Computer Simulates the Effect of Internal Restriction on Residuals in Linear Regression Model with First-order Autoregressive Procedures

Mei-Yu Lee*

Abstract

This paper evidences how the factors including the non-normal error distribution, constraints of the residuals, sample size, the multi-collinear values of independent variables and autocorrelation coefficient have impact on the distributions of the errors and the residuals to explain that the residuals become more centralized as normal distribution when linear requirement have more constraints on residuals from the linear regression analysis method, but less linear requirement cause that the shape of error distribution are more clearly shown on residuals. The paper finds out that if the errors are normal distribution, then the residuals are also normal distribution, but if the errors are U-quadratic distribution, then the residuals are the mixture of the error distribution and normal distribution because of the interaction of linear requirement and sample size. Thus, the increasing constraints on the residual from more independent variables causes the residuals become normal distribution and cannot be a good estimator of the errors which are non-normal distribution. Only the sample size is larger enough to eliminate the effect of linear requirement and multi-collinearity, the residuals can be viewed as an estimator of the errors.

Keywords: Computer simulation, autoregresive model, linear requirement

* Corresponding author. Department of Applied Finance, Yuanpei University. NO. 306, Yuanpei Street, Hsinchu, 30015, Taiwan. Mail: mylee@gm.ypu.edu.tw, TEL: +886-3-6102366, FAX: +886-3-6102367.

1. INTRODUCTION

The paper focuses on why the business studies always use linear regression model but the engineering and quality management fields do not rely on the linear regression model when the researchers found out that the residual distribution is not the same as the error distribution, that obeyed the result of Box and Peirce (1970) who supposed that the residuals have a good fit should be the true errors and can be regarded as estimator of the errors in autoregressive process. The residuals that are viewed as a good estimator of the errors such as Durbin-Watson test statistic (Durbin and Watson, 1950,1951) and LaGrange Multiplier test statistic (Berusch and Pagan, 1980) in the linear regression model with the autoregressive error process.

In general, residuals are viewed as the highest representation of errors and are combined as estimator that is used to estimate the properties of errors such as serial correlation of errors or the error's distribution. In the literature, Yule (1921) was the pioneer to discuss the problem of serial correlation, then Roos (1936) provided the basic solutions about how independent variables are independent by the use of choosing lagged time and how the trend and fluctuation can be grabbed. Box and Pierce (1970) also investigated that residual autocorrelation can be approximated to the linear transformation of error autocorrelation and possess a normal distribution. Therefore, the residuals' distribution plays an important role that linear regression model argues cogently residuals can be used to test the distribution of error which is assumed at first. For example, errors follows normal distribution if residuals are tested as normal distribution. However, the time series data might have non-normal error distribution which contracts with one of assumptions in the linear regression model, and have autoregressive error procedure which contracts with the assumption of linear regression model which is independent errors with each other.

In fact, the paper intends to explain that the distributions of the residuals are away from the distributions of the errors by using a probability simulation approach, that is running computer simulation with random number table from uniform distribution with 0 and 1, from three parts to investigate that (1) if the errors are distributed as normality, then the residuals are also normal distribution; (2) if the errors are non-normal distribution, then the residuals are also non-normal distribution, and (3) if the sample size becomes larger, then the residuals distribution is approached to the errors distribution, that is the law of large number gradually has the influence on the distributions of the residuals. The second part is based in an example of the U-quadratic distributed errors. If augment of Box and Pierce were right, then the residuals should be U-quadratic distribution, but not normal distribution. However, the calculated residuals are represented as normal distribution and are contracted with Box and Pierce (1970). The paper discovers from the results of computer simulation that the normality of the residuals results from the number of independent variables which decides the constraints of the residuals, $\mathbf{X}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$, with 1 plus the number of independent variables. The constraints of the residuals, $\mathbf{X}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$ is called as linear requirement of linear regression models and its number of constraints is the same as the degree of freedom Lee (2014a, b, c). The third part shows that how the change of the residual distribution with fixed number of independent variables and U-quadratic error distribution when the sample sizes become larger.

The paper is structured as follows. Section 2 describes the model setting and simulation procedure. Section 3 results the three cases that the error is normal distribution, the error is U-quadratic distribution, and the sample size is changed. Section 4 concludes.

2. MODEL

Consider a linear regression model with *k* independent variables and *T* sample sizes, as

$$\underset{(T\times 1)}{\mathbf{Y}} = \underset{(T\times k)}{\mathbf{X}}\underset{(k\times 1)}{\boldsymbol{\beta}} + \underset{(T\times 1)}{\boldsymbol{\varepsilon}}$$

on the conditions of $E(\boldsymbol{\varepsilon}) = 0, E(\mathbf{X}^T\boldsymbol{\varepsilon}) = 0$ and the first-order autoregressive procedure is

$$\varepsilon_{t+1} = \rho\varepsilon_t + \mu_t,$$

where $t = 1,2,\dots,T-1$ and $|\rho| < 1$. The estimators are

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} \quad \text{and} \quad \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}},$$

then the residuals are

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\varepsilon} = \left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\boldsymbol{\varepsilon},$$

and linear requirement are $\mathbf{X}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$ where $\mathbf{Y} = \hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}$ (Baltagi 2011).

