

## Claremont Colleges Scholarship @ Claremont

---

CMC Senior Theses

CMC Student Scholarship

---

2015

# Applications of Monte Carlo Methods in Statistical Inference Using Regression Analysis

Ji Young Huh

*Claremont McKenna College*

---

### Recommended Citation

Huh, Ji Young, "Applications of Monte Carlo Methods in Statistical Inference Using Regression Analysis" (2015). *CMC Senior Theses*. Paper 1160.

[http://scholarship.claremont.edu/cmc\\_theses/1160](http://scholarship.claremont.edu/cmc_theses/1160)

This Open Access Senior Thesis is brought to you by Scholarship@Claremont. It has been accepted for inclusion in this collection by an authorized administrator. For more information, please contact [scholarship@cuc.claremont.edu](mailto:scholarship@cuc.claremont.edu).

**CLAREMONT MCKENNA COLLEGE**

**Applications of Monte Carlo Methods in  
Statistical Inference Using Regression Analysis**

**SUBMITTED TO**

**Professor Manfred Werner Keil**

**AND**

**Professor Mark Huber**

**AND**

**Dean Nicholas Warner**

**BY**

**Ji Young Huh**

**for**

**SENIOR THESIS**

**Spring 2015**

**April 27, 2015**



## **ABSTRACT**

This paper studies the use of Monte Carlo simulation techniques in the field of econometrics, specifically statistical inference. First, I examine several estimators by deriving properties explicitly and generate their distributions through simulations. Here, simulations are used to illustrate and support the analytical results. Then, I look at test statistics where derivations are costly because of the sensitivity of their critical values to the data generating processes. Simulations here establish significance and necessity for drawing statistical inference. Overall, the paper examines when and how simulations are needed in studying econometric theories.

## **ACKNOWLEDGEMENTS**

This thesis has a special meaning to me as the first product of my independent academic endeavor that I particularly felt passionate about. I feel very fortunate to have had such a valuable opportunity in my undergraduate career.

I would first like to thank Professor Keil for his continuous encouragement and advices, both academic and personal, throughout my college years. I am glad that I reached him at the end of my freshman year for an advice on joining Lowe Institute of Political Economy.

I would also like to express my gratitude to Professor Huber for his incredible course on Monte Carlo methods, which became the very topic of my paper. Furthermore, I am thankful to Professor De Pace for his incredibly helpful comments and guidance on this paper.

Four years at CMC granted me phenomenal education and wonderful memories. I would like to thank every professor, faculty member, friend, and study labs that made this journey possible.

Most importantly, I am hugely indebted to my parents for their unlimited love and support that guided me to become who I am right now.

# TABLE OF CONTENTS

<b>ABSTRACT</b> .....	iii
<b>ACKNOWLEDGEMENT</b> .....	iv
<b>1. INTRODUCTION</b> .....	2
<b>2. BACKGROUND</b> .....	4
2.1 Statistical Inference from a Distribution .....	4
2.2 History of Monte Carlo Simulation .....	7
<b>3. DERIVATION OF ESTIMATOR PROPERTIES</b> .....	9
3.1 Assumptions.....	9
3.2 Estimation of a Mean .....	10
3.3 Estimation of a Fixed (Non-Random) Independent Variable .....	12
3.4 Estimation of a Random Independent Variable .....	19
3.5 Estimation of Autoregressive (AR) Model .....	22
<b>4. DISTRIBUTION OF TEST STATISTICS</b> .....	25
4.1 Durbin-Watson Test for Autocorrelation .....	26
4.2 Dickey-Fuller Test for Nonstationarity .....	32
<b>5. CONCLUSION</b> .....	42
<b>6. APPENDIX</b> .....	44
<b>7. BIBLIOGRAPHY</b> .....	51

# 1. INTRODUCTION

In most economic applications, researchers are interested in statistical inference, in other words, the ability to infer relationships in a population from a sample. In order to do so, it is crucial to understand to what extent parameters in a model vary across samples. For example, you may not be particularly interested in the sample behavior of specific election polls carried out by a variety of organizations. The actual objective is to draw general conclusions about the population—who the population will actually vote for—from the election polls. Because it is impractical to ask all the potential voters, due to the time, cost, and unexpected circumstances, the election polls serve only as a practical starting point.

Statistical theory has provided an answer to the issue of a sample variation through analytical derivations of sample distributions in many applications. Sample means or regression coefficients follow a normal distribution when certain population parameters, such as population variances, are known. This allows you to construct confidence intervals, which, in turn, provide a range of probable population values. Even in more complicated set-ups where such population parameters are unknown, some of the resulting distributions converge to a normal distribution as long as certain assumptions, for example regarding the sample size, are met. However, there are other circumstances where analytical derivations are impossible, at least with the mathematical knowledge

that is available to us currently. In these situations, the only way to determine sample variation comes from a simulation technique called Monte Carlo methods.

The purpose of this paper is to investigate the various usages of Monte Carlo methods in drawing statistical inferences when mathematical derivations of sample distributions are both possible and unavailable. While it may seem at first that applications of Monte Carlo methods are not useful when analytical derivations are available, usage of Monte Carlo methods represents both a pedagogical device to explain sample variation as well as creating an illustration.

I proceed as follows. First, I analyze how properties of estimators are derived analytically and also through simulations. I focus on Ordinary Least Squares (OLS) and several other estimators when there is single unknown population parameter, unbiasedness as a representative property of an estimator. Then, two representative statistics used to test characteristics of sample distributions are investigated to see how analytical derivations become computationally expensive and simulation becomes necessary. The importance of a test statistic is established as well through Monte Carlo simulations. A final section concludes.



## 2. BACKGROUND

In this section, I motivate the use of Monte Carlo methods by introducing the importance of distributions in drawing statistical inferences. Then, I proceed with a brief history of Monte Carlo methods.

### *2.1 Statistical Inference from a Distribution*

Jan Kmenta, one of the first highly successful authors in econometrics, categorizes statistics into two: “descriptive statistics and statistical inference”.<sup>1</sup> While descriptive statistics try to come up with explanations for a phenomenon, inferential statistics provide tools to assess the reliability of the proposed explanations, such as how well the sample estimator represents the population parameters. For example, consider a model that generates “ideal” weights for females based on heights, which gives a convenient way to calculate the medication dosages for obese patients:

$$\text{Ideal Body Weight(Female)} = 45.5\text{kg} + 2.3\text{kg/m} * (\text{Height} - 1.5\text{m})$$

In this model, weight is estimated based on one’s height. Setting up the equation to study weight is a part of descriptive statistics. Here, the  $2.3\text{kg/m}$  is an estimate of the population parameter of the relationship between height and weight.

An estimator is a rationale of how we will set up an equation to explain a variable of interest, which we call dependent variable. Using a different estimator may produce

---

<sup>1</sup> Kmenta (1971)

different estimator distribution properties which may lead to contrasting judgments on statistical inferences. Properties of an estimator distribution provide information about the sample distribution and how it changes with varying sample size, which include unbiasedness, efficiency, skewness, kurtosis, and asymptotic normality.

- Unbiasedness: the estimator has the sampling distribution with a mean equal to the population parameter.
- Efficiency: the estimator has the lowest variance among other estimators.
- Skewness: measure of the asymmetry of the estimator distribution
- Kurtosis: measure of the “peakedness” of the distribution.
- Asymptotic Normality: distribution of the estimator becomes normal with an increase in sample size.

To get a better understanding of the importance of distributions on statistical inference, take the simple regression model as an example. You might also be interested in interpreting certain sample statistics relative to the population parameters. If the sample statistic does not behave as expected from the model, how “unexpected” is it likely to be observed? Statistics such as the  $t$ -statistic are used to address the deviation of a sample statistic from the population parameter. Nevertheless, these numbers are hard to interpret because their implications vary significantly. It can be easily assumed that  $t$ -statistics are distributed according to the standard  $t$ -distribution which approaches the normal distribution with increasing sample size. If one of the assumptions regarding the standard  $t$ -distribution fails, however, then the distribution changes, and critical values are no longer valid.

In a regression model, the  $t$ -statistic is calculated as follows:

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i - \beta_i}{s.e.(\hat{\beta}_i)}$$

Consider two models with different distributions for the  $t$ -statistic, one being normally distributed while the other is Cauchy distributed. If the calculated  $t$ -statistic was 1.96 in both situations, then we may draw wildly different inference about the population. A Cauchy distribution has thicker tails compared to a normal distribution, and hence 1.96 standard deviations away from the mean implies lower significance. In fact, a  $t$ -statistic of 1.96 implies that there is only a 2% chance that such a sample statistic will be found in the model under a normal distribution; the same sample statistic implies that there is an almost 20% chance under a Cauchy distribution. To obtain statistical significance as high as you get from a  $t$ -statistic of 1.96 under a normal distribution, the  $t$ -statistic under a Cauchy distribution requires a  $t$ -statistic of 12.7.<sup>2</sup>

So how do we figure out appropriate critical values for the models if distributions are unknown? For models that satisfy certain assumptions, distributions of statistics can be analytically derived; if not, numerical simulation allows us to draw statistical inference from the numbers generated. Because it is almost impossible to obtain sufficient data points from the real world to develop the distribution, simulation techniques have become a popular way to improve understanding of unknown distributions.

Analytical derivations of distributions are convenient because they can result in numerous applications that can be utilized in other models; however, they are often impractical to generate as most of the models in the real world do not share the

---

<sup>2</sup> See Appendix A1; standard  $t$ -distribution has Cauchy distribution when degree of freedom equals 1, and normal distribution when degree of freedom equals infinity.

assumptions necessary for the derivations, given that we have enough information about the data to assess the assumptions. With the technology to run thousands of simulations within a second, numerically producing the distribution with huge number of data points became a more time-efficient and effective solution for statistical inference.

In addition, these estimators require several crucial assumptions to make the process of constructing the model possible. With slight modifications to the assumptions, the critical values become invalid. Even worse, not all test statistics are independent of the data generating processes. One key usage of simulations is to validate these assumptions for the models. Tests are constructed with hypotheses related to the assumptions, and with the resulting test statistics, the distribution of the test statistic allows us to determine the validity of the assumptions. By examining two test statistics – the Durbin-Watson statistic and the Dickey-Fuller statistic – I study the degree to which they are sensitive to the specific data generating processes.

## *2.2 History of Monte Carlo Simulation*

Although simulation methods using randomness were used by some mathematicians and scientists,<sup>3</sup> the Monte Carlo method was first officially introduced in 1946 by Stanislaw Ulam when he was working as a physicist at Los Alamos Scientific Laboratory.<sup>4</sup> As they were trying to calculate the density of nuclear particles which determine the energy, Ulam proposed a method using random experiments, which was ultimately introduced in the Journal of the American Statistical Association as the Monte Carlo method.

---

<sup>3</sup> Harrison (2010)

<sup>4</sup> Wikipedia, (“Monte Carlo Methods”, accessed on January 3<sup>rd</sup>, 2015)

Metropolis and Ulam (1949) established the Monte Carlo method with its usefulness in solving physical problems. In its introduction, the study discusses how the method enables them to answer problems that cannot be solved through conventional deterministic approach. For problems that involve intermediary parts, the traditional approach of statistical or analytical mechanics cannot provide a practical solution. Similar problems in other fields of study are also presented, such as computing a volume of a region in a higher dimension, where one needs to count  $10^n$  lattice points in a unit cube with points satisfying  $n$  number of inequalities. Instead, with laws of large numbers and asymptotic properties of proposed models, it becomes possible to avoid impractical procedures to solve the problems, and this is where the Monte Carlo method plays an important role.<sup>5</sup>

With the advent of advanced computer technology, running Monte Carlo simulations on computers became much less expensive; it is also getting widely used in other fields of study including Econometrics. This paper further investigates Monte Carlo's necessity and significance specifically in Econometrics by looking analytically at different estimation models and statistics.

---

<sup>5</sup> Metropolis and Ulam (1949)

### 3. DERIVATION OF ESTIMATOR PROPERTIES

The current section looks at properties of different estimators with both mathematical derivations and computer simulations.

#### *3.1 Assumptions*

For the estimation models throughout this paper, I construct regression equations with one regressor; for the autoregressive models, I construct regression equations with one lag of the dependent variable. So I assume the standard form of a regression equation to be  $Y_i = \beta_0 + \beta_1 X_i + u_i$  and that for an autoregressive model to be  $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$ .

I take Gauss-Markov assumptions to be valid in this paper unless stated otherwise. Gauss-Markov assumptions are imposed with the OLS estimator which follows in the next section – they include  $E(\varepsilon) = 0$ ,  $Var(\varepsilon) = \sigma_\varepsilon^2$ , and  $cov(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$ , where  $\varepsilon$  is the error term. The first property insures that the mean of the error terms is zero; second the homoscedasticity in error; and third no autocorrelation in the residuals. The Durbin-Watson statistic which tests for autocorrelation is discussed in section 4.

