, as year-to-year variations in the weather conditions or variations due to change in economic factors of the countries.

4. Empirical parametric models for the net emission data

In the previous section we noticed that the consecutive values in the emission sequence might be correlated. To better model this property, in this section we consider a set of values x_i which can be considered as a time series consisting of N elements. We introduce both a difference model and a differential model to describe the time evolution of the data. Then we motivate the choice of the model and finally present some results for fitting the model to the emission data for some countries.

4.1. Difference model

As we assumed that x_i are positive we can define a new time series

$$g_i := \frac{x_{i+1}}{x_i} - 1 = \frac{x_{i+1} - x_i}{x_i} \quad , \quad i = 0, \ldots, \quad N-1$$

Each element g_i of a new time series can be interpreted as a relative difference of the two consecutive elements x_{i+1} and x_i.

From the latter relation we can now formulate the following difference equation

$$x_{i+1} - x_i = g_i \, x_i, \qquad x_0 = x(t_0) \tag{7}$$

which can be then easily solved giving

$$x_i = x_0 \prod_{j=0}^{i-1} (1+g_j)$$

As all x_i are positive we can convert this solution to an additive form

$$X_i = \sum_{j=0}^{i-1} \ln(1+g_j) \cong \sum_{j=0}^{i-1} g_j \tag{8}$$

where we inserted the variable $X_i = \ln x_i / x_0$.

As we have $y_i = (1+u_i)x_i$, then (7) can be transformed to

$$y_{i+1} = (1+g_i)\frac{1+u_{i+1}}{1+u_i} y_i$$

Dividing both sides by y_0 and taking logarithms yields

$$Y_{i+1} = \ln(1+g_i) + \ln\frac{1+u_{i+1}}{1+u_i} + Y_i$$

or approximately

$$Y_{i+1} - Y_i \approx g_i + u_{i+1} - u_i$$

from where the estimator $\hat{g}_i$ can be designed as

$$\hat{g}_i = Y_{i+1} - Y_i \tag{9}$$

Under our assumption on u_i's we have

$$E(\hat{g}_i) = E(Y_{i+1} - Y_i + u_i - u_{i+1}) = X_{i+1} - X_i = \ln(1+g_i) \approx g_i$$

Thus the estimator is unbiased (up to the approximation done). Its variance is

$$\text{var}(\hat{g}_i) = E(Y_{i+1} - X_{i+1} - Y_i + X_i)^2 =$$
$$= E(u_{i+1} - u_0 - u_i + u_0)^2 = E(u_{i+1} - u_i)^2 = \sigma_{i+1}^2 - 2\gamma_{i,i+1} + \sigma_i^2 .$$

4.2. Differential model

A similar expression can be provided directly for the function $x(t)$. Starting with the equation

$$\frac{dx(t)}{dt} = g(t)x(t), \qquad x(t_0) = x_0 \tag{10}$$

we obtain the solution for $x(t_i)$ which depends on the function $g(t)$

$$x(t_i) = x_0 \exp\left(\int_{t_0}^{t_i} g(\tau)d\tau \right)$$

This provides us with the formula

$$\ln\frac{x(t_i)}{x_0} = \int_{t_0}^{t_i} g(\tau)d\tau \cong \sum_{j=0}^{i-1} g(t_j)$$

Denoting $X(t_i) = \ln\frac{x(t_i)}{x_0}$ we get

$$X(t_i) = \int_{t_0}^{t_i} g(\tau)d\tau \cong \sum_{j=0}^{i-1} g(t_j)$$

This is approximately equivalent to (8).

The actual solution of the problem relies, as it was already mentioned, on the function g. From (10) this function can be written in the form

$$g(t)=\frac{1}{\frac{x(t)}{x_0}}\frac{d\frac{x(t)}{x_0}}{dt}=\frac{d}{dt}\ln\frac{x(t)}{x_0}=\frac{dX(t)}{dt} \tag{11}$$

From this expression we find that the function g can be conceived as the rate of change of the variable $X(t)$.

The estimate of $g(t)$ can be then calculated as

$$\hat{g}(t)=\frac{d\hat{X}(t)}{dt} \tag{12}$$

where $d\hat{X}(t)/dt$ is obtained from the smoothing spline approximation. This can be also expressed in the form

$$\hat{X}(t)=\int_{t_0}^{t_i}\hat{g}(\tau)d\tau \tag{13}$$

Estimated value of the signal in the year t_N, after modelling by the smoothing splines, is given by

$$\frac{\hat{x}(t_N)}{\hat{x}_0}=\exp\left(\hat{X}(t_N)\right)=\exp a_N$$

or

$$\frac{\hat{x}(t_N)}{\hat{x}_0}=\exp\left(\int_{t_0}^{t_N}\frac{d\hat{X}(\tau)}{d\tau}d\tau\right)=\exp\left(\int_{t_0}^{t_N}\hat{g}(\tau)d\tau\right)\cong\exp\sum_{i=0}^{N-1}b_i \tag{14}$$

4.3 Estimation of the parameter $g(t)$

Both expressions (9) and (12) were used to estimate the function $g(t)$ for few countries from the previously mentioned data of CO_2 emission from the fossil fuels published by Marland et al. (1999). The results are presented in Appendix A. The smoothing splines were used both to smooth the points obtained from (9) and to calculate the derivatives in (12), using the formulae (3). For each country, in the left panel the observations (dots) and their smoothing spline approximations (solid lines) are depicted. The right panel shows the estimates of the function $g(t)$. The dots represent the points calculated using the formula (9). The bold dashed line is obtained by smoothing these points. The solid line is calculated from (12). The normal thickness dashed lines on both panels show the 95% confidence intervals of the estimates.

Comparison of the estimates reveals that the curves obtained from smoothing are very similar in many cases. However, there is also a number of case where they differ, in a way that those from (9) are more smooth than those from (12). Thus, the method using the formula (12) is more sensitive to variations in data and more often shows the oscillations of the value $g(t)$. The method using (9) shows more often slow trends in the estimate, neglecting its quicker variations. However, it is more sensitive to the choice of the time interval for smoothing. A practical observation is that when smoothing of the points obtained from (9) there were less of pathological values of $\hat{\lambda}$, as happened in smoothing the points Y_i.

Table 2 depicts the values $\hat{\lambda}$ and estimates of the standard deviation of the errors u_{i+1}-u_i, both for the "edge" and the "middle" points. From the comparison with the values presented in Table 1 it can be seem that both estimates of the standard deviations are of the same order, although not always very close to each other. Notice, however, that Table 1 shows the standard deviations of the errors u_i-u_0, which might cause the differences. The values $\hat{\lambda}$ in Table 2 are bigger and often much than those in Table 1. This is related with bigger smoothing of the method connected with the formula (9).

Table 2. Optimal values of λ in smoothing estimates $\hat{g}_i$ for different countries in two time periods

Years: 1950 - 1998

Country	$\hat{\lambda}$	Std. Dev. "edge" [%]	Std. Dev. "middle" [%]
ARGENTINA	3,7E+10	1,4	0,7
AUSTRALIA	5,6E+06	0,9	0,5
AUSTRIA	3,3E+05	1,7	0,9
BELGIUM	2,7E-01	4,4	3,3
BRAZIL	4,8E+03	2,0	1,1
CANADA	1,8E+03	1,6	0,8
CHINA	1,1E-01	8,8	7,1
CUBA	3,7E+10	4,3	2,2
EGYPT	3,7E+10	2,7	1,4
FINLAND	3,7E+10	2,6	1,3
FRANCE	1,0E+00	4,4	3,0
GREECE	1,3E+05	1,7	0,9
ICELAND	3,7E+10	2,7	1,4
IRELAND	3,7E+10	2,4	1,2
ISRAEL	6,9E+01	4,0	2,2
ITALY	3,9E-01	3,1	2,3
JAPAN	1,3E-02	5,1	4,8
LUXEMBOURG	2,0E-01	5,5	4,3
MEXICO	9,0E+00	3,5	2,1
NETHERLANDS	3,3E+05	1,8	0,9
NEW ZEALAND	3,7E+10	1,6	0,8
NORWAY	3,7E+10	3,8	2,0
POLAND	7,6E+00	2,9	1,8
PORTUGAL	3,7E+10	1,7	0,9
ROMANIA	2,0E+00	3,7	2,4
SPAIN	4,3E+03	2,3	1,2
SWEDEN	2,5E+05	2,2	1,1
SWITZERLAND	6,4E-01	6,2	4,3
TURKEY	3,7E+10	1,8	0,9
UNITED KINGDOM	3,7E+10	1,0	0,5
USA	6,8E+04	0,9	0,5

Years: 1970 - 1998

Country	$\hat{\lambda}$	Std. Dev. "edge" [%]	Std. Dev. "middle" [%]
ARGENTINA	1,2E-10	0,1	0,1
AUSTRALIA	1,3E+10	1,1	0,5
AUSTRIA	1,3E+10	2,0	1,0
BELGIUM	4,6E-01	4,6	3,3
BRAZIL	1,2E+05	2,6	1,7
CANADA	1,4E-03	1,8	1,8
CHINA	1,7E+05	2,5	1,7
CUBA	1,6E+09	2,8	1,4
EGYPT	1,0E+06	2,2	1,1
FINLAND	5,5E+05	5,8	3,6
FRANCE	1,3E+10	2,1	1,1
GREECE	4,9E+05	1,7	0,9
ICELAND	1,6E+09	2,8	1,4
IRELAND	1,9E+05	3,8	2,2
ISRAEL	1,6E+09	1,7	0,9
ITALY	1,6E+09	1,3	0,7
JAPAN	5,2E-01	3,3	2,4
LUXEMBOURG	6,0E-01	5,6	4,0
MEXICO	2,8E+05	3,5	2,0
NETHERLANDS	1,6E+09	2,8	1,4
NEW ZEALAND	2,8E+05	3,6	2,1
NORWAY	1,6E+09	6,3	3,3
POLAND	6,9E+05	3,5	2,2
PORTUGAL	1,6E+09	2,4	1,2
ROMANIA	2,3E+05	4,4	2,9
SPAIN	1,9E+04	1,8	1,0
SWEDEN	1,3E+10	2,8	1,4
SWITZERLAND	1,3E+10	1,9	1,0
TURKEY	1,6E+09	2,2	1,1
UNITED KINGDOM	1,3E+10	1,4	0,7
USA	5,1E-02	2,5	2,1

4.4. Piecewise exponential model

Although the estimated functions $\hat{g}(t)$ in the previous section vary in time, in many instances their patterns resembles the constant value lines. To better investigate this question let us start with examining of few curves. Fig. 2 contains emission curves $y(t_i)$ and logarithmic curves $Y(t_i) = \ln y(t_i)/y(t_0)$ for the global emission data from Marland (1990). Similar data for Poland are depicted on Fig. 3. It can be seen that the data evolve approximately along piecewise exponential curve, and the logarithmic curves are approximately linear. Even better indication of the exponential relation can be inferred from Fig. 4. It shows the dependence of $y_i = y(t_i)$ on $y_{i-1} = y(t_{i-1})$. Referring to (7) this dependence should have the form $y_i = (1+g_i)\, y_{i-1}$. Similar relation can be inferred from the logarithmic data depicted on Fig. 5 for which it approximately holds $Y_i = \ln(1+g_i) + Y_{i-1}$

. The curves indicate that this dependence holds with g_i slightly bigger than 0.

However, looking at Fig. 3 we can easily notice periods where this simple constant parameter g (and therefore the growth along the exponential curve) does not hold. This is particularly visible in the periods of the Great Crisis of 1930s, the 2[nd] World Word, and the collapse of the communist regime.