When $\hat{Y}_i$ is normal distribution, the regression coefficient point estimators also are normal distribution and $\hat{\varepsilon}_i$ is normal distribution, $\hat{\mathbf{Y}}$ approximates to normal distribution, if *k* is enough large (The linear combination of independent variables). The probability distribution of $\hat{\mathbf{Y}}$ is transferred from the error probability distribution and the estimated condition of independent variable coefficient, then the mean-square-error (MSE) is

$$MSE = \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$$

that is from $E(\hat{\boldsymbol{\varepsilon}}) = 0$ and $E(\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}^T) = (\boldsymbol{X}^T\boldsymbol{X})^{-1} \times MSE$. Thus, there are two main factors that affect the probability distribution of the residuals, one is the assumption of the error distribution, the other is the linear requirement of $\mathbf{X}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$, which has *k+1* constraints. Besides, the above sample size, multi-collinear values of independent variables and autocorrelation coefficient affect the distribution of the residuals. However, the distribution of the residuals in the above model is difficult to formula, the paper only use computer simulation to evidence how the factors affect the relation between the errors and the residuals.

2.1. The simulator method

The sampling distribution of test statistic might be existed or non-existed, especially, some sampling distribution of test statistic cannot be transferred by the traditional mathematical method such calculus method or Monte Carlo method. The concept of Monte Carlo method is good simulation method and the continuous type data cannot be done in computer program. Thus, this simulation method is not suitable for this paper because the probability theory has been created by the basic concept of probability simulator and then the distribution functions of continuous random variables can be transferred from uniform distribution with parameters of 0 and 1, *U(0,1)*. This simulation method may be useful in these situations, but the probability distribution simulator is not created now and this simulated skill is not easy now. Thus, the paper runs computer simulation with a software program that can works any probability distribution transformation, generates the data and computing the coefficients and images.[1] The computer simulation is from the steps as follows.

(1) generating data from random number table of *U(0,1)*. Each value of U-quadratic distribution can be gotten when the value is from the inverse function of U-quadratic distribution.

(2) Collecting the values that number is match number of error, of course, the values follow independently identically distributed U-quadratic distribution. That is the set values as μ_t from the serial correlation model where

$$E(\mu_t)=0, Var(\mu_t)=\sigma^2, t=1,\ldots T,, \mu_1,\ldots\mu_T \overset{iid}{\sim} N(0,\sigma^2),$$
$$\varepsilon_{t+1}=\rho\times\varepsilon_t+\mu_{t+1}, t=1,2,\ldots,T-1, |\rho|<1.$$

When the ρ is known, the ε_{t+1} is be found.

(3) The residuals is followed the point estimated requirements of linear model.

$$Y_t=\beta_0+\beta_1\times X_{1,t}+\ldots+\beta_k\times X_{k,t}+\varepsilon_t, t=1,\ldots T,$$

If number and value of independent variables are known, the $\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_k$ can be

[1] The software program is named as "White model I" that can be download from http://goo.gl/oUDpsp. The distributions of the T^{th} error and residual, the distributions of sum of the errors and residuals can be simulated by the software. The simulation technology is from C.C.C. Ltd. (http://psccc.com.tw/en/product). And the U-quadratic distribution formula can reference at http://psccc.com.tw/uploads/files/probability/1/Chapter_one_02.pdf.

estimated. $\mathbf{X}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$ will be ruling the $\hat{\boldsymbol{\varepsilon}}$. The simulator is obeyed the linear model method and created the residual values ($\hat{\boldsymbol{\varepsilon}}$).

Thus, the paper is following the simulation process that is included as

Step 1: Giving the intercept and slope value and the data set of independent variables.

Step 2: Using the simulation method to get the error data set from probability distribution which sample size is T.

Step 3: According the linear model and computing the data set of dependent variable. $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

Step 4: Calculating the point estimator values of regression coefficient and getting the estimated values of dependent variable. $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

Step 5: Calculating the data set of residual. $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$.

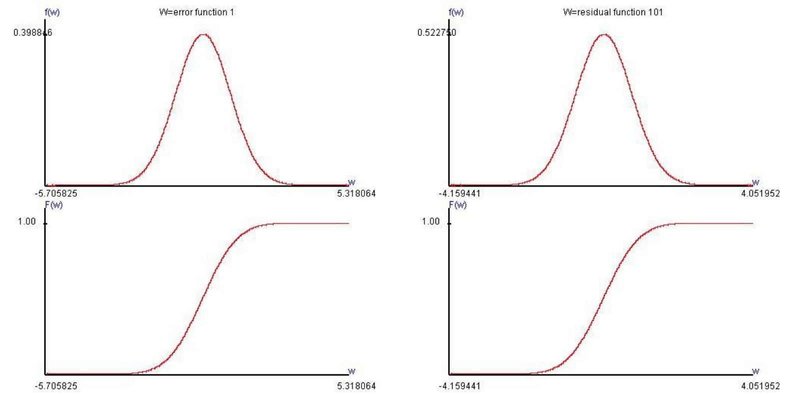
ε_t is simulated by the Step 1 and 2 and $\hat{\varepsilon}_t$ is simulated by the Step 1 to 5. ε_t and $\hat{\varepsilon}_t$ are all repeated 32768*2*1024 times and then produced 32768*2*1024 values to form the frequency distribution that can reach to the real ε_t and $\hat{\varepsilon}_t$ distributions.

