Bootstrap-Based Measurement of Serial correlation in Time Series Objects

by

Cun Wang

(Under the Direction of Cheolwoo Park)

Abstract

Serial correlation is a fundamental problem in time series analysis. Since financial data are collected every few minute and due to the large-scale of data, finding the correlation of time series data accurately becomes a crucial problem. This thesis explores the monthly correlation of time series data when daily values are observed. In this case, we call daily values within a month a time series object, and the goal of this study is to measure the correlation in time series objects. Traditionally, researchers often take the last day value, but this approach may lose information within the month. This thesis proposes an approach based on resampling techniques, bootstrap and wild bootstrap, and compare their finite sample performances with the last day approach.

Index words: Bootstrap, Serial correlation, Time series object, Wild bootstrap

Bootstrap-Based Measurement of Serial correlation in Time Series Objects

by

Cun Wang

B.B.A., Beijing Technology and Business University, 2014

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Master of Science

Athens, Georgia

2017

Bootstrap-Based Measurement of Serial correlation in Time Series Objects

by

Cun Wang

Approved:

Major Professor: Cheolwoo Park

Committee: Rai Bay
Sue Ryung Chang

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

Acknowledgments

I take this opportunity to express gratitude to my advisor Dr. Cheolwoo Park for his valuable guidance on this thesis. I also thank Dr. Rai Bay and Dr. Sue Ryung Chang for their time serving as the committee. Special thanks to my families and friends for their support through out my Master's study.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Objective of the Study 2

2 Background **4**

2.1 Autocorrelation 4

2.2 ARMA Model 7

2.3 ARCH and GARCH Models 8

2.4 FIGARCH Model 10

2.5 Bootstrap Method 12

3 Proposed Method **14**

3.1 Estimation of Monthly Serial Correlation Based on Bootstrap 14

3.2 Estimation of Monthly Serial Correlation Based on Wild-bootstrap 15

4 Simulation **17**

4.1 Simulation Study 17

4.2 Simulation Results 18

5 Conclusion **28**

Bibliography **29**

Appendix **32**

0.1 Algorithm for Bootstrap of time series objects and calculate correlation values 32

0.2 Substitute time series objects 34

List of Tables

4.1 Estimated ACFs of randomly generated samples from different models under the assumption that there are 36 months and 20 days in each month B=1000. 20

4.2 Estimated ACFs of randomly generated samples from different models under the assumption that there are 48 months and 20 days in each month B=1000. 21

4.3 Estimated ACFs of randomly generated samples from different models under the assumption that there are 60 months and 20 days in each month B=1000. 22

4.4 Estimated ACFs of randomly generated samples from different models under the assumption that there are 72 months and 20 days in each month B=1000. 23

4.5 Estimated ACFs of randomly generated samples from different models under the assumption that there are 36 months and 20 days in each month B=2000. 24

4.6 Estimated ACFs of randomly generated samples from different models under the assumption that there are 48 months and 20 days in each month B=2000. 25

4.7 Estimated ACFs of randomly generated samples from different models under the assumption that there are 60 months and 20 days in each month B=2000. 26

4.8 Estimated ACFs of randomly generated samples from different models under the assumption that there are 72 months and 20 days in each month B=2000. 27

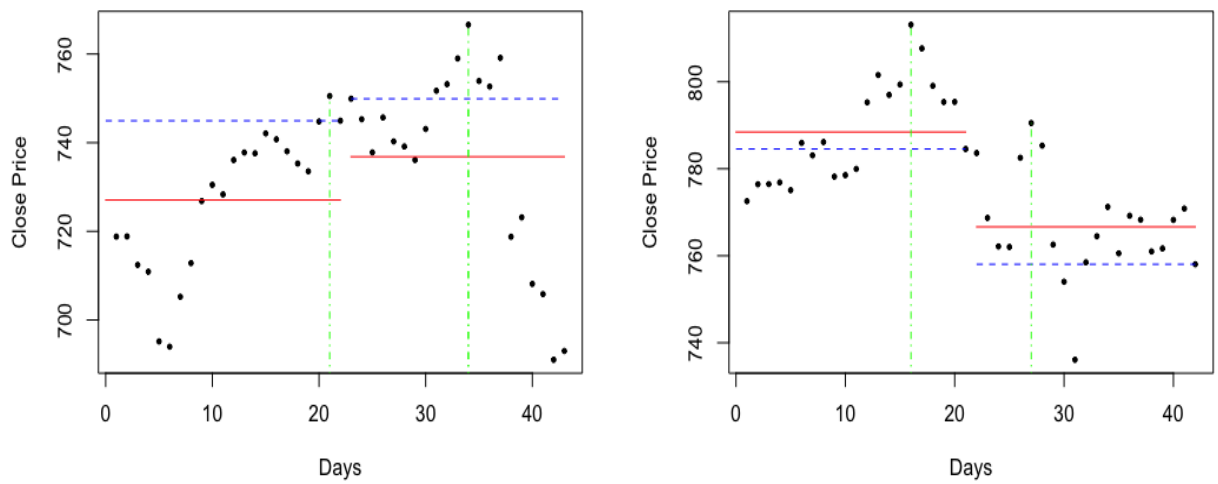
Chapter 1

Introduction

1.1 Motivation

Data collected sequentially over time induces a correlation between measurements because observations near each other in time will tend to be similar. Therefore, studying serial correlation is a fundamental problem in time series analysis. Serial correlation serves as an elementary tool for analyzing and modeling time series data with statistical methods. In finance, some researchers (e.g. Adrangi et al. (1999) and Chiang and Doong (2001)) study the daily, weekly and monthly stock returns or index serial correlation, and sometimes find somewhat paradoxical results (e.g. Campbell (1998)). There could be several possible explanations for those results, and in order to solve this problem, we first focus on the basic question: how do we measure daily, weekly, or monthly autocorrelation? More generally, how do we measure the serial correlation for a sequence of time series objects? Time series object refers to a set of time series (e.g. stock prices recorded every five minutes in a day or every business day in a month).

In this thesis, we are interested in measuring monthly serial correlation when daily data are available. One simple approach is using a representative value from each month, such as the last day's value or average value and measure the correlation of the time series. Or, one can use the within-variation by creating interval-value data using the maximum and minimum for each month.



(a) March-April (b) October-November

Figure 1.1: Alphabet Inc. stock price

In Figure 1.1, we plot some examples of stock prices of Alphabet Inc in two different time periods. In the plots, the red solid line represents the mean value of each month. The blue dotted horizontal line indicates the value of the last day for each month. The green vertical dashed line shows the maximum value of each month. It is obvious from the plot that using different statistics results in different correlations between months.

