Estimating the number of mean shifts under long memory*

Philipp Sibbertsen† and Juliane Willert

Institute of Statistics, Faculty of Economics and Management
Leibniz Universität Hannover, 30167 Hannover, Germany

Abstract

Detecting the number of breaks in the mean can be challenging when it comes to the long memory framework. Tree-based procedures can be applied to time series when the location and number of mean shifts are unknown and estimate the breaks consistently though with possible overfitting. For pruning the redundant breaks information criteria can be used. An alteration of the BIC, the LWZ, is presented to overcome long-range dependence issues. A Monte Carlo Study shows the superior performance of the LWZ to alternative pruning criteria like the BIC or LIC.

Keywords: long memory, mean shift, regression tree, ART, LWZ, LIC.

JEL-Codes: C14, C22

1 Introduction

The detection of changes in the mean is a fundamental issue for many areas of time series analysis. To specify the number and location of a mean shift can be even more challenging when the underlying framework consists of long memory behavior (see Sibbertsen (2004)). The high persistence in the time series with local trends and long cycles makes it hard for every

*Financial support by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

†Corresponding author, Phone: +49- 511- 762-3783, Fax: +49- 511- 762-3923, e-mail: sibbertsen@statistik.uni-hannover.de

breakpoint estimator. Therefore, the biggest challenge is distinguishing between the true long memory behavior and regular mean shifts. Undetected mean shifts can lead to misleading conclusions e.g. by biased estimation of the long memory parameter (see Granger and Hyung (1999) and Diebold and Inoue (2001) for further details).

Bai and Perron (1998) developed a method to specify number and location of mean shifts which is performing well in a short memory framework. Rea (2008) investigated that the Bai and Perron procedure does not work properly when it comes to long memory data. It tends to fail when high persistent behavior becomes too severe. To overcome this problem we adopt the fast approach of Breiman et al. (1984) via atheoretical regression trees (ART).

Regression trees split a time series into a left and right partition and continue by splitting the subpartitions recursively. The split choice is based on the location where the highest reduction in the residual sum of squares can be made. In this first phase the tree is spanned and builds a well overfitted tree of potential partitions and breakpoints (see Rea et al. (2010)). In the second phase the pruning technique tries to cut back branches with low contribution to the deviance reduction to locate the optimal partition of the time series.

The application of ART to time series analysis by Cappelli and Reale (2005) shows the enormous utility regarding break point analysis and opens a new perspective when it comes to structural break estimators. They showed that regression trees have reasonable performance in detecting and locating structural breaks. In comparison with Bai and Perron (1998) the least squares regression trees perform convincingly even in short-memory time series.

To locate the redundant mean shifts during the pruning phase of ART information criteria are used. Common pruning techniques such as the BIC fail when it comes to long memory behavior. Lavielle and Moulines (2002) suggested the LIC for the long memory case, which takes the long memory parameter into account. However, this requires a pre-specification that the underlying process is indeed long memory and an estimation of the long memory behavior when there are potential mean shifts coexistent. Thus a new information criteria, also Schwarz information criteria based, will be used to overcome this problem and still maintain the good properties of the regression trees to specify the number of mean shifts. The LWZ information criterion, first suggested by Liu, Wu and Zidek (1997), retains consistency but is constructed in a more flexible way with two parameters that are determined throughout the data generating process. It will be shown that it performs also in the long memory framework with superior results in comparison to the alternative pruning criteria.

The remainder of this paper is organized as follows. Section 2 outlines the tree-based procedure and their characteristics. Section 3 describes the new LWZ based pruning procedure and section 4 presents the results of the simulation study. It compares the LWZ with the procedure of Bai and Perron (1998, 2003) and the LIC (Laville and Moulines (2002)). Section 5 provides the conclusion.

2 Atheoretical regression trees

Atheoretical regression trees are used to detect and locate structural breaks. Using a nonparametric approach no distributional assumptions are required and a good fit to any kind of time series can be expected. Our break point model is defined by

$$\begin{aligned} y_t &= \mu + \varepsilon_t \\ \mu &= (\mu_1, \ldots, \mu_m) \\ \mu_k &= I_{(T_k+1<\ldots<T_{k+1})}\delta_k \text{ with } \delta_k \in \mathbb{R} \end{aligned}$$

where y_t is the value of the time series at time t, ε_t is the error term which is assumed to be stationary and μ_k is the mean of the time series in regime k up to the breakpoint m. The indicator function is 1 if you are in the regime k and 0 otherwise. $k = 1, \ldots, m$ are the breakpoints with the mean of the regime μ_k.

