INFREQUENT PERMANENT SHOCKS AND SIGNAL EXTRACTION IN MACROECONOMIC TIME SERIES*

Marco BIANCHI
C.O.R.E and I.R.E.S
Université Catholique de Louvain
B-1348, Louvain-la-Neuve

Preliminary version: June, 1992.

ABSTRACT: A procedure based on density estimation is suggested in the paper to discriminate trend stationary processes about *local* linear time trends from difference stationary processes. A 'rule of thumb' is constructed to detect the suitability of a segmented trend representation, and a regression analysis is used to identify the number and the dates of structural breaks. The U.S. series of nominal wages over the period 1900-1970 is analysed according to the assumption the dynamics are driven by exogenous shocks which occur infrequently. In a multivariate domain, implications of segmented trend modeling for cointegration theory are also briefly considered.

*The paper is the result of research activity I carried out while at the *Université Catholique de Louvain* (academic years 1989-1990 and 1991-1992) and the *London School of Economics* (academic year 1990-1991). At the UCL I especially benefited the comments of professors Luc Bauwens, Wolfgang Härdle, José Paris, and Henri Sneessens, while at the LSE my acknowledgment goes to prof. Charles Bean and Dr. George Evans for helpful suggestions. A special thank I like to address to prof. Nobuhisa Kashiwagi for encouragement and patient explanations. Within a project of collaboration with the University of Bologna, financial support from "Unipol Assicurazioni" is gratefully acknowledged. The usual disclaimer applies.

1. Introduction

In recent times much effort has been devoted in the literature to test for unit roots in macroeconomic time series. The main interest in this literature was that in presence of unit roots random shocks have persistent effects on economic variables, with potentially important implications for business cycle theorizing (see Nelson and Plosser, 1981, section 5). Former unit root tests were developed by Fuller (1976) and Dickey and Fuller (1979, 1981) according to simulation-based critical values, while complementary approaches, based on estimates of spectral densities at frequency zero, were thereafter developed by Campbell and Mankiw (1987a,b) and Cochrane (1988). Results persuaded in most cases economists about the strong empirical evidence of unit roots in macroeconomic time series, as recently reviewed for example by Campbell and Mankiw (1987a,b; 1989) and Perron (1988) in the domain of classical inference, and, even more recently, by Phillips (1991) in the domain of Bayesian inference.

Despite the broad agreement on unit roots and persistence of shocks, Perron (1989) published a paper in which he argued to what extent the acceptance of the null hypothesis of a unit root may be the result of an inadequate specification of the alternative hypothesis in the form of a *global* linear time trend, rather than the identification of the 'true' underlying data generating process.[1] Perron provided in fact a test of difference stationarity against the alternative hypothesis of trend stationarity in the form of *local* time trend with one break in the trend, and found then that the unit root hypothesis was rejected for 8 of the 11 series originally considered by Nelson and Plosser (1982): Thus, he concluded that "*all series analyzed have a unit root if the trend function is not allowed to change*" (p. 1385), that is "*trend stationary processes with a break are nearly observational equivalent to unit root processes with strong mean reversion and a fat-tailed distribution for the error sequence*" (p. 1389).

Along similar lines of research, Hamilton (1989) proposed a model in which changes in regime were allowed for, through the assumption the mean growth rate of a series is subject to *occasional* shifts which "*follow a nonlinear stationary process rather than a linear stationary process*" (p.357). Rappoport and Reichlin (1989) and Balke and Fomby (1991) also advocated that shocks may occur infrequently with 'strong' persistent effects (permanent effects on the growth rates of the series), rather than frequently with 'weaker' persistent effects (permanent effects on the levels but not on the growth rates of the series). Balke and Fomby, in particular, explicitly pointed out that Dickey-Fuller unit root tests and Cochrane (1988) variance ratio statistics, despite they are useful for determining the persistence of shocks, cannot discriminate in fact between frequent permanent shocks with small variance from

[1] Indeed, after the contribution of Nelson and Kang (1981), there are now few doubts that fitting pure deterministic time trends to most economic time series would deliver spurious regressions in the sense of Granger and Newbold (1974).

infrequent permanent shocks with large variance.

In this context, where only occasionally shocks have persistent effects, the aim of a statistical analysis goes clearly beyond the simple response 'yes' or 'no' to the null hypothesis of a unit root against the alternative of a global linear time trend, or to the estimate of some measures of persistence; rather, the real problem is to identify an array of historical dates at which infrequent permanent shocks are suspected to occur. In the unit root literature, as far as shocks were assumed to occur frequently with persistent effects on the levels of the series, there were in practice no reason to discriminate among different types of shocks, that is to attribute them to different historical dates. In Balke and Fomby (1991), on the contrary, the discrimination was implemented through an outlier analysis of ARIMA residuals, along the intervention method proposed by Tsay (1988).

In this paper, the possibility that events that have important and long-lasting effects on economic time series do not occur every period (that is precisely the hypothesis that shocks occur infrequently) lead us to consider in further details Perron's segmented trend framework. In particular, we propose a procedure for discriminating difference stationary versus trend stationary processes when breaks are allowed for, while no particular restriction is imposed on their number. The basic intuition behind the procedure is that when a random walk model with drift is estimated, but the underlying process is a segmented trend plus noise, residuals display, according to Perron, *"fat tails or multiple modes density distributions"*: Hence, useful information can be gained in principle by an accurate estimate of the density distribution of the residuals. More generally, nonetheless, the problem of the identification of the 'true' data generating process under the hypothesis of infrequent permanent shocks is equivalent to a statistical problem of signal extraction in time series, where the signal itself is restricted to be a step function of time. A solution to this problem is considered based on a regression strategy proposed by Kashiwagi (1991), that we will later discuss in details.

The plan of the paper is as follows. In section 2 alternative data generating processes are introduced and compared, and a simple procedure for discriminating among them is suggested on the basis of density estimation. An application of such a procedure is provided with a simulated time series. In section 3 nonparametric methods for signal extraction are briefly considered, with a short reference to the main advantages and disadvantages associated with these methods. The conclusion that nonparametric regressions produce satisfactory results in time series only at the condition extra information concerning the number and the location of knots (or change points) is previously incorporated, leads us to consider in section 4 a regression approach for the detection of change points in a sequence of observations. In section 5 an application to the U.S. series of nominal wages over the sample 1900-1970 is carried out, while in section 6 some implications for cointegration theory are briefly considered. The last section summarizes and concludes with some remarks for economic modeling.

2. Random walks versus segmented trends

Because of widespread empirical evidence, unit roots are now very popular in macroeconomics. On the economic front, however, but we will see in this section on the statistical front as well, random walk modeling is not immune from critiques and objections of various type. Some of them are for example the following:

a) What is the economic sense of the proposition: "The economy evolves as a random walk with drift" ? If the economy were a random walk, there would be no hope for predicting future values; furthermore, despite current innovations would produce permanent effects on the levels, the growth rate of the economy would be postulated constant, that is the possibility of shifts in the natural rate would be excluded *a priori.*

b) Why should exogenous shocks have permanent effects on the levels of the series at *each* point of the time ? And, on the other hand: Why should be neglected the possibility that shocks can have, but infrequently, permanent effects on the *growth rates* of the series ?

c) What is of the Lucas critique, in presence of unit roots ? The random walk model, as any model in which parameters are assumed constant over time, is incompatible with the Lucas critique.

It is perhaps accordingly to the above objections that Rappoport and Reichlin proposed in 1989 a segmented trend model (ST hereafter) which they recognized as a flexible representation in between the two extreme cases of 'trend stationary' (TS hereafter) and 'difference stationary' (DS hereafter) processes. On the economic side, their motivation was that "*economists are accustomed to attributing changes in trend rates of growth to event that occur infrequently*" (p. 169), that is "*the economy can persist in a particular steady-state for some time, and should therefore produce segmented trend data*" (p. 176). To the extent the simple TS model revealed however inappropriate, because inconsistent with the view that trends are subject to permanent shocks, Rappoport and Reichlin did also make clear their considerations "*do not conflict with this view, but suggest that, in contrast to the standard DS model, these shocks are more accurately characterised as occurring infrequently*" (p. 176). Along a similar line, Hamilton (1989) explained why a DS model can be better replaced by a stochastic model in which changes in regime follow a first order Markov process, rather than a simple linear process, while Balke and Fomby (1991) assumed a stochastic process of the Bernoulli type - but they also admitted a deterministic or a stochastic modelization is just a way to approach the same problem from an *ex-post* or *ex-ante* view (see Balke and Fomby, 1991, p. 67).

As far as we are concerned in this paper, our goal is to assess and compare the general performance of four alternative data generating processes (DGP) for a generic non stationary time series y_t. The models we consider are described by the equations

TS: $$y_t = \alpha_1 + \mu_1 t + v_{1t} \tag{1}$$

RW: $$y_t = \mu_2 + y_{t-1} + v_{2t}, \tag{2}$$

DS: $$y_t = \mu_3 + y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + v_{3t}, \tag{3}$$

ST: $$y_t = \alpha_{4i} + \mu_{4i} t + \varepsilon_t \quad \varepsilon_t \sim N(0,\sigma^2), \quad j(i)+1 \le t \le j(i+1) \tag{4}$$

where: y_t is a series in natural logarithms non stationary in mean (and eventually in variance); t is the time index, with $t = 1,2,...,T$; i is the index for the generic change point (structural break), with $i = 0,1,2,...n$; $j(i)$ are the dates at which change points are located, and by definition $j(0)=0$ and $j(n+1)=T$. The error terms generated by the first two models are not necessarily normally distributed or white noise processes; residuals generated by the DS model are white noise by construction, but not necessarily normally distributed; residuals generated by the ST model are assumed white noise *and* normally distributed.

Other main features of the models above may be described shortly as follows. According to equation (1), the pure TS model fits a *global* linear time trend to the actual series in levels, that is shocks are not supposed here to affect the trend component of the series. Equation (2) suggests y_t evolves as a random walk with drift, while equation (3) represents a generalization of the simple RW model in which lagged differences are accounted for, to ensure the white noise nature of the innovation term. Contrary to the TS case, both models imply shocks occur at each period of time which have permanent effects on the *levels* of the series. The segmented trend model of equation (4) suggests only infrequently shocks have persistent effects (either on the levels and on the *growth rates* of the series), while most of the time they have only transitory effects.[2]

In order to see the consequences of the above representations on the growth rate series, it is useful to rewrite (1)-(4) in first differences. We have then:

TS: $$\Delta y_t = \mu_1 + \Delta v_{1t} \tag{5}$$

RW: $$\Delta y_t = \mu_2 + v_{2t}, \tag{6}$$

DS: $$\Delta y_t = \mu_3 + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + v_{3t}, \tag{7}$$

ST: $$\Delta y_t = \mu_{4i} + \Delta\varepsilon_t \quad \Delta\varepsilon_t \sim N(0, 2\sigma^2), \quad j(i)+1 \le t \le j(i+1). \tag{8}$$

It is easily seen from equations (5) and (6) that TS and RW models predict a constant average growth rate, that is they assume Δy_t to have a unimodal density centred respectively at μ_1 and μ_2. Such representations obviously do not allow for shocks to have permanent effects on the growth rates of the series, and moreover, to the extent estimates of μ_1 and μ_2 are not too different, they are in fact observationally equivalent for the growth rate series.[3] On the other hand, DS and ST models predict *fitted* growth rates shift over time, i.e. both DS and ST models assume Δy_t to be characterized in fact as a multimodal density: The only difference is then just in the way the average growth rate is supposed to shift, the shift being continuous (frequent) in the DS model and discrete (infrequent) according to the ST model. Notice finally the ST model is the only one to be *consistent with the Lucas critique*.

