Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

9-2-2016 12:00 AM

Advances in Portmanteau Diagnostic Tests

Jinkun Xiao, *The University of Western Ontario*

Supervisor: Dr. A. Ian McLeod, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

© Jinkun Xiao 2016

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Applied Statistics Commons, Longitudinal Data Analysis and Time Series Commons, and the Statistical Methodology Commons

Recommended Citation

Xiao, Jinkun, "Advances in Portmanteau Diagnostic Tests" (2016). *Electronic Thesis and Dissertation Repository*. 4100.
https://ir.lib.uwo.ca/etd/4100

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

Portmanteau test serves an important role in model diagnostics for Box-Jenkins Modelling procedures. A large number of portmanteau test based on the autocorrelation function are proposed for a general purpose goodness-of-fit test. Since the asymptotic distribution for the statistics has a complicated form which makes it hard to obtain the p-value directly, the gamma approximation is introduced to obtain the p-value. But the approximation will inevitably introduce approximation errors and needs a large number of observation to yield a good approximation. To avoid some pitfalls in the approximation, the Lin-Mcleod Test is further proposed to obtain a numeric solution to this problems based on Monte Carlo Simulation.

In this thesis, we first identify the problem of nuisance parameters for Autoregressive Fractionally Integrated Moving Average Model (ARFIMA model) in the Lin-McLeod Test; the size would be distorted, leading to an inaccurate level of type I error. We solve the problem through a modification of Lin-McLeod Test: Wild Monte Carlo Test, borrowing the idea of Wild Dependent Bootstrapping. In order to validate the algorithm, we derive the asymptotic distribution for the bootstrapped statistics in ARFIMA cases. By perturbing the estimated residuals, the Wild Monte Carlo Test outperforms a wide spread of Portmanteau test for this type of model. It solves the problem of the size underestimation and improves the test power for ARFIMA cases.

Later, we consider the general variance Portmanteau test on Autoregressive Moving Average Model with Generalized Autoregressive Conditional Heteroskedasticity Error (ARMA − GARCH Model), as a special case of weak ARMA model . When we have the null hypothesis of an ARMA − GARCH process, the asymptotic distribution of general variance are derived. With the complicated structure of the asymptotic distribution, the Lin-McLeod Test can serve a better solution to obtain the p-value rather than the gamma approximation. However, the test will still suffer from the size distortion due to the nuisance parameter issues. In this chapter, we mainly derive the asymptotic distribution for general variance Portmanteau tests on the ARMA − GARCH models and propose to use the Wild Monte Carlo Test to reduce the effect of nuisance parameters. The simulation and practical examples show the power of the newly test compared to the results of Francq et al. (2005).

Additionally, we shall consider the data generating process as the ARMA with infinite variance errors. In order to valid the general variance Portmanteau test and Fisher-Gallagher Test under this setup, we use the idea of the autocorrelation of the trimmed time series to construct the modified Portmanteau test and derive the asymptotic distribution for these two kind of Portmanteau tests on trimmed time series. Still, when we use the Lin-McLeod Test to obtain the p-value, Wild Monte Carlo Test can correct the nuisance parameter distortion for the size and improve the model performance.

Finally, we revisit the cross correlation test for two independent time series. A mistake in Hong (1996) simulation is pointed out and the corrected size for the test is provided later. Then we identify a spurious correlation for the time series with GARCH type errors.

Keywords: Time series analysis, Portmanteau test, Monte Carlo Test, ***R***, Weak ARMA Model, Infinite Variance Model, V-Statistics, Cross Correlation Test.

Acknowledgements

This thesis would never be completed without the help from lots of people. With these five year living in this warm family at Department of Statistical and Actuarial Sciences at Western University, I gained a big family where I can get energy to conquer all the future obstacles.

Firstly, I intend to express my sincerely gratitude to my doctoral supervisor, Dr. A. Ian McLeod. His persistence and support triggers me to continue my academic career and make me understand the statistics and related knowledge further and deeper. I am deeply touched by his devotion to academic and teaching. Without his encourage could I never accomplish the completion of my Ph.D journey. Special thanks are also devoted to my examiners, Dr. Yu Hao, Dr. Serge Provost, Dr. Allan MacIsaac and Dr. Zhou Zhou for their insightful comments and suggestions.

I am also indebted to my instructors at Western University, including but not limited to Dr. Wenqing He, Dr. Reg Kulperger, Dr. Rogemar Mamon, Dr. Xiaoming Liu. And I also have the opportunity to work with, Dr. Duncan Murdoch, Dr. Bethany White and Dr. John Braun as their teaching assistant. I would express my thanks to the staff member, Ms. Jennifer Dungavell and Ms. Jane Bai.

Secondly, I would like to express my deep love for my father, Shuzhi Xiao, and mother, Zhumei Xiong and sister Lu Xiao. Their encouragement from thousand miles away served as my internal dynamics to help me go through these valuable five years.

Last but not least, I would like to say thanks to all my friends here at Western, including but not limited to, Shan Shen, Chen Yang, Tianpei Jiang, Kexin Luo, Nikita Zhang, Xin Wang, and others. You really make London become my second hometown.

Contents

Certificate of Examination **ii**

Abstract **iii**

Contents **vii**

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**

2 Improved Portmanteau Test **4**
2.1 Introduction 4
2.2 ARFIMA Model and Diagnostic Check 7
2.3 Lin-McLeod Test based on Portmanteau Test 8
2.3.1 Test Size Comparison 9
2.4 Wild Monte-Carlo Test 10
2.4.1 Re-sampling Bootstrap for Time Series 10
2.4.2 Algorithm for the Wild Monte-Carlo Test 12
2.5 Wild Monte Carlo Test Simulation Experiments 15
2.5.1 Size Comparisons 15
2.5.2 Power Comparisons 16
2.6 Application to Beveridge Wheat Prices 20
2.7 Conclusion 22

3 Improved Portmanteau Test for ARMA − GARCH **Models** **25**
3.1 Introduction 25
3.2 Model Specification 27
3.3 Extension to Univariate General Variance Portmanteau Test 29
3.3.1 Asymptotic Distribution for Partial Autocorrelation Function 30
3.3.2 Asymptotic Distribution for General Variance Portmanteau Statistics $\hat{D}_m$ 32
3.4 Lin-McLeod Test and its Extension 33
3.4.1 Lin-McLeod Test and its Limitations 33
3.4.2 Wild Monte Carlo Test and its Asymptotical Properties 35
3.5 Simulation Results 37
3.5.1 Size Comparison 37

3.5.2 Power Comparison 39
3.6 Illustrative Application 42
3.7 Conclusion 45

4 Trimmed Portmanteau Test for ARMA Model with Infinite Variance 46
4.1 Introduction 46
4.2 Improved Portmanteau Test for ARMA Model with Infinite Variance 49
4.3 Extension to the Diagnostic Check for Model Adequacy of ARMA Model with Infinite Variance 52
4.4 Simulation Study on ARMA with Infinite Variance 54
4.4.1 Randomness Evaluation 56
4.4.2 Model Adequacy Checking 57
4.4.3 Further Power Comparison 59
4.5 Illustrative Application 62
4.6 Conclusion 66
4.7 Appendix 67
4.7.1 Lin-McLeod Test for Randomness on Trimmed Time Series 67
4.7.2 Lin-McLeod Test for Model Adequacy on Trimmed Time Series 67

5 Independent Test between Time Series 69
5.1 Introduction 69
5.2 Corrected Empirical Results and Spurious Correlation 71
Bibliography75

List of Figures

2.1 Beveridge(1925) Wheat Price Index Data Description with Autocorrelation Plot for Original Data and the Difference of the Original Data 23

2.2 Power Comparison for Beveridge Wheat Prices for ARFIMA $(2,1,0)$ model and ARFIMA $(2,0,0)$ model for Various Portmanteau Tests. 24

3.1 Daily Return Series for $S\&P$ 500 Index from January 3, 1979-December 31, 2001 and Autocorrelation, Partial Autocorrelation Function 42

4.1 Comparison between Time Series with Paretian errors and its trimmed time series. Choosing series length $n = 250$ and trimming level $\lambda = 5\%$. 55

4.2 Comparison between Estimated Residuals after fitting AR(1) Model to AR(1) Model with Paretian Errors. Original Residuals Series and its Trimmed Series Presented below. Choosing $\phi = \pm 0.1, \pm 0.5$ and ± 0.9 and Stable Parameter $\alpha = 1.5, \beta = 1, \gamma = 0$. Series length $n = 1000$ and trimming level $\lambda = 5\%$. . . . 58

4.3 Daily Index Returns of $S\&P500$ From 2 January 1999 to 29 December 2006, Autocorrelation Plot (ACF) and Partial Autocorrelation Plot (PACF). Length of Series $n = 2010$ 62

4.4 QQ Plot for Original Time Series and Trimmed Time Series with Trimmed Level $\lambda = 0.01$ 63

4.5 Month Simple Returns of CRSP Value-Weighted Index From January 1926 to December 1997, Autocorrelation Plot (ACF) and Partial Autocorrelation Plot (PACF). Length of Series $n = 864$ 65

List of Tables

2.1 Size Comparison for Monte-Carlo Test for ARFIMA $(0,0,1)$(%) with $\theta = \pm 0.3, \pm 0.6$ and ± 0.9, $d_f = 0.2, 0.4$ with the series length $n = 200, 500$ and lag $m = 10, 30$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3 9

2.2 Size Comparison for Lin-McLeod Test for ARFIMA $(1,0,0)$(%) with $\phi = \pm 0.3, \pm 0.6$ and ± 0.9 , $d_f = 0.2, 0.4$ with the series length $n = 200, 500$ and lag $m = 10, 30$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3 10

2.3 Size Comparison for Wild Monte-Carlo Reject Rate for ARFIMA $(0,0,1)$(%)with $\theta = \pm 0.3, \pm 0.6, \pm 0.9, d_f = 0.2, 0.4$ with the series length $n = 200, 500$ and lag $m = 10, 30$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3. 16

2.4 Size Comparison for Wild Monte Carlo Reject Rate for ARFIMA $(1,0,0)$(%)with $\theta = \pm 0.3, \pm 0.6, \pm 0.9, d_f = 0.2, 0.4$ with the series length $n = 200, 500$ and lag $m = 10, 30$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3. 17

2.5 Empirical Size Significance Levels Comparison between the Lin-McLeod Test $(\hat{D}_m)$ and Wild Monte Carlo Test $(\tilde{D}_m)$ under ARMA $(1,0)$ Model with $\phi = 0.1, 0.3, 0.5, 0.7, 0.9$, series length $n = 10^3$ and lag $m = 10, 20$. Choosing nominal level $\alpha = 1\%, 5\%$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3 17

2.6 Power Comparison for ARMA Model (Peňa and Rodríguez, 2002) between Lin-McLeod Test $(\hat{D}_m)$ and Wild Monte Carlo Test $(\tilde{D}_m)$ with series length $n = 10^3$ and lag $m = 1, 2, 3, 4, 5$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3 18

2.7 Power Comparison for ARMA Model (Peňa and Rodríguez, 2002) between Lin-McLeod Test $(\hat{D}_m)$ and Wild Monte-Carlo Test $(\tilde{D}_m)$ with series length $n = 10^3$ and lag $m = 10, 15, 20, 25$ and 30. Bootstrapping replication number is 10^2 and each time simulations number is 10^3 19

2.8 Power Comparison between MC and Wild MC test with $m = 20$ and $b = 5, 10, 20$. Bootstrapping iteration $N = 10^3$ 20

2.9 Power Comparison between MC Test $(\hat{D}_m)$ and Wild MC Test $(\tilde{D}_m)$ in Long Memory Structure (Lin and McLeod, 2006). Choosing $d_f = \pm 0.1, \pm 0.2, \pm 0.3$, series length $n = 128, 256, 512$ and lag $m = 1, 2, 3, 4, 5$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3 21

2.10 Power Comparison between MC Test ($\hat{D}_m$) and Wild MC Test ($\tilde{D}_m$) in Long Memory Structure (Lin and McLeod, 2006). Choosing $d_f = \pm 0.1, \pm 0.2, \pm 0.3$, series length $n = 128, 256, 512$ and lag $m = 10, 15, 20, 25, 30, 40$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3 22

2.11 Power Comparison Wild MC in Long Memory Structure with b Change with $m = 40$ and $b = 10, 20, 40$. Test for $d = \pm 0.3, \pm 0.1$ and $n = 128, 256, 512$ . . . 23

2.12 p-value for Fitting ARFIMA Model on Beveridge Data between Box-Pierce, Mahdi-McLeod, Fisher-Gallagher, Lin-McLeod and Wild MC Test. Choosing the lag $m = 5, 10, 20, 30, 40, 50$. 24

3.1 Size Comparison between ARMA $(1, 0)$ – GARCH $(1, 1)$ for Box-Pierce Q_{BP}, Modified Box-Pierce Q^*_{BP}, Monte Carlo $\hat{D}_m$and Wild Monte Carlo Test $\tilde{D}_m$. Choosing lag $m = 2, 4, 12, 24$ and 36 and Nominal Level $\alpha = 5\%$. 38

3.2 Power Comparison between Lin-McLeod Test $\hat{D}_m$, Wild MC Test $\tilde{D}_m$ in GARCH $(1, 1)$ Model. Choosing series length $n = 250, 500, 1000$, lag $m = 12, 24, 32$ and Nominal Level $\alpha = 5\%$. $\hat{D}_m$ stands for Monte Carlo Test and $\tilde{D}$ stands for Wild Monte Carlo Test. The p-value is based on 10^3 simulation and the power is evaluated at 10^3 iteration. 39

3.3 Power Comparison for Box-Pierce Test Q_{BP}, Modified Box-Pierce Test Q^*_{BP}, Lin-McLeod Test $\hat{D}_m$ and Wild Monte Carlo Test $\tilde{D}_m$. Lag $m = 2, 4, 12, 24, 36$ and Nominal Level $\alpha = 5\%$. 40

3.4 P Value Comparison on Evaluating GARCH type errors for return of $S\&P$ 500 Index for Randomness. Series Length $n = 5807$. Monte Carlo Iteration $N = 10^3$. MC represent the Monte Carlo Randomness test in Lin (2006) and WMC represent the newly developed Wild Monte Carlo Test for randomness test 43

3.5 p-value Comparison on Evaluating the GARCH type error for Return of $S\&P$ 500 Index. Series Length $n = 5807$. Monte Carlo Iteration $N = 10^3$. MC represent the tradition Monte Carlo Randomness test in Lin (2006) and WMC represent the newly developed Wild Monte Carlo Test for randomness test . . . 44

3.6 p-value Comparison on Evaluating the Residuals of GARCH $(1, 1)$ Model for Squares of $S\&P$ 500 Index Returns. Series Length $n = 5807$. Monte Carlo Iteration $N = 10^3$. MC represent the tradition Monte Carlo Randomness test in Lin (2006) and WMC represent the newly developed Wild Monte Carlo Test for randomness test 44

4.1 Size Comparison between Fisher-Gallagher Test (both Ljung-Box type and Monti Type) and its Application on the Trimmed Time Series. Choosing Stable Distribution Parameters $\alpha = 1.1, 1.3, 1.5, 1.7, 1.9, \beta = 1, \gamma = 0$. Trimmed Level $\lambda = 5\%$, series length $m = 5, 10, 15$. 56

4.2 Size Comparison between Monte Carlo Test $\hat{D}_m$, Box-Pierce Test Q_{BP}, Fisher Gallagher Ljung-Box Type Test Q_W and Fisher Gallagher Monti Type Test M_W. Choosing Stable Distribution Parameters as $\alpha = 1.1, 1.3, 1.5, 1.7, 1.9, \beta = 1, \gamma = 0$, Trimmed Level as $\lambda = 5\%$ and lag $m = 5, 10, 15$ 56

4.3 Size Comparison between Fisher-Gallagher Test (both Ljung-Box Q_W and Monti M_W) as well as Their Application on Trimmed Time Series, and General Variance Portmanteau Test $\hat{D}_m$ and its Application on Trimmed Time Series. Choosing $\phi = \pm 0.1, \pm 0.3, \pm 0.5, \pm 0.7, \pm 0.9$ and lag $m = 5/10/15$ and Stable Distribution Parameters $\alpha = 1.5, \beta = 1, \gamma = 0$. Series length $n = 10^3$ and Nominal Level 5% 58
4.4 Power Comparison between Portmanteau Test (Ljung-Box Fisher-Gallagher Test Q_W and its Application on Trimmed Time Series Q_W^*, Monti Fisher-Gallagher Test M_W and its Application on Trimmed Time Series M_W^*). Choosing Trimmed Level $\lambda = 0.01$, Series Length $n = 100$. 61
4.5 p-value Comparison on Evaluating Randomness for Squares of $S\&P500$ Index Returns and the trimmed time series with Ljung-Box Test. Series Length $n = 2011$. Monte Carlo iteration $N = 1000$. $*$ represent the trimmed level $\lambda = 0.01$ and $**$ represent the trimmed level $\lambda = 0.05$ 63
4.6 P Value Comparison on Evaluating Randomness for Squares of $S\&P$ 500 Index Returns and the trimmed time series with Fisher-Gallagher Test. Series Length $n = 2011$. Monte Carlo Iteration $N = 1000$. $*$ represent the trimmed level $\lambda = 0.01$ and $**$ represent the trimmed level $\lambda = 0.05$ 64
4.7 P Value Comparison on Goodness-of-fit for Monthly Simple Return of CRSP and the trimmed time series with Ljung-Box and Fisher-Gallagher Test against normal and stable distribution error assumption. Series Length $n = 864$. Monte Carlo Iteration $N = 1000$. $*$ represent the trimmed time series with trimmed level $\lambda = 0.01$ 64

5.1 Size Comparison between Hong Statistics (Hong, 1996) and Haugh Statistics (Haugh, 1976) under Corrected Formula. Choosing $X_t = 0.5X_{t-1} + u_t, Y_t = 0.5Y_{t-1} + v_t$ where $u_t, v_t \sim N(0,1)$ and $\rho_{uv} = 0$ for all j. Series Length $n = 1000$, Nominal Level $\alpha = 0.5\%$. Q_N and $Q_N^\star$ stand for Hong Statistics and its adjusted statistic. S_N represent the Haugh Test. $S_R^\star$ stands for the right adjusted Haugh statistic and $S_W^\star$ Stands for the wrong adjusted Haugh statistic. We underline the statistics closest to the nominal level α. 71
5.2 Power Comparison under Three Alternative Comparison of $\rho_{uv}(j)$. Choosing $X_t = 0.5X_{t-1} + u_t, Y_t = 0.5Y_{t-1} + v_t$ where $u_t, v_t \sim N(0,1)$. Series Length $n = 1000$, Nominal Level $\alpha = 5\%$ 72
5.3 Power Comparison between Hong's test and Haugh's test under the GARCH error. Series Length $N = 100, 200$. Choosing lag $M = \log n$ for Comparison. Q_N and $Q_N^\star$ stand for Hong Statistics and its adjusted statistic. S represent the Haugh Test. $S^\star$ stands for the right adjusted Haugh statistic. The table shows the power of identification under two models. 73

Chapter 1

Introduction

As an important part of Box-Jenkins modelling procedure, the Portmanteau Test is used to perform diagnostic checks on various models. Box and Pierce (1970) and Ljung and Box (1978) proposed these tests initially and derive the asymptotic distribution as chi-square distribution when applied on linear process. Later various modifications to the original tests are provided in order to improve the performance under specific models (Monti, 1994; Kheoh and McLeod, 1992). Then Peňa and Rodríguez (2002, 2006) give another type of portmanteau test based on the general variance of the autocorrelation matrix, hereafter the general variance test or the Peňa-Rodríguez Portmanteau Test. However, Lin and McLeod (2006) pointed out the defects in the test and proposed a Monte Carlo based Test to obtain the p-value to avoid the approximation errors with Gamma Approximation proposed by Peňa and Rodríguez (2002), hereafter, the Lin-McLeod Test, and achieve a better performance. Mahdi and McLeod (2011) extend the General Variance portmanteau test to multivariate Autoregressive Moving Average Model (ARMA Model) and propose an modification of the General Variance Portmanteau Test for univariate linear time series (hereafter the Mahdi-McLeod Test). Later, Fisher and Gallagher (2012a) provide another type of the portmanteau test based the autocorrelation matrix, hereafter the Fisher-Gallagher Test. Various comparisons are provided in Gallagher and Fisher (2013) and compare the powers of these tests. In general, the Mahdi-McLeod Test and the Fisher-Gallagher Test would yield best performance among all the tests. From these simulation, these tests could achieve satisfactory results under linear time series, like ARMA . However, When we consider the application of these tests on some specific models, like Autoregressive Fractionally Integrated Moving Average Model (ARFIMA Model) and Autoregressive Moving Average Model with Generalized Autoregressive Conditional Heteroskedasticity Error (ARMA − GARCH model), the size level is inaccurate and the powerfulness of the test is undermined, as we demonstrated in chapter 2, 3, 4.

Initially, Chapter 2 discovers the size distortion for ARFIMA model when we implement the Lin-McLeod Test (Lin and McLeod, 2006) on General Variance Portmanteau test. The underestimation of the size appears due to the nuisance parameters in the models, which is partially mentioned by Dufour (2006). Dufour (2006) proposed to used the Maximized Monte Carlo test (MMC Test) to mitigate this problem but the calculation is cumbersome. In order to fur-

ther diminish the size distortion issue more efficiently, we borrow the idea of Wild Dependent Bootstrap (Shao, 2010) and disturb the residuals by a kernel random variable. The asymptotic distribution of the bootstrapped statistics is derived with complicated structure, and we develop the Wild Monte Carlo Test to improve the performance of the Lin-McLeod test with ARFIMA model specification. Simulation comparisons are based on the various time series model such as ARMA, ARFIMA with strong white noise and show the mitigation of the size distortion issue. After resolving the size distortion, simulation experiments shows the power for the Wild Monte Carlo Test and it outperforms the Lin-McLeod Test. In the simulation study, we show that the Wild Monte Carlo Test could achieve improvements in the power test for these time series model as well as correcting the type I error level. A real data example is shown at the end of the chapter.

Secondly, in Chapter 3, we extend the univariate General Variance Portmanteau test (Mahdi and McLeod, 2011) to the ARMA model with GARCH type errors. The main idea is stimulated by Francq et al. (2005). They point out the convergence distribution when traditional portmanteau test applied in ARMA − GARCH model and the inconsistence of χ^2 approximation to the Box-Pierce Test as well as provided a modified Box-Pierce Test to perform the goodness-of-fit test in ARMA model with GARCH error, as one of the important type of weak ARMA model.

In chapter 3, we consider the General Variance Portmanteau Tests in time domain, applied on ARMA − GARCH. Initially we derive the asymptotic distribution of the General Variance Portmanteau Test on ARMA − GARCH model. Due to the complicated structure of the asymptotic distribution, we propose to use the Wild Monte Carlo Test to evaluate the p-value, addressing potential issues for the Lin-McLeod test. In order the validate the portmanteau tests, the asymptotic distributions of the bootstrapped statistics are derived. Further, we point out the similarity with the methods proposed by Harris and Kew (2014) involving the trimmed autocorrelation on the portmanteau test but with different implementation approaches. The simulations results show the advantages over the modified Box-Pierce Portmanteau test and other possible alternatives (Francq et al., 2005). A real example is presented to show the power of the test.

Thirdly, in Chapter 4, the study is based on Lin and McLeod (2008) on the linear time series model with infinite variance error. In the paper, they modify the portmanteau test on this type of model and implement the Lin-McLeod test to evaluate the p-value, compared to the other existing tests. The asymptotic distribution for the test is presented here and compared with other alternatives.

In order to validate the portmanteau test on linear time series model with infinite variance errors, we consider the use of the trimmed autocorrelation on infinite variance time series and apply the Lin-McLeod test for the power evaluation. The idea of trimmed autocorrelation is raised by Lee and Ng (2010) and they derive the asymptotic distribution for trimmed autocorrelation with application on the Ljung-Box Portmanteau Test. We extend this idea to the General Variance Portmanteau test and the Fisher-Gallagher test, resulting the asymptotic distributions

of these tests. Still, to obtain the p-value and avoid the nuisance parameter issue from the Lin-McLeod test, the newly developed Wild Monte Carlo test is employed as mentioned in the previous chapters. The results are compared to Cui and Wu (2014) which consider the distribution of the Monti portmanteau test for ARMA model with infinite variance. Simulation results and practical examples are used to prove the validity of the test.

In the end, we revisit carefully Hong (1996) and point out a mistake in the simulation study. After correcting the mistake here, we show a more accurate result to compare Hong Test and Haugh Test for independence of two time series. Later, a special type of spurious correlation is pointed out.

This thesis reviews the general practices for the portmanteau test in Box-Jenkins Modeling process and considers several special models with their portmanteau tests. With some defects for tradition portmanteau test as well as those later modifications, further modifications are proposed for specific models to improve the test performances and accuracy under these model specfication. In addition, Wild Monte Carlo Test is proposed during the process as an modification to the Lin-McLeod Test in order to solve the issues caused by the nuisance parameter and gains improvement for these applications.

Chapter 2

Improved Portmanteau Test

2.1 Introduction

Goodness-of-fit tests are treated as one of the important parts in model diagnostics steps for Box-Jenkins modelling procedure. Usually, one would use portmanteau tests to test for the randomness for the model residuals in order to check the adequacy of the fitted model.

Among all the portmanteau tests, the test was first introduced by Box and Pierce (1970),

$$Q_{BP} = n \sum_{k=1}^{m} \hat{r}_k^2, \tag{2.1}$$

where $\hat{r}_k$ is defined as the autocorrelation for the residuals of lag k, and

$$\hat{r}_k = \frac{\sum_{t=k+1}^{n} \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^{n} \hat{a}_t^2}. \tag{2.2}$$

The asymptotic distribution of the Box-Pierce test is derived as $\chi^2_{m-(p+q)}$, where p and q are the parameters in ARMA (p, q) model. Considering an adequate fitted model, if the series length n is large enough, we could state $\hat{r}_1 \simeq \hat{r}_2 \simeq ... \simeq \hat{r}_m \simeq 0$ indicating the lack of the autocorrelation in the residuals.

Ljung and Box (1978) proposed a modified version of the Portmanteau test,

$$Q_{LB} = n(n+2) \sum_{k=1}^{m} \frac{\hat{r}_k^2}{n-k}. \tag{2.3}$$

The statistics Q_{LB} has a finite sample distribution closer to that of $\chi^2_{m-(q+p)}$. However, according to Davis et al. (1977), the variance of Q_{LB} will be significantly larger than its asymptotic distribution.

Kwan et al. (2005) has pointed out that various Portmanteau tests such as (Box and Pierce, 1970; Ljung and Box, 1978; Peña and Rodríguez, 2002) would have a distorted empirical size

when the residuals follow a non-gaussian distribution. Therefore, they suggest to use the non-parametric test on the rank of the autocorrelation for the estimated residuals (Dufour and Roy, 1986) to fix the problem. The Portmanteau test is based on asymptotic distribution provided that the innovations in the time series are statistically independent and identically distributed. Today it is realized that in some areas of application this assumption is violated by conditional heteroscedasticity. In this chapter we will assume strong white noise, that is, we assume that the innovations are statistically independently distributed with mean zero and constant variance. An improved Portmanteau test for dealing with some departures from the strong white noise assumption, including conditional heteroscedasticity, will be discussed in Chapter 3.

Kheoh and McLeod (1992) showed that Q_{LM} perform slightly better than Q_{LB} in many cases. To deal with the incorrect size problem, Li and McLeod (1981) proposed another modification of the Q_{LB},

$$Q_{LM} = Q_{LB} + \frac{m(m+1)}{2n}. \tag{2.4}$$

Monti (1994) proposed another Portmanteau test based on the partial autocorrelation for the residuals. Let $\hat{\pi}_k$ be the kth partial autocorrelation which is approximately distributed as a normal random variable with mean zero and variance $(n-k)/(n(n+2))$. The test statistics can be defined as follows,

$$Q_{MT} = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{\pi}_k. \tag{2.5}$$

The asymptotic distribution of Q_{MT} is χ^2 with $m-(p+q)$ degrees of freedom. A detailed introduction is presented in Li (2004).

