ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Department of Statistics

Econometrics

Chapter 9 : Structural Break - Point Models and Tests

Ioannis Vrontos
Associate Professor in Statistics
Department of Statistics
School of Information Sciences and Technology
Athens University of Economics & Business
Email: vrontos@aueb.gr
Office: Troias & Kimolou, 4^{th} floor, 412
Tel.: 210 8203927

Chapter 9: Structural Break-Point Models

- Structural break-point models
  - Introduction and motivation
  - Structural break-point regression models
  - Structural break-point time series models
- Testing for structural break-points
- Examples and applications using R

Structural Break-point models:

Introduction and motivation

- So far, we have estimated regression models:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_k X_{k,i} + u_i, \qquad i = 1, \dots, n$$

or

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + u_t, \qquad t = 1, \dots, T$$

and time series models:

$$Y_t = \delta + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \; t = 1, \dots, T$$

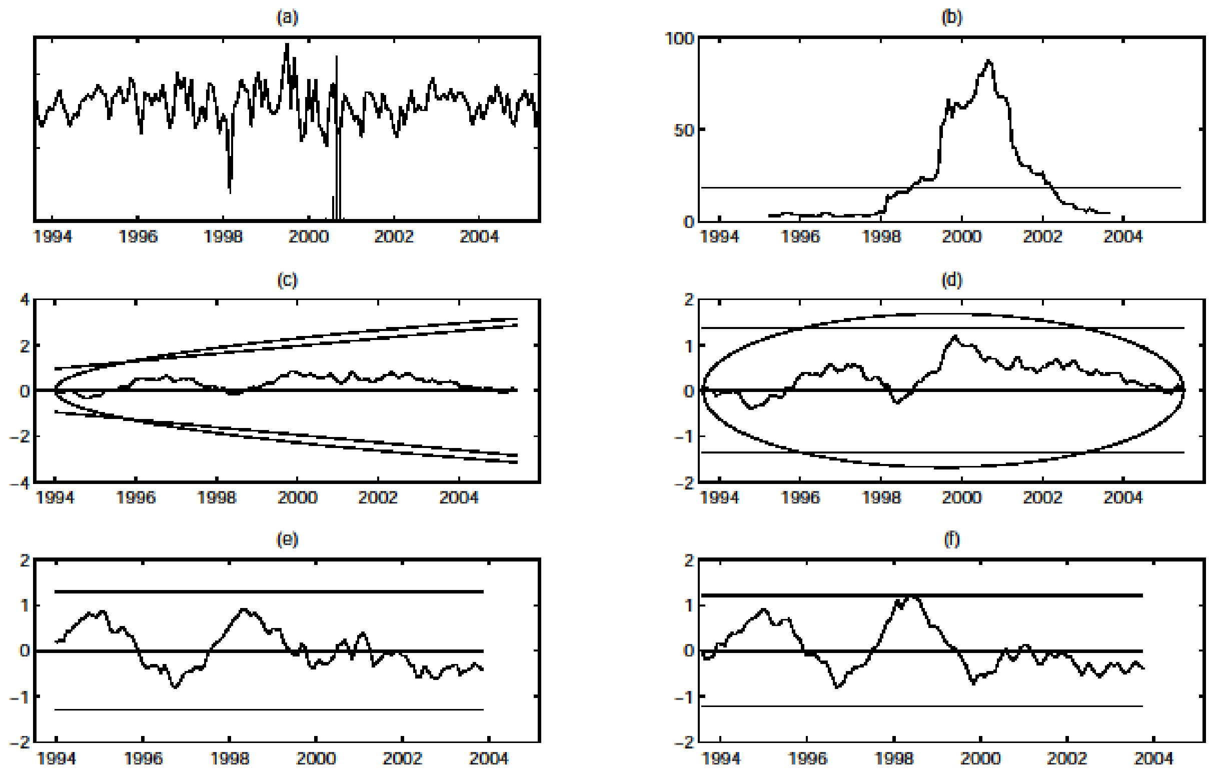
- We have implicitly assumed that the parameters of these models are constant for the entire sample period. However, there are time series data where this assumption in not valid. There is empirical evidence that there are single or multiple structural break-points in the underlying series

- We can test this implicit assumption using parameter stability tests

Structural Break-point models:

