WIRTSCHAFTS UNIVERSITÄT WIEN

A Unified Approach to Testing, Monitoring and Dating Structural Changes

Achim Zeileis

http://www.ci.tuwien.ac.at/~zeileis/

Overview

- Motivation
  - Model stability: testing, monitoring, dating
  - Frankel-Wei regression: Chinese currency regime
  - Error correction model: German M1 money demand
- Model frame
  - Parametric regression models
  - Objective function, estimating functions
  - Central limit theorem
- Fluctuation tests
  - Empirical M-fluctuation processes
  - Functional central limit theorem
  - Aggregation functionals

Overview

  - Special cases: OLS-based CUSUM, Nyblom-Hansen, supLM test
  - Other fluctuation processes
  - How to choose process and functional?
- Monitoring with fluctuation tests
  - Sequential testing
  - Boundary functions
- Dating structural breaks
  - Dynamic programming algorithm
  - Confidence intervals
- Software
- Summary

Motivation

Consider n ordered observations (y_i, x_i) for $i = 1, \dots, n$ in a regression setup which can be described by a regression model with parameter θ_i.

Ordering is typically with respect to time in time-series regressions, but could also be with respect to income, age, etc. in cross-section regressions.

Testing: Given that a model with parameter $\widehat{\theta}$ has been estimated for these n observations, the question is whether this is appropriate or: *Are the parameters stable or did they change through the sample period $i = 1, \dots, n$?*

Motivation

Monitoring: Given that a stable model could be established for these n observations, the question is whether it remains stable in the future or: *Are incoming observations for $i > n$ still consistent with the established model or do the parameters change?*

Dating: Given that there is evidence for a structural change in $i = 1, \ldots, n$, it might be possible that stable regression relationships can be found on subsets of the data. *How many segments are in the data? Where are the breakpoints?*

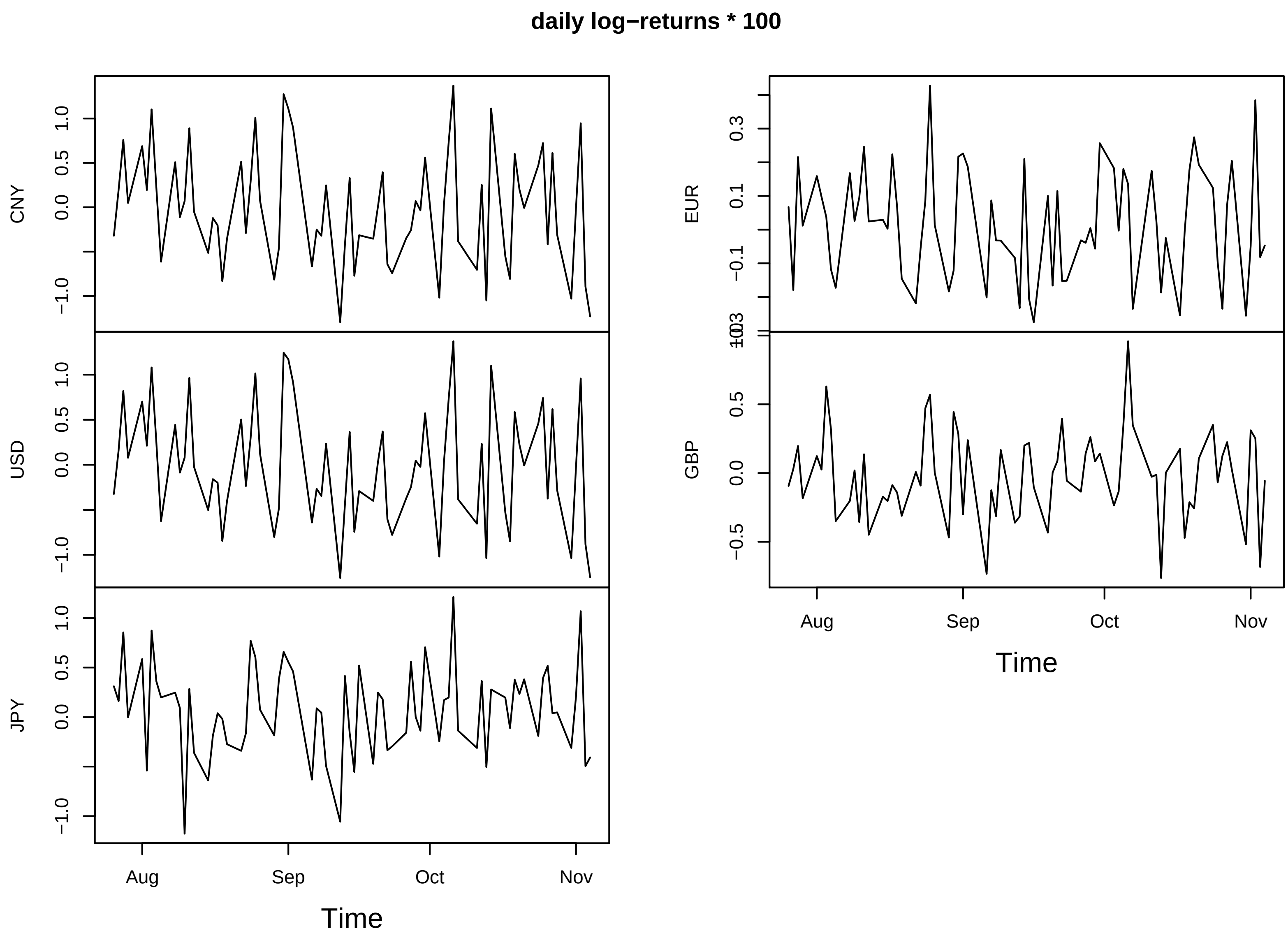
Motivation

Shah, Zeileis, Patnaik (2005) investigate the *de facto* currency regime for the Chinese yuan (CNY) after China gave up on a fixed exchange rate to the US dollar (USD) on 2005-07-21.

The People's Bank of China announced that the CNY would no longer be pegged to the USD but to a basket of currencies so that the CNY exchange rate would be improved with greater flexibility.

Using a Frankel-Wei regression model for the log-returns of the exchange rates of CNY, USD, JPY, EUR and GBP (wrt the base currency CHF) based on data up to 2005-10-31 ($n = 68$), it can be determined that a plain USD peg is still in operation.

Motivation


daily log−returns * 100
CNY
USD
JPY
EUR
GBP
1.0
0.5
0.0
−1.0
0.3
0.1
−0.1
−0.5
Aug
Sep
Oct
Nov
Time

Motivation

The Frankel-Wei regression uses ordinary least squares (OLS) estimation and gives:

$$\begin{aligned}\text{CNY}_i &= -0.005 + 0.9997\,\text{USD}_i + 0.005\,\text{JPY}_i \\ &\quad -0.014\,\text{EUR}_i - 0.008\,\text{GBP}_i + \widehat{u}_i,\end{aligned}$$

where only the USD coefficient is significantly different from zero (HAC-corrected t statistic 144.73).

The error standard deviation is tiny with 0.028 and $R^2 = 0.998$.

Motivation

Questions:

1. Is this model for the period 2005-07-26 to 2005-10-31 stable or is there evidence that China kept changing its currency regime after 2005-07-26? (*testing*)

2. Depending on the answer to the first question:
   - Does the CNY stay pegged to the USD in the future (starting form November 2005? (*monitoring*)
   - When and how did the Chinese currency regime change? (*dating*)

Motivation

Lütkepohl, Teräsvirta, Wolters (1999) investigate the linearity and stability of German M1 money demand: stable regression relation for the time before the monetary unification on 1990-06-01 but a clear structural instability afterwards.

Data: seasonally unadjusted quarterly data, 1961(1) to 1995(4)

Error Correction Model (in logs) for data up to 1990(2) with variables: M1 (real, per capita) m_i, price index p_i, GNP (real, per capita) y_i and long-run interest rate R_i:

$$\begin{aligned} \Delta m_i &= -0.30\Delta y_{i-2} - 0.67\Delta R_i - 1.00\Delta R_{i-1} - 0.53\Delta p_i \\ &\quad -0.12 m_{i-1} + 0.13 y_{i-1} - 0.62 R_{i-1} \\ &\quad -0.05 - 0.13 Q1 - 0.016 Q2 - 0.11 Q3 + \hat{u}_i, \end{aligned}$$

