Stony Brook University

APPROXIMATION OF LONG MEMORY PROCESS BY SHORT MEMORY PROCESS - with Application to Option Valuation

A Dissertation Presented

by

Pengyuan Shao

to

The Graduate School

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in

Applied Mathematics & Statistics

Quantitative Finance

Stony Brook University

December 2013

Stony Brook University

The Graduate School

Pengyuan Shao

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Andrew Mullhaupt – Dissertation Advisor

Professor

Dept. of Applied Mathematics and Statistics, SUNY Stony Brook

Svetlozar Rachev – Chairperson of Defense

Professor

Dept. of Applied Mathematics and Statistics, SUNY Stony Brook

Brent Lindquist – Committee Member

Professor and Chair

Dept. of Applied Mathematics and Statistics, SUNY Stony Brook

Christopher Bishop – External Committee Member

Professor

Dept. of Mathematics, SUNY Stony Brook

Young Shin Aaron Kim – External Committee Member

Assistant Professor

College of Business, SUNY Stony Brook

This dissertation is accepted by the Graduate School

Charles Taber

Interim Dean of the Graduate School

Abstract of the Dissertation

APPROXIMATION OF LONG MEMORY PROCESS BY SHORT MEMORY PROCESS - with Application to Option Valuation

by

Pengyuan Shao

Doctor of Philosophy

in

Applied Mathematics and Statistics

Quantitative Finance

Stony Brook University

2013

Options on an asset which follow a long memory process are difficult to value by conventional methods, due to the existence of arbitrage opportunities. Here we show how to avoid the problem of arbitrage opportunities and value vanilla European options when underlying asset returns follow a FARIMA(p,d,q) processes with $d > 0$ which is widely used as an model of long memory price processes.

We use information distance to prove that stationary ARMA processes are dense in all FARIMA processes in the total variation distance. As a consequence, statistical tests with finite sample size fail to distinguish a FARIMA process from ARMA processes. As option values are a special case of statistical test, the well understood option values for a sufficiently close stationary ARMA process can be taken as option values for the FARIMA process, with very low probability of error. We provide Monte Carlo experiments that confirm that long memory processes are not easily distinguished from our approximate ARMA processes with finite sample sizes

using a variety of well known statistical tests. We examine how long memory affects the option values and implied volatility surface. Finally we examine high frequency data for equities and spot foreign exchange rates for evidence of long memory effects.

Contents

List of Tables

List of Figures

Acknowledgements

In my three and half years doctoral study, I have received enormous help and support from many people inside and outside Stony Brook University. I really want to thank all of them before the starting of this thesis.

Firstly, I would like to thank Dr. Ann Tucker, who admitted me to the Quantitative Finance program at Stony Brook University and gave me a lot of help since I arrived in United States. It was also Ann that brought me into the real finance world from school in the winter of 2010. Without her help, I could not adjust to the new environment so quickly and realize how much I enjoy the work I am doing right now.

I would like to thank my advisors, Prof. Andrew Mullhaupt and Prof. Svetlozar Rachev. Honestly I was shocked when they gave me this thesis topic since it is such a challenging topic. Beyond the knowledge I have learned from them, Prof. Mullhaupt brought me into the HFT world and instructed me on how to apply mathematics on systematic trading; and Prof. Rachev taught me about the entrepreneurship and "how to think big", which will play an important role in my long term career. Besides, both of them taught me the importance of balanced life more than a thousand times and I am really getting into it now. I would also like to thank Prof. John Pinezich's help on reviewing my working papers and thesis. Also I would like to thank Prof. Frey. Without Prof. Frey's efforts, there would not be such a unique and creative Quantitative Finance program as there is now.

I would like to thank Greg Frank, Joseph Haggenmiller and Kendal Beer at Presagium LLC. We have been working together since Nov. 2010 and I have learned a lot from their invaluable experience in trading, firm construction etc. We have been working very closely on building the systematic foreign exchange rates trading system in the last three years and also gone through many happiness and sadness.

I would also like to thank the help from my friends and the time. Without them, I could not accomplish the balanced life and enjoy living in long island so much.

Specifically, I would like to thank one PhD student Xiang Shi's help on proof of one of the Lemmas.

Last but most, I very appreciate the help and support from my parents, without whom I would not succeed in any accomplishments. It is also them that teach me how to get through various difficult situations and appreciate these difficulties.

1 Introduction

When the underlying asset returns allow long memory or long range dependence exist, it is difficult to have a closed-form option pricing formula. Instead of using geometric Brownian motion to model the asset prices, fractional Brownian motion has been proposed to model the asset prices while capture the existence of long memory processes. However, there exists the arbitrage opportunities by using fractional Brownian motion model the asset prices, Rogers [59] show it by constructed such an arbitrage. FARIMA models are able to model the fractional Gaussian noises, which are the unit increments of fractional Brownian motion. To solve this problem, we use short memory ARMA processes to approximate the FARIMA processes valuate the European vanilla option based on the approximated ARMA processes. We show that FARIMA and approximated ARMA processes are indistinguishable with finite sample size, which means the option valuations are indistinguishable in discrete case. So the main concept of this thesis is how to valuate option while underlying asset returns are following long memory processes.

In this first chapter of the dissertation, we give a brief introduction to long memory processes and some pioneering works concerning the existence and modeling of long memory processes in financial time series. Also we review the previous works on the connection between long memory processes and other non-long memory stochastic processes, which can pass statistical tests for long memory.

1.1 Dissertation Organization

We briefly describe the organization of dissertation here:

In Chapter 1, "Introduction", we review long memory processes and their application to financial time series; the connection between long memory processes and other stochastic processes. We state the main results of this work.

In Chapter 2, "Distance between distributions", we propose to use information distance to measure the relation between different distributions and processes. We

review the connection among total variations, Hellinger distance and information distance to make the connection clear to statistical inference. Two processes close in information distance, then they are close in Hellinger distance and total variation.

In Chapter 3, "Fractional autoregressive integrated moving average model", we give the information geometry of the classical econometric model FARIMA models, which are widely used to model long memory processes. We use short memory process ARMA process to approximate the FARIMA model with long memory property. Also we use the connection proposed in chapter 2 to state the long memory FARIMA model can not be distinguished from the short memory ARMA processes by statistical inference with finite sample size.

In Chapter 4, "Simulation", we present our connection between long and short memory processes numerically. Based on different classical long memory statistical tests and the Monte Carlo evidences, we show that long memory processes cannot be distinguished from an approximated short memory processes with finite sample size. Also we show the connection between statistical inference and information distance numerical in this section.

In Chapter 5, "Option pricing with long memory processes", we valuate the European vanilla options based on Black-Scholes-ARMA model by approximating the long memory processes by stationary ARMA processes. Also we show that how the long memory processes affect the option valuation and the implied volatility calculated in Black-Scholes model. Also we show the change of term structure of volatility by adding a short-term volatility model to the long memory model.

In Chapter 6, "Empirical finance", we give an empirical investigation to show whether long memory process exists at high frequency financial time series firstly, which is hard to find in the literatures.

In Chapter 7, "Conclusion", we draw our conclusion of this dissertation and a brief summary of the problems we solved.

1.2 Long memory process

Whether a stochastic process has long memory or short memory is usually defined based on its autocovariance functions. Short memory stationary process is usually referring to the stochastic process whose autocovariance functions decay fast and its spectral density is bounded everywhere. One of the most widely applied time series model with short memory is the stationary autoregressive moving average model (ARMA), which we will describe more and make the connection between long memory process later in this dissertation. Compared to short memory process, long memory process is the stochastic process with slowly decaying autocovariance functions, which is not integrable or summable (see Mcleod and Hippel [14] in 1978). Equivalently, spectrum density of long memory process is unbounded.

Then the definition of the long memory is:

Definition 1. *A time series r_t, whose $E(x_t) = \mu$ and $Cov(x_t, x_{t+j}) = \gamma_j$ do not depend on time t, is said to be a covariance stationary time series [16]. Given the autocovariance functions γ_k, autocorrelation function ρ_k is defined as $\rho_k = \gamma_k/\gamma_0$. Then the time series r_t is a long memory process when the absolute value of autocorrelation functions is not finite:*

$$\lim_{T\to\infty} \sum_{k=0}^{T} |\rho_k| = \infty \tag{1}$$

1.3 Hurst exponent

Hurst exponent is a widely used measure of the long memory in observable time series, which is also related to the fractional differencing parameter in FARIMA model. The classical range statistics (R/S) was developed by Hurst [26]:

$$R/S(n) = \frac{1}{S(n)} \left[\max_{0\le t\le n} \left(Y(t) - \frac{t}{n} Y(n) \right) - \min_{0\le t\le n} \left(Y(t) - \frac{t}{n} Y(n) \right) \right], n \ge 1 \tag{2}$$

where $S(n)$ is the standard deviation estimator and $Y(n)$ is the sample returns.

$$S(n) = \sqrt{\frac{1}{n} \sum_{0 \leq t \leq n} \left(Y(t) - \bar{Y}_n\right)^2} \tag{3}$$

The relation between Hurst exponent and the R/S statistics is

$$Q(n) = R/S(n) = an^H \tag{4}$$

where H is the estimated Hurst exponent. Traditionally, processes with long memory are identified with Hurst exponents $\frac{1}{2} < H < 1$.

Lo [6] shows that the normalized R/S statistic, $V_c = \frac{1}{\sqrt{n}} Q(n)$, has the asymptotic distribution as follows

$$F_V(v) \sim 1 + 2 \sum_{k=1}^{\infty} \left(1 - 4k^2 \nu\right) e^{-2(kv)^2} \tag{5}$$

Also Lo [6] states that the classic R/S statistic could be biased caused by the short term memory and he proposes the modified R/S statistic $\tilde{Q}(n, q)$ in the following form:

$$\tilde{Q}(n, q) = R(n) / \tilde{S}(n, q) \tag{6}$$

where

$$\tilde{S}(n, q) = S(n) + 2 \sum_{i=1}^{q} \omega_i(q) \hat{\gamma}_i, \ \omega_i(q) = 1 - \frac{i}{q+1}, q < n \tag{7}$$

where $\hat{\gamma}_i$ are the sample estimated autocorrelations and we denote the normalized modified rescaled range statistics as

$$V(q) = \frac{1}{\sqrt{n}} \tilde{Q}(n, q) \tag{8}$$

Modified rescaled range statistics will exclude part of long term dependence implied by short term dependence. However, the shortcoming of modified R/S statistic is the need to choose parameter lag q, difficult to do robustly [6]. Lo [4] gives the fractiles of the distribution $F_V(v)$, which will be used to implement the statistical test in the empirical finance section.

Also the well-known fractional Brownian motion is defined based on the Hurst exponent by Mandelbrot [60]:

$$B_H(t) = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)} \int_{-\infty}^{t} (t-\tau)^{H-\frac{1}{2}} \, dB(\tau) \tag{9}$$

where $\Gamma(\cdot)$ is the gamma function and H is the Hurst parameter, $0 < H < 1$.

From the definition, we can also obtain the covariance function for fractional Brownian motion described by Hurst exponent H,

$$R_H(t,s) = E\left[B_H(t) B_H(s)\right] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right) \tag{10}$$

$$B_H(t) = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)} \int_{0}^{t} (t-s)^{H-\frac{1}{2}} \, dB(s) \tag{11}$$

We will show the connection between Hurst exponent and fractional differencing parameter later and use Hurst exponent in the Chapter "Empirical finance".

1.4 Long memory in financial time series

The existence of long memory process in financial time series has been discussed by numerous paper for a long time. We briefly review some of the pioneering works and important literatures here. The main discussions can be classified by return series and volatility series

One of the most widely used detection of long memory process is rescaled range analysis, R/S, which is proposed by Mandelbrot [5] in 1972. The reason we mention R/S first is that many pioneering researches are based on R/S or the modified R/S in the following years. Lo [6] in 1989, proposed the modified R/S analysis and found little evidence of long memory in the U.S. stock market returns by this measurement. Nawrocki [8] in 1995 claimed that there existed the long memory from the view of R/S statistics and suggest that the dependence raised from general economic cycle.

in 1996, Chow et al. [9] did the research on 22 international equity market indexes based on modified R/S test and rescaled variance ratio test. They found evidence against the random walk hypothesis is very weak, in agreement with Lo's result [6]. Mandelbrot [4] in 1997 claimed that there existed the long memory phenomenon in stock prices. In 1998, Lobato and Savin [10] analyzed the subperiods of returns and individual stocks in U.S. stock market and found no long memory in these returns. Willinger et al. [11] in 1999 argued that Lo [6] had a bias to accept the null hypothesis and they used the CRSP[1] daily data to show empirical evidence of long memory, but with low Hurst exponent (around 0.6). In 2001, Sadique and Silvapulle [12] found long memory, suggesting that some countries are not efficient markets. In currency markets, Cheung and Lai [13] in 2001 gave an alternative explanation to puzzling behavior of yen exchange rate based on long memory.

Compared to debate of the long memory process existence in financial returns, volatility series has been shown to have more pronounced long memory process by numerous literatures. Ding et al. [7] in 1993 did the research on the absolute return $|r_t|^d$, and claimed that this power transformation had very high autocorrelation for long lags. Bollerslev and Mikkelsen [19] in 1996 show the long memory process existence in U.S. stock market volatility and proposed to use fractionally integrated EGARCH model for characterizing the long memory. Breidt et al. [18] in 1998 show the empirical evidence of the long memory existence in an extensive set of U.S. stock return indexes and proposed the long memory stochastic volatility (LMSV) model, which incorporates FARIMA process in a standard stochastic scheme, to model the value-weighted CRSP market index. In 1998, Lobato and Savin [10] show that squared returns have strong long memory in individual stock returns.

Besides the financial returns and volatility series, long memory process has also been investigated in other parts of financial markets. For example, Baillie and

[1] CRSP: Center for Research in Security Prices is one of the major providers of historical stock market data. CRSP covers the U.S. traded stock data back to 1926. For more information, readers are referred to: http://www.crsp.chicagobooth.edu/index.html.

Bollerslev [23] in 1994 show that in foreign exchange market, the forward premium had long memory. Bouchaud et al. [21] reported the long memory effect in market order flow in the Paris Stock Exchange in 2004.

From the perspective of classical financial theory, the main reason that we want to know whether a financial time series is long memory or short memory process is related to the efficient market hypothesis (EMH). The efficient market hypothesis has a long history in finance and the existence long memory process means that the information from the distant past can still affect current value significantly, which would cast doubt on EMH.

Also in the Black-Scholes option pricing model, one important assumption is that stock returns will follow the exponential Brownian motion. Pricing financial derivatives using martingales does not deal with the long memory property easily. CAPM, or arbitrage pricing theory also has problems with long memory process.

1.5 Long memory process and other stochastic processes

There are many theoretical literatures discussing about whether non-long memory processes can present the properties of long memory process, which they call spurious long memory process. In 1974, Klemeš [24] show that the infinite memory of a process could be caused by the nonstationarity. Granger [1] in 1980 proved that some specific aggregation of AR models could have long memory properties. In 1997, Taqqu et al. [3] found that the aggregation of stationary process with superposition of many ON/OFF was converging to fractional Brownian motion with long memory property.

Diebold and Inoue [2] in 2001 show that, various structural break models could be confused with long memory process. However, most of these "analytical connections" have very strong assumptions and we can construct the spurious long memory processes with these assumptions as many as we want.

In reality, when we have finite samples instead of infinite length, the analytical

difference between the spurious and real ones might not be too important. For example, Diebold and Inoue [2] did the statistical tests on simulation from short-memory process and concluded that it was hard to clear the confusion between long memory process and spurious one given small number of samples. Specifically in their empirical part, he show the GPH result with a static markov-switching model, which is a short memory process. He said that when sample size was not large (5,000 here), the test would reject the process is a short memory process. But when sample size is large, this kind of fake process eventually wont be rejected as a short memory process. If the null hypothesis is $d = 0$ for $I(d)$ process, finding the relation between the rate of rejection and sample size is important for identifying long memory process with finite sample size. We will state this issue in this dissertation.

1.6 Statistical tests

We briefly describe the statistical tests we will use to compare two distributions. Since the connection between total variation and information distance is bridged, these statistical tests are also used to examine the effectiveness of information distance numerically.

1.6.1 Kolmogorov-Smirnov test

Kolmogorov-Smirnov statistic is a supremum class statistic to determine if two empirical distribution differ significantly and it. The statistic is based on the maximum absolute distance between two empirical distribution function, whose definition is as follows:

Definition 2. *Kolmogorov-Smirnov statistic of two samples:*

$$D_n = \sup_x |F_1(x) - F_2(x)| \tag{12}$$

where $F_1(x)$ and $F_2(x)$ are the empirical distribution functions of the first and second sample respectively.

1.6.2 Anderson-Darling Test

Anderson-Darling is a statistical test to tell whether a group of data are coming from a given probability distribution

$$n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{[F(x)(1 - F(x))]} dF(x) \tag{13}$$

where $F_n(x)$ is the empirical cumulative distribution and $F(x)$ is the hypothesized distribution.

However, we want to compare the empirical distributions from two samples rather than a given probability distribution, we apply the methodology proposed by Scholz and Stephens [29]. To simplify the problem, we get the two sample Anderson-Darling statistic from the k-sample Anderson-Darling test statistics [29]:

$$A_{2N}^2 = \frac{1}{N} \sum_{i=1}^{2} \frac{1}{n_i} \sum_{j=1}^{N-1} \frac{(NM_{ij} - jn_i)^2}{j(N - j)} \tag{14}$$

where M_{ij} is the i_{th} sample which are not greater than Z_j and N is total observations number of two samples. As for the Z_j, Z_j is the ordered sample in the pool of both sample observations, $Z_1 < Z_2 < \cdots < Z_N$.

As for the variance formula of A_{kN}^2, Scholz and Stephens [29] give the following formula

$$
\begin{aligned}
\sigma_N^2 &= \frac{aN^3 + bN^2 + cN + d}{(N-1)(N-2)(N-3)} \\
a &= (4g-6)(k-1) + (10-6g)H \\
b &= (2g-4)k^2 + 8hk + (2g-14h-4)H - 8h + 4g - 6 \\
c &= (6h+2g-2)k^2 + (4h-4g+6)k + (2h-6)H + 4h \\
d &= (2h+6)k^2 - 4hk \\
H &= \sum_{i=1}^{k} \frac{1}{n_i} \\
h &= \sum_{i=1}^{N-1} \frac{1}{i} \\
g &= \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \frac{1}{(N-i)\,j}
\end{aligned}
\tag{15}
$$

The Anderson-Darling test statistic is standardized as follows:

$$
T_{kN} = \frac{A_{kN}^2 - (k-1)}{\sigma_N} \tag{16}
$$

The Table 1 is the critical points for different significant levels [29]:

α	.25	.10	.05	.025	.01
Critical Point	.326	1.225	1.960	2.719	3.752

Table 1: Critical points for different significant levels of two samples Anderson-Darling test

2 Total variation, Hellinger and information distances

In this section, we briefly review the well known relationship between the total variations and Hellinger distances [34] and the relationship between Hellinger and information distances [33].

2.1 Total variation and Hellinger distances

In statistical hypothesis test, there are a hypotheses P_0 and alternative P_1, where P_0 and P_1 are probability measures. There are two errors of the statistical test: one is to reject P_0 while P_0 is true and the other one is to accept P_1 while P_1 is false. The sum of these two error probabilities of test ϕ can be written as [34]:

$$S_{P_0,P_1}(\phi) = \int_{\chi} \phi(x)\, dP_0(x) + \int_{\chi} (1 - \phi(x))\, dP_1(x) \tag{17}$$

So the best statistical hypothesis test is the test ϕ^* , which minimizes the sum of two errors probabilities (eq. 17), and the minimized errors probability can be represented by the total variation distance $\|P_0 - P_1\|_1$ (see Lehamann and Romano [34] in details):

$$S_{P_0,P_1}(\phi^*) = 1 - \frac{1}{2}\|P_0 - P_1\|_1 \tag{18}$$

Definition 3. *The total variation distance between P_0 and P_1 is defined as (see Lehmann and Romano [34]):*

$$\|P_0 - P_1\|_1 = \int |p_1 - p_0|\, d\mu \tag{19}$$

where p_i is the density of P_i with respect to any measure μ dominating both P_0 and P_1.

The connection between statistical error probabilities and total variation distance (eq. 18) shows that the smaller the total variation distance is, the worse the optimal statistical test is. For example, if total variation distance between P_0 and

P_1 is small enough, which means that the density functions of P_0 and P_1 are close enough, then we have $S_{P_0,P_1}(\phi^*)$ to be close to 1, which means the sum of the error probabilities of the optimal statistical test for P_0 and P_1 is close to 1.

Total variation distance is very hard to calculate based on its definition, but there is the well known relationship between total variation and Hellinger distances, which has the following definition:

Definition 4. *The square of the Hellinger distance $H(P_0, P_1)$ between two probability distributions P_0 and P_1 can be defined as [34]:*

$$H^2(P_0, P_1) = \frac{1}{2}\int_{\chi}\left[\sqrt{p_1(x)} - \sqrt{p_0(x)}\right]^2 d\mu(x) \tag{20}$$

where p_i is the density of P_i with respect to any measure μ dominating both P_0 and P_1.

Theorem 1. *There exists the following relationship between total variation and Hellinger distances [34]:*

$$H^2(P_0, P_1) \leq \frac{1}{2}\|P_0 - P_1\|_1 \leq H(P_0, P_1)\left[2 - H^2(P_0, P_1)\right]^{1/2} \tag{21}$$

2.2 Hellinger and information distances

Mullhaupt [33] shows that Hellinger distance is bounded by the information distance. Before going into the relationship between Hellinger distance and information distance, we review the Fisher information matrix. The information distance is defined on a family of distributions and we assume that the distributions of the family have densities $p(x, \theta)$ parameterized by θ.

Definition 5. *The Fisher information matrix is defined as [33]:*

$$F(\theta) = E\left(\left(\partial_\theta \log p(x,\theta)\right)\left(\partial_\theta \log p(x,\theta)\right)^T |\theta\right) \tag{22}$$

$$= \int p(x,\theta)\left(\partial_\theta \log p(x,\theta)\right)\left(\partial_\theta \log p(x,\theta)\right)^T d\mu(x) \tag{23}$$

Definition 6. *The information distance $I(\theta_0, \theta_1)$ is the Riemannian distance defined by choosing the Fisher information matrix as the metric tensor and it is defined by*

$$\begin{aligned} I(\theta_0, \theta_1) &= \min \int_0^1 \left(\dot{\theta}(t)^T F(\theta(t)) \dot{\theta}(t) \right)^{1/2} dt \\ s.t. \theta(0) &= \theta_0,\ \theta(1) = \theta_1 \ \textit{and}\ \theta(t) \in C[0,1] \end{aligned} \tag{24}$$

Mullhaupt [33] shows that the information distance multiplied by a constant is a upper bound for Hellinger distance. The proof is not given here, it can be found in Mullhaupt [33].