The regression tree determines breakpoints through fitting piecewise constant functions in an OLS regression framework. The exogenous predictor variable is the time t which works more like a counter than a predictor. At each regression step the best split of the time series is determined and an estimated breakpoint is not reconsidered but set fix in the further analysis. The determination of the best split is identified with a node impurity measurement. Usually the sum of squared residuals (RSS) is used as the risk function. The mean squared error is given by

$$R(t) = \frac{1}{n(t)} \sum_{x_i \in t} (y_i - \bar{y}(t))^2$$

with

$$\bar{y}(t) = \frac{1}{n(t)} \sum_{x_i \in t} y_i.$$

The predictor variable x_i represents the time points which belong to one regime and $n(t)$ is the number of elements in node t. A node symbolizes a part of the time series with length $n(t)$ i.e. the root node reflects the whole time series. To construct the tree a node t is split into a left child node t_L and a right child node t_R where the sum of the RSS of the left side and the right side of the node is minimized. That means we start by cutting the time series into two parts where the minimization of the RSS is highest. The minimization problem describes as follows.

$$\min_t (R(t_L) + R(t_R)) = \min_t \left(\frac{1}{n(t_L)} \sum_{x_i \in t_L} (y_i - \bar{y}(t_L))^2 + \frac{1}{n(t_R)} \sum_{x_i \in t_R} (y_i - \bar{y}(t_R))^2 \right)$$

The total sum of squares can be rewritten as a minimization of the within child nodes sum of squares. This can also be written as a maximization problem regarding the improvement through the splitting into t_L and t_R which maximally distinguishes the time series in the left and right nodes by generating the highest drop in deviance (see Rea et al. (2010)).

$$\max_t (R(t) - R(t_L) - R(t_R)) =$$
$$\max_t \left(\tfrac{1}{n(t)} \sum_{x_i \in t} (y_i - \bar{y}(t))^2 - \tfrac{1}{n(t_L)} \sum_{x_i \in t_L} (y_i - \bar{y}(t_L))^2 - \tfrac{1}{n(t_R)} \sum_{x_i \in t_R} (y_i - \bar{y}(t_R))^2 \right)$$

Each splitting process is a binary decision whether a node is found or not. This is applied separately to each subgroup recursively until no improvement of the criterion can be achieved. Thereby a hierarchical structure is build through the recursive partitionment of the time series into nodes and terminal nodes (leaves), where every terminal node represents a final regime with a shifted mean.

The growing process of the tree continues until no further improvement by splitting the time series can be made. In practice this would lead to as many terminal nodes as observations and therefore a minimum number of observations in each child node or a minimum within-node deviance is set. Denote in what follows the estimated break points by $\hat{\kappa} = (\hat{\kappa}_1, \ldots, \hat{\kappa}_m) = (\hat{T}_1/T, \ldots, \hat{T}_m/T)$ with true values $\kappa^0 = (\kappa_1^0, \ldots, \kappa_m^0)$. Under assumption 1 that b_T with $T \geq 1$ are nonnegative constants with probability one, we show adopting arguments similar to those in Bai and Perron (1998).

Assumption 1:

$$P_T(t) \geq b_T \frac{\log T}{T} \text{ for } T \geq 1 \text{ and } t \in \hat{T}_T \tag{1}$$

$P_T(t)$ denotes the empirical distribution of a random sample.

Lemma 1: Let ε_t be $I(d)$ with $d \in [0, 1/2)$. Then under assumption 1, $\hat{\kappa} \to \kappa^0$.

