Computational Statistics and Data Analysis 82 (2015) 137–151

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust heart rate variability analysis by generalized entropy minimization

Davide La Vecchia [a,b,*], Lorenzo Camponovo [b], Davide Ferrari [c]

[a] *Department of Econometrics and Business Statistics, Monash University, Australia*
[b] *School of Economics and Political Science, University of St. Gallen, Switzerland*
[c] *Department of Mathematics and Statistics, University of Melbourne, Australia*

ARTICLE INFO

Article history:
Received 24 February 2014
Received in revised form 18 August 2014
Accepted 3 September 2014
Available online 16 September 2014

Keywords:
Heart rate variability
Time series analysis
Influence function
M-estimators
Robust estimation

ABSTRACT

Typical heart rate variability (HRV) times series are cluttered with outliers generated by measurement errors, artifacts and ectopic beats. Robust estimation is an important tool in HRV analysis, since it allows clinicians to detect arrhythmia and other anomalous patterns by reducing the impact of outliers. A robust estimator for a flexible class of time series models is proposed and its empirical performance in the context of HRV data analysis is studied. The methodology entails the minimization of a pseudo-likelihood criterion function based on a generalized measure of information. The resulting estimating functions are typically re-descending, which enable reliable detection of anomalous HRV patterns and stable estimates in the presence of outliers. The infinitesimal robustness and the stability properties of the new method are illustrated through numerical simulations and two case studies from the Massachusetts Institute of Technology and Boston's Beth Israel Hospital data, an important benchmark data set in HRV analysis.

1. Introduction

An electrocardiogram (ECG) is a representation of the heart electrical activity over a period of time, as detected by electrodes attached to the outer surface of the skin. The ECG is an important diagnostic tool for noninvasive investigation of heart rate and beat regularity, heart damages, effect of drugs or artificial devices on the cardiovascular system and many other medical applications. ECG signals are periodic waves characterized by the presence of sharp upward spikes, called R-peaks; the time between two consecutive R-peaks is called an RR interval (see Fig. 1). This paper is concerned with the analysis of HRV time series where an observation is given by HRV $= 60/(\text{RR interval length})$, measured in beats per minute. Over the years, HRV analysis has gained a major role in modern medicine for both diagnosis and modeling of the neuro-cardiovascular system; e.g., see Morillo (2012) for an introduction on this topic. Despite the importance of HRV analysis, however, relatively little has been written on robust statistical analysis of HRV times series. Relevant contributions in this area include Guo et al. (2001), Li (2008), Spangl and Dutter (2012), and Lu (2012).

Typically, HRV times series are cluttered with anomalous observations, some of which have medical significance. Outliers can be divided into two main types: representative and non-representative outliers (Chambers, 1984). We regard representative outliers as atypical observations pertaining to the actual HRV process; they are correctly recorded values with medical meaning and there is no reason to exclude the chance of observing again similar values. Non-representative outliers are incorrectly recorded values or observations corresponding to unique events. In HRV analysis, measurement errors

* Corresponding author at: Department of Econometrics and Business Statistics, Monash University, Australia.
E-mail address: davide.la-vecchia@monash.edu (D. La Vecchia).

