Computational Statistics and Data Analysis 54 (2010) 2580–2593

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Wavelet-based detection of outliers in financial time series

Aurea Grané[a,*], Helena Veiga[a,b,1]

[a] Statistics Department, Universidad Carlos III de Madrid, Spain

[b] Finance Research Center, ISCTE Business School, Portugal

ARTICLE INFO

Article history:
Received 22 April 2009
Received in revised form 17 December 2009
Accepted 18 December 2009
Available online 2 January 2010

Keywords:
Outliers
Outlier patches
Prediction intervals
Volatility models
Wavelets

ABSTRACT

Outliers in financial data can lead to model parameter estimation biases, invalid inferences and poor volatility forecasts. Therefore, their detection and correction should be taken seriously when modeling financial data. The present paper focuses on these issues and proposes a general detection and correction method based on wavelets that can be applied to a large class of volatility models. The effectiveness of the new proposal is tested by an intensive Monte Carlo study for six well-known volatility models and compared to alternative proposals in the literature, before it is applied to three daily stock market indices. The Monte Carlo experiments show that the new method is both very effective in detecting isolated outliers and outlier patches and much more reliable than other alternatives, since it detects a significantly smaller number of false outliers. Correcting the data of outliers reduces the skewness and the excess kurtosis of the return series distributions and allows for more accurate return prediction intervals compared to those obtained when the existence of outliers is ignored.

1. Introduction

Return series of financial assets typically exhibit high kurtosis, higher order dependence and volatility clustering. Autoregressive conditional heteroscedastic models (ARCH and GARCH models) are the most popular models in parameterizing the higher order dependence and the evolution of volatility. Since their introduction to the literature by Engle (1982) and by Bollerslev (1986), respectively, they have been extended in several directions. The first extension, proposed by Bollerslev (1987), allowed the error of the GARCH model to follow a t-Student distribution in order to accommodate the high kurtosis of the data. However, it has been observed that the estimated residuals from this extended model still register excess kurtosis (see Baillie and Bollerslev, 1989; Teräsvirta, 1996). One possible reason for this occurrence is that some observations on returns are not fitted by a Gaussian GARCH model and not even by a t-distributed GARCH model. These observations may be influential (see Zhang, 2004, for a detailed definition of influential observation) since they can affect undesirably the estimation of parameters (see for example Fox, 1972; Van Dijk et al., 1999; Verhoeven and McAleer, 2000), the tests of conditional homoscedasticity (see Carnero et al., 2007; Grossi and Laurini, 2009) and the out-of-sample volatility forecasts (see for instance Ledolter, 1989; Chen and Liu, 1993a; Franses and Ghijsels, 1999; Grané and Veiga, 2009). When this is the case, some authors denote them by outliers and distinguish between additive and innovational (or innovations) outliers. The first type is classified into two categories: additive level outliers (ALO), which exert an effect on the level of the series but not on the evolution of the underlying volatility, and additive volatility outliers (AVO), that also affect the conditional variance (see Hotta and Tsay, 1998; Sakata and White, 1998). In order to illustrate the importance of outlier

* Corresponding address: Statistics Department, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe, Spain.
E-mail addresses: aurea.grane@uc3m.es (A. Grané), mhveiga@est-econ.uc3m.es (H. Veiga).

1 Finance Research Center, ISCTE Business School, Avenida das Forças Armadas, 1600-083 Lisboa, Portugal.

0167-9473/$ – see front matter
doi:10.1016/j.csda.2009.12.010

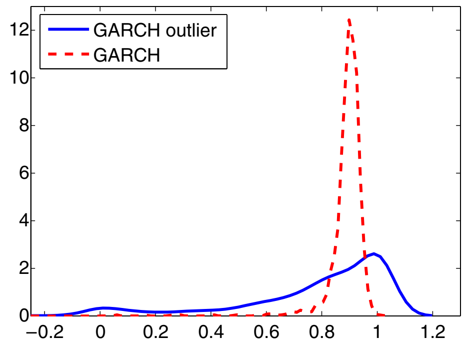

Fig. 1. Kernel density of $\hat{\beta}_1$ obtained from 1000 samples of size $n = 500$ from a GARCH(1, 1) with a true parameter value $\beta_1 = 0.912$.

detection, in Fig. 1 we depict the anomalies caused by one outlier of moderate magnitude on the distribution of one of the parameters of a GARCH(1, 1) model.