The above remarks induce to consider a natural procedure for the statistical discrimination of alternative DGP, which is based on the density estimation of Δy_t. Suppose in fact that estimating the density function of Δy_t, this has a multimodal shape; in such a case both the TS and the RW models are immediately ruled out, and the statistical task simply reduces to identify which is more appropriate between the DS and the ST model. At the same time, as far as the TS and the RW models are particular cases of the segmented trend plus noise model, in which $\mu_{4i} = \mu_1$ (or μ_2) $\forall\, i=1,2,\ldots,m=n+1$, under the hypothesis the 'true' data generating process is a segmented trend, it must be the case that outliers and/or multiple modes characterize the density function of the

---

[2] In the case of the ST model, as it has been noticed by Perron (1989), a particular postulate is introduced which differenciates the ST approach from previous approaches. Indeed, according to the ST model *"only few events (shocks) have permanent effects on various macroeconomic variables"*, that is shocks are not *"realizations of the underlying data-generating mechanism"*. It is then in this sense shocks themselves are postulated exogenous, but to use the exogeneity assumption *as a device to remove the influence of these shocks from the noise function*. (Perron, 1989, p. 1362).

[3] This can perhaps explain the well-known difficulty of discriminating between TS and RW representations in samples of small size (see for example Blough, 1988), but also revises in part the attractiveness of unit root tests, especially when the null hypothesis is modelled as a pure random walk with drift.

residuals Δv_{1t} (or v_{2t}). To understand how such a simple discrimination can be carried out, we consider in the rest of this section a time series artificially generated as a segmented trend plus noise, with one *'level'* and one *'growth'* effect.[4]

The series displayed in fig. 1 have been generated according to model (4) with n=4, with change points at $j(i)$=10, 12, 17, 22, and the following structure of growth: $\mu_{41}=0.05$ for $t\in[2\text{-}10]$, $\mu_{42}=0.10$ for $t\in[11\text{-}12]$, $\mu_{43}=0.05$ for $t\in[13\text{-}17]$, $\mu_{44}=-0.01$ for $t\in[18\text{-}22]$, and $\mu_{45}=0.05$ for $t\in[23\text{-}35]$. The initial condition is y_1=10, and v_{4t}~N(0,0.0002).[5]

Because the first break occurring at t=10 determined a transitory shift in the (let us call it) 'natural' rate (two periods only), this is the level effect in the series, while the second break, occurred at t=17, is the growth effect. The 4 change points split the sample into 5 separate sub-samples or regimes.

A careful analysis of the simulated series, reveals then the following points:

a) Estimating the density distribution of the simulated series in first differences, the multimodal shape reported in fig. 2 (left picture) indicates the invalidity of both the TS and the RW model.

b) Estimates of the growth rate parameters μ_1 and μ_2, derived running the regression equations (1) and (2), are found to be 0.041 in the TS case, and 0.044 in the RW case. This is a mean value of the true local growth rates, as it is also clear from fig. 2, right picture.

c) The inadequacy of the TS and the RW models is fully confirmed by the spectral estimate reported in fig. 3 (left picture), in which most of the variability in Δy_t is explained at the low frequencies. Δy_t has a trend structure which is not allowed for by those models.

---

[4] The terminology is borrowed from Lucas (1988), who emphasised "*the distinction between 'growth effects' - changes in parameters that alter growth rates along balanced path - and 'level effects' - changes that raise or lower balanced growth path without affecting their slope*" (Lucas, 1988, p. 12). Balke and Fomby (1991) designed the same effects under the names of segmented and shifting trends respectively.

[5] The simulation was carried out by replicating 500 times equation (4) for a given deterministic segmented trend structure. For each local linear time trend a column vector of 500 pseudo independent and normally distributed error terms, with zero mean and variance 0.1, was generated; subsequently, the mean value was taken for each column, which was summed to the corrisponding deterministic local time trend. Because of averaging the outcome of 500 replications, theoretical variance within each subsample is given by 0.1/500 = 0.0002.

d) The large persistence of shocks signaled by the spectral estimate is confirmed by Campbell and Mankiw's (1989) measures of persistence. The pairs $\{V_k, A_k\} = \{2.61, 2.20\}$ for k=3, {3.06, 2.38} for k=5, {3.07, 2.39} for k=7, and {1.90, 1.88} for k=14, are in fact estimated for a wide spectrum of lags.[6] The persistence is however the consequence in our case of infrequent, rather than frequent permanent shocks.

e) Running a Dickey-Fuller regression of the type $y_t = \alpha + \mu t + \rho\, y_{t-1} + \Sigma\, \gamma_j\, \Delta y_{t-j}$ + error, usual diagnostics indicate that 2 first differences lagged terms have to be included to get white noise residuals, what again confirms the inappropriateness of the simple RW model for our simulated series.

f) The t-ratio for the estimate of ρ in the Dickey-Fuller regression is smaller than the critical value tabulated in Fuller (1976) (-2.63 against 5% critical value of -3.45), that is the unit root hypothesis is largely accepted against the alternative of a global linear time trend. To the extent the series was generated as a ST process, however, it is evident the unit root test provides in our case more information about the inadequacy of the simple TS model, rather than identifying the true DGP in the form of a DS process.[7]

g) The relative performance of the DS model compared to the ST model is reported in fig. 3 (right picture), in which the distinction between frequent shocks with small variance (dashed line) and infrequent shocks with large variance (solid line) is particularly visible. It is quite clear from the picture that under the segmented trend hypothesis the DS provides imprecise indications about the dates of the breaks (for example the true signal of the series is shifted one period forward).

---

[6] Measures of persistence V_k and A_k are measures of trend reverting behaviour of the series, based on the sample autocorrelations of the process in first differences. k is the number of autocorrelations included to investigate the degree of trend reverting. The interested reader is referred to Campbell and Mankiw (1989) for precise definitions.

[7] On the one hand, this is completely in line with Balke and Fomby's proof that standard unit root tests cannot discriminate between frequent shocks with small variance and infrequent shocks with large variance; on the other hand the result is not surprising at all once it is considered how unit root tests were originally designed to cope with the alternative hypothesis of a *global*, but not local, linear time trend.

h) Analysing the residuals generated by the DS model, the density estimate reveals (see fig. 5, centre picture) a certain departure from the normality assumption. Jarque and Bera test statistic is 5.30, against 0.59 of the ST model, with 5% critical value of 5.99. Other residual-based regression diagnostics also penalize the DS model with respect to the ST model: the adjusted R^2 is 0.43 against 0.93, the log-likelihood 29.58 against 122.92, the Akaike information criteria -47 against -226, the Schwarz selection criteria -7.38 against -9.23.

i) Plotting the residuals of the different models (fig. 4), and estimating the corresponding density distributions (fig. 5), the conclusion that the structure missed (captured) by the systematic components of the different representations is expected to be found (not be found) in the error component is fully confirmed by the density estimates. Only in the ST case residuals are completely depurated from the 'signal' (the trend component) of the series.

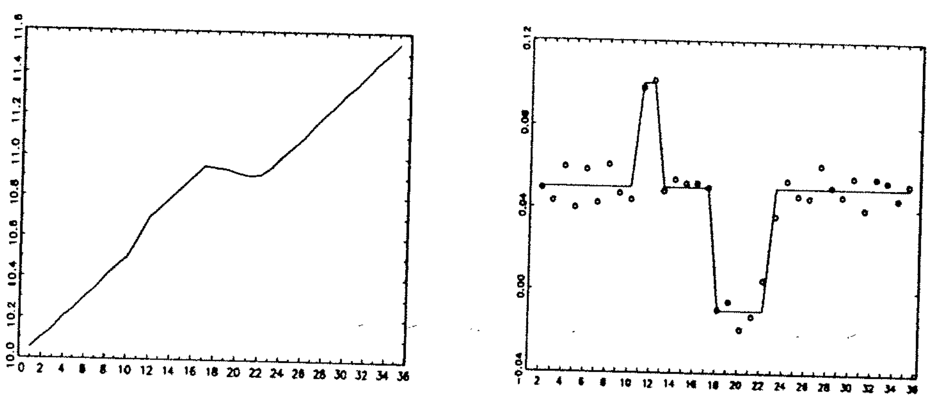

Fig. 1. Left: Simulated series in levels. Right: Simulated series in first differences. Note: the error term of equation (4) has been imposed normally distributed with zero mean and variance 0.0002, within each separate regime.

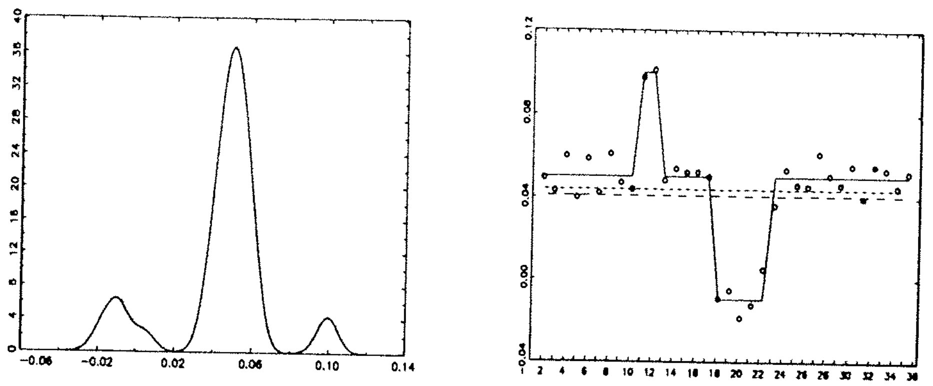

Fig. 2. Left: Density estimate of the simulated series in first differences. Smoothing parameter 0.005 for a Gaussian kernel, selected using Sheater and Jones' (1991) method. Right: Simulated series (circles), true ST structure (solid line), and structure suggested by the RW (dashed line) and the TS model (dotted line). Note: the structure suggested by the TS and the RW models are just the plots of $\hat{\mu}_1$ and $\hat{\mu}_2$, estimated by running respectively the regression equations (1) and (2).

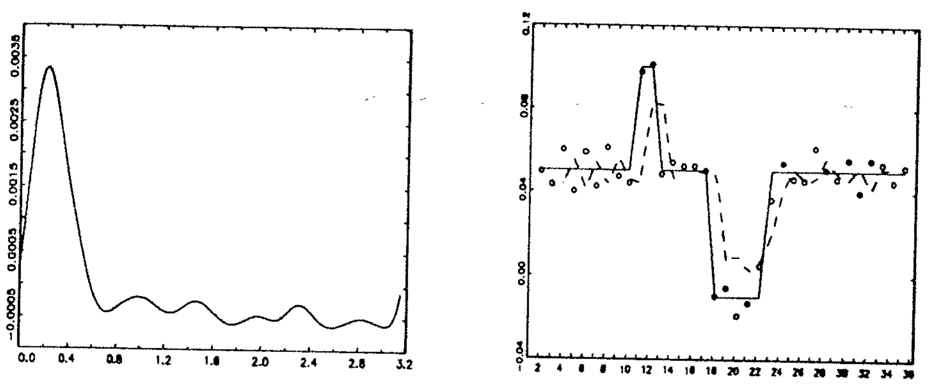

Fig. 3. Left: Spectral estimate of the simulated series in first differences computed by averaging the periodogram through a moving average of order 7. Right: Simulated series in first differences (circles), true ST structure (solid line), and structure suggested by the DS model for $p=2$ (dotted line).

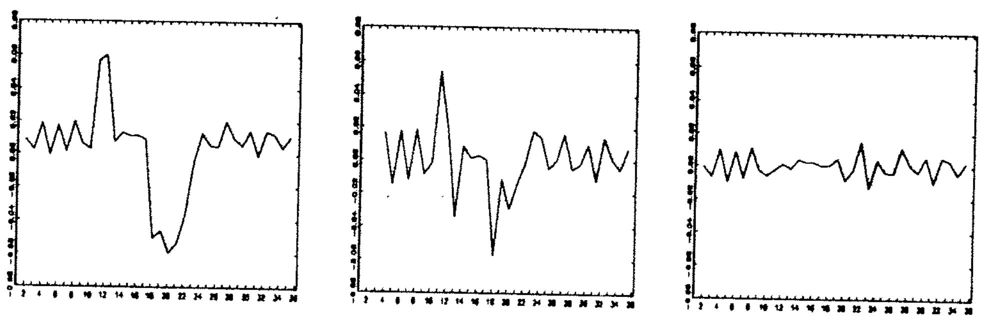

Fig. 4. Graph of the residuals derived as the difference between actual and fitted, with respect to the simulated series in first differences. In order from the left to the right: TS, DS, and ST representations.

