QUANTITATIVE METHODS IN ECONOMICS
Vol. XVI, No. 2, 2015, pp. 33 – 42

THE APPLICATION OF QUANTILE REGRESSION TO THE ANALYSIS OF THE RELATIONSHIPS BETWEEN THE ENTREPRENEURSHIP INDICATOR AND THE WATER AND SEWERAGE INFRASTRUCTURE IN RURAL AREAS OF COMMUNES IN WIELKOPOLSKIE VOIVODESHIP

Izabela Kurzawa, Jarosław Lira
Department of Finance and Accounting, University of Life Sciences in Poznan
e-mail: kurzawa@up.poznan.pl; jlira@up.poznan.pl

Abstract: The article presents the usefulness of quantile regression for the analysis of diversification in entrepreneurship in rural areas of communes in Wielkopolskie Voivodeship. The dependence between the entrepreneurship indicator value and the density and availability of the water and sewerage infrastructure was determined for individual quantiles of the entrepreneurship indicator distribution. This approach enables estimation of different quantile functions of the conditional cumulative distribution function of the entrepreneurship indicator. This analysis enables atypical observations when the conditional cumulative distribution function is diversified and does not have a standard form.

Keywords: quantile regression, entrepreneurship indicator, economic infrastructure, rural areas

INTRODUCTION

The demand for infrastructure and its services is related to the degree of socioeconomic development. The greater the growth and development of a county or commune is, the greater the demand for infrastructural services is [Ratajczak 1999].

The development of entrepreneurship in rural areas is significantly related to the distance between a county or commune and a major economic centre. As the distance from the centre increases, the number of business entities per 10,000 rural inhabitants at the working age decreases [Salamon 2009]. The development

of economic infrastructure in rural areas significantly influences both the number of business entities in the national economy and the growth of this number per 10,000 inhabitants at the working age. Therefore, increasing the infrastructural equipment in rural areas is a sine qua non, because it is a factor enabling their further growth and development [Lira 2014].

Microenterprises and small enterprises (about 99% of the total number of enterprises) play a key role in rural development. Therefore, the development of entrepreneurship plays a significant role in the development of local economy. It increases the production of goods, employment, people's income, the commune's budget and better satisfies local needs. It is necessary to stress the significance of the development of entrepreneurship for the functioning and socioeconomic development of the commune. It seems significant to research the influence of various traits, such as the influence of the density and availability of the water and sewerage infrastructure in rural areas on entrepreneurship in communes.

The aim of the study was to assess the usefulness of quantile regression for the analysis of dependence between the entrepreneurship indicator and the water and sewerage infrastructure in rural areas in Wielkopolskie Voivodeship. The article analyses communes in 2013 according to the density and availability of the water and sewerage infrastructure and registered business entities of the national economy (the REGON business entity register) in rural areas of Wielkopolskie Voivodeship.

METHOD

The article uses quantile regression because it is a method that enables the use of atypical observations (outlying observations). It is an advantage that the method uses the whole sample, so there is no problem of burdening parameter estimators. This problem would occur if the method of least squares was applied to subsamples identified on the basis of the dependent variable, i.e. the entrepreneurship indicator [Koenker 2005]. The authors of this method, Koenker and Basset [1978], observed that in the case of heteroscedasticity the estimation of a 0.5 quantile regression may prove to be a more effective method of searching for parameter values than the traditional regression based on the expected value of a dependent variable.

The quantile regression model can be presented in the following form [Koenker 2005,Trzpiot 2012, Davino at al. 2014]:

$$y_i = \beta_0^{(p)} + \sum_{j=1}^{J} \beta_j^{(p)} x_{ij} + \varepsilon_i^{(p)} \quad (1)$$

where:

$$Q_p(y_i|x_{ij}) = \beta_0^{(p)} + \sum_{j=1}^{J} \beta_j^{(p)} x_{ij} \quad (2)$$

conditional *p*-th quantile of dependent variable Y with known values of X_j variables, y_i – dependent variable values,

x_{ij} – values of independent variables (j=1, 2, …, J),
$0 < p < 1$ – the index defining regression parameters for *p*-th quantile of variable *Y* distribution.

