Mathematisches Forschungsinstitut Oberwolfach

Report No. 12/2008

Mini-Workshop: Time Series with Sudden Structural Changes

Organised by
Richard Davis, New York
Jürgen Franke, Kaiserslautern

February 24th – March 1st, 2008

Abstract. The workshop was concerned with models and tools for nonstationary time series, in particular for those which are piecewise stationary, but where the stochastic structure of the data generating process sometimes changes suddenly. The problem of analyzing such time series data was investigated from various angles, e.g. from the viewpoint of changepoint analysis, of Markov switching models and of nonparametric statistics.

Mathematics Subject Classification (2000): 62xxx.

Introduction by the Organisers

Time series with nonstationarities can be found in a variety of applications from economics to engineering to biological and medical sciences to geosciences. During the past 20 years, much effort in investigating nonstationary time series has been focused on models with slowly changing structure. Such examples include autoregressive processes with slowly changing parameters and, more generally, locally stationary processes with a time-dependent spectrum. Research on these models has led to considerable insight into the structure of stochastic processes as well as parametric and nonparametric statistical methods. In particular, the fact that time series data are no longer assumed to be identically distributed has resulted in some challenging and interesting statistical problems.

Typically locally stationary models are not well suited to cover the situation in which a time series may experience a sudden change in its dynamics. That is, there may be changes in the mean, variance, dependence structure, or some other feature that occur at random times. Often it is reasonable to assume that the time series is stationary between these *changepoints*. This assumption is tacitly made in the applications of classical changepoint analysis techniques to stochastic processes

of dependent data which have been proposed recently. It also forms the basis for segmentation techniques which partition the time axis such that the data in each subinterval may be modeled by a stationary process. Another approach explicitly takes the piecewise stationarity into account in the modeling step. There, the sudden changes in the dynamic structure of the data-generating process are controlled by a hidden Markov chain with finitely many states resulting in hidden Markov models, or, more generally, in Markov switching models. Such models are becoming increasingly popular in the analysis of economic, environmental and other data.

Some progress has been recently made in the theory regarding the stochastic structure of such piecewise stationary time series models as well as regarding the properties of parametric and nonparametric statistical procedures for estimating the system characteristics, of estimating locations of changepoints and of reconstructing the unobservable state process, i.e. of finding an appropriate segmentation of the data. Much of this work also requires the development of algorithms that are numerical feasible for carrying out these procedures. Nevertheless, there are still many open questions, and one of the main goals of the workshop beyond the exchange of current results between different statistical communitites working on those problems, was to join forces to discuss ideas on how to attack various open problems and to investigate the relations between different approaches.

A combination of review talks as well as presentations of recent progress in theory and of challenging applications formed the basis for the exchange of ideas. Altogether, there were 16 talks which invariably were accompanied by a very lively discussion. The presented results and the raised questions served as inspiration for future research and as stimulation for joint projects. An open problem session also contributed to the fruitful communication between the different research communities. At the end of the conference, the participants decided to continue the efforts started at the workshop and to further communication by creating a website which serves as a platform for researchers working on time series with change points.

Mini-Workshop: Time Series with Sudden Structural Changes

Table of Contents

Abstracts

Detecting changes in the mean of functional observations

Horváth, Lajos
(joint work with István Berkes, Robertas Gabrys and Piotr Kokoszka)

A main tool of FDA is the Principal component analysis (PCA). It represents the functional observations $X_i(t), t \in T, i = 1, 2, \ldots, n$, in the form $X_i(t) = \mu(t) + \sum_{1 \le \ell < \infty} \eta_{i,\ell}\phi_\ell(t)$, where μ is the mean, ϕ_ℓ are the eigenfunctions of the covariance operator, and the $\eta_{i,\ell}$ are the scores. The set T can be interpreted as a time or a spatial domain, the methodology we develop requires merely that it be a compact subset of a Euclidean space. To perform the functional PCA, the functional mean $\mu(t)$, approximated by sample mean of the $X_i(t)$, is first be subtracted from the data. The first principal component $\phi_1(t)$ is then interpreted as the main pattern of deviation of the observations from the mean $\mu(t)$, or equivalently, as the direction in a function space of the largest variability away From the mean function. The subsequent eigenfunction define analogous directions orthogonal to the previous eigenfunctions. This interpretation and inferential procedures based on it assume that that the mean function $\mu(t)$ is the same for all values of i. If, in fact, the mean changes at some index(es) i, the results of PCA are confounded by the change(s). Issues of this type are most likely to emerge if the data are collected sequentially over time. Application we have in mind abound in climatology, environmental science and economics; detecting and locating changes in mean can be interpreted, for example, as climate shifts, a baseline change in a pollution level, or a shift in a long–term rate of growth.

It is thus useful to develop a methodology for the detection of changes in the mean of functional observations that is both easy to apply and justified by an clear large sample argument. We propose a significance test for testing the null hypothesis of a constant functional mean against the alternative of a changing mean. We also show how to locate the change points if the null hypothesis is rejected. Our methodology is readily implemented in the `R` package `fda`. The null distribution of the test statistic is asymptotically pivotal with a well-known asymptotic distribution going back to the work of Kiefer [5].

References

[1] D. Bosq, *Linear Processes in Function Spaces*, Springer, 2000, New York.
[2] R. Gabrys and P. Kokoszka, *Portmanteau test of independence for functional observations*, Journal of the American Statistical Association **102** (2007), 1338-1348.
[3] J. O. Ramsay and B. W. Silverman, *Applied Functional Data Analysis*, Springer, 2002, New York.
[4] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, 2005, New York.
[5] J. Kiefer, *K-sample analogues of the Kolmogorov-Smirnov and Cramér-v.Mises tests*, Annals of Mathematical Statistics **30** (1959), 420-447.

Clinical Monitoring

Roland Fried
(joint work with Ursula Gather, Karen Schettlinger, Michael Imhoff)

Modern technical devices allow to collect and store many variables describing the state of a patient with high sampling frequencies. In intensive care, hemodynamic variables like several blood pressures, heart rate, temperature and pulsoximetry are measured at least once a minute. These data provide essential information on the current state of the patient. The goal is to extract and analyze this information in real time to provide adequate bedside decision support. For this we need automatic procedures which reliably detect sudden level shifts and onsets of trends in the time series with only short delays, resisting the many and often large measurement artifacts.

Our approach uses local parametric modeling and robust filtering techniques. Based on comparisons of robust regression techniques, Davies, Fried, and Gather [2] and Gather, Schettlinger, and Fried [9] propose application of Siegel's [14] repeated median to a moving time window for robust approximation of a local linear trend, both in delayed and in full online analysis. Double window (Bernholt et al. [1]) and repeated median hybrid filters (Fried, Bernholt, and Gather [4]) improve the preservation of local extremes, trend changes and sudden shifts in delayed analysis. Gather and Fried [8] suggest a simple and robust rule for the data-adaptive choice of the window width, which is modified by Schettlinger, Fried, and Gather [13] for full online use without any time delay. Weighted repeated medians (Fried, Einbeck, and Gather [5]) achieve further improvements by weighting the observations according to their temporal distance to the target point at which we want to estimate the signal value.

Based on a reliable estimate of the signal trend, tests can be performed to detect sudden changes like level shifts with only short delays. Several classes of tests are investigated by Fried [3] under different assumptions on the error distribution. It turns out that comparison of local medians, obtained from different time windows and standardized by a highly robust and quite efficient scale estimator like Q_n (Rousseeuw and Croux [12]) result in good power for detection of level shifts within locally constant signals even in the presence of outliers. As opposed to this, modified rank tests (Fried and Gather [7]), which also use a local estimate of the variability, are to be preferred for shift detection within local linear trends if we can use data from an estimation period which is not very short. Both types of tests are applicable to data with time-varying volatility when using local estimates of the variability as discussed in Nunkesser et al. [11].

A further challenge is to adjust the methods for possible autocorrelations in the data. The basic filtering procedures perform well in the presence of autocorrelations (Fried and Gather [6]), but large positive autocorrelations result in monotone sequences which can easily be confused with trends or shifts within short time windows. A large data study provides evidence that it is sufficient to

use low order autoregressive models for incorporating autocorrelations into test procedures (Imhoff et al. [10]).

References

[1] T. Bernholt, R. Fried, U. Gather, I. Wegener, *Modified repeated median filters*, Statistics and Computing **16** (2006), 177–192.

