



Calibration tests for multivariate Gaussian forecasts

Wei Wei^a, Fadoua Balabdaoui^{b,c}, Leonhard Held^{a,*}

^a Department of Biostatistics at the Epidemiology, Biostatistics and Prevention Institute, Universität Zürich, Switzerland

^b Université Paris-Dauphine, PSL Research University, CNRS, CEREMADE, 75016 Paris, France

^c Seminar für Statistik, ETH Zürich, 8092 Zürich, Switzerland

ARTICLE INFO

Article history:

Received 25 June 2015

Available online 14 November 2016

AMS subject classifications:

62H15

62E20

62P10

62J05

Keywords:

Calibration test

Multivariate forecast

Dawid–Sebastiani score

Logarithmic score

Gaussian forecast

Uniform correlation matrix

ABSTRACT

Forecasts by nature should take the form of probabilistic distributions. Calibration, the statistical consistency of forecast distributions and observations, is a central property of good probabilistic forecasts. Calibration of univariate forecasts has been widely discussed, and significance tests are commonly used to investigate whether a prediction model is miscalibrated. However, calibration tests for multivariate forecasts are rare. In this paper, we propose calibration tests for multivariate Gaussian forecasts based on two types of the Dawid–Sebastiani score (DSS): the multivariate DSS (mDSS) and the individual DSS (iDSS). Analytic results and simulation studies show that the tests have sufficient power to detect miscalibrated forecasts with incorrect mean or incorrect variance. But for forecasts with incorrect correlation coefficients, only the tests based on mDSS are sensitive to miscalibration. As an illustration, we apply the methodology to weekly data on Norovirus disease incidence among males and females in Germany, in 2011–2014. The results further show that tests for multivariate forecasts are useful tools and superior to univariate calibration tests for correlated multivariate forecasts.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

One main task for statistical analysis is to predict the future. In the past two decades, probabilistic predictions have become routine in applied forecasting. Probability forecasts were first commonly used for binary endpoints, and later extended to more general types of variables. They are widely discussed and applied in many areas: in weather forecasting; in economics and finance risk management; in clinical, chronic and infectious disease epidemiology; in health care management; in atmospheric science and many other areas. They usually take the form of ensemble forecasts, interval forecasts or density forecasts. Here we will focus on the density forecast which provides the most information. In the case of a binary event, the density forecast is the probability that the event will occur; in the case of continuous variable, a probabilistic forecast is the predictive density of the outcome of interest.

How to evaluate the performance of probabilistic forecasts is an essential question in forecast research. Gneiting et al. [19] contend that the goal of probabilistic forecasting is to *maximize the sharpness of the predictive distributions subject to calibration*. In this context, calibration refers to the statistical consistency between the probabilistic forecasts and the actual observations. Sharpness refers to the concentration of the predictive distributions. If the true data generator follows the predictive distribution, we say the forecast is ideal.

* Corresponding author.

E-mail addresses: fadoua@ceremade.dauphine.fr (F. Balabdaoui), leonhard.held@uzh.ch (L. Held).

Interest in density forecasting has spurred the development of methods for their evaluation [1,14]. Much of the literature in forecast evaluation focuses on the forecasting of univariate quantities or events. A comprehensive overview of probabilistic forecasting is given by Gneiting and Katzfuss [20], including discussion of diagnostic checks and methods for the evaluation of probabilistic forecasts. The evaluation of forecasts heavily depends on the distribution of the forecasts. The methodology began with binary outcomes (whether it will rain tomorrow), later extended to categorical events, count data and continuous quantities [9,26]. Many diagnostic tools have been developed for model evaluation and model selection. For univariate continuous forecasts, Dawid [12] and Diebold et al. [14] propose the use of the probability integral transform (PIT). For ideal forecasts, the PIT values are uniformly distributed. Therefore, a PIT histogram is typically used as a diagnostic tool. Gneiting et al. [19] propose proper scoring rules which can evaluate calibration and sharpness simultaneously. Three proper scores are commonly used: The Dawid–Sebastiani score (DSS) [13], the logarithmic score (LS) [10,22] and the ranked probabilistic score (RPS) [6,19]. Calibration tests have been developed to investigate whether the forecasts are miscalibrated. Held et al. [23] develop two types of calibration tests based on proper scoring rules. Alternatively, Mason et al. [30] suggest the use of the conditional exceedance probability (CEP) in a logistic regression framework to assess calibration of continuous probabilistic forecasts.

