THE SHANGHAI STOCK EXCHANGE: STATISTICAL PROPERTIES AND SIMULATION

by

Xiongzhi Chen

Master of Science in Mathematics
Sichuan University, China

Submitted to the Department of Mathematics
in Partial Fulfillment of the Degree of

Master of Arts in Mathematics
at the
University of Hawaii
April, 2009

THESIS COMMITTEE:
Thomas Ramsey
George Wilkens
We certify that we have read this document and that, in our opinion, it is satisfactory in scope and quality as a thesis for the degree of Master of Arts in Mathematics.

THESIS COMMITTEE

_________________________________

_________________________________

_________________________________
THE SHANGHAI STOCK EXCHANGE: STATISTICAL
PROPERTIES AND SIMULATION BY GENERALIZED HYPERBOLIC
DIFFUSION

by

Xiongzhi Chen

Submitted to the Department of Mathematics on April 21, 2009
in partial fulfillment of the requirements for the degree of
Master of Arts in Mathematics

Abstract

Empirical data of asset returns, including that of stock market index series, always exhibit non-stationarity, heavy tails, dependence structures, and other stylized features as well, among which the tail indices have wide applications in assessing the distribution of outliers and the Hurst index reveals roughly how strong and how long the correlation structure retains itself.

While estimating these statistics is usually hard, simulating financial markets by continuous-time models is even more challenging since the proposed model should be able to at least generate the empirical distribution, capture the intrinsic dependence structure, and simulate the volatilities, of the financial data, under appropriate statistical settings.

This thesis contains two independent papers that summarize the author's study of the statistical properties of the returns series of the daily price index series of the Shanghai Stock Exchange (SSE)\(^1\) for trading days between 01/01/1991 and 14/11/2007, and the author's simulation of the detrended logarithmic index series by the ergodic, generalized hyperbolic (GH) diffusion whose invariant density is the generalized hyperbolic (GH) density.

More specifically, the first paper herein gives estimates of the right tail index, the left tail index, and Hurst index for the returns series\(^2\) of the daily SSE index, all of which are consistent with what most researchers obtained for stock market index series and tend to argue that extreme care should be taken when making assumptions about the underlying stochastic process that generates the stock price index series. These estimates are obtained by the simplest method – the traditional ordinary least squares (OLS) method, to avoid unnecessary or even wrong assumptions on stationarity or dependence structures of the index series, but some comparative estimates via different methods are also provided.

\(^1\)The "SSE index" will specifically denote the "daily price index series of the SSE".

\(^2\)The "returns series" will denote specifically the "returns series of the daily price index series of the SSE".
The second paper is devoted to testing whether an ergodic GH diffusion process can simulate well the linearly detrended logarithmic SSE index series. The ergodic GH diffusion model was proposed by several researchers in the mid 1990s and has successfully simulated some stock price index series. Its main advantages are: its capability to generate the generalized autoregressive conditional heteroscedastic (GARCH) effect that is compatible with the correlation structure of most financial time series, and its capability to approximate the empirical densities of the financial time series by the model's invariant density which is a generalized hyperbolic (GH) density.

This second paper describes for the detrended index an ergodic GH diffusion whose invariant density is proportional to the GH density and whose parameters were estimated by Markov chain Monte Carlo (MCMC) methods. Six possible models were found since six independent, convergent Markov chains are obtained by the Monte Carlo standard error (MCSE) convergence diagnostic. Unfortunately the Gelman-Rubin convergence diagnostic failed to converge to one even after 350,000 iterations of the Metropolis-Hastings algorithm, since each of the six chains converged to its own stationary distributions at around 50,000 iterations, (which might be the consequence of trapped iterations for the chains in the simulation).

After six Monte Carlo estimates of the unknown parameters were obtained, we simulated the detrended index by the six competing models. However, none of them successfully generated paths that match the data well. Also, a uniform residual test was carried out for only four of the six possible models due to the computational capability of Mathematica and Matlab. The test results statistically reject all four selected models, thus rejecting the null hypothesis that the underlying stochastic process for the detrended index is an ergodic GH diffusion.

One reason for such a failure might be that the invariant density is inappropriate for the empirical density of the detrended index, since its empirical density is tri-modal whereas the GH density is unimodal.

---

3The "detrended index" will specifically denote the "linearly detrended logarithmic SSE index series".
Acknowledgements

I am deeply indebted to my adviser, Professor Thomas Ramsey, whose constant guidance, inspiring suggestions and supports motivated my study on the statistical properties of financial markets and on the simulation of financial time series by continuous-time models. Without his help, this thesis would not have been possible.

I would like to thank Professor George Wilkens for his help on my study in the theory of stochastic control, and, Professor James B. Nation, Professor Wayne Smith, Professor William Lampe for their decisive support to my pursuit of further academic opportunities in the near future.

I am grateful to other faculty members and stuff, particularly, the departmental secretaries, for their help that made my life and study here at Hawaii much more easier and enjoyable. I am also grateful to fellow graduate students in the department, particularly, to Shashi Bhushan Singh and Gretel Sia for their kind help.

Finally, I would like to express my appreciation to my family, especially my wife, my brother, and to my friends, especially Jin Li, whose love and contribution paved the way for my dream in mathematics.
Please notice that the two papers incorporated into this thesis have their own paging schemes. The next page will be the starting page of the first paper, and the second starts right after the ending page of the first.

Please go to the next page for the start of the first paper and then to the second paper after around 23 pages ......
THE SHANGHAI STOCK EXCHANGE I: STATISTICAL PROPERTIES

XIONGZHI CHEN

ABSTRACT. This paper studies the statistical properties, mainly, the tail behaviour, skewness, correlation structure, of returns series of the SSE index for trading days from 01/01/1991 to 14/11/2007, and gives their corresponding statistical estimates.

The estimated right tail index is around 4.6, that of the left tail index around 2.19 and the Hurst index around 0.61, for the returns series. These statistics are obtained by the simplest estimation method: linear regression via ordinary least squares (OLS), to avoid unnecessary or even wrong assumptions about the stationarity or dependence structures of the SSE index.

Results provided herein comply with what most researchers have found for stock market indices (see Cont01 [5] and Mills99 [21]). The tail indices show that statistically the underlying stochastic process that generates the SSE index can be assumed to be of finite variance, whereas the Hurst index shows that the underlying process is not a random walk but has long-range dependence.

Key words and phrases. Hill’s Estimator, Hurst Index, Lo’s Modified R/S Statistic, R/S Statistic, R/S Analysis, Short/Long-range Dependence, Skewness, Tail Index.
Contents

List of Figures 3

1. Introduction 4
2. Tail Index 5
  2.1. Hill’s Estimator 8
  2.2. Linear Regressor 10
  2.3. Mandelbrot’s Visual Check 13
3. Kurtosis and Skewness 14
4. Dependence 16
  4.1. Hurst’s R/S Statistic 16
  4.2. Mandelbrot’s R/S Analysis 19
  4.3. Direct Regression 22
5. Conclusions 23
References 23
SSE RETURNS: TAIL INDICES, SKEWNESS, HURST INDEX

LIST OF FIGURES

1 Original SSE Price Index 6
2 Returns Series 6
3 Empirical PDF against Normal 10
4 L-L Plot of Right Tail 11
5 L-L Plot of Left Tail 11
6 Regressed L-L Right Tail 12
7 Regressed L-L Left Tail 12
8 Sample Variance 13
9 Asymptotic Sample 4th Moment 13
10 Sample 5th Absolute Moment 14
11 Skewness 15
12 Zoom-in Kernel Density Estimate of EPDF 15
13 Classic R/S Statistics 17
14 L-L Plot of Classic R/S Statistic 17
15 Regressed R/S Statistic for Small Sample Size 18
16 Regressed R/S Statistic for Larger Sample Size 18
17 Pox Plot of R/S Analysis 19
18 Autocorrelation till Lag Order 1000 20
19 Lo’s Modified R/S Statistic 22
20 L-L Plot of Absolute Average Returns 23
1. Introduction

It’s widely acknowledged that most asset returns series have the following main stylized features revealed by empirical studies: leptokurtosis\(^1\), skewness, and power-decaying correlations in absolute returns (see Cont01 [5], Mandelbrot01a [18]).

As a reference for interested readers, the eleven such features of asset returns given in Cont01 [5] are listed below:

**F1: Absence of autocorrelations:** (linear) autocorrelations of asset returns are often insignificant, except for very small intraday time scales (about 20 minutes) for which microstructure effects come into play.

**F2: Heavy tails:** the (unconditional) distribution of returns seems to display a power-law or Pareto-like tails, with a tail index which is finite, higher than two but less than five for most data sets studied. In particular this excludes stable laws with infinite variance and the normal distribution. However the precise form of the tails is difficult to determine.

**F3: Gain/loss asymmetry:** one observes large drawdowns in stock prices and stock index values but not equally large upward movements.

**F4: Aggregational Gaussianity:** as one increases the time scale \(\Delta t\) over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales.

**F5: Intermittency:** returns display, at any time scale, a high degree of variability. This is quantified by the presence of irregular bursts in time series of a wide variety of volatility estimators.

**F6: Volatility clustering:** different measures of volatility display a positive autocorrelation over several days, which quantifies the fact that high-volatility events tend to cluster in time.

**F7: Conditional heavy tails:** even after correcting returns for volatility clustering (e.g., via GARCH-type models), the residual time series exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution of returns.

**F8: Slow decay of autocorrelation in absolute returns\(^2\):** the autocorrelation function of absolute returns decays slowly as a function of the time lag, roughly as a power law with an exponent \(\beta \in [0.2, 0.4]\). This is sometimes interpreted as a sign of long-range dependence.

**F9: Leverage effect:** most measures of volatility of an asset are negatively correlated with the returns of that asset.

**F10: Volume/volatility correlation:** trading volume is correlated with all measures of volatility.

**F11: Asymmetry in time scales:** coarse-grained measures of volatility predict fine-scale volatility better than the other way.

---

\(^1\)A leptokurtic distribution is one that is peaked near the center and has fat tails

\(^2\)Absolute returns are the absolute values of the returns
These characteristics are extremely helpful for researchers who seek to understand asset returns\(^3\) statistically and simulate them efficiently, but they seem to have frustrated any attempts to build mathematically and computationally feasible continuous-time models that are able to incorporate almost all of them.

To verify and better understand these empirical features and construct good simulations to stock market index series, we take the returns series of the SSE index\(^4\) and study its empirical statistical properties, such as skewness, kurtosis, tail index, and long-term dependence structure. These statistics provide statistical support for simulating this returns series by a generalized hyperbolic (GH) diffusion model (for a hyperbolic diffusion model for stock prices, see Bibby97 [2] and Rydberg99 [27]). In order to avoid unnecessary, even wrong assumptions about this returns series, all statistical estimates herein are obtained nonparametrically by linear regression with ordinary least squares (OLS). We also provide comparisons with other popular estimation techniques of these statistics that would be valid under the independent and identically distributed (i.i.d.) assumption. However no non-parametric check on the stationarity of this return series is done because of the lack of assumptions and the reluctance of the author to impose i.i.d or a normality assumption on the background noise of this series as proposed by most researchers, even though it is widely agreed that most financial time series are not strictly stationary or even non-covariance stationary (Mills95 [20], Mills99 [21], Loretan94 [11]).

The statistical properties estimated in this paper comply with their corresponding stylized empirical features and tend to argue against the common tradition of assuming stationarity for returns series of stock price indices and against the possibly inappropriate stable-distribution assumption for financial time series.

2. Tail Index

Figure 1 and Figure 2 depict the SSE index series \(\hat{P} = \{P_t : 1 \leq t \leq 4129\}\) and its return series \(\hat{R} = \{R_t : 1 \leq t \leq 4129\}\), where the horizontal axis denotes the number of trading days from Jan 1, 1991 and the vertical axis the price index and returns of the price index, respectively. It is clear that the original stock index series has an increasing trend for most part of the period considered, while the log-return series looks more like trendless noise but displays more volatility during the period roughly from Jan 1991 to Aug 2006.

Since under most circumstances only observations of a stochastic process are available, it is reasonable to start from the empirical cumulative distribution function (ecdf) of a process to obtain some of the required statistics. Under the traditional assumption that all random variables of the underlying stochastic process that generate the observations are identically distributed, the ecdf is even more crucial to statistical inference about the process. Studies show that the epdf of most financial time series display power-decaying

\(^3\)Here "asset returns" denotes "the asset returns series", and "return" means conventionally "log return".

\(^4\)Unless otherwise specified, notationally we will not distinguish a random process \(X = \{X_t, t \in I\}\) from an observation \(X = \{X_t, t \in I_0\}\), where \(I\) is a nonempty index set and \(I_0\) is usually nonempty, that is, \(\{X_t, t \in I_0\}\) also denotes the observation \(X\).
or semi-heavy tails (see, for example, Cont01 [5], Mantegale95 [22], Loretan [11]) with tail index $\alpha$ ranging between 0 and 5 (see, for example, Mills99 [21]).