Introduction and motivation: Single break

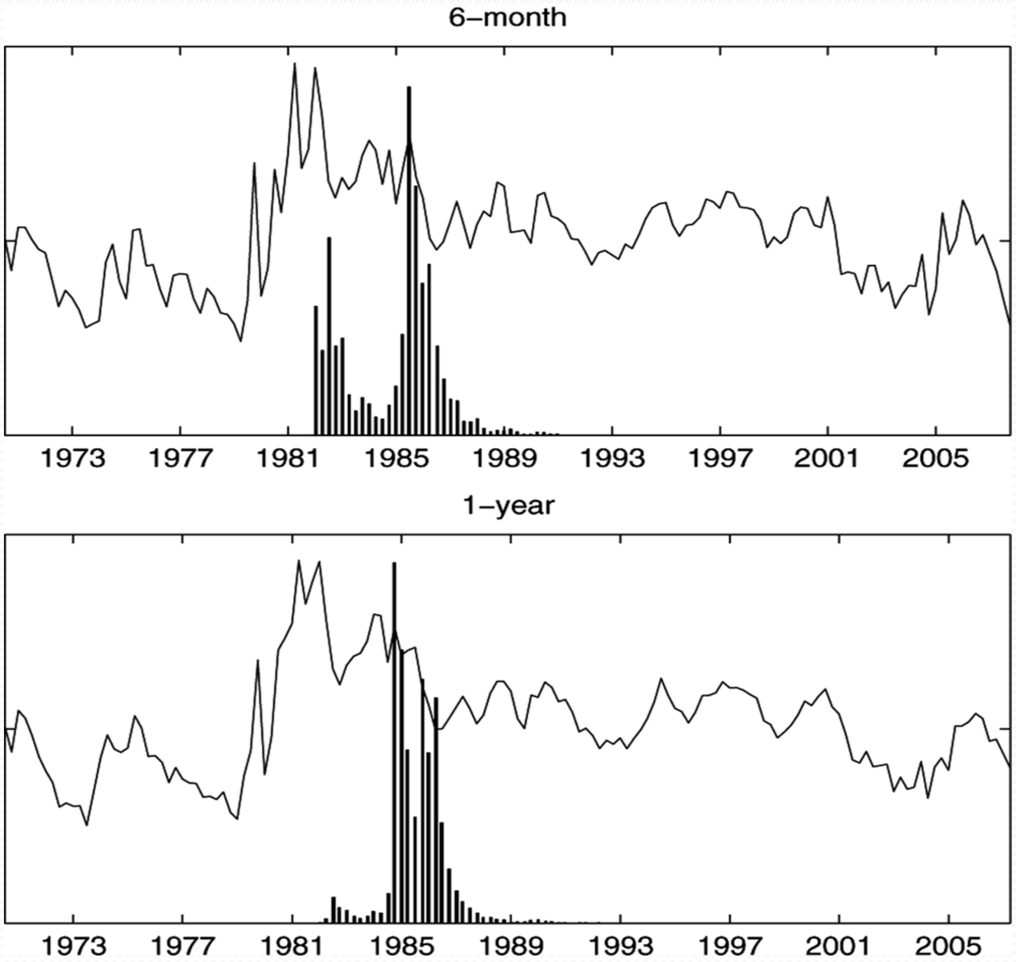
Hedge fund investments: composite index



Structural Break-point models:

Introduction and motivation: Single break

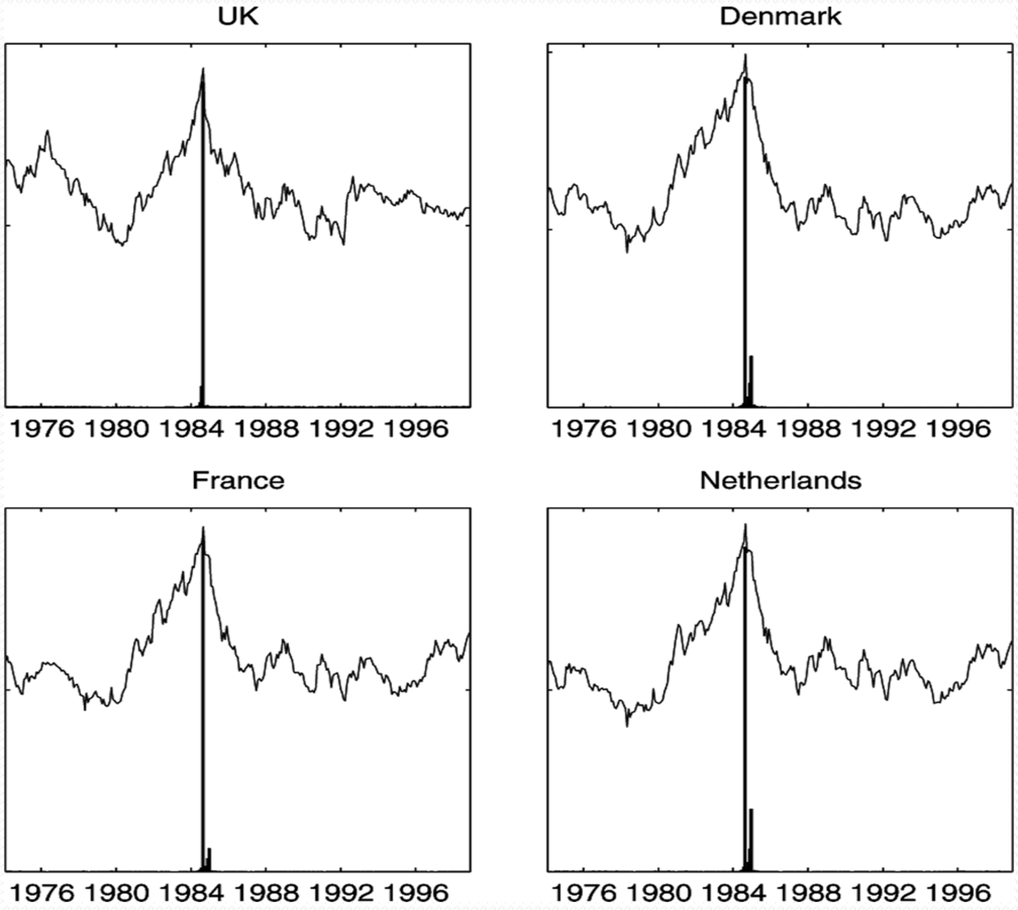
Real Interest Rates series (6-month, 1-year)