[2] P.L. Davies, R. Fried, U. Gather, *Robust signal extraction for on-line monitoring data*, Journal of Statistical Planning and Inference **122** (2004), 65–78.

[3] R. Fried, *On robust shift detection in time series*, Computational Statistics & Data Analysis **52** (2007), 1063–1074.

[4] R. Fried, T. Bernholt, U. Gather, *Repeated median and hybrid filters*, Computational Statistics & Data Analysis **50** (2006), 2313–2338.

[5] R. Fried, J. Einbeck, U. Gather, *Weighted repeated median smoothing and filtering*, Journal of the American Statistical Association **480** (2007), 1300–1308.

[6] R. Fried, U. Gather, *Robust trend estimation for AR(1) disturbances*, Austrian Journal of Statistics **34** (2005), 139–151.

[7] R. Fried, U. Gather, *On rank tests for shift detection in time series*, Computational Statistics & Data Analysis **52** (2007), 221–233.

[8] U. Gather, R. Fried, *Methods and algorithms for robust filtering*, in: Proceedings in Computational Statistics COMPSTAT 2004, Antoch, J. (ed.), Physica-Verlag, Heidelberg (2004), 159–170.

[9] U. Gather, K. Schettlinger, R. Fried, *Online signal extraction by robust linear regression*, Computational Statistics **21** (2006), 33–51.

[10] M. Imhoff, M. Bauer, U. Gather, R. Fried, *Pattern detection in intensive care monitoring time series with autoregressive models: Influence of the model order*, Biometrical Journal **44** (2002), 746–761.

[11] R. Nunkesser, R. Fried, K. Schettlinger, U. Gather, *Robust Time Series Analysis by the Qn-estimator*, Computational Statistics & Data Analysis, to appear.

[12] P.J. Rousseeuw, Ch. Croux, *Alternatives to the median absolute deviation*, Journal of the American Statistical Association **88** (1993), 1273–1283.

[13] K. Schettlinger, R. Fried, U. Gather, *Real time signal processing by adaptive repeated median filters*, Preprint, Department of Statistics, TU Dortmund, Germany (2008).

[14] A.F. Siegel, *Robust regression using repeated medians*, Biometrika **69** (1982), 242–244.

CHARME Models for Nonparametric Nonlinear Regression and Time Series Analysis

Joseph Tadjuidje Kamgaing

(joint work with J. Franke, W.K. Li and J.-P. Stockis)

We consider conditional heteroskedastic mixture of experts (CHARME) models and present the asymptotic behavior of kernel type estimates of the regression functions. First, we consider a regression setting which may be represented as a mixture of K different nonparametric regression models.

We assume that $X_1, \ldots, X_N \in \mathbb{R}$ are i.i.d. random variables and set

$$Y_t = \sum_{k=1}^{K} S_{t,k}\{m_k(X_t) + \sigma_k \epsilon_t\} \text{ with } S_{tk} = \begin{cases} 1 & \text{for } Q_t = k, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $\epsilon_1, \ldots, \epsilon_N \in \mathbb{R}$ are i.i.d. real-valued random variables with mean 0 and variance 1, $m_1(x), \ldots, m_K(x)$, $\sigma_1, \ldots, \sigma_K$ are unknown regression functions and residual variances for the K regression models. Moreover, Q_t is assumed to be a stationary α-mixing hidden Markov chain with finite state space $\{1, \ldots, K\}$ and stationary distribution $\pi = (\pi_1, \ldots, \pi_K)$. Q_t is is duplicated via the K-dimensional vector process $S_t = (S_{t1}, \ldots, S_{tK})$, for which at each time instant one and only entry is different from 0.

Given this assumption, this formulation implies a classification problem where one has K different subsets of data and for each of them the regression functions and the variance of the residuals need to be estimated, i.e.,

$$Y_t = m_k(x_t) + \sigma_k \epsilon_t \quad t \in T_k = \{n \leq N;\ S_{nk} = 1\},\ k = 1, \ldots, K. \tag{2}$$

In the homogeneous situation, where there is only one state ($K = 1$), kernel estimates and, more generally, local polynomial estimates, have been successfully applied for estimating the regression functions, see e.g. Fan and Gijbels [1].

Here, we extend the idea introduced by Wong and Li [2] for parametric mixture models to the following Nadaraya-Watson type estimates for the regression functions $m_k, k = 1, \ldots, K$, allowing for different bandwidth $h_k, k = 1, \ldots, K$,

$$\hat{m}_k^0(x, h_k) \quad = \quad \frac{\sum_{t=1}^N K_{h_k}(x - x_t)\ Y_t S_{tk}}{\sum_{t=1}^N K_{h_k}(x - x_t) S_{tk}}. \tag{3}$$

As the S_{tk} are either 1 or 0, the vector of function estimates

$$(\hat{m}_1^0(x, h_1), \ldots, \hat{m}_K^0(x, h_K))^T$$

solves the weighted least-squares problem

$$\sum_{t=1}^N \sum_{k=1}^K (Y_t - \mu_k)^2 S_{tk} K_{h_k}(x - x_t) = \min_{\mu_1, \ldots, \mu_K \in \mathbb{R}} !$$

Allowing for different bandwidth is motivated by the fact the amount of the data available for a given subset T_k, $N_k = \sum_{t=1}^N S_{tk}$ is a random variable. However, N_k can be approximated asymptotically as a function of the overall sample data size N and the stationary distribution of the hidden process π, using a Bernstein inequality for α-mixing processes.

Given the above considerations we derive under mild conditions the consistency of kernel type estimates of the regression functions for all $k = 1, \ldots, K$

$$\hat{m}_1^0(x, h_k) \stackrel{i.p.}{\to} m_k(x).$$

For practical applications, we follow the idea in Wong and Li [2] and use a non-parametric version of their EM algorithm to numerically calculate kernel estimates of the regressions functions m_k, the residual variances σ_k^2 and the stationary distribution $\pi_k, k = 1, \ldots, K$.

The i.i.d. assumption fits well some applications, like exploring EEG data of patient suffering from particular heart diseases.

In the more general time series setting, we consider

$$(4)\quad X_t = \sum_{k=1}^{K} S_{t,k}(m_k(\mathbb{X}_{t-1}) + \sigma_k(\mathbb{X}_{t-1})\epsilon_t) \text{ with } S_{t,k} = \begin{cases} 1 & \text{for } Q_t = k, \\ 0 & \text{otherwise,} \end{cases}$$

where $\{\epsilon_t\}$ are i.i.d. real random variables with mean 0 and variance 1, independent of $\mathbb{X}_{t-1} = (X_{t-1}, \cdots, X_{t-p})$.We discuss the probabilistic behavior as well as statistical properties of such models where details can be found in Stockis , Tadjuidje and Franke [3], [4]. In particular, we investigate the geometric ergodicity of the model in (4) and the identifiability in the case that the nonparametric trend functions $m_k, k = 1, \ldots, K$ are approximated by outputs of suitable single layer feedforward neural networks, i.e. by functions of the form

$$(5)\qquad f_k(x, H) = \nu_{0k} + \sum_{h=1}^{H} \nu_{hk}\psi(< \alpha_{hk}, x > + b_{hk}),\ k = 1, \cdots, K,$$

where, H is a given positive integer and ψ is the logistic function. We propose to estimate m_k, σ_k^2 by fitting functions of the form $f_k(x, H)$ to them, and we describe a corresponding estimation algorithm as well as an Viterbi-type algorithm for estimating the change points, i.e. the time instants where the hidden Markov chain switches from one value to another.

Consistency of the parameter estimates is shown in Tadjuidje [5]. We illustrate the usefulness of this approach for asset management and risk analysis by an application to some financial data.

References

[1] J. Fan and I. Gijbels, *Local polynomial modelling and its application*, Chapman and Hall, London (1996).

[2] C.S. Wong and W.K. Li, *On a Mixture Autoregressive Conditional Heteroscedastic Model*, Journal of the American Statistical Association **96** (2001), 982–995.

[3] J.-P. Stockis, J. Tadjuidje-Kamgaing and J. Franke, *On Geometric Ergodicity of CHMARME Models*, Manuscript, University of Kaiserslautern (2007).

[4] J.-P. Stockis, J. Tadjuidje-Kamgaing and J. Franke, *A note on the identifiability of the conditional expectation for the mixtures of neural networks*, To appear in S& PL.

