Computational Statistics and Data Analysis 56 (2012) 3366–3380

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

A new class of independence tests for interval forecasts evaluation

P. Araújo Santos [a,*], M.I. Fraga Alves [b]

[a] *Escola Superior de Gestão e Tecnologia de Santarém and CEAUL, Portugal*

[b] *Faculdade de Ciências da Universidade de Lisboa and CEAUL, Portugal*

ARTICLE INFO

Article history:
Received 15 December 2009
Received in revised form 30 September 2010
Accepted 1 October 2010
Available online 8 October 2010

Keywords:
Hit sequence
Backtesting
Quantitative risk management

ABSTRACT

Interval forecasts evaluation can be reduced to examining the unconditional coverage and independence properties of the hit sequence. A new class of exact independence tests for the hit sequence and a definition for tendency to clustering of violations are proposed. The tests are suitable for detecting models with a tendency to generate clusters of violations and are based on an exact distribution that does not depend on an unknown parameter. The asymptotic distribution is also derived. The choice of one test within the class is studied. Moreover, a simulation study provides evidence that, in order to test the independence hypothesis, the suggested tests perform better than other tests presented in the literature. An empirical application is given for a period that includes the 2008 financial crisis.

1. Introduction

One of the core topics of quantitative financial risk management is the accurate calculation of the Value at Risk (VaR), which amounts to a tail quantile of the forecast profit and loss distribution over a specified time horizon. Owing to the non-i.i.d. and non-Gaussian nature of financial asset returns data, the calculation of VaR is not trivial; see, e.g., Kuester et al. (2006) and the references therein for a survey of competing methods. The primary tool for assessing its accuracy is to monitor the binary sequence generated by observing if the return on day $t+1$ is in the tail region specified by the VaR at time-t, or not. This is referred to as the *hit sequence*. In mathematical terms we consider a time series of daily log returns, $R_{t+1} = \log(V_{t+1}/V_t)$, where V_t is the value of the portfolio at time-t. The corresponding one-day-ahead VaR forecasts made at time-t for time $t+1$, $\mathrm{VaR}_{t+1|t}(p)$, are defined by

$$P[R_{t+1} \leq \mathrm{VaR}_{t+1|t}(p)|\Omega_t] = p,$$

where Ω_t is the information set up to time-t and p is the *coverage rate*. Considering that a *violation* occurs when the daily return on the portfolio is lower than the reported VaR, we define the hit function as

$$I_{t+1}(p) = \begin{cases} 1 & \text{if } R_{t+1} < \mathrm{VaR}_{t+1|t}(p) \\ 0 & \text{if } R_{t+1} \geq \mathrm{VaR}_{t+1|t}(p). \end{cases} \quad (1)$$

Christoffersen (1998) showed that evaluating interval forecasts can be reduced to examining whether the hit sequence, $\{I_t\}_{t=1}^T$, satisfies the unconditional coverage (UC) and independence (IND) properties. UC hypothesis means $P[I_{t+1}(p) = 1] = p, \forall t$. IND hypothesis means that past information does not hold information about future violations.

Clustering of violations is one problematic infraction to the IND hypothesis, which corresponds to several large losses occurring in a short period of time. As noted by Campbell (2007), the IND property represents a more subtle yet equally

* Corresponding address: Departamento de Informática e Métodos Quantitativos, Escola Superior de Gestão e Tecnologia, Instituto Politécnico de Santarém, Complexo Andaluz, Apartado 295, 2001-904 Santarém, Portugal. Tel.: +351 243 303 200; fax: +351 243 332 152.
E-mail addresses: paulo.santos@esg.ipsantarem.pt (P. Araújo Santos), isabel.alves@fc.ul.pt (M.I. Fraga Alves).

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.10.002

important property. However, some authors argue that a certain amount of moderate clustering may not be harmful, so that a correct UC is somewhat more important than independence (e.g. Jorion, 2002). When both properties are valid we say that forecasts have a correct conditional coverage (CC) and we write

$$P[I_{t+1}(p) = 1|\Omega_t] = p, \quad \forall t. \tag{2}$$

In Lemma 1 of Christoffersen (1998) it is shown that condition CC (2) is equivalent to $I_{t+1}(p) \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$. In a recent paper, Berkowitz et al. (2009) extended and unified the existing tests by noting that the de-meaned hits $\{I_{t+1} - p\}$ form a martingale difference sequence. Eqs. (1) and (2) imply that $E[(I_{t+1} - p)|\Omega_t] = 0$ and then for any variable Z_t in the time-t information set, we must have

$$E[(I_{t+1} - p)Z_t] = 0. \tag{3}$$

This is the motivation for tests based on the martingale property.

