MPRA

Munich Personal RePEc Archive

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

Mei-Yu Lee

Department of Applied Finance, Yuanpei University

2014

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

Mei-Yu Lee[1]

Abstract

This paper demonstrates the impact of particular factors – such as a non-normal error distribution, constraints of the residuals, sample size, the multi-collinear values of independent variables and the autocorrelation coefficient – on the distributions of errors and residuals. This explains how residuals increasingly tend to a normal distribution with increased linear constraints on residuals from the linear regression analysis method. Furthermore, reduced linear requirements cause the shape of the error distribution to be more clearly shown on the residuals. We find that if the errors follow a normal distribution, then the residuals do as well. However, if the errors follow a U-quadratic distribution, then the residuals have a mixture of the error distribution and a normal distribution due to the interaction of linear requirements and sample size. Thus, increasing the constraints on the residual from more independent variables causes the residuals to follow a normal

---

[1] Department of Applied Finance, Yuanpei University, Hsinchu, Taiwan.
  E-mail: mylee@gm.ypu.edu.tw

distribution, leading to a poor estimator in the case where errors have a non-normal distribution. Only when the sample size is large enough to eliminate the effects of these linear requirements and multi-collinearity can the residuals be viewed as an estimator of the errors.

Mathematics Subject Classification: 37M10; 62M10; 37M05; 68U20

Keywords: Time series; Autoregressive model; Computer simulation; Non-normal distribution

1   Introduction

It is reasonable to question why business studies always use linear regression models, but the engineering and quality management fields do not rely on such models. Researchers face a residual distribution different to the error distribution, and this diverges from the result of Box and Pierce (1970). Box and Pierce suppose that residuals that have a good fit should be the true errors and can be regarded as estimators of the errors in an autoregressive process. Thus, some residuals that are viewed as good estimators of errors include the Durbin-Watson test statistic (Durbin and Watson, 1950, 1951) and the LaGrange Multiplier test statistic (Berusch and Pagan, 1980) in the linear regression model with an autoregressive error process.

In general, residuals are viewed as the highest representation of errors and are combined as an estimator that is used to estimate the properties of errors such as the serial correlation of errors or the error distribution. In the literature, Yule (1921) first discusses the problem of serial correlation, and then Roos (1936) provides basic solutions regarding 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 note that residual autocorrelation can be approximated

to the linear transformation of error autocorrelation and possesses a normal distribution. Therefore, the residual distribution plays an important role – residuals can be used to test the initially assumed error distribution. For example, the errors have a normal distribution if the residuals are identified as having a normal distribution. However, time series data may have a non-normal error distribution which contradicts an assumption in the linear regression model. Then, the autoregressive error procedure contradicts the assumption of the linear regression model regarding the independence of errors.

This paper aims to explain that the distributions of the residuals are separate from the distributions of the errors by using a probability simulation approach. That is, we run a computer simulation with a random number table with a uniform distribution between 0 and 1 to investigate the following three points: (1) if the errors are distributed normally, then the residuals also have a normal distribution, (2) if the errors have a non-normal distribution, then the residuals also have a non-normal distribution, and (3) as the sample size becomes larger, the residual distribution approaches to the error distribution, that is, the law of large numbers gradually has influence over the distribution of the residuals. The second point is based on an example of the U-quadratic distributed errors. If the argument of Box and Pierce were right, then the residuals should have a U-quadratic distribution, and not a normal distribution. However, the calculated residuals appear to have a normal distribution, contradicting Box and Pierce. This paper presents computer simulation results showing that the normality of the residuals results from the number of independent variables that determine 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, $\mathbf{X}^T \hat{\boldsymbol{\varepsilon}} = \mathbf{0}$, are referred to as the linear requirements of the linear regression model and its number of constraints is the same as the degree of freedom (Lee, 2014a, 2014b, 2014c). The third point refers to the change in the residual distribution with a fixed number of independent variables and a 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 gives the results of the three cases where the error has a normal distribution, the error has a U-quadratic distribution, and the sample size is changed. Section 4 concludes this paper.

