COMSTA: 5877 Model 3Gsc pp. 1–17 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A family of autoregressive conditional duration models applied to financial data

Q1 Víctor Leiva[a,*], Helton Saulo[b], Jeremias Leão[c], Carolina Marchant[a]

[a] Instituto de Estadística, Universidad de Valparaíso, Chile
[b] Instituto de Matemática e Estatística, Universidade Federal de Goiás, Brazil
[c] Departamento de Estatística, Universidade Federal do Piauí, Brazil

ARTICLE INFO

Article history:
Received 11 November 2013
Received in revised form 9 May 2014
Accepted 23 May 2014
Available online xxxx

MSC:
primary 62P20
secondary 62F99
60E05

Keywords:
Birnbaum–Saunders distribution
EM algorithm
High-frequency data
Maximum likelihood estimator
Monte Carlo simulation

ABSTRACT

The Birnbaum–Saunders distribution is receiving considerable attention due to its good properties. One of its extensions is the class of scale-mixture Birnbaum–Saunders (SBS) distributions, which shares its good properties, but it also has further properties. The autoregressive conditional duration models are the primary family used for analyzing high-frequency financial data. We propose a methodology based on SBS autoregressive conditional duration models, which includes in-sample inference, goodness-of-fit and out-of-sample forecast techniques. We carry out a Monte Carlo study to evaluate its performance and assess its practical usefulness with real-world data of financial transactions from the New York stock exchange.

1. Introduction

Q2

High-frequency data have gained increasing attention with the advances in computer technology and data recording and storage; see Engle (2000). The availability of high-frequency financial transaction (trade duration—TD) data has given an impulse to the research in business, economics and finance. The family of autoregressive conditional duration (ACD) models proposed by Engle and Russell (1998) has been the primary tool used for analyzing TD data, which are irregularly time-spaced and convey meaningful information. The importance of this type of data, and of their modeling, is stressed by the market microstructure literature; see, e.g., Pacurar (2008) and references therein.

Although TD data (i) have usually a unimodal hazard rate (HR) and (ii) follow an asymmetric distribution with heavy tails (see Grammig and Maurer, 2000; Bhatti, 2010), generalizations of the original ACD model are based on assumptions that do not necessarily comply with the characteristics (i) and (ii). Generalizations of the ACD model should be based on assumptions that take into account (A1) the shape of the HR of TD data; (A2) the symmetry or asymmetry of their distribution; and (A3) the conditional dynamics established in terms of their mean or median, depending on their symmetry or asymmetry; see Bauwens and Giot (2000), De Luca and Zuccolotto (2006), Fernandes and Grammig (2006) and Allen et al. (2008).

* Correspondence to: Instituto de Estadística, Universidad de Valparaíso, Gran Bretaña 1111, Playa Ancha, Valparaíso, Chile.
E-mail address: victor.leiva@uv.cl (V. Leiva).
URL: http://www.deuv.cl/leiva (V. Leiva).

http://dx.doi.org/10.1016/j.csda.2014.05.016
0167-9473/

Please cite this article in press as: Leiva, V., et al., A family of autoregressive conditional duration models applied to financial data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.05.016

Birnbaum and Saunders (1969) introduced a distribution to model fatigue life data, assuming that the failure follows from the development and growth of a dominant fissure produced by stress. The Birnbaum–Saunders (BS) distribution has been widely studied because of its good properties and its relation with the normal distribution; see, e.g., Cysneiros et al. (2008), Balakrishnan et al. (2009a), Balakrishnan et al. (2011), Kotz et al. (2010), Vilca et al. (2010), Vilca et al. (2011), Villegas et al. (2011), Ferreira et al. (2012), Leiva et al. (2012), Li and Xie (2012), Vanegas et al. (2012), Fierro et al. (2013), Lemonte (2013) and Barros et al. (in press). In addition, although it has its genesis from engineering, its applications have been considered in other fields, including business, economics, finance and quality control; see Jin and Kawczak (2003), Balakrishnan et al. (2007), Ahmed et al. (2010), Bhatti (2010), Leiva et al. (2011b), Leiva et al. (2014a), Leiva et al. (2014b), Leiva et al. (2014c), Paula et al. (2012) and Marchant et al. (2013). The BS distribution is asymmetrical, has positive skewness and a unimodal HR, and has been successfully applied to model lifetime data; see e.g. Leiva et al. (2007) and Leiva et al. (2008a). Thus, it can be a good model for describing TD data. Bhatti (2010) proposed an extension of the ACD model based on the BS distribution (BS–ACD model), which provides (B1) a realistic distributional assumption (in terms of the shape of its probability density function - PDF - and of its HR); (B2) an easy parameter estimation (because it is simple, converges fast and has initial values for the numerical procedure that can be straightforwardly obtained); and (B3) a natural parameterization in terms of a conditional median duration, which is expected to improve the model fit, instead of using the conditional mean duration; see (A3). This is because the median is often considered as a better measure of central tendency than the mean, when the data follow asymmetric and heavy-tailed distributions, such as is the case of TD data.