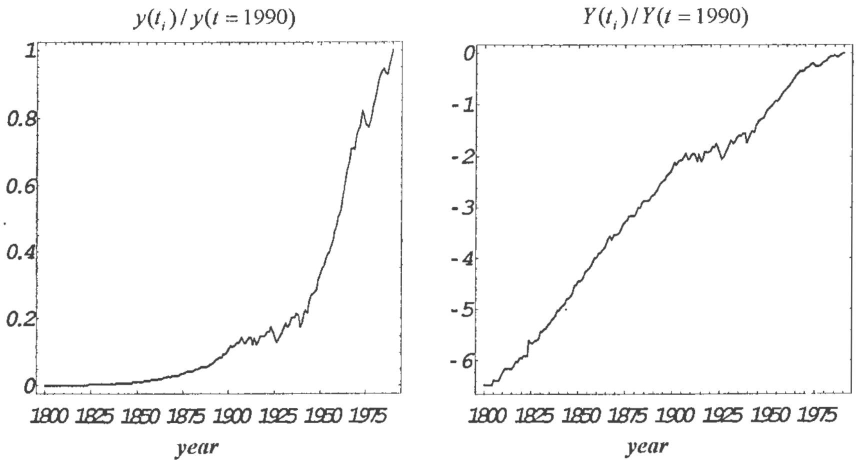


Figure 2. **Global data:** CO_2 total emissions (1800 – 1990)

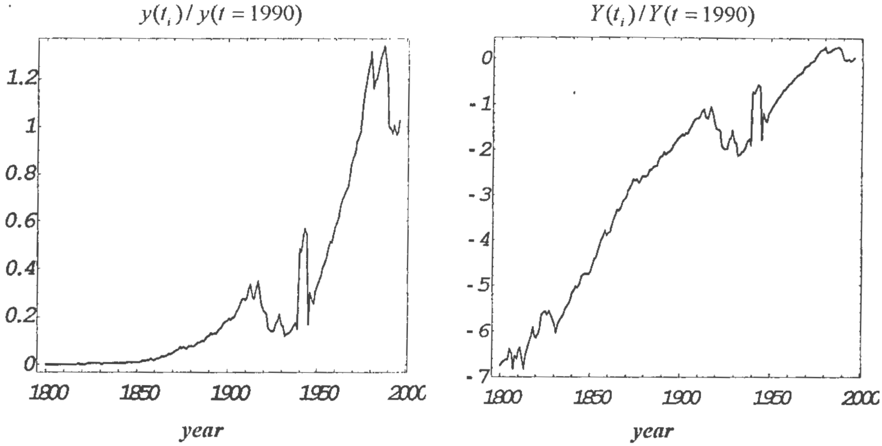
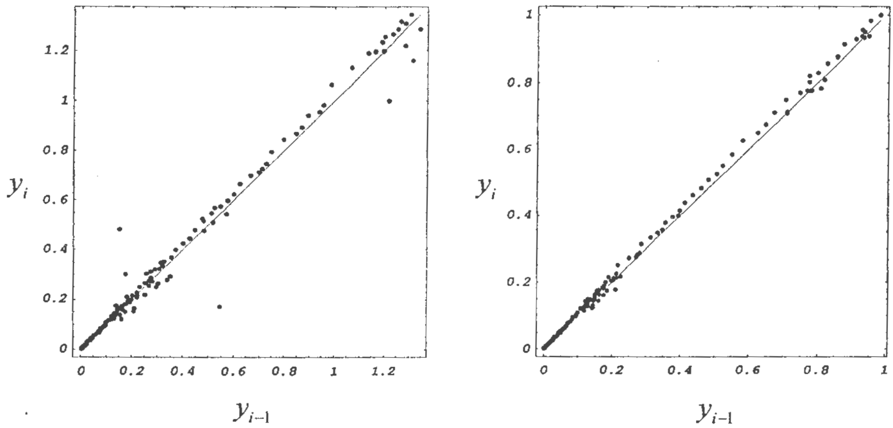


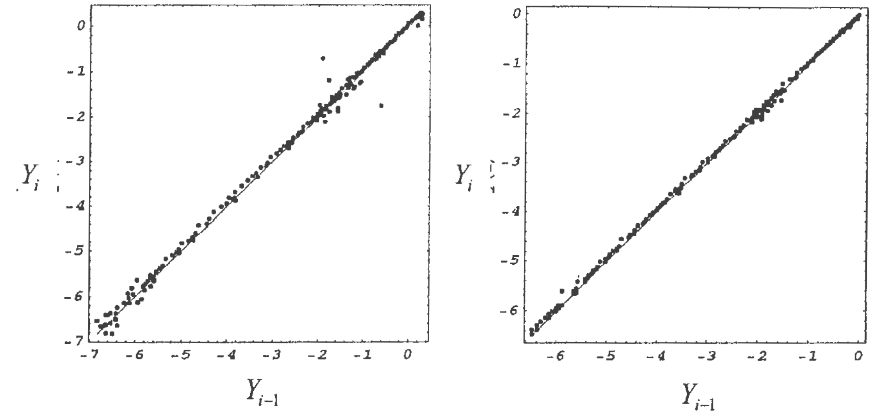
Figure 3. **Poland's data:** CO_2 emissions (1800 – 1998)

Phase plane



Poland's data　　　　Global data

Figure 4. The phase plane portraits of the data



Poland's data   Global data

Figure 5. The phase plane portraits of the logarithmic data

Thus the exponential growth models describe quite well development of data only in some intervals, as shown in the Appendix B. These intervals seem to be the periods of constant development conditions. One can easily distinguish on the figures the period of the 19^{th} century industrial revolution, the periods of the World Wars and Great Crisis of 1930^{th}, the periods of the post-war prosperity of 1950^{th}-1960^{th} and the energy shocks of 1970^{th}-1980^{th}. Also smaller ripples can be distinguished and explained, like for example in the case of the Polish transformation period.

The fit of this simple piecewise exponential model is quite good in the periods of growth or decay. In the period of steady growth it is almost perfect. In the decay periods the emission is often more volatile. War and transition periods, like those of 1970s in the West Europe or 1980s in Poland, are highly irregular and were skipped from fitting.

The results obtained are generally quite similar for both methods. The error variance estimates calculated by the regression method (parametric model) turn out to be usually small, although mostly greater than those calculated by the smoothing splines. This seems to be connected with too big simplicity of the exponential model used. The good fit of the piecewise exponential model seems to be an important observation. It means that the emissions follow approximately the exponential functions in defined longer periods. The jump from one such segment to another is mostly connected with a big political or economic change.

5. Conclusions

Nonparametric and parametric methods for modelling the greenhouse gas emission phenomena and for estimating the parameters are proposed in the paper. They differ in degree of smoothing and precision of fitting the observations. Comparison of the methods made up to now reveals that the method using the formula (9) gives more smooth curves in many instances, although it is more sensitive to the smoothing interval. The method using the formula (12) is more accurate and better emphasizes the ripples in data. The parametric piecewise exponential model gives the most rough but also most simple description, showing general trends in evolution of emission data.

One of the main goals of the paper was to estimate the standard deviation of the errors. Good estimate could help to answer a quite important Apart of confirmation of the uncertainty estimates obtained by aggregation of the errors, an important question for verification of the Kyoto obligations is how big part of the uncertainty pertains to the systematic error, which biases equally all observations. During the verification, when calculation of the difference $x_N - x_0$ is performed, the systematic parts of the errors cancel. Thus, this kind of uncertainty is perhaps less harmful for the verification results. This paper aims at elaboration of a method for estimating the stochastic part of the error. Some signal processing methods are proposed and preliminary results are presented. They are based on the published observations of the emissions from the fossil fuels (Marland at al., 1999) and therefore do not cover the whole emissions, which is reported within the Kyoto agreement. Moreover, they may be biased by the method of preparation of the data. Thus, the results are only partial and addition of the rest of data may change the estimates and conclusions. Although the estimators are consistent, nothing is known on their small sample properties. The early results show that the variability of the estimates is rather high.

The empirical approach proposed in the paper heavily relies on the assumption of the smoothness of the emission process. But no doubt there exists some volatility in observations, which may be related not only to the observation errors but also to such factors as changing weather conditions and rapidly changing economic situation of the country. These phenomena may increase the estimated variance and the variance connected with the stochastic errors may be actually smaller. And although the results obtained up to now and presented in this paper do not indicate existence of big systematic errors, a more precise answer to this question requires analysis of more data.

Under all this reservations, the calculations performed for the fossil fuels indicate that the empirical approach gives reasonable estimates, comparable to the aggregate ones. The partial results obtained here do not falsify the procedure applied up to now. However, the present knowledge does not allow us to state definite conclusions as yet.

An interesting result connected with the relation between the piecewise exponential character of the emission curve and the economic development of the country may extend to some other components of the emission. But probably this will not concern removal of the greenhouse gases by sinks, also included in the full calculation of the greenhouse gas balance of countries. Evolution of this type of data in time will be possible to analyse when longer historical records will be available.

The proposed approach can be used to better estimate the real emissions, by filtering out the errors, and possibly for prediction. The latter may be, however, a little risky until more will be known on how much the abrupt economic and weather condition changes can influence the emissions.

References

Cox A.B., Bryson Jr., A.E. (1980) Identification by a combined smoothing nonlinear programming algorithm. *Automatica*, **16**, 689-694.

Debnath L. (2002) *Wavelet Transforms and Their Applications*. Birkhäuser.

FCCC (1998). *Report of the Conference of the Parties on Its Third Session, Held at Kyoto From 1 to 11 December 1997. Addendum.* Document FCCC/CP/1997/7/Add.1. United Nations Framework Convention on Climate Change (FCCC). Available on the Internet: http://unfccc.int/index.html.

FCCC (2001). *Implementation of the Buenos Aires Plan of Action: Adoption of the Decisions Giving Effect to the Bonn Agreements. Draft Decisions Forwarded for Elaboration,* *Completion and Adoption. National Systems, Adjustments and Guidelines Under Articles 5, 7 and 8 of the Kyoto Protocol.* Document FCCC/CP/2001/L.18. United Nations Framework Convention on Climate Change (FCCC). Available on the Internet: http://www.unfccc.de/.

IPCC (2000a). *Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories.* J. Penman, D. Kruger, I. Galbally, T. Hiraishi, B. Nyenzi, S. Emmanuel, L. Buendia, R. Hoppaus, T. Martinsen, J. Meijer, K. Miwa and K. Tanabe (eds.), Intergovernmental Panel on Climate Change (IPCC) National Gas Inventories Program, Technical Support Unit, Institute for Global Environmental Strategies, Hayama, Kanagawa, Japan.

IPCC (2000b). *Land Use, Land-Use Change, and Forestry.* Special Report of the Intergovernmental Panel on Climate Change (IPCC). R.T. Watson, I.R. Noble, B. Bolin, N.H. Ravindranath, D.J. Verardo and D.J. Dokken (eds.), Cambridge University Press, Cambridge, United Kingdom.

Gu Ch. (2002) *Smoothing Spline ANOVA Models.* Springer.

Marland G., R. J. Andres, T. A. Boden, C. A. Johnston, and A. L. Brenkert (1999) *Global, regional, and national CO2 emission estimates from fossil fuel burning, cement production, and gas flaring: 1751-1996 (revised March 1999).* Carbon Dioxide Information Analysis Center, Available on the Internet http://cdiac.esd.ornl.gov/trends/emis/em_cont.htm

Nahorski Z., Jęda W., Jonas M. (2003) Coping with uncertainty in verification of the Kyoto obligations. In: Studziński J., Drelichowski L., Hryniewicz O.: *Zastosowania informatyki i analizy systemowej w zarządzaniu*, IBS PAN, Warszawa 2003, 305-317.

Nilsson, S., M. Jonas, M. Obersteiner and D. Victor (2001). Verification: The Gorilla in the Struggle to Slow Global Warming. *The Forestry Chronicle*, **77**, 475–478.

Nilsson, S., M. Jonas and M. Obersteiner (2002). COP 6: A Healing Shock. *Climatic Change,* **52**: 25–28.

Wahba G.: (1990) *Spline Models for Observational Data.* SIAM.

