Universität zu Köln

Department Mathematik/Informatik
Prof. i.R. Dr. Josef G. Steinebach

Weyertal 86 - 90
50931 Köln, Germany
Telefon +49 221 470-2275
Telefax +49 221 470-5908

24.11.2019

Opponent's Report on the Habilitation Thesis
by
Michal Pešta (Prague)
on
"Dynamics and Instabilities in Time Series and Panel Data"

(ČJ: UKMFF/286521/2019-6)

The habilitation thesis under review is an extended collection of nine original research papers and one erratum, which have already been published, mostly in high-level mathematical journals. Two papers are single-authored works and the others are joint papers together with well-known coauthors in the field. Altogether they cover a considerably broad spectrum of topics and methods in mathematical and applied statistics as well as in risk theory and insurance mathematics. Focus is on the discussion of dynamics and instabilities in various stochastic frameworks like, for example, change-point problems in time series and panel data or developments in claim amount samples. Goal is to extend and further generalize procedures, which are capable of detecting and handling structural breaks or predicting important quantities in statistical data of different types. The latter are relevant in a variety of applied fields, e.g., in engineering, finance, insurance, economics, medicine, biology, hydrology, climatology, or ecology, to give just an incomplete list. The thesis essentially consists of three main chapters according to the different types of statistical framework under investigation.

After an introductory Chapter 1 describing the motivation and the structure of the thesis in more detail, Chapter 2 deals with change-point problems in time series. It is based on four original publications from 2016 to 2019, two of which are single-authored and the other two are joint works. The time series under consideration forms a triangular array with *atmost one abrupt change in the constant mean* and a single, α- mixing sequence of observation errors (cf. (2.1)). In order to test for a possible change in the mean, a ratio type test statistic based on $\mathcal{M}$-residuals is suggested (cf. (2.5)), which is a robustified version of a ratio type test statistic earlier considered by Hušková (2007) and Horváth et al. (2008). The main results derive the limit distribution of the test statistic (Theorem 2.1.1), which allows to determine asymptotic critical values (cf. Tab. 2.1), and show consistency of the procedure under alternatives of a certain order (Theorem 2.1.2). In addition, based on a (so-called) *circular moving block bootstrap method*, it is shown that a bootstrapped ratio type test statistic, conditioned on the original observations, has exactly the same limit behavior as the original test statistic under the null (Theorem 2.1.3). Here it does not matter whether the observations come from the null hypothesis or the alternative, which also means that the procedure provides asymptotically correct critical values in any case.

Chapter 2.2 discusses the testing of an abrupt change in a time series by making use of *self-normalized test statistics*. The latter avoid the problem of estimating possible nuisance parameters, such as, e.g., long-run variances, which may result in mis-estimation and thus lack of precision of the asymptotics. Two versions of self-normalized test statistics are suggested (cf. (2.10)-(2.11)) and their limit distributions under the null hypothesis are derived under certain assumptions (Theorem 2.2.1), which allow for heteroscedasticity and weak dependencies. Moreover, consistency of the procedures is verified (Theorem 2.2.2), so that the asymptotic distributions from Theorem 2.2.1 can be used to construct tests possessing asymptotic power 1. Since, however, the latter asymptotics cannot be given in an explicit form, *wild bootstraps* of the test statistics are studied which avoid and overcome the problem of estimating a nuisance function involved. The validity of these bootstraps is also verified (cf. Theorem 2.2.3). Finally, it can be shown that, under the alternative, the self-normalized test statistic may be used to consistently estimate the underlying change-point (Theorem 2.2.4).

The discussion of change-points in time series is completed in Chapter 2.3 by studying the detection of a possible change in the linear relation parameter β of an *errors-in-variable* (EIV) model (see p. 28), which also allows for non-stationarity. The *supremum* or *integral-type self-normalized test statistics*, based on a *goodness-of-fit* comparison over two different ranges of observations (cf. (2.17)-(2.18)), can be shown to converge to non-degenerate limit distributions under the null hypothesis (Theorem 2.3.2). The latter asymptotics are based on a *spectral weak invariance principle* provided in Proposition 2.3.1. Again, the test procedures turn out to be consistent (under local alternatives), cf. Theorem 2.3.3. Asymptotic critical values have to be determined via simulation (see Tab. 2.2) and an argmax-criterion provides a consistent estimator for the unknown change-point under the alternative (Corollary 2.3.4).