Given a stochastic process\(^5\) $X = \{X_t, t \in T\}$ defined on some underlying probability space $(\Omega, \mathcal{F}, P)$ such that all $X_t, t \in T$ are identically distributed (i.d.)\(^6\), the tail index $\alpha$ of $X$ is usually defined as (see, Samorodnitsky94 [28])

$$\alpha = \sup \left\{ \beta > 0 : E_P \left( |X_t|^{\beta} \right) < \infty \right\}$$

\(^5\)For more details on the definition of stochastic process, please see Karatzas98[10]

\(^6\)For most financial time series, the random variables of the underlying stochastic process that generates the observation is not identically distributed.
where $E_P$ denotes the mathematical expectation with respect to the probability measure $P$. Tail index is extremely important in risk management since it helps measure the probability of outliers.

Several methods for tail index estimation have been proposed (see, for example, Hill75 [8], Loretan94 [11], Samorodnitsky94 [28], Taqqu97 [30], and their properties and performances have been compared (see, Taqqu97 [30], McCulloch [19])). But it seems that the classical Hill’s estimator, the linear regression method (based on ordinary least squares (OLS)) and maximum likelihood estimation (MLE) perform better in an overall fashion than the others (It should be noted that for stable distributions Hill’s estimator behaves somewhat badly (see, Taqqu97 [30], McCulloch [19]). Based on these results, we basically base our estimation of the tail index $\alpha$ on linear regression by OLS but also provide the estimate obtained by the modified Hill’s estimator introduced in Loretan94 [11].

Before estimating the tail index for the ecdf of returns series $R = \{R_t : 1 \leq t \leq 4129\}$ (of $P$), we provide a slightly different version of the definition for the generalized tail index (compared to Loretan94 [11]):

**Definition 1.** Let $X = \{X_t : t \in T\}$ be a stochastic process such that $X_t$, $t \in T$ is identically distributed (i.d.) with $F(x) = P(X_{t_0} < x)$ where $t_0 \in T$ is fixed. If

$$
\lim_{x \to \infty} P(X_{t_0} > x) = C x^{-\alpha} (1 + B_R(x)) , x > 0
$$

holds for some constants $\alpha > 0$, $C$ and for some $B_R(x)$ with

$$
\lim_{\lambda \to \infty} B_R(x) = 0
$$

Then $\alpha$ is called the right tail index of $X$.

The left tail index $\beta$ is defined similarly. The above definition actually relaxes the i.i.d. assumption introduced in Loretan94 [11], Mills95 [20] and such relaxation of assumptions may be important for analyzing returns series, in that for most financial stochastic processes $X = \{X_t : t \in T\}$ the random variables $X_t, t \in T$ are not i.i.d or even i.d.. Thus in practice we would have to content ourselves with the following interpretation of what is meant by saying "the empirical distributions of a stochastic process having heavy tails". (For details on how to construct stochastic process from a consistent family of finite dimensional marginal distributions, please see Karatzas98 [10] for the the Kolmogorov consistency theorem):

**Definition 2.** Let $X = \{X_t : t \in T\}$ be a stochastic process with underlying probability space $(\Omega, \mathcal{F}, P)$ and let

$$
O_N = \{X_{t_i}(\omega_i) : t_i \in T, \omega_i \in \Omega, 1 \leq i \leq N\} , N \in \mathbb{N}
$$

be a discrete observation of $X$. We say that the epdf of $X$ has right tail index $\alpha$ if and only if

$$
\lim_{N \to \infty} \left[ \lim_{x \to \infty} \hat{P}_N \left( \bigcup_{i=1}^{N} \{X_{t_i} (\omega_i) > x\} \right) \right] = C x^{-\alpha} (1 + B_R(x)) , x > 0
$$

holds for any such observation, for some constants $\alpha > 0$, $C \in \mathbb{R}$ and for some $B_R(x) \in \mathbb{R}$ with

$$
\lim_{x,N \to \infty} B_R(x) = 0
$$
where \( P_N \) denotes the empirical density obtained from the observation \( O_N \).

The above definition describes more realistically what we see and do when plotting the empirical cdf of samples by smoothing techniques, and it is almost equivalent to the assumption that the process \( X \) is ergodic.

### 2.1. Hill’s Estimator.

In econometrics, the most popular estimator of the constants \( C \) and \( \alpha \) (see, Hill75 [8]) in (2.1) are obtained by using rank statistics of the observations \( \{X_i: 1 \leq i \leq N\} \), that is, if \( \{X_{(i)}: 1 \leq i \leq N\} \) are the order statistics of \( \{X_i: 1 \leq i \leq N\} \) such that \( X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(N)} \) then

\[
\hat{\alpha} = \left( s^{-1} \sum_{j=1}^{s} \log X_{(N-j+1)}^{(N-s)} - \log X_{(N-s)}^{(N-s)} \right)^{-1},
\]

(2.3)

is the right tail index estimator for \( \{X_t: t \in T\} \) (assuming that all \( X_t, t \in T \) are the identically distributed). Further, the estimator for \( C \) is given by

\[
\hat{C} = \frac{S}{N} X_{(N-s)}^{\hat{\alpha}}
\]

(2.4)

Empirical study shows that the tails of ecdf of asset returns are not asymptotically so heavy as those of stable distributions but mostly are power-law decaying or semi-heavy. McCulloch97 [19] has pointed that the Hill’s estimator is usually up-biased (in that the estimate can be around 2 even for stable distributions with tail index \( \alpha = 1.65 \)) and behaves badly for stable distributions with tail index \( \alpha \) close to 2, and that an estimate \( \hat{\alpha} > 2 \) obtained by Hill’s estimator only rejects the joint hypothesis that the underlying innovations are i.i.d. and have stable distributions.

Due to the appeal of the Hill’s estimator, we still carried the estimate for \( \alpha \) by using (2.3), (2.4) and followed the suggestions from Philips96 [24] that, according to minimizing the mean squared error (MSE) of the limiting distribution of \( \hat{\alpha} \), we should set

\[
s(N) = \left\lceil \lambda N^{2/3} \right\rceil
\]

(2.5)

while \( \lambda \) is adaptively estimated by

\[
\hat{\lambda} = \left( \frac{\bar{\alpha}_1 N}{\sqrt{2 s_2}} (\bar{\alpha}_1 - \bar{\alpha}_2) \right)^{2/3}
\]

(2.6)

and \( \bar{\alpha}_1, \bar{\alpha}_2 \) are preliminary estimates of \( \hat{\alpha} \) using truncates \( s_1 = \lceil N^\sigma \rceil, s_2 = \lceil N^\tau \rceil \) with \( 0 < \sigma < 2/3 < \tau < 1 \) and \( \sigma = 0.6, \tau = 0.9 \) for the actual implementation. Note that these estimates are for the right or the left tail of the distribution respectively.

The most devastating shortcoming of the above method is that it is based on the i.i.d. assumption on the innovations of the sample. Since sufficient evidence show that most
returns series are not i.i.d., we can assume that this is also true for $\tilde{R}$ and consequently expect that the estimator $\hat{\alpha}$ will be bigger than 2.

Hereunder we provide the detailed estimation procedure for the Hill’s estimator for $\tilde{R}$: by setting $s_1 = [4129^{0.6}] = 147$, $s_2 = [4129^{0.9}] = 1795$, we have

$$\hat{\alpha}_1 = \left[ s_1^{-1} \sum_{j=1}^{s_1} \log X_{(N-j+1)} - \log X_{(N-s_1)}^{-1} \right]^{-1}$$

$$= \left( \sum_{j=1}^{147} \log X_{(4129-j+1)} - \log X_{(4129-147)}^{-1} \right)^{-1}$$

$$= \left( \frac{1237.917}{147} - 8.118 \right)^{-1} = 3.2320$$

and

$$\hat{\alpha}_2 = \left( s_2^{-1} \sum_{j=1}^{s_2} \log \left( X_{(N-j+1)} - \log X_{(N-s_1)} \right) \right)^{-1}$$

$$= \left( \sum_{j=1}^{1795} \log X_{(4129-j+1)} - \log X_{(4129-1795)}^{-1} \right)^{-1}$$

$$= \left( \frac{13473.64}{1795} - 7.1903 \right)^{-1} = 3.1655$$

which means

$$\hat{\lambda} = \left| \frac{3.2320 \cdot 4129}{\sqrt{2} \cdot 1795} (3.2320 - 3.1655) \right|^{2/3}$$

$$= 0.4963$$

and

$$s(N) = \left[ 0.4963 \cdot (4129)^{2/3} \right] = [1366.1] = 1366$$

Thus

$$\hat{\alpha} = \left( s^{-1} \sum_{j=1}^{s} \log X_{(N-j+1)} - \log X_{(N-s)} \right)^{-1}$$

$$= \left( \sum_{j=1}^{1366} \log X_{(4129-j+1)} - \log X_{(4129-1366)} \right)^{-1}$$

$$= \left( \frac{10366.64}{1366} - 7.2958 \right)^{-1} = 3.4100$$

and

$$\hat{C} = \frac{s}{N} X_{(N-s)}^{\hat{\alpha}} = \frac{1366}{4129} (7.2958)^{3.4100} = 290.19$$
Consequently, assuming $\tilde{R}_t$, $1 \leq t \leq 4139$ are identically distributed (which might also not be true) and the right tail index is constant in the sense that the sample size for $\tilde{R}$ is large enough to stabilize the estimate (see, Mills99 [21]), we have asymptotically

$$P (R_1 > x) = 290.19^{3.4100} x^{-3.4100} (1 + \zeta_R (x))$$

where $\lim_{x \to +\infty} \zeta_R (x) = 0$, i.e.,

$$\log P (X > x) \approx 19.337 - 3.41 \log x$$

Thus the Hill estimator of the right tail index is 3.41.

2.2. Linear Regressor. Now we will use visual examination and linear regression to estimate the tail index $\alpha$ without assumptions on the underlying innovations for the return series $\tilde{R}$. The following Figure 3 shows the comparison between the empirical probability density function (epdf) obtained by a kernel estimator and the normal density whose mean and standard deviation are the same as those of $\tilde{R}$, which shows clearly that the epdf has heavier tail, bigger kurtosis, compared to that of a normal density $N(0.00091, 0.00289)$.

We also provide in Figure 4 and Figure 5 the log-log plots of both right and left tails. A visual check on these two figures reveals that there are some points for $x$ right beyond which $\log F_N (x)$ and $\log (1 - F_N (x))$ are strongly linear against $\log x$, respectively. These points are $x \geq 0.024$ for the left tail and $x \geq 0.03$ for the right tail, where $F_N (\cdot)$ is the empirical cdf of $\tilde{R}$ and $N$ is the size of $\tilde{R}$.

Since all we want are the asymptotic properties (assuming that a sample size $N = 4129$ is sufficiently large for the estimation of asymptotic statistics), we can safely choose $x \geq 0.024$ and $x \geq 0.03$ respectively and then use linear regression by OLS to estimate the right and left tail index. The following Figure 6 and Figure 7 display where the log-linearity are the strongest for sufficiently large $|x|$ with reference linear regressors. Both of the figures show that both ends of the two tails diverge a lot from their linear
regressors. This is the consequence of a few outliers in the sample $\bar{R}$ due to relatively large gains or losses that happened approximately in Oct 2007.

The regression equation for the right tail is

$$\ln \left(1 - F_N(x)\right) = -0.9766 - 4.670 \ln x, \ x \geq 0.030061$$

and that of left tail is

$$\ln F_N(x) = \ln P(X < -x) = -10.370 - 2.192 \ln x, \ x \geq 0.024045$$

Consequently, the OLS estimate for the right tail index is

$$\hat{\alpha}_R = 4.670$$

and that for left tail index is

$$\hat{\alpha}_L = 2.192$$
These indices are consistent with both the fact that usually the ecdf of stock return series are not stable distributions but Paretian type with power-decaying tails, and the abundant evidence against assuming infinite variance for stock return series whose innovations are identically distributed (see, Mills99 [21]). They are also in accordance with the tail range most researcher have found empirically and the observation that the ecdf has a relatively heavier left tail than the right tail (see Mills99 [21]). (The reason why the linear regression estimator is smaller than Hill’s estimator is unknown to me and will be examined in future work.)
2.3. Mandelbrot’s Visual Check. Supplementary to the tail estimations conducted above, we plot the sample moments computed by

$$\hat{\sigma}^p(n) = \frac{1}{n} \sum_{i=1}^{n} \left( R_i - \frac{1}{n} \sum_{j=1}^{n} R_j \right)$$

of order $p = 2, 4$ for $R$ for relatively large sample size $n$, $1000 \leq n \leq 4129$ in Figure 8 and Figure 9.