2   Model and simulation procedures

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

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

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

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

where $t = 1,2,\cdots,T-1$ and $|\rho| < 1$. The estimators are $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}$ and $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, 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 $\hat{Y}_t$ is normal distribution, the regression coefficient point estimators also are normal distribution and $\hat{\varepsilon}_t$ is normal distribution, $\hat{\mathbf{Y}}$ approximates to normal distribution, if k is enough large. We calculate the mean-square-error (MSE) as

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

from $E(\hat{\boldsymbol{\varepsilon}}) = 0$ and $E(\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}^T) = \left(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\right)^{-1} \times MSE$. Thus, the two main factors that

affect the probability distribution of the residuals are the assumption of the error distribution and the linear requirement of $\mathbf{X}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$, which has $k+1$ constraints. Moreover, the sample size, multi-collinear values of independent variables and the autocorrelation coefficient affect the distribution of the residuals. However, as the distribution of the residuals in the above model is difficult to formulate, this paper only uses computer simulations to show how factors affect the relation between the errors and the residuals.

2.1 The simulator method

The sampling distribution of a test statistic may or may not be known. In particular, some sample distribution of the test statistic cannot be transferred using traditional mathematical techniques such as calculus or Monte Carlo methods. The concept of the Monte Carlo method is a good simulation method, but the use of continuous data isn't possible in our computer program. To manage the continuous normal distribution (or U-quadratic distribution), we run a computer simulation using a software program that can work with any probability distribution transformation. The probability theory of this paper has been created using the basic concept of a probability distribution simulator, and the functions of continuous random variables can be transferred from a uniform distribution with parameters of 0 and 1, $X \sim U(0,1)$. Thus, we can generate the data and compute the coefficients and images.[2] The computer simulation is based on the following steps.

---

[2] The software program is named as "White model I" that can be download from http://goo.gl/oUDpsp. The distributions of the Tth 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.

(1) Generate data from a random number table of $U(0,1)$. Each value of the U-quadratic distribution can be obtained when the value is from the inverse function of a Normal distribution (or U-quadratic distribution).

(2) Collect the values whose number matches the number of errors. Clearly, these values are i.i.d. Normal distribution (or U-quadratic distribution). That is the set values of μ_t from the serial correlation model where

$$E(\mu_t) = 0, Var(\mu_t) = \sigma^2, t = 1,...,T,$$

and $\mu_1,...,\mu_T \overset{iid}{\sim} N(0,\sigma^2)$ or $\mu_1,...,\mu_T \overset{iid}{\sim} U\text{-quadratic distribution,}$

$$\varepsilon_{t+1} = \rho \times \varepsilon_t + \mu_{t+1}, t = 1,2,....,T-1, |\rho| < 1.$$

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

(3) The residuals follow the point estimate requirements of the linear model.

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

If the number and values of independent variables are known, then $\hat{\beta}_0, \hat{\beta}_1,..., \hat{\beta}_k$ can be estimated, whereas $\hat{\boldsymbol{\varepsilon}}$ is constrained by $\mathbf{X}^T \hat{\boldsymbol{\varepsilon}} = \mathbf{0}$. Meanwhile, the simulator obeys the linear model method and creates the residual values ($\hat{\boldsymbol{\varepsilon}}$).

Thus, the simulation process includes:

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 the probability distribution with sample size, T.

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

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

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

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 simulated totally 32768　2　1024 times, and then generated 32768　2　1024 values, respectively, to form the frequency distribution that can reach to the real ε_t and $\hat{\varepsilon}_t$ distributions.

3　Main Results

3.1 The errors follow normal distribution

Section 3.1 discusses the condition that the distribution of the errors is a normal distribution with 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 a standard normal distribution, and the second column of Figure 1 illustrates the shape and coefficient of the 7th residual distribution where the coefficients of the mean, skewness, and kurtosis represent a normal distribution and are the same as the error distribution.

The residuals can be viewed as an estimator of the errors because Figure 1 guarantees that the distribution of the residuals is the same as the distribution of the errors. Thus, Figure 1 supports the conclusion of Box and Pierce (1970) when the errors are normally distributed. The reason is as follows: the residuals are a combination of the errors in the linear regression model, that is $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon}$, 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 a normal

distribution impacts the linear combination of the errors. Thus, the residuals show the normally distributed property of the errors.