1.2 Objective of the Study

As technology evolves, the availability of large-scale computational power and storage space increases and financial data are recorded every second. Mining accurate information from such a large amount of data becomes a problem. The objective of the proposed work is to study the serial correlation of a sequence of time series objects. Examples include monthly correlation when time series are observed on a daily basis. Lee and Park (2017) developed a test for examining the serial correlation in mean and variance of time series objects. The main contribution of this thesis is to explore an appropriate approach to measuring the serial correlation of the time series objects by utilizing the internal variation information within a time series object via bootstrap and wild bootstrap sampling . As mentioned in Gonçalves and Kilian (2004), standard bootstrap procedures could be invalid in the presence of conditional heteroscedasticity, therefore, we apply the wild bootstrap sampling method.

This thesis is organized as follows. Chaper 2 introduces background knowledge of autocorrelation, some popular time series models and bootstrap methods. Chapter 3 illustrates the proposed method. Chapter 4 shows the simulation results. Finally, Chapter 5 provides the conclusion.

Chapter 2

Background

2.1 Autocorrelation

In statistics, the autocorrelation of a random process is the correlation between values of the process at different times, as a function of two times of the time lag. Let X be a stochastic process, and t be any point in time. Then X_t is the value produced by a given run of the process at time t with mean μ_t. Then the definition of the autocorrelation between times s and t is:

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

where $\gamma(s,t) = cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)]$ for all s and t is the autocovariance function, and $\rho(s,t)$ is some value between -1 and 1 with 1 indicating perfect correlation and -1 indicating perfect anti-correlation. The autocovariance function depends on the separation of X_s and X_t, say, $h = |s - t|$, and not on where the points are located in time. As long as the points are separated by h units, it does not matter where the location of the two points are. This is called weak stationarity when the mean is constant.

Although there is no special assumptions about the behavior of the time series, many of the existing examples hinted that there exists a sort of regularity over time in a time series. Here, the regularity is the concept called stationarity.

A strictly stationary time series is one for which the probabilistic behavior of every collection of values

$$X_{t1}, X_{t2}, \ldots, X_{tk}$$

is identical to that of the time shifted set

$$X_{t1+h}, X_{t2+h}, \ldots, X_{t3+h}.$$

That is,

$$P\{X_{t_1} \leqslant c_1, \ldots, X_{t_k} \leq c_k\} = P\{X_{t1+h} \leqslant c_1, \ldots, X_{t+h} \leq c_k\}.$$

For all $k = 1, 2, \ldots$, all time points $t_1, t_2, \ldots, t_k$, all numbers $c_1, c_2, \ldots, c_k$, and all time shifts $h = 0, \pm 1, \pm 2, \ldots$. If a time series is strictly stationary, all of the multivariate distribution functions for subsets of variables must agree with their counterparts in the shifted set for all values of the shift parameter h. Thus if the variance function of the process exists, the autocovariance function of the series X_t satisfies

$$\gamma(s, t) = \gamma(s + h, t + h)$$

for all s and t and h. We can interpret this by saying the autocovariance function of the process depends only on the time difference between s and t, and not on the actual times.

A weakly stationary time series, X_t, is a finite variance process such that the mean value μ_t is constant and does not depend on time t, and the autocovariance function, $\gamma(s, t)$, depends on s and t only through their difference $|s - t|$.

We usually use the term stationary to mean weakly stationary, if a process is stationary in the strict sense, we will use the term strictly stationary. Because the autocovariance function, $\gamma(s,t)$, of a stationary time series, X_t, depends on s and t only through their difference $|s-t|$, we can simplify the notation:

$$\gamma(t+h,t) = cov(X_{t+h}, x_t) = cov(X_h, X_0) = \gamma(h,0)$$

since the time difference between times $t+h$ and t is the same as the time difference between times h and 0. Thus for convenience, we can write it as $\gamma(h)$.

Then, the autocorrelation function of a stationary time series can be shown as:

$$\rho(h) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t+h,t+h)\gamma(t,t)}} = \frac{\gamma(h)}{\gamma(0)}.$$

Although the theoretical autocorrelation function is useful for describing the properties of certain hypothesized methods, most of the analyses must be performed using sampled data. This causes a problem because we will typically not have i.i.d copies of x_t that are available for estimating the covariance and correlation functions. Somehow, we must use averages over this single realization to estimate the population means and covariance functions.

Accordingly, if a time series is stationary, the mean function $\mu_t = \mu$ is constant so that we can estimate it by the sample mean form the observed time series X_t,

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

The theoretical autocorrelation function is estimated by the sample autocorrelation function as follows:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

where:

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})$$

with $\hat{\gamma}(-h) = \hat{\gamma}(h)$ for $h = 0, 1, \ldots, n-1$ and $t + h \leq n$.

2.2 ARMA Model

ARMA models combine autoregressive (AR) and moving averages (MA) into a composite model of the time series. Autoregressive models are based on the idea that the current value of the series, x_t, can be explained as a function of p past values, X_{t-1},X_{t-2},. . . ,X_{t-p}, where p determines the number of steps into the past needed to forecast the current value. This model can be written in the form:

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \omega_t$$

where the terms in α are autocorrelation coefficients at lags $1, 2 \ldots, p$ and ω_t is a random error term and it often assumed to be white Gaussian noise with mean 0 and variance σ_ω^2. Note that this error term specifically relates to the current time period t. The autoregressive representation in which the X_t on the left-hand side of the equation is assumed to be combined linearly. As an alternative, the moving average model of order q, abbreviated as MA(q), assumes the white noises ω_t on the right-hand side of the defining equation are combined linearly. It can be written as:

$$X_t = \omega_t + \beta_1\omega_{t-1} + \beta_2\omega_{t-2} + \ldots + \beta_q\omega_{t-q}$$

where the β_i terms are the weights applied to prior values in the time series. We assume that ω_t is a Gaussian white noise series with mean zero and variance σ^2_ω. We can combine these two models by simply adding them together as a model of order (p,q), where we have p AR terms and q MA terms:

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \omega_t + \beta_1\omega_{t-1} + \ldots + \beta_q\omega_{t-q}$$

with $\alpha_p \neq 0$, $\beta_q \neq 0$, and $\sigma^2_\omega > 0$. The parameters p and q are called the autoregressive and moving average orders, respectively.

In general, this form of combined ARMA model can be used to model a time series with fewer terms overall than either an MA or an AR model by themselves. It expresses the estimated value at time t as the sum of q terms that represent the average variation of random variation over q previous periods (the MA component), plus the sum of p AR terms that compute the current value of X as the weighted sum of the p most recent values.

2.3 ARCH and GARCH Models

ARMA models assume a constant variance. However, recently, in finance researchers have motivated the study of the volatility, or variability, of a time series. Models such as the autoregressive conditionally heteroscedastic (ARCH) process, first introduced by Engle (1982), is a model for the variance of a time series. It has constant unconditional variance but non-constant conditional variance. ARCH forecasts current period value depending on the changing conditional variance.

Let ϵ_t denote the error terms, and then these error terms are split into a stochastic piece z_t

and a time-dependent standard deviation σ_t:

$$\epsilon_t = \sigma_t z_t$$

The random variable z_t is a white noise process. The series σ_t^2 is modelled by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$$

where $\alpha_0 > 0$ and $\alpha_i \geqslant 0$, $i > 0$.