In recent years, the evaluation of multivariate forecasts came into focus with the proliferation of multivariate probabilistic forecasting. Tools have been developed for multivariate ensemble forecasts, for example, the multivariate rank histogram [21] and the band depth rank histogram [39]. However, for multivariate density forecasts, only a limited number of methods can be applied. Firstly, many tools for univariate forecasts do not apply to multivariate forecasts. The PIT approach fails in that it is not uniform even when the observation is drawn from the predictive distribution [21]. Although some alternative transforms are proposed to retain uniformity, they do not work perfectly. For example, a step-wise procedure for PIT is proposed by Diebold et al. [15], in which the univariate PIT values are computed sequentially based on the conditional cumulative distribution function (CDF). More specifically, the univariate PIT is first applied to the first component, then to the conditional CDF of the second given the first, and so on. However, this approach depends on the order of the components. Methods based on CEP fail as well because the quantile of a multivariate forecast is not unique. In addition, most existing methods for the evaluation of multivariate forecasts encounter the issue of low efficacy for high dimensionality, which is usually referred to as the curse of dimensionality [3]. Gneiting et al. [21] give an overview on methods for diagnostic checking and recommend proper scoring rules to evaluate multivariate forecasts. Among all the proper scoring rules available, we have decided to use DSS, which is equivalent to LS under normality of the forecast. DSS is easy to compute, straightforward to interpret, and reported to be sensitive to mis-specified correlations [36]. In addition, DSS is a standardized score which avoids problems arising from components with incommensurable or incomparable magnitude, for example if one component has values between -1 and 1 , while another component is in the range -100 to 100 . Finally, it is possible to derive the first two moments of the DSS for ideal Gaussian forecasts, as we will show in Section 2.

In this paper, we develop calibration tests for multivariate Gaussian density forecasts based on the DSS. The structure of the paper is as follows. In Section 2 we introduce two types of the DSS and their properties for ideal Gaussian forecasts. In Section 3 we develop calibration tests based on the DSS to check calibration of multivariate predictions. We evaluate the power of the proposed tests analytically and via simulations in Section 4, where we pay particular attention to uniform correlation matrices [16]. As a practical application, in Section 5 we apply the tests to evaluate predictive models for the weekly number of reported *Norovirus* infections among males and females in Germany. We end with a discussion in Section 6. In this paper, boldface symbols represent multivariate quantities or distributions, whereas standard typeface symbols are univariate quantities.

2. David–Sebastiani scores for multivariate Gaussian forecasts

A scoring rule assigns a numerical score based on the predictive distribution and the realization. It can be viewed as a penalty of the statistical difference between an observation and a prediction and is usually negatively oriented, i.e., the smaller, the better. A proper scoring rule ensures that quoting the true predictive distribution as forecast distribution is an optimal strategy in expectation. In this paper, we consider the David–Sebastiani and the logarithmic score, which are equivalent for a normal distribution. For a univariate observation x and a normal prediction $P = \mathcal{N}(\mu_P, \sigma_P^2)$, the logarithmic score is defined as

$$\text{LS}(x, P) = \frac{1}{2} \ln \sigma_P^2 + \frac{(x - \mu_P)^2}{2\sigma_P^2}.$$

The Dawid–Sebastiani score $\text{DSS}(x, P) = 2 \text{LS}(x, P)$ is the same up to the multiplicative constant of 2 and can also be used for other (non-normal) multivariate predictions.

For an m -dimensional normal prediction, there are two types of DSS: the multivariate score (mDSS) and the individual score based on each component (iDSS). Denoting the multivariate prediction as $\mathbf{P} = \mathcal{N}_m(\boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P)$, the mDSS is computed as

$$\text{mDSS}(\mathbf{x}, \mathbf{P}) = \ln |\boldsymbol{\Sigma}_P| + (\mathbf{x} - \boldsymbol{\mu}_P)^\top \boldsymbol{\Sigma}_P^{-1} (\mathbf{x} - \boldsymbol{\mu}_P),$$

where $|\boldsymbol{\Sigma}_P|$ denotes the determinant of $\boldsymbol{\Sigma}_P$. The component $\ln |\boldsymbol{\Sigma}_P|$ measures the sharpness and the quadratic form $Q(\mathbf{x}, \mathbf{P}) = (\mathbf{x} - \boldsymbol{\mu}_P)^\top \boldsymbol{\Sigma}_P^{-1} (\mathbf{x} - \boldsymbol{\mu}_P)$ is the standardized difference between the observation $\mathbf{x}$ and the predicted mean $\boldsymbol{\mu}_P$. The multivariate

Download English Version:

<https://daneshyari.com/en/article/5129431>

Download Persian Version:

<https://daneshyari.com/article/5129431>

[Daneshyari.com](https://daneshyari.com)