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

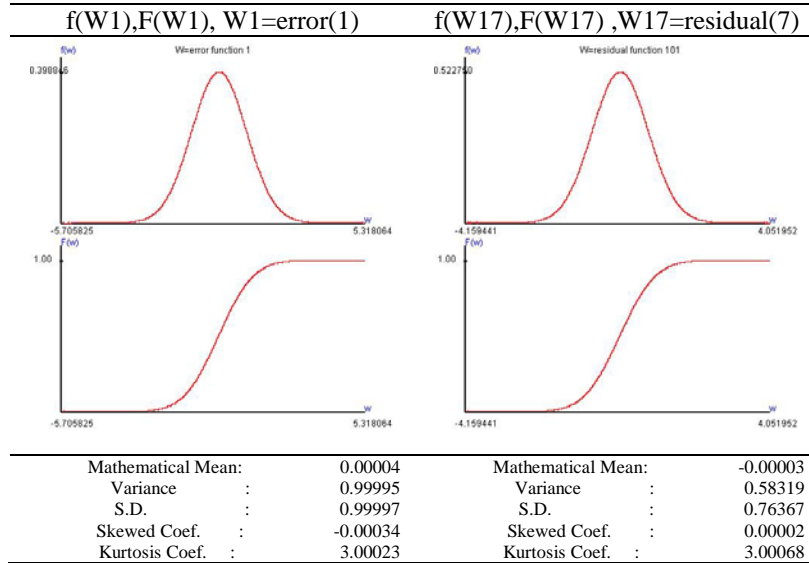| Mathematical Mean:		0.00004	Mathematical Mean:		-0.00003
Variance	:	0.99995	Variance	:	0.58319
S.D.	:	0.99997	S.D.	:	0.76367
Skewed Coef.	:	-0.00034	Skewed Coef.	:	0.00002
Kurtosis Coef.	:	3.00023	Kurtosis Coef.	:	3.00068

Figure 1: The errors and residuals distributions as the errors are normal distribution

Another reason is that the mathematical formula of the normal distribution includes *sin* and *cos* functions, which are cyclical functions. Therefore, the residuals follow the normal distribution when the error distribution is a normal distribution. On the other hand, if the error distribution has no properties of *sin* or *cos* functions, such as logistic, uniform, U-quadratic, or an exponential distribution, then the errors have no cyclical property and may not be shaped as per the normal distribution, but like other distributions instead. This paper confirms the proposition below.

Proposition 1 *The residuals are normally distributed when the errors are normally distributed because a normal distribution with a sin and cos function form is subjected to the additive property.*

However, it is hard to promise that the errors will always be normally distributed when researchers do not have the population data. Samples should always be tested to classify what distributions they 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 columns are 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.

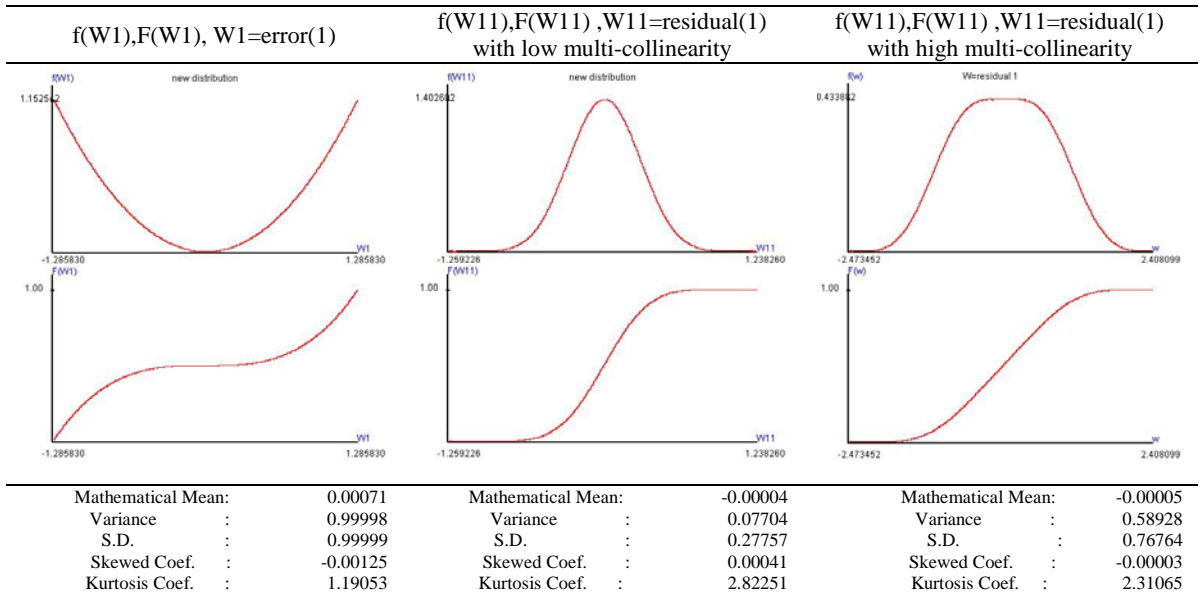| f(W1),F(W1), W1=error(1)		f(W11),F(W11) ,W11=residual(1) with low multi-collinearity		f(W11),F(W11) ,W11=residual(1) with high multi-collinearity	
Mathematical Mean:	0.00071	Mathematical Mean:	-0.00004	Mathematical Mean:	-0.00005
Variance :	0.99998	Variance :	0.07704	Variance :	0.58928
S.D. :	0.99999	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 main difference between Figure 1 and Figure 2 is the assumption of the error distribution, but the residuals show a normal distribution in Figure 2. There must be some special factors not yet discovered, independent of the error distribution which has the property of *sin* and *cos* functions or the property of addition from the normal distribution. Lee (2014b) discovered that the number of independent variables from 1 to 6 causes the residual distribution to tend towards a normal distribution in the linear regression model with an autoregressive procedure. Therefore, the paper supposes that errors are not restricted, but that the residuals are restricted by $\mathbf{X}^T\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$ which has $k+1$ constraints. This linear requirement of the residuals in the linear regression model distorts the error distribution away

