Hierarchical long-memory time series models

Nan-Jung Hsu
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/12207

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700  800/521-0600
Hierarchical long-memory time series models

by

Nan-Jung Hsu

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Statistics
Major Professor: F. Jay Breidt

Iowa State University
Ames, Iowa
1997
This is to certify that the Doctoral dissertation of

Nan-Jung Hsu

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

\textit{Major Professor}

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College
# TABLE OF CONTENTS

## 1 INTRODUCTION

1.1 Long-Memory Processes .......................... 1

1.2 Estimation ......................................... 5

1.3 Model Checking ...................................... 8

1.4 Applications ....................................... 8

1.4.1 Nile River Data .................................. 8

1.4.2 Value-Weighted CRSP Stock Returns .............. 9

1.5 Dissertation Organization .......................... 9

## 2 LONG-MEMORY TIME SERIES MODELS

2.1 Fractionally Integrated ARMA Processes ............ 10

2.2 Semi-Markov Processes .............................. 11

2.2.1 Properties ...................................... 13

2.2.2 Examples ....................................... 16

2.2.3 Embedding an AR(p) Process in a Semi-Markov Process .... 20

2.3 Hierarchical Long-Memory Processes ............... 31

2.3.1 ARFIMA with Additive Noise .................... 31

2.3.2 Long-Memory Stochastic Volatility Models ........ 32

2.3.3 Random Coefficient Models with Long-Memory Coefficients .... 33

## 3 ESTIMATION

3.1 ARFIMA(0, d, 0) and Related Models ............... 36

3.1.1 Overview ...................................... 36

3.1.2 Quality of the ARMA Approximation ............ 38
LIST OF TABLES

Table 3.1 The bias and the root mean squared error of the posterior mean and also the actual coverage of the 95% credible interval based on the posterior distribution for ARFIMA(0, d, 0) process (T=100, 500 replications). The values in parentheses are the relative efficiencies of the posterior means compared to the MLEs. .............................. 52

Table 3.2 The bias and the root mean squared error of the posterior mean and also the actual coverage of the 95% credible interval based on the posterior distribution for ARFIMA(1, d, 0) process (T=100, 100 replications). The values in parentheses are the relative efficiencies of the posterior means compared to the MLEs. .............................. 53

Table 3.3 The bias and the root mean squared error of the posterior mean and also the actual coverage of the 95% credible interval based on the posterior distribution for ARFIMA(0, d, 1) process (T=100, 100 replications). The values in parentheses are the relative efficiencies of the posterior means compared to the MLEs. .............................. 54

Table 3.4 The bias and the root mean squared error of the posterior mean of d and also the actual coverage of the 95% credible interval based on the posterior distribution for the ARFIMA(0, d, 0) plus noise model with the signal-to-noise ratio $\rho$ (T=100, 100 replications). The case with $\rho = 0$ corresponds to pure ARFIMA(0, d, 0) misspecified as ARFIMA plus noise. ................................................................. 55
Table 5.1  The posterior mean and quantiles of parameters in ARFIMA(0,d,0) model for the standardized Nile River minimum water levels. The values in parentheses are the posterior standard deviations.

Table 5.2  The posterior mean and quantiles of parameters in the SV model incorporated with an ARFIMA(0,d,0) for the value-weighted CRSP market index. The values in parentheses are the posterior standard deviations.

Table 5.3  The posterior mean and quantiles of parameters in the SV model incorporated with an ARFIMA(0,d,0) plus AR(1) for the value-weighted CRSP market index. The values in parentheses are the posterior standard deviations.

Table 5.4  The posterior mean and quantiles of parameters in the SV model incorporated with a semi-Markov process for the value-weighted CRSP market index ($T = 1000$). The values in parentheses are the posterior standard deviations.
LIST OF FIGURES

Figure 2.1 A trajectory of a semi-Markov process. ........................................ 13
Figure 2.2 Given $(\xi_1, \xi_2)' \in (0,1) \times (0, 1)$ but $\xi_1 \neq \xi_2$, the value of $\lambda$ (contour) is
the mean of the exponential sojourns for which the constructed semi-
Markov process has the same autocovariances as the AR(2) with distinct
roots at all integer lags. .............................................................. 24
Figure 2.3 Given $\xi > 0$ (horizontal axis), the value of $\lambda$ (shown in log scale) is
the mean of the exponential sojourns for which the constructed semi-
Markov process has the same autocovariances as the AR(2) with equal
roots at all integer lags. .............................................................. 26
Figure 2.4 Given an $\text{AR}(2)$ with complex conjugate roots $re^{\pm i\theta}$, Region $A$ is the
subspace of the $(r, \theta)$ parameter space for which the $\text{AR}(2)$ is embed-
dable in a semi-Markov process. Region $B$ and $C$ are not embeddable. 28
Figure 2.5 Given $\tilde{\xi}_1 = re^{i\theta}$ and $(\theta, r) \in (0, \pi/2) \times (0, 1)$, the value of $\lambda$ (con-
tour) is the mean of the exponential sojourns for which the constructed
semi-Markov process has the same autocovariances as the $\text{AR}(2)$ with
complex conjugate roots at all integer lags. ........................................ 29
Figure 2.6 The shaded area is the set of all $\phi_1$ and $\phi_2$ for which the corresponding
$\text{AR}(2)$ process is embeddable in a semi-Markov process. All of the
parabola is included in the shaded region. The entire triangular area is
the region for an $\text{AR}(2)$ to be stationary. ........................................ 30
Figure 3.1 The relation between $d$ in the ARFIMA$(0,d,0)$ model and the coefficients of the approximate ARMA$(2,2)$ model satisfying $(1 + a_1B + a_2B^2)\alpha_t = (1 + b_1B + b_2B^3)\eta_{1t}$. The functions are evaluated at $d = -0.48, -0.46, \ldots, 0.46, 0.48$, and then interpolated using cubic splines. 39

Figure 3.2 The minimal orders of the truncated AR process to achieve the same Kullback-Leibler discrepancy produced by our approximate ARMA$(2,2)$ process for various $d$. 40

Figure 3.3 The difference between the log spectral densities of the approximating processes and the target ARFIMA$(0,0.4,0)$ process. 40

Figure 3.4 The log differences between the one-step prediction mean squared errors of the ARFIMA$(0, d, 0)$ model and the approximation ARMA$(2,2)$ model for $d$ equal to $0.05, 0.10, \ldots, 0.45$ and $T = 2, 3, \ldots, 1000$. 42

Figure 3.5 The values of $E(\alpha_t - \hat{\alpha}_t(1))^2/\hat{v}_t^{(1)} - 1$ under ARFIMA$(0, d, 0)$ model for $d = 0.05, 0.10, \ldots, 0.45$ and $T = 2, 3, \ldots, 1000$. 43

Figure 5.1 The yearly minimum water levels of the Nile River at the Roda Gauge for the years 622AD to 1284AD. 63

Figure 5.2 The log-log plot of the sample autocorrelations versus time lags for the Nile River data. The solid line is the least squares fit with slope $-0.533$ which is smaller than $-1$ for the short-memory processes. 64

Figure 5.3 The log-log plot of the rescaled range statistic $R/S$ versus $T$ for the Nile River data. The solid line is the least squares fit with slope 0.985 and the dotted line is the reference line with slope 0.5 for short-memory processes. 64

Figure 5.4 The log-log plot of the periodogram ordinates versus frequencies for the Nile River data. The decreasing tendency near zero frequency shows long memory. 65
Figure 5.5 The log-log plot of $\text{var}(\hat{y}_T)$ versus $T$ for the Nile River data. The solid line is the least squares fit with slope $-0.991$ and the dotted line is the reference line with slope $-1$ for short-memory processes. ............... 65

Figure 5.6 The values of $R^{1/2}$ for $d$ and $\sigma_n^2$ based on five independent Markov chains with length $2n$. ................................................. 66

Figure 5.7 The autocorrelations between draws in the Metropolis-Hastings algorithm for $d$ (denoted by o) and $\sigma_n^2$ (denoted by x). Horizontal reference lines are the Bartlett bounds $\pm 1.96T^{-1/2}$ for testing significance of estimated autocorrelations. ............................. 66

Figure 5.8 The estimated marginal posterior density of the parameter $d$ (solid line) in the ARFIMA(0,$d$,0) model for the Nile River data. The dotted line indicates the analytic result. ............................ 67

Figure 5.9 The estimated marginal posterior density of the parameter $\sigma_n^2$ (solid line) in the ARFIMA(0,$d$,0) model for the Nile River data. The dotted line indicates the analytic result. ............................ 67

Figure 5.10 The sample autocorrelations of the Nile River minimum water levels. The dotted lines and the dashed lines are the 90% and 95% pointwise prediction bands based on 2500 simulated datasets under the estimated ARFIMA(0,$d$,0) model. ................................................. 69

Figure 5.11 The kurtosis based on the Nile River data compared to the predictive distribution of those based on 2500 simulated datasets under the estimated ARFIMA(0,$d$,0) model. ................................. 69

Figure 5.12 The total of prediction mean squared errors based on the Nile River data for the years 1285AD to 1400AD compared to the predictive distribution of those based on 2500 simulated datasets under the estimated ARFIMA(0,$d$,0) model. ................................................. 70

Figure 5.13 The histogram of the relocated log importance weights $\{c_j^*\}$ for the Nile River data. ................................................................. 71
Figure 5.14  The contour plot of the joint posterior density for the Nile River data under the the ARFIMA(0,d,0) model, the approximating ARMA(2,2) model and the truncated AR(p) models.  

Figure 5.15  The daily returns data for the value-weighted CRSP market index from July 1962 to December 1987. The smoothed volatilities from the SV model incorporating an ARFIMA are superimposed on the CRSP data. The observation $y_{6355} = -0.17$, which corresponds to the October 1987 market crash, has been truncated.  

Figure 5.16  The sample autocorrelations of the log squared value-weighted CRSP data. The fitted autocorrelations (solid line) under the SV model incorporated with an ARFIMA(0,d,0) are superimposed. The dotted lines and dashed lines are the 90% and the 95% pointwise posterior prediction confidence bands, respectively.  

Figure 5.17  The densities of $\log \chi^2$ and the mixture Gaussian approximation.  

Figure 5.18  The sample autocorrelations of the log squared value-weighted CRSP data. The fitted autocorrelations (solid line) under the SV model incorporated with an ARFIMA plus AR(1) are superimposed. The dotted lines and the dashed lines are the 90% and the 95% pointwise posterior prediction confidence bands, respectively.  

Figure 5.19  The kurtosis based on the log squared value-weighted CRSP compared to the predictive distribution of those based on simulated datasets under ARFIMA(0,d,0) plus AR(1) model.  

Figure 5.20  The skewness based on the log squared value-weighted CRSP compared to the predictive distribution of those based on simulated datasets under ARFIMA(0,d,0) plus AR(1) model.  

Figure 5.21  The smoothed volatilities from the stochastic volatility model incorporating a long-memory semi-Markov process, superimposed on the value-weighted CRSP data for $T = 1000$.  

1 INTRODUCTION

1.1 Long-Memory Processes

The long-memory phenomenon, in which correlations between distant observations decay extremely slowly, has been recognized in many different areas. It was first observed by Hurst (1951) in analyzing the yearly Nile River flow minima along with many other geophysical time series. Since then, more empirical evidence was found in other fields such as hydrology (Lawrance and Kottegoda, 1977; Hipel and McLeod, 1978), meteorology (Mandelbrot and Wallis, 1968), textile engineering (Cox and Townsend, 1947), computer science (Willinger et al., 1995) and economics (Granger, 1980; Breidt et al., 1997). In addition, the phenomenon appears not only in time series data but also in higher dimensional spatial data (Whittle, 1956; Haslett and Raftery, 1989).

The memory property of a second order stationary process is characterized by its dependence structure. The duration of the second order dependence relies on the decay rate of the autocovariance function. Based on the decay behavior of the autocovariance function, the processes can be grouped into three categories: short-memory processes, intermediate-memory processes and long-memory processes. A process is said to be short memory if its autocovariance function decays exponentially and therefore the absolute values of the autocovariances are summable. In contrast, a process is said to be long memory if the autocovariance function decays hyperbolically with the rate \( h^{-\alpha} \) as the lag \( h \) goes to infinity for some \( \alpha \in (0, 1) \) and therefore the absolute values of the autocovariances are not summable. In between, an intermediate-memory process has hyperbolic decaying autocovariance function with the rate \( h^{-\alpha} \) as \( h \) goes to infinity for some \( \alpha > 1 \) but the absolute values of the autocovariances are summable. Sometimes, intermediate-memory processes are referred to the long memory cate-
Notice that the definition of long memory is asymptotic in the sense that it is only related to the asymptotic behavior of the autocovariances, but not relevant to their magnitudes at any finite lags. Based on this fact, it seems to be very difficult to detect the existence of long memory in real applications since we only have a finite number of observations. Therefore, from the application point of view, someone might question: why not just use short-memory processes for modeling if we have no information about the memory property of the underlying process? Long-memory processes are useful for three reasons. First, it is quite difficult to have a parsimonious short-memory model such that the long-memory feature can be captured with only few parameters. Second, in long-memory processes, there exists strong persistence such that the autocorrelations are not summable; therefore, the behaviors of many estimators and test statistics are quite different from the same test statistic under the usual short-memory processes. For instance, the sample mean converges to the true mean at a rate slower than \( T^{-1/2} \) which can be achieved by any short-memory process (Taqqu, 1975). As another example, the rescaled range statistic (Mandelbrot, 1975) behaves like \( T^{1/2 + \delta} \) asymptotically for some \( \delta > 0 \) in contrast to \( \delta = 0 \) in short memory cases. Due to these dramatically different asymptotic results, the model assumption (long memory or short memory) becomes crucial in order to make correct asymptotic inferences, in particular for the precision of the estimates. Third, there are many physical mechanisms that can actually generate long-memory phenomena, including thinning and pulling procedures in textile engineering (Cox, 1984), thermodynamic systems (Cassandro and Jona-Lasinio, 1978) and turbulence (Kolmogorov, 1941) in physics, diffusion processes driven by partial differential equations (Whittle, 1956, 1962) and aggregation procedures in economics (Granger, 1980). However, for most long-memory processes found in empirical studies, the background mechanisms are too complicated to understand and model. Therefore, it is necessary and important to establish long-memory models with simple structures, which can represent a variety of behaviors.

In time series analysis, the most commonly used autoregressive moving-average (ARMA) models (Box and Jenkins, 1976) are all short-memory processes. Unlike short-memory pro-
cesses, a long-memory process exhibits very strong persistence between observations and therefore it sometimes behaves more like a non-stationary process. Based on this observation, a class of continuous-time long-memory processes, called fractional Brownian motions, was proposed by Mandelbrot and van Ness (1968). However this class of models is quite restricted because it depends on only one parameter. The most commonly used class of long-memory processes is the fractionally integrated ARMA (ARFIMA) processes, first introduced by Granger and Joyeux (1980) and Hosking (1981). Unlike fractional Brownian motion, ARFIMA models can represent a variety of dependence structures, including both short memory and long memory, and therefore are more interesting in applications. Besides, seasonal effects can be incorporated into an ARFIMA process in many ways to construct even more flexible models (Porter-Hudak, 1990; Ray, 1993).

Another class of long-memory models, called generalized ARMA processes (Chung, 1996), are constructed based on the Gegenbauer polynomials (Gray, Zhang and Woodward, 1989). This class allows for both hyperbolic and sinusoidal decay in autocovariances. In addition, the corresponding spectral density could have a pole at a given frequency which indicates strong persistence in cycles. Therefore, this class is useful for modeling persistent seasonality and business cycles. In particular, an ARFIMA process is a special case in this class.

In the continuous-time context, Viano, Deniau and Oppenheim (1994) and Comte and Renault (1996) both investigated the class of long-memory models driven by stochastic differential equations. The former study considered a class of processes constructed by a particular linear filtering of Brownian motions. The latter considered a class of continuous-time fractionally integrated moving average processes with respect to Brownian motions, which is equivalent to a continuous-time moving average process with respect to a fractional Brownian motion.

Besides the linear processes described above, long-memory features can also be produced by many nonlinear mechanisms. In this dissertation, we propose a class of semi-Markov processes constructed from an underlying discrete-time Markov process and a sojourn distribution with positive support. This generalizes the model discussed by Taqqu and Levy (1986). The semi-Markov processes were first introduced by Lévy (1954) and Smith (1955) independently.
Roughly speaking, a semi-Markov process is a stochastic process which moves from one state to another state according to a Markov chain, and then remains in that state for a random length of time (the sojourn time) until the next state transition. The trajectory of a semi-Markov process is piecewise constant and continuous from the right. In general, a semi-Markov process does not satisfy the Markov property except when the sojourn times are exponentially distributed and independent of the states. Much research was done from 1950's to 1980's on this topic, but most of this research involves investigating the limiting behavior and the probabilistic properties of this class (Çinlar, 1969; Kesten, 1974; Athreya, McDonald and Ney, 1978).

Semi-Markov models have been used for applications in queuing systems and reliability problems. But there seems to be very few applications of semi-Markov processes in the time series context. In this dissertation, some properties associated with these models are discussed, including stationarity and memory properties. The semi-Markov process exhibits long memory if the distribution of the sojourn times is heavy-tailed, i.e., a Pareto-like distribution.

Because these semi-Markov processes are derived in continuous time, it is of interest to study their relationship to discrete-time models. One means of doing this is through the concept of “embedding”.

A discrete-time process is said to be embedded in a continuous-time process if the autocovariances of the two processes are equal at all integer lags. This concept is discussed most often in embedding of a discrete-time ARMA process into a continuous-time ARMA (CARMA) process. Chen and Tong (1987) concluded that a stationary AR(1) process is always embeddable in a CAR(1). The more general case for ARMA processes was partially discussed by He and Wang (1989). They claimed that a discrete-time ARMA(p, q) process with q < p, whose autoregressive polynomial has r negative real roots, is always embeddable in a CARMA(p', q') with q' < p' = p+r. This topic was investigated in detail by Brockwell (1995) especially for the case of AR(2). Brockwell also pointed out the result by He and Wang is incorrect by showing a counterexample that a discrete-time AR(2) with complex roots cannot always be embedded in a CARMA(2,1). In this situation, it can be embedded in a CARMA(4,2) instead.
We will investigate the embedding of a discrete-time AR process into a continuous-time semi-Markov process constructed by an AR and an exponential sojourn distribution, focusing on the case of discrete-time AR(1) and AR(2).

In general, the long-memory phenomenon might show up in more complicated forms or may not be observed directly. For instance, the long-memory component can be incorporated into a linear regression model (Yajima, 1989; Beran, 1993); the long-memory feature can show up in the mean structure, the seasonality or the noise disturbance. As another example, the long-memory component or its transformation might occur as unobservable states in a generalized state space model, e.g., stochastic volatility models, long memory with additive outliers, and random coefficient models. In order to describe these complicated phenomenon, we are motiviated to consider a more general class of hierarchical long-memory time series models and the estimation methods associated with these models.

1.2 Estimation

In the early stages of statistical analysis of long memory, the long-memory parameter was estimated by several heuristic methods such as the rescaled range statistic $R/S$, the log-log plot of $\text{var}(y_T)$ versus $T$ and the least squares regression in the frequency domain. These methods were all motivated by the special features associated with long-memory processes, e.g., slower convergence rate for statistics and infinite spectrum at zero frequency. Generally speaking, they are model-free estimation procedures. For parametric models, more efficient estimation methods should be considered. However, these simple procedures can still be useful diagnostic tools to check for the existence of long memory. We describe the procedures using periodograms in the following paragraph; other methods are introduced in Chapter 4 for model checking.

The semi-parametric estimation procedure proposed by Geweke and Porter-Hudak (1983) is a two-step method in which the properties of the spectral density around frequency zero are used. In the first step, the estimate of the long-memory parameter is obtained as the estimated slope in a regression of the log periodogram ordinates on the log frequencies. This estimate is
very easy to derive but varies considerably according to different choices of frequencies used in the regression. In addition, it converges to the true value at a rate slower than $T^{-1/2}$ (Robinson, 1994) and therefore the estimator is asymptotically inefficient. The rest of parameters in the model are estimated in the second step based on the transformed time series in which the long-memory component is filtered out.

The maximum likelihood estimator (MLE) has also been investigated and shown to be asymptotically efficient and normal under mild regularity conditions by Yajima (1985) and Dahlhaus (1989). However, the maximization procedure, involving the computation of a complicated autocovariance matrix and its inverse, is difficult to implement and suffers from convergence problems. The problem is particularly serious for parameters near the boundary. Sowell (1992) derived an algorithm to evaluate the exact likelihood function recursively, which is still computationally intensive and numerically unstable. An alternative to obtaining exact MLE is to maximize an approximation of the likelihood function, such as Whittle’s approximation considered by Fox and Taqqu (1986), or the approximation based on a truncated AR representation considered by both Li and McLeod (1986) and Haslett and Raftery (1989). Both approximate MLEs have the same asymptotic properties as the exact MLE under some mild conditions (Fox and Taqqu, 1986; Giraitis and Surgailis, 1990; Heyde and Gay, 1993). The MLE-type estimation procedures are currently only applicable for ARFIMA models but not for more complicated hierarchical cases.

In addition to the frequentist approach, several Bayesian approaches have been investigated (Carlin, Dempster and Jonas, 1985; Pai and Ravishanker, 1996). In Carlin, Dempster and Jonas’s paper, a linear model, containing both random components and fixed effects, was assumed for economic time series. Month-to-month and season-to-season changes were modeled as random components with known long-memory characteristics (i.e., the fractional differencing parameters are given and fixed). Under this completely specified long-memory structure, the estimation procedure follows the usual Bayesian setup. They considered and compared the same model using different values of the fractional differencing parameters. In Pai and Ravishanker’s study, ARFIMA models were considered and Bayesian analysis was made using
Markov chain Monte Carlo methods. In their algorithm, draws of the differencing parameter \( d \) were generated through a Metropolis-Hastings algorithm using a Gaussian proposal since the conditional distribution of \( d \) is nonstandard and complicated. This step is extremely time consuming since the inverse of the covariance matrix for the fractionally integrated process has to be evaluated in each Metropolis step in order to compute the acceptance probability. The algorithm works only for non-hierarchical setup.

For more complicated models, e.g. hierarchical long-memory processes, the likelihood function becomes a high dimensional integral and does not have a closed form in general. Therefore it causes major difficulties in maximum likelihood type estimation. The classical estimation procedure for this type of model is a two-step procedure. In the first step, the integral is evaluated analytically or numerically and then maximized in order to find the maximum likelihood type estimates for parameters. In the second step, the estimated parameters are treated as the true values and are plugged into the model to make further inference for the hidden long-memory process. A disadvantage of this estimation procedure is that the precision for the inferences on the hidden states in the second step are doubtful, most likely underestimated, because the variation of estimation in the first step is not taken into account. Therefore, for more complicated hierarchical models, we consider a Bayesian approach to unify the estimation problem (for the parameters) and the prediction problem (for the hidden process) at the same time.

In this dissertation, we propose a new Bayesian approach for hierarchical long-memory time series which is time-efficient, has good estimation and prediction properties and can be extended easily to a variety of interesting variations on ARFIMA processes, including ARFIMA with additive noise, ARFIMA with outliers, stochastic volatility model associated with ARFIMA and common long-memory component models. The methodology can also be extended to multivariate cases.
1.3 Model Checking

Model checking is important both before and after the analysis. In applications, the basic features of the observed data should be explored to provide modeling information. In this dissertation, long memory is the key feature to be checked before the analysis. Several graphical tools can be used for detecting long memory, including the rescaled range statistic $R/S$, $\text{var}(\tilde{y}_T)$ and the sample autocovariances and periodograms. See, for example, Beran (1994).

After Bayesian inference, the goodness of fit for the estimated model can be examined by comparing the observed data to the posterior predictive distribution of some characteristic via simulation (see Gelman et al., 1995). The checking procedure is first to simulated many replications from the assumed model. Then, given a characteristic of interest (a statistic), the observed data is compared to the posterior predictive distribution based on those simulated replications in order to check if the behavior of the observed data is consistent with the assumed model.

In this study, we consider the long-memory feature, the normality of the distribution and the prediction mean squared errors as the characteristics for detecting model failure in the diagnostic procedures.

1.4 Applications

In this dissertation, two data sets are presented for illustration. One is the famous Nile River data and the other is stock returns data.

1.4.1 Nile River Data

The initial interest in studying long-memory processes was motivated by the investigation of river flows in hydrology. Among early empirical studies, the Nile River data perhaps are the most well-known example. These data were used by Hurst when he first found evidence of long memory in 1951. The Nile River data are yearly minimum water levels as measured at the Roda Gauge near Cairo (Beran, 1994). This data set, which contains 663 observations for
the years 622AD to 1284AD, is used as the first example. The same data set is also studied by Beran (1994).

1.4.2 Value-Weighted CRSP Stock Returns

In economics, after the pioneering work by Granger (1980), long memory has drawn increasing attention and empirical evidence has been found in areas including stock returns (Lo, 1991), exchange rates (Baillie and Bollerslev, 1989; Cheung, 1993), interest rates (Shea, 1991, Crato and Rothman, 1994) and a variety of applications in macroeconomics (Diebold and Rudebusch, 1989). Recently, long-memory processes have also been considered as good descriptions for the volatilities of many financial market time series because of their strong persistence (Ding, Granger and Engle, 1993; Breidt, Crato and de Lima, 1994; Baillie, Bollerslev and Mikkelson, 1996). We consider the daily returns data (first differences of log prices) for the value-weighted Center for Research in Securities Prices (CRSP) market index as the second example to investigate their volatility behaviors.

1.5 Dissertation Organization

This dissertation is organized as follows. In Chapter 2, several different long-memory models are introduced, including ARFIMA models, long-memory processes constructed from semi-Markov processes, and three hierarchical models: ARFIMA with additive noise, long-memory stochastic volatility model and a random coefficient model with long-memory coefficients. In Chapter 3, algorithms for Bayesian estimation and prediction for long-memory time series are developed. In Chapter 4, model selection techniques and some diagnostic tools are described. In Chapter 5, two applications are given to demonstrate the methodology. Finally, conclusions and directions for further research are addressed in Chapter 6. Proofs of theorems and propositions and the extensions of the algorithm are provided in detail in the appendices. All references are listed at the end.
2 LONG-MEMORY TIME SERIES MODELS

2.1 Fractionally Integrated ARMA Processes

Definition 2.1 The process \{\alpha_t : t \in \mathbb{Z}\} is said to be an ARFIMA\((p, d, q)\) process with \(d \in (-0.5, 0.5)\) if \{\alpha_t\} is stationary and satisfies the difference equations

\[
\Phi(B)(1 - B)^d \alpha_t = \Theta(B) \eta_t,
\]

where \{\eta_t\} are independent and identically distributed (iid) as \(N(0, \sigma^2_\eta)\), \(B\) denotes the backshift operator, \(\Phi(B) = 1 - \sum_{j=1}^p \phi_j B^j\) and \(\Theta(B) = 1 - \sum_{j=1}^q \theta_j B^j\) are polynomials with no common roots and with all roots outside the unit circle.

These processes, which generalize the classic ARMA processes \((d = 0)\), were introduced independently by Granger and Joyeux (1980) and Hosking (1981). The operators \(\Phi(B)\) and \(\Theta(B)\) produce geometrically decaying autocovariance structure and characterize the short-memory behavior of the process. In contrast, the operator \((1 - B)^d\) produces hyperbolically decaying autocovariance structure and characterizes the long-memory behavior of the process. Therefore, the class of ARFIMA processes is a very rich family for modeling both short-memory and long-memory behaviors.

The autocovariance function of ARFIMA\((p, d, q)\) is complicated and cannot be expressed in closed form except for \(p = 0, q = 0\). This causes major difficulty in making inference for ARFIMA processes. For the simplest case, an ARFIMA\((0, d, 0)\) has autocovariance function given by

\[
\gamma_\alpha(0) = \sigma^2_\eta \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)}, \\
\gamma_\alpha(h) = \gamma_\alpha(0) \frac{\Gamma(h + d) \Gamma(1 - d)}{\Gamma(h - d + 1) \Gamma(d)},
\]
(e.g., Brockwell and Davis, 1991, §13.2). However, for all ARIMA processes, the spectral densities have the following simple form

\[ f_\omega(\omega) = \frac{\sigma_n^2}{2\pi} \left| \frac{\Theta(e^{i\omega})}{\Phi(e^{i\omega})} \right|^2 \left| 1 - e^{i\omega} \right|^{-2d}; \quad \omega \in [-\pi, \pi]. \]

It can be shown that for some constants \( C_1 \) and \( C_2 \), and \( d \neq 0 \),

\[ \gamma_\omega(h) \sim C_1 h^{2d-1}; \quad \text{as } h \to \infty, \]

\[ f_\omega(\omega) \sim C_2 \omega^{-2d}; \quad \text{as } \omega \to 0. \]

An ARIMA(0, d, 0) process can also be represented as an AR(\infty), which is

\[ \sum_{j=0}^{\infty} \pi_j \alpha_{t-j} = \eta; \quad \text{where } \pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}. \quad (2.2) \]

### 2.2 Semi-Markov Processes

The semi-Markov processes were first introduced by Lévy (1954) and Smith (1955) independently. We discuss a particular class of semi-Markov processes which actually are Markov processes transformed randomly in the time scale (Yackel, 1968). Before defining semi-Markov processes, we first define some key terms.