All simulations are replicated 10,000 times unless specified otherwise. The lines appearing on top of most histograms presented illustrate the normal density with the same mean and standard deviation as the data. The distribution of  $\hat{\beta}_0$  is omitted since we are

interested in the slope rather than the intercept, and the distribution properties of the intercepts are described in the table in detail.  $t$ -statistics and Jarque-Bera test statistics are added. The  $t$ -statistic measures how far a sample statistic is away from the population parameter in units of standard errors; the Jarque-Bera test statistic measures how much the distribution differs from a normal distribution with the same mean and standard deviation as the data.<sup>6</sup>

### 3.2 Estimation of a Mean

Consider a situation where we need to estimate the population mean from a set of samples. For a population,  $Y_i = \mu + \varepsilon_i$ , we estimate the mean  $\mu$  from the equation derived from the sample mean  $\beta_0$  in  $Y_i = \beta_0 + e_i$ .

First consider the Ordinary Least Squares (OLS) estimator. The least squares estimation method minimizes the squared distance between the actual data point and the estimate, or the sum of squares of error terms,  $\sum \hat{e}_i^2$ .

To minimize squared sums of the error terms, we take the first order derivative with respect to  $\beta_0$ :

$$\begin{aligned}\sum e_i^2 &= \sum (Y_i - \beta_0)^2 \\ FOC: \quad -2 \sum (Y_i - \hat{\beta}_0) &= 0 \\ n\hat{\beta}_0 &= \sum Y_i \\ \hat{\beta}_0 &= \frac{\sum Y_i}{n} = \bar{Y}\end{aligned}$$

---

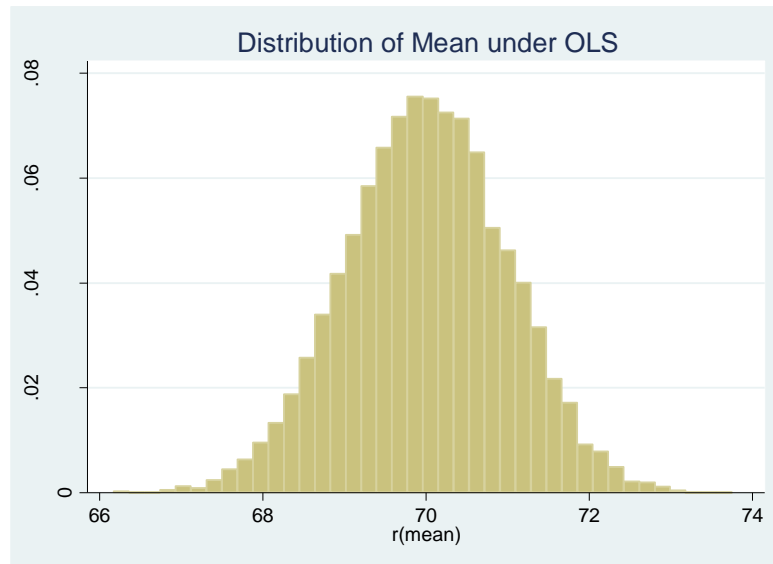
<sup>6</sup> See Appendix A1 for standard  $t$ -table, and A2 for Jarque-Bera table.

Here, we can see that the sample mean is the least squares estimator of the population mean.

To test its unbiasedness, consider the expected value of  $\hat{\beta}_0$ .

$$\begin{aligned} E(\hat{\beta}_0) &= E\left(\frac{\sum Y_i}{n}\right) = \frac{E(\sum \beta_0 + e_i^2)}{n} \\ &= \frac{n\beta_0}{n} = \beta_0 \end{aligned}$$

It is clear that we do not need to run a simulation to find out our estimator is unbiased. If we run Monte Carlo experiment to get the distribution of the estimator for pedagogical purposes, we get the following.<sup>7</sup>

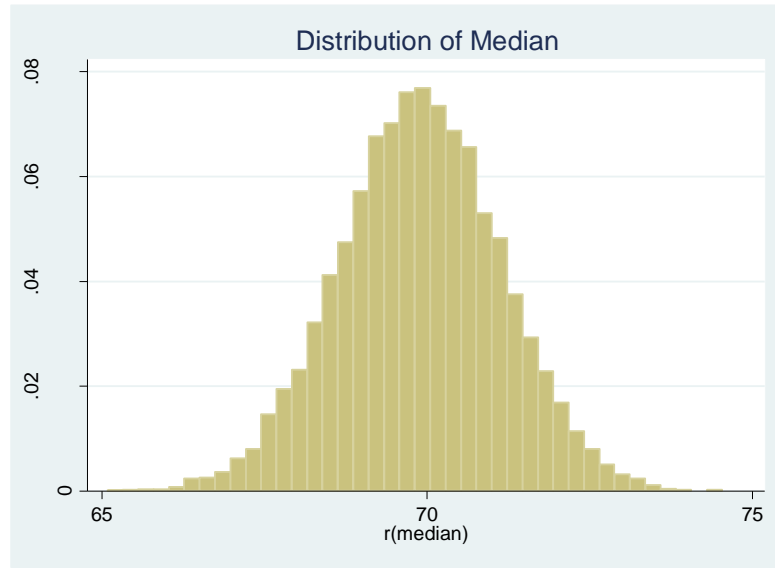


In addition to the least squares estimator, we can always take another estimator, such as a median. Result for the distribution of a median shows that the sample median is not an unbiased estimator of the population mean – the  $t$ -statistic is statistically significant to reject the null hypothesis of the median being equal to the mean.

---

<sup>7</sup> For STATA codes of the simulations, see Appendix A3.





	True Mean	Obs	Mean	StDev	Skewness	Kurtosis <sup>8</sup>	t- statistic	JB- statistic
<b>mean</b>	70	100	70.00	0.99	-0.04	2.96	0	0.030
<b>median</b>	70	100	69.89	1.25	-0.04	3.05	9.12	0.032

This leads to an introduction of median-unbiasedness. In general when bias is concerned, it is bias from the mean; bias can also be measured with respect to median. Median-unbiased estimators have different properties than mean-unbiased estimators, such as invariance under linear transformations.<sup>9</sup> In this paper, I use the term bias with respect to the mean.

### 3.3 Estimation of a Fixed (Non-Random) Independent Variable

Now, assume that we have an explanatory variable for our dependent variable and we would like to test how well different estimators provide different sample distributions that estimate the population parameters.

<sup>8</sup> A normal distribution has a kurtosis of 3.

<sup>9</sup> Van der Vaart (1962)

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

First, I look at situations where  $X_i$  is a fixed in repeated sample. When we drop this assumption in the next section,  $X_i$  will hold different values for different replications. Under this condition, we look at three different estimators to see how we can derive the unbiasedness of the estimator distributions analytically and run simulations to replicate the results. The three estimators are least squares estimator, least absolute errors estimator, and arbitrary 2 points estimator.

### 3.3.1 Ordinary Least Squares (OLS) estimator

From  $\sum u_i^2 = \sum (Y_i - \beta_0 - \beta_1 X_i)^2$ , we solve for the first order condition with respect to  $\beta_0$  and  $\beta_1$ .

$$\frac{\partial \sum u_i^2}{\partial \beta_0} = -2 \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

$$\begin{aligned} \hat{\beta}_0 &= \frac{\sum Y_i}{n} - \frac{\hat{\beta}_1 \sum X_i}{n} \\ &= \bar{Y} - \hat{\beta}_1 \bar{X} \end{aligned}$$

$$\frac{\partial \sum u_i^2}{\partial \beta_1} = -2 \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0$$

$$\sum X_i Y_i - n \hat{\beta}_0 \bar{X} - \hat{\beta}_1 \sum X_i^2 = 0$$

Since  $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ ,

$$\sum X_i Y_i - n(\bar{Y} - \hat{\beta}_1 \bar{X}) \bar{X} - \hat{\beta}_1 \sum X_i^2 = 0$$

$$\hat{\beta}_1 \left( \sum X_i^2 - n \bar{X}^2 \right) = \sum X_i Y_i - n \bar{X} \bar{Y}$$

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

To prove its unbiasedness, we show that  $E(\hat{\theta}) = \theta$ .

$$\begin{aligned} E(\hat{\beta}_1) &= E\left(\frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}\right) \\ &= E\left(\frac{\sum (X_i - \bar{X}) Y_i}{\sum (X_i - \bar{X})^2}\right) - E\left(\frac{\bar{Y} \sum (X_i - \bar{X})}{\sum (X_i - \bar{X})^2}\right) \\ &= E\left(\frac{\sum (X_i - \bar{X}) (\beta_0 + \beta_1 X_i + u_i)}{\sum (X_i - \bar{X})^2}\right) \\ &= E\left(\frac{\beta_1 (\sum X_i^2 - n \bar{X}^2) + \sum X_i u_i}{\sum (X_i - \bar{X})^2}\right) \\ &= \beta_1 + E\left(\frac{\sum X_i u_i}{\sum (X_i - \bar{X})^2}\right) \end{aligned}$$

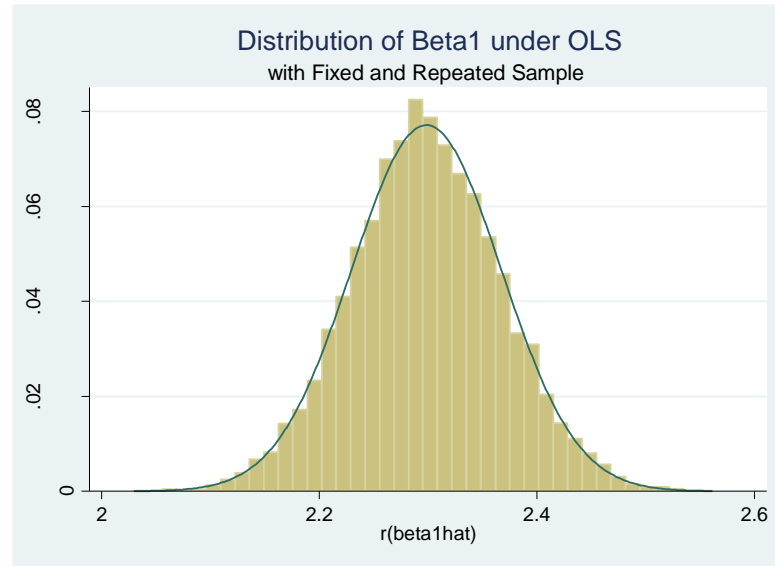
Here, note that  $X_i$  is a fixed in repeated sample, so  $\frac{\sum X_i}{\sum (X_i - \bar{X})^2}$  is a constant. Therefore,

$$E(\hat{\beta}_1) = \beta_1$$

Consequently,

$$\begin{aligned} E(\hat{\beta}_0) &= E(\bar{Y} - \hat{\beta}_1 \bar{X}) \\ &= E(\beta_0 + \beta_1 \bar{X} + \bar{u} - \hat{\beta}_1 \bar{X}) \\ &= \beta_0 + E(\beta_1 - \hat{\beta}_1) \bar{X} \\ &= \beta_0 \end{aligned}$$

The unbiased shape of the estimator distribution is illustrated from the following simulation results.



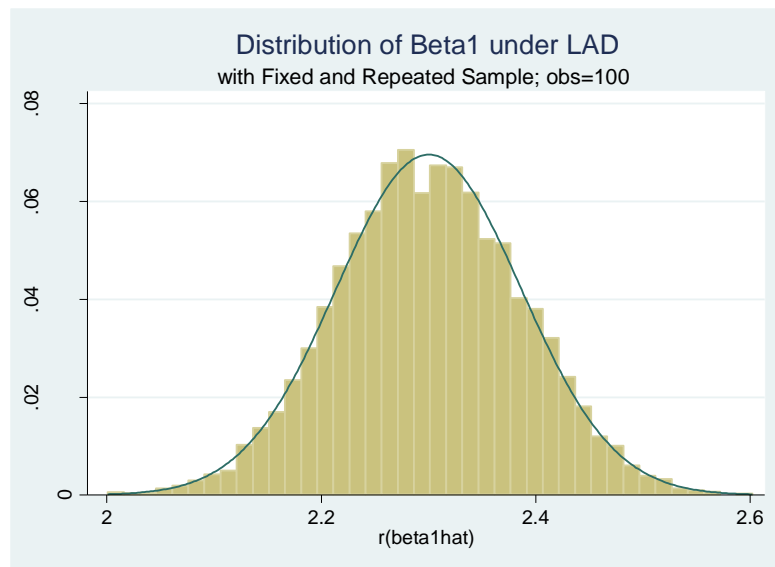
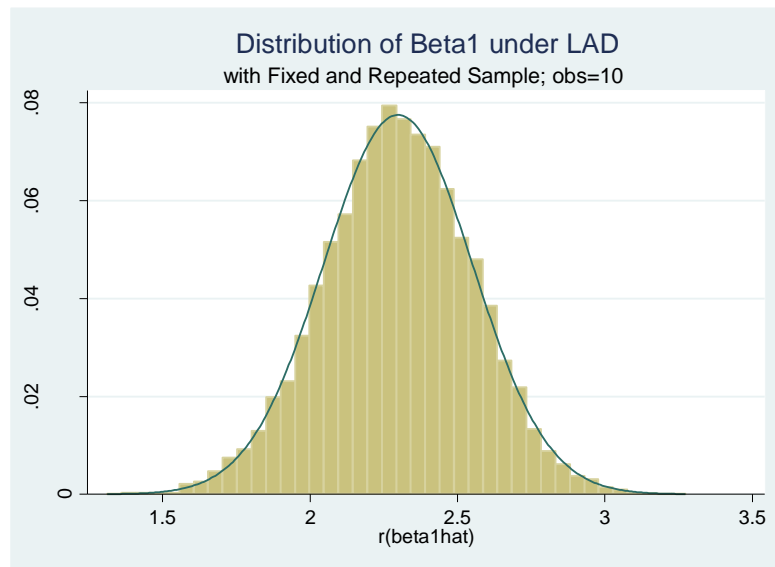
OLS, fixed	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	t- statistic	JB- statistic
$\beta_0$	100	42	42.26	12.09	-0.03	3.03	-2.15	0.021
$\beta_1$	100	2.3	2.30	0.07	0.04	3.03	1.45	0.024

### 3.3.2 Least Absolute Deviation (LAD) Estimator

Instead of minimizing  $\sum u_i^2$ , consider minimizing  $\sum |u_i|$ . This estimator is called a least absolute deviation (LAD) estimator. An LAD estimator exhibits asymptotic normality<sup>10</sup> - with a bigger sample size, the distribution of the sample estimators approaches closer to normality. The analytical proof is not shown here.