Theorem 2. *The Hellinger distance between two probability distributions P_0 and P_1 is bounded by:*

$$H(P_0, P_1) \leq \frac{1}{\sqrt{8}} I(P_0, P_1) \tag{25}$$

Using transfer function instead of the density function, Mullhaupt [39] shows that the Fisher information matrix element of a minimum phase transfer function $f(z)$ is

$$\frac{1}{2\pi i} \int_{|z|=1} (\partial_u \log f(z)) (\partial_v \log f(z))^* \frac{dz}{z} \tag{26}$$

The logarithm of the transfer function can be parameterized in the form

$$\log f(z) = a_0 + a_1 z + a_2 z^2 + \ldots$$

and the Fisher information matrix elements (j, k) are

$$\frac{1}{2\pi i} \int_{|z|=1} \left(\partial_{a_j} \log f(z) \right) (\partial_{a_k} \log f(z))^* \frac{dz}{z} = \delta_{jk} \tag{27}$$

which shows that the information distance between white noise and any stochastic processes with minimum phase transfer function is the norm of the Hardy space of the disc $H^2(D)$:

$$I(f, 1) = \|\log f\|_{H^2}^2 = \sum_{k=0}^{\infty} |a_k|^2 \tag{28}$$

The information distance between two transfer functions $f(z)$, $g(z)$ with minimum phase can be written as

$$I(f(z), g(z)) = I\left(\frac{f(z)}{g(z)}, 1\right) = \left\|\log\left(\frac{f(z)}{g(z)}\right)\right\|_{H^2}^2 = \|\log f(z) - \log g(z)\|_{H^2}^2 \quad (29)$$

2.3 Statistical inference and information distance

After showing the relationship between total variation distance, Hellinger distance and information distance, we can also get the bounded relationship between total variation distance and information distance.

We know that $2H(P_0, P_1)\left[2 - H^2(P_0, P_1)\right]^{1/2}$ is a nondecreasing function of $H(P_0, P_1)$ when $0 < H(P_0, P_1) < 1$. Then we have

$$\begin{aligned} \|P_0 - P_1\|_1 &\leq 2H(P_0, P_1)\left[2 - H^2(P_0, P_1)\right]^{1/2} \\ &\leq \frac{1}{\sqrt{2}} I(P_0, P_1)\left[2 - \frac{1}{8} I^2(P_0, P_1)\right]^{1/2} \quad (30) \\ &\leq I(P_0, P_1) \quad (31) \end{aligned}$$

We can see that when the total variation is bounded by the information distance.

2.4 Connection with option valuation

The total variation distance controls the error of statistical tests based on bounded measurable functions; however option values are based on payouts that are not bounded, even though the value (discounted expected payout) may still be bounded. We do not know *a priori* that the processes we study will have bounded option values, but it is very unlikely that processes for which the value of a simple call option is infinite will have sensible values. We give here proofs of the intuitively clear results that if any call option has a finite value, then all call options have finite values, and call options decay to zero as the strike increases to infinity. This allows us to approximate the value any finite basket of options with as closely as we like with a finite basket of options with a bounded payout. This allows us to apply the theory of bounded statistical tests to option values.

Lemma 1. *If there exists a finite strike price K such that the vanilla European call option with strike price K has finite value, then the value of vanilla European call options of all strikes have finite values, and the value C of the call option goes to* 0 *as the strike price goes to infinity.*

Proof: *Let X denote the underlying asset price and the value of the vanilla European call option with strike K is $C(S_0, K, T) = e^{-rT} E(S_T | S_T \geq K)$. We have*

$$C(S_0, K, T) \leq C(S_0, 0, T) \leq K + C(S_0, K, T)$$

which proves the first claim, since

$$\begin{aligned} e^{-rT} E(S_T | S_T \geq K) &= e^{-rT} [E(S_T) - E(S_T | S_T < K)] \\ &= e^{-rT} [E(S_T) - E(S_T \cdot 1(S_T < K))] \end{aligned} \tag{32}$$

then the monotone convergence theorem shows

$$\lim_{K \to \infty} E(S_T \cdot 1(S_T < K)) = E(S_T) \tag{33}$$

so we have

$$\lim_{K \to \infty} C(S_0, K, T) = e^{-rT} [E(S_T) - E(S_T)] = 0 \tag{34}$$

Lemma 2. *The value of any basket of finitely many European vanilla options can be approximated as closely as desired by the value of an options basket with bounded payout.*

Proof: *The slope of the payout of a basket of finitely many European vanilla options is a constant for prices above the highest strike in the basket. We buy or sell a call in the appropriate amount so that the call payout has the opposite slope as the basket slope for high prices. The payout resulting from adding the call to the basket is bounded since the slope for high prices is zero. By choosing the strike price of the call as high as we like, by the preceding lemma, the value of the call is as close to*

zero as we like. Therefore the basket with the call has bounded payout, and value as close as we like to the original basket.

Accordingly we restrict our attention to option positions with bounded payouts. The difference between values of one option position with bounded payout for two processes P_0 and P_1, is special case of a statistical test between P_0 and P_1, and it may or may not be a best possible test. The sum of Type 1 and Type 2 errors for the best test satisfies:

$$\alpha + \beta \geq 1 - \frac{1}{2} \|P_0 - P_1\|_1 \tag{35}$$

Since the total variation distance is bounded by the Hellinger distance and Hellinger distance is bounded by the information distance, when the information distance between P_0 and P_1 is small enough, $\alpha + \beta$ will be close to 1 and it is extremely difficult to distinguish these option prices.

When the underlying return process is a long memory process, it is impossible to value the option price using non-arbitrage condition. However if we can approximate the long memory process by a sequence of short memory processes converging in information distance, then the option valuations based on short memory processes are well defined, and cannot be distinguished based on any given finite amount of data from option prices that make sense for the long memory process. We will go into this discussion in the section "Option pricing with long memory".

3 Fractional autoregressive integrated moving average model

We use FARIMA to represent the fractional autoregressive integrated moving average model, while some other acronyms are also used in literature, such as ARFIMA and ARIMA. For example, Hosking [25] used ARIMA to represent the fractional ARIMA process by allowing the order of differencing d to be fractional. To make it more clear and less ambiguous, FARIMA is used to represent the fractional autoregressive integrated moving average process in this dissertation.

3.1 Fractional difference operator

The fractional difference operator is one of the fundamental parts in the FARIMA model and it is also the part that adds the long memory property to the short classic memory ARMA process. The definition of the fractional difference operator is defined as follows (see Hosking [25] in 1981):

Definition 7. *Fractional difference operator ∇^d can be defined by a binomial series of the backward-shift operators:*

$$\nabla^d = (1-L)^d = \sum_{k=0}^{\infty} \begin{pmatrix} d \\ k \end{pmatrix} (-L)^k = \sum_{k=0}^{\infty} \frac{\prod_{a=0}^{k-1}(d-a)}{k!} (-L)^k \tag{36}$$

where L is the backward-shift operator, such that $Lx_t = x_{t-1}$, and we also have the following expression: $\nabla^{-d} = \sum_{k=0}^{\infty} \begin{pmatrix} d \\ k \end{pmatrix} L^k$

3.2 FARIMA process

FARIMA process can be defined as follows:

Definition 8. *$FARIMA(p,d,q)$ process is defined to have the following form [25, 28]:*

$$\phi(L)\nabla^d x_t = \theta(L)\varepsilon_t \tag{37}$$

where ∇^d *is the fractional difference operator;* $\phi(L)$ *and* $\theta(L)$ *are polynomials of the backward-shift operator* L*:*

$$\begin{aligned} \phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \\ \theta(L) &= 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q \end{aligned}$$

p *and* q *define the order of* AR *and* MA *parts respectively and they are integers.* d *represents the order of fractional difference and it can take fractional value.*

Hosking [25] showed that without AR or MA effects in $FARIMA$ process, $-\frac{1}{2} < d < \frac{1}{2}$ would ensure $FARIMA(0,d,0)$ to be stationary and invertible. To be consistent with other parts of this dissertation, we will use z to represent the backward-shift operator L here.

3.3 Information distance between white noise and FARIMA(0,d,0)

The simplest stochastic process is white noise, which is also a short memory process. Before getting into more complicated short memory processes, we compare the white noise process and the $FARIMA(0,d,0)$ process.

The transfer function of white noise process is simply:

$$g(z) = 1$$

and we know that the transfer function of $FARIMA(0,d,0)$ is:

$$f(z) = \nabla^{-d} = (1-z)^{-d} = \left(\frac{1}{1-z}\right)^d \tag{38}$$

The logarithm of the transfer function of $FARIMA(0,d,0)$ is:

$$\log f(z) = d\sum_{k=1}^{\infty} \frac{z^k}{k} \tag{39}$$

So the information distance, described in previous section, between the two processes can be written as

$$\|\log f(z) - \log g(z)\|_{H^2}^2 = \sum_{k=1}^{\infty} \frac{d^2}{k^2} = \frac{\pi^2 d^2}{6} \tag{40}$$

and the shortest information distance between these two processes is 0, when $d = 0$.

According to the binomial expansion of the fractional difference operator (see equation 36), the transfer function of $FARIMA(0, d, 0)$ can also be expressed as:

$$f(z) = (1-z)^{-d} = \sum_{k=0}^{\infty} \begin{pmatrix} d \\ k \end{pmatrix} z^k = 1 + \sum_{k=1}^{\infty} \frac{\prod_{a=0}^{k-1}(d+a)}{k!} z^k \tag{41}$$

Equivalently, the impulse response function $h(n)$ of $FARIMA(0, d, 0)$ can be written as:

$$h(n) = \frac{\prod_{a=0}^{n-1}(d+a)}{n!} \tag{42}$$

3.4 Information distance between ARMA(p,q) and FARIMA(0,d,0)

In this part, we consider the more general class of stochastic processes, autoregressive moving average process, ARMA. Compared to the FARIMA process with $d > 0$, the stationary ARMA process is a short memory process, whose autocovariance function decays exponentially. Here through two more generalized theorems, we show that 1) for fixed p, q, the $ARMA\left(p,q\right)$ processes are not information dense in the processes with finite information length; 2) with unbounded p, q, the $ARMA\left(p,q\right)$ processes are not information dense with the processes with finite information length.

3.4.1 FARIMA(0,d,0) and ARMA(p,q) with bounded p and q

To show that for bounded p, q, the $ARMA(p,q)$ processes are not information dense with the processes with finite information length, we provide two Lemmas first. Let S be a subset of the complex plane, and $R_{pq}\left(S\right)$ denote the set of rational functions which have p poles and q zeros, and the poles and zeros are in S. For example, with D as the unit disc $\{z \text{ s.t. } |z| < 1\}$, then $R_{pq}\left(\mathbb{C}\backslash D\right)$ includes the stable and minimum phase $ARMA\left(p,q\right)$ processes, and some limiting cases.

Lemma 1. *$R_{pq}\left(\mathbb{C}\backslash D\right)$ is a subset of $\exp\left(H^2\left(D\right)\right)$, that is, the information length of any $f \in R_{pq}\left(\mathbb{C}\backslash D\right)$ is finite.*

Proof: *Write the logarithm of $f \in R_{pq}\left(\mathbb{C}\backslash D\right)$ as*

$$\log f\left(z\right) = \log f\left(0\right) + \sum_{j=1}^{p} \log\left(\frac{1}{1-\lambda_j z}\right) - \sum_{k=1}^{q} \log\left(\frac{1}{1-\mu_k z}\right)$$

then the triangle inequality provides

$$\left\|\log f\right\|_{H^2(D)} \leq \left|\log f\left(0\right)\right| + \sum_{j=1}^{p} \left\|\log\left(\frac{1}{1-\lambda_j z}\right)\right\|_{H^2(D)} + \sum_{k=1}^{q} \left\|\log\left(\frac{1}{1-\mu_k z}\right)\right\|_{H^2(D)}.$$

Since for $|a| \leq 1$ we have

$$\left\|\log\left(\frac{1}{1-az}\right)\right\|_{H^2(D)}^2 = \sum_{j=1}^{\infty} \frac{|a|^2}{k^2} \leq \sum_{j=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

then

$$\|\log f\|_{H^2(D)} \leq |\log f(0)| + \frac{\pi(p+q)}{\sqrt{6}}.$$

Corollary 1. *An immediate corollary to the lemma is that* $R_{pq}\left(\mathbb{C}\backslash\overline{D}\right) \subset R_{pq}(\mathbb{C}\backslash D) \subset \exp\left(H^2(D)\right)$.

Lemma 2.

$$\left|\log \frac{f(0)}{g(0)}\right|^2 \leq I(f,g)^2.$$

Proof: *Since* $\log \frac{f(z)}{g(z)} \in H^2(D)$, *we have*

$$\left(\log \frac{f(0)}{g(0)}\right)^2 = \int_{-\pi}^{\pi} \left(\log \frac{f\left(re^{i\theta}\right)}{g\left(re^{i\theta}\right)}\right)^2 \frac{d\theta}{2\pi}$$

and

$$\left|\log \frac{f(0)}{g(0)}\right|^2 \leq \left|\int_{-\pi}^{\pi} \left(\log \frac{f\left(re^{i\theta}\right)}{g\left(re^{i\theta}\right)}\right)^2 \frac{d\theta}{2\pi}\right| \leq \int_{-\pi}^{\pi} \left|\log \frac{f\left(re^{i\theta}\right)}{g\left(re^{i\theta}\right)}\right|^2 \frac{d\theta}{2\pi} = I(f,g)^2.$$

After having the two Lemmas, we give our theorem for bounded p, q.

Theorem 3. *Let* f *be given with* $0 = \inf\left\{I(f,g) \text{ s.t. } g \in R_{pq}\left(\mathbb{C}\backslash\overline{D}\right)\right\}$. *Then* $f \in R_{pq}(\mathbb{C}\backslash D)$.

Proof: *Let* $g \in R_{pq}\left(\mathbb{C}\backslash\overline{D}\right)$ *satisfy* $I(f,g) \leq 1$, *then* $\log \frac{f(z)}{g(z)} \in H^2(D)$, *and*

$$\left|\log \frac{f(0)}{g(0)}\right| \leq I(f,g) \leq 1.$$

Since $g(0)$ *is finite, so is* $f(0)$. *By hypothesis there exists* g_m, *a sequence in* $R_{pq}\left(\mathbb{C}\backslash\overline{D}\right)$ *such that* $I(g_m, f) \to 0$ *as* $m \to \infty$. *Since*

$$I(g_m, g_n) \leq I(g_m, f) + I(f, g_n)$$

must also tend to zero for all $n \geq m$ *as* $m \to \infty$, *then* $\log g_m$ *is a Cauchy sequence in the Hilbert space* $H^2(D)$, *and must converge to a limit* $\log g_\infty \in H^2(D)$. *Since* $I(g_\infty, f) = 0$, $\log g_\infty = \log f$, *and* $\lim_{m\to\infty} g_m(0) = f(0)$. *Let* $\theta_m \in D^{p+q}$ *be*

the vector of poles and zeros of g_m*, then the sequence* θ_m *must have a point of accumulation* θ^∞ *in the closed polydisc* $\overline{D}^{p+q}$ *and there is a subsequence* m_k *such that* $\theta_{m_k} \to \theta^\infty$ *as* $k \to \infty$*. Therefore*

$$f(z) = f(0) \frac{\prod_{j=1}^{q} \left(1 - \mu_j^\infty z\right)}{\prod_{k=1}^{p} \left(1 - \lambda_k^\infty z\right)} \in R_{pq}\left(\mathbb{C} \backslash D\right).$$

This theorem shows that for fixed p, q, the $ARMA(p, q)$ processes are not information dense in the processes with finite information length $\exp\left(H^2(D)\right)$, since there are many functions which are analytic but not rational. In particular the fractionally differenced process has transfer function $(1 - z)^d$ which is rational if and only if d is an integer. However since

$$\left\| \log \left(\frac{1}{1 - z} \right)^d \right\|_{H^2(D)} = \frac{\pi |d|}{\sqrt{6}}$$

all the fractionally differenced and fractionally integrated noises have finite information length. By the theorem or any particular noninteger d, and fixed p, q, we have

$$\inf \left\{ I\left(\left(\frac{1}{1 - z} \right)^d, g \right) \text{ s.t. } g \in R_{pq}\left(\mathbb{C} \backslash D\right) \right\} > 0.$$

On the other hand, for the random walk, we have

$$\lim_{r \nearrow 1} \left\| \log \left(\frac{1}{1 - z} \right) - \log \left(\frac{1}{1 - rz} \right) \right\|_{H^2(D)} = 0$$

since the theorem of mean convergence to boundary function [31]. This is consistent with the theorem since $\frac{1}{1-z} \in R_{pq}\left(\mathbb{C} \backslash D\right)$.

In the case of the fractionally differenced or integrated noises, we will have to use sequences with unbounded p, q if we want to approximate these noises to arbitrary information distance precision. In fact, the processes with unbounded p, q are information dense in the finite information length processes, as we will show in the next theorem.

3.4.2 FARIMA(0,d,0) and ARMA(p,q) with unbounded p and q

Here we show that the $ARMA(p,q)$ processes are information dense with the processes with finite information length when p and q are unbounded by the following theorem:

Theorem 4. *Rational functions are information dense with the transfer functions that have finite information distance from the unit variance white noise.*

Proof: *Let $R_{pq}(z)$, $f(z)$ represent the rational functions and the transfer functions that have finite information distance from the unit variance white noise respectively. So we have*

$$R_{pq}(z) = \frac{P(z)}{Q(z)},\ \|\log f(z)\|_{H^2} < \infty \tag{43}$$

where 1) P and Q are polynomials in z; 2) poles and zeros of $R_{pq}(z)$ are in $\mathbb{C}\backslash D$ with D as the unit disc $\{z,\ s.t.\ |z|<1\}$.

So the proposition is equivalent to prove that:

$$\|\log R_{pq}(z) - \log f(z)\|_{H^2} < \varepsilon,\ \forall \varepsilon > 0 \tag{44}$$

To prove this proposition, we need to use the dilated process of $f(z)$, which is denoted as $f(rz)$ and $0<r<1$. Then we have

$$\begin{aligned} & \|\log R_{pq}(z) - \log f(z)\|_{H^2} \\ = \ & \|\log R_{pq}(z) - \log f(rz) + \log f(rz) - \log f(z)\|_{H^2} \qquad (45) \\ \leq \ & \|\log R_{pq}(z) - \log f(rz)\|_{H^2} + \|\log f(rz) - \log f(z)\|_{H^2} \qquad (46) \end{aligned}$$

In order to prove that $\|\log R_{pq}(z) - \log f(z)\|_{H^2}$ can be sufficiently small, we will show that both $\|\log R_{pq}(z) - \log f(rz)\|_{H^2}$ and $\|\log f(rz) - \log f(z)\|_{H^2}$ can be sufficiently small.

1) From the theorem of mean convergence to boundary function [31], we know that if an analytic function $f(z) \in H_p$,

$$\int_{-\pi}^{\pi} \left| f\left(re^{i\theta}\right) - f\left(e^{i\theta}\right) \right|^p d\theta \to 0 \tag{47}$$

as $r \uparrow 1$*, where* $f\left(e^{i\theta}\right)$ *is the boundary value of* $f(z)$ *which exists a.e.*

Since $\|\log f(z)\|_{H^2} < \infty$*, then we have*

$$\lim_{r\uparrow 1} \|\log f(z) - \log f(rz)\|_{H^2}^2 < \varepsilon_1, \forall \varepsilon_1 > 0 \tag{48}$$

2) Recall Runge's theorem [32]: let K *be a compact subset of* $\mathbb{C}$ *and* g *be analytic on a neighborhood of* K*, and let* $P \subset \mathbb{C}\backslash K$*. Then for any* $\varepsilon > 0$*, there exists a rational function* $r(z)$ *with poles in* P *such that*

$$\max_{z\in K} |g(z) - r(z)| < \varepsilon \tag{49}$$

We know that $f(rz)$ *is analytic on* $|z| < r^{-1}$*. Then Runge's theorem tells us that there exists a exists a rational function* $R_{pq}(z)$*, that has poles outside* $|z| < r^{-1}$*, such that*

$$\max_{|z|<r^{-1}} |f(rz) - R_{pq}(z)| < \varepsilon_2 \tag{50}$$

which is equivalent to

$$\|f(rz) - R_{pq}(z)\|_{H^\infty} < \varepsilon_2 \tag{51}$$

Since $0 < |f(rz)| < \infty$*,* $0 < |R_{pq}(z)| < \infty$*, we have* $0 < |\log f(rz)| < \infty$*,* $0 < |\log R_{pq}(z)| < \infty$*. Then we can get*

$$\|\log f(rz) - \log R_{pq}(z)\|_{H^\infty} \le M \|f(rz) - R_{pq}(z)\|_{H^\infty} < M\varepsilon_2 \tag{52}$$

where M *is a positive constant.*

Based on the definition of norms in Hardy space, we know that $\|\cdot\|_{H^2}^2 \le \|\cdot\|_{H^\infty}^2$*, then we have*

$$\|\log f(rz) - \log R_{pq}(z)\|_{H^2} \le \|\log f(rz) - \log R_{pq}(z)\|_{H^\infty} \le M\varepsilon_2 \tag{53}$$

After having (Eq. 48) and (Eq. 53), there exists $\varepsilon = \varepsilon_1 + M\varepsilon_2$*, such that*

$$\|\log f(z) - \log R_{pq}(z)\|_{H^2}^2 \le \varepsilon_1 + M\varepsilon_2 = \varepsilon \tag{54}$$

Based on this theorem, we derive the following corollaries which connect stationary ARMA processes with other processes from the perspective of information distance.

Corollary 2. *The information distance between stationary ARMA(p,q) and unit root processes can be sufficiently small.*

Proof: *Recall the definition of* $R_{pq}(z)$ *in the proposition,* $R_{pq}(z) = \frac{P(z)}{Q(z)}$ *and both poles and zeros of* $R_{pq}(z)$ *are outside* $|z| < 1$. *We know that the transfer functions of stationary ARMA(p,q) processes satisfy it. The transfer function of unit root process is* $f(z) = \frac{1}{1-z}$ *and we know that transfer functions of unit root processes have finite information distance from the unit variance white noise, since*

$$\|\log f(z)\|_{H^2} = \left\| \sum_{k=1}^{\infty} \frac{z^k}{k} \right\|_{H^2} = \frac{\pi}{\sqrt{6}} < \infty \tag{55}$$

So we know that there exists a stationary ARMA(p,q) process with transfer function $R_{pq}(z)$ *such that*

$$\left\| \log R_{pq}(z) - \log\left(\frac{1}{1-z}\right) \right\| < \varepsilon, \forall \varepsilon > 0 \tag{56}$$

Corollary 3. *The information distance between stationary ARMA(p,q) and FARIMA(0,d,0) processes can be sufficiently small.*

Proof: *Similarly, the transfer functions of FARIMA(0,d,0) processes,* $f(z) = \left(\frac{1}{1-z}\right)^d$, *have finite information distance from the unit variance white noise, since*

$$\|\log f(z)\|_{H^2} = d \left\| \sum_{k=1}^{\infty} \frac{z^k}{k} \right\|_{H^2} = \frac{\pi d}{\sqrt{6}} < \infty \tag{57}$$

So we know that there exists a stationary ARMA(p,q) process with transfer function $R_{pq}(z)$ *such that*

$$\left\| \log R_{pq}(z) - \log\left(\left(\frac{1}{1-z}\right)^d\right) \right\| < \varepsilon, \forall \varepsilon > 0 \tag{58}$$

Corollary 4. *The information distance between stationary ARMA($\hat{p}$,$\hat{q}$) and FARIMA(p,d,q) processes can be sufficiently small.*

Proof: *The transfer function of FARIMA(p,d,q) can be written as:*

$$f(z) = \frac{1 + \sum_{k=1}^{q} \theta_k z}{1 - \sum_{k=1}^{p} \phi_k z} \left(\frac{1}{1-z} \right)^d \tag{59}$$

and we can easily get $\|\log f(z)\|_{H^2} < \infty$*, since*

$$\|\log f(z)\|_{H^2} \leq \left\| \log \left(\frac{1 + \sum_{k=1}^{q} \theta_k z}{1 - \sum_{k=1}^{p} \phi_k} \right) \right\|_{H^2} + \left\| d \log \left(\frac{1}{1-z} \right) \right\|_{H^2} < \infty$$

So there exists a stationary ARMA($\hat{p}$,$\hat{q}$) process have sufficiently small information distance from FARIMA(p,d,q) processes.

