European Journal of Operational Research 234 (2014) 140–154

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk

R.T. Rockafellar [a], J.O. Royset [b,*], S.I. Miranda [c]

[a] University of Washington, Seattle, WA, United States
[b] Naval Postgraduate School, Monterey, CA, United States
[c] Portuguese Navy, Lisbon, Portugal

ARTICLE INFO

Article history:
Received 13 March 2013
Accepted 17 October 2013
Available online 30 October 2013

Keywords:
Generalized regression
Superquantiles
Conditional value-at-risk
Uncertainty quantification
Buffered failure probability
Stochastic programming

ABSTRACT

The paper presents a generalized regression technique centered on a superquantile (also called conditional value-at-risk) that is consistent with that coherent measure of risk and yields more conservatively fitted curves than classical least-squares and quantile regression. In contrast to other generalized regression techniques that approximate conditional superquantiles by various combinations of conditional quantiles, we directly and in perfect analog to classical regression obtain superquantile regression functions as optimal solutions of certain error minimization problems. We show the existence and possible uniqueness of regression functions, discuss the stability of regression functions under perturbations and approximation of the underlying data, and propose an extension of the coefficient of determination R-squared for assessing the goodness of fit. The paper presents two numerical methods for solving the error minimization problems and illustrates the methodology in several numerical examples in the areas of uncertainty quantification, reliability engineering, and financial risk management.

Published by Elsevier B.V.

1. Introduction

Analysts and decision makers are often concerned with a random variable describing possible ‘cost,’ ‘loss,’ or ‘damage.’ The interest may be focused on a single ‘system’ or could involve study and comparison across a multitude of systems and designs. In either case, it may be beneficial to attempt to approximate such a *loss random variable* Y in terms of an n-dimensional *explanatory random vector* X that is more accessible in some sense. This situation naturally leads to least-squares regression and related models that estimate conditional expectations. While such models are adequate in many situations, they fall short in contexts where a decision maker is risk averse, i.e., is more concerned about upper-tail realizations of Y than average loss, and views errors asymmetrically with underestimating losses being more detrimental than overestimating. We focus on such contexts and therefore maintain an orientation of Y that implies that high realizations are unfortunate and low realizations are favorable. Of course, a parallel development with an opposite orientation of the random variable Y, focused on profits and gains, and concerns about overestimating instead of underestimating is also possible but not pursued here.

Quantile regression (see Gilchrist, 2008; Koenker, 2005 and references therein) accommodates risk-averseness and an asymmetric view of errors by estimating conditional quantiles at a certain probability level such as those in the tail of the conditional distribution of Y. While suitable in some contexts, quantile regression only deals with the signs of the errors and therefore is overly ‘robust’ in the sense that large portions of a data set can change dramatically without impacting the best-fitting regression function. A quantile corresponds to ‘value-at-risk’ (VaR) in financial terminology and relates to ‘failure probability’ in engineering terms. Quantile regression informs the decision maker about these quantities conditional on values of the explanatory random vector X. However, a quantile is not a *coherent* measure of risk in the sense of Artzner, Delbaen, Eber, and Heath (1999) (see also Delbaen, 2002); it fails to be subadditive. Consequently, a quantile of the sum of two random variables may exceed the sum of the quantiles of each random variable at the same probability level, which runs counter to our understanding of what ‘risk’ should express. Moreover, quantiles cause computational challenges when incorporated into decision optimization problems as objective function, failure probability constraint, or chance constraint. The use of quantiles and the closely related failure probabilities is therefore problematic in risk-averse decision making; see (Artzner et al., 1999; Krokhmal, Zabarankin, & Uryasev, 2011; Rockafellar & Royset, 2010; Rockafellar & Uryasev, 2000, 2013) for a detailed discussion.

A *superquantile* of a random variable, also called conditional value-at-risk, average value-at-risk, and expected shortfall,[1] is an ‘average’ of certain quantiles as described further below. It is a

* Corresponding author. Tel.: +1 831 656 2578; fax: +1 831 656 2595.
E-mail address: joroyset@nps.edu (J.O. Royset).

[1] We prefer the application-neutral name ‘superquantile’ when deriving methods applicable broadly.

0377-2217/$ - see front matter Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.ejor.2013.10.046

coherent measure of risk well suited for risk-averse decision making and optimization; see (Wang & Uryasev, 2007) for its application in financial engineering, (Kalinchenko, Veremyev, Boginski, Jeffcoat, & Uryasev, 2011) for military applications, and (Rockafellar & Royset, 2010) for use in reliability engineering. While this risk measure has reached prominence in risk-averse optimization, there has been much less work on regression techniques that are consistent in some sense with it. In this paper, we derive such a *superquantile regression* methodology, study its properties, and propose means to assess the goodness-of-fit. The importance of such a regression methodology becomes apparent by considering the following two situations.

Suppose that a loss is given by a random variable Y, but our primary concern is with the conditional loss given that an explanatory random vector X takes on specific values. We aim to select these values judiciously in an effort to minimize the conditional loss. We denote by $Y(x)$ the conditional random variable Y given that $X = x \in \mathbb{R}^n$. Of course, 'minimizing' $Y(x)$ is not well-defined and a standard approach is to minimize a risk measure of $Y(x)$; see for example (Krokhmal et al., 2011; Rockafellar & Uryasev, 2013). An attractive choice is to use a superquantile measure of risk, which as mentioned above is coherent and also computationally approachable. While in some contexts a superquantile of $Y(x)$ can be evaluated easily for any $x \in \mathbb{R}^n$, there are numerous situations, especially beyond the financial domain, where only a data base of realizations of $Y(x)$ is available for various x. In the latter situation, there is a need for building an approximating model, based on the data, for the relevant superquantile of $Y(x)$ as a function of x. We refer to this as *superquantile tracking*. In comparison, if the goal were to minimize the expectation of $Y(x)$, then least-squares regression would yield a model that approximates that conditional expectation. Likewise, if the goal were to minimize a quantile of $Y(x)$, quantile regression would provide a model of the conditional quantile. While these models are valuable for analysts and decision makers focused on the expectation and quantile risk measures, they do not provide estimates of conditional superquantiles. In essence, the same need for estimating conditional superquantiles arises in reliability engineering when the goal is to determine a 'design' x with buffered failure probability of $Y(x)$ being no larger than a given probability level, which corresponds to a constraint on a superquantile of $Y(x)$ (Rockafellar & Royset, 2010).