Each time the estimation is carried out on the total sample. However, a different beta parameter is estimated for each quantile of the dependent variable. The estimation of quantile regression parameters consists in minimisation of the weighted sum of absolute remainder values, where appropriate weights are assigned to them[1]:

$$min \sum_{i=1}^{n} \rho_p \left(\left| y_i - \left(\beta_0^{(p)} + \sum_{j=1}^{J} \beta_j^{(p)} x_{ij} \right) \right| \right) \tag{3}$$

where:

$$\rho_p(z) = \begin{cases} pz & for\ z \geq 0 \\ (1-p)z & for\ z < 0 \end{cases} \tag{4}$$

$$z = y_i - \left(\beta_0^{(p)} + \sum_{j=1}^{J} \beta_j^{(p)} x_{ij} \right) \tag{5}$$

The estimated quantile regression parameters are interpreted similarly to classical regression estimators, i.e. parameter $\beta_j^{(p)}$ indicates variation in a particular quantile *p* of dependent variable *Y* as a result of unit variation of *j*-th independent variable X_j, where we assume that the other variables do not change. This enables us to show the diversified influence of independent variables on individual quantiles of dependent variable distribution. On the other hand, an absolute term can be interpreted as an approximate conditional distribution of the quantile function for dependent variable *Y*, where we assume that the values of independent variables equal zero.

The following measures are usually used to assess the quality of estimated quantile regression:

1. The Wald test is used to measure the significance of parameter assessment (the zero hypothesis assumes the insignificance of each individual parameter in the model) [Koenker, Machado 1999]:

[1] Atypical observations receive lower weights and thus, the problem of including them in the model is solved. Depending on the phenomenon character and data distribution, in empirical applications 3-9 different quantile regressions are usually estimated (these regressions correspond to consecutive quantiles or deciles in the distribution). The phenomenon is analysed according to all the models obtained. The bootstrap method is usually applied to obtain the estimators of standard errors of coefficients for the quantile regression. The STATA 12 software was used to estimate models in the article.

$$\begin{cases} H_0: \beta_j^{(p)} = 0 \\ H_1: \beta_j^{(p)} \neq 0 \end{cases} \qquad z = \frac{\hat{\beta}_j^{(p)}}{D(\hat{\beta}_j^{(p)})} \qquad \text{(j=0,1,...,J)} \tag{6}$$

2. Pseudo-R^2 [Davino et al. 2014]:

$$\text{pseudo-}R^2 = 1 - \frac{\sum_{y_i \geq \hat{y}_i} p \cdot |y_i - \hat{y}_i| + \sum_{y_i < \hat{y}_i} (1-p) \cdot |y_i - \hat{y}_i|}{\sum_{y_i \geq \hat{y}_i} p \cdot |y_i - \hat{Q}| + \sum_{y_i < \hat{y}_i} (1-p) \cdot |y_i - \hat{Q}|} \tag{7}$$

and [Koenker, Machado 1999]:

$$R^1 = 1 - \frac{\sum_{y_i \geq \hat{y}_i} \tau \cdot |y_i - \hat{y}_i| + \sum_{y_i < \hat{y}_i} (1-\tau) \cdot |y_i - \hat{y}_i|}{\sum_{y_i \geq \hat{y}_i} \tau \cdot \left|y_i - \hat{\beta}_0^{(p)}\right| + \sum_{y_i < \hat{y}_i} (1-\tau) \cdot \left|y_i - \hat{\beta}_0^{(p)}\right|} \tag{8}$$

where:
y_i – dependent variable values,
$\hat{y}_i$ – theoretical values of dependent variable,
$0 < p < 1$ – the index defining regression parameters for *p*-th quantile of the distribution of the variable Y – it is used as a weight,
$\hat{Q}$ – estimated quantile from the sample,
$\hat{\beta}_0$ – quantile for dependent variable Y from the estimated model, where we assume that the values of independent variables equal zero.
In theory these measures assume values from the interval [0,1], but they cannot be interpreted as coefficients of determination from classical linear regression. They are only a local measure of goodness of fit between the model and a particular quantile rather than the global measure of goodness of fit in the total conditional distribution. The higher the value of the measures is, the better the model was estimated.
In this study the authors suggest that the assessment of the model should be additionally supplemented with a quantile coefficient of determination and quantile coefficient of variation, adapted to quantile regression. They prove the goodness of fit between the model and empirical data [Rousseeuw, Leroy 1987]:

$$\text{quantile } R^2 = 1 - \left[\frac{Med\left|r_i^{(p)}\right|}{Med\left|y_i - \hat{\beta}_0^{(p)}\right|}\right]^2 \tag{9}$$

and

$$v = \frac{\hat{\sigma}}{\hat{\beta}_0^{(p)}} \tag{10}$$

$$\hat{\sigma} = \left\{ \sum_{i=1}^{n} w_i \cdot r_i^{(p)\,2} \Big/ \left(\sum_{i=1}^{n} w_i - 2\right) \right\}^{0,5} \tag{11}$$

$$w_i = \begin{cases} 1 & if \left|\frac{r_i^{(p)}}{s^0}\right| \leq 2.5 \\ 0 & in\ other\ case \end{cases} \quad (12)$$

$$s_0 = 1.4826 \cdot (1 + \tfrac{5}{n-2}) \cdot \sqrt{Med_i {r_i^{(p)}}^2} \quad (13)$$

where:

$Med\left|r_i^{(p)}\right|$ – median of absolute values of residuals,

$Med\left|y_i - \hat{\beta}_0^{(p)}\right|$ – median of deviations of real values of dependent variable from quantile $\hat{\beta}_0^{(p)}$ for dependent variable Y from the estimated model, where we assume that the values of independent variables equal zero,

$\hat{\sigma}$ – quantile standard deviation,

w_i – weights.

The higher the value of the quantile coefficient of determination is and the lower the value of the quantile coefficient of variation is, the better the model was estimated.

DATA

The study was based on data from 207 rural communes and isolated rural areas of urban-rural communes in Wielkopolskie Voivodeship in 2013 [Local Data Bank, Central Statistical Office, Warsaw].

The entrepreneurship indicator in the communes was determined by calculating the number of business entities in the REGON business entity register per 10,000 inhabitants at the working age[2]. The indicator was assumed as dependent variable *Y*. In the first variant of the model independent variables were related with the density of the water and sewerage infrastructure and they were expressed as follows:

X_1 – density of water supply distribution network (km/100 km^2),

X_2 – density of sewerage distribution network (km/100 km^2),

X_3 – percentage of rural inhabitants[3] with access to sewage treatment plants (%).

In the second variant independent variables were related with access to the water and sewerage infrastructure and they were expressed as follows:

X_4 – percentage of rural inhabitants with access to water supply network (%),

X_5 – percentage of rural inhabitants with access to sewerage network (%),

X_3 – percentage of rural inhabitants with access to sewage treatment plants (%).

Table 1 shows basic descriptive statistics of the independent variables and dependent variable under analysis. The sewerage network density was characterised

[2] Inhabitants at the working age: men (15-64 years), women (15-59 years).

[3] The actual population residing in the area was taken into consideration (as of 31 December 2013).

by the greatest diversification (about 108%). Apart from that, the network density was also characterised by relatively high right-sided asymmetry (3.39). Some of the communes under analysis did not have a sewerage network (about 7% of the total number of communes under study). This resulted in diversification between the communes in terms of the percentage of inhabitants with access to sewage treatment plants (the standard deviation was 71% of the mean value of the variable). As far as the entrepreneurship indicator is concerned, there were rather considerable differences between the communes. The lowest value of the indicator was noted in the commune of Wysoka in Piła County (439.9 business entities per 10,000 inhabitants at the working age). High values of the indicator were noted in the enterprising rural communes of Suchy Las, Tarnowo Podgórne and Komorniki in Poznań County, where the number of business entities exceeded 2,500).

Table 1. Descriptive statistics for components of the water and sewerage infrastructure and the entrepreneurship indicator in rural areas of communes in Wielkopolskie Voivodeship in 2013

Descriptive statistics	Density of distribution network (km/100 km^2)		Percentage of rural inhabitants with access to (%)			Entrepreneur-ship indicator
	water supply	sewerage	sewage treatment plants	water supply network	sewerage network	
minimum	8.80	0.00	0.00	52.80	0.00	439.90
0.25 quantile	68.00	10.20	15.50	84.00	17.80	874.20
0.50 quantile	92.20	18.70	32.80	89.60	34.40	1 042.40
0.75 quantile	116.00	33.80	52.80	93.30	47.60	1 210.50
maximum	236.20	222.50	96.80	98.90	77.00	3 224.60
coefficient of variation based on mean (%)	41.22	107.64	70.57	8.49	59.62	34.70
skewness	0.55	3.39	-1.36	-0.07	0.46	2.20

Source: own calculation based on Local Data Bank, Central Statistical Office, Warsaw