Structural Break-point models:

Introduction and motivation: Single break

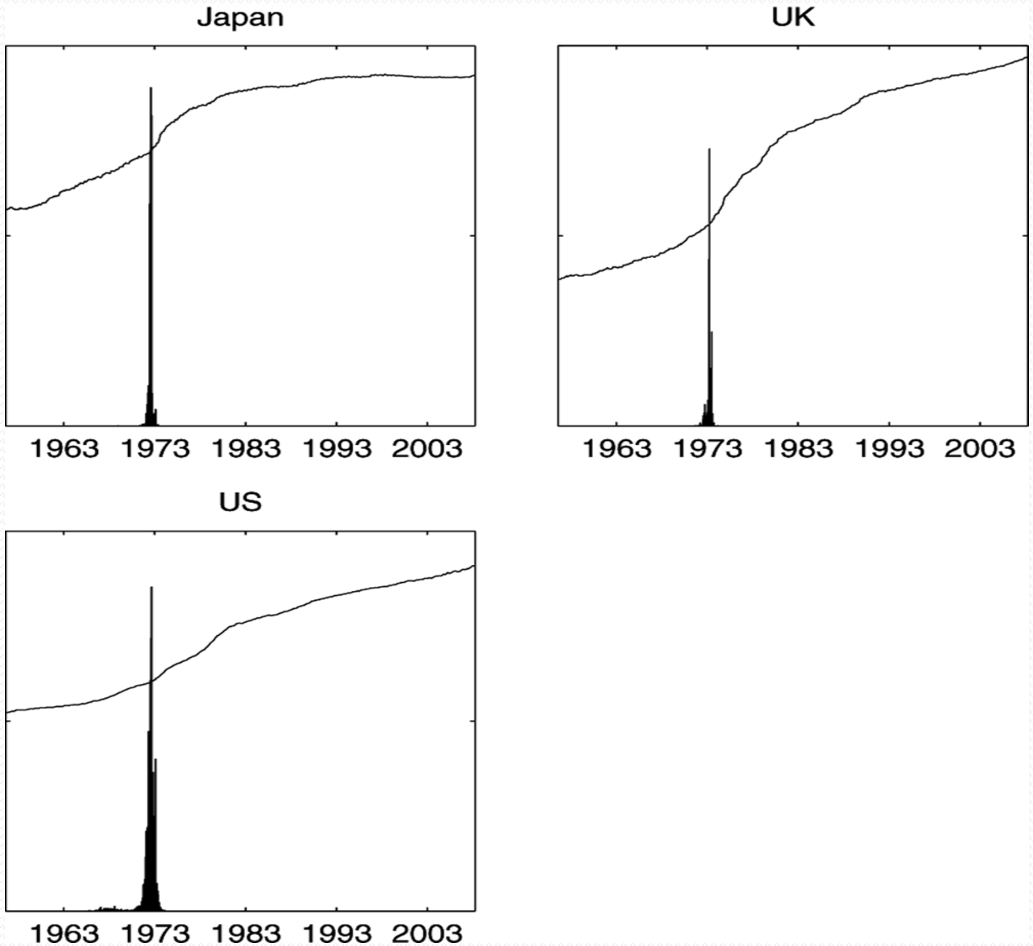
Real Exchange Rates series



Structural Break-point models:

Introduction and motivation: Single break

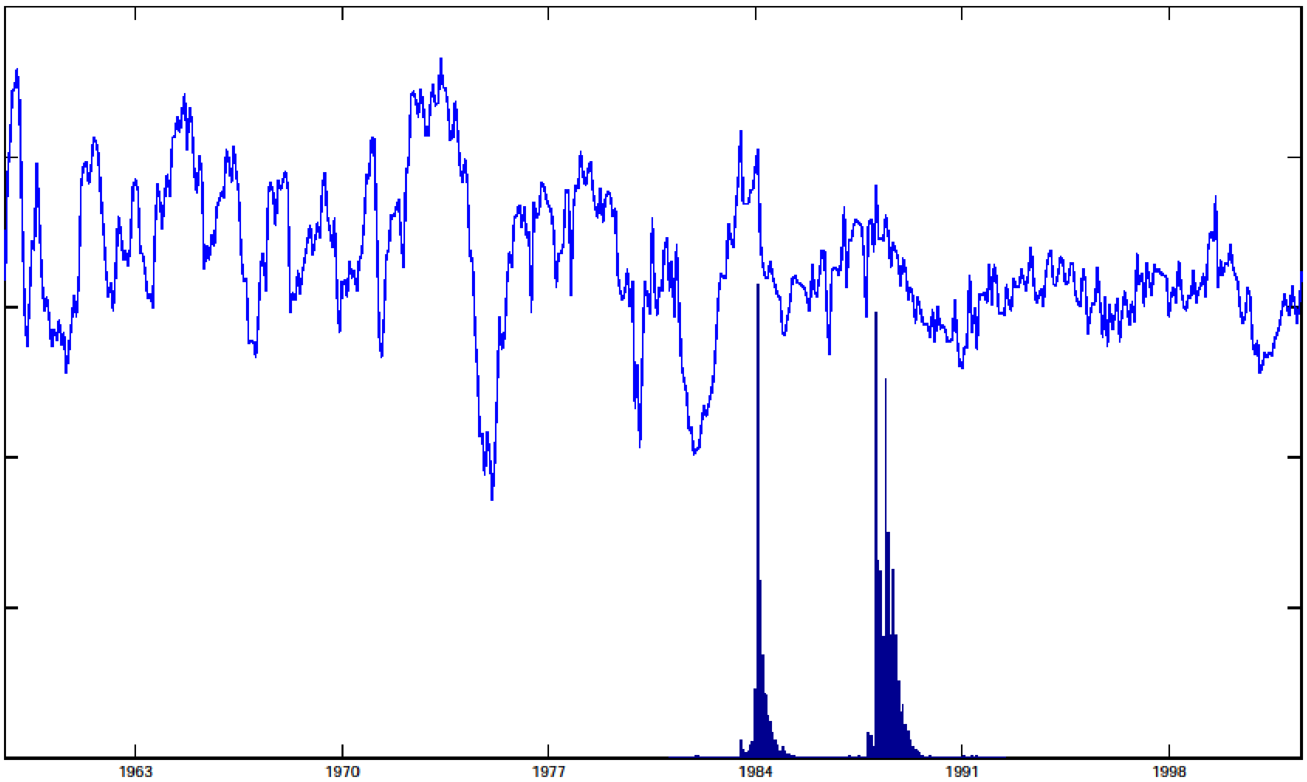
Consumer Price Index



Structural Break-point models:

Introduction and motivation: Single break

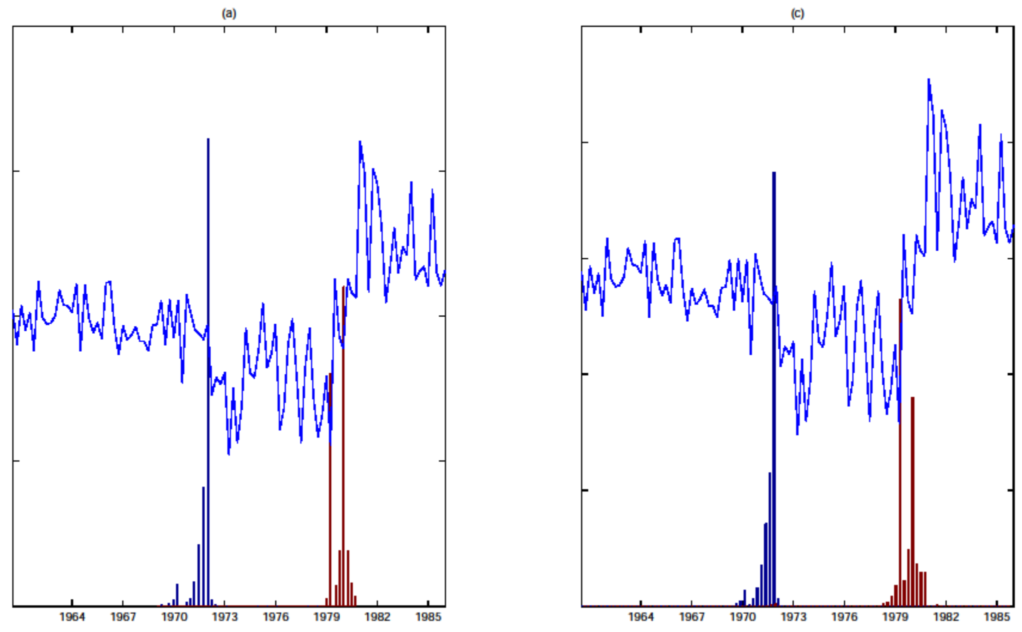
Inventories Index



Structural Break-point models:

Introduction and motivation: Multiple break-points

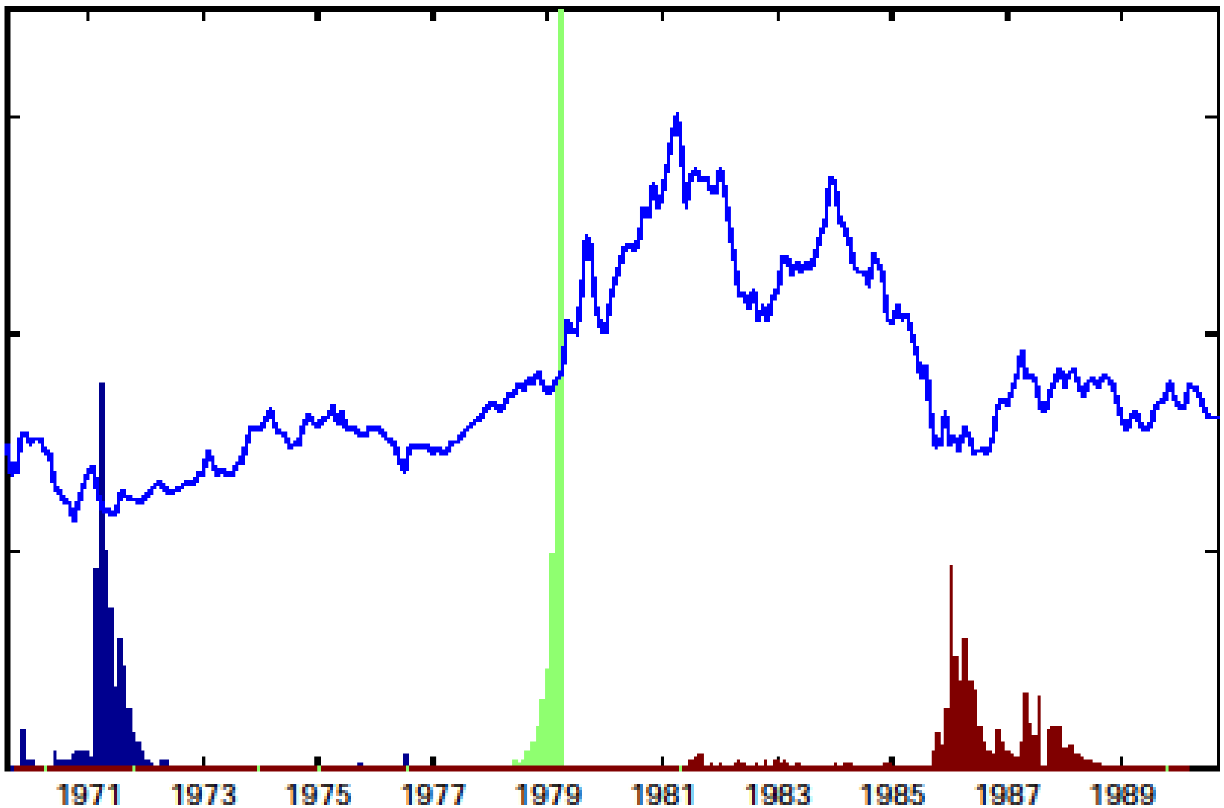
US ex-post real interest rates



Structural Break-point models:

Introduction and motivation: Multiple break-points

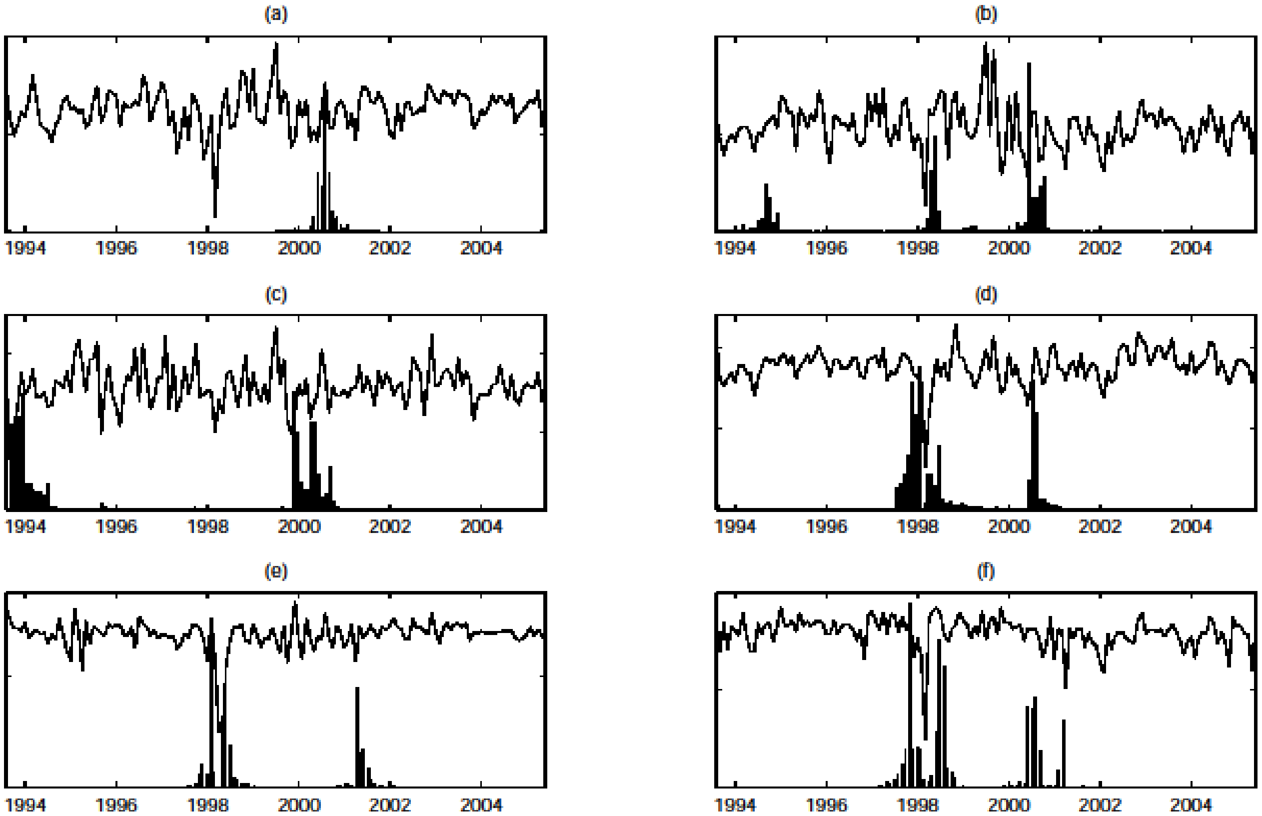
Nominal Interest Rates (10 year)