3. RESULTS

3.1. The errors follow normal distribution

The section 3.1 is on the condition that the distribution of the errors are normal distribution, 6 independent variables, 15 samples and the autocorrelation coefficient of the errors is zero, thus, the first column of Figure 1 is the distribution of the errors, which is normal distribution with standard normal distribution. Figure 1 of the paper illustrates the shape and coefficient of the 7th residual distribution where the coefficients of mean, skewness and kurtosis represents as normal distribution and as the same as the error distribution.

f(W1),F(W1), W1=error(1)	f(W17),F(W17) ,W17=residual(7)



Mathematical Mean:	0.00004	Mathematical Mean:	-0.00003
Variance :	0.99995	Variance :	0.58319
S.D. :	0.99997	S.D. :	0.76367
Skewed Coef. :	-0.00034	Skewed Coef. :	0.00002
Kurtosis Coef. :	3.00023	Kurtosis Coef. :	3.00068

Figure 1. The errors and residuals distributions when the errors follow normal

distribution

The residuals can be viewed as the estimator of the errors because Figure 1 guarantees the distribution of the residuals is the same as the distribution of the errors. Thus, Figure 1 evidences that the conclusion of Box and Pierce (1970) is correct when the errors are normal distribution. The reason is that as follows. The residuals are the combination of the errors in the linear regression model, that is $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{X}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \boldsymbol{\varepsilon}$, and the autoregressive procedure takes the errors toward the autocorrelation with each other without changing the property between the residuals and the errors, at the same time, the additive property of normal distribution has impact on the linear combination of the errors, thus, the residuals can show the normal distributed property of the errors. Another reason is that the mathematical formula of normal distribution includes *sin* and *cos* functions that is cycle functions, therefore, the residuals follow normal distribution when the errors' distribution is normal distribution. On the other hand, if the error distribution has no property of *sin* and *cos* functions, such as logistic, uniform, U-quadratic or exponential distribution, then the errors have no cycle property and might not be shaped as normal distribution, but other distributions. The paper can evidence the proposition below.

Proposition 1.

The residuals are normal distribution when the errors are normal distribution because the normal distribution with sin and cos function form has additive property.

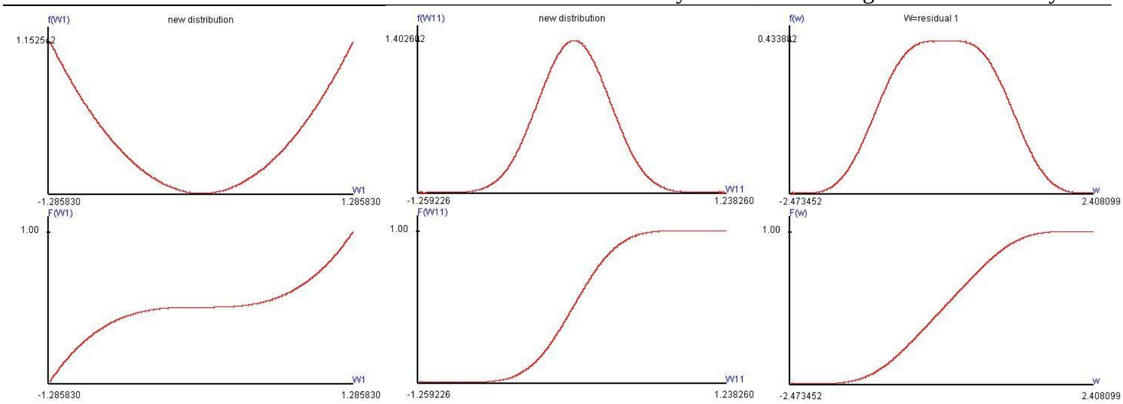
However, it is hard to promise that the errors are always normal distribution when the researchers do not have the population data. The samples should be tested to classify what distribution it follow.

3.2. The errors follow U-quadratic distribution

The paper gives an example of non-normal error distribution that the errors, μ_t, were U-quadratic distribution, then the computer simulation offers evidence about the distributions of the errors and residuals on the condition of 6 independent variables, 15 samples, the 1 lagged period, variance of error is 1 and the autocorrelation coefficient of the errors is zero. Lee (2014c) supposed that the values of independent variables have serious impact on the residuals, thus the paper simultaneously discusses the distributions of the residuals at two cases where the values of independent variables are separately with low and high multi-collinearlity in the linear regression model with the first-order autoregressive error procedure. The third column of Figure 2 presents the distribution of the first residual that is generated from the independent variables with the population correlation coefficient is 0.99.