from the original distribution to a normal distribution.

The linear requirement of the residuals can produce two sources in ANOVA table when the number of independent variables is k and the sample size is T. The first source is the sum of squares in regression (SSR) whose degree of freedom is k. The second source is the sum of squares of error (SSE) whose degree of freedom is T-k-1. These two sources decide the shape of the residual distribution. As k becomes larger, the residuals are restricted by more constraints, similar to a linear combination of random variables. If more random variables are added to the linear combination, the new random variable tends toward the normal distribution. The residuals that are restricted by more constraints also represent similar states such that the distribution of the residuals tends toward a normal distribution.

If we consider the collinear values of independent variables, then the collinear effect affects the convergence to a normal distribution when the error distribution is a U-quadratic distribution. The three columns of Figure 2 show that the residuals with convergence to normality are weakened by 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, $S(\hat{\varepsilon})$. MSE increases as the values of independent variables change from low to high multi-collinearity. The shape of the distribution in the third column of Figure 2 shows that the distribution of the first residual has a flat region around its mean and has more left-skewness and less centralization. Thus, the high multi-collinearity causes a big problem with the residuals, and 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 multi-collinearity of the independent variables results in the calculation and combination of the residuals becoming more complex. Moreover, the variance and distribution of the residuals are both disturbed by the collinearity when the errors follow a non-normal distribution. When the errors are asymmetric, the residuals

are delayed in their convergence to normality. To address this, 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 and weaken the collinear effect. Meanwhile, the degree of freedom also decreases.[3] This paper proposes a second proposition as follows.

Proposition 2.

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

(2)   Higher multi-collinear values of independent variables cause slower convergence to normality for the residuals.

(3)   Higher multi-collinear values and a smaller number of independent variables cause the distribution of the residuals to follow neither a non-normal distribution nor the error distribution, but a mixture of the error distribution and a normal distribution according to the constraints of the residuals.

The multi-collinear property and constraints on the residuals have opposite effects, simultaneously disturbing the distribution of the residuals which becomes a mixed distribution between a normal distribution and the error distribution.

3.3 Sample size is changed

The sample size effect plays a very important role in time series models that can represent the law of large numbers. Thus, the residuals may be regarded as a good estimator of the errors. The simulation case only changes the sample size from 9

---

[3] 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.

to 107 on the condition of 6 independent variables. The population variance is 1, the autocorrelation coefficient is zero, the values of independent variables are from the front T of the data set, and the simulated setting follows section 3.2. The residuals gradually revealed the property of the errors when the sample sizes increased from 9 to 107. Without loss of generalization, this paper only shows the shapes and coefficients of distributions for the $0.5 \times T^{th}$ residual in Figure 3 and Table -1, which are in Appendix A.