http://dx.doi.org/10.1016/j.csda.2014.09.001
0167-9473/

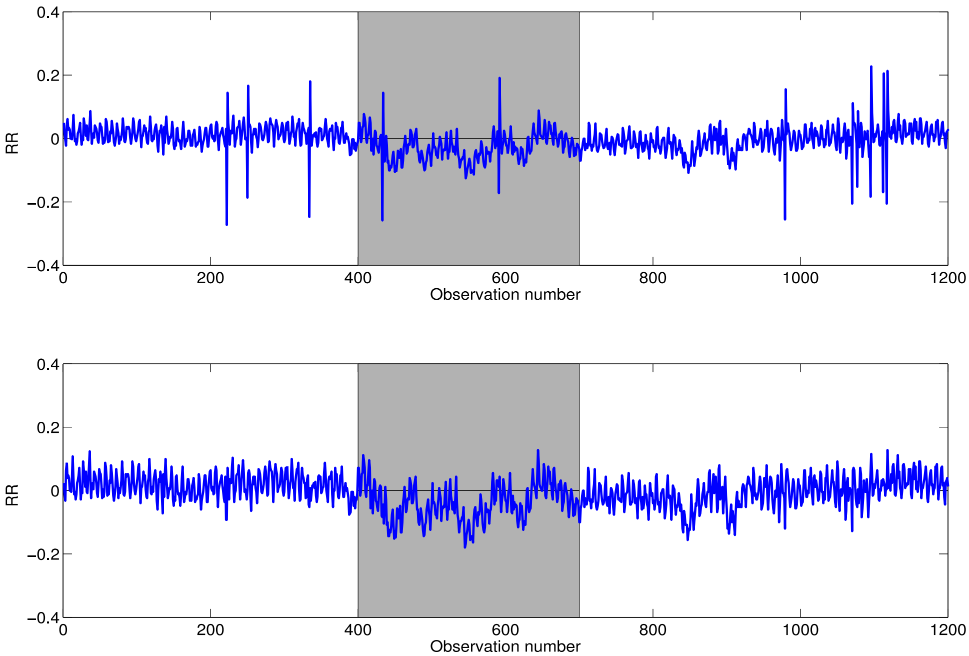

Fig. 1. RR intervals excerpt for Patient #100 from the MIT-BIH data set. Top panel: raw data (minus their mean) corresponding to 10 min from the central part of the ECG recording. Bottom panel: the same observations filtered by the `nguess` routine available on the MIT-BIH website.

(e.g., due to sudden movements of the patient) and artifacts (anomalous observations generated by the recording device) are non-representative outliers; ectopic beats (extra beats between two regular beats) can be either non-representative or representative outliers and can potentially convey information about meaningful but rare HRV patterns. In the context of time series models, such as the AR, ARMA and ARIMA models, an important example of representative outliers are the innovation outliers (IO), which affect the current observation as well as subsequent observations. Additive outliers (AO), defined as errors affecting a single observation, are examples of non-representative outliers. See Fox (1972) and Maronna et al. (2006, Chapter 8) for a more detailed discussion on IOs and AOs in the context of time series.

Clinicians are well-aware of these types of outliers and usually attempt to deal with them by data preprocessing, often involving visual and ad hoc outlier detection techniques. For example, currently available R and MATLAB statistical software packages for HRV analysis mostly deal with outliers by allowing the user to manipulate directly the HRV records (Perakakis et al., 2010; Rodríguez-Liñares et al., 2011). This kind of pre-processing, however, comes with a major risk: the filtered trajectory might be clear of evident outliers, but other atypical observations incompatible with the assumed model remain hidden within the processed sample. For example, in Fig. 1 we show a sample for Patient #100 from the Massachusetts Institute of Technology and Boston's Beth Israel Hospital (MIT-BIH) data set filtered by a standard routine.[1] In the bottom panel, the filtered records between 400 and 700 (shaded area) have shifted location and large variability compared to the rest of the observations. In stationary time series, this anomalous behavior is typically produced by outliers (Maronna et al., 2006, Chapter 8).

Autoregressive (AR), autoregressive moving average (ARMA), or bilinear models are standard model choices for HRV data (Christini et al., 1995; Spangl and Dutter, 2005). One of the most popular estimators for these models is the Gaussian pseudo maximum likelihood estimator (PMLE), where the Gaussian likelihood is used for the true but possibly unknown likelihood (Gourieroux et al., 1984). However, it is well-known that a handful of observations deviating from the parametric assumptions (e.g., observations violating the assumptions about either the first or the second conditional moment) can affect the PMLE, potentially leading to severely biased estimates (Mancini et al., 2005). Li (2008), Kaufmann et al. (2011), Clifford et al. (2012), and Behar et al. (2013) analyzed ECG records, showing that bias arises due to different types of anomalous records, such as measurement errors, artifacts, and ectopic beats.

In this paper, we advocate the use of robust inference methods for HRV data. Robust methods have a clear advantage over non-robust alternatives in this context: they are capable of producing parameter estimates resistant to outliers; hence, they enable clinicians to detect anomalous physiological patterns with increased confidence. From this perspective, we study a new estimator in the context of time series models, which we refer to as tilted and conditional maximum Lq-likelihood estimator (TC-MLqE). Then, we use the TC-MLqE to analyze the MIT-BIH data set. The MIT-BIH data set has been studied extensively for assessing arrhythmia detectors by manufacturers, as well as for comparing basic heart-related research, and it represents one of the most important benchmarks in HRV analysis (Lin and Tian, 2012).

[1] The MIT-BIH data set and the filtering routine `nguess` are freely available on http://www.physionet.org.