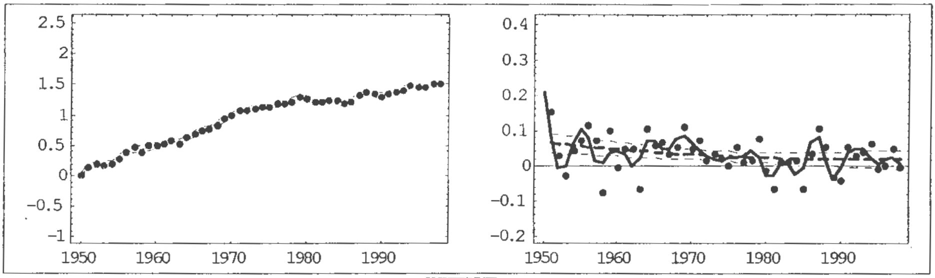
Walter G.G. (1994) *Wavelets and Other Orthogonal Systems With Applications.* CRC Press.

APPENDIX A

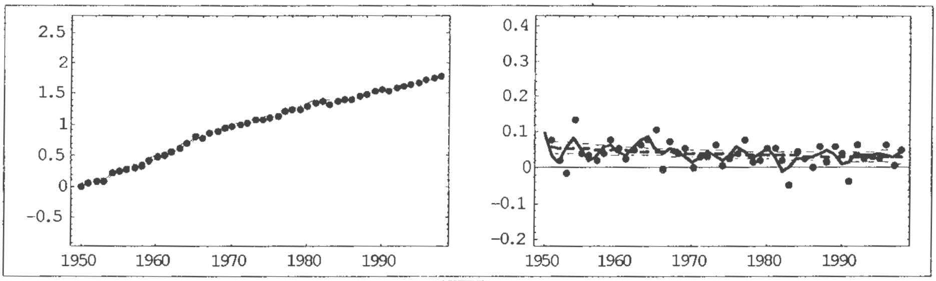
Results of smoothing and estimation of $g(t)$ for different countries.