It is interesting that in a fixed in repeated samples set-up, the LAD estimator is normal even for small sample size cases. As shown below using the Jarque-Bera test statistic, the distribution of the estimator does not become more normal with an increase in sample size.

<sup>10</sup> Pollard (1991)



<b>LAD, fixed</b>	<b>Obs</b>	<b>True Mean</b>	<b>Mean</b>	<b>StDev</b>	<b>Skewness</b>	<b>Kurtosis</b>	<b>t- statistic</b>	<b>JB- statistic</b>
$\beta_0$	10	42	42.35	44.34	0.06	3.02	-0.78	0.0052
	100	42	41.98	15.15	-0.007	2.91	0.13	0.033
$\beta_1$	10	2.3	2.30	0.25	-0.06	3.03	0.67	0.0056
	100	2.3	2.30	0.09	0.009	2.91	0	0.035

### 3.3.3 Arbitrary 2 Points (A2P) Estimator

Consider a line connecting two points (P and Q) chosen arbitrarily from a graph with explanatory variable on the x-axis and independent variable on the y-axis. We will call the estimator that takes this line as its regression equation an Arbitrary 2 Points Estimator, or an A2P estimator.  $\tilde{\beta}_0$  will be the intercept, and  $\tilde{\beta}_1$  will be the slope of the line. Then the A2P estimator is unbiased.

$$\begin{aligned}
Y_i &= \tilde{\beta}_0 + \tilde{\beta}_1 X_i + u_i \\
\tilde{\beta}_0 &= \frac{X_Q Y_P - X_P Y_Q}{X_Q - X_P} \\
&= \frac{(\tilde{\beta}_0 X_Q + \tilde{\beta}_1 X_P X_Q + u_P X_Q) - (\tilde{\beta}_0 X_P + \tilde{\beta}_1 X_Q X_P + u_Q X_P)}{X_Q - X_P} \\
&= \tilde{\beta}_0 + \frac{u_P X_Q - u_Q X_P}{X_Q - X_P} \\
\tilde{\beta}_1 &= \frac{Y_Q - Y_P}{X_Q - X_P} \\
&= \frac{(\tilde{\beta}_0 + \tilde{\beta}_1 X_Q + u_Q) - (\tilde{\beta}_0 + \tilde{\beta}_1 X_P + u_P)}{X_Q - X_P} \\
&= \tilde{\beta}_1 + \frac{1}{X_Q - X_P} (u_Q - u_P)
\end{aligned}$$

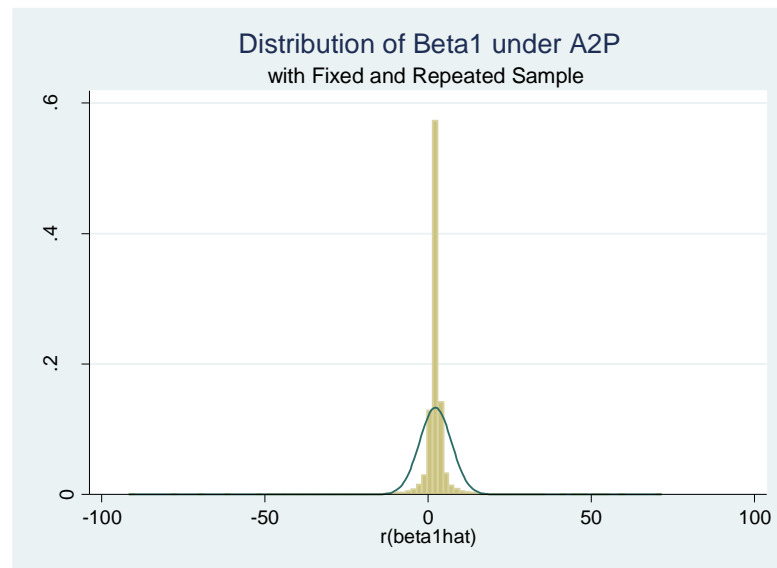
For unbiasedness, we look at expected values of the two.

$$\begin{aligned}
E(\tilde{\beta}_0) &= E\left(\tilde{\beta}_0 + \frac{u_P X_Q - u_Q X_P}{X_Q - X_P}\right) \\
&= \tilde{\beta}_0 + E\left(\frac{u_P X_Q}{X_Q - X_P}\right) - E\left(\frac{u_Q X_P}{X_Q - X_P}\right)
\end{aligned}$$

Since we assumed that there is no correlation between the regressor and the error term, they are independent.

$$\begin{aligned}
E(\tilde{\beta}_0) &= \tilde{\beta}_0 + E(u_P)E\left(\frac{X_Q}{X_Q - X_P}\right) - E(u_Q)E\left(\frac{X_P}{X_Q - X_P}\right) \\
&= \tilde{\beta}_0 \\
E(\tilde{\beta}_1) &= E\left(\tilde{\beta}_1 + \frac{u_Q - u_P}{X_Q - X_P}\right) \\
&= \tilde{\beta}_1 + E(u_Q)E\left(\frac{1}{X_Q - X_P}\right) - E(u_P)E\left(\frac{1}{X_Q - X_P}\right) \\
&= \tilde{\beta}_1
\end{aligned}$$

When we look at the simulation result, it shows an interesting shape. It has a huge variation, which makes sense considering that we are randomly choosing P and Q. So even though the mean actually seems to be significantly off compared to other estimators, the  $t$ -statistic is not significant, which shows that we cannot reject the null that the sample estimator is equal to the population parameter. Its distribution is also significantly different from the normal distribution, demonstrating a huge peak and flat tails.



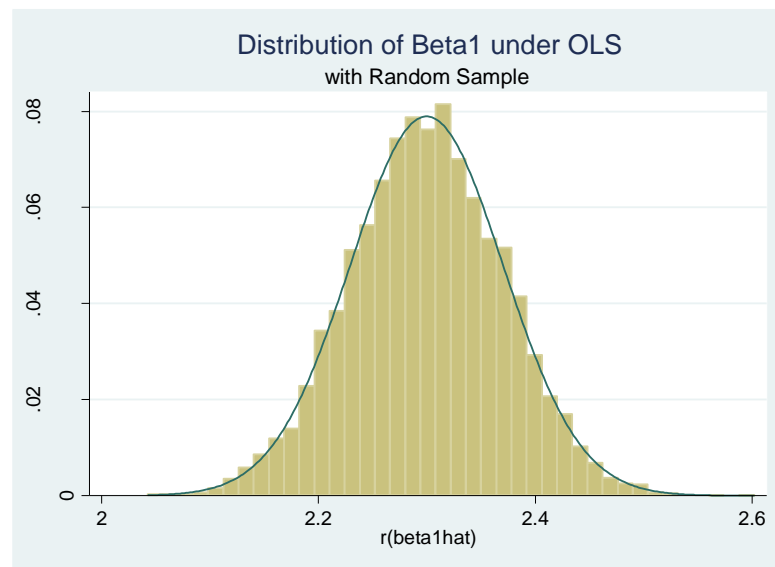
A2P, fixed	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	t- statistic	JB- statistic
$\beta_0$	100	42	46.59	868.61	1.84	78.78	-0.53	23984
$\beta_1$	100	2.3	2.28	4.89	-1.44	67.61	0.47	17428

### 3.4 Estimation of a Random Independent Variable

If our regressor is randomly generated,  $X_i$  will be different for each replication. This section confirms that the properties of the estimators do not change even if we draw  $X_i$  randomly from a normal distribution.<sup>11</sup>

#### 3.4.1 OLS Estimator

The same proof we used in the previous section applies here if we take expected values with the condition on the  $X$ s. For an elaborated version of the proof, see Stock and Watson (2015).



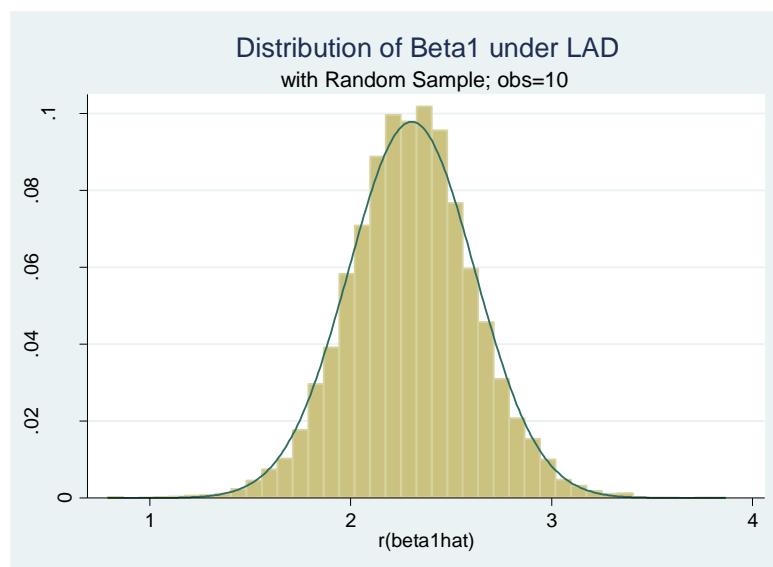
<sup>11</sup> There is a slight change with LAD estimator – it shows asymptotic normality with a random sample while it did not with a fixed sample.

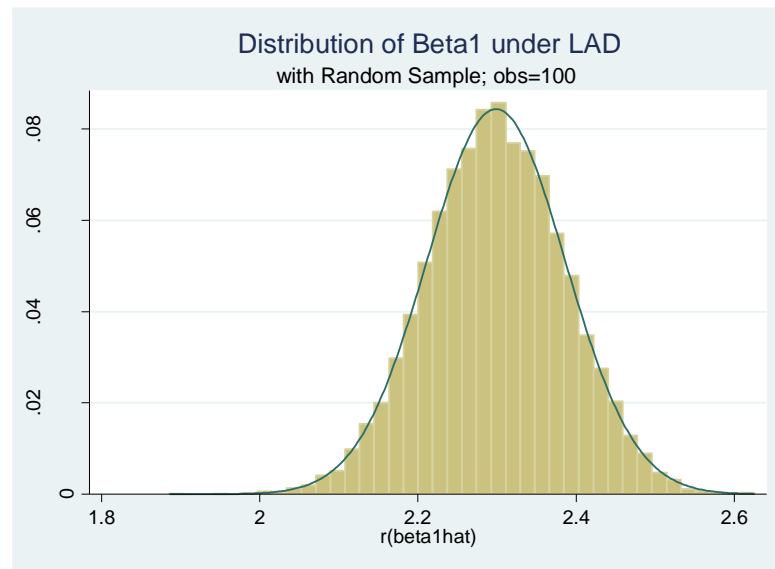


OLS, random	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	t- statistic	JB- statistic
$\beta_0$	100	42	42.04	12.42	0.015	2.96	-0.32	0.011
$\beta_1$	100	2.3	2.30	0.071	-0.011	2.96	0	0.010

### 3.4.2 Least Absolute Errors Estimator

Recall from the previous section with fixed in repeated samples – the LAD estimator showed normality for both small and big sample. In a random sample, however, we can see the asymptotic normality more clearly. The Jarque-Bera test statistic is smaller with 100 observations than 10 observations, even though they are both too small to reject the null hypothesis of normality.