These proposed Portmanteau tests are based on the sum of the weighted autocorrelations or partial autocorrelation functions. Later, in Peña and Rodríguez (2002), they raise the idea of another new Portmanteau test as defined,

$$\hat{D}_m = n\left[1 - |\hat{\mathbf{R}}_m|^{1/m}\right], \tag{2.6}$$

where we define the $\hat{\mathbf{R}}_\mathbf{m}$ as the residual correlation matrix

$$\hat{\mathbf{R}}_\mathbf{m} = \begin{bmatrix} 1 & \hat{r}_1 & \dots & \hat{r}_m \\ \hat{r}_1 & 1 & \dots & \hat{r}_{m-1} \\ \vdots & \vdots & \ddots & \dots \\ \hat{r}_m & \hat{r}_{m-1} & \dots & 1 \end{bmatrix}.$$

The asymptotic distribution of the $\hat{D}_m$ is derived in Peña and Rodríguez (2002).

Theorem 2.1. *If the model is correctly defined,$\hat{D}_m$ is asymptotically distributed as $\sum_{i=1}^{m}\lambda_i\chi^2_{1,i}$, where $\chi^2_{1,i}$ $(i = 1,...m)$ are independent χ^2_1 random variables and λ_i $(i = 1,2,...m)$ are the eigenvalues of $(\mathbf{I}-\mathbf{Q_m})\mathbf{W_m}$, where $\mathbf{W_m}$ is a diagonal matrix with elements $\omega_i = (m-i+1)/m(i = 1,2,...m)$ and $\mathbf{Q_m} = \mathbf{I_m}\mathbf{V}^{-1}\mathbf{I_m^T}$,$\mathbf{V}$ is the information matrix for the parameters ϕ and θ and $\mathbf{I_m}$ is diagnostic identify matrix.*

Since this Portmanteau test is mainly based on the general variance of the autocorrelation matrix, we consider these type of tests as the Generalized Variance Portmanteau test. Peña and Rodríguez (2002) replace $\hat{r}_k^2$ with $\tilde{r}_k^2 = (n+2)(n-k)^{-1}\hat{r}_k^2$ in autocorrelation matrix $\hat{\mathbf{R}}_m$, defined as $\tilde{\mathbf{R}}_m$ in order to improve the performance of the gamma approximation.

Later Peña and Rodríguez (2006) proposed a modified portmanteau,

$$\hat{D}_m^* = -\frac{n}{m+1}\log|\hat{\mathbf{R}}_\mathbf{m}|, \tag{2.7}$$

Under the corrected identification of the model, the $\hat{D}_m^*$ is asymptotically distributed as the same distribution as the previous $\hat{D}_m$. However several limitations for the test $\hat{D}_m^*$ have been pointed out in Lin and McLeod (2006). Firstly, the approximation to $\hat{D}_m$ proposed in Peña and Rodríguez (2002) would not agree with the original distribution if m is small based on the gamma approximation. Secondly, the modified $\tilde{\mathbf{R}}_\mathbf{m}$ is not always positive definitive.

Later a novel portmanteau statistics for both vector and univariate ARMA model is proposed by Mahdi and McLeod (2011). The new Portmanteau test is an extension of the General Variance Portmanteau test. For univariate time series, the Portmanteau test on standardized residuals could be written as,

$$\tilde{D}_m = -\frac{3n}{2m+1}\log\left|\hat{\mathbf{R}}_\mathbf{m}\right|. \tag{2.8}$$

Its asymptotic distribution and more accurate approximation is derived in Mahdi and McLeod (2011). And the performance is evaluated via simulation which shows its preferable properties in testing. Still, the Lin-McLeod test is recommended unless the series is very long for estimation.

Another types of important Portmanteau tests are proposed by Fisher and Gallagher (2012a). The two test statistics are defined as follows,

$$\tilde{Q}_W = n(n+2)\sum_{k=1}^{m}\frac{m-k+1}{m}\frac{\hat{r}_k^2}{n-k}. \tag{2.9}$$

$$\tilde{M}_W = n(n+2)\sum_{k=1}^{m}\frac{m-k+1}{m}\frac{\hat{\pi}_k^2}{n-k}. \tag{2.10}$$

The above two statistics could be treated as the weighted Ljung-Box test. The asymptotic distribution for this statistic is identical with the distribution of Peña and Rodríguez (2002). Another similar statistic could be constructed based on the weighted Monti Test,

According to Gallagher and Fisher (2013), all the above statistics could be expressed as weighted Ljung-Box or Monti Test.

$$Q_W = n\sum_{k=1}^{m} w_k\hat{r}_k^2. \tag{2.11}$$

The above $\hat{r}_k$ could be replaced by $\hat{\pi}_k$. The Portmanteau test in Mahdi and McLeod (2011) is essentially a weighted Monti Test.

2.2 ARFIMA Model and Diagnostic Checks

Let $X_t, t = 1, 2, \ldots$ be a stationary and invertible Autoregressive Moving Average Model ARMA (p, q) model in Box et al. (2008),

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) X_t = (1 - \theta_1 B - \ldots - \theta_q B^q) a_t, \tag{2.12}$$

where B is the backshift operator on t and a_t follow an identical and independent normal distribution with mean of zero and constant variance, σ_a^2. For convenience, we write the ARMA model,

$$\phi(B) X_t = \theta(B) a_t, \tag{2.13}$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B = \theta_2 B^2 - \ldots - \theta_q B^q$. Since the stationarity and casuality of the model, all the roots of $\phi(B) = 0$ and $\theta(B) = 0$ is outside of the unit root circle. [1]

We could employ maximum likelihood or least square method to estimate the parameters $\beta = (\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q)$ as $\hat{\beta} = (\hat{\phi}_1, \hat{\phi}_2, \ldots \hat{\phi}_p, \hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_q)$. Then we can have the residuals estimator $\hat{a}_t$,

$$\hat{a}_t = \frac{(1 - \hat{\phi}_1 B - \ldots - \hat{\phi}_p B^p)}{(1 - \hat{\theta}_1 B - \ldots - \hat{\theta}_q B^q)} X_t. \tag{2.14}$$

The most widely used long memory model is the Autoregressive Fractionally Integrated Moving Average model (ARFIMA),

$$\phi(B)(1 - B)^{d_f} X_t = \theta(B) a_t, \tag{2.15}$$

where $\phi(B)$, $\theta(B)$ and a_t are defined the same as previously defined for the ARMA and d_f is defined in the range of $-1/2 < d_f < 1/2$. The parameter d_f determines the long memory property as defined in Granger and Joyeux (1980) and Hosking (1981). For the fractional ARMA model, maximum likelihood could still be used to estimate the parameter $\beta = (\phi_1, \phi_2, \ldots, \phi_p, d, \theta_1, \theta_2, \ldots, \theta_q)$. However, the directly Maximized Likelihood Estimation (MLE) is slow in computation due to the complicated structure in the autocovariance. Therefore, we could use approximate maximum likelihood to estimate the parameters in the model. The properties of the parameter estimators are discuss in Li and McLeod (1986) and Beran (1994). The parameters could be estimated via the R package **FGN** (McLeod and Veenstra, 2007) and **arfima** (Veenstra and McLeod, 2013). A detailed introduction to the model estimation and these packages are presented given by (Veenstra, 2013).

In order to consider the Portmanteau test in ARFIMA , first we consider the distribution of the autocorrelations of the residuals. The distribution is derived in Li and McLeod (1986).

Theorem 2.2. *For any fixed $m \geq 1$, $\sqrt{n}\,\{\hat{r}_1, \hat{r}_2, \ldots \hat{r}_m\}$ is asymptotically normal with mean zero and covariance matrix $\mathbf{I_m} - \mathbf{X}\mathbf{I}^{-1}\mathbf{X}^{\mathbf{T}}$,where $\mathbf{I_m}$ is the $m \times m$ identity matrix and*

$$\mathbf{X} = (-\phi'_{\mathbf{i-j}} | \theta'_{\mathbf{i-j}} | \mathbf{K})_{\mathbf{m}},$$

[1] As in Box et al. (2008), the symbol B is over-loaded in the sense that it may represent either the backshift operator on t or a complex variable depending on how it is used.

where $K' = \left\{1, \frac{1}{2}, \ldots, \frac{1}{m-1}\right\}$*, and the parameters* θ'_{i-j} *and* ϕ'_{i-j} *are defined as in the* ARMA *model.*

According to this theorem, we can see the inner structure of the autocovariance matrix for the estimate residual autocorrelations in the ARFIMA model is more complicated than that in the ARMA model.

However, the same undersize problem for the p-value happens when we apply the new Portmanteau test (Peña and Rodríguez, 2002; Mahdi and McLeod, 2011) to the ARFIMA model.

This section addresses this problem. In order to fix the test inaccuracy of Lin-McLeod Test on ARFIMA Model, a more advanced Monte Carlo test using the idea of wild dependent bootstrap (Shao, 2010) is developed. Simulations and real examples demonstrate the effectiveness of the new test.

2.3 Lin-McLeod Test based on Portmanteau test

We revise the Lin-McLeod Test proposed by Lin and McLeod (2006) and use it on the ARFIMA model to check the goodness-of-fit in the residuals. The algorithm is restated below. In the statement of this algorithm, the Portmanteau test statistic, General Variance Portmanteau test D_m, is used but any of the other Portmanteau test statistics could be substituted.

Algorithm 2.1. *For an observed series,*

- **Step One:** *Fit the predetermined model with appropriate number of* p, q*. Obtain the estimated parameter space* $\hat{\Lambda} = (\hat{\phi},\ \hat{d}_f,\ \hat{\theta})$ *as well as the residuals* $\hat{a}_t,\ t = 1, 2, \ldots, n$*. Calculate the* $\hat{D}_m$ *based on* $\hat{a}_t$*.*

- **Step Two:** *Re-simulate the* $X_t^*, t = 1, 2, \ldots, n$ *based on the estimated parameter space* $\hat{\Lambda} = (\hat{\phi}, \hat{d}_f, \hat{\theta})$ *as the parameters in equation 2.12 and random variable* $\tilde{a}_t$ *following normal distribution with the mean and variance as* $\hat{a}_t$*. Refit the simulated* $X_t^i, t = 1, 2, \ldots, n$ *based on the number of parameters and obtain the residuals* $\hat{a}_t^i$*. Calculate the statistics* $\hat{D}_m^i$*.*

- **Step Three:** *Repeat Step Two N times. The p-value could be calculated as followed:*

$$p\text{-}value = \frac{\#\left\{\hat{D}_m^i > \hat{D}_m\right\} + 1}{N + 1}. \tag{2.16}$$

For simulation study, the above algorithm could be applied to a predetermined time series with known parameters. According to Lin and McLeod (2006), the Lin-McLeod Test can outperform the original Peña-Rodríguez test both in power and providing a correct type I error rate. In order to fully explore effectiveness of the Lin-McLeod Test, we apply the algorithm on the ARFIMA model. However, in this case, severe size distortions problems happened resulting in an invalid diagnostic test.

We will fit the algorithm for the following model to see the performance. For simplicity, we will test for the simple model: ARFIMA $(0,0,1)$ and ARFIMA $(1,0,0)$ under corrective specifying model and incorrectly specifying model. All of the following simulations are based on 10^2 bootstrap iteration of 10^3 simulations for each parameter setting. In order to improve the calculation efficiency, we employ the High Performance Computing Center (`https://www.sharcnet.ca/`), to run the simulation as well as parallel computing technique. The parallel package **Rmpi** (Yu, 2014) in R. The simulations was conducted using 30 core cluster. The CPU Running time for the calculation is around 70 Days.

2.3.1 Test Size Comparison

In the first model, $X_t = (1-\theta_1 B)\nabla^{d_f} a_t$, we test for ARFIMA $(0,0,1)$. Table 2.1 exhibits the test output where the true value of parameter is $\theta_1 = \pm 0.3, \pm 0.6, \pm 0.9$, $d_f = 0.2, 0.4$.

Table 2.1: Size Comparison for Monte-Carlo Test for ARFIMA $(0,0,1)$(%) with $\theta = \pm 0.3, \pm 0.6$ and ± 0.9, $d_f = 0.2, 0.4$ with the series length $n = 200, 500$ and lag $m = 10, 30$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3

		m = 10		**m = 30**	
	θ/d_f	**0.2**	**0.4**	**0.2**	**0.4**
n = 200	−0.9	4.7	4.8	4.8	6.5
	−0.6	4.2	6.0	4.1	4.8
	−0.3	5.6	5.8	3.4	5.0
	0.3	6.9	6.1	4.6	6.0
	0.6	4.8	5.9	5.5	6.7
	0.9	4.5	4.9	4.9	5.1
n = 500	−0.9	3.9	3.9	4.7	5.8
	−0.6	4.9	5.7	5.5	5.9
	−0.3	5.6	4.5	6.2	5.8
	0.3	5.2	3.8	3.6	5.0
	0.6	5.1	5.3	5.4	4.7
	0.9	5.2	4.8	5.3	4.9

From Table 2.1, showing the results for ARFIMA $(0,0,1)$, we see that the type I errors maintain at a reasonable level close to the nominal level of 5%.

In Table 2.2 the simulation results are shown for the second model, $(1-\phi_1 B)\nabla^{d_f} Z_t = a_t$, where $\phi_1 = \pm 0.3, \pm 0.6, \pm 0.9$, $d_f = 0.2, 0.4$

The Lin-McLeod Test (Lin and McLeod, 2006) shows its inefficiency especially when the predetermined model has autoregressive components and the parameters for it are near the boundary of stationarity. The type I error is drastically underestimated. Increasing the number of sample size does not help improve the performance of the tests. As previously pointed

Table 2.2: Size Comparison for Lin-McLeod Test for ARFIMA$(1,0,0)$(%) with $\phi = \pm 0.3, \pm 0.6$ and ± 0.9 , $d_f = 0.2, 0.4$ with the series length $n = 200, 500$ and lag $m = 10, 30$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3

		m = 10		**m = 30**	
	ϕ/d_f	**0.2**	**0.4**	**0.2**	**0.4**
n = 200	−0.9	*	*	*	*
	−0.6	*	*	*	*
	−0.3	*	*	*	*
	0.3	4.6	3.0	3.2	3.2
	0.6	3.1	2.9	3.2	3.3
	0.9	*	3.8	2.6	*
n = 500	−0.9	*	*	*	*
	−0.6	*	*	*	*
	−0.3	*	*	*	*
	0.3	3.0	4.2	4.4	4.9
	0.6	4.5	4.7	3.3	3.5
	0.9	3.6	2.2	4.4	3.0

out by Dufour (2006), the results of the Lin-McLeod Test may be severely distorted by nuisance parameters which the asymptotic distribution depends on especially for type I error and Table 2.2 confirms that this problem may exist when the usual Lin-McLeod test is applied to the ARFIMA . Based on this distortion, Dufour (2006) proposed the Maximized Monte Carlo test (MMC) to remove the effect of the nuisance parameters. But the MMC is quite cumbersome, difficult to implement in practice and no examples to demonstrate its effectiveness were reported either in Dufour (2006) or in any other literature of which we are aware. MMC is similar to the double bootstrap method (McCullough and Vinod, 1998) but the Wild Monte Carlo test is computationally more efficient and easier to implement.

2.4 Wild Monte-Carlo Test

In order to accurately test for the goodness-of-fit under the ARFIMA structure, we develop a new Wild Monte Carlo test (WMC) for the Portmanteau test, which perturbs the residuals by the random variable defined in Shao (2010).

2.4.1 Re-sampling Bootstrap for Time Series

As an important extension to bootstrap by Efron (1979), the block-based bootstrap methods can be used to approximate the sampling distribution of complicated data structure especially in time series. The Moving Block Bootstrap (MBB) was first introduced by Kunsch (1989) and Liu and Singh (1992). In order to improve the performance for bootstrap on temporal

dependence series, based on the idea of re-sampling block bootstrap, a large number of other similar bootstraps were developed by various researchers including the circular block bootstrap (CBB) (Politis and Romano, 1992), the stationary bootstrap (SB) (Politis and Romano (1994) and Tapered Block Bootstrap (TBB) (Paparoditis and Politis, 2001a) and Paparoditis and Politis (2001b). In the book of Lahiri (2003), detailed comparison between these methods are presented as well as detailed derivations.

Another type of bootstrap is called Dependent Wild Bootstrap proposed by Shao (2010) which is considered as an extension of traditional Wild Bootstrap (Wu, 1986) to time series. The traditional Wild Bootstrap would allow the auxiliary variables to mimic the weak dependence in the series, which will improve the test power. Several dependent random variables are recommended to capture the dependence in the series. The newly proposed Dependent Wild Bootstrap has extended the Wild Bootstrap to time series and allowed more complicated dependence structure. Dependent random variables with kernel are used to make this technique applicable to weakly dependent series. Another advantage for Dependent Wild Bootstraps compared to the block-based bootstrap methods is that it can avoid the choice of block size. The window length, similar to the block-size, is used to control the number of autocorrelations entering into the statistics, and should be well-specified and the choice of windows length will affect the performance of the Portmanteau test. The dependence is conveyed by the kernel smoothing function.

First we should define the dependent wild bootstrap pseudo-observation as

$$X_i^* = \bar{X}_n + (X_i - \bar{X}_n)W_i, \tag{2.17}$$

where $i = 1, 2, \ldots, n$, $\bar{X}_n$ is the sample mean of the original data. Without loss of generality, we can consider the data with mean of zero, then the pseudo-observations are changed into the simplified version,

$$X_i^* = X_i \times W_i, i = 1, 2, \ldots, n, \tag{2.18}$$

where the W_i is the random variable to consider the dependence in the series which is defined as followed assumption from Shao (2010),

Assumption 2.1. *The random variables $\{W_i\}_{t=1}^n$ are independent of the series $X_t, t = 1, 2, \ldots, n$. $E(W_t) = 0, Var(W_t) = 1$ for $t = 1, 2, \ldots, n$. Assume that W_t is a stationary process with $Cov(W_t, W_{t'}) = a\{(t - t')/l\}$, where $a(\cdot)$ is a kernel function and $l = l_n$ is a bandwidth parameter. Furthermore, assume that*

$$K_a(x) = \int_{-\infty}^{+\infty} a(u)e^{-iux}du \geq 0, \tag{2.19}$$

for $x \in \mathbb{R}$.

According to Theorem 3.1 in Shao (2010), we can see the statistics from the pseudo-observations X_t^* will convergence in law to the true value of the statistics. Proof of the convergence will be discussed in the next section.

With this assumption, we can see the lag window l will serve as the same function as the the block size. The condition for the kernel function ensures the nonnegative definitiveness of the

covariance matrix of W_t. According to Andrews (1991), we know several kernel functions such as Bartlett, Parzen and Quadratic Spectral Kernels, can be used here. These independent variable with these kernel functions can use reject methods or the methods proposed in Mammen (1993) to simulate the kernel for the bootstrap.

2.4.2 Algorithm for the Wild Monte-Carlo Test

We modify the Lin-McLeod Test in order to incorporate the Dependent Wild Bootstrapping into the test to convey the dependence in the residuals of ARFIMA Model.

Algorithm 2.2. *The algorithm for the Wild Monte-Carlo Test is stated as follows. For a given time series data,*

- **Step One:** *Fit the series X_t with appropriate p, d_f, q and obtain the estimates $\hat{\phi}_1, \ldots, \hat{\phi}_p$, $\hat{d}_f, \hat{\theta}_1, \ldots, \hat{\theta}_q$ as well as the residuals $\hat{a}_t$. Simulate dependent random variable $W_t, t = 1, 2, \ldots, n$ based on the kernel. Generate the pseudo-observations,*

$$\hat{a}_t^* = \hat{a}_t \times W_t, \tag{2.20}$$

  and then calculate the D_M^* statistics based on $\hat{a}_t^*$.

- **Step Two:** *Re-simulate the model based on the estimated $\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{d}_f, \hat{\theta}_1, \ldots, \hat{\theta}_q$ as well as simulated random variable $\tilde{a}_t$ following normal distribution with same mean as $\hat{a}_t$. Then we refit the model gaining the residuals $\hat{a}_t^j$. Generate the dependent random variable W_t^* based on the same kernel in* **Step One**. *Generate the pseudo-observations $\hat{a}_t^{j*}$,*

$$\hat{a}_t^{j*} = \hat{a}_t^j \times W_t^*, \tag{2.21}$$

  and then calculate the statistics D_m^{j*}.

- **Step Three:** *Repeat the* **Step Two** *for N times and obtain the estimated p-value.*

$$p\text{-}value = \frac{\#\left\{D_m^{j*} > D_m^*\right\} + 1}{N + 1}. \tag{2.22}$$

The above algorithm is proposed based on the idea of Shao (2010). We perturb residuals for the observed and simulated series by the random variable as defined in assumption 3.1. Based on Shao (2010), several eligible kernel, such as Parzen, Bartlett and Quadratic Spectral Kernel, could be used here. For illustrative purpose, we restate the Parzen kernel here.

Definition 2.1. *Parzen Kernel*

$$K_P(x) = \begin{cases} 1 - 6(x/6l)^2 + 6\,|x/6l|^3 & 0 \le |x/l| \le 1/2 \\ 2(1 - |x/6l|^3) & 1/2 \le |x/l| \le 1 \\ 0 & \text{Otherwise} \end{cases}. \tag{2.23}$$

The Parzen kernel can be used to adjust the underestimating problems in the Portmanteau test. We need to establish the asymptotic distributions for the bootstrapped statistics to validate the Wild Monte Carlo Test. Here we use the parameter m as the bandwidth, and $x = |i - j|$ as the distance between i and j. As we could see from the Parzen Kernel would consider exact number of autocorrelation in the bandwidth m.

We incorporate the Wild Monte Carlo Test in the **portest** function in the packages **portes** (Mahdi and McLeod, 2013) by setting the argument **Kernel = TRUE**. Here we use the **DLSimulate** function in package **tsar** McLeod et al. (2007) to simulate a stationary random variable following a given autocovariance function.

Therefore, we can derive the asymptotic distribution for the bootstrapped statistics.

Theorem 2.3. *The bootstrapped determinant based Test D_m^* (Mahdi and McLeod, 2011), under the Parzen kernel function $K_n(|i - j|/m)$, would hold for the asymptotic distribution as follows,*

$$D_m^* \rightarrow_d \sum_{i=1}^{m} \lambda_i \chi_{1,i}^2, \tag{2.24}$$

where $\lambda_i, i = 1, 2, \ldots, m$ are the eigenvalues of the matrix $(\mathcal{I} - \mathcal{Q})\mathcal{W}^*$. $\mathcal{W}^* = \mathcal{K}^T \mathcal{W} \mathcal{K}$, where $\mathcal{W}$ and $\mathcal{K}$ are both diagonal matrices with the following definition. And $\mathcal{Q}$ is the projection matrix defined as $X^T(X^TX)^{-1}X$.

$$\mathcal{W} = \begin{pmatrix} \frac{3m}{2m+1} & 0 & \cdots & 0 \\ 0 & \frac{3(m-1)}{2m+1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{3}{2m+1} \end{pmatrix}, \tag{2.25}$$

and

$$\mathcal{K} = \left(K_n(|i - j|)_{|i-j|=1,2,\ldots,m} \right). \tag{2.26}$$

Proof:

According to Mahdi and McLeod (2011), the determinant-based Portmanteau test is defined as,

$$D_m = -\frac{3n}{2m+1} \log |\mathcal{R}_m| . \tag{2.27}$$

According to Ramsey et al. (1974),

$$|\mathcal{R}_m| = \prod_{i=1}^{m} (1 - \pi_i^2)^{m+1-i}.$$

Therefore, we could write the test statistics as:

$$D_m = -n \sum_{i=1}^{m} \frac{3(m+1-i)}{2m+1} \log(1 - \pi_i^2). \tag{2.28}$$

The wild dependent bootstrapped test statistics could be defined as,

$$D_m^* = -n \sum_{i=1}^{m} \frac{3(m+1-i)}{2m+1} \log(1-(\pi_i^*)^2), \tag{2.29}$$

where π_i^* is the Parzen Kernel adjusted partial autocorrelation with $\pi_i^* = K_n(i/m)\pi_i$. Following the steps of Peña and Rodríguez (2002), assuming $(n\pi_1^2, n\pi_2^2, \ldots, n\pi_m^2)$ is asymptotically distributed as Y, based on the matrix form, we can have $n(\pi_1^*)^2, n(\pi_2^*)^2, \ldots, n(\pi_m^*)^2$ is asymptotically distributed as $\mathcal{K}^T Y \mathcal{K}$,where $\mathcal{K}$ is the diagonal matrix with $K_n(i/m), i = 1, 2, \ldots, m$. Then, we could apply the multivariate δ-method in Arnold (1990) to equation 2.29 to give,

$$D_m^* \to \left(\frac{3m}{2m+1}, \frac{3(m-1)}{2m+1}, \ldots, \frac{3}{2m+1}\right) \mathcal{K}^T Y \mathcal{K}. \tag{2.30}$$

We know the fact that,

$$\left(\frac{3m}{2m+1}, \frac{3(m-1)}{2m+1}, \ldots, \frac{3}{2m+1}\right) \left(n(\pi_1^*)^2, n(\pi_2^*)^2, \ldots, n(\pi_m^*)^2\right)^T \to \tag{2.31}$$

$$\left(\frac{3m}{2m+1}, \frac{3(m-1)}{2m+1}, \ldots, \frac{3}{2m+1}\right) \mathcal{K}^T Y \mathcal{K}, \tag{2.32}$$

based on the fact that $n^{1/2}\pi_{(m)} \to N(0, \mathcal{I}_m - Q_m)$, where Q_m is the project matrix and the quadratic forms by Box et al. (1954):

$$\left(\frac{3m}{2m+1}, \frac{3(m-1)}{2m+1}, \ldots, \frac{3}{2m+1}\right) \left(n(\pi_1^*)^2, n(\pi_2^*)^2, \ldots, n(\pi_m^*)^2\right)^T \tag{2.33}$$

$$= n\pi_{(m)}^T \mathcal{K}^T \mathcal{W} \mathcal{K} \pi_{(m)} \to \sum_{i=1}^{m} \lambda_i \chi_{1,i}^2, \tag{2.34}$$

where $\lambda_i, i = 1, 2, \ldots, m$ are the eigenvalues of matrix $(\mathcal{I}_m - Q_m)\mathcal{W}^*$ and $\mathcal{W}^* = \mathcal{K}^T \mathcal{W} \mathcal{K}$.

From Theorem 4.1, after perturbing the residuals, the bootstrapped statistics would follow as the weighted χ^2 distributions. The perturbed residuals would have less sensitivity to the estimated parameters and further give a more stabilized type I error.

Ljung and Box (1978) originally proposed the Ljung-Box test to perform the goodness-of-fit test for time series models. Still based on Li (2004), some biasness are possibly incurred due to the weak dependence in the residuals. Therefore, we can still perturb the estimated residuals by the dependent random variable w_t via the Wild Monte Carlo test. The test statistics can be restated as,

$$Q_{LB} = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} \hat{r}_k^2, \tag{2.35}$$

where $\hat{r}_k, k = 1, 2, \ldots, m$ represent the autocorrelation of the residuals. The bootstrapped statistics Q_{LB}^* will converge to asymptotic distribution.