4 Simulation

In previous section, we have shown that the information distance between short memory processes and long memory processes or the unit root process can be sufficiently small. Based on the connection between information distance, Hellinger distance and total variation, we know the total variation between the two processes can be sufficiently small. So sufficiently small information distance means that these processes can not be distinguished from each other with finite sample size. To show these results numerically, we implement different statistical tests to see whether these processes can be distinguished from each other.

4.1 White noise and FARIMA(0,d,0)

Like the idea from Diebold and Inoue [2], we are doing a Monte Carlo simulation (the number of trials is $1,000$ here) with different sample size N and fractional differencing parameter d.

So given sample size N and fractional differencing parameter d, two samples with sample size N are generated from white noise and FARIMA(0,d,0) processes. The two samples are represented by S_1 and S_2 here and the null hypothesis for the statistical test is:

$$H_0 : S_1 \text{ and } S_2 \text{ are sampled from the same distribution}$$

We show the fraction of $1,000$ trials where inference based on different statistics leads to rejection of the null hypothesis H_0 at 95% confidence interval. The results of Kolmogorov-Smirnov and Anderson-Darling tests are shown in Table 2 and Table 3 respectively. A few observations from these numerical tests are consistent with the information distance between white noise and FARIMA(0,d,0) processes:

1. When the same size N is small, such as 50 or 100, it is hard to reject the null hypothesis even if d is significantly different from 0.

2. A smaller value of information distance, $\frac{\pi^2 d^2}{6}$ here, needs more samples to reject the null hypothesis. It's also consistent with the underlying relationship between total variation distance and information distance: that the total variation distance is bounded by the information distance.

White noise process is one of the simplest stochastic process with short memory property. However, when sample size is small, statistical test cannot distinguish it from a long memory process, FARIMA(0,d,0) with $d > 0$ here. We will explore the long memory process with more complicated structure and higher dimensions in the following part of the dissertation.

Table 2: Kolmogorov-Smirnov statistic of white noise process and FARIMA(0,d,0) process

	N						
d	50	100	500	$1,000$	$1,500$	$2,000$	$2,500$
-0.40	0.750	0.993	1.000	1.000	1.000	1.000	1.000
-0.35	0.367	0.835	1.000	1.000	1.000	1.000	1.000
-0.30	0.188	0.477	1.000	1.000	1.000	1.000	1.000
-0.25	0.081	0.228	0.995	1.000	1.000	1.000	1.000
-0.20	0.044	0.080	0.793	0.998	1.000	1.000	1.000
-0.15	0.031	0.047	0.316	0.759	0.962	1.000	1.000
-0.10	0.030	0.032	0.098	0.211	0.425	0.563	0.718
-0.05	0.026	0.042	0.031	0.057	0.088	0.091	0.107
0.00	0.044	0.045	0.052	0.050	0.049	0.045	0.038
0.05	0.076	0.062	0.085	0.113	0.131	0.163	0.164
0.10	0.113	0.091	0.184	0.262	0.333	0.402	0.479
0.15	0.153	0.168	0.312	0.416	0.553	0.615	0.739
0.20	0.188	0.229	0.419	0.564	0.662	0.747	0.810
0.25	0.227	0.327	0.544	0.616	0.722	0.780	0.826
0.30	0.302	0.381	0.618	0.698	0.726	0.781	0.789
0.35	0.394	0.463	0.661	0.741	0.757	0.797	0.783
0.40	0.449	0.535	0.734	0.819	0.833	0.906	0.910

d represents the fractional differencing parameter and N represents sample size. The null hypothesis is that two samples with sample size N generated from white noise and FARIMA(0,d,0) processes come from the same distribution. The numbers in the table represent the fraction of $1,000$ trials where inference based on the Kolmogorov-Smirnov statistic leads to reject the null hypothesis.

Table 3: Anderson-Darling statistic of white noise process and FARIMA(0,d,0) process

	N						
d	50	100	500	$1,000$	$1,500$	$2,000$	$2,500$
-0.40	0.981	1.000	1.000	1.000	1.000	1.000	1.000
-0.35	0.758	0.993	1.000	1.000	1.000	1.000	1.000
-0.30	0.433	0.894	1.000	1.000	1.000	1.000	1.000
-0.25	0.159	0.587	1.000	1.000	1.000	1.000	1.000
-0.20	0.080	0.224	0.994	1.000	1.000	1.000	1.000
-0.15	0.039	0.079	0.750	0.993	1.000	1.000	1.000
-0.10	0.025	0.035	0.202	0.578	0.843	0.945	0.989
-0.05	0.030	0.044	0.048	0.091	0.109	0.193	0.251
0.00	0.046	0.048	0.047	0.061	0.034	0.048	0.056
0.05	0.078	0.075	0.099	0.139	0.187	0.216	0.219
0.10	0.112	0.138	0.253	0.383	0.481	0.618	0.720
0.15	0.158	0.214	0.424	0.595	0.744	0.865	0.915
0.20	0.231	0.289	0.530	0.724	0.846	0.900	0.953
0.25	0.284	0.398	0.621	0.735	0.850	0.884	0.934
0.30	0.366	0.452	0.663	0.763	0.787	0.853	0.855
0.35	0.441	0.513	0.713	0.782	0.782	0.834	0.826
0.40	0.474	0.611	0.759	0.852	0.868	0.916	0.939

d represents the fractional differencing parameter and N represents sample size. The null hypothesis is that two samples with sample size N generated from white noise and FARIMA(0,d,0) processes come from the same distribution. The numbers in the table represent the fraction of $1,000$ trials where inference based on the Anderson-Darling statistic leads to reject the null hypothesis.

4.2 Unit root and short memory processes

In proposition 4, we show that the information distance between a stationary ARMA(p,q) process and unit root process can be sufficiently small. Here we construct different stationary ARMA processes, which are short memory processes, and apply the statistical test to identify whether they are unit root processes. Here, we apply two statistical tests, Augmented Dickey-Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests, to examine whether the approximated short memory processes and unit root processes can be distinguished.

4.2.1 Augmented Dickey-Fuller test

Augmented Dickey-Fuller (ADF) test is a statistical test for unit root in the sample of a time series. Let y_t represents the samples from a stochastic time series and the ADF test is based on the OLS-estimated coefficients of the following equation [38]:

$$\Delta y_t = f(t) + (a-1)\, y_{t-1} + \sum_{i=1}^{p} a_i \Delta y_{t-i} + \varepsilon_t \tag{60}$$

where $\Delta y_t = y_t - y_{t-1}$ and ε_t is a mean zero innovation process. The null hypothesis of a unit root is

$$H_0 : a = 1 \tag{61}$$

So if the null hypothesis cannot be rejected, the ADF test concludes that y_t is a unit root process.

Two short memory processes, which are approximated the unit root processes, are tested by the ADF test. One is to approximate the impulse response, or the characteristic equation, of the unit root processes by high dimensional ARMA processes; the other one is the dilated process of the unit root process, which is a short memory process. Table 4 and Table 5 show the ADF test results of approximated ARMA processes and dilation processes respectively.

Table 4: ADF tests of samples from stationary ARMA process

	n						
N	1	2	10	20	50	100	200
500	1.000	0.467	0.068	0.075	0.049	0.061	0.052
$1,000$	1.000	0.993	0.062	0.055	0.052	0.054	0.048
$5,000$	1.000	1.000	0.297	0.060	0.085	0.057	0.065
$10,000$	1.000	1.000	0.905	0.112	0.052	0.048	0.064

n represents the dimensions of approximated ARMA processes and N represents sample size. The null hypothesis is that the samples from approximated ARMA processes are from the unit root processes. The numbers in the table represent the fraction of $1,000$ trials where inference based on the ADF statistic leads to reject the null hypothesis.

Table 5: ADF tests of samples from dilation processes

	d				
N	$1-10^{-1}$	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$
500	1.000	0.311	0.064	0.049	0.047
$1,000$	1.000	0.780	0.083	0.052	0.064
$5,000$	1.000	1.000	0.292	0.082	0.053
$10,000$	1.000	1.000	0.742	0.071	0.047

d represents the dilation parameter in the dilation process of the unit root processes and N represents sample size. The null hypothesis is that the samples from the dilation processes are from the unit root processes. The numbers in the table represent the fraction of $1,000$ trials where inference based on the ADF statistic leads to reject the null hypothesis.

4.2.2 KPSS tests

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests is an alternative statistical test for the unit root process. The KPSS tests [52] are used for testing the null hypothesis that a time series is stationary around a deterministic trend. Using y_t to represent the observable time series and the null hypothesis of KPSS tests can be derived from the following equations [52]:

$$y_t = \xi t + r_t + \varepsilon_t \tag{62}$$

$$r_t = r_{t-1} + u_t \tag{63}$$

where ε_t are $iid\left(0, \sigma_\varepsilon^2\right)$ and u_t are $iid\left(0, \sigma_u^2\right)$. The null hypothesis of KPSS tests is:

$$H_0 : \sigma_u^2 = 0 \tag{64}$$

So y_t is a stationary process if the null hypothesis can not be rejected.

Similar to the ADF tests, approximated ARMA and dilation processes are tested by the KPSS test. Table 6 and Table 7 show the ADF test results of approximated ARMA processes and dilation processes respectively.

Table 6: KPSS tests of samples from stationary ARMA process

	n						
N	1	2	10	20	50	100	200
500	0.968	1.000	1.000	1.000	1.000	1.000	1.000
1,000	0.972	1.000	1.000	1.000	1.000	1.000	1.000
5,000	0.963	1.000	1.000	1.000	1.000	1.000	1.000
10,000	0.964	1.000	1.000	1.000	1.000	1.000	1.000

n represents the dimensions of approximated ARMA processes and N represents sample size. The null hypothesis is that the samples from approximated ARMA processes are from stationary processes. The numbers in the table represent the fraction of 1,000 trials where inference based on the KPSS statistic leads to reject the null hypothesis.

Table 7: KPSS tests of samples from dilation processes

	d				
N	$1-10^{-1}$	$1-10^{-2}$	$1-10^{-3}$	$1-10^{-4}$	$1-10^{-5}$
500	1.000	1.000	1.000	1.000	1.000
1,000	1.000	1.000	1.000	1.000	1.000
5,000	1.000	1.000	1.000	1.000	1.000
10,000	1.000	1.000	1.000	1.000	1.000

d represents the dilation parameter in the dilation process of the unit root processes and N represents sample size. The null hypothesis is that the samples from the dilation processes are from stationary processes. The numbers in the table represent the fraction of 1,000 trials where inference based on the KPSS statistic leads to reject the null hypothesis.

4.3 FARIMA(0,d,0) and short memory processes

In the previous sections, we show that there exists short memory processes, stationary ARMA processes, to be close enough to the FARIMA(0,d,0) processes ($d > 0$) in information distance. Here we approximate the FARIMA(0,d,0) processes ($d > 0$) with stationary ARMA processes and examine whether we can find the long memory properties in these approximated ARMA processes.

We examine our approximated short memory processes and the long memory processes from two perspectives. One is from the serial correlation of the time series samples; the other one is from the conditional distributions of these processes. The first kind of test is more strict than the second one since it's based on the characteristic equation of the stochastic processes. The reason we are also interested in the second test perspective is that conditional distributions are the most important factor in European vanilla option valuation.

To examine FARIMA(0,d,0) and short memory processes from the perspective of the serial dependency, we apply the widely used statistical estimations for long memory processes, correlogram and Geweke-Porter-Hudak (GPH) estimations. Then we construct the statistical hypothesis tests on these estimations of long memory processes.

4.3.1 Correlogram estimation

Recall that the definition of long memory processes is based on the decay speed of the autocorrelation function of the underlying processes. Let ρ_k be the autocorrelation function of a time series r_t and the long memory processes have the absolute value of autocorrelation not summable: $\lim_{n\to\infty} \sum_{i=1}^{n} |\rho_k| = \infty$.

Therefore another way to examine the long memory is through the limiting decay rate of the autocorrelation function [53, 54]. Many previous works compare the estimated autocorrelation at lag k to the 95% confidence interval, which is $\pm 2/\sqrt{k}$ in this case, and argue whether there exists long memory in the observable

time series.

Beran [15] shows that for a stationary long memory process, there exists a real number $\alpha \in (0,1)$ and a constant $c_\rho > 0$ such that

$$\lim_{k\to\infty} \frac{\rho_k}{c_\rho k^{-\alpha}} = 1 \tag{65}$$

So one alternative way to test for long memory processes is to see how the $\log|\rho_k|$ for the processes decays asymptotically. A linear relation can be obtained by taking logarithm of the asymptotic relationship as $k \to \infty$:

$$\log\left(|\rho(k)|\right) \sim \log c_\rho - \alpha \log(k) \tag{66}$$

To test whether the estimated α from long and short memory processes' samples are statistically indistinguishable, we construct the following hypothesis testing.

Let Y represent logarithm of estimated autocorrelation values and X represent the corresponding logarithm of lags. Use D to be the dummy variable (with value equal to 0 or 1) to identify the estimations from FARIMA(0,d,0) processes and approximated ARMA processes. Then we have the following linear equations to regress:

$$Y_i = c + \gamma D_i + \alpha X_i + \beta\left(D_i X_i\right) + \varepsilon_i \tag{67}$$

where ε_i are iid Gaussian innovations.

So the null hypothesis in our statistical inference is:

$$H_0 : \beta = 0 \tag{68}$$

and we say the samples from FARIMA(0,d,0) and approximated ARMA processes are indistinguishable if we can not reject the null hypothesis at a given significant level.

After constructing the null hypothesis of statistical inference, the procedures and parameters in our simulation are described as follows:

Step 1: Choose a positive fractional differencing parameter d in the FARIMA(0,d,0) processes, dimension of approximated ARMA process n, sample size N.

Step 2: Approximate the impulse response of FARIMA(0,d,0) process by the n dimensions ARMA process.

Step 3: Simulate 500 trials from the two processes in Step 2 and estimate the sample autocorrelation functions of two samples in each trial.

Step 4: Implement the hypothesis testing on these estimated autocorrelation functions in each trials and count the fraction of these trials where the statistical inference leads to reject the null hypothesis.

We show the simulation results of $d = 0.3$ and $d = 0.4$ in Table 8 and Table 9 respectively. In these tables, we find that when the dimension of approximations increase, the information distance is smaller and it is harder to reject the null hypothesis that samples from FARIMA(0,d,0) and approximated ARMA processes are from the same process. Also when the long memory processes are more significant (larger d), it's easier for the statistical test to distinguish data from long memory and approximated short memory processes. But as we show in the previous propositions, there exists an approximated short memory ARMA processes to be close enough to a FARIMA(0,d,0) process when $|d| < \frac{1}{2}$. So the statistical test can not distinguish the data from FARIMA(0,d,0) and approximated process even though with a large fractional differencing parameter d.

To show how the information distance and statistical inference are connected numerically, Fig 1Fig 2 and Fig 3 show the cases for FARIMA(0,0.2,0), FARIMA(0,0.3,0) and FARIMA(0,0.4,0) respectively.

Figure 1: Hypothesis testing based on correlogram estimations and information distance between FARIMA(0,0.2,0) and approximated processes

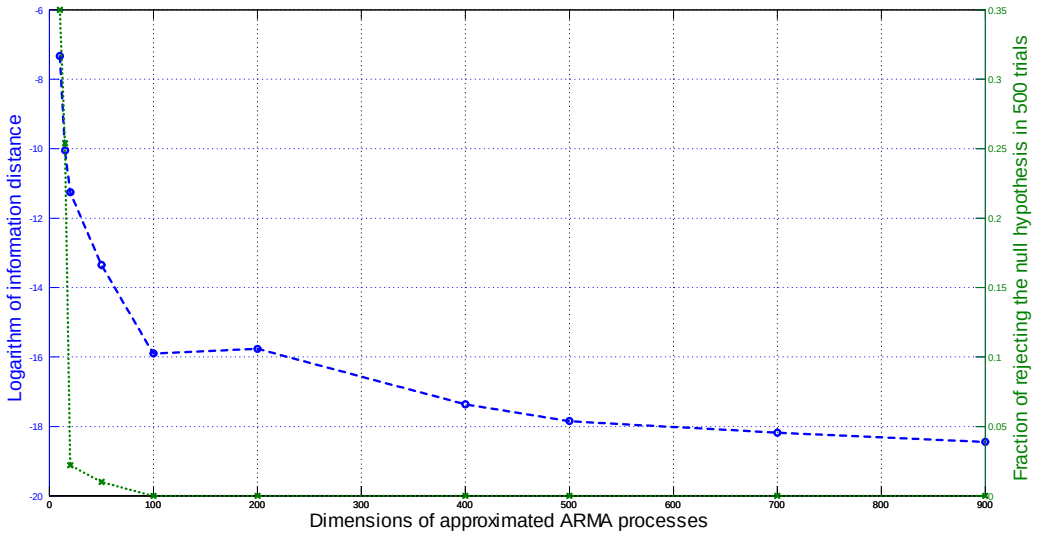


Figure 2: Hypothesis testing based on correlogram estimations and information distance between FARIMA(0,0.3,0) and approximated processes

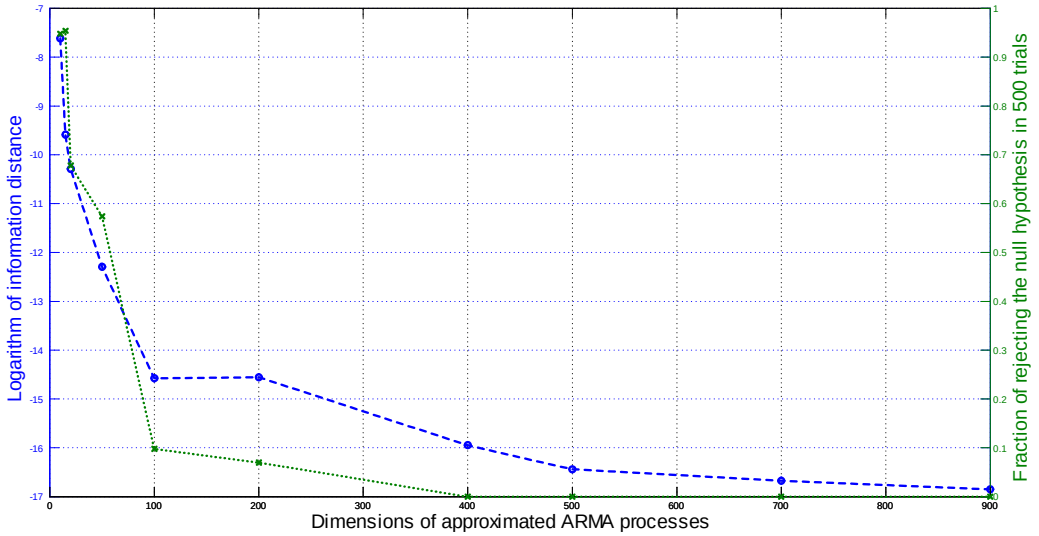
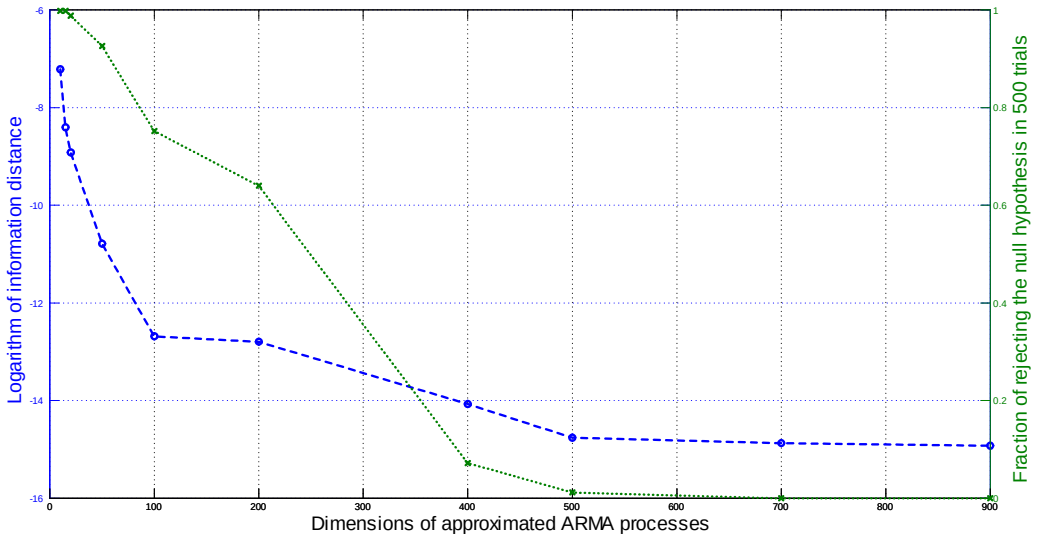


Figure 3: Hypothesis testing based on correlogram estimations and information distance between FARIMA(0,0.45,0) and approximated processes



In Fig 1, Fig 2 and Fig 3, we find that: 1) the fraction of rejecting the null hypothesis is highly related to the information distance and the fraction is going to zero when information distance is small enough; 2) information distance is not always decreasing with higher order of approximation, since the approximation is not implemented based on the measure of information distance; 3) higher fractional differencing parameter d leads to higher fraction of rejecting the null hypothesis with the same dimensions of the approximated ARMA processes.

Table 8: Hypothesis testings based on correlogram estimation of FARIMA(0,0.3,0) and its approximation

	n				
N	10	20	50	100	500
$5,000$	0.138	0.020	0.000	0.000	0.000
$10,000$	0.194	0.092	0.000	0.000	0.000
$50,000$	0.838	0.392	0.698	0.000	0.000
$100,000$	0.954	0.624	0.356	0.066	0.000
Distance	5×10^{-4}	3×10^{-5}	4×10^{-6}	5×10^{-7}	7×10^{-8}

Table 9: Hypothesis testings based on correlogram estimation of FARIMA(0,0.4,0) and its approximation

	n				
N	10	20	50	100	500
$5,000$	0.440	0.128	0.066	0.000	0.000
$10,000$	0.496	0.404	0.082	0.000	0.000
$50,000$	0.966	0.798	0.762	0.134	0.000
$100,000$	0.978	0.964	0.916	0.620	0.000
Distance	6×10^{-4}	8×10^{-5}	1×10^{-5}	1×10^{-6}	2×10^{-7}

n and N represent the dimensions of approximated ARMA processes and sample size respectively. The numbers in the table represent the fraction of 500 trials where inference correlogram estimations described in (Eq.67 and 68) leads to reject the null hypothesis. The last row in each table represents the information distance between these two processes.