Motivation

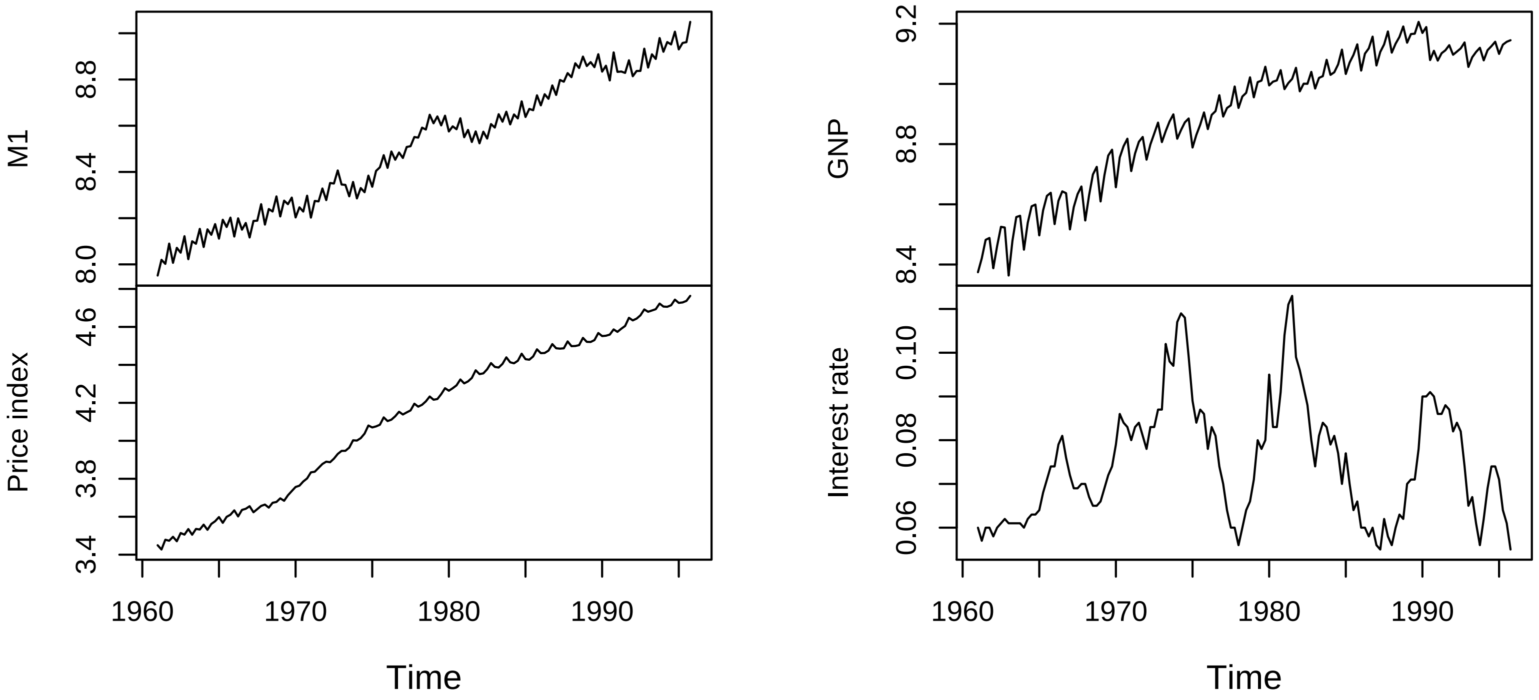
Lütkepohl, Teräsvirta, Wolters (1999) investigate the linearity and stability of German M1 money demand: stable regression relation for the time before the monetary unification on 1990-06-01 but a clear structural instability afterwards.

Data: seasonally unadjusted quarterly data, 1961(1) to 1995(4)

Error Correction Model (in logs) for data up to 1990(2) with variables: M1 (real, per capita) m_i, price index p_i, GNP (real, per capita) y_i and long-run interest rate R_i:

$$\begin{aligned}\Delta m_i \;=\; & -0.30\Delta y_{i-2} - 0.67\Delta R_i - 1.00\Delta R_{i-1} - 0.53\Delta p_i \\ & +1.00\hat{\varepsilon}_{i-1} \\ & -0.05 - 0.13Q1 - 0.016Q2 - 0.11Q3 + \hat{u}_i,\end{aligned}$$

Motivation


M1
8.0
8.4
8.8
Price index
3.4
3.8
4.2
4.6
1960
1970
1980
1990
Time
GNP
8.4
8.8
9.2
Interest rate
0.06
0.08
0.10
1960
1970
1980
1990
Time

Motivation

The cointegration residuals $\widehat{\varepsilon}_i$ are used to exclude trending regressors in the model.

Questions

1. Are the new observations after the monetary unification still consistent with this model? (*monitoring*)
2. Using the full data up to 1995:
   - Can a stable regression relationship be fitted to the full sample? (*testing*)
   - If not: When and how did the regression relationship change? (*dating*)

Model frame

Consider a regression model with k-dimensional parameter θ_i for $y_i \,|\, x_i$. To fit the models to observations $i = 1, \dots, n$ an objective function $\Psi(y, x, \theta)$ is used such that

$$\widehat{\theta} \quad = \quad \underset{\theta}{\text{argmin}} \sum_{i=1}^{n} \Psi(y_i, x_i, \theta).$$

This can also be defined implicitly based on the corresponding score function (or estimating function) $\psi(y, x, \theta)$

$$\psi(y, x, \theta) \quad = \quad \frac{\partial \Psi(y, x, \theta)}{\partial \theta}$$

via

$$\sum_{i=1}^{n} \psi(y_i, x_i, \widehat{\theta}) \quad = \quad 0.$$

Model frame

This class of M-estimators includes OLS and maximum likelihood (ML) estimation as well as IV, Quasi-ML, robust M-estimation and is closely related to GMM.

Under parameter stability and some mild regularity conditions, a central limit theorem holds

$$\sqrt{n}(\widehat{\theta} - \theta_0) \quad \xrightarrow{\text{d}} \quad \mathcal{N}(0, V(\theta_0)),$$

where the covariance matrix

$$V(\theta_0) = \{A(\theta_0)\}^{-1} B(\theta_0) \{A(\theta_0)\}^{-1}$$

and A and B are the expectation of the derivative of ψ and its variance respectively.

Model frame

For the standard linear regression model

$$y_i = x_i^\top \beta + u_i$$

with coefficients β and error variance σ^2 one can either treat σ^2 as a nuisance parameter $\theta = \beta$ or include it as $\theta = (\beta, \sigma^2)$.

In the former case, the estimating functions are $\psi = \psi_\beta$

$$\psi_\beta(y, x, \beta) = (y - x^\top \beta)\, x$$

and in the latter case, they have an additional component

$$\psi_{\sigma^2}(y, x, \beta, \sigma^2) = (y - x^\top \beta)^2 - \sigma^2.$$

and $\psi = (\psi_\beta, \psi_{\sigma^2})$.

Fluctuation tests

To assess the stability of the fitted model with $\widehat{\theta}$, we want to test the null hypothesis

$$H_0: \quad \theta_i = \theta_0 \qquad (i = 1, \dots, n)$$

against the alternative that θ_i varies over "time" i.

Various patterns of deviation from H_0 are conceivable: single/multiple break(s), random walks, etc.

To test this null hypothesis, the basic idea is to assess wether the empirical estimating functions $\widehat{\psi}_i = \psi(y_i, x_i, \widehat{\theta})$ deviate systematically from their theoretical zero mean.

Fluctuation tests

- empirical fluctuation processes reflect fluctuation in
  - estimating functions
  - F statistics
  - parameter estimates
  - recursive residuals
- theoretical limiting process is known
- choose boundaries which are crossed by the limiting process (or some functional of it) only with a known probability α.
- if the empirical fluctuation process crosses the theoretical boundaries the fluctuation is improbably large $\Rightarrow$ reject the null hypothesis.

Fluctuation tests

To capture systematic deviations the empirical fluctuation process of scaled cumulative sums of empirical estimating functions is computed:

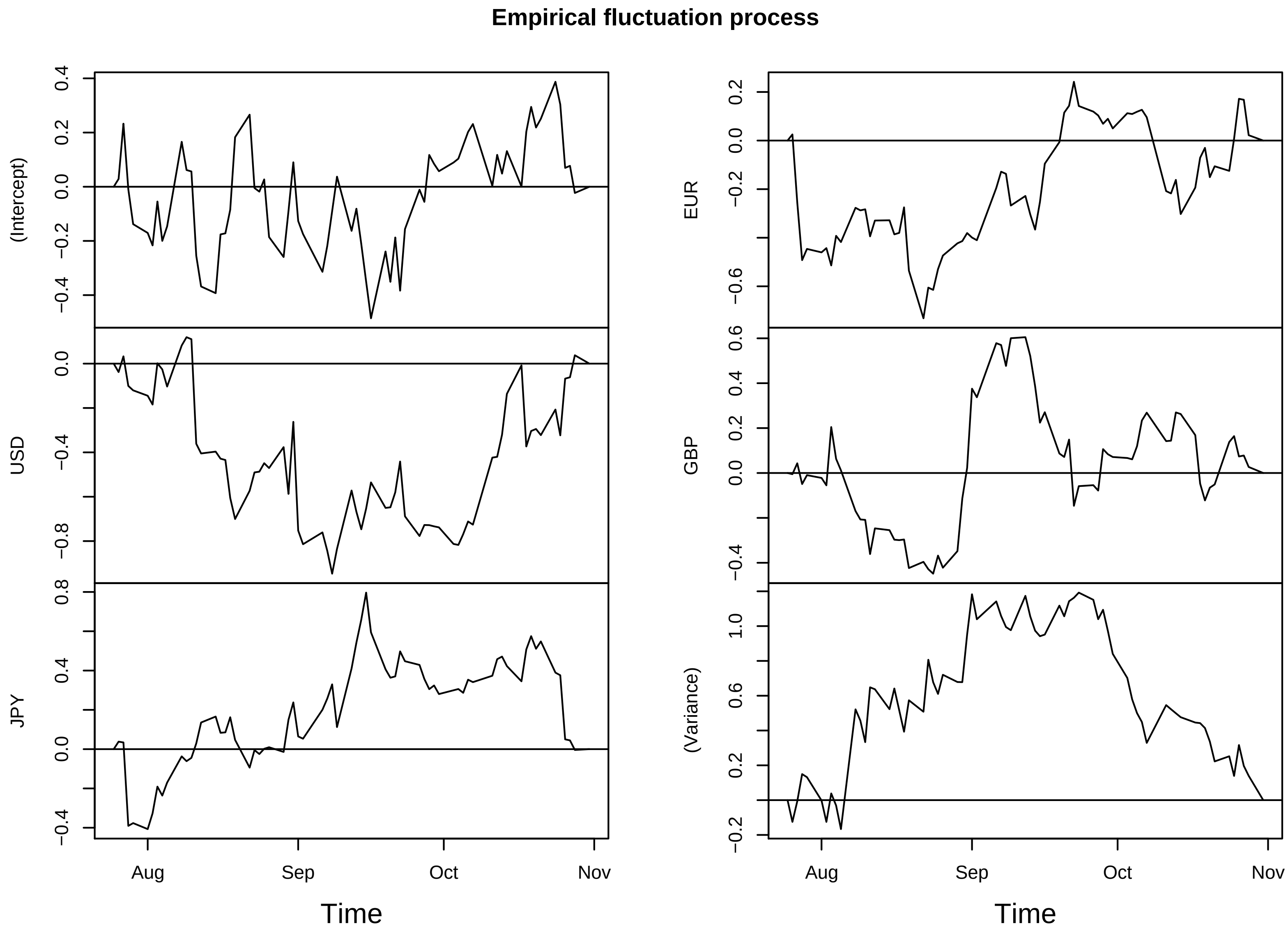
$$efp(t) \quad = \quad \widehat{B}^{-1/2} \, n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} \widehat{\psi}_i \qquad (0 \le t \le 1).$$

Under H_0 the following functional central limit theorem (FCLT) holds:

$$efp(\cdot) \quad \xrightarrow{\text{d}} \quad W^0(\cdot),$$

where W^0 denotes a standard k-dimensional Brownian bridge.