These graphs seem to show that as the sample size increases to infinity, the sample variance and 4th order sample moment stabilize themselves around certain values respectively. Thus by the idea of Mandelbrot63 [15], we seem to be able to infer that $R$
does have finite variance and 4th order moment \(^9\) (if \(\hat{R}\) has a unique law of probability\(^{10}\)), which supports the estimates of the tail indices here. Actually we also plotted the sample 5th absolute sample moment for \(\tilde{R}_1 = \{R_t, 4129 \geq t \geq 1500\}\) in Figure 10 to check whether the right-tail index is low-biased or not (since if it’s low-biased then the 5th absolute sample moment might diverge eventually) and it seems that the 5th absolute moment is also finite by Mandelbrot’s idea.

It should be pointed out that this visually checked convergence might be false due to insufficient sample size. But proof or disproof of this seems to be impossible here in this paper since we are not able estimate the exact asymptotic properties of the tail of the empirical cumulative distribution function (ecdf) without assuming much, much more on the underlying noise for \(\tilde{R}\). Also visual check might be misleading particularly when the sample is from a distribution with tail index 2, as justified by the sample variance of the Cauchy distribution. However under very stringent assumptions about the mixing properties and the order of moments of the innovations, the finiteness of 4th moment can be tested as done in Loretan94 \([11]\).

3. Kurtosis and Skewness

Kurtosis is easily recognized from the figure of epdf (Figure 3). To check the skewness, the easiest way is plot \(X_{(N-p)} - X_{\text{med}}\) against \(X_{\text{med}} - X_{(p)}\) since for symmetric epdf’s

\[
(X_{(N-p)} - X_{\text{med}}, X_{\text{med}} - X_{(p)})
\]

will be points one a line of slope -1, where \(X_{(i)}\) are rank statistics of the sample \(\{X_i : 1 \leq i \leq N\}\) and \(X_{\text{med}}\) is the median and \(1 \leq p \leq \lfloor N/2 \rfloor + 1\). Empirical evidence on financial time series show that within a fairly wide range around its mean the epdf

\(^9\)We say that "a stochastic process \(X = \{X_t, t \in T\}\) has some property" if (it is assumed that) all \(X_t, t \in T\) have the same property.

\(^{10}\)A stochastic process \(X = \{X_t, t \in T\}\) is said to have as unique law of probability if all \(X_t, t \in T\) have the same law of probability.

Figure 10. Sample 5th Absolute Moment
Figure 11. Skewness

Figure 12. Zoom-in Kernel Density Estimate of EPDF

is symmetric, and skewness appears seemingly in the tails (see, Mills99 [21], Mantega95 [22]). Moreover, most epdf’s are skewed to the left (see, Mills95 [20], Badrinath98 [1]).

Figure 11 plots $R_{(T-p)} - R_{med}$ against $R_{med} - R_{(p)}$ for $\tilde{R} = \{R_t : 1 \leq t \leq 4129\}$ An OSL linear regression estimate of $R_{(p)}$ against $R_{(T-p)}$ gives

$$R_{(N-p)} = -1.0629 * R_{(p)} + 0.0004$$

Surprisingly it shows that the epdf is skewed to more to the right than to the left (also, see the "zoom in" Figure 12 of a kerneal density estimate of the epdf), although the epdf of the returns $\tilde{R}$ is symmetric in a relatively wide range around its median. This strange observation might help explain why the ergodic generalized hyperbolic diffusions can not fit the SSE index in the second paper, but it complies with the fact that for most of the period from July 1994 to Oct 2007 the SSE index was increasing, and that
the probability of a large decrease in the returns is relatively lower than that of a large
increase in the returns.

4. Dependence

Dependence, long-term or short-term, exists in time series observed from a lot of natural
phenomena (see, Cont01 [5], Mandelbrot62, [14], Mandelbrot01a [18]) and has been
paid a lot of attention in recent years. For discrete time model that incorporates such
dependence, FIGARCH (i.e., fractional integrated GARCH) model has been introduced,
whereas for continuous model, fractional Brownian motion (fBm) (see, Mandelbrot62
[14], Mandelbrot01a [18]) has been constructed, both of which have wide applications.

For a discretely sampled stochastic process, the Hurst Index \( H \) can be used to measure
the strength and duration of its dependence structure. Hereunder we describe three
methods to estimate this index and provide the estimates for for \( R \).

4.1. Hurst’s R/S Statistic. Hurst (see, Hurst51 [9]), Kolmogorov and Mandelbrot are
the first few pioneers who discovered the long-term memory effects in natural pheno-
mena, defined the scaling properties of probability distribution, and applied self-similarity
mathematically. Actually, what Hurst found out empirically or how the Hurst index is
defined is that

\[
E(\frac{R}{S}(n)) \propto Cn^H
\]

holds for sufficiently large sample size \( n \) for some constants \( C > 0 \) and some \( H \in (1/2, 1) \),
where \( E \) is the expectation with respect to the probability measure \( P \) of the underlying
probability space \((\Omega, \mathcal{F}, P)\) on which the random process that generates the sample
\( X = \{x_i, i = 1, 2, ..., n\} \) is defined and

\[
\frac{R}{S}(n) = \frac{R(n)}{S(n)}
\]

is called the classic rescaled range statistic or \( R/S \) statistic with

\[
\mu = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2, \quad S_n = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2}
\]

and

\[
R(n) = \max_{1 \leq j \leq n} \left\{ \sum_{i=1}^{j} x_i - \frac{j}{n} \sum_{i=1}^{n} x_i \right\} - \max_{1 \leq j \leq n} \left\{ \sum_{i=1}^{j} x_i - \frac{j}{n} \sum_{i=1}^{n} x_i \right\}
\]

One major disadvantage of the classic Hurst \( R/S \) estimator is that it is too sensitive to
short-range dependence\(^{11}\), but one key advantage it has is that, if we define a long-range
dependence parameter \( d \) by

\[
d = H - \beta
\]

with \( \beta = 1/2 \) for finite variance processes or \( 1/\alpha \) for infinite variance processes with tail
index \( 0 < \alpha < 2 \), then \( R/S \) statistic consistently provides an estimate of

\[
\hat{H} = d + \frac{1}{2}
\]

\(^{11}\)See Mandelbrot01a [18] for details on “short/long-range dependence”.

independent of the value of $\alpha$ since $R(n) \propto n^H = n^{d+1/\alpha}$ and $S(n) \propto n^{1/\alpha - 1/2}$ asymptotically (see, Taququ98 [31], Mandelbrot75 [17]).

Figure 13 is a plot of the estimated classic $R/S(n)$ against $n$ for

$$\tilde{R}_n := \{ R_t : t = 1, \ldots, n \}, 1 \leq n \leq 4025$$

and the corresponding log-log plot in provided in Figure 14.

As can been seen from the log-log plot Figure 14, linearity is very strong for two sections: $\exp(0.8) \leq n \leq \exp(6)$ and $8 \geq n \geq \exp(6)$. Consequently it’s reasonable to

12* := * means * defined as *
estimate $H$ for each section. An OLS regressor for $(\log R/S(n), \log n)$ for $2 \leq n \leq 410$ is given as

$$L_1(n) = 0.3366 \ln n + 1.6223$$

and that for $410 \leq n \leq 4025$ is given as

$$L_2(2) = 0.4265 \ln n + 0.9152$$

which are combined into

$$\log R/S(n) = \begin{cases} 
1.0833 \ln n + 3.2191 & \text{if } 4 \leq n \leq 410 \\
0.4265 \ln n + 0.9152 & \text{if } 410 \leq n \leq 4025 
\end{cases}$$

The corresponding plots of $\log R/S(n)$ against $\log n$ with regressors are provided in Figure 15 and Figure 16.

Even though for the last 100 sample points, the linearity is strong with a much larger slope of regression than have been previously given, but they contribute little when incorporated into the second group for regression.
Thus focusing on large sample properties, we can just take 0.4265 to be the estimate of \( d + \frac{1}{2} \) since we showed that \( \hat{R} \) has finite variance by its tail index estimate (by explicitly assuming that \( \hat{R} \) has a unique probability law). This suggests that the estimated \( \hat{d} = -0.08 \) and that \( \hat{R} \) is a process with short-range dependence.

However, the \( R/S \) analysis and modified \( R/S \) statistic in the following subsection both show the existence of long-range dependence in \( \hat{R} \), contradicting the short-range dependence revealed by classic \( R/S \) statistics. This might be a consequence of the excessive sensitivity of the classic \( R/S \) statistic to short-range dependences in various durations. We present the \( R/S \) analysis next.

4.2. Mandelbrot’s \( R/S \) Analysis. The actual and more realistic implementaion of the classic \( R/S \) statistic is through the so called \( R/S \) analysis (see, Mandelbrot69 [16]) and is decribed briefly as follows (see, Rachev00 [26]): let \( N \) be the sample size. Divide the sample in \( K \) blocks each of size \( N/K \) and define \( k_m = 1 + (m - 1) \frac{N}{K} \) with \( m = 1, \cdots, K \). Then estimate \( R_n(k_m) \) and \( S_n(k_m) \) for each lag \( n \) such that \( k_m + n < N \). Finally, plot \( \log \frac{R_n(k_m)}{S_n(k_m)} \) against \( \log n \) for each \( m = 1, \cdots, K \), in one figure. The resulting graph is called "rescaled adjusted range plot" or "pox plot of \( R/S \) (see, Rachev00 [26])".

To allow flexibility in grouping, we only selected the first 4000 entries from \( \hat{R} \) with group size \( S = 400 \) and provide the pox plot of \( R/S \) in Figure 17.

Note that in the Figure 17 the red line has slope 0.62 and the blue one has slope 0.5, the latter of which corresponds to the Hurst index of a standard Brownian motion. As can be seen clearly from the plot that for large sample size \( n \geq \exp(5) \), all \( \frac{R_n(k_m)}{S_n(k_m)} \) stays within the cone formed by \( L_1 = 0.62 \cdot \log x \) and \( L_2 = 0.5 \cdot \log x \) (which is relatively sufficient for asymptotic property). The most striking feature is that the line

\[
L = 0.57 \cdot \log n
\]
Figure 18. Autocorrelation till Lag Order 1000

gives a good least squares fit to these scattered $R/S$ statistics, which suggests the estimated Hurst index is

$$\hat{H} = 0.57$$

(4.3)

, in consistency with what most researchers have found as the Hurst index estimates for financial time series (see, Cont01 [5]).

Thus we can reject the hypothesis that the returns series of the SSE index follows a standard Brownian motion. Moreover, if the process $R$ has finite variance we then can conclude that the returns $R$ has a relatively shorter memory among long-range dependence processes as justified by the estimated long-range dependence parameter

$$\hat{d} = 0.07$$

(4.4)

(Note that this is also supported by the autocorrelations of $R$ for the maximal order of lag $L = 1000$ in Figure 18)

To reduce the excessive sensitivity of $R/S$ analysis to short-range dependences (see Figure 14), Lo91 ([12]) introduced the so called "Lo’s modified $R/S$ statistic" to detect overall long-range dependence without giving any estimation of the Hurst index $H$ and provided the asymptotic distribution for this statistic. Lo’s modified $R/S$ statistic is defined as

$$V_N(q) = \frac{1}{\sqrt{N}} \frac{R_N}{S_N(q)}$$

(4.5)

where

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} X_i, \omega_j(q) = 1 - \frac{j}{q+1}, q < N$$

(4.6)
and

\[ R_N = \max_{1 \leq q \leq N} \left( \sum_{j=1}^{q} X_j - q\hat{\mu} \right) - \min_{1 \leq q \leq N} \left( \sum_{j=1}^{q} X_j - q\hat{\mu} \right) \]  \tag{4.7}

and

\[ S_N^2(q) = \frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{\mu})^2 + \frac{2}{N} \sum_{j=1}^{q} \omega_j(q) \left( \sum_{i=j+1}^{N} (X_i - \hat{\mu}) (X_{i-j} - \hat{\mu}) \right) \]  \tag{4.8}

To use the Lo’s modified \( R/S \) statistic, we cite from in Lo91 [12] the following theorem

**Theorem 1** (Lo91 [12]). Suppose the random process \( \varepsilon = \{ \varepsilon_t : t \geq 0 \} \) satisfies the following conditions

\( a1) \): \( E(\varepsilon_t) = 0 \)

\( a2) \): \( \sup_t E \left( |\varepsilon_t|^{2\beta} \right) < \infty \) for some \( \beta > 2 \)

\( a3) \): \( 0 < \sigma^2 = \lim_{n \to \infty} E \left( \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j^2 \right) < \infty \)

\( a4) \): \( \{ \varepsilon_t \} \) is strongly mixing with mixing coefficients \( \alpha_k \) that satisfy

\[ \sum_{j=0}^{\infty} \alpha_j^{1-2/\beta} < \infty \]

\( a5) \): \( q \sim o(n^{1/4}) \) as both \( n \) and \( q \) approach to \( +\infty \)

Then

\[ V_N(q) \to^d W_1 = \max W_0(t) - \min W_0(t) \]  \tag{4.9}

if \( \{ \varepsilon_t \} \) has no long-range dependence, where \( W_0(t) \) is the standard Brownian bridge and

\[ P(W_1 \in [0.809, 1.862]) = 0.95 \]

Note that in the above theorem, at least uniformly the fourth moment is required to be finite in order to get the asymptotic distribution. Assumption \( a3) \) is a weak condition on the ergodicity of the variance and assumption \( a4) \) requires that the random variable \( \varepsilon_t \) can not be so dependent.