Figure 2 shows that the first column is the error distribution which is U-quadratic distribution, the second column is the distribution of the first residual with low multi-collinearity and the third column is the distribution of the first residual with high multi-collinearity. The distributions of the first residual in the second and third column is as similar as normal distribution while the distribution of the errors is U-quadratic distribution. The residual distributions in Figure 2 are different from the error distribution, thus, the residuals cannot be regarded as an estimator of the errors when the errors are non-normal distribution, moreover, the serial correlation test for autocorrelation of the errors is not suitable to use the mathematic combination of the residuals because the difference between the distributions of the errors and residuals. The researchers should first investigate the distribution of the data to classify what distribution the data is or they always obtain the result that the errors follow normal distribution from the residuals.

f(W1),F(W1), W1=error(1)	f(W11),F(W11) ,W11=residual(1) with low multi-collinearity	f(W11),F(W11) ,W11=residual(1) with high multi-collinearity



Mathematical Mean:	0.00071	Mathematical Mean:	-0.00004	Mathematical Mean:	-0.00005
Variance :	0.99998	Variance :	0.07704	Variance :	0.58928
S.D. :	0.99999	S.D. :	0.27757	S.D. :	0.76764
Skewed Coef. :	-0.00125	Skewed Coef. :	0.00041	Skewed Coef. :	-0.00003
Kurtosis Coef. :	1.19053	Kurtosis Coef. :	2.82251	Kurtosis Coef. :	2.31065

Figure 2. The error and residual distributions when error follows U-quadratic distribution (T=15)

The most difference between Figure 1 and Figure 2 is the assumption of the error distribution, but the residuals are similar normal distribution in Figure 2. There must be some special factors that did not be discovered before and the factor is certainly not from the error distribution which has the property of sin and cos functions or the property of addition from normal distribution. Lee (2014b) discovered that the number of independent variables from 1 to 6 causes the residual distribution toward normal distribution in the linear regression model with autoregressive procedure, therefore, the paper supposes that the errors are not restricted but the residuals are restricted by the restriction of $\mathbf{X}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$ which has *k+1* constraints. This linear requirement of the residuals in linear regression model distorts the error distribution away from the original distribution to normal distribution.

The linear requirement of the residuals can produce two forces out when the number of independent variables is k and sample size is T. The first part is sum of square in regression (SSR) whose degree of freedom is k. The second part is sum of square of error (SSE) whose degree of freedom is T-k-1. Two forces decide the shape of the residual distribution. When k becomes larger the residuals are restricted by more constraints just alike the linear combination of random variables. If the more random variables are added in the linear combination then the new random variable become toward to normal distribution. The residuals that are restricted by more constraints also represent the similar states so that the distribution of the residuals become similar to normal distribution.

If consider the collinear values of independent variables, then collinear effect affects the convergence to normal distribution when the error distribution is U-quadratic distribution. The three columns of Figure 2 show the residuals with the convergence to normality will be weaken by the high multi-collinear values of independent variables when the number of independent variables is fixed. The coefficients of Figure 2 express the variance of the residuals, $S(\hat{\boldsymbol{\varepsilon}})$, and mean-square of error (MSE) rise up when the values of independent variables are from low to high multi-collinearity. The shape of distribution in third column of Figure 2 presents that the distribution of the first residual has a flat region around the mean of the first residual and has more left-skewness and less centralization. Thus, the high mutli-collinarity brings a big problem of the residuals, so the distribution of the first residual is not the same as the shape of the distribution in the second column of Figure 2.

The multicollinearity of the independent variables results in that the calculation and the combination of the residuals are more complex, moreover, the variance and the distribution of the residuals is both disturbed by the collinearity when the errors follow non-normal distribution whose errors is asymmetric will cause that the residuals are slowed down the convergence to normality. The improvement of above problem is that the linear regression model adds more independent variables to accelerate the residuals' convergence to normality because more and more independent variables can restrict the residuals then weak the collinear effect, meanwhile, the degree of freedom also falls down.[2]

The paper proposes the second proposition as follows.

[2] In fact, the multicollinear case of the computer simulation implies that only the number of independent variables is more than 20 and the degree of freedom is more than 2, then the residuals and coefficients of regression model with first-order autoregressive error process will have normal distributed point estimators when the low multicolliearity of independent variables exists and the errors follow independently identically distribution with symmetric at zero.

Proposition 2.

(1) ***The larger number of independent variables, k, brings to the faster convergence to normality on the residuals when the errors are non-normal distribution.***

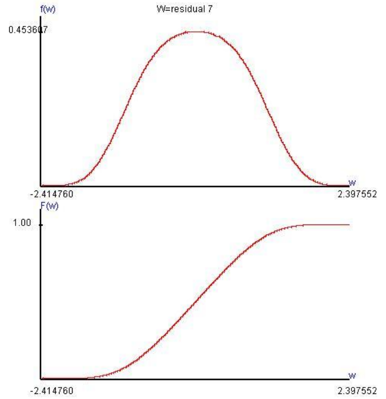
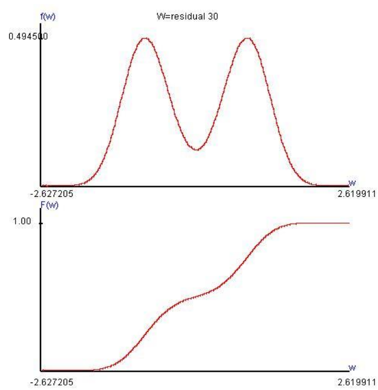
(2) ***The higher multi-collinear values of independent variables brings to the slower convergence to normality on the residuals.***

(3) ***The high multi-collinear values and small number of independent variables cause that the distribution of the residuals is not normal distribution nor the error distribution, but mixture of the error distribution and normal distribution by the constraints of the residuals.***

The multi-collinear property and the constraints on the residuals have opposite force to simultaneously disturb the distribution of the residuals which becomes a mixed distribution between normal distribution and the error distribution. However, there is one factor that the sample size, T, in the internal constraints of linear regression model that have two forces out, is fixed and is not discussed.