Corollary 2.1. *Under the Parzen kernel function K_n, the asymptotic distribution of the Wild Monte Carlo Test based on Ljung-Box Statistics could be specified as followed:*

$$Q^*_{LB} \rightarrow_d \sum_{i=1}^{m} \lambda_i \chi^2_{1,i}, \tag{2.36}$$

where $\lambda_i, i = 1, 2, \ldots, m$ are the eigenvalues of the matrix $(\mathcal{I} - \mathcal{Q})\mathcal{W}^*$. $\mathcal{W}^* = \mathcal{K}^T \mathcal{W} \mathcal{K}$ where $\mathcal{W}$ and $\mathcal{K}$ are both diagonal matrices with the following definition, and $\mathcal{Q}$ is the projection matrix defined as $X^T(X^TX)^{-1}X$,

$$\mathcal{W} = \begin{pmatrix} \frac{n-2}{n-1} & 0 & \cdots & 0 \\ 0 & \frac{n-2}{n-2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{n-2}{n-m} \end{pmatrix}, \tag{2.37}$$

and

$$\mathcal{K} = \left(K_n(|i-j|)_{|i-j|=1,2,\ldots,m} \right). \tag{2.38}$$

The proof of this corollary follows similar steps like theorem 4.1 by adding the weights for the autocorrelation function which leading to dissimilar asymptotic distribution. This algorithm could be extended to various Portmanteau test including the Box-Pierce test (Box and Pierce, 1970), the Fisher-Gallagher test (Fisher and Gallagher, 2012a). The asymptotic distribution could be obtained based on the derivations already presented but still we recommend the numeric approaches to obtain the p-values since we have demonstrated already that the asymptotic distribution often has incorrect size.

2.5 Wild Monte Carlo Test Simulation Experiments

We have already validated the Wild Monte Carlo test by providing the asymptotic distribution. Now, we redo the size comparison for ARFIMA model and provide some further power comparisons for numerous models are implemented here to show the validation of the Wild Monte Carlo test compared with the Portmanteau test as well as the Lin-McLeod Test.

2.5.1 Size Comparisons

The serious problem for the the Lin-McLeod Test (Lin and McLeod, 2006) is the underestimation of the type I error. Though the underestimation of the type I error would lead to a comparatively conservative model, somehow, this problem could lead to an inadequate model. Table 2.3 showing the size may be compared with Table 2.2 for ARFIMA $(0, 0, 1)$ model.

From the Table 2.3, we can see the Wild Monte Carlo adjusts the type I error for ARFIMA $(0, 0, 1)$ model especially when the parameters are near the stationarity boundary. Compared with Table 2.1, with various choice of parameters, the type one errors for Wild Monte Carlo Test are

Table 2.3: Size Comparison for Wild Monte-Carlo Reject Rate for ARFIMA $(0,0,1)$(%)with $\theta = \pm 0.3, \pm 0.6, \pm 0.9, d_f = 0.2, 0.4$ with the series length $n = 200, 500$ and lag $m = 10, 30$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3.

		m = 10		**m = 30**	
	θ/d_f	**0.2**	**0.4**	**0.2**	**0.4**
n = 200	−0.9	3.4	4.1	2.8	2.2
	−0.6	4.2	3.5	2.5	3.3
	−0.3	4.0	3.7	3.6	2.9
	0.3	4.7	4.4	4.6	4.2
	0.6	4.2	4.6	4.6	4.5
	0.9	4.5	3.5	4.7	4.3
n = 500	−0.9	3.6	4.1	4.3	3.7
	−0.6	4.5	4.6	4.1	3.9
	−0.3	3.6	3.7	3.2	3.7
	0.3	4.7	4.4	3.0	3.6
	0.6	3.9	3.7	2.9	3.4
	0.9	4.9	4.3	5.5	4.8

stabilized, closer to the nominal level of test 5%. In Table 2.3, we see the increasing of series length does not help the performance of the test.

The improvement for the type I errors for ARFIMA $(1,0,0)$ would be much noticeable as shown in Table 2.4.

Originally, in Table 2.2, severe type I error underestimations would take place: some of type I error would even shrink near zero. Although conservative models are recommended by some authors, such as Dufour (2006), accurate estimated type one error would be informative for us to identify the appropriate models for the data. Under the framework of Wild Monte Carlo test, we can see the test is not sensitive to the length series and lag number.

In order to further compare the Wild Monte Carlo Test here with the original Lin-McLeod Test, we redo the simulation study of Lin and McLeod (2006) for ARMA $(1,0)$ model.

Table 2.5 compares the Wild Monte Carlo Test ($\tilde{D}_m$) with the Lin-McLeod Test $\hat{D}_m$ under the first-order autoregressive models. With various choice of autoregressive parameter ϕ, under different nominal level $\alpha = 1\%, 5\%$, the type one errors for various models would stay at the same level for both the Lin-McLeod Test and the Wild Monte Carlo Test which gives credence to the validity of the Wild Monte Carlo Test.

2.5.2 Power Comparisons

Further, we wish to compare the performance for Wild Monte Carlo Test on various ARMA model. We choose the data generating process in Peňa and Rodríguez (2002) and compare the results between the Lin-McLeod Test and the Wild Monte Carlo test.

Table 2.4: Size Comparison for Wild Monte Carlo Reject Rate for ARFIMA $(1,0,0)$(%)with $\theta = \pm 0.3, \pm 0.6, \pm 0.9, d_f = 0.2, 0.4$ with the series length $n = 200, 500$ and lag $m = 10, 30$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3.

		m = 10		**m = 30**	
	θ/d_f	**0.2**	**0.4**	**0.2**	**0.4**
n = 200	−0.9	3.5	4.0	2.1	2.9
	−0.6	3.2	4.5	2.8	2.1
	−0.3	3.2	4.1	2.6	2.7
	0.3	3.5	3.7	3.6	2.5
	0.6	5.4	4.6	4.8	3.6
	0.9	3.5	3.6	4.6	4.7
n = 500	−0.9	4.6	4.3	3.0	4.5
	−0.6	3.8	3.4	3.1	3.4
	−0.3	3.3	3.6	3.7	4.2
	0.3	3.9	3.0	4.7	4.5
	0.6	3.9	4.2	3.3	4.4
	0.9	3.8	4.6	3.9	3.7

Table 2.5: Empirical Size Significance Levels Comparison between the Lin-McLeod Test ($\hat{D}_m$) and Wild Monte Carlo Test ($\tilde{D}_m$) under ARMA $(1,0)$ Model with $\phi = 0.1, 0.3, 0.5, 0.7, 0.9$, series length $n = 10^3$ and lag $m = 10, 20$. Choosing nominal level $\alpha = 1\%, 5\%$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3

		$\phi = 0.1$	$\phi = 0.3$	$\phi = 0.5$	$\phi = 0.7$	$\phi = 0.9$
$\alpha = 5\%$	m=10	5.0, 5.1	5.0, 5.3	4.9, 5.5	5.5, 4.7	5.6, 5.5
	m=20	5.2, 5.4	4.1, 4.8	4.0, 4.4	5.3, 5.1	5.5, 5.5
$\alpha = 1\%$	m=10	1.3, 1.1	1.0, 0.9	0.6, 0.9	1.4, 1.0	0.9, 1.2
	m=20	1.0, 1.2	0.8, 0.8	1.2, 1.2	0.9, 0.9	1.1, 1.1

For convenience, we state the data generating process as followed: ARMA (p,q) model is defined as:

$$\phi(B)Z_t = \theta(B)a_t, \tag{2.39}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p, \tag{2.40}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q. \tag{2.41}$$

Here we use a series of ARMA model to verify the power of Wild Monte Carlo test. The model parameters are set as Peňa and Rodríguez (2002) and are listed below:

- Model 1: $\theta_1 = -0.5$
- Model 2: $\theta_1 = -0.8$
- Model 3: $\theta_1 = -0.6$, $\theta_2 = 0.3$

- Model 4: $\phi_1 = 0.1, \phi_2 = 0.3$
- Model 5: $\phi_1 = 1.3, \phi_2 = -0.35$
- Model 6: $\phi_1 = 0.7, \theta_1 = -0.4$
- Model 7: $\phi_1 = 0.7, \theta_1 = -0.9$
- Model 8: $\phi_1 = 0.4, \theta_1 = -0.6, \theta_2 = 0.3$
- Model 9: $\phi_1 = 0.7, \theta_1 = 0.7, \theta_2 = -0.15$
- Model 10: $\phi_1 = 0.7, \phi_2 = 0.2, \theta_1 = 0.5$
- Model 11: $\phi_1 = 0.7, \phi_1 = 0.2, \theta_1 = -0.5$
- Model 12: $\phi_1 = 0.9, \phi_2 = -0.4, \theta_1 = 1.2, \theta_2 = -0.3$

We generate the data through the above ARMA models and fit the model by AR(1). We exam the residuals and evaluate the power of the Lin-McLeod Test and Wild Monte Carlo Test here.

Table 2.6: Power Comparison for ARMA Model (Peña and Rodríguez, 2002) between Lin-McLeod Test ($\hat{D}_m$) and Wild Monte Carlo Test ($\tilde{D}_m$) with series length $n = 10^3$ and lag $m = 1, 2, 3, 4, 5$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3

Model	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$
1	97.1	89.1	94.7	92.1	88.9	88.4	85.4	91.3	84.3	93.7
2	100	90.1	100	100	100	100	100	100	100	100
3	95.3	84.3	84.3	78.3	71.1	62.7	63.7	65.2	53.4	59.3
4	82.3	97.3	84.6	98.6	97.5	98.5	96.2	97.9	96.1	97.8
5	87.8	100	91.2	100	93.5	100	95.2	100	96.2	100
6	89.3	84.4	88.5	91.8	84.8	77.7	54.7	60.9	34.9	47.0
7	60.5	40.4	60.5	44.4	59.5	52.4	50.8	58.1	61.0	67.9
8	85.4	75.6	84.7	65.7	73.4	81.2	71.8	71.0	63.6	69.0
9	100	90.2	100	100	100	100	100	100	100	100
10	100	98.7	100	99.2	100	98.1	98.0	99.7	98.2	99.9
11	98.3	99.9	99.2	99.7	99.9	99.9	99.8	99.7	99.6	99.7
12	95.7	100	100	100	100	100	100	100	100	100

From Table 2.7, we can see the Wild Monte Carlo would achieve better results with regarding to the power. In general, the performances of the test are improved and in some cases, the powers are increasing with large level. Such as, when the data generating process is the MA (1) and we test for the lag $m = 20$, the Lin-McLeod Test would get 64.5% power while the Wild Monte Carlo test gains 80.4% power. The power increases by $(80.4 - 64.5)/64.5 = 24.2\%$. Another case should be mentioned is the model 8 as ARMA $(1, 2)$ with $\phi_1 = 0.4, \theta_1 = -0.6, \theta_2 = -0.3$, for lag $m = 20$, the power is increased by $(70.8 - 57.7)/57.7 = 22.7\%$. This exemplifies the

Table 2.7: Power Comparison for ARMA Model (Peña and Rodríguez, 2002) between Lin-McLeod Test ($\hat{D}_m$) and Wild Monte-Carlo Test ($\tilde{D}_m$) with series length $n = 10^3$ and lag $m = 10, 15, 20, 25$ and 30. Bootstrapping replication number is 10^2 and each time simulations number is 10^3

Model	$m = 10$		$m = 15$		$m = 20$		$m = 25$		$m = 30$	
	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$
1	79.9	89.9	73.0	84.8	64.5	80.4	61.7	77.8	59.3	75.9
2	100	100	100	100	100	100	100	100	100	100
3	47.3	51.9	43.0	53.6	39.9	51.9	37.7	49.0	33.1	45.6
4	94.2	97.6	90.6	96.8	87.5	95.1	84.0	93.4	83.1	93.5
5	99.7	100	100	100	100	100	100	100	100	100
6	28.6	35.9	27.1	36.8	22.9	31.6	23.5	32.8	18.6	26.3
7	68.2	69.5	69.1	69.9	62.8	66.3	62.7	72.3	56.6	69.9
8	68.3	76.6	61.0	72.3	57.7	70.8	53.0	69.2	53.3	67.4
9	100	100	100	100	100	100	100	100	100	100
10	94.3	99.2	98.0	99.7	88.7	97.8	88.0	99.7	82.6	94.1
11	99.8	100	99.9	100	99.4	99.9	98.9	99.9	98.9	99.1
12	100	100	100	100	100	100	100	100	100	100

advantages of the Wild Monte Carlo test in ARMA model here.

In order to fully investigate the effect of the kernel bandwidth b here, we compare the Lin-McLeod Test with Wild Monte Carlo Test when we have fixed number of autocorrelation, m and change the kernel bandwidth, l.

From Table 2.8 we see that when we fixed the number of autocorrelation m and change the kernel bandwidth l, the power has not change at a marginal level and generally when b gets reasonably large, the power will be optimal of the models.

As we can see, generally the Wild Monte Carlo Test would have higher power compare to the Lin-McLeod Test here. Another important model considered in Lin and McLeod (2006) is the long memory model. As one of the important long memory model, Fractionally Differenced White Noise (FD) model and its extension the ARFIMA model are evaluated. For comparison, we apply the Lin-McLeod Test and Wild Monte Carlo Test on these models to evaluate the performance on randomness test.

In the Tables 2.9 and 2.10, we generate the data based on FD model and test for its randomness. We employ the randomness test algorithm in Lin (2006). The FD model can be specified as followed:

$$\nabla^{d_f} Z_t = a_t, t = 1, 2, \cdots, n, \tag{2.42}$$

where a_t follows a standard normal distribution. Here, we choose the $d_f = \pm 0.1, \pm 0.2, \pm 0.3$ and various lags $m = 5, 10, 20, 30, 40$. We consider the series lengths $n = 128$ and $n = 256$ following the convention of Lin and McLeod (2006) to see the impact of the serious length on

Table 2.8: Power Comparison between MC and Wild MC test with $m = 20$ and $b = 5, 10, 20$. Bootstrapping iteration $N = 10^3$

Model	$b = 5$		$b = 10$		$b = 20$	
	$\tilde{D}_m$	$\hat{D}_m$	$\hat{D}_m$	$\hat{D}_m$	$\hat{D}_m$	$\hat{D}_m$
1	83.6	84.6	85.2	83.9	84.7	83.5
2	100	100	100	100	100	100
3	50.6	49.6	53.1	48.3	52.3	49.2
4	95.3	91.3	97.1	90.8	96.2	91.4
5	100	97.2	100	98.6	100	96.3
6	31.2	28.3	33.6	29.8	32.5	30.2
7	66.8	57.6	70.1	59.3	68.6	61.2
8	65.2	56.8	72.1	60.9	69.3	65.8
9	100	100	100	100	100	100
10	99.6	96.9	100	97.4	100	97.9
11	100	98.8	100	96.2	100	95.7
12	100	100	100	100	100	100

the test.

From Tables 2.9 and 2.10, we can see when the lag number m is relatively small $m < 5$, the difference for the performance between the Lin-McLeod Test and the Wild Monte Carlo is not noticeable. However, when the lag number m gets relatively large, the Wild Monte Carlo Test would outperform the Lin-McLeod Test. In some cases, the improvement is not marginal. For example, when $d = 0.2$ with lag $m = 20$, the Wild Monte-Carlo Test could achieve $(72.6 - 59.6)/59.6 = 21.8\%$ improvement. When $d = -0.2$ and lag $m = 30$, the achievement in power could be $(83.6 - 69.4)/69.4 = 20.5\%$ and $(50.2 - 36.6)/36.6 = 37.2\%$ with $m = 30, n = 128$. Therefore, we could safely obtain the conclusion that the Wild Monte Carlo Test could improve the power for long memory models such as fractional difference for randomness test.

Still, we are considering for the effect of the kernel bandwidth l in the performance for the long memory model.

From Table 2.11, we can see the kernel bandwidth l will have less impact on the Wild Monte Carlo Test with the Lin-McLeod Portmanteau test $\hat{D}_m$ than its on the Ljung-Box Test $\hat{Q}_m$ if we fixed the number of autocorrelation m.

In comparing the above models for its power, we can see the Wild Monte Carlo Test would be more powerful especially when the tested series have some dependence left. The simulation study shows the Wild Monte Carlo Test fix problem of underestimations for the type one errors and also improve the test power for various models.

Table 2.9: Power Comparison between MC Test ($\hat{D}_m$) and Wild MC Test ($\tilde{D}_m$) in Long Memory Structure (Lin and McLeod, 2006). Choosing $d_f = \pm 0.1, \pm 0.2, \pm 0.3$, series length $n = 128, 256, 512$ and lag $m = 1, 2, 3, 4, 5$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3

d	n	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
		$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$
−0.3	128	80.2	72.3	85.4	80.3	87.3	81.4	85.9	80.5	87.5	83.8
−0.2	128	49.7	49.2	54.9	52.0	55.5	51.6	53.3	53.2	54.8	53.9
−0.1	128	20.0	18.7	24.1	24.6	20.9	21.7	20.3	19.9	20.3	21.3
0.1	128	19.6	18.2	18.2	17.2	16.7	17.2	18.6	19.1	14.8	15.9
0.2	128	83.8	76.8	82.1	81.9	80.2	82.9	77.9	83.0	77.7	93.9
0.3	128	99.2	96.5	99.6	99.8	98.9	99.3	98.5	99.0	98.4	98.4
−0.3	256	96.3	91.1	99.0	96.7	99.3	97.1	99.3	97.9	99.6	98.8
−0.2	256	80.4	80.4	79.6	84.3	77.7	84.0	84.9	89.7	85.2	83.8
−0.1	256	33.2	29.1	31.0	30.4	34.4	34.0	29.8	30.6	33.5	34.2
0.1	256	36.9	38.8	34.8	33.7	34.4	34.7	32.8	34.6	31.3	35.8
0.2	256	99.4	97.3	99.2	99.2	98.7	98.8	98.9	98.9	97.8	98.0
0.3	256	97.9	100	100	100	100	100	100	100	100	100
−0.3	512	100	93.9	100	99.9	100	100	100	100	100	100
−0.2	512	96.6	90.6	99.0	96.7	99.2	98.3	99.6	98.2	99.2	98.9
−0.1	512	55.9	49.6	57.7	55.4	57.8	51.4	58.4	56.6	56.1	59.8
0.1	512	60.4	58.9	67.6	64.6	64.4	63.8	63.9	65.2	63.3	65.6
0.2	512	100	97.9	100	100	100	100	100	100	100	100
0.3	512	100	98.4	100	100	100	100	100	100	100	100

2.6 Application to Beveridge Wheat Prices

In order to compare the Portmanteau test, we apply the methods to the famous Beveridge Wheat Price Index, which presents the annual price data from 1500 through 1869, averaged over many locations in Western and Central Europe.

According to the Figure 2.1, we can see the data presents an obviously increasing trend with a persistent long memory autocorrelation function. After first differencing, the sample autocorrelations are very small and are randomly positive and negative so neither long memory or its negative counterpart , anti-persistence, seems to be present. Nevertheless Baillie (1996) analyzed the data and fit the ARFIMA $(2, 1, 0)$ and the ARFIMA $(2, 0, 0)$ models. We compare the performance of different Portmanteau tests on these two models with the change of the lag m.

From Table 2.12, we can see clearly that, when we fit the ARFIMA $(2, 0, 0)$ model, we underestimate the difference number which indicates the rejection of the null hypothesis for the Portmanteau test on the estimated residuals. We list various Portmanteau test p-value and find

Table 2.10: Power Comparison between MC Test ($\hat{D}_m$) and Wild MC Test ($\tilde{D}_m$) in Long Memory Structure (Lin and McLeod, 2006). Choosing $d_f = \pm 0.1, \pm 0.2, \pm 0.3$, series length $n = 128, 256, 512$ and lag $m = 10, 15, 20, 25, 30, 40$. Bootstrapping replication number is 10^2 and each time simulations number is 10^3

d	n	$m = 10$		$m = 15$		$m = 20$		$m = 25$		$m = 30$		$m = 40$	
		$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$
−0.3	128	83.5	87.4	78.5	86.5	77.6	88.0	73.8	84.5	73.1	84.1	64.9	77.1
−0.2	128	47.8	54.8	44.3	53.9	38.8	49.7	37.9	48.9	36.6	50.2	32.9	44.0
−0.1	128	16.3	19.8	13.4	18.4	14.6	17.2	12.8	17.6	11.2	14.7	12.9	15.5
0.1	128	11.9	13.8	12.6	16.4	11.0	14.6	10.3	13.6	9.0	9.9	8.5	10.3
0.2	128	71.1	79.8	65.7	78.3	59.6	72.6	58.7	73.3	58.8	70.5	53.7	65.4
0.3	128	97.8	98.8	96.7	98.9	95.3	98.1	93.7	96.7	93.5	96.6	90.5	95.5
−0.3	256	99.0	99.2	99.3	99.7	98.9	99.6	98.3	99.6	97.9	99.5	97.3	99.3
−0.2	256	82.0	86.1	78.6	86.5	74.3	85.0	73.0	83.8	69.4	83.6	62.7	76.0
−0.1	256	30.5	36.2	26.5	35.1	25.1	30.7	22.2	28.7	21.8	29.4	17.1	24.2
0.1	256	28.6	35.2	24.4	32.3	21.6	31.0	19.6	27.3	17.9	26.2	17.3	24.4
0.2	256	97.7	98.9	95.9	97.8	95.2	97.3	93.3	97.1	92.5	96.5	88.8	95.2
0.3	256	100	100	100	100	100	100	100	100	99.8	100	100	100
−0.3	512	100	100	100	100	100	100	100	100	100	100	100	100
−0.2	512	99.4	99.7	98.7	99.2	98.5	99.3	98.8	99.7	97.1	99.3	98.4	98.6
−0.1	512	51.0	57.9	52.4	60.3	42.5	54.5	42.3	53.7	40.9	53.4	38.3	51.2
0.1	512	56.0	62.7	53.3	64.1	46.9	59.3	43.8	55.2	39.3	53.7	39.4	51.2
0.2	512	100	100	100	100	99.9	99.9	100	100	99.9	100	100	100
0.3	512	100	100	100	100	100	100	100	100	100	100	100	100

the Wild MC Test could achieve the most consistent results with the change of lag m. Other Portmanteau tests, including Box-Pierce test (Box and Pierce, 1970), General-Variance Portmanteau test (Mahdi and McLeod, 2011), Fisher-Gallagher Test (Fisher and Gallagher, 2012a) and Lin-McLeod Test (Lin and McLeod, 2006), would achieve satisfactory results only when the lag m increases to enough and need more information to do the goodness-of-fit test. When we fit the appropriate ARFIMA $(2, 1, 0)$ model, we should fail to reject the hypothesis, which indicates the adequacy of models. The Wild MC Test would still give the most consistent results compare to other goodness-of-fit test. In this case, we can say the Wild MC test would improve our test performance. The performance of the various tests is visualized in Figure 2.2

The above examples shows the power of Wild MC Test in goodness-of-fit in practical dataset especially when the fitted models involve some long memory component such as the ARFIMA , the FD models here. Especially the Wild MC method shows a clear rejection for the inadequate model fitting as ARFIMA $(2, 0, 0)$ for the data.

Table 2.11: Power Comparison Wild MC in Long Memory Structure with b Change with $m = 40$ and $b = 10, 20, 40$. Test for $d = \pm 0.3, \pm 0.1$ and $n = 128, 256, 512$

d	n	$b = 10$		$b = 20$		$b = 40$	
		$\hat{D}_m$	Q_m	$\hat{D}_m$	Q_m	$\hat{D}_m$	Q_m
−0.3	128	77.5	71.1	82.1	70.9	79.2	68.3
−0.1	128	13.9	9.0	17.9	11.1	15.7	10.2
0.1	128	9.2	4.2	10.9	5.8	9.6	6.8
0.3	128	91.9	81.7	95.2	84.2	93.6	82.5
−0.3	256	99.7	88.9	100	86.3	100	85.6
−0.1	256	22.6	18.5	27.8	21.3	25.9	20.4
0.1	256	24.6	21.1	27.1	20.9	26.3	18.7
0.3	256	99.1	95.3	100	92.5	100	91.6
−0.3	512	98.5	96.1	100	95.9	100	93.1
−0.1	512	51.6	40.8	55.4	39.1	52.8	37.8
0.1	512	49.8	34.1	54.7	37.8	51.9	36.2
0.3	512	100	100	100	100	100	100

2.7 Conclusion

In this chapter, we could see the pitfalls of the Lin-McLeod Test (Lin, 2006) on the ARFIMA model could be successfully alleviated by the Wild Monte Carlo Test with application of Wild Dependent bootstrap (Shao, 2010).. The simulations and real world data examples shows the Wild MC Test would enjoy more consistent and reliable test results and is less sensitive to the lag m.

Figure 2.1: Beveridge(1925) Wheat Price Index Data Description with Autocorrelation Plot for Original Data and the Difference of the Original Data

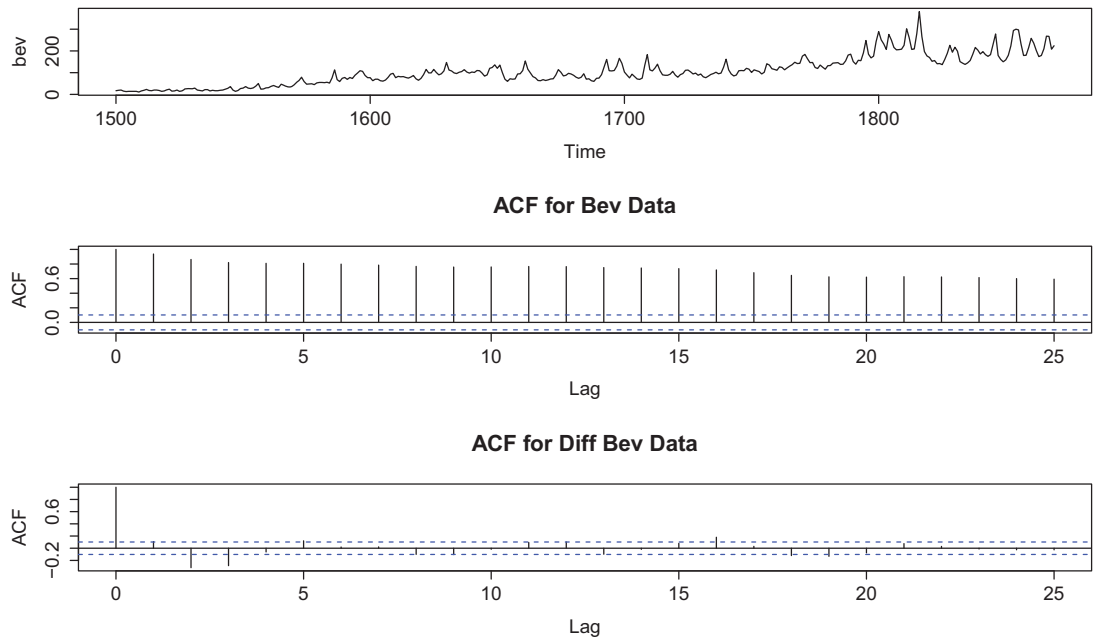


Figure 2.2: Power Comparison for Beveridge Wheat Prices for ARFIMA $(2, 1, 0)$ model and ARFIMA $(2, 0, 0)$ model for Various Portmanteau tests.

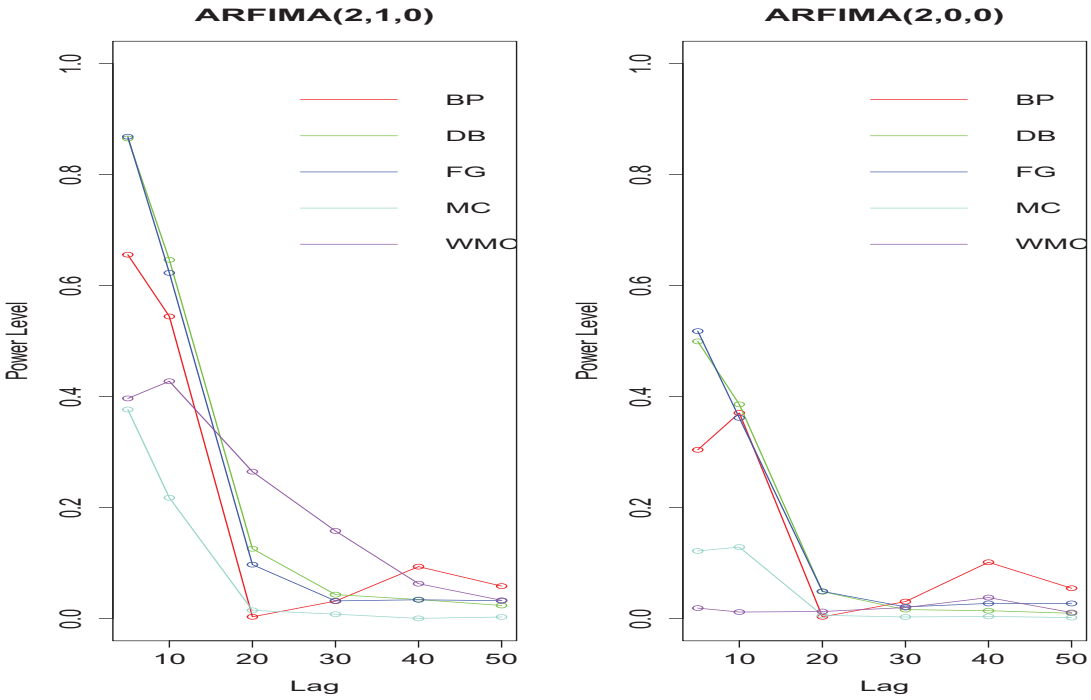


Table 2.12: p-value for Fitting ARFIMA Model on Beveridge Data between Box-Pierce, Mahdi-McLeod, Fisher-Gallagher, Lin-McLeod and Wild MC Test. Choosing the lag $m = 5, 10, 20, 30, 40, 50$.