Proof: Denote by $\hat{\varepsilon}_t$ the estimated residuals

$$\hat{\varepsilon}_t = y_t - \hat{\mu}_k \qquad \text{for} \qquad t \in [\hat{T}_{k-1}+1, \hat{T}_k]. \tag{2}$$

Here, $\hat{\mu}_k = \bar{y}(t) = \frac{1}{n(t)} \sum_{t \in [\hat{T}_{k-1}+1, \hat{T}_k]} y_i$ and $n(t)$ gives the number of time points t in $[\hat{T}_{k-1}+1, \hat{T}_k]$. Thus in our model the mean is piecewise estimated with the arithmetic mean of the respective observations. It holds

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t \leq \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t. \tag{3}$$

Furthermore, we have with $d_t = \hat{\mu} - \mu^0$ for $t \in [\hat{T}_{k-1}+1, \hat{T}_k]$ and $\hat{\varepsilon}_t = \varepsilon_t - d_t$

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2 + \frac{1}{T}\sum_{t=1}^{T} d_t^2 - 2\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t d_t. \tag{4}$$

Using Lemma 1 in Bai and Perron (1998) which holds also in the long-memory context for $d < 1/2$ (Beran et al. (1998)) and states that $\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t d_t = o_P(1)$ and the equations (3) and (4) it can be seen that $\frac{1}{T}\sum_{t=1}^{T} d_t^2 \xrightarrow{P} 0$. This states that $\hat{\kappa}$ contains the correct break points among possible other incorrectly estimated mean shifts. Therefore, the regression tree is overfitted. However, pruning the tree by any under $I(d)$ consistent information criteria gives the desired consistency for the number and location of the mean shifts. $\diamondsuit$

3 Pruning by means of the LWZ information criterion

The process of pruning is the ex post discarding of branches whose proportion to the error reduction is negligible. In order to find out the optimal sequence of partitions and breakpoints of all candidates a model selection criteria can be employed. The well-established BIC fails in the presence of long-range dependencies. It retains its consistency but is outperformed in finite sample studies (Bai and Perron (2004)).

Lavielle and Moulines (2002) suggested an information criterion based on the bayesian information criterion that penalizes the estimation with a term including the long memory parameter d. The LIC is defined by

$$LIC = \min_{1 \leq k \leq m} \min_{\kappa_1,\ldots,\kappa_m} \sum_{k=1}^{m+1} \sum_{t=[\kappa_{k-1}T]+1}^{[\kappa_k T]} (y_t - \hat{\mu}_k)^2 + \frac{4k \log T}{T^{1-2d}}.$$

The penalization is chosen in order to obtain a consistent estimator for the change-point and balances the number of over- and underestimation (see Lavielle and Moulines (2002)). The information criterion is built exclusively for the long memory case and leads to a necessary pre-specification of the underlying framework. Also the long memory parameter has to be estimated without being biased through potential mean shifts.

Liu, Wu and Zidek (1997) suggested a modified Schwarz criterion to estimate the number of sections of their multivariate regression model which is denoted as LWZ. This criterion takes the form

$$LWZ(m) = \ln(S_T(\hat{T}_1,\ldots,\hat{T}_m)/(T-p^*)) + (p^*/T)c_o(\ln(T))^{2+\delta_0},$$

where $c_0 > 0$ and $\delta_0 > 0$ are some constants and p^* describes the total number of fitted parameters. T denotes the total number of observations and $\hat{T}_i$ the number of observations of regime i. The idea is to change the well-established Schwarz criterion as little as possible to retain consistency but also to embrace the desire to construct a more flexible information criterion accordingly.

By minimizing the sum of squares of the residuals a model dependent best criterion is given. A reasonable choice of c_0 and δ_0 is suggested by Liu, Wu and Zidek (1997) for short memory processes. They set a small δ_0 (=0.1) to reduce the potential risk of underestimation with a normal noise distribution and estimate $c_0 = 0.299$ by equalizing the LWZ to the Schwarz information criterion, but call for further research to develop a globally optimal pair of c_0 and δ_0 under a variety of specifications which will be done in section 4.

Bai and Perron (2004) show that the LWZ outperforms the BIC in all short memory cases including serial correlation. Under long memory the BIC is generally outperformed (see Rea (2008) and Rea et al. (2010) for demonstrative comparison). The classic BIC is therefore no competitor when it comes to performance questions.

4 Monte Carlo study

In the long memory context a simulation to specify a globally optimal pair for (c_0, δ_0) of the LWZ is done. Based on an ARFIMA(p,d,q) process with negligible short memory components for differentiation reasons no, one and two shifts in the mean of the time series is considered. For ten different values of the long memory parameter d (stationary and non-stationary) and a level shift height equally to the variance of the noise distribution (constantly 1) an overall distribution regarding the percentage of correctly specified break points is computed. Under normal, t- and double exponential noise all combinations are examined. Through a two-dimensional grid search procedure for all considered cases the optimal parameter pair ($c_0 = 0.26$, $\delta_0 = 3.76$, marked with a dot in figure 1) leads to 83% correct specifications. The performance deficit of 17% is based on high (nonstationary) d values and challenging break patterns when there are two mean shifts in the data.