Fluctuation tests



Fluctuation tests

In empirical samples, $efp(\cdot)$ is a $k \times n$ array. For significance testing, aggregate it to a scalar test statistic by a functional $\lambda(\cdot)$

$$\lambda\left(efp_j\left(\frac{i}{n}\right)\right),$$

where $j = 1, \dots, k$ and $i = 1, \dots n$.

Typically, $\lambda(\cdot)$ can be split up into

- $\lambda_{\mathsf{comp}}(\cdot)$ aggregating over components j (e.g., absolute maximum, Euclidian norm),
- $\lambda_{\mathsf{time}}(\cdot)$ aggregating over time i (e.g., max, mean, range).

The limiting distribution is given by $\lambda(W^0)$ and can easily be simulated (or some closed form results are also available).

Fluctuation tests

The generalized fluctuation test framework ...

"... includes formal significance tests but its philosophy is basically that of data analysis as expounded by Tukey. Essentially, the techniques are designed to bring out departures from constancy in a graphic way instead of parametrizing particular types of departure in advance and then developing formal significance tests intended to have high power against these particular alternatives." (Brown, Durbin, Evans, 1975)

Fluctuation tests

Aggregating over time first, yields k independent statistics, each associated with one parameter $\Rightarrow$ *component of change can be identified*.

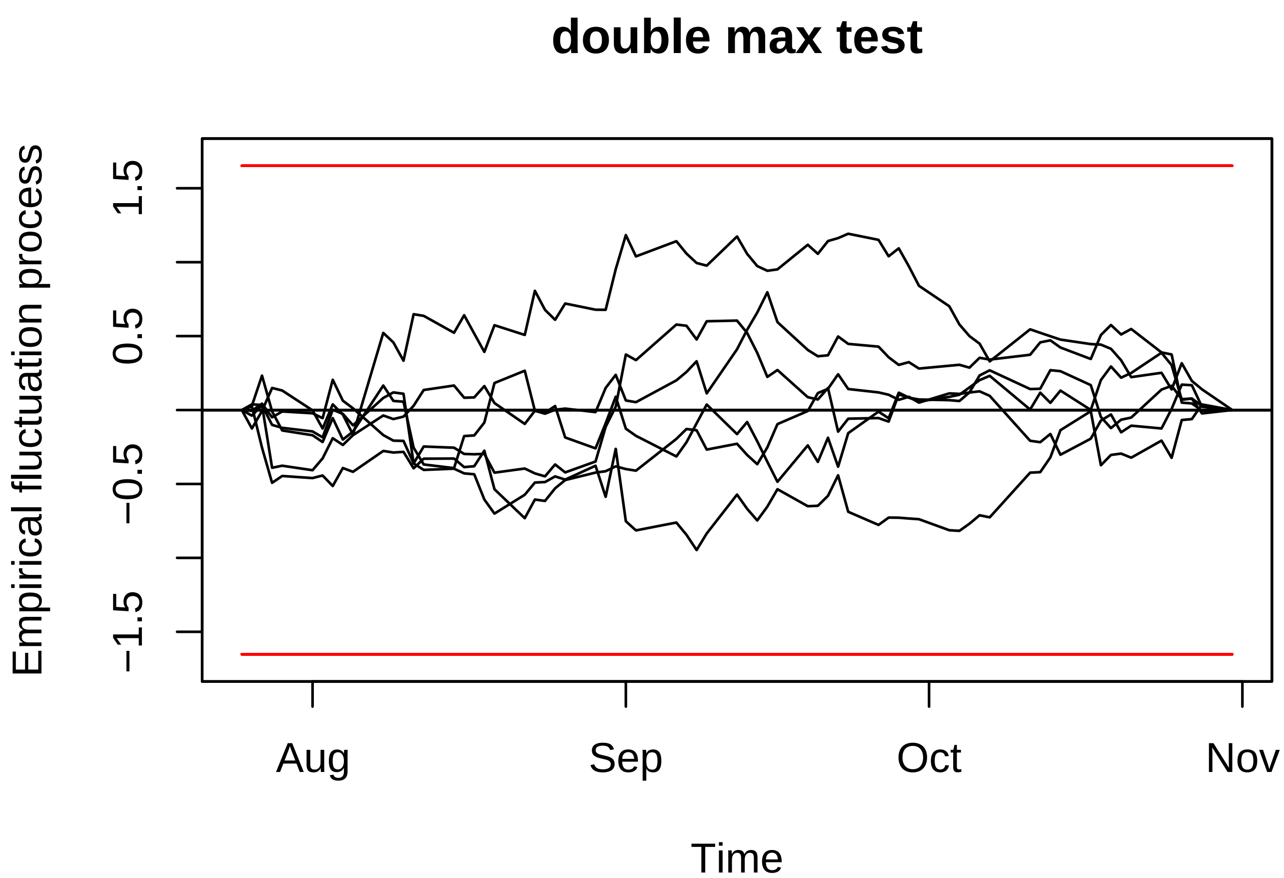
Aggregating over components first, yields a single process that can be inspected for instabilities over time $\Rightarrow$ *timing of change can be identified*.

The only functional that allows for both interpretations is the double maximum functional

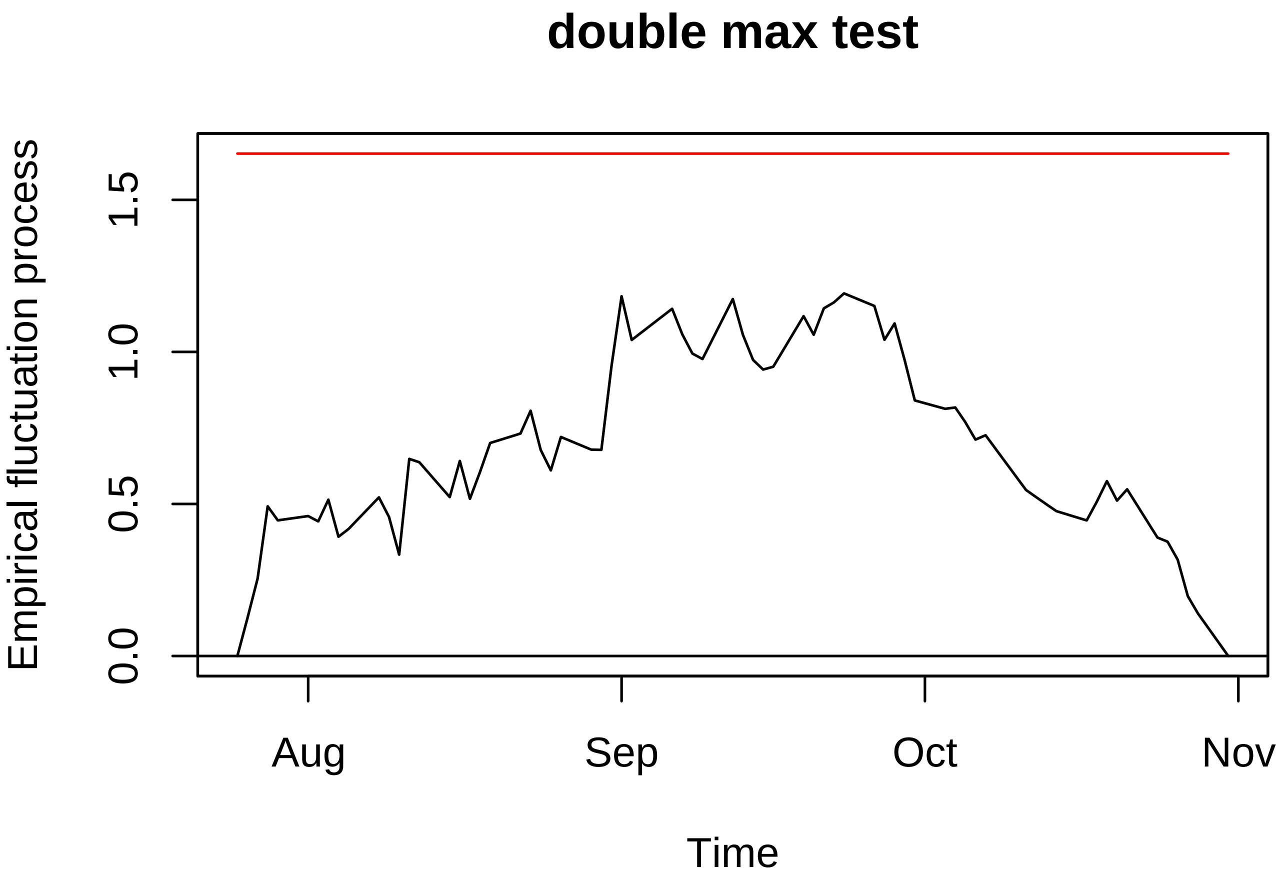
$$\max_{j=1,\dots,k} \max_{i=1,\dots,n} |efp_j(i/n)|$$

which is 1.192 for the Chinese currency data ($p = 0.524$).