Another situation arises when the explanatory random vector X is beyond our direct control, but the dependence between the loss random variable Y and X makes us hopeful that, for a carefully selected *regression function* $f : \mathbb{R}^n \to \mathbb{R}$, the random variable $f(X)$ may serve as a surrogate for Y. When the distribution of X is known, at least approximately, and f has been determined, then the distribution of $f(X)$ is usually easily accessible. That distribution may then serve as input to further analysis, simulation, and optimization in place of the unknown distribution of Y. Such *surrogate estimation* may arise in numerous contexts. 'Factor models' in financial investment applications (see for example Conner, 1995; Knight & Satchell, 2005), where Y may be the loss associated with a particular asset and X a vector describing a small number of macroeconomic 'factors,' is a result of surrogate estimation. 'Uncertainty quantification' (see for example Eldred, Swiler, & Tang, 2011; Lee & Chen, 2009) considers the output of a system described by a random variable Y, for example measuring damage, and estimates its moments and distribution from observed realizations as well as knowledge about the distribution of the input to the system characterized by a random vector X. A main approach here centers on surrogate estimation with $f(X)$ serving as an estimate of Y. In this situation, an essential question is what criterion should be used for selecting f. Clearly, one would like the *error random variable* $Z_f := Y - f(X)$ to be small in some sense. However, minimizing the mean-squared error of Z_f would not reflect a greater concern about underestimating Y, i.e., underestimating losses, than overestimating. We may want to assess the error of Z_f in a manner that is 'consistent' with our use of a superquantile as risk measure and weigh large levels of underestimation more heavily than smaller levels.

In this paper, we develop a 'generalized' regression technique that addresses the issue of superquantile tracking and surrogate estimation. The technique is an extension of least-squares and quantile regression, which center on expectations and quantiles, respectively, to one that focuses on superquantiles.

The foundation of least-squares and quantile regression is the fact that mean and quantiles minimize the expectation of certain convex random functions. A natural extension to superquantile regression could then possibly involve determining a random function that when minimizing its expectation, we obtain a superquantile. However, such a random function does not exist (Chun, Shapiro, & Uryasev, 2012; Gneiting, 2011), which has lead to studies of indirect approaches to superquantile tracking grounded in quantile regression.

For a random variable with a continuous cumulative distribution function, a superquantile equals a conditional expectation of the random variable given realizations no lower than the corresponding quantile. Utilizing this fact, studies have developed kernel-based estimators for the conditional probability density functions, which are then integrated and inverted to obtain estimators of conditional quantiles. An estimator of the conditional superquantile is then finally constructed by integrating the density estimator over the interval above the quantile (Cai & Wang, 2008; Scaillet, 2005) or forming a sample average (Kato, 2012). These studies also include asymptotic analysis of the resulting estimators under a series of assumptions, including that the data originates from certain time series.

A superquantile of a random variable is defined in terms of an integral of corresponding quantiles with respect to the probability level. Since the integral is approximated by a weighted sum of quantiles across different probability levels, an estimator of a conditional superquantile emerges as the sum of conditional quantiles obtained by quantile regression; see (Leorato, Peracchi, & Tanase, 2012; Peracchi & Tanase, 2008), which also show asymptotic results under a set of assumptions including the continuous differentiability of the cumulative distribution function of the conditional random variables. Similarly, (Chun et al., 2012) utilizes the integral expression for a superquantile, but observes that a weighted sum of quantiles is an optimal solution of a certain minimization problem; see (Rockafellar & Uryasev, 2013). Analogously to the situation in least-squares and quantile regression, an optimization problem therefore yields an estimator of a conditional superquantile. Though, in contrast to the case of least-squares and quantile regression, the estimator is 'biased' due to the error induced by replacing an integral by a finite sum. Under a linear model assumption, (Chun et al., 2012) also constructs a conditional superquantile estimator using an appropriately shifted least-squares regression curve based on quantile estimates of residuals. In both cases, asymptotic results are obtained for a homoscedastic linear regression model. Under the same model, (Trindade, Uryasev, Shapiro, & Zrazhevsky, 2007) studies 'constrained' regression, where the error random variable $Z_f = Y - f(X)$ is minimized in some sense, for example in terms of least square or absolute deviation, subject to a constraint that limits a superquantile of Z_f. While this approach does not lead to superquantile regression in the sense we derive below, it highlights the need for alternative techniques for regression that incorporate superquantiles in some manner.

The need for moving beyond classical regression centered on conditional expectations is therefore now well recognized and has driven even further research towards estimating conditional distribution function, i.e., $\text{Prob}(Y(x) \leqslant y)$ for all $y \in \mathbb{R}$, using nonparametric kernel estimators (see for example Hall & Muller,

Download English Version:

https://daneshyari.com/en/article/479872

Download Persian Version:

https://daneshyari.com/article/479872

Daneshyari.com