Engle (1982) mentions several attractive characteristics of the ARCH model. It can be used to describe a changing, possibly volatile variance. Although an ARCH model could possibly be used to describe a gradually increasing variance over time, most often it is used in situations in which there may be short periods of increased variation. Gradually increasing variance related to a gradually increasing mean level might be better handled by transforming the variable. ARCH models are created in the context of econometric and finance problems having to do with the amount that investments or stocks increase (or decrease) per time period, so there is a tendency to expound them as models for that type of variable.

A more general process, generalized autoregressive conditional heteroscedasticity (GARCH) model, has also been a popular tool to analyze financial time series data since it was introduced by Bollerslev (1986). It takes a weighted average of lagged squared returns and the lagged conditional variance. An univariate GARCH(p,q) model satisfies:

$$X_t = \sigma_t \epsilon_t, \sigma_t^2 = \omega^o + \sum_{i=1}^{q} \alpha_i^o X_{t-i}^2 + \sum_{j=1}^{p} \beta_j^o \sigma_{t-j}^2$$

where the innovation $\{\epsilon_t\}_{t \in Z}$ is a sequence of standard i.i.d. random variables. It is also assumed that $\omega^o > 0$, $\alpha^o \geq 0$ for all $i = 1, \ldots, q$ and $\beta_j^o \geq 0$ for all $j = 1, \ldots, p$. The GARCH models are conditionally heteroscedastic, but have a constant unconditional variance. The volatility is highly persistent when the sum of the estimated coefficients of the squared lagged returns and the lagged conditional variance terms in a GARCH model are close to 1.

Practically, both ARCH and GARCH models can successfully model financial time series with clustering of volatilities and in finance, volatility clustering can be understood as information clustering. While asset pricing attempts to reflect the market news on the price as accurately as possible, successfully modeling the volatility clustering makes ARCH and GARCH models attractive to researchers.

2.4 FIGARCH Model

It has long been recognized that asset returns determined in speculative markets are possibly uncorrelated but not independent through time, the reason is that most return processes tend to underlie temporal bursts of volatility. Therefore, modeling the volatility of a time series can improve the efficiency in parameter estimation and the accuracy of forecast. In particular, after learning the ARCH model by Engle (1982) and GARCH model by Bollerslev (1986), many researchers have noted the extreme degree of persistence of shocks to the conditional variance process; for a study of extensive literature on ARCH modeling in finance see Bollerslev et al. (1992). Analogous to the problems relating to the proper modeling of the long-run dependencies in the conditional mean of economic time series, similar issues can be seen as modeling of conditional variances. Engle and Bollerslev (1986) formulate the Integrated GARCH or IGARCH, this class of model possesses many of the features of the unit root. For example, the underlying effect of a shock for optimal forecast of the future conditional variance will cause the corresponding cumulative impulse response weights tend toward a nonzero constant, and the forecasts will increase linearly with the forecast horizon. This signifies that the pricing of risky securities, including long-term options and futures contracts, may show extreme dependence on the initial conditions or the current state of the economy. However, this extreme degree of dependence seems to conflict with observed pricing behavior.

The Fractionally Integrated GARCH (FIGARCH) model of Baillie et al. (1996) combines many of the features of the fractionally integrated process for mean together with the regular GARCH process for the conditional variance. A key feature of the FIGARCH model is that it develops a more flexible class of processes for the conditional variance, that can represent and explain the observed temporal dependencies in financial market volatility. Especially, the FIGARCH model allows absolute innovations in the conditional variance function or only a slow hyperbolic rate of decay for the lagged squared.

We see from above that a GARCH(p,q) process may also be expressed as an ARMA(m,q) process in ω_t^2,by writing

$$[1-\alpha(L)-\beta(L)]\omega_t^2 = \alpha_o + [1-\beta(L)]\sigma_t,$$

where $m = \max\{p, q\}$ and $\sigma_t = \omega_t^2 - h_t$. The $\{\sigma_t\}$ process can be interpreted as the "innovations" for the conditional variance, as it is a zero-mean martingale. Therefore, an integrated GARCH(p,q) process can be written as

$$[1-\alpha(L)-\beta(L)](1-L)\omega_t^2 = \alpha_0 + [1-\beta(L)]\sigma_t.$$

The fractionally integrated GARCH or FIGARCH class of models is obtained by replacing the first difference operator (1-L) with the fractional differencing operator $(1-L)^d$, where d is a fraction $0 < d < 1$. Thus, the FIGARCH class of models can be obtained by considering:

$$[1-\alpha(L)-\beta(L)](1-L)^d\omega_t^2 = \alpha_0 + [1-\beta(L)]\sigma_t.$$

2.5 Bootstrap Method

The bootstrap method, initially introduced by Efron (1979) for independent variables, is a class of nonparametric methods that allows practitioners to implement statistical inference on a wide range of problems without imposing many structural assumptions on the underlying data-generating random process. The method involves generation of a large number of independent resamples or bootstrap samples, each drawn from the original sample with replacement. Later this method has been extended to deal with more complex dependent variables.

By now, there are many books, e.g. Efron94 (1994), Tu and Shao (1995), Davison and Hinkley (1997), that discuss various aspects of the bootstrap methodology at different situations of sophistication. Furthermore, several researchers such as Berkowitz and Kilian (2000), Carey (2005), Härdle et al. (2003), Hongyi Li and Maddala (1996), and Politis et al. (2003) give overviews of different aspects of bootstrapping on time series data, such as parametric or nonparametric time series models, autoregressive and Markov processes, long range dependent time series and nonlinear time series. Relevant bootstrap methods including residual bootstrap, Markovian bootstrap and the prominent block bootstrap approaches.

The basic idea behind the bootstrap methods can be described in general terms as follows. Let $X_1, \ldots, X_n$ be a stretch of a time series with joint distribution P_n. For estimating a population parameter θ, suppose that we have constructed an estimator $\hat{\theta}_n$ (e.g., using the generalized method of moments) based on $X_1, \ldots, X_n$. There is a common problem that statisticians need to address to assess the accuracy of $\hat{\theta}_n$, for example, by using an estimate of its mean square error (MSE) or an interval estimate of a given confidence level. However, any such measure of accuracy depends on the sampling distribution of $\hat{\theta}_n - \theta$, which is typically unknown in practice and often very complicated. Bootstrap methods provide a general recipe for estimating the distribution of $\hat{\theta}_n$ and its functionals without restrictive model assumptions on the time series.

In this thesis, we aim to provide an idea about resample time series data by bootstrap and wild bootstrap methods, and then compare their measured performance with the last day value method.