Recently, based on the relationship between the BS and normal distributions, Balakrishnan et al. (2009b) introduced the scale-mixture BS (SBS) distributions; see also Díaz-García and Leiva (2005) and, for a recent TD data analysis using kernel estimation based on SBS models with independent data, see Marchant et al. (2013). The class of SBS distributions (C1) inherits the good properties of the BS distribution discussed in (B1)–(B3); and (C2) permits the maximum likelihood (ML) estimates of the model parameters to be computed in an efficient way, using the expectation–maximization (EM) algorithm, and the corresponding estimation procedure to be robust for some of its members.

The main objective of this work is to propose a methodology based on ACD models generated from SBS distributions (SBS–ACD models). This methodology includes in-sample techniques considering estimation of the model parameters via the EM algorithm, inference about these parameters and model checking based on residual analysis, and out-of-sample forecast techniques. We evaluate our methodology with Monte Carlo (MC) methods and apply it to TD data, which have unique features absent in data with low frequencies. For example, as mentioned, TD data (D1) are collected in irregular time intervals; (D2) possess a large number of observations; (D3) their trading activities exhibit some diurnal patterns, so that activity is higher in the beginning and closing than in the middle of the trading day; and (D4) present a unimodal HR; see Engle and Russell (1998) and Bhatti (2010).

The remaining of this paper unfolds as follows. In Section 2, we present SBS distributions. In Sections 3 through 6, we propose the methodology based on SBS–ACD models, which includes their formulation, estimation and inference of their parameters, robustness and in-sample model checking analyses, and out-of-sample forecast techniques. In Section 7, we conduct an MC study to evaluate the performance of this methodology. In Section 8, we apply it to six real-world data sets of NYSE securities. Finally, in Section 9, we discuss some conclusions and future studies.

2. Scale-mixture Birnbaum–Saunders distributions

As mentioned, BS distributions have been considered for analysis of TD data with independent and dependent structures; see Jin and Kawczak (2003) and Bhatti (2010), respectively. A random variable (RV) X follows a BS distribution if it can be represented by

$$X = \sigma\left[\kappa Z/2 + \{(\kappa Z/2)^2 + 1\}^{1/2}\right]^2, \tag{1}$$

where $Z \sim \mathrm{N}(0, 1)$ and $\kappa > 0, \sigma > 0$ are shape and scale parameters, respectively. In this case, the notation $X \sim \mathrm{BS}(\kappa, \sigma)$ is used and the corresponding PDF is given by

$$f_{\mathrm{BS}}(x; \kappa, \sigma) = \frac{1}{[2\pi]^{1/2}} \exp\left(-\frac{1}{2\kappa^2}\left[\frac{x}{\sigma} + \frac{\sigma}{x} - 2\right]\right) \frac{x^{-3/2}[x+\sigma]}{2\kappa\sigma^{1/2}}, \quad x > 0.$$

Note that, as the shape parameter κ goes to zero, the BS distribution tends to be symmetrical, degenerating at σ, when $\kappa = 0$ (see Kundu et al., 2008), whereas the scale parameter σ is also the median of the distribution. The BS model holds proportionality and reciprocal properties given by $bX \sim \mathrm{BS}(\kappa, b\sigma)$, with $b > 0$, and $1/X \sim \mathrm{BS}(\kappa, 1/\sigma)$, respectively.

Scale mixture of normal (SMN) distributions have been considered for analysis of financial time series; see Tsay (2002, pp. 12–13). An RV Y follows an SMN distribution (which is symmetrical) if it can be represented by $Y = \mu + \{g(U)\}^{1/2}V$, where $V \sim \mathrm{N}(0, \vartheta^2)$, $\mu \in \mathbb{R}$, $\vartheta^2 > 0$, U is a positive RV independent of V, with cumulative distribution function (CDF) $H(\cdot)$, and $g(\cdot)$ is a positive function associated with $H(\cdot)$. This is denoted by $Y \sim \mathrm{SMN}(\mu, \vartheta^2, H)$ and its PDF is

$$\phi_{\mathrm{SMN}}(y; \mu, \vartheta, H) = \int_0^\infty \phi(y; \mu, g(u)\vartheta^2)\, \mathrm{d}H(u), \quad y \in \mathbb{R}, \tag{2}$$