DISSERTATION

SOME OBSERVATION DRIVEN MODELS FOR TIME SERIES

Submitted by

Sarah Streett

Department of Statistics

In partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2000

UMI Number: 3002102

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

August 31, 2000

WE HEREBY RECOMMEND THAT THE DISSERTATION SOME OBSERVATION DRIVEN MODELS FOR TIME SERIES PREPARED UNDER OUR SUPERVISION BY SARAH STREETT BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Committee Member

Committee Member

Committee Member

Co-adviser

Adviser

Department Head

ABSTRACT

SOME OBSERVATION DRIVEN MODELS FOR TIME SERIES

We begin by reviewing generalized state-space models and the two categories into which they are typically divided, parameter driven and observation driven models. Since the models considered throughout the remainder of the thesis are observation driven, several examples of processes of this type are given. In Chapter 2 stationarity properties for two families of observation driven models are derived using results from Meyn and Tweedie (1993). The first family of models, BIN models, were developed by Rydberg and Shephard (1999) to analyze the number of trades occuring within a given time interval. We also consider a class of GLARMA models for modeling time series of counts. We show that a particular variant of a GLARMA model is uniformly ergodic. This enables us to use a procedure known as exact sampling to sample from the stationary distribution. In Chapter 3 we develop stationarity properties for a process used by Rydberg and Shephard (1998) for modeling stock prices. In the final Chapter, we return to the GLARMA models of Chapter 2. We calculate the maximum likelihood estimates of the model parameters and derive their asymptotic distribution. We also look at simulations as well as fit this model to a data set of asthma counts in order to determine how the theory applies in practice.

Sarah Streett
Department of Statistics
Colorado State University
Fort Collins, Colorado 80523
Fall 2000

ACKNOWLEDGEMENTS

Richard Davis, my advisor, deserves my deepest gratitude for guiding me through the doctoral process. He has provided answers to my questions, supported my ideas and challenged my intellect. I feel very lucky to have had him as my mentor.

I would also like to express my thanks to Richard Tweedie, my co-advisor, for his invaluable insight. His help and support have been greatly appreciated.

William Dunsmuir has provided contributions to the thesis for which I would like to thank and acknowledge him.

While attending Colorado State University, I have had the opportunity to learn from many professors, two of whom I feel especially grateful for having met. I want to thank Duane Boes and Hari Iyer for always taking the time to answer questions and to offer encouragement; they have made my time at CSU all the more enjoyable.

I would like to thank the Department of Statistics, the National Science Foundation (DMS 9803682 and DMS 9972015) and the United States Department of Agriculture (GC-111953) for supporting me for five years through teaching, consulting and research.

Lastly, I want to thank Adam for being my inspiration.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Motivation

Recent developments in models for describing non-Gaussian time series have been the motivating factor behind much of the thesis. Generalized state space models, which provide a flexible framework for representing data of this type, have become popular for modeling time series. State-space models are characterized by two equations: the observation equation and the state equation. The observation equation describes the evolution of the observed variable at time t as a function of the state variable at time t. If we let Y_t represent the observation variable, W_t represent the state variable, $\mathbf{Y}^{(t)} = (Y_1, \ldots, Y_t)'$ and $\mathbf{W}^{(t)} = (W_1, \ldots, W_t)'$, then the observation equation satisfies the following relationship:

$$f(y_t|w_t) = f(y_t|\mathbf{w}^{(t)}, \mathbf{y}^{(t-1)}), \quad t = 1, 2, \ldots .$$

That is, Y_t, conditioned on the past and present states and the past observations, depends only on the present state, and is "independent" of the past states and observations.

The state equation governs the manner in which the state variable evolves based on past states and observations. The form of this equation depends on whether the model is parameter driven or observation driven, the two categories into which generalized state-space models are typically divided as discussed by Cox (1981). For observation driven models, the evolution of the state vector depends on past observations. This is not the case for parameter driven models. For these models, the dependence results from an underlying latent process. Although the basic properties for each model type will be described in this introduction, the focus of the thesis is on models which are observation driven.

1.2 Parameter Driven Models

A similar independence assumption as that used in defining the conditional density of the observation variable is made for the state variable in a parameter driven model. Here we assume that the conditional density of W_{t+1} given W_t is the same as the conditional density of W_{t+1} given $(\mathbf{W}^{(t)}, \mathbf{Y}^{(t)})$, i.e.,

$$f(w_{t+1}|w_t) = f(w_{t+1}|\mathbf{w}^{(t)}, \mathbf{y}^{(t)}), \quad t = 1, 2, \ldots .$$

Note that this implies that the state process, $\{W_t\}$, is Markov.

