ICIC Express Letters
Volume **11**, Number **6**, June **2017**

ICIC International ©2017 ISSN 1881-803X
pp. 1143–1151

RESEARCH ON RELIABILITY TEST METHOD BASED ON REGRESSION VOLATILITY

Fachao Li, Xiaoxiao Su and Chenxia Jin

School of Economics and Management
Hebei University of Science and Technology
No. 26, Yuxiang Street, Shijiazhuang 050018, P. R. China
lifachao@tsinghua.org.cn; 1301141201@qq.com; jinchenxia2005@126.com

Received November 2016; accepted February 2017

Abstract. *Regression analysis is a classic prediction method. Determining the regression equation with the minimum sum of all the sample deviations is a core of regression analysis. The classical regression model is based on the assumption that residual error is normal distribution. However, real error term structure is complex and does not obey the normal distribution. Although the assumption of residual error has nothing to do with the least square method, it is very important to the statistical inference and test. It means that the classical test method cannot verify the regression equation without following the assumption. In the paper, we firstly analyze the characteristics of classical regression test models, and put forward the volatility value reflecting the distance between two regression equations. Secondly we analyze the volatility value and build the volatility statistic. Thirdly we analyze the characteristics of the volatility statistic, then we establish a regression test model based on the volatility statistics (denoted by VR-TM), and further analyze the characteristics of VR-TM from theory and application. The results show that VR-TM not only has good structure characteristics and interpretability, but also extends and perfects the existing regression test methods.*

Keywords: Regression analysis, The volatility statistic, Test method, Reliability

1. **Introduction.** Regression analysis, as one of the three branches of statistics, helps people to study the dependent relationship between the reason variables and result ones. Regression analysis has been applied successfully in many management and prediction problems. For example, [1] built a sensor-based forecasting model using Support Vector Regression and applied it to an empirical data-set from a multi-family residential building in New York City. For the problem that support vector regression has deficiency to solve the problem that electric load forecasting appears highly nonlinear characteristics, [2] proposed a multidimensional regression model based on support vector to generate color scales which provided a tool to distinguish the stage of phenolic ahead of the autumn harvest. [3] established a fuzzy linear regression model based on the influence of sensory evaluation to fried doughnuts sales. [4] presented a regression model about nitrogen dioxide concentration and local wind direction, and then predicted the nitrogen dioxide concentration precisely in future. [5] built an hourly cooling load forecasting regression model based on time index. [6] applied Gaussian Process Regression (GPR) to probabilistic stream flow forecasting. People in the study found that the new forecasting method and model constantly perfect the existing model. [7] presented the least squares regression is based on some basic assumptions; any deviation hypothetical situation will affect the regression results. The article discussed the main problems in regression analysis deviating from the basic assumption which may affect the results of regression analysis and give the corresponding ways to find and remedy problems. [8] built a regression model based on the quasi linear function (QRM), and discussed the parameter estimation of QRM strategies; this paper give the parameters estimation method based on the genetic algorithm

and the least squares estimation method, and the error test method based on the residual. [9] established a high dimensional nonlinear regression model to the physical combustion model and the main variables are selected by the principal component analysis method. [10] proposed an improved auto correlation kernel regression method, which can provide early prediction of industrial parts. [11] extended some fuzzy linear regression methods to polynomial form which are applied to financial problems. [12] improved the frequency domain regression model to calculate the coefficient of the multilayer structure of the conduction transfer function. [13] developed novel structural break tests to distinguish breaks in intercept from slope parameters in linear regression models and to significantly improve the power of these tests, the process from which they are derived is weighted and exploits higher moments of the residual process.

From the above analysis, we know that regression analysis is widely used in practice. The premise of the regression analysis application is through the tests. Current regression test models are mostly established on the basis of the error term obeying normal distribution, while a lot of sample sets are hard to adjust to obey the assumption. So it is necessary to put forward a testing method whose application is wider. We mainly do the work as follows. 1) We introduce the classic regression analysis and common test models. 2) We construct the volatility statistic, and then establish the VR-TM based on the law of large numbers. The VR-TM can test the regression functions whether its residual error obeys the classical regression model hypothesis or not. 3) Using a concrete case, we show the application value of VR-TM.

2. Main Test Models of Regression Function.

Regression analysis is a method to study the correlation between variables. Its core content is to determine the relationship between variables in the sense of average. And the basic form is:

$$y = \mu(x) + \varepsilon(x). \tag{1}$$

Here, x denotes explanatory variable, y denotes explained variable, $\mu(x)$ (called ***regression function***, and it denotes the mathematical expectation of $y(x)$ intuitively) is the deterministic relationship of x, and $\varepsilon(x)$ is the error term.

Current regression analysis theories are mostly established on the basis of $\varepsilon(x)$ obeying normal distribution $N(0, \sigma^2)$. So after estimating the $\mu(x)$, we should verify the reliability on the basis of $\varepsilon(x)$ obeying $N(0, \sigma^2)$. There are two common types of method to verify the reliability of regression function. One is to verify the confidence interval of the regression function to predict under a certain confidence level. The other is to verify the fitting degree, for example:

1) F-test: Let $S_R = \sum(\hat{y}_i - \bar{y})^2$ and $S_R = \sum(y_i - \hat{y}_i)^2$, and then $F = \frac{S_R}{S_e/(n-2)}$.