T=14, the 7th residual	T=60, the 30th 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
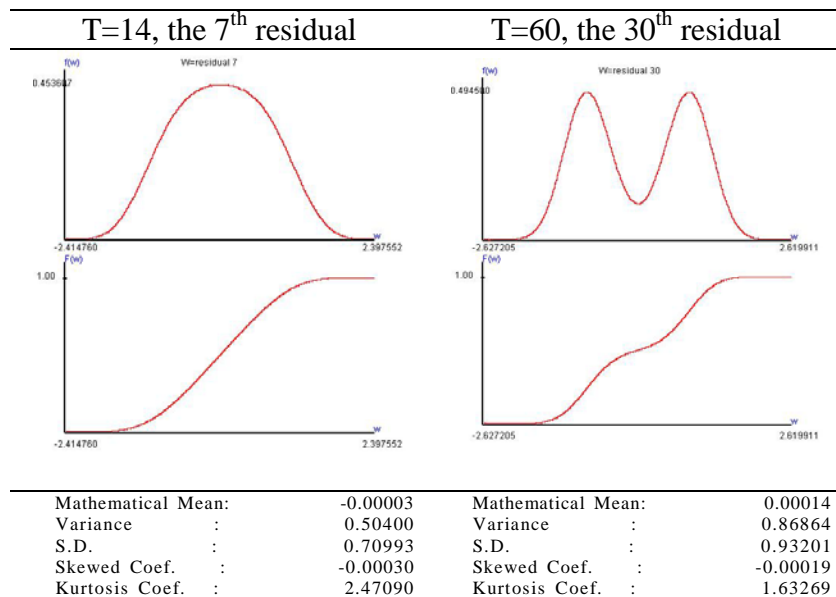
Figure 3: the $0.5T^{th}$ residual's distribution

Figure 3 explores the distributions of the $0.5T^{th}$ residual at $T = 14$ and 60. The second column of Figure 3 has a less regular distribution shape than the first column at $T = 14$. Meanwhile, the shape of the 30th residual at $T = 60$ tends toward a U-quadratic distribution more than the distribution of the 7th residual at $T = 14$. Figure 3 clearly shows that the larger sample sizes cause the $0.5T^{th}$ residual to tend towards the error distribution. Thus, the sample size effect can limit the effect

of the linear requirements on the residuals. At the same time, it allows the distribution of the residuals to represent more properties of the errors. The coefficients in Table   -1 show that the $0.5T^{th}$ residual has a mean of around zero, skewed coefficients, larger variances, and decreasing kurtosis when the sample sizes increased from 9 to 107. In comparison with the first column in Figure 2, the coefficients in Table   -1 approach the coefficients in the first column in Figure 2 when the sample sizes increased, especially for the variance and kurtosis coefficients. The coefficients that change with different sample sizes also show that the distribution of the residuals can be an estimator of the errors if the sample size effect is sufficiently larger than the effect of constraints on the residuals. The interaction between the sample size and the linear requirement causes the different shape and coefficients of the distribution of the $0.5T^{th}$ 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 tend toward the error distribution when the number of constraints is not large enough. In other words, the residual distribution is vastly different from the error distribution because the residuals are affected by the linear requirement of linear regression models, the error distribution, sample sizes, and the values of independent variables. It is generally assumed that errors are normally distributed, which is a symmetric distribution. However, the residuals can be an estimator of the errors when *k+1* is greater than 20, *T* ranges from 23 to *23+(k-19)*, and there is 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. We discuss the case where the autocorrelation coefficient of the errors is 0.7 and other

conditions are as the same as section 3.2. There are also two cases of low and high multi-collinear values of independent variables. We show the shapes and coefficients of distribution of first and 15th errors and residuals in Figure 4 and Figure 5, respectively.

the 1st error	the 1st residual Low multi-collinearity	the 1st residual High multi-collinearity