Assuming that the initial state W_1 has probability density f_1, the joint density of the observation and state variables can be computed as follows:

$$\begin{aligned} f(y_1, \ldots, y_n, w_1, \ldots, w_n) &= f(y_n|w_n, \mathbf{w}^{(n-1)}, \mathbf{y}^{(n-1)}) f(w_n, \mathbf{w}^{(n-1)}, y^{(n-1)}) \\ &= f(y_n|w_n) f(w_n|\mathbf{w}^{(n-1)}, \mathbf{y}^{(n-1)}) f(\mathbf{y}^{(n-1)}, \mathbf{w}^{(n-1)}) \\ &= f(y_n|w_n) f(w_n|w_{n-1}) f(\mathbf{y}^{(n-1)}, \mathbf{w}^{(n-1)}) \\ &\;\;\vdots \\ &= \left(\prod_{j=1}^{n} f(y_j|w_j)\right)\left(\prod_{j=2}^{n} f(w_j|w_{j-1})\right) f_1(w_1), \end{aligned}$$

and since W_t is Markov,

$$f(y_1, \ldots, y_n|w_1, \ldots, w_n) = \left(\prod_{j=1}^{n} f(y_j|w_j)\right).$$

Therefore, $Y_1, \ldots Y_n$ are conditionally independent given $W_1, \ldots W_n$. Equivalently, the dependence of the observation variables arises from an unobservable latent process involving the state variables.

One of the drawbacks in choosing a parameter driven model is that the computation of the likelihood function is usually quite complex. To see this, consider the forecast equation

$$f(y_{t+1}|\mathbf{y}^{(t)}) = \int f(y_{t+1}|w_{t+1})f(w_{t+1}|\mathbf{y}^{(t)})d\mu(w_{t+1}). \tag{1.1}$$

To compute this, we need a means of computing $f(w_{t+1}|\mathbf{y}^{(t)})$. Using the independence assumption made above, this conditional density can be computed as follows:

$$f(w_{t+1}|\mathbf{y}^{(t)}) = \int f(w_{t+1}|w_t)f(w_t|\mathbf{y}^{(t)})d\mu(w_t).$$

However, this requires calculation of $f(w_t|\mathbf{y}^{(t)})$. Using the independence assumption again, as well as Bayes' theorem, we arrive at a recursive solution.

$$f(w_t|\mathbf{y}^{(t)}) = f(y_t|w_t)f(w_t|\mathbf{y}^{(t-1)})/f(y_t|\mathbf{y}^{(t-1)}).$$

To solve these recursions, it is assumed that $f(w_1|\mathbf{y}^{(0)}) = f_1(w_1)$. It is easily seen how computation of the likelihood function can be complicated by these recursions. As a result, estimation of the model parameters as well as prediction of future observations can be difficult for parameter driven models. Although there are simulation based techniques for calculating the recursions involved in computing the likelihood, they are not easily implemented, complicating the model selection process for this class of models.

One of the benefits from using a parameter driven model is that the observation process typically inherits properties such as stationarity and ergodicity from the state process. Asymptotic results for estimates of the parameters in the model are therefore usually easier to derive for parameter driven rather than observation driven models.

Another benefit arising from the use of this class of models is that the parameters often have meaningful interpretations. This is a result of the marginal expectation of the observation variable often being a function involving only the parameters of interest to the researcher.

1.3 Observation Driven Models

For models from this class, we allow the state variable to depend on past observations. The conditional density of the state variable is specified as follows:

$$f(w_{t+1}|\mathbf{y}^{(t)}) = f_{W_{t+1}|\mathbf{Y}^{(t)}}(w_{t+1}|\mathbf{y}^{(t)}), \quad t = 0, 1, \ldots,$$

where $f(w_1|\mathbf{y}^{(0)}) := f_1(w_1)$ for some specified initial density $f_1(w_1)$. By defining the conditional density of the state variable in this manner, we are able to compute the forecast density (1.1) directly. This simplifies computation of the likelihood function,

$$f(y_1, \ldots, y_n) = \prod_{i=1}^{n} f(y_i \mid \mathbf{y}^{(n-i)}).$$

Therefore, estimation of the model parameters using maximum likelihood and prediction of future observations are generally easier for observation rather than parameter driven models.

