European Journal of Operational Research 243 (2015) 852–864

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

Clustering financial time series: New insights from an extended hidden Markov model

José G. Dias [a,*], Jeroen K. Vermunt [b], Sofia Ramos [c]

[a] Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (UNIDE-IUL), Portugal
[b] Department of Methodology and Statistics, Tilburg University, The Netherlands
[c] NEOMA Business School, France & Business Research Unit (UNIDE-IUL), Portugal

ARTICLE INFO

Article history:
Received 30 December 2010
Accepted 23 December 2014
Available online 30 December 2014

Keywords:
Data mining
Hidden Markov model
Stock indexes
Latent class model
Regime-switching model

ABSTRACT

In recent years, large amounts of financial data have become available for analysis. We propose exploring returns from 21 European stock markets by model-based clustering of regime switching models. These econometric models identify clusters of time series with similar dynamic patterns and moreover allow relaxing assumptions of existing approaches, such as the assumption of conditional Gaussian returns. The proposed model handles simultaneously the heterogeneity across stock markets and over time, i.e., time-constant and time-varying discrete latent variables capture unobserved heterogeneity between and within stock markets, respectively. The results show a clear distinction between two groups of stock markets, each one characterized by different regime switching dynamics that correspond to different expected return-risk patterns. We identify three regimes: the so-called bull and bear regimes, as well as a stable regime with returns close to 0, which turns out to be the most frequently occurring regime. This is consistent with stylized facts in financial econometrics.

1. Introduction

In recent years dealing with unobserved heterogeneity has become a predominant topic in many research areas. As Heckman emphasized in his Nobel lecture, one of the most important discoveries in microeconometrics is the pervasiveness of heterogeneity and diversity in economic life: “When a full analysis of heterogeneity in response was made, a variety of candidate averages emerged to describe the average person, and the long-standing edifice of the representative consumer was shown to lack empirical support. (Heckman, 2001, p. 674)”. In finance research, heterogeneity has been mostly assumed observed (e.g., based on countries), where groups and their boundaries are delineated without regarding the intrinsic information on the observed data. However, there are plenty of examples in the academic and professional finance literature that show that heterogeneity exists among capital market participants, business managers, fund managers, among others.

The correct modeling of the dynamics of stock market returns has been an important challenge in modern financial research. Though the dominant approach followed by both academics and practitioners has been to assume that returns follow a normal distribution (see, e.g., Lundblad, 2007 and Fu, 2009), it has also been recognized that stock market returns and returns of financial assets contain skewness and excessive kurtosis. A common conclusion is that the normal distribution is inadequate for short period returns of financial assets (Fama, 1965; Mandelbrot, 1963; Praetz, 1972). Several alternative distributions have therefore been suggested for modeling returns, one of which is the Laplace distribution. These alternatives have in common that they try to accommodate for the excessive kurtosis in the empirical distribution of the returns. Whereas excess kurtosis of financial return distributions has been well addressed in the financial literature, the asymmetry of the distribution has not received much attention, and the few studies available tend to be inconclusive (Peiró, 1999; Simkowitz & Beedles, 1980; Singleton & Wingender, 1986).

Latent class or finite mixture modeling has proven to be a powerful tool for analyzing a wide range of social and behavioral science data (see, for example, Clogg, 1995 and Vermunt, 2003). The identification of distinct dynamics in time series data has been an important topic of research from a substantive point of view. We propose a latent class model for financial data analysis that takes into account unobserved heterogeneity by means of time-constant and time-varying discrete latent variables. A feature of latent class modeling is that it yields a model-based clustering of observational units that is especially attractive to the typical analysis in finance research, where it

* Corresponding author at: Av. das Forças Armadas, 1649-026 Lisboa, Portugal. Tel.: +351 217 903 208, 351 21 479 33 88; fax: +351 217 903 004.
E-mail addresses: jose.dias@iscte.pt (J. G. Dias), J.K.Vermunt@uvt.nl (J. K. Vermunt), sofia.ramos@neoma-bs.fr (S. Ramos).

http://dx.doi.org/10.1016/j.ejor.2014.12.041
0377-2217/

is of interest to identify subpopulations of firms, investors, markets or countries that differ in their propensities to specific characteristics (regulation, company governance characteristics, etc.). The challenging task of clustering this type of observations is that one has to incorporate data dependency in the clustering process (Kakizawa, Shumway, & Taniguchi, 1998). Therefore, clustering of time series has attracted large attention in statistics and data mining literature. However, most of the proposals developed have been based on classic clustering algorithms in such a way that time series data can be handled. Interested readers may refer to Liao (2005) and Esling and Agon (2012) who provide a detailed discussion of time series clustering in the data mining literature. Contrary to alternative heuristic clustering techniques for financial time series analysis which operate directly on the correlations (e.g., Basalto et al., 2007; Mantegna, 1999) or other measures of similarity between time series being clustered (e.g., Bastos & Caiado, 2013), the approach proposed here is a model-based clustering technique that accommodates for serial dependencies and unobserved heterogeneity by assuming a regime-switching model (RSM), also known as hidden Markov model, underlying each cluster. There is a vast body of literature on RSMs in economics and empirical finance, including Hamilton (1989), Hamilton and Susmel (1994), and Gray (1996) to name just a few. Hidden Markov models and regime-switching models as discrete state models can also be connected to stochastic volatility models (see, e.g., Langrock, MacDonald, & Zucchini, 2012; Rossi & Gallo, 2006). The autoregressive conditional root (ACR) model (Bec, Rahbek, & Shephard, 2008) is another econometric model that connects to regime-switching models. Many extensions of the regime-switching models have been suggested, adding new possibilities and modeling additional stylized facts of financial time series. For instance, Lux and Morales-Arias (2010) proposes a model that takes long memory and heavy tails of return time series into account, whereas Guidolin and Timmermann (2007), Fu, Wei, and Yang (2014), and Bae, Kim, and Mulvey (2014) apply RSMs to asset allocation and portfolio optimization. RSMs have been applied to pension funds optimization (Hainaut, 2014) and weather derivatives (Elias, Wahab, & Fang, 2014). For recent surveys on the application of RSMs in empirical finance, we refer to Lange and Rahbek (2009) and Guidolin (2011). These models have broader fields of application, covering manpower systems, where both observable and latent sources of dynamic heterogeneity should be accounted for (Guerry, 2011), and reliability analysis (Zhou, Hu, Xu, Chen, & Zhou, 2010).