The wild bootstrap method, originally proposed by Wu (1986) is suited in heteroscedastic situation. The idea is to leave the regressors at their sample value and resample the response variable based on the residual values. That is, for each replicate, one computes a new y based on

$$y_i = \hat{y}_i + \hat{\epsilon}_i v_i$$

so the residuals can randomly multiply a random variable v_i with mean 0 and variance 1. Different forms can be used for the random variable v_i, such as

- The standard normal distribution
- A distribution suggested by Mammen (1993):

$$v_i = \begin{cases} -(\sqrt{5}-1)/2 & \text{with probability}(\sqrt{5}+1)/(2\sqrt{5}), \\ (\sqrt{5}+1)/2 & \text{with probability}(\sqrt{5}-1)/(2\sqrt{5}). \end{cases}$$

- Rademacher distribution:

$$\omega_t = \begin{cases} 1 & \text{with probability}\frac{1}{2}, \\ -1 & \text{with probability}\frac{1}{2}. \end{cases}$$

Chapter 3

Proposed Method

3.1 Estimation of Monthly Serial Correlation Based on Bootstrap

We denote a sequence of time series objects with random length $\{D_t\}_{t=-\infty}^{\infty}$ by $\{X_t\}_{t=-\infty}^{\infty}$ with $X_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,D_t})'$, because in real life, there are different business days within each month, obtaining random length rather than fixed length in the definition of monthly time series objects makes more sense. A monthly time series object $X_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,D_t})'$ contains D_t business days at the tth month, and its ith component $X_{t,i}$ of X_t is the value of the ith business day of the tth month. Throughout the thesis, we use the following assumption on $\{D_t\}_{t=-\infty}^{\infty}$:

(D) A sequence of random length $\{D_t\}_{t=-\infty}^{\infty}$ is i.i.d discrete random variables on a finite sample space $S_D = \{d_1, \ldots, d_K\}$ with $d_1 < d_2 \ldots < d_K$ and independent of $\{X_{t,1}, \ldots, X_{t,l_t}\}_{t=-\infty}^{\infty}$ for any given $l_t \in S_D, -\infty < t < \infty$.

In this thesis, we assume that there are 20 business days in each month for convenience of simulation study, so we consider fixed length of time series data instead of random length but it would not affect the purpose of our study.

Let $\{X_t\}_{t=1}^n$ be randomly sampled monthly data with $X_t = (X_{t1}, X_{t2}, \ldots, X_{tm})$ for t=1,...,n and m is a number of data in t-th month. Now we are interested in a definition of an autocorrelation function ρ for $\{X_t\}_{t=1}^n$. We propose a nonparametric bootstrap-like resampling method to fit an autocorrelation function ρ as follows: For $b = 1, \ldots$,B,

1. For each $t = 1, \ldots, n$, we sample X_t^{*b} from X_t and denote the bootstrap sample $(X_1^{*b}, X_2^{*b}, \ldots, X_n^{*b})$.

2. Obtain the bth autocorrelation function estimate $\hat{\rho}^{*b}(k)$ for k=0,1,...,n-1 with the time series sample X_t^{*b} in Step 1 above. The final estimate can be obtained by taking the average of B estimates, e.g.,

$$\bar{\hat{\rho}}^*(k) = \frac{1}{B}\sum_{b=1}^{B} \hat{\rho}^{*b}(k)$$

for $k = 1, \ldots, n-1$.

3. Calculate bias of $\bar{\hat{\rho}}(k)$, standard deviation of $\hat{\rho}(k)$ and root mean square error (RMSE) of $\hat{\rho}(k)$.

3.2 Estimation of Monthly Serial Correlation Based on Wild-bootstrap

Standard bootstrap measures for serial correlation fail to consider conditional heteroscedasticity, according to Ahlgren and Catani (2012). This is because the standard bootstrap samples fail to mimic the conditional heterosedasticity of the original data and in order to solve this issue, we apply the wild bootstrap-based sampling method.

For the standard bootstrap test, we consider a bootstrap time series sample $\{X_t^*\}_{t=-\infty}^{\infty}$ from a sequence of time series objects $\{X_t\}_{t=-\infty}^{\infty}$ by sampling a single value X_t^* from X_t independently from the empirical distribution P_t defined by $\{X_{t,1}, X_{t,2}, \ldots, X_{t,D_t}\}$. Under **(D)** and the assumption $\{X_{t,1}, X_{t,2}, \ldots, X_{t,l_t}\}_{t=-\infty}^{\infty}$ for any given $l_t \in S_D$, $-\infty < t < \infty$, is weakly stationary with mean μ and finite variance σ^2. Then we define consistency of the h

lag sample autocorrelation function of a bootstrap time series sample $\{X_t^*\}_{t=1}^n$ as

$$\hat{\rho}^*(h) = \frac{\frac{1}{n-h}\sum_{t=1}^{n-h}(X_t^* - \bar{X}_n^*)(X_{t+h}^* - \bar{X}_n^*)}{\frac{1}{n-1}\sum_{t=1}^{n}((X_t^*)^2 - \bar{X}_n^*)^2}$$

However, the standard bootstrap sampling does not consider the conditional heteroscedasticity and in order to deal with this problem, we now employ the wild bootstrap method. The wild bootstrap samples do not have serial correlation in mean, but can mimic conditional hetereoscedasticity as the original data. In our paper, we apply the Rademacher distribution because this distribution works well as suggested by Davidson and Flachaire (2008).

We consider the Rademacher distribution for the wild bootstrap method:

$$\omega_t = \begin{cases} 1 & \text{with probability}\frac{1}{2}, \\ -1 & \text{with probability}\frac{1}{2}. \end{cases}$$

The wild bootstrap is now implemented to obtain the wild bootstrap time series samples using the observed time series objects $\{X_t\}_{t=1}^n$ as follows:

1. For each $b = 1, \ldots, B$, we sample X_{tt}^{*b} from X_t and obtain a wild bootstrap sample as $X_{t,WB}^{*b} = X_t^{*b}\omega_t$, where ω_t is a random draw from the Rademacher distribution, denoting the wild bootstrap time series sample $\{X_{t,WB}^{*b}\}_{t=1}^T$.

2. Calculate the b-th sample autocorrelation function estimate $\hat{\rho}_{WB}^{*b}(k)$ for $k = 1, \ldots, n-1$ using $\{X_{T,WB}^{*b}\}_{t=1}^T$. The final estimate can be obtained by taking the average of B estimated, e.g.,

$$\bar{\hat{\rho}}^*(k) = \frac{1}{B}\sum_{b=1}^{B}\hat{\rho}^{*b}(k)$$

for $k = 1, \ldots, n-1$.

3. Calculate bias of $\bar{\hat{\rho}}(k)$, standard deviation of $\hat{\rho}(k)$ and root mean square error (RMSE) of $\hat{\rho}(k)$.