Figure 1: Contour lines for correct specifications over all parameter combinations

See figure 1 for the contour plot of all considered parameter combinations. Yellow lines represent a low percentage of correct specifications and the more red the contour level line the higher the percentage of correct specification over all considered cases. The parameter combination with the highest percentage (83%) is marked with a dot in figure 1 and lies at $c_0 = 0.26$ and $\delta_0 = 3.76$. The LWZ would be accordingly

$$LWZ(m) = ln(S_T(\hat{T}_1, \ldots, \hat{T}_m)/(T - p^*)) + (p^*/T)0.26(ln(T))^{5.76}.$$

Not surprisingly, the penalization is typically higher than in the BIC (see Yao (1988)). As the BIC was constructed based on the iid case, the penalty term has to be somewhat stronger to balance the long-range dependance structure.

Besides an optimal parameter pair the graph also tells us that there is a rather wide central corridor for results of roughly equally good quality. That implies that the exact parameter combination is subordinate because of the stability of the results. The combination suggested by Liu, Wu and Zidek (1997) ($\delta_0 = 0.1$, $c_0 = 0.299$) is situated at the edge of the red corridor. Due to the fact that this combination leads to good results in the short memory case and outperforms the BIC, in general the LWZ is supposed to lead to good specification results as long as the penalty term is higher than the BIC.

In the short memory case the optimal parameter pair for long memory ($c_0 = 0.26$, $\delta_0 = 3.76$) leads to 89% correct specifications which makes the criterion safe to use for both frameworks without previous specification analysis. In the short memory case the *optimal* parameter pair would be a smaller value for c_0 with the same constant δ_0 or vice versa.

The Monte Carlo study serves as a comparison between the new adjusted LWZ criterion, the ordinary BIC as a benchmark information criterion and the LIC which is specialized in long memory cases. For ART we used tree growing procedures as implemented in 'tree' (Ripley (2005)) as a contributed package in the 'R' software. A time series with a length of 500 observations will be used and 100,000 replications are made since the computation time is not an issue.

The question that needs to be addressed after applying regression trees to time series according to Rea et al. (2010) is whether the pruning method under- or overestimates mean shifts and is robust against e.g. serial correlation. When there is no mean shift present in the data the results for the estimated number of mean shifts is given in table 1.

Table 1: Simulation results for pruning criteria when there are no mean shifts present

	LWZ			BIC			LIC		
d	% correct	mean	s.d.	% correct	mean	s.d.	% correct	mean	s.d.
0,05	100,00%	0,00	0,00	59,41%	0,65	0,95	50,23%	0,83	1,03
0,15	99,99%	0,00	0,00	18,86%	1,99	1,48	15,92%	2,15	1,49
0,25	99,94%	0,00	0,03	0,04%	3,36	1,52	0,03%	3,45	1,50
0,35	96,13%	0,04	0,19	0,00%	4,33	1,35	0,04%	4,36	1,33
0,45	77,62%	0,23	0,44	0,00%	4,87	1,18	0,10%	4,48	1,29
0,55	50,38%	0,57	0,65	0,00%	5,11	1,11	1,37%	3,26	1,45
0,65	27,21%	1,02	0,86	0,00%	5,17	1,10	6,14%	2,05	1,22
0,75	13,18%	1,56	1,09	0,00%	5,12	1,12	14,90%	1,34	0,93
0,85	5,91%	2,17	1,26	0,00%	4,99	1,15	24,92%	0,96	0,74
0,95	2,66%	2,70	1,32	0,00%	4,82	1,16	34,03%	0,76	0,65

The LWZ performs well when it comes to low and moderate long memory. For high values of d the increasing process variance of the underlying long memory tends to cover the true behavior of the mean. The BIC fails and tends to find at least one mean shift. The LIC develops a valley distribution. The shape of the estimation with the LIC is conditioned on the penalty term. With T^{2d-1} it degenerates for d values close to 0.5 and increases very strong for higher d values. For very small d values it performs well again because of the negligible long-range dependency. That's why for the LIC rather good results can be observed for low and high d values but not for moderate ones.