[5] J. Tadjuidje-Kamgaing, *Competing Neural Networks as Models for Non stationary Financial Time Series*, Ph.D. Dissertation, University of Kaiserslautern, http://kluedo.ub.uni-kl.de/volltexte/2005/1818/pdf/dfin.pdf (2005).

Some Theory Behind AutoPARM (Automatic Piecwise AR Modeling)

Richard A. Davis
(joint work with Stacey Hancock and Yi-Ching Yao)

Much of the recent interest in time series modeling has focused on data from financial markets, from communications channels, from speech recognition and from engineering applications, where the need for non-Gaussian, non-linear, and nonstationary models is clear. With faster computation and new estimation algorithms, it is now possible to make significant in-roads on modeling more complex-phenomena. In this talk, we describe some theory associated with a structural break detection procedure called *Automatic Piecewise AutoRegressive Modeling (AutoPARM)* developed by Davis, Lee, and Rodriguez-Yam [1]. The novelty of AutoPARM is to combine the use of genetic algorithms with the principle of minimum description length (MDL), an idea developed by Rissanen in the 1980s, to find "optimal" models over a potentially large class of models. If the true model consists of piecewise autoregressions, then we show that AutoPARM produces consistent estimation of the breakpoints and in some cases, consistent estimation of the number of breakpoints.

In order to describe the structural break modeling setup, let τ_j denote the break point between the j-th and $(j+1)$-th segments, where $j = 1, \ldots, m$ and m is the number of breakpoints. Setting $\tau_0 = 1$ and $\tau_{m+1} = n+1$, the j-th piece of the series is modeled as an AR process

$$Y_t = X_{t,j}, \quad \tau_{j-1} \le t < \tau_j, \tag{1}$$

where $\{X_{t,j}\}$ is the AR(p_j) process

$$X_{t,j} = \gamma_j + \phi_{j1} X_{t-1,j} + \ldots + \phi_{j,p_j} X_{t-p_j,j} + \sigma_j \varepsilon_t,$$

$\boldsymbol{\psi}_j := (\gamma_j, \phi_{j1}, \ldots, \phi_{j,p_j}, \sigma_j^2)$ is the parameter vector corresponding to this AR(p_j) process, and the noise sequence $\{\varepsilon_t\}$ is iid with mean 0 and variance 1. Given an observed series $\{y_i\}_{i=1}^n$, the objective is to obtain a "best" fitting model from this class of piecewise AR processes. This is equivalent to finding the "best" combination of the number of pieces $m+1$, the break point locations $\tau_1, \ldots, \tau_m$, and the orders $p_1, \ldots, p_{m+1}$ of the segmented autoregressions.

The above problem of finding a "best" combination of m, τ_j's and p_j's can be treated as a statistical model selection problem, in which candidate models may have different numbers of parameters. To solve this selection problem we apply the minimum description length (MDL) principle of Rissanen [5] to *define* a best fitting model (see Hansen and Yu [2] for a comprehensive review of MDL). The basic idea behind the MDL principle is that the best fitting model is defined as the one that produces maximum compression of the data. Successes in applying MDL to a variety of practical problems have been widely reported in the literature; see, for example, Lee [4] and Hansen and Yu [3].

For the piecewise AR model, the MDL objective function reduces to

$$\mathrm{MDL}(m, \tau_1, \ldots, \tau_m, p_1, \ldots, p_{m+1}) = \log m + (m+1)\log n + \sum_{j=1}^{m+1} \log p_j + \sum_{j=1}^{m+1} \frac{p_j + 2}{2} \log n_j + \sum_{j=1}^{m+1} n_j \log \hat{\sigma}_j^2 \,,$$

where n_j is the length of the j-th segment and $\hat{\sigma}_j^2$ is the estimated one-step prediction error over the j-th segment. To establish consistency, we assume that the true values τ_j^0 of the change point locations take the form $\tau_j^0 = [\lambda_j^0 n]$, where $\epsilon < \lambda_1^0 < \cdots < \lambda_m^0 < 1 - \epsilon$ for numbers $\lambda_j^0 \in (0,1)$, which have minimal separation that exceeds $\epsilon > 0$. It was shown in Davis, et al. [1], that if the number of change points m is known, then with probability one,

$$\hat{\lambda}_j \to \lambda_j^0,$$

where $\boldsymbol{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m)' = \mathrm{argmin}_{\boldsymbol{\lambda}} \min_{p_1,\ldots,p_{m+1}} \mathrm{MDL}(m, n\boldsymbol{\lambda}, p_1, \ldots, p_{m+1})$

The assumption that the true number of changepoints is known is certainly a limitation of this result. It remains to show that the estimate $\hat{m}$ of m found by also minimizing MDL with respect to m is consistent. It is easy to show that with probability one, $\hat{m} \geq m$ for all n large. That is, asymptotically, one cannot underestimate m. Partial results on this consistency problem can be established. For example, if there are no change points under the true model, i.e., $m = 0$, then $\hat{m} \to 0$ a.s. In this case, the main idea of the proof , which draws on an application of the law of the iterated logarithm, is to show that

$$\sum_{k=1}^{K+1} (\hat{\lambda}_k - \hat{\lambda}_{k-1}) \log \hat{\sigma}_k^2 - \log \hat{\sigma}^2 = O((\ln \ln n)/n)\,,$$

where $\hat{\sigma}^2$ is the estimate of σ^2 based on fitting one autoregression to the full data set. In some cases, the extension to the general case when $m > 0$ can be reduced to this instance since if we over estimate m, then there must be a homogeneous segment of the data which is subdivided into at least two pieces.

References

[1] Davis, R. A., Lee, T.C.M. and Rodriguez-Yam, G. A. (2006). "Structural Breaks Estimation for Non-stationary Time Series Models," *Journal of the American Statistical Association* 101, 223-239.

[2] Hansen, M. H., and Yu, B. (2000), "Wavelet Thresholding via MDL for Natural Images", *IEEE Transactions on Information Theory*, 46, 1778-1788.

[3] Hansen, M. H., and Yu, B. (2001), "Model Selection and the Principle of Minimum Description Length," *Journal of the American Statistical Association*, 96, 746-774.

[4] Lee, Thomas C. M. (2000), "A Minimum Description Length Based Image Segmentation Procedure, and its Comparison with a Cross-Validation Based Segmentation Procedure", *Journal of the American Statistical Association*, 95, 259-270.

[5] Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry*, Singapore: World Scientific.

Local change point analysis with applications to volatility modeling

Vladimir Spokoiny

Since the seminal papers of Engle [6] and Bollerslev [1], modeling the dynamic features of the variance of financial time series has become one of the most active fields of research in econometrics. The aim of the present paper is to develop a new estimation approach based on the local parametric modeling and applies the idea of pointwise adaptive choice of the interval of homogeneity. Similarly to Grama and Spokoiny [11] the choice of such an interval is done by a local multiscale change point analysis. This means that for every historical time point we test on a structural change at this point for the corresponding scale. The largest interval does not containing any change is used for estimation of the parameters of the return distribution. This approach has a number of important advantages of being easy to implement and very sensitive to the structural changes in the return process. We carefully address the question of selecting the tuning parameters of the procedure which is extremely important for practical applications. The proposed "propagation" approach suggests to tune the parameters under the simple time homogeneous situation to provide the prescribed performance of the procedure.

The change point detection problem for financial time series was considered in Mikosch and Starica [16] but they focused on asymptotical properties of the test if only one change point is present. Kitagawa [13] applied non-Gaussian random walk modeling with heavy tails as the prior for the piecewise constant mean for one-step-ahead prediction of nonstationary time series. However, the mentioned modeling approaches require some essential amount of prior information about the frequency of change points and their size. The new local change point (LCP) approach proposed in this article does not assume smooth or piecewise constant structure of the underlying process and does not require any prior information. That allows to proceed in a unified way with smoothly varying coefficient models and change point models. The approach is quite general and can be applied to many different problems. Grama and Spokoiny [11] studied the problem of Pareto tail estimation, Giacomini et al. [10] considered time varying copulae estimation, Čížek et al. [20] applied it to compare the performance of global and time varying ARCH and GARCH specifications. A comprehensive study of the general LCP procedure is to be given in the forthcoming monograph Spokoiny [19].