From the estimates already obtained, we know that \( \tilde{R} = \{ R_t : 1 \leq t \leq 4129 \} \) does not satisfy assumption \( a2) \) even though it might satisfy assumption \( a4) \) since the daily returns series have very weak dependence (see, Cont01 [5]). Based on the previous tail index estimates and running the risk that \( \tilde{R} \) might violate the fourth moment requirement \( a4) \), we still plotted the Lo’s \( R/S \) statistic \( V_N(q) \) in Figure 19 for the detrended \( R = \{ R_t - \mu : 1 \leq t \leq 4139 \} \) where \( \mu \) is the sample mean of \( \tilde{R} \)
Note in Figure 19 the confidence interval of the Lo’s $R/S$ statistic are given by the red and green horizontal lines (under the assumption of the previous theorem). Obviously, approximately for $q \geq 1500$, Lo’s $R/S$ statistic stays outside the 95% confidence interval and thus suggests the existence of long-range dependence. (It should be noted that Lo’s modified $R/S$ statistic is widely used even though not all assumptions in the above theorem are satisfied. See Rachev00 [26].)

4.3. **Direct Regression.** As Rachev00 [26] has pointed out that the $R/S$ analysis is still sensitive to the presence of short-range dependence, we adopt the method of direct regression therein to estimate the Hurst index, that is, for discrete observations \( \{X_t: 1 \leq i \leq N\} \) of a random process \( X = \{X_t: t \in T\} \) we can regress

\[
\hat{\rho}_k = \log \hat{E}(|X_{t+k} - X_t|)^{13}
\]

against $\log k$ to estimate the Hurst index $H$, where $\hat{E}$ is the sample mean and $i = 1, \cdots, N$. For the returns \( \tilde{R} = \{R_t: 1 \leq t \leq 4129\} \), the results are presented in Figure 20 in which the linear regressor is given by

\[
\hat{\rho}_k = \log \hat{E}(|R_{t+k} - R_t|) = 0.6133 \log k - 4.6021
\]

Thus the estimated Hurst index for the returns series $\tilde{R}$ is

\[
\hat{H}^* = 0.6133
\]

This result confirms the existence of long-range dependence in $\tilde{R}$ and is consistent with similar results obtained in Oh06 [23].

---

13If \( \{X_t, t = 1, 2, \ldots, N\} \) is the logarithmic SSE index, then $\hat{E}(|X_{t+k} - X_t|)$ is called the average absolute returns over $k$ quotes. (See, Rachev00 [26]).
5. Conclusions

We have mainly estimated the tail indices, Hurst index of the returns series of the SSE index, by non-parametric linear regression. These statistics are consistent with empirical estimates obtained by most researchers, except that returns series of SSE index skews more to the right. An imperfection of this paper is that no statistical tests have been carried out for the estimates, since we do not pre-assume any specific properties of the stochastic process that generates this index series. However, the results will definitely help to measure the appropriateness of the ergodic generalized hyperbolic diffusion as a model for the detrended logarithmic SSE index series in the next paper.

Please continue to the second part of the thesis after the references

References


SHANGHAI STOCK EXCHANGE II: NOT AN ERGODIC GENERALIZED HYPERBOLIC DIFFUSION

XIONGZHI CHEN

ABSTRACT. Motivated by recent successes of modeling stock prices by hyperbolic distributions, (generalized) hyperbolic diffusions (Bibby97 [5], Elerian98 [18], Rydberg99 [37]), this paper attempts to simulate the linearly detrended logarithmic SSE index series by an ergodic generalized hyperbolic (GH) diffusion which induces a returns series whose tail index and correlation structure match the estimates obtained in part one of this thesis and whose invariant density is proportional to some generalized hyperbolic (GH) density.

We adopt methodologies from Bibby97 [5], Bibby04 [7], and Sorensen02 [40] to build such a model and estimate the six parameters involved in the model by Markov chain Monte Carlo (MCMC) methods whose convergence criteria are the Monte Carlo standard error (MCSE) (see, Roberts96 [35]) and the Gelman-Rubin (GR) diagnostic (see, Gelman92 [20], Brooks98 [10], Brooks00 [11]), while retaining a good acceptance rate (see, Chib95 [12], Johannes03 [25]).

We run 6 independent chains in the simulation whose initial values are widely dispersed but centered around the nonlinear least squares estimate of the parameters of a GH density based on the empirical squared diffusion function of the linearly detrended logarithmic SSE index series. MCSEs suggests that all six independent chains have converged to their own stable distributions at around 50,000 iterations and remains at around $10^{-3}$ afterwards. Unfortunately the GR statistic does not converge to 1 even after 350,000 iterations and the acceptance rates became too low for six chains after around 50,000 iterations, so the simulation has to be terminated since further iterations will not give statistically significant moves of the chains.

We take the last 10% of the first 50,000 simulated samples from each chain in the simulation to obtain 6 Monte Carlo estimates for the parameter vector which suggest 6 possible models. Due to the computational inability of both Mathematica and Matlab resulted from the 2 badly mixed chains, only 4 chains are selected for the uniform residual tests (see, Bibby97 [5]).

Tests results show that all 4 possible models are rejected statistically. Thus under the scheme of approximating the unknown posterior for the parameter vector by the likelihood function obtained from the Milstein's discretization, we reject the hypothesis that "the underlying stochastic process for the linearly detrended logarithm SSE index in the period from 01/01/1991 to 14/11/2007 is an ergodic, GH diffusion", i.e., the linearly detrended logarithmic SSE index is not an ergodic GH diffusion.

Key words and phrases. Generalized Hyperbolic Distribution, Ergodicity, Invariant Measure, Markov chain Monte Carlo Methods, Metropolis-Hastings Algorithm, Gelman-Rubin Convergence Diagnostic/Statistic, Uniformal Residual Test.
List of Tables 3
List of Figures 4
1. INTRODUCTION 5
2. CONSTRUCTION OF ERGODIC, GH DIFFUSION 6
2.1. General Theory on Solutions of One Dimensional SDEs 6
2.2. Ergodic Diffusion with GH Invariant Density 8
3. MCMC ESTIMATION OF PARAMETERS 10
3.1. Milstein Discretization Scheme 11
3.2. MCMC Methods and Metropolis-Hastings Algorithm 15
3.3. Convergence Checking 17
3.4. Implementation and Estimates 19
3.5. Test of Model 22
4. CONCLUSION AND DISCUSSION 24
References 24
5. APPENDIX 27
5.1. Numerical Results 27
List of Tables

1  Monte Carlo Estimates for 6 Chains  
2  MCSEs  
3  Test Statistics for 4 Models
LIST OF FIGURES

1 Linearly Detrended Logarithmic SSE Index 11
2 Empirical Density of Detrended Index 20
3 Squared Volatility from Model 2 20
4 Maximal MCSEs for Individual Chains 21
5 Acceptance Rates for Individual Chains 21
6 QQ Plot of Uniform Residuals for Model 2 23
7 Sample Path for $\alpha$ for 6 Chains 27
8 Sample Path for $\beta$ for 6 Chains 28
9 Sample Path for $\delta$ for 6 Chains 29
10 Sample Path for $\lambda$ for 6 Chains 30
11 Sample Path for $\mu$ for 6 Chains 31
12 Sample Path for $\sigma$ for 6 Chains 32
13 Autocorrelation for $\alpha$ for 6 Chains 33
14 Autocorrelation for $\beta$ for 6 Chains 34
15 Autocorrelation for $\delta$ for 6 Chains 35
16 Autocorrelation for $\lambda$ for 6 Chains 36
17 Autocorrelation for $\mu$ for 6 Chains 37
18 Autocorrelation for $\sigma$ for 6 Chains 38
19 Model Based Squared Volatilities 39
20 Simulated Detrended Index from Models 1, 2, 3 40
21 Simulated Detrended Index from Models 4, 5, 6 41
22 QQ Plot of Uniform Residuals for 4 Tested Models 42
1. INTRODUCTION

In view of recent successful applications of generalized hyperbolic (GH) distributions in modeling asset returns, stock prices and option prices (see Bibby97 [5], Karsten99 [27], and Eberlein95 [16]), we tentatively assume that the underlying stochastic process

$$X_t = f_{X_t}(t)$$

for the linearly detrended logarithmic SSE index is an ergodic GH diffusion whose invariant density is some GH density and which for the trading days considered herein induces a returns series that has the tail indices and dependence structures as previously estimated in part one of this thesis. To be more specific, we try to solve following simulation

**Problem 1.** Construct an ergodic, weak solution

$$X_t = f_{X_t}(t)$$

for the parametric stochastic differential equation (SDE)

$$dX_t = b(X_t, \theta) dt + v(X_t, \theta) dB_t$$

with appropriate drift function $$b(\cdot)$$, diffusion function $$v(\cdot)$$, where the parameter $$\theta$$ lies in some bounded, open subset of $$\mathbb{R}^m$$ for some $$m \in \mathbb{N}$$, $$B = \{B_t, t \geq 0\}$$ is a standard Brownian motion, such that

1. $$X$$ has the generalized hyperbolic (GH) density

$$f_{gh}(x; \lambda, \alpha, \beta, \delta, \mu)$$

as its invariant or initial or asymptotic density, where the domain of variation of the parameters is

$$\lambda \in \mathbb{R}$$

and

$$\delta > 0, |\beta| < \alpha$$

if $$\lambda > 0$$

$$\delta > 0, |\beta| < \alpha$$

if $$\lambda = 0$$

$$\delta > 0, |\beta| \leq \alpha$$

if $$\lambda < 0$$

(1.3)

and the modified Bessel function $$K_\lambda$$ of the second kind is given by

$$K_\lambda(x) = \frac{1}{2} \int_0^\infty y^{\lambda-1} \exp \left( -\frac{x}{2} (y + y^{-1}) \right) dy, x > 0$$

(1.4)

2. and discretized on the trading days considered in part one of this thesis, $$X$$ induces a returns series that has tail indices and dependence structures as estimated in part one of this thesis.

We adopt the traditional notations and rewrite $$f_{gh}$$ compactly as

$$f_{gh}(x, \theta) = a(\lambda, \alpha, \beta, \delta) f_{gh-1/2}(\alpha \sqrt{\delta^2 + (x - \mu)^2})$$

$$\times \left( \delta^2 + (x - \mu)^2 \right)^{(\lambda-1)/2} \exp(\beta (x - \mu))$$

(1.5)

\[1\] "Detrended index" will denote "the linearly detrended logarithmic SSE index series", and the exact linear detrending will be explained in later sections.

\[2\] For a definition of "diffusion", please see Shreve98 [38]. For a definition of "generalized hyperbolic diffusion", please see Rydberg99 [37]
with parameter vector $\theta = (\lambda, \alpha, \beta, \delta, \mu)$ by letting

$$a(\lambda, \alpha, \beta) = \frac{1}{\sqrt{2\pi} \alpha^{1/2} \delta^{3/2} K_0 \left( \delta \sqrt{\alpha^2 - \beta^2} \right)} \left( \alpha^2 - \beta^2 \right)^{\lambda/2}$$

(1.6)

Further, we simply denote the unnormalized GH density $f_{gh}(x, \theta)$ by

$$f(x, \theta) = K_{\lambda-1/2} \left( \alpha \sqrt{\delta^2 + (x - \mu)^2} \right) \left( \delta^2 + (x - \mu)^2 \right)^{(\lambda-1)/2} \exp(\beta(x - \mu))$$

(1.7)

We adopt the method in Bibby95 [4] and Bibby97 [5] to build such an ergodic GH diffusion for the detrended index. As for parameter estimation we use Markov chain Monte Carlo (MCMC) methods driven by Metropolis-Hastings (MH) algorithm to sample from the unknown posterior of the parameter vector (see, Johannes03 [25] and Tes04 [42]. Note however that in Tse04 [42] at least two mistakes are found by the author of this thesis).

It is somehow sad that for the detrended index, due to the intense computational complexity, the method of martingale estimating function (see, Bibby95 [4] and Bibby97 [5]) has not been successfully implemented, and that due to lack of time, estimation methods introduced in Ait02 [3] have not been applied.

2. CONSTRUCTION OF ERGODIC, GH DIFFUSION

There are many references to building diffusion-type models with given marginal distributions and (exponentially-decreasing) autocorrelation functions (see Bibby97 [5], Bibby04 [7], Madan02 [30] and Ait02 [3]). All models constructed fall into three categories: (i) $X$ is a diffusion with a given initial density; (ii) $X$ is an ergodic diffusion whose asymptotic density or invariant density is the given density; (iii) $X$ is a diffusion with a given family of marginal densities. Herein, we only concentrate on the construction of a driftless, ergodic GH diffusion with a GH invariant density.

