9/26/2013

**Simulation and Risk Analytics: Homework 1**

# Monte Carlo Simulation as a device to verify the properties of a model or a theory

Orange Team 6:

*Team Lead:* Phillip **Domschke**

*Other Contributors:* Marc Zimmerman, Steve Neola, Wes Ledebuhr, Jacob Frost

# TABLE OF CONTENTS

Monte Carlo Simulation as a device to verify the properties of a model or a theory 0

TABLE OF CONTENTS 1

EXECUTIVE SUMMARY 2

ASSUMPTIONS 3

Model Assumptions 3

ANALYSIS 3

Part A – Is the distribution of the betas the one suggested in theory? 3

Part B – How many times is incorrectly rejected? 4

Part C – Change in Variance and Introduction of Heteroscedasticity 4

Part D – Omitting X3 to Explore Possible Bias 4

Part E – Introducing Correlation between Variables 5

CONCLUSION 5

APPENDIX A: Distribution Graphs 6

# 

# EXECUTIVE SUMMARY

In this report, we will show through Monte Carlo simulations that the parameter estimates of linear regression models will follow a normal distribution following the central limit theorem and the law of big numbers. We will also show that heteroscedasticity among the error will not affect the normality assumption, however it will affect the associated t-test in some degree. Further, we demonstrate that strong correlation between two variables will heavily influence the rejection rate of our parameter estimates to unacceptable levels of more than 99%.

We will break down the individual tests into five distinct parts A, B, C, D, and E for readability.

# ASSUMPTIONS

## Model Assumptions

True regression equation, which will be verified through simulation:

Where,

# ANALYSIS

## Part A – Is the distribution of the betas the one suggested in theory?

In order to answer this and the other question we rely on Monte Carlo simulations, in which we simulate sets of 100 random observations, which follow the above described distributions. To get reliable results, we repeat the process 20,000 times and evaluate the aggregate results.

As shown in the results table below, all measures indicate that the parameter estimates follow a normal distribution. The mean and median are almost equal and very close to the true parameter. Furthermore, excess kurtosis and skewness are approximately zero, which is another indicator for a normal distribution of the parameter estimates. Finally, the Anderson-Darling test verifies these indicators. In the Appendix are additional QQ plots as well as parameter distribution histograms to further illustrate the validity of these results.

Table 1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | True Parameter | Mean | Median | Variance | Excess Kurtosis | Skewness | Anderson-Darling | |
| x1 | 0.21 | 0.20936 | 0.21133 | 0.12273 | -0.0214 | -0.0027 | 0.29346 | p>0.25 |
| x2 | -0.9 | 0.89775 | 0.89649 | 0.05363 | 0.10742 | -0.0067 | 1.87194 | p>0.089 |
| x3 | 3.45 | 3.44804 | 3.44661 | 0.06948 | 0.13759 | -0.0045 | 0.69287 | p>0.074 |

## Part B – How many times is incorrectly rejected?

Using a traditional alpha level of for the t-test, we find that 1038 records are misclassified and rejected. That equates to approximately 5.19% of all parameter estimates. This outcome was completely expected considering the alpha level controls the risk of committing a type I error. Therefore, we can expect the test to falsely reject the true parameter at the same level as alpha.

## Part C – Change in Variance and Introduction of Heteroscedasticity

In part C, we show the effect of the violation of the OLS assumption of independent errors. In order to simulate this effect, we change the variance of the error distribution to:

Table 2

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | True Parameter | Mean | | Median | Variance | Excess Kurtosis | Skewness | Anderson-Darling | |
| x1 | 0.21 | | 0.20874 | 0.21076 | 0.12241 | 0.00864 | -0.0047 | 0.26671 | p>0.25 |
| x2 | -0.9 | | -0.8977 | -0.896 | 0.07181 | 0.04452 | -0.0053 | 0.4815 | p>0.237 |
| x3 | 3.45 | | 3.44758 | 3.44685 | 0.06929 | 0.17352 | 0.00518 | 0.88809 | p>0.024 |

As we can see from Table 2, all the assumptions regarding normality still hold as well as the unbiased estimate of the true population parameter. However, we cannot trust the t-test anymore. This becomes evident in the fact that now 1879 out of all 20,000 simulations falsely reject the null hypothesis. This amounts to almost double the set alpha level of 5% with approximately 9.4%, respectively.

## Part D – Omitting X3 to Explore Possible Bias

In order to explore the effects of completely leaving out one of the explanatory variables in the parameter estimation process, we ran another set of 20,000 simulations with 100 observations each without this time. The results are very similar to the one we found in Part C. Table 3 illustrates that the normality and unbiased parameter estimates are still valid assumptions. As before, we cannot trust the results of our t-test since it falsely rejects the null hypothesis about 9.55% of the time, while it truly should commit this mistake only 5% of the time. This amounts to 1909 misclassifications.

Table 3

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | True Parameter | Mean | Median | Variance | Excess Kurtosis | Skewness | Anderson-Darling | |
| x1 | 0.21 | 0.20955 | 0.21107 | 0.12258 | 0.04844 | -0.0058 | 0.43387 | p>0.25 |
| x2 | -0.9 | -0.8992 | -0.9008 | 0.071 | 0.02756 | 0.00981 | 0.44047 | p>0.25 |

## Part E – Introducing Correlation between Variables

To show the effect of high correlation between two variables, we simulated the case in which have a correlation of 0.6. Despite the addition of correlation, the standard indicators of normality remained strong as shown in the table below. The key difference was that the mean of X2 shifted from -.9 to .86 as a result of the correlation that was introduced. Lastly, the Type I error rate was 99.91% at an alpha level of .05. This is clearly an unacceptable rate that would need to be resolved before using this aspect of the model.

Table 4

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | True Parameter | Mean | Median | Variance | Excess Kurtosis | Skewness | Anderson-Darling | |
| X2 | -.9 | 0.863756 | 0.86030 | 0.11868 | 0.088754 | 0.067314 | 1.414563 | <0.005 |

# CONCLUSION

In conclusion, through the Monte Carlo simulation technique, we were able to test various distributions and the effects on them when different factors are introduced to the model. This method proved to be effective at simulating large data sets without actually needing to have the data – assuming certain characteristics like the distributions are known.

# APPENDIX A: Distribution Graphs

