3.3. Sample size is changed

The sample size effect plays very important role in time series models that can represent the law of large number, then the residuals might be regarded as a good estimator of the errors. The simulation case only change the sample size from 9 to 107 on the condition of 6 independent variables, population variance is 1, zero autocorrelation coefficient, and the values of independent variables are from the front T of data set and the simulated setting is followed section 3.2. The residuals can gradually reveal the property of the errors when the sample sizes are increasing from 9 to 107. Without loss of generation, the paper only shows the shapes and coefficients of distributions of $0.5 * T^{th}$ residual in Figure 3 and Table 1 that is in Appendix A.

T=14, the 7^{th} residual		T=60, the 30^{th} residual	
			
Mathematical Mean:	-0.00003	Mathematical Mean:	0.00014

Variance	:	0.50400	Variance	:	0.86864
S.D.	:	0.70993	S.D.	:	0.93201
Skewed Coef.	:	-0.00030	Skewed Coef.	:	-0.00019
Kurtosis Coef.	:	2.47090	Kurtosis Coef.	:	1.63269

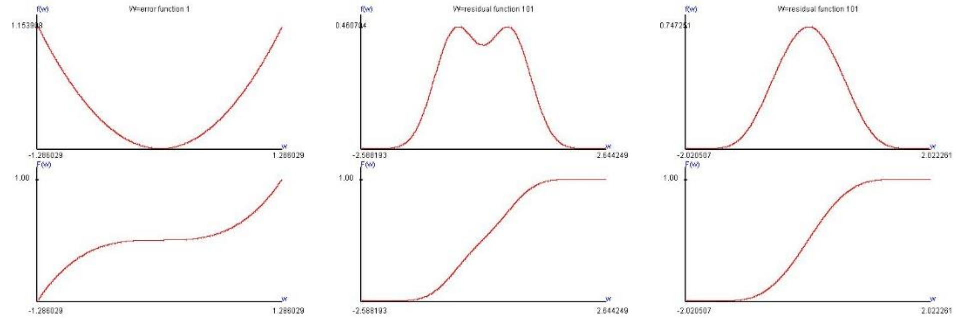
Figure 3. the $0.5T^{th}$ residual's distribution

Figure 3 explores the distributions of the $0.5 * T^{th}$ residual at T=14 and 60. The second column of Figure 3 has less regular shape of distribution than the first column at T=14, meanwhile, the 30^{th} residual at T=60 gets the shape of distribution more toward U-quadratic distribution than distribution of the 7^{th} residual at T=14. Figure 3 clearly shows that the large sample sizes cause the $0.5T^{th}$ residual toward the error distribution. Thus, the sample size effect can delimitate the effect of linear requirement on residuals, at the same time, can let the distribution of the residuals represent more property of the errors. The coefficients in Table 1 show that the $0.5 * T^{th}$ residual has around zero means and skewed coefficients, and larger variances and falling down kurtosis when the sample sizes are from 9 to 107. In comparison of the first column in Figure 2, the coefficients in Table 1 are approaching to the coefficients in the first column of Figure 2, especially variances and kurtosis coefficients, when the sample sizes are increasing. The coefficients that are changed with the different sample sizes also shows the distribution of the residuals can be an estimator of the errors if the sample size effect is larger enough than the effect of constraints on the residuals. The interaction between the sample size and linear requirement causes the different shape and coefficients of distribution of the $0.5 * T^{th}$ residual. The residuals become more centralized when the linear requirement has more constraints, nevertheless, the residuals are affected by the error distribution and are more likely to the error distribution when the number of constraints are not larger enough. In other words, the residuals' distribution is vastly different to the errors' distribution because the residuals are affected by linear requirement of linear regression models, the error distribution, sample sizes and the values of independent variables. It is general assumption that the errors are normal distribution which is symmetric distribution, meanwhile, the residuals can be an estimator of the errors by the residuals are normal distribution when $k+1$ constraints are more than 20, T is from 23 to $23+(k-19)$ and a little or no multi-collinearity.

3.4. Autocorrelation coefficient of the errors is 0.7

The above statements are on the condition of zero autocorrelation of the errors, however, nonzero autocorrelation cases are usually seen in the data. The paper discusses the case that the autocorrelation coefficient of the errors is 0.7 and other conditions are the same as section 3.2. There is also two cases of low and high multicollinear values of independent variables. The paper shows the shapes and coefficients of distribution of first and 15^{th} errors and residuals in Figure 4 and Figure 5.

the 1st error		the 1st residual Low multicollinearity		the 1st residual High multicollinearity	



Mathematical Mean:	0.00006	Mathematical Mean:	-0.00004	Mathematical Mean:	-0.00005
Variance :	0.99999	Variance :	0.52980	Variance :	0.25373
S.D. :	0.99999	S.D. :	0.72787	S.D. :	0.50372
Skewed Coef. :	-0.00012	Skewed Coef. :	-0.00005	Skewed Coef. :	-0.00013
Kurtosis Coef. :	1.19049	Kurtosis Coef. :	2.17951	Kurtosis Coef. :	2.62116

Figure 4. The first error and residual distributions when error follows U-quadratic distribution and autocorrelation coefficient is 0.7

The distribution of the first error is U-quadratic distribution, but the first residuals that are affected by the degree of multicollinearity show different shapes of the distribution of the residuals in Figure 4. The distributions of the first residuals cannot represent the property of the first error and the more multicollinear values of independent variables make more centralized distributions of the first residuals.