SELECTED RESEARCH FINDINGS

The analysis of individual estimated quantile regression models was started with statistical verification of the models. Particular attention was paid to the significance of structural parameters because only these parameters can be interpreted. Apart from that, the goodness of fit between the model and empirical data was determined by analysing pseudo-R^2, R^1, quantile R^2 and quantile coefficient

of variation v, which can be respectively treated as the local equivalents of the coefficient of determination and the coefficient of variation of the random component in a classical regression analysis estimated with the method of least squares. Tables 2 and 3 show the estimated parameter values and their errors for each model as well as probability *p* and coefficients describing the goodness of fit. It is noteworthy that quantile R^2 is characterised by much greater values than pseudo-R^2 and R^1. The values of pseudo-R^2 and R^1 are similar when the quartiles of the entrepreneurship indicator distribution are similar to the quartiles estimated on the basis of the model (the intercept).

The estimated quantile regression models enable determination of the diversification of the influence of individual traits referring to the density and availability of the water and sewerage infrastructure in rural areas of communes for the identified entrepreneurship indicator quantiles.

Table 2 shows the estimated 0.25, 0.5 and 0.75 quantile regression parameters of the entrepreneurship indicator. The analysis of the data in the table reveals that the density of the water supply distribution network had negative influence on the entrepreneurship indicator value in rural communes. For example, in 0.75 quantile of the entrepreneurship indicator distribution when the density of the water supply distribution network was increased by one ceteris paribus unit (i.e. 1 km per 100 km^2), the entrepreneurship indicator decreased by about 2.359. This means that in more entrepreneurial communes increasing the density of the water supply distribution network will not increase entrepreneurship, because this network is sufficient and it is treated as the basic network. Apart from that, in the 0.75 quantile, which referred to the communes with a relatively high entrepreneurship indicator, the influence was statistically significant ($p < 0.05$). On the other hand, the density of the sewerage network had statistically significant positive effect on entrepreneurship. The higher the value of this indicator was, the greater the entrepreneurship was. For example, an increase by one ceteris paribus unit in the communes with the highest entrepreneurship indicator values caused the entrepreneurship indicator to increase by about 11.705. This observation leads us to think that in more entrepreneurial communes the density of the sewerage distribution network is still unsatisfactory and the extension of the network causes a considerable increase in entrepreneurship. Usually this network is being developed in rural areas and it still is not sufficient. Apart from that, in each quantile of the distribution the influence was stronger by 7.757, 9.592 and 11.705, respectively. Thus, we can suppose that by increasing the density of the sewerage distribution network communes with high entrepreneurship indicator values will have greater chances for further significant development than less entrepreneurial communes and they will attract new investors. On the other hand, the percentage of rural inhabitants with access to sewage treatment plants had negative influence on the indicator under analysis and it proved to be statistically significant only in the 0.75 quantile.

Table 2. The dependence between the entrepreneurship indicator and the density of the water and sewerage infrastructure in rural areas of communes in Wielkopolskie Voivodeship in 2013

Conditional quantile for the entrepreneurship indicator	Constant	Density of distribution network (km/100 km^2)		Percentage of rural inhabitants with access to sewage treatment plants (%)
		water supply network	sewerage network	
0.25 quantile	**858.483**	**-0.957**	**7.757**	**-0.671**
standard error	71.084	0.747	3.662	1.952
p - value	0.000	0.202	0.035	0.732
goodness of fit (%)	7.8[a)]	7.9[b)]	28.7[c)]	29.8[d)]
0.50 quantile	**978.521**	**-1.073**	**9.592**	**-0.973**
standard error	90.706	0.883	2.274	1.235
p - value	0.000	0.226	0.000	0.432
goodness of fit (%)	14.2[a)]	16.1[b)]	32.7[c)]	20.5[d)]
0.75 quantile	**1,256.938**	**-2.359**	**11.705**	**-2.923**
standard error	124.291	1.012	1.816	1.360
p - value	0.000	0.021	0.000	0.033
goodness of fit (%)	20.5[a)]	20.9[b)]	44.0[c)]	22.9[d)]

[a)] pseudo-R^2, [b)] R^1, [c)] quantile R^2,[d)] quantile coefficient of variation v.