Chapter 4

Simulation

4.1 Simulation Study

We conduct a simulation study under various settings including AR(1), ARMA(1,1),GARCH(1,1) and FIGARCH models and set these models with different coefficients:

- AR(1) models:

$$X_t = \phi X_{t-1} + \epsilon_t$$

- ARMA(1,1) models:

$$X_t = \phi X_{t-1} + \theta \omega_{t-1} + \omega_t$$

- GARCH(1,1) models:

$$X_t = \sigma_t \epsilon_t$$

$$\sigma_t{}^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$$

- FIGARCH models:

$$[1 - \alpha(L) - \beta(L)](1 - L)^d \omega_t^2 = \alpha_0 + [1 - \beta(L)]\sigma_t.$$

For the proposed method, we simulate interval-valued data from the above models respectively and set B=1000 and 2000 when estimating the monthly correlation. We repeat this process 1000 times for each simulation setting. First, we assume that there are 20 days for each month and generate 36,48,60,72 months of data respectively, which results in 720,960,1200,1440 observations. We consider $\rho(20)$ as the true monthly autocorrelation for each time series model. For the parameter values of AR(1) models, we consider $\phi = 0.99, 0.95, 0.90$ and the corresponding correlation values are $\rho(20) = 0.8179, 0.3585, 0.1216$ respectively. For those of ARMA(1,1) models, we consider $(\alpha, \beta) = (0.9, 0.1), (0.8, 0.2), (0.7, 0.3)$ and the corresponding correlation values are $\rho(20) = 0.1237, 0.0123, 0.0009$, respectively. For those of GARCH(1,1) models, we consider $(\alpha, \beta) = (0.05, 0.90), (0.10, 0.85), (0.15, 0.80)$ with a fixed value $\omega = 10^{-6}$ and consequently, $\rho(20) = 0.0274, 0.0676, 0.1132$ for squared observations. The figures in Tables indicate bias, standard deviation (SD), and root mean square error (RMSE) of the estimates by two proposed bootstrap methods and the estimates by using the last day of each month.

4.2 Simulation Results

Tables 4.1-4.4 report the result when we set bootstrap sample size equal to 1000 and the length of months (36,48,60,72), respectively. We can see from tables that two bootstrap methods are overall better than the last day value method, but the wild bootstrap method's performance varies from setting to setting as for AR and FIGARCH models, the bias of the wild bootstrap is the largest, while for GARCH models, all three methods yield close biases and sometimes the wild bootstrap approach produces the smallest value. In additional, for ARMA models, the wild bootstrap method sometimes shows the similar result as the last day value. It seems that these trends are not affected by the length of months.

Tables 4.5-4.8 report the result when we set bootstrap sample size equal to 2000 and the length of months 36,48,60,72, respectively. The reason we make this setting is that we want to see whether different bootstrap sample sizes can change the result. As we can see from the tables that the two bootstrap methods are better overall than the last day value method, but the wild bootstrap method's performance varies from setting to setting. As for the AR and FIGARCH models, the bias of the wild bootstrap approach is the largest; for GARCH models, all three methods shows close biases. For ARMA models, sometimes the wild bootstrap method yields similar results as the last day value method. It seems that the bootstrapping sample size will not affect the trends.

Table 4.1: Estimated ACFs of randomly generated samples from different models under the assumption that there are 36 months and 20 days in each month B=1000.

Models	Correlation	Methods	Bias	SD	RMSE
AR(1)	$\rho(20) = 0.8179$	Boots	0.1332	0.1216	0.1803
		Wild Boots	0.8441	0.0071	0.8442
		Last Day	0.1396	0.1348	0.1940
	$\rho(20) = 0.3585$	Boots	0.0359	0.1152	0.1207
		Wild Boots	0.3857	0.0057	0.3857
		Last Day	0.0654	0.1587	0.1716
	$\rho(20) = 0.1216$	Boots	-0.0107	0.0895	0.0902
		Wild Boots	0.1492	0.0053	0.1492
		Last Day	0.0396	0.1618	0.1667
ARMA(1,1)	$\rho(20) = 0.1237$	Boots	-0.0058	0.0912	0.0913
		Wild Boots	0.1513	0.0055	0.1514
		Last Day	0.0423	0.1610	0.1665
	$\rho(20) = 0.0123$	Boots	-0.0068	0.0601	0.0605
		Wild Boots	0.0310	0.0052	0.0403
		Last Day	0.0321	0.1586	0.1618
	$\rho(20) = 0.0009$	Boots	0.0095	0.0426	0.0437
		Wild Boots	0.0288	0.0051	0.0292
		Last Day	0.0264	0.1611	0.1632
GARCH(1,1)	$\rho(20) = 0.0274$	Boots	0.0546	0.0095	0.0554
		Wild Boots	0.0553	0.0052	0.0556
		Last Day	0.0546	0.1671	0.1758
	$\rho(20) = 0.0676$	Boots	0.0949	0.0098	0.0944
		Wild Boots	0.0952	0.0053	0.0953
		Last Day	0.0918	0.1682	0.1916
	$\rho(20) = 0.1132$	Boots	0.1405	0.0010	0.1409
		Wild Boots	0.1409	0.0057	0.1410
		Last Day	0.1466	0.1713	0.2255
FIGARCH	$\rho(20) = 0.1301$	Boots	-0.0050	0.0120	0.0130
		Wild Boots	-0.0263	0.0284	0.0387
		Last Day	-0.0049	0.0326	0.0329

Table 4.2: Estimated ACFs of randomly generated samples from different models under the assumption that there are 48 months and 20 days in each month B=1000.

Models	Correlation	Methods	Bias	SD	RMSE
AR(1)	$\rho(20) = 0.8179$	Boots	0.0933	0.0990	0.1360
		Wild Boots	0.8379	0.0065	0.8380
		Last Day	0.0961	0.1046	0.1420
	$\rho(20) = 0.3585$	Boots	0.0164	0.1004	0.1017
		Wild Boots	0.3791	0.0051	0.3792
		Last Day	0.0448	0.1351	0.1423
	$\rho(20) = 0.1216$	Boots	-0.0179	0.0776	0.0796
		Wild Boots	0.1425	0.0048	0.1426
		Last Day	0.0296	0.1329	0.1361
ARMA(1,1)	$\rho(20) = 0.1237$	Boots	-0.0193	0.0773	0.0796
		Wild Boots	0.1445	0.0047	0.1446
		Last Day	0.0377	0.1411	0.1461
	$\rho(20) = 0.0123$	Boots	-0.01635	0.0512	0.0538
		Wild Boots	0.0330	0.0044	0.0333
		Last Day	0.0228	0.1340	0.1418
	$\rho(20) = 0.0009$	Boots	-0.0006	0.0370	0.0370
		Wild Boots	0.0216	0.0045	0.0221
		Last Day	0.0218	0.1395	0.1412
GARCH(1,1)	$\rho(20) = 0.0274$	Boots	0.0490	0.0085	0.0497
		Wild Boots	0.0483	0.0045	0.0485
		Last Day	0.0488	0.1480	0.1558
	$\rho(20) = 0.0676$	Boots	0.0883	0.0086	0.0887
		Wild Boots	0.0884	0.0047	0.0886
		Last Day	0.0893	0.1450	0.1703
	$\rho(20) = 0.1132$	Boots	0.1340	0.0089	0.1343
		Wild Boots	0.1340	0.0048	0.1341
		Last Day	0.1396	0.1512	0.2059
FIGARCH	$\rho(20) = 0.1162$	Boots	-0.0120	0.0108	0.0227
		Wild Boots	-0.0409	0.0260	0.0485
		Last Day	-0.0212	0.0276	0.0349

Table 4.3: Estimated ACFs of randomly generated samples from different models under the assumption that there are 60 months and 20 days in each month B=1000.