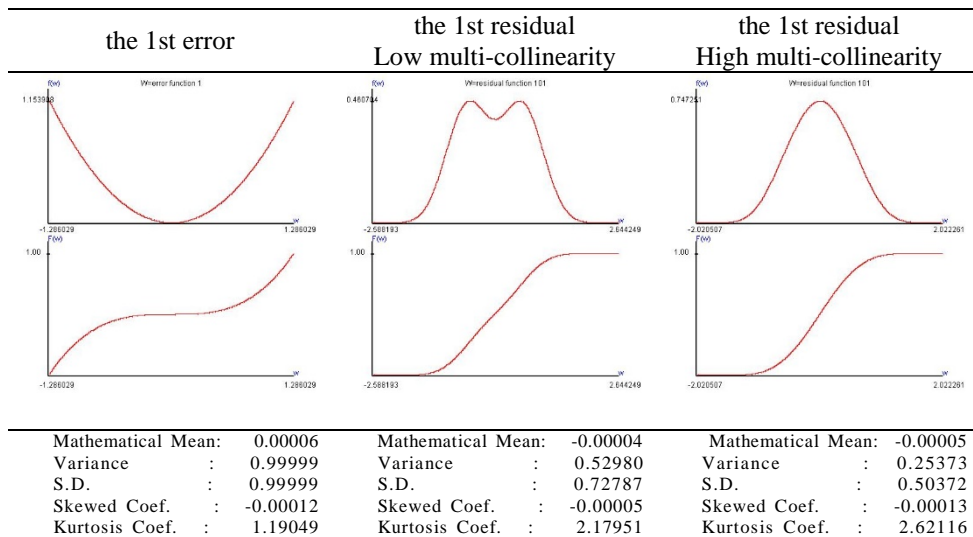|								
Mathematical Mean:	0.00006	Mathematical Mean:	-0.00004	Mathematical Mean:	-0.00005
Variance	:	0.99999	Variance	:	0.52980	Variance	:	0.25373
S.D.	:	0.99999	S.D.	:	0.72787	S.D.	:	0.50372
Skewed Coef.	:	-0.00012	Skewed Coef.	:	-0.00005	Skewed Coef.	:	-0.00013
Kurtosis Coef.	:	1.19049	Kurtosis 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 follows a U-quadratic distribution, but the first residuals that are affected by the degree of multi-collinearity show different shapes for the distribution of the residuals in Figure 4. The distribution of the first residuals cannot represent the property of the first error and greater multi-collinear values of independent variables cause more centralized distributions of the first residuals.

Another important view from coefficients in Figure 4 is that the coefficients are similar to each other among the three columns, but the population distribution of the first column and sampling distributions of the second and third columns have completely different shapes from each other. On the other hand, the coefficients

may highlight questionable results that are used to pass hypothesis testing when researchers only investigate the means and variances of residuals. Comparing the three columns, the linear requirement and multi-collinearity push the first residual towards centralization. Higher multi-collinear values of independent variables induce more centralized residuals.

This study also ran the distributions of the 15th error and residuals which are divided into two parts including low and high multi-collinearity in Figure 5.[4] The distribution of the 15th error is not the same as the U-quadratic distribution, but is more centralized as per the 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 show that there is a vast difference between the 15th residual and the 15th error.

Comparing Figure 4 and Figure 5, the coefficients of the errors show similar coefficients, except for the kurtosis coefficients. The diagrams of the first column from Figure 4 and Figure 5 are from the U-quadratic shape to the opposite of the U-quadratic shape because of the non-zero autocorrelation coefficient. The diagrams of low multi-collinear residuals show bulging from the first residual to the 15th residual, and the diagrams of high multi-collinear residuals are similar. This highlights that the nonzero autocorrelation coefficient and multi-collinearity interact on the residuals. Higher multi-collinearity decreases the effect of the autocorrelation coefficient on the residuals

---

[4] We also simulate the case of -0.7 autocorrelation coefficient in Appendix B.

the 15th error	the 15th residual Low multi-collinearity	the 15th residual High multi-collinearity

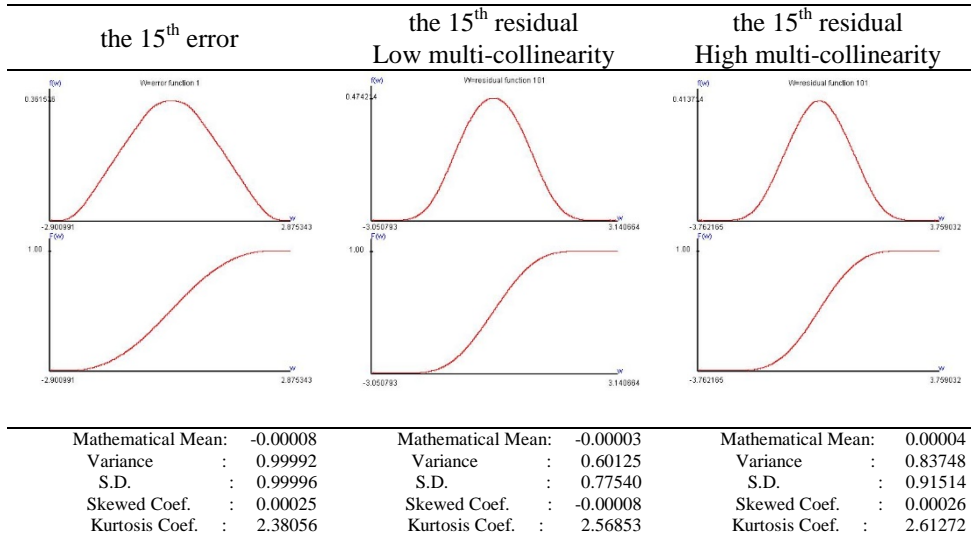|								
Mathematical Mean:	-0.00008		Mathematical Mean:	-0.00003		Mathematical Mean:	0.00004
Variance	:	0.99992	Variance	:	0.60125	Variance	:	0.83748
S.D.	:	0.99996	S.D.	:	0.77540	S.D.	:	0.91514
Skewed Coef.	:	0.00025	Skewed Coef.	:	-0.00008	Skewed Coef.	:	0.00026
Kurtosis Coef.	:	2.38056	Kurtosis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