I. Years 1950-1998

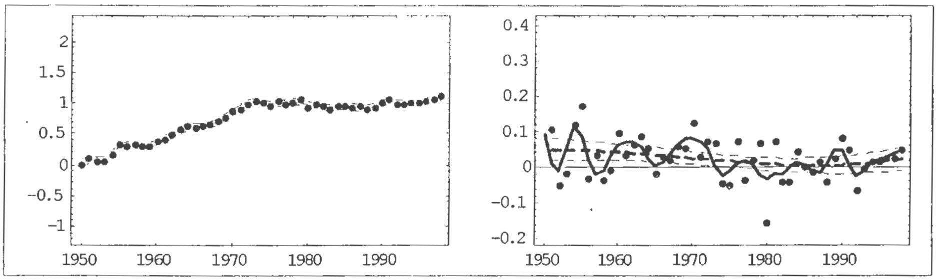
ARGENTINA



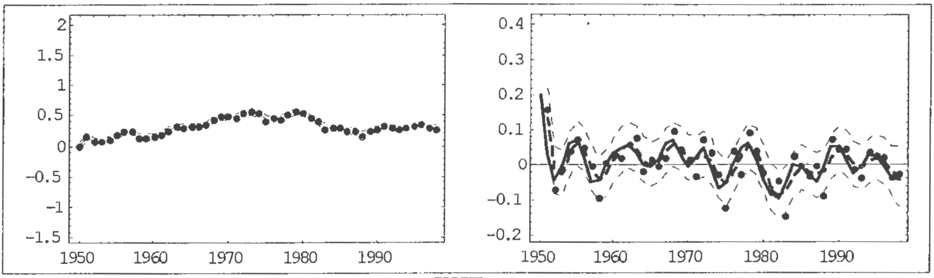
AUSTRALIA



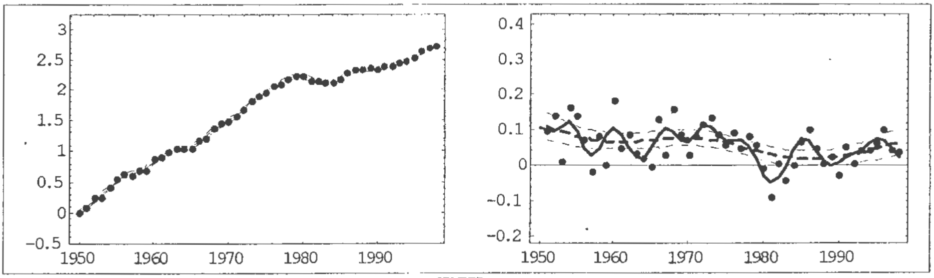
AUSTRIA



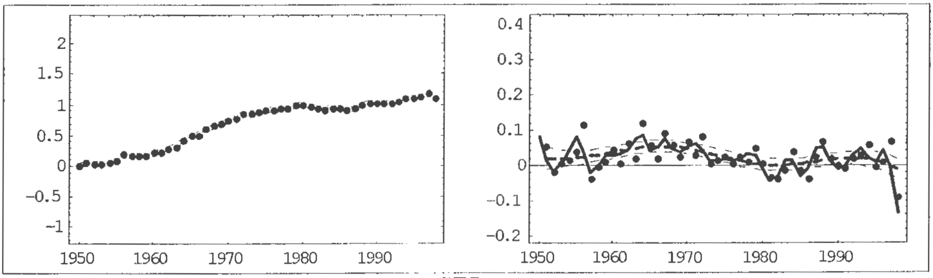
BELGIUM



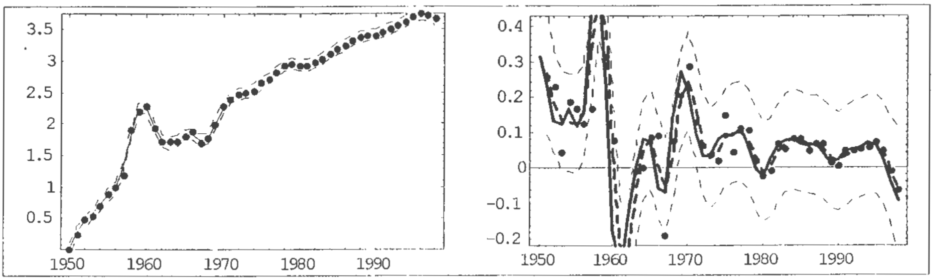
BRAZIL



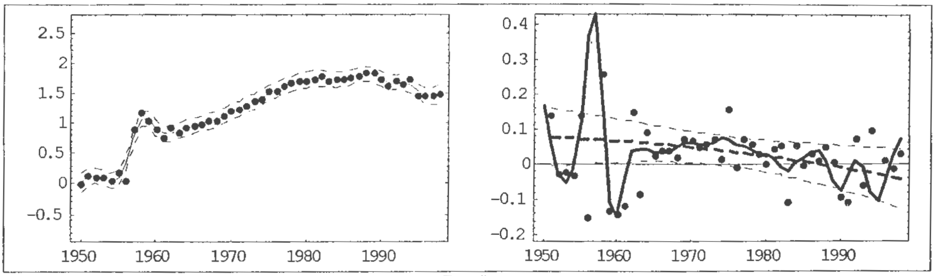
CANADA



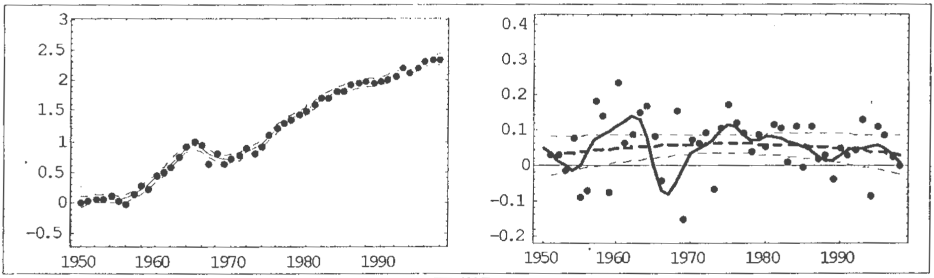
CHINA



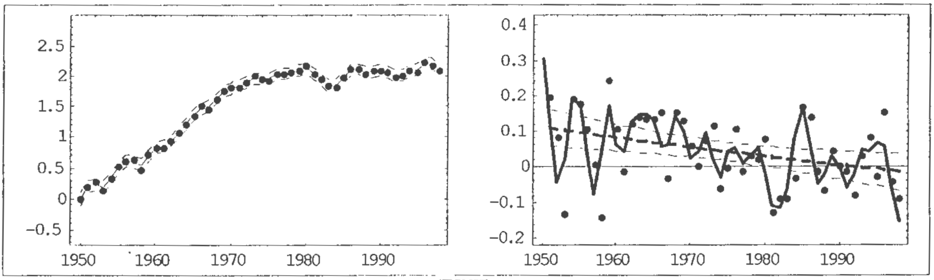
CUBA



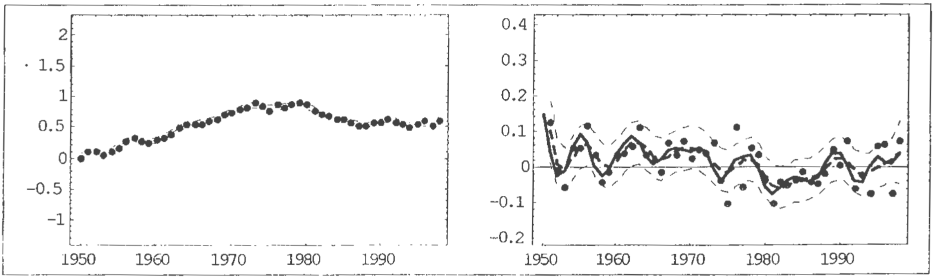
EGYPT



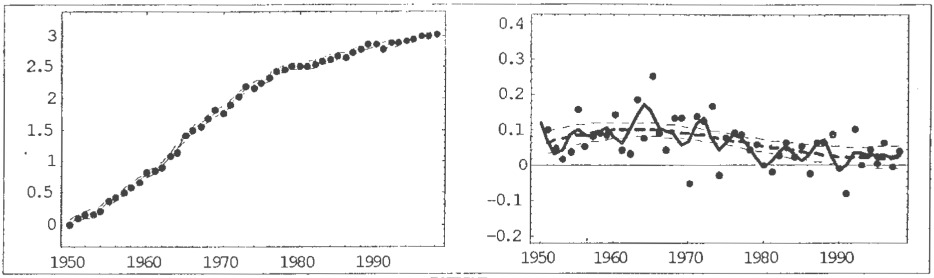
FINLAND



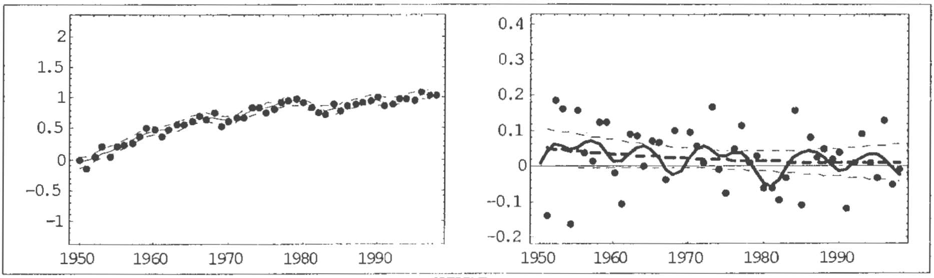
FRANCE



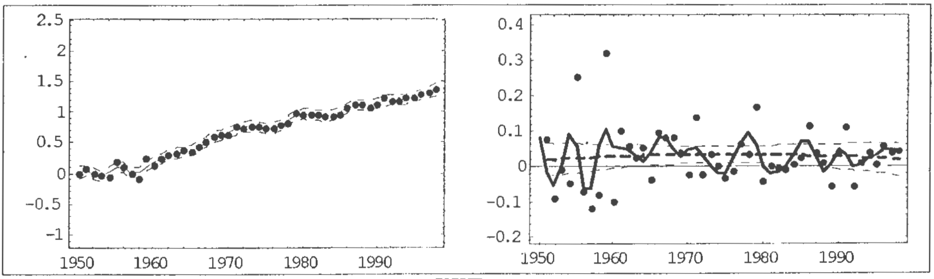
GREECE



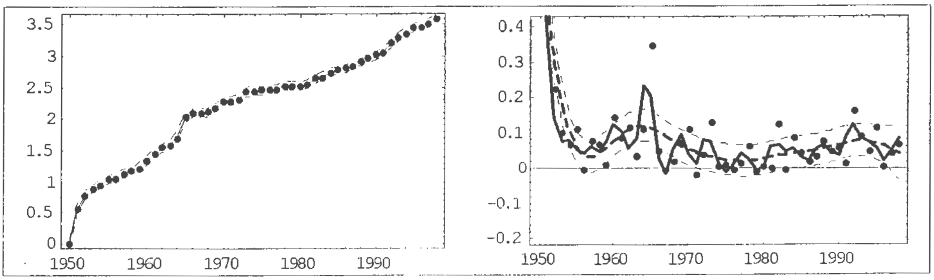
ICELAND



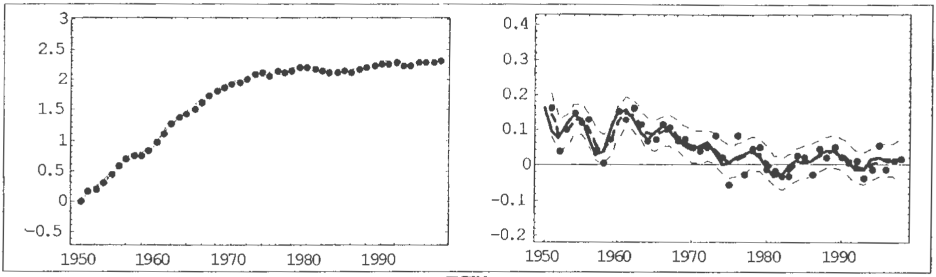
IRELAND



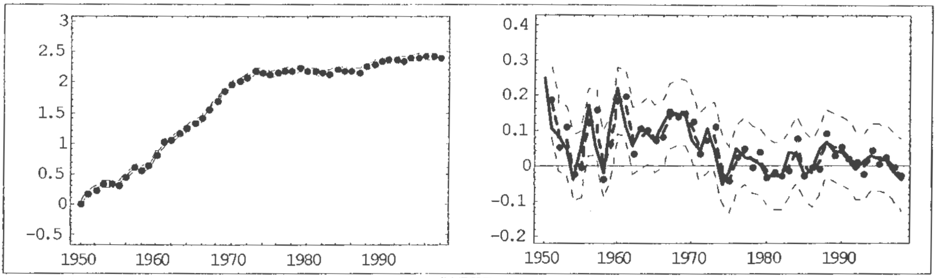
ISRAEL



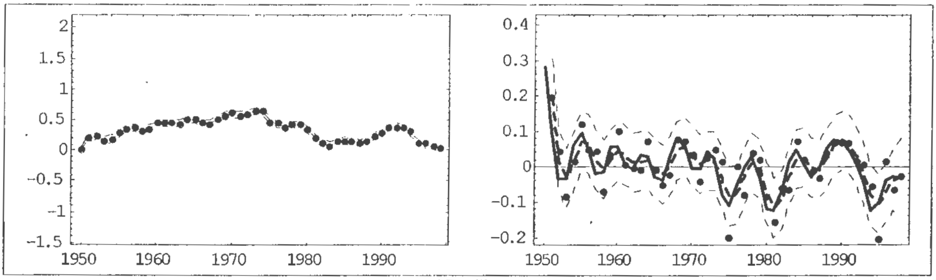
ITALY



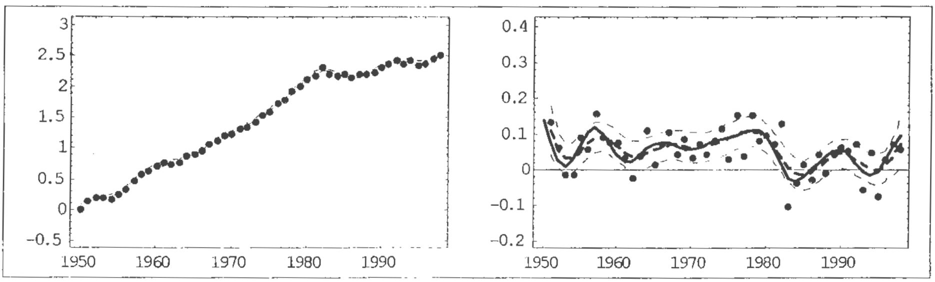
JAPAN



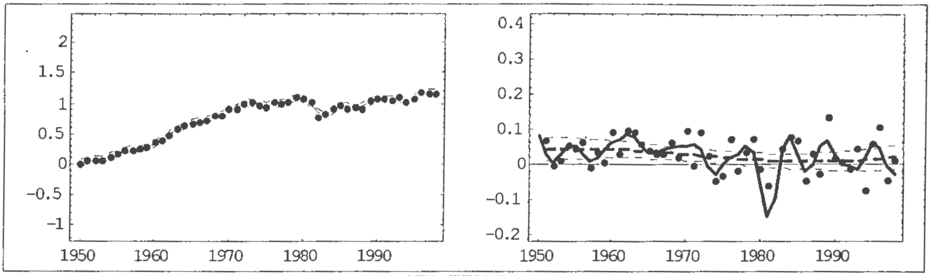
LUXEMBOURG



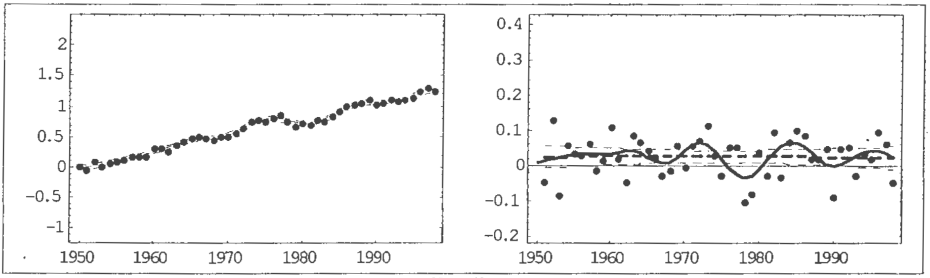
MEXICO



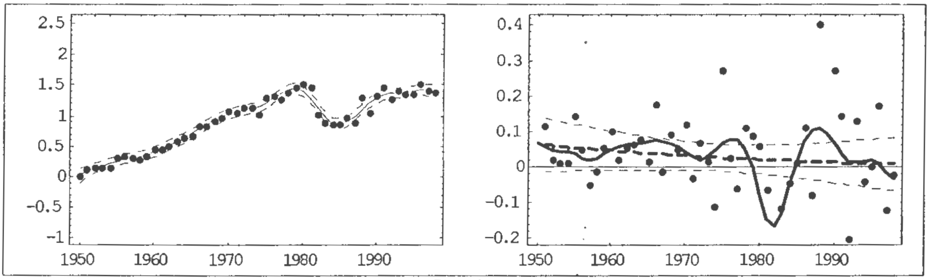
NETHERLANDS



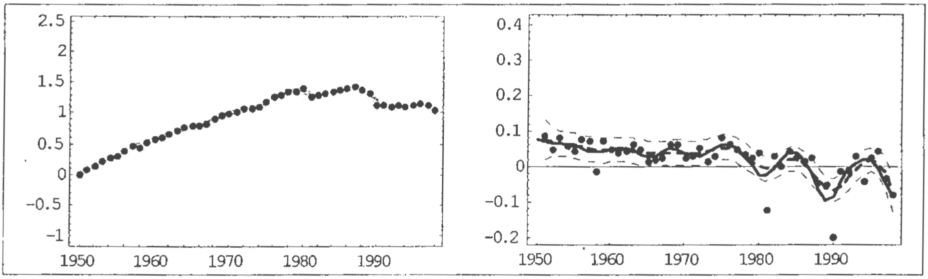
NEW ZEALAND



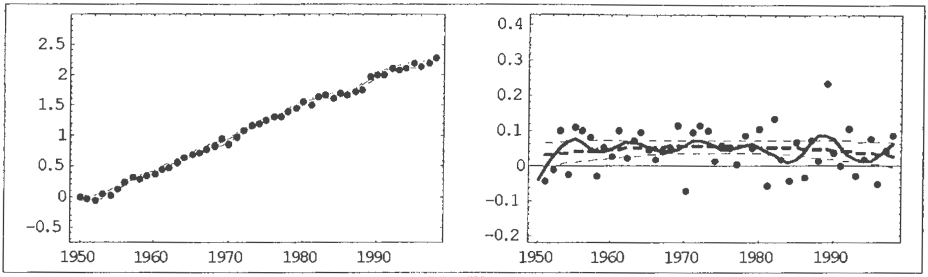
NORWAY



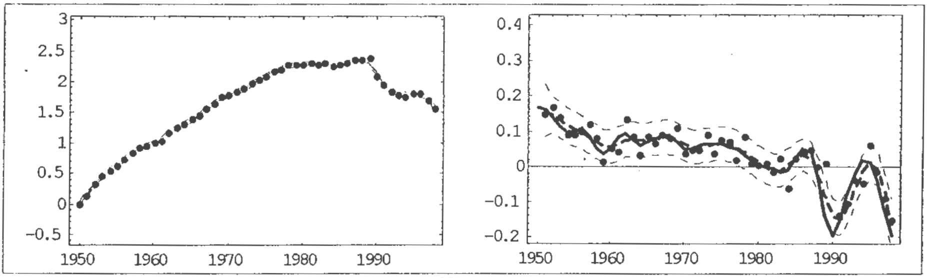
POLAND



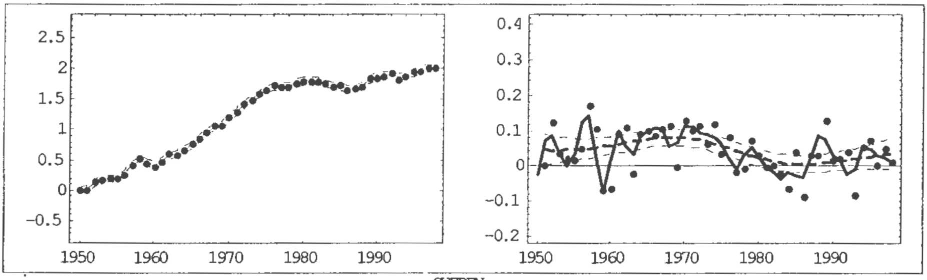