In Grané and Veiga (2009) we also found that the presence of outliers, even of small magnitude, biases the estimates of the minimum capital risk requirements (MCRRs), a measure of risk used by financial institutions that consists of a part of capital that should be reserved to absorb a pre-specified percentage of unforeseen losses. This reinforces the importance of outlier detection, since accurate estimates of MCRRs are crucial in avoiding a wasting of valuable resources.

This paper focuses mainly on the detection of additive (level and volatility) outliers by using wavelets. The effects of innovational outliers on the dynamic properties of the series are less important because they are propagated by the same dynamics, as in the rest of the series (see for example Peña, 2001). Wavelets are a family of basis functions that allows other functions to be expressed and approximated. In fact, wavelet coefficients are able to detect changes in variance, level changes and discontinuities in functions. For instance, wavelet variance analysis has been used by Gallegati and Gallegati (2007) to analyze the phenomenon of the "Great Moderation" (see Summers, 2005), a subject of an interesting debate in macroeconomics to denote a sharp, persistent and widespread fall in the volatility of aggregate economic activity over the last two decades. These authors were able to characterize the decline in volatility and to identify possible explanations for the moderation.

Our approach is inspired by Bilen and Huzurbazar (2002), who proposed an outlier detection method based on wavelets, but departs from theirs in the way the threshold limits are obtained. They used the proposals suggested by Donoho and Johnstone (1994) and Wang (1995), which rely on the assumption that the data is Gaussian. Moreover, since their threshold limits are quite conservative, their procedure leads to an extremely high average of false outliers. Our method for computing the threshold limits, on the other hand, is based on the distribution of the maximum of the detailed coefficients (in absolute value) obtained by Monte Carlo. In fact, the threshold is taken to be the 95th-percentile of the distribution of this maximum. In this way, our procedure can be applied to the estimated residuals of different volatility models with errors following any known distribution. The proposal deals with the estimated model residuals because we are interested in detecting if an observation is an outlier for a particular volatility model. It can also be seen as a mis-specification test, in the sense that if the number of outliers is substantially large it may indicate that the model is not appropriate.

Hotta and Tsay (1998), Franses and Ghijsels (1999), Doornik and Ooms (2005) and Zhang and King (2005) proposed several alternatives to detect outliers in GARCH models. The method of Hotta and Tsay (1998), that is based on a Lagrange Multiplier test suffers from the "masking phenomenon", as pointed out by Zhang and King (2005). This phenomenon occurs when one outlier prevents others from being detected and it is very common to encounter it while dealing with real data. The proposals by Franses and Ghijsels (1999) and Doornik and Ooms (2005) were inspired by the work of Chen and Liu (1993b), and are therefore susceptible to the same criticisms: a recursive procedure that needs to re-estimate a GARCH model several times, with the risk that the estimates of the parameters may be affected by the presence of a remaining outlier. Zhang and King (2005) developed a curvature-based directional diagnostic to examine the local influence of minor perturbations in a regression and a GARCH model, that can also be used to detect outliers.

The contributions of the present paper are several. First, we propose a method for outlier detection and correction that can be applied to the residuals of different volatility models whose errors may follow any known distribution. Secondly, it is well suited for one outlier or multiple outlier detections. Thirdly, it detects patches of outliers in different volatility models. Fourthly, our detection procedure can be extended to innovational outliers. Finally, the method is easy and quick to apply, which makes it an attractive tool for use by academic communities and/or by practitioners.

The effectiveness of our proposal is tested applying it to several volatility models, such as the GARCH(1, 1), the GJR(1,1) (see Glosten et al., 1993) and the autoregressive stochastic volatility model, ARSV(1), by Taylor (1986), with errors following a Gaussian or a Student's t distribution and comparing it, whenever it is possible, with the proposals of Bilen and Huzurbazar (2002), Franses and Ghijsels (1999) and Doornik and Ooms (2005). The intensive Monte Carlo study reveals that our proposal is not only as good as these alternatives in detecting outliers, but it is also much more reliable, since it detects a significantly smaller number of false outliers. Indeed, one may be sure that when one observation is detected as a possible outlier, it is an outlier.