Structural Break-point models:

Introduction and motivation: Multiple break-points

Hedge Fund Investment Indices: (a) Emerging markets, (b) Equity hedge, (c) Macro, (d) Distressed securities, (e)Fixed Income arbitrage, (f) Merger arbitrage



Structural Break-point models:

Regression-type models (Piecewise regression models)

- Single break-point regression models:

$$Y_t = \beta_{0j} + \beta_{1j}X_{1,t} + \beta_{2j}X_{2,t} + \cdots + \beta_{kj}X_{k,t} + u_t,$$

$$u_t \sim N(0, \sigma_j^2), \quad j = \begin{cases} 1, & for \;\; 1 \le t \le T_0 \\ 2, & for \;\; T_0 < t \le T \end{cases}$$

- Multiple break-point regression models:

$$Y_t = \beta_{0j} + \beta_{1j}X_{1,t} + \beta_{2j}X_{2,t} + \cdots + \beta_{kj}X_{k,t} + u_t,$$

$$u_t \sim N(0, \sigma_j^2), \qquad for \;\; \tau_{j-1} < t \le \tau_j$$

- $\tau_0 = 0 < \tau_1 < \tau_2 < \cdots < \tau_m < \tau_{m+1} = T$ are the m breaks, that define m+1 disjoint segments
- $\beta_{0j}, \beta_{1j}, \cdots \beta_{kj}, \sigma_j^2$ are the regression model parameters associated with segment j

Structural Break-point models:
Time series models (Piecewise AR models)

- Single break-point Autoregressive models:

$$Y_t = \mu_j + \beta_j t + \sum_{i=1}^{p} \varphi_{ij}(Y_{t-i} - \mu_j - \beta_j(t-i)) + u_t,$$

$$u_t \sim N(0, \sigma_j^2), \quad j = \begin{cases} 1, & for \;\; 1 \le t \le T_0 \\ 2, & for \;\; T_0 < t \le T \end{cases}$$

- Multiple break-point Autoregressive models :

$$Y_t = \mu_j + \beta_j t + \sum_{i=1}^{p} \varphi_{ij}(Y_{t-i} - \mu_j - \beta_j(t-i)) + u_t,$$

$$u_t \sim N(0, \sigma_j^2), \qquad for \;\; \tau_{j-1} < t \le \tau_j$$

- $\tau_0 = 0 < \tau_1 < \tau_2 < \cdots < \tau_m < \tau_{m+1} = T$ are the m breaks, that define m+1 disjoint segments

Testing for Structural Break-points:

Different types of tests

- There are different types of tests for structural breaks in the linear regression and time series models
- Classical methods/techniques
  - Based on dummy variables
  - Fluctuation tests
  - F-tests
  - Tests based on least squares (dynamic programming)
- Bayesian methods

Testing for Structural Break-points:

Based on Dummy variables

- **Testing for structural Break-point models using Dummy variables**
  - Dummy variables are used in the regression set up to examine the effect of a qualitative variable to the dependent variable. These variables have predetermined values of 0 and 1, and are technically constructed to examine if specific levels of a categorical variable are important (statistically significant) to explain the variability of the dependent variable.
  - In structural break-point regression models, they can be used to examine the stability of the model parameters, i.e. if the constant and/or the beta coefficients are statistically different in disjoint segments of the data.
  - They can also be used to identify seasonality effects and/or the effects of market events and crises to the dependent variable.