2.1. General Theory on Solutions of One Dimensional SDEs. First of all, we would like to cite some powerful theorems on the existence and uniqueness of weak solutions to one-dimensional stochastic differential equations (SDEs). These theorems give the sufficient or necessary and sufficient conditions on the existence, uniqueness, recurrence and ergodicity properties of solutions to such SDE's.

From Kutoyants04 [26] comes the following sufficient condition for the existence and ergodicity of a weak solution as a diffusion process.

Theorem 1 (Kutoyants04[26]). Suppose $b$ is locally bounded, $v$ is continuous and positive, and

$$xb(x) + v^2(x) \leq A(1 + x^2)$$

3Ait's method is the only one that gives explicit approximation to the transition densities of sufficiently good diffusions and estimates the unknown parameters only after a good approximation of the transition densities are obtained, thus distinguishing itself from all other methods that estimate the unknown parameters with a pre-specified model. For a comparison of estimation methods on discretely sampled diffusions, please see Jessen99 [24].
for some constant $A > 0$, then the stochastic differential equation
\[ dX_t = b(X_s) \, dt + v(X_t) \, dW_t \]
has a unique weak solution.

Moreover, if
\[ V(b, x) = \int_0^x \exp \left\{ -2 \int_0^y \frac{b(s)}{v^2(s)} \, ds \right\} \, dy \to \pm \infty, \text{ as } x \to \pm \infty \]
and
\[ G(b) = \int_{-\infty}^{+\infty} v^{-2}(x) \exp \left\{ 2 \int_0^x \frac{b(s)}{v^2(s)} \, ds \right\} \, dx < \infty \]
then the solution is ergodic with invariant density
\[ f_b(x) = G^{-1}(b) \, v^{-2}(x) \exp \left\{ 2 \int_0^x \frac{b(s)}{v^2(s)} \, ds \right\} \]

Remark 1. The conditions on existence of weak solutions in the above theorem is relatively strong but necessary for statistical inference.

From Shreve98 ([38]), we cite the elegant theorem proved by H. J. Engelbert and W. Schmidt (1984).

**Theorem 2** (Engelbert and Schmidt 1984). Equation
\[ dX_t = v(X_t) \, dW_t \]
has a non-exploding weak solution unique in probability law for every initial distribution $\mu_0$ if and only if for all real $x$
\[ v(x) = 0 \iff (\forall \varepsilon > 0) \int_{-\varepsilon}^\varepsilon \frac{dy}{v(x+y)^2} = \infty \]

Finally, we cite from Skorokhod89 [39] this powerful theorem on ergodicity:

**Theorem 3** (Skorokhod89). If $r_1 = -\infty, r_2 = +\infty, and \int_{-\infty}^{+\infty} b^{-2}(x) \, dx < \infty$, then $dX_t = b(X_t) \, dW_t$ is ergodic with ergodic distribution
\[ \pi(A) = \frac{\int_A v^{-2}(z) \, dz}{\int \nu^{-2}(z) \, dz} \]

Remark 2. In Theorem 3, $r_1, r_2$ denote respectively the left, right boundary for the state space of $X$. For an ergodic diffusion, its ergodic distribution corresponds to its invariant distribution. For more details on this, please see Skorokhod89 [39]. Here we briefly cite from Skorokhod89 [39] a definition that, a $\sigma$-finite measure $\mu$ is said to be invariant if $\mu T_t = \mu$ for all $t \geq 0$, where $\mu T_t(A) = \int P(t, x, A) \mu(dx) = E_\mu(I_A(x(t)))$ and $P(t, x, A)$ is the underlying transition density. The relationship between the diffusion coefficient and the invariant density can be also found in Bibby97 [5], where the above theorem is used.
2.2. Ergodic Diffusion with GH Invariant Density. With these theorems at hand, we are now ready to construct the proposed ergodic GH diffusion. We borrow the model from Bibby97 [5] and Rydberg99 [37] and assume that the underlying stochastic process $S$ for the SSE index satisfies
\[ S_t = \exp \left( X_t + \kappa t \right), \quad t \geq 0 \] (2.1)
where $\kappa$ is some constant to be determined and
\[ X_t = X_0 + \int_0^t v (X_s) \, dW_s \] (2.2)
for some $v (\cdot)$ satisfying the conditions which give existence and uniqueness in probability law of a weak solution to (2.2).

2.2.1. Model I. With such a specification in (2.1) $\sim$ (2.2), we have two choices: one is to estimate $\kappa$ by the ordinary least squares linear regression to obtain $\hat{\kappa}$ based on the observed SSE index $\tilde{P} = \{ P_t, t = 1, \ldots, 4131 \}$ and assume that
\[ \ln S_t - \hat{\kappa} t, \quad t = 1, \ldots, 4131 \] (2.3)
forms an observation of the ergodic diffusion
\[ X_t = X_0 + \int_0^t v (X_s) \, dW_s, \quad t \geq 0 \] (2.4)
whose invariant density is proportional to some unnormalized GH density $f (\cdot ; \theta)$ as defined in (1.7), then estimate the unknown parameter $\theta$ based on $\tilde{P}$; the other is to assume that
\[ \ln S_t = X_t + \kappa t = X_0 + \int_0^t v (X_s) \, dW_s + \int_0^t \kappa ds \] (2.5)
is itself an ergodic diffusion whose invariant density is some GH density defined in (1.2), then estimate the involved unknown parameter vector $\tilde{\theta} = (\alpha, \beta, \delta, \kappa, \lambda, \mu, \sigma)$ based on the observed SSE index $\tilde{P}$.

Hereunder we call (2.3) $\sim$ (2.4) the simplified Model I. Setting the diffusion function (or, diffusion coefficient) $v (\cdot)$ to be
\[ v \left( x; \tilde{\theta} \right) = \sqrt{\sigma^2 / f \left( x; \theta \right)} \] (2.6)
with the augmented parameter vector $\tilde{\theta} = (\theta, \sigma)$ yields
\[ X_t = X_0 + \int_0^t \sqrt{\sigma^2 / f \left( x; \tilde{\theta} \right)} \, dW_s \] (2.7)
Since $v \left( x; \tilde{\theta} \right) > 0$ for $x \in \mathbb{R}, \theta$ in (1.5), $\sigma \neq 0$ and $f$ has finite second order moment by the following moment generating function for the GH density.
\[ M (t) = e^{\lambda \mu} \left( \frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + t)^2} \right)^{\lambda/2} K_\lambda \left( \frac{\delta \sqrt{\alpha^2 - (\beta + t)^2}}{K_\lambda (\delta \sqrt{\alpha^2 - \beta^2})} \right), \quad |\beta + t| < \alpha \] (2.8)

\[ ^4 \text{Please see the first part of the thesis for the observed SSE index.} \]
(see, Karsten99 [27]), the conditions of Theorem 2, Theorem 3 are satisfied and \( X \) is weakly uniquely defined by (2.7) with ergodic density proportional to \( f (\cdot ; \theta) \).

2.2.2. Model II. However, if we take (2.5) \( \sim \) (2.6) as Model II and want to ensure that the invariant density is the GH density, then according to Theorem 1 we have to find \( v (\cdot ; \theta) \) such that

\[
V (x; \theta) = \int_0^x \exp \left\{ -2 \int_0^y \frac{\kappa}{v(s; \theta)^2} ds \right\} dy \to \pm \infty, \text{ as } x \to \pm \infty
\]

and

\[
G = \int_{-\infty}^{+\infty} v(x)^{-2} \exp \left\{ 2 \int_0^x \frac{\kappa}{v(s; \theta)^2} ds \right\} dx < \infty
\]

so that the invariant density is

\[
f (x; \theta) = \frac{v(x; \theta)^{-2} \exp \left\{ 2 \int_0^x \frac{\kappa}{v(s; \theta)^2} ds \right\}}{G}
\]

Remark 3. If the initial \( X_0 \) has the density \( f (\cdot ; \theta) \) then \( X \) is stationary.

More specifically, we have to solve for \( v (\cdot ; \theta) \) the following equation

\[
\begin{cases}
  f_{gh} (x; \theta) = G^{-1} v(x; \theta)^{-2} \exp \left\{ 2 \int_0^x \frac{\kappa}{v(s; \theta)^2} ds \right\} \\
  f_0^x \exp \left\{ -2 \int_0^y \frac{\kappa}{v(s; \theta)^2} ds \right\} dy \to \pm \infty, \text{ as } x \to \pm \infty
\end{cases}
\]

from which we know that \( G < \infty \) since \( f_{gh} \) has finite second moment.

From (2.12) we have

\[
v(x; \theta)^2 \frac{d \ln f_{gh} (x; \theta)}{dx} + 2v' (x; \theta) v(x; \theta) = 2\kappa
\]

where the superscript ‘ \( \text{'} \) denotes differentiation with respect to \( x \). Setting \( h (x; \theta) = v^2(x; \theta) \) and solving the equivalent

\[
\frac{d}{dx} \left\{ f_{gh} (x; \theta) h(x; \theta) \right\} = 2\kappa f_{gh} (x; \theta)
\]

yields

\[
h(x; \theta) = f_{gh}^{-1} (x; \theta) \left( 2\kappa \int_0^x f_{gh} (s; \theta) ds + G^{-1} \right)
\]

and

\[
v(x; \theta) = \sqrt{h(x; \theta)} = f_{gh}^{-1/2} (x; \theta) \left( 2\kappa \int_0^x f_{gh} (s; \theta) ds + G^{-1} \right)^{1/2}
\]

Next, we still have to show

\[
\lim_{x \to +\infty} \int_0^x \exp \left\{ -2 \int_0^y \frac{1}{v(s; \theta)^2} ds \right\} dy = +\infty
\]
By exploiting \( f_{gh}(x; \theta) = G^{-1} v(x; \theta)^{-2} \exp \left\{ 2 f_0^x \frac{1}{v(s; \theta)} ds \right\} \), we have

\[
I_1 : = \int_0^x \frac{dy}{\exp \left\{ 2 f_0^y v^{-2}(s; \theta) ds \right\}} = G^{-1} \int_0^x f_{gh}(y; \theta) v(y; \theta) dy = G^{-1} \int_0^x (2\kappa \int_0^y f_{gh}(s; \theta) ds + G^{-1})
\]
such that

\[
\lim_{x \to \pm \infty} I_1 = \pm \infty \text{ for } \kappa \geq 0
\]

Thus, the proposed ergodic diffusion model in (2.5) is

\[
S_t = \exp (\kappa t + X_t)
\]

and

\[
X_t = X_0 + \int_0^t \kappa ds + \int_0^t v(X_s; \theta) dW_s
\]

\[
= X_0 + \int_0^t \kappa ds + \int_0^t f_{gh}^{-1/2}(X_s; \theta) \left( 2\kappa \int_0^X f_{gh}(r; \theta) dr + G^{-1} \right)^{1/2} dW_s
\]

with invariant density \( f_{gh}(\cdot; \theta) \) for \( \kappa \geq 0 \).

Finally, applying Ito’s formula to \( H(t, x) = \exp (b(t) + x) \) to the model, we obtain \( S_t \) as

\[
dS_t = S_t \left\{ \left[ b'(t) + \frac{1}{2} v^2 (\ln S_t - b(t)) \right] dt + v (\ln S_t - b(t)) dW_t \right\}
\]

(2.18)

**Remark 4.** It is well known that under certain non-degeneracy and locally integrability conditions, a stochastic integral-differential equation can be transformed into a driftless equation (see, Shreve98 [38]) or a constant diffusion equation (see, Rogers79 [36]). But in most cases with the presence of drift functions, we will have to use the methods in Ait02 [3] or in Lo88 [28].