4.3.2 GPH estimations

An alternative way to say whether we can distinguish the specific FARIMA(0,d,0) processes and approximated short memory processes is to compare the estimations of fractional differencing parameter with finite sample size. There are different methodologies to estimate the fractional differencing parameter d in the FARIMA(p,d,q) models. As it is already shown in Definition 9, without the ARMA part in FARIMA(p,d,q) processes, x_t is said to follow FARIMA(0,d,0) process when $x_t = (1-z)^{-d}\varepsilon_t$. So its autocovariance-generating function is

$$g_x(z) = \sigma^2 (1-z)^{-d} \left(1-z^{-1}\right)^{-d} = \left[(1-z)\left(1-z^{-1}\right)\right]^{-d} \tag{69}$$

since $\sigma^2 = 1$ for standard Gaussian innovations. So the spectral density function of FARIMA(0,d,0) is:

$$f(\omega) = \frac{1}{2\pi}\left[\left(1-e^{i\omega}\right)\left(1-e^{-i\omega}\right)\right]^{-d} = \frac{1}{2\pi}\left(4\sin^2\frac{\omega}{2}\right)^{-d} \tag{70}$$

To estimate the fractional differencing parameter d, Geweke and Porter-Hudak [55] proposed to use the periodogram estimator to estimate the spectrum density function $f(\omega)$ and estimate the d through the linear regression of $\log f(\omega)$:

$$\log f(\omega) = -\log 2\pi - 2d\log\left(2\sin\frac{\omega}{2}\right) \tag{71}$$

Similar to the hypothesis testing in correlogram estimation part, the null hypothesis here is also based on a categorical regression. Let Y represent logarithm of estimated power values by periodogram and X represent the corresponding frequencies. Use D to be the dummy variable (with value equal to 0 or 1) to identify the estimations from FARIMA(0,d,0) processes and approximated ARMA processes. Then we have the following linear equations to regress:

$$Y_i = c + \gamma D_i + \alpha X_i + \beta\left(D_i X_i\right) + \varepsilon_i \tag{72}$$

where ε_i are iid Gaussian innovations.

Similarly the null hypothesis in our statistical inference is:

$$H_0 : \beta = 0 \tag{73}$$

and we say the samples from FARIMA(0,d,0) and approximated ARMA processes are indistinguishable if we can not reject the null hypothesis at a given significant level.

The procedures in this simulation are the same as the correlogram estimation's and it is not described again here. We show the simulation results of $d = 0.3$ and $d = 0.45$ in Table 10 and Table 11 respectively. Similar to the results in the statistical inference from correlogram estimation, the results show that statistical inference can not distinguish the samples from original and approximated processes with finite sample size when dimension of approximation is large enough. Also we find that statistical inferences based on GPH is weaker than the ones from correlogram estimation, which may be related with the raw periodogram regression. Other estimations of the spectral density are also be used to estimate the fractional differencing parameter d. For example, Reisen [57] applied smoothed periodogram to estimate d, Robinson [58] proposed a new of the log-periodogram regression etc.

Table 10: Hypothesis testings based on GPH estimation of FARIMA(0,0.3,0) and its approximation

	n				
N	2	5	10	20	50
$5,000$	0.566	0.000	0.000	0.000	0.000
$10,000$	0.990	0.000	0.000	0.000	0.000
$50,000$	1.000	0.008	0.000	0.000	0.000
$100,000$	1.000	1.000	0.000	0.000	0.000
$1,000,000$	1.000	1.000	1.000	0.000	0.000
Distance	5×10^{-2}	1×10^{-2}	5×10^{-4}	3×10^{-5}	4×10^{-6}

Table 11: Hypothesis testings based on GPH estimation of FARIMA(0,0.45,0) and its approximation

	n				
N	2	5	10	20	50
$5,000$	0.996	0.000	0.000	0.000	0.000
$10,000$	1.000	0.000	0.000	0.000	0.000
$50,000$	1.000	0.998	0.000	0.000	0.000
$100,000$	1.000	1.000	0.000	0.000	0.000
$1,000,000$	1.000	1.000	1.000	0.000	0.000
$2,000,000$	1.000	1.000	1.000	0.004	0.000
Distance	3×10^{-2}	5×10^{-3}	7×10^{-4}	1×10^{-4}	2×10^{-5}

n and N represent the dimensions of approximated ARMA processes and sample size respectively. The numbers in the table represent the fraction of 500 trials where inference correlogram estimations described in (Eq.72 and 73) leads to reject the null hypothesis.

4.3.3 Conditional distributions

With regard to the application in European vanilla option pricing, we care more about the conditional distributions of underlying asset prices instead of the serial dependency, which has been examined in the previous parts. For the valuation of European vanilla option, it's completely determined by the distributions of the underlying asset prices at the expiration time. So if statistical inference can not tell the difference between the distributions of underlying asset returns with long memory and short memory processes with finite sample size, the option valuations are not distinguishable neither.

To implement the statistical tests on conditional distributions, we have a "true" price distribution generated by the FARIMA(0,d,0) and an approximated price distribution. Here we apply the widely-used statistical tests for distributions, Kolmogorov-Smirnov test. Let $S_t^{(s,1)}$ and $S_t^{(s,2)}$ represent the s_{th} simulated price trajectories at time t from "true" and approximated price processes respectively. $1 \leq s \leq S$ and $1 \leq t \leq T$. Given S and T, we have the null hypothesis:

$$H_0 : S_T^{(s,1)} \text{ and } S_T^{(s,2)} \text{ are sampled from the same distribution} \tag{74}$$

In this simulation study, the static parameters are chosen as: $S_0 = 30$, $\sigma = 15\%$, $h = 1\text{min}$ and $T = 1/4$. We show the Kolmogorov-Smirnov statistics with varying fractional differencing parameters d, sample size S and approximated dimensions n.

Fig 4 shows how the Kolmogorov-Smirnov statistics change with different fractional differencing parameter d and approximated dimensions n. The sample size S is fixed to be $1,000$ here. In Fig 4, we find that 1) given the same approximated dimension, the Kolmogorov-Smirnov static is larger with larger d. Or in other word, when dimension of approximated ARMA processes is the same, it is easier to tell the difference between samples when d is larger; 2) when approximated ARMA dimensions increase, it is harder to tell difference between samples.

Figure 4: Kolmogorov-Smirnov statistics with varying fractional differencing parameter d and approximated dimensions n

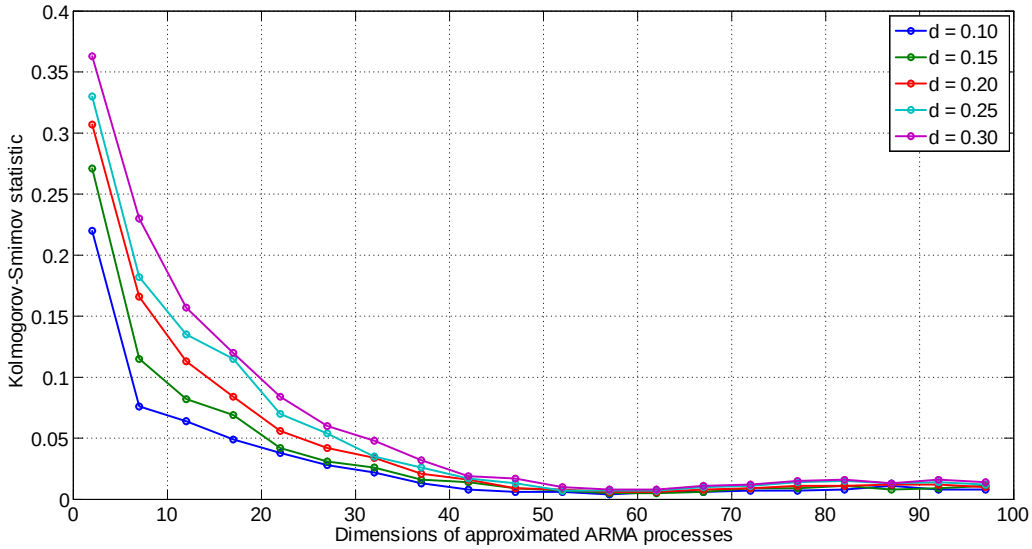


Table 12 and Table 13 are two numerical studies show the Kolmogorov-Smirnov statistics and the p-values to reject the null hypothesis when d is chosen to be 0.2 and 0.3 respectively. We find that 1) when the approximated dimension is low, K-S statistic distinguish with small amount of samples, i.e. $n = 2$ and $S = 1,000$ in Table 12; 2) when the approximated dimension is large, K-S statistics can not distinguish the samples even with large sample size, i.e. $n = 100$ and $S = 100,000$ in Table 12; 3) it requires higher dimensions of approximation to not to reject the null hypothesis when the fractional differencing parameter is larger. These results are consistent with our our theoretical analysis in previous sections that it is hard to reject the null hypothesis when the dimension of approximations is large with finite sample size.

Table 12: Hypothesis testings(K-S) on conditional distributions of FARIMA(0,0.2,0) and its approximations

	Dimensions n				
Sample size S	2	10	20	50	500
100	0.3200	0.1600	0.0900	0.0400	0.0100
	(0.0000)	(0.1400)	(0.7942)	(1.0000)	(1.0000)
$1,000$	0.2920	0.1380	0.0720	0.0090	0.0020
	(0.0000)	(0.0000)	(0.0106)	(1.0000)	(1.0000)
$10,000$	0.2899	0.1225	0.0590	0.0029	0.0004
	(0.0000)	(0.0000)	(0.0000)	(1.0000)	(1.0000)
$100,000$	0.2815	0.1174	0.0567	0.0018	0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.9964)	(1.0000)

Table 13: Hypothesis testings(K-S) on conditional distributions of FARIMA(0,0.3,0) and its approximations

	Dimensions n				
Sample size S	2	10	20	50	500
100	0.4000	0.2600	0.1600	0.0300	0.0200
	(0.0000)	(0.0018)	(0.1400)	(1.0000)	(1.0000)
$1,000$	0.3730	0.1990	0.0960	0.0110	0.0030
	(0.0000)	(0.0000)	(0.0002)	(1.0000)	(1.0000)
$10,000$	0.3555	0.1775	0.0902	0.0052	0.0005
	(0.0000)	(0.0000)	(0.0000)	(0.9992)	(1.0000)
$100,000$	0.3556	0.1786	0.0887	0.0036	0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.5499)	(1.0000)

5 Option pricing with long memory process

In the previous sections, we have already shown that a long memory process, such as FARIMA(p,d,q), can be close enough at information distance. Based on the connection between total variation distance and information distance, the total variation distance between these two processes can also be sufficiently small. Also the numerical simulations show that given finite number of samples (large enough), the samples from FARIMA process with long memory property and stationary ARMA process can not be distinguished statistically at a certain confidence level.

As for the pricing of derivatives, Black and Scholes in 1973 [40] and Merton in 1973 [41] gave the theoretical estimates of the European option price, which will be briefly reviewed in the following parts. The Black-Scholes model assumes the underlying asset returns follow geometric Brownian motion. But it is also well-know that financial returns have significantly serial correlation in empirical finance, which does not obey the Black-Scholes model assumption. Voluminous works have shown that financial returns are not independent from the historical data. Some influential pioneering works are Fama in 1965 [42] and 1970 [45], Lo and Mackinlay in 1988 [44], Fama and French in 1988 [46], Jegadeesh in 1990 [43] etc.

Even though financial returns have significant serial correlation, the Black-Scholes model can still be applied to the predictable underlying asset returns given some constraints. For example, Grundy in 1991 [47] shows that the Black-Scholes model holds under a Ornstein-Uhlenbeck(O-U) process, which generates the returns with serial correlation. Lo and Wang in 1995 [48] constructed an adjustment to the Black-Scholes model for the predictability of asset returns. Moreover, Wang et al. [49] showed the Black-Scholes model can be applied to the underlying asset returns which follow ARMA process and derived the closed-form formula.

However, it is extremely hard, or impossible, to price the option when the underlying asset return series follows a long memory process. Since the price process is required to be a semi-martingale to eliminate the arbitrage opportunities [50] and

long memory process, such as FARIMA process, is not a semi-martingale process, there will always exist the arbitrage opportunities for a long memory price process. So to price option with long memory process, firstly we approximate the long memory process by a high dimensions ARMA process, which is information distance close enough to the original process. Then the option price can be determined by the Black-Scholes model with ARMA process [49].

5.1 Review of Black-Scholes model

Black-Scholes option pricing model has been widely used in the derivatives pricing in practice. We briefly review the assumptions and formulas of Black-Scholes option pricing model here.

The fundamental assumptions of classic Black-Scholes model are 1) the option is European style option, which can only be exercised at expiration; 2) the returns of underlying asset is a geometric Brownian motion; 3) there is no transaction cost involved; 4) there is no dividends; 5) there is no arbitrage opportunity. With these assumptions, the well-known Black-Scholes equation is:

$$\frac{\partial V(S,t)}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V(S,t)}{\partial S} + rS\frac{\partial V(S,t)}{\partial S} - rV = 0 \tag{75}$$

where S is the price of stock, $V(S,t)$ is the price of a derivative, r is the risk-free rate and σ is the volatility of stock's returns.

With the initial condition and the payoff function of derivative, the European call option C and put option P have the following closed form prices:

$$C = N(d_1) S_t - N(d_2) K e^{-r(T-t)} \tag{76}$$

$$P = N(-d_2) K e^{-r(T-t)} - N(-d_1) S_t \tag{77}$$

where $d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[\ln(S_t/K) + \left(r + \sigma^2/2\right)(T-t)\right]$, $d_2 = d_1 - \sigma\sqrt{T-t}$, $N(\cdot)$ is the cumulative distribution function of standard normal distribution, S_t represents the stock price at t, K represents the stick price and $(T-t)$ is the time to maturity.

5.2 European option pricing with ARMA process

When returns of stock prices S_t follow a stationary ARMA(p,q) model, it can be written as

$$\ln \frac{S_t}{S_{t-1}} = \mu + \sum_{k=1}^{p} \phi_k \ln \frac{S_{t-k}}{S_{t-k-1}} + \sigma \sum_{k=0}^{q} \theta_k \varepsilon_{t-k} \tag{78}$$

where ϕ_k and θ_k are the AR and MA coefficients respectively, σ is the constant volatility coefficient, ε_t are the standard normal innovations. Let z represent the lag operator, we have

$$\ln \frac{S_t}{S_{t-1}} = \mu + \left(\sum_{k=1}^{p} \phi_k z^k \right) \ln \frac{S_t}{S_{t-1}} + \sigma \left(\sum_{k=0}^{q} \theta_k z^k \right) \varepsilon_t \tag{79}$$

$$\ln \frac{S_t}{S_{t-1}} = \frac{\mu}{\left(1 - \sum_{k=1}^{p} \phi_k\right)} + \sigma \frac{\left(\sum_{k=0}^{q} \theta_k z^k\right)}{\left(1 - \sum_{k=1}^{p} \phi_k z^k\right)} \varepsilon_t \tag{80}$$

and we know that it can be written in the $MA(\infty)$ form:

$$\ln \frac{S_t}{S_{t-1}} = \omega + \sigma \left(\sum_{k=1}^{\infty} f_k z^k \right) \varepsilon_t \tag{81}$$

where f_k is the k_{th} element in the impulse response of this ARMA process.

When the stock price S_t follows a stationary ARMA model, Wang et al. [49] gave the closed form solutions for its European option pricing by using martingale pricing method.

$$C = N\left(\tilde{d}_1\right) S_t - N\left(\tilde{d}_2\right) K e^{-r(T-t)} \tag{82}$$

$$P = N\left(-\tilde{d}_2\right) K e^{-r(T-t)} - N\left(-\tilde{d}_1\right) S_t \tag{83}$$

where

$$\tilde{d}_1 = \frac{1}{\sigma\sqrt{T-t}} \left[\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\tilde{\sigma}^2}{2}\right)(T-t) \right] \tag{84}$$

$$\tilde{d}_2 = \tilde{d}_1 - \tilde{\sigma}\sqrt{T-t} \tag{85}$$

So the only difference between European option price with ARMA process and classic Black-Scholes model is the calculation of volatility. As for the calculation of volatility function $\tilde{\sigma}$

$$\tilde{\sigma} = \sqrt{\frac{V_n}{T-t}} \tag{86}$$

Assume the sampling frequency is h, n is the integer part of $\frac{T-t}{h}$ and f_k is the coefficient of k_{th} lag operator in eq.(81), the V_n can be calculated by

$$V_n = \sigma^2 \sum_{j=0}^{n-1} \left(1 + \sum_{i=1}^{j} f_i\right)^2 h \tag{87}$$

So when the price process is a geometric Brownian motion, $f_k = 0$, for $k \geq 1$. Then the volatility (eq.86) is

$$\tilde{\sigma} = \sqrt{\frac{\sigma^2 \sum_{j=0}^{n-1} h}{T-t}} = \sigma \sqrt{\frac{T-t}{T-t}} = \sigma \tag{88}$$

and the option pricing formula is the same as Black-Scholes model.

5.3 Option valuation of long memory processes

To compare our vanilla European option valuation and Black-Scholes model, we show the calculated option prices based on our long memory option valuation approach with various long memory processes when $S_0 = 100$, $K = 100$, $r_f = 0.25\%$, $\sigma = 20\%$ and $h = 10$mins. Table 14 shows the ratio of our option valuation to the BSM prices. Numerically we show that 1) existence of long memory process affect the option valuation dramatically; 2) price valuations based on our approach converge with higher approximated dimension.

Table 14: ATM call option valuation of FARIMA(0,d,0)

	n					
d	2	10	50	100	200	1000
−0.40	71%	47%	47%	47%	47%	47%
−0.35	67%	47%	47%	47%	47%	47%
−0.30	62%	47%	47%	47%	47%	47%
−0.25	56%	47%	47%	47%	47%	47%
−0.20	50%	47%	48%	48%	48%	48%
−0.15	47%	51%	51%	51%	51%	51%
−0.10	47%	60%	59%	59%	59%	59%
−0.05	50%	75%	74%	74%	74%	74%
0.00	63%	97%	100%	100%	100%	100%
0.05	81%	130%	142%	142%	142%	142%
0.10	103%	176%	209%	208%	208%	208%
0.15	130%	240%	312%	310%	310%	310%
0.20	161%	328%	464%	459%	460%	460%
0.25	196%	444%	661%	653%	653%	653%
0.30	237%	585%	852%	844%	844%	844%
0.35	284%	738%	946%	943%	943%	943%
0.40	337%	867%	957%	957%	957%	957%

d represents the fractional differencing parameter and n represents the number of approximated ARMA processes dimensions. The static parameters for option pricing are: $S_0 = 100$, $K = 100$, $r_f = 0.25\%$, $\sigma = 20\%$, $h = 10$ mins. The entries in this table represent the ratio of approximated Black-Scholes-ARMA option valuation to BSM prices.

Since we are using high dimensions ARMA processes to approximate long memory processes, we calculate an ATM call option price with different approximated dimensions in ARMA model to see whether the option price converges when dimension increases. In this experiment, we select the static parameters $S_0 = 100$, $r_f = 0.25\%$, $\sigma = 20\%$, $h = 10$ mins and $d = 0.1$. Fig 5 and Fig 6 show how the call option prices changes with different strike prices or times to maturities. In Fig 5 and Fig 6, n represents the dimensions of approximated ARMA processes.

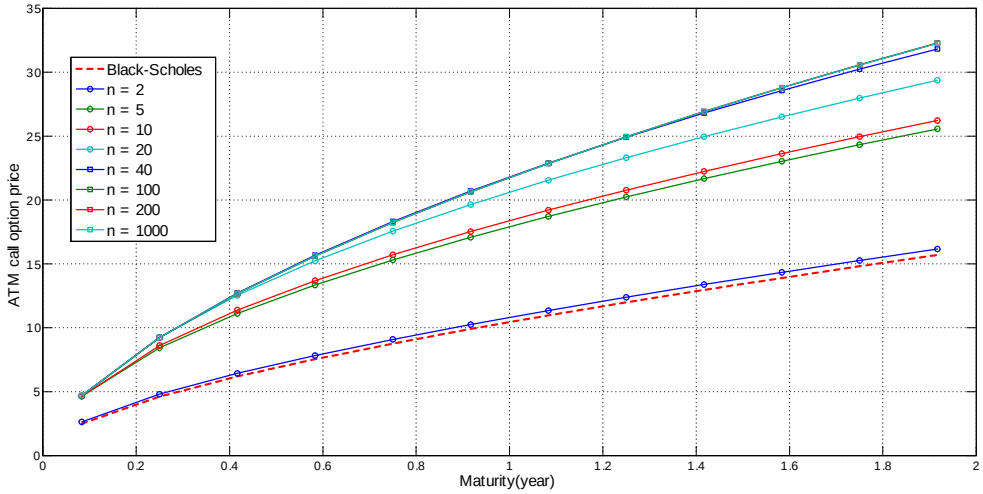


Figure 5: ATM call option valuations of FARIMA(0,0.1,0) with different approximated dimensions and time to maturity

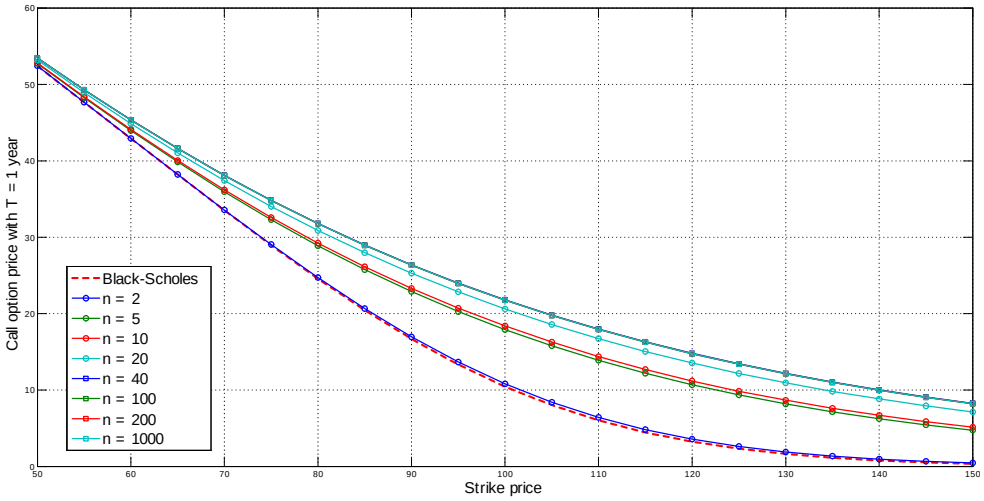


Figure 6: ATM call option valuations of FARIMA(0,0.1,0) with different approximated dimensions and strike prices

From Fig 5 and Fig 6, we can see that when dimensions of approximated ARMA processes is low, the option prices are close to the Black-Scholes option prices. When the dimension n is large enough ($n = 100$ here), the option prices are hardly to be distinguished.

