International Journal of Forecasting 28 (2012) 343–352

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

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

Efficient evaluation of multidimensional time-varying density forecasts, with applications to risk management

Arnold Polanski[a], Evarist Stoja[b,*]

[a] Queen's University Management School, 25 University Square, Queen's University Belfast, Belfast, BT7 1NN, UK
[b] School of Economics, Finance and Management, University of Bristol, 8 Woodland Road, Bristol, BS8 1TN, UK

ARTICLE INFO

Keywords:
Multivariate density forecast evaluation
Probability integral transformation
Multidimensional value at risk
Monte Carlo simulations

ABSTRACT

We propose two simple evaluation methods for time-varying density forecasts of continuous higher-dimensional random variables. Both methods are based on the probability integral transformation for unidimensional forecasts. The first method tests multinormal densities and relies on the rotation of the coordinate system. The advantages of the second method are not only its applicability to arbitrary continuous distributions, but also the evaluation of the forecast accuracy in specific regions of its domain, as defined by the user's interest. We show that the latter property is particularly useful for evaluating a multidimensional generalization of the Value at Risk. In both simulations and an empirical study, we examine the performances of the two tests.

1. Introduction

The evaluation of the accuracy of forecasts occupies a prominent place in the finance and economics literature. However, most of this body of literature (e.g., Diebold & Lopez, 1996) focuses on the evaluation of point forecasts, rather than interval or density forecasts. The driving force for this over-focus is that, until recently, point forecasts appeared to serve the requirements of the forecast users well. However, there is increasing evidence that a more comprehensive approach is needed. One example is the Value at Risk (VaR), which is defined as the maximum loss on a portfolio over a certain period of time that can be expected with a certain probability. When the returns are normally distributed, the VaR of a portfolio is a simple function of the variance of the portfolio.[1] In this case, normality justifies the use of point forecasts for the variance. However, when the return distribution is non-normal, as is now the general consensus, the VaR of a portfolio is determined not just by the portfolio variance, but by the entire conditional distribution of returns. More generally, decision making under uncertainty with an asymmetric loss function and non-Gaussian variables involves density forecasts (see Guidolin & Timmermann, 2005; Tay & Wallis, 2000, for a survey and discussion of density forecasting applications in finance and economics).

The increasing importance of forecasts of the entire (conditional) density naturally raises the issue of forecast evaluation. Although the relevant literature is developing at a rapid rate, it is still in its infancy. This is somewhat surprising, considering that the crucial tools which are employed date back a few decades. Indeed, a key contribution by Diebold, Gunther, and Tay (1998) relies on the probability integral transformation (PIT) result from the work of Rosenblatt (1952). Diebold et al. point out that the correct density is weakly superior to all forecasts. This suggests that the forecasts should be evaluated in terms of their correctness, as this is independent of the loss function. To this end, Diebold et al. (1998) employ the PITs of the univariate density forecasts, which, if accurate, are

---

* Corresponding author.
E-mail addresses: a.polanski@qub.ac.uk (A. Polanski), e.stoja@bristol.ac.uk (E. Stoja).

[1] When the mean return on an asset is assumed to be zero, as is commonly the case in practice when dealing with short-horizon returns, the VaR of a portfolio is simply a constant multiple of the square root of the variance of the portfolio.

0169-2070/$ – see front matter © 2011 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2010.10.007

i.i.d. standard uniform. They measure the forecast accuracy by the distance between the empirical distribution of the PITs and the 45° line, and argue that the visual inspection of this distance may provide valuable insights into the deficiencies of the model and possible ways of improving it. Obviously, standard goodness-of-fit tests can be applied to the PITs directly (see Noceti, Smith, & Hodges, 2003, for a comparison of the existing goodness-of-fit tests). Additional tests have been proposed by Anderson, Hall, and Titterington (1994), Bai (2003), Berkowitz (2001), Granger and Pesaran (1999), Hong (2001), Hong and Li (2003), Hong, Li, and Zhao (2007), Li (1996) and Li and Tkacz (2001).