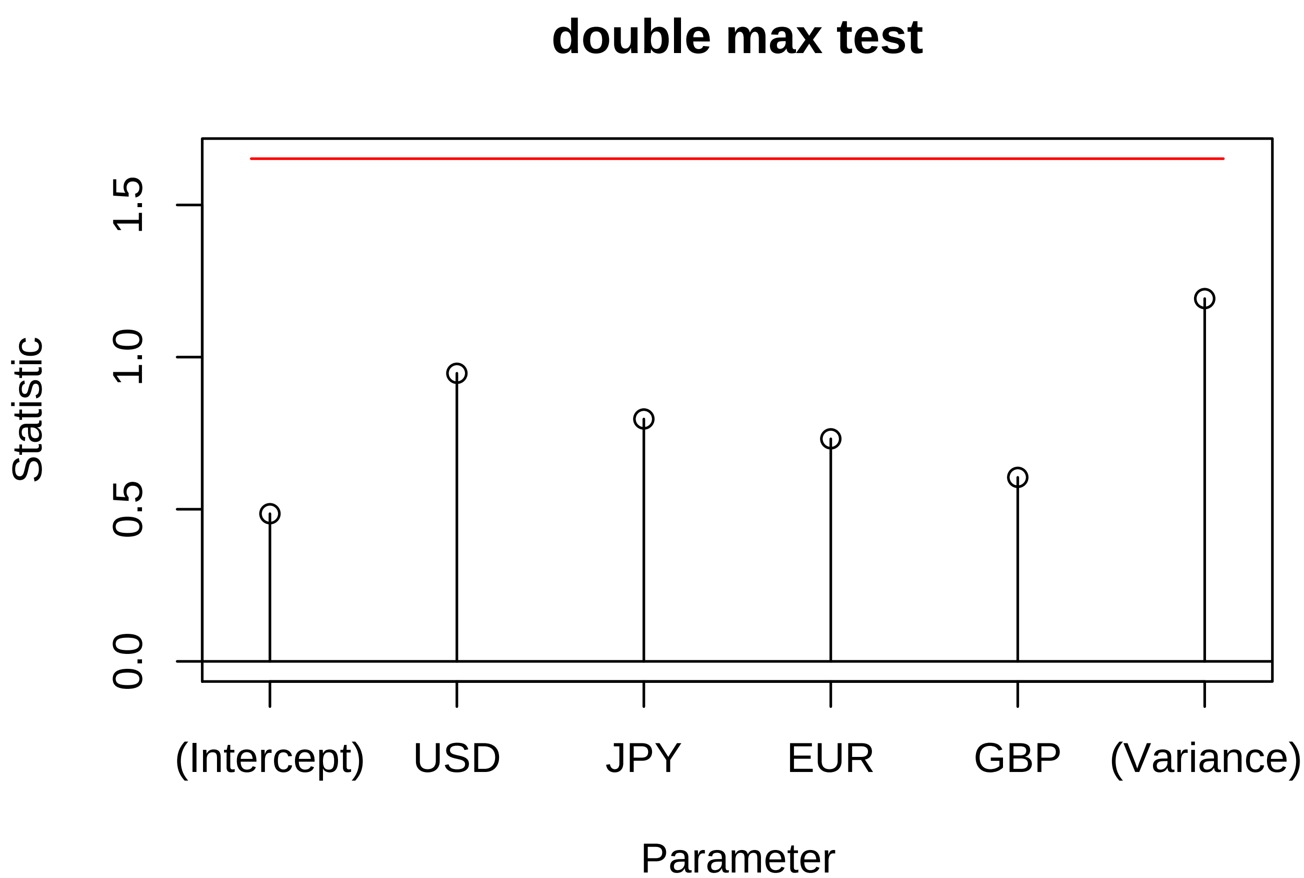
Fluctuation tests



Fluctuation tests



Fluctuation tests



Fluctuation tests

This class of generalized M-fluctuation tests is derived in Zeileis & Hornik (2003). It contains various well-known tests from the statistics and econometrics literature as special cases (Zeileis 2005).

OLS-based CUSUM test (Ploberger & Krämer 1992, 1996)
If the regression model contains an intercept the first component of the estimating functions corresponds to the OLS residuals

$$\psi(y, x, \theta)_1 \quad = \quad (y - x^\top \beta)$$

The process based on OLS residuals can capture changes in the conditional mean.

Fluctuation tests

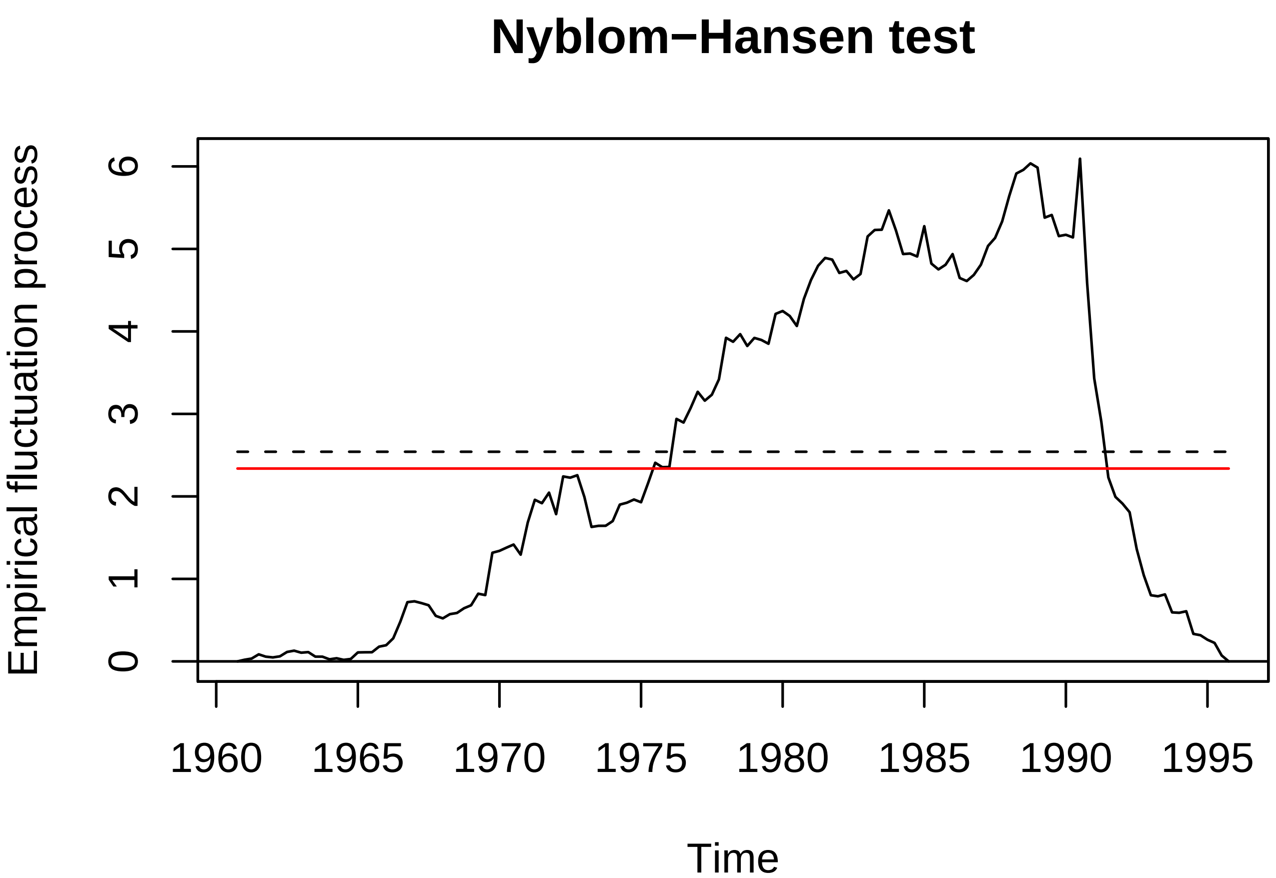
Nyblom-Hansen test (Nyblom 1989, Hansen 1992)
The test was designed for a random-walk alternative and employs a Cramér-von Mises functional.

It aggregates $efp(\cdot)$ over the components first, using the squared Euclidian norm, and then over time, using the mean.

$$\frac{1}{n}\sum_{i=1}^{n}\left\|efp\left(\frac{i}{n}\right)\right\|_2^2.$$

For the German M1 ECM this is 2.541 ($p = 0.026$).

