The Contribution of Edgeworth Approximation

in a Real Study Case

Edlira Donefski
Faculty of Natural and Human Sciences
Department of Mathematics and Physics
“Fan S. Noli” University, Korça, Albania

Tina Donefski
Faculty of Economy
Department of Finance-Accounting
“Fan S. Noli” University, Korça, Albania

Eljona Milo
Faculty of Natural and Human Sciences
Department of Mathematics and Physics
“Fan S. Noli” University, Korça, Albania

This article is distributed under the Creative Commons by-nc-nd Attribution License.
Copyright © 2021 Hikari Ltd.

Abstract

Edgeworth approximation has a considered contribution in the reduction of the sum of standard deviation of the independent variables’ coefficients in a Quantile Regression Model. Quantile regression is an extension of linear regression that estimates the conditional median or other quantiles and is used when the conditions of linear regression are not met.

In our paper, we have generated a quantile regression model for the relation of the customers’ demand and the prices of the products in a real study case during the past two years. This model is used instead of Least Square model of Linear Regression, because the Linear Regression of the Demand and Prices was not free of autocorrelation and heteroscedasticity, and our aim is to analyze in more details the relation between the variables taken into study: the demand and the prices and how to minimize the standard errors of the independent variable involved in this study, the price.
In our work we have applied the Bootstrap version, Edgeworth Approximation for IID cases and the Edgeworth approximation for Bootstrap Quantile Regression Model. All these applications are concretized with graphics that makes the output of identifying the best approximating model of this study more visualized.

Keywords: Quantile, Edgeworth approximation, IID, Bootstrap

Introduction

At the early ages of the history, after the proposal of Bošković and Edgeworth for the Quantile Regression Model, served as a significant prelude in Koenker’s contribution in this model, [13],[14],[15]. This model is widely used to study in details the relations between the variables taken for a study, when some of the conditions of the Linear Regression are not met, as the linearity and not being free of heteroscedasticity and autocorrelation. Edgeworth expansions for the studentized bootstrap quantile estimate [10], has an important contribution in this study. The method used for avoiding the cumbersome estimation is to employ bootstrapping techniques for the reduction of standard error. A brief overview of the various bootstrap methods, see [6], [16] and [8].

Our work is focused in the creation of a quantile regression for real data as the prices and the customers’ demand during the past two years through the Eviews10 software package. Although the demand is depended on various variables and the only influence of the prices is not strongly significant, this makes a data that is interesting to be study in the context of increasing the statistical importance, through reducing the standard errors of the independent variable and to bring some important conclusions. In our paper, we have applied the Bootstrap version, Edgeworth Approximation for IID cases and the Edgeworth approximation for Bootstrap Quantile Regression Model. All these applications are concretized with graphics that makes the output of identifying the best approximating model of this study more visualized.

Definition of the mean squared error estimate

Let $X = \{X_1, \ldots, X_n\}$ denote a random sample of size $n$ drawn from a distribution with distribution function $F$, and write

$$\hat{F}(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x)$$

for the empirical distribution function of the sample, [10]. The bootstrap estimate of the $p$th quantile of $F$, $\xi_p = \hat{F}^{-1}(p)$, is

$$\hat{\xi}_p = \hat{F}^{-1}(p) = \inf \{x : \hat{F}(x) \geq p\}$$

$$= X_{nr}$$

(1)

where $X_{n1} \leq \ldots \leq X_{nm}$ denote the order statistics of $X$ and $r = [np]$ is the largest integer not greater than $np$. The mean squared error of $\hat{\xi}_p$ is given by
The contribution of edgeworth approximation in a real study case

\[ \tau^2 = E \left\{ (\hat{\xi}_p - \xi_p)^2 \right\} \]
\[ = \frac{n!}{(r-1)!(n-r)!} \int_{-\infty}^{\infty} (x - \bar{\xi}_p)^2 F(x)^{r-1} \{1 - F(x)\}^{n-r} dF(x) \]
\[ = r \left( \frac{n}{r} \right) \int_0^1 \left\{ \hat{F}^{-1}(u) - \hat{F}^{-1}(p) \right\}^2 u^{r-1}(1-u)^{n-r} du, \]
of which the bootstrap estimate is
\[ \hat{\tau}^2 = r \left( \frac{n}{r} \right) \int_0^1 \left\{ \hat{F}^{-1}(u) - \hat{F}^{-1}(p) \right\}^2 u^{r-1}(1-u)^{n-r} du \]
\[ = \sum_{j=1}^n (X_{nj} - X_{nm})^2 w_j, \quad \text{(2)} \]
where
\[ w_j = r \left( \frac{n}{r} \right) \int_{(j-1)/n}^{j/n} u^{r-1}(1-u)^{n-r} du. \]
The bootstrap variance estimate differs in details, as follows. Bootstrap estimates of the mean and mean square of \( \hat{\xi}_p \) are respectively
\[ r \left( \frac{n}{r} \right) \int_0^1 \hat{F}^{-1}(u)u^{r-1}(1-u)^{n-r} du = \sum_{j=1}^n X_{nj}^2 w_j, \]
\[ r \left( \frac{n}{r} \right) \int_0^1 \hat{F}^{-1}(u)^2 u^{r-1}(1-u)^{n-r} du = \sum_{j=1}^n X_{nj}^2 w_j. \]
Therefore, the bootstrap variance estimate is
\[ \sum_{j=1}^n X_{nj}^2 w_j - (\sum_{j=1}^n X_{nj} w_j)^2. \]