To observe how the long memory processes affect the option valuations, we implement an experiment with various fractional differencing parameters d in FARIMA(0,d,0) processes. The other static parameters are chosen to be: $S_0 = 100$, $r_f = 0.25\%$, $\sigma = 20\%$, $h = 10$mins and $n = 1,000$. Similarly we have Fig 7 and 8 show the changes of ATM call option prices with different T and the changes of call option prices with different strike prices respectively.

Figure 7: ATM call option valuations of FARIMA(0,d,0) with different d and time to maturity

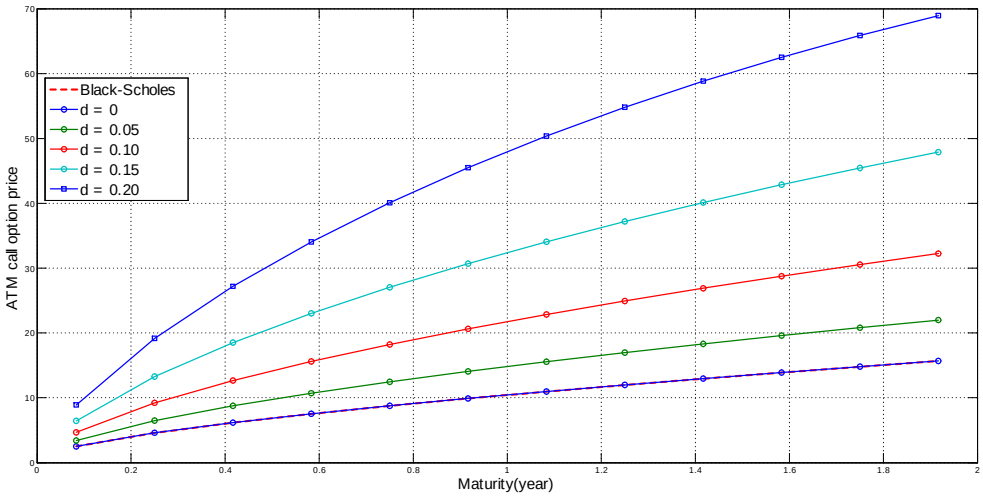
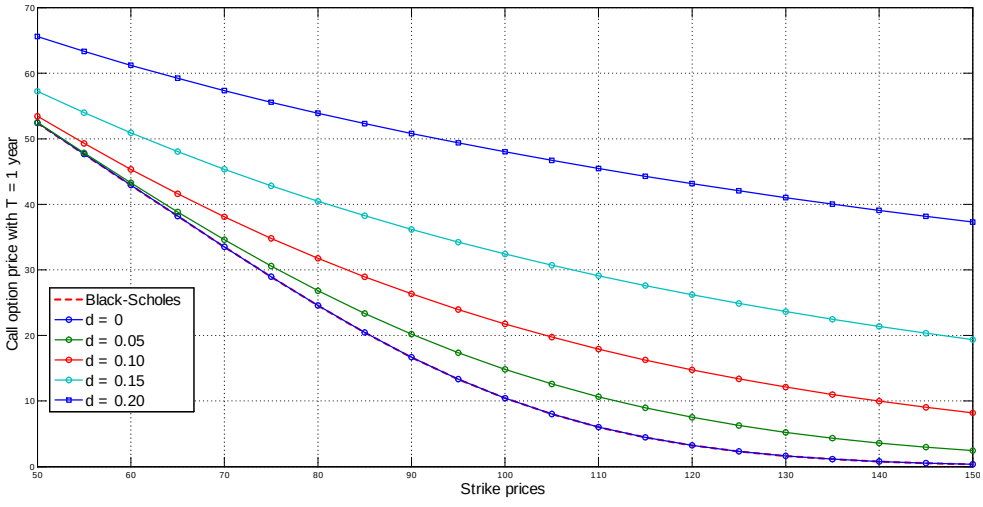


Figure 8: ATM call option valuations of FARIMA(0,d,0) with different d and strike prices



From Fig 7 and Fig 8, we can see that when $d = 0$, the underlying prices are

following geometric Brownian motion and the Black-Scholes-ARMA option prices are the same as Black-Scholes'. Also when d is higher, the option valuation is higher given maturity and strike price.

5.4 Implied volatility surface

In Black-Scholes model, the volatility is assumed to be a constant. But in empirical finance, volatility surface calculated from the market option prices are not flat and there exists both significant volatility smile and term structure of volatility. In this part, we analyze the implied volatility surface of our long memory option prices. The implied volatility is calculated calculated in the classic Black-Scholes model.

To distinguish it from Black-Scholes model, we use $C_{ARMA}\left(S_0, K, \sigma, r_f, T, h, \psi\left(z\right)\right)$ to represent the risk-neutral valuation of European call option price in Black-Scholes model with ARMA(p,q) process, which is described previously.

In $C_{ARMA}\left(S_0, K, \sigma, r_f, T, h, \psi\left(z\right)\right)$, S_0 represents the stock price at t_0, K represents the strike price, σ represents the annual volatility constant, r_f is the annual risk-free interest rate, T is the time from t_0 to the option expiration, h is the sampling frequency in the ARMA processes and $\psi\left(z\right)$ is the characteristic equation of approximated ARMA process. So the risk-neutral valuation of European call option price can be calculated given these parameters.

To get the implied volatility surface, including the volatility smile and term structure of volatility, the following procedures are implemented:

1. Assume the underlying asset price is following a stochastic model which is a long memory process. Here we are using FARIMA(p,d,q) here with $d > 0$;

2. Use the long memory stochastic model FARIMA(p,d,q) as the price trajectory generator and generate $S = 1,000$ price paths with fixed price length T^*. So now we have $S_t^{(s)}$, $1 \leq t \leq T^*$, $1 \leq s \leq S$, to represent the price path in each price path;

3. For each trajectory $S_t^{(s)}$, we fit our ARMA(p,q) process, which is equivalently a linear time invariant system, to the return series and estimate the characteristic equation $\psi^{(s)}(z)$, $1 \leq s \leq S$.

4. To get the evaluation of the option in the Black-Scholes-ARMA model, we use $S_{T^*}^{(s)}$ as the stock price at t_0. We have K, r_f, T, σ and h fixed and characteristic equations $\psi^{(s)}(z)$ are estimated in each trajectory. Then we can get the evaluation of the European call option price $C_{ARMA}\left(S_{T^*}^{(s)}, K, \sigma, r_f, T, h, \psi^{(s)}(z)\right)$ in the s_{th} trajectory with moneyness $\frac{S_{T^*}^{(s)}}{K}$.

5. After calculating the call option price in each trajectory, the implied volatility $\tilde{\sigma}^{(s)}$ can be calculated in the Black-Scholes models. So now we can see whether the volatility smile exists from the implied volatility $\tilde{\sigma}^{(s)}$ with different moneyness $\frac{S_{T^*}^{(s)}}{K}$, $1 \leq s \leq S$.

6. Similarly by changing the maturity T, we can get the term structure of volatility and the volatility surface.

Based on the above-mentioned procedure and different approximation processes (different dimensions of ARMA process), we will analyze the following questions numerically:

1. Whether our approximation of a long memory process, which is FARIMA(0,d,0) here, can produce an option valuation closer to the reality. In the previous chapters, we know that the higher dimensions the ARMA process is, the shorter information distance between ARMA and FARIMA(0,d,0) processes. So we observe the Black-Scholes implied volatilities with different dimensions to examine whether the underlying idea is correct.

2. How does the long memory return process affect the our option valuation? We examine the Black-Scholes implied volatility with different fractional differencing parameter d in the price path generator.

3. What is the term structure of volatility in a long memory process option valuation model? It is examined by changing the time to maturity T in our long memory option valuation model.

5.4.1 Implied volatility in different approximations

Using the FARIMA(0,d,0) processes with $d > 0$ as the underlying returns stochastic process in the simulation, we could generate the long memory return series. We are using a stationary ARMA(p,q) process to approximate the long memory process and the information distance is smaller between them when the dimensions of the ARMA model increases. Here we examine the impact on implied volatility with different approximated ARMA processes' dimensions. In this simulation, the variable is the dimensions of approximated ARMA processes and the other parameters are chosen in Table 15. We only examine the difference of implied volatility against the moneyness level here, rather than the term structure of volatility.

S_0	K	r_f	T(year)	h(min)	σ	d
100	100	0.25%	0.5	5	5%	0.1

Table 15: Parameter settings in simulation with different dimensions of approximations

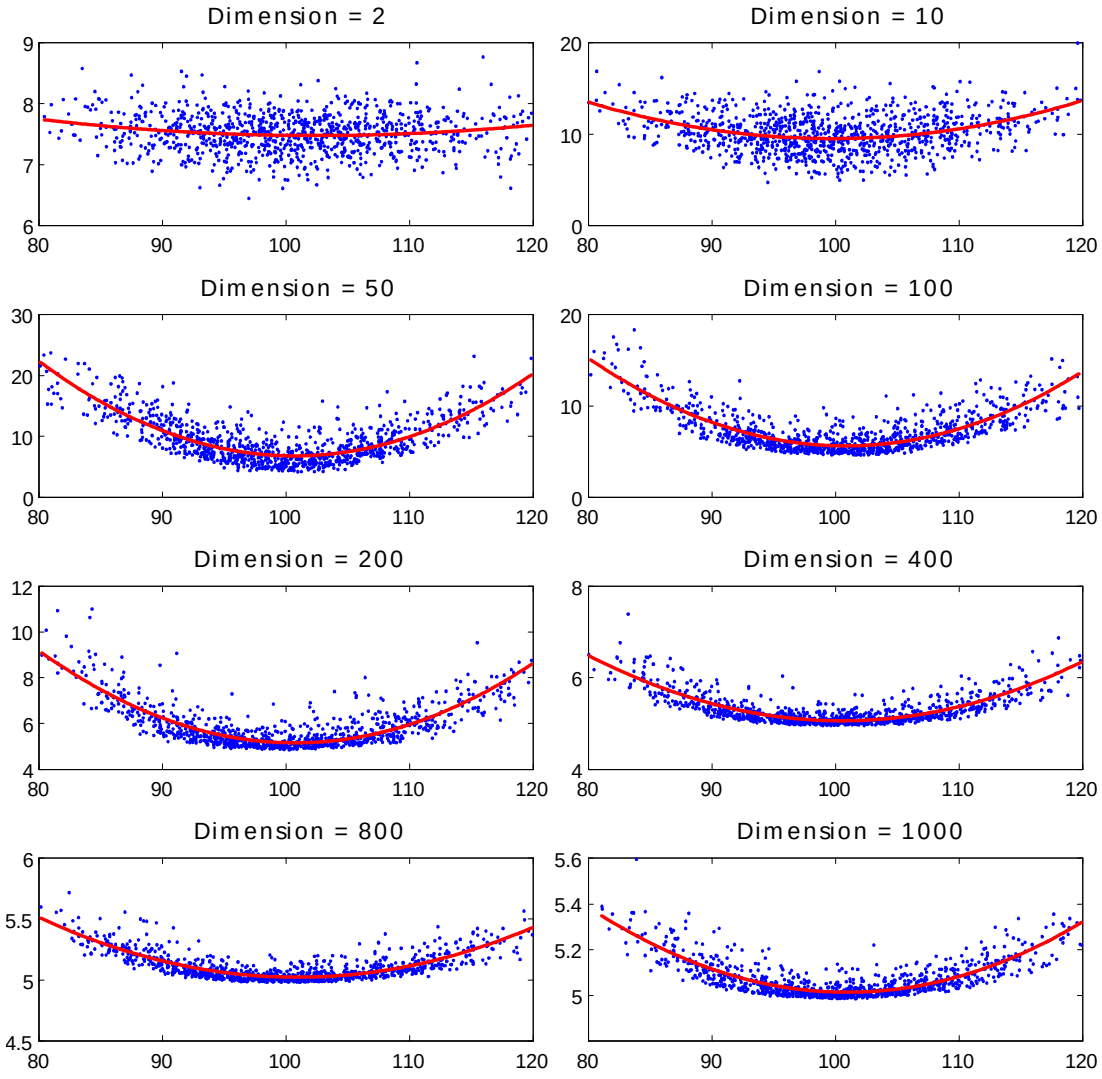
Table 16 displays how the implied volatility changes with different moneyness level and dimensions of approximated ARMA processes.

	Dimension of approximated ARMA processes							
moneyness(%)	2	10	50	100	200	400	800	1,000
80	7.75	13.48	22.34	15.21	9.19	6.49	5.51	5.38
82	7.70	12.72	19.49	13.46	8.45	6.22	5.42	5.32
84	7.66	12.04	16.92	11.88	7.79	5.98	5.34	5.26
86	7.62	11.44	14.64	10.48	7.20	5.77	5.27	5.20
88	7.59	10.92	12.65	9.26	6.68	5.59	5.21	5.16
90	7.56	10.48	10.95	8.21	6.24	5.43	5.16	5.12
92	7.54	10.13	9.54	7.34	5.87	5.30	5.11	5.08
94	7.52	9.85	8.43	6.65	5.58	5.20	5.08	5.05
96	7.50	9.66	7.60	6.14	5.37	5.13	5.05	5.03
98	7.49	9.55	7.06	5.80	5.23	5.08	5.03	5.02
100	7.48	9.51	6.81	5.64	5.16	5.06	5.03	5.01
102	7.48	9.56	6.86	5.66	5.17	5.07	5.03	5.01
104	7.48	9.69	7.19	5.85	5.26	5.10	5.03	5.02
106	7.49	9.91	7.82	6.22	5.41	5.16	5.05	5.04
108	7.50	10.20	8.73	6.77	5.65	5.25	5.08	5.06
110	7.51	10.58	9.93	7.50	5.96	5.37	5.12	5.08
112	7.53	11.03	11.43	8.40	6.34	5.51	5.16	5.12
114	7.55	11.57	13.21	9.48	6.80	5.68	5.22	5.16
116	7.58	12.19	15.29	10.73	7.33	5.88	5.28	5.21
118	7.61	12.88	17.65	12.17	7.94	6.10	5.35	5.26
120	7.65	13.67	20.31	13.78	8.63	6.35	5.43	5.32

Table 16: Implied volatility in different dimensions of approximated ARMA processes

Fig 9 shows the changes of volatility smiles with different dimensions of approximated ARMA processes.

Figure 9: Implied volatility with different dimensions of ARMA processes



In these subplots, the x-axis is the moneyness(%), the y-axis is the implied volatility(%) calculated in the Black-Scholes model. The blue dots are the data points in the simulation and the red line is the least square fitted polynomial function (degree is 2) to these data.

5.4.2 Implied volatility in different underlying long memory processes

In this part, using FARIMA(0,d,0) with $d > 0$ to generate the price paths with long memory property, we show that the implied volatility based on Black-Scholes model and our option valuation model has the volatility smile effect.

To investigate the impact of long memory process on option valuation, we test various fractional differencing parameters d to observe the changes of implied volatility. The way we calculate the implied volatility from Black-Scholes model and our valuation of option price has been described in the previous section.

In this simulation, the static parameters are chosen as follows

S_0	K	r_f	$T(year)$	$h(\min)$	σ	np
100	100	0.25%	0.5	5	5%	400

Table 17: Simulation parameters setting

	Fractional differencing parameter d							
moneyness(%)	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
80	6.32	6.42	6.62	7.16	7.56	7.78	11.36	17.70
82	6.07	6.17	6.36	6.86	7.31	7.69	11.03	16.90
84	5.85	5.95	6.13	6.59	7.07	7.62	10.73	16.17
86	5.66	5.75	5.92	6.34	6.87	7.56	10.46	15.51
88	5.48	5.58	5.74	6.13	6.69	7.50	10.22	14.90
90	5.34	5.43	5.59	5.95	6.54	7.46	10.02	14.36
92	5.22	5.31	5.46	5.80	6.41	7.43	9.85	13.88
94	5.12	5.21	5.36	5.68	6.30	7.40	9.72	13.47
96	5.06	5.14	5.28	5.59	6.23	7.39	9.62	13.11
98	5.01	5.09	5.23	5.53	6.17	7.39	9.55	12.82
100	4.99	5.07	5.20	5.50	6.15	7.39	9.52	12.60
102	5.00	5.07	5.20	5.50	6.15	7.41	9.52	12.43
104	5.04	5.10	5.23	5.54	6.17	7.43	9.55	12.33
106	5.10	5.15	5.28	5.60	6.22	7.47	9.62	12.29
108	5.18	5.23	5.35	5.69	6.30	7.51	9.72	12.32
110	5.29	5.33	5.45	5.82	6.40	7.57	9.86	12.40
112	5.43	5.46	5.58	5.97	6.52	7.63	10.03	12.55
114	5.59	5.61	5.73	6.16	6.68	7.70	10.23	12.77
116	5.78	5.79	5.91	6.37	6.86	7.79	10.47	13.04
118	5.99	6.00	6.12	6.62	7.06	7.88	10.74	13.38
120	6.23	6.22	6.35	6.90	7.29	7.99	11.04	13.78

Table 18: Implied volatility in different FARIMA(0,d,0) models

Figure 10: Comparison of volatility smile with different fractional differencing parameters

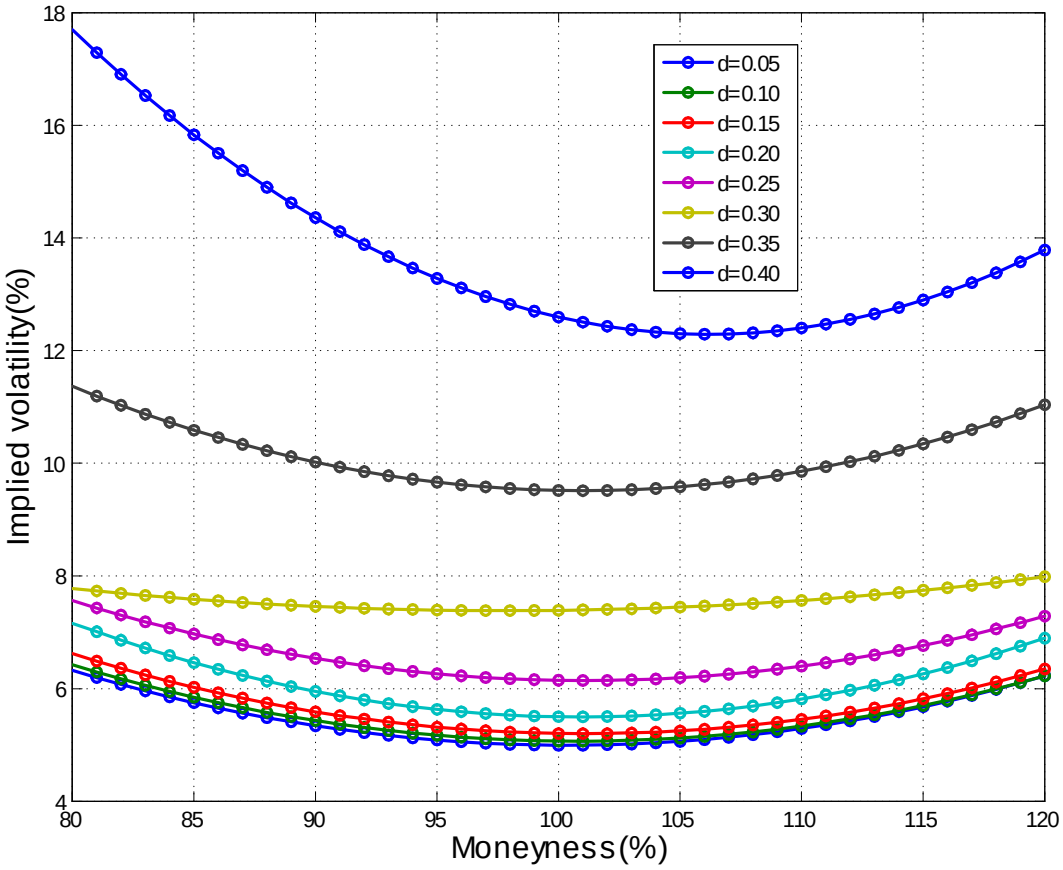
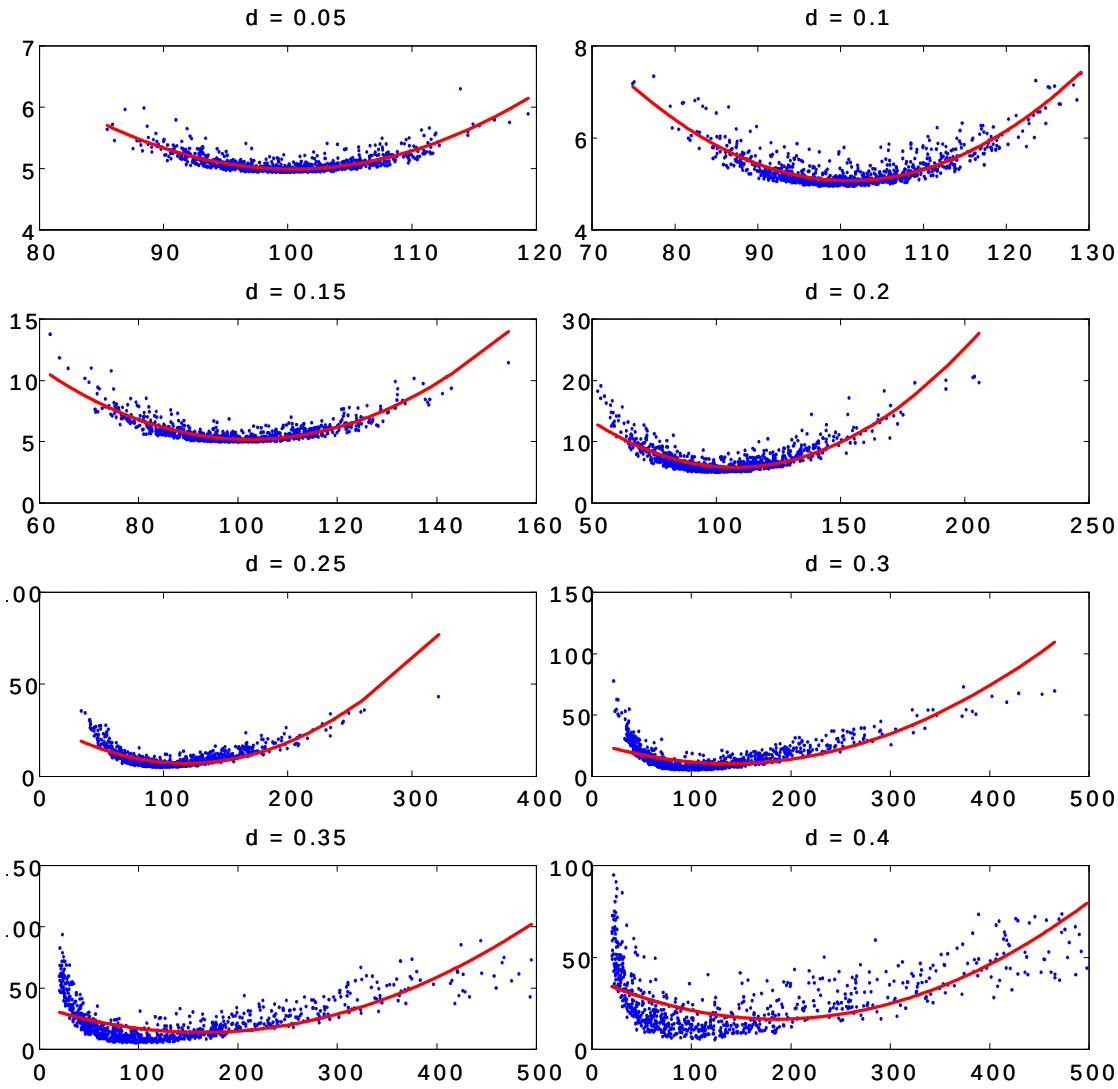


Figure 11: Implied volatility with different FARIMA(0,d,0) processes



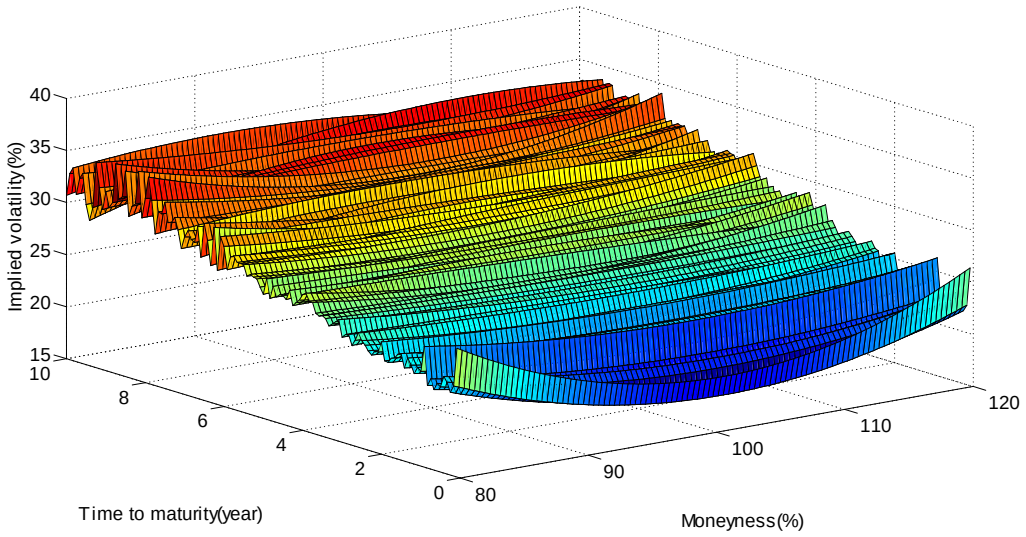
In these subplots, the x-axis is the moneyness(%), the y-axis is the implied volatility(%) calculated in the Black-Scholes model. The blue dots are the data points in the simulation and the red line is the least square fitted polynomial function (degree is 2) to these data.