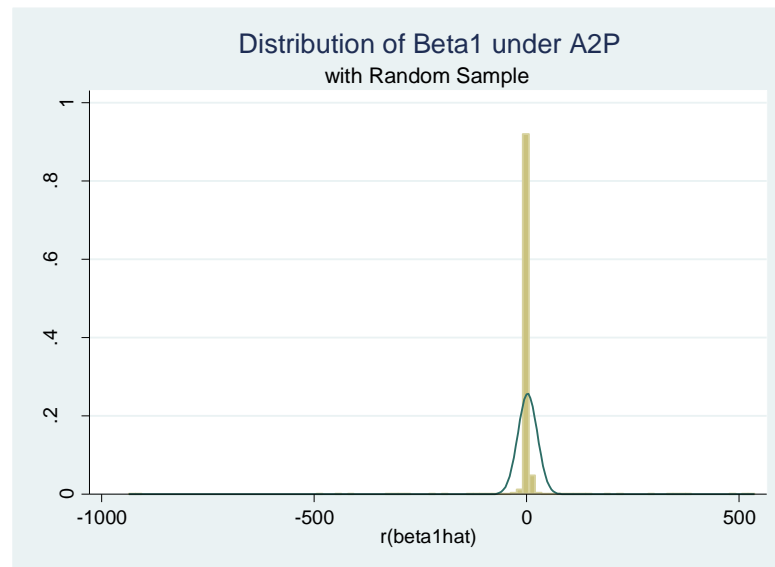




	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	t-statistic	JB-statistic
$\beta_0$	10	42	41.46	55.12	-0.0005	3.73	0.99	0.22
	100	42	42.19	15.38	0.03	3.01	-1.26	0.01
$\beta_1$	10	2.3	2.30	0.31	0.01	3.66	-1.12	0.18
	100	2.3	2.30	0.088	-0.03	3.02	0	0.02

### 3.4.3 A2P Estimator

Just as shown previously, A2P estimator has a big variation and kurtosis. With random samples, the variation and kurtosis rises even more outrageously.



A2P, random	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	t- statistic	JB- statistic
$\beta_0$	100	42	63.98	4216	14.22	843.33	-0.52	2,945,710
$\beta_1$	100	2.3	2.18	23	-11.49	722.25	0.51	2,157,701

### 3.5 Estimation of Autoregressive (AR) Model

An AR model is a model where the dependent variable can be predicted by its own past behaviors. So the explanatory variables are the lags of the dependent variable on the left hand side.

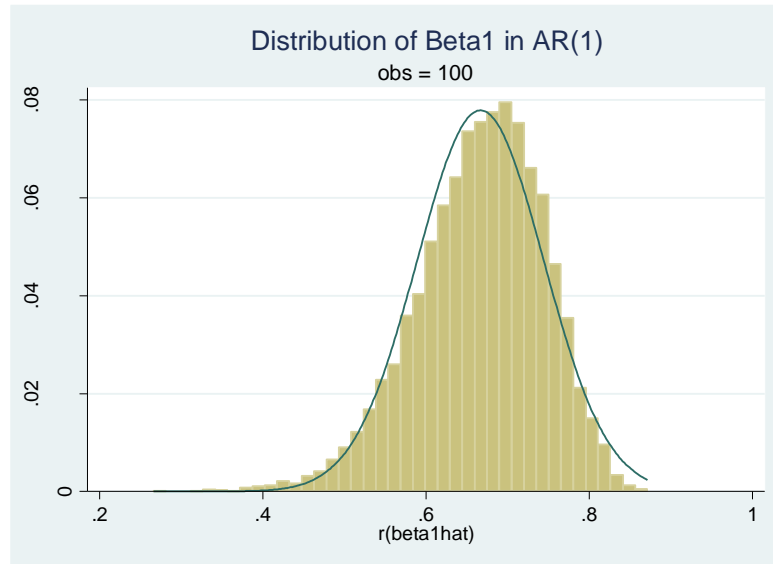
$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

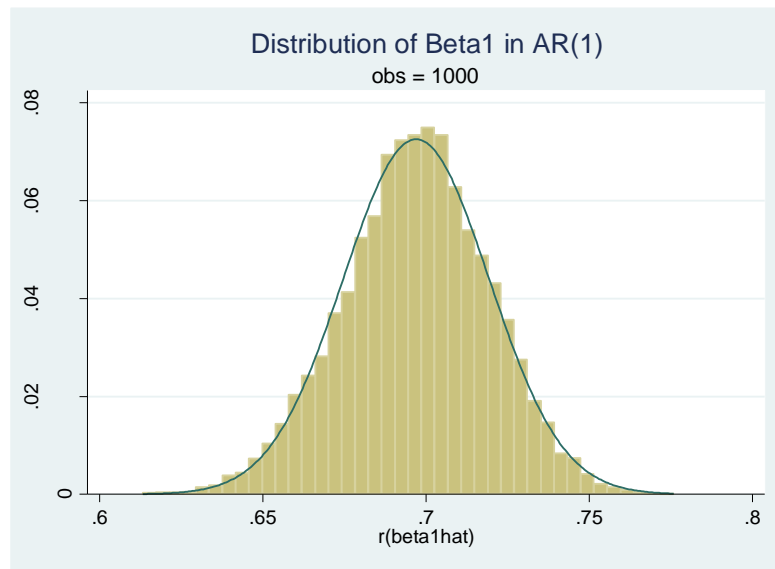
An AR model is different from the models we have been looking at so far because the explanatory variable and the independent variable are no longer independent. The regressor is no longer exogenous; the lagged variable is endogenous, or dependent, on the variable we are trying to study.

I examine the AR model using the least squares estimator. In the next section, I look at two tests that help us understand the model more in depth – the Durbin-Watson test and the Dickey-Fuller test.

### 3.5.1 Least Squares Estimator

For the model  $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$ , we consider the least squares estimator. With 100 observations, the distribution of  $\hat{\beta}_1$  is slightly skewed to the left. However, with 1000 observations,  $\hat{\beta}_1$  is much closer to normal with 20% of the skewness reported compared to the one with 100 observations. In addition, the kurtosis also decreases with a bigger sample size. This, along with the Jarque-Bera statistic, illustrates the asymptotic normality of  $\hat{\beta}_1$  distribution in the AR(1) under the least squares method.





AR(1)	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	<i>t</i> -statistic	<i>JB</i> -statistic
$\beta_0$	100	0	0.011	1.13	0.007	3.55	-0.97	1.27
	1000	0	0.0003	0.32	-0.04	3.05	-0.10	0.36
$\beta_1$	100	0.7	0.67	0.078	-0.51	3.39	42.31	4.97
	1000	0.7	0.70	0.022	-0.12	3.01	13.64	2.59

## 4. DISTRIBUTION OF TEST STATISTICS

Suppose that we have a correctly-specified model to explain a given data set, then the natural next task is to draw statistical inferences about the population parameter. In order to do so, we often need to assume several characteristics about the population characteristics, such as homoscedasticity, no autocorrelation, and stationarity for time series models. Then, statistical tests are proposed to check the validity of various assumptions. Failing to meet these assumptions may completely alter the inference about the parameters of interest, which makes performing these tests and interpreting the statistics crucial.

The Durbin-Watson Test and Dickey-Fuller Test both compare a sample statistic calculated to a distribution of the statistic. Monte Carlo simulation makes a significant contribution since most test statistic distributions are not standard distributed, and have to be generated through numerical simulations for each data generating process (DGP) and observation size. Without a simulation technique, the distribution of the statistic often cannot be created and thus cannot be interpreted. In this section, I show how estimators become biased when certain assumptions fail to hold. The demonstration of the significance of having the right test statistic distribution to detect nonstationarity follows.

#### 4.1 Durbin-Watson Test for Autocorrelation

Suppose the error terms of a model are autocorrelated, meaning they are dependent on their lags.

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

$$u_t = \rho u_{t-1} + v_t$$

$v_t$  has the usual properties.

This will influence our inference judgment about the regression coefficients,  $\beta_0$  and  $\beta_1$ , even though they remain unbiased. The Durbin-Watson test is used to detect if there is autocorrelation in the error terms; depending on the range the statistic falls into, we can detect if there is an autocorrelation present or not.<sup>12</sup>

$$d = \frac{\sum (\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2}$$

Durbin and Watson (1950, 1951, 1971) show that this statistic can be estimated by a simpler equation below.

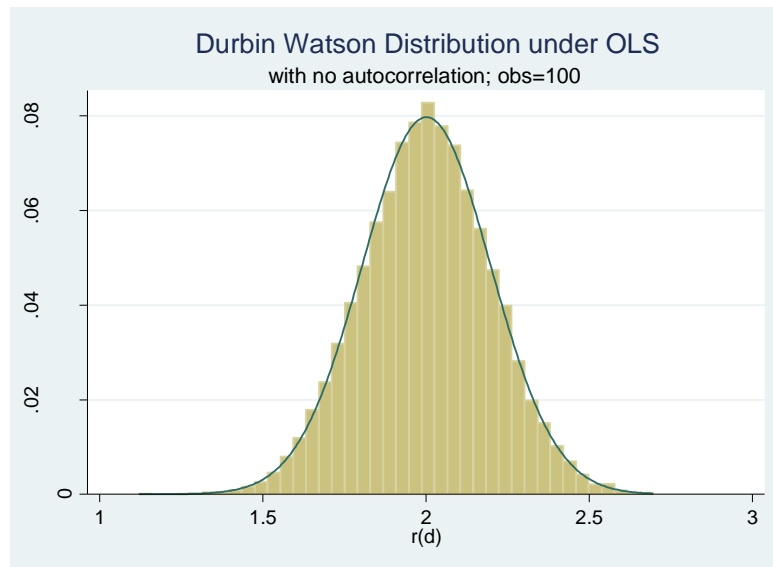
$$d \approx 2(1 - \rho)$$

Given that  $\rho = 0$  under the null hypothesis, we can conclude errors are not correlated if  $d = 2$ . Because the statistic is dependent on the DGP of  $Y_t$ , getting critical values require extensive simulations. Durbin and Watson (1951) suggests the upper bound and lower bound for critical values that are less sensitive of  $Y_t$ , and those values can be found in Durbin-Watson table.

Consider a model  $Y_t = \beta_0 + \beta_1 X_t + u_t$ . The Durbin-Watson distribution with no autocorrelation in errors and 100 observations looks like the following.

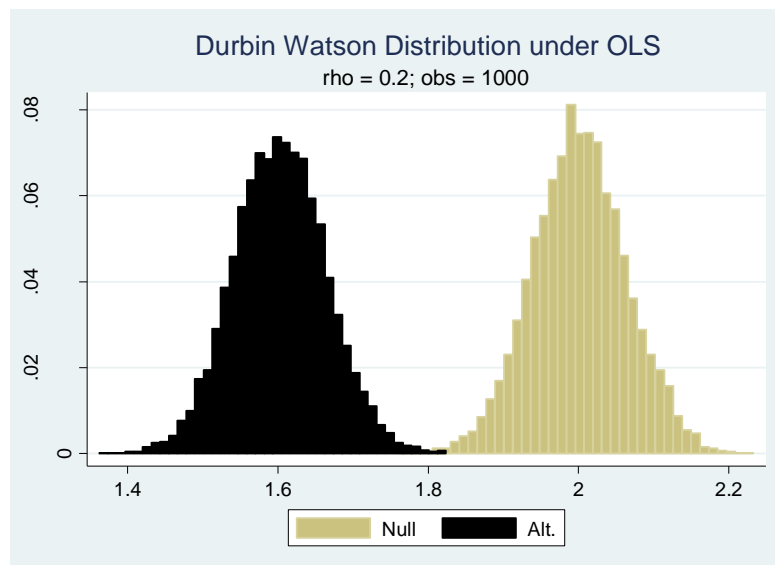
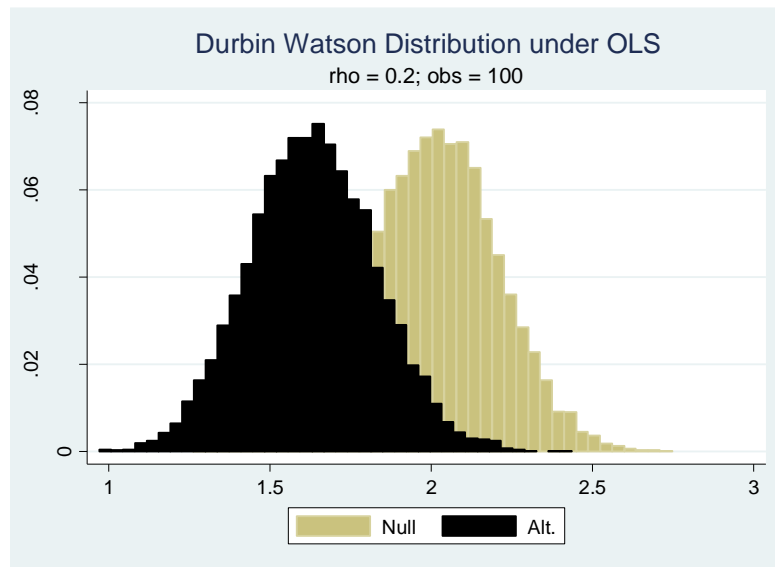
---