Edgeworth expansions for the studentized bootstrap quantile estimate

Let \( \hat{\xi}_p \) and \( \hat{\tau}^2 \) denote the bootstrap estimates of \( \xi_p \) and \( \tau^2 \), defined in (1) and (2). Since \( \hat{\xi}_p \) is asymptotically Normal \( N(0, \tau^2) \) [4], and since \( \hat{\tau}^2 / \tau^2 \to 1 \) in probability, \( (\hat{\xi}_p - \xi_p) / \hat{\tau} \) has a limiting Standard Normal distribution. The rate of convergence to this limit may be described by an Edgeworth expansion, although the properties of terms in that expansion are very different from those described in Chapter 2, [10]. Of course, this is principally due to the fact that \( \hat{\tau}^2 / \tau^2 - 1 \) is of order \( n^{-1/4} \) rather than \( n^{-1/2} \). Interestingly, the first term in the expansion is of size \( n^{-1/2} \), not \( n^{-1/4} \), but the polynomial in the coefficient of that term is neither even nor odd; in the cases encountered previously, the polynomial for the \( n^{-1/2} \) term was always even.
To describe the result more explicity, define

\[
q(x) = \frac{1}{2} \left\{ \pi p(1-p) \right\}^{-1/2} x(x^2 + 1 + 2^{3/2}) \\
+ \frac{1}{6} \left\{ p(1-p) \right\}^{-1/2} (1 + p)(x^2 - 1) \\
- [p(1-p)]^{-1/2} p + \left\{ p(1-p) \right\}^{1/2} f(\xi_p) f(\xi_p^{-2}) x^2 \\
- \left\{ p(1-p) \right\}^{-1/2} \left\{ \frac{1}{2} (1-p) + r - np \right\}.
\]

Then

\[
P \left\{ \left( \hat{\xi}_p - \xi_p \right) / \hat{\tau} \leq x \right\} = \Phi(x) + n^{-1/2} q(x) \phi(x) + O(n^{-3/4})
\]

as \( n \to \infty \); see [11]. The polynomial \( q \) is of degree 3 and is neither an odd function nor an even function.

This result may be explained intuitively, as follows. The Edgeworth expansion of a Studentized quantile consists, essentially, of a “main” series of terms decreasing in powers of \( n^{-1/2} \), arising from the numerator in the Studentized ratio, together with a “secondary” series arising from the denominator. On the present occasion the secondary series decreases in powers of \( n^{-1/4} \) since the variance estimate has relative error \( n^{-1/4} \). However, it may be shown that the first term in the secondary series vanishes. For both series, the /th term is even or odd according to whether \( j \) is odd or even, respectively. The term of order \( n^{-1/2} \) in the combined series includes the first, even term of the main series and the second, odd term of the secondary series.

Hall and Sheather [12], discussed Edgeworth expansion for the distribution of a Studentized quantile when the standard deviation estimate is based explicity on a density estimator, of the type introduced by [9] and [2].

**The Model**

While the great majority of regression models are concerned with analyzing the conditional mean of a dependent variable, there is increasing interest in methods of modeling other aspects of the conditional distribution. One increasingly popular approach, *quantile regression*, models the quantiles of the dependent variable given a set of conditioning variables.

As originally proposed by [14], quantile regression provides estimates of the linear relationship between regressors \( X \) and a specified quantile of the dependent variable \( Y \). One important special case of quantile regression is the least absolute deviations (LAD) estimator, which corresponds to fitting the conditional median of the response variable.