Models	Correlation	Methods	Bias	SD	RMSE
AR(1)	$\rho(20) = 0.8179$	Boots	0.0746	0.0840	0.1123
		Wild Boots	0.8342	0.0057	0.8342
		Last Day	0.0778	0.0898	0.1189
	$\rho(20) = 0.3585$	Boots	0.0102	0.0875	0.0881
		Wild Boots	0.3753	0.0046	0.3753
		Last Day	0.0433	0.1222	0.1296
	$\rho(20) = 0.1216$	Boots	-0.0258	0.0711	0.0757
		Wild Boots	0.1383	0.0042	0.1384
		Last Day	0.0208	0.1266	0.1283
ARMA(1,1)	$\rho(20) = 0.1237$	Boots	-0.0241	0.0727	0.0766
		Wild Boots	0.1403	0.0042	0.1404
		Last Day	0.0778	0.0898	0.1189
	$\rho(20) = 0.0123$	Boots	-0.0185	0.0445	0.0481
		Wild Boots	0.0288	0.0039	0.0290
		Last Day	0.0180	0.1290	0.1302
	$\rho(20) = 0.0009$	Boots	-0.0038	0.0334	0.0336
		Wild Boots	0.0174	0.0039	0.0178
		Last Day	0.0147	0.1259	0.1267
GARCH(1,1)	$\rho(20) = 0.0274$	Boots	0.0439	0.0078	0.0446
		Wild Boots	0.0441	0.0040	0.0442
		Last Day	0.0480	0.1333	0.1416
	$\rho(20) = 0.0676$	Boots	0.0840	0.0080	0.0847
		Wild Boots	0.0842	0.0042	0.0843
		Last Day	0.0821	0.1310	0.1546
	$\rho(20) = 0.1132$	Boots	0.1299	0.0078	0.1301
		Wild Boots	0.1297	0.0045	0.1290
		Last Day	0.1273	0.1371	0.1870
FIGARCH	$\rho(20) = 0.1233$	Boots	-0.0133	0.0102	0.0168
		Wild Boots	-0.0292	0.0236	0.0376
		Last Day	-0.0129	0.0266	0.0230

Table 4.4: Estimated ACFs of randomly generated samples from different models under the assumption that there are 72 months and 20 days in each month B=1000.

Models	Correlation	Methods	Bias	SD	RMSE
AR(1)	$\rho(20) = 0.8179$	Boots	0.0638	0.0751	0.0985
		Wild Boots	0.8314	0.0051	0.8314
		Last Day	0.0685	0.0807	0.1059
	$\rho(20) = 0.3585$	Boots	0.0037	0.0780	0.0781
		Wild Boots	0.3722	0.0041	0.3722
		Last Day	0.0357	0.1052	0.1111
	$\rho(20) = 0.1216$	Boots	-0.0286	0.0651	0.0711
		Wild Boots	0.1351	0.0039	0.1352
		Last Day	0.0260	0.1144	0.1173
ARMA(1,1)	$\rho(20) = 0.1237$	Boots	-0.0282	0.0635	0.0695
		Wild Boots	0.1377	0.0039	0.1377
		Last Day	0.0194	0.1135	0.1152
	$\rho(20) = 0.0123$	Boots	-0.0238	0.0442	0.0501
		Wild Boots	0.0262	0.0038	0.0265
		Last Day	0.0113	0.1144	0.1150
	$\rho(20) = 0.0009$	Boots	-0.0006	0.0317	0.0324
		Wild Boots	0.0150	0.0037	0.0154
		Last Day	0.0182	0.1182	0.1196
GARCH(1,1)	$\rho(20) = 0.0274$	Boots	0.0410	0.0069	0.0416
		Wild Boots	0.0413	0.0037	0.0414
		Last Day	0.0494	0.1171	0.1271
	$\rho(20) = 0.0676$	Boots	0.0816	0.0071	0.0819
		Wild Boots	0.0814	0.0039	0.0816
		Last Day	0.0919	0.1193	0.1506
	$\rho(20) = 0.1132$	Boots	0.1272	0.0077	0.1274
		Wild Boots	0.1270	0.0040	0.1270
		Last Day	0.1240	0.1247	0.1760
FIGARCH	$\rho(20) = 0.1413$	Boots	0.0036	0.0097	0.0104
		Wild Boots	-0.0126	0.0230	0.0262
		Last Day	0.0044	0.0243	0.0247

Table 4.5: Estimated ACFs of randomly generated samples from different models under the assumption that there are 36 months and 20 days in each month B=2000.

Models	Correlation	Methods	Bias	SD	RMSE
AR(1)	$\rho(20) = 0.8179$	Boots	0.1280	0.1164	0.1730
		Wild Boots	0.8441	0.0050	0.8441
		Last Day	0.1342	0.1300	0.1868
	$\rho(20) = 0.3585$	Boots	0.0399	0.1136	0.1204
		Wild Boots	0.3858	0.0041	0.3858
		Last Day	0.0715	0.1537	0.1696
	$\rho(20) = 0.1216$	Boots	-0.0286	0.0651	0.0711
		Wild Boots	0.1351	0.0039	0.1352
		Last Day	0.0260	0.1144	0.1173
ARMA(1,1)	$\rho(20) = 0.1237$	Boots	0.0550	0.0086	0.0557
		Wild Boots	0.0551	0.0036	0.0552
		Last Day	0.0567	0.1659	0.1754
	$\rho(20) = 0.0123$	Boots	-0.0058	0.0567	0.0570
		Wild Boots	0.0399	0.0036	0.0400
		Last Day	0.0280	0.1519	0.1544
	$\rho(20) = 0.0009$	Boots	0.0075	0.04457	0.0451
		Wild Boots	0.0288	0.0036	0.0290
		Last Day	0.0310	0.1600	0.1629
GARCH(1,1)	$\rho(20) = 0.0274$	Boots	0.0550	0.0086	0.0557
		Wild Boots	0.0551	0.0036	0.0552
		Last Day	0.0567	0.1659	0.1753
	$\rho(20) = 0.0676$	Boots	0.0949	0.0087	0.0953
		Wild Boots	0.0952	0.0038	0.0953
		Last Day	0.0967	0.1690	0.1947
	$\rho(20) = 0.1132$	Boots	0.1414	0.0089	0.1417
		Wild Boots	0.1406	0.0040	0.1406
		Last Day	0.1427	0.1721	0.2235
FIGARCH	$\rho(20) = 0.1205$	Boots	-0.0150	0.0116	0.0190
		Wild Boots	-0.0363	0.0287	0.0463
		Last Day	-0.0155	0.0314	0.0350

Table 4.6: Estimated ACFs of randomly generated samples from different models under the assumption that there are 48 months and 20 days in each month B=2000.