<sup>12</sup> Durbin and Watson (1950)

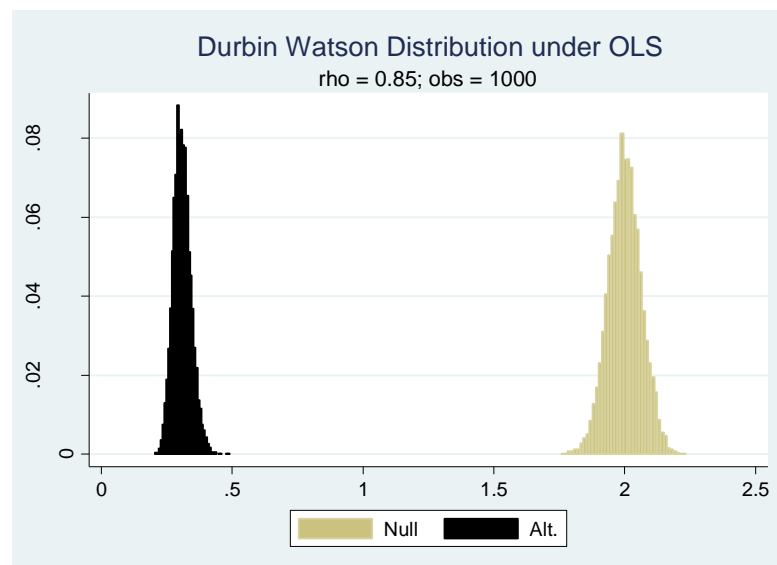
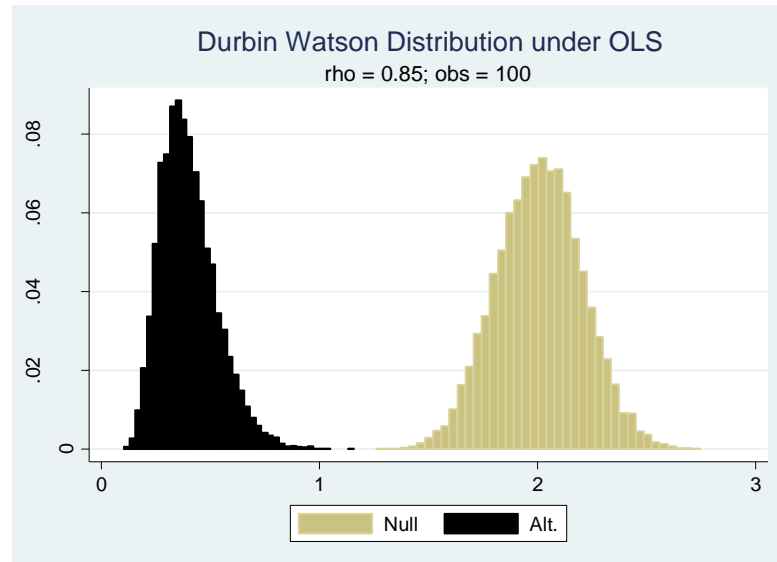


Now assume an autocorrelation  $u_t = \rho u_{t-1} + v_t$  where  $\rho = 0.2$ . Null hypothesis here is that there is no autocorrelation, or  $\rho = 0$ , and  $p$ -values for the test are shown in the table. It represents the likelihood that the observed samples can be found in the population under the null hypothesis. The null hypothesis is more likely to be rejected if it has a lower  $p$ -value. There is 8.86% chance of finding such data if there is no autocorrelation with sample size of 100, and 0.00042% with 1000. Consequently, we reject the null of no autocorrelation.





If we have a higher autocorrelation,  $\rho = 0.85$ , we have significantly lower  $p$ -values to reject the null hypothesis of no autocorrelation, as desired.



		Obs	Mean	StDev	Skewness	Kurtosis	p-value	JB-statistic
<b>Null</b>	<i>d</i>	100	2.00	0.20	-0.02	2.99		0.004
<b>Alt.</b>	<i>d</i>	100	1.63	0.20	0.09	2.95	0.09	0.14
$\rho =$	<i>d</i>	1000	1.60	0.06	0.03	3.04	0.000004	0.28
0.2	<i>d</i>	100	0.40	0.12	0.77	3.86	0	13.00
	<i>d</i>	1000	0.31	0.03	0.32	3.18	0	18.74

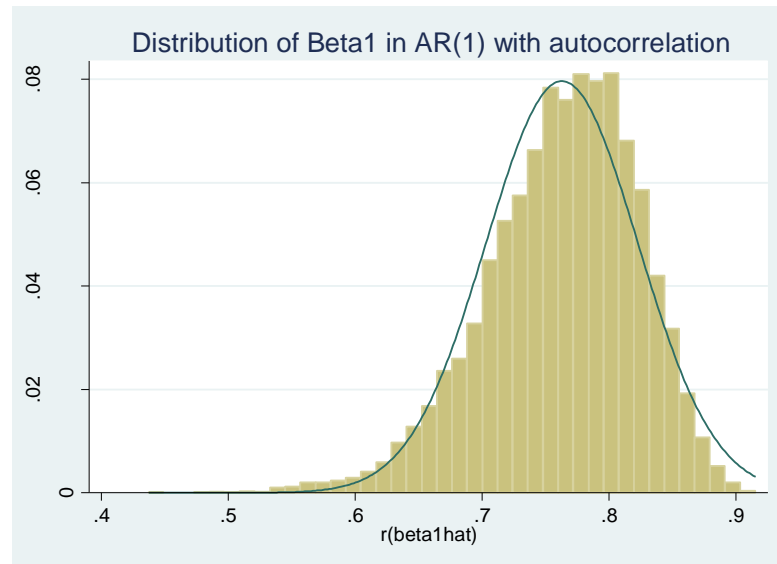
#### 4.1.1 Autocorrelation in AR model

Let's assume an AR(1) model for  $Y_t$  with autocorrelation in the error term.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + v_t$$

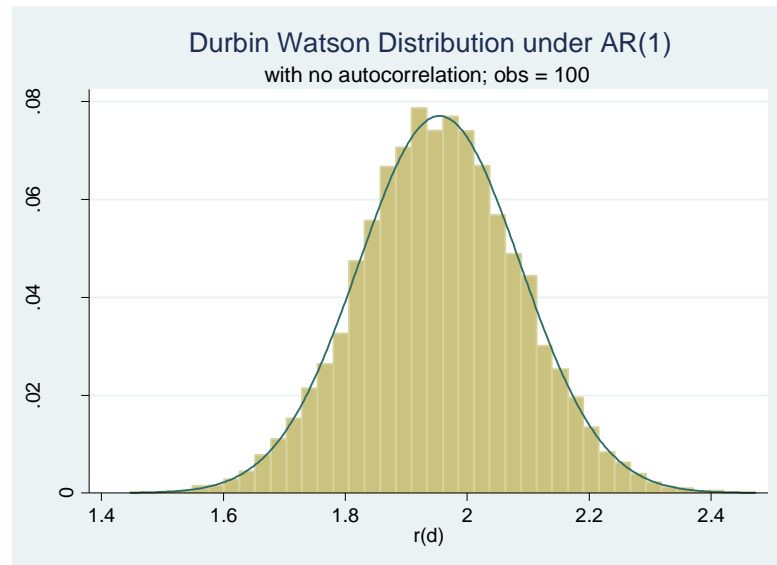
The following distribution illustrates the biasedness in the distribution of the estimator  $\beta_1$ .



$\rho =$	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	t-statistic	JB-statistic
<b>0.2</b>								
$\beta_0$	100	0.3	0.24	0.31	0.27	3.72	19.36	3.38
$\beta_1$	100	0.7	0.76	0.06	-0.60	3.56	-100	7.31

The distribution of the statistic with no autocorrelation under an autoregressive model is as follows: the distribution of the Durbin-Watson statistic is no longer unbiased

– its expected value is no longer 2. This means even if there is no autocorrelation, Durbin-Watson test will falsely report an autocorrelation.



	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	<i>t</i> -statistic	<i>JB</i> -statistic
<b><i>d</i></b>	100	2	1.95	0.13	0.02	3.10	38.46	0.05

The Durbin-Watson statistic is no longer accurate because in autoregressive models, a variable depends on its own previous values which include the errors – so the Durbin-Watson statistic underestimates the actual autocorrelation in the errors. Consequently, Durbin (1970) constructed the Durbin *h*-statistic which corrects for this inaccuracy.

## 4.2 Dickey-Fuller Test for Nonstationarity

### 4.2.1 Nonstationarity

The majority of macroeconomic variables, such as GDP, are trended. A variable is stationary if all moments remain constant over time. Any data set with a trend is no longer stationary, meaning the average and variance would change depending on time.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

With one lagged dependent variable, the regressor is stationary if  $|\beta_1| < 1$ ; it is not stationary if it has a unit root, or  $\beta_1 = 1$ . With more than one regressor,  $|\beta_1| < 1$  does indicate that it does not have a unit root, but it does not necessarily imply that it is stationary. Here, I just consider an autoregressive model with one regressor, so a unit root is equivalent to nonstationarity. I do not discuss the case when  $|\beta_1| > 1$ , where the variable exhibits an explosive behavior.

To test for the presence of a unit root, the Dickey-Fuller test is used. A Dickey-Fuller statistic does not have the usual  $t$ -distribution even in large samples. The distribution from simulations allows us to determine the critical values to accept or reject the null hypothesis, which is that there exists a unit root.

The test regression for Dickey-Fuller is as follows:

$$\Delta Y_t = (\beta_1 - 1)Y_{t-1} + u_t$$

or

$$\Delta Y_t = \varphi Y_{t-1} + u_t$$

and the equation for a Dickey-Fuller statistic is:

$$DF = \frac{\varphi}{SE(\hat{\beta})}$$

Augmented Dickey-Fuller (ADF) is a modified version of Dickey-Fuller that takes the possibility of an autocorrelation in the variable  $u_t$ . The test regression for Augmented Dickey-Fuller, therefore, is as follows:

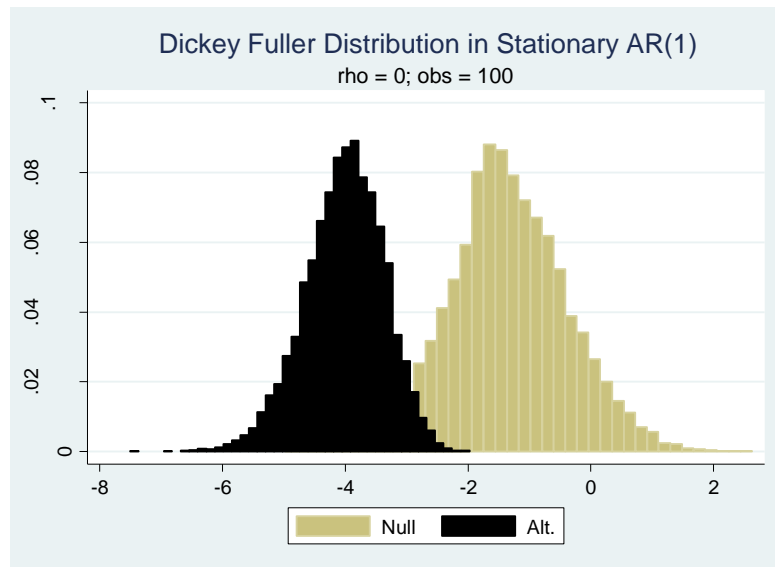
$$\Delta Y_t = a + bt + \varphi Y_{t-1} + (\theta_1 \Delta Y_{t-1} + \theta_2 \Delta Y_{t-2} + \dots + \theta_{p-1} \Delta Y_{t-p+1}) + u_t$$

where the second to the last element in parenthesis is the augmentation component, which removes the possible serial correlation. An information criterion, such as AIC and BIC, is used to select the appropriate value for  $p$ .