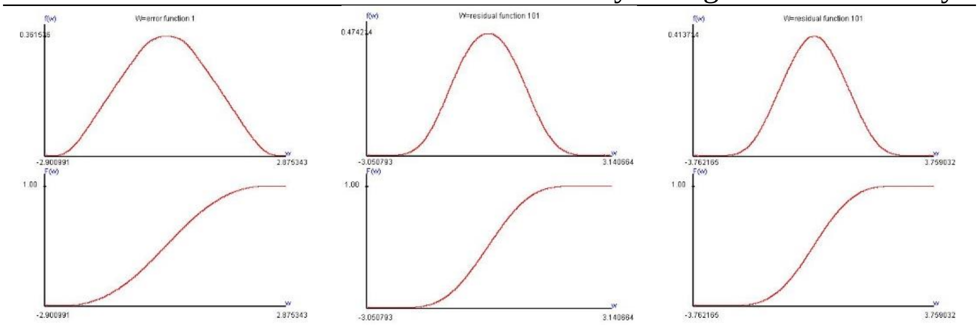
Figure 4 explores another very important point that the coefficients cannot show out. The coefficients are similar to each other among three columns, but the population distribution of the first column and sampling distributions of the second and third columns have completely different shapes with each other. In the other hand, the coefficients judgment might highlight a questionable results that is used to pass the hypothesis testing when the researchers only investigate the means and variances of the residuals. Comparing to the three columns, the linear requirement and multi-collinearity keep aside push to the first residual become more centralization. The higher multi-collinear values of independent variables induce more centralized residuals.

The paper also runs the distributions of the 15th error and residuals which are divided into two parts including low and high multicollinearity in Figure 5.[3] The distribution of the 15th error is not the same as U-quadratic distribution, but more centralized alike normal distribution because of the nonzero autocorrelation coefficient, meanwhile, the distributions of the 15th residual are more similar to the distribution of the 15th error in Figure 5 than the first residual in Figure 4. The means and variances at the first and third columns in Figure 5 highlight that the 15th residual is very similar to the 15th error, however, the diagrams and kurtosis coefficients shows that there is a vast different

[3] The case with -0.7 autocorrelation coefficient is in Appendix B.

between the 15th residual and the 15th error.

the 15th error	the 15th residual Low multicollinearity	the 15th residual High multicollinearity



Mathematical Mean:	-0.00008	Mathematical Mean:	-0.00003	Mathematical Mean:	0.00004
Variance :	0.99992	Variance :	0.60125	Variance :	0.83748
S.D. :	0.99996	S.D. :	0.77540	S.D. :	0.91514
Skewed Coef. :	0.00025	Skewed Coef. :	-0.00008	Skewed Coef. :	0.00026
Kurtosis Coef. :	2.38056	Kurtosis Coef. :	2.56853	Kurtosis Coef. :	2.61272

Figure 5. The 15th error and residual distributions when error follows U-quadratic distribution and autocorrelation coefficient is 0.7

With comparison between Figure 4 and Figure 5, the coefficients of the errors show the similar coefficients, except for kurtosis coefficients. The diagrams of the first column from Figure 4 to Figure 5 are from U-quadratic shape to opposite of U-quadratic shape because of the non-zero autocorrelation coefficient. The diagrams of low multicollinear residuals are bulging from the first residual to the 15th residual, while the diagrams of high mulitcollinear residuals are similar. This highlights that the nonzero autocorrelation coefficient and multicollinearity interact on the residuals. The higher multicollinearity will decrease the effect of autocorrelation coefficient on the residuals.

4. CONCLUSIONS

The purpose of this paper supposes an explanation why the residuals cannot perfectly represent the errors. The paper evidences that the residuals that should be on some specific conditions can be viewed as an estimator of the errors which may not be necessary to be assumed as normal distribution because of the property of data. But if the errors are normal distribution, the residuals can be a good estimator of the errors because of the property of normal distribution.

The paper also evidences how non-normal distribution, linear requirement, multicollinearity, sample size and autocorrelation coefficient affect the distributions of the errors and the residuals by computer simulation results. First, residuals are restricted but errors are not, hence, the values of the residuals are constrained by linear requirement that are from regression analysis, and causes the residuals are not perfectly representing errors. Second, the paper supposes that the linear requirement, sample sizes and the values of independent variables are interacted in the linear regression model. Thus, (1) the constraints of linear requirement are more enough then the residuals follow normal distribution whatever the error term is assumed when the sample sizes are fixed. (2) If the error term is assumed as normal distribution, then residuals follow normal distribution. (3) Larger sample sizes result in the residuals reveal the property of error term when the linear requirement is fixed.