**Definition 2.2** A process \( \{y_t \in \mathbb{R}^d : t \in T\} \) is said to be a Markov process if it satisfies

\[ P[y_{t_n} \in B | y_{t_1}, \ldots, y_{t_{n-1}}] = P[y_{t_n} \in B | y_{t_{n-1}}] \]

with probability 1, for any integer \( n \geq 1 \), \( B \in \mathcal{B}(\mathbb{R}^d) \) and \( t_1 < t_2 < \cdots < t_n \) where \( t_i \in T \).

**Definition 2.3** Let \( F \) and \( G \) be cumulative distribution distribution functions (cdf's) with positive support. A sequence \( \{\tau_n\} \) is said to be a renewal sequence if \( \tau_n = \tau_{n-1} + T_n \) and \( \tau_0 = 0 \), where \( \{T_i : i \geq 2\} \) is an iid sequence of random variables with cdf \( F \) and \( T_1 \) has the cdf \( G \) and is independent of \( \{T_i : i \geq 2\} \).

**Definition 2.4** A counting process \( N(t) \) with respect to a renewal sequence \( \{\tau_n\} \) is defined as

\[ N(t) = \sum_{n=1}^{\infty} 1_{[0,t]}(\tau_n) = \max\{n \in \mathbb{Z}^+ : \tau_n \leq t\}, \quad (2.3) \]

which is the number of transitions in \([0,t]\).
The definition of semi-Markov processes is introduced as follows.

**Definition 2.5** Let \( \{w_n : n \in \mathbb{Z}^+\} \) be a Markov process and \( \{T_n : n \in \mathbb{Z}^+\} \) be a sequence of independent, nonnegative random variables such that the distribution of \( T_n \) depends only on \( w_{n-1} \). Let \( \{S_n\} \) be the partial sums of the process \( \{T_n\} \); that is, \( S_n = \sum_{i=1}^{n} T_i \). Define
\[
\alpha_t = w_n; \quad \text{for } S_n \leq t < S_{n+1}.
\]

Then the process \( \{\alpha_t : t \in \mathbb{R}^+\} \) is called a semi-Markov process with states \( \{w_n\} \) and sojourn times \( \{T_n\} \).

Since the counting process \( N(t) \) keeps constant values on the intervals \([\tau_n, \tau_{n+1})\) and is continuous from the right, the semi-Markov process \( \{\alpha_t\} \) is also constant on the intervals \([\tau_n, \tau_{n+1})\) and continuous from the right.

We consider a process \( \{\alpha_t\} \) defined by
\[
\alpha_t = w_{N(t)},
\]
where \( \{w_n\} \) is Markovian and \( N(t) \) is a counting process defined by Definitions 2.2 and 2.4. Clearly, the process \( \{\alpha_t\} \) is a special case of a semi-Markov process. A simulated trajectory of a semi-Markov process \( \{\alpha_t\} \) is presented in Figure 2.1.

**Assumption 2.1**

(A1) \( \{w_n\} \) is weakly stationary with mean zero.

(A1') \( \{w_n\} \) is strictly stationary with mean zero.

(A2) The mean of \( F, \mu = \int_0^\infty t dF(t) \), exists.

(A3) \( G(t) = \mu^{-1} \int_0^t (1 - F(x)) dx \).

(A4) Under weak stationarity, the autocovariance function \( \gamma^w(h) \) of \( \{w_n\} \) is square summable; that is, \( \sum_{h=0}^{\infty} \gamma^w_2(h) < \infty \).

**Definition 2.6** The forward recurrence time \( B(t) \) at time \( t \), which is the waiting time from time \( t \) to the next transition, is given by
\[
B(t) = \tau_{N(t) + 1} - t.
\]
Lemma 2.1 Under Assumption 2.1 (A3), $B(t) \sim G$ for all $t \geq 0$.


In fact, the distribution $G$ defined in Assumption 2.1 (A3) is the limiting distribution of $B(t)$ as $t \to \infty$. Therefore, Lemma 2.1 essentially says that the forward recurrence time $B(t)$ has the same distribution as $B(0)$ for all $t$ if and only if the process is started from its limiting distribution (i.e., $B(0) \sim G$).

2.2.1 Properties

In this section, the conditions for the process $\{\alpha_t\}$ to be weakly stationary and strictly stationary are described. Under the weak stationarity assumption, the general form of the spectral density for the semi-Markov processes is also derived. The memory properties associated with the semi-Markov processes are also discussed. All proofs are provided in Appendix A.

Theorem 2.1 (Weak Stationarity) Under (A1)-(A3), the process $\{\alpha_t\}$ is weakly stationary with $E\alpha_t = 0$ and autocovariance function $\gamma_\alpha$ defined as $\gamma_\alpha(t) = \sum_{h=0}^{\infty} \gamma_\alpha(h) P[N(t) = h]$. 

Figure 2.1 A trajectory of a semi-Markov process.
Moreover, if \( \{w_n\} \) are uncorrelated then \( \gamma_\alpha(t) = \gamma_w(0)(1 - G(t)) \).

**Theorem 2.2 (Strict Stationarity)** Under \((A1')\), \((A2)\) and \((A3)\), the process \( \{\alpha_t\} \) is strictly stationary with \( E\alpha_t = 0 \), satisfying

\[
\begin{aligned}
P[\alpha_{t_1+s} \leq a_1, \alpha_{t_2+s} \leq a_2, \cdots, \alpha_{t_k+s} \leq a_k] &= P[\alpha_{t_1} \leq a_1, \alpha_{t_2} \leq a_2, \cdots, \alpha_{t_k} \leq a_k],
\end{aligned}
\]

for arbitrary \( k \in \mathbb{Z}^+ \), \( s \in \mathbb{R}^+ \), \( \alpha = (a_1, a_2, \cdots, a_k) \in \mathbb{R}^k \) and \( t = (t_1, t_2, \cdots, t_k) \in (\mathbb{R}^+)^k \).

The spectral density of the process \( \{\alpha_t\} \) can be expressed in terms of the autocovariance function of \( \{w_n\} \) and the characteristic function of \( F \). Before introducing the theorem, we have a lemma.

**Lemma 2.2** The characteristic function for \( G \) is

\[
\phi_G(\omega) \equiv E e^{i\omega t} = \frac{1}{i\omega \mu}(\phi_F(\omega) - 1).
\]

**Theorem 2.3** Under Assumption 2.1 \((A1)-(A4)\), the semi-Markov process \( \{\alpha_t\} \) has the spectral density

\[
\begin{aligned}
f_\alpha(\omega) &= \frac{1}{2\pi i\omega} \left\{ \left( \phi_G(\omega) - \phi_G(-\omega) \right) \gamma_w(0) + \left( 1 - \phi_F(-\omega) \right) \frac{\phi_G(-\omega)}{\phi_F(-\omega)} \sum_{h=1}^{\infty} \gamma_w(h) \phi_F(-\omega)^h \\
&\quad - \left( 1 - \phi_F(\omega) \right) \frac{\phi_G(\omega)}{\phi_F(\omega)} \sum_{h=1}^{\infty} \gamma_w(h) \phi_F(\omega)^h \right\},
\end{aligned}
\]

where \( \phi_F \) and \( \phi_G \) are the characteristic functions of \( F \) and \( G \).

The spectral density of \( \{\alpha_t\} \) is a function of \( \gamma_w \) and \( \phi_F \). Although the expression of \( f_\alpha \) in Theorem 2.3 still involves summations, they can be simplified further if the underlying discrete-time process \( \{w_n\} \) is uncorrelated or has a geometrically decaying autocovariance function, e.g. a discrete time autoregressive process. The results of simplification are in the following corollaries.

**Corollary 2.1** Under Assumption 2.1 \((A1)-(A4)\), if the process \( \{w_n\} \) is uncorrelated then the spectral density of \( \{\alpha_t\} \) is

\[
f_\alpha(\omega) = \frac{\gamma_w(0)}{2\pi \mu \omega^2} \left\{ 2 - \phi_F(\omega) - \phi_F(-\omega) \right\}.
\]
Corollary 2.2 Under Assumption 2.1 (A1)-(A4), if the process \{w_n\} is an AR(1) process with \( \gamma_w(h) = \gamma_w(0)e^{\xi|h|} \), then the spectral density of \{\alpha_t\} is

\[
f_\alpha(\omega) = \frac{(1 - \xi)\gamma_w(0)}{2\pi \mu^2} \left\{ \frac{1 - \phi_F(\omega)}{1 - \xi \phi_F(\omega)} + \frac{1 - \phi_F(-\omega)}{1 - \xi \phi_F(-\omega)} \right\}.
\]

The semi-Markov models can produce both short-memory and long-memory processes. The memory characteristic is only determined by the tail behavior of the sojourn distribution \( F \). If \( F \) is heavy-tailed then the corresponding semi-Markov process \( \{\alpha_t\} \) is long memory, otherwise \( \{\alpha_t\} \) is short memory. The results are described in the following theorems after giving some related definitions.

Definition 2.7 A function \( L(x) \) is said to be slowly varying at \( \infty \) if it satisfies \( L(cx)/L(x) \to 1 \) for large \( x \) and arbitrary \( c \).

Definition 2.8 A function \( U(x) \) is said to be regularly varying at \( \infty \) with exponent \( \beta \in \mathbb{R} \), if it satisfies \( U(x) = x^\beta L(x) \) where \( L(x) \) is slowly varying at \( \infty \).

Definition 2.9 A distribution \( F \) is said to be heavy-tailed with index \( \beta \), \( 1 < \beta < 2 \), if \( F \) satisfies

\[1 - F(y) = y^{-\beta} L(y) \text{ for large } y, \text{ where } L \text{ is slowly varying at } \infty.\]

Theorem 2.4 Assume \( F(\cdot) \) is a heavy-tailed distribution with index \( \beta \). If \( \{w_n\} \) are uncorrelated, then \( \gamma_\alpha(t) \sim t^{1-\beta} L_0(t) \) for large \( t \), where \( L_0 \) is slowly varying at \( \infty \). Otherwise, if \( \{w_n\} \) are correlated, the result still holds provided \( \sum_{h=0}^{\infty} \gamma_w(h)L_h(t) \) is slowly varying at \( \infty \), where \( L_h(\cdot) \) is the corresponding slowly varying function satisfying \( P[N(t) = h] \sim t^{1-\beta} L_h(t) \).

Equivalently, the tail behavior can be defined by using the density function of the cdf \( F \). Then the autocorrelation function and the spectral density of \( \{\alpha_t\} \) have the following asymptotic expressions.

Theorem 2.5 Let \( f(x) \) be the density function of \( F \) and \( \beta \in (1, 2) \). Assume \( f(x) \sim x^{-\beta-1} \) as \( x \to \infty \). Then

\[\begin{align*}
g_\alpha(t) &\sim t^{1-\beta}, \text{ as } t \to \infty, \quad (2.5) \\
f_\alpha(\omega) &\sim \omega^{\beta-2}, \text{ as } \omega \to 0. \quad (2.6)
\end{align*}\]
Equations (2.5) and (2.6) satisfy the definitions of long memory, and the range $\beta \in (1, 2)$ covers the entire space of all possible hyperbolically decaying rates.

### 2.2.2 Examples

Three examples are given in this section, including two examples of short-memory processes and one example of a long-memory process.

**Example 2.1 Stationary Process States with Degenerate Sojourns**

Assume $F(t) = 1_{t \geq c}$ so that $T_i = c$ for all $i \geq 2$. By Assumption 2.1 (A3) and Lemma 2.2, we have

\[
F_0(t) = \frac{1}{c} \int_0^t \{1 - F(x)\} dx = \begin{cases} t/c; & \text{if } t \leq c, \\ 1; & \text{otherwise}, \end{cases}
\]

\[
\phi_P(\omega) = e^{i\omega c}, \quad \phi_G(\omega) = \frac{1}{i\omega} \left( e^{i\omega} - 1 \right).
\]

Let $m \in \mathbb{Z}^+$ and consider $t \in [(m - 1)c, mc]$. Then the autocovariance function of $\{\alpha_t\}$ is

\[
\gamma_\alpha(t) = \sum_{h=0}^{\infty} \gamma_w(h) P[N(t) = h] = \sum_{h=0}^{\infty} \gamma_w(h) (P[r_{h+1} > t] - P[r_h > t]) = \left\{ \gamma_w(m-1) P[r_m > t] + \sum_{h=m}^{\infty} \gamma_w(h) \right\} - \left\{ \gamma_w(m) P[r_m > t] + \sum_{h=m+1}^{\infty} \gamma_w(h) \right\} = P[r_m > t] \gamma_w(m-1) + P[r_m \leq t] \gamma_w(m), = \gamma_w(m-1) + \{\gamma_w(m) - \gamma_w(m-1)\} \{t/c - (m - 1)\},
\]

which is the linear interpolation of $\gamma_w(h)$ with time unit $c$. When $c = 1$, the two autocovariance functions $\gamma_\alpha$ and $\gamma_w$ are matched at every integer lag, which means any given discrete-time Markov process can be embedded into a semi-Markov process by choosing $c = 1$. The corresponding spectral density $f_\alpha$ also has a simple expression in terms of the spectral density $f_w$, which is described in the following proposition.
Proposition 2.1

\[ f_\alpha(\omega) = \frac{\cos^2(\omega/2)}{(\omega/2)^2} f_w(\omega). \]

Since \( \sin z/z \to 1 \) as \( z \to 0 \), the function \( f_\alpha \) has similar behavior at the origin as the function \( f_w \) and is equal to zero at frequencies \( \omega = 2\pi k/c \) for all nonzero integers \( k \). The discretized process \( \{\alpha_{t_0+ck} : k \in \mathbb{N}\} \) from \( \{\alpha_t\} \) is reduced to the underlying \( \{w_n\} \) process for arbitrary choice of \( t_0 \).

Example 2.2 \( AR(p) \) Process States with Exponential Sojourns

Assume the sojourn times follow an exponential distribution, that is \( F(t) = 1 - \exp(-t/\lambda) \) with \( \lambda > 0 \). Then, the corresponding \( \{\tau_n\} \) process becomes a Poisson process and \( G = F \) because of the memoryless property of the exponential distribution. In particular, the process \( \{\alpha_t\} \) itself is also Markovian in this case. The characteristic functions are \( \phi_G(\omega) = \phi_F(\omega) = (1 - \lambda \omega i)^{-1} \). Further assume that \( \{w_n\} \) is an \( AR(p) \) process with

\[ \phi(B) = \prod_{j=1}^{p} (1 - \xi_j B); \quad \text{where } |\xi_j| < 1 \text{ for all } j. \]

Assume \( \{\xi_j, j = 1, 2, \cdots, p\} \) has \( p_r \) distinct real values and \( p_c \) distinct pairs of complex conjugates. Let \( \Gamma_r \equiv \{\xi_{rj}, j = 1, 2, \cdots, p_r\} \) be the set of the distinct real \( \xi_j \)'s, \( \Gamma_c \equiv \{\xi_{cj} = r_j \exp(i\theta_j), j = 1, 2, \cdots, p_c\} \) be the set of the distinct complex conjugates and \( (1 + \kappa_{ij}) \) be the multiplicity of \( \xi_{ij} \) for \( i = r, c \) (e.g., \( \kappa_{ij} = 0 \) for an unrepeated \( \xi_{ij} \); \( \kappa_{ij} = 1 \) for \( \xi_{ij} \) repeated once).

The values of \( \kappa_{ij} \)'s satisfy \( \sum_{j=1}^{p_r} (\kappa_{rj} + 1) + 2 \sum_{j=1}^{p_c} (\kappa_{cj} + 1) = p \). Then the autocovariance function of \( \{w_n\} \) can be decomposed into two components with respect to \( \Gamma_r \) and \( \Gamma_c \), that is \( \gamma_w(h) = \gamma_{wr}(h) + \gamma_{wc}(h) \), where

\[ \gamma_{wr}(h) = \sum_{j=1}^{p_r} r_j^h \left( \sum_{k=0}^{\kappa_{rj}} c_{jk} h^k \right), \quad (2.7) \]

\[ \gamma_{wc}(h) = \sum_{j=1}^{p_c} r_j^h \left( \sum_{k=0}^{\kappa_{cj}} d_{jk} \cos(h\theta_j + \beta_{jk}) h^k \right), \quad (2.8) \]

and the constants \( c_{jk}, d_{jk} \) and \( \beta_{jk} \) are functions of \( \xi_j \)'s (e.g., Brockwell and Davis, 1991). The autocovariance function of the constructed \( \{\alpha_t\} \) can also be decomposed into two parts with
respect to $\Gamma_r$ and $\Gamma_c$, that is $\gamma(t) = \gamma_{or}(t) + \gamma_{ac}(t)$ where

$$\gamma_{or}(t) = \sum_{h=0}^{\infty} \gamma_{ur}(h) P[N(t) = h], \quad (2.9)$$

$$\gamma_{ac}(t) = \sum_{h=0}^{\infty} \gamma_{wc}(h) P[N(t) = h]. \quad (2.10)$$

The following lemma allows further simplification of these expressions.

**Lemma 2.3** Given a constant vector $c = (c_0, c_1, \cdots, c_\kappa)'$, there exists $c^* = (c_0^*, c_1^*, \cdots, c_\kappa^*)'$, a linear transformation of $c$, satisfying $c^* = (A_{\kappa}^{-1})^T c$, such that

$$\sum_{k=0}^{\kappa} c_k h^k = \sum_{k=0}^{\kappa} c_k^* \frac{h!}{(h-k)!},$$

where $A_{\kappa} = (a_0, a_1, \cdots, a_\kappa)'$ is a $(\kappa + 1) \times (\kappa + 1)$ matrix defined by

$$a_0 = (1, 0, 0, \cdots, 0)',$$

$$a_i = (1, -1, a_{i-1}, \cdots, 0)', \quad \text{for } i = 1, 2, \cdots, \kappa,$$

and $B^{-1}$ is the forward shift operator which shifts the $i$-th element to the $(i+1)$-th element.

Applying Lemma 2.3, (2.9) and (2.10) can be expressed as the forms in the following theorem.

**Proposition 2.2**

$$\gamma_{or}(t) = \sum_{j=1}^{p_r} \xi^{*}_{rj} t \left( \sum_{k=0}^{r_{rj}} c_{jk}^{*} t^k \right),$$

$$\gamma_{ac}(t) = \sum_{j=1}^{p_c} r^{*j} t \left\{ \sum_{k=0}^{r_{cj}} d_{jk}^{*} \cos(t\theta_j^{*} + \beta_j^{*} t^k) \right\},$$

where

$$\xi^{*}_{rj} = \exp\{- (1 - \xi_{rj})/\lambda\},$$

$$c_{jk}^{*} = c_{jk}^{*} (\xi_{rj}/\lambda)^k,$$

$$r^{*j} = \exp\{- (1 - r_j \cos \theta_j)/\lambda\},$$

$$\theta_j^{*} = (r_j \sin \theta_j)/\lambda,$$

$$\beta_j^{*} = \theta_j + \tan^{-1}(d_{jk2}/d_{jk1}),$$

$$d_{jk}^{*} = (r_j/\lambda)^k \sqrt{d_{jk1}^2 + d_{jk2}^2},$$
in which

\[ \begin{align*}
  d_{jk1} &= d_{jk} \cos \beta_{jk}, \\
  d_{jk2} &= d_{jk} \sin \beta_{jk}, \\
  (c_{j0}, c_{j1}, \ldots, c_{j\kappa_j}) &= (c_{j0}, c_{j2}, \ldots, c_{j\kappa_j}) A_{rj}^{-1}, \\
  (d_{j0i}, d_{j2i}, \ldots, d_{j\kappa_{ji}}) &= (d_{j0i}, d_{j2i}, \ldots, d_{j\kappa_{ji}}) A_{cj}^{-1}; \text{ for } i = 1, 2,
\end{align*} \]

\[ A_{rj} \equiv A_{\kappa_j} \text{ and } A_{cj} \equiv A_{\kappa_c} \text{ are } (\kappa_{rj} + 1) \times (\kappa_{rj} + 1) \text{ and } (\kappa_{cj} + 1) \times (\kappa_{cj} + 1) \text{ matrices defined in Lemma 2.3.} \]

Notice that \( \gamma_{or} \text{ and } \gamma_{ac} \) have exactly the same forms as \( \gamma_w \text{ and } \gamma_{wc} \) respectively (see Proposition 2.2 and Equation (2.7) and (2.8)) but with different coefficients. Therefore, we can expect that the behaviors of \( \gamma_o \text{ and } \gamma_w \) are quite similar because of the same building units.

Analyzing in a similar way, the spectral density \( f_o \) can also be decomposed with respect to \( \Gamma_r \) and \( \Gamma_c \) as \( f_o(\omega) = f_{or}(\omega) + f_{ac}(\omega) \), where \( f_{or} \text{ and } f_{ac} \) can be calculated from \( \gamma_{or} \text{ and } \gamma_{ac} \), respectively. It turns out that \( f_{or} \text{ and } f_{ac} \) are the weighted averages of several spectral densities \( f_{rj} \)'s and \( f_{cj} \)'s, each of which is the spectral density of a CARMA process.

**Example 2.3 AR(1) Process States with Pareto Sojourns**

Assume \( F \) is a Pareto distribution with the density function

\[ f(x) = \frac{\beta x_0^\beta}{x^{\beta+1}} \mathbf{1}(x \geq x_0); \quad \beta \in (1, 2). \tag{2.11} \]

The value of \( \beta \) characterizes the tail behavior of the sojourn distribution. If \( \beta < 2 \), the distribution is heavy-tailed, but the mean of the distribution \( F \) does not exist if \( \beta \leq 1 \). Therefore, we only consider the case for \( 1 < \beta < 2 \) for modeling long-memory processes. By definition of \( G \) and Lemma 2.1, the distribution of the forward recurrence time, \( G \), can be obtained as follows.

\[ 1 - F(t) = \int_t^\infty f(x)dx \]

\[ = \begin{cases} 
  (x_0/t)^\beta; & \text{if } t \geq x_0, \\
  1; & \text{otherwise},
\end{cases} \]

\[ \mu = \int_{x_0}^\infty xf(x)dx = \frac{\beta x_0}{\beta - 1}. \]
Therefore,

\[
G(t) = \mu^{-1} \int_0^t \{1 - F(x)\} dx
\]

\[
= \begin{cases} 
1 - \beta^{-1}(x_0/t)^{\beta-1}; & \text{if } t \geq x_0, \\
(\beta x_0)^{-1}(\beta - 1)t; & \text{otherwise.}
\end{cases}
\]  

(2.12)

For the most simple case, \(\{w_n\}\) is assumed to be a sequence of iid random variables. Then the correlation function of \(\{\alpha_t\}\), \(\rho_\alpha(t)\), is exactly equal to \(1 - G(t)\) which decays linearly when \(t \leq x_0\) and then decays hyperbolically with the rate \(\beta - 1\). For more general cases, we need to calculate the characteristic function \(\phi_F\),

\[
\phi_F(\omega) = \int_{x_0}^{\infty} \frac{\beta x_0^\beta}{x^{\beta+1}} e^{i\omega x} dx
\]

\[
= \beta x_0^\beta \left\{ \int_{x_0}^{\infty} x^{-(\beta+1)} \cos \omega x dx + i \int_{x_0}^{\infty} x^{-(\beta+1)} \sin \omega x dx \right\}.
\]

This can be simplified by the following lemma.

**Lemma 2.4** Let \(\Gamma(u, x)\) denote the incomplete gamma function, which is defined as the integral \(\int_x^{\infty} t^{u-1} e^{-t} dt\). Then

(i) \(\int x^{u-1} \sin ax dx = -\frac{1}{2a^u} \left\{ \exp \left( \frac{\pi i}{2} (u - 1) \right) \Gamma(u, -iax) + \exp \left( \frac{\pi i}{2} (1 - u) \right) \Gamma(u, iax) \right\},\)

(ii) \(\int x^{u-1} \cos ax dx = -\frac{1}{2a^u} \left\{ \exp \left( \frac{\pi i}{2} u \right) \Gamma(u, -iax) + \exp \left( -\frac{\pi i}{2} u \right) \Gamma(u, iax) \right\}.\)

By using the lemma, we have

\[
\phi_F(\omega) = \beta(-ia\omega)^\beta \left\{ \Gamma(-\beta, -i\omega x) \right\}_{x=x_0}^{x=\infty}.
\]

The density function (2.11) satisfies the conditions in Theorem 2.5, therefore the semi-Markov processes under this construction are long memory for \(\beta \in (1, 2)\).

**2.2.3 Embedding an AR(p) Process in a Semi-Markov Process**

From Example 2.2 in the previous section, we observe the relation between the class of CARMA processes and the class of semi-Markov processes constructed by an AR(p) with an exponential distribution for sojourns. In this section, the embedding relation between the
constructed semi-Markov processes and the underlying AR(p) processes is investigated for two special cases, $p = 1$ and $p = 2$. The relation between the constructed semi-Markov processes and CARMA processes is also discussed. The derivations of all the results are provided in Appendix A. Before the discussion, the formal definition of CARMA processes is introduced.

**Definition 2.10** A process $\{y_t : t \in \mathbb{R}\}$ is said to be a CARMA$(p,q)$ process (with $0 < q < p$) if it is a stationary solution of the stochastic differential equation

$$a(D)y_t = \sigma b(D)DW_t,$$

where $D$ is the differential operator with respect to $t$, $W_t$ is the standard Brownian motion, $\sigma$ is a constant, and $a(\cdot)$ and $b(\cdot)$ are polynomials with no common roots, which have the form

$$a(\lambda) = \prod_{j=1}^{p}(\lambda - a_j),$$

$$b(\lambda) = \prod_{j=1}^{q}(\lambda - b_j),$$

in which all roots have negative real parts.

In general, the analytic solution of $y_t$ is very difficult to obtain explicitly except for the most simple cases ($p \leq 1, q = 0$), and so is its autocovariance function, but its spectral density has a simple expression:

$$f_y(\omega) = \frac{\sigma^2 |b(i\omega)|^2}{2\pi |a(i\omega)|^2}; \quad \omega \in \mathbb{R}. \quad (2.13)$$

The spectral density (2.13) is bounded at the origin; therefore, all CARMA$(p,q)$ processes are short-memory processes.

The case of AR(1) is described first. Assume the underlying process $\{w_n\}$ is an AR(1) with mean zero and variance one satisfying

$$(1 - \xi_1 B)w_n = \eta_n,$$

$|\xi_1| < 1$, and the sojourn times follow an exponential distribution with mean $\lambda > 0$ (denoted by Exp($\lambda$)). From Theorem 2.1 and Corollary 2.2, the autocovariance function and the spectral
The spectral density $f_\alpha(\omega)$ satisfies (2.13) with $p = 1$, $q = 0$ and $a_1 = (\xi_1 - 1)/\lambda < 0$ since $|\xi_1| < 1$. Therefore, under this construction, the constructed semi-Markov process has the same autocovariance structure as a CARMA(1,0). In fact, the class of semi-Markov processes under this construction and the class of CARMA(1,0) are equivalent in representing the same second order structures of stationary processes. That is, given a CARMA(1,0), we can always find a corresponding semi-Markov process under this construction with the same autocovariance structure, and vice versa. Moreover, $\gamma_{\bar{\alpha}}$ and $\gamma_w$ can be made identical at all integer lags if $0 < \xi_1 < 1$. That is, setting $\xi_1 = \exp\{-1 - \xi_1\}/\lambda$, the resulting solution is $\lambda = (\xi_1 - 1)/\log x_1$ which is always positive. Therefore, we have the following propositions.

**Proposition 2.3** The class of semi-Markov processes constructed from $AR(1)$ states and exponential sojourns and the class of CARMA(1,0) processes represent the same autocovariance structures.

**Proposition 2.4 (Embedding)** Given $\xi_1 \in (0, 1)$, an AR(1) can be embedded into a semi-Markov process such that the two processes have equal autocovariances at all integer lags by choosing $\lambda = (\xi_1 - 1)/\log x_1$ as the mean for the exponential sojourns.