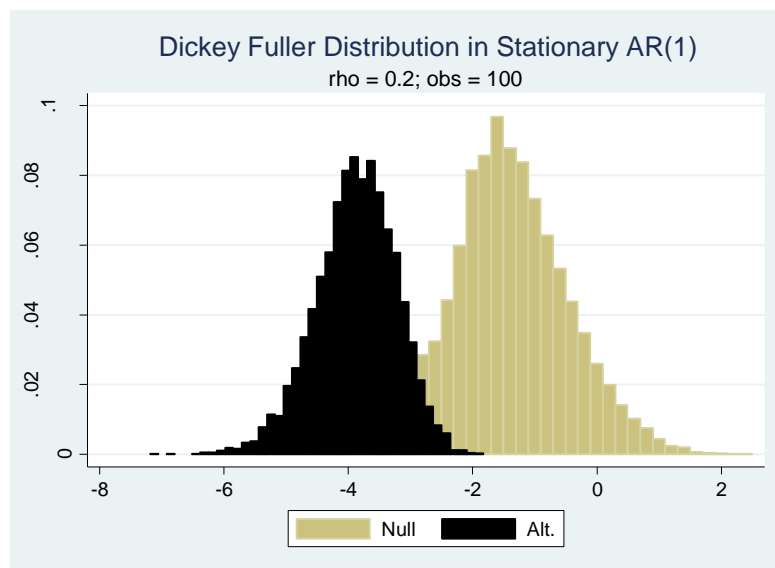
The statistic is usually negative, and more negative it is, the stronger the rejection becomes, and the more likely  $Y_t$  will be stationary. Because under the null hypothesis the nonstationarity of the data violates the assumption we established earlier, the Dickey-Fuller distribution does not resemble a standard  $t$ -distribution even though it is basically testing if  $\varphi$  is zero. In fact, the distribution is highly sensitive to the sample size and the DGP. For instance, the distribution will change in a presence of a drift term or and the absence of an intercept.

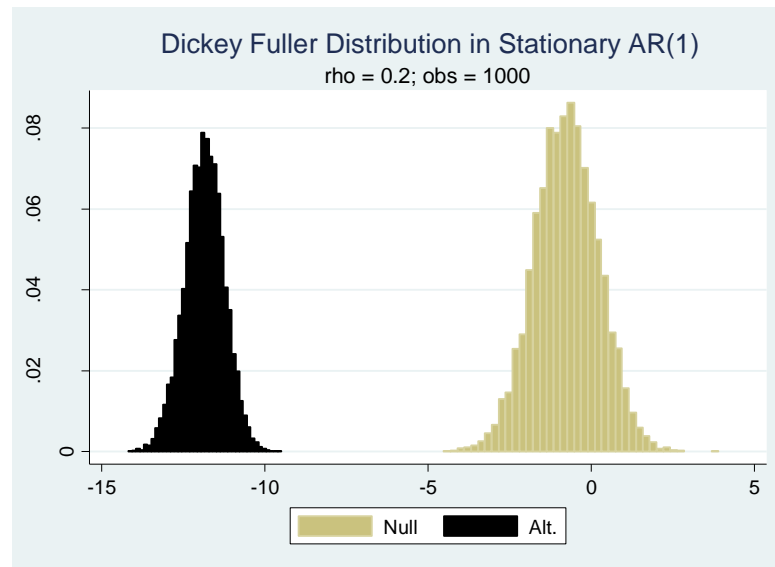
Consider the model  $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$  where  $\beta_1$  is 0.7 and 1 with initial condition  $Y_0 = 0$ . The null hypothesis will be that a unit root exists:  $\beta_1 = 1$ . Then, we graph the distribution of Dickey-Fuller statistic when  $\beta_1 = 0.7$  to see if it is significantly off the distribution under the null hypothesis – if it rejects the null.

Distributions for the null hypotheses are different for each DGP in the alternative hypotheses since the Dickey-Fuller distributions are distinct.



Autocorrelation does not seem to affect the distribution significantly, which confirms that ADF controls the autocorrelation. With a larger sample size, ADF is much more likely to reject the null if there is no unit root.





		Obs	$\beta_0$	$\beta_1$	$\rho$	Mean	StDev	Skewness	Kurtosis	<i>p</i> -value	<i>JB</i> statistic
<b>Null</b>	<i>ADF</i>	100	0.3	1	0	-1.32	0.94	0.18	3.14		0.62
	<i>ADF</i>	100	0.3	1	0.2	-1.39	0.92	0.19	3.26		0.88
	<i>ADF</i>	1000	0.3	1	0.2	-0.75	0.99	-0.0056	3.06		0.16
<b>Alt.</b>	<i>ADF</i>	100	0.3	0.7	0	-4.02	0.64	-0.35	3.28	0.0065	2.37
	<i>ADF</i>	100	0.3	0.7	0.2	-3.88	0.65	-0.34	3.27	0.011	2.23
	<i>ADF</i>	1000	0.3	0.7	0.2	-11.85	0.61	-0.09	3.07	0	1.55

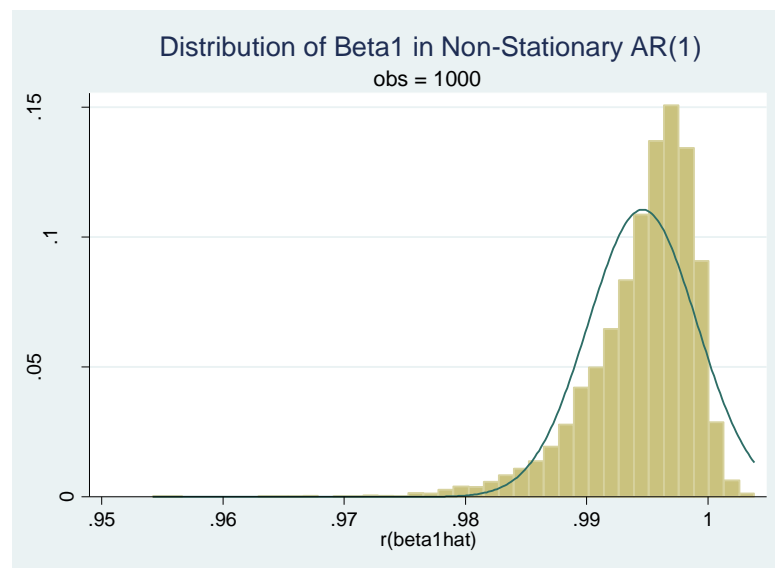
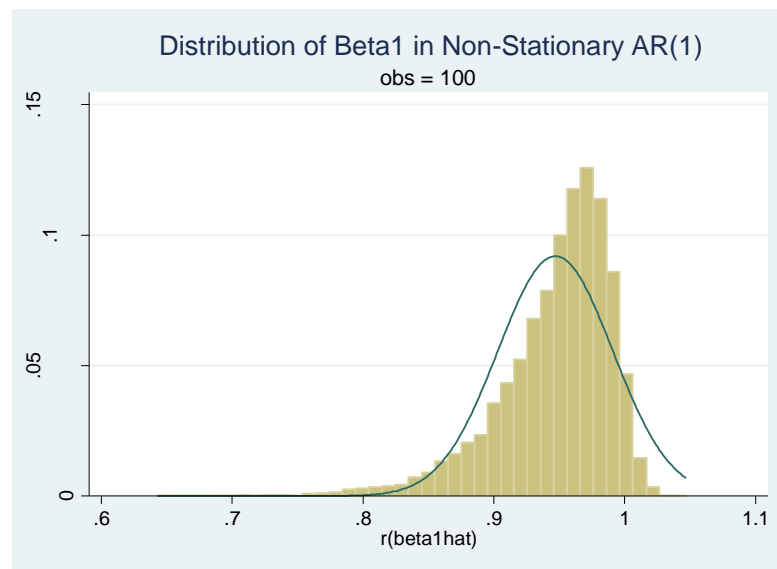
Nonstationarity is observed very often in a macroeconomic setting, and it causes substantial problems for three reasons:<sup>13</sup> the OLS estimator is no longer unbiased, and its coefficients become biased toward 0; the *t*-distribution used to draw statistical inference will no longer be normal; and a spurious regression might be observed where two independent variables may indicate false relationship due to the same trend. I simulate three cases to see how nonstationarity may lead us to draw inaccurate statistical inference.

<sup>13</sup> Stock and Watson (2014)



#### 4.2.2 OLS estimator no longer unbiased

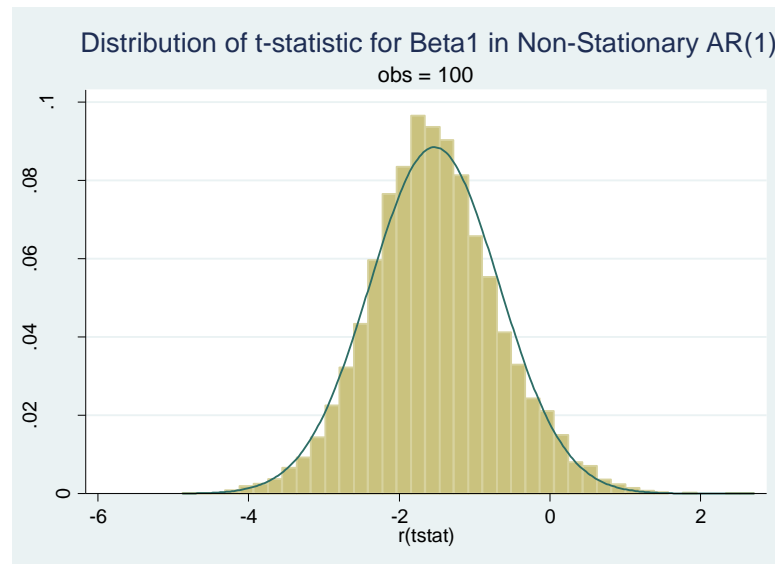
For a model  $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$  where  $\beta_1 = 1$  and  $Y_0 = 0$ , the distribution of the estimator  $\hat{\beta}_1$  looks as follows. The distribution no longer exhibits the normal shape like it did before, and the sample estimator for the population parameter is no longer unbiased, which can be shown from the high  $t$ -statistics.

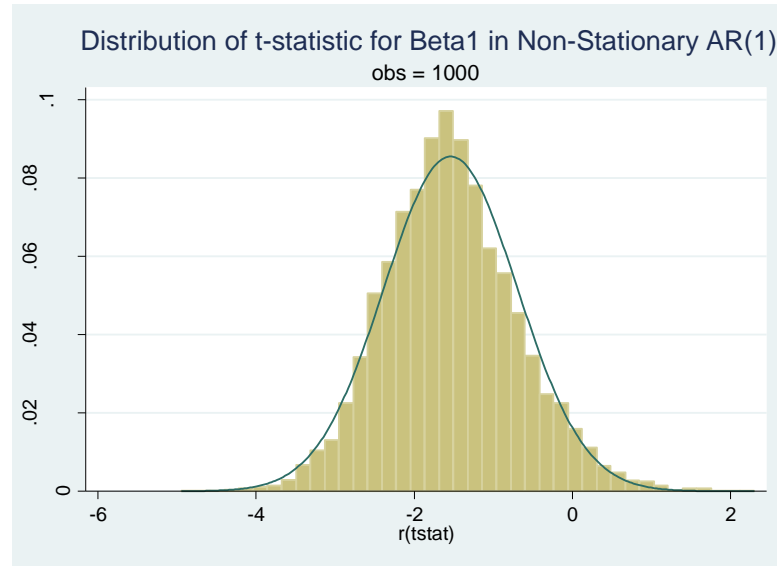


AR(1)	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	<i>t</i> -statistic	<i>JB</i> -statistic
$\beta_0$	100	0	0.05	3.59	0.011	3.94	-1.35	3.68
	1000	0	0.01	1.18	-0.03	4.18	-0.64	58.42
$\beta_1$	100	1	0.95	0.04	-1.48	6.55	121	88.89
	1000	1	0.99	0.0045	-1.55	7.40	120.8	1205

#### 4.2.3 *t*-statistics for Estimators No Longer Standard *t*-distributed

In addition to the OLS estimator being biased, the standard statistical inference using *t*-statistic is also problematic because its distribution is no longer centered at 0 and no longer *t*-distributed; increasing the sample size does not solve this problem. The *t*-statistics of the *t*-distribution with nonstationarity from the standard *t*-distribution indicate how different the distribution is from the correct mean 0.





AR(1)	Obs	True Mean	Mean	StDev	Skewness	Kurtosis	t-statistic	JB-statistic
<b>t-dist.</b>	100	0	-1.53	0.86	0.16	3.39	179.18	1.03
<b>t-dist.</b>	1000	0	-1.54	0.84	0.26	3.46	182.46	20.56

#### 4.2.4 Spurious Regression

Originally considered by Granger and Newbold (1974), the last, most common and obnoxious consequence of nonstationarity is a spurious regression. Consider two independent variables. First, only one of the two is stationary; second, neither are stationary.

For the first case, we have:

$$Y_{1,t} = \gamma + \theta Y_{1,t-1} + u_{1,t}$$

and

$$Y_{2,t} = \omega + Y_{2,t-1} + u_{2,t}$$

where  $\theta < 1$ . Here,  $Y_1$  and  $Y_2$  are independent of each other. However, regression results show that they are correlated with the regression equation  $Y_1 = \beta_0 + \beta_1 Y_2 + \varepsilon$ ,  $\beta_1 \neq 0$ . Fortunately, if the sample size increases,  $\beta_1$  approaches zero.