Chapter 3 deals with change-point problems in panel data and is based on three papers of Peštová and Pešta from 2016 to 2017. Main focus is an testing and estimating structural changes in a moderate or large number of panels, while the length of these panels is relatively short. Motivation comes, for example, from non-life insurance business where, for reliable predictions, it may be important to find out whether or not historical claim amount data, collected from an association of companies, are stable with respect to their underlying means.

The panel change-point model under investigation is described in Section 3.1 (cf. (3.1)). Key ingredients are the assumptions that the different panels i are independent, but may have individual means, and that within each panel the errors $\varepsilon_{i,t}$ form a weakly stationary sequence along t with a common correlation structure. The number N of panels tends to infinity, whereas the common panel length T is fixed. To test whether or not there is a change in the means of the panels at a common change-point τ, a ratio type test statistic $\mathcal{R}_N(T)$ is suggested (cf. p. 40). In Theorem 3.1.1 it is shown that, under the null hypothesis of no change, this test statistic has an asymptotically normal limit behavior, not depending on a variance nuisance parameter, but on the underlying (unknown) correlation structure, which would have to be estimated for testing purposes. Moreover, Theorem 3.1.2 shows that the asymptotic test is consistent under the given assumptions.

In order to possibly improve the determination of asymptotic critical values for the testing procedure, a residual bootstrap analogue $\mathcal{R}_N^*(T)$ of the above test statistic is introduced (see Chapter 3.1.5). Its validity is justified by showing that, under suitable assumptions, the asymptotic distribution of $\mathcal{R}_N(T)$ (under the null) and the conditional asymptotic distribution of $\mathcal{R}_N^*(T)|\mathbb{Y}$ (under null and alternative hypotheses) coincide, where $\mathbb{Y}$ denotes the observed panel data (cf. Theorems 3.1.4 and 3.1.5). Note that the bootstrap procedure does not require an estimation of the intra-panel correlation structure, which is a definite advantage compared to the asymptotic test based on Theorem 3.1.1.

A consistent change-point *estimator* is discussed in Section 3.1.3 requiring, however, the assumption that a change has actually occurred (cf. Theorem 3.1.3). Via modifying an approach of Bai (2010), another estimator $\hat{\tau}_N$ is later introduced in Section 3.2.2, turning out to be consistent if there is a change, but also resulting in the very last time-point T, if no change has occurred (cf. Theorem 3.2.1). The assumptions made are briefly discussed as well as possible generalizations are indicated. Moreover, various competing change-point estimators are suggested and a gap in the literature is also pointed out (cf. p. 51).

Motivated by relevant applications in insurance mathematics, Chapter 4 deals with the dynamic behavior in triangular data. Main focus is on the distribution-free (so-called) chain ladder approach introduced by Mack (1993), which can be used to estimate outstanding loss liabilities of an insurance company under quite mild assumptions on the mean structure, but under independence of the observations in different accident years. Various extensions and generalizations are provided, based on three papers of Pešta and coauthors from 2012 to 2014, essentially aiming at developing methods, which are capable of capturing dependencies among observations in the triangular setup.

After having introduced the classical claims reserving notation and terminology (Chapter 4.1.1), it is discussed why, from a statistical point of view, the often stressed unbiasedness of estimates $\hat{f}_j^{(n)}$ for the underlying developing factors f_j (cf. p. 55) is of less importance compared to their consistency (p. 56f). Consequently, Theorem 4.1.1 discusses necessary and sufficient conditions for the (conditional) consistency of the development factors' estimates, given the set of all past and future claims $C_{i,k}$. A more detailed discussion follows in Chapter 4.1.3, also including a convergence rate statement (cf. Propositions 4.1.2 - 4.1.3 and Corollary 4.1.4).