The case of $AR(2)$ satisfying

$$(1 - \xi_1 B)(1 - \xi_2 B)w_n = \eta_n,$$

with mean zero and variance one, is introduced for three situations—distinct real roots, equal real roots and complex conjugate roots. These three situations are discussed separately.

**Distinct Real Roots**

Assume that the semi-Markov process $\{\alpha_t\}$ is constructed by an AR(2) $\{w_n\}$ with $|\xi_1| < 1$, $|\xi_2| < 1$ and $\xi_1 \neq \xi_2$ for states and $Exp(\lambda)$ sojourns. Then, from Theorem 2.1 and Theorem 2.3,
the autocovariance function and the spectral density of \( \{ \alpha_t \} \) satisfy

\[
\gamma_\alpha(t) = c_1 \exp \left\{ - \left( \frac{1 - \xi_1}{\lambda} \right) |t| \right\} + c_2 \exp \left\{ - \left( \frac{1 - \xi_2}{\lambda} \right) |t| \right\},
\]

\[
f_\alpha(\omega) = \frac{(1 - \xi_1)(1 - \xi_2)}{\pi \lambda (1 + \xi_1 \xi_2)} \left| \frac{i \omega - b_1}{|i \omega - a_1| (i \omega - a_2)|^2} \right|,\]

where

\[
c_1 = \frac{\xi_1 (1 - \xi_2)}{(\xi_1 - \xi_2) (1 + \xi_1 \xi_2)},
\]

\[
c_2 = 1 - c_1,
\]

\[
a_1 = \frac{-1 - \xi_1}{\lambda}, \quad a_2 = \frac{-1 - \xi_2}{\lambda},
\]

\[
b_1 = -\lambda^{-1} \sqrt{1 - \xi_1 \xi_2 (\xi_1 + \xi_2 - 1)}.
\]

The autocovariance function \( \gamma_\alpha \) can be written as

\[
\gamma_\alpha(t) = c_1 \xi_1^{|t|} + c_2 \xi_2^{|t|}
\]

\[
= \frac{c_1}{\xi_1} \left( \hat{\xi}_1 \xi_1^{|t|} + \hat{\xi}_2 \xi_2^{|t|} \right) + \left( 1 - c_1 \frac{c_1 \hat{\xi}_2}{\xi_1} \right) \xi_2^{|t|}
\]

\[
= \frac{c_1}{\xi_1} \left( \hat{\xi}_1 \xi_1^{|t|} + \hat{\xi}_2 \xi_2^{|t|} \right) + \left( 1 - \frac{c_1}{\xi_1} \right) \xi_2^{|t|}, \tag{2.14}
\]

where

\[
\bar{\xi}_i = \exp \{- (1 - \xi_i)/\lambda \}; \quad i = 1, 2,
\]

\[
\hat{\xi}_1 = \frac{\xi_1 (1 - \xi_2)}{(\xi_1 - \xi_2) (1 + \xi_1 \xi_2)},
\]

\[
\hat{\xi}_2 = 1 - \hat{\xi}_1.
\]

Equation (2.14) can be reduced to the autocovariance function of a discrete-time AR(2) with positive distinct \( \bar{\xi}_1 \) and \( \bar{\xi}_2 \) if and only if \( c_1 = \hat{\xi}_1 \). Therefore, an AR(2) with distinct real roots is embeddable in a semi-Markov process if there exists \( 0 < \lambda < \min \{ -2/\log \bar{\xi}_1, -2/\log \bar{\xi}_2 \} \) satisfying \( c_1 = \hat{\xi}_1 \). The solutions exist for all AR(2) processes with positive roots, i.e., \( 0 < \xi_1 < 1 \) and \( 0 < \xi_2 < 1 \). The values of \( \lambda \) over the entire parameter space \( \{ (\tilde{\xi}_1, \tilde{\xi}_2)' \in (0, 1) \times (0, 1) : \tilde{\xi}_1 \neq \tilde{\xi}_2 \} \) are displayed by a contour plot in Figure 2.2.

**Proposition 2.5** *(Embedding)* An AR(2) with distinct real roots satisfying \( (\xi_1, \xi_2) \in (0, 1) \times (0, 1) \) is always embeddable into a semi-Markov process. The corresponding semi-Markov pro-
24

Figure 2.2 Given $(\xi_1, \xi_2)' \in (0, 1) \times (0, 1)$ but $\xi_1 \neq \xi_2$, the value of $\lambda$ (contour) is the mean of the exponential sojourns for which the constructed semi-Markov process has the same autocovariances as the AR(2) with distinct roots at all integer lags.

A process can be constructed by choosing $\xi_i = 1 + \lambda \log \xi_i$, $i = 1, 2$ for the AR(2) states, and $\lambda$ satisfying $c_1 = \tilde{c}_1$ as the mean for the exponential sojourns.

From $f_\alpha$, since $a_1$, $a_2$ and $b_1$ are all negative, therefore the corresponding semi-Markov process $\{\alpha_t\}$ has spectral density exactly identical to a CARMA(2, 1) process (see Equation (2.13)). In addition, the constructed semi-Markov processes can only represent part of the second order structures present in CARMA(2, 1) with $a_1 \neq a_2$. The subspace of CARMA(2, 1) which can be represented is given in Proposition 2.6.

**Proposition 2.6** Given a CARMA(2, 1) with $a_1 \neq a_2$, there exists a semi-Markov process constructed from an AR(2) with distinct real roots and $\text{Exp}(\lambda)$ sojourns such that two continuous-time processes have the same second order structure if and only if $(a_1, a_2)'$ is in one of the following regions: $S_0 \cap S_j$, $j = 1, 2, 3, 4$ where

$$S_0 = \{(a_1 + a_2)^2 + a_1 a_2 + b_1^2 - 2(2a_1 a_2)^{1/2}\lvert a_1 + a_2\rvert \geq 0\},$$
\[ S_1 = \{ b_1^2 < 3a_2 + a_1a_2 - a_1, \quad b_1^2 < 3a_1 + a_1a_2 - a_2 \}, \]
\[ S_2 = \{ b_1^2 < 3a_2 + a_1a_2 - a_1, \quad b_1^2 \geq 3a_1 + a_1a_2 - a_2, \quad a_1 < b_1 \}, \]
\[ S_3 = \{ b_1^2 \geq 3a_2 + a_1a_2 - a_1, \quad b_1^2 < 3a_1 + a_1a_2 - a_2, \quad a_2 < b_1 \}, \]
\[ S_4 = \{ b_1^2 \geq 3a_2 + a_1a_2 - a_1, \quad b_1^2 \geq 3a_1 + a_1a_2 - a_2, \quad a_1 < b_1, \quad a_2 < b_1 \}. \]

**Equal Real Roots**

Assume that the semi-Markov process \( \{ \alpha_t \} \) is constructed by an AR(2) \( \{ w_n \} \) with \( \xi_1 = \xi_2 = \xi \) for states and \( \text{Exp}(\lambda) \) sojourns. Then, from Theorem 2.1 and Theorem 2.3, the autocovariance function and the spectral density of \( \{ \alpha_t \} \) satisfy

\[
\gamma_\alpha(t) = \left\{ 1 + \xi \frac{(1 - \xi^2)}{\lambda(1 + \xi^2)} |t| \right\} \exp \left\{ -\left( \frac{1 - \xi}{\lambda} \right) |t| \right\},
\]
\[
f_\alpha(\omega) = \frac{(1 - \xi)^2 |\omega + \lambda^{-1} \sqrt{(1 + \xi + 2\xi^2)(1 - \xi)}|^2}{\pi \lambda (1 + \xi^2) |\omega + \lambda^{-1} (1 - \xi)|^4}.
\]

The autocovariance function can be written as

\[
\gamma_\alpha(t) = (1 + c |t|) \tilde{\xi}^{|t|}
\]
\[
= (1 + \tilde{c} |t|) \tilde{\xi}^{|t|} + (c - \tilde{c}) |t| \tilde{\xi}^{|t|}, \tag{2.15}
\]

where

\[
\tilde{\xi} = \exp\{-(1 - \xi)/\lambda\},
\]
\[
c = \frac{1 - \xi^2}{1 + \xi^2}, \quad \tilde{c} = \frac{\tilde{\xi}(1 - \xi^2)}{\lambda(1 + \xi^2)}.
\]

Equation (2.15) can be reduced to the autocovariance function of a discrete-time AR(2) with positive equal roots, i.e., \( 0 < \tilde{\xi} < 1 \), if and only if \( c = \tilde{c} \). Therefore, an AR(2) with equal real roots is embeddable in a semi-Markov process if there exists \( 0 < \lambda < -2/\log \tilde{\xi} \) satisfying \( c = \tilde{c} \).

It turns out the solution exists for all AR(2) process with equal roots. In fact, for each process, there are two solutions of \( \lambda \) satisfying \( c = \tilde{c} \). The values of \( \lambda \) for \( \tilde{\xi} \in (0, 1) \) are displayed on a log scale in Figure 2.3.

**Proposition 2.7** (Embedding) An AR(2) with equal roots and \( \tilde{\xi} \in (0, 1) \) is always embeddable into a semi-Markov process. The corresponding semi-Markov process can be constructed by
Figure 2.3 Given $\xi > 0$ (horizontal axis), the value of $\lambda$ (shown in log scale) is the mean of the exponential sojourns for which the constructed semi-Markov process has the same autocovariances as the AR(2) with equal roots at all integer lags.

choosing $\xi = 1 + \lambda \log \hat{\xi}$ for the AR(2) states, and $\lambda$ satisfying $c = \hat{c}$ as the mean for the exponential sojourns.

The function $f_a$ is also the spectral density of a CARMA(2,1) (see Equation (2.13)) with

$$a_1 = a_2 = -(1 - \xi)/\lambda,$$
$$b_1 = -\lambda^{-1} \sqrt{(1 + \xi + 2\xi^2)(1 - \xi)}.$$

Therefore, we have the following proposition.

**Proposition 2.8** Given a CARMA(2,1) with $a_1 = a_2$, there exists a semi-Markov process constructed from an AR(2) with equal roots such that these two continuous-time processes have the same second order structure if and only if $b_1 \leq (4\sqrt{2} - 5)^{1/2}a_1$.

**Complex Conjugate Roots**

Assume that the semi-Markov process $\{\alpha_t\}$ is constructed by an AR(2) $\{w_n\}$ with conjugate roots, that is $\xi_2 = \bar{\xi}_1$ for the states and $\text{Exp}(\lambda)$ sojourns. Then, from Theorem 2.1 and
Theorem 2.3, the autocovariance function and the spectral density of \( \{\alpha_t\} \) satisfy

\[
\gamma_\alpha(t) = \bar{c} \exp \left\{ -\left( \frac{1 - \xi_1}{\lambda} \right) |t| \right\} + c \exp \left\{ -\left( \frac{1 - \bar{\xi}_1}{\lambda} \right) |t| \right\},
\]

\[
f_\alpha(\omega) = \frac{\bar{c}(1 - \xi_1) + c(1 - \xi_1)}{\pi \lambda} \frac{\omega^2 + \lambda^{-2} |1 - \xi_1|^2}{\left( \omega^2 + \left( \frac{1 - \xi_1}{\lambda} \right)^2 \right) \left( \omega^2 + \left( \frac{1 - \bar{\xi}_1}{\lambda} \right)^2 \right)},
\]

where

\[
c = \frac{1}{2} \left\{ 1 - \frac{(\xi_1 + \bar{\xi}_1)(1 - \xi_1 \bar{\xi}_1)}{(\xi_1 - \bar{\xi}_1)(1 + \xi_1 \bar{\xi}_1)} \right\},
\]

and \( \bar{c} \) is the complex conjugate of \( c \). The autocovariance function \( \gamma_\alpha \) can be written as

\[
\gamma_\alpha(t) = \bar{c} \xi_1^{|t|} + c \bar{\xi}_1^{|t|} = \bar{c} \xi_1^{|t|} \left( \bar{\xi}_1 \xi_1 + \bar{\xi}_1 \bar{\xi}_1 \right) + \left( c - \frac{\bar{c} c}{\bar{c}} \right) \bar{\xi}_1^{|t|},
\]

(2.16)

where

\[
\xi_1 = \exp \left\{ -(1 - \xi_1) / \lambda \right\},
\]

\[
\bar{c} = \frac{1}{2} \left\{ 1 - \frac{(\xi_1 + \bar{\xi}_1)(1 - \bar{\xi}_1 \xi_1)}{(\xi_1 - \bar{\xi}_1)(1 + \bar{\xi}_1 \bar{\xi}_1)} \right\},
\]

and \( \bar{\xi}_1 \) and \( \bar{c} \) are the complex conjugates of \( \xi_1 \) and \( c \), respectively. Equation (2.16) can be reduced to the autocovariance function of a discrete-time AR(2) with positive complex conjugate roots if and only if \( \bar{c} \xi_1 = \bar{c} \bar{\xi}_1 \). This is equivalent to \( \bar{c} \xi_1 \in \mathbb{R} \), that is \( \bar{c} \) has to be a real multiple of \( c \). Since \( \bar{c} \) and \( c \) have the same real part, equal to 1/2, therefore, \( \bar{c} = c \) is required for an AR(2) with complex conjugate roots to be embeddable in a semi-Markov process.

It turns out the solutions only exist for some of the AR(2) processes with complex conjugate roots. The embeddable region is shown in Figure 2.4. For an AR(2) with \( \xi_1 = re^{i\theta} \), the region \( A \) in Figure 2.4 is embeddable in a semi-Markov process constructed by a stationary \( \{w_n\} \); the region \( B \) in Figure 2.4 is not embeddable in a semi-Markov process since the resulting \( \xi_1 \) for \( \{w_n\} \) is outside the stationary region, i.e., \( |\xi_1| > 1 \); the region \( C \) in Figure 2.4 is also not embeddable since the solution of \( \lambda \in \mathbb{R} \) does not exist. The values of \( \lambda \) in the region \( (\theta, r)' \in (0, \pi/2) \times (0, 1) \) are calculated and displayed by a contour plot in Figure 2.5.
Figure 2.4 Given an AR(2) with complex conjugate roots $re^{\pm i\theta}$, Region A is the subspace of the $(r, \theta)$ parameter space for which the AR(2) is embeddable in a semi-Markov process. Region B and C are not embeddable.

**Proposition 2.9** (Embedding) An AR(2) with complex conjugate roots, and $\xi_1 = re^{i\theta}$ and $\xi_2 = \xi_1$ satisfying $r < 1$ and $\theta \in (0, \pi/2)$, is always embeddable into a semi-Markov process. The corresponding semi-Markov process can be constructed by choosing $\xi_1 = 1 + \lambda \log \xi_1$ and $\xi_2 = \xi_1$ for the AR(2) states, and $\lambda$ satisfying $\hat{c} = \hat{c}$ as the mean for the exponential sojourns.

Compared to (2.13), $f_\alpha$ is also the spectral density of a CARMA(2, 1) with $a_1 = (\xi_1 - 1)/\lambda$, $a_2 = \bar{a}_1$, and $b_1 = -|a_1|$, yielding the following proposition.

**Proposition 2.10** Given a CARMA(2, 1) with $a_2 = \bar{a}_1$, there exists a semi-Markov process constructed from an AR(2) with complex conjugate roots and Exp(\lambda) sojourns such that they have the same covariance structure if and only if $b_1 = -|a_1|$.

To sum up, for an AR(2) process with $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2$, the region of the parameter space $(\phi_1, \phi_2)$ such that the AR(2) is embeddable in a semi-Markov process is displayed in Figure 2.6. In Figure 2.6, the entire triangular region corresponds to the parameter space of
Figure 2.5 Given $\xi_1 = re^{i\theta}$ and $(\theta, r) \in (0, \pi/2) \times (0, 1)$, the value of $\lambda$ (contour) is the mean of the exponential sojourns for which the constructed semi-Markov process has the same autocovariances as the AR(2) with complex conjugate roots at all integer lags.

Weakly stationary AR(2) processes and the shaded region corresponds to AR(2) processes which are embeddable in semi-Markov processes. The case of real roots corresponds to the shaded sub-region on the right hand side above the parabola. The case of equal roots corresponds to the parabola itself. The case of complex roots corresponds to the other shaded sub-region.

From the above discussion, all semi-Markov processes constructed from AR(2) states and exponential sojourns have the same second order covariance structures as CARMA(2,1) processes and therefore the shaded region in Figure 2.6 consists of AR(2) processes which are embeddable in CARMA(2,1). The region of $\phi_1$ and $\phi_2$ for embedding an AR(2) in a CARMA(2,1) through this semi-Markov construction is smaller than the region found by Brockwell (1994) for embedding an AR(2) in a CARMA(2,1). The difference of the regions between Brockwell's approach and ours comes from the case of complex roots. Not apparent from Figure 2.6 is the fact that the shaded region covers a very small part of the space left of $\phi_1 = 0$ and below the parabola. In Brockwell's result, the region for embedding an AR(2) with complex roots in a
Figure 2.6 The shaded area is the set of all $\phi_1$ and $\phi_2$ for which the corresponding AR(2) process is embeddable in a semi-Markov process. All of the parabola is included in the shaded region. The entire triangular area is the region for an AR(2) to be stationary.

CARMA(2,1) process covers a much larger sub-region left of $\phi_1 = 0$ and below the parabola in Figure 2.6.

In general, the embedding of discrete-time AR($p$) for $p > 2$ is a hard problem and has not been solved. In our approach, the case of AR($p$) can be studied using similar but more complicated derivations. We conjecture that at a minimum, the special case of AR($p$) with distinct real roots will be embeddable in a CARMA($p, p - 1$). Therefore, the semi-Markov construction may be useful in obtaining further results on the embedding of AR($p$) processes in CARMA processes.
2.3 Hierarchical Long-Memory Processes

As a more general class of models, we consider hierarchical long-memory processes with the following form

\[ y_t = g(Y_{t-1}, \alpha_t, \varepsilon_t), \]

where \( Y_{t-1} = (y_1, y_2, \ldots, y_{t-1})' \) is the data vector up to time \((t - 1)\), \(\{\alpha_t\}\) is a long-memory process which may or may not be observed directly, and \(\{\varepsilon_t\}\) is a Gaussian process with mean zero and variance \(\sigma^2\), independent of \(\{\alpha_t\}\).

Three special cases of hierarchical long-memory processes—ARFIMA with additive noise, long-memory stochastic volatility models and random coefficient models with long-memory coefficients—are introduced in the following sections.

2.3.1 ARFIMA with Additive Noise

Let

\[ y_t = g(\alpha_t, \varepsilon_t) = \alpha_t + \varepsilon_t, \]

where \(\{\alpha_t : t \in \mathbb{Z}\}\) is an ARFIMA process. Then the observed process \(\{y_t\}\) is a noisy version of an ARFIMA. In particular, if the noise \(\{\varepsilon_t\}\) follows a mixture distribution

\[ \varepsilon_t \sim \begin{cases} 0; & \text{with probability } 1 - p, \\ N(0, \sigma^2_\varepsilon); & \text{with probability } p, \end{cases} \]

with a given \(p \in (0, 1)\), then \(\{y_t\}\) becomes an ARFIMA with additive outliers. It is well known that an ARMA plus uncorrelated noise is still an ARMA, and more generally an ARMA plus ARMA is still an ARMA, but this closure does not hold for ARFIMA models.

The autocovariance function and the spectral density for ARFIMA plus noise are

\[ \gamma_y(h) = \gamma_\alpha(h) + \gamma_\varepsilon(h) \]

\[ = \begin{cases} \gamma_\alpha(0) + \sigma^2_\varepsilon; & \text{if } h = 0, \\ \gamma_\alpha(h); & \text{otherwise}, \end{cases} \]
For the additive outlier model, \( \sigma^2 = \rho \sigma^2_\epsilon \). The long memory feature is clearly preserved in this model.

### 2.3.2 Long-Memory Stochastic Volatility Models

Let

\[
y_t = g(\alpha_t, \varepsilon_t) \equiv e^{\alpha_t/2} \varepsilon_t,
\]

where \( \{\alpha_t\} \) is a long-memory process. Then the observed process \( \{y_t\} \) is called a long-memory stochastic volatility model. The cases for \( \{\alpha_t\} \) being an ARFIMA or a semi-Markov long-memory process are both discussed in the next chapter. Under stochastic volatility models, the autocovariance functions for the observed process \( \{y_t\} \) and the transformed process \( \{y^2_t\} \), which is a proxy for its volatility, are derived as follows.

\[
\gamma_y(h) = E(y_t y_{t+h})
\]

\[
= E\left(e^{\alpha_t/2} \varepsilon_t e^{\alpha_{t+h}/2} \varepsilon_{t+h}\right)
\]

\[
= E\left(e^{\alpha_t \varepsilon_t^2} 1_{(h=0)}\right)
\]

\[
= (E e^{\alpha_t^2}) \sigma^2_\epsilon 1_{(h=0)},
\]

\[
\gamma_{y^2}(h) = E(y^2_t y^2_{t+h}) - (E y^2_t)^2
\]

\[
= \left\{
\begin{array}{ll}
E \left(e^{2\alpha_t \varepsilon_t^2} - (E y^2_t)^2\right); & h = 0, \\
E \left(e^{\alpha_t+\alpha_{t+h}}\right) (E \varepsilon^2_t) (E \varepsilon^2_{t+h}) - (E y^2_t)^2; & \text{otherwise},
\end{array}
\right.
\]

\[
= \left\{
\begin{array}{ll}
3(E e^{2\alpha_t}) \sigma^4_\epsilon - (E e^{\alpha_t})^2 \sigma^4_\epsilon; & h = 0, \\
E \left(e^{\alpha_t+\alpha_{t+h}}\right) \sigma^4_\epsilon - (E e^{\alpha_t})^2 \sigma^4_\epsilon; & \text{otherwise}.
\end{array}
\right.
\]

Further assume \( \{\alpha_t\} \) is a mean-zero Gaussian process, then

\[
\gamma_y(h) = e^{\gamma_0(0)/2} \sigma^2_\epsilon 1_{(h=0)},
\]

\[
\gamma_{y^2}(h) = \left\{
\begin{array}{ll}
e^{\gamma_0(0)} \sigma^4_\epsilon \left(3e^{\gamma_0(0)} - 1\right); & h = 0, \\
e^{\gamma_0(0)} \sigma^4_\epsilon \left(e^{\gamma_0(h)} - 1\right); & \text{otherwise}.
\end{array}
\right.
\]
Since \( e^x - 1 \sim x \) for \( x \) close to zero,
\[
e^{\gamma_a(h)} - 1 \sim \gamma_a(h); \quad \text{as } h \to \infty.
\]

Therefore the process \( \{y_t^2\} \) is a long-memory process. The observed process \( \{y_t\} \) is actually a martingale difference sequence. The marginal distribution is a mixture Gaussian distribution with variance greater than \( \sigma^2 \).

2.3.3 Random Coefficient Models with Long-Memory Coefficients

Let
\[
y_t = g(Y_{t-1}, \alpha_t, \varepsilon_t) = \alpha_t y_{t-1} + \varepsilon_t,
\]
where \( \{\alpha_t\} \) is a long-memory process with \( E\alpha_t^2 < 1 \) and \( \{\varepsilon_t\} \) are iid \( N(0, \sigma^2) \). Then \( \{y_t\} \) is called a random coefficient AR(1) model with long-memory coefficients. Random coefficient models have been studied for short-memory coefficients (Nicholls and Quinn, 1980) but there seems to be no literature existing for long memory cases.

**Proposition 2.11** (Brandt, 1986) If the sequence \( \{(\alpha_t, \varepsilon_t)\} \) is stationary and ergodic satisfying \(-\infty \leq E\log|\alpha_t| < 0 \) and \( E(\log|\varepsilon_t|)^+ < \infty \), where \( x^+ = \max(0, x) \) for \( x \in \mathbb{R} \), then \( \{y_t\} \) has the unique stationary solution
\[
y_t = \sum_{j=0}^{\infty} \left( \prod_{k=0}^{j-1} \alpha_{t-k} \right) \varepsilon_{t-j}; \quad t \in \mathbb{Z}.
\]

Under the assumption \( E\alpha_t^2 < 1 \) and applying the Jensen inequality, we have
\[
E \log |\alpha_t| \leq \frac{1}{2} \log E\alpha_t^2 < 0.
\]

Therefore, by Proposition 2.11, the process \( \{y_t\} \) defined in (2.17) is strictly stationary and has the unique solution with an infinite moving average representation in (2.18). Let \( a_j = \prod_{k=0}^{j-1} \alpha_{t-k} \); that is \( y_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \). Assuming the process \( \{y_t\} \) has finite second moment, the autocovariance function of \( \{y_t\} \) can be expressed as
\[
\gamma_y(h) = E \left[ \prod_{j=1}^{h} a_{t+j} \left( \sum_{j=0}^{\infty} a_j^2 \right) \right].
\]
In general, the process \( \{y_t\} \) does not necessarily preserve the long memory feature of \( \{\alpha_t\} \). Preservation depends upon moment conditions for \( \{y_t\} \) and also the strength of dependence in \( \{\alpha_t\} \). For instance, the long memory feature will disappear in \( \{y_t\} \) if \( E y_t^4 < \infty \) because

\[
|\gamma_y(h)| \leq E \left[ \prod_{j=1}^{h} \alpha_{t+j} \right] \left( \sum_{j=0}^{\infty} a_j^2 \right)
\leq \left\{ E \left( \prod_{j=1}^{h} \alpha_{t+j}^2 \right) \right\} \left( \sum_{j=0}^{\infty} a_j^2 \right)^{1/2}
= \left\{ E \left( \prod_{j=1}^{h} \alpha_{t+j}^2 \right) \right\} \left( \frac{E y_t^4}{3\sigma_t^4} \right)^{1/2},
\]

in which

\[
E y_t^4 = E \left( \sum_{j=0}^{\infty} a_j^2 \right)^4
= \left( \sum_{j=0}^{\infty} E a_j^4 \right) \left( E \varepsilon_t^4 \right) + 3 \left( \sum_{j=0}^{\infty} \sum_{k \neq j} a_j^2 a_k^2 \right) \left( E \varepsilon_t^2 \right)^2
= 3\sigma_t^4 E \left( \sum_{j=0}^{\infty} a_j^2 \right)^2,
\]

and

\[
E \left( \prod_{j=1}^{h} \alpha_{t+j}^2 \right) \leq E \left( \frac{1}{h} \sum_{j=1}^{h} \alpha_{t+j}^2 \right)^h \to (E\alpha_t^2)^h \text{ as } h \to \infty,
\]

therefore, for sufficiently large \( h^* \),

\[
\sum_{h=h^*}^{\infty} |\gamma_y(h)| \leq \left( \frac{E y_t^4}{3\sigma_t^4} \right)^{1/2} \sum_{h=h^*}^{\infty} (E\alpha_t^2)^{h/2} < \infty,
\]

provided \( E\alpha_t^2 < 1 \).

As another example, assume \( \alpha_t \equiv \alpha \sim \text{Beta}(\beta_1, \beta_2) \); that is the probability density function of \( \alpha \) is

\[
f(x) = \frac{\Gamma(\beta_1 + \beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} x^{\beta_1-1} (1-x)^{\beta_2-1}; \quad x \in (0,1).
\]

The process \( \{y_t\} \) is long memory if \( \beta_2 < 1 \) and is short memory if \( \beta_2 > 2 \) because

\[
\gamma_y(h) = E \left[ \frac{\alpha^h}{1-\alpha^2} \right],
E\alpha^h \leq \gamma_y(h) \leq E \left[ \frac{\alpha^h}{2\alpha(1-\alpha)} \right],
\]
and for sufficiently large $h$,

$$E a^h = \int_0^1 x^h f(x) dx = \frac{\Gamma(\beta_1 + \beta_2)}{\Gamma(\beta_1)} \frac{\Gamma(h + \beta_1)}{\Gamma(h + \beta_1 + \beta_2)} \sim \frac{\Gamma(\beta_1 + \beta_2)}{\Gamma(\beta_1)} \frac{e^{-h-\beta_1+1}(h + \beta_1 - 1)^{h+\beta_1-1/2}}{e^{-h-\beta_1-\beta_2+1}(h + \beta_1 + \beta_2 - 1)^{h+\beta_1+\beta_2-1/2}} \sim h^{-\beta_2},$$

$$E \left[ \frac{\alpha^h}{2\alpha(1-\alpha^2)} \right] = \int_0^1 \frac{x^h}{2x(1-x)} f(x) dx = \frac{\Gamma(\beta_1 + \beta_2)\Gamma(\beta_2 - 1)}{2\Gamma(\beta_1)\Gamma(\beta_2)} \frac{\Gamma(h + \beta_1 - 1)}{\Gamma(h + \beta_1 + \beta_2 - 2)} \sim \frac{\Gamma(\beta_1 + \beta_2)}{2\Gamma(\beta_1)(\beta_2 - 1)} \frac{e^{-h-\beta_1+2}(h + \beta_1 - 2)^{h+\beta_1-3/2}}{e^{-h-\beta_1-\beta_2+3}(h + \beta_1 + \beta_2 - 3)^{h+\beta_1+\beta_2-5/2}} \sim h^{1-\beta_2},$$

in which the relation $\Gamma(x) \sim \sqrt{2\pi}e^{-x+1}(x-1)^{x-1/2}$ is used for large $x$.