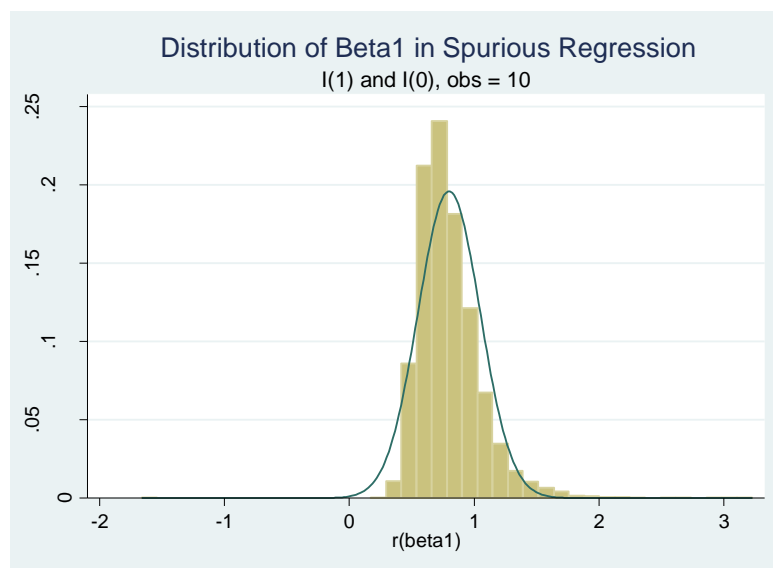
For the second case, the problem gets worse: increasing the sample size does not solve the problem. Now we have:

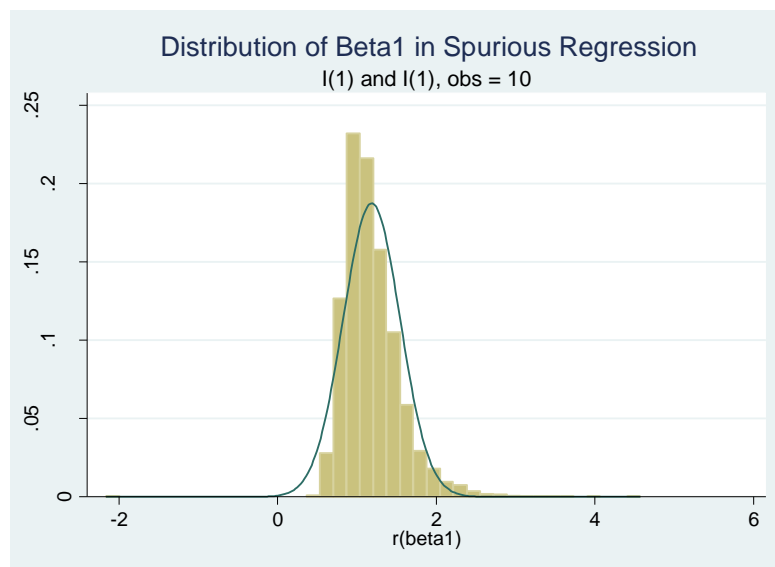
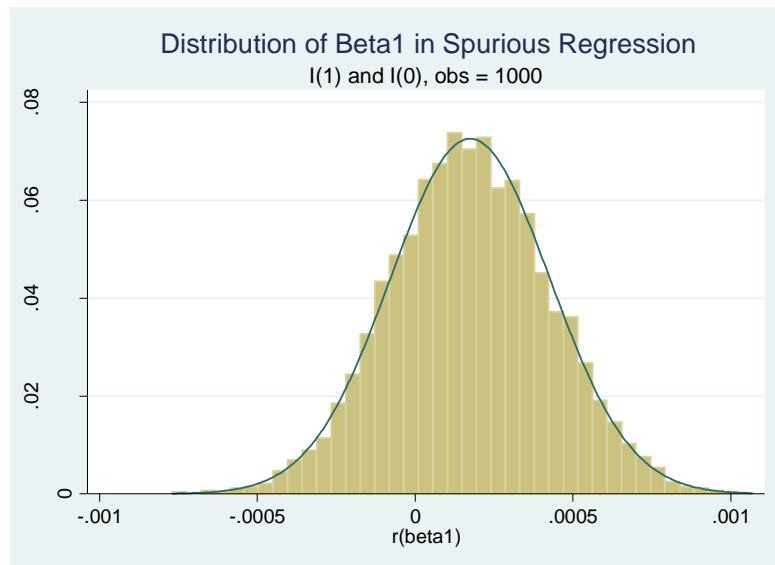
$$Y_{1,t} = \gamma + Y_{1,t-1} + u_{1,t}$$

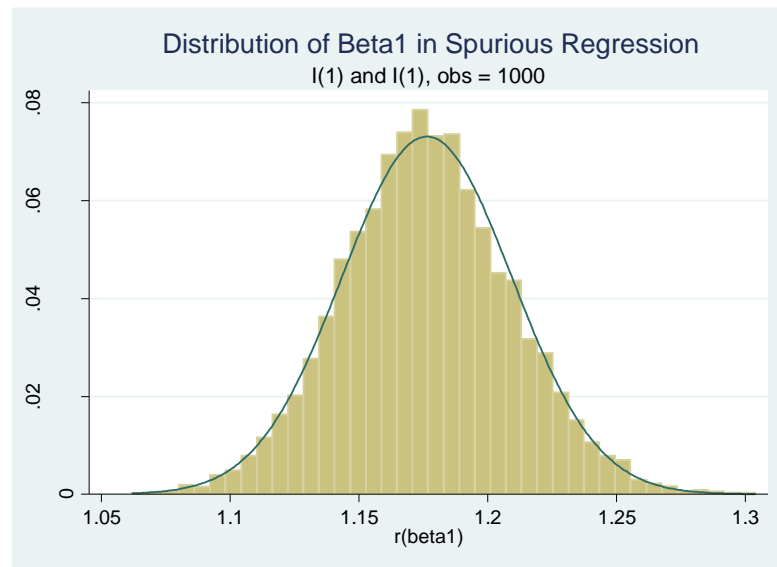
and

$$Y_{2,t} = \omega + Y_{2,t-1} + u_{2,t}$$

Here both of them have a unit root. When both of the variables exhibit stochastic trends, the asymptotic normality that we found earlier disappears. So when there is indeed no relationship between the two variables, the regression results will show that there is a statistically significant relationship.







$\beta_1$	Obs	True Mean	Mean	StdDev	Skewness	Kurtosis	t-statistic	JB statistic
<b>I(0), I(1)</b>	10	0	0.55	0.18	1.48	10.27	-300.83	25.71
	1000	0	0.0002	0.0003	0.01	2.96	-79.05	0.08
<b>I(1), I(1)</b>	10	0	1.19	0.36	1.50	9.36	-331.02	20.64
	1000	0	1.18	0.03	0.10	3.14	-3564	2.43

## 5. CONCLUSION

This paper studies a variety of applications of Monte Carlo simulation techniques in drawing statistical inference in regression analysis. One of the goals using econometric models is to draw statistical inferences about population parameters. Because the judgments on statistical inference are sensitive to the distribution of the estimators and statistics, analyses of the distributions are crucial. As shown with the OLS estimator, when the distribution of a population is known, it is possible to derive analytically how well a model explains the population and describe samples based on the model. When the distribution of the population is unknown, however, or hard to derive, simulation techniques are applied to create samples and draw statistical inferences based on their distributions. This paper analyzes both cases where analytical derivations are valid and invalid, and by producing the simulations, it reestablishes the significance of understanding the distribution to build statistical inferences.

With the advent of cheaper and faster computing resources, Monte Carlo techniques have taken on an increasingly important role in statistical analysis as well as in other fields, especially in providing a different route to look at traditional questions in mathematics. However, caution should always prevail. For instance, it is tempting to apply the method in finance to perform projection analysis on stock markets. In fact, it has been proposed that the peculiarities of the market might reduce the reliability of using

Monte Carlo simulations.<sup>14</sup> Although its applications for understanding econometric theories do not require such restraints, limitations of the simulations in studying econometric models and alternative solutions should be the topic of a future research related to the study of the technique in econometrics.

---

<sup>14</sup> Crawshaw (2003)



## 6. APPENDIX

A1.  $t$ -table<sup>15</sup>

one-tail two-tails	0.5 1	0.25 0.5	0.2 0.4	0.15 0.3	0.1 0.2	0.05 0.1	0.025 0.05	0.01 0.02	0.005 0.01	0.001 0.002	0.0005 0.001
df											
1	0.00	1.00	1.38	1.96	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	0.00	0.82	1.06	1.39	1.89	2.92	4.30	6.97	9.93	22.33	31.60
3	0.00	0.77	0.98	1.25	1.64	2.35	3.18	4.54	5.84	10.22	12.92
4	0.00	0.74	0.94	1.19	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	0.00	0.73	0.92	1.16	1.48	2.02	2.57	3.37	4.03	5.89	6.87
6	0.00	0.72	0.91	1.13	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	0.00	0.71	0.90	1.12	1.42	1.90	2.37	3.00	3.50	4.79	5.41
8	0.00	0.71	0.89	1.11	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	0.00	0.70	0.88	1.10	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	0.00	0.70	0.88	1.09	1.37	1.81	2.23	2.76	3.17	4.14	4.59
11	0.00	0.70	0.88	1.09	1.36	1.80	2.20	2.72	3.11	4.03	4.44
12	0.00	0.70	0.87	1.08	1.36	1.78	2.18	2.68	3.06	3.93	4.32
13	0.00	0.69	0.87	1.08	1.35	1.77	2.16	2.65	3.01	3.85	4.22
14	0.00	0.69	0.87	1.08	1.35	1.76	2.15	2.62	2.98	3.79	4.14
15	0.00	0.69	0.87	1.07	1.34	1.75	2.13	2.60	2.95	3.73	4.07
16	0.00	0.69	0.87	1.07	1.34	1.75	2.12	2.58	2.92	3.69	4.02
17	0.00	0.69	0.86	1.07	1.33	1.74	2.11	2.57	2.90	3.65	3.97
18	0.00	0.69	0.86	1.07	1.33	1.73	2.10	2.55	2.88	3.61	3.92
19	0.00	0.69	0.86	1.07	1.33	1.73	2.09	2.54	2.86	3.58	3.88
20	0.00	0.69	0.86	1.06	1.33	1.73	2.09	2.53	2.85	3.55	3.85
21	0.00	0.69	0.86	1.06	1.32	1.72	2.08	2.52	2.83	3.53	3.82
22	0.00	0.69	0.86	1.06	1.32	1.72	2.07	2.51	2.82	3.51	3.79
23	0.00	0.69	0.86	1.06	1.32	1.71	2.07	2.50	2.81	3.49	3.77
24	0.00	0.69	0.86	1.06	1.32	1.71	2.06	2.49	2.80	3.47	3.75
25	0.00	0.68	0.86	1.06	1.32	1.71	2.06	2.49	2.79	3.45	3.73
26	0.00	0.68	0.86	1.06	1.32	1.71	2.06	2.48	2.78	3.44	3.71
27	0.00	0.68	0.86	1.06	1.31	1.70	2.05	2.47	2.77	3.42	3.69
28	0.00	0.68	0.86	1.06	1.31	1.70	2.05	2.47	2.76	3.41	3.67
29	0.00	0.68	0.85	1.06	1.31	1.70	2.05	2.46	2.76	3.40	3.66
30	0.00	0.68	0.85	1.06	1.31	1.70	2.04	2.46	2.75	3.39	3.65
40	0.00	0.68	0.85	1.05	1.30	1.68	2.02	2.42	2.70	3.31	3.55
60	0.00	0.68	0.85	1.05	1.30	1.67	2.00	2.39	2.66	3.23	3.46
80	0.00	0.68	0.85	1.04	1.29	1.66	1.99	2.37	2.64	3.20	3.42
100	0.00	0.68	0.85	1.04	1.29	1.66	1.98	2.36	2.63	3.17	3.39
1000	0.00	0.68	0.84	1.04	1.28	1.65	1.96	2.33	2.58	3.10	3.30
z	0.00	0.67	0.84	1.04	1.28	1.65	1.96	2.33	2.58	3.09	3.29
Confid. Level	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.80%	99.90%

<sup>15</sup> Reproduced standard  $t$ -table which can be found from variety of sources

## A2. Jarque-Bera Test Distribution<sup>16</sup>

<b>One-tail</b>	<b>0.995</b>	<b>0.975</b>	<b>0.2</b>	<b>0.1</b>	<b>0.05</b>	<b>0.025</b>	<b>0.02</b>	<b>0.01</b>	<b>0.005</b>	<b>0.002</b>	<b>0.001</b>
<b><i>p</i>-value</b>	0.01	0.05	3.22	4.61	6.00	7.38	7.82	9.21	10.60	12.43	13.82

## A3. STATA Code for Algorithms

### A3.1. Estimation of a Mean

```
# delimit ;
set more off;
clear all;

program meantest, rclass;
    drop _all;
    set obs 100;
    scalar et_sd = 10;
    gen Y = 70+et_sd*invnormal(runiform());
    summarize Y;
    return scalar mean = r(mean);
end;

meantest;

simulate mean = r(mean),reps(10000) nodots seed(1000): meantest;
histogram mean, fraction
    title("Distribution of Mean under OLS")
    name("graph1");
```