	ARFIMA $(2, 1, 0)$					
Test/Lag	$m = 5$	$m = 10$	$m = 20$	$m = 30$	$m = 40$	$m = 50$
Box-Pierce	0.6555	0.5442	0.0032	0.0315	0.0936	0.0583
Mahdi-McLeod	0.8649	0.6462	0.1258	0.0432	0.03428	0.0234
Fisher-Gallagher	0.868	0.6227	0.0972	0.0319	0.0339	0.0322
Lin-McLeod	0.3766	0.2177	0.0149	0.0079	0.00399	0.0029
Wild MC	0.3996	0.4275	0.2647	0.1578	0.0629	0.0329
	ARFIMA $(2, 0, 0)$					
Box-Pierce	0.3040	0.3705	0.0032	0.0306	0.1015	0.0546
Mahdi-McLeod	0.4994	0.3859	0.0486	0.0162	0.0142	0.0096
Fisher-Gallagher	0.5178	0.3616	0.04907	0.0211	0.0273	0.0273
Lin-McLeod	0.1218	0.1288	0.0059	0.0029	0.00399	0.0019
Wild MC	0.0189	0.0118	0.0128	0.0199	0.0379	0.0109

Chapter 3

Improved Portmanteau Test for ARMA − GARCH Models

3.1 Introduction

Portmanteau tests plays an important role in evaluating the model performance in the Box-Jenkin modelling process to identify the appropriate model for the time series data. Based on the famous paper of Box and Pierce (1970) and Ljung and Box (1978), two portmanteau tests, Box-Pierce Test and Ljung-Box Test are defined as followed:

$$Q_{BP} = n \sum_{h=1}^{m} \hat{r}_k^2, \tag{3.1}$$

$$Q_{LB} = n(n+2) \sum_{h=1}^{m} \frac{\hat{r}_k^2}{n-k}. \tag{3.2}$$

where $\hat{r}_k$ is the autocorrelation function of lag $k, k = 1, 2, \ldots, m$. Under a certain lag m, we know these two portmanteau tests are following a χ^2 distribution. Later, Monti (1994) replaced the autocorrelation function $\hat{r}_k$ by the partial autocorrelation function $\hat{\pi}_k$ and found the asymptotic distribution for Monti Test:

$$Q_M = n(n+2) \sum_{k=1}^{m} \frac{\hat{\pi}_k^2}{n-k}. \tag{3.3}$$

where m is a certain number of lag window.

Peňa and Rodríguez (2002, 2006) suggested another type of portmanteau test based on the general variance of the autocorrelation matrix, hereafter, General Variance Portmanteau test,

$$\hat{\mathbf{R}}_{\mathbf{m}} = \begin{pmatrix} 1 & \hat{r}_1 & \cdots & \hat{r}_{m-1} \\ \hat{r}_1 & 1 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \hat{r}_1 \\ \hat{r}_{m-1} & \cdots & \hat{r}_1 & 1 \end{pmatrix}, \tag{3.4}$$

where $\hat{r}_k, k = 1, 2, \ldots, m$. Peňa and Rodríguez (2006) proposed the new portmanteau test statistic,

$$\hat{D}_m = -n \log \left| \hat{\mathbf{R}}_{\mathbf{m}} \right| . \tag{3.5}$$

and showed that the asymptotic distribution is the sum of weighted χ^2 distribution. In order to calculate the p-value for the test, they proposed to use the gamma approximation to obtain the approximate distributions. In order to avoid the inaccuracy of this approximation, Lin and McLeod (2006) proposed to use the Monte Carlo test to calculate p-value and showed its better performance over other tests, hereafter, the Lin-McLeod Test. Mahdi and McLeod (2011) further extended the general variance test to multivariate time series model and provide a optimal modification for Peňa and Rodríguez (2006) test as:

$$\hat{D}_m = -\frac{3n}{2m+1} \log \left| \hat{\mathbf{R}}_{\mathbf{m}} \right| . \tag{3.6}$$

The asymptotic distribution is still the sum of weighted χ^2 distributions. As we notice, one of the key assumptions for these statistics is that the series follows i.i.d. (independent and identical distribution) which enable the validity of asymptotic distribution.

In contrast to the above mentioned assumption, these statistics would become invalid once we replace the independent identical distribution (i.i.d.) assumption by the uncorrelated, but not independent assumption as weak white noise assumption, such as GARCH error. We will focus on ARMA − GARCH model in this chapter.

Recently, a large number of research papers attempted to weaken the assumption of strong white noise for the portmanteau test (Romano and Thombs, 1996; Francq and Zakoïan, 1998; Francq et al., 2005). In order to solve this problem of non-i.i.d. noise, some researchers (Chen and Deo, 2004; Deo, 2000) derived the spectral density test for the martingale difference hypothesis in the presence of conditional heteroscedasticity. Later, Lobato et al. (2001, 2002) addressed the problems to test the null hypothesis with time series to uncorrelated up to a fixed order m and proposed a modified Box-Pierce test for this null hypothesis. However, these tests do not solve the problem of weak white noise including the important special case of GARCH innovations. Francq et al. (2005) successfully derived the distribution of the residuals under the null hypothesis of uncorrelated, but not independent errors and then proposed another modification for Box-Pierce Test similar to McLeod (1978). They show the advantages and better performance of their tests compared to the original portmanteau tests.

Later, we consider the general variance portmanteau test on ARMA − GARCH model by using the asymptotic distribution of the autocorrelation and extend the general variance portmanteau

test (Mahdi and McLeod, 2011) to ARMA − GARCH model. Still, since the asymptotic distribution is the sum of weighted χ^2 distributions with complicated structure, using gamma approximation, like Peňa and Rodríguez (2002), still suffer from approximation errors. Therefore, we consider to use the Lin-McLeod Test (Lin and McLeod, 2006) to evaluate the test to avoid the inconsistency of the approximation. However, the nuisance parameters in the asymptotic distribution for the Lin-McLeod Test would still severely distort the size of the test when applied on the GARCH errors, as in the previous mentioned ARFIMA model. Therefore, we would use the Wild Monte Carlo Test to adjust the calculation for this potential issue and mitigate this limitation with potential performance improvement. The asymptotic distributions for the bootstrapped statistics are derived later in the chapter and the empirical evaluations show satisfactory results compared to the Lin-McLeod Test.

3.2 Model Specification

Since we are considering the portmanteau test for ARMA − GARCH models, we would follow the model specification as Francq et al. (2005) and Zhu and Li (2013). According to Box et al. (2008), ARMA (p, q) model,

$$\phi(B)X_t = \theta(B)a_t, \tag{3.7}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p, \tag{3.8}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q. \tag{3.9}$$

All the portmanteau tests (Box and Pierce, 1970; Ljung and Box, 1978; Peňa and Rodríguez, 2002, 2006; Mahdi and McLeod, 2011; Fisher and Gallagher, 2012a) and their asymptotic distributions are derived under the assumption that a_t is independent and uncorrelated series with mean of zero. Then according to Francq et al. (2005), we could classify the model into three categories based on the underlying assumption of error:

1. Strong ARMA model, $\{a_t\}$ is independent and uncorrelated sequence,
2. Weak ARMA model, $\{a_t\}$ is uncorrelated but not independent sequence. GARCH type of error served as the main type of errors,
3. Semi-strong ARMA model, $\{a_t\}$ is a martingale difference sequence.

For further discussion, we would restrict our discussions under the GARCH type error framework as an important kind of weak ARMA model. Therefore, we would define the GARCH type error as Francq and Zakoian (2011).

Definition 3.1. *A sequence* $\{a_t\}$ *is called a* GARCH (r, s) *process if* $E(a_t|a_u, u < t) = 0$ *for all* t, *and let* η_t *be an i.i.d. sequence with mean* 0 *and standard deviance* 1,

$$a_t = \sigma_t \eta_t, \tag{3.10}$$

$$\sigma_t^2 = \mathrm{Var}(a_t|a_u, u < t) = \omega + \sum_{i=1}^{s} \alpha_i a_{t-1}^2 + \sum_{j=1}^{r} \beta_j \sigma_{t-j}^2. \tag{3.11}$$

$\omega, \alpha_i, i = 1, 2, \ldots, s$ and $\beta_j, j = 1, 2, \ldots r$ and the first two conditional moments of $\{a_t\}$ are assumed to exist.

The GARCH type of error is firstly introduced by Bollerslev (1986) as an extension of ARCH (Engle, 1982), mainly used to model the financial market data. Moreover, according to Tsay (2005), many empirical studies show the economic and financial data following the ARMA model with uncorrelated but not independent error GARCH type errors (ARMA − GARCH). In addition, Francq et al. (2005) showed many nonlinear time series models admit the ARMA representation with these type of error terms, especially in financial market time series data.

For diagnostic checks, according to Francq et al. (2005), the distribution of autocorrelation function would be severely distorted under the GARCH errors with different asymptotic distribution. Therefore, some careful modifications for the original portmanteau test are needed to correct the distortion to achieve better diagnostic performance.

Therefore, Francq et al. (2005) proposed a modified Box-Pierce Test which gives better performance under the correct asymptotic distribution of autocorrelation for the purpose of goodness-of-fit test. From both theoretical and empirical perspectives, this statistics would have favorable features for ARMA − GARCH model and can yield better performance.

Since the asymptotic distribution for these tests embedded with complicated structure, usually it is difficult to obtain the p-value based on the asymptotic distributions. Using approximate distribution, like Peňa and Rodríguez (2002), will introduce approximation errors. Therefore, the Lin-McLeod Methods are proposed to evaluate the p-value using Monte Carlo test. However, as shown in later empirical experiments, the Lin-McLeod Test would still have the nuisance parameter impacts. In order to eliminate this potential issue, we would use the Wild Monte Carlo test in previous chapter to remove the effects of nuisance parameters.

For simplicity, we define $\hat{a}_t = a_t(\theta)$ as the residuals where $\theta = (p_i, q_j, \omega, \alpha, \beta)$ as a vector of parameters in the estimated model. Then we can define the following auto-covariance function for the residuals in the test.

$$\hat{\Gamma}(l, l') = \sum_{h=-\infty}^{\infty} E(\hat{a}_t \hat{a}_{t+l} \hat{a}_{t+h} \hat{a}_{t+h+l'}), \hat{c}_{ll} = \frac{1}{n} \sum_{l=1}^{n-l} \hat{a}_t \hat{a}_{t-l}, \tag{3.12}$$

for $(l, l') \neq (0, 0)$ and the autocorrelation function as:

$$\hat{r}_k = \hat{c}_k / \hat{c}_0, \tag{3.13}$$

where $\hat{c}_l$ is the auto-covariance function with lag k. Let $\hat{r}_m = (\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_m)$. Therefore, under the null hypothesis H_0, according to Francq et al. (2005), we know the following theorem:

Theorem 3.1. *Under the regulatory assumption, we have:*

$$\sqrt{n}\hat{r}_m \rightarrow_d N(0, \Sigma_{\hat{r}_m}), \tag{3.14}$$

where $\Sigma_{\hat{r}_m} = R_{m,m}$ *when* $p = q = 0$*, and when* $p > 0$ *or* $q > 0$

$$\begin{aligned}\Sigma_{\hat{r}_m} = R_{m,m} + \mathbf{\Lambda'_m}\left\{\mathbf{\Lambda}_\infty\mathbf{\Lambda'}_\infty\right\}\mathbf{\Lambda}_\infty\mathbf{R}^{-1}_{\infty,\infty}\mathbf{\Lambda'}_\infty\left\{\mathbf{\Lambda}_\infty\mathbf{\Lambda'}_\infty\right\}^{-1}\mathbf{\Lambda_m}\\ -\mathbf{\Lambda'_m}\left\{\mathbf{\Lambda}_\infty\mathbf{\Lambda'}_\infty\right\}^{-1}\mathbf{\Lambda}_\infty\mathbf{R}_{\infty,\mathbf{m}} - \mathbf{R}_{\mathbf{m},\infty}\mathbf{\Lambda'}_\infty\left\{\mathbf{\Lambda}_\infty\mathbf{\Lambda'}_\infty\right\}^{-1}\mathbf{\Lambda_m},\end{aligned}$$

where $\mathcal{R}(l, l') = \Gamma(l, l')/\sigma^4$ *and* $\mathbf{R}_{i,j} = \left(\mathcal{R}(l, l')_{i\le l\le l'\le i, 1\le l'\le j}\right)$ *for* $i, j = 1, 2, \cdots, \infty$ *and* $\mathbf{\Lambda_m}$ *is defined as the equation (5) in Francq et al. (2005),*

$$\mathbf{\Lambda_m} = (\lambda_\mathbf{1}, \lambda_\mathbf{2}. \ldots, \lambda_\mathbf{m}), \tag{3.15}$$

$$= \begin{pmatrix} -1 & -\phi^*_1 & \cdots & & & & -\phi^*_{m-1} \\ 0 & -1 & \ddots & & & & \vdots \\ \vdots & & \ddots & & & & \\ 0 & \cdots & & -1 & -\phi^*_1 & \cdots & -\phi^*_{m-p} \\ 1 & \psi^*_1 & \cdots & & \psi^*_p & \cdots & \psi^*_{m-1} \\ 0 & 1 & \ddots & & & & \vdots \\ \vdots & & \ddots & & & & \\ 0 & \cdots & & 1 & \psi^*_1 & \cdots & \psi^*_{m-q} \end{pmatrix}, \tag{3.16}$$

where ϕ^*_i *and* ψ^*_i *is defined as,*

$$\phi^{-1}(B) = \sum_{i=0}^{\infty}\phi^*_i B^i, \tag{3.17}$$

$$\psi^{-1}(B) = \sum_{i=0}^{\infty}\psi^*_i B^i, \tag{3.18}$$

and B *is the backward operator and* $\phi(B), \psi(B)$ *are the function with the parameters of* ARMA *model.*

Thanks to the above statement, we can successfully understand the distortion of GARCH type errors on the portmanteau test for the ARMA − GARCH model. Therefore, we are now developing a more accurate portmanteau with implementation of Mahdi and McLeod (2011) and Fisher and Gallagher (2012a).

3.3 Extension to Univariate General Variance Portmanteau Test

Based on the asymptotic distribution of the autocorrelation, we can extend its distribution to the partial autocorrelation function and then derive the distribution of the portmanteau test for Mahdi and McLeod (2011). With derived asymptotic distribution, the General Variance Portmanteau test on the ARMA − GARCH model is valid.

3.3.1 Asymptotic Distribution for Partial Autocorrelation Function

In order to further our study, we need to derive the asymptotic distribution for the General Variance Portmanteau statistics proposed by Peňa and Rodríguez (2002) and improved by Mahdi and McLeod (2011).

Under the the assumption of strong white noise, Mahdi and McLeod (2011) derived the asymptotic distribution as a sum of weighted χ^2 distributions. According to Gallagher and Fisher (2013), on various simulated data from models with strong white noise, the General Variance Portmanteau test outperforms other portmanteau tests such as Box-Pierce test (Box and Pierce, 1970), Ljung-Box test (Ljung and Box, 1978) and Monti test (Monti, 1994). However, under the GARCH type error assumptions, similar to the statement in Francq et al. (2005), the distribution of the autocorrelation and partial autocorrelation for residuals are dissimilar, leading to some potential issues as mentioned in previous chapter. Therefore, the asymptotic distribution for the General Variance Portmanteau test need to be adjusted.

Firstly, we need to extend the result of Francq et al. (2005) for the autocorrelation to the partial autocorrelation under ARMA − GARCH model. According to Francq et al. (2005), we can estimate the parameters of ARMA model using the Maximized Likelihood Estimation and obtain the residuals.

Theorem 3.2. *Under the* ARMA −GARCH *model, we define $\hat{\pi}_k$ as the partial autocorrelation function with lag k and $\hat{r}_k$ as the autocorrelation function of lag k for residuals, then we have:*

$$\sqrt{n}(\hat{\pi}_{(m)} - \pi_{(m)}) \to_d \mathcal{N}(0, \Lambda_m \mathbf{J}^{-1}\mathbf{I}_m\mathbf{J}^{-1}\Lambda_m^T\sigma^2), \tag{3.19}$$

where $\hat{\pi}_{(m)}$ is defined as $(\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_m)$.

Proof:

According to Monti (1994), equation 1, we can see the partial autocorrelation function π_k could be treated as a function of autocorrelation function $(r(k))$:

$$\pi_k = \Psi(r(k)) = \frac{r_k - r_{(k-1)}^T R_{(k-1)}^{-1} r_{(k-1)}^*}{1 - r_{(k-1)}^T R_{k-1}^T r_{(k-1)}}, \tag{3.20}$$

where $r_{(m)} = (r_1, r_2, \ldots, r_m)^T$, $\pi_{(m)} = (\pi_1, \pi_2, \ldots, \pi_m)^T$, $r_{(m)}^* = (r_m, r_{m-1}, \ldots, r_1)$,

$$R_{(k)} = \begin{pmatrix} r_1 & r_2 & \cdots & r_k \\ r_2 & r_1 & \cdots & r_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_k & r_{k-1} & \cdots & r_1 \end{pmatrix},$$

and the estimated partial correlation function $\hat{\pi}(k)$ has the same relationship with autocorrelation function $\hat{r}_k$. In addition, if the model is correct identified, we could see that $\hat{\pi}_{(k)} =$

$\hat{r}_{(k)} + O_p(n^{-1})$. Therefore, we could expand $\hat{\pi}_{(k)} = \Psi(\hat{r}_{(k)})$ around $r_{(k)}$, yielding:

$$\hat{\pi}_{(m)} = \pi_m + \frac{\partial \pi_{(k)}}{\partial r_{(k)}} (\hat{r}_{(m)} - r_{(m)}) + O_p(n^{-1}). \tag{3.21}$$

Under correct model identification, from equation 3.21, we could know that $\partial \pi_k / \partial r_k = I_m + O_p(n^{-1})$, according to Francq et al. (2005),

$$\hat{r}_{(k)} - r_{(k)} = \sigma^2 \Lambda_m^T (\hat{\theta} - \theta_0) + O_p(n^{-1}), \tag{3.22}$$

where Λ_m is as defined in equation 3.15:

$$\Lambda_m = (\lambda_1, \lambda_2, \ldots, \lambda_m)$$
$$= \begin{pmatrix} -1 & -\phi_1^* & \cdots & & & & -\phi_{m-1}^* \\ 0 & -1 & \ddots & & & & \vdots \\ \vdots & & \ddots & & & & \\ 0 & \cdots & & -1 & -\phi_1^* & \cdots & -\phi_{m-p}^* \\ 1 & \psi_1^* & \cdots & & \psi_p^* & \cdots & \psi_{m-1} \\ 0 & 1 & \ddots & & & & \vdots \\ \vdots & & \ddots & & & & \\ 0 & \cdots & & 1 & \psi_1^* & \cdots & \psi_{m-q}^* \end{pmatrix}.$$

where ϕ_k^*, ψ_k^* are defined as McLeod (1978):

$$\phi^{-1}(B) = \sum_{i=0}^{+\infty} \phi_i^* B^i, \tag{3.23}$$

$$\psi^{-1}(B) = \sum_{i=0}^{+\infty} \psi_i^* B^i. \tag{3.24}$$

Therefore, due to Francq and Zakoïan (1998), we know that

$$\sqrt{n}(\theta - \theta_0) \to_d \mathcal{N}(0, \mathbf{J^{-1}IJ^{-1}}), \tag{3.25}$$

where $\mathbf{I} = \mathbf{I}(\theta_\mathbf{0})$ and $\mathbf{J} = \mathbf{J}(\theta_\mathbf{0})$,

$$\mathbf{I}(\theta) = \lim_{n \to +\infty} \operatorname{var}\left\{ \sqrt{n} \frac{\partial \log L_n(\hat{\theta})}{\partial \theta} \right\}, \tag{3.26}$$

$$\mathbf{J}(\theta) = \lim_{n \to \infty} \frac{\partial^2}{\partial \theta \partial \theta^T} \log L_n(\hat{\theta}). \tag{3.27}$$

$\log L_n(\hat{\theta})$ is the log-likelihood function evaluated at $\theta = \hat{\theta}$. Combined with equation 3.22, we can have:

$$\sqrt{n}(\hat{\pi}_{(m)} - \pi_{(m)}) \cong \sqrt{n}(\hat{r}_{(m)} - r_{(m)}) \to_d \mathcal{N}(0, \Lambda_m \mathbf{J}^{-1} \mathbf{I}_m \mathbf{J}^{-1} \Lambda_m^T \sigma^2). \tag{3.28}$$

We can see under the ARMA $-$ GARCH model, generally, the partial autocorrelation function $\hat{\pi}_m$ is followed similar normal distribution to the autocorrelation function $\hat{r}_m$. The above proof uses the major results in Monti (1994) and Francq et al. (2005).

3.3.2 Asymptotic Distribution for General Variance Portmanteau Statistics $\hat{D}_m$

With the distribution for the partial autocorrelation function in ARMA − GARCH model, we can further attempt to derive the asymptotic distribution for the General Variance Portmanteau test.

Theorem 3.3. *Under the* ARMA − GARCH *model, the general variance portmanteau test* $\hat{D}_m$ *defined by Mahdi and McLeod (2011) has the following asymptotic distribution:*

$$\mathcal{D}_m \to_d \sum_{i=1}^{m} \lambda_i \chi^2_{1,i}, \tag{3.29}$$

where $\lambda_i, i = 1, 2, \ldots, m$ *are the eigenvalues of matrix* $\Lambda_m \mathbf{J}^{-1}\mathbf{I}_m\mathbf{J}^{-1}\mathbf{\Lambda}_m^T\mathbf{\Lambda}_m\mathbf{J}^{-1}\mathbf{I}_m\mathbf{J}^{-1}\mathbf{\Lambda}_m^T\sigma^4$.

Proof:

Based on the above distribution, we attempt to derive the asymptotic distribution of $\mathcal{D}_m$. The following proof follows the general steps of Peňa and Rodríguez (2002). According to Mahdi and McLeod (2011), the univariate General Variance Portmanteau test statistics $\mathcal{D}_m$ is defined as:

$$\mathcal{D}_m = -\frac{3n}{2m+1} \log \left|\hat{\mathbf{R}}_{\mathbf{m}}\right|. \tag{3.30}$$

According to Ramsey et al. (1974), we can rewrite:

$$\mathcal{D}_m = -n \sum_{i=1}^{m} \frac{3(m+1-i)}{2m+1} \log(1-\hat{\pi}_i^2). \tag{3.31}$$

Assume $(n\hat{\pi}_1, n\hat{\pi}_2, \cdots, n\hat{\pi}_m)$ is asymptotic distributed as Y. Based on the multivariate δ method in Arnold (1990), we can say:

$$\mathcal{D}_m \to_d \left(\frac{3m}{2m+1}, \frac{3(m-1)}{2m+1}, \ldots, \frac{3}{2m+1}\right) Y, \tag{3.32}$$

and from the Cramer-Wold theorem,

$$\begin{aligned}\left(\frac{3m}{2m+1}, \frac{3(m-1)}{2m+1}, \cdots, \frac{3}{2m+1}\right)&\left(n\hat{\pi}_1^2, n\hat{\pi}_2^2, \ldots, n\hat{\pi}_m^2\right)\\ &\to_d \left(\frac{3m}{2m+1}, \frac{3(m-1)}{2m+1}, \ldots, \frac{3}{2m+1}\right) Y.\end{aligned}$$

If the model is correctly identified, from the equation 3.28, we can see

$$\sqrt{n}\hat{\pi}_{(m)} \to_d \mathcal{N}(0, \Lambda_m \mathbf{J}^{-1}\mathbf{I}_m\mathbf{J}^{-1}\Lambda_m^T\sigma^2).$$

Therefore, we have $\mathcal{D}_m \rightarrow_d \sum_{i=1}^{m} \lambda_i \chi_1^2$ where $\lambda_i, i = 1, 2, \ldots, m$ are the eigenvalues of matrix $\Lambda_m \mathbf{J}^{-1} \mathbf{I}_m \mathbf{J}^{-1} \Lambda_m^T \Lambda_m \mathbf{J}^{-1} \mathbf{I}_m \mathbf{J}^{-1} \Lambda_m^T \sigma^4$.

Generally, under the framework of ARMA − GARCH model, the D_m will follow a sum of weighted χ^2 distribution. Therefore, as a typical ARMA − GARCH model (Francq et al., 2005), the above asymptotic distribution can be applied to ARMA (p, q)−GARCH (r, s) model:

$$(1 - \phi_1 B - \cdots - \phi_p B^p) X_t = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t, \tag{3.33}$$

where the error term a_t can be defined as:

$$a_t = \sigma_t \eta_t, \tag{3.34}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{r} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2. \tag{3.35}$$

And η_t is i.i.d. sequence with normal distribution of mean 0 and standard deviance 1. For an observed time series data, we can fit the data with the ARMA − GARCH model and test against the null hypothesis:

$$H_0\text{: } X_t \text{ admit a ARMA} - \text{GARCH model.}$$

For better illustrative purposes, we restrict the discussions to the residuals of ARMA −GARCH model, a typical weak ARMA model. Based on previous derivation, with a more complicated asymptotic distribution specified above, the p-value for the portmanteau test on ARMA − GARCH model is harder to implement with the gamma approximation in Mahdi and McLeod (2011). In order to decide the p-value of the general variance portmanteau test and avoid approximation errors , we choose to use the Lin-McLeod test (Lin and McLeod, 2006) for the test.

3.4 Lin-McLeod Test and its Extension

3.4.1 Lin-McLeod Test and its Limitations

Due to the complicated structure of the asymptotic distribution, two approaches are proposed to handle the p-value calculation: Peňa and Rodríguez (2002) suggest to use the gamma approximation and use the approximate distribution to obtain the p-value. This approach will inevitably introduce approximation errors and larger number of observations are needed to gain an acceptable level of approximation. In order to solve these problems, Lin and McLeod (2006) proposed to use the Lin-McLeod Test based on the Monte Carlo test to numerically calculate the p-value for the general variance portmanteau test (Peňa and Rodríguez, 2002). Later, Mahdi and McLeod (2011) still use the Lin-McLeod test to evaluate the newly proposed multivariate portmanteau test. For simplicity, we restate the algorithm here.