5.4.3 Volatility surface with GARCH(p,q) process

Financial return data are well-known to have volatility clustering, heavy tail properties which Gaussian distributions do not have. Volatility clustering increases the short-term volatility, which can be seen in the term structure of volatility. To produce the short-term volatility clustering in simulation, we combine the generalized autoregressive conditional heteroskedasticity model, GARCH model with the FARIMA(0,d,0) model.

The reason we want to add the GARCH model into the price generator in our simulation is to introduce the short term volatility clustering, which is closer to the reality. Fig 12 shows the volatility surface based on solely FARIMA(0,0.1,0) return generator. Because of the existence of long memory, the implied volatility is exploding when time to maturity is longer and it is opposite to the reality. So to increase the variety of the return generator in our simulation, we add GARCH model for the short-term volatility clustering.

Figure 12: Volatility surface with d = 0.1



GARCH model is widely used to model realized volatility of financial time series. Let σ_t be a time series of the realized volatility and GARCH(p,q) is defined as follows [51]:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{89}$$

where ε_t is the white noise process, $\alpha_i > 0$, $\beta_i > 0$ and $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$.

So instead of simulating the price paths $S_t^{(s)}$ from a FARIMA(0,d,0) process in s_{th} scenario, we split the return series (after removing the drift part) into a FARIMA(0,d,0) process and a GARCH(1,1) model:

$$\ln \frac{S_t^{(s)}}{S_{t-1}^{(s)}} = \sigma_t^{(s)} u_t^{(s)} \tag{90}$$

where $u_t^{(s)}$ are the fractional Gaussian noise generated from the FARIMA(0,d,0) process and $\sigma_t^{(s)}$ is modeled by GARCH(1,1) model(eq. 89).

In the simulation, we use the GARCH(1,1) parameters estimated from the daily return series of S&P 500 index (SPX), which are from 1/3/1990 to 10/21/2013. The estimated parameters and statistics are shown in Table 19, which shows the parameters estimated values are statistically significant.

Parameter	Value	Standard Error	T statistic
α_0	1.03×10^{-6}	1.26×10^{-7}	8.14
α_1	0.075	0.0044	16.96
β_1	0.917	0.0049	185.70

Table 19: GARCH(1,1) parameter estimation with SPX daily return series

Fig 13 shows the conditional standard deviation σ_t in GARCH(1,1) model with the SPX daily returns. Also the strong evidence of volatility clustering and short-term volatility explosion can been seen in this example.

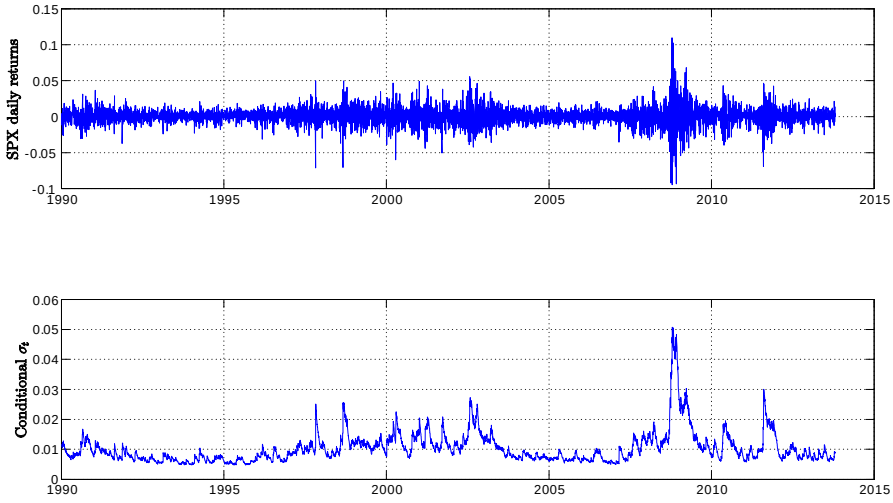


Figure 13: Conditional standard deviation in GARCH(1,1)

By adding the estimated GARCH(1,1) model with SPX daily returns, we implement the following steps to get the volatility surface from our option valuation model.

Step 1: The underlying asset returns are assumed to follow a stochastic model which combines a long memory process FARIMA(0,d,0) and a short term volatility model GARCH(1,1), which parameters are estimated from SPX daily returns data.

Step 2: Given the price path length T^*, we generate $S_1 = 100$ conditional volatility paths from the given GARCH(1,1) model and they are denoted as $\sigma_t^{(s_1)}$, where $1 \leq t \leq T^*$ and $1 \leq s_1 \leq S_1$.

Step 3: Given each conditional volatility path $\sigma_t^{(s_1)}$ generated in the previous step, we generate $S_2 = 100$ price paths $S_t^{(s_1,s_2)}$, where $1 \leq t \leq T^*$, $1 \leq s_1 \leq S_1$ and $1 \leq s_2 \leq S_2$. The price paths $S_t^{(s_1,s_2)}$ are simulated based on (eq. 90):

$$\ln \frac{S_t^{(s_1,s_2)}}{S_{t-1}^{(s_1,s_2)}} = \sigma_t^{(s_1)} u_t^{(s_2)}$$

where $u_t^{(s_2)}$ is generated from a FARIMA(0,d,0) process with $d > 0$.

Step 4: Now we have $S = S_1 S_2 = 10,000$ scenarios of price trajectories. For each trajectory $S_t^{(s_1,s_2)}$, we fit our ARMA(p,q) process, which is equivalently a linear time invariant system, to the return series and estimate the characteristic equation $\psi^{(s_1,s_2)}(z)$, where $1 \leq s_1 \leq S_1$ and $1 \leq s_2 \leq S_2$.

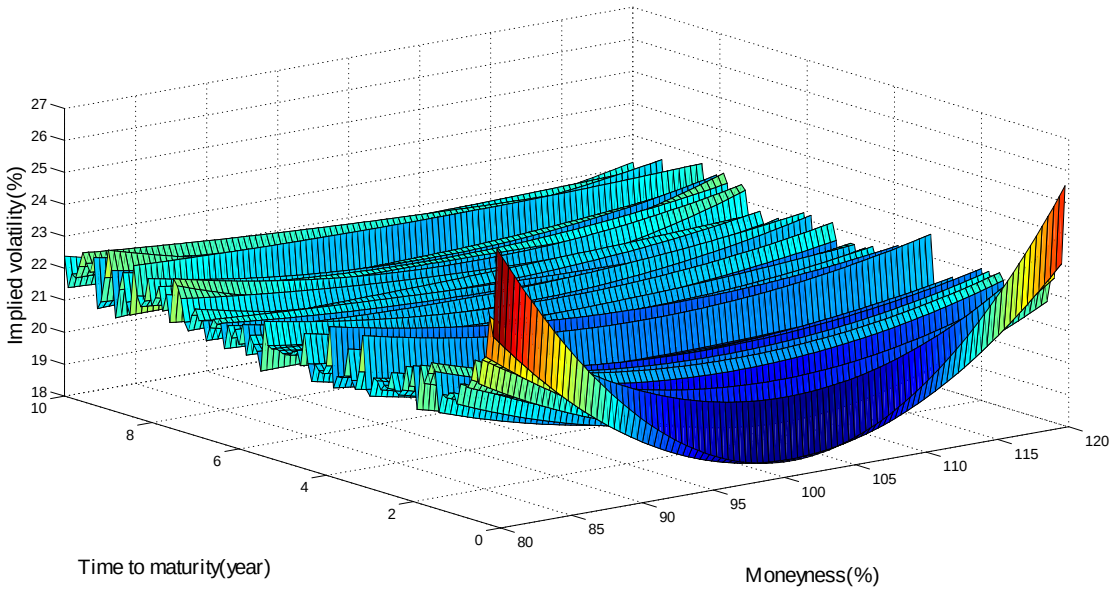
Step 5: To get the evaluation of the option in the Black-Scholes-ARMA model, we use $S_{T^*}^{(s_1,s_2)}$ as the stock price at t_0. We have K, r_f, σ and h fixed, characteristic equations $\psi^{(s_1,s_2)}(z)$ are estimated in each trajectory and time to maturity T is chosen to be the same as T^*. Then we can get the evaluation of the European call option price $C_{ARMA}\left(S_{T^*}^{(s_1,s_2)}, K, \sigma, r_f, T^*, h, \psi^{(s_1,s_2)}(z)\right)$ in the (s_1, s_2) trajectory with moneyness $\frac{S_{T^*}^{(s_1,s_2)}}{K}$.

Step 6: After calculating the call option price in each trajectory, the implied volatility $\tilde{\sigma}^{(s_1,s_2)}$ can be calculated from the Black-Scholes models.

Step 7: By using different time to maturity T^* in these simulations, we can get the volatility surface with different time to maturity and moneyness.

Fig 14 shows the implied volatility surface when the fractional differencing parameter d is chosen to be 0.1.

Figure 14: Volatility surface with d = 0.1 and GARCH(1,1)



In Fig 14, we can find that volatility smile becomes less significant when the time to maturity is further, which is consistent with the phenomenon observed in the financial data.

When we increase the fractional differencing parameter d in the simulations, we find that the long memory FARIMA(0,d,0) part dominates the short term volatility GARCH(1,1) part and the implied volatility increases when the time to maturity is further. However, the volatility smile is still less significant with longer time to maturity. We give an example with $d = 0.2$ in Fig 15.

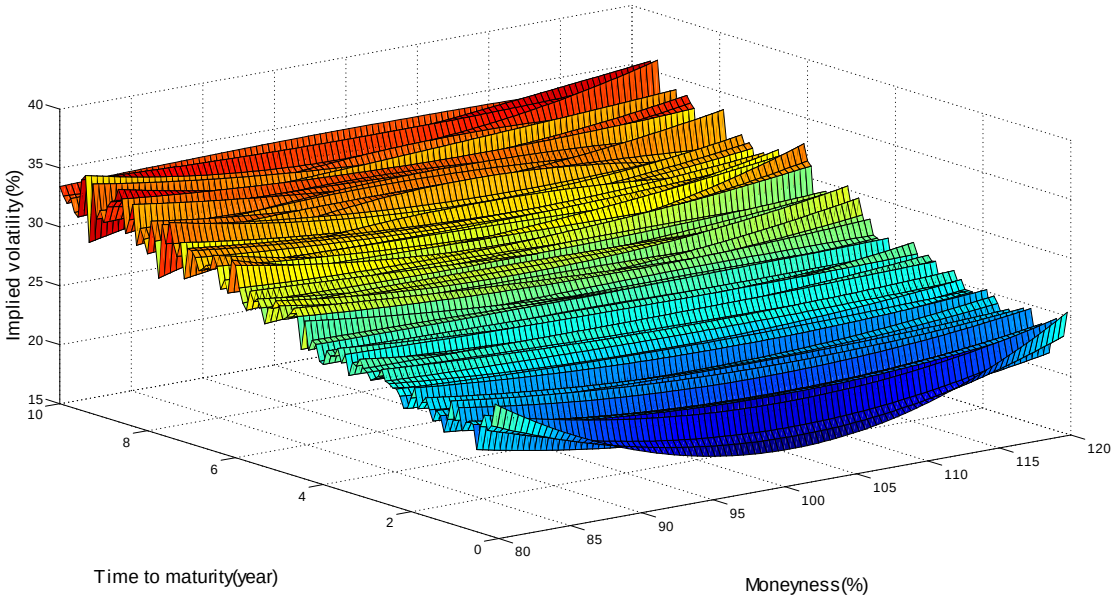


Figure 15: Volatility surface with d = 0.2 and GARCH(1,1)

6 Empirical finance at high frequency

We have reviewed some of the pioneering works and literatures of the existence of long memory process in financial time series in the "Introduction" section. However, few works have been done on the high frequency financial data. In this section, we examine the long memory phenomenon with high frequency financial data and we show our major findings of empirical finance results in this section.

Based on the high frequency financial returns and volatility time series, we apply these widely used statistical techniques, which have been mentioned previously, of the long memory process. Also we examine long memory phenomena on equity markets and spot foreign exchange (FX) markets. Because there are much fewer works about long memory process in FX market published than equity markets, we focus more on the FX part in this study.

6.1 Long memory in FX market

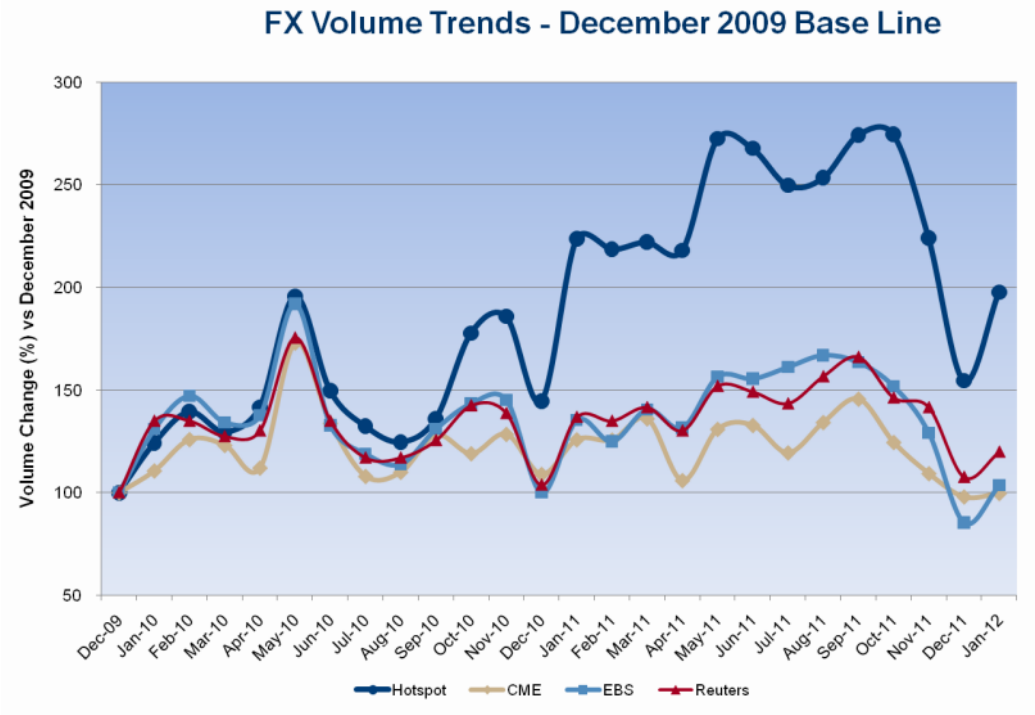
Data in this empirical research are coming from the tick by tick quote database, recorded in Hotspot trading venue[2]. The data set are the seven most liquid currencies in the world, which are called G7 currencies in the following sections. The G7 currencies are Australian Dollar (AUD), Canadian Dollar (CAD), Swiss Franc (CHF), Euro (EUR), British Pound (GBP), Japanese Yen (JPY) and US Dollar (USD). The largest factor that we only used G7 currencies here is the liquidity.

Compared to other trading instruments, FX trading is more liquid and multi-venues traded. Instead of a single electronic exchange, various trading venues exist, such as Reuters, Hotspot, FXCM and a lot of single-bank ecommerce platforms, such as Citi, Barclays, etc [63]. For this reason, it is almost impossible to get an completely aggregate FX order book among the entire existed trading venues. In this empirical study, we apply the tick data from Hotspot, which is very dense and has higher trading volumes than most other trading venues. The following graph is

[2]The Hotspot data for this study are provided and authorized by Presagium (USA) LLC.

the FX volume trend in some major trading venues[3]:

Figure 16: FX Volume Trends



This graph shows us that it is reasonable to use Hotspot quote data to represent the FX trading rates. The top five currency pairs by volume in Hotspot are EUR/USD(47.33%), AUD/USD(10.74%), USD/JPY(9.52%), GBP/USD(8.92%), USD/CAD(5.05%).[4] G10 currencies contribute the majority of trading volumes in FX trading and that is why only G10 currencies are under the consideration in this empirical study.

There are few published research concerning the long memory processes in ultra high frequency trading. We extract the last mid-price of ask price and bid price

[3]Sources: Hotspot Monthly Statistics, Jan 2012

[4]Source: Hotspot Monthly Statistics, Jan 2012

every second. The data range from 2008-01-01 00:00:01 to 2011-12-31-23:59:59, which include 4 years data. Also there exists 30 currency pairs generated from the G10 currencies, such as USD/CHF, AUD/CAD, AUD/NZD etc.

FX market is a 24×5 trading instrument and generally trading events do not occur during weekend (trading typically starts 2100GMT Sunday). We exclude all non-trading periods, such as weekend, holiday, etc. Even though FX is a 24 hours trading instrument, there is no guarantee that there will be new quotes placed every second, especially for some less liquid currency pairs. For example, considering EUR/CHF, there are far fewer quotes generated in Asia trading time than the New York and London trading time. If there are no new quotes during any one second, we exclude that second in our data.

6.1.1 R/S statistics

The null hypothesis for the test: long memory does not exist in the return series. Table 20 shows the classical and modified R/S statistics of selected FX pairs return series. The null hypothesis can not be rejected based on the results and there is no strong evidence to say that there exist long memory process in high frequency (second) FX return series.

Because it is hard to decide the lag parameter q in the modified R/S statistic, we apply some short memory processes to remove the short range dependence in original series. Then we use normalized classical range scaled statistics V_c to test the existence of long memory in the residuals. $AR\,(1)$, $MA\,(1)$, $ARMA\,(1,1)$ and $ARMA\,(1,1) - GARCH\,(1,1)$ processes are fitted to the data to remove the short range dependence components.

The critical values of the modified R/S statistic are given by Lo [6]: 90% confidence interval is $[0.861, 1.747]$; 95% confidence interval is $[0.809, 1.862]$ and 99% confidence interval is $[0.721, 2.098]$. Table **??** shows the results of the test and we find that none of the tested currency pairs can reject the null hypothesis. After

removing the short range dependence part by filtering through these short memory processes, we still can not get the conclusion that there exists long memory in return series of FX rate in ultra high frequency.

We now show the results for the same test, but based on absolute returns, which shows that strong long memory exists in the absolute return series. The result is completely different from the returns'. There is a very strong long memory process exist in the absolute second returns of G7 FX rates.

Table 20: Classic and modified R/S statistics of FX return series (second)

r_t	V_c	$V_c^{(1)}$	$V_c^{(2)}$	$V_c^{(3)}$	$V_c^{(4)}$	$V_{(5)}$	$V_{(10)}$	$V_{(20)}$	$V_{(40)}$	$V_{(80)}$
AUD/CAD	0.65	0.73	0.75	0.85	0.89	0.88	0.80	0.84	0.87	0.90
AUD/CHF	0.65	0.79	0.86	0.94	0.93	0.97	0.89	0.94	0.97	0.99
AUD/JPY	1.23	1.26	1.26	1.26	1.22	1.22	1.26	1.24	1.23	1.21
AUD/USD	1.22	1.25	1.26	1.29	1.19	1.28	1.29	1.29	1.28	1.28
CAD/JPY	1.09	1.15	1.09	1.17	1.09	1.18	1.17	1.18	1.18	1.18
CHF/JPY	1.03	1.18	1.23	1.33	1.20	1.35	1.28	1.32	1.34	1.37
EUR/AUD	0.77	0.86	0.89	1.00	1.08	1.05	0.95	1.00	1.04	1.07
EUR/CAD	0.77	0.86	0.89	0.99	1.01	1.05	0.94	0.99	1.03	1.08
EUR/CHF	1.04	1.13	1.15	1.24	1.29	1.27	1.21	1.24	1.26	1.28
EUR/GBP	0.86	1.00	0.86	1.04	0.98	1.09	1.03	1.06	1.08	1.11
EUR/JPY	0.97	0.99	0.99	1.02	0.97	0.99	1.00	1.00	0.99	1.00
EUR/USD	0.88	0.94	0.94	0.99	0.83	0.98	0.97	0.97	0.98	0.99
GBP/AUD	0.77	0.86	0.88	0.95	0.97	0.98	0.92	0.95	0.98	1.00
GBP/CAD	0.58	0.64	0.66	0.73	0.74	0.77	0.70	0.73	0.76	0.79
GBP/CHF	0.73	0.83	0.86	0.91	0.90	0.96	0.89	0.92	0.95	0.97
GBP/JPY	1.21	1.28	1.29	1.33	1.20	1.31	1.31	1.32	1.31	1.31
GBP/USD	1.34	1.37	1.38	1.37	1.27	1.37	1.38	1.38	1.37	1.37
USD/CAD	1.12	1.17	1.17	1.25	1.11	1.24	1.22	1.23	1.24	1.25
USD/CHF	1.47	1.48	1.48	1.56	1.44	1.55	1.53	1.54	1.55	1.56
USD/JPY	0.63	0.69	0.71	0.74	0.62	0.73	0.71	0.72	0.73	0.73

V_c: classic R/S statistics; $V_c^{(1)}$: classic R/S statistics after filtering by AR(1); $V_c^{(2)}$: classic R/S statistics after filtering by MA(1); $V_c^{(3)}$: classic R/S statistics after filtering by ARMA(1,1); $V_c^{(4)}$: classic R/S statistics after filtering by ARMA(1,1)-GARCH(1,1); $V_{(q)}$: modified R/S statistics with lag parameter q; (**) and (*) denote

the statistics is significant at 99% and 95% level respectively.