3. MCMC ESTIMATION OF PARAMETERS

To reduce the computational complexity resulted from computing the numerical integral \( \int_0^x f_{gh}(r; \theta) dr \) involved in Model II, we will only estimate the model

\[
\ln S_t - \tilde{\kappa} t = X_t = X_0 + \int_0^t v \left( X_s; \tilde{\theta} \right) dW_s, t \geq 0
\]

(3.1)
defined in (2.1) \sim (2.4) with

\[
v \left( \cdot; \tilde{\theta} \right) = \sqrt{\frac{\sigma^2}{f(\cdot; \theta)}}
\]

(3.2)

where \( S = \{S_t, t \geq 0\} \) here denotes the underlying stochastic process for the observed SSE index \( \tilde{P} \) and \( X = \{X_t, t \geq \} \) denotes that for the detrended index \( \tilde{D} = \{\ln S_t - \tilde{\kappa} t, t = 1, \ldots, 4131\} \) which is also denoted by \( X \) as a discrete observation of \( X \). (See Figure 1 for the plot of \( \tilde{D} \).)
3.1. Milstein Discretization Scheme. To approximate the likelihood function and the unknown posterior \( p(\theta | X) \) for the parameter vector \( \hat{\theta} = (\alpha, \beta, \delta, \lambda, \mu, \sigma) \) given the discrete observation \( X = \{X_t, t = 1, 2, ..., n\} \), we start from the Milstein approximation scheme (of strong convergence of order one) by discretizing model (3.1) for a small change \( \Delta t \) in \( t \) into

\[
X_{t+\Delta t} = X_t + v(X_t; \hat{\theta}) \Delta W_t + \frac{1}{2} v(X_t; \hat{\theta}) \frac{\partial v(X_t; \hat{\theta})}{\partial X_t} [ (\Delta W_t)^2 - \Delta t ]
\]

which can be rewritten into

\[
X_{t+\Delta t} - X_t + \frac{1}{2} v(X_t; \hat{\theta}) \frac{\partial v(X_t; \hat{\theta})}{\partial X_t} \Delta t + \frac{v(X_t; \hat{\theta})}{2 \partial \partial X_t v(X_t; \hat{\theta})} (3.3)
\]

\[
= \left( v(X_t; \hat{\theta}) \sqrt{\Delta t} \right) \epsilon + \left( \frac{1}{2} v(X_t; \hat{\theta}) \frac{\partial v(X_t; \hat{\theta})}{\partial X_t} \Delta t \right) \epsilon^2 + \frac{v(X_t; \hat{\theta})}{2 \partial \partial X_t v(X_t; \hat{\theta})}
\]

where \( \frac{\partial v(X_t; \hat{\theta})}{\partial X_t} := \lim_{\Delta x \to 0} \frac{v(X_t; \hat{\theta}) - v(x; \hat{\theta})}{\Delta x} \bigg|_{x=X_t} \) and \( \epsilon \) is standard normal.
In order to obtain the conditional density after discretization, we use the results from Elerian98 [18] and set
\[
\begin{align*}
A_t &= \frac{1}{2} v \left( X_t; \hat{\theta} \right) \frac{\partial v \left( X_t; \hat{\theta} \right)}{\partial X_t} \Delta t \\
B_t &= -\frac{v \left( X_t; \hat{\theta} \right)}{2 \partial \theta X_t v \left( X_t; \hat{\theta} \right)} + X_t - A_t \\
\Lambda_t &= \frac{\Delta t \left( \partial \theta X_t v \left( X_t; \hat{\theta} \right) \right)^2}{1} \\
y_{t+1, \Delta t} &= \frac{X_{t+1} - B_t}{A_t}
\end{align*}
\] (3.4)
to conclude that \( y_{t+1, \Delta t} = (\epsilon + \sqrt{\lambda})^2 \) and

\[
h \left( X_{t+\Delta t} | X_t \right) = \frac{1}{|A_t|} \chi^2 \left( \frac{X_{t+1} - B_t}{A_t}, 1, \Lambda_t \right)
\]
where \( \chi^2 (\cdot, d, l) \) is a non-central chi-squared density function with \( d \) degrees of freedom and location parameter \( l \). More specifically

\[
\chi^2 \left( z, d, l \right) = \frac{1}{2} \exp \left\{ -\frac{l + z}{2} \right\} \left( \frac{z}{\sqrt{1}} \right)^{d/4-1/2} I_{-1/2} \left( \sqrt{1} \right)
\]

where \( I_{-1/2} (w) = \sqrt{\frac{2}{\pi w}} \cosh (w) \).

In our case with \( d = 1 \), we have

\[
\chi^2 \left( z, 1, l \right) = \frac{1}{2} \exp \left\{ -\frac{l + z}{2} \right\} \left( \frac{z}{\sqrt{1}} \right)^{-1/4} I_{-1/2} \left( \sqrt{1} \right)
\]

\[
= \frac{1}{2} \exp \left\{ -\frac{l + z}{2} \right\} z^{-1/4} t^{1/4} \sqrt{\frac{2}{\sqrt{\pi \sqrt{1}}} \cosh \left( \sqrt{1} \right)}
\]

\[
= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{l + z}{2} \right\} z^{-1/2} \cosh \left( \sqrt{1} \right)
\]

and

\[
h \left( X_{t+\Delta t} | X_t; \hat{\theta} \right) = \exp \left\{ -0.5 \left( \Lambda_t + \frac{X_{t+\Delta t} - B_t}{A_t} \right) \right\} \cosh \left( \sqrt{\frac{\Lambda_t (X_{t+\Delta t} - B_t)}{A_t}} \right)
\]

\[
= \sqrt{2\pi} |A_t| \sqrt{\frac{(X_{t+\Delta t} - B_t)}{A_t}}
\]

which saves a bit of computation time in computing the uniform residuals for model test in later subsection.

With the Milstein’s discretization, the likelihood function \( \hat{L}_M \left( \hat{\theta} | \mathbf{X} \right) \) of \( \hat{\theta} \) given observation \( \mathbf{X} \) is approximated by

\[
L_M \left( \hat{\theta} | \mathbf{X} \right) := \prod_{t=1}^{n} h \left( X_{t+\Delta t} | X_t; \hat{\theta} \right)
\] (3.6)
which further approximates the unknown \( p(\hat{\theta}|\mathbf{X}) \) (by the Bayesian rule).

However, based on the observation \( \mathbf{X} \) (of the detrended index), the approximate likelihood function \( \tilde{L}_M (\hat{\theta}|\mathbf{X}) \) generates huge quantities which are even beyond the computational capabilities of Mathematica and Matlab. Thus we use another scheme by approximating the unknown posterior \( p(\hat{\theta}|\mathbf{Y}) \) based on the transformed observation \( \mathbf{Y} \) as follows.

We set

\[
\begin{align*}
g (X_t; \hat{\theta}) &= \frac{1}{2} v (X_t; \hat{\theta}) \frac{\partial v (X_t; \hat{\theta})}{\partial X_t} \\
c_t &= v (X_t; \hat{\theta}) \sqrt{\Delta t} \\
d_t &= g (X_t; \hat{\theta}) \Delta t = \frac{1}{2} v (X_t; \hat{\theta}) \frac{\partial v (X_t; \hat{\theta})}{\partial X_t} \\
Y_{t,\Delta t} &= X_{t+\Delta t} - X_t + g (X_t; \hat{\theta}) \Delta t = X_{t+\Delta t} - X_t + d_t
\end{align*}
\]

to obtain

\[
Y_{t,\Delta t} = c_t \epsilon + d_t \epsilon^2 = d_t \left[ \left( \epsilon + \frac{c_t}{2d_t} \right)^2 - \frac{c_t^2}{4d_t^2} \right]
\]

where \( \epsilon \) is standard normal. Since \( Z = \left( \epsilon + \frac{c_t}{2d_t} \right)^2 \) has density

\[
h (z; \hat{\theta}) = \frac{1}{2} \exp \left\{ - \frac{r_t + z}{2} \right\} \left( \frac{z}{r_t} \right)^{-1/4} I_{-1/2} (\sqrt{r_t} z)
\]

where \( r_t = c_t^2 / (4d_t^2) \) and \( I_{-1/2} (w) = \sqrt{\frac{2}{\pi w}} \cosh (w) \) (see, Elerian98 [18]), then the density of \( Y_{t,\Delta t} \) is

\[
h^* (y; t, \hat{\theta}) = \frac{1}{|d_t|} h \left( \frac{y}{d_t} + \frac{c_t^2}{4d_t^2} \right)
\]

and the approximate likelihood given the observations \( \mathbf{Y} = \{ Y_{t,\Delta t}; t = 1, ..., n \} \) is

\[
L_M (\hat{\theta}|\mathbf{Y}) = \prod_{t=1}^{n} h^* (y; t, \hat{\theta})
\]
Next, we compute the functions needed in the Milstein’s discretization. Firstly we have

\[ \frac{\partial}{\partial X_t} v(X_t; \hat{\theta}) = \frac{\sigma^2 f'(X_t; \hat{\theta})}{f(X_t; \hat{\theta})} \]

(3.11)

\[ = -\frac{1}{2} \frac{\sigma^2 f^{-2}(X_t; \hat{\theta})}{v(X_t, \hat{\theta})} \frac{d}{dX_t} f(X_t; \hat{\theta}) \]

\[ = -0.5v(X_t; \hat{\theta}) \frac{d}{dX_t} \ln f(X_t; \hat{\theta}) \]

and secondly we have

\[ \frac{d}{dx} \ln f_{gh}(x; \hat{\theta}) \]

\[ = \frac{d}{dx} \left\{ \ln K_{\frac{\lambda}{2}} \left( \alpha \sqrt{\delta^2 + (x - \mu)^2} \right) + \frac{\lambda - 1/2}{2} \ln \left( \delta^2 + (x - \mu)^2 \right) + \beta (x - \mu) \right\} \]

\[ = \beta + \frac{(\lambda - 1/2) (x - \mu)}{\delta^2 + (x - \mu)^2} + \frac{d}{dx} \ln K_{\frac{\lambda}{2}} \left( \alpha \sqrt{\delta^2 + (x - \mu)^2} \right) \]

\[ = \beta + \frac{(\lambda - 1/2) (x - \mu)}{l(x; \delta, \mu)} - \frac{K_{\frac{\lambda+1}{2}} \left( \alpha \sqrt{l(x; \delta, \mu)} \right) + K_{\frac{\lambda-3}{2}} \left( \alpha \sqrt{l(x; \delta, \mu)} \right)}{2 \times K_{\frac{\lambda}{2}} \left( \alpha \sqrt{l(x; \delta, \mu)} \right)} \frac{\alpha (x - \mu)}{\sqrt{l(x; \delta, \mu)}} \]

from the fact that

\[ \frac{d}{dx} \ln K_{\frac{\lambda}{2}} \left( \alpha \sqrt{\delta^2 + (x - \mu)^2} \right) \]

(3.12)

\[ = -\frac{1}{2} \left( K_{\frac{\lambda+1}{2}} \left( \alpha \sqrt{l(x; \delta, \mu)} \right) + K_{\frac{\lambda-3}{2}} \left( \alpha \sqrt{l(x; \delta, \mu)} \right) \right) \]

\[ \times \frac{\alpha (x - \mu)}{\sqrt{\delta^2 + (x - \mu)^2}} \]

\[ = -\frac{1}{2} \left( K_{\frac{\lambda+1}{2}} \left( \alpha \sqrt{l(x; \delta, \mu)} \right) + K_{\frac{\lambda-3}{2}} \left( \alpha \sqrt{l(x; \delta, \mu)} \right) \right) \frac{\alpha (x - \mu)}{\sqrt{l(x; \delta, \mu)}} \]

where

\[ l(x; \delta, \mu) = \delta^2 + (x - \mu)^2 \]

(3.13)

(see Karsten99 [27] for properties of the modified Bessel function of the third kind).

\[ ^{5} \text{It is obvious that for the observed detrended index, we have } \Delta t = 1. \]
Consequently
\[
\frac{\partial}{\partial x_t} v(X_t; \hat{\theta}) = -0.5 v(X_t; \hat{\theta}) \frac{d}{dx} \ln f(X_t; \hat{\theta})
\]
(3.14)
\[
= -0.5 v(X_t; \hat{\theta}) \times \left\{ \beta + \frac{(\lambda - 1/2) (X_t - \mu)}{I(X_t; \delta, \mu)} - \frac{K_{\lambda+1/2} (\alpha \sqrt{I(X_t; \delta, \mu)}) + K_{\lambda-3/2} (\alpha \sqrt{I(X_t; \delta, \mu)})}{2K_{\lambda-1/2} (\alpha \sqrt{I(X_t; \delta, \mu)})} (X_t - \mu) \right\}
\]

3.2. MCMC Methods and Metropolis-Hastings Algorithm. To estimate the parameter vector \( \hat{\theta} \) in Model I by Markov chain Monte Carlo (MCMC) methods we first remove the constraints on the ranges of the component parameters (except for \( |\beta| < \alpha \)) by a batch of transformations
\[
\begin{align*}
\delta &= \exp(\delta^*), \delta^* \in \mathbb{R} \\
\sigma &= \exp(\sigma^*), \sigma^* \in \mathbb{R} \\
\alpha &= \exp(\alpha^*), \alpha^* \in \mathbb{R} \\
\beta^* &= \ln \frac{\beta + \alpha}{\beta - \alpha} \Leftrightarrow \beta = \frac{\exp(\beta^*) - 1}{\exp(\beta^*) + 1} \exp(\alpha^*), \beta^* \in \mathbb{R}
\end{align*}
\]
(3.15)
Let the new parameter vector be denoted by \( \tilde{\theta} = (\alpha^*, \beta^*, \delta^*, \lambda, \mu, \sigma^*) = (\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4, \tilde{\theta}_5, \tilde{\theta}_6) \in \mathbb{R}^6 \) and still let \( \theta = (\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4, \tilde{\theta}_5, \tilde{\theta}_6) \) for notational simplicity, we then have
\[
v(x; \hat{\theta}) = \sqrt{\frac{\exp(\sigma^*)}{f(x; \hat{\theta})}}
\]

By the Bayesian formula, the posterior density \( \hat{p}(\tilde{\theta} | Y) \) for the parameter vector \( \tilde{\theta} \) given the observation \( Y \) is approximated by
\[
p(\tilde{\theta} | Y) = \frac{p(Y | \hat{\theta}) \pi(\hat{\theta})}{\int p(Y | \hat{\theta}) \pi(\hat{\theta}) d\hat{\theta}} \propto \pi(\hat{\theta}) L_M(\tilde{\theta} | Y)
\]
(3.16)
where \( \pi(\hat{\theta}) \) is the prior density for \( \hat{\theta}, L_M(\tilde{\theta} | Y) \) is as in (3.10).