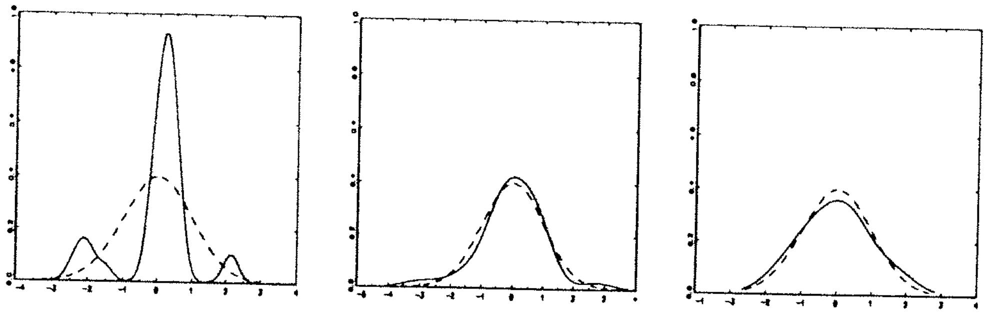

Fig. 5. Density estimates of the *standardized* residuals (solid line), compared with a normal density having zero mean and unitary variance. In all cases, smoothing parameter selected through the better rule of thumb criteria as defined in Härdle, 1991, p.91, or Silverman (1986), section 3.4. (0.21,0.52,0.51). In order from the left to the right: TS, DS, and ST representations.

3. ST, DS models, and signal extraction: A comparison with the nonparametric approach

TS and DS representations as defined in section 2 look both as parametric models, but they can be given a nonparametric interpretation which may reveal interesting for the discussion of signal extraction. On the basis of such an interpretation, in particular, it is possible to clearly point out why DS representations, despite their wide popularity in current empirical research, are likely to produce quite misleading results when adopted in presence of structural breaks. At the same time, it is also emphasized the nonparametric nature of the segmented trend model, in the sense of a *constrained* nonparametric regression.[8]

To introduce this issue, it is useful to rewrite the series in first differences (growth rates) under the following notation:

ST: $$\Delta y_t = \mu_i(t) + \text{error}, \tag{9}$$

DS: $$\Delta y_t = \mu(t) + \text{error}, \tag{10}$$

[8] With the speed actually available from modern computers, nonparametric techniques have got more and more attention from both theoretical and applied statisticians. In many cases, these techniques have produced quite satisfactory results, but in time series the diffusion of such methods have encountered some relevant problems. The basic problem has been that in time series few observations are usually available, and usually not satisfying the assumption of indenpendence and identical distribution. The logic of *local averaging*, then, which is indeed the essence of any nonparametric method, has revealed in time series much more costly than elsewhere. There are however signs of developments in the literature concerning nonparametric applications in time series, in which the problem of the poverty of sample information is balanced through the introduction of more effective constraints. The segmented trend model is just an example of this, as it represents the attempt to estimate the systematic component of the model as an *unknown* function of time, which is restricted however to be a *step function* of time. Under the hypothesis of smooth signals, a general approach has been recently developed by Rodriguez-Poo (1992).

where $\mu_i(t)$ i=1,2,...,m and $\mu(t)$ = const. + $\Sigma\, \gamma_j\, \Delta y_{t-j}$, j=1,2,...,p represent the *fitted* series based on the estimates of the two parametric models. It is convenient to interpret the fitted series of a particular regression model as the signal extracted by that specific model. It is then clear from fig. 3 (left picture) how both ST and DS representations extract a signal which is a nonlinear function of time, with the only difference the nonlinear function is a discrete function of time in the ST case, and a continuous function of time in the DS case - subscript i in equation (9) stands then to emphasizes discreteness. In the DS case, in particular, the fitted series is just a moving average of order p, with weights given by OLS estimates of the γs.

Consider now the nonparametric approach for a direct signal extraction, in place of the indirect signal extraction through estimates former derived by a parametric regression. The uncontroversial advantage of nonparametric regression with respect to the simple linear regression model is that in most practical situations the assumption of linearity may reveal inadequate, what makes more appropriate to let the data show the effective functional form. The basic idea behind nonparametric regression is indeed the idea of *scatterplot smoothing* (see Hastie and Tibshirani, 1990, chapters 1 and 2), where the mean dependence of a response variable y on a predictor variable x is modelled according to the regression

$$y_j = m(x_j) + \eta_j, \quad \eta_j \sim iid\left(0,\sigma_\eta^2\right) \;\; j = 1,2,\ldots,T,$$

where $\{y,x\}_j$ is a sample of independent and identically distributed (*i.i.d.*) observations, and $m(x_j)$ an arbitrary *unspecified* function. A *smoother* is defined in this framework as an estimate of $m(x)$ less variable than y itself, but capable of well fitting the scatterplot of the two variables.

Technically speaking, the problem of smoothing is determined by identifying a smoother matrix $\boldsymbol{S}$ that applied to the observations vector $\boldsymbol{y}$ is able to produce the smoothed curve in accordance to the relation $\boldsymbol{m}$=$\boldsymbol{Sy}$ (in which case the smoother is said to be linear). Being **S** a matrix of weights, nonparametric regression is just a procedure for *local averaging* the response variable at any given neighbourhood of the predictor variable.

Despite well-established methods such as *regressograms, running-lines, running medians, kernel smoothers, k nearest neighbourhood smoothers, cubic splines*, etc. are known in the nonparametric literature, which are employed in general with satisfactory results,[9] the main problem in time series is that observations usually do not match the *i.i.d.* assumption. This makes that 'boundary' observations (that is observations in between different regimes) are averaged sometimes which should not be averaged in fact for the 'true' signal

[9] See Hastie and Tibshirani (1990) and Härdle (1990*b*) for a good survey of the different possibilities, together with all major technical details.

of the series not to be masked.

To realise this point, take for an example the simulated series of section 2. The predictor variable is time, and the scatterplot is simply the plot of the series. Fitting a regressogram, a *k*-NN smoother, or a running median, all the results reported in fig. 6 indicate the difficulty of the different smoothers to extract the true signal out from the actual time series. For the same reason, the DS model, which is in fact a real nonparametric model in the form of a weighted moving average, is also badly performing in presence of structural breaks (remember the signal extracted for the simulated series, in fig. 3 - right picture).

A natural solution to prevent the problem of averaging observations generated by different regimes is to use nonparametric regression only after *previous* information has been gained about the number and the location of knots (change points). This can be interpreted as a constrained nonparametric regression, or piece-wise regression with known location of knots, and the goodness of such a strategy is particularly clear for economic time series, under the hypothesis of *infrequent* permanent shocks. This is the reason why we would consider in the next section a method for estimating the number and location of change points.

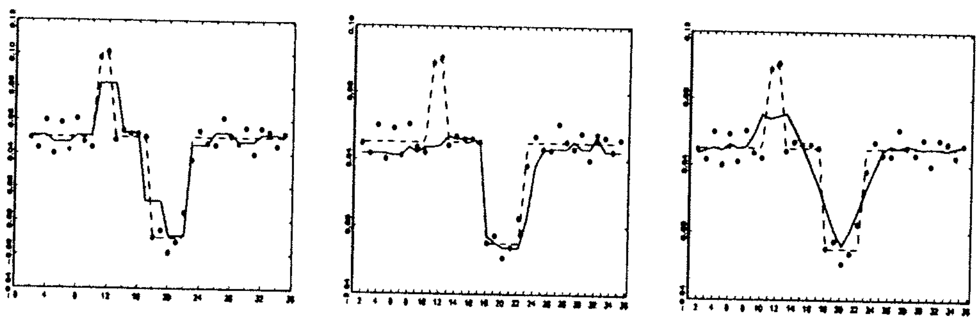

Fig. 6. Scatterplot of simulated series in first differences against time (circles) and nonparametric regression (solid line). Dashed line is the true signal. Left: Regressogram of order 3. Centre: Running-median of order 4. Right: *k*-Nearest Neighbourhood of order 4.

4. Regression strategy for the detection of change points

There is an old tradition in the statistical literature concerning how to fit segmented straight lines to a time series, which initiated on the *Journal of the American Statistical Association* with early developments of computer capabilities. Example of this literature are the contributions given by Quandt (1958,1960), Bellman and Roth (1969), McGee and Carleton (1970), in which the change point problem has been investigated from the maximum likelihood viewpoint. Subsequently, the change point problem has been also considered from the nonparametric and the Bayesian viewpoint.

We consider in this section an incidentally revised version of the method recently developed by Kashiwaghi (1991), in which a Bayesian approach is used to make inference about the number and location of change points in time series. In Kashiwagi (1991), the predictive log-likelihood is used as an estimate of the likelihood fuction conditional to a given number and location of change points, and inference about the number and the location of change points themselves is based on posterior probabilities, which are derived by combining the sample information with *flat* prior probabilities, through the Bayes theorem (see Kashiwagi, 1991, section 2). Prior probabilities are assigned to the marginal event that the series has $N=n$ change points and to the conditional event (conditional to $N=n$) that change points are located at $j(1), j(2), ..., j(n)$.

Despite the assumption of flat priors may appear neutral, it turns out that the detection of outlier observations as change points observations is heavily penalized within such a Bayesian analysis. For an intuitive understanding of this, consider in fact the case where there is only a change point ($n=1$); flat priors about the location of the change point means an equal probability is assigned a priori to each combination, i.e. the change point has the same probability to occur at observation no. 3, or, say, observation no. 7, no. 13, etc. Consider however the case of 5 change points; with respect to the case $n=1$, this case is generating a much greater number of combinations, most of whom are unlikely to be 'reasonable' combinations at all - for example all combinations of the type 1,2,3,4,5; 2,3,4,5,6 etc. Imposing the same prior probability to all the combinations is going therefore to penalize the detection of change points for higher values of n. This is why we take advantage in the following of the regression strategy proposed by Kashiwagi for the estimation of likelihoods, but do not make inference upon the number and the location of change points by elaborating the sample information in a Bayesian way.

4.1 Short review of Kashiwagi's statistical model

Consider the sequence $y_1, y_2, ... y_T$ of observations at equally spaced intervals, and define a to be the joint event that the sequence $y_1, y_2, ... y_T$ has $N=n$ change points at $j(1), j(2), ... j(n)$. The sequence of random variables $Y_1, Y_2, ..., Y_T$ is said to satisfy event a if and only if the density of $\mathbf{y}=(y_1, y_2, ... y_T)'$ can be factorized in the form:

$$p(y \mid a; \theta_0, \ldots, \theta_n, \sigma^2) = \prod_{i=0}^{n} p_i(y_i \mid \theta_i, \sigma^2)$$

$$= \prod_{i \in I_1} p_i(y_i \mid \theta_i, \sigma^2) \cdot \prod_{i \in I_2} p_i(y_i \mid \theta_i, \sigma^2) \qquad (11)$$

where $y_i=(y_{j(i)+1},\ldots,y_{j(i+1)})'$, $j(0)=0$, $j(n+1)=T$; $k_i=\dim(y_i)$, $I_1=\{i \mid k_i=1\ 0\le i\le n\}$ and $I_2=\{i \mid k_i\ge 2\ 0\le i\le n\}$; $p_i(y_i \mid \theta_i,\sigma^2)$ is the density of y_i with parameter θ_i and σ^2 (Kashiwagi, 1991, p. 77). Under the hypothesis of n change points, the integrated likelihood of a is given by

$$p(y \mid a)=\int\cdots\int p(y \mid a,\theta)\omega(\theta)d\theta,$$

where $\theta=(\theta_0,\theta_1,\ldots,\theta_n,\sigma^2)$, and $\omega(\theta)$ is a prior density for θ. The value of $p(y \mid a)$ can be obtained theoretically at the cost of specifying the prior density $\omega(\theta)$. In order to avoid such a difficult specification, Kashiwagi suggests to approximate it by the exponential of the predictive log-likelihood, defined as the maximum log-likelihood corrected for a bias term. For the switching regression model described by equation (4), what is called by Kashiwagi "*the simple regression model*", the maximum likelihood estimates of the parameters $\theta_i=(\alpha_i,\mu_i)'$ and σ^2 are given by:

$$\hat{\alpha}_i=y_{j(i)+1} \qquad i\in I_1$$

$$\hat{\theta}_i=(\mathbf{A}_i'\mathbf{A}_i)^{-1}\mathbf{A}_i'\mathbf{y}_i \quad i\in I_2$$

$$\hat{\sigma}^2=\frac{1}{T}\sum_{i\in I_2}\hat{\varepsilon}_i'\hat{\varepsilon}_i;$$

where

$$\hat{\varepsilon}_i=\mathbf{y}_i-\mathbf{A}_i\hat{\theta}_i: \quad \mathbf{y}_i=\begin{bmatrix} y_{j(i)+1}\\ y_{j(i)+2}\\ \cdot\\ y_t\\ \cdot\\ y_{j(i+1)}\end{bmatrix}, \quad \mathbf{A}_i=\begin{bmatrix}1 & j(i)+1\\ 1 & j(i)+2\\ \cdot & \cdot\\ 1 & t\\ \cdot & \cdot\\ 1 & j(i+1)\end{bmatrix}.$$