Algorithm 3.1. *The following algorithm is a simple restatement of Lin and McLeod (2006): for a given time series data,*

Step One: *Fit the predetermined model with appropriate number of p, q, r, s. Obtain the estimated parameter space $\hat{\Lambda}$ for equation 3.33 as well as the residuals $\hat{a}_t, t = 1, 2, \ldots, n$ following* GARCH *type errors with estimated parameters. Calculate the $\hat{D}_m$ based on $\hat{a}_t$, where $\hat{D}_m$ is defined as general variance portmanteau test in Lin and McLeod (2006).*

Step Two: *Re-simulate the $X_t^*, t = 1, 2, \ldots, n$ based on the estimated parameter space $\hat{\Lambda}$. Refit the simulated $X_t^i, t = 1, 2, \ldots, n$ based on the number of parameters and obtain the residuals ^i_t. Calculate the statistics $\hat{D}_m^i$.*

Step Three: *Repeat the* **Step Two** *for N times. The p-value could be calculated as followed:*

$$p\text{-value} = \frac{\#\left\{\hat{D}_m^i > \hat{D}_m\right\} + 1}{N + 1} \tag{3.36}$$

The usual Lin-McLeod test on ARMA − GARCH model suffers from the following two limitations: firstly, according to Section 4 in Dufour (2006), the nuisance parameters existing in the model would result in inaccurate type I error. With the proposed Maximized Monte Carlo algorithm, Dufour (2006) can solve alleviate issue, but this approach is hard to implement with cumbersome calculation caused by the optimization. Especially in ARMA − GARCH model, the heteroscedasticity lying in the residuals would change the observed statistic value. Due to the fact that p-value is calculated based on the empirical distribution of the statistics in the simulation and the observed statistic value, An inaccurate observed value would inevitably contribute to a inaccurate type I error and lower test power. Secondly, the consistency of parameter estimation would further lead to the distortion of type I error and lower power. Especially when the error follows an GARCH type white noise, the parameters estimated based on Maximized Likelihood Estimator (MLE) would be biased and inconsistency (see Li (2004)).

In order to further diminish the aforementioned issues, we consider the Wild Monte Carlo Test. The size distortion problem for the portmanteau tests when applied on the time series model with heteroscedasticity has been pointed out. Phillips and Xu (2006) derive the parameters inferences for AR(p) model with heteroscedasticity. Lobato et al. (2001) propose a modification of Box-Pierce test to be more robust to the model with GARCH type of errors. Godfrey and Tremayne (2005) solve the problems by implementing the wild bootstrap method in the estimating the regression error term. Their simulations show good power for Wild Bootstrap to remove the effect of heteroscedasticity and improve the test power. Recently, Harris and Kew (2014) propose to alleviate the problems with GARCH type errors through the modification of the autocorrelation function with weights. The weighted autocorrelation is defined as followed:

$$\hat{\rho}_j(\omega) = \frac{\sum_{t=1}^N \omega_t^2 X_{t-j} X_t}{\sum_{t=1}^N w_t^2 X_t^2}. \tag{3.37}$$

where ω_t is the fix weighted related to the time t. By using this weighted autocorrelation with kernel function, they showed the performance of the portmanteau test is improved with higher power in the test. In their cases, they choose the weights from a non-parametric estimator of the variance function of X_t. After carefully adjusting, the proposed Wild Monte Carlo Test can combine the ideas of Godfrey and Tremayne (2005) and Harris and Kew (2014) by using the Wild Dependent bootstrap method (Shao, 2010) to improve the performance of portmanteau test in the ARMA − GARCH model.

3.4.2 Wild Monte Carlo Test and its Asymptotical Properties

Several ways are proven to reduce these nuisance parameter effect on the Lin-McLeod test. Based on the previous chapter, we propose the Wild Monte Carlo Test through perturbing the estimated the residuals in the Lin-McLeod Test. This method is shown to be efficient to remove effect of nuisance parameters in ARFIMA model to handle the complicated asymptotic distributions. The underestimated size distortion (Francq et al., 2005) would be addressed by this algorithm and the improvement of the power on ARMA − GARCH models would be expected here.

The perturbed residuals would introduce weights when applying the portmanteau test similar to Harris and Kew (2014). Still, we restate the algorithm here:

Algorithm 3.2. *The algorithm for the Wild Monte Carlo Test is stated as follows: for a given time series data*

- **Step One:** *Fit the series X_t with appropriate $(p_i, q_j, \omega, \alpha_n, \beta_n)$ and obtain the the parameter estimates as well as the residuals $\hat{a}_t$. Simulate dependent random variable $W_t, t = 1, 2, \ldots, n$ based on the kernel outlined as followed. Generate the pseudo-observations,*

$$\hat{a}_t^* = \hat{a}_t \times W_t. \tag{3.38}$$

  and then calculate the D_M^* statistics based on $\hat{a}_t^*$. Here D_M^* can be defined as General Variance Portmanteau Test or other type of portmanteau test

- **Step Two:** *Re-simulate the model based on the estimated $(p_i, q_j, \omega, \alpha_n, \beta_n)$ and a_t follows a* GARCH *type errors with estimated parameters. Refit the model obtaining the residuals $\hat{a}_t^j$. Generate the dependent random variable W_t^* based on the same kernel in* **Step One**. *Generate the pseudo-observations $\hat{a}_t^{j*}$,*

$$\hat{a}_t^{j*} = \hat{a}_t^j \times W_t^* \tag{3.39}$$

  and then calculate the statistics D_m^{j*}.

- **Step Three:** *Repeat the* **Step Two** *for N times and obtain the estimated p-value.*

$$p\text{-}value = \frac{\#\left\{D_m^{j*} > D_m^*\right\} + 1}{N + 1}. \tag{3.40}$$

Then the dependent random variable W_t can be outlined as:

Assumption 3.1. *The random variables $\{W_i\}_{t=1}^n$ are independent of the series $X_t, t = 1, 2, \ldots, n$. $E(W_t) = 0, Var(W_t) = 1$ for $t = 1, 2, \ldots, n$. Assume that W_t is a stationary process with $Cov(W_t, W_{t'}) = a\{(t - t')/l\}$, where $a(\cdot)$ is a kernel function and $l = l_n$ is a bandwidth parameter. Furthermore, assume that*

$$K_a(x) = \int_{-\infty}^{+\infty} a(u)e^{-iux}du \geq 0, \tag{3.41}$$

for $x \in \mathbb{R}$

We can use the Parzen kernel to generate the dependent random variable:

Definition 3.2. *Parzen kernel could be stated as followed:*

$$K_a(x) = \begin{cases} 1 - 6(x/6l)^2 + 6\ x/6l|^3 & 0 \leq |x/l| \leq 1/2, \\ 2(1 - \ x/6l|^3) & 1/2 \leq |x/l| \leq 1, \\ 0 & \text{Otherwise}\,. \end{cases} \tag{3.42}$$

When we use the Wild Monte Carlo test on the residuals, the kernel applied on the critical value estimation procedure would control the observed statistics value and force it close to the natural level. While we simulate the series according to the fitted model and refit the model, the choice of kernels will not affect the statistics due to the lack of heteroscedasticity in the simulated residuals. Therefore, the type I error would be further corrected through use of the Wild Monte Carlo test.

When the model is correctly identified, if the observed statistics value is larger than it should be, the size estimated would be underestimated. Furthermore, if the model is incorrectly identified, the power would be inevitably lower than it should be. At this time, the Wild Monte Carlo Test would not only amend the nuisance parameters problem in the Lin-McLeod Test, but improve the power of the test.

Additionally, the perturbed estimated error would get implemented by the weights on the autocorrelation, leading to the performance improvement on the portmanteau test. The autocorrelation function for the bootstrapped statistics can be easily written as,

$$\hat{\rho}_t = \frac{\sum a_t^* a_{t-k}^*}{\sum (a_t^*)^2} = \frac{\sum W_t W_{t-k} a_t a_{t-k}}{\sum W_t^2 a_t^2}. \tag{3.43}$$

The autocorrelation works similar to the weighted autocorrelation function in equation 3.37 as fixed weights chosen from the Parzen kernel in the Wild Monte Carlo Test. This algorithm is implemented in the **portes** package (Mahdi and McLeod, 2013) in R. Therefore, we would need to employ the High Performance Computing in Canada with 36 cores parallel. The CPU running time is around 70 days.

3.5 Simulation Results

In order to fully evaluate the validity of the Wild Monte Carlo test compared with the General Variance Portmanteau test, the Wild MC test is proven to be a more stable approach to evaluate the goodness-of-Fit in the model.

Li and Mak (1994) theoretically derived the asymptotic distribution for the Box-Pierce Test on non-linear time series model with GARCH errors. Therefore, it is reasonable to compare these tests on the ARMA − GARCH models.

3.5.1 Size Comparison

First, we compare the sizes to of what correctly identified ARMA − GARCH model. In order to further evaluate the performance of the tests on the ARMA − GARCH model, we apply the Wild Monte Carlo Test with Parzen kernel and the Lin-McLeod Test on the AR(1) − ARCH(1) model to compare with the results of Francq et al. (2005). We defined the data generating process as follows:

$$X_t = \phi_1 X_{t-1} + a_t, \tag{3.44}$$

$$a_t = \sqrt{1 + \alpha_1 a_{t-1}^2 \eta_t}, \tag{3.45}$$

where η_t is defined as i.i.d normal distribution $N(0, 1)$. We simulate the data from the following model as Francq et al. (2005):

Model 1 : $\phi_1 = 0$ and $\alpha_1 = 0$,

Model 2 : $\phi_1 = 0$ and $\alpha_1 = 0.2$,

Model 3 : $\phi_1 = 0.5$ and $\alpha_1 = 0.2$,

Model 4 : $\phi_1 = 0.9$ and $\alpha_1 = 0.2$,

Model 5 : $\phi_1 = 0.9$ and $\alpha_1 = 0.4$.

Then we fit the these series with corresponding fitting models and applied the different tests on the residuals. Once the model is correct identified, the type I error for the tests can be evaluated. In the following table,the Q_{BP} denotes the Standard Box-Pierce Test with adjustment, Q^*_{BP} denotes the modified Box-Pierce Test (Francq et al., 2005), $\hat{D}_m$ denotes the Lin-McLeod Test (Lin, 2006), $\tilde{D}_m$ denotes the Wild Monte Carlo Test with Parzen kernel. We compare these tests on the above five models with various lag $m = 2, 4, 12, 24$ and 36. The simulation is based on 10^3 replications with series length of 10^3 as defined in Francq et al. (2005)

Table 3.1: Size Comparison between ARMA $(1,0)$ – GARCH $(1,1)$ for Box-Pierce Q_{BP}, Modified Box-Pierce Q^*_{BP}, Monte Carlo $\hat{D}_m$and Wild Monte Carlo Test $\tilde{D}_m$. Choosing lag $m = 2, 4, 12, 24$ and 36 and Nominal Level $\alpha = 5\%$.

	m	Q_{BP}	Q^*_{BP}	$\hat{D}_m$	$\tilde{D}_m$
	2	5.3	5.2	5.2	5.8
	4	5.6	5.2	4.5	5.1
Model 1	12	4.6	4.3	5.3	5.5
	24	5.6	4.9	5.5	5.0
	36	5.4	4.9	4.8	4.8
	2	7.2	5.8	6.2	5.7
	4	4.2	3.7	5.8	5.7
Model 2	12	5.9	5.0	5.4	5.3
	24	4.5	3.5	4.9	5.0
	36	4.6	5.6	4.7	5.8
	2	7.7	5.8	10.0	5.5
	4	5.9	4.3	6.6	3.1
Model 3	12	5.1	3.8	5.6	5.5
	24	5.1	3.9	4.9	6.0
	36	3.6	2.8	5.2	5.4
	2	15.4	5.7	6.4	4.6
	4	9.2	5.2	6.5	4.8
Model 4	12	6.4	5.0	9.0	5.3
	24	7.4	5.0	6.7	6.5
	36	4.7	4.0	5.7	5.5
	2	26.3	5.1	6.9	5.6
	4	20.0	5.6	6.2	4.6
Model 5	12	12.2	5.3	6.6	4.6
	24	10.7	5.7	2.3	5.8
	36	8.2	4.7	8.8	4.1

From the above simulation results, we can see the Box-Pierce Test and the Lin-McLeod Test would suffer from the over large size especially when the the lag m is a small number. When the parameters are more wildly defined like in model 5, the size distortion would be more severe. The modified Box-Pierce Test (Francq et al., 2005) and the Wild Monte Carlo Test would outperform the other two tests here. These two tests would perform at the nearly same level of size on various cases. This supports our conclusions about the validity of Wild Monte Carlo Test on ARMA – GARCH models.

3.5.2 Power Comparison

In order to further compare the Lin-McLeod Test and the Wild Monte Carlo test, we followed the structure of Peňa and Rodríguez (2002) and Lin (2006) to test the performance of the these tests for randomness diagnostic. First, We defined the data generating process,

$$Y_t = a_t \sigma_t, \tag{3.46}$$

$$\sigma_t^2 = 1 + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{3.47}$$

where $\alpha \geq 0, \beta \geq 0$ and $\alpha + \beta < 1$. This defines a typical covariance stationary GARCH $(1,1)$ model. Then we could choose the following two models.

Model A : $\alpha = 0.05, \beta = 0.9$.

Model B : $\alpha = 0.15, \beta = 0.8$.

We use the Lin-McLeod Test for randomness from Lin (2006) to evaluate the performance of the tests.

Table 3.2: Power Comparison between Lin-McLeod Test $\hat{D}_m$, Wild MC Test $\tilde{D}_m$ in GARCH $(1,1)$ Model. Choosing series length $n = 250, 500, 1000$, lag $m = 12, 24, 32$ and Nominal Level $\alpha = 5\%$. $\hat{D}_m$ stands for Monte Carlo Test and $\tilde{D}$ stands for Wild Monte Carlo Test. The p-value is based on 10^3 simulation and the power is evaluated at 10^3 iteration.

Model	n	$m = 12$		$m = 24$		$m = 32$	
		$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$	$\hat{D}_m$	$\tilde{D}_m$
A	250	28.9	24.6	29.1	31.1	27.6	30.0
B	250	84.1	82.6	80.7	82.1	79.4	83.7
A	500	49.9	43.7	52.5	52.5	48.9	50.4
B	500	99.2	98.3	98.4	98.1	98.4	99.0
A	1000	79.7	71.5	82.7	81.54	79.7	81.74
B	1000	100	100	100	100	100	100

From the simulation results in the table 3.2, The Wild Monte Carlo Test would still remains at the same level of power compare to the traditional Lin-McLeod Test. When we choose a large number for lag m, the Wild Monte Carlo Test would overpower the performance of the other two tests.

We intend to compare the Wild Monte Carlo Test with the Modified Box-Pierce Test (Francq et al., 2005). In the following simulation, we compare the power for goodness of fit on the ARMA $-$ GARCH model based on Francq et al. (2005). In the section, we simulate the data based on various ARMA $(1,1)$ - GARCH $(1,0)$ models with different $\theta, \phi, \alpha, \beta$ and fit the data

with AR(1) model. Then we compare the performance of various tests including Box-Pierce Test, Modified Box-Pierce Test (Francq et al., 2005), the Lin-McLeod Test (Lin, 2006). In the simulation study, we choose $n = 10^3$, $N = 10^3$ and choose the lag number $m = 2, 4, 12, 24$ and 36 to compare the power for different tests.

The data generating process is defined as followed:

$$X_t = \phi_1 X_{t-1} + a_t + \theta_1 a_{t-1}, \tag{3.48}$$

$$a_t = \sqrt{1 + \alpha_1 a_{t-1}^2 \eta_t}, \tag{3.49}$$

where η_t is i.i.d normal distribution $N(0, 1)$. Then the data is generating by the following five models.

- Model 1 $\phi_1 = 0, \theta_1 = 0.2, \alpha_1 = 0$.
- Model 2 $\phi_1 = 0, \theta_1 = 0.2, \alpha_1 = 0.2$.
- Model 3 $\phi_1 = 0.5, \theta_1 = 0.2, \alpha_1 = 0.2$.
- Model 4 $\phi_1 = 0.9, \theta_1 = 0.2, \alpha_1 = 0.2$.
- Model 5 $\phi_1 = 0.9, \theta_1 = 0.2, \alpha_1 = 0.4$.

In table 3.3, the Q_{BP} denotes the Standard Box-Pierce Test with adjustment, Q^*_{BP} denotes the modified Box-Pierce Test (Francq et al., 2005), $\hat{D}_m$ denotes the Lin-McLeod Test (Lin, 2006), $\tilde{D}_m$ denotes the Wild Monte Carlo Test with Parzen kernel. Best performance test is bolded in the table.

We can see from the table 3.3, the Wild Monte Carlo test with Parzen kernel is significantly better than other tests with larger power in the models listed above.

Table 3.3: Power Comparison for Box-Pierce Test Q_{BP}, Modified Box-Pierce Test Q^*_{BP}, Lin-McLeod Test $\hat{D}_m$ and Wild Monte Carlo Test $\tilde{D}_m$. Lag $m = 2, 4, 12, 24, 36$ and Nominal Level $\alpha = 5\%$.

	m	Q_{BP}	Q^*_{BP}	$\hat{D}_m$	$\tilde{D}_m$
	2	27.0	27.8	26.9	**87.5**
	4	19.3	18.0	21.6	**35.6**
Model 1	12	11.8	11.0	14.3	**20.0**
	24	9.9	8.20	13.0	**18.7**
	36	6.9	10.5	10.6	**15.4**
	2	19.1	23.1	28.3	**86.8**
	4	17.2	20.0	24.3	**37.2**
Model 2	12	10.1	9.1	17.2	**24.5**
	24	9.6	8.1	12.0	**18.2**
	36	10.5	7.5	11.9	**15.7**
	2	90.7	91.7	97.9	**100**
	4	89.7	89.5	95.2	**99.8**
Model 3	12	76.2	73.5	89.8	**96.3**
	24	56.8	55.9	99.7	**100**
	36	47.8	47.3	74.8	**89.0**
	2	99.7	99.6	**100**	**100**
	4	99.3	99.2	**100**	**100**
Model 4	12	97.7	97.5	99.8	**100**
	24	91.3	91.6	98.6	**99.5**
	36	88.7	89.0	98.9	**99.7**
	2	96.7	95.9	100	**100**
	4	96.2	95.0	99.9	**100**
Model 5	12	92.5	91.8	99.1	**99.5**
	24	86.1	85.0	97.7	**99.2**
	36	81.4	80.6	96.3	**98.7**

From the table 3.3, we see that the Wild Monte Carlo test generally has the highest power in model identification. When we have Data Generating Process (DGP) following ARMA − GARCH model, we fit them with the strong ARMA model. Then during the Lin-McLeod Test, the observed statistics would increase due to the heteroscedasticity in the residuals. However, we would re-simulate the data based on a strong ARMA model and fit it.

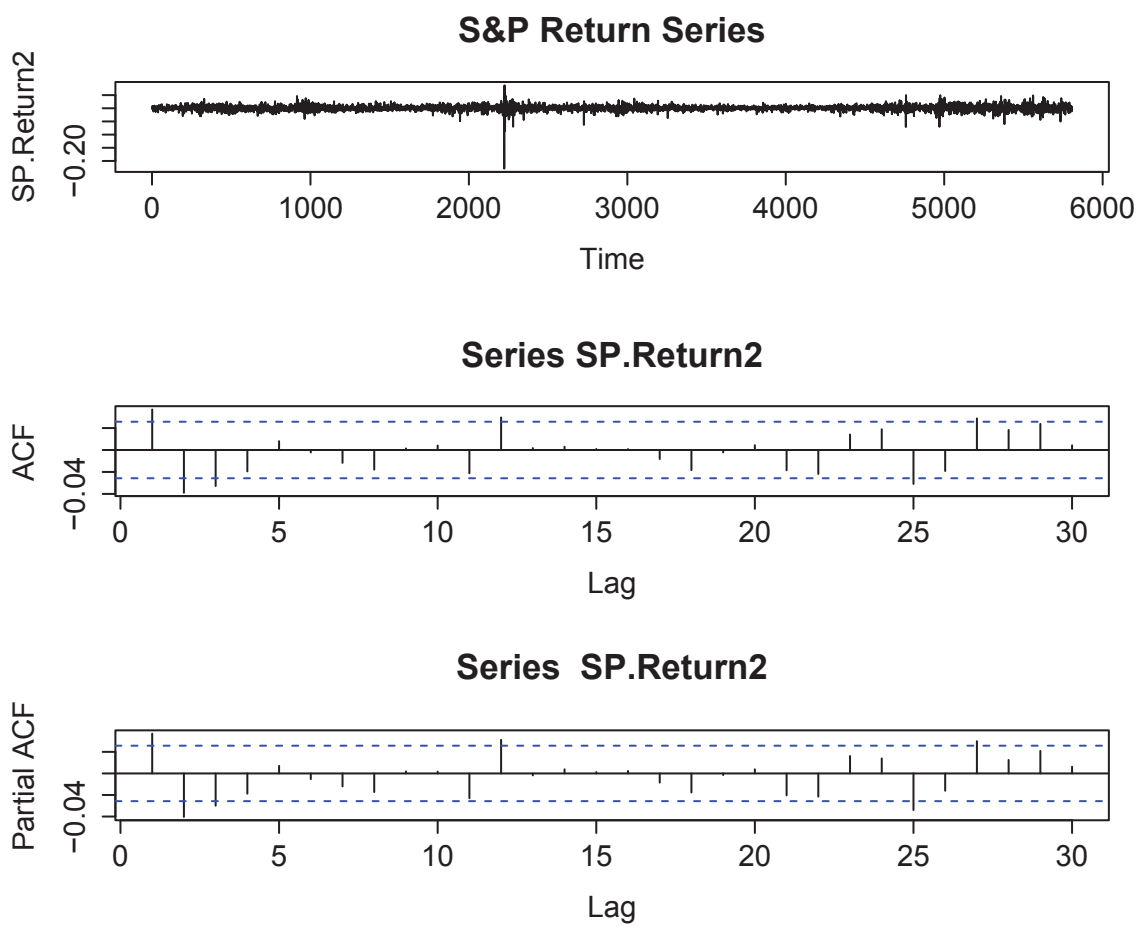
The Box-Pierce statistics on squared residuals reduces the sensitivity for the statistics on the parameters estimated; therefore, the power for this statistics on ARMA model will be higher. The Wild Monte Carlo Test reduces the overinflated observed statistics by the kernel function but remains the simulate statistics at the original level. Thereby, it achieves the improvement in the goodness-of-fit power.

Another point needs our attention is the effect of the lag m. In most of cases, increasing the lag m will not affect the performance of the tests. We would expect a more stable test especially when we intend to use the general variance portmanteau test on the Wild Monte Carlo Test.

3.6 Illustrative Application

We consider that the use of the newly developed portmanteau test with the Wild Monte Carlo Test to evaluate the null hypothesis of ARMA − GARCH Model. We use the daily return of $S\&P$ 500 Index for January 3, 1979 - December 31, 2001 as in Francq et al. (2005). The length of the series $n = 10^3$. Initially, based on the standard economic series, the return of such stock price should be martingale differenced (e.g. semi strong time series model) rather than generally independent time series model. We show the data in the Figure 3.1:

Figure 3.1: Daily Return Series for $S\&P$ 500 Index from January 3, 1979-December 31, 2001 and Autocorrelation, Partial Autocorrelation Function



From the figure 3.1, we can see the plot exhibits the features of differenced martingale. Then we compare the p-value of the newly developed portmanteau test with Wild Monte Carlo Test with the Ljung-Box test.

In order to evaluate the randomness of the time series, we use the algorithm proposed by Lin (2006). For simplicity, we restate the algorithm here:

Algorithm 3.3. *Assume we have the series of interest* $\{X_t\}, t = 1, 2, \ldots, n$,

Step One: *Calculate the sample autocorrelation function of the first m lags and use them to calculate the portmanteau test* D_m^o

Step Two: *Re-sample n observations with replacement from the original data set.*

Step Three: *Repeat Step Two for N times and then calculate the sample autocorrelation function each time and use them to calculate the corresponding portmanteau test* $D_{m,j}$, $j = 1, 2, \ldots, N$.

Step Four: *Then the p-value could be calculated based on the following formula:*

$$p\text{-}value = \frac{\#\left\{D_{m,j} > D_m^o\right\} + 1}{N + 1}. \tag{3.50}$$

Remark: *Step Two could be modified to incorporate the Wild Monte Carlo Test into the randomness test*

Step Two: *Re-sample n observations with replacement from the original data set as* X_t^*. *Generate the dependent random variable* $W_t, t = 1, 2, \cdots, n$ *and then generate the pseudo-sample:*

$$\tilde{X}_t^* = \bar{X}_t^* + (X_t^* - \bar{X}_t^*) \times W_t. \tag{3.51}$$

Where $\bar{X}_t^*$ *denote the sample mean of* X_t^*.

We could apply Step Three on the pseudo-sample $\tilde{X}_t^*$.

Table 3.4: P Value Comparison on Evaluating GARCH type errors for return of $S\&P$ 500 Index for Randomness. Series Length $n = 5807$. Monte Carlo Iteration $N = 10^3$. MC represent the Monte Carlo Randomness test in Lin (2006) and WMC represent the newly developed Wild Monte Carlo Test for randomness test

Lag	1	2	3	4	5	8	10	15	20
LB Stats	7.90	16.71	22.96	25.17	25.54	28.23	28.34	36.14	38.63
PValue(LB)	0.0049	0.0002	0.00005	0.00005	0.00001	0.0004	0.0002	0.0002	0.0007
PValue(MC)	0.007	0.003	0.002	0.001	0.001	0.001	0.002	0.001	0.002
PValue(WMC)	0.7033	0.6813	0.2937	0.1378	0.164	0.113	0.107	0.117	0.12

From the Table 3.4, we can see the tradition Lin-McLeod Test would behave similar to the Ljung-Box Test. The Wild Monte Carl Test would be more appropriate to check the weak white noise assumption here. Though in long term cases, the p-value would decrease, since too much information is included and the weak white noise would be partly undermined. Compared to Francq et al. (2005), though the test performance would be worse than their test if the lag m chooses large number. The Wild Monte Carlo Test would be more practical especially for short length time series data.

According to the Ljung-Box Test and the Lin-McLeod Test, we can see the strong white noise assumption gets rejected while the weak white noise fails to be rejected according to the Wild Monte Carlo test.

Then we are fit an ARMA $(1,1)$ Test to the squares of the $S\&P$ returns. According to the parameter estimation, We get an ARMA $(1,1)$ Model as $Z_t - 0.8313Z_{t-1} = a_t + 0.7269a_{t-1}$. Then we apply the statistics on the residuals here to evaluate the weak white noise assumptions.

Table 3.5: p-value Comparison on Evaluating the GARCH type error for Return of $S\&P$ 500 Index. Series Length $n = 5807$. Monte Carlo Iteration $N = 10^3$. MC represent the tradition Monte Carlo Randomness test in Lin (2006) and WMC represent the newly developed Wild Monte Carlo Test for randomness test

Lag	1	2	3	4	5	8	10	15	20
LB Stats	2.26	15.53	16.86	37.61	83.37	97.77	95.93	96.69	98.28
PValue(LB)	0.132	0.0004	0.0008	0.00005	0.0000	0.000	0.000	0.000	0.000
PValue(LM)	0.042	0.013	0.001	0.005	0.002	0.0001	0.0001	0.004	0.008
PValue(WMC)	0.782	0.449	0.2797	0.181	0.121	0.129	0.117	0.107	0.108

From the table 3.5, we can reject the strong white noise assumptions based on the Ljung-Box and Lin-McLeod Test here. But we fail to reject the weak white noise assumption for Wild Monte Carlo. The underestimation for the p-value of the Lin-McLeod Test can still be attributing to the nuisance parameters in the residuals due to the weak white noise assumptions.

We are attempting to evaluate the weak white noise by fitting GARCH Model. We first fit GARCH $(1,1)$ Model to the squares of $S\&P$ 500 returns. And then we apply the portmanteau test on the residuals of the GARCH $(1,1)$ Model.

Table 3.6: p-value Comparison on Evaluating the Residuals of GARCH $(1,1)$ Model for Squares of $S\&P$ 500 Index Returns. Series Length $n = 5807$. Monte Carlo Iteration $N = 10^3$. MC represent the tradition Monte Carlo Randomness test in Lin (2006) and WMC represent the newly developed Wild Monte Carlo Test for randomness test

Lag	1	2	3	4	5	8	10	15	20
LB Stats	4.87	6.29	7.28	7.34	7.39	9.85	12.34	24.43	36.79
PValue(LB)	0.027	0.043	0.063	0.118	0.193	0.2755	0.263	0.06	0.012
PValue(LM)	0.028	0.03	0.041	0.049	0.063	0.115	0.122	0.08	0.067
PValue(WMC)	0.067	0.026	0.022	0.024	0.029	0.041	0.041	0.073	0.0949

After fitting the GARCH $(1,1)$ Model, from Table 3.6, we could see based on the Ljung-Box Test, we fail to reject the strong white noise for most lags. The Wild Monte Carlo Test shows we should reject the assumption of weak white noise. Still, the Lin-McLeod Test would suffer test bias due to the nuisance parameters of the asymptotic distribution.