### A3.2. Estimation of a Median

```
# delimit ;
set more off;
clear all;

program mediantest, rclass;
    drop _all;
    set obs 100;
    scalar et_sd = 10;
    gen Y = 70+et_sd*invnormal(runiform());
    summarize Y;
    sort Y;
    scalar k=r(N);
    return scalar median = Y[k/2];
end;

mediantest;
simulate median = r(median),reps(10000) nodots seed(1000): mediantest;
```

---

<sup>16</sup> Reproduced chi-squared distribution with two degrees of freedom, which can be found from variety of sources

### A3.3. OLS Estimator

```
# delimit ;
set more off;
clear all;

program fixedx, rclass;
    drop _all;
    set obs 100;
    range X 150 200;          //for random sample, use: gen X = 150+50*runiform()
    scalar beta0hat = 42;
    scalar beta1hat = 2.3;
    scalar et_sd = 10;
    gen Y = beta0hat + beta1hat*X + et_sd*invnormal(runiform());
    summarize Y;
    reg Y X;
    return scalar beta0hat = _b[_cons];
    return scalar beta1hat = _b[X];
end;

fixedx;
simulate beta0 = r(beta0hat) beta1 = r(beta1hat),reps(10000) nodots seed(1000): fixedx;
```

### A3.4. LAD Estimator

```
# delimit ;
set more off;
clear all;

program fixedLAD, rclass;
    drop _all;
    set obs 100;
    range X 150 200;          //for random sample, use: gen X = 150+50*runiform()
    scalar beta0hat = 42;
    scalar beta1hat = 2.3;
    scalar et_sd = 10;
    gen Y = beta0hat + beta1hat*X + et_sd*invnormal(runiform());
    summarize Y;
    qreg Y X;
    return scalar beta0hat = _b[_cons];
    return scalar beta1hat = _b[X];
end;

fixedLAD;
simulate beta0 = r(beta0hat) beta1 = r(beta1hat),reps(10000) nodots seed(100): fixedLAD;
```

### A3.5. A2P Estimator

```
# delimit ;
set more off;
clear all;

program fixedA2P, rclass;
    drop _all;
    set obs 100;
    range X 150 200;          //for random sample, use: gen X = 150+50*runiform()
    scalar beta0hat = 42;
    scalar beta1hat = 2.3;
    scalar et_sd = 10;
    gen Y = beta0hat + beta1hat*X + et_sd*invnormal(runiform());
    summarize Y;
```

```

    scalar i = floor(100*runiform()+1);
    scalar j = floor(100*runiform()+1);
    while i==j {
        scalar i = floor(100*runiform()+1);
        scalar j = floor(100*runiform()+1);
    };

    return scalar beta0hat = (X[j]*Y[i] - X[i]*Y[j]) / (X[j] - X[i]);
    return scalar beta1hat = (Y[j] - Y[i]) / (X[j] - X[i]);
end;

fixedA2P;
simulate beta0 = r(beta0hat) beta1 = r(beta1hat),reps(10000) nodots seed(100): fixedA2P;

```

### A3.6. Estimation of an AR(1) model

```

# delimit ;
set more off;
clear all;

program arls, rclass;
    drop _all;
    set obs 100;
    scalar beta0hat = 0;
    scalar beta1hat = 0.7;
    scalar et_sd = 10;
    gen t=_n;
    tsset t;
    gen y`i' = 0 if _n==1;           //adjust the initial condition
    replace y`i' = beta0hat + beta1hat*y`i'[_n-1] + et_sd*invnormal(runiform()) if _n > 1;

    reg y L.y;
    return scalar beta0hat = _b[_cons];
    return scalar beta1hat = _b[L.y];
end;

arls;
simulate beta0 = r(beta0hat) beta1 = r(beta1hat),reps(10000) nodots seed(1000): arls;

```

### A3.7. Durbin-Watson Statistics

```

# delimit ;
set more off;
clear all;

program dw_ac, rclass;
    drop _all;
    set obs 100;
    scalar beta0hat = 42;
    scalar beta1hat = 2.3;
    scalar rho = 0.85;               //adjust the correlation coefficient
    scalar vt_sd = 3;
    gen t=_n;
    tsset t;
    gen X = invnormal(runiform());   //X generated from a normal distribution
    gen u`i' = 0 if _n==1;
    replace u`i' = rho * u`i'[_n-1] + vt_sd * invnormal(runiform()) if _n > 1;
    gen Y = beta0hat + beta1hat*X + u`i';
    summarize Y;
    reg Y X;

```

```

        estat dwatson;
        return scalar d = r(dw);
end;

dw_ac;
simulate d = r(d), reps(10000) nodots seed(1000): dw_ac;

```

### A3.8. Durbin-Watson Statistics for AR(1) model

```

# delimit ;
set more off;
clear all;

program arac_dw, rclass;
    drop _all;
    set obs 100;
    scalar beta0hat = 0.3;           //adjust the intercept
    scalar beta1hat = 0.7;
    scalar rho = 0;                 //adjust the correlation coefficient
    scalar vt_sd = 3;
    gen t=_n;
    tsset t;
    gen u`i' = 0 if _n==1;
    replace u`i' = rho * u`i'[_n-1] + vt_sd * invnormal(runiform()) if _n > 1;

    gen y`i' = 0 if _n==1;
    replace y`i' = beta0hat + beta1hat*y`i'[_n-1] + u`i' if _n > 1;

    reg y L.y;
    estat dwatson;
    return scalar d = r(dw);
end;

arac_dw;
simulate d = r(d), reps(10000) nodots seed(1000): arac_dw;

```

### A3.9. *p*-value for Durbin-Watson Statistics

```

# delimit ;
set more off;
clear all;

program dw_p, rclass;
    drop _all;
    set obs 100;
    scalar beta0hat = 42;
    scalar beta1hat = 2.3;
    scalar rho = 0;                 //null hypothesis : no autocorrelation
    scalar vt_sd = 3;
    gen t=_n;
    tsset t;
    gen X = invnormal(runiform());
    gen u`i' = 0 if _n==1;
    replace u`i' = rho * u`i'[_n-1] + vt_sd * invnormal(runiform()) if _n > 1;
    gen Y = beta0hat + beta1hat*X + u`i';
    summarize Y;
    reg Y X;
    estat dwatson;
    return scalar null_dw = r(dw);
end;

```

```

drop _all;
set obs 100;
scalar beta0hat = 42;
scalar beta1hat = 2.3;
scalar rho = 0.2;           //adjust the correlation coefficient
scalar vt_sd = 3;
gen t=_n;
tsset t;
gen X = invnormal(runiform());
gen u`i' = 0 if _n==1;
replace u`i' = rho * u`i'[_n-1] + vt_sd * invnormal(runiform()) if _n > 1;
gen Y = beta0hat + beta1hat*X + u`i';
summarize Y;
reg Y X;
estat dwatson;

return scalar alt_dw = r(dw);
end;

dw_p;
simulate null_dw = r(null_dw) alt_dw = r(alt_dw) , reps(10000) nodots seed(1000): dw_p;

scalar p=0;
forvalues i=1/10000 {;
    quietly count if alt_dw[`i'] > null_dw;
    scalar p = p + r(N);
};

scalar pvalue=p/10000^2;
di pvalue;

```

### A3.10. *p*-value for Dickey-Fuller Statistics

```

# delimit ;
set more off;
clear all;

program df_p, rclass;
drop _all;
set obs 100;
scalar beta0hat = 0.3;           //adjust the intercept
scalar beta1hat = 1;             //null hypothesis : nonstationary
scalar rho = 0;                 //adjust the correlation coefficient
scalar vt_sd = 3;
gen t=_n;
tsset t;
gen u`i' = 0 if _n==1;
replace u`i' = rho * u`i'[_n-1] + vt_sd * invnormal(runiform()) if _n > 1;

gen y`i' = 0 if _n==1;
replace y`i' = beta0hat + beta1hat*y`i'[_n-1] + u`i' if _n > 1;
dfuller y, lags(1);
return scalar null_df = r(Zt);

drop _all;
set obs 100;
scalar beta0hat = 0.3;           //adjust the intercept; consistent with the null
scalar beta1hat = 0.7;
scalar rho = 0;                 //adjust the correlation coeff; consistent with the null
scalar vt_sd = 3;
gen t=_n;
tsset t;
gen u`i' = 0 if _n==1;
replace u`i' = rho * u`i'[_n-1] + vt_sd * invnormal(runiform()) if _n > 1;

```

```

        gen y`i' = 0 if _n==1;           //adjust the initial condition
        replace y`i' = beta0hat + beta1hat*y`i'[_n-1] + u`i' if _n > 1;

        dfuller y, lags(1);

        return scalar alt_df = r(Zt);
end;

df_p;
simulate null_df = r(null_df) alt_df = r(alt_df) , reps(10000) nodots seed(1000): df_p;

scalar p=0;
forvalues i=1/10000 {;
    quietly count if alt_df[`i'] > null_df;
    scalar p = p + r(N);
};

scalar pvalue=p/10000^2;
di pvalue;

```

### A3.11. Spurious Regression

```

# delimit ;
set more off;
clear all;

program spurious2, rclass;
    drop _all;
    set obs 10;
    scalar beta0hat1 = 2;
    scalar beta0hat2 = 1.7;
    scalar beta1hat1 = 0.8;           //I(0); make it 1 to make it nonstationary
    scalar beta1hat2 = 1;             //I(1)
    scalar et_sd1 = 0.8;
    scalar et_sd2 = 1.2;
    gen t=_n;
    tsset t;
    gen y1`i' = 0 if _n==1;
    replace y1`i' = beta0hat1 + beta1hat1*y1`i'[_n-1] + et_sd1*invnormal(runiform()) if _n > 1;
    gen y2`i' = 0 if _n==1;
    replace y2`i' = beta0hat2 + beta1hat2*y2`i'[_n-1] + et_sd2*invnormal(runiform()) if _n > 1;

    reg y1 y2;
    return scalar beta1 = _b[y2];
end;

spurious2;
simulate beta1 = r(beta1),reps(10000) nodots seed(1000): spurious2;

```

## 7. BIBLIOGRAPHY

- Crawshaw, Julie. "Why Some Advisors Just Say No to Monte Carlo Simulations." *WealthManagement*. May 1, 2003. Accessed January 20, 2015. <http://wealthmanagement.com/practice-management/why-some-advisors-just-say-no-monte-carlo-simulations>.
- Durbin, J. "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables." *Econometrica* 38, no. 3 (1970): 410-21. Accessed March 19, 2015. <http://www.jstor.org/stable/1909547>.
- Durbin, J., and G. S. Watson. "Testing for Serial Correlation in Least Squares Regression: I." *Biometrika* 37, no. 3/4 (1950): 409-28. Accessed February 10, 2015. <http://www.jstor.org/stable/2332391>.
- Granger, C.W.J, and P. Newbold. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2, no. 2, 111-20. Accessed April 9, 2015. [http://ac.els-cdn.com/0304407674900347/1-s2.0-0304407674900347-main.pdf?\\_tid=5a58ea66-ec47-11e4-86da-00000aab0f6b&acdnat=1430075373\\_74734b9aa64195329babfe61a880fed8](http://ac.els-cdn.com/0304407674900347/1-s2.0-0304407674900347-main.pdf?_tid=5a58ea66-ec47-11e4-86da-00000aab0f6b&acdnat=1430075373_74734b9aa64195329babfe61a880fed8).
- Harrison, Robert L. "Introduction To Monte Carlo Simulation." *AIP conference proceedings*, 1204 (2010): 17-21. Accessed Feb 17, 2015.
- Kmenta, Jan. *Elements of Econometrics*. New York: Macmillan (1971).
- Metropolis, Nicholas and S. Ulam. "The Monte Carlo Method," *Journal of the American Statistical Association* 445 (1949): 335-341.
- Pollard, David. "Asymptotics for Least Absolute Deviation Regression Estimators." *Econometric Theory* 7, no. 2 (1991): 186. Accessed March 7, 2015. <http://www.jstor.org/stable/3532043>.
- Stock, James H., and Mark W. Watson. "5. Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals." In *Introduction to Econometrics, Update*, 180-181. 3rd ed. Boston: Pearson/Addison Wesley, 2014
- Stock, James, and Mark Watson. "18. The Theory of Multiple Regression." In *Introduction to Econometrics, Update*, 720-722. 3rd ed. Boston: Pearson/Addison Wesley, 2014.



Van der Vaart, H. R. "Some Extensions of the Idea of Bias." *The Annals of Mathematical Statistics* 32, no. 2 (1962): 436-47.

Wikipedia contributors, "Monte Carlo method," *Wikipedia, The Free Encyclopedia*, Accessed February 2, 2015. [http://en.wikipedia.org/wiki/Monte\\_Carlo\\_method](http://en.wikipedia.org/wiki/Monte_Carlo_method).