The maximum log-likelihood is given by

$$\log p(y \mid a;\hat{\theta},\hat{\sigma}^2)=-\frac{T}{2}\log(2\pi\hat{\sigma}^2)-\frac{T}{2}$$

which has however the bias

$$Bias: = \log p(y \mid a;\hat{\theta},\hat{\sigma}^2) - \mathrm{E}_Z \log p(Z \mid a,\hat{\theta},\hat{\sigma}^2) =$$

$$= \frac{1}{2}\left[\frac{T\sigma^2}{\hat{\sigma}^2} + \sum_{i \in I_1}\left(\frac{\alpha_i - \hat{\alpha}_i}{\hat{\sigma}}\right)^2 + \sum_{i \in I_2}\frac{(\theta_i - \hat{\theta}_i)' A'_i A_i (\theta_i - \hat{\theta}_i)}{\hat{\sigma}^2}\right] - \frac{T}{2}$$

(Kashiwagi, 1991, p. 86), where E_Z denotes the expectation conditional to Z. Because of the bias term, the maximum log-likelihood has a tendency to overestimate the number of change points. The expected bias is obtained by taking the expectation with respect to Y as

$$\mathrm{E}_Y(Bias) = \frac{T(T + m_1 + 2m_2)}{2(\sum_{i \in I_2} k_i - 2m_2 - 2)} - \frac{T}{2} = BC - \frac{T}{2},$$

where m_1, m_2 are the number of elements included in the set I_1 and I_2. (see Kashiwagi, 1991, p.86). Subtracted from the maximum log-likelihood, the approximation of the predictive log-likelihood is given by

$$\log p^{\text{pred}}(y \mid a) \approx -\frac{T}{2}\log(2\pi\hat{\sigma}^2) - BC. \qquad (12)$$

From the predictive log-likelihood, the likelihood $p(y \mid a)$ is finally derived by taking the exponential

$$p(y \mid a) = \exp\{\log p^{\text{pred}}(y \mid a)\}. \qquad (13)$$

For any given a combination of N=n change points, an high score of the maximum likelihood $p(y \mid a)$ means in general a small value of the estimated variance, that is a good fitting of that particular combination to the actual series. The predictive log - likelihood, however, it is also a negative function of the BC term, what implies a trade-off between goodness of fit and number of estimated parameters, as long as the BC term is an increasing function of the number of change points ($\partial BC/\partial m_1 > 0$ and $\partial BC/\partial m_2 > 0$). Then, a greater number of change points not necessarily imply an higher value of the likelihood function.

4.2 Using the shifting mean value model as the regression model

There is a problem with the simple linear regression model that suggests not to consider it in applications. The problem is the failure of the model to precisely detect the location of change points, what we call hereafter the 'split decision' problem. In order to understand this, consider the simple case in which the time series is just the deterministic sequence y_t={1,2,3,4,7,9,11,13,15,17}. This sequence is growing of one unit up to observation no. 4, then has a jump, and is growing of 2 units afterwards: Hence, it only contains a change point in correspondence of t=4.

Implementing the Bayesian procedure for detection of change points based on the simple linear regression model, within the case n=1 the combinations $j(1)$=3 and $j(1)$=4 associate the same predictive log-likelihood. The predictive log-likelihood is in fact only a function of $\hat{\sigma}^2$ and the *BC* term, and both combinations $j(1)$=3 and $j(1)$=4 associate the same $\hat{\sigma}^2$ and *BC* term. It is in this sense the simple regression model associate therefore a 'split decision', due to the fact that when local linear trends are fitted for different combinations of change points, the likelihood may be approximatively the same when the change point is fixed at the last observation of the previous regime, or at the first observation of the next regime.

For our simulated series, arising of the split decision problem is confirmed by the results presented in the upper part of table 1, where the top predictive log-likelihoods are reported for n=0,1,...,5.[10] Indeed, despite the highest top predictive log-likelihood is found for n=4, that is a correct inference is actually drawn on the number of change points, the location of the change points is not precisely detected. True change points are located in fact at t=10,12,17,22, whereas highest frequencies of occurrency are found for observations no. 10,11,16,22.

A natural way to avoid the split decision problem is to model the series in growth rates, rather than in levels: In this case jumps in the series are necessarily unmasked, that is estimated variances for different combinations should associate remarkably different likelihoods. For this reason, we consider in the following the ST model in the form of the *shifting mean value model* (equation (8) of section 2). This model associates the maximum likelihood estimates

$$\hat{\mu}_i = y_{j(i)+1} \qquad i \in I_1$$

$$\hat{\mu}_i = \text{mean}(y_i) \qquad i \in I_2$$

and the *BC* term

$$BC = \frac{T(T+m_1+m_2)}{2(\sum_{i \in I_2} k_i - m_2 - 2)},$$

[10] All the results derived in this paper are obtained through GAUSS programming on IBM-PC compatible machines. The codes of the programs are available on request from the author.

while estimate of the variance term is the same as for the simple regression model. The results in the bottom part of table 1 clearly confirm the good performance of the shifting mean value model. The highest predictive log-likelihood is found in correspondence of n=4, and the observations with higher frequency of occurrency are exactly the true change points of the series (remember first differencing implies the first observation is lost, so that a change point at observation no. 10 for the series in levels is a change point at observation no. 9 for the series in growth rates).

The top predictive log-likelihood associated with the case n=5, however, is also high, so that we might wish to check the performance of Kashiwagi's approximation by testing the hypothesis of the significance of an additional change point. This can be easily done by using the likelihood-ratio principle. The shifting mean value model can be written in fact as a standard linear regression model

$$y = X\mu + u$$

where X is a $T \times n$ design matrix of 0's and 1's. Under the null hypothesis of n structural breaks, the parameter space is $\Omega_0 = \{\mu_1, \mu_2, \ldots, \mu_{n+1}\}$, while under the alternative hypothesis of n+h structural breaks we have $\Omega_1 = \{\mu_1, \mu_2, \ldots, \mu_n, \mu_{n+1}, \ldots, \mu_{n+h+1}\}$. Estimating the model under both the null and the alternative hypothesis,[11] under the hypothesis of normality of the errors the maximum log-likelihoods l_0^* and l_1^* can be compared, and the likelihood ratio statistic computed as

$$LR = 2\left(l_1^* - l_0^*\right) \overset{H_0}{\sim} \chi^2_{(h)}, \quad \text{for } T \to \infty$$

When h additional change points are considered which are in fact 'not significative' change points, the likelihood of the model should not improve significantly, i.e. values of the LR statistic should not be significantly different from zero. In the case the hypothesis of 4 against 5 change points is tested for the simulated series, the likelihood ratio statistic is as low as 0.49 (against a 5% critical value of 3.84), that is the significance of observation 20 as a change point is highly rejected by the data.

To summarize, our procedure for change point detection is as follows: Given the whole set of all data points, this can contain in principle all as much as no change points. Through the regression analysis we cluster a subset of

[11] As far as X is by definition such that $(X'X)^{-1}$ is always a singular matrix, estimates of the μ's are derived by computing the Moore-Penrose generalized inverse.

observations that, more than others, may be suspected of being change points in the series. Then, we use the likelihood ratio testing strategy as a validatory or stopping rule procedure.

Table 1: Top predictive log-likelihoods for the simulated series.

levels	max pred	$j(1)$	$j(2)$	$j(3)$	$j(4)$	$j(5)$
n=0	4.55					
n=1	15.12	19				
n=2	22.18	16	22			
n=3	32.03	10	16	22		
n=4	**38.92**	10	11	16	22	
n=5	38.84	10	11	17	21	22

first diff.	max pred	$j(1)$	$j(2)$	$j(3)$	$j(4)$	$j(5)$
n=0	21.41					
n=1	21.86	16				
n=2	28.94	16	21			
n=3	28.91	9	16	21		
n=4	**36.21**	9	11	16	21	
n=5	35.83	9	11	16	20	21

Table 2: True structure and structure suggested by top predictive log-likelihood estimates

$j(i)$	**10**	**11**	**16**	17	19	21	**22**
freq.	3	2	3	1	1	1	4
true $j(i)$'s	10	12	17	22			

$j(i)$	**9**	**11**	**16**	20	**21**
freq.	3	2	4	1	4
true $j(i)$'s	9	11	16	21	

5. An application to the U.S. series of nominal wages, 1900-1970

In the previous sections, the example provided by the simulated series has served as a benchmark case in order to understand some basic properties of ST processes, and to evaluate possible regression strategies for estimating the number and the location of change points.

In this section we consider a real observed economic time series, which is given by the U.S. series of nominal wages over the period 1900-1970 (see fig. 7), and apply to this series the methods presented in sections 2-4. The reason for examining U.S. nominal wages is that the same series was previously analyzed by Nelson and Plosser (1982) and Perron (1989), what favours fair comparisons of alternative modelizations. The aim is in particular to reconsider earlier results from the point of view of infrequent permanent shocks.

5.1 Attained empirical evidence

In the pioneristic work of Nelson and Plosser (1982), the hypothesis of frequent permanent shocks was implicitly assumed for the nominal wage series. A unit root test was performed on the series, and the null hypothesis of a DS process with $p=2$ was accepted against the alternative hypothesis of a TS process (t-ratio statistic of -2.09 against a critical value of -3.45).[12] In Perron (1989), on the other hand, the hypothesis of infrequent shocks was considered for the first time in the form of a shifting trend model with one break in the trend at time $t=1929$, by estimating the regression equation $y_t = \alpha + \beta\, DU(t) + \mu\, t$ + error, with $DU(t) = 0$ if $t \leq 1929$ and $DU(t) = 1$ if $t > 1929$. In fig. 9 and 10 the fits of both models are reported, and compared with the simplest case of trend stationarity about a global linear time trend (fig. 8).

A main comment about the results refers to the different modelization of the trend component for the same observed time series. It is seen in fact the TS representation models the trend in a very inflexible way, while in the DS model the trend is moving at *each* point of the time. Perron's ST model suggests an intermediate position such that the trend is moving, but only at $t=1929$. Interpreting the residuals in figures 8-10 as representing the cyclical components of the series, it is with no surprise the way cycles are derived directly depend on the way trends themselves are modeled. Then, in the TS case a very inflexible trend associates 'large' fluctuations about (but these fluctuations are spurious in fact), while in the DS an extremely flexible trend associates no fluctuations about. Perron's ST suggests once again an intermediate position, where the magnitude of fluctuations is in between the

[12] Measures of persistence also indicate large persistence of shocks being the pair $\{V_k, A_k\}$ = {1.42, 1.25} for $k=3$, {1.42, 1.25} for $k=5$, {1.52, 1.30} for $k=7$, and {1.50, 1.28} for $k=10$.

two extremes.