To implement the MCMC methods to estimate \( \tilde{\theta} \), we set the prior \( \pi(\hat{\theta}) \) to be a radially symmetric uniform distribution centered at zero with compact support as large as required, and propose the next value of parameter vector \( \tilde{\theta}' \) by generating it from the multivariate normal proposal \( N(\tilde{\theta}_n, \Gamma) \) with mean vector \( \tilde{\theta}_n \) and a pre-assigned constant covariance matrix \( \Gamma \), where \( \tilde{\theta}_n \) is the current value of the parameter vector \( \hat{\theta} \) at the \( n \)-th iteration in the Metropolis-Hastings (MH) algorithm that drives the proposed Markov chain.
Under such settings, the acceptance probability for \( \hat{\theta}' \) given \( \hat{\theta}_n \) is simply

\[
\alpha \left( \hat{\theta}' \big| \hat{\theta}_n \right) = \min \left\{ 1, \frac{p \left( \hat{\theta}' \big| Y \right) q \left( \hat{\theta}_n \big| \hat{\theta}' \right)}{p \left( \hat{\theta}_n \big| Y \right) q \left( \hat{\theta}' \big| \hat{\theta}_n \right)} \right\} \tag{3.17}
\]

\[
= \min \left\{ 1, \frac{p \left( Y \big| \hat{\theta}' \right) \pi \left( \hat{\theta}' \right) q \left( \hat{\theta}_n \big| \hat{\theta}' \right)}{p \left( Y \big| \hat{\theta}_n \right) \pi \left( \hat{\theta}_n \right) q \left( \hat{\theta}' \big| \hat{\theta}_n \right)} \right\} \approx \min \left\{ 1, \frac{L_M \left( \hat{\theta}' \big| Y \right) \pi \left( \hat{\theta}' \right) q \left( \hat{\theta}_n \big| \hat{\theta}' \right)}{L_M \left( \hat{\theta}_n \big| Y \right) \pi \left( \hat{\theta}_n \right) q \left( \hat{\theta}' \big| \hat{\theta}_n \right)} \right\} = \min \left\{ 1, \frac{L_M \left( \hat{\theta}' \big| Y \right)}{L_M \left( \hat{\theta}_n \big| Y \right)} \right\}
\]

where \( p \left( Y \big| \hat{\theta} \right) \) is the joint posterior for the transformed observation \( Y \) given \( \hat{\theta} \) and \( q \left( \hat{\theta}' \big| \hat{\theta}_n \right) = N \left( \hat{\theta}_{n+1}; \hat{\theta}_n, \Gamma \right) \) (Note that \( N \left( \cdot ; \hat{\theta}_n, \Gamma \right) \) here also denotes a multivariate normal density with mean vector \( \theta_n \) and covariance matrix \( \Gamma \)).

The Metropolis-Hastings (MH) algorithm (see, for example, Chib95 [12]) is summarized below:

**Step 1:** Given the current state \( \hat{\theta}_n \), generate a candidate \( \hat{\theta}' \) from the proposal density \( q \left( \cdot \big| \hat{\theta}_n \right) \)

**Step 2:** Calculate the acceptance probability \( \alpha \left( \hat{\theta}' \big| \hat{\theta}_n \right) \)

**Step 3:** Generate a random number \( u \) from the uniform distribution on \((0, 1)\). If \( u \leq \alpha \left( \hat{\theta}' \big| \hat{\theta}_n \right) \), set \( \hat{\theta}_{n+1} = \hat{\theta}' \). Otherwise, set \( \hat{\theta}_{n+1} = \hat{\theta}_n \).

**Step 4:** Repeat the previous steps to obtain a chain \( \{ \hat{\theta}_0, \hat{\theta}_2, \ldots \} \), where \( \hat{\theta}_0 \) denotes the initial state of \( \hat{\theta} \). Discard the burn-in values \( \{ \hat{\theta}_0, \hat{\theta}_0, \ldots \hat{\theta}_m \} \) for some preassigned \( m \in \mathbb{N} \) (probably depending on the acceptance rates) obtained while the chain converges in distribution. Then the remaining values, \( \{ \hat{\theta}_{m+1}, \hat{\theta}_{m+1}, \ldots \} \), form a correlated Markov chain sampled from its stationary density which is \( p \left( \hat{\theta} \big| Y \right) \).

As long as the chain converges, the post burn-in values, say, \( \{ \hat{\theta}_{m+1}, \hat{\theta}_{m+1}, \ldots, \hat{\theta}_{n_1} \} \) are used to estimate the unknown parameters by Monte Carlo methods. In this paper, the ergodic mean

\[
\hat{\theta}^* = \frac{1}{n_1 - m} \sum_{i=m+1}^{n_1} \hat{\theta}_i \rightarrow E_{p(\theta \big| X)} \left( \hat{\theta} \big| Y \right), \text{ a.s.} \tag{3.18}
\]

\(^{6}\)Note that the acceptance criterion adopted here is the most widely used one.
which is exactly the Monte Carlo estimate\(^7\) of the unknown \(\hat{\theta}\), is used as an estimate of the unknown \(\hat{\theta}\), where "\(\rightarrow a.s.\)" denotes "strong convergence" and the expectation is taken with respect to the posterior density \(p(\hat{\theta} | Y)\).

(For more details on Monte Carlo methods in finance, see Glasserman04 \([32]\)).

3.3. Convergence Checking. In the actual implementation of the MCMC methods, the sampled path, denoted by \(\{\theta_i : i = 1, 2, \ldots, N\}\), forms a convergent Markov chain whose stationary density is the unknown posterior \(p(\hat{\theta} | Y)\), and the target quantity \(E_p(\hat{\theta} | Y) [\varphi(\hat{\theta})]\) is estimated by the ergodic average, i.e., the Monte Carlo (MC) estimate

\[
\hat{\varphi}_N = \frac{1}{N} \sum_{i=1}^{N} \varphi(\hat{\theta}^{(i)})
\]

where \(\varphi(\cdot)\) is a (pre-assigned) real-valued function, \(E_p(\hat{\theta} | Y) [\cdot]\) again denotes the expectation with respect to \(p(\hat{\theta} | Y)\). In this paper, \(\varphi\) is just the identity function.

Yet the most important part in MCMC simulation is convergence checking (since the best we can do is not to reject the null hypothesis that the chain has converged), which can be done by a combination of checking Monte Carlo standard errors (MCSEs) and the Gelman-Rubin convergence diagnostic.

3.3.1. MCSE Convergence Criterion. In Roberts96 \([35]\) is stated the following convergence theorem

\[
\sqrt{N} \left( \hat{\varphi}_N - E_p(\hat{\theta} | Y) [\varphi(\hat{\theta})] \right) \overset{d}{\to} N(0, s^2_{\varphi})
\]

for some \(s^2_{\varphi} \geq 0\), where \(\overset{d}{\to}\) denotes convergence in distribution and \(N(a, b)\) denotes a normal density with mean \(a\) and variance \(b\). Thus the accuracy of the ergodic average as an estimate of \(E_p(\hat{\theta} | Y) [\varphi(\hat{\theta})]\) is essentially measured by \(s^2_{\varphi}\); a quantity that can only be empirically estimated from the sample. In practice, \(s^2_{\varphi}\) is estimated by the so called "batch mean", where the MCMC algorithm is run for \(N = m \times n\) iterations for sufficiently large \(n\) so that

\[
y_k = \frac{1}{n} \sum_{i=(k-1)n+1}^{kn} \varphi(\hat{\theta}^{(i)}), k = 1, 2, ..., m
\]

are approximately independently distributed as

\[
N \left( E_p(\hat{\theta} | X) [\varphi(\hat{\theta})], s^2_{\varphi}/n \right)
\]

\(^7\)Monte Carlo estimate is essentially just the average of the sample from the target quantity, regardless of ergodicity. Theoretical convergence of the Markov chains in the simulation and ergodicity of the means obtained from the post burn-in samples here are ensured by the detailed balance condition built into the Metropolis-Hastings algorithm.
and then $s_\varphi^2$ is estimated by

$$
\hat{s}_\varphi^2 = \frac{n}{m-1} \sum_{k=1}^{m} (y_k - \tilde{\varphi}_N)
$$

(3.22)

Subsequently the standard error of $\tilde{\varphi}_N$ can be estimated by $\sqrt{\hat{s}_\varphi^2/N}$, which is called the Monte Carlo standard error (MCSE) and is used to measure the mixing performance of the MH algorithm and as a convergence diagnostic for the MCMC algorithm.

3.3.2. Gelman-Rubin Statistic. Since a single Markov chain may converge locally to some mode of the unknown posterior and the MSCE is not able to diagnose whether the proposed unknown posterior has been sufficiently sampled, we have to run multiple chains in the MCMC simulation to ensure a large proportion of the unknown posterior is to be sampled and check whether all chains have converged to the same stationary distribution. Consequently, we adopt the famous Gelman-Rubin (GR) statistic\(^8\) (see Gelman\(^9\) [20], Brooks\(^0\) [11], and Brooks\(^1\) [10]) for convergence diagnosis of multiple-sequence simulations.

Now we describe how to construct the GR statistic in both univariate and multivariate cases. We independently simulate $I \geq 2$ sequences of length $2N$, each beginning with different initial values that are sufficiently\(^9\) widely dispersed with respect to the unknown yet stationary distribution (to which the $I$ independent Markov chains are supposed to converge). Then we discard the first $N$ iterations of each simulated sequence and retain only the last $N$. For the purpose of our simulation here, we compute variance between the $I$ sequence means, which are denoted by $\hat{\theta}^{(i)}$, and define

$$
B = \frac{1}{I-1} \sum_{i=1}^{I} \left( \hat{\theta}^{(i)} - \hat{\theta} \right)
$$

(3.23)

where $\hat{\theta}^{(i)}_t$ denotes the $t$-th observation of $\hat{\theta}$ from the $i$-th chain and

$$
\hat{\theta}^{(i)} = \frac{1}{N} \sum_{t=N+1}^{2N} \hat{\theta}^{(i)}_t, \quad \hat{\theta} = \frac{1}{I} \sum_{i=1}^{I} \hat{\theta}^{(i)}
$$

Further, we define $W$, the mean of the $I$ within-sequence variances, $s_2^{(i)}$, by

$$
W = \frac{1}{I} \sum_{i=1}^{I} s_2^{(i)}
$$

(3.24)

where

$$
s_2^{(i)} = \frac{1}{N-1} \sum_{t=N+1}^{2N} \left( \hat{\theta}^{(i)}_t - \hat{\theta}^{(i)} \right)^2
$$

\(^8\)In the literature, the GR statistic is often called the "Potential Scale Reduction Ractor (PSRF)".

\(^9\)It is possible to measure how "sufficiently" the initial values are dispersed, see Dempster\(^7\) [14]. Also a method to choose sufficiently widely dispersed initials are given in this paper. However due to the time constraints for the project, the author was not able to use the method suggested therein.
Finally if $\tilde{\theta}$ is univariate, then we define
\begin{equation}
\hat{V} = \frac{N - 1}{N} W + \left(1 + \frac{1}{I}\right) B \quad (3.25)
\end{equation}
and
\begin{equation}
\hat{R} = \frac{\hat{V}}{W} \quad (3.26)
\end{equation}
Otherwise, we define
\begin{equation}
\hat{R}_p = \frac{N - 1}{N} + \left(1 + \frac{1}{I}\right) \rho \quad (3.27)
\end{equation}
where $\rho$ is the largest eigenvalue of the positive definite matrix $W^{-1}B$. (See Brooks98 [10] for more details on (3.26), (3.27).)

The key property of $\hat{R}, \hat{R}_p$ is that they both converge to 1 when all individual Markov chains in the MCMC simulation converge to the same stationary density. Consequently they can be used as convergence diagnostics. Another interesting relationship between them is that, if $\hat{R}(k)$ denotes the GR statistic for the $k$-th component of a $d$-dimensional multivariate parameter vector $\tilde{\theta}$ in multiple-chain MCMC simulation, then $\max_{1 \leq k \leq d} \left\{ \hat{R}(k) \right\} \leq \hat{R}_p$. Since each $\hat{R}(k) \geq 1, k = 1, ..., d$ if the initial values for the Markov chains in the simulation are sufficiently widely dispersed (see Brooks00 [11]), we do not have to check the convergence for each $\hat{R}(k), k = 1, ..., d$ when convergence of $\hat{R}_p$ is confirmed.