Table 21: Classic and modified R/S statistics of absolute values of FX return series (second)

$\|r_t\|$	V_c	$V_{(5)}$	$V_{(10)}$	$V_{(20)}$	$V_{(40)}$	$V_{(80)}$
AUD/CAD	631.52**	195.88**	353.84**	279.86**	214.46**	160.88**
AUD/CHF	486.14**	150.10**	272.33**	215.33**	164.58**	122.79**
AUD/JPY	658.89**	210.69**	374.09**	297.38**	230.01**	173.98**
AUD/USD	778.26**	226.29**	414.67**	325.05**	247.92**	185.66**
CAD/JPY	500.25**	158.98**	281.40**	224.45**	173.57**	131.32**
CHF/JPY	330.55**	102.42**	181.42**	143.79**	111.38**	85.56**
EUR/AUD	616.96**	203.31**	353.25**	283.54**	221.22**	169.17**
EUR/CAD	522.89**	180.47**	308.49**	249.45**	195.90**	151.08**
EUR/CHF	292.86**	98.86**	168.90**	136.25**	107.14**	83.11**
EUR/GBP	505.59**	148.47**	263.18**	206.26**	160.39**	127.23**
EUR/JPY	470.75**	164.86**	281.53**	227.19**	178.52**	139.15**
EUR/USD	385.32**	134.52**	219.96**	177.82**	143.78**	116.94**
GBP/AUD	646.62**	211.03**	369.53**	296.24**	230.08**	174.77**
GBP/CAD	516.26**	178.96**	307.40**	248.49**	194.65**	148.80**
GBP/CHF	423.04**	136.55**	237.09**	189.96**	148.36**	114.20**
GBP/JPY	432.19**	152.55**	256.01**	207.66**	164.70**	129.33**
GBP/USD	522.40**	179.73**	304.41**	245.98**	194.44**	151.32**
USD/CAD	662.87**	205.53**	366.00**	290.54**	224.27**	170.11**
USD/CHF	271.76**	102.81**	167.64**	138.37**	110.84**	87.32**
USD/JPY	393.40**	128.53**	214.05**	171.54**	137.48**	111.69**

$V_{(q)}$: modified R/S statistics with lag parameter q; (**) and (*) denote the statistics is significant at 99% and 95% level respectively.

6.1.2 Correlogram estimations

As discussed before, correlogram estimation is much less accurate when the long memory effect is less significant. So we only apply the correlogram estimation to the absolute return to see whether there is strong evidence of long memory in the absolute return series. To see whether the result is robust, or the tail of autocorrelation decays stably, we estimate it by different number of lags N. Table 22 and Table 23 show the estimated Hurst exponents when $N = 100$ and $N = 10,000$ respectively. These results also show that very strong long memory exists in the absolute return of FX rates at the frequency of second, which stands the same as modified rescaled range results. All of the selected currency pairs have strong long memory and most of the estimated Hurst exponents are even larger than 0.9. AUD/USD has the highest Hurst exponent estimation, which is 0.97. This level of long memory has not previously been reported in published research.

Table 22: Correlogram estimations of FX absolute returns (second) with the number of lags equal to 100

$N = 100$	$\hat{H}$	95% CI	99% CI
AUD/CAD	0.952310	[0.952270, 0.952350]	[0.952302, 0.952318]
AUD/CHF	0.961603	[0.961537, 0.961669]	[0.961590, 0.961616]
AUD/JPY	0.943674	[0.943594, 0.943753]	[0.943658, 0.943689]
AUD/USD	0.950410	[0.950359, 0.950460]	[0.950399, 0.950420]
CAD/JPY	0.940665	[0.940607, 0.940722]	[0.940653, 0.940676]
CHF/JPY	0.908390	[0.908293, 0.908488]	[0.908371, 0.908410]
EUR/AUD	0.925992	[0.925948, 0.926037]	[0.925983, 0.926001]
EUR/CAD	0.912354	[0.912326, 0.912383]	[0.912349, 0.912360]
EUR/CHF	0.900993	[0.900909, 0.901077]	[0.900976, 0.901010]
EUR/GBP	0.816872	[0.816493, 0.817252]	[0.816797, 0.816948]
EUR/JPY	0.894664	[0.894516, 0.894812]	[0.894634, 0.894694]
EUR/USD	0.828689	[0.828376, 0.829002]	[0.828626, 0.828752]
GBP/AUD	0.936937	[0.936873, 0.937001]	[0.936924, 0.936950]
GBP/CAD	0.932685	[0.932622, 0.932748]	[0.932672, 0.932698]
GBP/CHF	0.909436	[0.909398, 0.909473]	[0.909428, 0.909443]
GBP/JPY	0.892867	[0.892708, 0.893026]	[0.892835, 0.892899]
GBP/USD	0.905427	[0.905304, 0.905550]	[0.905403, 0.905452]
USD/CAD	0.935134	[0.935081, 0.935187]	[0.935123, 0.935145]
USD/CHF	0.886113	[0.886010, 0.886216]	[0.886092, 0.886133]
USD/JPY	0.818747	[0.818466, 0.819029]	[0.818691, 0.818803]

Table 23: Correlogram estimations of FX absolute returns (second) with the number of lags equal to 1000

$N = 10000$	$\hat{H}$	95% CI	99% CI
AUD/CAD	0.968843	[0.968840, 0.968846]	[0.968842, 0.968844]
AUD/CHF	0.929510	[0.929498, 0.929521]	[0.929507, 0.929512]
AUD/JPY	0.961892	[0.961890, 0.961894]	[0.961892, 0.961893]
AUD/USD	0.974766	[0.974764, 0.974768]	[0.974766, 0.974766]
CAD/JPY	0.929779	[0.929777, 0.929782]	[0.929779, 0.929780]
CHF/JPY	0.853738	[0.853724, 0.853751]	[0.853735, 0.853740]
EUR/AUD	0.956222	[0.956219, 0.956225]	[0.956221, 0.956222]
EUR/CAD	0.925398	[0.925391, 0.925405]	[0.925397, 0.925400]
EUR/CHF	0.900955	[0.900950, 0.900960]	[0.900954, 0.900956]
EUR/GBP	0.939420	[0.939408, 0.939431]	[0.939417, 0.939422]
EUR/JPY	0.919640	[0.919636, 0.919643]	[0.919639, 0.919640]
EUR/USD	0.922862	[0.922852, 0.922871]	[0.922860, 0.922864]
GBP/AUD	0.954527	[0.954525, 0.954529]	[0.954527, 0.954527]
GBP/CAD	0.939754	[0.939751, 0.939757]	[0.939753, 0.939754]
GBP/CHF	0.901365	[0.901360, 0.901369]	[0.901364, 0.901366]
GBP/JPY	0.924787	[0.924784, 0.924791]	[0.924787, 0.924788]
GBP/USD	0.952395	[0.952391, 0.952399]	[0.952394, 0.952396]
USD/CAD	0.963177	[0.963173, 0.963180]	[0.963176, 0.963177]
USD/CHF	0.881305	[0.881301, 0.881309]	[0.881304, 0.881306]
USD/JPY	0.929824	[0.929815, 0.929833]	[0.929822, 0.929826]

6.1.3 GPH estimation

We implement the GPH estimation on the high frequency FX returns and absolute return series. To see whether our data are sensitive to the choice of sample length

n in estimating the fractional differencing parameters, we try different n in this empirical study, $n = T^{0.45}, T^{0.5}, T^{0.55}$. Recalling the relationship between Hurst exponent H and fractional differencing parameter d, $H = 1/2 + d$, Table 24 and Table 25 show the estimated Hurst exponents of return series and absolute return series in GPH estimations respectively. Null hypothesis is: $d = 0$, or $H = 1/2$ equivalently.

In Table 24, only EUR/CHF and EUR/GBP have a $H \in (1/2, 0)$ statistically significant, but nulls are only marginally rejected for some specified choice of n. In consistent with other long memory test, there is no strong evidence to show the long memory process exists in FX return series at second frequency.

In Table 25, all of the currency pairs show a very strong long memory in the absolute return series and they are statistically significant at 99.99% level. The results agree to the empirical results of modified R/S statistics and correlogram results that there exist a very strong long memory process in absolute returns of FX rates at second frequency.

Table 24: GPH estimations of FX returns (second)

r_t	$T^{0.45}$	$T^{0.5}$	$T^{0.55}$		$T^{0.45}$	$T^{0.5}$	$T^{0.55}$
AUD/CAD	0.4348**	0.4477**	0.4433**	EUR/JPY	0.4803	0.4926	0.4935
	(0.0000)	(0.0000)	(0.0000)		(0.0953)	(0.3370)	(0.1876)
AUD/CHF	0.5122	0.4973	0.5011	EUR/USD	0.5138	0.4999	0.4969
	(0.3626)	(0.7503)	(0.8464)		(0.2706)	(0.9891)	(0.5373)
AUD/JPY	0.4782	0.4885	0.4981	GBP/AUD	0.4778	0.4759**	0.4844**
	(0.0895)	(0.1615)	(0.7239)		(0.0631)	(0.0019)	(0.0026)
AUD/USD	0.4946	0.4840	0.4865*	GBP/CAD	0.4656**	0.4791*	0.4793**
	(0.6703)	(0.0563)	(0.0117)		(0.0089)	(0.0121)	(0.0001)
CAD/JPY	0.4764	0.4856	0.5058	GBP/CHF	0.5090	0.5166*	0.5014
	(0.0859)	(0.1078)	(0.3191)		(0.4632)	(0.0352)	(0.7749)
CHF/JPY	0.4971	0.4926	0.4814**	GBP/JPY	0.0491	0.4839*	0.4898*
	(0.8179)	(0.3492)	(0.0002)		(0.4682)	(0.0368)	(0.0375)
EUR/AUD	0.4998	0.4898	0.4905	GBP/USD	0.4911	0.4956	0.4889*
	(0.9856)	(0.1919)	(0.0582)		(0.4873)	(0.5947)	(0.0357)
EUR/CAD	0.4916	0.4959	0.4947	USD/CAD	0.5111	0.5056	0.4994
	(0.4973)	(0.6064)	(0.3007)		(0.4043)	(0.5277)	(0.9242)
EUR/CHF	0.5276*	0.5051	0.4990	USD/CHF	0.5246	0.5074	0.4979
	(0.0371)	(0.5487)	(0.8580)		(0.0643)	(0.3774)	(0.6937)
EUR/GBP	0.5102	0.5186**	0.5061	USD/JPY	0.4707*	0.4777**	0.4852**
	(0.4108)	(0.0192)	(0.2367)		(0.0305)	(0.0099)	(0.0073)

Table 25: GPH estimations of absolute values of FX returns (second)

$\lvert r_t \rvert$	$T^{0.45}$	$T^{0.5}$	$T^{0.55}$		$T^{0.45}$	$T^{0.5}$	$T^{0.55}$
AUD/CAD	0.8863**	0.8698**	0.8859**	EUR/JPY	0.8163**	0.8545**	0.8939**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
AUD/CHF	0.9116**	0.9922**	1.0638**	EUR/USD	0.7802**	0.8242**	0.8058**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
AUD/JPY	0.8639**	0.9001**	0.9355**	GBP/AUD	0.8129**	0.8620**	0.8820**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
AUD/USD	0.8327**	0.8917**	0.9429**	GBP/CAD	0.7716**	0.8242**	0.8771**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
CAD/JPY	0.7898**	0.8376**	0.8637**	GBP/CHF	0.8229**	0.8252**	0.8508**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
CHF/JPY	0.7761**	0.7711**	0.8398**	GBP/JPY	0.8340**	0.8406**	0.8758**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
EUR/AUD	0.8619**	0.8774**	0.8875**	GBP/USD	0.7741**	0.8490**	0.8931**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
EUR/CAD	0.7546**	0.7671**	0.8067**	USD/CAD	0.7216**	0.8418**	0.8672**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
EUR/CHF	0.7936**	0.8349**	0.8287**	USD/CHF	0.7700**	0.8112**	0.8530**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
EUR/GBP	0.7936**	0.8349**	0.8287**	USD/JPY	0.8221**	0.8512**	0.8675**
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)

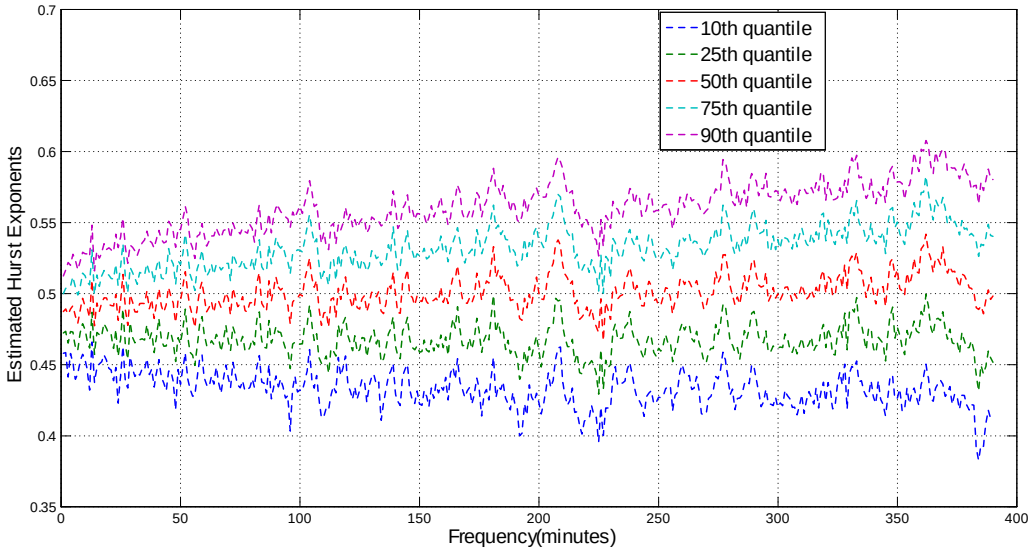
6.2 Long memory in equity market

There has been a long time history of the long memory research in equity market, such as Lo [6], Chow et al. [9], Mandelbrot [4], Lobato and Savin [10], Willinger et al. [11] etc. To examine the long memory effects in high frequency equity market, we are using S&P500 listed stocks data from the Trade and Quote, TAQ[5], database. We query 465 stocks out of S&P500 listed stocks and estimated the Hurst exponent at different frequencies varying from 1 minute to 390 minutes. There are numerous outliers in ultra high frequency data, which are usually caused by the misplacement of quotes or the quote recorded problem. Thompson's τ method [61,62] here is used to remove the statistically significant outliers.

After removing the ARMA-GARCH effect by filtering through ARMA(1,1)-GARCH(1,1) model, the Hurst exponents are estimated based on the innovations series. Fig 17 shows the estimated Hurst exponents (based on R/S statistics) quantities of the return series across 465 stocks. Based on the classic R/S estimation, more than 25% of the selected stocks show the long memory existence in their intraday return series. Also some of the stocks' Hurst estimation increase when the frequency is lower.

[5] TAQ: Trade and Quote database is a high frequency database at Stony Brook University. The data are originally collected from NYSE, extracted and maintained by Quantitative Finance Program at Stony Brook University.

Figure 17: Hurst exponents estimation(R/S) of S&P500 stocks r_t at different frequencies



Comparing to the debate on the existence of long memory in stock return series, realized volatility of the stock returns has been examined to pertain long memory property in many published works. Here we examine the Hurst exponents estimations of stocks realized volatility at a much higher frequency than most of the previous works. Fig 18 and Fig 19 show the estimated Hurst exponents quantities of realized volatilities across 465 stocks based on classic R/S and correlogram estimations respectively. Realized volatility series of most stocks are estimated to have long memory property in both studies and they show similar decaying pattern. The estimated Hurst exponents of realized volatility series decay with lower sampling frequencies, which is different from the pattern in return series.

Figure 18: Hurst exponents estimation(R/S) of S&P500 stocks $|r_t|$ at different frequencies

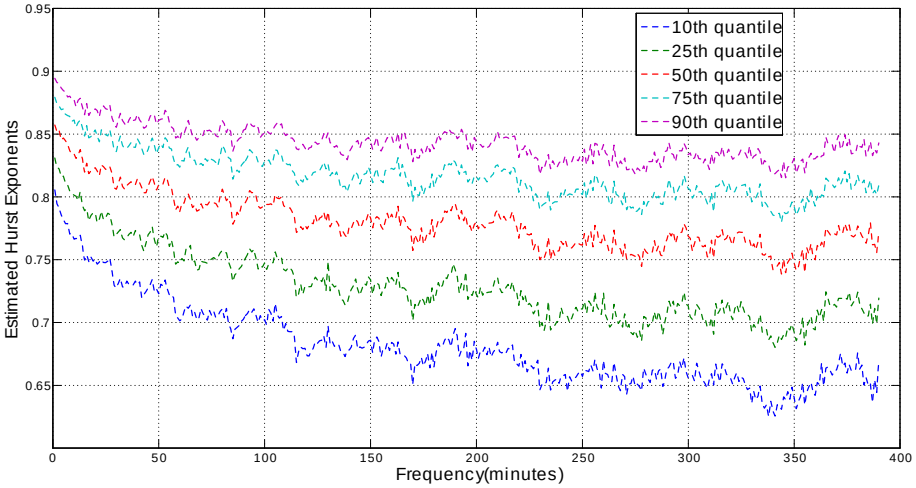
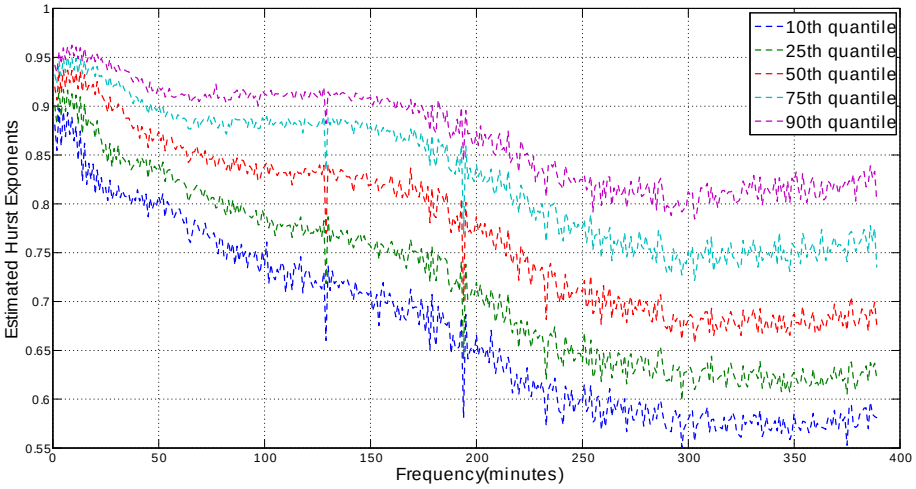


Figure 19: Hurst exponents estimation(Correlogram) of S&P500 stocks $|r_t|$ at different frequencies



7 Conclusion

In this thesis, we show that rational functions are information dense with the transfer functions that have finite information distance from the unit variance white noise. Under the connection among total variation, Hellinger and information distances, long memory, such as FARIMA(p,d,q), and short memory processes, such as ARMA(p,q), can not be distinguished by statistical inference with finite sample size.

Option pricing with long memory processes is difficult to achieve and we show that it can be valuated by Black-Scholes model based on approximated ARMA models. Also we show that volatility smile can be generated by this long memory option valuation approach. Adding short term volatility model to option valuation approach makes the implied volatility surface close the the reality numerically. Besides we examine the volatility in high frequency equity and FX world and classic statistical tests show strong evidence of long memory existence.

Appendix A. Equivalence of ARMA and LTI

.

ARMA representing form has numerical constraints in high dimensional calculations and state space representation is an equivalent representation of the ARMA stochastic process, but with better numerical stability. In this appendix, we briefly show the mathematical equivalence between ARMA and state space representations by one example.

Recall the definition of ARMA representation of $y(t)$ (eq. **??**),

$$y(t) = \sum_{k=1}^{p} \phi_k y(t-k) + x(t) + \sum_{k=1}^{q} \theta_k x(t-k) \tag{91}$$

where $x(t)$ are iid white noises.

Without loss of generality, assume $p = q = k$ since we can make the extra coefficients φ_k or θ_k to be zero, if $p \neq q$. Equivalently, we can represent $y(t)$ in (eq. 91) by the following state space form with various $z(t)$, A, B, C and D.

$$\begin{aligned} z(t+1) &= Az(t) + Bx(t) & (92)\\ y(t) &= Cz(t) + Dx(t) & (93) \end{aligned}$$

One possible solution to that is:

$$\begin{aligned} z(t) &= \left[\begin{array}{ccccccc} y(t) & \cdots & y(t-p) & x(t) & \cdots & x(t-q) \end{array} \right]^T_{2k\times 1}\\ x(t) &= x(t) \end{aligned}$$

and

$$
\begin{aligned}
A &= \begin{bmatrix} A_1 & A_2 \\ \mathbf{0} & A_3 \end{bmatrix}_{2k\times 2k} \\
B &= \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T_{2k\times 1} \\
C &= \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}^T_{2k\times 1} \\
D &= 1
\end{aligned}
$$

where

$$
A_1 = \begin{bmatrix} \varphi_1 & \cdots & \varphi_{p-1} & \varphi_p \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix}_{p\times p}, A_2 = \begin{bmatrix} \theta_1 & \cdots & \theta_{q-1} & \theta_q \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix}_{q\times q}, A_3 = \begin{bmatrix} \varphi_1 & \cdots & \varphi_{p-1} & \varphi_p \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix}_{p\times p}
$$

There are also other equivalent state space representations of ARMA model and readers are referred to [37] for more equivalent representations. The example we show is not the only solution for the state space representation of ARMA process. However, they have different numerical stabilities which we will show in the following part of appendix.

Appendix B. Triangular input balance representation of LTI

.

In this appendix, we briefly show the form of one state space representation, triangular input balanced representation, which has better numerical stability. Recall the state space representation of LTI:

$$z(t+1) = Az(t) + Bx(t) \tag{94}$$

$$y(t) = Cz(t) + x(t) \tag{95}$$

In a linear time invariant system, the Hankel matrix and impulse response can be factorized as

$$h(t) = \begin{pmatrix} CB & CAB & CA^2B & \cdots \end{pmatrix} \tag{96}$$

$$H = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix} \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \tag{97}$$

We say the input pairs is triangular input normal form if and only if

$$AA^* + BB^* = I \tag{98}$$

and A is a lower triangular matrix. We have the following by z transformation:

$$Y(z) = \left[C(zI - A)^{-1} B + I \right] X(z) \tag{99}$$

So the poles of linear time invariant system in triangular input balanced form are the eigenvalues of matrix A. Since A is a lower triangular matrix here, eigenvalues of A are the diagonal entries and it does not need any transformations to get the poles at all.