The rest of the paper is organized as follows. In Section 2 we review existent tests for evaluating interval forecasts. In Section 3 we present the new class of independence tests and exact and asymptotic distributions are derived for a random variable (rv) related with the test statistic. The choice of one test within the class is also studied. In Section 4, and through simulation experiments, we compare the performance with other tests under study. Section 5 presents an empirical application. Section 6 concludes.

2. Tests for interval forecasts evaluation

There are several backtesting procedures for evaluating interval forecasts; for a detailed review see Campbell (2007) and Berkowitz et al. (2009). The first procedures were mainly concerned with the UC property and the proportion of failures (POF) test proposed by Kupiec (1995) is a well known example. A simple autocorrelation based independence test was proposed by Granger et al. (1989). In the last ten years, several tests have been suggested to examine both the IND and the CC properties. The Christoffersen (1998) Markov tests are perhaps the most widely used in the literature. Therein π_{ij} is defined as $P[I_t = j|I_{t-1} = i]$, for $i, j \in \{0, 1\}$. In this context, the null hypothesis of the IND test is $H_{0,\text{IND}} : \pi_{01} = \pi_{11}$ and the null hypothesis of the CC test is $H_{0,\text{CC}} : \pi_{01} = \pi_{11} = p$. Denoting by π_1 the common value of π_{01} and π_{11} under $H_{0,\text{IND}}$, by T_0 the number of zeros in the hit sequence $\underset{\sim}{I}$, T_1 the number of ones, $T = T_0 + T_1$ and T_{ij} the number of observations with a j following an i, the maximum likelihood estimators are $\hat{\pi}_{01} = T_{01}/T_0$, $\hat{\pi}_{11} = T_{11}/T_1$ and $\hat{\pi}_1 = T_1/T$, the log-likelihood under the alternative hypothesis is

$$\log L(\underset{\sim}{I}, \pi_{01}, \pi_{11}) = (1 - \pi_{01})^{T_0 - T_{01}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_1 - T_{11}} \pi_{11}^{T_{11}},$$

the IND test statistic is

$$LR_{\text{IND}} = -2(\ln L(\underset{\sim}{I}, \hat{\pi}_1) - \ln L(\underset{\sim}{I}, \hat{\pi}_{01}, \hat{\pi}_{11})), \tag{4}$$

and the CC test statistic is

$$LR_{\text{CC}} = -2(\ln L(\underset{\sim}{I}, p) - \ln L(\underset{\sim}{I}, \hat{\pi}_{01}, \hat{\pi}_{11})). \tag{5}$$

The test statistics (4) and (5) are asymptotically distributed as chi-square with one degree of freedom. We use the notation M_{IND} for the Markov independence test. If in Eq. (3) we set Z_t to be the most recent de-meaned hit we have $E[(I_{t+1} - p)(I_t - p)] = 0$, the only condition explored by the Markov tests. If we set $Z_t = (I_{t-k} - p)$ for any $k \geq 0$, we have $E[(I_{t+1} - p)(I_{t-k} - p)] = 0$. Based on this broader condition Berkowitz et al. (2009) suggested the Ljung–Box statistic, for a joint test of whether the first m autocorrelations of $\{I_t\}$ are zero. The testing procedure is based on an asymptotic chi-square distribution with m degrees of freedom.

Considering other data in the information set such as past returns, under CC we have $E[(I_{t+1} - p)g(I_t, I_{t-1}, \ldots, R_t, R_{t-1}, \ldots)] = 0$ for any non-anticipating function $g(.)$. In the same line as Engle and Manganelli (2004) and Berkowitz et al. (2009) consider the autoregression

$$I_t = \alpha + \sum_{k=1}^{n} \beta_{1k} I_{t-k} + \sum_{k=1}^{n} \beta_{2k} g(I_{t-k}, I_{t-k-1}, \ldots, R_{t-k}, R_{t-k-1}) + \varepsilon_t, \tag{6}$$

with $n = 1$ and $g(I_{t-k}, I_{t-k-1}, \ldots, R_{t-k}, R_{t-k-1}) = \text{VaR}_{t-k+1|t-k}(p)$. These authors propose the logit model and test the CC hypothesis with a likelihood ratio test considering for the null $P(I_t = 1) = 1/(1 + e^{-\alpha}) = p$ and the coefficients β_{11} and β_{21} equal to zero. For the IND hypothesis the null is $\beta_{11} = \beta_{21} = 0$ and in this case the asymptotic distribution is chi-square with 2 degrees of freedom. We refer to these tests as the CAViaR tests of Engle and Manganelli (CAViaR).

A *duration-based* approach emerged in the literature. There are related works on testing duration dependence (e.g., Kiefer, 1988). As far as we know, the first authors that proposed this approach for interval forecast evaluation were Danielsson and