2) R-test: Let $l_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$ and $l_{xx}l_{yy} = \sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2$, and then $R = l_{xy}/\sqrt{l_{xx}l_{yy}}$. Here, $\hat{y}_i$ denotes the predicted value of x_i by the regression function, $\bar{y}$ denotes the mean value of the y_i and $\bar{x}$ denotes the mean value of the x_i.

Current regression test models are mostly established on the basis of $\varepsilon(x)$ obeying normal distribution $N(0, \sigma^2)$. However, the error term does not obey the normal distribution in many cases. It greatly limits the application of the regression analysis. We always can obtain the regression function by the OLS, whether error term obeys the normal distribution or not. In the case that the error term does not obey the normal distribution, the R-test can verify the fitting degree (F-test no longer applies), while the models to verify the confidence interval of the regression function lose efficacy. The above analysis shows that the existing regression test methods need improving further. In the following, we will focus on the shortcomings of current regression test method and discuss the VR-TM.

3. **The Volatility Statistic.** In this section, surrounding the reliability measurement of the regression function, we will discuss the construction strategy of the wave statistic. For convenience, we assume that: 1) $\Omega = \{(x_i, y_i)|i = 1, 2, \ldots, n\}$ denotes the sample data set. Here, (x_i, y_i) denotes the observed values of explanatory variables and explained variables. 2) Ω_k denotes the proper subset of Ω. 3) $\hat{y}(x, \Omega)$ (*called* ***parent function***) and $\hat{y}(x, \Omega_k)$ (*called* ***sub-function***) are corresponding short for the regression function based on the data set Ω and Ω_k.

It is easy to see, $\hat{y}(x, \Omega)$ has the exact meaning, and $\hat{y}(x, \Omega_k)$ has the same meaning with $\hat{y}(x, \Omega)$ to a certain degree. For the sample data set $\Omega = \{(x_i, y_i)|i = 1, 2, \ldots, n\}$, we randomly take some elements from Ω, and then repeat m times. Every time the elements taken are denoted by Ω_k $(k = 1, 2, \ldots, m)$ and there are enough elements in Ω_k $(k = 1, 2, \ldots, m)$. If Ω has a layered phenomenon or the managers classify Ω with an important feature, in each type of classes, the elements inside should be taken by the same opportunity. So we can understand Ω_k $(k = 1, 2, \ldots, m)$ as a cover of Ω and $\{\hat{y}(x, \Omega_k)|k = 1, 2, \ldots, m\}$ are the basic factors reflecting the local feature of $\hat{y}(x, \Omega)$. Then the reliability of $\hat{y}(x, \Omega)$ can be reflected by the distance between the parent function and the sub-functions.

The distance between different regression functions based on limited sample data sets can be measured by the average distance from the parent function to the sub-functions based on all variables, that is:

$$D_k = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}(x_i, \Omega) - \hat{y}(x_i, \Omega_k)| \tag{2}$$

is called ***the volatility value***. By the definition of D_k, we know that D_k denotes the volatility value between $\hat{y}(x, \Omega)$ and $\hat{y}(x, \Omega_k)$. Because of the randomness of selection, we can know that $\{D_k|k = 1, 2, \ldots, m\}$ is an independent and identically distributed sequence. Based on the above analysis, we can build the statistic $\bar{D}$ (called ***the volatility statistic***):

$$\bar{D} = \frac{1}{m} \sum_{k=1}^{m} D_k. \tag{3}$$

Combined with the definition of D_k, we know that $\bar{D}$ shows the mean distance between the parent function and the sub-functions. Let us take an extreme example, $\bar{D} = 0$ if and only if the relation between variables is certain. So, if the regression function based on the data set Ω is a certain curve, then its sub-functions are the same as the parent function. In such case $D_1 = D_2 = \cdots = D_m$, then $\bar{D} = 0$.

4. Volatility-Based Reliability Test Model.

4.1. The characteristic analysis of the volatility statistic and the construction of VR-TM.

In this section we will further discuss the value rule of the volatility statistic from the angle of quantification. To help us understand the volatility statistic, we introduce the theorems of [14].

Theorem 4.1. *Let $\{\varphi_n(t)\}$ denote the characteristic function sequence of distribution function sequence $\{F_n(x)\}$, and $\varphi(t)$ denote the characteristic function of distribution function $F(x)$. Then $\{F_n(x)\}$ weak converges to $F(x)$, if and only if $\{\varphi_n(t)\}$ converges to $\varphi(t)$.*

Theorem 4.2. *Let $\{D_m\}$ be independent and identically distributed sequence with the mean value μ and the variance σ^2, $Y_m = \frac{\bar{D}-\mu}{\sigma/\sqrt{m}}$. Then for any real number s,*

$$\lim_{m \to +\infty} P(Y_m \le s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{\frac{-t^2}{2}} dt \triangleq \Phi(s).$$