In this example, the parameter $\beta_2$ completely controls the memory property of $\{y_t\}$. The fact is that $\beta_2$ controls the amount of probability mass near $\alpha = 1$ which corresponds to the nonstationary case. As $\beta_2 \to 0$, the probability mass is shifted toward one, therefore producing stronger dependence for the $\{y_t\}$ process. This argument is similar to the result obtained by Granger (1980). Granger showed that the process aggregated from a large number of independent AR(1) processes may have the long memory feature. As an example in his paper, assuming the squared random coefficient is Beta($\beta_1, \beta_2$), the average of the aggregated process is an integrated process with differencing parameter $1 - \beta_2/2$. Therefore it is long memory when $\beta_2 \in (1, 2)$ and is short memory when $\beta_2 \geq 2$. For $\beta_2 \leq 1$, the aggregated process does not have finite variance.
3 ESTIMATION

In the introduction, we have described the possible difficulties for estimating long-memory processes, including nasty computation of the autocovariances, numerical imprecision of high-dimensional matrix inversion, and the problem of high-dimensional integration for hierarchical models. In our study, Bayesian analysis is considered to unify both the estimation and prediction problems.

3.1 ARFIMA(0, d, 0) and Related Models

In this section, the idea of the estimation procedure is introduced in Section 3.1.1. and some diagnostics are provided in Section 3.1.2. Then an algorithm is described in Section 3.1.3 focusing on ARFIMA(0, d, 0) processes and ARFIMA(0, d, 0) with additive noise because the methodology can be easily extended to other models. In addition, posterior inferences can be compared to analytical results in the simplest cases as well as to published simulation results for other estimation procedures, which are summarized in Section 3.1.4. The extensions for ARFIMA(p, d, 0), ARFIMA(0, d, q), ARFIMA(p, d, q) and ARFIMA(0, d, 0) with additive AR(p) noise can be derived in a straightforward way and are provided in the Appendix B.

3.1.1 Overview

We know that ARFIMA(0, d, 0) exhibits long memory and therefore its covariance structure is very different from the classical short memory ARMA(p, q) processes. Though all ARMA processes have geometrically decaying autocovariance functions, certain ARMA processes have autocovariance functions which decay slowly enough to resemble long memory (O'Connell, 1971). Therefore, our approach is to approximate an ARFIMA(0, d, 0) process by an ARMA
process with appropriate coefficients which are functions of \( d \). We then use importance sampling to re-weight posterior inferences from the approximate model into the correct posterior inferences. More specifically, for the parametric function \( g(\theta) \), the posterior mean is obtained as

\[
E_0[g(\theta) \mid \alpha_T] = \int g(\theta)p_0(\theta \mid \alpha_T)d\theta
= \frac{\int g(\theta)p_0(\theta \mid \alpha_T)p_1(\theta \mid \alpha_T)d\theta}{\int p_1(\theta \mid \alpha_T)d\theta}
= \frac{E_1[g(\theta)c(\theta \mid \alpha_T)]}{E_1[c(\theta \mid \alpha_T)]},
\]

where \( \theta = (d, \sigma_\theta^2)' \) is the parameter vector, \( \alpha_T = (\alpha_1, \alpha_2, \ldots, \alpha_T)' \) are the observations from an ARFIMA(0, \( d \), 0), \( p_0 \) is the posterior under the true ARFIMA(0, \( d \), 0) model, \( p_1 \) is the posterior under the approximate ARMA model and \( c(\theta \mid \alpha_T) = L_0(\theta \mid \alpha_T)/L_1(\theta \mid \alpha_T) \) is the likelihood ratio. Assume that \( \{\theta_j\} \) are independent draws from \( p_1 \). Then, a Monte Carlo estimate of \( E_0[g(\theta) \mid \alpha_T] \) is

\[
\hat{E}_0[g(\theta) \mid \alpha_T] = \left( \sum_{j=1}^{K} c(\theta_j \mid \alpha_T) \right)^{-1} \sum_{j=1}^{K} g(\theta_j)c(\theta_j \mid \alpha_T),
\]

where \( \{c(\theta_j \mid \alpha_T)\} \) are called the importance weights for \( \{\theta_j\} \).

By using the idea of importance sampling and the ARMA approximation, the problem of making inference for an ARFIMA(0, \( d \), 0) model is reduced to the problem of making inference for the corresponding ARMA model. A good approximation requires that the importance weights are as close as possible for all \( \alpha_T \) and \( \theta \). Since the importance weight \( c(\theta \mid \alpha_T) \) is a random quantity depending on \( \alpha_T \), we select the approximating ARMA process by minimizing an approximation to the Kullback-Leibler discrepancy, subject to the constraint of equal variance for the ARFIMA(0, \( d \), 0) model and the approximating ARMA model. The Kullback-Leibler discrepancy is defined by

\[
-E \left\{ \log \frac{L_1(\theta \mid \alpha_T)}{L_0(\theta \mid \alpha_T)} \right\} \approx \frac{1}{2} \sum_{\omega_j \in \mathcal{F}_T} \log \frac{f_0(\omega_j)}{f_1(\omega_j)} - \frac{1}{2} \sum_{\omega_j \in \mathcal{F}_T} \left\{ 1 - \frac{f_0(\omega_j)}{f_1(\omega_j)} \right\},
\]

where \( \mathcal{F}_T = \{ j : j = 0, \pm 1, \pm 2, \ldots, \pm[T/2] \} \) and \( \{\omega_j = 2\pi j/T : j \in \mathcal{F}_T\} \) is the set of Fourier frequencies. As \( T \to \infty \), minimizing this criterion is equivalent to minimizing the corresponding
integral expression:

\[ D(f_0, f_1) \equiv -\int_{-\pi}^{\pi} \log \frac{f_0(\omega)}{f_1(\omega)} \, d\omega - \int_{-\pi}^{\pi} \left\{ 1 - \frac{f_0(\omega)}{f_1(\omega)} \right\} \, d\omega, \]

which does not depend on the sample size \( T \).

In this study, we use \( p = 2 \) and \( q = 2 \). Without loss of generality, let \( \text{var}(\alpha_t) = 1 \); that is \( \sigma^2 = \Gamma^2(1 - d)/\Gamma(1 - 2d) \). Therefore, the coefficients of the ARMA(2,2) model, which satisfies \((1 + a_1 B + a_2 B^2)\alpha_t = (1 + b_1 B + b_2 B^2)\eta_t\), are chosen by the optimization criterion:

\[(a_1, a_2, b_1, b_2) = \text{argmin} D(f_0, f_1), \text{ subject to } \text{var}(\alpha_t) = 1, \quad (3.1)\]

where

\[ f_0(\omega) = \frac{\Gamma^2(1 - d)}{2\pi \Gamma(1 - 2d)} \left| 1 - e^{i\omega} \right|^{-2d}, \]

\[ f_1(\omega) = \frac{\rho \sigma^2}{2\pi} \left| \frac{1 + b_1 e^{i\omega} + b_2 e^{2i\omega}}{1 + a_1 e^{i\omega} + a_2 e^{2i\omega}} \right|^2, \]

and \( r \) is the scale factor such that \( \text{var}(\alpha_t) = 1 \). The minimization needs to be done only once to find the relation between \( d \) and the coefficients of the ARMA, which do not depend on the random vector \( \alpha_T \) or the sample size \( T \). The corresponding coefficients of the ARMA(2,2) are displayed in Figure 3.1. In Figure 3.1, the four coefficients were evaluated at \( d = -0.48, -0.46, \ldots, 0.46, 0.48 \) and then interpolated using cubic splines. The case at \( d = 0 \) corresponds to the independent process for which the coefficients are not identified. The relationship between each of these four coefficients and \( d \) is quite smooth and linear.

3.1.2 Quality of the ARMA Approximation

Since ARMA is not the only candidate model with a finite state-space representation which could be used to approximate ARFIMA, the performance of the ARMA approximation is evaluated by making a comparison with another candidate: the truncated AR approximation. The truncated AR approximation is obtained by truncating (2.2) at a fixed order \( p \) and rescaling the white noise variance to get the ARFIMA(0, \( d \), 0) process variance. Again, we consider the Kullback-Leibler discrepancy as the measure of difference between the approximations and
Figure 3.1 The relation between $d$ in the ARFIMA(0,d,0) model and the coefficients of the approximate ARMA(2,2) model satisfying $(1 + a_1 B + a_2 B^2)\alpha_t = (1 + b_1 B + b_2 B^2)\eta_t$. The functions are evaluated at $d = -0.48, -0.46, \ldots, 0.46, 0.48$, and then interpolated using cubic splines.

The ARFIMA(0,d,0) processes. The corresponding order of the truncated AR approximations which achieves the same Kullback-Leibler discrepancy as our ARMA(2,2) process is displayed in Figure 3.2 for various $d$. The required order is greater than 2000 for $d \geq 0.38$ and increases extremely fast as $d$ increases. From Figure 3.2, our four-coefficient approximation (depending on one parameter) has outstanding performance relative to the truncated AR approximations in which more than 35 coefficients are required even for very small nonzero $|d|$. The performance is even more dramatic when $|d|$ increases. In Figure 3.3, the differences between the log spectral densities of the approximating process and the target ARFIMA(0, d, 0) are given for the case $d = 0.4$.

We also evaluate the performance of the importance sampling in the estimation procedure.
Figure 3.2 The minimal orders of the truncated AR process to achieve the same Kullback-Leibler discrepancy produced by our approximate ARMA(2,2) process for various $d$.

Figure 3.3 The difference between the log spectral densities of the approximating processes and the target ARFIMA(0,0.4,0) process.
through examining the importance weights. The importance weight is the ratio of the likelihood $L_0$ under ARFIMA(0, $d$, 0) versus the likelihood $L_1$ under the ARMA(2,2). The likelihood function can be expressed as a product form using one-step predictions and the corresponding mean squared errors, that is

$$L_i(\theta | \alpha_T) = \left( \frac{1}{2\pi} \right)^{-T/2} \left( \prod_{t=1}^{T} \nu_t^{(i)} \right) \exp \left\{ -\frac{T}{2} \left( \frac{\alpha_t - \hat{\alpha}_t^{(1)}}{\nu_t^{(1)}} \right)^2 \right\},$$

for $i = 0, 1$, where $\nu_t^{(0)}$ and $\nu_t^{(1)}$ are the one-step prediction mean squared errors of the best linear predictors $\hat{\alpha}_t^{(0)}$ and $\hat{\alpha}_t^{(1)}$ under the correctly specified ARFIMA model and the misspecified ARMA model, respectively. Therefore, in order to compare the two likelihoods it is equivalent to compare the one-step prediction mean squared errors $\nu_t^{(0)}$, $\nu_t^{(1)}$ and $E_0 \left[ (\alpha_t - \hat{\alpha}_t^{(1)})^2 \right]$ where the expectation is taken under the ARFIMA model. The values of $\log \left\{ \nu_t^{(1)}/\nu_t^{(0)} \right\}$, which are nonrandom and independent of $\alpha_T$, are plotted in Figure 3.4 for different values of $d$ and $T$.

The log exponential term in $L_0(\theta | \alpha_T)/L_1(\theta | \alpha_T)$ is random with the expectation

$$E_0 \left\{ \frac{(\alpha_t - \hat{\alpha}_t^{(1)})^2}{2\nu_t^{(1)}} - \frac{(\alpha_t - \hat{\alpha}_t^{(0)})^2}{2\nu_t^{(0)}} \right\} = \frac{1}{2} \left\{ \frac{E(\alpha_t - \hat{\alpha}_t^{(1)})^2}{\nu_t^{(1)}} - 1 \right\},$$

where the expectation is taken under the ARFIMA model. These values are plotted in Figure 3.5 for $d = 0.05, 0.10, \ldots, 0.45$ and $T = 1, 2, \ldots, 1000$. The small variation in both plots indicate small differences between two likelihoods and thus the approximation is good.

### 3.1.3 Algorithms

#### 3.1.3.1 ARFIMA(0, $d$, 0)

Assume $\{y_1, y_2, \ldots, y_T\}$ are observations from an ARFIMA(0, $d$, 0) process defined in (2.1). We model the ARFIMA as an ARMA(2,2),

$$y_t = \alpha_t,$$

$$(1 + a_1 B + a_2 B^2)\alpha_t = (1 + b_1 B + b_2 B^2)\eta_{1t},$$

where $\{\eta_{1t}\}$ is iid $N(0, r(d)\sigma^2)$; the coefficients $a_1$, $a_2$, $b_1$ and $b_2$ are obtained from (3.1) (displayed in Figure 3.1); and the function $r(d)$ is the scale factor to satisfy the constraint of

...
Figure 3.4 The log differences between the one-step prediction mean squared errors of the ARFIMA$(0, d, 0)$ model and the approximation ARMA$(2,2)$ model for $d$ equal to $0.05, 0.10, \ldots, 0.45$ and $T = 2, 3, \ldots, 1000$. 
Figure 3.5. The values of $\text{E}(\alpha_i - \hat{\alpha}_i)$ under ARIMA(0,1,0) model for $\theta = 0.05, 0.10, \ldots, 0.45$ and $T = 1$ to 1000.
equal variance. The above ARMA(2, 2) model \( \{y_t\} \) can be expressed in a state-space form

\[
y_t = \alpha_t,
\]

\[
\alpha_t = (b_2(d), b_1(d), 1) X_t,
\]

\[
X_t = AX_{t-1} + \eta_t,
\]

(3.2)

(3.3)

where

\[
\begin{pmatrix}
x_{t-2} \\
x_{t-1} \\
x_t
\end{pmatrix} = \begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & -a_2(d) & -a_1(d)
\end{pmatrix}
\begin{pmatrix}
x_{t-1} \\
x_{t-2} \\
x_t
\end{pmatrix} + \begin{pmatrix}
0 \\
0 \\
1
\end{pmatrix} \eta_t.
\]

Let \( \theta = (d, \sigma_n^2)' \) and \( X_{0T} = (x_1, x_2, \ldots, x_T)' \). By adding latent variables \( (x_{-1}, x_0)' \) to augment the data, the joint density of \( Y_T, (x_{-1}, x_0)' \) and \( \theta \) becomes

\[
f(Y_T, x_{-1}, x_0, \theta) = f(Y_T, x_{-1}, x_0 | \theta) \pi(\theta)
\]

\[
= \prod_{t=1}^{T} f(x_t | x_{t-1}, x_{t-2}, \theta) f(x_{-1}, x_0 | \theta) \pi(\theta),
\]

where

\[
f(x_t | x_{t-1}, x_{t-2}, \theta) = (2\pi r(d) \sigma_n^2)^{-1/2} \exp \left\{ -\frac{(x_t + a_1 x_{t-1} + a_2 x_{t-2})^2}{2r(d) \sigma_n^2} \right\},
\]

\[
f(x_{-1}, x_0 | \theta) = (2\pi)^{-1} |\Sigma_{11}|^{-1/2} \exp \left\{ -\frac{(x_{-1}, x_0, \Sigma_{11}^{-1}(x_{-1}, x_0)')}{2} \right\},
\]

and \( \Sigma_{11} \) is the covariance matrix of \( (x_{-1}, x_0)' \). Because \( \{x_t\} \) in (3.3) is an AR(2) process, \( \Sigma_{11} \) is well known. Assume that the prior has the product form \( \pi(\theta) = \pi(d) \pi(\sigma_n^2) \) with

\[
\pi(d) = 1_{(-1/2, 1/2)}(d),
\]

\[
\pi(\sigma_n^2) \propto (\sigma_n^2)^{-1} 1_{(0, \infty)}(\sigma_n^2).
\]

The posterior distribution for \( \theta \) can be found via a Metropolis-Hastings algorithm which constructs a Markov chain with invariant distribution equal to the target posterior distribution by iteratively sampling from the proposal distributions and moving from one state to another.
state with a specified probability. In particular, it is called Gibbs sampling (Geman and Geman, 1984) when the proposal distributions are identical to the full conditional distributions. Under our setup, the conditional distributions for $\sigma^2_\eta$ and $(x_{-1}, x_0)'$ satisfy

$$f(\sigma^2_\eta | Y_T, x_{-1}, x_0, d) \propto (\sigma^2_\eta)^{-\frac{T+2}{2}-1} \mathbf{1}_{(0,\infty)}(\sigma^2_\eta) \exp \left\{ \frac{-\sum_{t=1}^T (x_t + a_1 x_{t-1} + a_2 x_{t-2})^2}{2r(d)\sigma^2_\eta} \right\} \times \exp \left\{ \frac{(1 - a_2^2)(x_{-1}^2 + x_0^2 + 2x_{-1}x_0a_1/(1 + a_2))}{2r(d)\sigma^2_\eta} \right\},$$

$$f(x_{-1}, x_0 | Y_T, d, \sigma^2_\eta) = f(x_{-1}, x_0 | x_1, x_2, d, \sigma^2_\eta). \quad (3.4)$$

The former is proportional to an inverse gamma distribution and the latter is a bivariate normal distribution. More precisely, the mean and covariance matrix for equation (3.4) is

$$E[(x_{-1}, x_0)' | x_1, x_2, d, \sigma^2_\eta] = \Sigma_{12} \Sigma_{11}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

$$\text{var}[(x_{-1}, x_0)' | x_1, x_2, d, \sigma^2_\eta] = \Sigma_{11} - \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{12},$$

where $\Sigma_{11} = \text{var}[(x_{-1}, x_0)']$ and $\Sigma_{12} = \text{cov}[(x_{-1}, x_0)', (x_1, x_2)]$.

The conditional distribution of $d$ is nonstandard; it is given by

$$f(d | Y_T, x_{-1}, x_0, \sigma^2_\eta) \propto r(d)^{-T/2} \exp \left\{ \frac{-\sum_{t=1}^T (x_t + a_1 x_{t-1} + a_2 x_{t-2})^2}{2r(d)\sigma^2_\eta} \right\} f(x_{-1}, x_0 | \theta). \quad (3.5)$$

In (3.5), $d$ is involved in the function $r(d)$, the coefficients $a_1(d)$, $a_2(d)$, and also the $\{x_t\}$ process since $X_{d_T}^T \equiv (x_{-1}, x_0, X_{d_T}'') = C^{-1} Y_T$ where

$$Y_T = \begin{pmatrix} x_{-1} \\ x_0 \\ Y_T \end{pmatrix}, \quad C = C(d) = \begin{pmatrix} 1 & 0 & b_2 \\ 0 & 1 & b_1 \\ \vdots & \vdots & \vdots \\ b_2 & b_1 & 1 \end{pmatrix}. \quad (3.6)$$

Because of the complexity of the conditional distribution of $d$, a Metropolis-Hastings algorithm is used to generate draws of $d$. The following truncated Gaussian proposal density $J(d | d^{(i-1)})$ is used for generating $d$:

$$J(d | d^{(i-1)}) \propto \exp \left\{ \frac{-\sum_{t=1}^T (x_t^* + a_1^* x_{t-1}^* + a_2^* x_{t-2}^*)^2}{2r(d^{(i-1)})\sigma^2_\eta} \right\} \mathbf{1}_{(-1/2,1/2)}(d),$$
where $d^{(i-1)}$ is the previous draw and

$$(x_{-1}, x_0, x_1^*, \ldots, x_T^*)' = \left[C(d^{(i-1)})\right]^{-1} Y_T,$$

$$a_1^* = a_{11}(d^{(i-1)}) + a_{12}(d^{(i-1)})d,$$

$$a_2^* = a_{21}(d^{(i-1)}) + a_{22}(d^{(i-1)})d,$$

$a_1^*$ and $a_2^*$ are linearizations of the functions $a_1(d)$ and $a_2(d)$ at $d = d^{(i-1)}$, and $\{a_{ij} : i, j = 1, 2\}$ are the corresponding intercepts and slopes. The proposal $J$ is a good approximation of the conditional distribution of $d$ for two reasons. First, the effect of ignoring $f(x_{-1}, x_0 | \theta)$ in $J$ is negligible as $T$ goes to $\infty$ (see (3.5)). Second, from Figure 3.1, the coefficients $a_1$ and $a_2$ are almost linear functions of $d$. This relation makes the local linearizations $a_1^*$ and $a_2^*$ good approximations of $a_1$ and $a_2$ for the entire domain of $d$.

The transition rule for the Markov chain is defined as

$$d^{(i)} = \begin{cases} 
    d^*; & \text{with probability } \delta, \\
    d^{(i-1)}; & \text{with probability } 1 - \delta,
\end{cases}$$

where $d^*$ is a random draw from $J(d | d^{(i-1)})$ and the acceptance probability is

$$\delta = \min \left\{1, \frac{p(d^*) J(d^{(i-1)} | d^*)}{p(d^{(i-1)}) J(d^* | d^{(i-1)})} \right\},$$

(3.7)

where $p(d) \equiv f(d | Y_T, x_{-1}, x_0, \sigma^2_\eta)$. Under this setup, the Markov chain is ergodic since the domain of $d$ in $J$ is $(-1/2, 1/2)$ and is reversible by the choice of $\delta$. Therefore, the limiting distribution of the chain is the target conditional distribution in (3.5) (Hastings, 1970). The advantage of this particular choice of $J$ is that this Metropolis-Hastings algorithm becomes very easy to implement because $J(d | d^{(i-1)})$ is Gaussian and also $\delta$ is just a ratio of Gaussian densities. In this algorithm, the matrix $[C(d)]^{-1}$ has an explicit, simple form and there is no other high-dimensional matrix inversion involved.

To sum up, our Markov chain Monte Carlo procedure has the following steps:

1. Give initial values for $d$ and $\sigma^2_\eta$.

2. Generate $(x_{-1}, x_0)'$ from a multivariate normal distribution.
3. Generate $\sigma^2_\eta$ from an inverse gamma distribution.

4. Generate $d$ from a Metropolis-Hastings algorithm with a Gaussian proposal (this step might be performed several times to reduce the dependence between draws).

5. Repeat steps 2–4 until the Markov chain converges.

6. Adjust to the correct inference by importance sampling. For example, the posterior distribution of $g(\theta)$ is estimated by the Monte Carlo estimator

$$
\hat{F}_{g(\theta)}(x) = \left( \sum_{j=1}^{K} c(\theta_j) \right)^{-1} \sum_{j=1}^{K} c(\theta_j) I_{g(\theta_j) \leq x}.
$$

3.1.3.2 ARFIMA(0, d, 0) with Additive White Noise

Assume the process $\{y_t\}$ satisfies $y_t = \alpha_t + \varepsilon_t$, where $\{\alpha_t\}$ is an ARFIMA(0, d, 0) process defined in (2.1) and $\{\varepsilon_t\}$ is a Gaussian white noise process with variance $\sigma^2_\varepsilon$, independent of $\{\alpha_t\}$. Using (3.2) and (3.3), $\{y_t\}$ can be modeled as an ARMA(2,2) plus noise process, which has the following state-space form:

$$
y_t = (b_2, b_1, 1)X_t + \varepsilon_t,
$$

$$
X_t = AX_{t-1} + R\eta_t.
$$

Let $\theta = (d, \sigma^2_\eta, \sigma^2_\varepsilon)'$. Because $\{x_t\}$ are unobservable, the likelihood function of the approximating process is a $T$-dimensional integral. Therefore, we augment the parameter vector $\theta$ with the entire vector $X_{0:T}^T = (x_{-1}, x_0, x_1, \ldots, x_T)'$. The joint density of $Y_T, X_{0:T}^T$ and $\theta$ becomes

$$
f(Y_T, X_{0:T}^T, \theta) = f(Y_T | X_{0:T}^T, \theta)f(X_{0:T}^T | \theta)\pi(\theta)
$$

$$
= \left\{ \prod_{t=1}^{T} f(y_t | X_t, \theta) \right\} \left\{ \prod_{t=1}^{T} f(x_t | x_{t-1}, x_{t-2}, \theta) \right\} f(x_{-1}, x_0 | \theta)\pi(\theta),
$$

(3.8)

where $f(x_t | x_{t-1}, x_{t-2}, \theta)$ and $f(x_{-1}, x_0 | \theta)$ are defined in Section 3.1.3.1, and

$$
f(y_t | X_t, \theta) = \frac{1}{\sqrt{2\pi} \sigma^2_{\varepsilon}} \exp \left\{ -\frac{(y_t - (b_2, b_1, 1)X_t)^2}{2\sigma^2_{\varepsilon}} \right\}.
$$
For the hierarchical case, the priors of the two variance components are chosen to be inverse gamma distributions satisfying

\[ \pi(\sigma^2_e) \propto (\sigma^2_e)^{-2} e^{-1/\sigma^2_e}, \]
\[ \pi(\sigma^2_\theta) \propto (\sigma^2_\theta)^{-2} e^{-1/\sigma^2_\theta}. \]

Then, the sampling algorithms for \( \sigma^2_e \) and \( \sigma^2_\theta \) in the Metropolis algorithm both follow inverse gamma distributions. For the process with noise, the parameter \( d \) is not only involved in the term \( f(X_{0T}^T | \theta) \) but also \( f(Y_T | X_{0T}^T, \theta) \). Therefore, the corresponding Metropolis-Hastings algorithm should be modified as follows. First, the term \( f(x_{-1}, x_0 | \theta) \) is ignored as before. Second, for the \( i \)th Metropolis step, not only \( a_1(d) \) and \( a_2(d) \) are linearized at \( d = d^{(i-1)} \) but also \( b_1(d) \) and \( b_2(d) \). Denote the corresponding linearizations for \( a_i \) and \( b_i \) to be

\[ a^*_i = a_{i1}(d^{(i-1)}) + a_{i2}(d^{(i-1)})d, \]
\[ b^*_i = b_{i1}(d^{(i-1)}) + b_{i2}(d^{(i-1)})d, \]

where \( \{a_{ij} : i,j = 1,2\} \) and \( \{b_{ij} : i,j = 1,2\} \) are the intercepts and slopes for the tangent lines of \( a_i \)'s and \( b_i \)'s at \( d = d^{(i-1)} \). Thirdly, the value of \( r(d) \) is replaced by \( r(d^{(i-1)}) \). The proposal distribution \( J \) is defined to be a truncated Gaussian distribution satisfying

\[ J(d | d^{(i-1)}) \propto \exp \left\{ -\frac{\sum_{t=1}^{T}(x_t + a^*_1 x_{t-1} + a^*_2 x_{t-2})^2}{2r(d^{(i-1)})\sigma^2_\theta} - \frac{\sum_{t=1}^{T}(y_t - (b^*_1, b^*_1, 1)X_t)^2}{2\sigma^2_e} \right\}, \quad (3.9) \]

where \( d \in (-1/2,1/2) \). The acceptance probability is defined in (3.7) with \( p(d) \) replaced by the current conditional distribution for \( d \).

Unlike the ARFIMA(0, \( d \), 0) case, we have to generate the whole latent vector \( X_{0T}^T \) instead of \( (x_{-1}, x_0)' \). The advantage of our ARMA approximation is its finite state-space representation, which allows us to draw the entire latent vector in one step via the Kalman smoothing algorithm (Koopman and Shephard, 1992). Drawing the entire latent vector at once produces better mixing in the Metropolis algorithm and hence faster convergence (Carter and Kohn, 1994).

An extension of this algorithm, in which the noise process is replaced by a short memory AR(\( p \)) process, is described in Appendix B.
3.1.4 Simulation Study

A simulation study was conducted to investigate the performance of the Bayesian analysis using our MCMC algorithm. Four models are considered: ARFIMA(0, d, 0), ARFIMA(1, d, 0), ARFIMA(0, d, 1), and ARFIMA(0, d, 0) with additive white noise. Priors for d, \(\phi\) and \(\theta\) are uniform and priors for variance parameters are proportional to \(1/(\text{variance})\) for pure ARFIMA case and follow inverse gamma distributions for hierarchical case (described in Section 3.1.3). We check the frequentist properties of the posterior means, including bias, mean squared error (MSE) and the actual coverage based on the 95% credible interval of the posterior distribution. The quantities of interest include the parameters d, \(\sigma^2\), \(\phi\), \(\theta\) and also the future outcome \(y_{T+1}\).