PORTUGAL



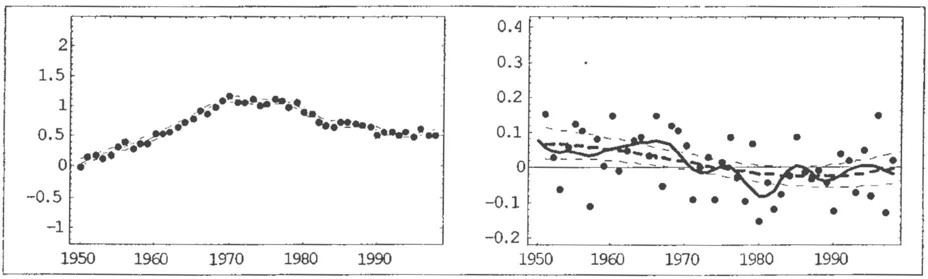
ROMANIA



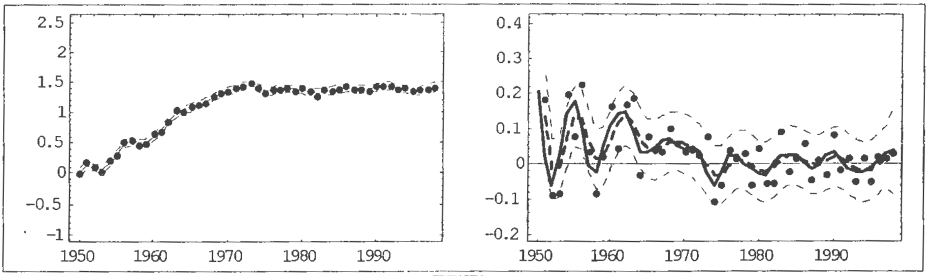
SPAIN



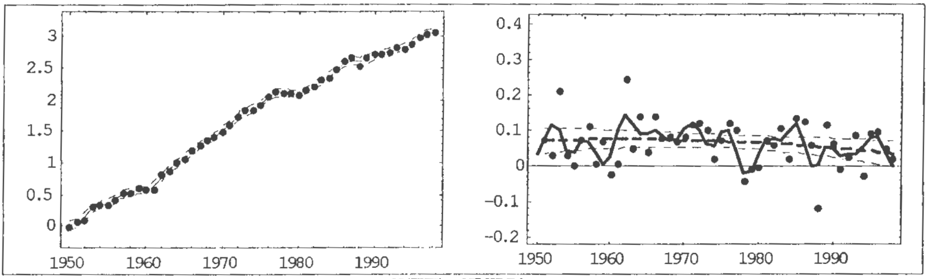
SWEDEN



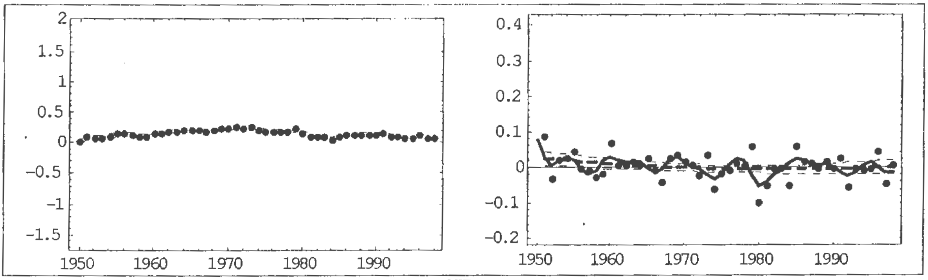
SWITZERLAND



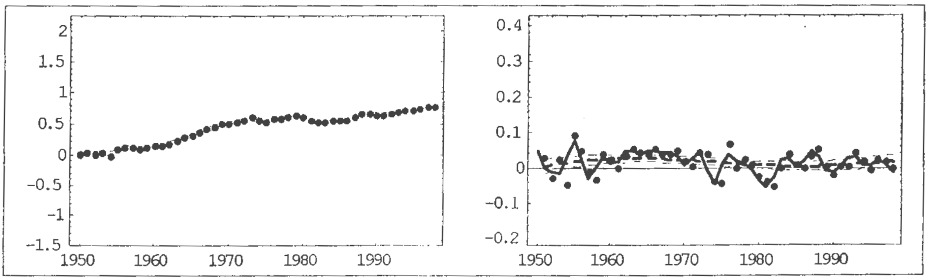
TURKEY



UNITED KINGDOM

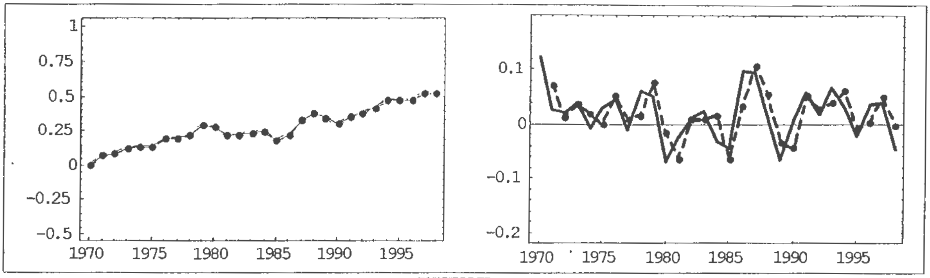


USA

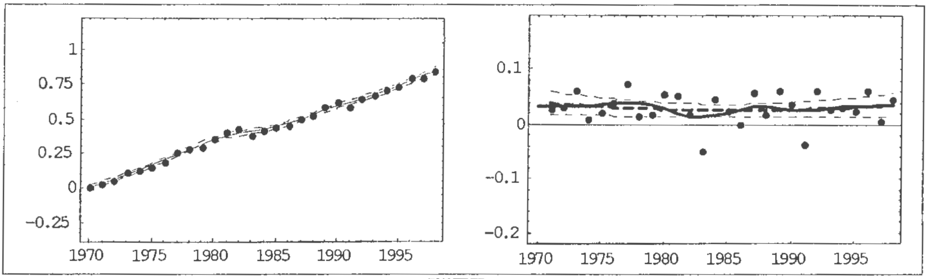


II. Years 1970-1998

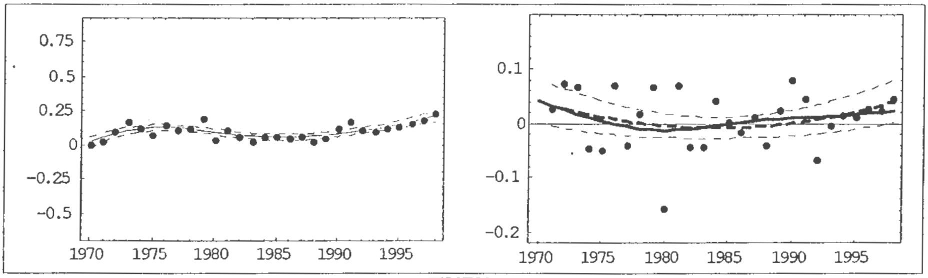
ARGENTINA



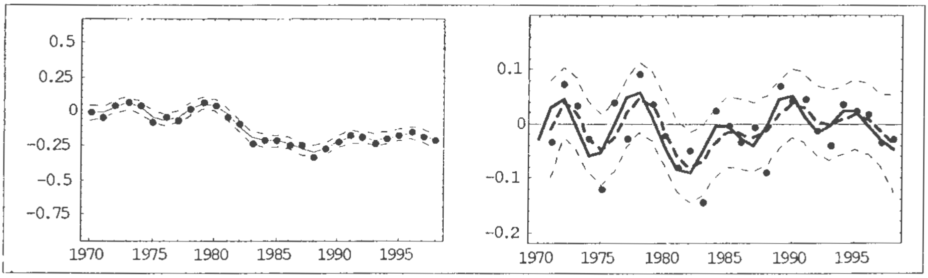
AUSTRALIA



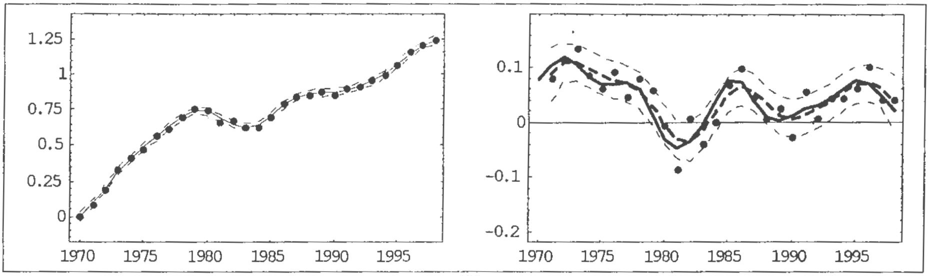
AUSTRIA



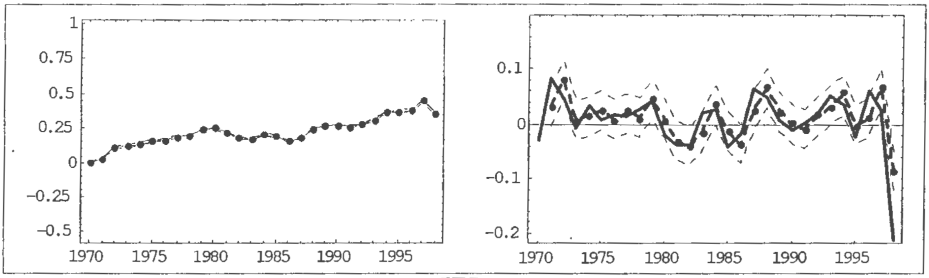
BELGIUM



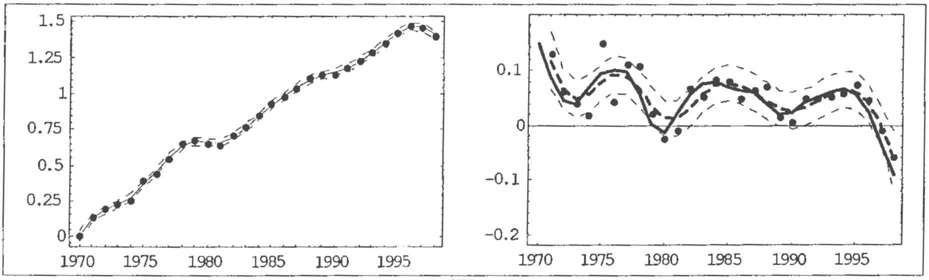
BRAZIL



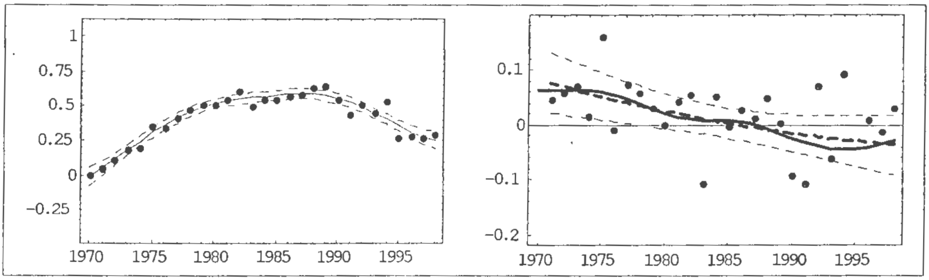
CANADA



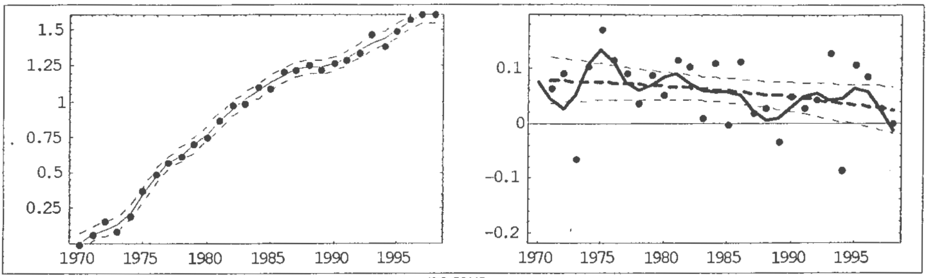
CHINA



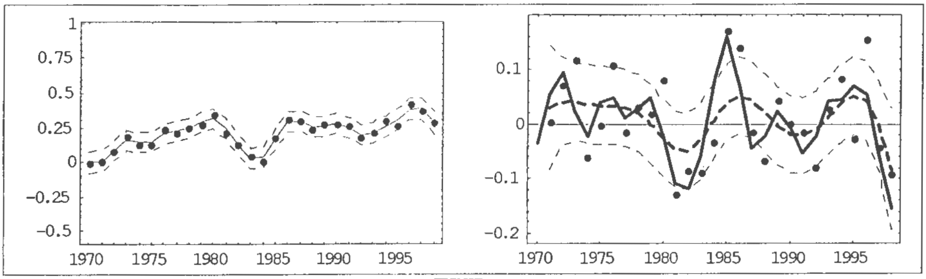
CUBA



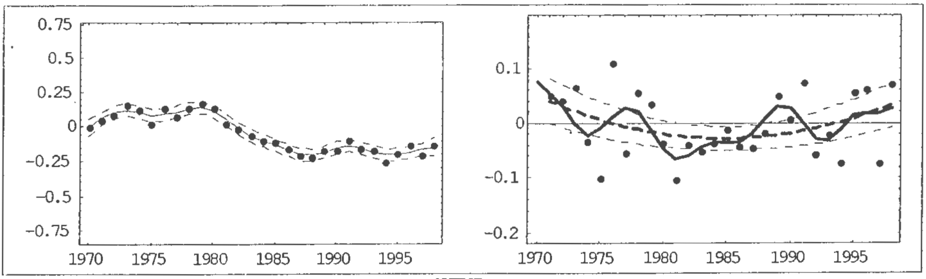
EGYPT



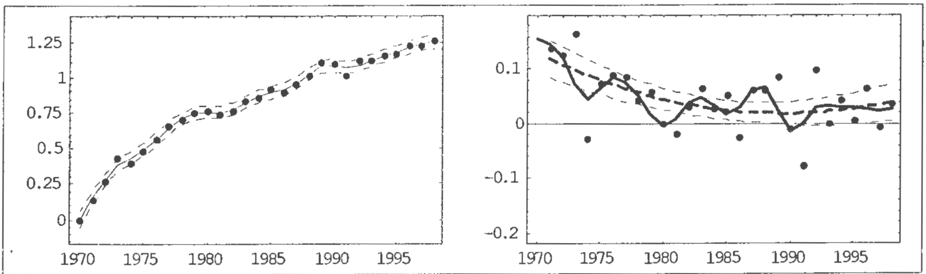
FINLAND



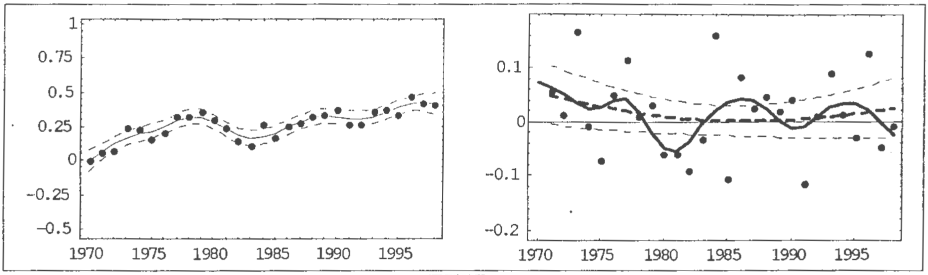
FRANCE



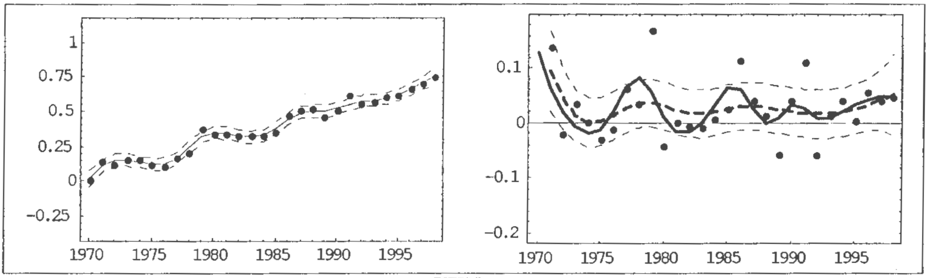
GREECE



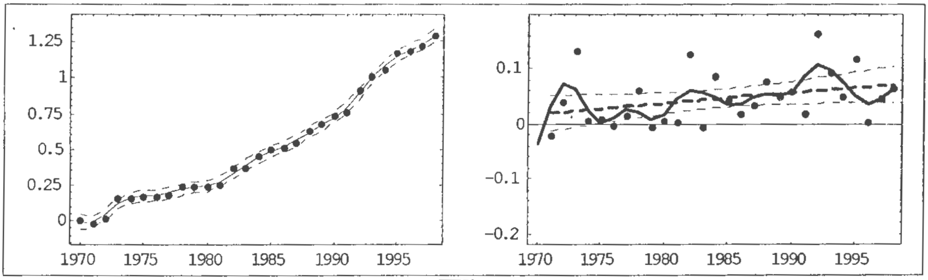
ICELAND



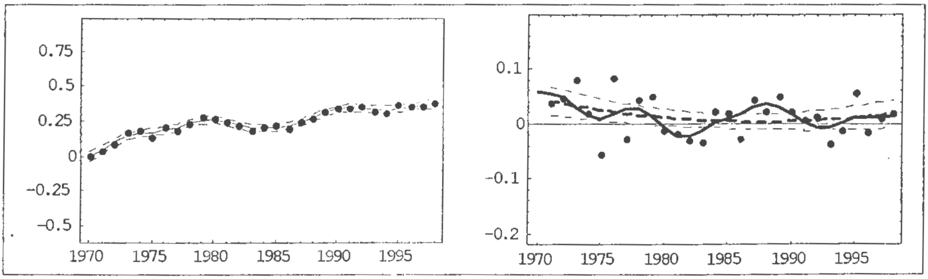
IRELAND



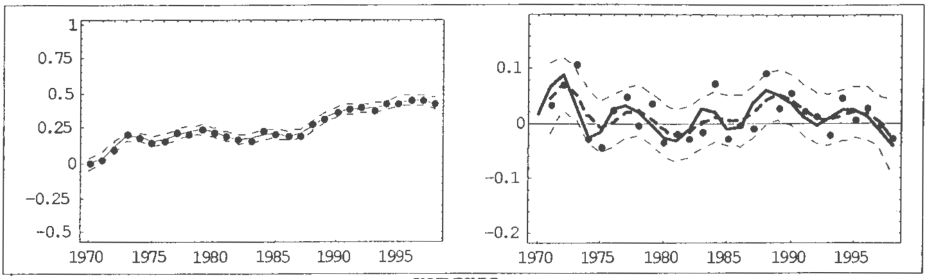
ISRAEL



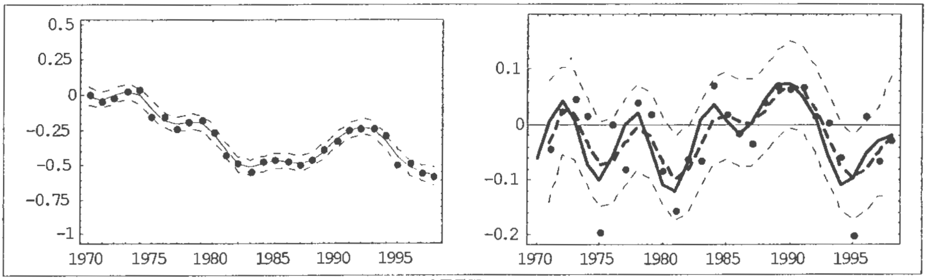
ITALY



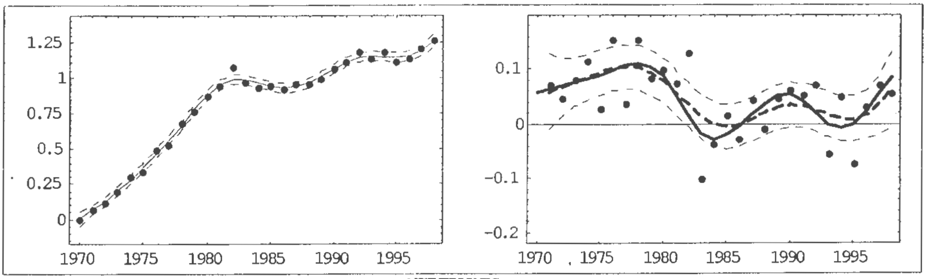