3.3.3. Confidence Interval under GR Scheme. Under the Gelman-Rubin diagnostic scheme, a $100(1 - p)\%$ central posterior confidence interval for the parameter $\tilde{\theta}$ is obtained based upon the final $N$ of $2N$ iterations as described below (see, Brooks98 [10]):

**S1:** From each individual chain, take the empirical $100(1 - p)\%$ interval, that is, the $100P \%$ and the $100(1 - p)\%$ points of the $N$ simulation draws to form $I$ within-sequence interval length estimates

**S2:** From the entire set of $IN$ observations gained from all chains, compute the empirical $100(1 - p)\%$ interval to obtain a total-sequence interval length estimate

**S3:** Evaluate $\hat{R}$ defined as
\begin{equation}
\hat{R}_{\text{Interval}} = \frac{\text{length of total-sequence interval}}{\text{mean length of within-sequence intervals}}
\end{equation}

3.4. Implementation and Estimates. Since the approximate posterior $p(\tilde{\theta} | Y)$ is proportional to $L_M(\tilde{\theta} | Y)$ and the prior $\pi(\tilde{\theta})$ is chosen to be a radially symmetric uniform distribution centered at zero, we see that the modes of $p(\tilde{\theta} | Y)$ depends on those of $L_M(\tilde{\theta} | Y)$. In order to choose initial values that are sufficiently widely dispersed with respect to $p(\tilde{\theta} | Y)$, we estimated all 3 modes of the empirical density of the detrended index (see, Figure 2 for this density and its nonlinear least squares fit by a GH density) and for each mode a group of two independent chains are simulated. Unfor-
Fortunately this simulation scheme gives negative $z$’s in (3.8), which forces the author to choose the following initialization strategy.

Firstly by fitting the empirical squared diffusion of the detrended index by the $v(\cdot; \tilde{\theta})$ in (2.6) in the least square sense (within the working precision of Matlab), we obtain the initial estimate

$$\tilde{\theta}_0 = (\alpha_0, \beta_0, \delta_0, \lambda_0, \mu_0, \sigma_0) = (-0.3090, -2.7787, -19.2004, 3.2323, 3.6693, -3.7583)$$

(Figure 3 plots the squared diffusion function for $X = \{X_t, 1 \leq t \leq 4131\}$ based on the estimated parameter vector from chain 2 in the simulation)

Then we simply propose 6 independent Markov chains whose initial states are the 6 different vectors generated by a symmetric uniform distribution centered at $\tilde{\theta}_0$ and start the Metropolis-Hastings algorithm. The tuning
paramter is set to be $\tau = 0.0015$, partially based on the dimension of the parameter space (see, Haario01 [23], Roberts08 [34]) and by the requirement to control the acceptance rates to be between 0.2 and 0.5 (see, Johannes03 [25], Chib95 [12], and Tse04 [42]).

At around 50,000 iterations MCSEs suggests that all six independent chains have converged to their own stationary distributions and remains at around $10^{-3}$ afterwards. (See Figure 4 for the plot of the maximum component MSCEs for 6 individual chains for the first 50,000 iterations.)

But the GR statistic does not convergence to 1 even after 350,000 iterations and the average acceptance rates are too low for six chains after around 50,000 iterations, so the simulation has to be terminated since the 6 chains might have converged to different stationary distributions and further iterations will not statistically move the chains forward. (See Figure 5 for the acceptance rates and the figures for the badly mixed sample paths in the appendix)

We take the last 10% of the first 50,000 simulated samples (which we call the "effective samples" ) from each chain in the simulation to reduce or remove the effects of initilization in the Markov chains, and obtain 6 MC estimates for the parameter vector $\hat{\theta}$, which suggests 6 possible models. Table 1 summarizes the MC estimates for each component parameter for each individual chain and the corresponding MSCEs for the 100-th batch are

<table>
<thead>
<tr>
<th>Chain</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\delta$</th>
<th>$\lambda$</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.633729</td>
<td>-3.003546</td>
<td>-19.157304</td>
<td>3.041060</td>
<td>4.578130</td>
<td>-3.509746</td>
</tr>
<tr>
<td>2</td>
<td>-0.358308</td>
<td>-2.672644</td>
<td>-20.055300</td>
<td>2.611961</td>
<td>4.382647</td>
<td>-5.049936</td>
</tr>
<tr>
<td>3</td>
<td>-0.412318</td>
<td>-1.712247</td>
<td>-18.745429</td>
<td>2.597826</td>
<td>3.120015</td>
<td>-5.662547</td>
</tr>
<tr>
<td>4</td>
<td>-0.423069</td>
<td>-1.950044</td>
<td>-18.693538</td>
<td>2.999596</td>
<td>3.721765</td>
<td>-4.576278</td>
</tr>
<tr>
<td>5</td>
<td>-0.277084</td>
<td>-1.368601</td>
<td>-19.774321</td>
<td>3.635318</td>
<td>3.143871</td>
<td>-4.097122</td>
</tr>
<tr>
<td>6</td>
<td>-0.389012</td>
<td>-2.301538</td>
<td>-19.248898</td>
<td>2.943287</td>
<td>4.338962</td>
<td>-4.459357</td>
</tr>
</tbody>
</table>

| Table 1. Monte Carlo Estimates for 6 Chains |
Figure 5. Acceptance Rates for Individual Chains

<table>
<thead>
<tr>
<th>Chain</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\delta$</th>
<th>$\lambda$</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain 1</td>
<td>0.0482</td>
<td>0.0635</td>
<td>0.0278</td>
<td>0.0544</td>
<td>0.2010</td>
<td>0.2914</td>
</tr>
<tr>
<td>Chain 2</td>
<td>0.0002</td>
<td>0.0006</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0004</td>
<td>0.0012</td>
</tr>
<tr>
<td>Chain 3</td>
<td>0.0003</td>
<td>0.0012</td>
<td>0</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0018</td>
</tr>
<tr>
<td>Chain 4</td>
<td>0.0004</td>
<td>0.0016</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0007</td>
<td>0.0024</td>
</tr>
<tr>
<td>Chain 5</td>
<td>0.0843</td>
<td>0.4031</td>
<td>0.0190</td>
<td>0.0774</td>
<td>0.1154</td>
<td>0.1500</td>
</tr>
<tr>
<td>Chain 6</td>
<td>0.1628</td>
<td>0.9247</td>
<td>0.1062</td>
<td>0.0403</td>
<td>0.2049</td>
<td>0.7743</td>
</tr>
</tbody>
</table>

TABLE 2. MCSEs

provided in Table 2 (Note that all numbers in Table 2 should be multiplied by $10^{-3}$).

From Table 2, the 95% Bayesian confidence interval (CI) for each estimate in Table 1 can be computed by (3.20), but is unnecessary to be computed here because the next subsection rejects all 6 possible models based on these estimates given in Table 1, even though the MCSEs in Table 2 must give good estimates for the CIs.

3.5. Test of Model. To test the proposed model

$$X_t = X_0 + \int_0^t v \left( X_s; \hat{\theta} \right) dW_s, t \geq 0$$

for the detrendex index $\{ \ln \hat{P}_t - \hat{\kappa}t, t = 1, \ldots, 4131 \}$, we simply follow the methods in Bibby97 [5] and compute the uniform residuals

$$U_t \left( \hat{\theta}^{(j)}_{est} \right) = P_{0^{(j)}_{est}} (X_{t+1} \leq y | X_t = x), x, y \in \mathbb{R}, j = 1, \ldots, 6, t = 1, \ldots, 4130$$
for each estimate $\hat{\theta}_{est}^{(j)}$ of $\hat{\theta}$ obtained from the 6 chains, where $P_{\hat{\theta}_{est}^{(j)}}(\cdot|\cdot)$ is exactly the conditional (transition) density in (3.5). If the models are appropriate, then $U_t\left(\hat{\theta}_{est}^{(j)}\right), t = 1, \ldots, 4130$ for each fixed $j$ should be independent and identically uniformly distributed in $(0, 1)$ (which is easily seen from the Milstein’s discretization, the property of the standard Brownian motion, and the invariance property of independence (between random vectors) under Borel transformations).

Due to the computational inability of both Mathematica and Matlab resulted from badly mixed chain 3 and chain 5 (as revealed by the sample paths given in the appendix), we have to discard models obtained from them and test only the other four models obtained from chain 1, chain 2, chain 4 and chain 6 in the simulation by the uniform residual test (see, Bibby95 [4], Bibby97 [5]). We directly test the null hypothesis: $U_t\left(\hat{\theta}_{est}^{(j)}\right), t = 1, \ldots, 4130$ for each fixed $j$ are independent and identically uniformly distributed in $(0, 1)$ \(^{10}\) (rather than testing $-2 \sum_{t=1}^{4130} \ln \left(U_t\left(\hat{\theta}_{est}^{(j)}\right)\right)$ for fixed $j$ as done in Bibby97 [5]).

Tests results for these 4 chains are summarized in Table 3, where "p-Value" denotes the significance leve at which the null hypothesis is rejected, "Chi2 Stat" the chi-squared statistic, and "DF" the degrees of freedom. Supplementary to these statistics, the quantile plot of $U_t\left(\hat{\theta}_{est}^{(j)}\right), t = 1, \ldots, 4130$ for $j = 2$ is provided in Figure 6.

These test results clearly show that all 4 possible models are rejected statistically because of the low chi-squared statistics (compared to the sample size 4131) and degrees of freedom. Thus under the scheme of approximating the unknown posterior for the parameter vector by the likelihood function obtained from the Milstein’s discretization, we reject the null hypothesis that the underlying stochastic process for the linearly detrended logarithmic SSE in the period from 01/01/1991 to 14/11/2007 is an ergodic, GH diffusion with some GH invariant density", i.e., the detrended index is not an ergodic GH diffusion.

4. CONCLUSION AND DISCUSSION

It is no doubt that the ergodic GH diffusion can not fit the detrended index since the for the ergodic GH diffusion its invariant density, the GH density, is unimodal while the empirical density of the detrended index is

\(^{10}\)This is easily done by Matlab command "chi2gof", i.e., chi-squared goodness of fit test, designed to test whether the sample are independent and identically uniformly distributed in $(0, 1)$ or not.
tri-modal. Uniform residual test results also reject the assumption that the underlying stochastic process for the detrended index is an ergodic GH diffusion since obviously the residuals from the model are not independent and identically uniformly distributed in \((0, 1)\) and the detrended index still has a strong linear trend. For more evidence on rejecting the ergodic GH diffusion, please see the appendix for other numerical results such as autocorrelations, simulated detrended index, etc..

The high acceptance rates in the first 50,000 iterations might be the result of correctly estimated initial states for the Markov chains in the simulation or due to the small tuning parameter which results in a narrow scope of sampling. Since we approximate the unknown posterior by the likelihood function obtained from the Milstein’s discretization scheme and the initial states for the Markov chains are obtained based on the empirical squared diffusion, it might be that these initial states capture the modes of the unknown posterior sufficiently well before the chains wandered away from the high acceptance region. Also, a small tuning parameter might be the cause of small MCSEs in the simulation, since the proposal states for the Markov chains in the simulation might be concentrated locally around some region of the unknown posterior.

Finally, due to the high volatility of the SSE index, it unknown to the author yet whether the posterior of the unknown parameter vector based on the transformed observation approximates the exact unknown posterior well or not.

References


5. APPENDIX

5.1. Numerical Results. We present some numerical results in the MCMC simulation. The effective sample paths are the last 10% of the first 50,000 iterations.

Figure 7. Sample Path for $\alpha$ for 6 Chains
Figure 8. Sample Path for $\beta$ for 6 Chains
Figure 9. Sample Path for $\delta$ for 6 Chains
Sample paths for $\lambda$ for 6 Chains

Figure 10. Sample Path for $\lambda$ for 6 Chains
Sample paths for $\mu$ for 6 Chains

Figure 11. Sample Path for $\mu$ for 6 Chains
Figure 12. Sample Path for $\sigma$ for 6 Chains
Figure 13. Autocorrelation for $\alpha$ for 6 Chains
Figure 14. Autocorrelation for $\beta$ for 6 Chains
Figure 15. Autocorrelation for $\delta$ for 6 Chains
Figure 16. Autocorrelation for $\lambda$ for 6 Chains
Figure 17. Autocorrelation for $\mu$ for 6 Chains
Figure 18. Autocorrelation for $\sigma$ for 6 Chains
Figure 19. Model Based Squared Volatilities
Figure 20. Simulated Detrended Index from Models 1, 2, 3
Figure 21. Simulated Detrended Index from Models 4, 5, 6
Figure 22. QQ Plot of Uniform Residuals for 4 Tested Models