Testing for Structural Break-points:

Fluctuation tests

- **Fluctuation processes based on cumulative sums (CUSUM) of recursive residuals**
  - In this case, testing for structural breaks in linear regression models is performed using fluctuation processes based on cumulative sums (CUSUM) of recursive residuals. Linear and alternative non-linear bounds, which are more sensitive to early and late breaks, have been proposed for the recursive CUSUM process. The process crossing these boundaries suggests deviation from parameter constancy.
- **Fluctuation processes based on cumulative sums (CUSUM) of OLS residuals**
  - A CUSUM test is based on standard OLS residuals, which usually change whenever a new observation is added to the sample. Again, linear and alternative boundaries are available for the OLS CUSUM fluctuation process.
- **Fluctuation processes based on moving sums (MOSUM) of recursive and OLS residuals**
  - Testing for structural breaks is based on moving sums (MOSUM) of recursive and OLS residuals of a fixed-size window. This approach seems intuitively appealing because moving sums are more sensitive to parameter changes than the CUSUM tests.

Testing for Structural Break-points:

F-tests

- The **F-tests** are tests for the null hypothesis of parameter constancy against the alternative of a single structural change.
- Chow (1960) first suggested an F-test for the case that the break-date is known. A straightforward generalization of the Chow test was to calculate the F statistics for all possible break-dates and reject the null hypothesis if any of those statistics were too large (Quandt, 1960).
- Andrews (1993) and Andrews and Ploberger (1994) proposed F-tests based on quantities obtained from the sequence of the F statistics. The time at which the F-statistic becomes maximum provides an estimate for the break-date.
- The class of F-tests for a structural break of unknown position enjoy optimality properties in the case of a single break alternative (Andrews, 1993). On the other hand the MOSUM tests have greater power at detecting double structural changes (Chu et al. 1995). In general, the above classical tests for structural breaks are complementary and can be used together as a first attempt to visualize possible deviations from parameter constancy and test for the presence of structural breaks.

Testing for Structural Break-points:

Dynamic programming based on sum of squared residuals

- Bai and Perron (1998, 2003) considered theoretical and computational issues regarding the estimation of and testing for multiple structural breaks in linear regression models.
- Their method is a dynamic programming approach based on the calculation of the sums of squared residuals in all possible segments of the data, with the minimum size of each segment being pre-specified. Then a sequential procedure is used for the global minimization of the overall sum of squared residuals. The method produces a single optimal m-breaks partition for each given value of m.
- Bai and Perron method selects the optimal number of breaks given the estimated break-dates, by sequentially testing the hypothesis of l breaks versus l + 1 breaks using significance tests, or using information based criteria such as the AIC and the BIC.

Testing for Structural Break-points:

Bayesian methods

- Bayesian techniques have also been used to examine the presence of a single or of multiple break points in regression and/or time series models (Meligkotsidou, Tzavalis and Vrontos, 2011, 2017, Meligkotsidou and Vrontos, 2008, 2011).
- Bayesian methods allow the analyst to search for the relevant explanatory variables which best describe the underlying dependent variable, taking into account the presence of single or multiple structural breaks in the series.
- Bayesian model comparison allows to compute posterior model probabilities for a large number of competing models (without and/or with breaks), and enables to obtain the posterior distributions of the model parameters, i.e. the alphas, the betas and the variances in different segments (sub-periods in the observed sample).
- Thus, we are able to compute reliable estimates for the model parameters, and assess the parameter uncertainty

Regression models with Dummy variables (Reminder)

- Recall the use of dummy variables in the standard linear regression framework:

1. Simple Linear Regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

2. Linear Regression model with dummy variable (test if the constant is different):

$$Y_i = \beta_0 + \beta_1 X_i + \gamma Z_i + u_i$$

3. Linear Regression model with dummy variable (test if the slope is different):

$$Y_i = \beta_0 + \beta_1 X_i + \delta (ZX)_i + u_i$$

4. Linear Regression model with dummy variable (test if the constant and the slope are different):

$$Y_i = \beta_0 + \beta_1 X_i + \gamma Z_i + \delta (ZX)_i + u_i$$

Estimate Linear model 1:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

```
fit<-lm(consumption ~ income)
summary(fit)

#Call:
#  lm(formula = consumption ~ income)

#Coefficients:
#                 Estimate  Std. Error t value  Pr(>|t|)
#(Intercept) 66.50939   54.84543   1.213    0.231
#income         0.82285     0.02999  27.436   <2e-16 ***

#Residual standard error: 130.1 on 48 degrees of freedom
#Multiple R-squared:  0.9401,      Adjusted R-squared:  0.9388
#F-statistic: 752.8 on 1 and 48 DF,  p-value: < 2.2e-16
```

Estimate Linear Model 2:

$$Y_i = \beta_0 + \beta_1 X_i + \gamma Z_i + u_i$$

```
fit2<-lm(consumption ~ income + z)
summary(fit2)

#Call:
#  lm(formula = consumption ~ income + z)

#Coefficients:
#                Estimate   Std. Error   t value    Pr(>|t|)
#(Intercept)  97.88728    29.39477      3.33      0.0017 **
#  income       0.70345      0.01932    36.42     < 2e-16 ***
#  z          272.36597    24.69108    11.03    1.23e-14 ***

#Residual standard error: 69.41 on 47 degrees of freedom
#Multiple R-squared:  0.9833,       Adjusted R-squared:  0.9826
#F-statistic:  1384 on 2 and 47 DF,  p-value: < 2.2e-16
```