The theoretical study focuses on two important features of the proposed procedure: stability in the homogeneous situation and sensitivity to spontaneous changes of the model parameter(s). We particularly show that the procedure provides the optimal sensitivity to changes for the prescribed "false alarm" probability. Note that the classical asymptotic methods for stationary time series do not apply in the considered nonstationary situation with possibly small samples requiring to develop new approaches and tools. Our way of analysis is based on the so called "small modeling bias" condition which generalizes the famous bias-variance trade-off. The main result claims that the procedure delivers the estimation accuracy corresponding to the largest historical interval of homogeneity as if this interval were known.

References

[1] Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

[2] Brodskij, B. and Darkhovskij, B. (1993). *Nonparametric methods in change point problems*. Mathematics and its Applications. Kluwer Academic Publisher, Dordrecht.

[3] Dahlhaus, R. and Rao, S. S. (2003). Statistical inference of time-varying ARCH processes. DFG Schwerpunktprogramm 1214, Nr. 31.

[4] Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica*, 68:1343–1376.

[5] Eberlein, E. and Prause, K. (2002). The generalized hyperbolic model: financial derivatives and risk measures. In Geman, H., Madan, D., Pliska, S., and Vorst, T., editors, *Mathematical Finance-Bachelier Congress 2000*, pages 245–267, Berlin. Springer Verlag.

[6] Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008.

[7] Engle, R. F., editor (1995). *ARCH, selected readings*. Oxford University Press, Oxford.

[8] Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometrc Reviews*, 5:1–50, 81–87.

[9] Fan, J. and Gu, J. (2003). Semiparametric estimation of Value at Risk. *Econometrics Journal*, 6:260–289.

[10] Giacomini, E., Härdle, W., and Spokoiny, V. (2007). Inhomogeneous dependency modelling with time varying copulae. *Journal of Business and Economic Statistics*.

[11] Grama, I. and Spokoiny, V. (2008). Adaptive estimation of the excess of a distribution function. *Ann. Statist.*

[12] Härdle, W., Herwartz, H., and Spokoiny, V. (2003). Time inhomogeous multiple volatility modelling. *Journal of Financial Econometrics*, 1:55–95.

[13] Kitagawa, G. (1987). Non-Gaussian state-space modelling of nonstationary time series. *Journal of the American Statistical Association*, 82:1032–1063.

[14] McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroskedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300.

[15] Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics*, 32:577–602.

[16] Mikosch, T. and Starica, C. (2000a). Change of structure in financial time series, long range dependence and the garch model. University of Aarhus, Aarhus school of Business, Discussion Paper Nr. 58.

[17] Mikosch, T. and Starica, C. (2000b). Is it really long memory we see in financial returns? In Embrechts, P., editor, *Extremes and Integrated Risk Management*, Risk Books. UBS Warburg.

[18] Pollak, M. (1985). Optimal detection of a change in distribution. *Annals of Statistics*, 13:206–227.

[19] Spokoiny, V. (2008). *Local Parametric Estimation*. Springer Verlag, Berlin.

[20] Čížek, P., Härdle, W., and Spokoiny, V. (2007). Adaptive pointwise estimation in time-inhomogeneous time-series models. *unpublished manuscript*.

Statistical inference for locally stationary processes

Rainer Dahlhaus

We consider inference for locally stationary processes (cf. Dahlhaus [1, 2]), that is for processes $X_{t,n}$ $(t = 1, \dots, n)$ which have a slowly-varying moving average representation

$$X_{t,n} = \sum_{j=-\infty}^{\infty} a_{t,n}(j)\, \varepsilon_{t-j} \tag{1}$$

where the $a_{t,n}(j)$ can be approximated by $a(\frac{t}{n}, j)$ with a function $a(u, j)$ of bounded variation in u. The ε_t are assumed to be independent and identically distributed with $E\varepsilon_t \equiv 0$, $E\varepsilon_t^2 \equiv 1$ and $\kappa_4 := \mathrm{cum}_4(\varepsilon_t)$. Details on the assumptions can be found in Dahlhaus [3]. The function

$$f(u, \lambda) = \frac{1}{2\pi} |A(u, \lambda)|^2$$

with $A(u, \lambda) = \sum_{j=-\infty}^{\infty} a(u, j) \exp(-i\lambda j)$ is the time varying spectral density.

The goal now is to make statistical inference about the process - for example to estimate the coefficient functions of a time varying AR-process

$$X_{t,n} \, + \, a(\frac{t}{n})\, X_{t-1,n} = \sigma(\frac{t}{n})\, \epsilon_t$$

which can be represented in the form (1).

For such problems the so-called empirical spectral measure plays a major role. It is an estimate of

$$F(\phi) = \int_0^1 \int_{-\pi}^{\pi} \phi(u, \lambda)\, f(u, \lambda)\, d\lambda\, du$$

defined by

$$F_n(\phi) = \frac{1}{n} \sum_{t=1}^{n} \int_{-\pi}^{\pi} \phi(\frac{t}{n}, \lambda)\, J_n(\frac{t}{n}, \lambda)\, d\lambda$$

with the pre-periodogram

$$J_n\Big(\frac{t}{n}, \lambda\Big) = \frac{1}{2\pi} \sum_{k:1 \le [t+1/2 \pm k/2] \le n} X_{[t+1/2+k/2],n} X_{[t+1/2-k/2],n} \exp(-i\lambda k).$$

Many statistics occurring in the analysis of locally stationary time series are of the form $F_n(\phi)$:

1.	$\phi(u, \lambda) = K_n(u_0 - u)\, \cos(\lambda k)$	local covariance estimator
2.	$\phi(u, \lambda) = K_n(u_0 - u)\, K_n(\lambda_0 - \lambda)$	spectral density estimator
3.	$\phi(u, \lambda) = K_n(u_0 - u)\, \nabla f_\theta(u, \lambda)^{-1}$	local Whittle estimator
4.	$\phi(u, \lambda) \approx K_n(u_0 - u)\, \nabla f_\theta(u, \lambda)^{-1}$	local least squares
5.	$\phi(u, \lambda) = \nabla f_\theta(u, \lambda)^{-1}$	parametric Whittle estimator

6. $\phi(u,\lambda) = \big(I_{[0,u_0]}(u) - u_0\big) I_{[0,\lambda_0]}(\lambda)$ testing stationarity
7. $\phi(u,\lambda) = \cos(\lambda k)$ stationary covariance
8. $\phi(u,\lambda) = \nabla f_\theta(\lambda)^{-1}$ stationary Whittle estimator

Examples 5,6,7,8 can be treated with the results in Dahlhaus and Polonik [5] where a central limit theorem and an exponential inequality have been proved. A more complex example is nonparametric maximum likelihood estimation under shape restrictions (Dahlhaus and Polonik [4]). The asymptotic properties of the estimates in examples 1,2,3,4 follow from the following central limit theorem. Suppose $\phi_{n1}, \ldots, \phi_{nd}$ are index functions. Let

$$\rho_2(\phi) := \Big(\int_0^1 \int_{-\pi}^{\pi} \phi(u,\lambda)^2 \, d\lambda \, du \Big)^{1/2}$$

and

$$c_E(\phi_j, \phi_k) := 2\pi \int_0^1 \int_{-\pi}^{\pi} \phi_j(u,\lambda) \left[\phi_k(u,\lambda) + \phi_k(u,-\lambda) \right] f^2(u,\lambda) \, d\lambda \, du$$
$$+ \kappa_4 \int_0^1 h_n^4(u) \Big(\int_{-\pi}^{\pi} \phi_j(u,\lambda_1) f(u,\lambda_1) d\lambda_1 \Big) \Big(\int_{-\pi}^{\pi} \phi_k(u,\lambda_2) f(u,\lambda_2) d\lambda_2 \Big) du.$$

Theorem *If the limit*

$$\lim_{n\to\infty} \frac{c_E(\phi_{nj}, \phi_{nk})}{\rho_2(\phi_{nj}) \, \rho_2(\phi_{nk})} = \Sigma_{j,k}$$

exists then

$$\left(\frac{\sqrt{n}}{\rho_2(\phi_{nj})} \Big(F_n(\phi_{nj}) - F(\phi_{nj}) \Big) \right)_{j=1,\ldots,d} \to \mathcal{N}(0, \Sigma).$$

The proof and details about the assumptions can be found in Dahlhaus [3].