Quantile regression permits a more complete description of the conditional distribution than conditional mean analysis alone, allowing us, for example, to describe how the median, or perhaps the 10th or 95th percentile of the response variable, are affected by regressor variables. Moreover, since the quantile regression approach does not require strong distributional assumptions, it offers a robust method of modeling these relationships.
EViews 10
EViews is a statistical and econometric software package. The most current professional version is EViews 10, because it helps in creating the aiming model and studying it through a various commands and scripts.

Independent And Identical
Koenker and Bassett, [14] derive asymptotic normality results for the quantile regression estimator in the \( i.i.d \) setting, showing that under mild regularity conditions,
\[
\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \sim N(0, \tau(1-\tau)s(\tau)^2 J^{-1}) \tag{3}
\]
where
\[
J = \lim_{n \to \infty} (\sum_i X_i X_i' / n) = \lim_{n \to \infty} (X' X / n)
\]
and \( s(\tau) = F^{-1}(\tau) = 1/ f(F^{-1}(\tau)) \)
and \( s(\tau) \), which is termed the sparsity function or the quantile density function, may be interpreted either as the derivative of the quantile function or the inverse of the density function evaluated at the \( \tau \)-quantile [1]. Note that the \( i.i.d \). error assumption implies that \( s(\tau) \) does not depend on \( X \) only in location, hence all conditional quantile planes are parallel.

Given the value of the sparsity at a given quantile, direct estimation of the coefficient covariance matrix is straightforward. In fact, the expression for the asymptotic covariance in (3) is analogous to the ordinary least squares covariance in the \( i.i.d \) setting, with \( \tau(1-\tau)s(\tau)^2 \) standing in for the error variance in the usual formula. The ordinary \( i.i.d \) setting is related to Edgeworth approximation and it is very useful in the minimizing the standard error of the independent coefficient.

Sparsity Estimation
We have seen the importance of the sparsity function in the formula for the asymptotic covariance matrix of the quantile regression estimates for \( i.i.d \). data. Unfortunately, the sparsity is a function of the unknown distribution \( F \), and therefore is a nuisance quantity which must be estimated.
EViews provides three methods for estimating the scalar sparsity \( s(\tau) \): two Siddiqui [9] difference quotient methods ([15]; Bassett and [3]) and one kernel density estimator ([5]; [7]; [6])

Bootstrapping
The direct methods of estimating the asymptotic covariance matrices of the estimates require the estimation of the sparsity nuisance parameter, either at a single point, or conditionally for each observation. One method of avoiding this cumbersome estimation is to employ bootstrapping techniques for the estimation of the covariance matrix.
EViews supports four different bootstrap methods: the residual bootstrap (Residual), the design, or XY-pair, bootstrap (XY-pair), and two variants of the Markov Chain Marginal Bootstrap (MCMB and MBMB-A).

The following discussion provides a brief overview of the various bootstrap
methods. For additional detail, see [6], [16] and [8].

**Residual Bootstrap**
The residual bootstrap, is constructed by resampling (with replacement) separately from the residuals $u_i(\tau)$ and from the $X_i$.

Let $u^*$ be an $m$-vector of resampled residuals, and let $X^*$ be a $m \times p$ matrix of independently resampled $X$. (Note that $m$ need not be equal to the original sample size $n$.) We form the dependent variable using the resampled residuals, resampled data, and estimated coefficients, $Y^* = X^* \hat{\beta}(\tau) + u^*$, and then construct a bootstrap estimate of $\beta(\tau)$ using $Y^*$ and $X^*$.

This procedure is repeated for $M$ bootstrap replications, and the estimator of the asymptotic covariance matrix is formed from:

$V(\hat{\beta}) = n \left( \frac{m}{n} \right) \frac{1}{B} \sum_{j=1}^{B} (\hat{\beta}_j(\tau) - \bar{\beta}(\tau))(\hat{\beta}_j(\tau) - \bar{\beta}(\tau))^\prime$  \hspace{1cm} (4)

where $\bar{\beta}(\tau)$ is the mean of the bootstrap elements. The bootstrap covariance matrix $V(\hat{\beta})$ is simply a (scaled) estimate of the sample variance of the bootstrap estimates of $\beta(\tau)$.

Note that the validity of using separate draws from $u_i(\tau)$ and $X_i$ requires independence of the $u$ and the $X$.

**XY-Pair (Design) Bootstrap**
The XY-pair bootstrap is the most natural form of bootstrap resampling, and is valid in settings where $u$ and $X$ are not independent. For the XY-pair bootstrap, we simply form $B$ randomly drawn (with replacement) subsamples of size $m$ from the original data, then compute estimates of $\beta(\tau)$ using the $(y^*, X^*)$ for each subsample. The asymptotic covariance matrix is then estimated from sample variance of the bootstrap results using (4).