In Chapter 4.2, the common approach to the claims reserving problem, based on generalized linear models (GLM) and assuming that the claims in different origin and development years are independent, is extended, since, in case of violation of the independence assumptions, the classical techniques may provide incorrect predictions of the claims reserves or even misleading estimates of the prediction error. An application of generalized estimating equations (GEE) for the estimation of the claims reserves is shown to be an effective alternative to an earlier suggested generalized linear mixed models (GLMM) approach. Three main pillars of the GEE are introduced in Chapter 4.2.1, two of which are common with the usual GLM framework, but the third one is different. Because, in contrast to the GLM, the GEE approach does not require any specification of the whole distribution for the outcomes, but just assumes that it belongs to an exponential family of distributions together with a specification of the underlying variance-covariance structure. A detailed discussion of the GEE estimation procedure is given, including a selection of the working correlation structure, which may be closer to the true one. Application of the GEE approach to claims reserving, allowing for dependencies, is provided in Chapter 4.2.2. Moreover, model selection criteria for the GEE reserving method are proposed in Chapter 4.2.3 and the mean square error (MSE) of prediction for the claims reserves within this framework is derived in Chapter 4.2.4.

Main content of Chapter 4.3 is the discussion of a generalized time series model for a triangular collection $\{Y_{i,j}\}$ of claims amount data that allows for dependencies. The idea is to overcome estimation and prediction problems arising in earlier approaches, which assume independent claims in different years and may not be realistic. In addition to the GLMM or GEE approaches for handling possible dependencies among the incremental claims in successive development, which extend the classical GLM and are frequently used in panel data analyses, a possible alternative is investigated here, namely a conditional mean-variance (CMV) model including a copula function (cf. Definition 4.3.1 and *Assumption* $\mathcal{A}.4.29$). Aim is to predict the ultimate claims' amount $Y_{i,n}$ and the outstanding claims' reserve $R_i^{(n)} = Y_{i,n} - Y_{i,n+1-i}$ for all years $i = 2, \ldots, n$. In addition, estimation of the whole distribution of the reserves is intended to provide important distributional quantities like, e.g., quantiles for the value at risk. Conditional least squares (CLS) estimators (cf. Definition 4.3.2) are used to approximate the CMV model parameters and it is shown that the latter estimates are consistent (confer Theorems 4.3.1 - 4.3.2 and Corollary 4.3.3). Moreover, an algorithmic Procedure 4.3.1 is developed in order to provide a computational way for obtaining the CMV parameter estimates. The algorithm also provides the fitted residuals as a side product. The latter can then be used to estimate the unknown copula parameter (cf. Procedure 4.3.2). Finally, based on the CMV framework, a semiparametric bootstrap is suggested to predict the unobserved claims $\hat{Y}_{i,j}$ in a *telescopic* way (see p. 86). When the errors in this procedure are simulated sufficiently many times, the empirical (bootstrap) distribution of $\hat{Y}_{i,n}$ can mimic the true unknown distribution of $Y_{i,n}$ and may thus allow for calculating some imported risk quantities like, e.g., mean, variance, or quantiles of the reserves.

In conclusion, this comprehensive habilitation work under review shows a broad and sound competence of the candidate in his field. He proves to be familiar with deep and sophisticated methods in probability theory, mathematical statistics and their applications as well as to be able to provide significant contributions to the state of the art with new ideas and suggestions. Moreover, the habilitation thesis is very well written and organized, although there are highly involved techniques behind. Altogether it is an outstanding piece of scientific work, which certainly proves the candidate's ability for doing independent and successful research work in his field and, moreover, for dealing with relevant applications. So, I can only strongly recommend to accept this thesis as the author's cumulative habilitation work in probability and mathematical statistics.

Cologne, November 24, 2019

[redacted]

(Prof. i.R. Dr. J. G. Steinebach)