Models	Correlation	Methods	Bias	SD	RMSE
AR(1)	$\rho(20) = 0.8179$	Boots	0.0950	0.1014	0.1390
		Wild Boots	0.8379	0.0044	0.8379
		Last Day	0.0989	0.1090	0.1472
	$\rho(20) = 0.3585$	Boots	0.0211	0.1002	0.1204
		Wild Boots	0.3788	0.0036	0.3788
		Last Day	0.0552	0.1335	0.1444
	$\rho(20) = 0.1216$	Boots	-0.0156	0.0758	0.0773
		Wild Boots	0.1424	0.0033	0.1425
		Last Day	0.0334	0.1417	0.1456
ARMA(1,1)	$\rho(20) = 0.1237$	Boots	-0.0163	0.0788	0.0805
		Wild Boots	0.1444	0.0032	0.1444
		Last Day	0.0377	0.1397	0.1447
	$\rho(20) = 0.0123$	Boots	-0.0146	0.0499	0.0520
		Wild Boots	0.0331	.0031	0.0332
		Last Day	0.0229	0.1423	0.1441
	$\rho(20) = 0.0009$	Boots	0.0011	0.0374	0.0374
		Wild Boots	0.0217	0.0032	0.0219
		Last Day	0.0233	0.1392	0.1412
GARCH(1,1)	$\rho(20) = 0.0274$	Boots	0.0480	0.0077	0.0486
		Wild Boots	0.0484	0.0031	0.0485
		Last Day	0.0522	0.1414	0.1508
	$\rho(20) = 0.0676$	Boots	0.0881	0.0081	0.0884
		Wild Boots	0.0886	0.0033	0.0887
		Last Day	0.0866	0.1467	0.1703
	$\rho(20) = 0.1132$	Boots	0.1341	0.0083	0.1343
		Wild Boots	0.1340	0.0034	0.1340
		Last Day	0.1370	0.1546	0.2066
FIGARCH	$\rho(20) = 0.1214$	Boots	-0.0146	0.0109	0.0182
		Wild Boots	-0.0366	0.0277	0.0459
		Last Day	-0.0141	0.0282	0.0315

Table 4.7: Estimated ACFs of randomly generated samples from different models under the assumption that there are 60 months and 20 days in each month B=2000.

Models	Correlation	Methods	Bias	SD	RMSE
AR(1)	$\rho(20) = 0.8179$	Boots	0.0758	0.0874	0.1156
		Wild Boots	0.8339	0.0041	0.8339
		Last Day	0.0794	0.0941	0.1231
	$\rho(20) = 0.3585$	Boots	0.0114	0.0903	0.0910
		Wild Boots	0.3750	0.0033	0.3749
		Last Day	0.0407	0.1210	0.1276
	$\rho(20) = 0.1216$	Boots	-0.0224	0.0684	0.0720
		Wild Boots	0.1382	0.0023	0.1383
		Last Day	0.0295	0.1231	0.1266
ARMA(1,1)	$\rho(20) = 0.1237$	Boots	-0.0266	0.0695	0.0744
		Wild Boots	0.1406	0.0031	0.1406
		Last Day	0.0235	0.1189	0.1212
	$\rho(20) = 0.0123$	Boots	-0.0207	0.0461	0.0505
		Wild Boots	0.0290	.0029	0.0291
		Last Day	0.0217	0.1247	0.1266
	$\rho(20) = 0.0009$	Boots	-0.0025	0.0334	0.0335
		Wild Boots	0.0175	0.0028	0.0177
		Last Day	0.0168	0.1274	0.1285
GARCH(1,1)	$\rho(20) = 0.0274$	Boots	0.0436	0.0070	0.0442
		Wild Boots	0.0440	0.0029	0.0441
		Last Day	0.0439	0.1258	0.1332
	$\rho(20) = 0.0676$	Boots	0.0842	0.0071	0.0845
		Wild Boots	0.0841	0.0030	0.0842
		Last Day	0.0782	0.1335	0.1547
	$\rho(20) = 0.1132$	Boots	0.1297	0.0075	0.1299
		Wild Boots	0.1298	0.0032	0.1299
		Last Day	0.1280	0.1453	0.1936
FIGARCH	$\rho(20) = 0.1259$	Boots	-0.0114	0.0098	0.0150
		Wild Boots	-0.0262	0.0221	0.0343
		Last Day	-0.0111	0.0257	0.0280

Table 4.8: Estimated ACFs of randomly generated samples from different models under the assumption that there are 72 months and 20 days in each month B=2000.

Models	Correlation	Methods	Bias	SD	RMSE
AR(1)	$\rho(20) = 0.8179$	Boots	0.0597	0.0726	0.0940
		Wild Boots	0.8314	0.0038	0.8314
		Last Day	0.0638	0.0790	0.1016
	$\rho(20) = 0.3585$	Boots	0.0029	0.0807	0.0807
		Wild Boots	0.3721	0.0031	0.3721
		Last Day	0.0341	0.1074	0.1127
	$\rho(20) = 0.1216$	Boots	-0.0275	0.0647	0.0703
		Wild Boots	0.1354	0.0027	0.1354
		Last Day	0.0218	0.1175	0.1195
ARMA(1,1)	$\rho(20) = 0.1237$	Boots	-0.0298	0.0663	0.0727
		Wild Boots	0.1374	0.0028	0.1374
		Last Day	0.0199	0.1084	0.1103
	$\rho(20) = 0.0123$	Boots	-0.0237	0.0465	0.0488
		Wild Boots	0.0263	0.0025	0.0264
		Last Day	0.0186	0.1114	0.1130
	$\rho(20) = 0.0009$	Boots	-0.0061	0.0310	0.0316
		Wild Boots	0.0148	0.0026	0.0151
		Last Day	0.0142	0.1143	0.1152
GARCH(1,1)	$\rho(20) = 0.0274$	Boots	0.0412	0.0063	0.0416
		Wild Boots	0.0414	0.0027	0.0415
		Last Day	0.0437	0.1160	0.1239
	$\rho(20) = 0.0676$	Boots	0.0815	0.0065	0.0818
		Wild Boots	0.0816	0.0030	0.0817
		Last Day	0.0780	0.1213	0.1443
	$\rho(20) = 0.1132$	Boots	0.1271	0.0071	0.1273
		Wild Boots	0.1270	0.0029	0.1270
		Last Day	0.1270	0.1217	0.1759
FIGARCH	$\rho(20) = 0.1237$	Boots	-0.0143	0.0091	0.0170
		Wild Boots	-0.0304	0.0225	0.0378
		Last Day	-0.0150	0.0252	0.0287

Chapter 5

Conclusion

In this thesis, we aim to find a better way of measuring the serial correlation of time series objects by comparing the bootstrap and wild bootstrap methods with the last day method. We apply our idea in the following way:

1. Simulate time series data under different models, AR, ARMA,GARCH and FIGARCH.

2. Resample the data from each time series object by applying bootstrap, wild bootstrap and last day value approaches.

3. Calculate the correlation of the sample data and take the average for the bootstrap method.

4. Calculate bias, standard deviation and root mean square error of each method.

We observe that the standard bootstrap method is overall better than the last day value approach in all settings. The wild bootstrap method can be better under very specific situations, but it not as good as both the bootstrap and last day value methods. This conclusion is consistent and not affected by bootstrap sampling size and original sample size.