References

[1] B.H. Baltagi, "Econometrics," 5th ed., Springer, 2011.

[2] C.F. Roos, "Annual Survey of Statistical Techniques: The Correlation and Analysis of Time Series--Part II," Econometrica. vol. 4, 1936, pp. 368–381.

[3] G.E. Box, D.A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-integrated Moving Average Time Series Models," Journal of the American Statistical Association. vol. 65, 1970, pp. 1509–1526.

[4] G.U. Yule, "On the Time-Correlation Problem, with Especial Reference to the Variate- Difference Correlation Method," Journal of the Royal Statistical Society. vol. 84, 1921, pp. 497–537.

[5] J. Durbin, G.S. Watson, "Testing for Serial Correlation in Least Squares Regression: I," Biometrika. vol. 37, 1950, pp. 409–428.

[6] J. Durbin, G.S. Watson, "Testing for Serial Correlation in Least Squares Regression. II," Biometrika. vol. 38, 1951, pp. 159–177.

[7] M.Y. Lee, “The Pattern of R-Square in Linear Regression Model with First-Order Autoregressive Error Process and Bayesian property: Computer Simulation,” Journal of Accounting & Finance Management Strategy, forecoming, 2014a.

[8] M.Y. Lee, “Limiting Property of Durbin-Watson Test Statistic,” manuscript, 2014b.

[9] M.Y. Lee, “The Conflict of Residual and Error Simulated in Linear Regression Model with AR(1) Error Process,” manuscript, 2014c.

[10] T.S. Breusch, L.G. Godfrey, "A review of recent work on testing for autocorrelation in dynamic economic models," University of Southampton, 1980.

Appendix A.

Section 3.3 shows how the change of sample size affect the distribution of each residual. However, each different sample size case has T residuals, the paper shows the coefficients of the $0.5T^{th}$ residual whose $0.5T^{th}$ is half T. The coefficients of Table 1 almost have the same means of the $0.5T^{th}$ residual and cannot let readers know what the difference among those distributions of the $0.5T^{th}$ residual from different sample sizes. Thus, the paper put the coefficients of the $0.5T^{th}$ residual in Appendix and the diagrams of distribution of the $0.5T^{th}$ residual in the main content.

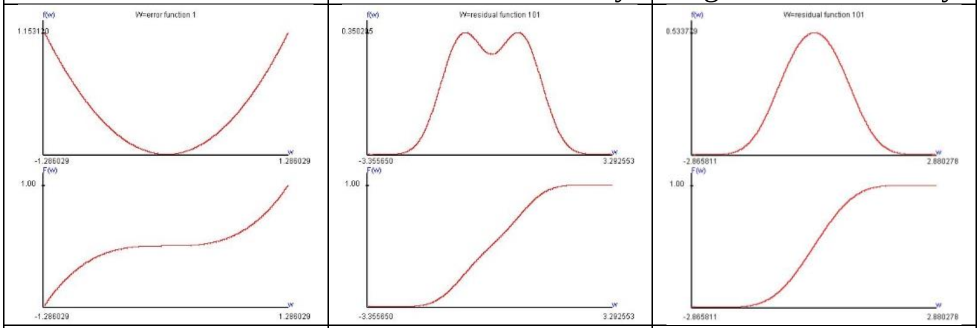
Table 1. The coefficients of the $0.5T^{th}$ residual in different sample sizes