Known the poles of linear time invariant system, λ_k, we can get the explicit band fraction representation for the triangular input balance form(see Mullhaupt [35, 36]):

$$A = M^{-1}N, B = \rho_1 M^{-1} e_1$$

where M and N are sparse matrices and the form also increases the calculation speed dramatically:

$$M = \begin{pmatrix} 1 & & & \\ \gamma_1 & 1 & & \\ & \gamma_2 & \ddots & \\ & & \ddots & \ddots \end{pmatrix}, N = \begin{pmatrix} \lambda_1 & & & \\ \mu_1 & \lambda_2 & & \\ & \mu_2 & \ddots & \\ & & \ddots & \ddots \end{pmatrix} \tag{100}$$

where

$$\rho_k = \sqrt{1 - |\lambda_k|^2} \tag{101}$$

$$\mu_k = \frac{\rho_{k+1}}{\rho_k} \tag{102}$$

$$\gamma_k = \lambda_k^* \mu_k \tag{103}$$

Appendix C. Numerical stability of triangular input balance representationI

.

We compare numerical stability of ARMA and triangular input balance representations here. Rewrite the ARMA(p,q) by the lag operator:

$$\left(1-\varphi_1 L-\varphi_2 L^2-\cdots-\varphi_p L^p\right) y_t=\left(1+\theta_1 L+\theta_2 L^2+\cdots+\theta_q L^q\right) x_t \tag{104}$$

where x_t are iid white noises.

Then we can get

$$y_t = \frac{\left(1+\theta_1 L+\theta_2 L^2+\cdots+\theta_q L^q\right)}{\left(1-\varphi_1 L-\varphi_2 L^2-\cdots-\varphi_p L^p\right)} x_t \tag{105}$$

$$y_t = \frac{D(L)}{N(L)} x_t \tag{106}$$

where $D(L)$ and $N(L)$ are polynomial functions.

So ARMA models can be described by the two polynomial functions, $D(L)$ and $N(L)$. ARMA model is reduced ARMA form when the $D(L)$ and $N(L)$ do not have coprime.

Now considering the state space representation,

$$z(t+1) = Az(t)+Bx(t) \tag{107}$$

$$y(t) = Cz(t)+Dx(t) \tag{108}$$

Laplace-transform of the linear time invariant system will lead to:

$$Z(s) = (sI-A)^{-1} BX(s) \tag{109}$$

$$Y(s) = \left[C(sI-A)^{-1} B+I\right] X(s) \tag{110}$$

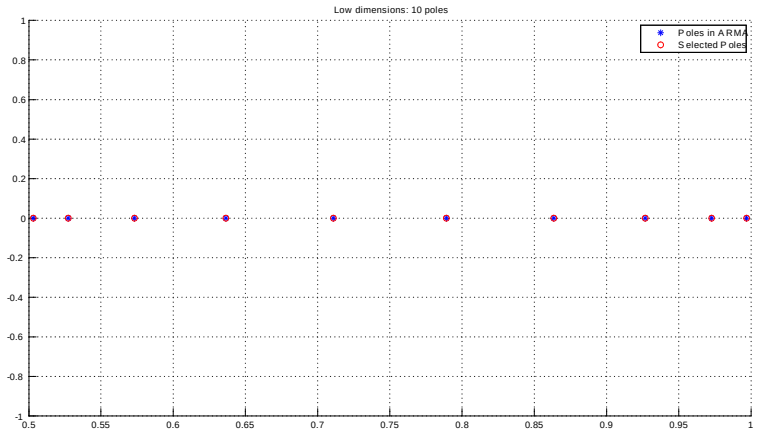
So the poles of the state space representation should be the solution set of:

$$\det|sI-A|=0 \tag{111}$$

which means that the poles of LTI system are the eigenvalues of matrix A. According to the equivalence between LTI systems and ARMA processes, the poles of LTI systems should be equal to the poles of ARMA process. So the eigenvalues of A should be the same as the roots of $N(L)$ in the characteristic function of ARMA processes.

One simple numerical test for instability in ARMA processes is to compare the location of given poles and corresponding roots of $N(L)$. Firstly let's take a look at the low dimension system shown in Fig 20, when the number of poles is small. In this example, we choose 10 poles to ARMA process and examine the roots of $N(L)$. We can see that in low dimension, roots of $N(L)$ are very stable and close to the selected poles. The most important thing is that all of the poles are still lying inside the unit disc.

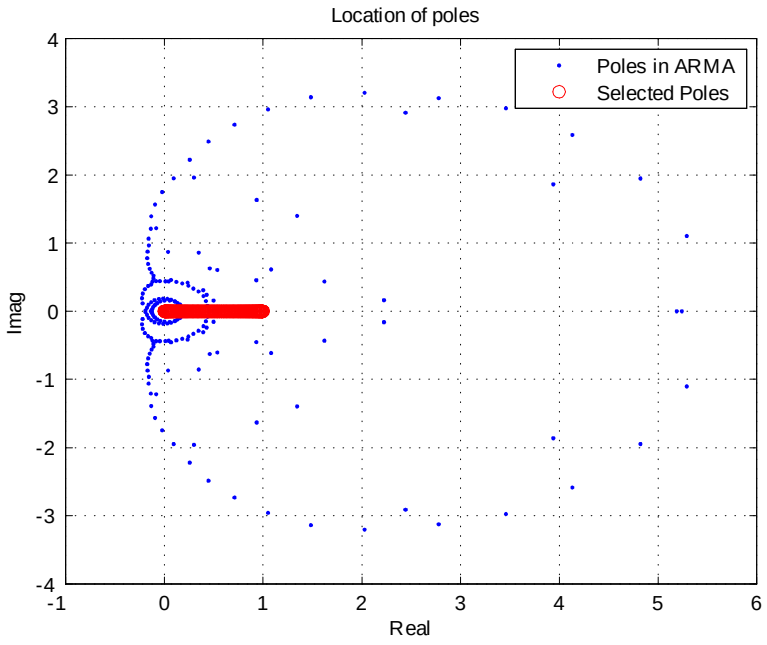
Figure 20: Location of poles of low dimension ARMA process



However, the stable phenomenon will change when we increase the dimensions of system. Fig 21 is one example for the high dimensions system. 1) Now 100 selected poles, inside the unit disc, are given to the ARMA process; 2) For the state space represented by triangular input balance form, it has exactly the same poles as what

we assign, because A is a lower triangular matrix; 3) Then we can observe that the roots of $N(L)$ lie outside the unit disc and will lead the system to be very unstable.

Figure 21: Location of poles of high dimension ARMA process



From these two simply numerical examples, it can be easily seen that ARMA process will become very unstable when the dimensions, or number of poles, increases. However, triangular input balance representation has a very stable poles location when dimension is high.

Appendix D. Example of information distance between FARIMA(0,d,0) and ARMA with fixed p or q

.

Proposition 1. *The information distance between any stationary AR(1) and FARIMA(0,d,0) with d>0 processes has a lower bound.*

Proof: *The information distance between AR(1) and FARIMA(0,d,0) processes is:*

$$\begin{aligned}\left\|\log f(z)-\log g(z)\right\|_{H^2}^2 &= \sum_{n=1}^{\infty}\left(\frac{\lambda^n}{n}-\frac{d}{n}\right)^2 \\ &= \sum_{n=1}^{\infty}\frac{\left(\lambda^2\right)^n}{n^2}-2d\sum_{n=1}^{\infty}\frac{\lambda^n}{n^2}+\sum_{n=1}^{\infty}\frac{d^2}{n^2} \\ &= Li_2\left(\lambda^2\right)-2dLi_2(\lambda)+\frac{\pi^2d^2}{6} \qquad (112)\end{aligned}$$

where $Li_n(\lambda)$ is the polylogarithm and it has these properties which will be used later:

$$\begin{aligned}\frac{\partial}{\partial\lambda}Li_n(\lambda) &= \frac{1}{\lambda}Li_{n-1}(\lambda) \qquad (113)\\ Li_0(\lambda) &= \frac{\lambda}{1-\lambda} \qquad (114)\\ Li_1(\lambda) &= -\log(1-\lambda) \qquad (115)\end{aligned}$$

To show the information distance between these two processes has a lower bound, we show the information distance is a convex function of λ.

$$\begin{aligned}&\frac{\partial^2}{\partial\lambda^2}\left\|\log f(z)-\log(z)\right\|_{H^2}^2 \\ =\ &\frac{\partial^2}{\partial\lambda^2}\left(Li_2\left(\lambda^2\right)-2dLi_2(\lambda)+\frac{\pi^2d^2}{6}\right) \\ =\ &\frac{2}{\lambda^2}\log\left(1-\lambda^2\right)+\frac{4}{\left(1-\lambda^2\right)}-\frac{2d}{\lambda^2}\log(1-\lambda)-\frac{2d}{\lambda(1-\lambda)} \\ =\ &\frac{2}{\lambda^2}\left(\frac{(2-d)\lambda^2-d\lambda}{1-\lambda^2}-\log\left(\frac{(1-\lambda)^d}{1-\lambda^2}\right)\right) \qquad (116)\end{aligned}$$

Because $\log\left(\frac{(1-\lambda)^d}{1-\lambda^2}\right) \leq \frac{(1-\lambda)^d}{1-\lambda^2} - 1$, *then we have*

$$\begin{aligned} & \frac{\partial^2}{\partial\lambda^2}\left\|\log f(z) - \log(z)\right\|_{H^2}^2 \\ \geq & \frac{2}{\lambda^2}\left(\frac{(2-d)\lambda^2 - d\lambda}{1-\lambda^2} - \frac{(1-\lambda)^d}{1-\lambda^2} + 1\right) \\ \geq & \frac{2}{\lambda^2\left(1-\lambda^2\right)}\left((1-d)\lambda^2 + (1-d\lambda) - (1-\lambda)^d\right) \end{aligned} \tag{117}$$

Let $\kappa(\lambda) = (1-d)\lambda^2 + (1-d\lambda) - (1-\lambda)^d$ *and* $|\lambda| < 1$, $0 < d < \frac{1}{2}$. *It can be easily shown that* $\kappa(\lambda) \geq \kappa(0) = 0$ *since* $\frac{\partial^2\kappa(\lambda)}{\partial\lambda^2} = (1-d)\left(2 + d(1-\lambda)^{d-2}\right) \geq 0$ *and* $\frac{\partial\kappa(\lambda)}{\partial\lambda}|_{\lambda=0} = 0$ *when* $|\lambda| < 1$, $0 < d < \frac{1}{2}$.

So now we show that

$$\frac{\partial^2}{\partial\lambda^2}\left\|\log f(z) - \log(z)\right\|_{H^2}^2 \geq 0 \tag{118}$$

and the information distance between these two processes is a convex function of λ. *Since we know that*

$$\begin{aligned} \frac{\partial}{\partial\lambda}\left\|\log f(z) - \log(z)\right\|_{H^2}^2 &= \frac{2}{\lambda}\left(Li_1\left(\lambda^2\right) - dLi_1(\lambda)\right) \\ &= \frac{2}{\lambda}\left((1-\lambda)^d - \lambda^2 + 1\right) \end{aligned} \tag{119}$$

so the lower bound can be obtained by numerically solving the equation $(1-\lambda)^d + \lambda^2 - 1 = 0$.

Proposition 2. *The information distance between any stationary MA(1) processes and FARIMA(0,d,0) with d>0 processes has a lower bound.*

Proof: *It is easier to show the lower bound of MA(1) than the AR(1) process. The information distance between these two processes is:*

$$\begin{aligned} \left\|\log f(z) - \log g(z)\right\|_{H^2}^2 &= \sum_{n=1}^{\infty}\left(\frac{\mu^n}{n} + \frac{d}{n}\right)^2 \\ &= Li_2\left(\mu^2\right) + 2dLi_2(\mu) + \frac{\pi^2 d^2}{n^2} \end{aligned} \tag{120}$$

and the first derivative of information distance is:

$$\begin{aligned}\frac{\partial}{\partial\mu}\left\|\log f(z)-\log g(z)\right\|_{H^2}^2 &= \frac{\partial}{\partial\mu}\left(Li_2\left(\mu^2\right)+2dLi_2(\mu)+\frac{\pi^2d^2}{n^2}\right)\\ &= \frac{2}{\mu}Li_1\left(\mu^2\right)+\frac{2d}{\mu}Li_1(\mu) \qquad (121)\end{aligned}$$

To show the convexity of information distance between MA(1) and FARIMA(0,d,0) processes, the second derivative of information distance is shown to be nonnegative:

$$\begin{aligned}&\frac{\partial^2}{\partial\mu^2}\left\|\log f(z)-\log g(z)\right\|_{H^2}^2\\ =\;& \frac{2}{\mu^2}\left(dLi_0(\mu)-dLi_1(\mu)+2Li_0\left(\mu^2\right)-Li_1\left(\mu^2\right)\right)\\ =\;& \frac{2}{\mu^2}\left(\frac{d\mu+d\mu^2+2\mu^2}{1-\mu^2}+d\log\left(\left(1-\mu^2\right)(1-\mu)\right)\right)\\ \geq\;& \frac{2}{\mu^2}\left(\frac{d\mu+d\mu^2+2\mu^2}{1-\mu^2}+\frac{d\mu^3-d\mu^2-d\mu}{(1-\mu^2)(1-\mu)}\right)\\ \geq\;& \frac{2(2-d-2\mu)}{(1-\mu^2)(1-\mu)}\geq 0 \qquad (122)\end{aligned}$$

and the lower bound of the distance can be obtained by numerically solving the equation $\left(1-\mu^2\right)(1-\mu)^d=0$.

.

References

[1] Granger, Clive WJ. "Long memory relationships and the aggregation of dynamic models." Journal of econometrics 14.2 (1980): 227-238.

[2] Diebold, Francis X., and Atsushi Inoue. "Long memory and regime switching."Journal of Econometrics 105.1 (2001): 131-159.

[3] Taqqu, Murad S., Walter Willinger, and Robert Sherman. "Proof of a fundamental result in self-similar traffic modeling." ACM SIGCOMM Computer Communication Review 27.2 (1997): 5-23.

[4] Mandelbrot, Benoit B. Fractals and scaling in finance: Discontinuity and concentration. Springer Verlag, 1997.

[5] Mandelbrot, Benoit. "Statistical methodology for nonperiodic cycles: from the covariance to RS analysis." Annals of Economic and Social Measurement, Volume 1, number 3. NBER, 1972. 259-290.

[6] Lo, Andrew W. Long-term memory in stock market prices. No. w2984. National Bureau of Economic Research, 1989.

[7] Ding, Zhuanxin, Clive WJ Granger, and Robert F. Engle. "A long memory property of stock market returns and a new model." Journal of empirical finance 1.1 (1993): 83-106.

[8] Nawrocki, David. "R/S analysis and long term dependence in stock market indices." Managerial Finance 21.7 (1995): 78-91.

[9] Chow, K. Victor, Ming-Shium Pan, and Ryoichi Sakano. "On the long-term or short-term dependence in stock prices: Evidence from international stock

markets." Review of Quantitative Finance and Accounting 6.2 (1996): 181-194.

[10] Lobato, Ignacio N., and Nathan E. Savin. "Real and spurious long-memory properties of stock-market data." Journal of Business & Economic Statistics 16.3 (1998): 261-268.

[11] Willinger, Walter, Murad S. Taqqu, and Vadim Teverovsky. "Stock market prices and long-range dependence." Finance and stochastics 3.1 (1999): 1-13.

[12] Sadique, Shibley, and Param Silvapulle. "Long-term memory in stock market returns: international evidence." International Journal of Finance & Economics 6.1 (2001): 59-67.

[13] Cheung, Yin-Wong, and Kon S. Lai. "Long memory and nonlinear mean reversion in Japanese yen-based real exchange rates." Journal of International Money and Finance 20.1 (2001): 115-132.

[14] McLeod, Angus Ian, and Keith William Hipel. "Preservation of the rescaled adjusted range: 1. A reassessment of the Hurst Phenomenon." Water Resources Research 14.3 (1978): 491-508.

[15] Beran, Jan. Statistics for long-memory processes. Vol. 61. CRC Press, 1994.

[16] Robinson, Peter M., ed. Time series with long memory. Oxford University Press, 2003.

[17] Hamilton, James Douglas. Time series analysis. Vol. 2. Princeton: Princeton university press, 1994.

[18] Breidt, F. Jay, Nuno Crato, and Pedro De Lima. "The detection and estimation of long memory in stochastic volatility." Journal of Econometrics 83.1 (1998): 325-348.

[19] Bollerslev, Tim, and Hans Ole Mikkelsen. "Modeling and pricing long memory in stock market volatility." Journal of Econometrics 73.1 (1996): 151-184.

[20] Rostek, Stefan. Option Pricing in Fractional Brownian Markets. Vol. 622. Springer, 2009.

[21] Bouchaud, Jean-Philippe, et al. "Fluctuations and response in financial markets: the subtle nature of 'random'price changes." Quantitative Finance 4.2 (2004): 176-190.

[22] Baillie, Richard T., Tim Bollerslev, and Hans Ole Mikkelsen. "Fractionally integrated generalized autoregressive conditional heteroskedasticity." Journal of econometrics 74.1 (1996): 3-30.

[23] Baillie, Richard T., and Tim Bollerslev. "The long memory of the forward premium." Journal of International Money and Finance 13.5 (1994): 565-571.

[24] Klemeš, V. "The Hurst phenomenon: a puzzle?." Water Resources Research 10.4 (1974): 675-688.

[25] Hosking, Jonathan RM. "Fractional differencing." Biometrika 68.1 (1981): 165-176.

[26] Hurst, Harold Edwin. "Long-term storage capacity of reservoirs." Trans. Amer. Soc. Civil Eng. 116 (1951): 770-808.

[27] Billingsley, Patrick. Convergence of probability measures. Vol. 493. Wiley. com, 2009.

[28] Granger, Clive WJ, and Roselyne Joyeux. "An introduction to long-memory time series models and fractional differencing." Journal of time series analysis 1.1 (1980): 15-29.

[29] Scholz, F. W., and M. A. Stephens. "K-sample Anderson–Darling tests." Journal of the American Statistical Association 82.399 (1987): 918-924.

[30] Katznelson, Yitzhak. An introduction to harmonic analysis. Cambridge University Press, 2004.

[31] Koosis, Paul. Introduction to Hp spaces. Vol. 115. Cambridge University Press, 1998.

[32] Leibowitz, Gerald M. Lectures on complex function algebras. Vol. 384. Glenview, IL: Scott, Foresman, 1970.

[33] Andrew Mullhaupt, Christopher Bishop, Approximation of long memory process by short memory process, to be published. 2013

[34] Romano, Joseph P. Testing statistical hypotheses. Springer, 2005.

[35] Mullhaupt, Andrew P., and Kurt S. Riedel. "Fast adaptive identification of stable innovation filters." Signal Processing, IEEE Transactions on 45.10 (1997): 2616-2619.

[36] Mullhaupt, Andrew P., and Kurt S. Riedel. "Banded matrix fraction representation of triangular input normal pairs." Automatic Control, IEEE Transactions on 46.12 (2001): 2018-2022.

[37] de Jong, Piet, and Jeremy Penzer. "The ARMA model in state space form." Statistics & probability letters 70.1 (2004): 119-125.

[38] Rachev, Svetlozar T., et al. Financial econometrics: from basics to advanced modeling techniques. Vol. 150. Wiley. com, 2007.

[39] Mullhaupt, unpublished result

[40] Black, Fischer, and Myron Scholes. "The pricing of options and corporate liabilities." The journal of political economy (1973): 637-654.

[41] Merton, Robert C. "Theory of rational option pricing." The Bell Journal of Economics and Management Science (1973): 141-183.

[42] Fama, Eugene F. "The behavior of stock-market prices." The journal of Business 38.1 (1965): 34-105.

[43] Jegadeesh, Narasimhan. "Evidence of predictable behavior of security returns." The Journal of Finance 45.3 (1990): 881-898.

[44] Lo, Andrew W., and Archie Craig MacKinlay. "Stock market prices do not follow random walks: Evidence from a simple specification test." Review of financial studies 1.1 (1988): 41-66.

[45] Malkiel, Burton G., and Eugene F. Fama. "Efficient Capital Markets: A Review Of Theory And Empirical Work*." The journal of Finance 25.2 (1970): 383-417.

[46] Fama, Eugene F., and Kenneth R. French. "Permanent and temporary components of stock prices." The Journal of Political Economy (1988): 246-273.

[47] Grundy, Bruce D. "Option prices and the underlying asset's return distribution." The Journal of Finance 46.3 (1991): 1045-1069.

[48] Lo, Andrew W., and Jiang Wang. "Implementing option pricing models when asset returns are predictable." The Journal of Finance 50.1 (1995): 87-129.

[49] Chou-Wen Wang, Chin-Wen Wu, Shyh-Weir Tzang, Implementing option pricing models when asset returns follow an autoregressive moving average process, International Review of Economics & Finance, Volume 24, October 2012, Pages 8-25

[50] Harrison, J. Michael, and Stanley R. Pliska. "Martingales and stochastic integrals in the theory of continuous trading." Stochastic processes and their applications 11.3 (1981): 215-260.

[51] Bollerslev, Tim. "Generalized autoregressive conditional heteroskedasticity." Journal of econometrics 31.3 (1986): 307-327.

[52] Kwiatkowski, Denis, et al. "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?." Journal of econometrics 54.1 (1992): 159-178.

[53] Beran, Jan. "Statistical methods for data with long-range dependence." Statistical science (1992): 404-416.

[54] Hiemstra, Craig, and Jonathan D. Jones. "Another look at long memory in common stock returns." Journal of Empirical Finance 4.4 (1997): 373-401.

[55] Geweke, John, and Susan Porter-Hudak. "The estimation and application of long memory time series models." Journal of time series analysis 4.4 (1983): 221-238.

[56] Hurvich, Clifford M., Rohit Deo, and Julia Brodsky. "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series." Journal of Time Series Analysis 19.1 (1998): 19-46.

[57] Reisen, Valderio A. "Estimation of the fractional difference parameter in the ARIMA (p, d, q) model using the smoothed periodogram." Journal of Time Series Analysis 15.3 (1994): 335-350.

[58] Robinson, Peter M. "Log-periodogram regression of time series with long range dependence." The annals of Statistics (1995): 1048-1072.

[59] Rogers, L. Chris G. "Arbitrage with fractional Brownian motion." Mathematical Finance 7.1 (1997): 95-105.

[60] Mandelbrot, B. B. and J. W. Van Ness (1968). "Fractional Brownian motions, fractional noises and applications." SIAM review 10(4): 422-437.

[61] Abernethy, R. B., R. P. Benedict, and R. B. Dowdell. "ASME measurement uncertainty." ASME Journal of Fluids Engineering 107.2 (1985): 161-164.

[62] Thompson, R. "A note on restricted maximum likelihood estimation with an alternative outlier model." Journal of the Royal Statistical Society. Series B (Methodological) (1985): 53-55.

[63] Shao, B. and G. Frank (2012). "Aggregation of an FX order book based on complex event processing." Investment Management and Financial Innovation: 2012(1).