Estimate Linear Model 3:

$$Y_i = \beta_0 + \beta_1 X_i + \delta(ZX)_i + u_i$$

```
fit3<-lm(consumption ~ income + zx)
summary(fit3)

#Call:
#  lm(formula = consumption ~ income + zx)

#Coefficients:
#                  Estimate   Std. Error  t value  Pr(>|t|)
#(Intercept)   290.88749   32.52130    8.945   1.03e-11 ***
#  income          0.55372     0.02635    21.017     < 2e-16 ***
#  zx                0.18876     0.01532    12.323   2.50e-16 ***

#Residual standard error: 63.93 on 47 degrees of freedom
#Multiple R-squared:  0.9858,      Adjusted R-squared:  0.9852
#F-statistic:  1635 on 2 and 47 DF,  p-value: < 2.2e-16
```

Estimate Linear Model 4:

$$Y_i = \beta_0 + \beta_1 X_i + \gamma Z_i + \delta(ZX)_i + u_i$$

```
fit4<-lm(consumption ~ income + z + zx)
summary(fit4)

#Call:
#  lm(formula = consumption ~ income + z + zx)

#Coefficients:
#                Estimate   Std. Error  t value   Pr(>|t|)
#(Intercept)   243.24425   53.99953    4.505    4.54e-05 ***
#  income        0.58848    0.04102   14.345     < 2e-16 ***
#  z            74.56259   67.55380    1.104     0.27544
#zx              0.14145    0.04550    3.109     0.00322 **

#Residual standard error: 63.78 on 46 degrees of freedom
#Multiple R-squared:  0.9862,     Adjusted R-squared:  0.9853
#F-statistic:  1096 on 3 and 46 DF,  p-value: < 2.2e-16
```

Testing for Structural Break-points:

Based on Dummy variables

Testing for structural Break in the intercept (only) of the model using Dummy variable

- Define the dummy variable $d_t = \begin{cases} 1, & \text{if the characteristic happens} \\ 0, & \text{if the characteristic does not happen} \end{cases}$

- Examine if there is a structural change in the constant of the model

$$Y_t = \beta_0 + \delta d_t + \beta_1 X_{1,t} + u_t, \qquad u_t \sim N(0, \sigma^2)$$

- If $d_t = 0$, then the model can be written:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + u_t$$

- If $d_t = 1$ and δ is statistically significant, then the model can be written:

$$Y_t = (\beta_0 + \delta) + \beta_1 X_{1,t} + u_t$$

Testing for Structural Break-points:

Based on Dummy variables

Testing for structural Break in the beta coefficient (only) of the model using Dummy variable

- Define the variable $\mathrm{d_t X_{1,t}} = \begin{cases} \mathrm{X_{1,t}}, & \text{if the characteristic happens } (\mathrm{d_t} = 1) \\ 0, & \text{if the characteristic does not happen } (\mathrm{d_t} = 0) \end{cases}$

- Examine if there is a structural change in the beta coefficient of the model
$$\mathrm{Y_t} = \beta_0 + \beta_1 \mathrm{X_{1,t}} + \gamma(\mathrm{d_t X_{1,t}}) + \mathrm{u_t}, \qquad \mathrm{u_t} \sim \mathrm{N}(0, \sigma^2)$$

- If $\mathrm{d_t} = 0$, then the model can be written:
$$\mathrm{Y_t} = \beta_0 + \beta_1 X_{1,t} + \mathrm{u_t}$$

- If $\mathrm{d_t} = 1$ $i.e.$ $\mathrm{d_t X_{1,t}} = \mathrm{X_{1,t}}$, and γ is statistically significant, then the model can be written:

$$\mathrm{Y_t} = \beta_0 + (\beta_1 + \gamma) X_{1,t} + \mathrm{u_t}$$

Testing for Structural Break-points:

Based on Dummy variables

Testing for structural Break in the intercept and the beta coefficient of the model using Dummy variables

- Examine if there is a structural change in the beta coefficient of the model

$$\mathrm{Y_t} = \beta_0 + \beta_1 \mathrm{X}_{1,t} + \delta \mathrm{d_t} + \gamma(\mathrm{d_t} C) + \mathrm{u_t}, \qquad \mathrm{u_t} \sim \mathrm{N}(0, \sigma^2)$$

- If $\mathrm{d_t} = 0$, then the model can be written:

$$\mathrm{Y_t} = \beta_0 + \beta_1 X_{1,t} + \mathrm{u_t}$$

- If $\mathrm{d_t} = 1$ $i.e.$ $\mathrm{d_t X_{1,t}} = \mathrm{X_{1,t}}$, and γ and δ are statistically significant, then the model can be written:

$$\mathrm{Y_t} = (\beta_0 + \delta) + (\beta_1 + \gamma) X_{1,t} + \mathrm{u_t}$$