**Quantile Process Testing**
The focus of our analysis thus far has been on the quantile regression model for a single quantile, $\tau$. In a number of cases, we may instead be interested in forming joint hypotheses using coefficients for more than one quantile. We may, for example, be interested in evaluating whether the location-shift model is appropriate by testing for equality of slopes across quantile values. Consideration of more than one quantile regression at the same time comes under the general category of quantile process analysis.

While the EViews equation object is set up to consider only one quantile at a time, specialized tools allow you to perform the most commonly performed quantile process analyses.

Before proceeding to the hypothesis tests of interest, we must first outline the required distributional theory. Define the process coefficient vector:

$\beta = (\beta(\tau_1)', \beta(\tau_2)', ..., \beta(\tau_K))'$

Then

$\sqrt{n}(\hat{\beta} - \beta) \sim N(0, \Omega)$
The contribution of edgeworth approximation in a real study case

where $\Omega$ has blocks of the form:

$$\Omega_{ij} = \left[ \min(\tau_i, \tau_j) - \tau_i \tau_j \right] H^{-1}(\tau_i) J H^{-1}(\tau_j)$$  \hspace{1cm} (5)

In the i.i.d. setting, $\Omega$ simplifies to,

$$\Omega = \Omega_0 \otimes J$$  \hspace{1cm} (6)

where $\Omega_0$ has representative element:

$$\omega_{ij} = \frac{\min(\tau_i, \tau_j) - \tau_i \tau_j}{f(F^{-1}(\tau_i))(f(F^{-1}(\tau_j)))}$$ \hspace{1cm} (7)

Estimation of $\Omega$ may be performed directly using (5), (6) and (7), or using one of the bootstrap variants.

The application of the methods above in a real case

The study case is based on the real data taken from one of the most known businesses in Korçë city, Albania for 129 observations of the prices of different products and the particular customers’ demand for 2019 and 2020. We have used EViews10 for our simulations.

Above is given the model of quantile regression generated for the relation between the prices of the different products in EuroMarket in Korçë city, Albania and the customers’ demand during the 2019. From the table, it is easy to identify that the relation of prices and demand is strong but inverse, so when the prices have an increase of 1%, the demand decreases approximately in 5 units. The influence of Prices in the Demand is not statistically important because the probability is 0.1417 and is greater than 5%. Also, the explanation of Demand Variance is done at 2.7% from the prices, so it can be concluded that this variance should be better explained by other variance at the 97.3% measure. So, the variance of this model is high (1549.915).

Also at the table it can be identified the standard error of the C, that is the free term of the model and the Prices. The standard error of C is considerably high, whereas the standard error of Prices is evidently low, because of the lower
importance that Prices have in the Demand of customers. The aim of every study is to optimize the model, and this optimization is through decreasing/minimizing the standard errors of every coefficient included in the model.

To achieve this result, the best way is to apply the Edgeworth approximation in the IID case, Bootstrap version, and the Edgeworth for Bootstrap version. The methodology followed consists in selecting 30 quantiles and identifying the errors of the Prices based on them. This methodology is used for the Quantile Regression Model, Edgeworth of the Quantile Regression Model, Bootstrap of Quantile Regression Model and the Edgeworth of the Bootstrap applied for the Quantile Regression Model (QRM).

Above, in the graph is presented the distribution of the standard errors (S.E.) of the Prices.

Edgeworth approximation in the case of IID

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1.234</td>
<td>0.234</td>
<td>5.678</td>
<td>0.0001</td>
</tr>
<tr>
<td>Prices</td>
<td>-0.345</td>
<td>0.123</td>
<td>3.456</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

After the application of the Edgeworth in the Model, the standard errors of the Prices are reduced. Also the Prices are statistically important in this generation, because the prob 0.0017 < 0.05. So the model after the Edgeworth approximation is improved.

So, the distribution of the standard errors of the Coefficients of C and Prices for every quantile, is improved, because of the better approximation generated by the Edgeworth of the IID case.

Below is presented the comparison of the standard errors distributions of the Quantile Regression Model and the Edgeworth application in QRM.
The contribution of edgeworth approximation in a real study case

After the application of the Bootstrap Version for 100,000 replications in the QRM the standard error of the Prices’ coefficient is decreased compared with the first generations of QRM.

Above, in graph is given the distribution of the standard errors for the Price’s Coefficient.