Estimating the spectrum of the nominal wage series in first differences, and the density distribution, the inappropriateness of the TS and the RW representations is revealed in fig. 11. The DS process identified by Nelson and Plosser (1982), as well as the shifting trend suggested by Perron (1988), are however also likely to be problematic modelizations of the true DGP. In Perron (1988), in particular, μ is estimated to be 0.051; since fitted values between 1929 and 1930 associate a growth rate of -0.057, Perron's shifting trend implicitly assumes for Δy_t a bimodal density with a big mode centred around 0.051 and a smaller mode centred around -0.057, which is not confirmed by our density estimation. This suggests almost for sure a deterministic trend with only one break is not flexible enough to avoid the appearance of spurious cycles.[13]

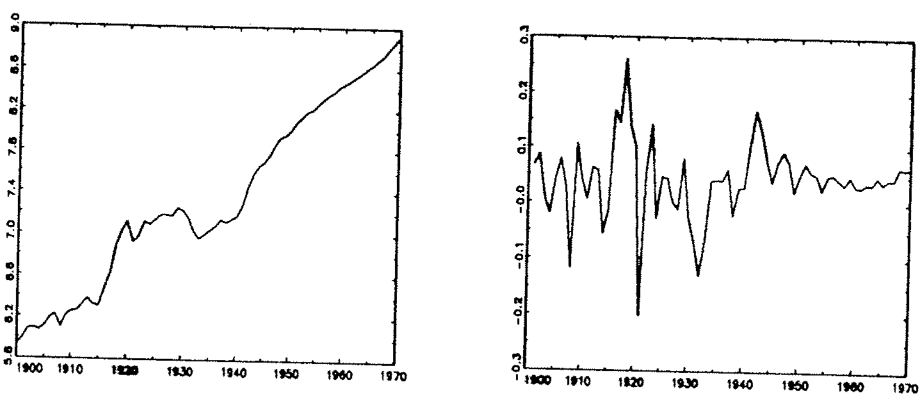

Fig. 7. Left: Natural logarithm of nominal wages. Right: growth rates of nominal wages.

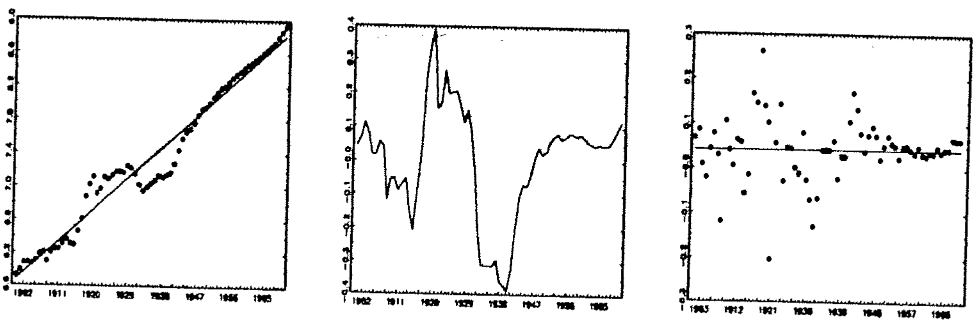

Fig. 8. TS modelization of the nominal wage series (left) and associated residuals (centre). Right: Implication of the TS model for the series in first differences.

[13] Indeed, in Perron ST model as in the pure TS case, regression diagnostics indicate an high value of the R^2 (0.98 - 0.96 in the TS case) associated with a low value of the Durbin-Watson statistic (0.48 - 0.16 in the TS case), that is the typical symptoms of spurious regressions according to Granger and Newbold (1974).

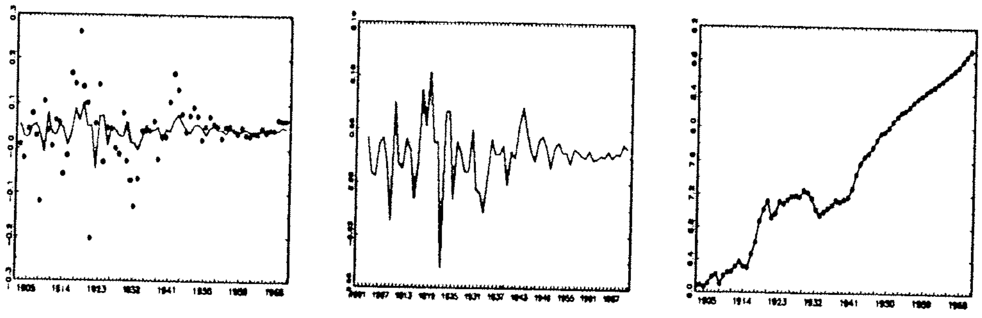

Fig. 9. Nelson and Plosser modelization of a DS process with $p=2$ (left) and associated residuals (centre). Right: Implication of the DS model for the series in levels. Note: the fitted series in levels has been reconstructed from the series of growth rates derived by fitting the DS model, considered the initial condition obtained as the arithmetic mean of the first two observations of the nominal wage series in levels.

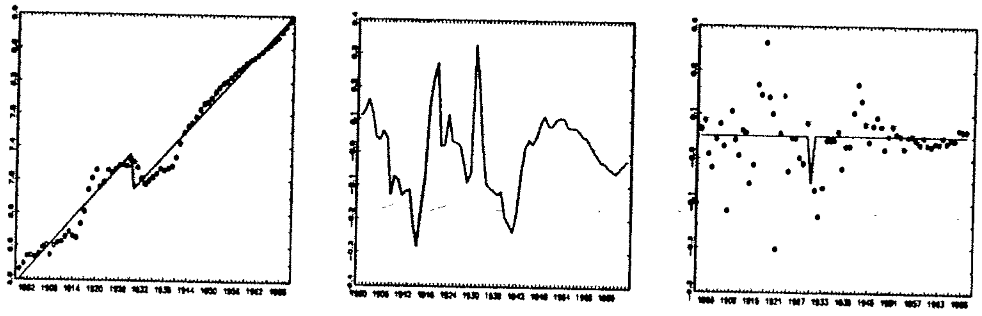

Fig. 10. Perron's ST model with break in 1929 (left) and associated residuals (centre). Right: Implication of the ST model for the series in first differences.

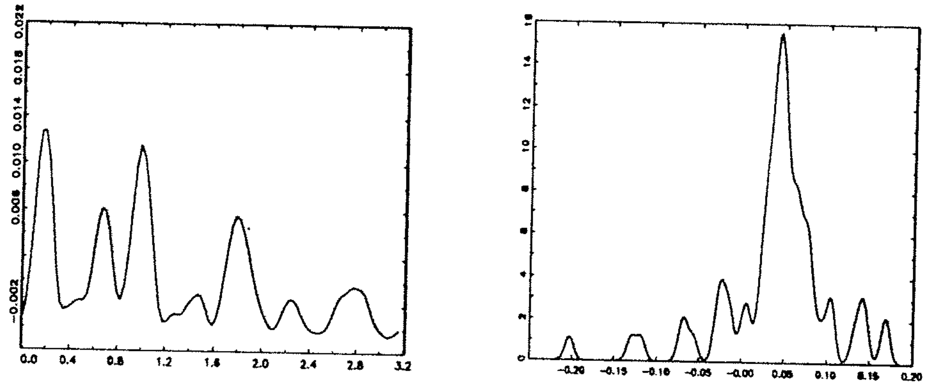

Fig. 11. Left: Spectral estimate of the nominal wage series in first differences computed by averaging the periodogram through a moving average of order 5. Right: Density estimate of the nominal wage series in first differences (standardized). Smoothing parameter computed by Sheater and Jones' (1991) method: 0.0052 for a Gaussian kernel.

5.2 Infrequent shocks and change point detection

According to the hypothesis that shocks occur infrequently, a running mean smoother can be estimated for the nominal wage series in first differences, after the detection of change points. A visual inspection of the density estimate clearly suggests the appearence of 9 modes, that is the presence of *at least* 9 change points (10 regimes) in the growth rates of the nominal wage series. The growth rate occurring more frequently is without doubt about 5% per year.

Before we proceed towards change point detection, an important remark concerns the best conditions under which the statistical model presented in section 4 correctly performs. Indeed, Kashiwagi's procedure successfully detected both the number and location of change points in the simulated series, but because the assumption of homoschedasticity among different regimes was respected. The nominal wage series, however, clearly displays a bigger variability in the first period of the sample, compared to the second period (see fig. 7). A *preliminary* analysis of variance patterns is therefore recognized, in order to possibly identify a cut-point in the sample.[14]

In fig. 12, two plots of variability for the series in first differences are reported. The former is obtained by graphing the square of the distance between any observation and its right-hand neighbour (first sample observation is lost); the latter is derived by plotting the square of the distance between each observation and the mean value of the series. According to the first plot, the splitting date would be 1923, while from the second plot it is clear a much lower variability characterizes the series after 1933. In what follows the analysis is carried out by splitting at 1933 (i.e. analysing the subsamples 1901-1933, and 1934-1970), that is at the date of the through of the 'big crash'. Analoguous results has been found however splitting the sample at 1923.

The results of the regression analysis are presented in tables 3 and 4, for the subsamples 1901-1933 and 1934-1970 respectively. In the former period, suspect change points are found in correspondence of observations no. 7, 8, 15, 20, 21, 23, and 29, that is at time t=1907, 1908, 1915, 1920, 1921, 1923, and 1929. In the latter subsample, on the other hand, suspect change points are found in correspondence of observations 4, 5, 7, 10, and 15, that is at the historical dates of 1937, 1938, 1940, 1943, and 1948.

The results of the likelihood ratio tests reported in table 5 basically confirm the accurancy of the change point estimation, except for the date of 1937, which is insignificant as a change point at 5% confidence level. In the subsample 1901-1933, instead, change points at 1907 and 1908 are found to be quite

[14] Notice how selecting a cut point for the original series, not only ensures a better consistency with respect to the homoschedasticity assumption involved by our statistical model, but allows at the same time for the feasibility of the full computation. Indeed, the computational effort required for the series as a whole, say for a maximum number of 10 change points, is the estimate of 70!/(10!60!) = 4.72E+11 likelihoods, while splitting the sample heavily reduces the amount of computations.

significative, and the test statistic for the date of 1923 is very close to the rejection border too.

Based on the regression analysis, the likelihood ratio testing procedure, and the residual diagnostics obtained in the various cases, change points are imposed finally at 1907-1908, 1915, 1920-1921, 1923, 1929, 1933, 1940, 1943, 1948. Data suggests then the presence of 9 permanent shocks over 70 years of economic history about the U.S. nominal wage series - remember in fact despite there are 11 change points, whenever two change points are adjacent, they are generated by a single shock. This is on average an evidence of about one 'permanent' shock every 7 years.

All the dates are associated moreover with well identifiable economic events. In Friedman and Schwarz (1963), for example, October 1907 is referred to as the date of the "*banking panic, culminating in the restriction of payments by the banking system, i.e., in a concerted refusal by the banking system to convert deposit into currencies or specie at the request of the depositors*"; this was basically due to the "*reversal in gold movements from net imports to net exports*" (p.157). The "*contraction of 1920-1921*" (Friedman and Schwartz, p. 231) is also referred to as a depression caused by the banking System, as a consequence of the member banks decision to rise interest rates at "*the higher rate that has ever been imposed by the System, before or since*" (p. 233), in order to ensure profitability of their borrows. All the other dates are also relatively easily accounted for, being 1915 the date of the beginning of World War I, 1923 the peak of capital inflows (see for example the same Friedman and Schwartz, chart 17, p.201), 1929 the date of the 'big crash', 1933 the through of the Great Depression; 1940 the beginning of World War II, culminating in 1943 with the peak of industrial production (chart 45, p.547, in Friedman and Schwartz), and 1948 the date of the post-war era initiated with the Marshall plan and with other forms of financial and economic aids for the recovery of Western Europe economies.[15]

On the statistical front, given the above identification of change points, a segmented trend model (equation (4) of section 2) and a shifting mean value model (equation (8) of section 2) are estimated, yielding the signal extractions reported in fig. 14.

---

[15] Of course a much deeper analysis would be requested here from the viewpoint of institutional factors, economic history, economic theory arguments, etc. This would without doubt play an important role for a better understanding of the working of the economic system from an historical perspective, what would probably much involve authorities in charge of economic policies. But this is also something which goes beyond my specific task, which is in the following rather confined to the viewpoint of 'data-analysis'.