For further study, we plan to improve the wild bootstrap method by applying different distributions or by using designed sampling rather than random sampling.

Bibliography

Bahram Adrangi, Arjun Chatrath, and Kambiz Raffiee. Volatility characteristics and persistence in latin american emerging markets. *International Journal of Business*, 4(1):19–37, 1999.

Niklas Ahlgren and Paul Catani. Wild bootstrap tests for autocorrelation in vector autoregressive models. *Statistical Papers*, pages 1–28, 2012.

Richard T Baillie, Tim Bollerslev, and Hans Ole Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 74(1):3–30, 1996.

Jeremy Berkowitz and Lutz Kilian. Recent developments in bootstrapping time series. *Econometric Reviews*, 19(1):1–48, 2000.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, pages 307–327, 1986.

Tim Bollerslev, Ray Y Chou, and Kenneth F Kroner. Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1-2):5–59, 1992.

Vincent J Carey. Resampling methods for dependent data, 2005.

Thomas C Chiang and Shuh-Chyi Doong. Empirical analysis of stock returns and volatility: Evidence from seven asian stock markets based on tar-garch model. *Review of Quantitative Finance and Accounting*, 17(3):301–318, 2001.

Russell Davidson and Emmanuel Flachaire. The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1):162–169, 2008.

Anthony Christopher Davison and David Victor Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.

B Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, pages 1–26, 1979.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

Robert F Engle and Tim Bollerslev. Modelling the persistence of conditional variances. *Econometric reviews*, 5(1):1–50, 1986.

Silvia Gonçalves and Lutz Kilian. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 123(1):89–120, 2004.

Wolfgang Härdle, Joel Horowitz, and Jens-Peter Kreiss. Bootstrap methods for time series. *International Statistical Review*, 71(2):435–459, 2003.

GS Hongyi Li and Maddala. Bootstrapping time series models. *Econometric reviews*, 15(2): 115–158, 1996.

Taewook Lee and Cheolwoo Park. Tests for serial correlation in mean and variance of a sequence of time series objects. *Journal of Statistical Computation and Simulation*, 87(3): 478–492, 2017.

Enno Mammen. Bootstrap and wild bootstrap for high dimensional linear models. *The annals of statistics*, pages 255–285, 1993.

Dimitris N Politis et al. The impact of bootstrap methods on time series analysis. *Statistical Science*, 18(2):219–230, 2003.

Dongsheng Tu and Jun Shao. The jackknife and bootstrap. *Springer Series in Statistics, New York*, 85(486-492):8, 1995.

Chien-Fu Jeff Wu. Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics*, pages 1261–1295, 1986.

Appendix

0.1 Algorithm for Bootstrap of time series objects and calculate correlation values

In order to generate time series data, we need R packages "rugarch" and "FGN".

```
R=1000; ## Number of repetitions

B1=1000; ## Number of bootstrap sample

t=36; ## Time duration:

n=20; ## Number of intraday data

######################################################################

w <- c(-1,1)  ## Rademacher distribution

rho.boot=seq(1:B1)

rho.w.boot=seq(1:B1)

rho.last=seq(1:R)

rho.boot.mean=seq(1:R)

rho.w.boot.mean=seq(1:R)

corr.true=seq(1:t)

for(iter in 1:R){

x<- list()

###### AR models

phi=0.9
```

```
y<-arima.sim(model=list(ar=phi),n=n*t)
corr=ARMAacf(ar=(phi),lag=20)[21]
x <- lapply(1:t, function(z) y[(n*(z-1)+1) : (n*z)])
## bootstrap samples
boot <- lapply(1:B1, function(j) unlist(lapply(1:t, function(z) sample(x[[z]],1,replace=
## Wild bootstrap samples
w.boot <- lapply(1:B1, function(j) unlist(lapply(1:t, function(z) sample(x[[z]],1,replac
## last day value
newx=seq(1:t)
for (i in 1:t) {
newx[i]=x[[i]][n]
}
rho.last[iter]=acf(newx,type="correlation",plot=F)$acf[2]
for(i in 1:B1){
rho.boot[i]=acf(boot[[i]],type="correlation",plot=F)$acf[2]
rho.w.boot[i]=acf(w.boot[[i]],type="correlation",plot=F)$acf[2]
}
rho.boot.mean[iter]=mean(rho.boot)
rho.w.boot.mean[iter]=mean(rho.w.boot)
}
## calcualte bias
bias.boot=corr-mean(rho.boot.mean)
bias.w.boot=corr-mean(rho.w.boot.mean)
bias.last=corr-mean(rho.last)
## calcuate standard deviation
sd.boot=sd(rho.boot.mean)
sd.w.boot=sd(rho.w.boot.mean)
```

```
sd.last=sd(rho.last)
## calcuate RMSE
rmse.boot=sqrt(bias.boot^2+sd.boot^2)
rmse.w.boot=sqrt(bias.w.boot^2+sd.w.boot^2)
rmse.last=sqrt(bias.last^2+sd.last^2)
output=matrix(nrow=3,ncol=3)
output[1,]=c(bias.boot,sd.boot,rmse.boot)
output[2,]=c(bias.w.boot,sd.w.boot,rmse.w.boot)
output[3,]=c(bias.last,sd.last,rmse.last)
colnames(output)=c("Bias","SD","RMSE")
rownames(output)=c("Boots","Wildboots","Last value")

output;corr
```

0.2 Substitute time series objects

```
###### ARMA models
y=arima.sim(model=list(ar=0.7,ma=0.3),n=n*t)
corr=ARMAacf(ar=0.7,ma=0.3,lag=20)[21]

x <- lapply(1:t, function(z) y[(n*(z-1)+1) : (n*z)])

###### GARCH models
omega=10^-6; alpha=0.1; beta=0.85
spec <- ugarchspec(variance.model=list(garchOrder=c(1,1)),mean.model=list(armaOrder=c(0,
distribution.model="norm")
```

```
y <- ugarchpath(spec, n.sim=n*t)@path$seriesSim

rho=(alpha*(1-alpha*beta-beta^2))/(1-2*alpha*beta-beta^2)

corr=((alpha+beta)^19)*rho

x <- lapply(1:t, function(z) y[(n*(z-1)+1) : (n*z)])


##### FIGARCH models

H=0.8

y=SimulateFGN(n*t,H)

corr=(var(y)/2)*(19^(2*H)-2*(20^(2*H))+21^(2*H))


x <- lapply(1:t, function(z) y[(n*(z-1)+1) : (n*z)])
```