One of the major drawbacks to using observation driven models is that establishing properties such as stationarity and ergodicity for the observation process is not a trivial matter for this class of models. A large part of the thesis is devoted to proving asymptotic results of this nature for observation driven models where the conditional density of the observation variable given the state variable is a member of the exponential family.

Another drawback to models from this class is that the marginal expectation of the observation variable often involves unknown parameters of the state process. Thus, meaningful interpretation of the model parameters, a key component of interest to the researcher, may be quite difficult.

1.4 Exponential Family Models

A convenient and flexible choice for the conditional distribution in the observation equation is the exponential family of distributions. Many well-known statistical distributions such as the binomial, Poisson, exponential, gamma and normal distributions belong to this family. The following canonical form for the density of an exponential family member will be used throughout the dissertation:

$$f(y_t|w_t) = \exp\left[(y_t w_t - b(w_t))\phi^{-1} + d(y_t)\right].$$

With this parameterization, it follows that (see McCullagh and Nelder (1989))

$$E(Y_t|W_t) = \mu_t = \dot{b}(W_t), \quad V(Y_t|W_t) = \phi\ddot{b}(W_t),$$

where $\dot{b}$ and $\ddot{b}$ denote the first and second derivatives of the function $b(\cdot)$.

1.5 Examples of Observation Driven Models

In subsequent chapters, the concentration will be on developing conditions under which the properties of stationarity and ergodicity may be obtained for observation driven models. For this reason, several examples of observation driven models and their applications are given below. In these examples, the distribution of the observation variable is a member of the exponential family. The dependence of the state variable on the observation variable is specified through a link function, $g(\cdot)$, on $\dot{b}(W_t)$. Thus, in the examples that follow, the state equation has the following structure:

$$g\{\dot{b}(W_t)\} = \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t,$$

where $\mathbf{x}_t$ is a vector of covariates and ϵ_t is a function of the observation vector, $\mathbf{Y}^{(t)}$, and possibly also of the state vector, $\mathbf{W}^{(t-1)}$, and past covariate values. The link function used in most of the examples is the canonical link, the special case where $g\{\dot{b}(x)\} = x$.

1.5.1 Observations with Constant Coefficient of Variation

For modeling outcomes with constant coefficient of variation, Zeger and Qaqish (1988) assume that the conditional density of the observation variable at time t is gamma with mean μ_t and variance μ_t^2/r. They use the canonical link for the gamma distribution, i.e., $g(x) = 1/x$, to define the state equation. Thus,

$$f(y_t|w_t) = e^{(y_t w_t - \ln w_t)(-r) + r\ln(ry_t) - \ln y_t - \ln\Gamma(r)} \tag{1.2}$$

and

$$W_t = \mu_t^{-1} = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{p} \theta_i[(\max(Y_{t-i}, c))^{-1} - \mathbf{x}_{t-i}'\boldsymbol{\beta}].$$

where c is a positive constant. Realizations of the observation process (Y_t) with $\mathbf{x}_t'\boldsymbol{\beta} = 1$, $p = 1, c = 0.05, \theta = 0.8$ and scale parameter $r = 1, 25, 100$ are shown in Figure 1.1. Notice that as r increases, the realizations become closer to Gaussian.

Zeger and Qaqish applied this model in analyzing the interspike times of a motor cortex neuron of a monkey as studied by Dr. A. Georgopoulos of Johns Hopkins University. Their objective was to estimate the average time between spikes as well as the time dependency among successive spikes. The Gaussian assumption was clearly not met for this time series. A plot of the data (see Zeger and Qaqish (1988)) shows that the variance is much larger when the interspike times are large and that the data is skewed toward greater time lapses between spikes. The data also appears to be correlated. They note that although a logarithmic transformation validates the assumption of normality in the case of studying a single neuron, a typical study involves several neurons and this approach is not reasonable. Instead, they suggest applying the model given by equation (1.2) using a quasi-likelihood approach to estimate the parameters. They fit two second order Markov models with $c = 0$, one with just an intercept term, the other with an intercept and linear trend regression term. Due to the structure of the state process, the parameters have meaningful interpretations in this modeling framework. Here, the intercept is interpreted as