Source: as in Table 1

Table 3. The dependence between the entrepreneurship indicator and the availability of the water and sewerage infrastructure in rural areas of communes in Wielkopolskie Voivodeship in 2013

Conditional quantile for the entrepreneurship indicator	Constant	Percentage of rural inhabitants with access to (%)		
		water supply network	sewerage network	sewage treatment plants
0.25 quantile	**529.759**	**2.739**	**3.190**	**0.736**
standard error	187.754	2.129	3.359	2.197
p - value	0.005	0.200	0.344	0.738
goodness of fit (%)	6.1[a)]	42.3[b)]	88.6[c)]	45.6[d)]
0.50 quantile	**724.186**	**1.991**	**6.397**	**-1.700**
standard error	346.362	4.165	2.782	2.612
p - value	0.038	0.633	0.022	0.516
goodness of fit (%)	7.2[a)]	41.4[b)]	77.6[c)]	30.5[d)]
0.75 quantile	**1,297.940**	**-3.604**	**11.485**	**-3.556**
standard error	289.171	3.486	2.370	1.836
p - value	0.000	0.302	0.000	0.054
goodness of fit (%)	8.3[a)]	10.3[b)]	39.1[c)]	25.0[d)]

[a)] pseudo-R^2, [b)] R^1, [c)] quantile R^2,[d)] quantile coefficient of variation v.

Source: as in Table 1

Table 3 shows the estimated 0.25, 0.5 and 0.75 quantile regression parameters of the entrepreneurship indicator depending on the availability of the water and sewerage infrastructure. The analysis of the data in the table revealed that the percentage of the population with access to the sewerage network had positive influence on entrepreneurship in the communes. However, the influence increased for consecutive quantiles of the distribution of variable Y. In turn, there was low statistical significance between the entrepreneurship indicator and the percentage of the population with access to the water supply network. On the other hand, the percentage of the population with access to sewage treatment plants had positive influence on the increase in the entrepreneurship indicator only in communes with low entrepreneurship indicator values. Its negative influence was observed in more entrepreneurial communes. For example, in 0.75 entrepreneurship distribution quantile an increase in the percentage of rural population with access to sewage treatment plants by one ceteris paribus unit (by one per cent) caused the entrepreneurship indicator to decrease by 3.556, on average.

CONCLUSIONS

1. The quantile regression method complements the classical regression of least squares and it diversifies the influence of independent variables in the conditional quantiles of the dependent variable.
2. The quantile R^2 and quantile coefficient of variation v complement the possibilities to verify quantile regression models statistically.
3. The density of the water supply distribution network (km/100 km^2) had negative influence on the entrepreneurship indicator in the communes. The influence decreased in more entrepreneurial communes.
4. The density of the sewerage distribution network (km/100 km^2) had positive influence on the entrepreneurship indicator in the communes. The influence was the greatest in the most entrepreneurial communes.
5. The percentage of rural population with access to sewage treatment plants (%) had negative influence on the entrepreneurship indicator in the communes. Most likely, it indicates certain saturation and further increase in the percentage will decrease entrepreneurship.
6. As far as entrepreneurship is concerned, the density of the water and sewerage infrastructure is more significant than rural inhabitants' access to it.

REFERENCES

Davino C. Furno. M. Vistocco. D. (2014) Quantile Regression. Theory and Applications, John Wiley & Sons. Ltd. Oxford.

Koenker R. & Bassett G. Jr. (1978) Regression quantiles, Econometrica, Vol. 46. No. 1. pp. 33-50.

Koenker R. (2005) Quantile Regression, Econometric Society Monographs, No. 38, Cambridge University Press.

Koenker R. & Machado J. (1999) Goodness of fit and related inference processes for quantile regression, Journal of the American Statistical Association, 94. pp. 1296-1310.

Lira J. (2014) Rozwój infrastruktury gospodarczej a wskaźniki przedsiębiorczości na obszarach wiejskich województwa wielkopolskiego w latach 2004-2012, Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu, Vol. XVI, Issue 6, Warszawa-Poznań-Lublin, pp. 320-235.

Ratajczak M. (1999) Infrastruktura w gospodarce rynkowej, Akademia Ekonomiczna w Poznaniu, Poznań.

Rousseeuw P. J. Leroy A. M. (1987) Robust regression and outlier detection. John Wiley and Sons. New York.

Salamon J. (2009) Przestrzenne zróżnicowanie wartości wskaźnika przedsiębiorczości na obszarach wiejskich województwa świętokrzyskiego, Infrastruktura i Ekologia Terenów Wiejskich, No. 5, PAN, Oddział Kraków, pp. 231-239.

Trzpiot G. (2012) Ekstremalna regresja kwantylowa, Studia Ekonomiczne Uniwersytetu Ekonomicznego w Katowicach, Zeszyty Naukowe Wydziałowe, No. 91, Katowice 2012, pp. 11-20.