When it comes to a single mean shift at midpoint of the series the characteristics of the pruning criteria hold. For different break sizes that correspond to the standard deviation of the noise distribution ($s_{\varepsilon_t} = 1$) see table 2. The position of the mean shift does not affect the estimations strongly though mean shifts in the boundary area weaken every criterion.

Table 2: Simulation results for pruning criteria when there is one mean shift present

$\mu_1 = 1, \mu_2 = 3$	LWZ			BIC			LIC		
d	% correct	mean	s.d	% correct	mean	s.d.	% correct	mean	s.d.
0,05	100,00%	1,00	0,00	86,76%	1,14	0,38	86,76%	1,14	0,38
0,15	100,00%	1,00	0,00	49,20%	1,69	0,81	49,20%	1,69	0,81
0,25	99,90%	1,00	0,03	16,40%	2,63	1,13	16,40%	2,63	1,13
0,35	96,52%	0,97	0,18	3,92%	3,60	1,24	3,92%	3,60	1,24
0,45	82,65%	0,89	0,41	0,85%	4,35	1,23	1,04%	4,27	1,23
0,55	65,85%	0,91	0,64	0,27%	4,82	1,17	7,33%	3,54	1,37
0,65	54,00%	1,16	0,89	0,12%	5,03	1,14	29,05%	2,26	1,25
0,75	42,95%	1,63	1,11	0,09%	5,06	1,14	50,13%	1,42	0,96
0,85	29,96%	2,18	1,26	0,07%	4,96	1,15	58,60%	0,99	0,76
0,95	18,83%	2,71	1,32	0,07%	4,81	1,16	57,99%	0,77	0,65
$\mu_1 = 1, \mu_2 = 2$									
0,05	38,39%	0,38	0,49	60,40%	1,51	0,72	59,78%	1,52	0,72
0,15	40,94%	0,41	0,49	23,57%	2,40	1,11	22,99%	2,41	1,10
0,25	41,48%	0,41	0,49	6,36%	3,42	1,27	6,14%	3,42	1,27
0,35	42,73%	0,43	0,50	1,59%	4,22	1,26	1,54%	4,23	1,26
0,45	45,51%	0,48	0,53	0,45%	4,77	1,19	1,13%	4,49	1,25
0,55	50,50%	0,67	0,66	0,15%	5,04	1,13	10,22%	3,36	1,44
0,65	51,64%	1,06	0,87	0,09%	5,13	1,11	32,14%	2,10	1,23
0,75	43,09%	1,58	1,10	0,06%	5,10	1,13	50,64%	1,36	0,94
0,85	30,11%	2,17	1,27	0,07%	4,98	1,15	58,33%	0,98	0,75
0,95	19,08%	2,71	1,33	0,05%	4,82	1,16	57,50%	0,76	0,65

The BIC again performs inferior with an average break estimation higher than 1. The LIC holds its shape and outperforms the LWZ for several combinations. The problem of the LIC still holds that d has to be estimated first and therefore can lead due to the deviance of the criterion easily to false results in a practical setting. The LWZ stays comparably constant when the long memory parameter changes and tends to underestimate the number of mean shifts for stationary long memory. Tree-based procedures in general overfit for small breaks and short observation length (see Rea et al. (2010)), hence a criterion which does not exceed this behavior could be a more than welcome technique.

For more than one mean shift the criteria weaken and are highly dependent on the break size but fortunately not on the break pattern.

5 Empirical Application

To exemplify the benefits of break point detection with regression trees we analyze the monthly current account deficit reported in the balance of payment (BoP) including January 2012. The balance of payment is calculated according to the IMF (International Money Fund) and therefore defined as all transactions between a domestic economy and the rest of the world in a given time period (see IMF (2010)). It consists of the visible trade (total balance of goods), the invisible trade and transfers of funds which balances out even. These trades will be considered in the current account (deficits or surpluses) and the capital account excluding central's band reserve account (see Milesi-Ferretti and Razin (1996)).