Although MLE and some approximate MLEs are asymptotically efficient, their small sample properties may be poor. Sowell (1992) investigated the exact MLE, the approximate MLE proposed by Fox and Taqqu (1986), and the estimator proposed by Geweke and Porter-Hudak (1983). Sowell’s results show that the exact MLE generally performs better than others for ARFIMA(0, d, 0), ARFIMA(1, d, 0) and ARFIMA(0, d, 1) processes. We use the same three models as Sowell. Several different parameter values are examined for each model. For each process, 500 realizations for ARFIMA(0, d, 0) and 100 realizations for other models; each realization has sample size \(T = 100\). For each realization, a Metropolis algorithm with 500 iterations is run in which the Metropolis step for generating d is repeated 50 times within each iteration to reduce the dependence between draws. These chain lengths will not be adequate for all problems, but were chosen here to make the simulation study feasible. The initial value of d is chosen using the Geweke and Porter-Hudak procedure and the initial values of other parameters are chosen by the method of moments. The final inference is made using the values only from the second half of the draws in the Metropolis algorithm.

The first model considered is ARFIMA(0, d, 0). The results are presented in Table 3.1. Compared to the results by Sowell, our estimates of d have competitive performance both in bias and \(\sqrt{\text{MSE}}\). The posterior mean of d has negative bias for almost all cases which is similar to the results for other estimation procedures. The values in parentheses in the \(\sqrt{\text{MSE}}\) columns are the ratio of the estimated \(\sqrt{\text{MSE}}\) for the MLE of d (Sowell, 1992) to the estimated
\( \sqrt{\text{MSE}} \) for the posterior mean of \( d \), the ratio of the asymptotic \( \sqrt{\text{MSE}} \) for the MLE of \( \sigma_n^2 \) to the estimated \( \sqrt{\text{MSE}} \) for the posterior mean of \( \sigma_n^2 \), and the ratio of the theoretical \( \sqrt{\text{MSE}} \) for the best predictor of \( y_{101} \) given \( Y_{100} \) (computed with \( \theta \) known) to the estimated \( \sqrt{\text{MSE}} \) for the posterior mean of \( y_{101} \). In addition, the asymptotic \( \sqrt{\text{MSE}} \) for the MLE of \( d \) is 0.078. These ratios indicate the relative efficiency of the posterior means compared with the MLEs. Most of the ratios in Table 1 are around one which means our procedure is comparable with the MLEs. In particular, the ratios for \( y_{101} \) are close to one which indicates our procedure has good predictive ability for ARFIMA(0, \( d \), 0) processes. Besides, the actual coverages for all quantities are very close to the target coverage 0.95, suggesting that the Markov chains have reached equilibrium.

The second model considered is the ARFIMA(1, \( d \), 0) process. The results are presented in Table 3.2. Similarly, the relative efficiency for the parameters are reported in the parentheses. Compared to the results by Sowell, the biases in \( d \) and \( \phi \) in our procedure are competitive with those for MLE. However, the \( \sqrt{\text{MSE}} \) in \( d \) and \( \phi \) in our procedure are smaller than those for MLE for most of the cases, in particular for negative \( d \). The attained coverages of the credible intervals are, for the most part, close to the nominal 0.95, though there is a slight tendency to undercover in some cases.

The third model considered is the ARFIMA(0, \( d \), 1) process. The results are presented in Table 3.3. For this model, the estimated posterior means have relatively larger biases compared with the MLEs but none of the differences is significant. The magnitudes of \( \sqrt{\text{MSE}} \) are competitive. The attained coverages are not as accurate as for the previous two models. This indicates slower convergence of the sampling procedure for ARFIMA(0, \( d \), 1), possibly caused by the slow mixing of draws for \( \theta \). Of course, this can be improved by considering longer chains or a higher order ARMA approximation.

For the fourth model, four cases with different signal-to-noise ratios (\( \rho = 2, 5, 10 \) and \( \infty \)) are considered, where \( \rho \) is defined to be the ratio of the variance of the ARFIMA(0, \( d \), 0) process to the variance of the additive noise process. The setup for the Metropolis algorithm

\(^1\)The estimated \( \sqrt{\text{MSE}} \) for the MLE of \( \sigma_n^2 \) was not reported in Sowell's paper, therefore the asymptotic \( \sqrt{\text{MSE}} \) is used instead.
is the same as in the first simulation. The results for $d$ are summarized in Table 3.4. The case with $p = \infty$ corresponds to pure ARFIMA$(0, d, 0)$ misspecified as ARFIMA plus noise. Generally speaking, for the cases with stronger dependence ($|d| = 0.3, 0.4$), the performance of the posterior mean of $d$ is improved as the signal-to-noise ratio increases. However, the performance among different ratios are similar for the cases with almost independent structure ($|d| = 0, 0.1$). The tendency holds similarly for $\sigma^2_{\eta}$ but the magnitudes in bias and $\sqrt{\text{MSE}}$ are substantially larger than those cases without noise. It is difficult to estimate $\sigma^2_{\eta}$ and $\sigma^2_{\epsilon}$ accurately since the two variance components are hardly separated from each other, especially when the signal-to-noise ratio is small. The values of $\sigma^2_{\eta}$ are always underestimated and the values of $\sigma^2_{\epsilon}$ are always overestimated however the overall variance of the process is fairly accurately estimated.

3.2 Semi-Markov Processes and Related Models

The class of semi-Markov processes is useful for describing the phenomenon of regime switches because of its piecewise-constant structure. In this section, the semi-Markov process, defined in Example 2.3 of Section 2.2.2, is considered to model long-memory regime switches. The hierarchical version of such processes is also discussed briefly.

3.2.1 AR(1) with Pareto Sojourns

Let $\{t_j : j = 1, 2, \ldots, T\}$ be the index set of observed times, which are equally spaced. Let

$$Y_T \equiv (y_{t_1}, y_{t_2}, \ldots, y_{t_T}) = \alpha_T \equiv (\alpha_{t_1}, \alpha_{t_2}, \ldots, \alpha_{t_T})',$$

be the observed data vector and

$$\alpha_{t_j} = w_{N(t_j)},$$

$$(1 - \phi B)w_n = \eta_n,$$

where $\{\eta_n : n \in \mathbb{Z}\}$ are iid $N(0, \sigma^2_{\eta})$ and the sojourn times $\{T_n : n = 2, 3, \ldots\}$ are from a Pareto distribution satisfying (2.11).
Table 3.1 The bias and the root mean squared error of the posterior mean and also the actual coverage of the 95% credible interval based on the posterior distribution for ARFIMA(0, \(d\), 0) process (\(T=100, 500\) replications). The values in parentheses are the relative efficiencies of the posterior means compared to the MLEs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias</th>
<th>(\sqrt{\text{MSE}})</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d)</td>
<td>(\hat{d})</td>
<td>(\hat{\sigma}_n^2)</td>
<td>(\hat{y}_{T+1})</td>
</tr>
<tr>
<td>(\sigma_n^2)</td>
<td>(\hat{\sigma}_n^2)</td>
<td>(\hat{y}_{T+1})</td>
<td>(\hat{\sigma}_n^2)</td>
</tr>
<tr>
<td>-0.4</td>
<td>1.690</td>
<td>0.015</td>
<td>0.035</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.801</td>
<td>-0.010</td>
<td>0.020</td>
</tr>
<tr>
<td>-0.2</td>
<td>1.900</td>
<td>-0.006</td>
<td>0.025</td>
</tr>
<tr>
<td>-0.1</td>
<td>1.971</td>
<td>-0.017</td>
<td>0.050</td>
</tr>
<tr>
<td>0.0</td>
<td>2.000</td>
<td>-0.049</td>
<td>-0.055</td>
</tr>
<tr>
<td>0.1</td>
<td>1.962</td>
<td>-0.005</td>
<td>0.040</td>
</tr>
<tr>
<td>0.2</td>
<td>1.820</td>
<td>0.035</td>
<td>0.167</td>
</tr>
<tr>
<td>0.3</td>
<td>1.519</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>0.4</td>
<td>0.966</td>
<td>-0.029</td>
<td>0.066</td>
</tr>
</tbody>
</table>
Table 3.2  The bias and the root mean squared error of the posterior mean and also the actual coverage of the 95% credible interval based on the posterior distribution for ARFIMA(1,d,0) process (T=100, 100 replications). The values in parentheses are the relative efficiencies of the posterior means compared to the MLEs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias</th>
<th>( \sqrt{\text{MSE}} )</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>( \phi )</td>
<td>( \sigma_n^2 )</td>
<td>( \hat{d} )</td>
</tr>
<tr>
<td>-0.3</td>
<td>-0.7</td>
<td>1.801</td>
<td>0.147</td>
</tr>
<tr>
<td>-0.3</td>
<td>-0.2</td>
<td>1.801</td>
<td>0.008</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.3</td>
<td>1.801</td>
<td>-0.022</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.8</td>
<td>1.801</td>
<td>-0.018</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.7</td>
<td>2.000</td>
<td>0.019</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.2</td>
<td>2.000</td>
<td>-0.141</td>
</tr>
<tr>
<td>0.0</td>
<td>0.3</td>
<td>2.000</td>
<td>-0.044</td>
</tr>
<tr>
<td>0.0</td>
<td>0.8</td>
<td>2.000</td>
<td>-0.021</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.7</td>
<td>1.519</td>
<td>-0.044</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.2</td>
<td>1.519</td>
<td>-0.255</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>1.519</td>
<td>-0.069</td>
</tr>
<tr>
<td>0.3</td>
<td>0.8</td>
<td>1.519</td>
<td>-0.015</td>
</tr>
</tbody>
</table>
Table 3.3 The bias and the root mean squared error of the posterior mean and also the actual coverage of the 95% credible interval based on the posterior distribution for ARFIMA(0, d, 1) process (T=100, 100 replications). The values in parentheses are the relative efficiencies of the posterior means compared to the MLEs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias</th>
<th>√MSE</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d$</td>
<td>$\hat{d}$</td>
<td>$\hat{\sigma}_n^2$</td>
<td>$\hat{\sigma}_{n+1}$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>$\hat{\theta}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>$\hat{\sigma}^2$</td>
<td>$\hat{\sigma}_{T+1}$</td>
<td>$\hat{\sigma}_{T+1}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$\hat{\gamma}$</td>
<td>$\hat{\gamma}_{T+1}$</td>
<td>$\hat{\gamma}_{T+1}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$d$</th>
<th>$\theta$</th>
<th>$\sigma^2$</th>
<th>$d$</th>
<th>$\hat{d}$</th>
<th>$\hat{\sigma}_n^2$</th>
<th>$\hat{\sigma}_{n+1}$</th>
<th>$\hat{\sigma}_{T+1}$</th>
<th>$\hat{\sigma}_{T+1}$</th>
<th>$d$</th>
<th>$\hat{d}$</th>
<th>$\hat{\sigma}_n^2$</th>
<th>$\hat{\sigma}_{n+1}$</th>
<th>$\hat{\sigma}_{T+1}$</th>
<th>$\hat{\sigma}_{T+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.3</td>
<td>-0.8</td>
<td>1.801</td>
<td>0.183</td>
<td>0.165</td>
<td>0.046</td>
<td>0.128</td>
<td>0.237 (0.91)</td>
<td>0.388 (0.66)</td>
<td>0.161 (1.20)</td>
<td>1.362 (0.99)</td>
<td>0.73</td>
<td>0.84</td>
<td>0.71</td>
<td>0.89</td>
</tr>
<tr>
<td>-0.3</td>
<td>-0.3</td>
<td>1.801</td>
<td>0.454</td>
<td>0.046</td>
<td>-0.413</td>
<td>-0.147</td>
<td>0.502 (0.45)</td>
<td>0.278 (0.92)</td>
<td>0.479 (0.52)</td>
<td>1.348 (1.00)</td>
<td>0.45</td>
<td>0.94</td>
<td>0.41</td>
<td>0.92</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.5</td>
<td>1.801</td>
<td>-0.036</td>
<td>0.069</td>
<td>0.019</td>
<td>0.082</td>
<td>0.120 (0.80)</td>
<td>0.337 (0.76)</td>
<td>0.123 (0.90)</td>
<td>1.211 (1.11)</td>
<td>0.82</td>
<td>0.85</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.9</td>
<td>1.801</td>
<td>-0.057</td>
<td>0.112</td>
<td>0.003</td>
<td>-0.115</td>
<td>0.102 (0.95)</td>
<td>0.322 (0.79)</td>
<td>0.048 (1.00)</td>
<td>1.399 (0.96)</td>
<td>0.71</td>
<td>0.86</td>
<td>0.82</td>
<td>0.87</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.8</td>
<td>2.000</td>
<td>0.154</td>
<td>0.117</td>
<td>-0.075</td>
<td>0.087</td>
<td>0.233 (0.96)</td>
<td>0.324 (0.87)</td>
<td>0.135 (1.53)</td>
<td>1.560 (0.91)</td>
<td>0.69</td>
<td>0.93</td>
<td>0.57</td>
<td>0.89</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.3</td>
<td>2.000</td>
<td>0.109</td>
<td>0.073</td>
<td>-0.115</td>
<td>-0.364</td>
<td>0.218 (0.91)</td>
<td>0.311 (0.91)</td>
<td>0.264 (0.90)</td>
<td>1.486 (0.95)</td>
<td>0.84</td>
<td>0.92</td>
<td>0.76</td>
<td>0.88</td>
</tr>
<tr>
<td>0.0</td>
<td>0.5</td>
<td>2.000</td>
<td>-0.060</td>
<td>0.123</td>
<td>0.028</td>
<td>0.312</td>
<td>0.140 (0.71)</td>
<td>0.342 (0.83)</td>
<td>0.135 (0.87)</td>
<td>1.676 (0.84)</td>
<td>0.85</td>
<td>0.91</td>
<td>0.84</td>
<td>0.87</td>
</tr>
<tr>
<td>0.0</td>
<td>0.9</td>
<td>2.000</td>
<td>-0.070</td>
<td>0.231</td>
<td>0.007</td>
<td>-0.168</td>
<td>0.130 (0.62)</td>
<td>0.521 (0.54)</td>
<td>0.050 (0.98)</td>
<td>1.585 (0.89)</td>
<td>0.76</td>
<td>0.77</td>
<td>0.75</td>
<td>0.84</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.8</td>
<td>1.519</td>
<td>0.006</td>
<td>0.030</td>
<td>0.000</td>
<td>-0.018</td>
<td>0.131 (1.87)</td>
<td>0.251 (0.86)</td>
<td>0.141 (1.82)</td>
<td>1.279 (0.96)</td>
<td>0.88</td>
<td>0.85</td>
<td>0.76</td>
<td>0.83</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.3</td>
<td>1.519</td>
<td>-0.017</td>
<td>0.047</td>
<td>0.025</td>
<td>-0.004</td>
<td>0.119 (1.19)</td>
<td>0.268 (0.80)</td>
<td>0.164 (1.12)</td>
<td>1.192 (1.03)</td>
<td>0.86</td>
<td>0.91</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>0.3</td>
<td>0.5</td>
<td>1.519</td>
<td>-0.053</td>
<td>0.157</td>
<td>0.058</td>
<td>-0.006</td>
<td>0.115 (0.77)</td>
<td>0.328 (0.66)</td>
<td>0.126 (0.87)</td>
<td>1.462 (0.84)</td>
<td>0.81</td>
<td>0.81</td>
<td>0.70</td>
<td>0.80</td>
</tr>
<tr>
<td>0.3</td>
<td>0.9</td>
<td>1.519</td>
<td>-0.040</td>
<td>0.224</td>
<td>-0.063</td>
<td>-0.080</td>
<td>0.105 (0.62)</td>
<td>0.409 (0.53)</td>
<td>0.198 (0.24)</td>
<td>1.386 (0.89)</td>
<td>0.83</td>
<td>0.73</td>
<td>0.66</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Table 3.4  The bias and the root mean squared error of the posterior mean of $d$ and also the actual coverage of the 95% credible interval based on the posterior distribution for the ARFIMA$(0, d, 0)$ plus noise model with the signal-to-noise ratio $\rho$ ($T=100$, 100 replications). The case with $\rho = 0$ corresponds to pure ARFIMA$(0, d, 0)$ misspecified as ARFIMA plus noise.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias</th>
<th>(\sigma^2_n)</th>
<th>(\sqrt{\text{MSE}})</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>$\rho = 2$</td>
<td>$\rho = 5$</td>
<td>$\rho = 10$</td>
<td>$\rho = \infty$</td>
</tr>
<tr>
<td>-0.4</td>
<td>1.690</td>
<td>0.296 0.288</td>
<td>0.283 0.273</td>
<td>0.398 0.390</td>
</tr>
<tr>
<td>-0.3</td>
<td>1.801</td>
<td>0.186 0.190</td>
<td>0.188 0.191</td>
<td>0.303 0.293</td>
</tr>
<tr>
<td>-0.2</td>
<td>1.900</td>
<td>0.163 0.202</td>
<td>0.186 0.189</td>
<td>0.221 0.207</td>
</tr>
<tr>
<td>-0.1</td>
<td>1.971</td>
<td>0.113 0.106</td>
<td>0.116 0.117</td>
<td>0.122 0.112</td>
</tr>
<tr>
<td>0.0</td>
<td>2.000</td>
<td>0.028 0.048</td>
<td>0.045 0.063</td>
<td>0.069 0.083</td>
</tr>
<tr>
<td>0.1</td>
<td>1.962</td>
<td>-0.019 0.058</td>
<td>0.016 0.052</td>
<td>0.062 0.123</td>
</tr>
<tr>
<td>0.2</td>
<td>1.820</td>
<td>0.015 0.029</td>
<td>0.040 0.057</td>
<td>0.132 0.126</td>
</tr>
<tr>
<td>0.3</td>
<td>1.519</td>
<td>0.014 0.026</td>
<td>0.016 0.043</td>
<td>0.121 0.125</td>
</tr>
<tr>
<td>0.4</td>
<td>0.966</td>
<td>0.005 0.013</td>
<td>0.043 0.030</td>
<td>0.089 0.085</td>
</tr>
</tbody>
</table>
Set \( x_0 = t_2 - t_1 \), the time interval between two adjacent observations, to allow no more than one transition in each interval. Define \( M \) to be the number of distinct values of \( y_{t_i} \) in \( Y_T \). Then the maximal number of transitions given \( Y_T \) is equal to \( M \) or \((M - 1)\) depending on \( \tau_1 < x_0 \) or \( \tau_1 \geq x_0 \).

Define

\[
\begin{align*}
    u_k &= \max\{j : y_{t_j}, y_{t_{j+1}}, \ldots, y_{t_{j+k-1}} \text{ only have } k \text{ distinct values}\}, \\
    w_{k-1} &= y_{t_{u_k}}, \\
    U_j &= (u_1, u_2, \ldots, u_j), \\
    W_j &= (w_0, w_1, w_2, \ldots, w_j), \\
    \tau_j &= (\tau_1, \tau_2, \ldots, \tau_j), \quad \text{and } \tau_0 = 0,
\end{align*}
\]

where \( u_k \) is the lag such that \( t_{u_k} \) is the last observed time before the \( k \)-th observed transition occurs, \( W_M \) or \( W_{M-1} \) is the ordered set of the distinct values of \( \{y_{t_j} : j = 1, 2, \ldots, T\} \) depending on \( \tau_1 < x_0 \) or \( \tau_1 \geq x_0 \). Let the parameter vector \( \theta = (\sigma^2, \phi, \beta)' \) have prior given by

\[
\pi(\theta) \propto 1_{\{\phi \leq 0\}} 1_{\{\beta < 0\}} \sigma^{-2} 1_{\{\sigma^2 > 0\}}.
\]

The joint density of \( Y_T, \tau_M \) and \( \theta \) is

\[
f(Y_T, \tau_M, \theta) = f(\tau_M | Y_T, \theta) f(Y_T | \theta) \pi(\theta)
\]

\[
= f(\tau_M | U_M, \theta) f(W_M | \theta) \pi(\theta),
\]

where

\[
f(\tau_M | U_M, \theta) \propto \prod_{k=2}^{M} \left\{ (\tau_k - \tau_{k-1})^{-\beta-1} I_{\{t_{u_k} < t_{u_{k+1}}\}} \right\} f(\tau_1 | \theta),
\]

\[
f(\tau_1 | \theta) = \left\{ \begin{array}{ll}
    \frac{\beta-1}{\beta x_0}; & \tau_1 \geq x_0, \\
    \frac{\beta-1}{\beta x_0} \left( \frac{x_0}{\beta} \right)^\beta; & \tau_1 < x_0,
\end{array} \right.
\]

\[
f(W_M | \theta) = \prod_{k=1}^{M} \frac{1}{\sqrt{2\pi \sigma_n^2}} \exp \left\{ \frac{-(w_k - \phi w_{k-1})^2}{2\sigma_n^2} \right\} f(w_0 | \theta),
\]

\[
f(w_0 | \theta) = \frac{\sqrt{1 - \phi^2}}{\sqrt{2\pi \sigma_n^2}} \exp \left\{ \frac{- (1 - \phi^2) w_0^2}{2\sigma_n^2} \right\}.
\]
Then the full conditional distributions can be derived as

\[
\begin{align*}
    f(\phi \mid Y_T, \tau_M, \sigma^2_n, \beta) &\propto \sqrt{1 - \phi^2} \exp \left\{ -\frac{\sum_{k=1}^{M}(w_k - \phi w_{k-1})^2 + (1 - \phi^2) w_0^2}{2\sigma^2_n} \right\} 1_{\{\phi < 1\}}, \\
    f(\sigma^2_n \mid Y_T, \tau_M, \phi, \beta) &\propto (\sigma^2_n)^{-(M+1)/2-1} \exp \left\{ -\frac{\sum_{k=1}^{M}(w_k - \phi w_{k-1})^2 + (1 - \phi^2) w_0^2}{2\sigma^2_n} \right\} 1_{\{\sigma^2_n > 0\}}, \\
    f(\beta \mid Y_T, \tau_M, \sigma^2_n, \phi) &\propto \left\{ \begin{array}{ll}
        \left( \prod_{k=2}^{M-1} \frac{\tau_k - \tau_{k-1}}{\tau_k} \right)^{\beta} & \beta \text{M-2}(\beta - 1)1_{\{1 < \beta\}}; \quad \text{if } \tau_1 \geq \tau_0, \\
        \left( \prod_{k=1}^{M} \frac{\tau_k - \tau_{k-1}}{\tau_k} \right)^{\beta} & \beta \text{M-2}(\beta - 1)1_{\{1 < \beta\}}; \quad \text{otherwise,}
    \end{array} \right. \\
    f(\tau_k \mid Y_T, \tau_M(-k), \theta) &\propto \left( \tau_k - \tau_{k-1} \right)^{-\beta-1} \left( \tau_{k+1} - \tau_k \right)^{-\beta-1} 1_{\{\tau_{k} < \tau_{k+1}\}}; \quad \text{for } k \neq 1, \\
    f(\tau_1 \mid Y_T, \tau_M(-1), \theta) &\propto \left\{ \begin{array}{ll}
        (\tau_2 - \tau_1)^{-\beta-1} 1_{\{0 < \tau_1 < \tau_0\}}; & \text{if } \tau_1 \geq \tau_0, \\
        (\tau_2 - \tau_1)^{-\beta-1} \tau_1^{-\beta} 1_{\{0 < \tau_1 < \tau_0\}}; & \text{otherwise,}
    \end{array} \right.
\end{align*}
\]

where \( \tau_M(-k) \) is the vector \( \tau_M \) without its \( k \)th element.

Posterior distributions of all unknown quantities are approximated via Gibbs sampling. The procedures are described as follows.

1. Give initial values for \( \tau_n, \beta \) and \( \sigma^2_n \).
2. Generate \( \phi \) from a rejection algorithm with a Gaussian proposal (ignore \( \sqrt{1 - \sigma^2} \) in (3.10)).
3. Generate \( \sigma^2_n \) from an inverse gamma distribution.
4. Generate \( \beta \) from a rejection algorithm. Since (3.11) is a log-concave function therefore a bounding function can be found.
5. Generate \( \tau_k \) from a rejection algorithm. Since both functions in (3.12) (3.13) are defined on bounded domains, the maximum can be found to form bounding functions.
6. Repeat steps 2–5 until the corresponding Markov chain converges.

### 3.2.2 Hierarchical Semi-Markov Processes

Assume that the observed process satisfies

\[ y_{t_j} = \alpha_{t_j} + \epsilon_{t_j}, \]
where \( \{ \alpha_t : t \in \mathbb{R} \} \) is the semi-Markov process defined in (2.3) and \( \{ \epsilon_j : j \in \mathbb{Z} \} \) are iid \( N(0, \sigma_j^2) \), independent of \( \{ \alpha_t \} \). Unlike the noiseless processes, the number of transitions in 
\((0, t_T)\) is unknown in the hierarchical case. We still set \( x_0 = t_2 - t_1 \) and further set \( M = \lceil t_T/x_0 \rceil + 1 \) which is the maximal number of transitions which might happen in 
\((0, t_T)\). The value of \( M \) is fixed through the entire analysis. Let \( \theta = (\sigma_x^2, \sigma_n^2, \phi, \beta)' \) with prior \( \pi(\theta) = \pi(\sigma_x^2)\pi(\sigma_n^2)\pi(\phi, \beta) \), where \( \pi(\phi, \beta) \) is defined as before and \( \pi(\sigma_x^2) \) and \( \pi(\sigma_n^2) \) are inverse gamma distributions satisfying
\[
\pi(\sigma_x^2) \propto (\sigma_x^2)^{-2} e^{-1/\sigma_x^2},
\]
\[
\pi(\sigma_n^2) \propto (\sigma_n^2)^{-2} e^{-1/\sigma_n^2}.
\]
Since the states \( W_M \) and the sojourns \( \tau_M \) are both unobservable, we consider the following joint density
\[
f(Y_T, W_M, \tau_M, \theta) = f(Y_T|W_M, \tau_M, \theta)f(W_M, \tau_M, \theta)\pi(\theta)
\]
\[
= \left\{ \prod_{k=1}^M f(w_k, \tau_k|w_{k-1}, \tau_{k-1}, \theta) \right\} \left\{ \prod_{j=1}^{T} f(y_{t_j}|w_{N(t_j)}, \theta) \right\} \pi(\theta),
\]
where
\[
f(y_{t_j}|w_{N(t_j)}, \theta) = \frac{1}{\sqrt{2\pi \sigma_x^2}} \exp \left\{ - \frac{(y_{t_j} - w_{N(t_j)})^2}{2\sigma_x^2} \right\},
\]
\[
f(w_k, \tau_k|w_{k-1}, \tau_{k-1}, \theta) = f(w_k|w_{k-1}, \theta)f(\tau_k|\tau_{k-1}, \theta)
\]
\[
= \frac{1}{\sqrt{2\pi \sigma_n^2}} \exp \left\{ - \frac{(w_k - \phi w_{k-1})^2}{2\sigma_n^2} \right\} \frac{\beta x_0^\beta}{(\tau_k - \tau_{k-1})^{\beta+1}} 1(\tau_k - \tau_{k-1} > x_0),
\]
for \( k \neq 1 \). The conditional distributions for \( \phi, \sigma_n^2 \) and \( \beta \) are essentially the same as before.

Other conditional distributions are derived as follows.
\[
f(\sigma_x^2|Y_T, W_M, \tau_M, \theta) \propto (\sigma_x^2)^{-T/2-2} \exp \left\{ \frac{1}{2\sigma_x^2} \sum_{j=1}^{T} (y_{t_j} - w_{N(t_j)})^2 \right\} 1(\sigma_x^2 > 0),
\]
\[
f(w_k|Y_T, W_M(-k), \tau_M, \theta) \propto \exp \left\{ - \frac{1}{2\sigma_x^2} \sum_{j=1}^{M} (y_{t_j} - w_k)^2 1(N(t_j) = k) \right\}
\times \exp \left\{ - \frac{(w_k - \phi w_{k-1})^2 + (w_{k+1} - \phi w_k)^2}{2\sigma_n^2} \right\}, \text{ for } k \neq 1,
\]
\[
f(\tau_k|Y_T, W_M, \tau_M(-k), \theta) \propto \left\{ \prod_{j=1}^{T} f(y_{t_j}|w_{N(t_j)}) 1(N(t_j) = k, k-1) \right\} (\tau_k - \tau_{k-1})^{-\beta-1}
\times (\tau_{k+1} - \tau_k)^{-\beta-1} 1(\tau_{k+1} - x_0 < \tau_k < \tau_{k+1} - x_0),
\]
where $W_{M(-k)}$ is the vectors $W_M$ without its $k$th element, $\theta_{(-\sigma_2^2)}$ is the vector of $\theta$ without the element $\sigma_2^2$, and so on. The conditional densities for $w_1$ and $\tau_1$ require modification.