4 Conclusion

The purpose of this paper is to explain why residuals cannot perfectly represent errors. This paper confirms that residuals that require specific conditions can be viewed as an estimator of the errors, but it may not be appropriate to assume a normal distribution because of the properties of the data. However, if the errors are normally distributed, the residuals can be a good estimator of the errors because of the properties of a normal distribution.

This paper also shows, via computer simulation results, how a non-normal distribution, linear requirements, multi-collinearity, sample size, and the autocorrelation coefficient affect the distributions of the errors and the residuals. First, residuals are restricted, but errors are not. Hence, the values of the residuals are constrained by a linear requirement from the regression analysis, and this causes the residuals to not perfectly represent errors. Second, this paper supposes that the linear requirement, sample sizes, and the values of independent variables

interact in the linear regression model. Thus: (1) the constraints of the linear requirements are high enough that the residuals follow a normal distribution when the sample sizes are fixed, regardless of the error term assumed. (2) If the error term is assumed to follow a normal distribution, then the residuals follow a normal distribution. (3) Larger sample sizes result in the residuals revealing properties of the error term when the linear requirement is fixed.

References

[1] B.H. Baltagi, *Econometrics*, Fifth edition, Springer: New York, 2011.

[2] C.F. Roos, Annual Survey of Statistical Techniques: The Correlation and Analysis of Time Series--Part II, *Econometrica*, **4**(4), (1936), 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*, **65**(332), (1970), 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*, **84**(4), (1921), 497-537.

[5] J. Durbin, G.S. Watson, Testing for Serial Correlation in Least Squares Regression: I, *Biometrika*, **37**(3/4), (1950), 409-428.

[6] J. Durbin, G.S. Watson, Testing for Serial Correlation in Least Squares Regression. II, *Biometrika*, **38**(1/2), (1951), 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*, **9**(1), (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 affects 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 graphs 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	9	10	12
$0.5T^{th}$	5	5	6
Coefficients	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	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	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
T	14	16	18
$0.5T^{th}$	7	8	9
Coefficients	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	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	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
T	20	30	40
$0.5T^{th}$	10	15	20
Coefficients	Mathematical Mean: -0.00001 Variance : 0.57802 S.D. : 0.70280 Skewed Coef. : 0.00022 Kurtosis Coef. : 2.35887 MAD : 0.63243 Range : 5.39457 Median : -0.87208 IQR : -0.44459	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	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
T	80	107	
$0.5T^{th}$	40	50	
Coefficients	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	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.