JAPAN



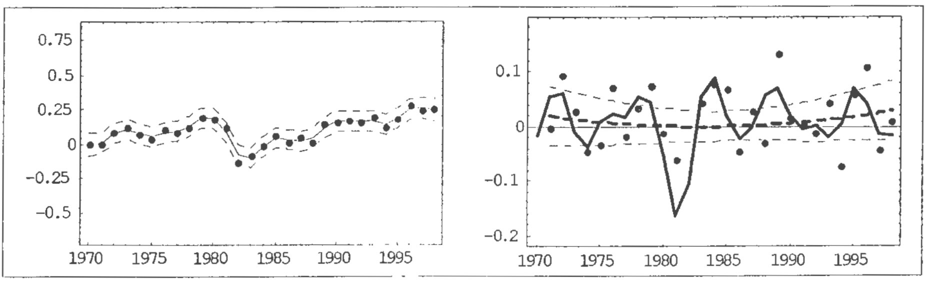
LUXEMBOURG



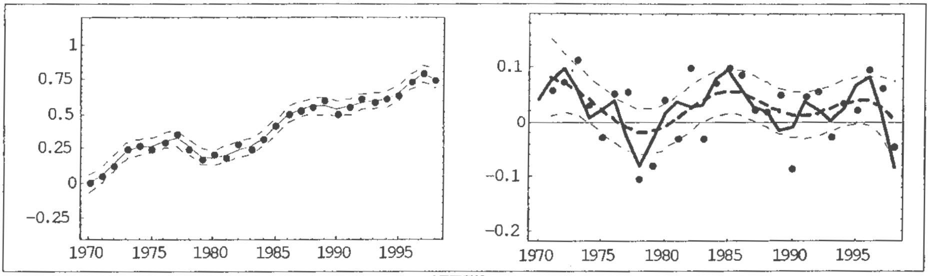
MEXICO



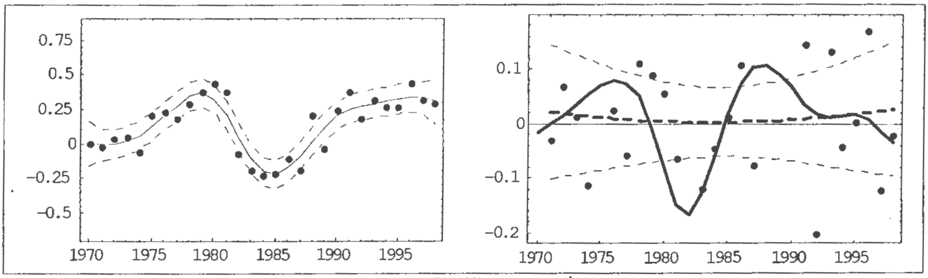
NETHERLANDS



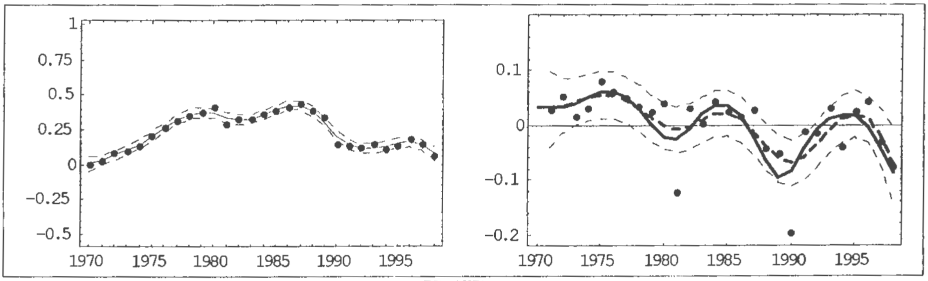
NEW ZEALAND



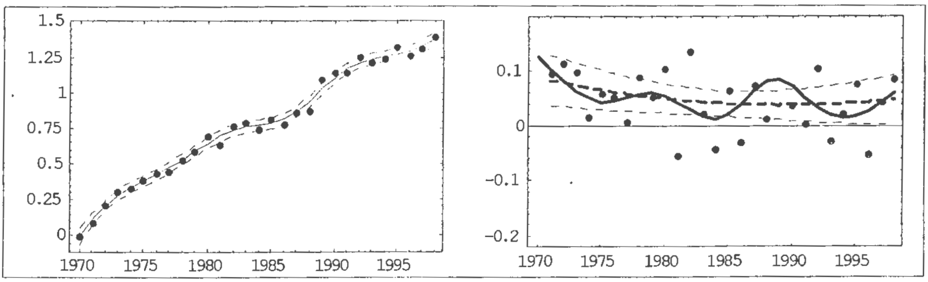
NORWAY



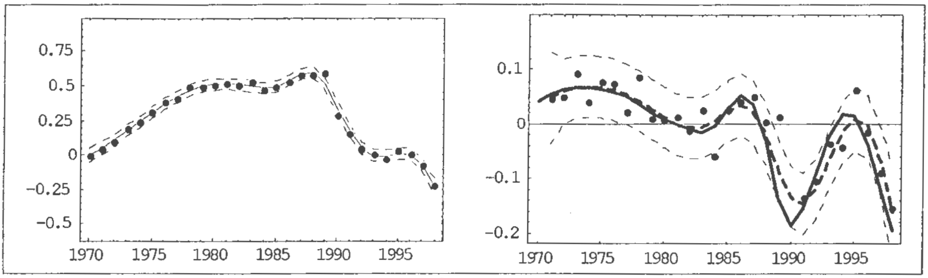
POLAND



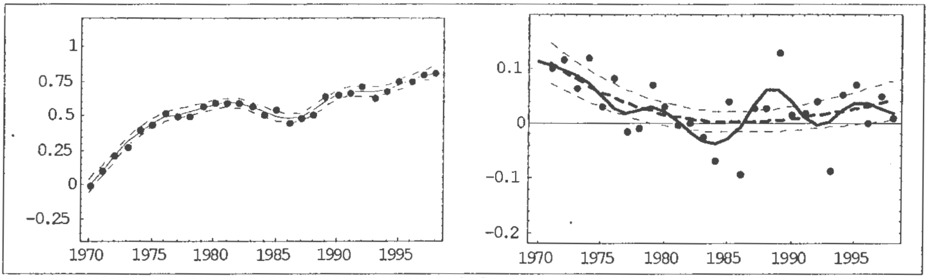
PORTUGAL



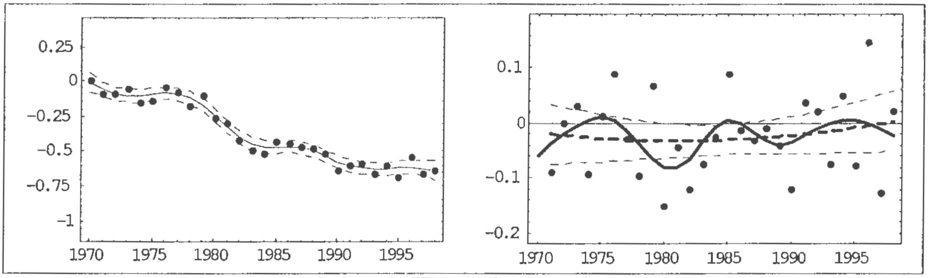
ROMANIA



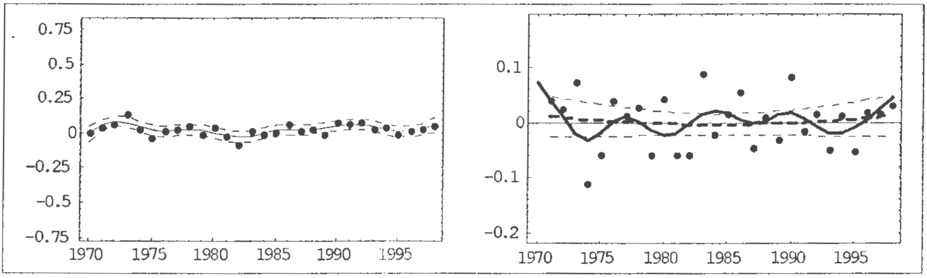
SPAIN



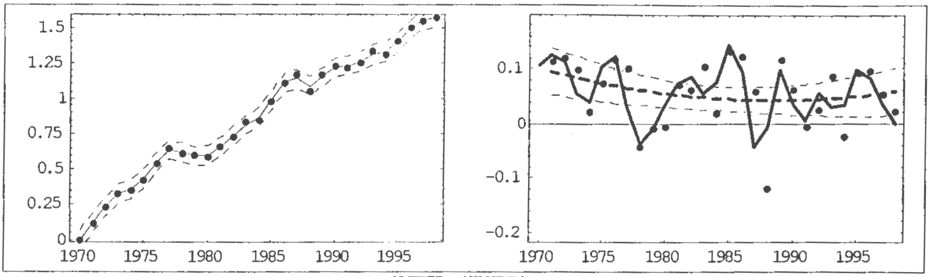
SWEDEN



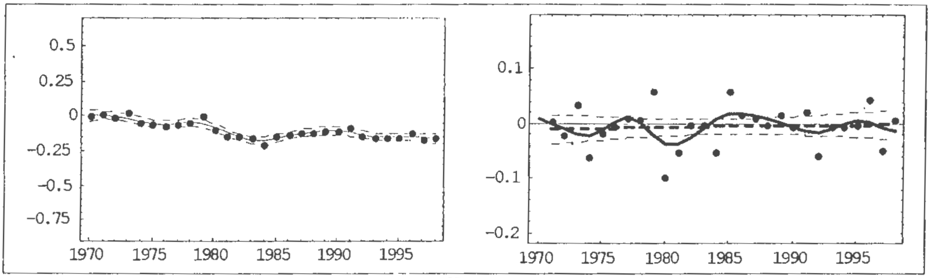
SWITZERLAND



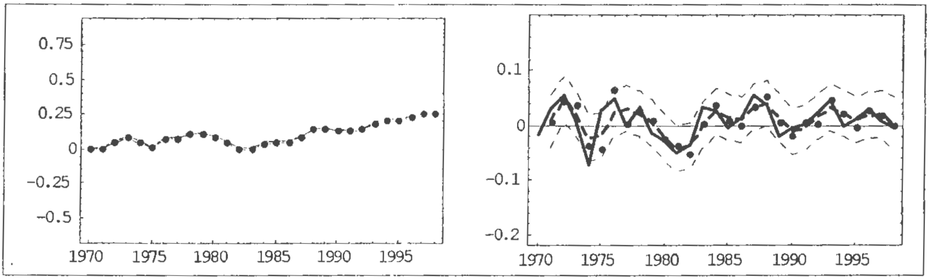
TURKEY



UNITED KINGDOM



USA



APPENDIX B

Results of fitting the piecewise exponential models.

POLAND 1870 - 1998

Estimated parameters and model verification statistics for the logarithmic regression models.

YEARS 1870-1914

	Estimate	SE	TStat	PValue
1	-29.9226	0.535618	-55.8655	0.
t	0.0179755	0.000283089	63.4976	0.

E.Var. 0.000608259

YEARS 1918-1938

	Estimate	SE	TStat	PValue
1	25.2475	5.74289	4.3963	0.000310319
t	-0.0109162	0.00297866	-3.66481	0.00164665

E.Var. 0.00683176

YEARS 1947-1978

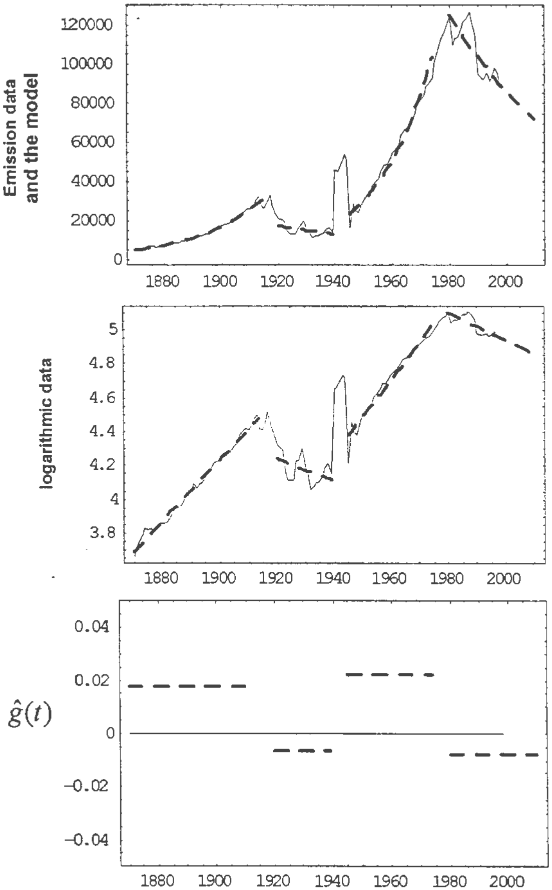
	Estimate	SE	TStat	PValue
1	-35.7734	0.935288	-38.2486	0.