References

[1] R. Dahlhaus, *Fitting time series models to nonstationary processes*, Ann. Statist. **25** 1-37 (1997).

[2] R. Dahlhaus, *A likelihood approximation for locally stationary processes*, Ann. Statist. **28** 1762-1794 (2000).

[3] R. Dahlhaus, *Local inference for locally stationary time series based on the empirical spectral measure*, preprint (2008).

[4] R. Dahlhaus and W. Polonik, *Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes*, Ann. Statist. **34** 2790-2824 (2006).

[5] R. Dahlhaus and W. Polonik, *Empirical spectral processes for locally stationary time series*, preprint (2007).

Sequential procedures for detection of changes

Marie Hušková
(joint work with A. Aue, I. Berkes, L. Horváth, P. Kokoszka, A. Koubková, J. Steinebach)

The talk concerns procedures for detection of a change in linear models and times series models when data arrives sequentially. Moreover, it is assumed that training (historical) data with no change are available. Such problems occur in a number of applications, e.g., in economics and finance, geophysical sciences, statistical quality control, medical care.

Several classes of sequential test procedures for detection of a change are developed and their asymptotic properties are studied. Particularly, the test procedure is determined by a sequence of statistics say $\{Q(m+k)\}_k$, boundary function and tuning parameter. Here $Q(m+k)$ is based on the first $m+k$ observations and m is the number of historical data with no change. The boundary function is chosen in a way that the change is detected early. The tuning constant is chosen in such that the probability of false alarm is approximately is smaller than prechosen number (in terms of hypothesis testing it means level of the test procedure) and that if the change occurs it should be detected with probability one. To get a suitable approximation for the tuning constant one has to study asymptotic distribution of related stopping rule. It leads to a study of extremes of certain Gaussian processes.

Theoretical (asymptotic) properties of the test procedures are studied. Various modifications are discussed Also theoretical results are checked on a simulation study.

References

[1] Aue, A. and L. Horváth, (2004) *Delay time in sequential detection of change*, Statistics & Probability Letters, Vol.**67**, 221 – 231.
[2] Aue, A. and L. Horváth, M. Hušková and P. Kokoszka, (2006) *Change-point monitoring in linear models*, Econometrics Journal, Vol. **9**, 1-31.
[3] Chu, C.-S.J., M. Stinchcombe and H. White, (1996) *Monitoring Structural Change*, Econometrica, Vol. **64**, No. **5**, 1045–1065.
[4] Horváth, L., M.Hušková , P. Kokoszka and J. Steinebach, (2004) *Monitoring Changes in Linear Models*, Journal of statistical Planning and Inference, **126**, 225-251.
[5] Hušková, M. and Koubková, A. (2005): Monitoring jump changes in linear regression.Journal of Statistical Research, 59–78.
[6] Hušková, M and Koubková (2006): Sequential procedures for detection of changes in autoregressive sequences. In Prague Stochastics, eds. Hušková, M. and Lachout P., 437–447.
[7] Koubková, A. (2006) Sequential detection of changes, PhD thesis, Charles University, Prague.
[8] Leisch, F., K. Hornik and Ch.-M. Kuan, (2000) *Monitoring Structural Changes with the Generalized Fluctuation Test*, Econometric Theory, **16**, 835-854.

Resampling methods in change-point analysis

Claudia Kirch

In change-point analysis critical values for testing procedures are usually obtained by distributional asymptotics. However, convergence is often rather slow. Using resampling methods one can often obtain better approximations for small sample sizes or non-normal data.
Recently permutation methods have been applied to a variety of change-point models. These models include location models (diverse sum- and maximum-type statistics as well as certain moving sum procedures which are specifically suitable for the detection of multiple changes), general U-statistics as well linear regression and auto-regression models. The relevant results and further references can be found in the recent survey paper by Hušková [1].

All of these results deal with independent errors. Some first results for dependent data have been derived by Kirch and Steinebach [5] for error sequences fulfilling strong invariance principles. The statistic they are using is based on increments of growing intervals which makes blocking arguments unnecessary since they are in a sense already included by the model.

Kirch [3] proves the validity of block resampling methods in a classical location model where the errors are no longer independent but rather form a linear process. In this situation resampling methods have some additional advantages in comparison with the asymptotic methods. First of all the small sample behavior of the statistics does depend rather strongly on the underlying dependency structure, whereas the asymptotic distribution is independent of it. Secondly a good nonparametric estimator for the asymptotic variance correction term is difficult to get but not needed for the proposed resampling methods.
In a similar fashion Hušková and Kirch [2] use block resampling techniques to obtain asymptotic confidence intervals for the change-point, in a forthcoming work the results will be improved using studentizing.

Recently sequential change-point analysis has become more and more popular. In this setup, one gets the observations 'online', i.e. sequentially one-by-one, after having observed a historic data set without change. For each new observation one checks whether one can still assume the null hypothesis. This is becoming more and more important in such diverse fields as medicine, material science or finance. In such a setting it is not clear how one should best do the bootstrap. The simplest method certainly is to just use the historic sample for the bootstrap. This is easy and computationally fast but since one draws a much larger (possibly in theory infinite) sample from a relatively small set of observations, this is not optimal. So the second choice essentially is to use in each step a different critical values obtained from a bootstrap based on all observations available up to this point. In fact such a procedure yields much better results as some simulations show, but it is computationally very expensive. This is why we develop a third procedure which exploits the idea that we already have generated bootstrap samples and that only the old ones are not good enough anymore whereas the newer ones are still reasonably good. In a simulation study we see that this procedure is in fact

much faster than the second one and yields almost identical results. The theoretic details as well as a simulation study can be found in Kirch [4].

References

[1] M. Hušková, *Permutation principle and bootstrap in change point analysis*, Fields Inst. Commun., **44** (2004), 273–291.

[2] M. Hušková and C. Kirch, *Bootstrapping confidence intervals for the change-point of time series* , J. Time Ser. Anal., (2008+), submitted.

[3] C. Kirch, *Block permutation principles for the change analysis of dependent data*, J. Statist. Plann. Inference, **137** (2007), 2453–2474.

[4] C. Kirch, *Bootstrapping sequential change-point tests*, Sequential Anal., (2008+), submitted.

[5] C. Kirch, und J. Steinebach, *Permutation principles for the change analysis of stochastic processes under strong invariance*, J. Comput. Appl. Math., **186** (2006), 64–88.

Multiple change point detection by binary segmentation for time series

Rainer von Sachs

(joint work with Piotr Fryzlewicz)

The method of *binary segmentation* has been introduced by Sen and Srivastava [6], with theory provided by Vostrikova [8], for testing on the change of the mean of a series of, say N, i.i.d. normal random variables Y_t . In a nutshell, let $S_k = \sum_{t=1}^{k} Y_t$, and define a (CUSUM-like) statistic $Z_t = (t\ (1 - t/N))^{-1/2} \cdot (t\ S_N/N - S_t)$ as well as an appropriate threshold b_N (critical value under the null of no change). For any k in a given segment $[k_1 + 1, k_2]$ of $[0, N]$ calculate $Z_k = Z^k_{k_1,k_2}$. If $\max_{k_1<k<k_2} |Z_k| > b_N$ decide for a change point at any position k such that $|Z_k| = \max_{k_1<t<k_2} |Z_t|$.

While Venkatraman [7] shows that, with $b_N = N^{3/8}$, under a model of an asymptoticaly growing number of change points, both the number and the location of the change points can be estimated consistently, we are concerned with generalisations of his set-up, using a tree-structured variant of proceeding down on shorter and shorter segments of $[0, N]$ as long as a change point is found.

The aforementioned method for detection of a "signal change" embedded in (additive and independent) noise is akin to estimating a piecewise constant signal (with an asymptotically growing number of constant pieces). In this project we want to apply this methodology to *multiplicative* models, where we face the problem of heteroscedasticity, and in particular to time series models of *correlated* data, such as: autoregression with a time-change in the autoregressive parameters (see below), time-varying ARCH-processes (Fryzlewicz and Subba Rao [4]) and models in the frequency domain (time-varying spectral representations). Hereby the change is assumed to occur in the variance-covariance structure of the data, and not in its mean. Hence we will need to adapt the method to detection of changes in the mean of *squared* data.