Fluctuation tests



Fluctuation tests

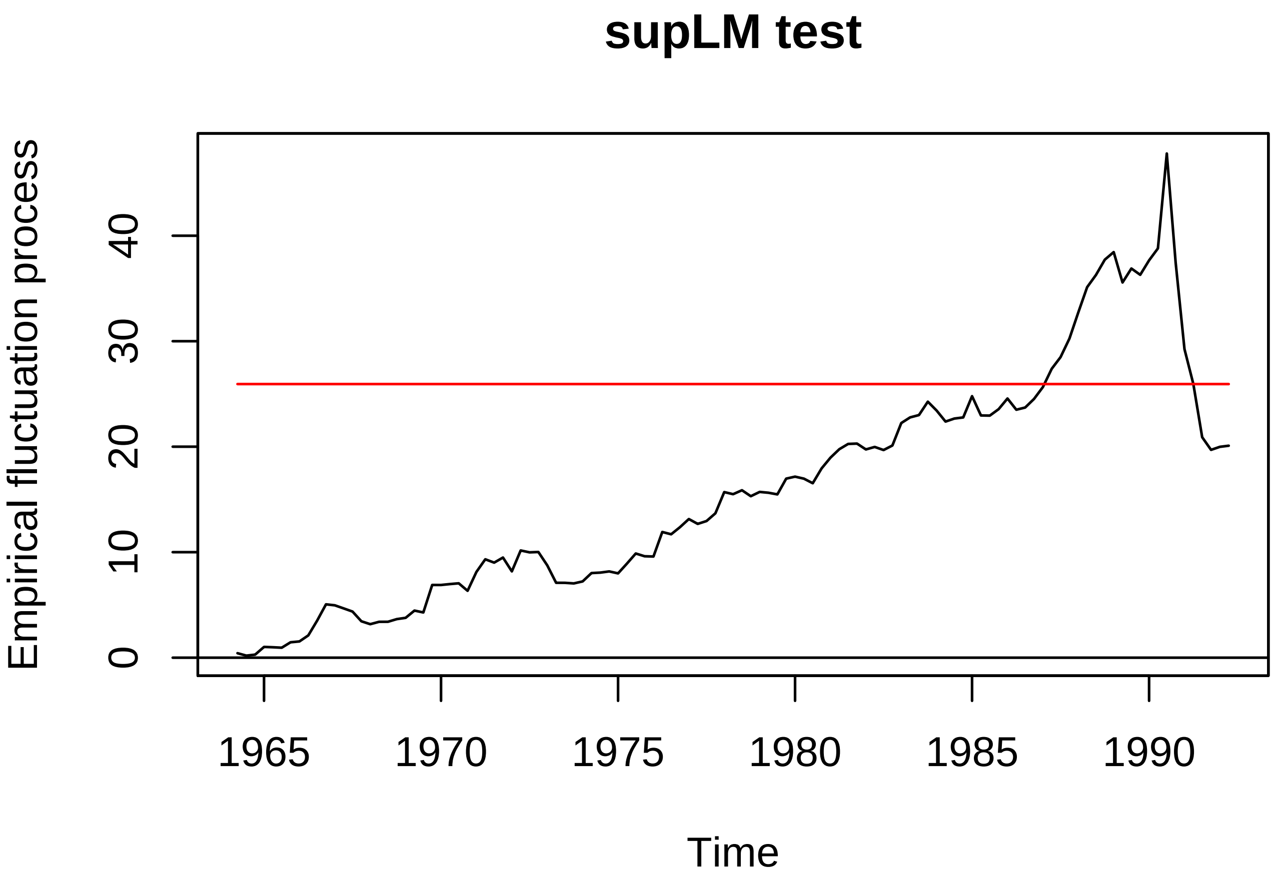
supLM test (Andrews 1993)
The test was designed for a single shift alternative (with unknown timing) and employs the supremum of the LM statistics for this alternative.

It aggregates $efp(\cdot)$ over the components first, using a weighted squared Euclidian norm, and then over time, using the maximum (over a compact interval $\Pi \subset [0, 1]$).

$$\sup_{t \in \Pi} LM(t) \quad = \quad \sup_{t \in \Pi} \frac{||efp(t)||_2^2}{t\,(1-t)}.$$

For the German M1 ECM this is 47.794 ($p < 0.001$).

Fluctuation tests



Fluctuation tests

Similarly, aveLM and expLM statistics can be computed which have some optimality properties (Andrews & Ploberger 1994).

The same type of processes can be derived based on Wald and LR statistics. These are not special cases of the M-fluctuation framework because they require re-estimation of the parameters on sub-samples. However, the asymptotic behaviour is the same.

Similarly, fluctuation processes can be based on recursive or rolling parameter estimates (Ploberger, Krämer, Kontrus 1989; Chu, Hornik, Kuan 1995) which are also not special cases but have the same asymptotic properties.

Fluctuation tests

Furthermore, fluctuation processes can be computed from recursive residuals (Brown, Durbin, Evans 1975; Bauer & Hackl 1978; Krämer, Ploberger, Alt 1988) which are similar in spirit, but have slightly different asymptotic properties.

In practice, this multitude of processes and functionals is often a curse rather than a blessing. *Which combination of process and functional should be used?*

Fluctuation tests

Which process?

The M-fluctuation tests only depend on one fitted model and are hence very easy to compute and interpret $\Rightarrow$ ideal for diagnostic checking, especially with no particular alternative in mind.

If a single shift alternative is plausible, it might be worth the effort to compute a sequence of Wald (or LR) statistics.

Recursive and rolling approaches are most plausible in a monitoring setup where a model should be simultaneously updated and tested.

Fluctuation tests

Which functional?

For explorative purposes, (double) maximum tests are most attractive and allow identification of component and timing of changes.

For random walk or single shift alternatives respectively (that affect all parameters), the Nyblom-Hansen or supLM functionals are most suitable.

For multiple shift alternatives, statistics based on moving sums (increments of $efp(\cdot)$) perform very well.

Monitoring with fluctuation tests

Fluctuation tests can be applied sequentially to monitor regression models (Chu, Stinchcombe, White 1996; Leisch, Hornik, Kuan 2000; Zeileis *et al.* 2005).

Basic assumption: The model parameters are stable $\theta_i = \theta_0$ in the history period $i = 1, \dots, n$ ($0 \le t \le 1$).

To assess whether the model remains stable we want to test the null hypothesis

$$H_0: \quad \theta_i = \theta_0 \qquad (i > n)$$

against the alternative that θ_i changes at some time in the future. The observations $i > n$ correspond to $t > 1$.

Monitoring with fluctuation tests

In the monitoring period $1 \leq t \leq T$:

- use the same empirical fluctuation processes,
- update $efp(t)$,
- re-compute $\lambda(efp(t))$.

For this sequential procedure not only a single critical value is needed, but a full boundary function $b(t)$ that satisfies

$$1 - \alpha \quad = \quad \mathsf{P}(\lambda(W^0(t)) \leq b(t) \mid t \in [1, T])$$

Various boundary functions (or weighting functions) are conceivable (see Horváth *et al.* 2004; Zeileis *et al.* 2005) that can direct power to early or late changes or try to spread the power evenly.

Monitoring with fluctuation tests

Here, we use again the double maximum functional and employ a simple boundary that spreads the power rather evenly:

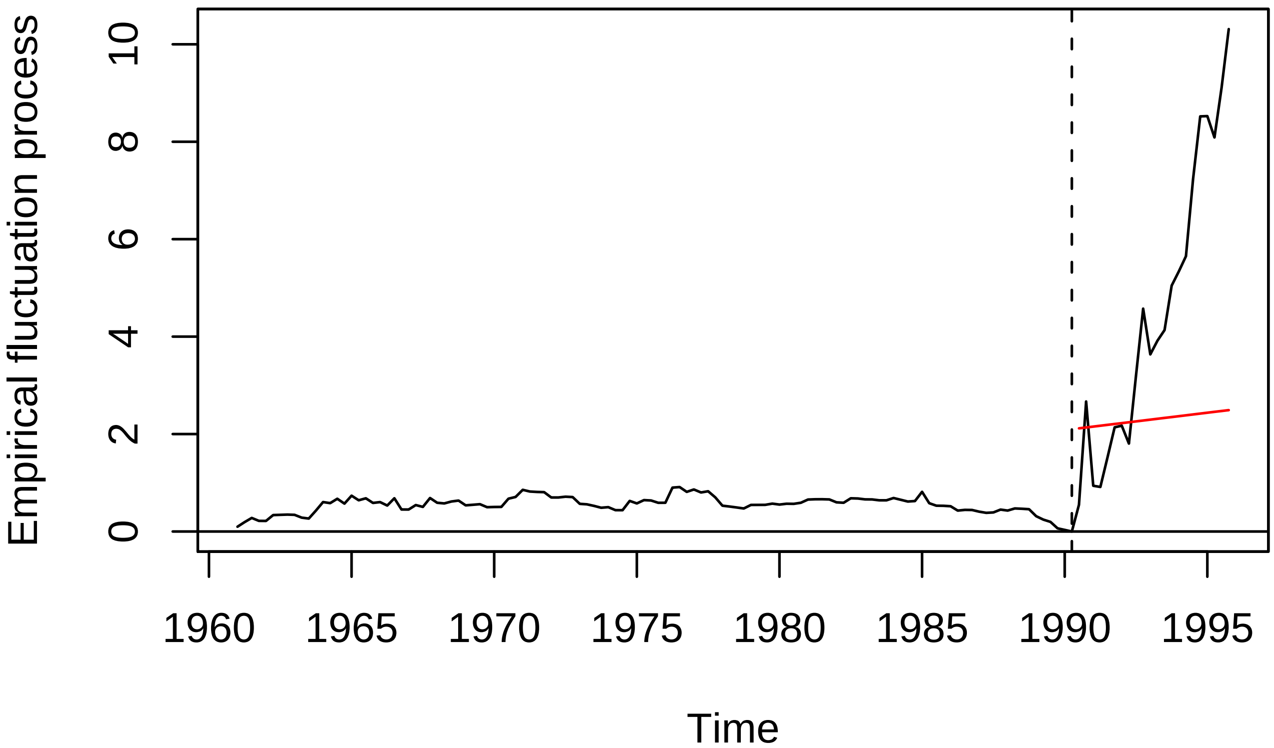
$$b(t) \quad = \quad c \cdot t,$$

where c controls the significance level.