In the Gibbs sampling, $\sigma_2^2$ can be generated from an inverse gamma distribution and $w_k$ can be generated from a Gaussian distribution. The conditional density of $\tau_k$ is a piecewise convex function defined on a finite interval. A bounding function can be constructed by connecting the two ends of each convex function for all pieces. Hence, a rejection algorithm can be used.

A similar algorithm can be applied directly to other hierarchical semi-Markov processes with the form: $g_1(y_t) = g_2(\alpha_t)g_3(Y_{t-1}) + g_4(Y_{t-1})\epsilon_t$, e.g., stochastic volatility model associated with a long-memory semi-Markov process in volatilities (Hsu and Breidt, 1997) or the random coefficient model.

One drawback for regime switching models is that, in the estimating procedure, it takes a long time to achieve convergence because $\{\tau_n\}$ mixes slowly in the Markov chain sampler. Therefore, as the number of observations increases the implementation becomes inefficient due to the slow convergence.
4 MODEL CHECKING AND DIAGNOSTICS

4.1 Long Memory Feature Checking

As mentioned in the introduction, there are several heuristic methods for detecting the existence of long memory features. One is to calculate the sample autocovariances to see how slowly the function decays. The other tool is to calculate the rescaled range statistic \( R/S \) defined as \( R_T/s_T \) with

\[
R_T = \max_{1 \leq i \leq T} \left\{ \sum_{j=i}^{t} (y_j - \bar{y}_T) \right\} - \min_{1 \leq i \leq T} \left\{ \sum_{j=i}^{t} (y_j - \bar{y}_T) \right\},
\]

\[
s_T = \left( \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y}_T)^2 \right)^{1/2},
\]

where \( \bar{y}_T = \frac{1}{T} \sum_{t=1}^{T} y_t \) is the sample mean and \( s_T \) is the sample standard deviation. For sufficiently large values of \( T \),

\[
\log E(R_T/s_T) \approx a + \delta \log T, \text{ for some } \delta \in (1/2, 1),
\]

which is proved by Mandelbrot (1975). Therefore, if the slope is greater than 1/2 in the log-log plot of \( R/S \) versus \( T \), the long memory feature might exist. The other useful graphical tool for long memory checking is the log-log plot of \( \text{var}(\bar{y}_T) \) versus \( T \). For long memory processes, \( \text{var}(\bar{y}_T)=O(T^{\delta-1}) \) for some \( \delta \in (0, 1) \) which is larger than \( O(T^{-1}) \) in short memory cases. Therefore, the slope greater than \( -1 \) in the log-log plot of \( \text{var}(\bar{y}_T) \) indicates long memory.

In addition, the log-log plot of the periodogram ordinates versus the Fourier frequencies and the log-log plot of the sample autocovariances versus the time lags are also useful for detecting long memory.
4.2 Goodness of Fit

Given a model, the goodness of fit can be examined by comparing the observed data to the posterior predictive distribution via simulation. Simply speaking, the idea is to simulate many replications (with the equal sample size as the observed data) from the assumed model with either known or estimated parameters. The characteristics of the observed data are compared to those of the simulated replications with respect to some chosen statistics in order to detect differences between the given model and the data set. The existence of discrepancies indicates model failure.

In our application, we use several different statistics to check the goodness of fit, including the sample autocovariances, sample kurtosis and skewness and prediction mean squared errors.

The autocovariances are useful for checking the consistency of dependence structures between the assumed model and the observed data. Hosking (1996) shows that the sample autocovariances and autocorrelations under ARFIMA have a negative asymptotic bias, the magnitude of which depends on the long memory parameter, that is \( Ec(h) - \gamma(h) \sim O(T^{2d-1}) \) and \( Er(h) - \rho(h) \sim O(T^{2d-1}) \) as \( T \to \infty \) where \( d \) is the fractional differencing parameter. \( \gamma(h) \) and \( \rho(h) \) (\( c(h) \) and \( r(h) \)) are the (sample) autocovariance and the (sample) autocorrelation at lag \( h \), respectively. Attempts to adjust for this bias may be difficult due to parameter estimation uncertainty and finite-sample deviations from the asymptotic values. Posterior predictive checking provides another way of assessing the fit of the autocovariance structure, which handles parameter estimation uncertainty without appealing to asymptotic arguments.

Besides the long memory feature, the kurtosis and skewness are useful in checking the normality assumption. Prediction mean squared errors are checked to look for important failures in modeling the short memory dependence structure.
5 APPLICATIONS

We illustrate the methodology by fitting a long-memory model to the famous Nile River data, and long-memory stochastic volatility models to the Center for Research in Securities Prices value-weighted market index data, using both an ARFIMA and a semi-Markov process.

5.1 Nile River Data

The Nile River data are yearly minimum water levels as measured at the Roda Gauge near Cairo (Beran, 1994). The data set, which contains 663 observations for the years 622AD to 1284AD, is displayed in Figure 5.1. In Figure 5.2, the log-log plot of the sample autocorrelations versus time lags shows the slow decay suggestive of long memory. In Figure 5.3 and Figure 5.4, the estimated slope is 0.985 in the log-log plot of the rescaled range statistic \( R/S \) and the decreasing tendency appears near zero frequency in the log-log plot of the periodogram ordinates versus frequencies. Both strongly suggests long memory. However, in Figure 5.5, the slope is estimated as -0.991 in the log-log plot of \( \text{var}(\hat{y}_T) \) versus \( T \) shows no evidence of long memory.

Table 5.1 The posterior mean and quantiles of parameters in ARFIMA\((0,d,0)\) model for the standardized Nile River minimum water levels. The values in parentheses are the posterior standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior Mean 2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>0.394 (0.030)</td>
<td>0.334</td>
<td>0.373</td>
<td>0.393</td>
<td>0.416</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.626 (0.036)</td>
<td>0.563</td>
<td>0.601</td>
<td>0.624</td>
<td>0.648</td>
</tr>
</tbody>
</table>
Figure 5.1 The yearly minimum water levels of the Nile River at the Roda Gauge for the years 622AD to 1284AD.

We considered an ARFIMA(0,d,0) model for the standardized data (the data are centered about the sample mean and divided by the sample standard deviation). Five parallel Markov chains with widely spread initial values were run independently in which the Metropolis step for generating $d$ is repeated 50 times to complete one iteration. The convergence was checked by monitoring the values of $\sqrt{\hat{R}}$ (Gelman and Rubin, 1991) based on the five independent chains. As shown in Figure 5.6, the values converge to one within 200 iterations to suggest the Markov chain converges extremely fast. Figure 5.7 shows that the draws within one chain are nearly uncorrelated, indicating good mixing in the Metropolis-Hastings algorithm. The final posterior inferences were made based on a total number of 2500 draws from the five Markov chains after the first 500 iterations. The posterior means, standard deviations and quantiles for $d$ and $\sigma^2_\eta$ are summarized in Table 5.1. These results are consistent with the results obtained by Beran (1994, §6.1), in which the estimate of $d$ from Whittle's approximation is 0.40 with the 95% confidence interval (0.34, 0.46). The estimated marginal posterior distributions are displayed in Figures 5.8 and 5.9.
Figure 5.2 The log-log plot of the sample autocorrelations versus time lags for the Nile River data. The solid line is the least squares fit with slope $-0.533$ which is smaller than $-1$ for the short-memory processes.

Figure 5.3 The log-log plot of the rescaled range statistic $R/S$ versus $T$ for the Nile River data. The solid line is the least squares fit with slope $0.985$ and the dotted line is the reference line with slope $0.5$ for short-memory processes.
Figure 5.4  The log-log plot of the periodogram ordinates versus frequencies for the Nile River data. The decreasing tendency near zero frequency shows long memory.

Figure 5.5  The log-log plot of $\text{var}(\hat{y}_T)$ versus $T$ for the Nile River data. The solid line is the least squares fit with slope $-0.991$ and the dotted line is the reference line with slope $-1$ for short-memory processes.
Figure 5.6 The values of $\hat{H}^{1/2}$ for $d$ and $\sigma^2_n$ based on five independent Markov chains with length $2n$.

Figure 5.7 The autocorrelations between draws in the Metropolis-Hastings algorithm for $d$ (denoted by o) and $\sigma^2_n$ (denoted by x). Horizontal reference lines are the Bartlett bounds $\pm 1.96T^{-1/2}$ for testing significance of estimated autocorrelations.
Figure 5.8 The estimated marginal posterior density of the parameter $d$ (solid line) in the ARFIMA(0, $d$, 0) model for the Nile River data. The dotted line indicates the analytic result.

Figure 5.9 The estimated marginal posterior density of the parameter $\sigma_n^2$ (solid line) in the ARFIMA(0, $d$, 0) model for the Nile River data. The dotted line indicates the analytic result.
Under the ARFIMA(0,d,0) model, the goodness of fit is checked by comparing the sample autocorrelations, the sample kurtosis and the prediction mean squared errors for the Nile River data to the corresponding statistics for simulated data. The results are summarized in Figure 5.10-5.12. In Figure 5.10, the 90% and 95% pointwise posterior prediction band for the sample autocorrelations is calculated based on 2500 simulated replications (each has sample size equal to 663) and superimposed on the sample autocorrelations of the real data. The sample autocorrelations are inside both bands. The same simulated datasets are used for other statistics. In Figure 5.11, the kurtosis based on the Nile River data is equal to 3.11. Comparing to the predictive distribution of the kurtosis based on the simulated replications, the probability that the simulated data could be more extreme than the observed data is \( p = 0.20 \). The predictions were checked for another hundred years (1285AD-1400AD). These data are actually recorded (Tousson, 1925) except six data points are missing. Combining these extra data, the total of prediction squared errors for the years 1286AD to 1400AD based on the real data is equal to 113.03. Comparing to its posterior predictive distribution based on the simulated data under the model, displayed in Figure 5.12 (in which the probability for values greater than 500 is shown in one category), the probability that the simulated data could be more extreme than the observed data is \( p = 0.52 \). All three diagnostics show no evidence of model failure.

The performance of the ARMA approximation can be assessed in two ways. First, the importance weights \( \{c_j : j = 1, \ldots, K\} \) would be equal to one if the approximating process was exactly equal to the true process. Thus, the quantity \( c_j^* \equiv \log c_j - \log \sum c_j + \log K \) should be near zero for good approximation. In this example, the histogram of the relocated log importance weights \( \{c_j^*\} \), displayed in Figure 5.13, is centered around zero (slightly to the left because of the log transformation) and has only few extremes.

Second, for this two-parameter model, we can actually calculate the joint posterior analytically under the ARFIMA(0,d,0) model and compare with the joint posterior under the approximating ARMA(2,2) model. We evaluate the joint density \( f(Y_T, \theta) \) on a grid of \( (d, \sigma_n^2) \) under each model. After the appropriate normalization, the contour plots for the posterior...
Figure 5.10  The sample autocorrelations of the Nile River minimum water levels. The dotted lines and the dashed lines are the 90% and 95% pointwise prediction bands based on 2500 simulated datasets under the estimated ARFIMA(0,d,0) model.

Figure 5.11  The kurtosis based on the Nile River data compared to the predictive distribution of those based on 2500 simulated datasets under the estimated ARFIMA(0,d,0) model.
Figure 5.12 The total of prediction mean squared errors based on the Nile River data for the years 1285AD to 1400AD compared to the predictive distribution of those based on 2500 simulated datasets under the estimated ARFIMA(0,d,0) model.

densities under each model are displayed in Figure 5.14. From Figure 5.14, the two joint posteriors have very similar shape. But the posterior based on the approximating ARMA(2,2) model is less flat and the parameter space supporting most of the probability is slightly shifted to the left with respect to the ARFIMA(0,d,0) model. However, the marginal posteriors after the application of importance sampling, displayed in Figure 5.8 and 5.9, are shifted back to the correct location.

Similarly, the posterior distributions can be computed for the truncated AR(p) approximations. The contour plots of the corresponding posteriors under several choices of p for the Nile River data are also shown in Figure 5.14. Clearly, the shape of the posterior densities are destroyed. Although the truncated AR(p) model converges to the ARFIMA(0,d,0) as $p \to \infty$ for any fixed $d$, the convergence rates are substantially different for different values of $d$. That is why the corresponding contour plot is extremely distorted for every choice of fixed $p$. Thus, the ARMA(2,2) approximation performs extremely well relative to the truncated AR approximations, even for very large $p$. 
5.2 Long-Memory Stochastic Volatility in Stock Returns

The data are the daily returns (first differences of log prices) for the Center for Research in Securities Prices (CRSP) value-weighted market index starting on the first trading day of July 1962 and ending on the last trading day of December 1987. For illustration purposes, day of the week and month of the year effects and serial correlation in the mean, which account for a very small proportion of the overall variation in the data, have been removed by regressing on dummies and filtering with an AR(1). The resulting 6408 filtered observations, denoted \( \{y_t\} \), are displayed in Figure 5.15. The sample autocorrelations of \( \{\log y_t^2\} \) in Figure 5.16 show evidence of slowly-decaying nonlinear dependence.

One way to describe this dependence is with a long-memory stochastic volatility model (Breidt, Crato and de Lima, 1994):

\[
y_t = \sigma \exp(\alpha_t/2) \epsilon_t,
\]

where \( \{\epsilon_t\} \) are iid \( N(0,1) \) and \( \{\alpha_t\} \) is a long-memory process. We consider two types of models
Figure 5.14 The contour plot of the joint posterior density for the Nile River data under the ARFIMA(0,d,0) model, the approximating ARMA(2,2) model and the truncated AR(p) models.
Figure 5.15  The daily returns data for the value-weighted CRSP market index from July 1962 to December 1987. The smoothed volatilities from the SV model incorporating an ARFIMA are superimposed on the CRSP data. The observation \( y_{6355} = -0.17 \), which corresponds to the October 1987 market crash, has been truncated.

for \( \{\alpha_t\} \): one is an ARFIMA related model and the other is a regime switching model.

5.2.1  ARFIMA and Related Models

Assume

\[
(1 - B)^d \alpha_t = \eta_t,
\]

where \( \{\eta_t\} \) are iid \( N(0, \sigma^2) \) independent of \( \{\varepsilon_t\} \). After the log squared transformation, the series has the linear representation

\[
z_t \equiv \log y_t = \mu + \alpha_t + \varepsilon_t^*,
\]

where \( \mu = \log \sigma^2 + E(\log \varepsilon_t^2) \) and \( \varepsilon_t^* = \log \varepsilon_t^2 - E(\log \varepsilon_t^2) \). The distribution of \( \varepsilon_t^* \) is a centered \( \log \chi^2_1 \) \( (E\varepsilon_t^* = 0) \) which can be accurately approximated by a given mixture Gaussian distribution with seven components (Kim, Shephard and Chib, 1996) shown on Figure 5.17. By
Figure 5.16 The sample autocorrelations of the log squared value-weighted CRSP data. The fitted autocorrelations (solid line) under the SV model incorporated with an ARFIMA(0,d,0) are superimposed. The dotted lines and dashed lines are the 90% and the 95% pointwise posterior prediction confidence bands, respectively.

Figure 5.17 The densities of $\log \chi^2$ and the mixture Gaussian approximation.
treating the mixture Gaussian distribution as the exact distribution for \( \epsilon_i^* \), the transformed process \( \{ z_t \} \) become an ARFIMA(0, \( d \), 0) with additive mixture Gaussian noise. Define the vector of indicators \( w_t = (w_{t1}, w_{t2}, \ldots, w_{tr})' \) where \( w_{tk} \) is one if the \( k \)-th component is selected in the mixture distribution and is zero otherwise. As a result, \( w_t \) follows a multinomial distribution. To modify the algorithm for the mixture noise, the set of independent latent vectors \( \{ w_1, w_2, \ldots, w_T \} \) was introduced to indicate which Gaussian component in the mixture distribution was selected. Given \( \{ w_t \} \), the algorithm in Section 3.1.3.1 can be applied directly and the extra vectors \( \{ w_t : t = 1, 2, \ldots, T \} \) can be drawn independently from the corresponding \( f(w_t | Z_t, \alpha_t, \theta) \) which is a multinomial distribution for each \( t \).

In this application, we actually use \( \theta = (d, \sigma^2_\theta, \mu_0)' \) where \( \mu_0 = (1 + b_1(d) + b_2(d))^{-1} \mu \) rather than other parameterizations because it produces a Markov chain with better mixing in the Metropolis-Hastings algorithm. The importance weight \( c(\theta_j) \) is proportional to \( p_0(\alpha_T | \theta_j) / p_1(\alpha_T | \theta_j) \) where \( \alpha_t = (\alpha_1, \alpha_2, \ldots, \alpha_t)' \), \( p_0 \) and \( p_1 \) are the sampling distributions under ARFIMA(0, \( d \), 0) and the approximating ARMA(2,2) respectively. For these high dimensional data, the values of \( p_0(\alpha_T | \theta_j) \) can be approximated by truncating the AR(\( \infty \)) representation of an ARFIMA(0, \( d, 0 \)) process in Equation (2.2), thus avoiding the direct inversion of a 6408-dimensional covariance matrix. That is, we approximate

\[
p_0(\alpha_{6408} | \theta) = p_0(\alpha_{1000} | \theta) \prod_{t=1001}^{6408} p_0(\alpha_t | \alpha_{t-1}, \theta),
\]

by

\[
p_0(\alpha_{1000} | \theta) \prod_{t=1001}^{6408} (2\pi \sigma_n^2)^{-1/2} \exp \left\{ -\frac{1}{2} \left( \sum_{j=0}^{t-1} \pi_j \epsilon_{t-j} \right)^2 \right\}.
\]

As in the first example, the convergence of the Markov chain was checked which was achieved much slower than the situation in the first example because of the high-dimensional parameter space. Finally, the inference was made based on a total number of 5000 draws from five independent chains after the first 1500 iterations. The results are shown in Table 5.2. The smoothed volatilities, which are the posterior means of \( \{ \sigma \exp(\alpha_t/2) \} \), are superimposed on the value-weighted CRSP data in Figure 5.15.
Table 5.2  The posterior mean and quantiles of parameters in the SV model incorporated with an ARFIMA(0,d,0) for the value-weighted CRSP market index. The values in parentheses are the posterior standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d)</td>
<td>0.466 (0.010)</td>
<td>0.447</td>
<td>0.460</td>
<td>0.462</td>
<td>0.480</td>
<td>0.480</td>
</tr>
<tr>
<td>(\sigma_d^2)</td>
<td>0.267 (0.018)</td>
<td>0.237</td>
<td>0.267</td>
<td>0.269</td>
<td>0.272</td>
<td>0.309</td>
</tr>
<tr>
<td>(\sigma \times 10^3)</td>
<td>6.731 (0.786)</td>
<td>5.319</td>
<td>6.076</td>
<td>6.448</td>
<td>7.663</td>
<td>7.968</td>
</tr>
</tbody>
</table>

Similarly, the goodness of fit of the assumed model is checked by comparing the sample autocorrelations of \(\{\log y_t^2\}\) based on the observed data and 5000 simulated replications from the assumed model. The 90% and 95% pointwise prediction bands and the posterior means for the sample autocorrelations based on the simulation is superimposed on the actual sample autocorrelations shown in Figure 5.16. Obviously, the sample autocorrelations based on the observed data are larger than those expected from the assumed model, in particular at small lags. This indicates that more short memory structure should be added.

Therefore, we consider an ARFIMA(0,d,0) with additive AR(1) model for \(\{\alpha_t\}\) to increase the short range dependence. The model is then

\[
\alpha_t = \alpha_{1t} + \alpha_{2t},
\]

\[
(1 - B)^d \alpha_{1t} = \eta_t,
\]

\[
(1 - \phi B) \alpha_{2t} = \eta_{\phi t},
\]

where \(\{\eta_t\} \sim N(0,\sigma_\eta^2), \{\eta_{\phi t}\} \sim N(0,\sigma_\phi^2)\) is independent of \(\{\eta_t\}\). The resulting inference is summarized in Table 5.3. The values of \(d\) are slightly smaller than the first assumed model, however \(\phi\) is very close to one providing another source of persistence for the short range dependence.

Similar diagnostics for model checking based on the sample autocorrelations are done and displayed in Figure 5.18, which shows that this model has better fit than the pure ARFIMA(0,d,0) model, in particular for the short range dependence. The kurtosis and the
Table 5.3 The posterior mean and quantiles of parameters in the SV model incorporated with an ARFIMA(0,d,0) plus AR(1) for the value-weighted CRSP market index. The values in parentheses are the posterior standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>Quantile 2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>0.468 (0.010)</td>
<td>0.456</td>
<td>0.459</td>
<td>0.462</td>
<td>0.478</td>
<td>0.479</td>
</tr>
<tr>
<td>φ</td>
<td>0.905 (0.016)</td>
<td>0.874</td>
<td>0.893</td>
<td>0.905</td>
<td>0.917</td>
<td>0.934</td>
</tr>
<tr>
<td>σ²</td>
<td>0.247 (0.006)</td>
<td>0.227</td>
<td>0.243</td>
<td>0.243</td>
<td>0.250</td>
<td>0.256</td>
</tr>
<tr>
<td>σ²φ</td>
<td>0.056 (0.006)</td>
<td>0.045</td>
<td>0.051</td>
<td>0.055</td>
<td>0.060</td>
<td>0.069</td>
</tr>
<tr>
<td>σ × 10³</td>
<td>6.412 (0.437)</td>
<td>5.465</td>
<td>6.120</td>
<td>6.397</td>
<td>6.687</td>
<td>7.300</td>
</tr>
</tbody>
</table>

Skewness of \{\log y_t\} are checked and summarized in Figure 5.19 and Figure 5.20. Both observed values are reasonable under the assumed model.

Clearly, the autocorrelations are underestimated under either model. One possible reason is that the variance \(\sigma_N^2\) tends to be underestimated (as we found in the simulation study). Another explanation is that the process is not stationary, then non-stationary models (e.g., random walk) might provide better fit.

5.2.2 Regime Switching Semi-Markov Models

Assume

\[ \alpha_t = w_{N(t)}, \]

which defined in (2.4) with AR(1) states and Pareto sojourns as described in Example 2.3 in Section 2.2.2. Only the first 1000 observations of the CRSP data are used.

The algorithm described in Section 3.2.2 was used consisting 3000 iterations in Gibbs sampler. By comparing several independent chains, we decided the convergence is achieved after 1000 iterations and used the rest of the chain to estimate the posteriors. Posterior means and quantiles are shown in Table 5.4 and the smoothed volatilities for this model are superimposed on the observed data shown in Figure 5.21.
5.2.3 Comparisons and Summary

The posteriors for the long-memory characteristic, $d$ in Table 5.2 and Table 5.3, $1 - \beta/2$ in Table 5.4 are all massed far away from zero, showing strong posterior evidence that the volatility of the CRSP index has long memory from both model specifications. The smoothed volatilities in Figure 5.15 for the ARFIMA model seem to track even small variance changes. In Figure 5.21, on the other hand, the volatilities are relatively smooth and some regimes can be clearly identified.
Figure 5.19 The kurtosis based on the log squared value-weighted CRSP compared to the predictive distribution of those based on simulated datasets under ARFIMA(0, d, 0) plus AR(1) model.

Figure 5.20 The skewness based on the log squared value-weighted CRSP compared to the predictive distribution of those based on simulated datasets under ARFIMA(0, d, 0) plus AR(1) model.
Table 5.4  The posterior mean and quantiles of parameters in the SV model incorporated with a semi-Markov process for the value-weighted CRSP market index ($T = 1000$). The values in parentheses are the posterior standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior Mean</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.046 (0.032)</td>
<td>1.006</td>
<td>1.022</td>
<td>1.038</td>
<td>1.062</td>
<td>1.125</td>
</tr>
<tr>
<td>$1 - \beta/2$</td>
<td>0.477 (0.016)</td>
<td>0.438</td>
<td>0.469</td>
<td>0.481</td>
<td>0.489</td>
<td>0.497</td>
</tr>
<tr>
<td>$\phi$</td>
<td>0.131 (0.291)</td>
<td>-0.469</td>
<td>-0.073</td>
<td>0.153</td>
<td>0.350</td>
<td>0.637</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1.561 (0.501)</td>
<td>0.813</td>
<td>1.198</td>
<td>1.476</td>
<td>1.848</td>
<td>2.735</td>
</tr>
<tr>
<td>$\sigma \times 10^3$</td>
<td>4.366 (0.692)</td>
<td>3.162</td>
<td>3.873</td>
<td>4.243</td>
<td>4.796</td>
<td>5.831</td>
</tr>
</tbody>
</table>

Figure 5.21  The smoothed volatilities from the stochastic volatility model incorporating a long-memory semi-Markov process, superimposed on the value-weighted CRSP data for $T = 1000$. 
6 CONCLUSIONS

6.1 Summary

The long-memory feature of data has been well-known in many sciences, especially in the physical sciences and increasingly in economics and finance. Many useful models have been proposed over the years. For more complicated models (e.g., hierarchical long-memory models) however, most existing estimation procedures (e.g., MLE type estimations) involve intensive computations due to the complexity of the underlying covariance structure and also suffer from convergence problem. Therefore, searching for more feasible and time-efficient estimation procedures is necessary and important for practical applications.

In this thesis, a new Bayesian approach is proposed to make inference for ARFIMA related long-memory processes. An ARFIMA\((p, d, q)\) process is approximated by an ARMA\((p+2, q+2)\) process in a Metropolis-Hastings algorithm, and importance sampling is used to correct for the approximation. Moreover, the approximation can always be improved by increasing the order of \(p\) and \(q\) for the ARMA process. Compared to the previous Bayesian approach (Pai and Ravishanker, 1996), this new algorithm is fast since most of the complicated computations are eliminated. In addition, the Markov chain in the Metropolis-Hastings algorithm achieves its equilibrium faster in our approach since the problem has been simplified to a short-memory ARMA process.

The other advantage of our method is that the algorithm allows for a number of model extensions, such as the case of ARFIMA\((0, d, 0)\) plus short-memory AR\((p)\) noise. In particular, it handles the case of additive noise, which need not be Gaussian (e.g., mixture Gaussian in the second example of Chapter 5). This allows for the treatment of additive outlier models as well as the transformed stochastic volatility model. The same idea of the approximation can
also be applied to multivariate ARFIMA(0, d, 0) processes.

Since we are also interested in the performance of our estimation method under repeated samples, the methods of frequentist statistics are considered to justify the resulting Bayesian inference. Under the frequentist criteria of bias and MSE over repeated sampling, the posterior means are found via a simulation study to be competitive with exact MLEs for small samples of ARFIMA(0, d, 0), ARFIMA(1, d, 0) and ARFIMA(0, d, 1) processes. The method also has good ability of prediction, producing point predictors with prediction MSEs close to the best possible with known parameters.

Another contribution in this thesis is the study of the second order properties for the class of semi-Markov processes. The general forms of the autocovariance function and the spectral density associated with a stationary semi-Markov process are derived. The memory property of these processes, which only depends on the tail behavior of the distribution of sojourn times, is also investigated. In particular, a class of semi-Markov long-memory processes constructed by a discrete-time AR process for the states and a Pareto distribution for the sojourns is studied. A Markov chain Monte Carlo approach to Bayesian inference for this particular class and its related processes is also provided.

Another class of short-memory semi-Markov processes constructed by an AR process for the states and an exponential distribution for the sojourns is also discussed. This class has a close relation to the continuous-time ARMA processes in terms of similar second order properties. In addition, the embedding issue is discussed for this class of semi-Markov processes relating to their underlying AR processes.

6.2 Further Research

The estimation methodology described here can be generalized to multivariate settings. The generalization for the multivariate ARFIMA(0, d, 0) with diagonal form is straightforward. The extension for the non-diagonal situation is still under investigation.

Another interesting application of our method is to the common long-memory component models. In unpublished work, Ray and Tsay (1997) have proposed test procedures for identify-
ing multiple time series with common long-memory components, but the question of estimating that component is open.

Several modeling issues warrant further investigation. In particular, long-memory models with time-varying long-memory characteristics might be of interest. For example, the differencing parameter itself could be modeled as a stochastic component in an ARFIMA(0, d, 0). Since d should be inside the range (−1/2, 1/2) for a causal invertible ARFIMA(0, d, 0), there will be constraints on parameters which will need to be investigated to ensure the stationarity of the resulting process. It is also interesting to study what features can be produced under this model and its relation to the random coefficient models with long-memory coefficients.