The "new" key ingredients of this project are the use of a fast and non-dyadic tree algorithm based on the "Unbalanced Haar" (UH) transform (see, e.g., Delouille, Franke and von Sachs [1]), with control of the distributional properties in the coefficient domain of the transform. Hereby we borrow strength from existing results on denoising by non-linear wavelet thresholding in the context of nonparametric function estimation: Fryzlewicz [2] shows that by the use of UH, a generalised Haar transform allowing to perfectly encode the position of jumps in a piecewise constant function at arbitrary non-dyadic grid points of the unit interval of the regression design, one can denoise the underlying signal as efficiently as with classical wavelet thresholding. Moreover, interpreting as a hypothesis test the comparison of each wavelet (UH) coefficient with a threshold (proportional to $\sqrt{2\log(N)}$) based on asymptotic normality, one can also extract from the method a sequence of multiple hypothesis tests on the significance of the coefficient for the presence of a jump (change point). It is in this sense that UH can be directly interpreted as hierarchical (*multiscale*) binary segmentation.

In preliminary experiments we found out that the thresholds used in the UH approach work better than the aforementioned thresholds b_N as the latter ones turn out to be too conservative in practice (in particular if smaller though significant jumps are "masked" by some bigger ones).

A concrete idea for an algorithm taylored to the problem of detecting changes in a piecewise stationary (zero mean) autoregressive (AR) process is as follows. (1) "Estimate" its parameters as if the AR process were stationary (select them by a grid search such that they "best expose" change points in the *residuals*): we get a sequence of residuals which are stationary if and only if the AR process is stationary. (2) *Square* the residuals, take their mean: AR-parameter changes correspond to change points in their mean. (3) Apply binary segmentation and use (less conservative) thresholds taylored to approximately Gaussian data after studentizing and "gaussianizing" the data by the UH followed by the "Unbalanced Haar Fisz" (UHF) transform (successfully applied also in Fryzlewicz, Nason and von Sachs [3]): take the UH transform of the data, divide each coefficients by an estimate of its own standard deviation (the "Fisz" transform, in our case just the local mean of the data incorporated in the UH "detail" coefficient) and take the inverse UH transform. (Note that in our multiplicative model set-up, the local standard deviation is proportional to the local mean.) This "local studentizing" (as it operates on genuinely local transforms) renders the data more homoscedastic and pulls their distribution close to Gaussian.

We believe that our new approach has the following advantages over classical binary segmentation (or CUSUM-based approaches such as MOSUM): the method is truly multiscale as we grow down the whole UH tree before thresholding, hence the risk to ignore small jumps in the presence of dominating ones is reduced; the method is fast (as based on existing fast code coming from signal estimation and denoising); an overall "Gaussian" threshold proportional to $\sqrt{2\log(N)}$ applied to

(automatically correctly scaled) coefficients replaces the critical values under the null hypothesis of no change point on each scale.

Finally our methodology should give an interesting alternative to existing dyadic time series *segmentation* procedures (such as the SLEX approach of Ombao, Raz and von Sachs [5]).

References

[1] V. Delouille, J. Franke and R. von Sachs. *Nonparametric stochastic regression with design-adapted wavelets*. Sankhya **63** (2001), Series A, 328–366.

[2] P. Fryzlewicz. *Unbalanced Haar technique for nonparametric function estimation*. J. Amer. Stat. Assoc. **102** (2007). 1318–1327.

[3] P. Fryzlewicz, G. Nason and R. von Sachs. *A wavelet-Fisz approach to spectrum estimation*. J. Time Series Analysis (2008), to appear.

[4] P. Fryzlewicz and S. Subba Rao. *BaSTA: consistent multiscale change-point detection for piecewise-stationary ARCH models*. Mimeo under development, 2008.

[5] H. Ombao, J. Raz and R. von Sachs. *Automatic Statistical Analysis of Bivariate Non-Stationary Time Series*. J. Amer. Stat. Assoc. **96** (2001), 543–560.

[6] A. Sen and M.S. Srivastava. *Some one-sided tests for changes in level*. Technometrics **17** (1975), 61–65.

[7] E.S. Venkatraman. *Consistency results in multiple change-point problems*. PhD thesis, Stanford University, 1993.

[8] L.J. Vostrikova. *Detecting "disorder" in multidimensional random processes*. Soviet Math. Dokl. **24** (1981), 55–59.

Exploiting the duality between runs and change points

John A. D. Aston

(joint work with Donald E. K. Martin and Michael Jyh-Ying Peng)

A method for calculating exact change point and feature distributions for changing regimes in Markov switching models is proposed. Markov switching models are a rich class of models which generalise Hidden Markov Models (HMMs) by incorporating greater dependence structure both within the data and between the data and the underlying hidden switching states (Cappe [4], Frühwirth-Schnatter [7]). The change point and other related distributions given here are generated from an algorithm to calculate the exact marginal distributions of change point locations in finite sample data, in contrast to Markov Chain Monte Carlo methods of sampling the underlying states to characterise the distributions as in Chib [5]. The proposed algorithm is derived from the fact that change points and runs in the underlying state sequence, such as runs of the type defined in Balakrishnan and Koutras [2], are intrinsically linked. Through the ideas of finite Markov chain imbedding (Fu and Koutras [8]) applied to HMMs (Aston and Martin [1]) and Markov switching models, it is possible to efficiently determine waiting time distributions of runs in the underlying state sequence. Parameter estimation effects can also be incorporated using suitable priors on the parameters in ways such as those of Fearnhead and Liu [6]. Distributions for many regime features can then be calculated from these waiting times, including the start and end of a regime period

(the change points), the number of change points, and the length of the longest regime period between change points. These distributions help quantify many of the features and implied change points routinely mentioned by practitioners when discussing Markov switching models. For example, Markov switching models have been used to assess recessions (Hamilton [9]) and changes in financial volatility (Cai [3], Hamilton and Susmel [10]), and the proposed distributions allow for the quantification of uncertainty in the estimates of the regimes given by the models.

References

[1] Aston, J. A. D. and D. E. K. Martin (2007). Distributions associated with general runs and patterns in hidden Markov models. *Annals of Applied Statistics 1*, 585–611.

[2] Balakrishnan, N. and M. Koutras (2002). *Runs and Scans with Applications*. New York: Wiley.

[3] Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business and Economic Statistics 12*, 309–316.

[4] Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer.

[5] Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*, 221–241.

[6] Fearnhead, P. and Z. Liu (2007). Online inference for multiple changepoint problems. *Journal of the Royal Statistical Society, Series B 69*, 589–605.

[7] Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

[8] Fu, J. C. and M. V. Koutras (1994). Distribution theory of runs: A Markov chain approach. *Journal of the American Statistical Association 89*, 1050–1058.

[9] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

[10] Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics 64*, 307–333.

Autocovariance Structure of Markov-Switching ARMA and GARCH Processes

Christian Francq

(joint work with Jean-Michel Zakoian)

The Markov-Switching Models (MSM) introduced by Hamilton [16] have become increasingly popular in econometrics and time series (see Cai [2], Dueker [4], Gray [14], Hamilton and Lin [17], Hamilton and Susmel [18], Kim [20], Klaassen [21] for relevant applications of MSM in finance and econometrics and Lange and Rahbek [23] for a recent review). A MSM is a model in which the coefficient varies according to the state of a non observed Markov chain. In the econometric literature, each state of the Markov chain, usually referred to as the regime, may correspond to a state of the economy. The dynamics of each regime can be specified by, for instance, standard ARMA or GARCH models. The changes in regime can be smooth or abrupt, and they can occur frequently or occasionally depending on the transition probabilities of the chain, which makes the model very flexible. An interesting feature of MSM's is that, at the same time, they can be stationary and also exhibit

sudden changes in the dynamics. Moreover, the process can be globally stationarity even when some regimes are explosive (see *e.g.* Yang [26], Francq and Zakoian [10]), but the stationarity within each regime is neither sufficient nor necessary to obtain global stationarity (see Francq and Zakoian [9]). For the MS-GARCH recently introduced in Haas, Mittnik and Paolella [15], the stationarity conditions are studied by Liu [24]. It has been shown that the L^2-structures of powers of several classes of MSM are those of ARMA models, with non-independent linear innovations (see Francq and Zakoian [11, 12]). Concerning the estimation of the parameters, the direct application of maximum-likelihood method is only possible when the regimes are markovian, due to a problem of path dependence (see *e.g.* Klaassen [21]). Condition ensuring the consistency of the maximum-likelihood estimator can be found in Francq and Zakoian [6, 7, 8], and conditions for asymptotic normality are given in Douc, Moulines, and Ryden [3]. When the regimes are not markovian, simulated methods, such as MCMC, must be employed to approximate the likelihood (see Bauwens, Preminger, Rombouts [1] and the references therein). In Francq and Zakoian [13], an alternative method is proposed for estimating MS-GARCH, using the above-mentioned ARMA representations of different powers within the GMM approach of Hansen [19]. Interesting subclasses of MS-GARCH models, namely the standard GARCH, the independent-switching GARCH studied in Fong, Li, and An [5] and Wong and Li [25], the Hidden-Markov Models and the Markov-switching ARCH, can also be discriminated by considering the ARMA autocovariance structures of several powers.