Proof: Let $\varphi(t)$ be the characteristic function of $D_m - \mu$. And we know that the characteristic function of $D_m - \mu$ is the same as D_m. Then by $\bar{D} = \frac{1}{m}\sum_{k=1}^{m} D_k$, we know the characteristic function of Y_m is $\varphi_{Y_m}(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{m}}\right)\right]^m$. If we assume that $\varphi'(t)$ denotes the first derivative function of $\varphi(t)$ and $\varphi''(t)$ denotes the second derivative function of $\varphi(t)$, and then using $E(D_m-\mu) = 0$ and $Var(D_m-\mu) = \sigma^2$, we have $\varphi'(0) = 0$, $\varphi''(0) = -\sigma^2$, and

$$\varphi(t) = \varphi(0) + \varphi'(0) + \varphi''(0)\frac{t^2}{2} + o\left(t^2\right) = 1 - \frac{1}{2}\sigma^2 t^2 + o\left(t^2\right).$$

Taking $m \to \infty$ on the characteristic function $\varphi_{Y_m}(t)$, we can obtain

$$\lim_{m\to\infty} \varphi_{Y_m}(t) = \lim_{m\to\infty}\left[1 - \frac{t^2}{2m} + o\left(\frac{t^2}{m}\right)\right]^m = e^{-t^2/2}. \tag{4}$$

$e^{-t^2/2}$ is the characteristic function of the standard normal distribution $N(0,1)$. According to Theorem 4.1, we know Theorem 4.2 holds.

According to Theorem 4.2, we know the following: 1) $\bar{D}$ weak converges to $N(\mu, \sigma^2/m)$. This implies that the smaller μ is, the more reliable the parent function is; the smaller σ is, the more stable the parent function is. For example, according to the 3σ principle of the normal distribution $N(\mu, \sigma^2)$, we know that the percentage of values from normal distribution is 68.26% in $(\mu - \sigma, \mu + \sigma)$, 95.45% in $(\mu - 2\sigma, \mu + 2\sigma)$ and 99.73% in $(\mu - 3\sigma, \mu + 3\sigma)$. The length of the interval represents the reliability of the prediction. 2) If m is big enough, then the distribution of $\bar{D}$ is $N(\mu, \sigma^2/m)$. According to the above analysis, when the deviation standard of the parent function is δ, then the following formula can measure the percentage:

$$P\left(0 \le \bar{D} \le \delta\right) \approx \Phi\left(\frac{\delta - \mu}{\sigma/\sqrt{m}}\right) + \Phi\left(\frac{\mu}{\sigma/\sqrt{m}}\right) - 1 \triangleq \beta. \tag{5}$$

So, $[\delta, \beta]$ can be used to describe the regression function in qualitative. Obviously, 1) the intuitive meaning of $[\delta, \beta]$ is that the probability β of average deviation between the parent function and the sub-function is less than δ. It also means the probability that the predicted value $\hat{y}$ falls in $[\hat{y} - \delta, \hat{y} + \delta]$ is β. For example, let the $[\delta, \beta]$ of a regression function be $[0.5, 0.9]$. That means that the predicted deviation of the regression function is less than 0.5 with probability 0.9. 2) When β is certain, the smaller δ is, the higher reliability of the regression equation is. When δ is in a certain range, the bigger β is, the higher reliability of the regression equation is. In practice, δ dose not exceed 10% of the predicted value and β should be greater than 90%. For μ and σ are unknown in Formula (5), we can use the sample mean and the sample standard deviation as the estimations to compute. Then, for any regression function, we can get a quantitative description about the reliability by Formula (5). For convenience Formula (5) is called ***the volatility-reliability test model*** (shorthand for VR-TM).

4.2. **The application of VR-TM in the regression model conforming to the assumptions.** In the zoology, sometimes people need to know the relationship between the volume and weight of animals. It is relatively easy to measure the weight of the animal to the volume. So people want to predict the volume by the weight of animals. There are 18 samples data in Table 1 about the volume and the weight of some animals. x_i denotes the weight of the animal and the unit of x_i is kg; y_i denotes the volume of the animal and the unit of y_i is dm^3.

The specific steps are stated as follows.

Step 1 Get the parent function of the 18 sample data by the least square method: $\hat{y} = 0.988x - 0.105$.

Table 1. The sample data about the weight x_i and volume y_i of the animals

x_i	10.4	10.5	11.9	12.1	13.8	15.0	15.1	15.1	15.1
y_i	10.2	10.4	11.6	11.9	13.5	14.5	14.8	15.1	14.5
x_i	15.7	15.8	16.0	16.5	16.7	17.1	17.1	17.8	18.4
y_i	15.7	15.2	15.8	15.9	16.6	16.7	16.7	17.6	18.3

Table 2. The 30 sub-functions

$y = 0.991x - 0.136$	$y = 0.943x + 0.460$	$y = 1.002x - 0.351$	$y = 0.945x + 0.371$	$y = 0.990x - 0.153$
$y = 0.995x - 0.150$	$y = 0.937x + 0.553$	$y = 0.999x - 0.264$	$y = 0.917x + 0.036$	$y = 0.990x - 0.121$
$y = 0.980x - 0.035$	$y = 0.989x - 0.089$	$y = 1.000x - 0.206$	$y = 0.976x + 0.056$	$y = 0.984x - 0.099$
$y = 0.999x - 0.134$	$y = 0.968x + 0.186$	$y = 0.955x + 0.280$	$y = 0.986x - 0.075$	$y = 0.981x + 0.008$
$y = 0.988x - 0.023$	$y = 0.974x + 0.081$	$y = 0.978x - 0.031$	$y = 0.970x + 0.114$	$y = 0.998x - 0.242$
$y = 1.003x - 0.334$	$y = 0.983x - 0.060$	$y = 0.998x - 0.144$	$y = 0.983x - 0.078$	$y = 1.006x - 0.404$