It is non-controversial that the structural characteristics of current account deficits are persistent based on the concept that structural imbalances are not easy to suppress in the economy. Structural changes can be external crisis, wars and political instabilities, shifts in demand at key industries, changes in the rate of international capital movements or elementary technological innovations. This affects not only short term cycles but also long-run behavior of the deficit or surplus in the balance of payments of a country (see McCombie (1997), Moreno-Brid (1998) and Araujo and Lima (2007)). In addition Cunado et al. (2004) investigated the usage of fractional integrated models in this framework and could embrace their thoughts through testing and application to the US deficit.

Of interest is the presence of sudden changes in the structure of the mean. We allow for an endless number of breaks as long as the segment size contains at least 50 observations. The estimation of the break point via regression trees leads to the following results (see table 3).

Table 3: Break positions in current account deficits

Country	Start date	Break position (Month/Year)
Germany	1956	11 / 2003
Japan	1985	03 / 2003
Norway	1981	02 / 1999
UK	1955	11 / 1998
US	1960	02 / 1999

The results indicate that these countries mostly reacted to two different events. UK and US, among other countries such as Norway, Australia and South Korea responded to the dot-com bubble in 1999 (up to March 2000). And Germany and Japan struggled in the course of the third oil crisis due to their large focus of high-end (oil-dependent) technology exports. McCombie (1997) (investigating Japan, UK and US) pointed out that there has to be an awareness for structural breaks for current account deficits. Furthermore he could not rule out an I(1) behavior in the data. Looking at the persistence of the series before and after the structural break in the mean is taken into account it becomes clear that the observed non-stationary behavior can 'turned' into (stationary) long memory (see table 4).

Table 4: Long memory estimation in current account deficits

Country	d estimation	d estimation after demeaning
Germany	1,005	0,564
Japan	0,843	0,432
Norway	0,946	0,455
UK	0,688	0,428
US	1,212	0,635

Milesi-Ferretti and Razin (1996) point out that in the long-run the dynamics are mostly driven by the development of productivity differentials between traded and non-traded goods in an economy in comparison to the rest of the world. These persistent current account imbalances, refereed as the Balassa-Samuelson effect, are not compulsory a weakness of a domestic economy as long as the export sector and savings are large enough. Therefore it is also not implied that a large deficit leads to a crisis or that a crisis can only occur if a large current account deficit is present (see Summers (2000) and Araujo and Lima (2007)).

6 Conclusion

Estimating the number of mean shifts in a long-memory time series can be challenging. Tree-based procedures are presented as a powerful yet simple technique (see De'ath, G., Fabricius, K. (2000)) and are therefore useful for the practitioner (Rea et al. (2010)). To prune the over-fitting of atheoretical regression trees the BIC is widely used in a short memory framework and surprisingly outperformed by the LWZ under multiple specifications (Bai and Perron (2004)). The LIC which was derived for long memory shows good properties as well and partially outperforms the LWZ for some combinations of d. Though the disadvantages of the LIC to depend on the true value of d last. The LWZ keeps reasonable results even when the framework contains long memory and thus needs no beforehand knowledge of the data generating process. It is therefore preferable to the BIC and LIC.

References

Araujo, R.A., Lima, G.T. (2007): „A structural economic dynamics approach to balance-of-payments-constrained growth." *Cambridge Journal of Economics* 31(5), 755 -– 774.

Bai, J., and Perron, P. (1998): „Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica* 66, 47 -– 78.

Bai, J., and Perron, P. (2003): „Computation and analysis of multiple structural change models." *Journal of Applied Econometrics* 18, 1 -– 22.

Bai, J., and Perron, P. (2004): „Multiple Structural Change Models: A Simulation Analysis", In: *Econometric theory and practice: frontiers of analysis and applied research by Corbae, D., Durlauf, S.N. and Hansen B.E.* Cambridge, 212 – 237.

Beran, J. Bhansali, R.J. and Ocker, D. (1998): „On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes”, *Biometrika* 85(4), 921 – 934.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1993): *Classification and Regression Trees.* Chapman & Hall, New York.

Brown, R.L., Durbin, J. and Evans, J.M. (1975): „Techniques for testing the constancy of regression relationships over time.” *Journal of the Royal Statistical Society B* 37, 149 – 163.

Cappelli, C., Penny, R.N., Rea, W.S. and Reale, M. (2008): „Detecting multiple mean breaks at unknown points in official time series.” *Mathematics and Computers in Simulation* 78, 351 – 356.