3.7 Conclusion

In this chapter, we considered the goodness-of-fit test in $\mathrm{ARMA}-\mathrm{GARCH}$ model. To weaken the assumption of strong white noise, we introduce the GARCH type errors in the ARMA model and extended the conclusion of Mahdi and McLeod (2011) to the $\mathrm{ARMA}-\mathrm{GARCH}$ model and derived the asymptotic distribution for the general variance Portmanteau test. The asymptotic distribution for the Portmanteau test is still a sum of weighted χ^2 distributions.

Due to the complicated structure of the asymptotic distribution, the gamma approximation adopted by Peña and Rodríguez (2002) and Mahdi and McLeod (2011) is not a convenient way to calculate the p-value. Therefore, we choose the Lin-McLeod Test (Lin and McLeod, 2006) and provide accuracy estimation of the p-value for the test. However, due to the null hypothesis of $\mathrm{ARMA}-\mathrm{GARCH}$ model, when we fit the ARMA model for the series, the nuisance parameters in the GARCH type error would distort the power and the size of the Lin-McLeod Test for the Portmanteau test. We considered using the Wild Monte Carlo test to get rid of the effect of nuisance parameters here, as we had done the previous chapter for the ARFIMA model.

After we applying the Wild Monte Carlo Test on the $\mathrm{ARMA}-\mathrm{GARCH}$, the power of model mis-identification is significantly improved, while the size still remains at a stable level. Due to the wide existence for GARCH type series in the economic and financial data, we would use this proposed method as a more accurate detection for the existing GARCH type of errors.

Chapter 4

Trimmed Portmanteau Test for ARMA Model with Infinite Variance

4.1 Introduction

Several new portmanteau tests have been developed recently. Under the assumption of infinite variance, Adler et al. (1998) claim that the Box-Jenkins modelling process could be applied to the ARMA model with infinite variance with careful modifications. The process X_t can be a strictly stationary process defined by:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j a_{t-j}, t = 1, 2, \ldots, n, \tag{4.1}$$

where parameter ψ_i satisfy the following condition:

$$\sum_{j=-\infty}^{\infty} |j| \left|\psi_j\right|^{\delta} < \infty, \delta \in (0, \alpha) \cap [0, 1], \tag{4.2}$$

and a_t, is defined as the stable distribution with its characteristic function in the following type:

$$E(e^{ita}) = \begin{cases} \exp\left\{-\sigma |t|^{\alpha} \left(1 - i\beta \times \mathrm{Sign}\,(t) \tan(\pi\alpha/2)\right) + i\mu t\right\} & \alpha \neq 1, \\ \exp\left\{-\sigma |t| \left(1 - i\beta \times \mathrm{Sign}\,(t) \log |t|\right) + i\mu t\right\} & \alpha = 1. \end{cases} \tag{4.3}$$

and $i^2 = -1$, $\alpha \in (0, 2]$ is the stability index, $\beta \in [-1, 1]$ is the skewness parameter, $\mu \in \mathbf{R}$ define the location and

$$\mathrm{Sign}\,(t) = \begin{cases} 1 & t > 0, \\ 0 & t = 0, \\ -1 & t < 0. \end{cases}$$

Under this model, the stable analogue of autocorrelation function at lag k is defined as:

$$\rho_k = \frac{\sum_j \psi_j \psi_{j+k}}{\sum_j \psi_j^2}, \; k = 1, 2, \ldots \tag{4.4}$$

The autocorrelation function ρ_k can be estimated by the sample autocorrelation as,

$$r_k = \frac{\sum_{t=1}^{n-k} X_t X_{t+k}}{\sum_{t=1}^{n} X_t^2}, \ k = 1, 2, \ldots \tag{4.5}$$

For $\alpha > 1$, the mean-correct sample autocorrelation $\hat{r}_k$ at lag k can be defined as,

$$\hat{r}_k = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}, \tag{4.6}$$

where $\bar{X}$ is the sample average of the series $_t$. Based on Davis and Resnick (1986), for any lag k, the limiting distribution for the sample autocorrelation function is,

$$\left[\frac{n}{\log n}\right]^{1/\alpha} (r_1 - \rho_1, \cdots, r_k - \rho_k)^T \to_d (Y_1, \cdots, Y_k)^T, \tag{4.7}$$

where Y_h is,

$$Y_h = \sum_{j=1}^{\infty} (\rho_{k+j} + \rho_{k-j} - 2\rho_j \rho_k) \frac{S_j}{S_0}, \tag{4.8}$$

and $S_0 \sim Z_{\alpha/2}(C_{\alpha/2}, 1, 0)$, $S_j \sim Z_\alpha(C_\alpha^{-1/\alpha}, 0, 0)$.

$$C_\alpha = \begin{cases} (1-\alpha)/\left(\Gamma(2-\alpha)\cos(\pi\alpha/2)\right) & \alpha \neq 0, \\ 2/\pi & \alpha = 0. \end{cases}$$

From Davis and Resnick (1986), we know the limiting distribution of $\hat{r}_k$ is the same as that of r_k. Under the null hypothesis of X_t are a sequence of i.i.d stable Paretian random variable, then $\rho_0 = 1, \rho_k = 0$ which can simplify the limiting distribution for sample autocorrelation function:

$$\left[\frac{n}{\log n}\right]^{1/\alpha} (\hat{r}_1, \cdots, \hat{r}_k)^T \to_d (W_1, \cdots, W_k)^T \tag{4.9}$$

where W_j, $j = 1, 2, \ldots, m$ are defined by $W_j = S_j / S_0$.

Based on this conclusion, Runde (1997) developed the asymptotic distribution of the Box-Pierce Statistics. His result shows:

$$Q_{BP} = \left(\frac{n}{\log(n)}\right)^{2/\alpha} \sum_{j=1}^{m} \hat{r}_j^2 \to_d \sum_{j=1}^{m} W_j^2, \tag{4.10}$$

Later, Lin and McLeod (2008) study both Box-Pierce and Peňa-Rodríguez test under the ARMA model with infinite variance and proposed a better modification based on,

$$\hat{D}_m = \left(\frac{n}{\log n}\right)^{2/\alpha} (1 - |\mathcal{R}_m|)^{1/m}, \tag{4.11}$$

where $\mathcal{R}_m$ is the autocorrelation matrix of X_t, $t = 1, 2, \ldots, n$. The asymptotic distribution of the new portmanteau statistics is:

$$\hat{D}_m \rightarrow_d \sum_{j=1}^{m} \frac{m+1-i}{m} W_j^2, \tag{4.12}$$

where Wj is as defined in 4.9. Still, according to their simulation, the Lin-McLeod Test is recommended based on the better simulation performance.

Another type of modification on portmanteau test for ARMA models with infinite variance is proposed by Lee and Ng (2010). The process can be defined as equation 4.1 and other parameters have the same meaning as defined before. They proposed to replace the original autocorrelation function $\hat{r}_k$ by the autocorrelation of the trimmed time series (e.g. trimmed autocorrelation $\tilde{r}_k$). In order to calculate the trimmed autocorrelation, first we defined the trimmed series:

$$\tilde{X}_t = \begin{cases} X_t & M^L < X_t < M^U, \\ 0 & \text{Otherwise}. \end{cases} \tag{4.13}$$

Then $\tilde{r}_k$ is defined as the sample autocorrelation function of the trimmed time series.

M^L and M^U are the $n\lambda^U$ and $n\lambda^L$-th order statistics, $0 < \lambda^L < \lambda^U < 1$, where λ^L is the lower trimming portion and $(1 - \lambda^U)$ is the upper trimming proportion.

$$\tilde{r}_k = \frac{\sum_{t=1}^{n-k} \tilde{X}_t \tilde{X}_{t+k}}{\sum_{t=1}^{n} \tilde{X}_t^2}, k = 1, 2, \ldots \tag{4.14}$$

For simplicity, we restate the key result in Lee and Ng (2010):

Theorem 4.1. *For any positive integer m, the asymptotic distribution of* $(\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_m)$ *is given by:*

$$\sqrt{n}(\tilde{r}_1, \ldots, \tilde{r}_m) \rightarrow_d N(0, I_m), \tag{4.15}$$

where I_m *is the m-dimensional identity matrix.*

From the theorem 4.1, they derived the asymptotical distribution of Box-Pierce Test:

$$\tilde{Q}_{BP} = n(n+2) \sum_{k=1}^{m} \frac{\tilde{r}_k^2}{n-k} \rightarrow_d \chi_m^2.$$

They claim that though the trimming autocorrelation function $\tilde{r}_k$ may lose part of the information in the series, this problem can be overcome by appropriately choosing the trimming threshold. This idea actually simplifies the distribution of the autocorrelation function $\hat{r}_k$ in the ARMA model with infinite variance to the case with finite variance. Further, Cui and Wu (2014) derived the asymptotic distribution of the trimmed partial autocorrelation function:

Corollary 4.1. *For any positive integer m, the asymptotic distribution of* $(\tilde{\pi}_1, \tilde{\pi}_2, \ldots, \tilde{\pi}_m)$ *is given by:*

$$\sqrt{n}(\tilde{\pi}_1, \ldots, \tilde{\pi}_m) \to_d N(0, I_m). \tag{4.16}$$

Therefore, the Monti-type of the portmanteau test can be defined in the trimming partial autocorrelation function:

$$\tilde{M} = n(n+2) \sum_{k=1}^{m} \frac{\tilde{\pi}_k^2}{n-k} \to_d \chi_m^2. \tag{4.17}$$

Therefore, we can use the idea to improve the performance of the portmanteau test.

4.2 Improved Portmanteau Test for ARMA Model with Infinite Variance

Now, we intend to incorporate the idea of trimmed autocorrelation into the general variance portmanteau test (Peña and Rodríguez, 2002):

$$\hat{D}_m = n\left(1 - |\mathcal{R}_m|^{1/m}\right), \tag{4.18}$$

where $\mathcal{R}_m$ is the autocorrelation matrix of the series X_t, $t = 1, 2, \ldots, n$.

The asymptotic distribution of the portmanteau test is derived in Peña and Rodríguez (2002) and its application on the ARMA model with infinite variance has been developed in Lin and McLeod (2008). Now we are trying to investigate its trimmed effect on the portmanteau test. By convention, we defined the trimmed autocorrelation function ($\tilde{r}_k$) and partial autocorrelation function $\tilde{\pi}_k$ (Lee and Ng, 2010; Cui and Wu, 2014). We denote the original autocorrelation function and partial autocorrelation function as $\hat{r}_k$ and $\hat{\pi}_k$.

Definition 4.1. *Based on the trimmed autocorrelation function,we define the trimmed version of general variance Test as:*

$$\tilde{D}_m^* = n\left(1 - \left|\tilde{\mathcal{R}}_m\right|^{1/m}\right), \tag{4.19}$$

where $\tilde{\mathcal{R}}_m$ *is defined as the matrix of trimmed autocorrelation function,*

$$\left|\tilde{\mathcal{R}}_m\right| = \begin{vmatrix} 1 & \tilde{r}_1 & \cdots & \tilde{r}_{m-1} \\ \tilde{r}_1 & 1 & \cdots & \tilde{r}_{m-2} \\ \vdots & \ddots & \ddots & \vdots \\ \tilde{r}_{m-1} & \cdots & \tilde{r}_1 & 1 \end{vmatrix}, \tag{4.20}$$

where $\tilde{r}_t$, $t = 1, 2, \ldots, m-1$ *are the trimmed autocorrelation function.*

Then we could have the following theorem:

Theorem 4.2. *For a series X_t, $t = 1, 2, .., n$ which follows a stable distribution with infinite variance, then we have:*

$$\tilde{D}^*_m \rightarrow_d \sum_{j=1} \frac{m+1-j}{m} \chi^2_{1,j}. \tag{4.21}$$

Proof: Generally the proof process would follow the procedure of Theorem 1 in Lin and McLeod (2008) and Peňa and Rodríguez (2002). Since we have the distribution of the trimmed autocorrelation, $\tilde{r}_k$,

$$\sqrt{n}(\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_m) \rightarrow_d N(0, I_m).$$

then we have the asymptotic distribution here.

A similar adjustment can be made to Fisher-Gallagher Portmanteau Test (Fisher and Gallagher, 2012a). As its defined:

$$Q_W = n(n+2) \sum_{k=1}^{m} \frac{m-k+1}{m} \frac{\hat{r}_k^2}{n-k}$$

and

$$M_W = n(n+2) \sum_{k=1}^{m} \frac{m-k+1}{m} \frac{\hat{\pi}_k^2}{n-k}$$

by replacing the standardized autocorrelation and partial autocorrelation $(n+2/n-k)\hat{r}_k$ and $(n+2/n-k)\hat{\pi}_k$ with the trimmed counterpart $\tilde{r}_k$ and $\tilde{\pi}_k$. Firstly, following the ideas in Lin and McLeod (2008), we can get the asymptotic distribution for the exact autocorrelation function under slight modification to the origin Q_W test.

Theorem 4.3. *In* ARMA *model with infinite variance errors, we define the Fisher-Gallagher statistics as follows:*

$$\hat{Q}_W = \left(\frac{n}{\log(n)}\right)^{2/\alpha} \sum_{k=1}^{m} \frac{m-k+1}{m} \hat{r}_k^2. \tag{4.22}$$

We have the asymptotic distribution under the stable distribution:

$$\hat{Q}_W \rightarrow_d \sum_{k=1}^{m} \frac{m-k+1}{m} W_k^2, \tag{4.23}$$

where W_k , $i = 1, 2, \ldots, m$ are as defined in 4.9.

Interestingly, we could see the asymptotic distribution for the Fisher-Gallagher Test is exactly the same as the one in Lin and McLeod (2008) and the performance equivalence is evaluated later by simulations. Actually, Fisher and Gallagher (2012a) show the general variance portmanteau test (Mahdi and McLeod, 2011) is asymptotic equivalent to Fisher-Gallagher Portmanteau test (Fisher and Gallagher, 2012a) . However the derivation would be different.

Proof:

First we rewrite the modified Fisher-Gallagher Statistics in its quadratic form:

$$\hat{Q}_W = \left(\frac{n}{\log n}\right)^{2/\alpha} \hat{\mathbf{r}}^T \mathbf{K} \hat{\mathbf{r}},$$

where $\hat{\mathbf{r}} = (\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_m)$ and

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \frac{m-1}{m} & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{m} \end{bmatrix}.$$

Then under the null hypothesis of independent and identical random variable with a stable Paretian distribution, the limiting distribution of the sample ACF is:

$$\left(\frac{n}{\log n}\right)^{1/\alpha} \mathbf{r}^T \to_d \mathbf{W}^T, \tag{4.24}$$

where $\mathbf{W} = (W_1, \ldots, W_m)$ and $W_j = S_j/S_0$, $j = 1, 2, \ldots, m$. According to the distribution of the quadratic form, we could get the asymptotic distribution

$$\hat{Q}_W \to_d \sum_{j=1}^{m} \frac{m-j+1}{m} W_j^2. \tag{4.25}$$

We can apply the ideas of trimmed modification on the modified Fisher-Gallagher Portmanteau Test and derive its asymptotic distribution.

Corollary 4.2. *Let $\tilde{r}_k$ and $\tilde{\pi}_k$, $k = 1, 2, \ldots, m$ be the trimmed autocorrelation function and partial autocorrelation function for the series. Then we could have the modified portmanteau statistics:*

$$\tilde{Q}_W = n \sum_{k=1}^{m} \frac{m-k+1}{m} \tilde{r}_k^2, \tag{4.26}$$

and

$$\tilde{M}_W = n \sum_{k=1}^{m} \frac{m-k+1}{m} \tilde{\pi}_k^2. \tag{4.27}$$

with the asymptotic distribution

$$\tilde{Q}_W \to_d \sum_{k=1}^{m} \frac{m-k+1}{m} \chi_{1,k}^2, \tag{4.28}$$

$$\tilde{M}_W \to_d \sum_{k=1}^{m} \frac{m-k+1}{m} \chi_{1,k}^2. \tag{4.29}$$

The proof of this corollary is basically the same as the previous theorem except we would consider the limit distribution of the trimmed autocorrelation function. According to Lee and Ng (2010), we could know that: $\sqrt{n}(\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_m) \to_d N(0, I_m)$ Then obviously we could get the asymptotic distribution above.

4.3 Extension to the Diagnostic Check for Model Adequacy of ARMA Model with Infinite Variance

In order to explore this question, according to Lin and McLeod (2008), we consider the case of AR(p) model with infinite variance and extend the results to the ARMA case based on the statements in Lin (2006). We specify the model as followed:

Definition 4.2. *Define the general* AR(p) *model as followed:*

$$\phi(B)X_t = a_t, \tag{4.30}$$

where a_t, $t = 1, 2, \ldots, n$ are the i.i.d random variables and follow the stable Paretian distribution. As in convention, $\phi(B)$ is the polynomial function with backward operator B, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$. Then the residuals can be defined as follows:

$$\hat{a}_t = (1 - \hat{\phi}_1 B - \hat{\phi}_2 B^2 - \ldots - \hat{\phi}_p B^p)X_t = \hat{\phi}(B)X_t, \tag{4.31}$$

and $\hat{\phi}(B) = (\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p)$ are the estimation of the $\phi(B)$.

Under this definition, we define the autocorrelation function of the $\hat{a}_t$:

$$\hat{r}_k = \frac{\sum \hat{a}_t \hat{a}_{t-k}}{\sum \hat{a}_t^2}, \tag{4.32}$$

under the assumption of Paretian stable distribution, we have the estimator $\hat{\phi}_j$, $j = 1, 2, \ldots, p$ as follows:

$$\hat{\phi}_j = \phi_j + O_p\left(\left[\frac{n}{\log n}\right]^{-2/\alpha}\right), \tag{4.33}$$

and we have

$$\hat{r}_k = r_k + \sum_{j=1}^{p}(\hat{\phi}_j - \phi_j)\psi_{k-j} + O_p\left(\left[\frac{n}{\log n}\right]^{-2/\alpha}\right), \tag{4.34}$$

where ψ_j is the j-th term of the impulse response coefficients at lag i

$$r_k = \frac{\sum a_t a_{t-k}}{\sum a_t^2},$$

Under the null hypothesis of the portmanteau test, $r_1 = r_2 = \ldots = r_k = 0$. Then we can write the equation 4.34 in its matrix form:

$$\hat{\mathbf{r}} = \mathbf{r} + \mathbf{\Psi}(\phi - \hat{\phi}), \tag{4.35}$$

where

$$\mathbf{\Psi} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \psi_1 & 1 & \cdots & 0 \\ \psi_2 & \psi_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{m-1} & \psi_{m-2} & \cdots & \psi_{m-p} \end{bmatrix}.$$

Then based on above assumption and equation 4.35,we could get the asymptotic distribution for modified Fisher-Gallagher Portmanteau Test for AR(p) Model with infinite variance.

Theorem 4.4. *Under the* AR(p) *model with infinite variance, assume* $\hat{a}_t$ *is the estimated residuals after estimation, then we have:*

$$\tilde{Q}_W = \left(\frac{n}{\log(n)}\right)^{2/\alpha} \sum_{k=1}^{m} \frac{m-k+1}{m} \hat{r}_k^2 \to_d \mathbf{W}_m^T \mathbf{U}_m \mathbf{W}_m, \tag{4.36}$$

where $\mathbf{W}_m = (W_1, W_2, \ldots, W_M)^T$, $\mathbf{U}_m = (\mathbf{I}_m - \mathbf{Q})^T \mathbf{K}_m (\mathbf{I}_m - \mathbf{Q})$, $\mathbf{Q}_m = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, $\mathbf{I}_m$ *is the identity matrix, and*

$$\mathbf{K}_m = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \frac{m-1}{m} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{m} \end{pmatrix}.$$

Proof:

First, we write the quadratic form of the modified Fisher-Gallagher Portmanteau Test:

$$\tilde{Q}_W = \left(\frac{n}{\log(n)}\right)^{2/\alpha} \hat{\mathbf{r}}^T \mathbf{K}_m \hat{\mathbf{r}}. \tag{4.37}$$

According to equation 4.33, we can have:

$$\hat{\phi}_j - \phi_j = \left(\frac{n}{\log n}\right)^{-1/\alpha}. \tag{4.38}$$

According to Box and Pierce (1970), the estimated residuals $\hat{Z}_t$ satisfy the orthogonal condition and to the order of $O_p\left(1/\sqrt{n}\{n/\log n\}^{1/\alpha}\right)$, we have:

$$\hat{\mathbf{r}}^T \mathbf{X} = 0. \tag{4.39}$$

Multiple both side of 4.35 by $\mathbf{Q}_m$ given by:

$$\mathbf{Q}_m = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T, \tag{4.40}$$

we get the relationship between autocorrelation $\mathbf{r}$ and sample autocorrelation $\hat{\mathbf{r}}$ as

$$\hat{\mathbf{r}} = (\mathbf{I}_m - \mathbf{Q}_m)\mathbf{r}. \tag{4.41}$$

Then according to equation 8 in Lin and McLeod (2008):

$$\left[\frac{n}{\log n}\right]^{1/\alpha} \mathbf{r}^T \to_d \mathbf{W}_m^T, \tag{4.42}$$

where $\mathbf{W}_m = (W_1, W_2, \ldots, W_M)^T$. We have the quadratic form of Fisher-Gallagher Portmanteau test convergence:

$$\hat{Q}_W = \left(\frac{n}{\log(n)}\right)^{2/\alpha} \hat{\mathbf{r}}^T \mathbf{K}_m \hat{\mathbf{r}} \to_d \mathbf{W}_m^T \mathbf{U}_m \mathbf{W}_m, \tag{4.43}$$

where $\mathbf{U}_m$ is defined as $(\mathbf{I}_m - \mathbf{Q}_m)\mathbf{K}_m$.

According to Lee and Ng (2010), if the trimming proportion λ is close to zero, then we have $\tilde{r}_k \to \hat{r}$. Therefore, the above theorem can be extended:

Corollary 4.3. *If λ is close to 0 then we could have the trimmed portmanteau test on the residuals of* AR(p) *with infinite variance convergence:*

$$\tilde{Q}_W = n\tilde{\mathbf{r}}^T\mathbf{K}_m\tilde{\mathbf{r}} \to_d \sum_{j=1}^{m} \lambda_i \chi_1^2, \tag{4.44}$$

where λ_j, $j = 1, 2, \ldots, m$ is the eigenvalues of the matrix of $(\mathbf{I}_m - \mathbf{Q}_m)\mathbf{K}_m$.

Remark 1: According to Lin (2006) and Lin and McLeod (2008), the residuals in the fitted ARMA model would be asymptotically equivalent to a certain AR(p) model under the general linear process form. Then the behavior of the residuals and trimmed residuals in AR(p) could be generalized to ARMA model with infinite variance.

Remark 2: According to Cui and Wu (2014), the trimmed partial autocorrelation function will hold the same convergence as its trimmed autocorrelation function, hence one can generalized this result to the weighted Monti-type Portmanteau Test (Fisher and Gallagher, 2012a) which would have the similar convergence results.

4.4 Simulation Study on ARMA with Infinite Variance

We investigate the power of new portmanteau test in Fisher and Gallagher (2012a) and apply on the ARMA Model with infinite variance. The powers of the tests are estimated by the Monte Carlo Algorithm in Lin and McLeod (2006) (Lin-McLeod Test). The new portmanteau tests are listed as follows:

$$Q_W = n(n+2)\sum_{k=1}^{m} \frac{m-k+1}{m}\frac{\hat{r}_k^2}{n-k}, \tag{4.45}$$

and

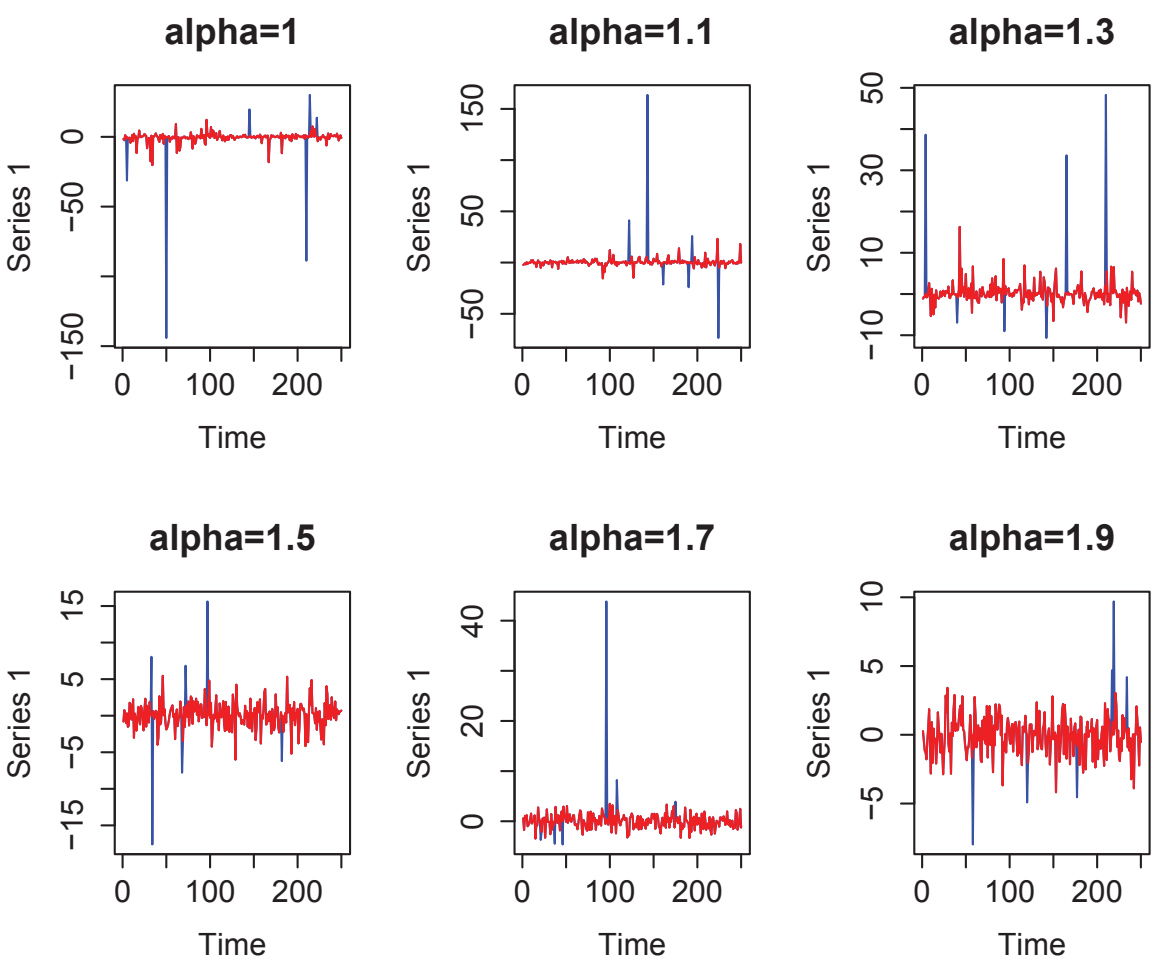
$$M_W = n(n+2)\sum_{k=1}^{m} \frac{m-k+1}{m}\frac{\hat{\pi}_k^2}{n-k}, \tag{4.46}$$

The asymptotic distribution is the same as defined in Peňa and Rodríguez (2002) as the weighted χ^2 distribution. First, we investigate the empirical size of Lin-McLeod Test for randomness of stable white noise with index α. We set the nominal size as 5% to compare. The simulations are presented in Lin and McLeod (2008) for improved portmanteau test in Peňa and Rodríguez (2002) and Box-Pierce Test (Box and Pierce, 1970).