Table 3. The 30 volatility values

0.013	0.137	0.038	0.176	0.019	0.059	0.146	0.022	0.325	0.013
0.013	0.081	0.037	0.077	0.059	0.024	0.027	0.031	0.110	0.049
0.051	0.030	0.078	0.031	0.055	0.135	0.041	0.120	0.004	0.037

Step 2 Select randomly 12 data from the Table 1, and then get the regression function of the 12 sample data. Repeat it 30 times. The 30 sub-functions are shown in Table 2.

Step 3 Calculate the volatility values of the 30 sub-functions by Formula (2). The volatility values are shown in Table 3.

Step 4 Calculate the mean value $\bar{\mu}$ and the variance $\hat{\sigma}^2$ of the 30 volatility values: $\bar{\mu} = 0.068$ and $\hat{\sigma}^2 = 0.004$.

Step 5 If the weight of one animal is 17.6kg, then the estimated volume of this animal is 17.284dm^3 according to the parent function: $\hat{y} = 0.988x - 0.105$. When the percentage about the reliability of the estimated volume is 95%, the prediction interval is $(16.808, 17.763)$. The length of the prediction interval is 0.955. So the reliability of this regression function is high, and we can rely on the predictive results.

Step 6 Calculate the probability by Formula (5). Y_m denotes the deviation from the actual value to estimated values including positive deviation and minus deviation. So $\delta = 0.955/2 = 0.478$. If $\delta = 0.478$, then $P(0 \leq \bar{D} \leq \delta) \approx 0.918$. If we want to know the interval with the confidence level 95%, we do it like the following: δ_1 denotes the actual deviation of the estimated value. When $P(0 \leq \bar{D} \leq \delta_1) = \frac{1}{\sqrt{2\pi}\sigma/\sqrt{m}} \int_0^{\delta_1} e^{\frac{-(t-\mu)^2}{2(\sigma^2/m)}} dt = 0.95$, then $\delta_1 = 0.418$. Form the above, if the weight of the animal is 17.6kg, then the interval of the real volume with the confidence level 95% is $[16.866, 17.702]$. So the reliability of this regression function is high, and we can rely on the predictive results.

The main contribution of the regression function is prediction. From the above analysis, if the function conforms to the classical assumptions, the interval under a certain confidence level by VR-TM is basically the same as that got by the classical test method. So the VR-TM is feasible. Since VR-TM can not only apply to the standard regression functions but also those regression functions whose error terms do not obey the classical assumptions, VR-TM is the promotion of classical test model.

5. **Application Example.**

In Section 4.2, we introduce the implementation strategy of the VR-TM simply by a regression model in line with the assumption. In this section, we will combine a concrete case to show the special thing about the VR-TM compared with the current test models. The speciality is that VR-TM is not though adjusting the error term to obey the assumption to test the regression functions without obeying the assumption. The case whose sample set does not obey the assumption, is different from the case in Section 4.2. And we also can combine with this concrete case to further illustrate the effectiveness and the specific implementation process of VR-TM.

Case description: The purpose of investors to create the company is profitable. For the established companies, they have a variety of objectives: improve treatment of workers, improve working conditions, expand market share, improve product quality, reduce environmental pollution and so on, but the basis is profit. The goal of shareholders' capital to enterprises is to maximize the wealth of shareholders. Rate of return on common stockholders' equity (shorted for ROE) is an important indicator to measure earnings. ROE reflects the relation of net profit and owner's equities. Net sales divided by total assets is called total asset turnover. It is a basic driving force of ROE.

How to develop the total assets turnover to achieve the aim of shareholders' ROE is important in academics and applications. A new type of enterprise plans to increase shareholders' wealth by a a period of operation plan. The key of plan is to find the relationship between ROE and total asset turnover based on the records of 90 from the similar established companies. The 90 samples data are shown in Table 4. Here, x_i denotes the total asset turnover; y_i denotes the ROE. This shows that the correlation of the total asset turnover and the ROE can be summarized as a regression problem based on data sample $\{(x_i, y_i)|i = 1, 2, \ldots, n\}$. We will use VR-TM to verify the reliability of the regression function of the total asset turnover and ROE. The specific process is stated as follows.

Step 1 Take the first 82 data in Table 4 as the regression set Ω, the last 8 as the test sample data.

Step 2 Calculate the parent function based on Ω. $\hat{y}(x,\ \Omega)$ denotes the parent function. $\hat{y}(x,\ \Omega)$ is shown in Table 5.