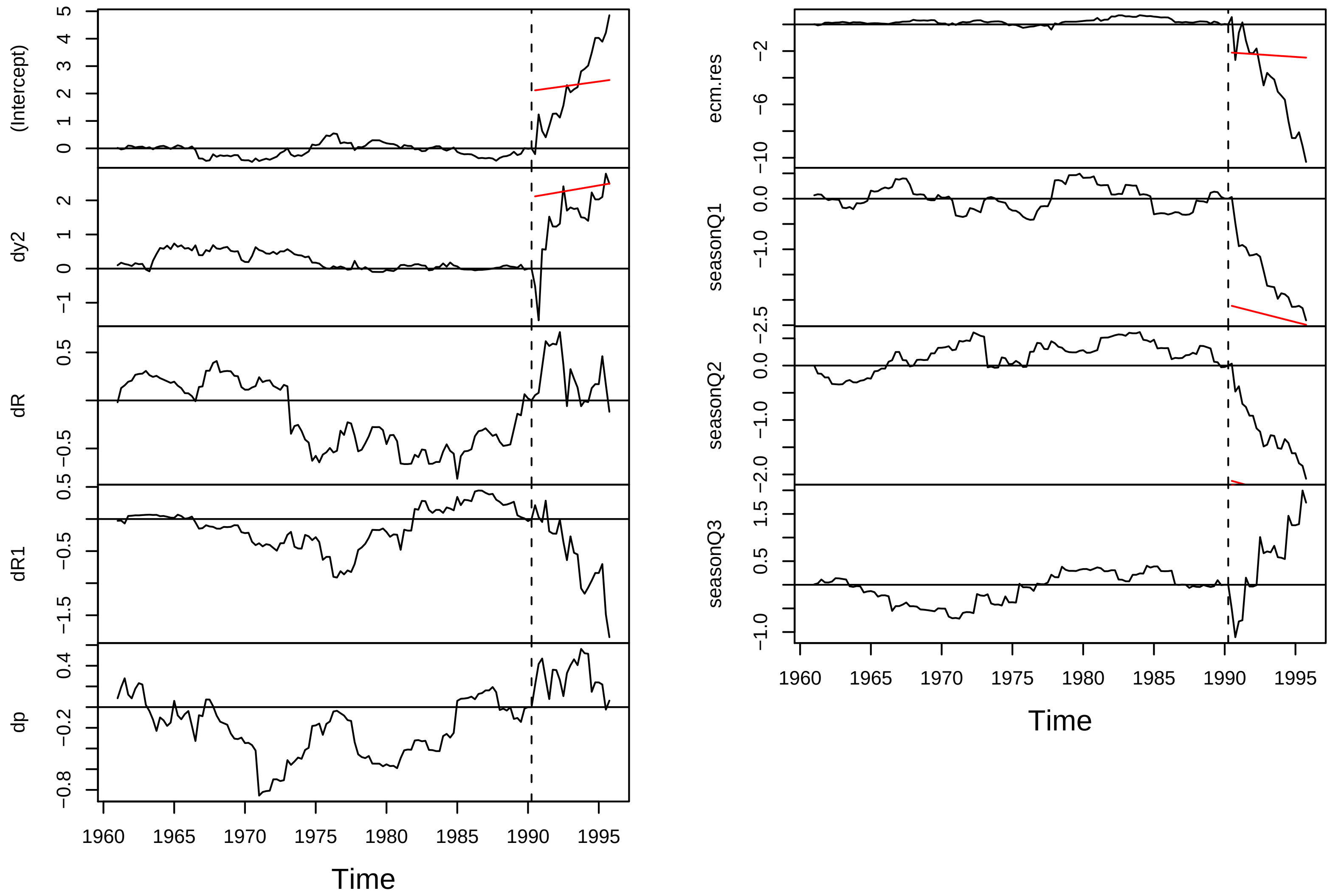
The fluctuation process used is the same M-fluctuation process used for the historical samples.

For the German M1 ECM, there is one peak already in 1990(4) and the processes clearly crosses its boundary again in 1992(3).

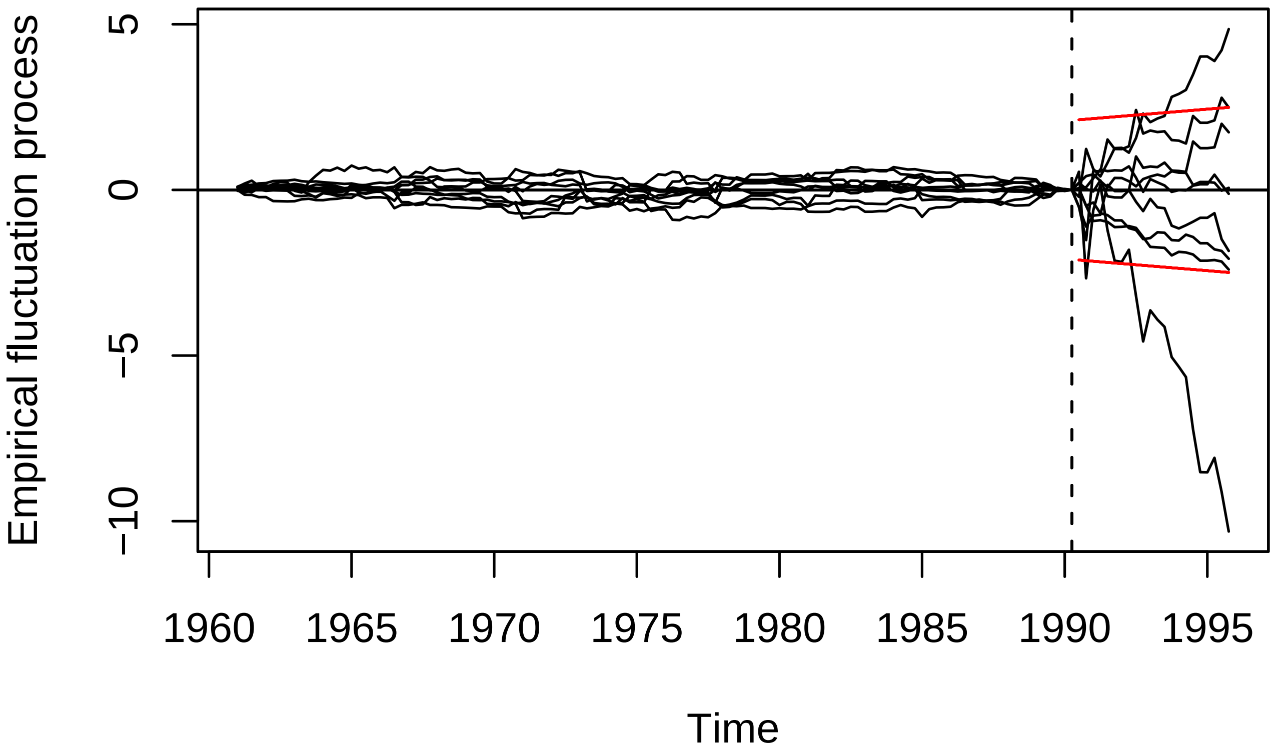
Monitoring with fluctuation tests



Monitoring with fluctuation tests



Monitoring with fluctuation tests

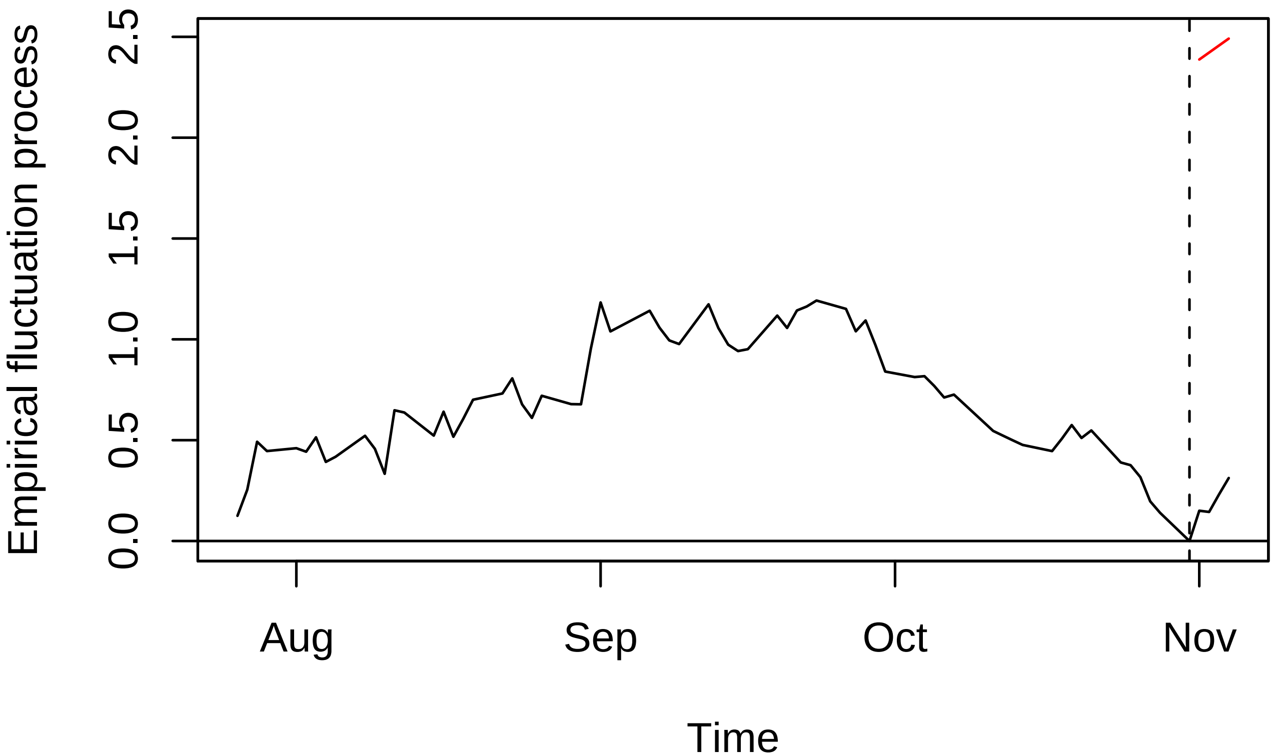


Monitoring with fluctuation tests

For the Chinese currency regime regression, there is not yet much data in the monitoring period.