References

[1] Bauwens, L., Preminger, A. and Rombouts, J.V.K., *Theory and Inference for a Markov Switching GARCH Model*, CORE Discussion Paper No. 2007/55.

[2] Cai, J., *A Markov model of switching-regime ARCH*, Journal of Business & Economic Statistics **12** (1994) 309–316.

[3] Douc R., Moulines, E. and Ryden, T., *Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime*, Annals of Statistics **32** (2004) 2254–2304.

[4] Dueker, M. J., *Markov switching in GARCH processes and mean-reverting stock-market volatility*, Journal of Business & Economic Statistics **15** (1997) 26–34.

[5] Fong, P.W., Li, W.K. and An, H.Z., *A simple multivariate ARCH model specified by random coefficients*, Computational Statistics & Data Analysis **51** (2006) 1779–1802.

[6] Francq, C. and Roussignol, M., *On white noises driven by hidden Markov chains*, Journal of Time Series Analysis **18** (1997) 553–578.

[7] Francq, C. and Roussignol, M., *Ergodicity of autoregressive models with Markov-switching and consistency of the maximum-likelihood estimator*, Statistics **32** (1998) 151–173.

[8] Francq, C., Roussignol, M. and Zakoian, J-M., *Conditional heteroskedasticity driven by hidden Markov chains*, Journal of Time Series Analysis **22** (2001) 197–220.

[9] Francq, C. and Zakoian, J-M., *Stationarity of multivariate Markov-switching ARMA models*, Journal of Econometrics **102** (2001) 339–364.

[10] Francq, C. and Zakoian, J-M., *Comments on the paper by Minxian Yang "Some properties of vector Autoregressive processes with Markov-Switching coefficients*, Econometric Theory **18** (2002) 815–818.

[11] Francq, C. and Zakoian, J-M., *Autocovariance Structure of Powers of Switching-Regime ARMA Processes*, ESAIM P&S **6** (2002) 259–270.

[12] Francq, C. and Zakoian, J-M., *The L^2-structures of standard and switching-regime GARCH models*, Stochastic Processes and their Applications **115** (2005) 1557–1582.
[13] Francq, C. and Zakoian, J-M., *Deriving the autocovariances of powers of Markov-switching GARCH models, with applications to statistical inference*, Computational Statistics & Data Analysis **52** (2008) 3027–3046.
[14] Gray, S.F., *Modeling the conditional distribution of interest rates as a regime-switching process*, Journal of Financial Economics **42** (1996) 27–62.
[15] Haas, M., Mittnik, S. and Paolella, M., *A new approach to markov-switching GARCH models*, Journal of Financial Econometrics **2** (2004) 493–530.
[16] Hamilton, J.D., *A new approach to the economic analysis of nonstationary time series and the business cycle*, Econometrica **57** (1989) 357–384.
[17] Hamilton, J.D. and Lin, G., *Stock market volatility and the business cycle*, Journal of Applied Econometrics **11** (1996) 573–593.
[18] Hamilton, J.D. and Susmel, R., *Autoregressive conditional heteroskedasticity and changes in regime*, Journal of Econometrics **64** (1994) 307–333.
[19] Hansen, L.P., *Large Sample Properties of Generalized Method of Moments Estimators*, Econometrica **50** (1982) 1029–1054.
[20] Kim, C.J., *Dynamic linear models with Markov-switching*, Journal of Econometrics **60** (1994) 1–22.
[21] Klaassen, F., *Improving GARCH volatility forecasts with Regime-Switching GARCH*, Empirical Economics **27** (2002) 363–394.
[22] Krishnamurthy, V. and Ryden, T., *Consistent estimation of linear and non-linear autoregressive models with markov regime*, Journal of Time Series Analysis **19** (1998) 291–307.
[23] Lange, T. and and Rahbek, A. *An Introduction to Regime Switching Time Series Models*, To appear in *Handbook of Financial Time Series*, Edts: T. G. Andersen, R.A. Davis, J-P. Kreiss, T. Mikosch. Springer Statistics.
[24] Liu, J.C., *Stationarity of a Markov-Switching GARCH Model*, Journal of Financial Econometrics **4** (2006) 573–593.
[25] Wong, C.S. and Li, W.K., *On a Mixture Autoregressive Conditional Heteroskedastic Model*, Journal of the American Statistical Association **96** (2001) 982–995.
[26] Yang, M., *Some Properties of Vector Autoregressive Processes with Markov-Switching Coefficients*, Econometric Theory **16** (2000) 23–43.

On a dynamic mixture GARCH model

W.K. Li
(joint work with Xixin Cheng, Philip L.H. Yu)

After its inception in Engle [1], the GARCH model has been commonly used to model the conditional variance of many financial data. One problem is raised, however that, the GARCH model seemed incapable to capture a sudden and dramatic increase in the volatility of data in a very short period. Gray [2] in his paper, tried to use GARCH to model some interest rate data, but found that a regime-switching process fitted interest rate data much better than a single GARCH process. Chen, So and Gerlach [3] also found that regime-switching GARCH is more suitable than a single GARCH or ARCH process in modeling the huge volatility change driven by news arrival in financial markets.

Further, Mikosch and Starica [4] pointed out that fitting a normal GARCH model to data with structural change would exhibit the integrated GARCH

(IGARCH) effect, where the sum of the ARCH coefficient and the GARCH coefficient is close to 1. This IGARCH effect is widely showed in stock market data and other financial market data. In order to explain this IGARCH effect and capture the characteristic of dramatic volatility change in a short time, two major approaches have been proposed, the threshold model and the mixture model.

In the direction of mixture model, Schwert [5] proposes a model in which observations have either a higher volatility or a lower volatility, and this depends on a hidden variable Z_t which follows a two stage Markov process. Wong and Li [6] propose a similar model but the hidden volatility state variable Z_t is linked to other variables (may be past observations) with a logistic linking function. These approaches, however, can be regarded as an extension of Tong and Lim [7]'s threshold autoregressive model. Indeed, threshold time series with the threshold indicator variable satisfying a Markovian process was first pointed out by Tong and Lim [7] in their rejoinder to the discussion.

As an attempt to extend the mixture model to cater for conditional heteroscedasticity case, Wong and Li [8] proposes a mixture ARCH model with a constant mixture probability. An early attempt to model changing structure in a time series is the Markov-switching model. Goldfeld and Quandt [9] began the Markov-switching model. Hamilton and Susmel [10], and Cai [11] are the first few authors who propose the Markov-switching ARCH model.

The mixture GARCH model with a classical GARCH specification is a path-dependent model. Gray [2] proposes a traceable mixture GARCH model, in which the $\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta\sigma_{t-1}^2$ in the GARCH model is replaced by $\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta E(\sigma_{t-1}^2 \mid \mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1}$ is the information set up to time $t-1$. However, this modification made the mixture model incapable of capturing the dramatic changes in volatility in a short period. This is because the expectation of volatility in the two regimes mingles the volatilities, so that the 'mingled volatility' can not describe the two distinct volatility levels in the mixture data. Haas, Mittnik and Paolella [12] propose another Markov-switching mixture GARCH model, where a hidden state variable Z_t determines which volatility regime is exhibited in the observation.

In this paper, we propose a dynamic mixture GARCH model (DMGARCH), in which the mixture proportion is linked with other variables by means of a logistic link function. Generally, the link function can be any inverse cumulative distribution function. Because of its simplicity, we choose the logistic function as the link function in our model. Section 2 gives a detailed model specification.