Cappelli, C. and Reale, M. (2005): „Detecting Changes in Mean Levels with Atheoretical Regression Trees.” *Research Report UCMSD* 2005/2, Department of Mathematics and Statistics, University of Canterbury.

Chow, G.C. (1960): „Tests of equality between sets of coefficients in two linear regressions.” *Econometrica* 28, 591 – 605.

Cooper, S.J. (1998): „Multiple Regimes in US Output Fluctuations.” *Journal of Business and Economic Statistics* 28(3), 92 – 100.

Corvoisier, S. and Mojon, B. (2005): „Breaks in the Mean of Inflation: How they happen and what to do with them.” ECB Working Paper No. 451.

Cunado, J., Gil-Alana, L.A. and Pérez de Gracia (2004): „Is the US fiscal deficit sustainable?: A fractional integrated approach.” *Journal of Economics and Business* 56(6), 501 – 526.

da Rosa, J.C., Veiga, A. and Medeiros, M.C. (2008): „Tree-Structured Smooth Transition Regression Models Based on CART Algorithm.” *Journal of Computational Statistics and Data Analysis* 52, 2469 – 2488.

De'ath, G. and Fabricius, K. (2000): „Classification and regression trees: a powerful yet simple technique for ecological data analysis." *Ecology* 81(11), 3178 – 3192.

Diebold, F.X. and Inoue, A. (2001): „Long memory and regime switching." *Journal of Econometrics* 105, 131–159.

Granger, C. and Hyung, N. (1999): „Occasional Structural Breaks and Long Memory." *University of California, San Diego, Discussion Paper* 99.14.

Hansen, B.E. (1997): „Approximate asymptotic p values for structural-change tests." *Journal of Business & Economic Statistics* 15, 60 – 67.

Hsu, C.-C. (2005): „Long memory or structural changes: An empirical examination on inflation rates." *Economics Letters* 88, 289 – 294.

International Monetary Fund (2010): „Balance of payments and international investment position manual." *Washington, D.C., 6th ed.* (BMP6).

Kokoszka, P. and Leipus, R. (2002): „Detection and estimation of changes in regime" In: *Long-range Dependence: Theory and Applications by P. Doukhan, G. Oppenheim and M. S. Taqqu, eds.* Birkhauser, Boston, 325 – 337.

Lavielle, M. and Moulines, E. (2002): „Least-squares estimation of an unknown number of shifts in a time series." *Journal of Time Series Analysis* 21(1), 33 – 59.

Liu, J., Wu, S. and Zidek, J. (1997): „On segmented multivariate regression." *Statistica Sinica* 7, 497 – 525.

McCombie, J.S.L. (1997): „On the Empirics of Balance-of-Payments-Constrained Growth." *Journal of Post Keynesian Economics* 19(3), 345 – 375.

Milesi-Ferretti, G.M., Razin, A. (1996): „Sustainability of persistent current account deficits." *National Bureau of Economic Research Working Paper* 5467.

Moreno-Brid, J.C. (1998): „On capital flows and the balance-of-payments-constrained growth model." *Journal of Post Keynesian Economics* 21(2), 283 – 298.

Ploberger, W. and Krämer, W. (1992): „The CUSUM test with OLS residuals." *Econometrica* 60(2), 271 – 285.

R Development Core Team (2008): „R: A language and environment for statistical computing." Available at www.r-project.org.

Rea, W.S. (2008): „The Application of Atheoretical Regression Trees to Problems in Time Series Analysis." *PhD Thesis.* Department of Mathematics and Statistics, University of Canterbury.

Rea, W.S., Reale, M., Cappelli, C. and Brown, J.A. (2010): „Identification of Changes in Mean with Regression Trees: An Application to Market Research." *Econometric Reviews.* 29(5), 754 – 777.

Ripley, B. (2005): „tree: Classification and regression trees." *R package version 1.0-19.*

Schwarz, G. (1978): „Estimating the dimension of a model." *The Annals of Statistics* 6, 461 – 464.

Sibbertsen, P. (2004): „Long-memory versus structural change: An overview." *Statistical Papers* 45, 465 – 515.

Summers, L.H. (2000): „International Financial Crisis: Causes, Prevention, and Cures." *The American Economic Review* 90(2), 1 – 16.

Yao, Y. (1988): „Estimating the number of change-points via Schwarz criterion." *Statistical and Probability Letters* 6, 181 – 189.