Edgeworth Approximation for Bootstrap Quantile Regression Model

After the application of the Edgeworth version in Bootstrap Model applied in QRM, the statistical importance of the Prices and C is increased. Also, the standard errors are sensitively decreased and consequently the model is improved.

Above is given the distribution of the standard errors of Prices. It can be identified that the tendence of standard errors is to be approximated with the 0 values. So, the Edgeworth Version in Bootstrap Model applied in Quantile Regression Model has improved the model, where the statistical importance of Prices in the customers’ demand is increased and the standard errors are sensitively decreased. So, the relation between the prices and the customer demand is successfully estimated and approximated by this method.

Below is shown the graphic where are presented the overlapped the Standard errors of the QRM, Edgeworth in QRM, Bootstrap in QRM and EW in Bootstrap Version. It looks clearly that the Edgeworth version in Bootstrap in QRM is the best solution.
The tendency of 2020 is quietly different compared with 2019. If the price increases with 1%, the demand decreases with approximately 3.5 units. So, the prices are identical, but the demand has been more inelastic, because the pandemic situation had increased the consuming tendency. Also, the explanation of the Demand through the prices is improved compared with the previous year (4%). Also the variance of the model has been improved (1371).

During the following phases are followed the same steps as for 2019.

Edgeworth approximation for IID case
Dependent Variable: Demand
Method: Quantile-Regression (Median)
Date: 02/05/21 Time: 12:44
Sample: 1 129
Included observations: 129
Ordinary (SD) Standard Errors & Covariance
Sparky metod: Bootstrapping using residuals
Bandwidth method: Hall-Hammer, bandwidth: 80 10229
Estimation successfully identifies unique optimal solution

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1026.348</td>
<td>176.125</td>
<td>5.813912</td>
<td>0.0000</td>
</tr>
<tr>
<td>price</td>
<td>-3.453567</td>
<td>1.387726</td>
<td>-2.515458</td>
<td>0.0135</td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.499662</td>
<td>Mean dependent var</td>
<td>1371.000</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.424179</td>
<td>S.D. dependent var</td>
<td>1754.567</td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>1028.890</td>
<td>Objective</td>
<td>61862.81</td>
<td></td>
</tr>
<tr>
<td>Quantile dependent</td>
<td>639.000</td>
<td>t-statistic</td>
<td>64822.00</td>
<td></td>
</tr>
<tr>
<td>Sparky</td>
<td>2878.466</td>
<td>Quasi-LR statistic</td>
<td>12.3372</td>
<td></td>
</tr>
<tr>
<td>Prob(Quasi-LR stat)</td>
<td>0.000444</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Edgeworth approximation for Bootstrap QR
Dependent Variable: Demand
Method: Quantile-Regression (Median)
Date: 02/05/21 Time: 12:48
Sample: 1 129
Included observations: 129
Ordinary (SD) Standard Errors & Covariance
Sparky metod: Kernel (Edgeworth) using residuals
Bandwidth method: Hall-Hammer, bandwidth: 80 10229
Initial Values: C(1)=0.00000 C(2)=0.00000
Estimation successfully identifies unique optimal solution

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1026.348</td>
<td>176.125</td>
<td>5.813912</td>
<td>0.0000</td>
</tr>
<tr>
<td>price</td>
<td>-3.453567</td>
<td>1.387726</td>
<td>-2.515458</td>
<td>0.0135</td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.406062</td>
<td>Mean dependent var</td>
<td>1371.000</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.424179</td>
<td>S.D. dependent var</td>
<td>1754.567</td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>1026.880</td>
<td>Objective</td>
<td>61602.81</td>
<td></td>
</tr>
<tr>
<td>Quantile dependent</td>
<td>639.000</td>
<td>t-statistic</td>
<td>64822.00</td>
<td></td>
</tr>
<tr>
<td>Sparky</td>
<td>2578.466</td>
<td>Quasi-LR statistic</td>
<td>12.3372</td>
<td></td>
</tr>
<tr>
<td>Prob(Quasi-LR stat)</td>
<td>0.000444</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Even in this case, the Edgeworth Approximation for Bootstrap QRM is the optimal solution that reduces the standard error of the prices.
Conclusions

In our paper, we have treated a Quantile Regression Model for the real data, in which we have applied Edgeworth Approximation in the case of IID, Bootstrap Version and the Edgeworth Approximation for the Bootstrap QRM, with the help of EViews10. Based on the results of the study we conclude that Edgeworth Approximation for Bootstrap Quantile Regression Model is the optimal solution that reduces the standard error of the independent variable’s coefficient.

References

https://doi.org/10.1198/016214502388618591

Received: February 7, 2021; Published: February 25, 2021