t	0.0206553	0.000476575	43.3411	0.

E.Var. 0.000619593

YEARS 1979-1998

	Estimate	SE	TStat	PValue
1	20.1063	2.60431	7.72036	$4.05251' 10^{-7}$
t	-0.00758245	0.00130968	-5.78953	0.0000173922

E.Var. 0.00114065



AUSTRALIA 1870 - 1998

Estimated parameters and model verification statistics for the logarithmic regression models.

YEARS 1870- 1914

	Estimate	SE	TStat	PValue
1	- 65.4169	1.4499	- 45.1182	0.
t	0.0361896	0.000766314	47.2255	0.

E.Var. 0.00445713

YEARS 1915- 1934

	Estimate	SE	TStat	PValue
1	- 4.22192	4.06755	- 1.03795	0.313035
t	0.00418793	0.00211355	1.98147	0.0630237

E.Var. 0.00297062

YEARS 1935- 1979

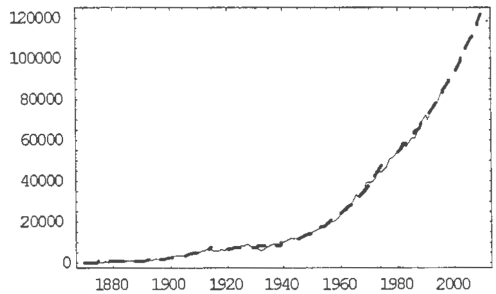
	Estimate	SE	TStat	PValue
1	- 32.8581	0.487671	- 67.3777	0.
t	0.0190004	0.000249188	76.2495	0.

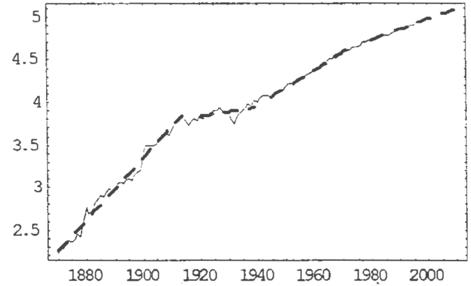
E.Var. 0.000471297

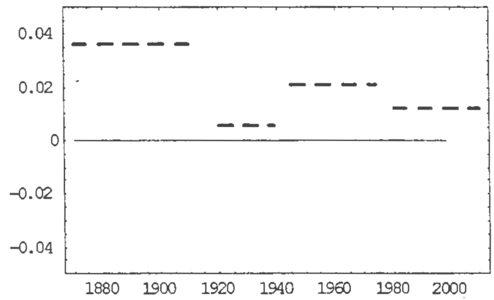
YEARS 1980- 1998

	Estimate	SE	TStat	PValue
1	- 18.5516	1.00012	- 18.5494	1.02063´ 10^{-12}
t	0.0117601	0.000502822	23.3882	2.28706´ 10^{-14}

E.Var. 0.000144113







AUSTRIA 1870 - 1998

Estimated parameters and model verification statistics for the logarithmic regression models.

YEARS 1870 - 1914

	Estimate	SE	TStat	PValue
1	- 34.0125	2.05615	- 16.5418	0.
t	0.0199313	0.00108673	18.3406	0.