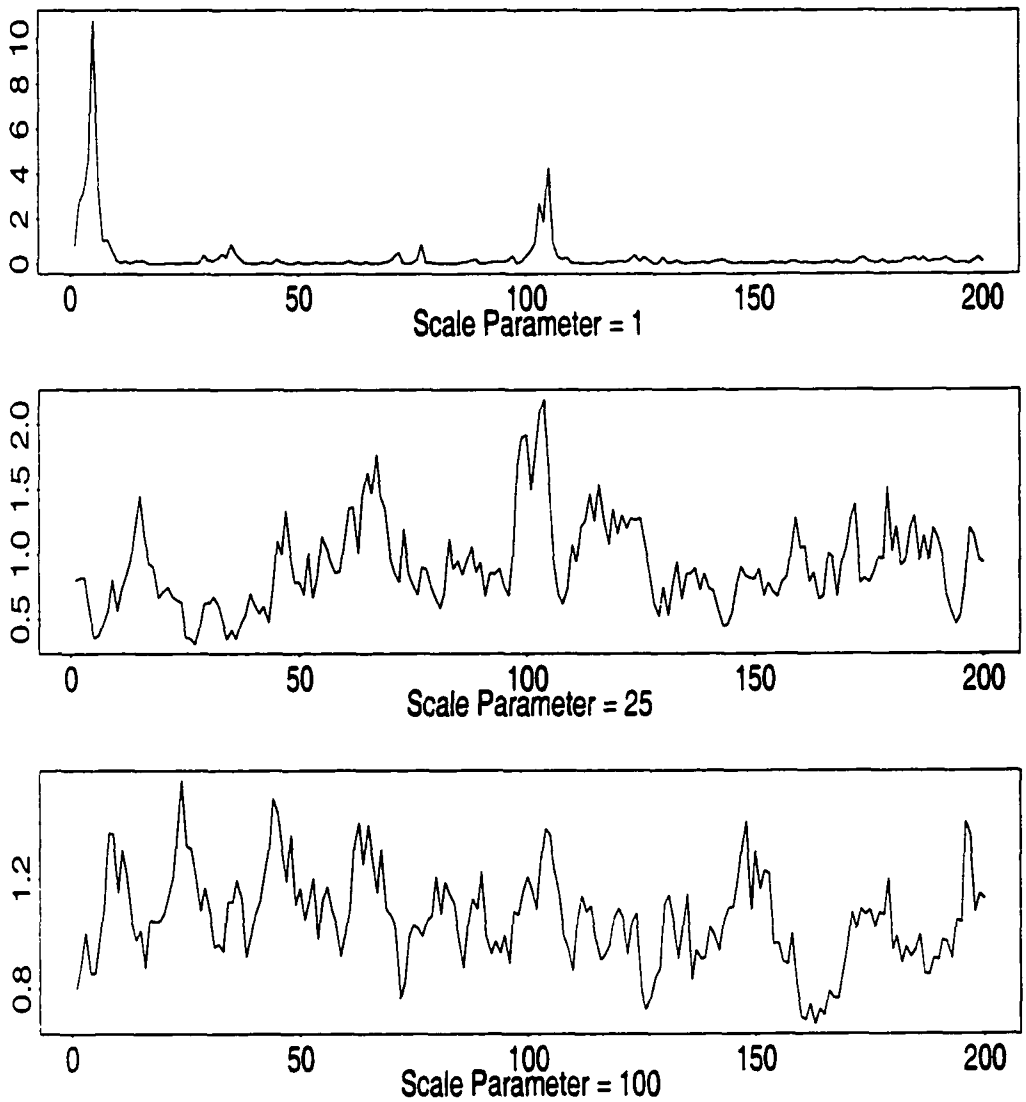

Figure 1.1: Realizations of an observation process with constant coefficient of variation. Equation 1.2 with $\mathbf{x}_t'\boldsymbol{\beta} = 1$, $p = 1, c = 0.05, \theta = 0.8$ and scale parameter $r = 1, 25, 100$.

the expected neuron firing rate and the coefficient of the covariate as the change in firing rate due to an additional spike.