Step 3 The scatter plot of Ω is shown in Figure 1. Obviously, stratification phenomenon shows in Figure 1. So the sample data set is not homogeneous. It means that the sample data set does not obey the classical assumptions. Divide Ω into two sub-samples: Ω_1 and Ω_2, according to the difference value between the real value and the predicted value by the corresponding parent function. The elements in Ω_1 are those whose real value is greater than or equal to the predicted value by the parent function. The elements in Ω_1 are on or up the curve of the parent function. Then the remaining samples are in Ω_2. The elements of Ω_1 and Ω_2 are shown in Table 6 and Table 7. Select randomly 38 data from Ω_1 and 22 data from Ω_2. Ω_{λ_1} denotes the set of the selected 60 data. If we repeat it 71 times, then we

TABLE 4. The total assert turnover x_i and the ROE y_i(%)

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
1.05	10.1	5.00	46.9	2.20	23.6	3.00	37.0	2.80	35.6	2.60	34.3	1.50	22.0	4.20	43.3
1.10	10.4	5.10	47.9	2.40	25.1	3.10	37.7	4.30	43.9	1.70	25.3	1.85	20.4	1.10	17.5
1.25	11.6	5.20	48.1	2.45	25.8	3.20	37.9	4.40	44.4	1.90	27.9	1.30	18.9	1.65	18.7
1.00	10.2	5.30	47.9	2.65	27.0	3.30	39.1	4.50	45.1	2.00	28.0	1.70	19.4	1.60	17.9
1.15	11.5	5.40	48.3	2.80	28.3	3.40	39.7	4.60	45.4	2.10	29.6	1.80	20.0	6.10	53.0
1.30	12.2	5.50	48.7	3.10	30.6	3.50	40.3	4.70	46.2	2.20	29.7	1.40	21.5	1.20	17.7
1.35	13.2	5.60	49.1	3.25	31.7	3.60	40.4	4.80	46.7	2.40	32.5	1.60	24.3	4.10	43.0
1.40	16.2	5.70	49.5	3.40	32.9	3.70	41.3	4.90	46.8	2.50	32.6	2.15	23.4	2.90	35.9
1.45	13.4	5.80	50.7	3.65	34.2	3.80	41.5	1.90	20.9	1.80	25.4	2.10	22.6	2.70	33.9
1.50	14.6	5.90	51.2	3.75	35.2	3.90	41.7	2.05	22.0	2.30	30.1	1.55	17.6	6.00	52.1
1.00	15.2	4.00	42.7	–	–	–	–	–	–	–	–	–	–	–	–