E.Var. 0.00896373

YEARS 1920 - 1939

	Estimate	SE	TStat	PValue
1	37.439	9.09654	4.11574	0.000649088
t	- 0.0175225	0.00471443	- 3.71679	0.00157852

E.Var. 0.0147802

YEARS 1945 - 1974

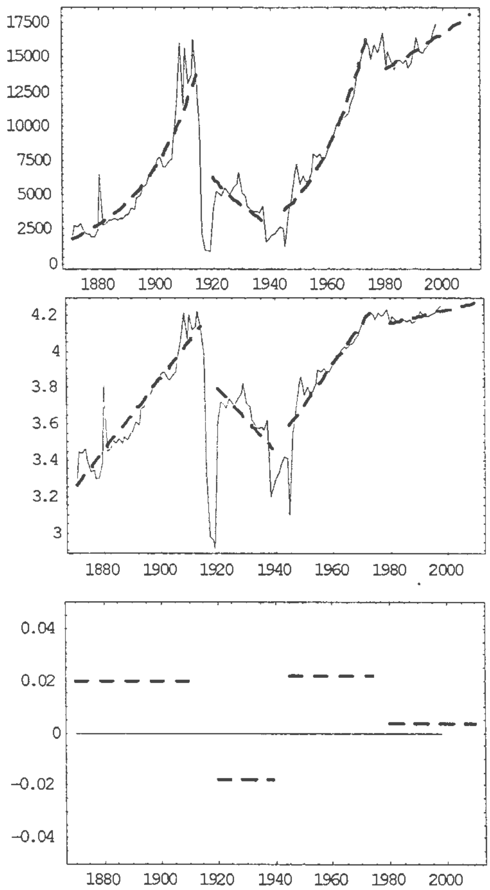
	Estimate	SE	TStat	PValue
1	- 39.0009	4.59309	- 8.49121	3.12151 ´ 10-9
t	0.0219002	0.00234399	9.34312	4.21976 ´ 10-10

E.Var. 0.0123484

YEARS 1984 - 1998

	Estimate	SE	TStat	PValue
1	- 5.33736	1.76231	- 3.02861	0.0096906
t	0.00478445	0.000885136	5.40532	0.000120043

E.Var. 0.000219371



UNITED STATES OF AMERICA
1870 - 1998

Estimated parameters and model verification statistics for the logarithmic regression models.

YEARS 1870-1914

	Estimate	SE	TStat	PValue
1	-45.7851	0.770813	-59.3985	0.
t	0.0268672	0.000407397	65.9485	0.

E.Var. 0.00125973

YEARS 1915-1940

	Estimate	SE	TStat	PValue
1	5.08135	2.70856	1.87603	0.0728651
t	0.000290048	0.00140521	0.206409	0.838212

E.Var. 0.00288787

YEARS 1941-1979

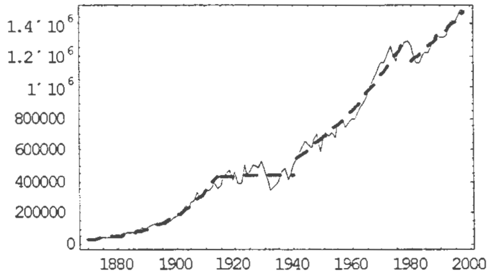
	Estimate	SE	TStat	PValue
1	-13.6495	0.690558	-19.7659	0.
t	0.00999051	0.00035232	28.3563	0.

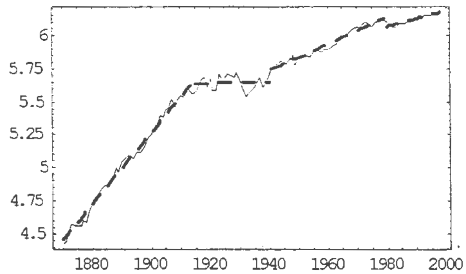
E.Var. 0.000613199

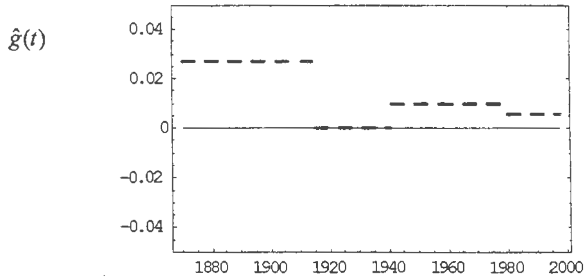
YEARS 1980-1998

	Estimate	SE	TStat	PValue
1	-5.38114	1.11315	-4.83417	0.000155248
t	0.00578105	0.000559649	10.3298	9.6028´ 10^{-9}

E.Var. 0.000178528







CANADA 1870 – 1998

Estimated parameters and model verification statistics for the logarithmic regression models.

YEARS 1870- 1914

	Estimate	SE	TStat	PValue
1	- 70.7008	2.42254	- 29.1845	0.
t	0.0392146	0.00128038	30.6272	0.
E.Var.	0.0124429			

YEARS 1918- 1939

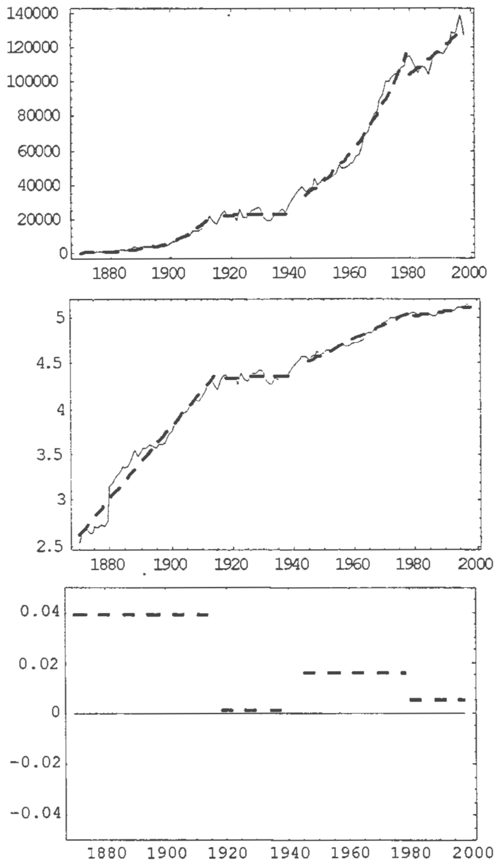
	Estimate	SE	TStat	PValue
1	2.2835	3.14261	0.726627	0.475874
t	0.00107509	0.00162955	0.659746	0.516943
E.Var.	0.00235139			

YEARS 1945- 1979

	Estimate	SE	TStat	PValue
1	- 25.9854	0.976209	- 26.6187	0.
t	0.0156893	0.000497551	31.5331	0.
E.Var.	0.00088378			

YEARS 1980- 1998

	Estimate	SE	TStat	PValue
1	- 5.7115	1.61099	- 3.54534	0.00248709
t	0.0054192	0.000809947	6.69081	3.79577×10^{-6}
E.Var.	0.000373928			



FRANCE (INCLUDING MONACO) 1870-1998

Estimated parameters and model verification statistics for the logarithmic regression models.

YEARS 1870-1914

	Estimate	SE	TStat	PValue
1	-15.4856	0.67664	-22.886	0.
t	0.0105217	0.000357624	29.4211	0.

E.Var. 0.000970722

YEARS 1918-1929

	Estimate	SE	TStat	PValue
1	-54.8631	6.14916	-8.92204	4.47441´ 10^{-6}
t	0.0309564	0.00319685	9.6834	2.13301´ 10^{-6}

E.Var. 0.00146144

YEARS 1945-1979

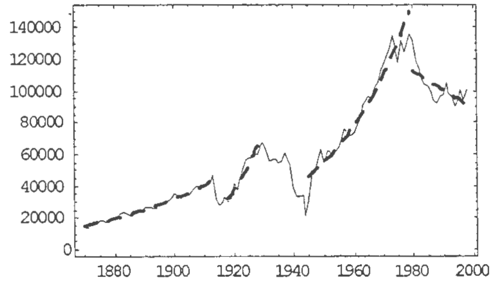
	Estimate	SE	TStat	PValue
1	-24.442	1.39318	-17.544	0.
t	0.0149665	0.000710073	21.0774	0.

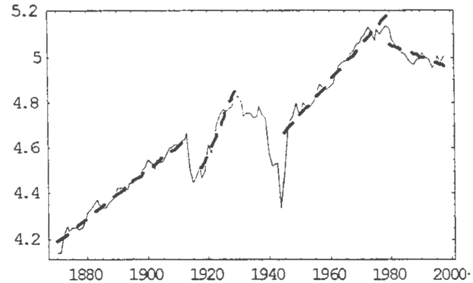
E.Var. 0.00180001

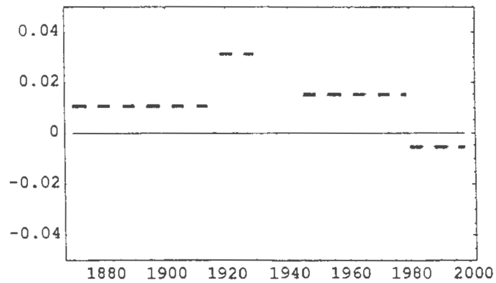
YEARS 1980-1998

	Estimate	SE	TStat	PValue
1	14.9833	2.53428	5.91227	0.0000170544
t	-0.00501543	0.00127414	-3.93632	0.00106445

E.Var. 0.00092536







JAPAN 1880 - 1998

Estimated parameters and model verification statistics for the logarithmic regression models.

YEARS 1880-1940

	Estimate	SE	TStat	PValue
1	-52.5721	1.99232	-26.3873	0.
t	0.0295688	0.00104306	28.3482	0.

E.Var. 0.0205735

YEARS 1880-1940

	Estimate	SE	TStat	PValue
1	-52.5721	1.99232	-26.3873	0.
t	0.0295688	0.00104306	28.3482	0.

E.Var. 0.0205735

YEARS 1945-1974

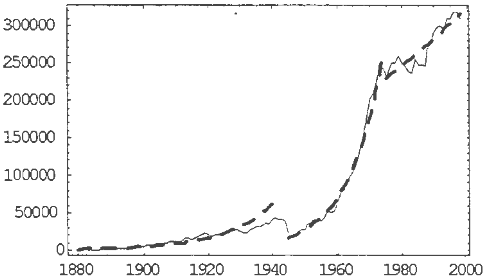
	Estimate	SE	TStat	PValue
1	-75.012	1.85253	-40.4916	0.
t	0.0407414	0.000945402	43.0943	0.

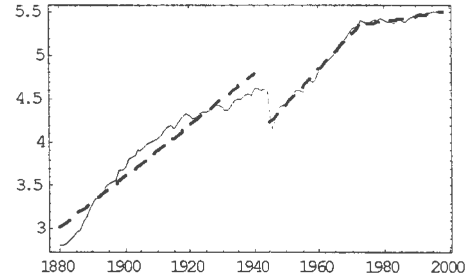
E.Var. 0.00200878

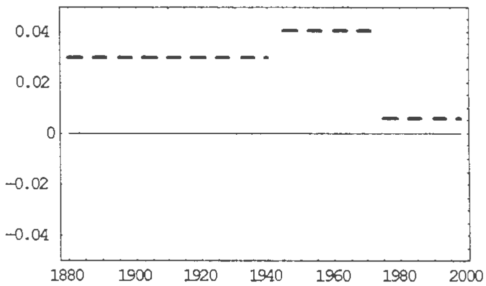
YEARS 1975-1998

	Estimate	SE	TStat	PValue
1	-6.45422	1.21319	-5.32004	0.0000243589
t	0.00598255	0.000610713	9.796	1.7541´ 10^{-9}

E.Var. 0.000428916







NETHERLANDS 1980 - 1998

Estimated parameters and model verification statistics for the logarithmic regression models.

YEARS 1880-1914

	Estimate	SE	TStat	PValue
1	-23.8167	0.875393	-27.2069	0.
t	0.0144343	0.000461455	31.2799	0.

E.Var. 0.000760199

YEARS 1920-1939

	Estimate	SE	TStat	PValue
1	-20.8725	3.81226	-5.4751	0.0000335984
t	0.0128839	0.00197577	6.52098	3.94722×10^{-6}

E.Var. 0.00259593

YEARS 1945-1974

	Estimate	SE	TStat	PValue
1	-40.5533	3.273	-12.3903	6.97553×10^{-13}
t	0.0228893	0.00167031	13.7036	6.12843×10^{-14}

E.Var. 0.00627036

YEARS 1980-1998

	Estimate	SE	TStat	PValue
1	-7.60261	2.98052	-2.55077	0.0206771
t	0.00612994	0.00149849	4.09073	0.00076184

E.Var. 0.00127993