1.5.2 ARCH Models

Engle (1982) developed the Autoregressive Conditional Heteroscedastic (ARCH) regression model for modeling volatility in economic data. An ARCH model is defined by the following equation:

$$(Y_t - \mathbf{x}_t'\boldsymbol{\beta}) = \sqrt{\alpha_0 + \sum_{i=1}^{p} \alpha_i (Y_{t-i} - \mathbf{x}_{t-i}'\boldsymbol{\beta})^2} \cdot Z_t, \tag{1.3}$$

where $\{Z_t\}$ is a sequence of independent, identically distributed standard normal random variables. In addition to the ability of the ARCH process to model changing variability in the data, it also models other behaviors common to financial time series well. These behaviors include little or no correlation present among observations but high correlation among the absolute values and squares of the observations, heavy-tailedness, and threshold exceedances which appear in clusters.

When $p = 1$, various stationarity results hold for $\{Y_t - \mathbf{x}_t'\boldsymbol{\beta}\}$ depending on the values of α_1. For $\alpha_1 = 0$, $\{Y_t - \mathbf{x}_t'\boldsymbol{\beta}\}$ is iid noise. For $|\alpha_1| \in (0, 1)$, $\{Y_t - \mathbf{x}_t'\boldsymbol{\beta}\}$ is strictly stationary with finite variance; hence, it is also weakly stationary. Lastly, for $1 \le \alpha_1 < 2e^{\gamma}$, ($\gamma$ represents Euler's constant), $\{Y_t - \mathbf{x}_t'\boldsymbol{\beta}\}$ is strictly stationary with infinite variance. Stationarity properties are much more difficult to establish when $p > 1$. The reader is referred to Engle, (1982) for further background and discussion of ARCH processes.

A simulated sample path from Equation 1.3 with $p = 1, \alpha_0 = 0.75, \alpha_1 = 1.5$ and $\mathbf{x}_t'\boldsymbol{\beta} = 1$ is shown in Figure 1.2. The corresponding sample autocorrelations for the simulated values, the absolute value of the simulated values and the square of the simulated values are given in the figure as well. Upon comparison of the autocorrelations, one can see that there is very little correlation among the observed values but quite strong correlation among the absolute values and squares of the data.

1.5.3 GLARMA Models

In a recent paper, Rydberg and Shephard (1998) consider modeling price movement for a given asset traded on a stock exchange. They decompose this process