T	The $0.5T^{th}$ residual	Coefficients	T	The $0.5T^{th}$ residual	Coefficients
9	5	Mathematical Mean: 0.00002 Variance : 0.12050 S.D. : 0.34712 Skewed Coef.: 0.00008 Kurtosis Coef.: 2.61692 MAD : 0.28164 Range : 2.18595 Median : 0.10410 IQR : -1.09240	10	5	Mathematical Mean: 0.00010 Variance : 0.66584 S.D. : 0.81599 Skewed Coef. : 0.00007 Kurtosis Coef. : 2.14263 MAD : 0.69024 Range : 4.36179 Median : 0.05433 IQR : -1.67194
12	6	Mathematical Mean: 0.00003 Variance : 0.41115 S.D. : 0.64121 Skewed Coef. : -0.00049 Kurtosis Coef. : 2.52345 MAD : 0.52335 Range : 4.03883 Median : -0.59530 IQR : -0.01178	14	7	Mathematical Mean: 0.00003 Variance : 0.50400 S.D. : 0.70993 Skewed Coef. : -0.00030 Kurtosis Coef.: 2.47090 MAD : 0.58384 Range : 4.96694 Median : -0.49846 IQR : 0.60169
15	1	Mathematical Mean: -0.00005 Variance : 0.33153 S.D. : 0.57579 Skewed Coef. :-0.00019 Kurtosis Coef. : 2.70090 MAD : 0.46598 Range : 4.47935 Median : 0.08346 IQR : -0.49092	16	8	Mathematical Mean:0.00015 Variance : 0.69658 S.D. : 0.83462 Skewed Coef. : 0.00016 Kurtosis Coef.: 2.09787 MAD : 0.71287 Range : 5.26536 Median : 1.13480 IQR : 1.81239
18	9	Mathematical Mean: -0.00007 Variance : 0.67610 S.D. : 0.82225 Skewed Coef. : -0.00021 Kurtosis Coef. : 2.15551 MAD : 0.69899 Range : 5.76724 Median : 0.51382 IQR : 0.72251	20	10	Mathematical Mean:-0.00001 Variance : 0.57802 S.D. : 0.76028 Skewed Coef. : 0.00022 Kurtosis Coef. : 2.35887 MAD : 0.63243 Range : 5.39457 Median : -0.87208 IQR : -0.44459
30	15	Mathematical Mean: -0.00011 Variance : 0.74874 S.D. : 0.86530 Skewed Coef. : 0.00018 Kurtosis Coef. : 1.97507 MAD : 0.75134 Range : 5.66672 Median : -0.36794 IQR : -0.17231	40	20	Mathematical Mean: -0.00011 Variance : 0.90912 S.D. : 0.95348 Skewed Coef.: 0.00022 Kurtosis Coef.: 1.50373 MAD : 0.88326 Range : 4.90625 Median : -0.30457 IQR : -0.51299
80	40	Mathematical Mean: 0.00004 Variance : 0.97787 S.D. : 0.98888 Skewed Coef. : -0.00034 Kurtosis Coef.: 1.26972 MAD : 0.94699 Range : 3.91584 Median : 0.88092 IQR : -2.35631	107	50	Mathematical Mean: 0.00028 Variance : 0.92229 S.D. : 0.96036 Skewed Coef.: -0.00034 Kurtosis Coef.: 1.46089 MAD : 0.89520 Range : 4.92764 Median : 1.02281 IQR : -0.48487

Appendix B.

The paper also simulates the situation that the autocorrelation of the errors is -0.7. The residuals have the smaller mean and variance and more negative skewness on the condition of high multicollinearity.

the 1^{st} error	the 1^{st} residual Low multicollinearity	the 1^{st} residual High multicollinearity

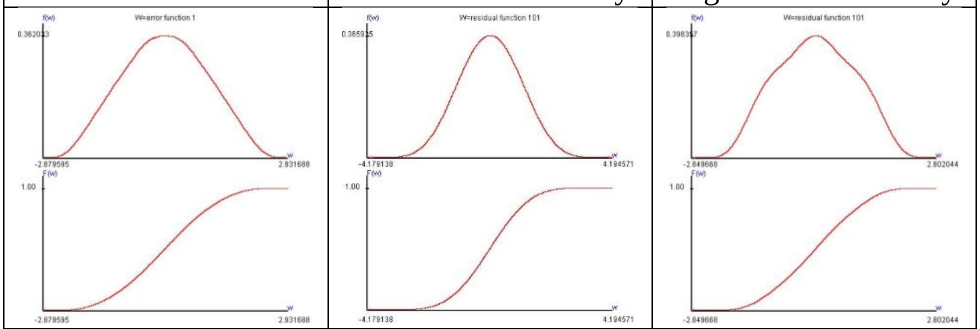


Mathematical Mean: -0.00000 Variance : 0.99996 S.D. : 0.99998 Skewed Coef. : 0.00008 Kurtosis Coef. : 1.19052	Mathematical Mean: -0.00005 Variance : 0.92580 S.D. : 0.96219 Skewed Coef. : -0.00003 Kurtosis Coef. : 2.13261	Mathematical Mean: -0.00016 Variance : 0.49797 S.D. : 0.70567 Skewed Coef. : -0.00069 Kurtosis Coef. : 2.66838

Figure B-1. The first error and residual distributions when error follows U-quadratic distribution and autocorrelation coefficient is -0.7

Comparison of the autocorrelation coefficients of the errors that are 0.7 and -0.7, the means of the first error and the 15^{th} error have less means when autocorrelation coefficient is -0.7. Second, the probability distribution of the first error in Figure B-1 is as the same as in the left side of Figure 4, so does the 15^{th} error in Figure B-2 and Figure 5. Third, the low multicollinearity situation shows that the first residual has larger variance and less centralization in Figure B-1 than in Figure 4. However, the 15^{th} residual is more centralized in Figure B-2 than in Figure 5. Finally, the high multicollinearity situation explores that the first residual in Figure B-1 has larger variance and more centralization than in Figure 4. However, the 15^{th} residual has smaller variance and less centralization in Figure B-2 than in Figure 5.

the 15^{th} error	the 15^{th} residual Low multicollinearity	the 15^{th} residual High multicollinearity



Mathematical Mean: -0.00024 Variance : 0.99998 S.D. : 0.99999 Skewed Coef. : 0.00036 Kurtosis Coef. : 2.38057	Mathematical Mean: -0.00012 Variance : 1.08709 S.D. : 1.04264 Skewed Coef. : 0.00057 Kurtosis Coef. : 2.68820	Mathematical Mean: -0.00017 Variance : 0.82336 S.D. : 0.90739 Skewed Coef. : 0.00021 Kurtosis Coef. : 2.27220

Figure B-2. The 15^{th} error and residual distributions when error follows U-quadratic distribution and autocorrelation coefficient is -0.7