Another modeling issue is to compare the multiplicative long-memory models and the additive long-memory models which are explained as follows. A multiplicative model is constructed through composition of the short memory operators (i.e., AR filter and MA filter) and the fractional differencing operator, e.g., ARFIMA(p, d, q) models. Under this multiplicative construction, the log spectral density of the resulting process is a weighted sum of the log spectral densities of the AR component, MA component and the fractional differencing component. However, its autocovariance function is quite complicated to derive due to the non-trivial operator compositions. In contrast, it might be more natural to consider additive models which consist of short-memory components and the fractionally integrated noise component. Under this additive construction, the resulting autocovariance function is simply the sum of the autocovariance functions of the components. Therefore, the relation between the constructed process and its building blocks is established through the dependence structure in the time domain directly instead of through the spectral density in the frequency domain as in the multiplicative cases. Both constructions can produce a variety of dependence structures. Their common features and individual specialties are worthy of further investigation in order to provide more strategies for modeling long-memory processes.
APPENDIX A  PROOFS

Proof of Theorem 2.1: Under Assumption 2.1 (A1),

\[ E\alpha_t = Ew_{N(t)} = 0, \]

\[ E(\alpha_s \alpha_{s+t}) = E\mathbb{E}[w_{N(s)}w_{N(s+t)}] = \sum_{h=0}^{\infty} \sum_{k=0}^{\infty} E(w_k w_{k+h}) P[N(s) = k, N(s+t) = k+h] = \sum_{h=0}^{\infty} \gamma_w(h) \left\{ \sum_{k=0}^{\infty} P[N(s) = k, N(s+t) = k+h] \right\}, \]

where

\[ \sum_{k=0}^{\infty} P[N(s) = k, N(s+t) = k+h] = \sum_{k=0}^{\infty} P[N(s+t) = k+h | N(s) = k] P[N(s) = k] = \sum_{k=0}^{\infty} P[B(s) + \sum_{i=k+2}^{k+h} T_i \leq t < B(s) + \sum_{i=k+2}^{k+h+1} T_i] P[N(s) = k] = \sum_{k=0}^{\infty} P[B(s) + S_{h-1} \leq t < B(s) + S_h] P[N(s) = k] = P[B(s) + S_{h-1} \leq t < B(s) + S_h] \left\{ \sum_{k=0}^{\infty} P[N(s) = k] \right\} = P[B(s) + S_{h-1} \leq t < B(s) + S_h], \]

and \( S_h \) is distributed as the \( h \)-th convolution of \( F \) and is independent of \( B(s) \). Under Assumption 2.1 (A2) and (A3), the distribution of \( B(s) \) is \( G \) which is independent of \( s \) (Lemma 2.1) and therefore \( E(\alpha_s \alpha_{s+t}) \) is also independent of \( s \). Hence, the process \( \{\alpha_t\} \) is weakly stationary with the autocovariance function

\[ \gamma_\alpha(t) = E(\alpha_s \alpha_{s+t}) \]
Moreover, if the process \( \{w_n\} \) is uncorrelated, then

\[
\gamma_\alpha(t) = \gamma_\alpha(0) P[N(t) = 0] = \gamma_\alpha(0) P[\tau_1 > t] = \gamma_\alpha(0)(1 - G(t)).
\]

Proof of Theorem 2.2: Under Assumption 2.1 \((A1')\), \( E\alpha_t = Ew_{N(t)} = 0 \). For any \( k \in \mathbb{Z}^+ \), the joint characteristic function of \((\alpha_{t_1}, \alpha_{t_2}, \ldots, \alpha_{t_k})\) is defined as

\[
\phi_{\alpha_{t_1}, \alpha_{t_2}, \ldots, \alpha_{t_k}}(\lambda_1, \lambda_2, \ldots, \lambda_k)
\]

\[
= E \exp \{i(\lambda_1 \alpha_{t_1} + \lambda_2 \alpha_{t_2} + \cdots + \lambda_k \alpha_{t_k})\}
\]

\[
= EE \left\{ \exp \{i(\lambda_1 w_{N(t_1)} + \lambda_2 w_{N(t_2)} + \cdots + \lambda_k w_{N(t_k)})\} \right\} \left\{ N(t_1), N(t_2), \ldots, N(t_k) \right\}
\]

\[
= \sum_{h_k = 0}^{\infty} \sum_{h_{k-1} = 0}^{h_k} \cdots \sum_{h_1 = 0}^{h_2} E \left\{ \exp \{i(\lambda_1 w_{h_1} + \lambda_2 w_{h_1+h_2} + \cdots + \lambda_k w_{h_1+h_2+\cdots+h_k})\} \right\}
\]

\[
\times P[N(t_1) = h_1, N(t_2) = h_1 + h_2, \ldots, N(t_k) = h_1 + h_2 + \cdots + h_k].
\]

Since \( \{w_n\} \) is strictly stationary from \((A1')\), the joint characteristic function \( \phi_{\alpha_{t_1}, \alpha_{t_2}, \ldots, \alpha_{t_k}}^{w}(\lambda_1, \lambda_2, \ldots, \lambda_k) \) of \((w_{h_1}, w_{h_1+h_2}, \ldots, w_{h_1+h_2+\cdots+h_k})'\) is independent of \( h_1 \). Then the previous equation can be simplified as

\[
\phi_{\alpha_{t_1}, \alpha_{t_2}, \ldots, \alpha_{t_k}}^{w}(\lambda_1, \lambda_2, \ldots, \lambda_k)
\]

\[
= \sum_{h_k = 0}^{\infty} \sum_{h_{k-1} = 0}^{h_k} \cdots \sum_{h_2 = 0}^{h_1} \phi_{0, h_2, h_3, \ldots, h_2 + \cdots + h_k}^{w}(\lambda_1, \lambda_2, \ldots, \lambda_k)
\]

\[
\times \left\{ \sum_{h_1 = 0}^{\infty} P[N(t_1) = h_1, N(t_2) = h_1 + h_2, \ldots, N(t_k) = h_1 + h_2 + \cdots + h_k] \right\}
\]

\[
= \sum_{h_k = 0}^{\infty} \sum_{h_{k-1} = 0}^{h_k} \cdots \sum_{h_2 = 0}^{h_1} \phi_{0, h_2, h_3, \ldots, h_2 + \cdots + h_k}^{w}(\lambda_1, \lambda_2, \ldots, \lambda_k)
\]

\[
\times \left\{ \prod_{i=2}^{k} P[B(t_{i-1}) + S_{h_{i-1}} \leq t_i - t_{i-1} < B(t_{i-1}) + S_{h_i}] \right\}
\]
\[ \times \left\{ \sum_{h_1=0}^{\infty} P[N(t_1) = h_1] \right\} \]
\[ = \sum_{h_2=0}^{\infty} \sum_{h_2+h_3=\cdots}^{\infty} \phi_0^{h_2} \phi_0^{h_2+h_3} \cdots \phi_0^{h_k} (\lambda_1, \lambda_2, \cdots, \lambda_k) \]
\[ \times \left\{ \prod_{i=2}^{k} P[B(0) + S_{h_i-1} \leq t_i - t_{i-1} < B(0) + S_{h_i}] \right\}, \]

where \( S_{h_i} \) is distributed as the \( h_i \)-th convolution of \( F \) and is independent of \( B(t_{i-1}) \). The last equation holds because of Assumption 2.1 (A3) which provides \( B(t_i) \) has the same distribution for all \( t_i \) (Lemma 2.1). From the last expression, the joint distribution of \( (\alpha_1, \alpha_2, \cdots, \alpha_k)' \) only depends on \( \{t_i - t_{i-1} : i = 2, 3, \cdots, k\} \) but not \( t_1 \) directly. Therefore, \( \{\alpha_i\} \) is strictly stationary.

\[ \square \]

**Proof of Lemma 2.2:**

\[ \phi_G(\omega) = \int_0^\infty e^{i\omega x} F_0(dx) \]
\[ = \int_0^\infty e^{i\omega x} \frac{1}{\mu} \{1 - F(x)\} dx \]
\[ = \frac{1}{\mu} \int_0^\infty e^{i\omega x} \left\{ \int_x^\infty f(t) dt \right\} dx \]
\[ = \frac{1}{\mu} \int_0^\infty f(t) \left\{ \int_0^t e^{i\omega x} dx \right\} dt \]
\[ = \frac{1}{i\omega \mu} \int_0^\infty (e^{i\omega t} - 1) f(t) dt \]
\[ = \frac{1}{i\omega \mu} (\phi_F(\omega) - 1). \]

\[ \square \]

**Lemma A.1** \( \int_0^\infty P[\tau_h > t] e^{-i\omega t} dt = \frac{1}{i\omega} \left\{ 1 - \int_0^\infty e^{-i\omega t} dF(h)(t) \right\}, \) where \( F(h) \) is the cdf of \( \tau_h \).

**Proof:** Using integration by parts,

\[ -\frac{1}{i\omega} \int_0^\infty e^{-i\omega t} dF(h)(t) = -\frac{1}{i\omega} \left\{ e^{-i\omega t} \left( F(h)(t) - 1 \right) \right\}_0^\infty + \int_0^\infty e^{-i\omega t} \left\{ 1 - F(h)(t) \right\} dt \]
\[ = -\frac{1}{i\omega} + \int_0^\infty P[\tau_h > t] e^{-i\omega t} dt. \]

\[ \square \]

**Proof of Theorem 2.3:** Under Assumption 2.1 (A1), \( \{\alpha_t\} \) is weakly stationary. The spectral density of \( \{\alpha_t\} \) is the Fourier transformation of \( \gamma_{\alpha} \), which is

\[ f_\alpha(\omega) = \frac{1}{2\pi} \int_{-\infty}^\infty \gamma_{\alpha}(t) e^{-i\omega t} dt \]
\[ = \frac{1}{2\pi} \left\{ \int_0^\infty \gamma_{\alpha}(t) e^{i\omega t} dt + \int_0^\infty \gamma_{\alpha}(t) e^{-i\omega t} dt \right\} , \]
where

\[
\int_0^\infty \gamma_\alpha(t)e^{-it\omega}dt = \int_0^\infty \left\{ \sum_{h=0}^\infty \gamma_w(h)P[N(t) = h] \right\} e^{-it\omega}dt
\]

\[
= \sum_{h=0}^\infty \gamma_w(h) \int_0^\infty \left\{ P[\tau_{h+1} > t] - P[\tau_h > t] \right\} e^{-it\omega}dt
\]

\[
= \sum_{h=1}^\infty \gamma_w(h-1) \left\{ \int_0^\infty P[\tau_h > t]e^{-it\omega}dt \right\}
- \sum_{h=0}^\infty \gamma_w(h) \left\{ \int_0^\infty P[\tau_h > t]e^{-it\omega}dt \right\}
\]

\[
= \sum_{h=1}^\infty \{\gamma_w(h-1) - \gamma_w(h)\} \int_0^\infty P[\tau_h > t]e^{-it\omega}dt. \tag{A.1}
\]

The second equality in (A.1) is justified by Fubini's theorem because

\[
\sum_{h=0}^\infty |\gamma_w(h)P[N(t) = h]| \leq \left\{ \sum_{h=0}^\infty \gamma_w^2(h) \right\}^{1/2} \left\{ \sum_{h=0}^\infty P^2[N(t) = h] \right\}^{1/2}
\]

\[
\leq \left\{ \sum_{h=0}^\infty \gamma_w^2(h) \right\}^{1/2} \left\{ \sum_{h=0}^\infty P[N(t) = h] \right\}^{1/2}
\]

\[
< \infty.
\]

Applying Lemma A.1 in (A.1), we have

\[
\int_0^\infty \gamma_\alpha(t)e^{-it\omega}dt = \frac{1}{i\omega} \sum_{h=1}^\infty \{\gamma_w(h-1) - \gamma_w(h)\} \left\{ 1 - \int_0^\infty e^{-it\omega}dF^{(h)}(t) \right\}
\]

\[
= \frac{1}{i\omega} \sum_{h=1}^\infty \{\gamma_w(h-1) - \gamma_w(h)\} \left\{ 1 - E\left( e^{-i\tau_h\omega} \right) \right\}
\]

\[
= \frac{1}{i\omega} \sum_{h=1}^\infty \{\gamma_w(h-1) - \gamma_w(h)\} \left\{ 1 - \phi_F^{(h)}(-\omega) \right\},
\]

where \(\phi_F^{(h)}(\omega)\) is the characteristic function of the cdf \(F^{(h)}\). Since \(\tau_h = T_1 + \sum_{i=2}^h T_i\) and \(\{T_i\}\) are independent random variables, \(\phi_F^{(h)}(\omega) = \phi_G(\omega)\phi_F(\omega)^{h-1}\). Then,

\[
\int_0^\infty \gamma_\alpha(t)e^{-it\omega}dt = \frac{1}{i\omega} \sum_{h=1}^\infty \{\gamma_w(h-1) - \gamma_w(h)\} - \frac{1}{i\omega} \phi_G(-\omega) \sum_{h=1}^\infty \{\gamma_w(h-1) - \gamma_w(h)\} \phi_F(-\omega)^{h-1}
\]

\[
= \frac{\gamma_w(0)}{i\omega} - \frac{1}{i\omega} \phi_G(-\omega) \left\{ \sum_{h=1}^\infty \gamma_w(h-1)\phi_F(-\omega)^{h-1} - \sum_{h=1}^\infty \gamma_w(h)\phi_F(-\omega)^{h-1} \right\}
\]

\[
= \frac{\gamma_w(0)}{i\omega} - \frac{1}{i\omega} \phi_G(-\omega) \left\{ \gamma_w(0) + [\phi_F(-\omega) - 1] \sum_{h=1}^\infty \gamma_w(h)\phi_F(-\omega)^{h-1} \right\}
\]

\[
= \frac{1}{i\omega} \left\{ \frac{1 - \phi_G(-\omega)}{\phi_F(-\omega)} \gamma_w(0) + \frac{1 - \phi_F(-\omega)}{\phi_G(-\omega)} \sum_{h=1}^\infty \gamma_w(h)\phi_F(-\omega)^{h} \right\}.\]
Therefore,

\[ f_\omega(\omega) = \frac{1}{2\pi i \omega} \left\{ \{\phi_G(\omega) - \phi_G(-\omega)\} \gamma_w(0) + \{1 - \phi_F(-\omega)\} \phi_G(-\omega) \sum_{h=1}^{\infty} \gamma_w(h) \phi_F(-\omega)^h \right\} - \{1 - \phi_F(\omega)\} \phi_F(\omega) \sum_{h=1}^{\infty} \gamma_w(h) \phi_F(\omega)^h \right\} \] .

Proof of Corollary 2.1: Since the \{w_n\} are uncorrelated, that is \( \gamma_w(h) = 0 \) for all \( h \neq 0 \), the spectral density (see Theorem 2.3) becomes

\[ f_\omega(\omega) = \frac{\gamma_w(0)}{2\pi i \omega} \left\{ \phi_G(\omega) - \phi_G(-\omega) \right\} \]

\[ = \frac{\gamma_w(0)}{2\pi i \omega \mu^2} \{2 - \phi_F(\omega) - \phi_F(-\omega)\} \] (by Lemma 2.2).

Proof of Corollary 2.2: Since \( \gamma_w(h) \) is a geometrically decaying sequence, the last two terms in the spectral density (see Theorem 2.3) can be simplified, and we have

\[ f_\omega(\omega) = \frac{\gamma_w(0)}{2\pi i \omega} \left\{ \{\phi_G(\omega) - \phi_G(-\omega)\} + \{1 - \phi_F(-\omega)\} \phi_G(-\omega) \frac{\xi \phi_F(-\omega)}{\phi_F(-\omega) 1 - \xi \phi_F(-\omega)} \right\} \]

\[ = \frac{\gamma_w(0)}{2\pi i \omega} \left\{ \phi_G(\omega) \left\{1 - \frac{\xi}{1 - \phi_F(\omega)} \right\} + \phi_G(-\omega) \left\{\frac{\xi}{1 - \phi_F(-\omega)} - 1\right\} \right\} \]

\[ = \frac{\gamma_w(0)}{2\pi i \omega} \left\{ \phi_G(\omega) \left\{\frac{1}{1 - \xi \phi_F(\omega)} \right\} + \phi_G(-\omega) \left\{\frac{\xi - 1}{1 - \phi_F(-\omega)} \right\} \right\} \]

\[ = \frac{(1 - \xi) \gamma_w(0)}{2\pi i \omega} \left\{ \frac{1}{1 - \xi \phi_F(\omega)} \left\{\frac{\phi_F(\omega) - 1}{1 - \phi_F(\omega)} + \frac{1}{\phi_F(-\omega) - 1} \right\} \right\} \] (by Lemma 2.2)

\[ = \frac{(1 - \xi) \gamma_w(0)}{2\pi i \omega^2} \left\{ \frac{1 - \phi_F(\omega)}{1 - \xi \phi_F(\omega)} + \frac{1 - \phi_F(-\omega)}{1 - \xi \phi_F(-\omega)} \right\} \]

Before proving Theorem 2.4, we need the following three lemmas.

Lemma A.2 Let \( L > 0 \) vary slowly. Let \( z_p(x) = \int_0^x y^p L(y)dy, \: z^*_p(x) = \int_x^\infty y^p L(y)dy \). If \( p \geq 1 \), then \( z_p \) varies regularly with exponent \( (p + 1) \). If \( p < -1 \), then \( z^*_p \) varies regularly with exponent \( (p + 1) \) and this remains true for \( p = -1 \) if \( z^*_p \) exists.


Lemma A.3 If \( F_1 \) and \( F_2 \) are two distribution functions such that \( 1 - F_1(x) = x^{-\beta} L_1(x) \) as \( x \to \infty \), where \( L_1(\cdot) \) is slowly varying, then the convolution \( G = F_1 * F_2 \) has a regularly varying tail such that \( 1 - G(x) \sim x^{-\beta} \{L_1(x) + L_2(x)\} \).
Proof: See Feller (1971), page 278. 

Lemma A.4 Assume $F$ is a heavy-tailed distribution function with index $\beta$ where $1 < \beta < 2$. Then, $P[\tau_h \leq t < \tau_{h+1}]$ is regularly varying with index $(1 - \beta)$ where $\tau_{h+1} = \tau_h + T_{h+1}$, \{T_n : n = 2, 3, \ldots\} is an iid sequence from $F$ and $T_1$ is distributed as $G$.

Proof: By Lemma A.3. 

Proof of Theorem 2.4: For large $t$,

$$\gamma_\alpha(t) \sim \int_t^\infty y^{-\beta} L(y) dy = z_{-\beta}(t),$$

where $1 < \beta < 2$. From Lemma A.2, $\gamma_\alpha(t)$ is regularly varying with exponent $(1 - \beta)$ at $\infty$, and can be written as $\gamma_\alpha(t) \sim t^{1-\beta} L_0(t)$ where $L_0$ is slowly varying at $\infty$. Applying Lemma A.4, for large $t$, $P[\tau_h \leq t < \tau_{h+1}] \sim t^{1-\beta} L_h(t)$ where $L_h$ is slowly varying at $\infty$. Then $\gamma_\alpha(t) \sim t^{1-\beta} \sum_{h=0}^{\infty} \gamma_\omega(h) L_h(t)$ is regularly varying with exponent $(1 - \beta)$ provided $\sum_{h=0}^{\infty} \gamma_\omega(h) L_h(t)$ varies slowly. 

Proof of Proposition 2.1:

$$f_\alpha(\omega) = \frac{1}{2\pi} \int_{-\infty}^\infty \gamma_\alpha(t) \cos(\omega t) dt$$

$$= \frac{1}{\pi} \int_0^\infty \gamma_\alpha(t) \cos(\omega t) dt$$

$$= \frac{1}{\pi} \sum_{m=1}^{\infty} \int_{(m-1)c}^{mc} \gamma_\alpha(t) \cos(\omega t) dt$$

$$= \frac{1}{\pi} \sum_{m=1}^{\infty} \int_{(m-1)c}^{mc} \left\{ \gamma_\alpha(m-1) \left( m - \frac{t}{c} \right) + \gamma_\alpha(m) \left( \frac{t}{c} - (m-1) \right) \right\} \cos(\omega t) dt$$

$$= \frac{1}{\pi} \sum_{m=1}^{\infty} \int_0^m \left\{ \gamma_\alpha(m-1)(m-t) + \gamma_\alpha(m)(t-(m-1)) \right\} \cos(\omega t) dt$$

$$= \frac{1}{\pi} \sum_{m=1}^{\infty} \int_0^1 \left\{ \gamma_\alpha(m-1)(1-t) + \gamma_\alpha(m) t \right\} \cos(\omega t + (m-1)\omega) dt$$

$$= \frac{1}{\pi} \sum_{m=1}^{\infty} \int_0^1 \left\{ \gamma_\alpha(m-1)(1-t) \cos(\omega t - 1) + mc\omega \right\} dt$$

$$+ \gamma_\alpha(m) t \cos(\omega t + (m-1)\omega) dt$$

$$= \frac{c}{\pi} (B_1 + B_2),$$
where

\[ B_1 = \sum_{m=1}^{\infty} \int_0^t \gamma_0(m-1)(1-t) \left\{ \cos[\omega(t-1)] \cos(m\omega) - \sin[\omega(t-1)] \sin(m\omega) \right\} dt \]

\[ = \left\{ \sum_{m=1}^{\infty} \gamma_0(m-1) \cos(m\omega) \right\} \int_0^t (1-t) \cos[\omega(1-t)] dt \]
\[ + \left\{ \sum_{m=1}^{\infty} \gamma_0(m-1) \sin(m\omega) \right\} \int_0^t (1-t) \sin[\omega(1-t)] dt, \]

\[ B_2 = \sum_{m=1}^{\infty} \int_0^t \gamma_0(m) t \left\{ \cos(\omega t) \cos[(m-1)\omega t] - \sin(\omega t) \sin[(m-1)\omega] \right\} dt \]

\[ = \left\{ \sum_{m=1}^{\infty} \gamma_0(m) \cos[(m-1)\omega] \right\} \int_0^t t \cos(\omega t) dt \]
\[ + \left\{ \sum_{m=1}^{\infty} \gamma_0(m) \sin[(m-1)\omega] \right\} \int_0^t t \sin(\omega t) dt. \]

Since

\[ \int_0^1 (1-t) \cos[\omega(1-t)] dt = \int_0^1 t \cos(\omega t) dt = \frac{\sin \omega}{\omega} \frac{1 - \cos \omega}{(\omega)^2}, \]
\[ \int_0^1 (1-t) \sin[\omega(1-t)] dt = \int_0^1 t \sin(\omega t) dt = \frac{-\cos \omega}{\omega} + \frac{\sin \omega}{(\omega)^2}, \]

then

\[ B_1 + B_2 = \left\{ \int_0^t t \cos(\omega t) dt \right\} \left\{ \sum_{m=1}^{\infty} \gamma_0(m-1) \cos(m\omega) + \sum_{m=1}^{\infty} \gamma_0(m) \cos[(m-1)\omega] \right\} \]
\[ + \left\{ \int_0^t t \sin(\omega t) dt \right\} \left\{ \sum_{m=1}^{\infty} \gamma_0(m-1) \sin(m\omega) - \sum_{m=1}^{\infty} \gamma_0(m) \cos[(m-1)\omega] \right\} \]
\[ = \left\{ \int_0^t t \cos(\omega t) dt \right\} (\cos \omega) \sum_{m=-\infty}^{\infty} \gamma_0(m) \cos(m\omega) \]
\[ + \left\{ \int_0^t t \sin(\omega t) dt \right\} (\sin \omega) \sum_{m=-\infty}^{\infty} \gamma_0(m) \cos(m\omega) \]
\[ = \sum_{m=-\infty}^{\infty} \gamma_0(m) \cos(m\omega) \]
\[ \times \left\{ \cos \omega \left( \frac{\sin \omega}{\omega} - \frac{1 - \cos \omega}{(\omega)^2} \right) + \sin \omega \left( -\frac{\cos \omega}{\omega} + \frac{\sin \omega}{(\omega)^2} \right) \right\} \]
\[ = \frac{1 - \cos \omega}{(\omega)^2} \sum_{m=-\infty}^{\infty} \gamma_0(m) \cos(m\omega) \]
\[ = \frac{\pi \sin^2(\omega/2)}{(\omega/2)^2} f_\omega(\omega). \]
Finally, we have

\[ f_\alpha(\omega) = \frac{c \sin^2(\omega/2)}{(\omega/2)^2} f_\omega(\omega). \]

**Proof of Lemma 2.3:** Let

\[ c = (c_0, c_1, \ldots, c_\kappa)', \]

\[ y = (1, h, h^2, \ldots, h^\kappa)', \]

and \( A_\kappa = (a_0, a_1, \ldots, a_\kappa)' \) which is a \((\kappa + 1) \times (\kappa + 1)\) matrix defined by

\[ a_0 = (1, 0, 0, \ldots, 0)', \]

\[ a_i = B^{-1} a_{i-1} - (i-1)a_{i-1}, \]

for \( i = 1, 2, \ldots, \kappa \), and \( B^{-1} \) is the forward shift operator which shifts the \( i \)-th element to the \((i+1)\)-th element, e.g. \( B^{-1} a_0 = (0, 1, 0, \ldots, 0)' \). Define \( c^* = (c_0^*, c_1^*, \ldots, c_\kappa^*)' = (A_\kappa^{-1})^T c \). The inverse matrix of \( A_\kappa \) is well defined since \( A_\kappa \) is a lower triangular matrix with ones on the main diagonal, so \( \det(A_\kappa) = 1 \neq 0 \). Therefore \( c^*A_\kappa y = c'y \). Since

\[ A_\kappa y = \begin{pmatrix} a_0'y \\ a_1'y \\ \vdots \\ a_\kappa'y \end{pmatrix} = \begin{pmatrix} 1 \\ h \\ h(h-1) \\ \vdots \\ h!/(h-\kappa)! \end{pmatrix}, \]

the result follows. \( \square \)