Table 5. The parent function and the sub-functions

$\hat{y}(x,\Omega)$	$y=23.275\ln x+9.203$	$\hat{y}(x,\Omega_{\lambda_1})$	$y=24.916\ln x+6.492$	$\hat{y}(x,\Omega_{\lambda_2})$	$y=24.067\ln x+7.237$
$\hat{y}(x,\Omega_{\lambda_3})$	$y=24.828\ln x+6.395$	$\hat{y}(x,\Omega_{\lambda_4})$	$y=24.771\ln x+6.526$	$\hat{y}(x,\Omega_{\lambda_5})$	$y=24.556\ln x+6.827$
$\hat{y}(x,\Omega_{\lambda_6})$	$y=25.008\ln x+6.826$	$\hat{y}(x,\Omega_{\lambda_7})$	$y=26.181\ln x+5.161$	$\hat{y}(x,\Omega_{\lambda_8})$	$y=24.968\ln x+6.500$
$\hat{y}(x,\Omega_{\lambda_9})$	$y=25.222\ln x+6.219$	$\hat{y}(x,\Omega_{\lambda_{10}})$	$y=24.477\ln x+6.639$	$\hat{y}(x,\Omega_{\lambda_{11}})$	$y=24.239\ln x+7.457$
$\hat{y}(x,\Omega_{\lambda_{12}})$	$y=24.960\ln x+6.569$	$\hat{y}(x,\Omega_{\lambda_{13}})$	$y=24.848\ln x+6.649$	$\hat{y}(x,\Omega_{\lambda_{14}})$	$y=24.815\ln x+6.378$
$\hat{y}(x,\Omega_{\lambda_{15}})$	$y=24.434\ln x+1.394$	$\hat{y}(x,\Omega_{\lambda_{16}})$	$y=24.376\ln x+7.335$	$\hat{y}(x,\Omega_{\lambda_{17}})$	$y=25.852\ln x+5.691$
$\hat{y}(x,\Omega_{\lambda_{18}})$	$y=25.587\ln x+5.411$	$\hat{y}(x,\Omega_{\lambda_{19}})$	$y=25.470\ln x+5.918$	$\hat{y}(x,\Omega_{\lambda_{20}})$	$y=24.460\ln x+1.144$
$\hat{y}(x,\Omega_{\lambda_{21}})$	$y=25.060\ln x+6.579$	$\hat{y}(x,\Omega_{\lambda_{22}})$	$y=25.089\ln x+6.293$	$\hat{y}(x,\Omega_{\lambda_{23}})$	$y=24.390\ln x+7.113$
$\hat{y}(x,\Omega_{\lambda_{24}})$	$y=24.587\ln x+6.877$	$\hat{y}(x,\Omega_{\lambda_{25}})$	$y=24.753\ln x+6.642$	$\hat{y}(x,\Omega_{\lambda_{26}})$	$y=25.006\ln x+6.364$
$\hat{y}(x,\Omega_{\lambda_{27}})$	$y=24.560\ln x+6.825$	$\hat{y}(x,\Omega_{\lambda_{28}})$	$y=25.055\ln x+6.580$	$\hat{y}(x,\Omega_{\lambda_{29}})$	$y=24.000\ln x+7.823$
$\hat{y}(x,\Omega_{\lambda_{30}})$	$y=24.533\ln x+6.745$	$\hat{y}(x,\Omega_{\lambda_{31}})$	$y=24.691\ln x+6.175$	$\hat{y}(x,\Omega_{\lambda_{32}})$	$y=24.717\ln x+12.252$
$\hat{y}(x,\Omega_{\lambda_{33}})$	$y=25.488\ln x+5.724$	$\hat{y}(x,\Omega_{\lambda_{34}})$	$y=24.194\ln x+7.160$	$\hat{y}(x,\Omega_{\lambda_{35}})$	$y=24.837\ln x+6.403$
$\hat{y}(x,\Omega_{\lambda_{36}})$	$y=24.247\ln x+6.927$	$\hat{y}(x,\Omega_{\lambda_{37}})$	$y=24.966\ln x+6.338$	$\hat{y}(x,\Omega_{\lambda_{38}})$	$y=24.305\ln x+7.281$
$\hat{y}(x,\Omega_{\lambda_{39}})$	$y=25.527\ln x+5.795$	$\hat{y}(x,\Omega_{\lambda_{40}})$	$y=25.051\ln x+6.341$	$\hat{y}(x,\Omega_{\lambda_{41}})$	$y=24.662\ln x+7.097$
$\hat{y}(x,\Omega_{\lambda_{42}})$	$y=24.056\ln x+7.749$	$\hat{y}(x,\Omega_{\lambda_{43}})$	$y=25.292\ln x+6.149$	$\hat{y}(x,\Omega_{\lambda_{44}})$	$y=25.150\ln x+6.098$
$\hat{y}(x,\Omega_{\lambda_{45}})$	$y=24.261\ln x+6.811$	$\hat{y}(x,\Omega_{\lambda_{46}})$	$y=25.496\ln x+5.791$	$\hat{y}(x,\Omega_{\lambda_{47}})$	$y=25.235\ln x+5.954$
$\hat{y}(x,\Omega_{\lambda_{48}})$	$y=25.022\ln x+6.177$	$\hat{y}(x,\Omega_{\lambda_{49}})$	$y=25.528\ln x+5.995$	$\hat{y}(x,\Omega_{\lambda_{50}})$	$y=24.681\ln x+6.271$
$\hat{y}(x,\Omega_{\lambda_{51}})$	$y=25.274\ln x+5.802$	$\hat{y}(x,\Omega_{\lambda_{52}})$	$y=24.858\ln x+6.661$	$\hat{y}(x,\Omega_{\lambda_{53}})$	$y=25.229\ln x+5.303$
$\hat{y}(x,\Omega_{\lambda_{54}})$	$y=24.708\ln x+6.454$	$\hat{y}(x,\Omega_{\lambda_{55}})$	$y=25.294\ln x+6.133$	$\hat{y}(x,\Omega_{\lambda_{56}})$	$y=24.714\ln x+6.954$
$\hat{y}(x,\Omega_{\lambda_{57}})$	$y=24.780\ln x+6.275$	$\hat{y}(x,\Omega_{\lambda_{58}})$	$y=23.509\ln x+7.650$	$\hat{y}(x,\Omega_{\lambda_{59}})$	$y=25.037\ln x+6.190$
$\hat{y}(x,\Omega_{\lambda_{60}})$	$y=24.891\ln x+5.896$	$\hat{y}(x,\Omega_{\lambda_{61}})$	$y=24.524\ln x+6.590$	$\hat{y}(x,\Omega_{\lambda_{62}})$	$y=24.994\ln x+6.179$
$\hat{y}(x,\Omega_{\lambda_{63}})$	$y=25.027\ln x+5.879$	$\hat{y}(x,\Omega_{\lambda_{64}})$	$y=25.081\ln x+6.301$	$\hat{y}(x,\Omega_{\lambda_{65}})$	$y=24.576\ln x+7.328$
$\hat{y}(x,\Omega_{\lambda_{66}})$	$y=24.149\ln x+7.940$	$\hat{y}(x,\Omega_{\lambda_{67}})$	$y=25.119\ln x+6.368$	$\hat{y}(x,\Omega_{\lambda_{68}})$	$y=24.527\ln x+6.367$
$\hat{y}(x,\Omega_{\lambda_{69}})$	$y=25.106\ln x+6.287$	$\hat{y}(x,\Omega_{\lambda_{70}})$	$y=25.043\ln x+6.147$	$\hat{y}(x,\Omega_{\lambda_{71}})$	$y=24.843\ln x+0.294$