Section 3 gives the procedure of parameter estimation with the EM algorithm, and section 4 discusses the BIC approach in the DMGARCH case.

In section 5, we apply this model to the S&P500 index and the Hang Seng Index on their daily log-return data. An interesting result from S&P500 and Hang Seng Index estimation is that the IGARCH effect in the normal GARCH model seems to be the result of a mixture of a stationary volatility component and a non-stationary volatility component. We also find that the Value-at-Risk(VaR)

based on the DMGARCH model performs better than the traditional GARCH based VaR.

References

[1] R.F. Engle, *Autoregressive conditional heteroscedasticity with estimation of the variance of United Kingkom inflation*, Econometrica **50** (1982), 987–1007.

[2] S.F. Gray, *Modeling the conditional distribution of interest rates as a regime-switching process*, Journal of Financial Economics **42** (1996), 27–62.

[3] C.W.S. Chen, M.K.P. So and R.H. Gerlach, *Asymmetric response and interaction of U.S. and local news in financial markets*, Appl. Stochastic Models Bus. Ind. **21**, (2005), 273–288.

[4] T. Mikosch and C. Starica, *Nonstationarities in financial time series, the long-range dependence and the IGARCH effects*, Review of Economics and Statistics **86** (2004), 378–390.

[5] G. Schwert, *Why does stock market volatility change over time?* Journal of Finance **44** (1989), 1115-1153.

[6] C.S. Wong and W.K. Li, *On a logistic mixture autoregressive model*, Biometrika **88(3)** (2001a), 833–846.

[7] H. Tong and K.S. Lim, *Threshold autoregression, limit cycles and cyclical data*, J. R. Statist. Soc. B **42** (1980), 245–292.

[8] C.S. Wong and W.K. Li, *On a mixture autoregressive conditional heteroscedastic model*, Journal of the American Statistical Associaton **96(455)** (2001b), 982–995.

[9] S.M. Goldfeld and R.E. Quandt, *A Markov model for switching regression*, Journal of Econometrics **1** (1973), 3–15.

[10] J. Hamilton and R. Susmel, *Autoregressive conditional heteroskedasticity and changes in regime*, Journal of Econometrics **64** (1994), 307–333.

[11] J. Cai, *Markov model of unconditional variance in ARCH*, Journal of Business and Economics Statistics **12** (1994), 309–316.

[12] M. Haas, M., S. Mittnik and M.S. Paolella, *A new approach to Markov-Switching GARCH models*, Journal of Financial Econometrics **2(4)** (2004), 493–530.

Unit Roots and Level Shifts

Alexander Aue

(joint work with Lajos Horváth, Marie Hušková and Shiqing Ling)

We study test procedures that detect structural breaks in underlying data sequences. In particular, we wish to discriminate between different reasons for these changes, such as (1) shifting means, (2) random walk behavior, and (3) constant means but innovations switching from stationary to difference stationary behavior. Almost all procedures presently available in the literature are simultaneously sensitive to all three types of alternatives (see, for example, Andreau and Spanos [1]; Busetti and Taylor [3]; and Nyblom and Mäkeläinen [4]).

The test statistics under investigation are based on functionals of the partial sums of observations. These CUSUM–type statistics have limit distributions if the mean remains constant and the errors satisfy the central limit theorem, but tend to infinity in the caseany of the alternatives (1), (2) or (3) holds. On removing the effect of the shifting mean, however, divergence of the test statistics will only occur under the random walk behavior, which in turn enables statisticians to not only detect structural breaks but also to specify their causes.

The results are underlined by a simulation study and an application to returns of the German stock index DAX.

References

[1] Andreau, E., and A. Spanos. Statistical adequacy and the testing of trend versus difference stationarity. *Econometric Reviews* 22 (2003) , 217–237.
[2] Aue, A., Horváth, L., Hušková, M., and S. Ling. On Distinguishing between Random Walk and Change in the Mean Alternatives. *Econometric Theory*, to appear.
[3] Busetti, F., and Taylor, A.M.R. (2004). Tests of stationarity against a change in persistence. *J*ournal of Econometrics 123 (2004), 33–66.
[4] Nyblom, J., and T. Mäkeläinen. Comparisons of tests for the presence of random walk coefficients in a simple linear model. *J*ournal of the American Statistical Association 78 (1982), 856–864.

Toward Automatic Local Spectral Envelope

David Stoffer
(joint work with Ori Rosen, Sally Wood)

The concept of spectral envelope for the scaling and analysis of categorical time series in the frequency domain was first introduced in Stoffer et al. [2]. There, we addressed the basic question of how to efficiently discover periodic components in categorical time series. Our present focus is on developing a better method for estimating a local spectral envelope. The initial method we are proposing is based on fitting local splines. The first step was to develop a method to estimate a spectral matrix function of a stationary vector process via smoothing splines. This step was accomplished in Rosen and Stoffer [1].

In many practical problems, time series are realizations of nonstationary random processes. Consequently, the next is to develop a method for fitting local spectra for univariate series. These processes can often be modeled as processes with slowly changing dynamics or as piecewise stationary processes. In these cases, various approaches to estimating the time varying spectral density have been proposed. Our approach in this paper is to estimate the log of the Dahlhaus-local spectrum using a Bayesian mixture of splines. The basic idea of our approach is to first partition the data into small sections. We then assume that the log spectral density of the evolutionary process in any given partition is a mixture of individual log spectra. We use a mixture of smoothing splines model to estimate the evolutionary log spectrum. The mixture model is fit using Markov chain Monte Carlo techniques that yield estimates of the log spectra of the individual subsections, as well as pointwise credible intervals for the unknown log spectrum. We use a reversible jump step to automatically determine the number of different spectral components.

Our final goal will be to combine all the method to obtain a local spectral envelope.

References

[1] Rosen, O. and Stoffer, D.S. (2007). Automatic estimation of multivariate spectra via smoothing splines. *Biometrika*, 94, 335–345.

[2] Stoffer, D. S., Tyler, D. E. & McDougall, A. J. (1993). Spectral analysis for categorical time series: Scaling and the spectral envelope. *Biometrika*, 80, 611–622.

Reporter: Claudia Kirch

Participants

Prof. Dr. John Aston
Institute of Statistical Science
Academia Sinica
Taipei , 11529
TAIWAN

Prof. Dr. Alexander Aue
Department of Statistics
University of California, Davis
One Shields Avenue
Davis CA 95616
USA

Prof. Dr. Rainer Dahlhaus
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg

Prof. Dr. Richard A. Davis
Department of Statistics
Columbia University
1255 Amsterdam Avenue, MC 4690
Room 1010 SSW
New York , NY 10027
USA

Prof. Dr. Christian Francq
GREMARS, UFR MSES
Universite Lille 3, Charles de Gaulle
Domaine Uinversitaire du Pont de Bois,
BP 60 149
F-59653 Villeneuve d'Ascq Cedex

Prof. Dr. Jürgen Franke
Fachbereich Mathematik
T.U. Kaiserslautern
Erwin-Schrödinger-Straße
67653 Kaiserslautern

Dr. Roland Fried
Fachbereich Statistik
Technische Universität Dortmund
44221 Dortmund

Prof. Dr. Lajos Horvath
Department of Mathematics
University of Utah
155 South 1400 East
Salt Lake City , UT 84112-0090
USA

Marie Huskova
Department of Probability and
Mathematical Statistics
Charles University
Sokolovska 83
18675 Praha 8
CZECH REPUBLIC

Dr. Claudia Kirch
Fachbereich Mathematik
T.U. Kaiserslautern
Erwin-Schrödinger-Straße
67653 Kaiserslautern

Prof. Dr. Wai-keung Li
Department of Statistics and
Actuarial Science
The University of Hong Kong
Hong Kong
HONG KONG

Rainer von Sachs
UCL-Institut de Statistique
Voie du Roman Pays 20
B-1348 Louvain-la-Neuve

Prof. Dr. Vladimir Spokoiny
Weierstrass-Institute for Applied
Analysis and Stochastics
Mohrenstr. 39
10117 Berlin

Dr. Joseph Tadjuidje-Kamgaing
Fachbereich Mathematik
T.U. Kaiserslautern
Erwin-Schrödinger-Straße
67653 Kaiserslautern

Prof. Dr. David S. Stoffer
Department of Statistics
University of Pittsburgh
Pittsburgh , PA 15260
USA