Here, RSMs are extended to take the clustering structure of the returns of 21 European stock market indexes into account. Stock markets are well-known for presenting cycles, however country idiosyncrasies are also likely to make them differ in their transition between boom and bust. As is illustrated below, the proposed approach is flexible in the sense that it can deal with the specific features of financial time series data, such as asymmetry, kurtosis, and unobserved heterogeneity, an aspect that tends to be neglected. Because we selected a rather large and heterogeneous sample of countries including both developed and emerging countries and both EMU (European Monetary Union) and non-EMU countries, we expect that heterogeneity in market returns due to country idiosyncrasies will show up in the results. For instance, emerging market return distributions tend to show larger deviations from normality, i.e., they are more skewed and have fatter tails (Harvey, 1995). In addition, stock markets are also known to contain asymmetry of volatility, i.e., volatility is higher in negative shocks than positive ones (see, e.g., Ang & Bekaert, 2002) and structural breaks in time, or in other words, regime switching.

The results show that stock markets are better described by three regimes: A high returns, a negative returns, and close to zero regimes that we interpret as a bull, a bear, and a stable regime. This, however, challenges the simplistic view that stock markets should be characterized by two regimes. The characterization of regimes is consistent with several stylized facts such as asymmetry in volatility, i.e., bear regimes are associated with larger volatility than bull regimes. Stock markets are clustered into two groups that are mainly distinguished by the propensity to switch to the bear regime, which includes countries that were more affected by crisis during the period of analysis as well as Eastern European emerging countries that were less integrated with the other European countries.

The paper is structured as follows. Section 2 describes the 21 country financial time series data set that is used throughout this paper. Section 3 presents the statistical framework for the analysis of heterogeneous financial time series. It also discusses parameter estimation by maximum likelihood and model selection issues. Section 4 reports the results obtained for the data set at hand. The paper concludes with a summary of the main findings and a description of possible implications.

2. Description of the data set

The data used in this article are daily closing prices from 27 January 1998 to 31 July 2013 for 21 European stock market indexes drawn from Datastream database.[1] The series are denominated in US dollars. In total, we have 4010 end-of-the-day observations per country. Let P_{it} be the observed daily closing price of market i on day t, $i = 1, \ldots, n$ and $t = 0, \ldots, T$. The daily rates of return are defined as the percentage log-return by $y_{it} = 100 \times \log(P_{it}/P_{i,t-1})$, $t = 1, \ldots, T$. This definition which is commonly used in the literature is justified by the fact that for expected small increases (decreases) of value, say r, $\log(1 + r) \simeq r$.

The 21 stock markets are listed in Table 1. Figs. 1 and 2 depict for six distinct countries the index and returns time series, respectively. As is well known, stock markets follow cycles. In the sample period there were two main periods of global stock market crises. The dot-com bubble bursting that started at the end of 1999 and went on until 2003, and the subprime crisis that had its first signs in the summer of 2007, and made stock markets plummet in September 2008 after the Lehman Brothers bankruptcy. From 2004 to 2007, stock markets registered a strong growth.

Fig. 2 depicts stock market returns. Russia and Turkey have the highest level of volatility, which is typical of emerging markets. It is worth to note that in August 1998 Russia defaulted a sovereign bond payment triggering the “ruble crisis” in financial markets. Market features like these seem well-suitable to test our econometric model.

Table 1 provides the relevant descriptive statistics for the 21 stock-return series. All markets show non-negative median returns. However, only 18 out of 21 had positive mean returns; that is, Greece, Italy, and Portugal showed negative mean returns. Emerging markets such as Czech Republic and Russia show larger positive mean return. These figures confirm that stock market distributions tend to be negatively skewed.

The 21 analyzed markets show very diverse patterns of dispersion, where the largest standard deviations are found for Russia (3.134) and Turkey (2.983) – both emerging markets – which are almost three times larger than for Switzerland, the stock market with the smallest dispersion with a standard deviation of 1.210. Moreover, the excess kurtosis shows values above 0, indicating heavier tails and more peakness than the normal distribution (which has a kurtosis of 0). The Jarque–Bera test rejects the null hypothesis of normality for each of the 21 stock markets’ returns.

Figs. 3 and 4 depict rolling means and standard deviations (30-day window) for these markets. Although moving averages tend to smooth trends, the main booms and peaks in the stock markets’ returns are still visible. The ruble crisis is visible in Russia and it propagates to neighbor markets such as Poland and Hungary (not shown). All stock markets show a volatility peak during the subprime crisis.

[1] Observations from different time zones can create problems of non-synchronization on the analysis, to eliminate such problems we focus on European markets.