**Proof of Proposition 2.2:** Using Lemma 2.3, the component \( \gamma_{ar} \) for real roots in (2.9) can be simplified:

\[ \gamma_{ar}(t) = \sum_{h=0}^{\infty} \left\{ \sum_{j=1}^{pr} \xi_{rj} \left( \sum_{k=0}^{\infty} c_{jk} h^k \right) \right\} \left\{ \frac{1}{h!} \left( \frac{t}{\lambda} \right)^h e^{-t/\lambda} \right\} \]

\[ = e^{-t/\lambda} \sum_{j=1}^{pr} \sum_{h=0}^{\infty} \frac{(\xi_{rj} t/\lambda)^h}{h!} \left( \sum_{k=0}^{\infty} c_{jk} h^k \right) \]

\[ = e^{-t/\lambda} \sum_{j=1}^{pr} \sum_{h=0}^{\infty} \frac{(\xi_{rj} t/\lambda)^h}{h!} \left\{ \sum_{k=0}^{\infty} c_{jk} \frac{h!}{(h-k)!} \right\} \]
\[
\begin{align*}
\gamma_{oc}(t) &= \sum_{h=0}^{\infty} \sum_{j=1}^{\nu} \sum_{k=0}^{n} r_j^k \left\{ \sum_{k=0}^{n} \left( d_{jk} \cos(h\theta_j + \beta_{jk}) h^k \right) \right\} \frac{1}{h!} \left( \frac{t}{\lambda} \right)^h e^{-t/\lambda} \\
&= \sum_{h=0}^{\infty} \sum_{j=1}^{\nu} \sum_{k=0}^{n} r_j^k \left\{ (\cos h\theta_j) \sum_{k=0}^{n} d_{jk1} h^k - (\sin h\theta_j) \sum_{k=0}^{n} d_{jk2} h^k \right\} \frac{1}{h!} \left( \frac{t}{\lambda} \right)^h e^{-t/\lambda} \\
&= e^{-t/\lambda} \sum_{j=1}^{\nu} \sum_{k=0}^{n} d_{jk1}^* t^k \sum_{h=0}^{\infty} \left( \cos h\theta_j \right) \frac{(r_j t/\lambda)^{h-k}(r_j/\lambda)^k}{(h-k)!} \\
&\quad - e^{-t/\lambda} \sum_{j=1}^{\nu} \sum_{k=0}^{n} d_{jk2}^* t^k \sum_{h=0}^{\infty} \left( \sin h\theta_j \right) \frac{(r_j t/\lambda)^{h-k}(r_j/\lambda)^k}{(h-k)!} \\
&= e^{-t/\lambda} \sum_{j=1}^{\nu} \sum_{k=0}^{n} d_{jk1}^* t^k B_{k1} - \sum_{k=0}^{n} d_{jk2}^* t^k B_{k2},
\end{align*}
\]

where

\[
\begin{align*}
d_{jk1} &= d_{jk} \cos \beta_{jk}, \quad d_{jk2} = d_{jk} \sin \beta_{jk}, \\
(d_{j01}, d_{j21}, \ldots, d_{j\kappa_{\nu1}}) &= (d_{j01}, d_{j21}, \ldots, d_{j\kappa_{\nu1}}) A_{j1}^{-1}, \\
(d_{j02}, d_{j22}, \ldots, d_{j\kappa_{\nu2}}) &= (d_{j02}, d_{j22}, \ldots, d_{j\kappa_{\nu2}}) A_{j2}^{-1},
\end{align*}
\]
and $A_{c_j} \equiv A_{\kappa_{c_j}}$ is a $(\kappa_{c_j} + 1) \times (\kappa_{c_j} + 1)$ matrix defined in Lemma 2.3. In (A.2), $B_{k1}$ and $B_{k2}$ can be further simplified as follows

\[
B_{k1} = \frac{1}{2} \left\{ \sum_{h=0}^{\infty} \left( e^{ih\theta_j} + e^{-ih\theta_j} \right) \frac{(r_{jt}/\lambda)^{h-k}(r_{jt}/\lambda)^k}{(h-k)!} \right\} \\
= \frac{1}{2} \left\{ \sum_{h=0}^{\infty} \frac{d^k}{dt^k} \left( e^{i\theta_j r_{jt}/\lambda} \right)^h \right. + \left. \sum_{h=0}^{\infty} \frac{d^k}{dt^k} \left( e^{-i\theta_j r_{jt}/\lambda} \right)^h \right\} \\
= \frac{1}{2} \frac{d^k}{dt^k} \left\{ \exp \left( \frac{r_{jt}}{\lambda} e^{i\theta_j} \right) + \exp \left( \frac{r_{jt}}{\lambda} e^{-i\theta_j} \right) \right\} \\
= \frac{1}{2} \left\{ \exp \left( \frac{r_{jt}}{\lambda} e^{i\theta_j} \right) \left( r_{jt}/\lambda e^{i\theta_j} \right)^k + \exp \left( \frac{r_{jt}}{\lambda} e^{-i\theta_j} \right) \left( r_{jt}/\lambda e^{-i\theta_j} \right)^k \right\} \\
= \frac{1}{2} \frac{r_{jt}}{\lambda} \left\{ \exp \left( \frac{r_{jt}}{\lambda} e^{i\theta_j} + i \theta_j k \right) + \exp \left( \frac{r_{jt}}{\lambda} e^{-i\theta_j} - i \theta_j k \right) \right\} \\
= \frac{1}{2} \frac{r_{jt}}{\lambda} \left\{ \exp \left( \frac{r_{jt}}{\lambda} \cos \theta_j \right) \left\{ \exp \left( i \left( \theta_j k + \frac{r_{jt}}{\lambda} \sin \theta_j \right) \right) \right\} \right\} \\
\right. \\
= \left( \frac{r_{jt}}{\lambda} \right)^k \exp \left( \frac{r_{jt}}{\lambda} \cos \theta_j \right) \cos \left( \theta_j k + \frac{r_{jt}}{\lambda} \sin \theta_j \right),
\]

\[
B_{k2} = \frac{i}{2} \left\{ \sum_{h=0}^{\infty} \left( e^{ih\theta_j} - e^{-ih\theta_j} \right) \frac{(r_{jt}/\lambda)^{h-k}(r_{jt}/\lambda)^k}{(h-k)!} \right\} \\
= \frac{i}{2} \left\{ \sum_{h=0}^{\infty} \frac{d^k}{dt^k} \left( e^{i\theta_j r_{jt}/\lambda} \right)^h \right. - \left. \sum_{h=0}^{\infty} \frac{d^k}{dt^k} \left( e^{-i\theta_j r_{jt}/\lambda} \right)^h \right\} \\
= \frac{i}{2} \frac{d^k}{dt^k} \left\{ \exp \left( \frac{r_{jt}}{\lambda} e^{i\theta_j} \right) - \exp \left( \frac{r_{jt}}{\lambda} e^{-i\theta_j} \right) \right\} \\
= - \left( \frac{r_{jt}}{\lambda} \right)^k \exp \left( \frac{r_{jt}}{\lambda} \cos \theta_j \right) \sin \left( \theta_j k + \frac{r_{jt}}{\lambda} \sin \theta_j \right).
\]

Let

\[
\begin{align*}
r_j^* & \equiv \exp \{- \left( 1 - r_j \cos \theta_j \right)/\lambda \}, \\
\theta_j^* & \equiv (r_j \sin \theta_j)/\lambda, \\
d_{jk}^* & \equiv (r_j/\lambda)^k \sqrt{d_{jk1}^2 + d_{jk2}^2}, \\
\theta_{djk} & \equiv \tan^{-1} \left( \frac{d_{jk2}^*}{d_{jk1}^*} \right), \\
\beta_{jk}^* & \equiv \theta_j k + \theta_{djk}.
\end{align*}
\]
Plugging $B_{k1}$ and $B_{k2}$ into (A.2), we have
\[
\gamma_{oc}(t) = \sum_{j=1}^{p_c} \left\{ \exp \left( -\frac{(1 - r_j \cos \theta_j)}{\lambda} \right) \right\} \sum_{k=0}^{\kappa_j} \left\{ d_{j,k1}^* \left( \frac{r_j}{\lambda} \right)^k \right\} t^k \cos \left( \theta_j k + \left( \frac{r_j}{\lambda} \sin \theta_j \right) t \right\} \\
- \sum_{j=1}^{p_c} \left\{ \exp \left( -\frac{(1 - r_j \cos \theta_j)}{\lambda} \right) \right\} \sum_{k=0}^{\kappa_j} \left\{ d_{j,k2}^* \left( \frac{r_j}{\lambda} \right)^k \right\} t^k \sin \left( \theta_j k + \left( \frac{r_j}{\lambda} \sin \theta_j \right) t \right\} \\
= \sum_{j=1}^{p_c} r_j^t \left\{ \sum_{k=0}^{\kappa_j} \left\{ d_{j,k1}^* \left( \frac{r_j}{\lambda} \right)^k \right\} t^k \cos(\theta_j k + \theta^*_j t) - \sum_{k=0}^{\kappa_j} \left\{ d_{j,k2}^* \left( \frac{r_j}{\lambda} \right)^k \right\} t^k \sin(\theta_j k + \theta^*_j t) \right\} \\
= \sum_{j=1}^{p_c} r_j^t \sum_{k=0}^{\kappa_j} \frac{d_{j,k1}^*}{\sqrt{d_{j,k1}^* + d_{j,k2}^*}} \sqrt{d_{j,k1}^* + d_{j,k2}^*} \cos(\theta_j k + \theta^*_j t) - \frac{d_{j,k2}^*}{\sqrt{d_{j,k1}^* + d_{j,k2}^*}} \sin(\theta_j k + \theta^*_j t) \\
= \sum_{j=1}^{p_c} r_j^t \sum_{k=0}^{\kappa_j} d_{j,k}^* \cos(t \theta^*_j + \beta^*_j) t^k.
\]

**Proof of Proposition 2.6:** Given $a_1 < 0$, $a_2 < 0$, $b_1 < 0$ and $a_1 \neq a_2$ for a CARMA(2, 1) process, its spectral density is also the spectral density of a semi-Markov process constructed by an AR(2) with distinct real roots ($\xi_1, \xi_2$) and $\text{Exp}(\lambda)$ sojourns if
\[
\xi_1 = 1 + a_1 \lambda, \quad \xi_2 = 1 + a_2 \lambda, \quad \lambda = -b_1^{-1} \sqrt{1 - \xi_1 \xi_2 (\xi_1 + \xi_2 - 1)}, \quad (A.3)
\]
subject to the constraints $\lambda > 0$ and $|\xi_1| < 1$, $|\xi_2| < 1$. From (A.3), $\lambda$ has to satisfy
\[
\lambda < \min\{-2/a_1, -2/a_2\}. \quad (A.5)
\]
Plugging (A.3) in (A.4),
\[
b_1 \lambda = \sqrt{1 - (1 + a_1 \lambda)(1 + a_2 \lambda)(1 + a_1 \lambda + a_2 \lambda)}.
\]
Squaring both sides and dividing by $\lambda$, we obtain
\[
a_1 a_2 (a_1 + a_2) \lambda^2 + \left\{(a_1 + a_2)^2 + a_1 a_2 + b_1^2\right\} \lambda + 2(a_1 + a_2) = 0,
\]
which has two solutions $\lambda_+$ and $\lambda_-:
\[
\lambda_{\pm} = \frac{(a_1 + a_2)^2 + a_1 a_2 + b_1^2}{2a_1 a_2 (a_1 + a_2)} \pm \frac{\sqrt{\left\{(a_1 + a_2)^2 + a_1 a_2 + b_1^2\right\}^2 - 8a_1 a_2 (a_1 + a_2)^2}}{2a_1 a_2 (a_1 + a_2)}.
\]
The inequality (A.6) is equivalent to
\[ S_0 \equiv \{(a_1 + a_2)^2 + a_1 a_2 + b_1^2 - 2\sqrt{2a_1 a_2 |a_1 a_2|} \geq 0\}. \]

At the same time, the solution \( \lambda_+ \) has to satisfy (A.5). Since \( a_1 \) and \( a_2 \) are symmetric in both \( \lambda_+ \) and (A.5), we only need to check the constraint \( \lambda_+ < -2/a_1 \) and the same condition can be applied for \( a_2 \) by interchanging \( a_1 \) and \( a_2 \). In order to satisfy \( \lambda_+ < -2/a_1 \), it requires
\[ \{(a_1 + a_2)^2 + a_1 a_2 + b_1^2\} - \sqrt{\{(a_1 + a_2)^2 + a_1 a_2 + b_1^2\}^2 - 8a_1 a_2(a_1 + a_2)^2} > -4a_2(a_1 + a_2), \]
that is
\[ a_1^2 - 3a_2^2 - a_1 a_2 + b_1^2 < \sqrt{\{(a_1 + a_2)^2 + a_1 a_2 + b_1^2\}^2 - 8a_1 a_2(a_1 + a_2)^2}. \] (A.7)

In (A.7), if the left hand side is negative then the inequality always holds, that is
\[ a_1^2 - 3a_2^2 - a_1 a_2 + b_1^2 < 0, \] (A.8)
Otherwise, taking squares to both sides in (A.7), we have
\[ \{(a_1^2 - 3a_2^2 - a_1 a_2 + b_1^2)^2 < \{(a_1 + a_2)^2 + a_1 a_2 + b_1^2\}^2 - 8a_1 a_2(a_1 + a_2)^2, \]
which can be simplified as
\[ a_2^2 < b_1^2, \quad \text{provided } a_1^2 - 3a_2^2 - a_1 a_2 + b_1^2 \geq 0. \] (A.9)

Because of the symmetry, another pair of constraints can be obtained by interchanging \( a_1 \) and \( a_2 \) in (A.8) and (A.9):
\[ a_2^2 - 3a_1^2 - a_1 a_2 + b_1^2 < 0, \] (A.10)
\[ a_1^2 < b_1^2, \quad \text{provided } a_2^2 - 3a_1^2 - a_1 a_2 + b_1^2 \geq 0. \] (A.11)

To sum up, the parameters \((a_1, a_2, b_1)\) \( \in S_0 \) should satisfy either (A.8) or (A.9) and satisfy either (A.10) or (A.11) at the same time such that \( \lambda_+ \) is the solution of (A.3)-(A.5). The
similar derivation can be done for finding constraints for \( \lambda_- \). It turns out the results have been covered in the results for \( \lambda_+ \).

\[ \square \]

**Proof of Proposition 2.8:** Given \( a_1 = a_2 < 0, b_1 < 0 \) for a CARMA(2,1) process, its spectral density is also the spectral density of a semi-Markov process constructed by an AR(2) with equal roots \( \xi \) and \( \text{Exp}(\lambda) \) sojourns if

\[
\begin{align*}
\xi &= 1 + a_1 \lambda, \\
\lambda &= -b_1^{-1} \sqrt{(1 - \xi)(1 + \xi + 2\xi^2)},
\end{align*}
\]

subject to the constraints \( \lambda > 0 \) and \( |\xi| < 1 \). Plugging (A.12) in (A.13), we have

\[
(b_1 \lambda)^2 = -a_1 \lambda(2a_1^2\lambda^2 + 5a_1\lambda + 4),
\]

which can be simplified as

\[
2a_1^2\lambda^2 + (5a_1^2 + b_1^2)\lambda + 4a_1 = 0.
\]

Therefore, it has two solutions \( \lambda_+ \) and \( \lambda_- \):

\[
\frac{-((5a_1^2 + b_1^2) \pm \sqrt{b_1^4 + 10a_1^2b_1^2 - 7a_1^4})}{4a_1^2},
\]

provided

\[
b_1^4 + 10a_1^2b_1^2 - 7a_1^4 \geq 0. \tag{A.14}
\]

The condition in (A.14) is always required for the existence of a solution. This condition is equivalent to

\[
b_1^2 \geq (4\sqrt{2} - 5)a_1^2. \tag{A.15}
\]

Under this condition, the constraint \( \lambda_+ < -2/a_1 \) is checked, which requires

\[
b_1^2 - 3a_1^2 < \sqrt{b_1^4 + 10a_1^2b_1^2 - 7a_1^4}.
\]

It turns out this inequality always holds under (A.15). Then the result follows. \[ \square \]
APPENDIX B  EXTENSIONS OF THE ALGORITHMS

In this Appendix, the extensions for ARFIMA(0, d, q), ARFIMA(p, d, 0), ARFIMA(p, d, q) and ARFIMA(0, d, 0) with additive AR(p) noise are derived. For convenience, we only describe the case with p = 1 and q = 1. For higher order of p and q, the algorithm can be extended in a straightforward manner.

ARFIMA(0, d, q)

Assume \{y_t\} is an ARFIMA(0, d, 1) process satisfying (2.1). Let \( y_t = (1 + dB)\alpha_t \), then \{\alpha_t\} becomes an ARFIMA(0, d, 0) process. Applying our approximate ARIMA(2, 2) process defined in (3.2) and (3.3) for \{\alpha_t\}, the corresponding approximating process for \{y_t\} is an ARIMA(2, 3) satisfying the following equations:

\[
y_t = (\theta b_2, \theta b_1 + b_2, \theta + b_1, 1)X_t^\dagger,
\]

\[
X_t^\dagger = A^\dagger X_{t-1}^\dagger + R^\dagger \eta_t.
\]

where

\[
X_t^\dagger = \begin{pmatrix} x_{t-3} \\ x_{t-2} \\ x_{t-1} \\ x_t \end{pmatrix}, \quad A^\dagger = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -a_2(d) & -a_1(d) \end{pmatrix}, \quad R^\dagger = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.
\]

Let \( \theta = (d, \sigma_\theta^2, \theta)' \) and \( Y_T = (y_1, y_2, \ldots, y_T)' \). The data are augmented by three latent variables \( (x_{-2}, x_{-1}, x_0)' \). The joint density of \( Y_T, (x_{-2}, x_{-1}, x_0)' \) and \( \theta \) has the form

\[
f(Y_T, x_{-2}, x_{-1}, x_0, \theta) = f(Y_T, x_{-2}, x_{-1}, x_0 | \theta) \pi(\theta)
\]
This joint density has exactly the same structure as ARFIMA(0, d, 0) models except \( f(x_{-1}, x_0 \mid \theta) \) is replaced by \( f(x_{-2}, x_{-1}, x_0 \mid \theta) \). Therefore, the same sampling techniques can be used for generating \( d, \sigma_n^2 \) and \( (x_{-2}, x_{-1}, x_0)' \). For generating \( \theta \), the coefficient for the moving average component, assume that the prior for \( \theta \) is uniform on \((-1, 1)\). The conditional distribution of \( \theta \) satisfies

\[
\begin{align*}
\pi(\theta \mid Y_T, x_{-2}, x_{-1}, x_0, d, \sigma_n^2) & \propto \left\{ \prod_{t=1}^{T} f(x_t \mid x_{t-1}, x_{t-2}, \theta) \right\} \pi(\theta) \\
& \propto \exp \left\{ -\frac{\sum_{t=1}^{T} (x_t + a_1 x_{t-1} + a_2 x_{t-2})^2}{2r(d) \sigma_n^2} \right\} \beta_{(-1, 1)}(\theta),
\end{align*}
\]

where the \( \{x_t\} \) are polynomials in \( \theta \). As in the Bayesian approach for ARMA process (Chib and Greenberg, 1994), a Metropolis-Hastings algorithm is conducted for generating \( \theta \).

For \( q > 1 \), an ARFIMA(0, d, q) process is approximated by an ARMA(2, 2 + q) model satisfying \( y_t = (b_2, b_1, 1) \theta(B) X_t \). The only change in the algorithm is to augment with \( (2 + q) \) latent variables \( (x_0, x_{-1}, \ldots, x_{-q-1})' \).

**ARFIMA(p, d, 0)**

Assume \( \{y_t\} \) is an ARFIMA(1, d, 0) process satisfying (2.1). Let \( (1 + \phi B)y_t = \alpha_t \), so that \( \{\alpha_t\} \) becomes an ARFIMA(0, d, 0) process. We consider the approximating ARMA (3, 2) model for \( \{y_t\} \) which satisfies

\[
(1 + \phi B)y_t = (b_2, b_1, 1) X_t,
\]

in which \( \{\alpha_t\} \) is approximated by (3.2) and (3.3). Let \( \theta = (d, \sigma_n^2, \phi)' \). The data are augmented with two latent variables \( (x_0, x_1)' \). The joint density of \( Y_T, (x_0, x_1)' \) and \( \theta \) becomes

\[
\begin{align*}
f(Y_T, x_0, x_1, \theta) &= f(Y_T, x_0, x_1 \mid \theta) \pi(\theta) \\
&= f(\alpha_2, \alpha_3, \ldots, \alpha_T, y_1, x_0, x_1 \mid \theta) \pi(\theta)
\end{align*}
\]
\[ f(x_{0T}, x_0, y_1 | \theta) \pi(\theta) \]
\[ = \left\{ \prod_{t=2}^{T} f(x_t | x_{t-1}, x_{t-2}, \theta) \right\} f(x_1, x_0, y_1 | \theta) \pi(\theta). \]

(B.1)

Since (B.1) also has the same structure as the joint density for ARFIMA(0,d,0), the same sampling procedures for \( d \) and \( \sigma^2 \) can be applied. The corresponding parameters for the inverse gamma distribution and the transition probability in the Metropolis-Hastings algorithm for \( d \) should be modified slightly according to their difference. The conditional distribution of \( \phi \) is proportional to (B.1) where \( \phi \) is involved in both terms: the product term where \( \{x_t\} \) are linear functions of \( \phi \) and the remaining term \( f(x_1, x_0, y_1 | \theta) \) where the components of the covariance matrix of \( (x_1, x_0, y_1)' \) are polynomials of \( \phi \). Since the linear relation makes the former term Gaussian and the remaining term is relatively negligible in (B.1), a Metropolis-Hastings algorithm is conducted using only the product term to form the proposal distribution \( J \). Specifically, in the \( i \)th Metropolis step, \( J(\phi) \) is Gaussian and does not depend on the previous \( \phi^{(i-1)} \). It satisfies

\[
J(\phi) \propto \prod_{t=2}^{T} f(x_t | x_{t-1}, x_{t-2}, \theta) \\
\propto \exp \left\{ -\sum_{t=2}^{T} \left[ x_t + a_1 x_{t-1} + a_2 x_{t-2} \right]^2 \right\} \\
\propto \exp \left\{ -\frac{(DX_{0T})'(DX_{0T})}{2r(d)\sigma^2_\eta} \right\} \\
\propto \exp \left\{ -\frac{(DC^{-1}\alpha_T')'(DC^{-1}\alpha_T)}{2r(d)\sigma^2_\eta} \right\},
\]

where the matrix \( C \) is defined in (3.6) and

\[
X_{0T}^\dagger = \begin{pmatrix} x_0 \\ x_1 \\ y_2 + \phi y_1 \\ y_3 + \phi y_2 \\ \vdots \\ y_T + \phi y_{T-1} \end{pmatrix}, \quad \alpha_T^\dagger = \begin{pmatrix} x_0 \\ x_1 \\ y_2 + \phi y_1 \\ y_3 + \phi y_2 \\ \vdots \\ y_T + \phi y_{T-1} \end{pmatrix}, \quad D = \begin{pmatrix} a_2 & a_1 & 1 \\ \vdots & \vdots & \vdots \\ a_2 & a_1 & 1 \end{pmatrix}.
\]
The acceptance probability is defined as

\[ \delta = \min \left\{ 1, \frac{f(x_1, x_0, y_1 \mid d, \sigma_n^2, \phi^r)}{f(x_1, x_0, y_1 \mid d, \sigma_n^2, \phi^{(i-1)})} \right\}, \]

where \( \phi^r \) is a draw from \( J(\phi) \).

For \( p > 1 \), an ARFIMA(\( p, d, 0 \)) process is approximated by an ARMA(\( 2 + p, 2 \)) model satisfying \( \phi(B)y_t = (b_2, b_1, 1)X_t \). The only change in the algorithm is to augment with latent variables \( (x_p, x_{p-1})' \).

**ARFIMA(\( p, d, q \))**

The algorithm for ARFIMA(\( p, d, q \)) can be form by combining the previous two algorithms. Assume \( \{y_t\} \) is an ARFIMA(1, \( d, 1 \)) process defined in (2.1). We consider an ARMA(3,3) model to approximate \( \{y_t\} \), which satisfies the following state space form

\[
(1 + \phi B)y_t = (\theta b_2, \theta b_1 + b_2, \theta + b_1, 1)X_t^\dagger,
\]

where \( X_t^\dagger \) is defined in the ARFIMA(0, \( d, q \)) case. The algorithm follows the following steps:

1. Given \( \phi \), the algorithm for ARFIMA(0, \( d, 1 \)) can be applied to the transformed data \( \{(1 + \phi B)y_t\} \) for generating \( \theta \).

2. Given \( \theta \), the algorithm for ARFIMA(1, \( d, 0 \)) can be applied by considering \( \{X_t^\dagger\} \) instead of \( \{X_t\} \) for generating \( \phi \).

3. Given \( \phi \) and \( \theta \), \( (d, \sigma_n^2)' \) can be generated by the algorithm for ARFIMA(0, \( d, 0 \)). This step can also be combined into step 1 or step 2.

For \( p > 1 \) and \( q > 1 \), an ARFIMA(\( p, d, q \)) process is approximated by an ARMA(\( 2 + p, 2 + q \)) model satisfying \( \phi(B)y_t = \theta(B)(b_2, b_1, 1)X_t \).

**ARFIMA(0, \( d, 0 \)) with Additive AR(\( p \)) Noise**

Assume the process \( \{y_t\} \) is the sum of an AR(1) and an ARFIMA(0, \( d, 0 \)) which satisfies

\[ y_t = u_t + \alpha_t, \]
\[(1 + \phi B)u_t = \epsilon_t,\]
\[(1 - B)^d \alpha_t = \eta_t,\]

where \(\{\epsilon_t\} \) and \(\{\eta_t\} \) are iid \(N(0, \sigma_\epsilon^2) \) and \(N(0, \sigma_\eta^2) \), respectively. We consider the approximation of \(\{y_t\} \) in which \(\{\alpha_t\} \) is approximated by an ARMA\((2,2)\) defined in (3.2) and (3.3). Then the approximating process satisfies

\[(1 + \phi B)y_t = (1 + \phi B)u_t + (b_2, b_1, 1)(1 + \phi B)X_t
= \epsilon_t + (b_2, b_1, 1)(1 + \phi B)X_t
= \epsilon_t + (\phi b_2, \phi b_1 + b_2, \phi + b_1, 1)X_t^1, \tag{B.2}\]

where \(X_t = (x_{t-2}, x_{t-1}, x_t)\) and \(X_t^1 = (x_{t-3}, x_t^1)\). In fact, the transformed process \(\{(1+\phi B)y_t\}\) is the sum of the white noise process \(\{\epsilon_t\}\) and an ARFIMA\((0, d, 1)\) process and therefore the approximating process \(\{(1 + \phi B)y_t\}\) is the sum of the same white noise process and the approximating ARMA\((2,3)\) process. Let \(\theta = (d, \sigma_\epsilon^2, \sigma_\eta^2, \phi)'\). As in the additive white noise case in Section 3.1.3.2, the entire vector \(X_{0T}^1\) augments the parameter vector. From (B.2), the joint density of \(Y_T, X_{0T}^1\) and \(\theta\) is (compare (3.8))

\[
f(Y_T, X_{0T}^1, \theta) = f(Y_T \mid X_{0T}^1, \theta)f(X_{0T}^1 \mid \theta)\pi(\theta)
= \left\{\prod_{t=2}^{T} f(y_t \mid y_{t-1}, X_t^1, \theta)\right\}\left\{\prod_{t=1}^{T} f(x_t \mid x_{t-1}, x_{t-2}, \theta)\right\} f(y_1 \mid X_1, \theta)
\times f(x_{-1}, x_0 \mid \theta)\pi(\theta),
\]

where

\[
f(y_t \mid y_{t-1}, X_t^1, \theta) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp \left\{ -\frac{(y_t + \phi y_{t-1} - (\phi b_2^*, \phi b_1^* + b_2^*, \phi + b_1^*, 1)X_t^1)^2}{2\sigma_\epsilon^2} \right\},
\]

\[
f(y_1 \mid X_1, \theta) = \frac{1}{\sqrt{2\pi\sigma_\eta^2/((1 - \phi^2)}} \exp \left\{ -\frac{(1 - \phi^2)(y_1 - (b_2, b_1, 1)X_1)^2}{2\sigma_\eta^2} \right\}.
\]

The sampling scheme for \(\sigma_\eta^2\) is the same as before. For \(d\), the Metropolis-Hastings algorithm is also the same, except \(\{y_t - (b_2^*, b_1^*, 1)X_t\} \) is replaced by \(\{y_t + \phi y_{t-1} - (\phi b_2^*, \phi b_1^* + b_2^*, \phi + b_1^*, 1)X_t^1\} \) in (3.9). For \(\sigma_\epsilon^2\), the conditional distribution is still an inverse gamma distribution though an extra term \(f(y_1 \mid X_1, \theta)\) is involved. In this case, the state space form still holds so that the entire latent vector \(X_{0T}^1\) can be generated in one step as before.
Assume the prior for $\phi$ is uniform on $(-1, 1)$. The conditional distribution of $\phi$ satisfies

$$f(\phi \mid Y_T, X_{0,T}, d, \sigma_n^2, \sigma_2^2) \propto \left\{ \prod_{t=2}^{T} f(y_t \mid y_{t-1}, X_t^1, \Theta) \right\} \left\{ f(y_1 \mid X_1, \Theta) 1_{(-1,1)}(\phi) \right\}$$

$$\propto \exp \left\{ -\frac{\sum_{t=1}^{T} (y_t + \phi y_{t-1} - (\phi b_2, \phi b_1^1 + b_2, \phi + b_1^1, 1)X_t^1)^2}{2\sigma_2^2} \right\}$$

$$\times \sqrt{1 - \phi^2} \exp \left\{ -\frac{(1 - \phi^2)(y_1 - (b_2, b_1, 1)X_1)^2}{2\sigma_2^2} \right\}$$

(B.3)

Except for the term $(1 - \phi^2)$, the log of (B.3) is a quadratic function of $\phi$. Therefore, a rejection algorithm can be conducted using the corresponding Gaussian density in (B.3) (the term $1 - \phi^2$ is ignored) with the acceptance probability $1 - \phi^2$. The acceptance probability is 1 when $\phi = 0$ and decreases as $|\phi| \to 1$. This step is simpler than that for ARFIMA(1, d, 0) since $\phi$ is generated from its conditional distribution via simple rejection as opposed to being generated via the Metropolis algorithm. The generalization for $p > 1$ is straightforward.
BIBLIOGRAPHY


Hsu, N.-J. and Breidt, F.J. (1997), Bayesian analysis of long memory stochastic volatility models, Preprint 97–4, Statistical Laboratory, ISU.


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


ACKNOWLEDGEMENTS

I wish to express my deep gratitude to my major professor Dr. Jay Breidt for his patient guidance, constant encouragement and wonderful research ideas. I am also thankful to my committee members, Dr. Wayne Fuller, Dr. Soumendra N. Lahiri, Dr. Hal Stern and Dr. Barry Falk for their helpful suggestions.

I thank my family for their love and support all through my study. This dissertation is dedicated to them with my deep appreciation.

This research was supported by the Office of Naval Research Grant #N000149610279.