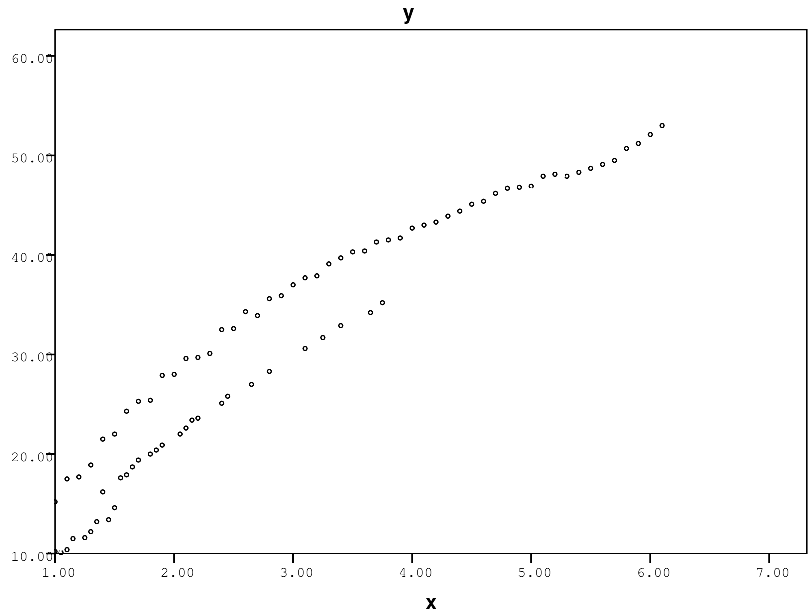

Figure 1. The scatter plot of Ω

have 71 sets denoted by $\Omega_{\lambda_1},\ \Omega_{\lambda_2},\ldots,\ \Omega_{\lambda_{71}}$. $\hat{y}(x,\ \Omega_{\lambda_1}),\ \hat{y}(x,\ \Omega_{\lambda_2}),\ldots,\ \hat{y}(x,\ \Omega_{\lambda_{71}})$ respectively represent the regression function based on the sample data set $\Omega_{\lambda_1},\ \Omega_{\lambda_2},\ldots,\ \Omega_{\lambda_{71}}$. The specific sub-functions are shown in Table 5.

Step 4 Calculate respectively the volatility values of the 71 sub-functions by Formula (2). The volatility values are shown in Table 8.

Step 5 Calculate the mean value $\bar{\mu}$ and the variance $\hat{\sigma}^2$ of the 71 volatility values: $\mu = 6.215$ and $\hat{\sigma} = 0.021$.

Table 6. The sample data of Ω_1

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
1.00	15.2	2.50	32.6	4.00	42.7	5.50	48.7	1.10	17.5	2.60	34.3	4.10	43.0	5.60	49.1
1.20	17.7	2.70	33.9	4.20	43.3	5.70	49.5	1.30	18.9	2.80	35.6	4.30	43.9	5.80	50.7
1.40	21.5	2.90	35.9	4.40	44.4	5.90	51.2	1.50	22.0	3.00	37.0	4.50	45.1	6.00	52.1
1.60	24.3	3.10	37.7	4.60	45.4	6.10	53.0	1.70	25.3	3.20	37.9	4.70	46.2	2.30	30.1
1.80	25.4	3.30	39.1	4.80	46.7	2.40	32.5	1.90	27.9	3.40	39.7	4.90	46.8	3.80	41.5
2.00	28.0	3.50	40.3	5.00	46.9	3.90	41.7	2.10	29.6	3.60	40.4	5.10	47.9	5.30	47.9
2.20	29.7	3.70	41.3	5.20	48.1	5.40	48.3	–	–	–	–	–	–	–	–

Table 7. The sample data of Ω_2

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
1.05	10.1	1.00	10.2	1.70	19.4	2.10	22.6	1.10	10.4	1.35	13.2	1.80	20.0	2.65	27.0
1.15	11.5	1.40	16.2	1.85	20.4	2.80	28.3	1.30	12.2	1.55	17.6	1.90	20.9	3.25	31.7
1.45	13.4	1.60	17.9	2.05	22.0	3.40	32.9	1.50	14.6	1.65	18.7	2.15	23.4	3.75	35.2
2.40	25.1	3.10	30.6	2.20	23.6	3.65	34.2	2.45	25.8	1.25	11.6	–	–	–	–

Table 8. The 71 volatility values (%)

6.298	4.401	6.921	7.240	7.204	6.124	9.146	6.003	7.154	3.697	6.561	5.194	5.191
0.103	8.846	8.172	6.850	6.203	4.173	5.793	3.853	8.267	8.317	8.083	5.882	4.719
8.331	6.624	6.318	7.534	4.052	10.649	7.484	3.174	5.176	6.745	5.188	5.993	6.845
6.526	5.395	6.866	7.166	5.138	6.514	6.192	8.133	6.840	6.876	6.218	5.573	10.395
2.965	10.342	6.006	5.343	4.528	2.065	6.919	6.405	7.464	5.208	4.275	7.771	6.782
6.865	4.548	5.947	6.215	3.162	12.337	–	–	–	–	–	–	–

Table 9. Prediction intervals with the confidence level 95%

x_i	1.6	1.7	1.9	2.0
y_i	24.3	25.3	27.9	28.0
$\hat{y}_i$	20.142	21.553	24.142	25.336
Prediction Interval	(15.342, 24.942)	(16.753,26.353)	(19.342,28.942)	(20.536,30.136)
x_i	2.1	2.2	2.4	2.5
y_i	29.6	29.7	32.5	32.6
$\hat{y}_i$	26.472	27.554	29.580	30.530
Prediction Interval	(15.342,24.942)	(16.753,26.353)	(19.342,28.942)	(20.536,30.136)

Step 6 Calculate the interval of the real volume with the confidence level 95% of the 8 sample data by Formula (5). These prediction intervals are shown in Table 9. Here, x_i denotes the sample data; $\hat{y}_i$ denotes the estimated value; y_i denotes the real value.