The existing evaluation methods of multidimensional density forecasts (MDF) rely on the advances made in the univariate case. Diebold, Hahn, and Tay (1999) extend the PIT idea to multivariate forecasts by factoring the multivariate probability density function (PDF) into its conditionals and computing the PIT for each conditional. As in the univariate case, the PIT of these forecasts is *i.i.d.* uniform if the sequence of forecasts is correct. Clements and Smith (2000, 2002) extend Diebold et al.'s (1999) idea and propose two tests based on the product and ratio of the conditionals and marginals. While the latter tests perform well when there is correlation misspecification, they perform worse than the original test by Diebold et al. (1999) when such misspecification is absent. However, both approaches rely on the factorization of each period's forecasts into their conditionals, which may not be practical for some applications (e.g., for numerical approximations of density forecasts). Moreover, these approaches assume that the forecasting model is correct under the null hypothesis. This assumption has important implications for the evaluation tools employed, particularly in relation to parameter estimation uncertainty. Recognising this issue, another strand of the MDF evaluation literature has recently gained momentum. This body of literature allows for dynamic misspecification and/or parameter estimation uncertainty, and includes important contributions by Bai and Chen (2008), Chen and Hong (2010) and Corradi and Swanson (2006b) *inter alia*. Corradi and Swanson (2006b) construct Kolmogorov-type conditional distribution tests in the presence of both dynamic misspecification and parameter estimation uncertainty. While their testing framework is flexible, it suffers from the fact that the limiting distribution is not free of nuisance parameters, and bootstrapping is needed to obtain valid critical values. Bai and Chen (2008) and Chen and Hong (2010) propose MDF evaluation tests that, under certain conditions, deal with the parameter estimation uncertainty. For example, Bai and Chen (2008) use the K-transformation of Khmaladze (1981) to remove the effect of parameter estimation, so that a distribution-free test can be constructed. However, they still rely on the factorization of the joint density, and only apply this procedure to the multivariate normal and multivariate-t distributions, in which case they obtain closed-form results. We discuss these issues in more detail in Section 3, and refer the interested reader to Corradi and Swanson (2006a) and Mecklin and Mundfrom (2004) for further insights into density forecast evaluation.

Broadly speaking, this paper belongs to the body of literature established by Clements and Smith (2000, 2002) and Diebold et al. (1998, 1999), which does not account for parameter estimation uncertainty. This approach also dominates the parametric-VaR area of the risk management literature, in which we are mainly interested (see for example Gourieroux & Jasiak, 2010, chap.10). Thus, in the simulations and empirical examples, we ignore the parameter estimation uncertainty and potential dynamic misspecification, but acknowledge that these could be important. Finally, we stress that forecasts may vary over time, making parameter estimation and forecast evaluation based on the laws of large numbers unfeasible.

This paper makes two important contributions. Firstly, it proposes two new tests for evaluating multidimensional, time-varying density forecasts, which — like their counterparts — may suffer from parameter estimation error and dynamic misspecification, although they are simpler and more flexible. Secondly, to the best of our knowledge, it is the first to formalise and propose a theoretical framework for testing the accuracy of a multidimensional VaR (MVaR). This framework is particularly important for examining multiple sources of tail risk.

The outline of the remainder of this paper is as follows. In Section 2, we discuss an evaluation procedure for multinormal density forecasts. Section 3 presents a test for arbitrary continuous densities, while Section 4 discusses the results of Monte Carlo simulations and an empirical application for the newly proposed tests. Finally, Section 5 concludes.

2. Evaluation procedure for multinormal density forecasts

Rosenblatt (1952) showed that, for the cumulative distribution function (CDF) $\widehat{F}_t$ (PDF $\widehat{f}_t$) which correctly forecasts the true data generating process (DGP) F_t of the observation x_t, i.e., for which $\widehat{F}_t(x_t) = F_t(x_t)$, the PIT

$$z_t = \int_{-\infty}^{x_t} \widehat{f}_t(u)\mathrm{d}u = \widehat{F}_t(x_t)$$

is *i.i.d.*, according to $U[0, 1]$. Therefore, the adequacy of the forecasts can easily be evaluated by examining the z_t series for violations of independence and uniformity.

The PIT idea was extended to the multivariate case by Diebold et al. (1999). Their test procedure (D-test hereafter) factors each period's MDF into the product of the conditionals

$$\begin{aligned}\widehat{f}_{t-1}&(x_{1,t}, x_{2,t}, \ldots, x_{N,t})\\ &= \widehat{f}_{t-1}(x_{N,t} \mid x_{1,t}, x_{2,t}, \ldots, x_{N-1,t})\\ &\quad \ldots \widehat{f}_{t-1}(x_{2,t} \mid x_{1,t}) \cdot \widehat{f}_{t-1}(x_{1,t}),\end{aligned}$$

and obtains the PIT for each conditional distribution, producing a set of Nz_t-series, which are *i.i.d.* $U[0, 1]$, both individually and as a whole whenever the MDF is correct.[2] Rejecting the null of *i.i.d.* $U[0, 1]$ for any of the z_t

---

[2] There are $N!$ different ways of factoring the MDF $\widehat{f}_t(x_{1,t-1}, \ldots, x_{N,t-1})$, giving us a wealth of z series with which to evaluate the forecast.