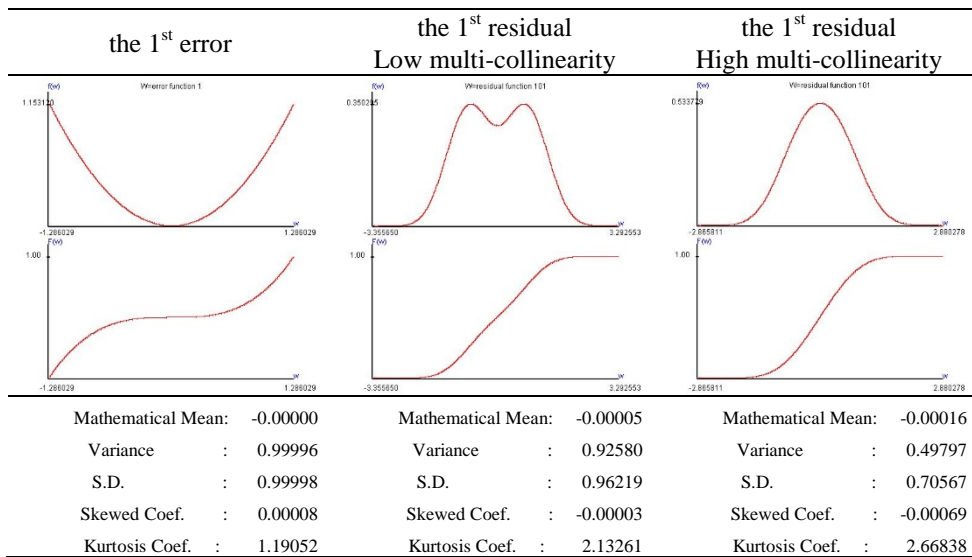| the 1st error	the 1st residual Low multi-collinearity		the 1st residual High multi-collinearity	
Mathematical Mean: -0.00000	Mathematical Mean:	-0.00005	Mathematical Mean:	-0.00016
Variance : 0.99996	Variance :	0.92580	Variance :	0.49797
S.D. : 0.99998	S.D. :	0.96219	S.D. :	0.70567
Skewed Coef. : 0.00008	Skewed Coef. :	-0.00003	Skewed Coef. :	-0.00069
Kurtosis Coef. : 1.19052	Kurtosis Coef. :	2.13261	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

With comparison of the autocorrelation coefficients of the errors that are 0.7 and -0.7, the means of the first error and the 15th 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 15th error in Figure B-2 and Figure 5. Third, the low multi-collinearity situation shows that the first residual has larger variance and less centralization in Figure B-1 than in Figure 4. However, the 15th residual is more centralized in Figure B-2 than in Figure 5. Finally, the high multi-collinearity situation explores that the first residual in Figure B-1 has larger variance and more centralization than in Figure 4,. However, the 15th residual has smaller variance and less centralization in Figure B-2 than in Figure 5.

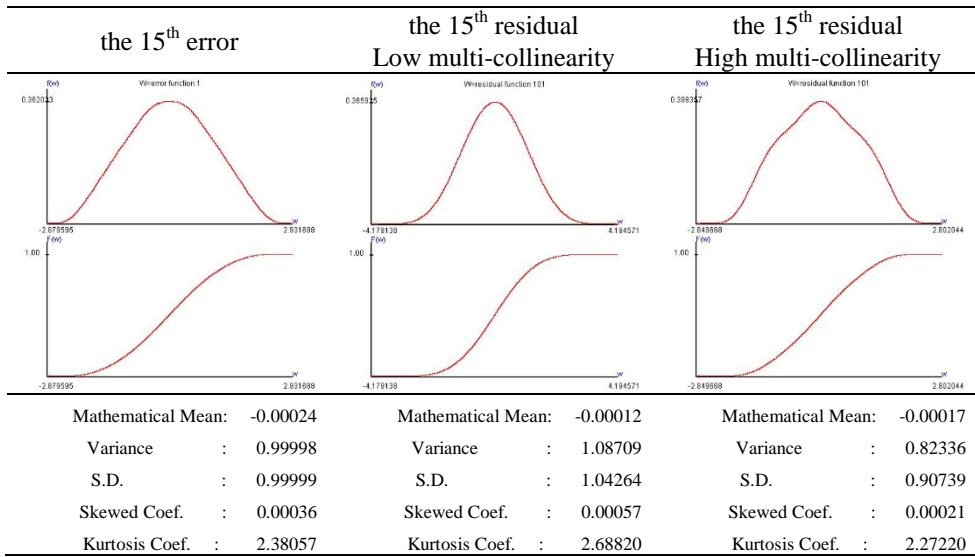| the 15th error	the 15th residual Low multi-collinearity	the 15th residual High multi-collinearity
Mathematical Mean: -0.00024	Mathematical Mean: -0.00012	Mathematical Mean: -0.00017
Variance : 0.99998	Variance : 1.08709	Variance : 0.82336
S.D. : 0.99999	S.D. : 1.04264	S.D. : 0.90739
Skewed Coef. : 0.00036	Skewed Coef. : 0.00057	Skewed Coef. : 0.00021
Kurtosis Coef. : 2.38057	Kurtosis Coef. : 2.68820	Kurtosis Coef. : 2.27220

Figure B-2.   The 15th error and residual distributions when error follows U-quadratic

distribution and autocorrelation coefficient is -0.7