From Table 9, we know that the estimated values are all in the corresponding interval, which shows the feasibility, validity, reliability, and practicality of VR-TM. So VR-TM can provide a good reference for decisions making. When $\beta = 0.9$, the $[\delta, \beta]$ of the regression function is $[6.608, 0.9]$; When $\beta = 0.95$, the $[\delta, \beta]$ of the regression function is $[6.694, 0.95]$. Obviously, δ in the two description ordered pair is bigger than the ten percent of the max predicted value. So, the reliability of this function is low. Then we should refuse the regression function as the main basis of decision-making. At the same time, it is easy to find that the length of the confidence interval is relatively long. Namely, the reference range is too big, which can reduce the reliability of the parent regression function to a certain degree. For this case, too large reference range means that the sensitive coefficient of total asset turnover to ROE is high. Also it shows its operating risk is high for a new enterprise.

6. **Conclusion.** Regression analysis is a common data analysis tool. It is convenient for people to make right decisions. Only when regression function is reliable, has it better application value. The common test models for reliability are based on that the residual error obeys the normal distribution. When this assumption does not hold in many applications, these test models cannot have application value, which will limit the applied range of regression function greatly. Our VR-TM makes up for its shortcoming to a large extent. Theoretical analysis and example calculation show that VR-TM not only has good structure and interpretability, but also extends and perfects the existing regression test methods. Also, VR-TM has a deficiency of complex computation. Our further work is to structure the application process of model so as to facilitate the users.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (71371064) and the Natural Science Foundation of Hebei Province (F2015208099, F2015208100).

REFERENCES

[1] R. K. Jain, K. M. Smith et al., Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy, *Applied Energy*, vol.123, pp.168-178, 2014.

[2] F. Arila, M. Mora, M. Ogarce, A. Zaniga and C. Fredes, A method to construct fruit maturity color scales based on support machines for regression: Application to olives and grape seeds, *Journal of Food Engineering*, vol.162, pp.9-17, 2015.

[3] Z. S. Zolfaghari, M. Mohebbi and M. Najarigah, Application of fuzzy linear regression method for sensory evaluation of fried donut, *Applied Soft Computing*, vol.22, pp.417-423, 2014.

[4] A. Donnelly, B. Missteur and B. Broderick, Application of nonparametric regression methods to study the relationship between NO_2 concentrations and local wind direction and speed at background sites, *Science of the Total Environment*, vol.409, pp.1134-1144, 2011.

[5] Y. Guo, E. Nazarian et al., Hourly cooling load forecasting using time-indexed ARX models with two-stage weighted least squares regression, *Energy Conversion and Management*, vol.80, pp.46-53, 2014.

[6] D. G. Kleinbaum and L. L. Kupper, *Applied Regression Analysis and Other Multivariate Methods*, Duxery Press, Boston, Massachussetts, 1978.

[7] F. C. Li, C. X. Jin, Y. Shi and K. Yang, Study on quasi-linear regression methods, *International Journal of Innovative Computing, Information and Control*, vol.8, no.9, pp.6259-6270, 2012.

[8] R. J. Jiang, C. F. Chen and S. Q. Zheng, The non-linear fitting method to analyze the measured M-S plots of bipolar passive films, *Electrochimica Acta*, vol.2, pp.2498-2504, 2010.

[9] B. J. Isaac, J. N. Thornock, J. Sutherland, P. J. Smith and A. Parente, Advanced regression methods for combustion medelling using principal components, *Combustion and Flame*, vol.162, pp.2592-2601, 2015.

[10] P. Barodi, F. D. Maio, P. Turati and E. Zio, Robust signal reconstruction monitoring of industrial components via a modified Auto Associative Kernel Regression method, *Mechanical Systems and Signal Processing*, vol.60, pp.29-44, 2015.

[11] S. Muzziodi, A. Ruggieri and B. De Bates, A comparison of fuzzy regression methods for the estimation of the implied volatility smile function, *Fuzzy Sets and Systems*, vol.266, pp.131-143, 2015.

[12] X. H. Xu, S. W. Wang and Y. M. Chen, An improvement to frequency-domain regression method for calculation conduction transfer of building walls, *Applied Thermal Engineering*, vol.28, pp.661-667, 2008.

[13] W. Pouliot, Robust tests for change in intercept and slope in linear regression models with application to manager performance in the mutual fund industry, *Economic Modeling*, vol.58, pp.523-534, 2016.

[14] B. V. Gnedenko, *Probability Theory Course*, Higher Education Press, Beijing, 1956.