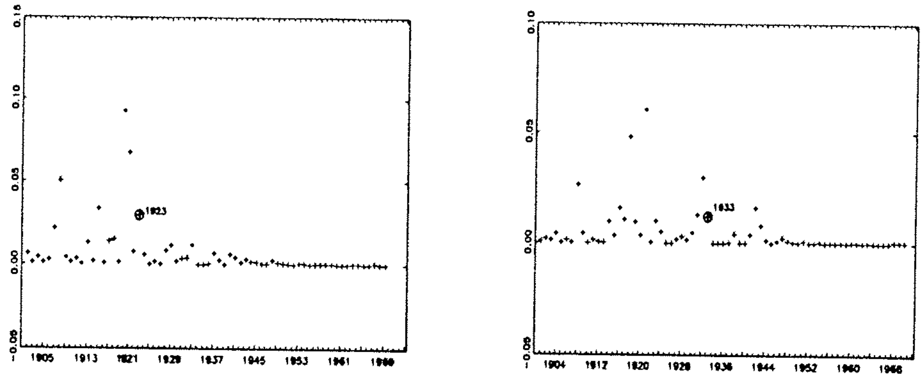

Fig. 12. Data variability plots for the U.S. series of nominal wages in first differences. Left: distances computed as the square of the difference between an observation and its right-hand neighbour (the first observation is lost). Right: distances computed in term of sample variances [i.e. $(y_t\text{-mean}(y))^2\ \forall\ t=1,2,\ldots,T$].

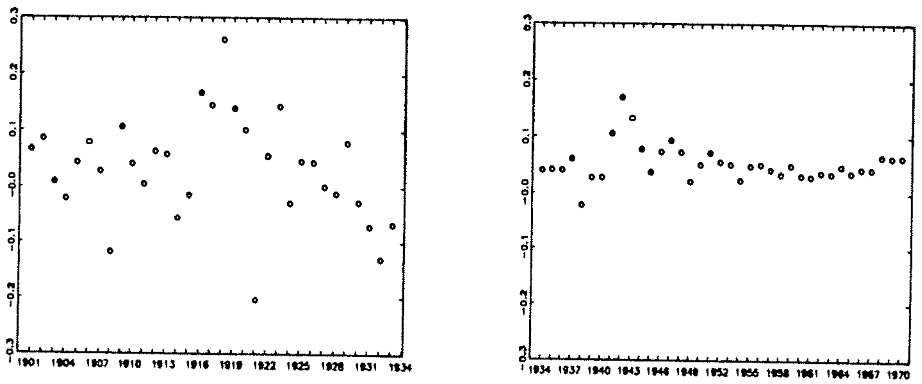

Fig. 13. Splitting the U.S. series of nominal wages in first differences. Left: subsample 1901-1933. Right: subsample 1934-1970.

Table 3: Top predictive log-likelihoods for the nominal wage series, 1901-1933.

	max pred	$j(1)$	$j(2)$	$j(3)$	$j(4)$	$j(5)$	$j(6)$
n=0	2.37						
n=1	2.57	1929					
n=2	3.58	1915	1920				
n=3	4.12	1915	1920	1921			
n=4	**4.96**	1915	1920	1921	1929		
n=5	4.16	1915	1920	1921	1923	1930	
n=6	3.62	1907	1908	1915	1920	1921	1929

Frequency of occurrency of observations suspected to be change points:

$j(i)$	1907	1908	**1915**	**1920**	**1921**	1923	**1929**	1930
freq.	1	1	**5**	**5**	**4**	1	**3**	1

Table 4: Top predictive log-likelihoods for the nominal wage series, 1934-1970.

	max pred	$j(1)$	$j(2)$	$j(3)$	$j(4)$	$j(5)$	$j(6)$
n=0	19.7						
n=1	19.5	1940					
n=2	25.1	1940	1943				
n=3	25.4	1940	1943	1948			
n=4	**25.6**	1937	1940	1943	1948		
n=5	25.5	1937	1938	1940	1943	1948	
n=6	25.4	1937	1938	1940	1941	1943	1948

Frequency of occurrency of observations suspected to be change points:

$j(i)$	**1937**	1938	**1940**	1941	**1943**	**1948**
freq.	**3**	2	**6**	1	**5**	**4**

Table 5: Likelihood ratio tests on the significance of change points, for the nominal wage series.

H_0	H_1	LR
1915,1920,1921 1929,1933,1940 1943,1948	1915,1920,1921 1929,1933,1940 1943,1948 1907,1908	7.56* (5.99)
1907,1908,1915 1920,1921,1929 1933,1940,1943 1948	1907,1908,1915 1920,1921,1929 1933,1940,1943 1948 1923	3.62 (3.84)
1907,1908,1915 1920,1921,1923 1929,1933,1940 1943,1948	1907,1908,1915 1920,1921,1923 1929,1933,1940 1943,1948 1937,1938	1.66 (3.84)
1907,1908,1915 1920,1921,1923 1929,1933,1940 1943,1948	1907,1908,1915 1920,1921,1923 1929,1933,1940 1943,1948 1930	0.00 (3.84)

Note: In the third column, the LR statistic is reported together with the 5% critical value (in parenthesis). In the second column, the date(s) underlined is (are) the date(s) tested as statistically significative change point(s).

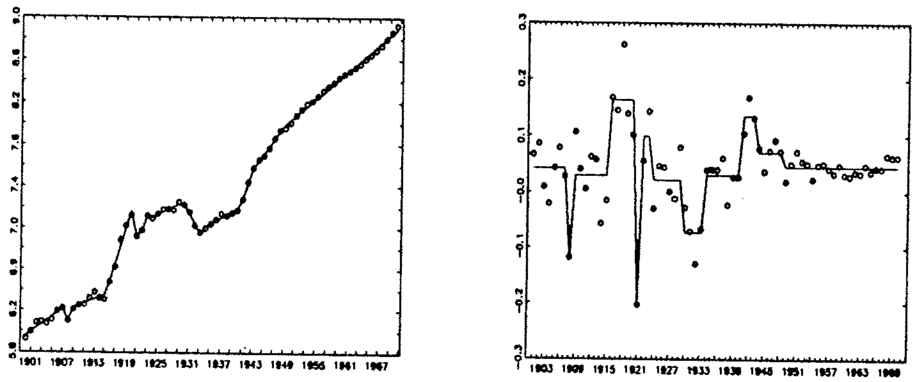

Fig. 14. Signal extraction on the nominal wage series in levels (left) and first differences (right).

5.3 Comparison of the results

The motivation for our analysis so far has been Perron's critique to standard unit root modeling. DS modelizations of time series extract the signal as a discontinuous, instead of a step function of time, that is they are by definition inconsistent with the assumption of infrequent permanent shocks. Nonetheless, the model fitted by Perron (1989) to the nominal wage series only imposes a structural break in 1929, while the same Perron advocates tests "*for structural changes in the trend function occurring at unknown dates*" (Perron, 1989, p. 1388). The above results are informative about the trend structure of the series, and also make use of the likelihood ratio test to validate the change points discovered by the regression analysis.

The results obtained under the hypothesis of infrequent permanent shocks can be compared now to existing modelizations. The comparison is carried out on the basis of the analysis of the residuals, i.e. according to the philosophy of residual diagnostic checking for the evaluation and validation of statistical models. In particular, the comparison is carried out with respect to Perron (1989) when analysing the residuals generated by fitting the series in levels, and with respect to Nelson and Plosser (1982) when analysing the residuals generated by fitting the series in first differences.

In the former case comparative plots are presented from fig. 15 to 17; in the latter, from fig. 18 to 20. In both cases, it can immediately be seen some major improvements are obtained, guaranteed by a more flexible modelization of the trend component of the series. For the segmented trend model with 12 regimes, in fact, residuals always appear white noise *and* normally distributed processes, while analogous good properties are not simultaneously associated with the competitive models. Specification diagnostics reported in table 6 validate the better fitting of the ST models, while estimates of the shifting mean rates, together with standard errors and t-values are finally reported in table 7. It is evident from all such results the superiority of the segmented trend model.

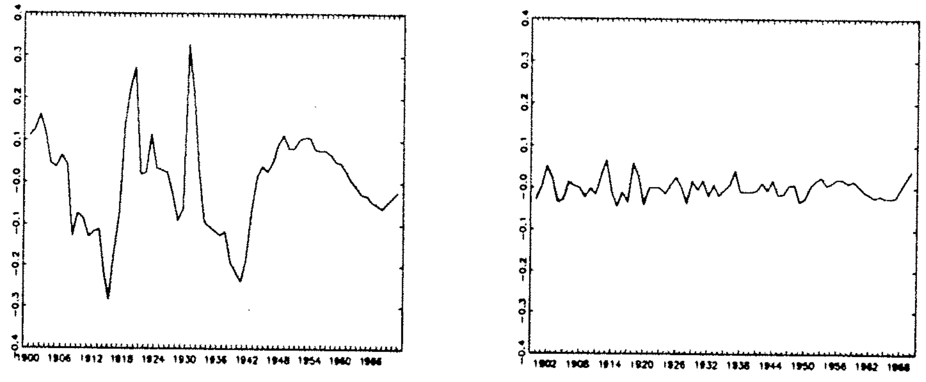

Fig. 15. Residuals generated by Perron's model (left), and by a ST model with m=12 local time trends (right), plotted in the same scale.

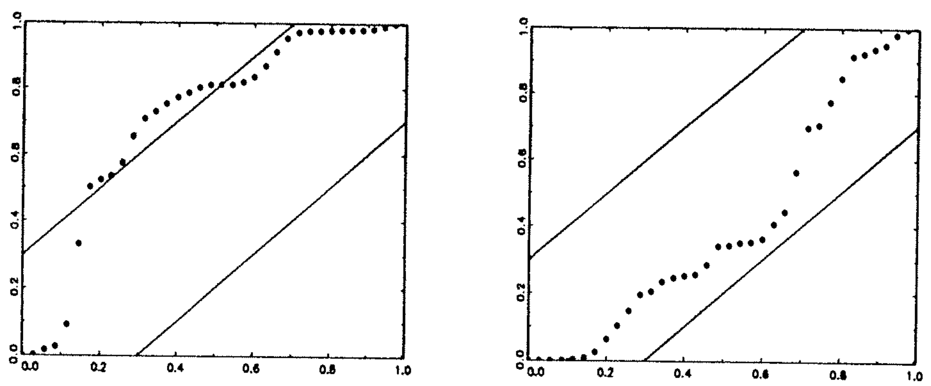

Fig. 16. Graphical test for white noise residuals. Cumulative periodograms with 95% confidence bands for the residuals generated by Perron's model (left), and by a ST model with m=12 local time trends (right).

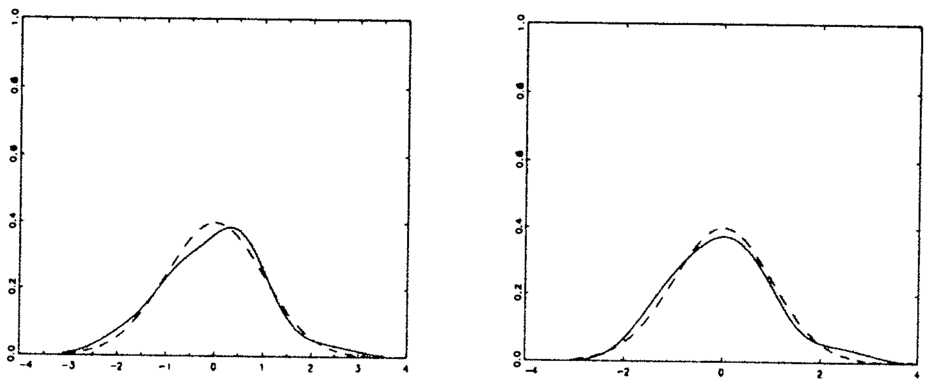

Fig. 17. Graphical inspection for normally distributed residuals: Kernel density estimation of the *standardized* residuals in the two cases (solid lines), compared with normal densities having the same sample means and variances. Bandwidth selected through the better rule of thumb, respectively 0.045 and 0.040.