In the figure 4.1, we simulate the series length of $n = 250$ and trim its upper and lower 1% of data. In this case, we assume $\beta = 0$, $\gamma = 1$. Figure 4.1 shows the comparison between the original data and the trimmed data. From Figure 4.1, we can see clearly that after trimming, the data has already been regulated and the behavior would be more like the finite variance data.

The blue line shows the original series with stable white noise innovation while the red line shows the trimmed series. The below plot show the comparison between the series following stable distribution and the series after trimming. Here, we choose the trimming rate as 0.01.

Figure 4.1: Comparison between Time Series with Paretian errors and its trimmed time series. Choosing series length $n = 250$ and trimming level $\lambda = 5\%$.



.

4.4.1 Randomness Evaluation

In order to evaluate and compare the empirical size of the trimmed Fisher-Gallagher Portmanteau test, we employ the Lin-McLeod Test here. First, in Table 4.1, we compare the empirical size between the original Fisher-Gallagher Test and the trimmed Fisher-Gallagher Test on Stable White Noise Innovation with index α.

Table 4.1: Size Comparison between Fisher-Gallagher Test (both Ljung-Box type and Monti Type) and its Application on the Trimmed Time Series. Choosing Stable Distribution Parameters $\alpha = 1.1, 1.3, 1.5, 1.7, 1.9, \beta = 1, \gamma = 0$. Trimmed Level $\lambda = 5\%$, series length $m = 5, 10, 15$.

	Q_W			M_W			$\tilde{Q}_W$			$\tilde{M}_W$		
α/m	5	10	15	5	10	15	5	10	15	5	10	15
1.9	**6.60**	5.10	5.30	**6.20**	5.40	5.50	4.90	5.00	5.70	5.50	5.00	5.20
1.7	4.80	5.20	**3.70**	5.20	4.70	**3.30**	5.10	5.50	5.50	4.70	5.00	5.40
1.5	4.90	5.90	**6.10**	5.10	5.70	**6.10**	5.60	5.70	4.90	5.10	5.30	4.80
1.3	4.50	5.50	**3.10**	4.40	5.30	**3.70**	4.20	5.50	4.70	4.40	5.30	4.80
1.1	4.80	5.00	4.80	4.90	5.20	4.70	5.20	4.20	4.20	5.30	4.30	4.60

The empirical size listed above was calculated based on $N = 10^3$ simulations. Each Monte Carlo p-value was estimated based on $B = 10^3$ simulation. Series length $n = 250$ and lags $m = 5, 10$ and 15 were investigated. In the Fisher-Gallagher Portmanteau Test table above, the bolded numbers mean significantly different from the nominal level 5% (Smaller than 4% and larger than 6%).

Remark 3: By comparing the Fisher-Gallagher Portmanteau Test on stable white noise innovation, we can see the trimmed portmanteau test would be more stable and accurate estimating the type one error especially for larger α value. Less exotic numbers appear after we invite the idea of trimming.

Secondly, we consider the comparisons between Fisher-Gallagher Test, general variance Portmanteau test and Box-Pierce Test with their trimmed tests. Still we consider the size with the change of α and lag number m. We choose the $\alpha = 1.1, 1.3, 1.5, 1.7, 1.9$, $m = 5, 10, 15$, series length is $n = 250$, Monte Carlo repetition time 10^3 and bootstrap iteration $N = 10^4$. Since the environment set up is the same as the Table 1 in Lin and McLeod (2008), then can compare to their results.

Table 4.2: Size Comparison between Monte Carlo Test $\hat{D}_m$, Box-Pierce Test Q_{BP}, Fisher Gallagher Ljung-Box Type Test Q_W and Fisher Gallagher Monti Type Test M_W. Choosing Stable Distribution Parameters as $\alpha = 1.1, 1.3, 1.5, 1.7, 1.9, \beta = 1, \gamma = 0$, Trimmed Level as $\lambda = 5\%$ and lag $m = 5, 10, 15$

Statistics	m	$\hat{D}_m$	Q_{BP}	Q_W	M_W	$\tilde{Q}_W$	$\tilde{M}_W$
	5	5.30	4.96	5.28	5.26	**4.99**	5.18
$\alpha = 1.9$	10	4.66	4.71	4.66	4.77	5.13	**5.10**
	15	4.78	4.84	**5.06**	5.12	4.78	4.86
	5	5.18	4.82	4.42	4.36	**5.06**	5.18
$\alpha = 1.7$	10	4.44	4.43	4.71	4.77	**4.89**	4.86
	15	4.44	4.41	4.68	4.50	**4.94**	5.05
	5	4.82	**5.07**	4.75	4.60	4.85	4.81
$\alpha = 1.5$	10	**4.99**	5.27	4.91	5.10	5.03	4.97
	15	5.13	5.30	5.23	5.11	**4.96**	5.24
	5	4.80	5.04	5.05	**5.01**	5.09	4.93
$\alpha = 1.3$	10	5.03	**5.00**	4.91	4.95	4.78	4.81
	15	5.18	5.27	4.72	4.79	**5.14**	5.20
	5	5.26	5.33	4.84	4.86	**4.95**	5.07
$\alpha = 1.1$	10	5.33	5.25	**5.03**	4.94	5.17	5.14
	15	5.12	5.15	4.86	4.84	**4.98**	5.25

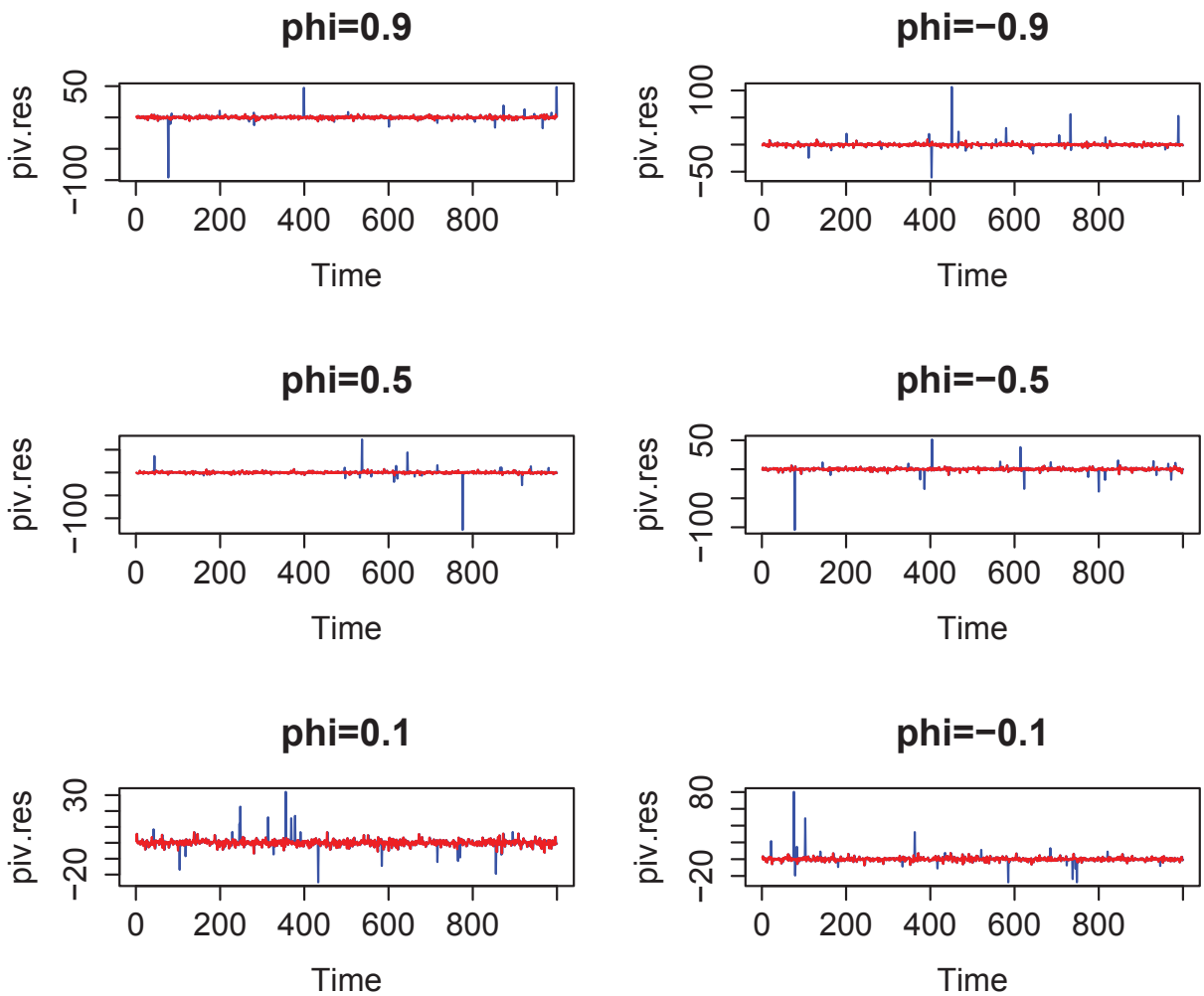
Remark 3: For comparison simplicity, we bolded the number with the smallest distance to the nominal rate 5%. As we can see, generally the trimmed Fisher-Gallagher portmanteau Test would dominate the test especially when α is close to 2. In a few cases where others tests perform better, the trimmed Fisher-Gallagher Portmanteau would perform close to the best performed tests. Therefore, we conclude that the trimmed tests could significantly increase the performance of the test in terms of size.

4.4.2 Model Adequacy Checking

Now we are comparing the various portmanteau test with their trimmed counterpart and see their performance on the model adequacy checking. The table 4.3 is simulated based on the Table III in Lin and McLeod (2008). We set the simulation based on AR(1) model and infinite variance ($\alpha = 1.5$). And we fit the model with burg estimation since it can always estimate the model in the stationary condition as the requirement of the Lin-McLeod Test. We fit the model with AR(1) model with $\phi_1 = \pm 1, \pm 3, \pm 5, \pm 7$ and ± 9. The empirical size of each test was calculated based on $N = 10^3$ and for each Lin-McLeod Test, we use 10^3 iterations to calculate the p-value. The series length is $n = 10^2$ and we calculated for lag $m = 5, 10, 15$.

Using Table 4.3, we can compare the Fisher-Gallagher Portmanteau Test (Weighted Ljung-Box (Q_M) and Weighted Monti (W_M)) (Fisher and Gallagher, 2012a), general variance Test $\hat{D}_m$ (Mahdi and McLeod, 2011) and their trimmed test ($Q_M^{\#}$, $W_M^{\#}$, $\hat{D}_m^{\#}$). When applying this tests on the ARMA model with infinite variance, we can see the performance of the original tests become unstable. The bolded numbers in Table 4.3 present the number that are significantly different from the nominal rate 5%, while their trimmed tests would be more stable than it.

Figure 4.2: Comparison between Estimated Residuals after fitting AR(1) Model to AR(1) Model with Paretian Errors. Original Residuals Series and its Trimmed Series Presented below. Choosing $\phi = \pm 0.1, \pm 0.5$ and ± 0.9 and Stable Parameter $\alpha = 1.5, \beta = 1, \gamma = 0$. Series length $n = 1000$ and trimming level $\lambda = 5\%$.



As we can see, the tests would be more stable when the lag m increases as it would include more information about the residuals in the tests and would therefore be a more accurate estimator. The Figure 4.3 show the residuals for various ϕ_1 after being fitted by the AR(1) Model. We can see the trimming could stabilize the residuals, therefore it can improve the performance of the portmanteau test.

Table 4.3: Size Comparison between Fisher-Gallagher Test (both Ljung-Box Q_W and Monti M_W) as well as Their Application on Trimmed Time Series, and General Variance Portmanteau Test $\hat{D}_m$ and its Application on Trimmed Time Series. Choosing $\phi = \pm 0.1, \pm 0.3, \pm 0.5, \pm 0.7, \pm 0.9$ and lag $m = 5/10/15$ and Stable Distribution Parameters $\alpha = 1.5, \beta = 1, \gamma = 0$. Series length $n = 10^3$ and Nominal Level 5% .

ϕ_1	m	Q_W	$Q_W^{\#}$	M_W	$M_W^{\#}$	$\hat{D}_m$	$\hat{D}_m^{\#}$
−0.9	5	**6.5**	5.9	**6.4**	5.5	**6.4**	5.5
−0.7	5	4.1	5.9	4.2	5.7	4.3	5.8
−0.5	5	5.0	4.4	4.9	4.4	5.0	4.4
−0.3	5	4.5	5.8	4.4	5.2	4.5	5.1
−0.1	5	5.2	5.1	5.4	5.3	5.4	5.3
0.1	5	4.3	4.9	4.6	5.1	4.6	5.2
0.3	5	**3.8**	5.4	4.3	5.2	4.3	5.3
0.5	5	**3.5**	5.7	**3.2**	5.6	**3.0**	5.6
0.7	5	4.5	5.1	4.8	5.3	4.7	5.2
0.9	5	**6.1**	5.7	5.9	5.9	5.9	5.9
−0.9	10	4.4	4.9	4.4	5.1	4.7	5.2
−0.7	10	4.6	4.8	4.7	4.4	4.8	4.4
−0.5	10	**2.6**	4.1	**3.1**	4.6	**3.0**	4.7
−0.3	10	5.9	4.0	4.0	5.5	4.0	5.7
−0.1	10	4.1	5.8	4.3	5.2	4.3	5.2
0.1	10	**3.9**	5.0	4.3	5.2	4.4	5.1
0.3	10	4.5	5.1	4.6	4.9	4.4	4.8
0.5	10	**2.6**	4.1	**3.1**	4.6	**3.0**	4.7
0.7	10	4.1	5.0	4.0	5.7	4.0	5.5
0.9	10	4.4	4.4	4.1	4.1	4.2	4.3
−0.9	15	4.9	5.2	5.0	5.3	4.9	5.0
−0.7	15	**3.7**	5.3	4.1	5.9	4.1	5.8
−0.5	15	5.1	4.8	4.1	5.3	3.9	5.2
−0.3	15	4.2	5.2	**3.9**	4.2	**3.8**	4.2
−0.1	15	4.1	5.7	4.5	5.1	4.5	5.5
0.1	15	4.4	5.5	4.0	5.2	4.0	5.6
0.3	15	5.1	4.5	4.8	5.5	5.0	5.0
0.5	15	4.5	5.3	4.6	4.9	4.7	5.1
0.7	15	4.1	5.0	4.0	5.7	4.0	5.5
0.9	15	4.8	4.9	4.1	5.2	4.1	5.3

4.4.3 Further Power Comparison

In order to fully explore the performance of the trimmed portmanteau test based on the Fisher and Gallagher (2012a). Since they have already shown the asymptotic equivalence between Mahdi-McLeod test Mahdi and McLeod (2011) and Fisher-Gallagher test Fisher and Gallagher

(2012a), then we will mainly consider the Fisher-Gallagher portmanteau test with Ljung-Box type and Moni Type. Here, we use the package **WeightedPortTest** (Fisher and Gallagher, 2012b) in R to perform the Fisher-Gallagher Test.

First, we simulate the following models from ARMA $(2, 2)$ with stable distribution ($\alpha = 1.5, \beta = 1, \gamma = 0$).

Model 1 : $\phi_1 = 0.1, \phi_2 = 0.3$

Model 2 : $\phi_1 = 1.3, \phi_2 = -0.35$

Model 3 : $\theta_1 = -0.5$

Model 4 : $\theta_1 = 0.8$

Model 5 : $\phi_1 = 0.7, \theta_1 = -0.4$

Model 6 : $\phi_1 = 0.4, \theta_1 = -0.6$

Model 7 $\phi_1 = 0.7, \phi_2 = 0.2, \theta_1 = 0.5$

Model 8 : $\phi_1 = 0.9, \phi_2 = -0.4, \theta_1 = 1.2, \theta_2 = -0.3$

In the table 4.4, we should have the simulate series from the above ARMA $(2, 2)$ model with stable distribution, and then fit the data with AR(1) Model. We evaluate the Fisher-Gallagher Test and their applications on trimmed time series to compare their power. Then we increase the sample size of the simulation.

As we can see from Table 4.4, under most cases of ARMA $(2, 2)$, the Fisher-Gallagher Test on the trimmed time series would outperform its original counterpart. Interestingly, as the lag m increases, the power would decrease since the variance in the residuals would disturb the portmanteau test. And further, we could see in the cases of ARMA $(2, 2)$ with moving average components of θ, the Monti type Fisher-Gallagher Test would perform better than its Ljung-Box type. Their applications on the trimmed time series preserve this property.

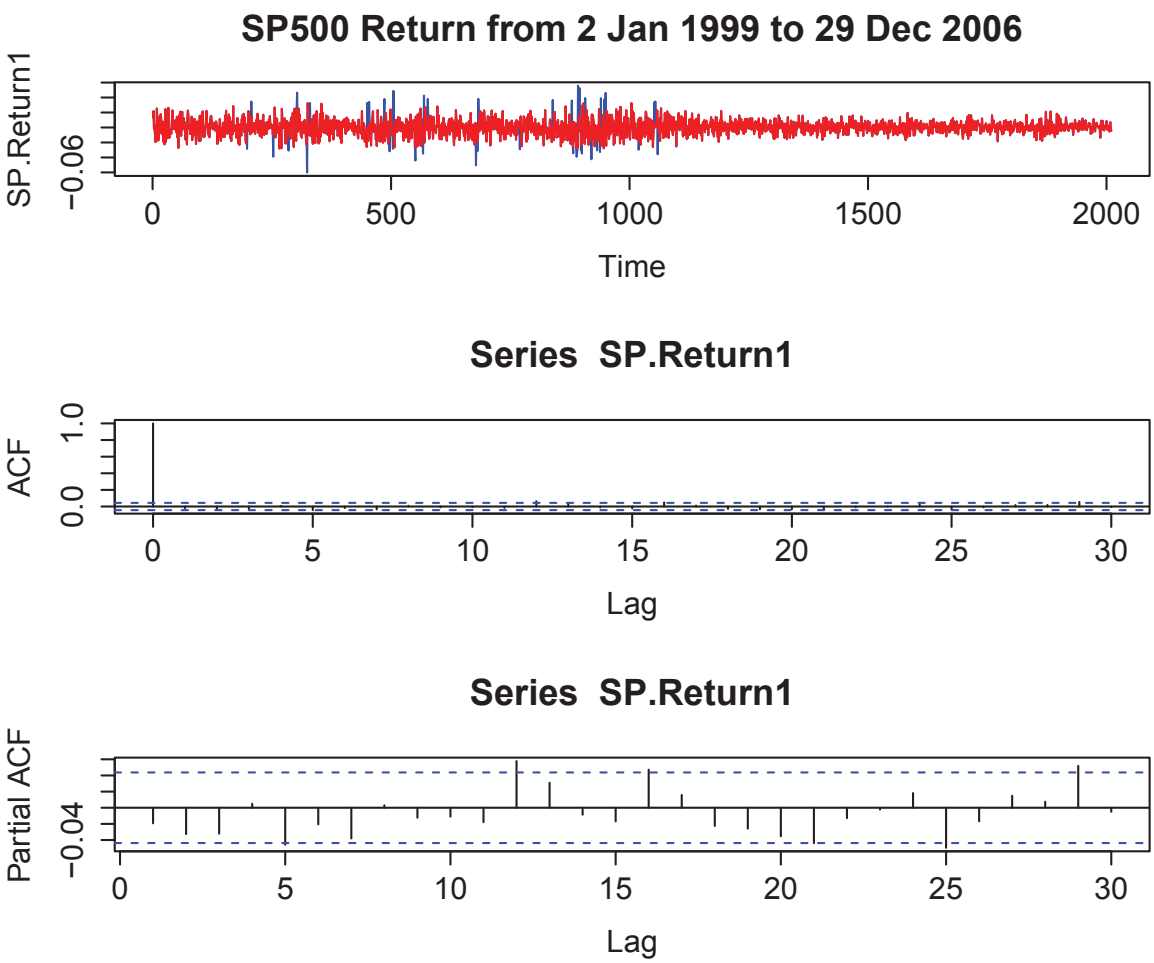
Table 4.4: Power Comparison between Portmanteau Test (Ljung-Box Fisher-Gallagher Test Q_W and its Application on Trimmed Time Series Q_W^*, Monti Fisher-Gallagher Test M_W and its Application on Trimmed Time Series M_W^*). Choosing Trimmed Level $\lambda = 0.01$, Series Length $n = 100$.

	m	Q_W	Q_W^*	M_W	W_m^*
	5	63.3	**75.2**	59.9	74.3
Model 1	10	57.0	**65.3**	53.8	57.1
	15	**53.7**	53.6	49.1	48.9
	5	62.1	91.2	69.8	**92.3**
Model 2	10	59.7	**89.7**	65.9	89.1
	15	53.2	**83.0**	56.6	77.7
	5	30.6	46.1	35.5	**52.3**
Model 3	10	22.4	31.4	27.9	**35.9**
	15	23.2	28.4	26.6	**30.8**
	5	69.5	98.9	76.8	**99.4**
Model 4	10	55.8	95.6	64.9	**98.3**
	15	**18.6**	16.4	16.4	16.1
	5	**18.6**	16.4	16.4	16.1
Model 5	10	**17.0**	12.8	15.4	12.2
	15	**14.7**	11.6	12.6	10.6
	5	13.4	**18.5**	15.5	21.3
Model 6	10	10.1	11.2	12.3	**13.4**
	15	9.9	11.0	11.3	**12.9**
	5	26.6	63.5	33.0	**72.6**
Model 7	10	20.7	45.4	26.1	**56.3**
	15	18.2	36.6	23.8	**44.0**
	5	96.9	**100**	98.1	**100**
Model 8	10	97.1	**99.9**	98.5	**99.9**
	15	97.4	**99.8**	97.4	98.9

4.5 Illustrative Application

Firstly, we consider the randomness test for the time series with infinite variance. As in Lin and McLeod (2008), we choose the $S\&P$ 500 daily index from 2 January ,1999 to 29 December ,2006. Then we calculate the log returns, $\log(X_{t+1}/X_t)$, and test the randomness for the series. We show the basic features of the data in the figure 4.3:

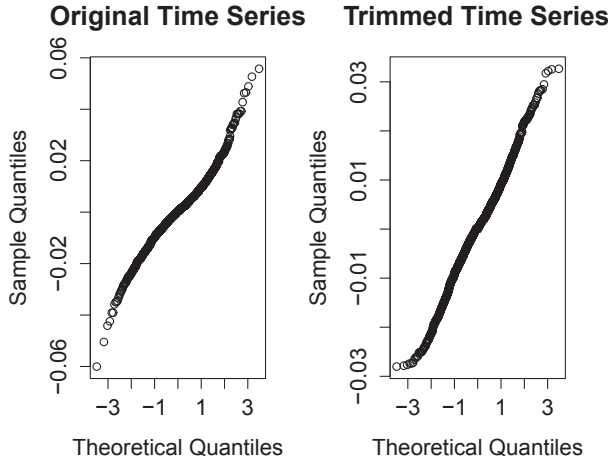
Figure 4.3: Daily Index Returns of $S\&P500$ From 2 January 1999 to 29 December 2006, Autocorrelation Plot (ACF) and Partial Autocorrelation Plot (PACF). Length of Series n = 2010



We plot the original time series as blue and the trimmed time series as red with trimmed level $\lambda = 0.01$. As we can see from the Figure 4.3, after trimming the time series, we have the time series with more stable and behave like strong white noise time series.

From Figure 4.4, after trimming the time series, we can have a time series behave much like

Figure 4.4: QQ Plot for Original Time Series and Trimmed Time Series with Trimmed Level $\lambda = 0.01$



the strong white noise case. However, we do need to choose the right level of trimming since the trimming will cause the loss of information.

Therefore, we apply the portmanteau test in Lin and McLeod (2008) as well as their application on trimmed time series to compare their performance. We use the algorithm for randomness test to evaluate the randomness and goodness-of-fit of the model adequacy.

Table 4.5: p-value Comparison on Evaluating Randomness for Squares of $S\&P500$ Index Returns and the trimmed time series with Ljung-Box Test. Series Length $n = 2011$. Monte Carlo iteration $N = 1000$. $*$ represent the trimmed level $\lambda = 0.01$ and $**$ represent the trimmed level $\lambda = 0.05$

Test Statistic/Lag	1	5	10	15	20	30	50
LB Normal	39.8	12.1	27.6	8.25	2.24	1.44	0.55
LB Stable	22.8	8.55	17.2	7.66	4.77	4.77	0.47
LB Stable(Trim)$*$	41.3	10.1	23.1	11.34	15.4	23.2	4.7
LB Stable(Trim)$**$	6.4	36.4	73.4	86.3	68.7	65.8	34.0

As we can see from Table 4.5, under the assumption of normality, the Ljung-Box Test would reject the null hypothesis of normality once the lag m is large enough. If we directly apply the Ljung-Box Test to test the null hypothesis of stable distribution, with small lag number m, the test works pretty well but it will reject the null hypothesis with large lag m. With trimmed time series, the test would still identify the stable distribution for the larger lag m.

Now, we test the performance of modified Fisher-Gallagher Test on the time series with infinite variance.

Table 4.6: P Value Comparison on Evaluating Randomness for Squares of $S\&P$ 500 Index Returns and the trimmed time series with Fisher-Gallagher Test. Series Length $n = 2011$. Monte Carlo Iteration $N = 1000$. $*$ represent the trimmed level $\lambda = 0.01$ and $**$ represent the trimmed level $\lambda = 0.05$

Test Statistic/Lag	1	5	10	15	20	30	50
FG Normal	37.5	21.0	27.6	17.6	5.71	2.64	4.47
FG Stable	22.6	11.7	17.2	11.0	7.11	5.52	3.23
FG Stable(Trim)$*$	39.4	15.3	12.1	10.9	10.4	13.1	9.99
FG Stable(Trim)$**$	7.51	14.9	41.2	59.3	64.6	65.9	58.9

From Table 4.7, after trimming, the null hypothesis of stable distribution would not be rejected at a large lag m for Fisher-Gallagher Test.

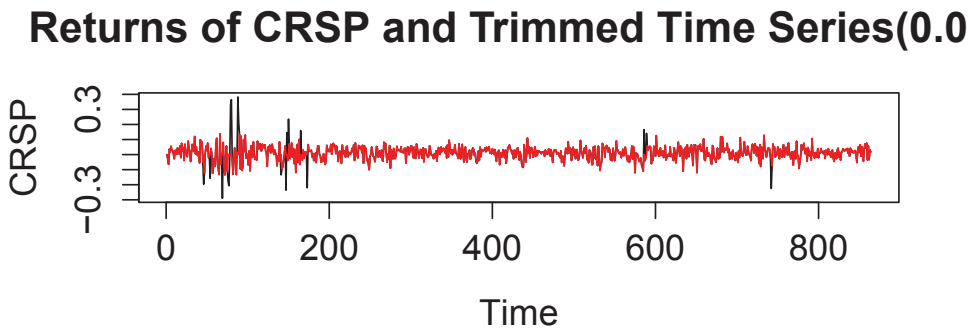
We follow the idea of Tsay (2005) and Lin and McLeod (2008) about the monthly simple return of CRSP Value-Weighted Index from Jan 1926 to Dec 1997. After fitting AR(5) model and using the algorithm of McCulloch (1986), we can get $\hat{\alpha} = 1.635$. Therefore, the infinite variance is plausible for this data according to Lin and McLeod (2008). Then we apply the Ljung-Box Test, Fisher-Gallagher Test and their applications on trimmed time series. We choose trimmed level $\lambda = 0.01$ and 0.05

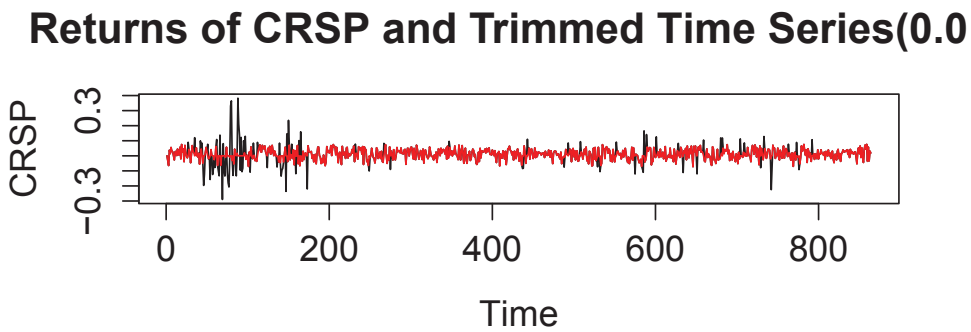
From Figure 4.5, when we choose trimmed level $\lambda = 0.01$, most information has already been included in the trimmed data but still some outlying data would affect the stability of the time series. While increasing the level of trimmed $\lambda = 0.05$, we will lose more information but the trimmed time series would appear to be much more stable. We should balance the trade-off between these two operations.