Shah, Zeileis, Patnaik (2005) set up a Web page on which the monitoring process will be updated weekly. See `http://www.mayin.org/ajayshah/papers/CNY-regime/`

Monitoring with fluctuation tests



Dating structural changes

Segmented regression model: A stable model with parameter vector θ_j holds for the observations in $i = i_{j-1} + 1, \dots, i_j$. The segment index is $j = 1, \dots, m + 1$.

The set of m breakpoints $\mathcal{I}_{m,n} = \{i_1, \dots, i_m\}$ is called m-partition. Convention: $i_0 = 0$ und $i_{m+1} = n$.

The value of the objective function Ψ is

$$PSI(i_1, \dots, i_m) \quad = \quad \sum_{j=1}^{m+1} psi(i_{j-1} + 1, i_j),$$

Dating structural changes

where $psi(i_{j-1}+1, i_j)$ is the minimal value of the objective function for the model fitted on the jth segment.

Dating tries to find

$$(\hat{i}_1, \dots, \hat{i}_m) \quad = \quad \underset{(i_1,\dots,i_m)}{\operatorname{argmin}} PSI(i_1, \dots, i_m)$$

over all partitions $(i_1, \dots, i_m)$ with $i_j - i_{j-1} \geq \lfloor nh \rfloor \geq k$

Bellman principle of optimality:

$$PSI(\mathcal{I}_{m,n}) \quad = \quad \min_{mn_h \leq i \leq n-n_h} [PSI(\mathcal{I}_{m-1,i}) + psi(i+1, n)]$$

Dating structural changes

It is long-known that this problem can be solved by a dynamic programming algorithm of order $O(n^2)$ that essentially relies on a triangular matrix of $psi(i, j)$ for all $1 \leq i < j \leq n$.

The algorithm has been re-discovered several times in the literature, good overviews are given by Hawkins (2001) for likelihood-based dating and by Bai & Perron (2003) for OLS-based dating.

Dating structural changes

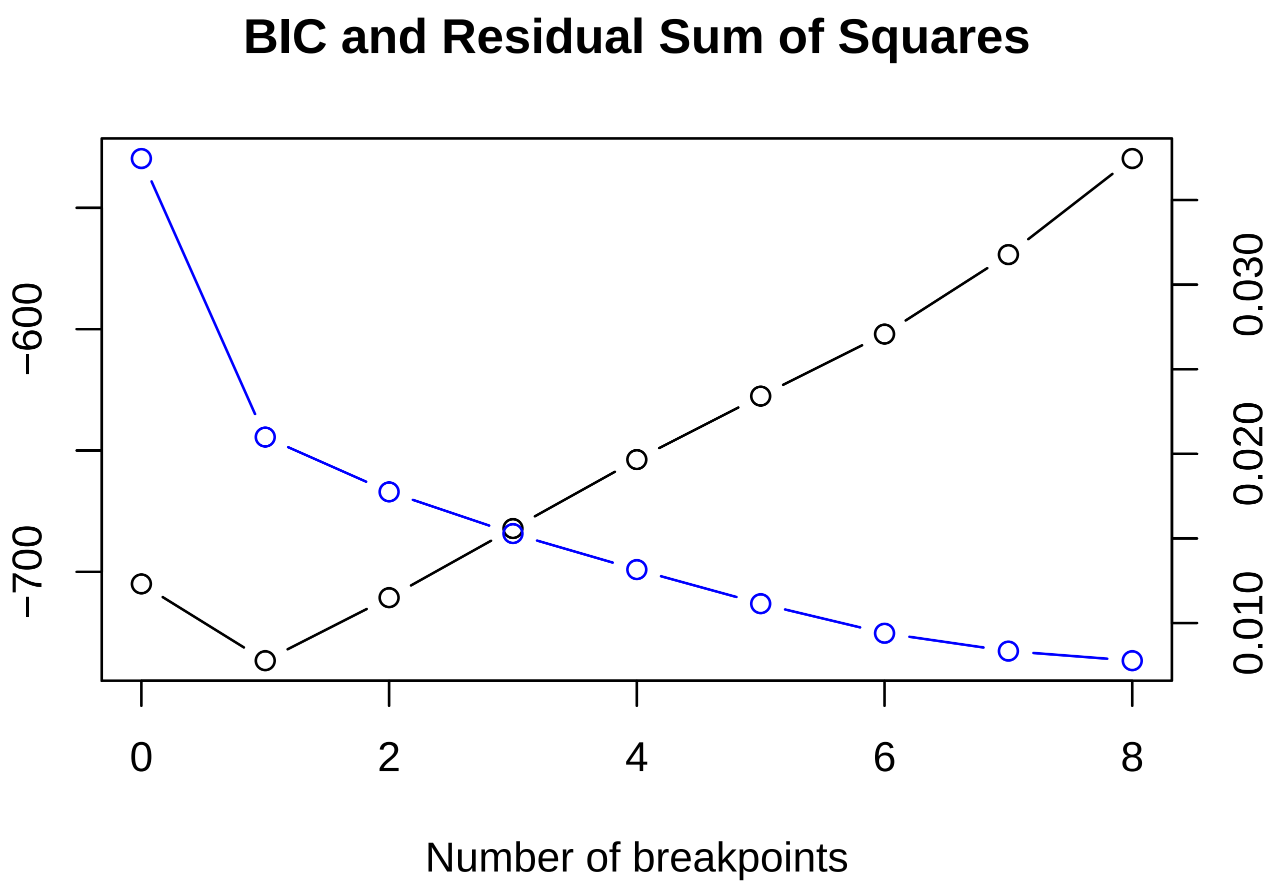
For a given number of breaks m, the optimal breaks can be found. But how should m be chosen?

The usual techniques for model selection can be applied here, e.g.

- information criteria,
- sequential tests.

Often, these do not work well out of the box, but should be handled with care and enhanced by other techniques.

Dating structural changes



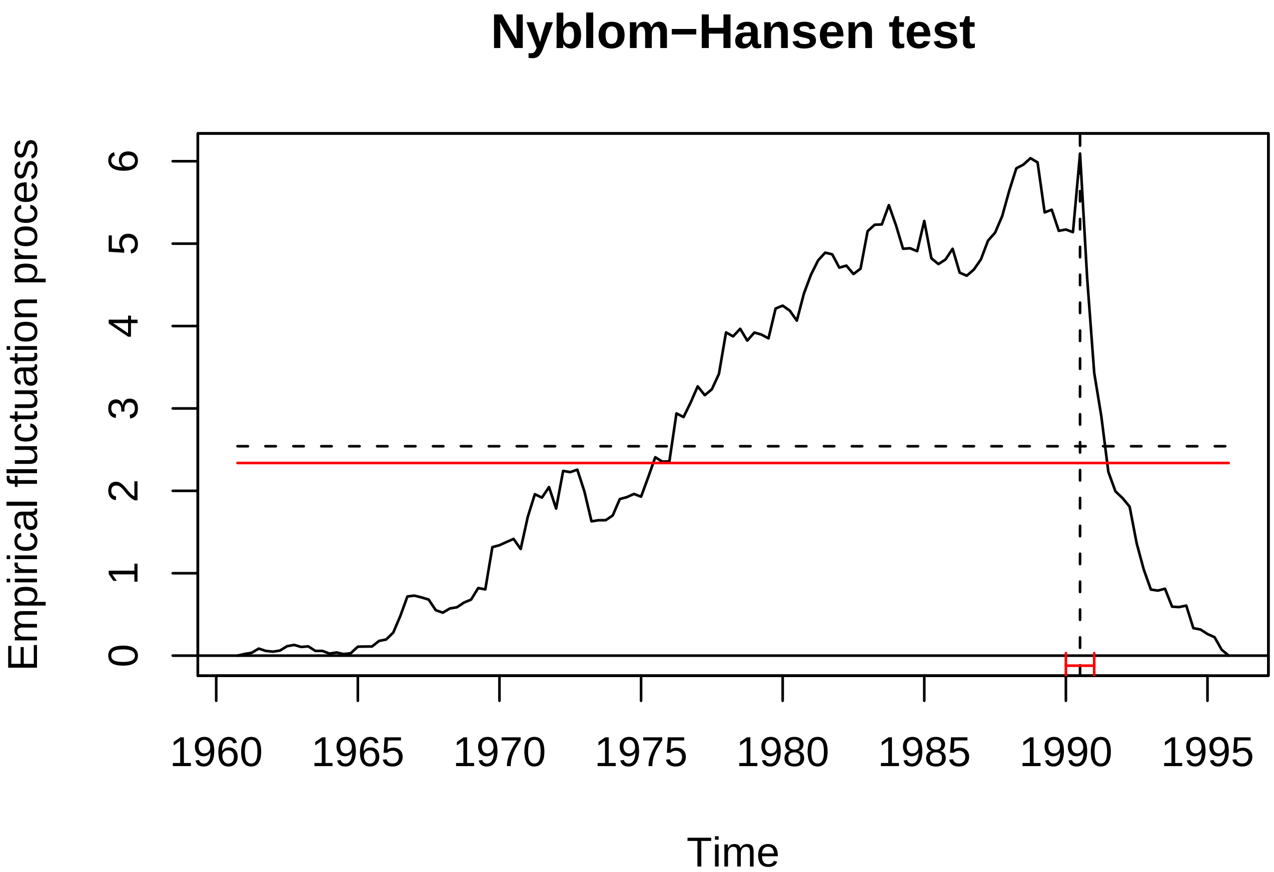
Dating structural changes

For the linear regression model, Bai & Perron (2003) provide the asymptotic distribution of the breakpoint estimators. It is the ratio of quadratic forms of Brownian bridges.