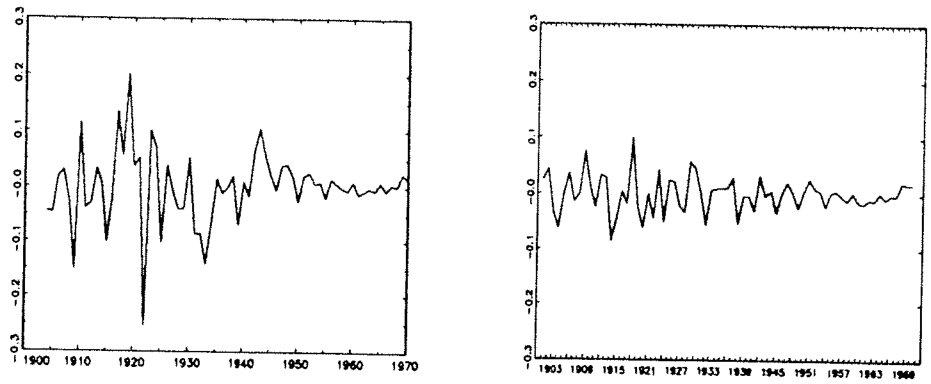

Fig. 18. Residuals obtained from a DS model with p=2 (left), and a ST model with m=12 local time trends (right), plotted in the same scale.

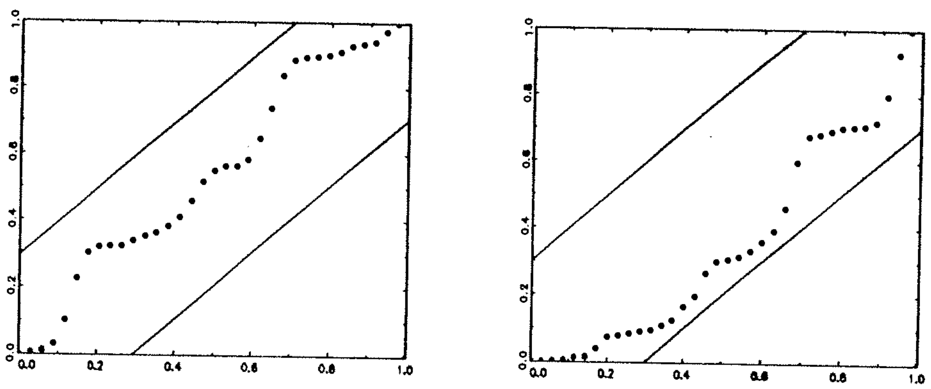

Fig. 19. Graphical test for white noise residuals. Cumulative periodograms with 95% confidence bands for the residuals generated by Perron's model (left), and by a ST model with m=12 local time trends (right).

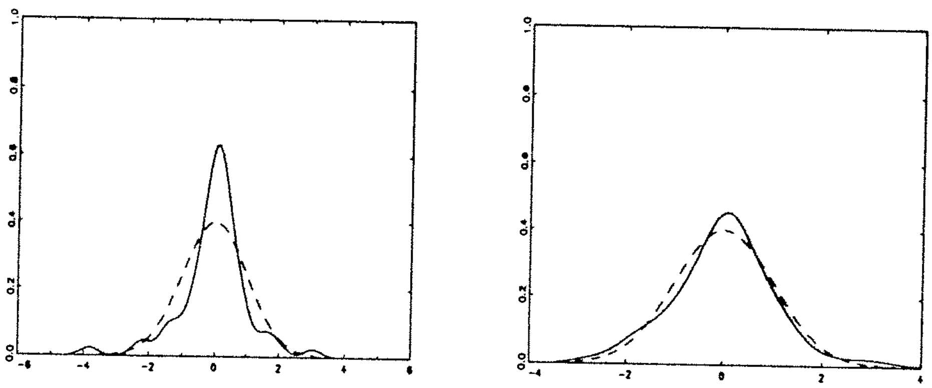

Fig. 20. Graphical inspection for normally distributed residuals: Kernel density estimation of the *standardized* residuals obtained from a DS model with p=2 (left), and a ST model with m=12 local time trends (solid lines), compared with a normal density having the same sample means and variances (dotted lines). Smoothing parameters: 0.26 and 0.41 respectively.

Table 6: Specification diagnostics for the evaluation of alternative models of the nominal wage series.

y_t	R^2	$\bar{R}^2$	log-lik.	AIC	SC	HQ	FPE	NT
ST (n=2)[a]	0.98	0.98	51.0	-90.1	-4.0	-21.1	0.016	0.05
ST (n=11)	0.99	0.99	166.6	-239.1	-5.3	-141.7	0.002	1.79
Δy_t								
DS (p=2)[b]	0.10	0.07	21.8	-31.6	-5.18	-22.1	0.005	32.5
ST (n=11)	0.78	0.74	141.9	-235.8	-5.79	-73.7	0.002	1.94

Note:[a] = Perron (1989) shifting trend model; [b] = Nelson and Plosser (1982) difference stationary model. AIC = Akaike information criteria; SC = Schwarz selection criteria; HQ = Hannan-Quinn selection criteria; FPE = final prediction error criteria. NT = Jarque and Bera normality test (5% critical value is 5.99).

Table 7: Regression results for the nominal wage series.

Period	μ_i	s.e.	t-value
1901-07	0.042	0.005	8.33
1908	-0.118	0.035	-3.38
1909-15	0.029	0.005	5.83
1916-20	0.163	0.007	23.9
1921	-0.203	0.035	-5.8
1922-23	0.099	0.018	5.68
1924-30	0.022	0.006	3.78
1931-33	-0.073	0.009	-8.38
1934-40	0.031	0.005	6.13
1941-43	0.136	0.012	11.6
1944-48	0.071	0.007	10.2
1949-70	0.045	0.002	28.5

6. ST in the multivariate domain: implications for prediction and cointegration theory

On the statistical front, whether economic time series are better characterized by ST rather than DS processes has intuitive implications for prediction and cointegration theory. For prediction purposes, for example, the signal extraction recommended by the shifting mean value model in the form of a step function of time represents a foremost information for the evaluation of the *current* natural rate. Take for an example an hypothetical economy which was growing about a (natural) rate of 5% per year up to ten or five years ago, but which is growing since then at a smaller (natural) rate of 2%: Such an economy cannot be expected to grow at 5% in the next period, *unless 'exceptional' events occur*. Past observations, as much as they were generated by a different regime, do not provide at all useful information to make better predictions under the *current* regime (i.e. under the current economic environment).

It is however about cointegration theory that ST representations may have interesting applications in macroeconomics, in order to check for co-movements or common trends in multiple time series. For this reason, we consider in the remaining of this section a concrete example in which the post-war nominal wage series is analysed together with the consumer price index series. An economist who wishes to explore the theoretical hypothesis that a close (positive) relation exists in the long run between growth in nominal wages and the cost-of-living, but who has at the same time the feeling that the dynamics in these series were driven by infrequent exogenous shocks, might proceed in accordance to the following strategy:

i) according to the hypothesis that shocks occurred infrequently, consider every single series separately, and extract the signal using the method described in sections 2-5. Two segmented trends would be recovered on the basis of the available data set;

ii) compare the dates of structural breaks, and, in the case dates of permanent shocks roughly coincide in the two series, interpret this as an empirical evidence of co-movements.

In fig. 21 (left picture) the series of nominal wages and the consumer price index in growth rates are plotted which give a clear impression of moving together. However, implementing the two-step cointegration regression strategy suggested by Engle and Granger (1987), the DW associated with the static regression $w(t) = \text{const.} + bp(t) + u(t)$ is 0.10, and the ADF statistic associated with the regression $\Delta u(t) = -\phi u(t-1) + \gamma \Delta u(t-1) + \text{error}$ is -1.26. The null hypothesis of absence of cointegration between nominal wages and the consumer price index is then highly accepted by the data.

Nonetheless, as long as the density estimate of fig. 21 (right picture) suggests the presence of several outliers, we consider the series of price index in first difference, and run the univariate analysis already implemented on the nominal wage series. Once again, the observation at time t=1933 appears to be the date at which splitting the sample (alternatively, we should split at 1920 - see fig. 22), so that the change point analysis previously implemented on the nominal wage series can be now repeated for the same subsamples (graphed in fig. 23). The results reported in terms of top predictive log-likelihoods (table 8) and likelihood ratio tests (table 9) indicate that 10 change points are overall detected at the historical dates of 1915, 1916, 1920, 1922, 1930, 1933, 1940, 1943, 1945, and 1948.[16]

This yields the regression results reported in table 10, in which all diagnostics confirm the goodness of the statistical model for the price index series in first differences, with the only exception of the t-value associated with the change point at time 1930, which appears non-significative. This can be due to the restriction imposed by the regression model in the form of shifting mean value model, as long as such model assumes observations stands locally on horizontal lines. Indeed, by looking at observations from 1923 to 1933, it can be seen perhaps a downward sloping line would better fit the data, in which case we would not need to recognize the date of 1930 as a change point. Applications in this case, which would require the use of the different regression model, called by Kashiwagi the "discrete spline" model (see Kashiwagi, 1991, sections 5.2 and 5.3), are left for future work in this area.

The signal extraction derived on the basis of the structural break analysis is reported in fig. 24 (left picture), and compared (right picture) with the corresponding signal extraction obtained for the nominal wage series. Some main comments are then in order:

a) A good correspondence between historical dates is found. Change in dynamics in the two series sometimes occur with one lag of adjustment, but most of the time without any lag of adjustment at all. This is particularly true starting from 1933. We have in fact changes at 1933, 1940, 1943, and 1948 in the nominal wage series are immediately matched by changes (in the same direction) in the dynamics of the price index series. Instantaneous adjustments take also place in 1915 and 1920, while changes in 1921 and 1929 in the nominal wage series are matched by the price index series only one period later.

b) Variations in the nominal wage and price index series are on average of the same size up to year 1920 (if we make the exception of the change in 1907 in the nominal wage series not matched by the price index series). Subsequently, reactions in the price indices appear to be more contained with respect to variations in the nominal wage series, with exception in 1945 (see also fig. 21, left picture), the date of the "*price peak*" (see Friedman and Schwarz, 1963, p. 574).

[16] The date of 1915 is also accepted to be a change point, given the value of the LR statistic extremely close to the rejection border.

c) In 1915, the adjustment of the price index series is partial, being only in the subsequent year it becomes fully effective. This suggests if a link of causality has to be established between the two series, this goes more probably from wages to price indices (i.e. dynamics in wages produce inflationary or deflationary effects), rather than the contrary. This is a general conclusion particularly true up to 1933. When the dates of change points exactly coincide, however, deeper inquires are requested in order to inference the direction of causality.

d) The breaks in 1907 and 1923 represent the only dates where the price index does not react to cuts in nominal wages. In 1945, on the contrary, a switch in the dynamics of the price index series does not associate a corresponding switch in the dynamics of the nominal wage series.

The above comments give an idea of how richer the amount of information gained from data analysis can be by fitting shifting trend models to actual macroeconomic series, rather than simple DS models. In the multivariate framework, as in the univariate domain, the assumption of infrequent permanent shocks forces in fact the researcher to go beyond a statistical test about the absence (or presence) of cointegrating relations; rather, a local analysis is needed from which co-movementes can be identified on the basis of the correspondence of historical dates.

For the U.S. economy we basically found evidence that wages cause prices most of the time, what would be in accordance for example with the economic theory of mark-up in the formation of prices. It would also be interesting (and not at all difficult) to check weather the same conclusion would hold for Western Europe economies as well, in particular those with an strong degree of wage indexation.

