Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 27 (2011) 197–207
international journal of forecasting
www.elsevier.com/locate/ijforecast

Quantiles as optimal point forecasts

Tilmann Gneiting

Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany

Abstract

Loss functions play a central role in the theory and practice of forecasting. If the loss function is quadratic, the mean of the predictive distribution is the unique optimal point predictor. If the loss is symmetric piecewise linear, any median is an optimal point forecast. Quantiles arise as optimal point forecasts under a general class of economically relevant loss functions, which nests the asymmetric piecewise linear loss, and which we refer to as generalized piecewise linear (GPL). The level of the quantile depends on a generic asymmetry parameter which reflects the possibly distinct costs of underprediction and overprediction. Conversely, a loss function for which quantiles are optimal point forecasts is necessarily GPL. We review characterizations of this type in the work of Thomson, Saerens and Komunjer, and relate to proper scoring rules, incentive-compatible compensation schemes and quantile regression. In the empirical part of the paper, the relevance of decision theoretic guidance in the transition from a predictive distribution to a point forecast is illustrated using the Bank of England's density forecasts of United Kingdom inflation rates, and probabilistic predictions of wind energy resources in the Pacific Northwest.

Keywords: Decision making; Density forecasts; Incentive-compatible compensation scheme; Loss function; Piecewise linear; Proper scoring rule; Quantile

1. Introduction

In many areas of human activity, a major desire is to make forecasts about an uncertain future. Consequently, forecasts ought to be probabilistic in nature, taking the form of probability distributions over future quantities or events. However, many practical situations require single-valued point forecasts, or point predictions, for reasons of decision making, market mechanisms, reporting requirements, communications, or tradition, among others.

For concreteness, suppose that we are to predict a real-valued future quantity Y, with verifying realization y. We represent the predictive distribution for the random variable Y in the form of a predictive cumulative distribution function, F. The predictive distribution is conditional on the forecaster's information set, consisting of data, together with expertise, theories and assumptions (Granger & Newbold, 1986, p. 120). However, the conditioning is immaterial to the transition from the predictive distribution to the point forecast, and thus will not be acknowledged in our notation. Our task is then to issue a point forecast, x,

E-mail address: t.gneiting@uni-heidelberg.de.

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2009.12.015

which we view as a decision or action whose monetary or societal consequences can be expressed by means of a *loss function*, L. The loss $L(x, y)$ constitutes the cost or penalty if we predict x and y materializes. The *optimal point forecast* or *Bayes rule* (Ferguson, 1967, p. 31; Elliott & Timmermann, 2008, p. 12) is any argument

$$\hat{x} = \arg\min_x \mathbb{E}_F \, L(x, Y) \tag{1}$$

that minimizes the expected loss. If the loss function is *quadratic*, $L(x, y) = (x - y)^2$, the *mean* of the predictive distribution is the unique optimal point predictor.[1] If the loss is *symmetric piecewise linear*, $L(x, y) = |x - y|$, any *median* of the predictive distribution is an optimal point forecast. The quadratic and the symmetric piecewise linear are the most widely studied loss functions; they are both *symmetric*, in that $L(x, y) = L(y, x)$, and of the *prediction error form*, in that the loss depends only on the prediction error $x - y$.

There is compelling evidence in the literature that empirically relevant loss functions tend to be asymmetric and may not be of the prediction error form (Britney & Winkler, 1974; Christoffersen & Diebold, 1996, 1997; Elliott & Timmermann, 2004; Elliott, Komunjer, & Timmermann, 2005, 2008; Granger, 1969; Patton & Timmermann, 2007a,b; Varian, 1974; Zellner, 1986). Under more general and more realistic loss functions, neither the mean nor the median remain optimal as point predictors. For instance, if the loss function is *asymmetric piecewise linear*,

$$PL_\alpha(x, y) = \begin{cases} \alpha \, |x - y| & \text{if } x \le y, \\ (1 - \alpha) \, |x - y| & \text{if } x \ge y, \end{cases} \tag{2}$$

of order $\alpha \in (0, 1)$, any α-quantile of the predictive distribution is an optimal point forecast (Raiffa & Schlaifer, 1961, p. 196).[2] This well-known result lies at the heart of quantile regression (Koenker & Bassett, 1978).

It is also well-known that quantiles are equivariant to nondecreasing transformations (Koenker, 2005, p. 39). Thus, the optimality of the α-quantile as a point forecast continues to hold under the class of *generalized piecewise linear* (GPL) loss functions of order $\alpha \in (0, 1)$, which are of the form

$$L(x, y) = \begin{cases} \alpha \, (g(y) - g(x)) & \text{if } x \le y, \\ (1 - \alpha) \, (g(x) - g(y)) & \text{if } x \ge y, \end{cases} \tag{3}$$

where g is a nondecreasing function. Furthermore, if a loss function is such that any α-quantile of the predictive distribution is an optimal point forecast, then it is necessarily GPL, subject to minor regularity conditions. Results of this type have been provided by Komunjer (2005), Saerens (2000) and Thomson (1979). We review them in the expository Section 2, where we also discuss proper scoring rules and incentive-compatible compensation schemes. In Section 3, we demonstrate the empirical relevance of decision theoretic guidance in the transition from the predictive distribution to the point forecast in a simulation experiment and case studies, based on the Bank of England's density forecasts of United Kingdom inflation rates, and probabilistic predictions of wind resources in the US Pacific Northwest. These experiments complement the empirical study of Ulu (2007), in that they apply to a richer class of economically relevant loss functions, rather then just asymmetric piecewise linear loss. The paper ends with a discussion in Section 4.

Table 1
Assumptions on a loss function L on a DO domain, $D = I \times I$, where $x \in I$ denotes the point forecast and $y \in I$ the verifying realization.

(A0)	$L(x, y) \ge 0$ with equality if $x = y$
(A1)	$L(x, y)$ is continuous
(A2)	$L(x, y)$ is twice continuously differentiable whenever $x \ne y$

2. Quantiles as optimal point forecasts

The basic notion of a *decision-observation* (DO) *domain* emphasizes our perception of a point forecast as a decision or action.

Definition 2.1. A subset D of the Euclidean plane, $\mathbb{R}^2$, is a DO domain if it is the Cartesian product, $D = I \times I$, of an interval, I, which might be finite or infinite, and open, half-open or closed, with itself.

We consider loss functions on general DO domains. The cases of primary interest are the real line, $I = \mathbb{R}$, and the nonnegative or positive half-axis, $I = [0, \infty)$ or $I = (0, \infty)$, which correspond to nonnegative

[1] This statement about uniqueness assumes that the predictive distribution has a finite second moment.

[2] Recall that an α-quantile ($0 < \alpha < 1$) of the cumulative distribution function F is any number x for which $\lim_{y \uparrow x} F(y) \le \alpha \le F(x)$.