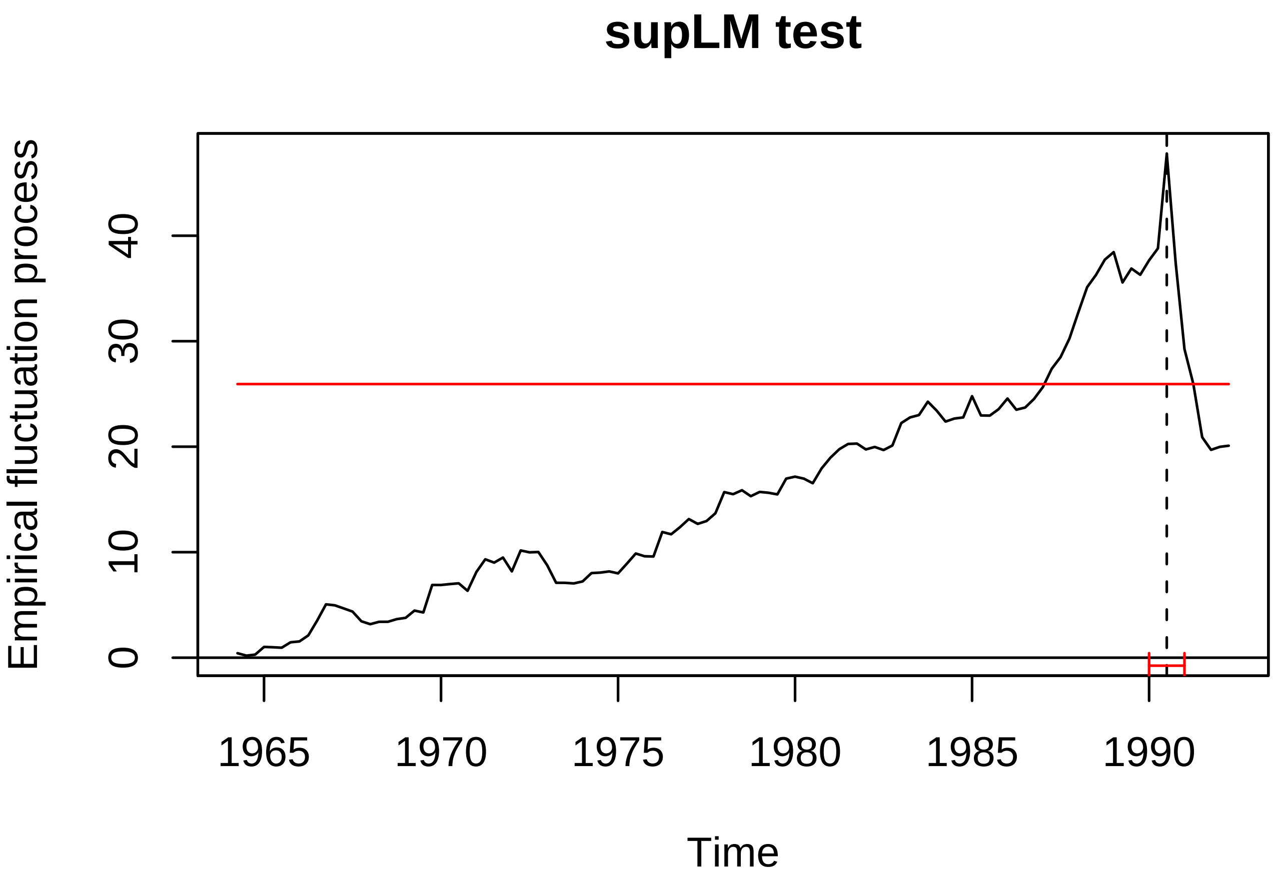
From this asymptotic distribution, confidence intervals can be computed.

For the German M1 ECM, the estimated breakpoint is 1990(3). A 95% confidence interval is 1990(1) to 1991(1).

Dating structural changes



Dating structural changes



Software

All methods are implemented in the R system for statistical computing and graphics (R Development Core Team 2005)

http://www.R-project.org/

in the contributed package **strucchange** (Zeileis *et al.* 2002).

Both are available under the GPL (General Public Licence) from the Comprehensive R Archive Network (CRAN):

http://CRAN.R-project.org/

Summary

- Given a model fitted by M-type estimator, parameter stability can be captured in partial sums of the empirical estimating functions.
- FCLT is the basis for inference.
- Significance tests can be constructed by aggregating the empirical fluctuation process to a scalar test statistic.
- Traditional significance tests can be enhanced by graphics that bring out the timing and/or the component of the structural change.
- Framework includes (representatives) from all important classes of structural change tests based on F statstistics, ML scores, OLS residuals.

Summary

- The same processes can be used for sequential monitoring of structural changes.
- Boundary functions can direct power against early/late changes.
- If an (additive) objective function exists (anti-derivative of estimating functions), the model can also be optimally segmented.
- Breakpoints are determined by a dynamic programming algorithm.
- Confidence intervals can also be computed.
- Software is freely available, works out of the box for linear regression, can be adapted to more general models.

References

Andrews DWK (1993). “Tests for Parameter Instability and Structural Change With Unknown Change Point.” *Econometrica*, **61**, 821–856.

Andrews DWK, Ploberger W (1994). “Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative.” *Econometrica*, **62**, 1383–1414.

Bai J, Perron P (2003). “Computation and Analysis of Multiple Structural Change Models.” *Journal of Applied Econometrics*, **18**, 1–22.

Bauer P, Hackl P (1978). “The Use of MOSUMS for Quality Control.” *Technometrics*, **20**, 431–436.

Brown RL, Durbin J, Evans JM (1975). “Techniques for Testing the Constancy of Regression Relationships over Time.” *Journal of the Royal Statistical Society B*, **37**, 149–163.

Chu CSJ, Hornik K, Kuan CM (1995). “The Moving-Estimates Test for Parameter Stability.” *Econometric Theory*, **11**, 669–720.

Chu CSJ, Stinchcombe M, White H (1996). “Monitoring Structural Change.” *Econometrica*, **64**(5), 1045–1065.

Hansen BE (1992). “Testing for Parameter Instability in Linear Models.” *Journal of Policy Modeling*, **14**, 517–533.

Hawkins DM (2001). “Fitting Multiple Change-Point Models to Data.” *Computational Statistics & Data Analysis*, **37**, 323–341.

References

Horváth L, Huškova M, Kokoszka P, Steinebach J (2004). "Monitoring Changes in Linear Models." *Journal of Statistical Planning and Inference*, **126**, 225–251.

Krämer W, Ploberger W, Alt R (1988). "Testing for Structural Change in Dynamic Models." *Econometrica*, **56**(6), 1355–1369.

Leisch F, Hornik K, Kuan CM (2000). "Monitoring Structural Changes With the Generalized Fluctuation Test." *Econometric Theory*, **16**, 835–854.

Lütkepohl H, Teräsvirta T, Wolters J (1999). "Investigating Stability and Linearity of a German M1 Money Demand Function." *Journal of Applied Econometrics*, **14**, 511–525.

Nyblom J (1989). "Testing for the Constancy of Parameters Over Time." *Journal of the American Statistical Association*, **84**, 223–230.

Ploberger W, Krämer W (1992). "The CUSUM Test With OLS Residuals." *Econometrica*, **60**(2), 271–285.

Ploberger W, Krämer W (1996). "A Trend-Resistant Test for Structural Change Based on OLS Residuals." *Journal of Econometrics*, **70**, 175–185.

Ploberger W, Krämer W, Kontrus K (1989). "A New Test for Structural Stability in the Linear Regression Model." *Journal of Econometrics*, **40**, 307–318.

R Development Core Team (2005). "R: A Language and Environment for Statistical Computing". R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

References

Shah A, Zeileis A, Patnaik I (2005). "What is the New Chinese Currency Regime?" *Report 23*, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Research Report Series. URL http://epub.wu-wien.ac.at/.

Zeileis A (2005). "A Unified Approach to Structural Change Tests Based on F Statistics, OLS Residuals, and ML Scores." *Report 13*, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Research Report Series. URL http://epub.wu-wien.ac.at/.

Zeileis A, Hornik K (2003). "Generalized M-Fluctuation Tests for Parameter Instability." *Report 80*, SFB "Adaptive Information Systems and Modelling in Economics and Management Science". URL http://www.wu-wien.ac.at/am/reports.htm#80.

Zeileis A, Leisch F, Hornik K, Kleiber C (2002). "**strucchange**: An R Package for Testing for Structural Change in Linear Regression Models." *Journal of Statistical Software*, **7**, 1–38. URL http://www.jstatsoft.org/v07/i02/.

Zeileis A, Leisch F, Kleiber C, Hornik K (2005). "Monitoring Structural Change in Dynamic Econometric Models." *Journal of Applied Econometrics*, **20**, 99–121.