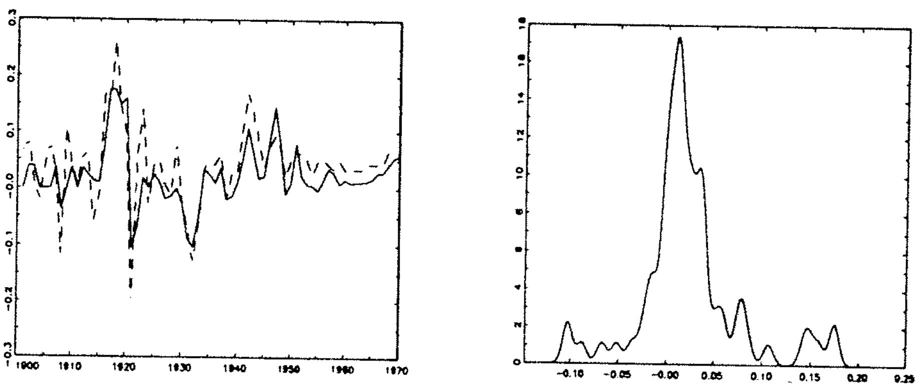

Fig. 21. Left: Nominal wage series (solid line) and price index series (dotted line) in growth rates: sample period 1901 - 1970. Right: Density estimate for the price index series in first differences. Smoothing parameter selected through Sheater and Jones' method: 0.0052. Gaussian kernel.

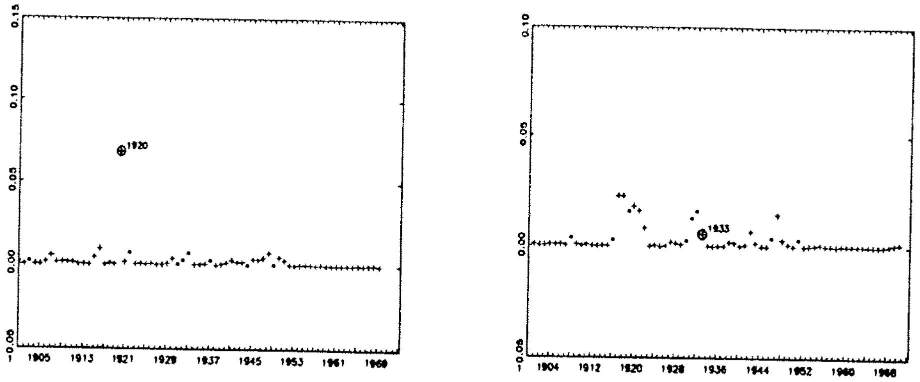

Fig. 22. Variability plot for the U.S. series of the price index in first differences. Left: Distances computed as the square of the difference between an observation and its right-hand neighbour (the first sample observation is lost). Right: variances computed as the squared differences from the sample mean.

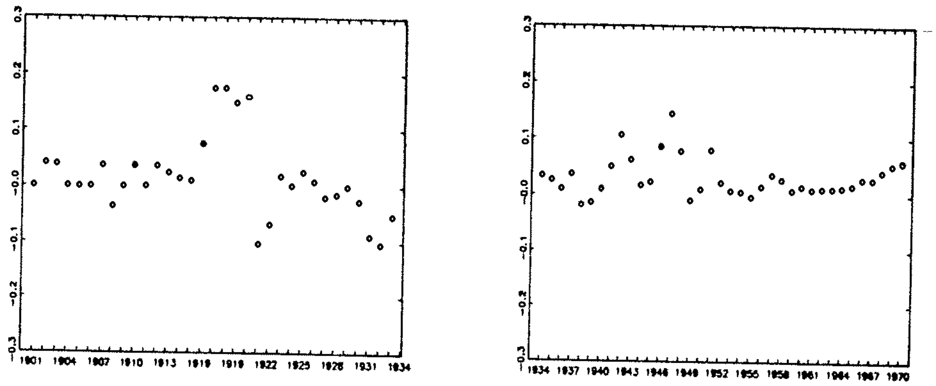

Fig. 23. Splitting of the U.S. price index series in first differences. Left: Subsample 1901-1933. Right: Subsample 1934-1970. Note: The series are plotted with the same scale as the nominal wage series.

Table 9: Top predictive log-likelihoods for the price index series, 1901-1933.

	max pred	$j(1)$	$j(2)$	$j(3)$	$j(4)$	$j(5)$	$j(6)$
$n=0$	6.63						
$n=1$	8.15	1920					
$n=2$	14.5	1916	1920				
$n=3$	15.3	1916	1920	1930			
$n=4$	17.9	1916	1920	1922	1930		
$n=5$	**18.2**	1915	1916	1920	1922	1930	
$n=6$	17.5	1915	1916	1920	1922	1926	1930

Frequency of occurrency of observations suspected to be change points:

$j(i)$	**1915**	**1916**	**1920**	**1922**	1926	**1930**
freq.	2	5	6	3	1	4

Table 8: Top predictive log-likelihoods for the price index series, 1934-1970.

	max pred	$j(1)$	$j(2)$	$j(3)$	$j(4)$	$j(5)$	$j(6)$
$n=0$	19.1						
$n=1$	18.7	1948					
$n=2$	20.9	1945	1948				
$n=3$	21.23	1940	1945	1948			
$n=4$	21.25	1940	1943	1945	1948		
$n=5$	**21.33**	1940	1943	1945	1948	1967	
$n=6$	20.1	1937	1940	1943	1945	1948	1967

Frequency of occurrency of observations suspected to be change points:

$j(i)$	1937	**1940**	**1943**	**1945**	**1948**	**1967**
freq.	1	4	3	5	6	2

Table 9: Likelihood ratio tests on the significance of change points, for the price index series.

H_0	H_1	LR
1916,1920,1922 1930,1933,1940 1943,1945,1948	1916,1920,1922 1930,1933,1940 1943,1945,1948 1915	3.71 (3.84)
1915,1916,1920 1922,1930,1933 1940,1943,1945 1948	1915,1916,1920 1922,1930,1933 1940,1943,1945 1948 1926	1.77 (3.84)
1915,1916,1920 1922,1930,1933 1940,1943,1945 1948	1915,1916,1920 1922,1930,1933 1940,1943,1945 1948 1937	2.20 (3.84)
1915,1916,1920 1922,1930,1933 1940,1943,1945 1948	1915,1916,1920 1922,1930,1933 1940,1943,1945 1948 1967	3.46 (3.84)

Note: In the third column, the LR statistic is reported together with the 5% critical value (in parenthesis). In the second column, the date(s) underlined is (are) the date(s) tested as statistically significative change point(s).

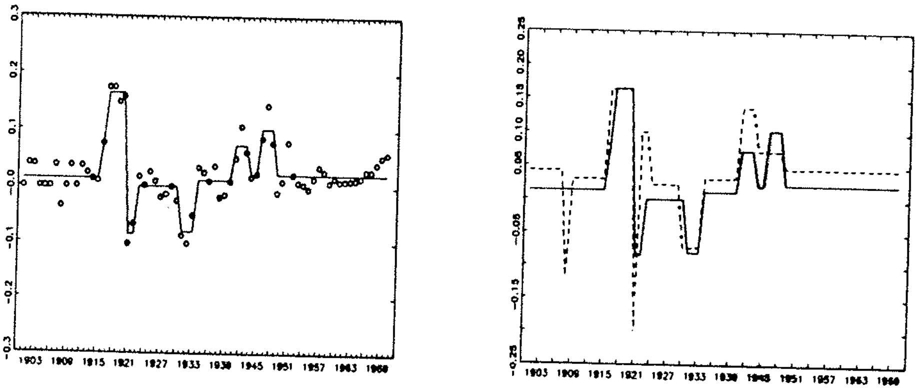

Fig. 24. Left: Signal extraction for the price index series in first differences (left). Right: Signal of the price index series (dotted line), compared to the corresponding trend structure of the nominal wage series.

7. Conclusions

In an article published in the book "*Unemployment, Hysteresis & the Natural Rate Hypothesis*", (see Rod Cross editor, 1988), Summers (1988) has recently summarized the present debate in macroeconomics in terms of the contra-position between Keynesian and New Classical theories. In trying to respond

Table 10: Regression results for the price index series.

Period	μ_i	s.e.	t-value
1901-14	0.013	0.001	9.18
1915	0.076	0.022	3.47
1916-20	0.164	0.005	30.1
1921-22	-0.085	0.011	-7.81
1923-30	-0.0004	0.003	-0.13*
1931-33	-0.08	0.007	-11.1
1934-40	0.0116	0.003	3.72
1941-43	0.073	0.007	10.0
1944-45	0.02	0.011	1.84
1946-48	0.102	0.007	14.1
1949-70	0.022	0.001	22.3

Note: Other regression diagnostics are the sample variance of the residuals, 0.0005; the coefficient of determination, 0.86, the adjusted R^2, 0.84; the Durbin-Watson statistic, 1.88; the log-likelihood, 174.5; the Akaike information criteria, -305 and the Schwarz selection criteria -6.81; Jarque and Bera normality test statistic, 1.60 (against a 5% critical value of 5.99)

to the challenges of New Classicals, Summers has argued Keynesians presently propose a disequilibrium explanation of business cycle fluctuations as a consequence of nominal rigidities like overlapping contracts, menu costs, slowly adaptive expectations, for which there is however mixed empirical evidence. A completely different approach characterizes on the other hand New Classical economics, that is a world made of intertemporal optimizing agents, rational expectations, and market clearing, but where the 'extreme' view of equilibrium often leads to propositions of neutrality of economic policies, in the long as well as in the short run. Despite deep divergences, Keynesians and New Classicals appear however to agree, according to Summers, on the view that "*economic fluctuations represent transitory movements away from equilibrium*" (Summers, p.15), and it is exactly in this sense both approaches look then *dispiriting* (p.xxx), as they neglect in practice that economic policies may have substantial effects on macroeconomic variables.[17] There seems nonetheless to be an "*empirical evidence of multiple equilibria in GDP behaviour*", according to which "*the economy does not fluctuate about a unique well defined equilibrium, but it is capable of setting at many different equilibria, one of which is the best*. (Summers, p. 23).

In this paper, our aim has been to highlight as much as possible (that is as

---

[17] "*The conclusion that policy cannot affect the average output is likely to be a feature of any model that postulates a unique equilibrium level of output and attributes fluctuations to disequilibrium situations*". (Summers, 1988, p. 15).

much as supported by the data) the empirical evidence of multiple equilibria in macroeconomic time series advocated by Summers in his article. The starting point for our research has been provided by major advances in the literature concerning unit roots, stochastic trends, and measures of persistence in macroeconomic time series.

Early contributions developed by Nelson and Plosser (1982) and others pointed out in fact for the first time a different nature of business cycle fluctuations, within a new framework in which the main attention was focused on forces explaining the growth path of the economy, rather then business cycles themselves. After the pioneristic contribution of Nelson and Plosser, however, it is also true there has been a tendency to view persistence of shocks as affecting the levels of the series, but not yet the growth rates. Empirical evidence of unit roots, in other words, has been more often interpreted in the sense of "*fluctuations in the 'natural rate*'" (Nelson and Plosser, 1982, p. 160), rather than in the sense of *multiple* natural rates (*multiple* local equilibria) advocated by Summers.

Now, the fact that economic time series may not exhibit trend reverting behaviour to previous *growth rates* opens the debate whether fluctuations in the natural rate have really to be considered transitory - i.e. whether the economy really fluctuates about a *constant* (unique) natural rate. In particular, the view of infrequent permanent shocks naturally calls for a closer analysis of models in which multiple equilibria arise in consequence of hysteresis effects - or path dependency - that is models in which movements in actual series do not simply represent transitory drifts from an underlying well defined (unique) natural rate. Instead, such movements may just reflect shifts in the natural rate itself.

On theoretical grounds, this is implicitly Blanchard and Summers' (1986b) definition of hysteresis effects in the European unemployment rate series: "*Most of the time the equilibrium rate is stable and unaffected by movements in the actual rate. But once a while a sequence of shocks pushes the equilibrium rate up or down, where it remains until another sequence of shocks dislodges it*". (Blanchard and Summers, 1986b, p. xxx). Such a definition of hysteresis exactly corresponds to the view that shocks may occur infrequently with quite persistent effects (effects which determine switches to *different* natural rates), rather than frequently with weaker effects (that is effects which only determine fluctuations about the *same* natural rate).

On the statistical front, segmented trends in the form of shifting mean value models may fit actual series much better than traditional difference stationary processes, as it appeared for example to be the case for the U.S. series of nominal wages. Indeed, modeling the series according to ST processes allow for signal extraction associating white noise *and* normally distributed residuals. The cost to be paid for such a significant improvement is just a computational cost that we can now easily afford in the era of modern computers.