The model adequacy is evaluated based on the algorithm in the Appendix 4.7. We are attempting to evaluate whether the model is following strong ARMA model or ARMA model with infinite variance. Here, we compare the Ljung-Box Test and Fisher-Gallagher Test performance. Then we have Table 4.7:

Table 4.7: P Value Comparison on Goodness-of-fit for Monthly Simple Return of CRSP and the trimmed time series with Ljung-Box and Fisher-Gallagher Test against normal and stable distribution error assumption. Series Length $n = 864$. Monte Carlo Iteration $N = 1000$. $*$ represent the trimmed time series with trimmed level $\lambda = 0.01$

Figure 4.5: Month Simple Returns of CRSP Value-Weighted Index From January 1926 to December 1997, Autocorrelation Plot (ACF) and Partial Autocorrelation Plot (PACF). Length of Series $n = 864$





Test Statistic/Lag	1	5	10	20	30	40	50
Ljung-Box Stable	0.1	0.1	4.3	1.5	1.6	2.7	3.0
Ljung-Box Normal	0.1	0.2	5.0	3.0	0.4	0.8	2.2
Ljung-Box Stable∗	4.0	4.2	1.8	1.9	3.0	1.4	1.9
Ljung-Box Normal∗	0.0	0.0	0.1	0.0	0.3	0.0	0.4
Fisher-Gallagher Stable	0.1	1.1	4.9	2.5	1.4	1.6	1.3
Fisher-Gallagher Normal	0.1	0.1	6.2	9.0	0.3	0.2	0.8
Fisher-Gallagher Stable∗	4.0	3.0	1.7	1.8	0.8	1.4	1.3
Fisher-Gallagher Normal∗	0.0	0.0	0.0	0.0	0.0	0.0	0.0

From the table 4.7, the p-value for these tests are small and under the predetermined nominal level $\alpha = 0.5$. Under a close look, we find after trimming, both Ljung-Box and Fisher-Gallagher Test on Normal and Stable distribution assumption would get a more rejective p-value while the traditional test on original time series will be inflated. Another notable point is Fisher-Gallagher tend to be more conservative when testing for the error distribution especially on stable distribution assumptions.

4.6 Conclusion

In this chapter, we consider the Fisher-Gallagher Test on the trimmed time series. We derive the asymptotic distribution for the portmanteau test and with this approach, the size distortion problem alleviated. Various simulations show the better performance of the test on trimmed time series especially when we have the infinite variance error assumptions. We apply this test on various ARMA model and discover the power improvement in this case. Finally, we use some practical examples to show the power of the tests here.

4.7 Appendix

4.7.1 Lin-McLeod Test for Randomness on Trimmed Time Series

Here, we modify the Lin-McLeod Test (Lin and McLeod, 2008) on the trimmed time series for randomness test:

Algorithm 4.1. *We have a series $X_t, t = 1, 2, \ldots, n$. We have the null hypothesis for the series following a stable/normal distribution.*

Step One: *Calculate the portmanteau test statistics $\hat{Q}_W^0/\hat{M}_W^0$ from the series/trimmed series with predetermined trimmed level λ.*

Step Two: *Calculate the observed distribution estimation. For Stable Distribution Assumption, calculate the parameters $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$. For normal distribution assumption, calculate the parameters $\hat{\mu}, \hat{\sigma}$.*

Step Three: *Simulate the series $\hat{X}_{t,j}, t = 1, 2, \ldots, n$ based on the observed parameters calculated before. then calculate the portmanteau test $\hat{Q}_W^j/\hat{M}_W^j$ from series and trimmed series with the predetermined level λ.*

Step Four: *Repeat the Step Three for N times as $\hat{X}_{t,j}, t = 1, 2, \ldots, n, j = 1, 2, , \ldots, N$. Calculate the p value as*

$$pvalue = \frac{\#\left\{\hat{Q}_W^j > \hat{Q}_W^0\right\} + 1}{N + 1}. \tag{4.47}$$

Under this test, we could test whether the series would follow the null hypothesis here. Then the goodness-of-fit could be obtained by modified this algorithm slightly.

4.7.2 Lin-McLeod Test for Model Adequacy on Trimmed Time Series

Algorithm 4.2. *We have a series $X_t, t = 1, 2, \ldots, n$. And then we fit the data with models and get the residuals. Then we evaluate the the model adequacy according to the residuals series/trimmed residuals series.*

Step One: *Fit the model to the series X_t and get the estimated parameters $\hat{\lambda}$ and the residuals $\hat{\epsilon}_t^0, t = 1, 2, \ldots, n$. Calculate the portmanteau test statistics $\hat{Q}_W^0/\hat{M}_W^0$ on the residuals series $\hat{\epsilon}_t^0$ or trimmed residuals series $\tilde{\epsilon}_t^0$.*

Step Two: *Fit the residuals* $\hat{\epsilon}_t^0, t = 1, 2, \ldots, n$ *to get the distribution parameters. For stable distribution assumption, we have* $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$*. For normal distribution assumption, we have* $\hat{\mu}, \hat{\sigma}$*.*

Step Three: *Simulate the series* $\hat{X}_{t,j}, t = 1, 2, \ldots, n$ *based on the estimated parameter* $\hat{\lambda}$ *and distribution estimator from the previous step. Then we could fit the simulated series* $\hat{X}_{t,j}$ *and get the residuals* $\hat{\epsilon}_{t,j}$*.*

Step Four: *Calculate the portmanteau test* $\hat{Q}_{M,j}, \hat{M}_{W,j}$ *from the series* $\hat{X}_{t,j}$ *or trimmed series* $\tilde{X}_{t,j}$ *with a predetermined trimmed level* λ*.*

Step Five: *Repeat the Step Three and Four for N times as* $\hat{X}_{t,j}, t = 1, 2, \ldots, n, j = 1, 2, \ldots, N$*. Calculate the p value as*

$$pvalue = \frac{\#\left\{\hat{Q}_W^j > \hat{Q}_W^0\right\} + 1}{N + 1}. \tag{4.48}$$

Chapter 5

Independent Test between Time Series

5.1 Introduction

In order to test the independence between two time series, Hong (1996) proposed an important portmanteau test based on correlation:

$$Q_N = \frac{N \sum_{j=1-N}^{N-1} k^2(j/M)\hat{\rho}_{uv}^2 - S_N(k)}{\{2D_N(k)\}^{1/2}}, \tag{5.1}$$

where $M/N \to 0$ and

$$S_N(k) = \sum_{1-N}^{N-1}(1 - |j|/N)k^2(j/M), \tag{5.2}$$

$$D_N(k) = \sum_{2-N}^{N-2}(1 - |j|/N)(1 - (|j| + 1)/N)k^4(j/M), \tag{5.3}$$

$k(|j|/M)$ is the kernel function. As proposed by Hong (1996), some common kernel, such as Parzen, Bartlett, Daniell, and quadratic spectral kernel, can be used here. Under the regularity condition, Q_N would converge to a normal distribution and the p-value could be calculated accordingly. This test can be treated as an important modification of the portmanteau test proposed by Haugh (1976) where the test statistics is defined as:

$$S = N \sum_{j=-M}^{M} \hat{\rho}_{uv}(j). \tag{5.4}$$

We know this statistics would converge to a chi-squared distribution. Based on the simulation results of Hong (1996), the Hong's Test would outperform the Haugh's Test under the hypothesis of independence for the AR model with $\phi = 0.5$ by providing a corrected type one error and larger power level (Hong, 1996). However, after revisiting the paper. we discovered that there is an error in the simulation process. When compared with the Haugh Test, according to

the definition, the correct formula for the adjusted test is:

$$S^* = N^2 \sum_{j=-M}^{M} (N - |j|)^{-1} \rho_{uv}^2(j), \tag{5.5}$$

However, when Hong simulates the results, mistakenly he put the formula as:

$$S^* = N^2 \sum_{j=-M}^{M} (N - j)^{-1} \rho_{uv}^2(j), \tag{5.6}$$

Analytically, this changes would produce a lower level of statistics S^* which contribute to a larger level of type one error. Therefore, we reproduce and compare the type one error for both correct and incorrect Haugh's Test.

Further, we investigate the independent test of two time series with GARCH type of error. The test on this type of time series is rarely discussed in the literature. Hong (1999) point out the tests based on correlation would be inappropriate since it would miss the nonlinear process with zero autocorrelation. Then he theoretically propose an independent test based on the generalized spectral density function and theoretically justify the power of the test. His method is extremely theoretical and hard to implement for this type of problems.

When we explore this type of tests here, we find out when the time series involved GARCH type of error, not just inappropriately but incorrectly identify independence by providing a spurious correlation between the series here. We explore the reasons for this phenomenon and modify the independent test with the Lin-McLeod Test (Lin, 2006; Mahdi and McLeod, 2011) and its modified version. The Lin-McLeod Test would give more adaption for the tests and allow them to have more complicated residuals structure in the time series. The size and power advantages will be shown in later the later simulation study.

5.2 Corrected Empirical Results and Spurious Correlation

Hong (1996) develop the correlation-based portmanteau test and its adjusted statistics. According to his simulation results, we find out the Hong's test would have a better performance on type one error. However, the equation of Haugh's adjust statistics on page 622 is clearly wrong due to an omitted absolute operation. This would lead to different simulation results.

Table 5.1: Size Comparison between Hong Statistics (Hong, 1996) and Haugh Statistics (Haugh, 1976) under Corrected Formula. Choosing $X_t = 0.5X_{t-1} + u_t, Y_t = 0.5Y_{t-1} + v_t$ where $u_t, v_t \sim N(0,1)$ and $\rho_{uv} = 0$ for all j. Series Length $n = 1000$, Nominal Level $\alpha = 0.5\%$. Q_N and $Q_N^\star$ stand for Hong Statistics and its adjusted statistic. S_N represent the Haugh Test. $S_R^\star$ stands for the right adjusted Haugh statistic and $S_W^\star$ Stands for the wrong adjusted Haugh statistic. We underline the statistics closest to the nominal level α.

lag	$N = 100$						$N = 200$					
	$M = 5$		$M = 8$		$M = 12$		$M = 5$		$M = 9$		$M = 15$	
α	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q_N	10.3	7.6	10.3	6.2	9.5	6.3	<u>10.1</u>	6.4	10.5	7.0	11.4	6.5
$Q_N^\star$	<u>9.9</u>	6.7	10.2	5.9	5.9	3.2	10.9	6.5	9.3	5.7	7.5	3.9
S	7.7	3.1	8.7	4.2	6.2	1.9	8.1	3.8	7.7	3.6	6.0	3.1
$S_R^\star$	9.3	<u>4.6</u>	<u>9.9</u>	4.8	<u>9.9</u>	<u>4.9</u>	9.5	<u>4.8</u>	<u>10.0</u>	<u>4.6</u>	<u>10.0</u>	<u>5.0</u>
$S_W^\star$	12.2	5.6	8.4	<u>5.0</u>	10.2	4.9	10.9	5.2	7.6	3.7	10.5	4.7

From the table 5.1, we can easily see that with the correct form of Haugh adjusted statistics. Actually, it gets close to the nominal level α which contradicts to the statement in Hong (1996). Still, we explore the power of these tests under the null hypothesis and alternative hypothesis following Hong (1996) simulation.

For convenience, we list the three hypothesises used in Hong (1996).

Alternative 1:

$$\rho_{uv}(j) = \begin{cases} 0.2 & j = 0 \\ 0 & \text{Otherwise} \end{cases} \tag{5.7}$$

Alternative 2:

$$\rho_{uv}(j) = \begin{cases} 0.125 & j = 0 \\ \sin(0.125\pi j)/\pi j & 1 \le j \le 8 \\ 0 & \text{Otherwise} \end{cases} \tag{5.8}$$

Alternative 3:

$$\rho_{uv}(j) = \begin{cases} 0.3 & j = 3 \\ 0 & \text{Otherwise} \end{cases} \tag{5.9}$$

We simulate two processes X_t and Y_t as: $X_t = 0.5X_{t-1} + u_t$ and $Y_t = 0.5Y_{t-1} + v_t$ where the error term we assumed they are under these three hypothesises. In the simulation we produce 10^4 replications.

Under Alternative 1, u_t and v_t are correlated simultaneously but not otherwise. In contrast, the cross-correlation function of Alternative 2 has a maximum value at $j = 0$, and then decays sinusoidally and smoothly to 0 at $j = 8$. For Alternative 3, u_t and v_t are correlated only at lag $j = 3$. These three types of correlation present the basic linear correlations.

Table 5.2: Power Comparison under Three Alternative Comparison of $\rho_{uv}(j)$. Choosing $X_t = 0.5X_{t-1} + u_t, Y_t = 0.5Y_{t-1} + v_t$ where $u_t, v_t \sim N(0,1)$. Series Length $n = 1000$, Nominal Level $\alpha = 5\%$

		$N = 100$			$N = 200$		
		$M = 5$	$M = 8$	$M = 12$	$M = 5$	$M = 9$	$M = 15$
Alter 1	Q_N	34.24	28.65	23.77	63.32	53.32	43.39
	$Q_N^\star$	32.65	26.66	18.72	63.48	51.08	37.87
	S	14.95	10.65	6.71	34.71	25.09	15.56
	$S^\star$	17.33	13.37	10.32	35.87	26.47	20.12
Alter 2	Q_N	39.63	39.62	35.83	71.78	73.19	66.30
	$Q_N^\star$	37.12	36.00	27.91	70.71	71.37	60.67
	S	25.51	16.49	10.24	60.74	43.71	26.70
	$S^\star$	26.97	20.57	16.07	62.48	45.67	31.70
Alter 3	Q_N	17.84	42.15	43.80	46.22	84.49	84.26
	$Q_N^\star$	15.54	35.67	36.98	44.52	82.61	81.02
	S	37.98	28.08	21.02	82.96	71.39	57.37
	$S^\star$	41.64	34.17	28.27	83.92	73.66	63.36

As shown in Table 5.2, in general cases, Hong's and its adjusted version outperform the Haugh's Test. However, in some cases with smaller lag m, the Haugh's test shows its power there. After we revise the Haugh's Test formula, we see less differences lie between Hong and Haugh Test. This gives evidences that Haugh's Test still have good power for these types of correlation between errors. Another point is that the adjust made on the original statistics is not always improving the performance of the statistics.

In order to further explore and compare the power of the statistics, we reapply these tests on the model with GARCH type of error. The data generating mechanics could be described

[0]In the table 5.2, Q_N and $Q_N^\star$ stand for Hong Statistics and its adjusted statistic. S represent the Haugh Test. $S^\star$ stands for the right adjusted Haugh statistic. The table shows the power of identification under three alternatives.

as followed: X_t and Y_t can still be defined as an autoregressive model with lag of 1, $X_t = 0.5X_{t-1} + u_t$, and $Y_t = 0.5X_{t-1} + v_t$. Replacing the usual error assumption of normality, we assume the following possible cases:

- u_t follows $N(0,1)$ and v_t follows GARCH $(1,1)$ model.
- Both u_t and v_t follow GARCH $(1,1)$ model.

In these cases, the errors are simulated independently but with weak autocorrelation embedded in. For illustrative purpose, we choose the parameters in GARCH model as $\alpha_0 = 10^{-6}, \alpha_1 = 0.2, \beta = 0.7$ and a skewed generalized error distribution with skewness of 1.25 and shape of 4.8. And the GARCH model defined as follows:

$$a_t = v_t \sqrt{h_t}, \tag{5.10}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{i=1}^{p} \beta_i a_{t-i}. \tag{5.11}$$

Under the above data generating process, we produce the power table with 10^4 repetition iteration for 95% confidence level.

Table 5.3: Power Comparison between Hong's test and Haugh's test under the GARCH error. Series Length $N = 100, 200$. Choosing lag $M = \log n$ for Comparison. Q_N and $Q_N^\star$ stand for Hong Statistics and its adjusted statistic. S represent the Haugh Test. $S^\star$ stands for the right adjusted Haugh statistic. The table shows the power of identification under two models.

		$N = 100$			$N = 200$		
		$M = 5$	$M = 8$	$M = 12$	$M = 5$	$M = 9$	$M = 15$
Model 1	Q_N	5.00	3.62	2.00	5.40	3.48	1.60
	$Q_N^\star$	5.32	2.81	1.38	5.78	2.88	1.28
	S	1.23	0.5	0.13	1.85	0.69	0.29
	$S^\star$	1.49	0.85	0.44	1.99	0.93	0.36
Model 2	Q_N	95.97	95.18	94.39	95.90	95.96	95.31
	$Q_N^\star$	95.74	94.97	93.94	96.39	95.78	94.82
	S	93.25	91.04	88.98	94.67	93.47	90.65
	$S^\star$	93.51	91.79	88.41	94.94	93.94	90.91

The drastic difference between these two models result alerts us about the usefulness of the independent tests. As we can clearly see from the model, when the tests are applied on the model 1, the rejection rate is usually less than 5%, the nominal level it should be. Especially when lag M gets increased, the level would decrease significantly even lower than 1%. This gives us an idea that the statistics would tend to the null hypothesis of independence. However, when the tests are applied on the second model, the situation becomes reversed. The rejection rates are usually no less than 88% which gives us a results in favor the alternative hypothesis. This form

a kind of spurious autocorrelation with overestimated p-value. After careful exploration, we find the test statistics of both Haugh and Hong tests would become enlarged after introducing the GARCH type of error. Then, the asymptotic distribution of the test statistics would move to the right significantly. Therefore, the p-value would increase significantly which contribute to this inaccuracy of the tests. However, when we closely look at these results, Hong's results still outperform Haugh's test.

Bibliography

Adler, R. J., R. E. Feldman, and C. Gallagher (1998). Analyzing stable time series. In R. J. Adler, R. E. Feldman, and M. S. Taqqu (Eds.), *A practical guide to heavy tails: statistical techniques and application*, pp. 133–158. Boston, MA: Birkhauser.

Andrews, D. W. K. (1991). Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Arnold, S. F. (1990). *Mathematical Statistics*. Prentice-Hall.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics 73*(1), 5–59.

Beran, J. (1994). *Statistics for Long Memory Process*. Chapman and Hall.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Box, G. E. et al. (1954, 09). Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics 25*(3), 484–498.

Box, G. E., G. M. Jenkins, and G. C. Reinsel (2008). *Time Series Analysis: Forecasting and Control*. Wiley&Sons Inc. Publication.

Box, G. E. P. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive- intergrated moving average time series models. *Journal of American Statistical Association 65*(332), 1509–1526.

Chen, W. W. and R. S. Deo (2004). A generalized portmanteau goodness-of-fit test for time series models. *Econometric Theory 20*(02), 382–416.

Cui, Y. and R. Wu (2014). A portmanteau test for arma processes with infinite variance. *Communications in Statistics-Simulation and Computation 43*(3), 597–614.

Davis, N., C. M. Triggs, and P. Newbold (1977). Significance level of the box-pierce portmanteau statistics in finite sample. *Biometrika 64*(3), 517–522.

Davis, R. and S. Resnick (1986). Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics 2*(14), 533–558.

Deo, R. S. (2000). Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *Journal of Econometrics 99*(2), 291–315.

Dufour, J.-M. (2006). Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics 133*(2), 443–477.

Dufour, J. M. and R. Roy (1986). Generalized portmanteau statistics and tests of randomness. *Communications in Statistics - Theory and Methods 15*(10), 2953–2972.

Efron, B. (1979). Bootstrap methods:another look at jackknife. *The Annals of Statistics 7*(1), 1–26.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society 50*(4), 987–1007.

Fisher, T. J. and C. M. Gallagher (2012a). New weighted portmanteau statistics for time series goodness of fit testing. *Journal of American Statistical Association 107*(498), 777–787.

Fisher, T. J. and C. M. Gallagher (2012b). *R Package: Weighted Portmanteau Tests for Time Series Goodness-of-Fit* (1.0 ed.).

Francq, C., R. Roy, and J.-M. Zakoïan (2005). Diagnostic checking in arma models with uncorrelated errors. *Journal of the American Statistical Association 100*(470), 532–544.

Francq, C. and J.-M. Zakoïan (1998). Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference 68*(1), 145–165.

Francq, C. and J.-M. Zakoian (2011). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.

Gallagher, C. M. and T. J. Fisher (2013, July). A comparison of various weighted portmanteau test for time series goodness-of-fit. Technical report, University of Missouri-Kansas City.

Godfrey, L. G. and A. R. Tremayne (2005). The wild bootstrap and heteroskedasticity-robust tests for serial correlation in dynamic regression models. *Computational Statistics & Data Analysis 49*(2), 377–395.

Granger, W. J. and R. Joyeux (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis 1*(1), 15–29.

Harris, D. and H. Kew (2014). Portmanteau autocorrelation tests under q-dependence and heteroskedasticity. *Journal of Time Series Analysis 35*(3), 203–217.

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. *Journal of the American Statistical Association 71*(354), 378–385.

Hong, Y. (1996). Testing for independence between two covariance stationary time series. *Biometrika 83*(3), 615–625.

Hong, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *Journal of the American Statistical Association 94*(448), 1201–1220.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika 68*(1), 165–176.

Kheoh, T. S. and A. I. McLeod (1992). Comparison of two modified portmanteau tests for model adequacy. *Computational Statistics & Data Analysis 14*(1), 99–106.

Kunsch, H. (1989). The jackknife and the bootstrap for general stationary observation. *The Annals of Statistics 17*(3), 1217–1241.

Kwan, A. C., A.-B. Sim, and Y. Wu (2005). A comparative study of the finite-sample of some portmanteau test for randomness of a time series. *Computational Statistics & Data Analysis 48*(2), 391–413.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. New York: Spinger.

Lee, S. and C. T. Ng (2010). Trimmed portmanteau test for linear processes with infinite variance. *Journal of Multivariate Analysis 101*(4), 984–998.

Li, W. K. (2004). *Diagnostic Check in Time Series*. Chapman & Hall.

Li, W. K. and T. K. Mak (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis 15*(6), 627–636.

Li, W. K. and A. I. McLeod (1981). Distribution of the residual autocorrelation in multivariate arma time series model. *Journal of The Royal Statistical Society, Series B 43*(2), 231–239.

Li, W. K. and A. I. McLeod (1986). Fractional time series modelling. *Biometrika 73*(1), 217–221.

Lin, J.-W. (2006). *Essays on Diagnostic Checks in Time Series*. Ph. D. thesis, University of Western Ontario.

Lin, J.-W. and A. I. McLeod (2006). Improved Peňa Rodríguez portmanteau test. *Computational Statistics & Data Analysis 51*(3), 1731–1738.

Lin, J.-W. and A. I. McLeod (2008). Portmanteau test for ARMA models with infinite variance. *Journal of Time Series Analysis 29*(3), 600–617.

Liu, R. Y. and K. Singh (1992). Moving blocks jackknife and bootstrap capture weak dependence. In R. Lepage and L. Billard (Eds.), *Exploring the Limits of the Bootstrap*, pp. 225–248. Wiley.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

Lobato, I., J. C. Nankervis, and N. E. Savin (2001). Testing for autocorrelation using a modified box-pierce q test. *International Economic Review 42*(1), 187–205.

Lobato, I. N., J. C. Nankervis, and N. E. Savin (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory 18*(03), 730–743.

Mahdi, E. and A. I. McLeod (2011). Improved multivariate portmanteau test. *Journal of Time Series Analysis 33*(2), 211–222.

Mahdi, E. and A. I. McLeod (2013). *portes :Portmanteau Tests for Univariate and Multivariate Time Series Models*.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annal of Statistics 21*(1), 255–285.

McCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics-Simulation and Computation 15*(4), 1109–1136.

McCullough, B. and H. Vinod (1998). Implementing the double bootstrap. *Computational Economics 12*(1), 79–95.

McLeod, A. I. (1978). On the distribution of residual autocorrelations in Box - Jenkins models. *Journal of Royal Statistical Association,Series B 40*(3), 297.

McLeod, A. I. and J. Veenstra (2007). *FGN: Fractional Gaussian Noise, estimation and simulaton*.

McLeod, A. I., H. Yu, and Z. L. Krougly (2007). Algorithms for linear time series analysis: With r package. *Journal of Statistical Software 5*(23), 1–26.

Monti, A. C. (1994). A proposal for residual autocorrelation test in linear models. *Biometrika 81*(4), 776–780.

Paparoditis, E. and D. N. Politis (2001a). Tapered block bootstrap. *Biometrika 88*(4), 1105–1119.

Paparoditis, E. and D. N. Politis (2001b). The tapered block bootstrap for general statistics from stationary sequences. *Econometrics Journal 5*(1), 131–148.

Peňa, D. and J. Rodríguez (2002). A powerful portmanteau test of lack of fit for time series. *Journal of The American Statistical Association 97*(458), 601–610.

Peňa, D. and J. Rodríguez (2006). The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series. *Journal of Statistical Planning and Inference 136*(8), 2706–2718.

Phillips, P. C. B. and K.-L. Xu (2006). Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis 27*(2), 289–308.

Politis, D. N. and J. P. Romano (1992). A circular block bootstrap resampling procedure for stationary data. In R. Lepage and L. Billard (Eds.), *Exploring the limits of bootstrap*, pp. 263–270. Wiley.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of American Statistical Association 89*(428), 1303–1313.

Ramsey, F. L. et al. (1974). Characterization of the partial autocorrelation function. *The Annals of Statistics 2*(6), 1296–1301.

Romano, J. P. and L. A. Thombs (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association 91*(434), 590–600.

Runde, R. (1997). The asymptotic null distribution of the box-pierce q-statistic for random variables with infinite variance an application to german stock returns. *Journal of Econometrics 78*(2), 205–216.

Shao, X. (2010). The dependent wild bootstrap. *Journal of American Statistical Association 105*(489), 218–235.

Tsay, R. S. (2005). *Analysis of financial time series*, Volume 543. John Wiley & Sons.

Veenstra, J. (2013). *Persistence and Anti-Persistence:Theory and Software*. Ph. D. thesis, University of Western Ontario.

Veenstra, J. and A. I. McLeod (2013). *The arfima R package: Exact Methods for Hyperbolic Decay Time Series*. Version 1.2-5.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics 14*(4), 1261–1295.

Yu, H. (2014). *Rmpi: Interface (Wrapper) to MPI (Message-Passing Interface)*. Version 0.6.5.

Zhu, K. and W.-K. Li (2013). A bootstrapped spectral test for adequacy in weak arma models. *Munich Personal RePEc Archive*.

Jinkun Xiao

Curriculum Vitae

EDUCATION

Ph.D Candidate in Statistics 2012-2016

Western University, London, Ontario, Canada

Research Area: Time Series Analysis, Data Mining, Scientific Computing
Thesis: Advances in Portmanteau Diagnostic Checks
Supervisor: Dr. A. Ian McLeod

Master of Science in Statistics 2011-2012

Western University, London, Ontario, Canada

Relevant Courses: Advanced Regression, Advanced Data Analysis, Advanced Probability, Topics in Statistical Inference, Time Series, Experimental Design, etc.

Bachelor in Actuarial Science 2007-2011

Hunan University, Changsha, Hunan, China

Relevant Courses: Life Contigency, Risk Management, Financial Market and Derivatives, etc.

WORK EXPERIENCE

Teaching and Research Assistant 2011-2016

Western Universiy, London, Ontario, Canada

Sessional Lecturer 2008-2011

New Oriental Edcuational Group, Changsha, Hunan, China

SKILLS

Languages	Mandarin(Native)
	English (fluent)
	German (fluent)
Software	R, LaTeX, SAS, SQL

REFERENCES

Available upon request