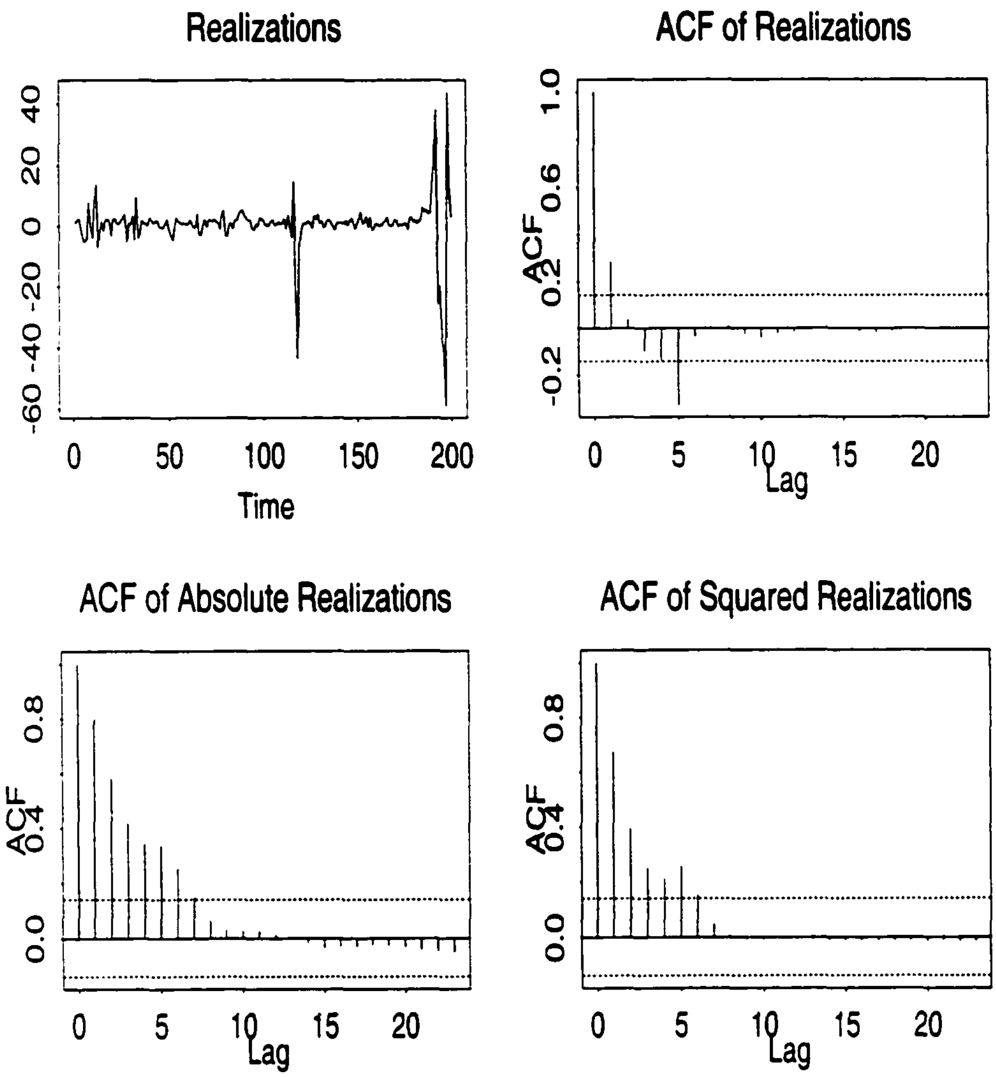

Figure 1.2: Realizations and ACF of an ARCH process. Equation 1.3 with $p = 1, \alpha_0 = 0.75, \alpha_1 = 1.5$ and $\mathbf{x}_t'\boldsymbol{\beta} = 1$. The autocorrelation function for the series, the absolute value of the series and the squared values of the series are also given.

into three pieces: a process modeling activity (whether or not the price changes), a process modeling the direction of the price changes, and a process modeling the magnitude of the price changes.

They use an autologistic model for the parameter of the Bernoulli process which describes whether or not a price change has occurred. Thus, if $N_t \in \{0, 1\}$ is the event of a price change at time t, then

$P(N_t = 1|\theta_t) = p(\theta_t) =$ the probability that the price moves at time t, and

$P(N_t = 0|\theta_t) = 1 - p(\theta_t) =$ the probability that the price does not move at time t,

where

$$p(\theta_t) = \frac{e^{\theta_t}}{1 + e^{\theta_t}}, \quad \theta_t = \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t.$$

Note that the canonical link for the Bernoulli distribution, the logit function is used as the link function for this case.

Rydberg and Shephard suggest using a generalized linear autoregressive moving average (GLARMA) model to model ϵ_t. The GLARMA they employ is as follows:

$$\epsilon_t = \sum_{i=1}^{p} \phi_i \epsilon_{t-i} + \sigma V_t + \sigma \sum_{j=1}^{q} \delta_j V_{t-j}, \quad \sigma > 0$$

where

$$V_t = \frac{N_{t-1} - p(\theta_{t-1})}{\sqrt{p(\theta_{t-1})(1 - p(\theta_{t-1}))}}.$$

This model has the desirable property that $\{V_t\}$ is a martingale difference sequence with unit conditional variance. At this point, it is natural to question under what conditions a stationary solution, π, exists for the process $\{\epsilon_t\}$. In addition, if such a solution exists, at what rate does the chain converge to this distribution? Answers to these questions will be discussed in Chapter 3 for a variation of this model.

1.5.4 BIN Models

The next model we introduce was also proposed by Rydberg and Shephard (1999) as part of their modeling framework for analyzing the return rate of an investment. Defining $p(t)$ to be the price of an asset at time t, they model $p(t)$ as follows:

$$p(t) = p(0) + \sum_{i=1}^{N(t)} Z_i,$$

where $\{N(t)\}_{t\geq 0}$ gives the number of trades recorded up to time t and Z_i is the price change associated with the i^{th} transaction. Their proposed model for the process $\{Z_t\}$ was introduced in Section 1.5.3 and is discussed in more detail in Chapter 3. A function of the process $\{p(t)\}$ which is of particular interest is

$$\begin{aligned} p_t &= p((t+1)\Delta-) - p(t\Delta) \\ &= \sum_{i=1}^{N[(t+1)\Delta-]} Z_i - \sum_{i=1}^{N(t\Delta-)} Z_i \\ &= \sum_{i=N(t\Delta-)+1}^{N[(t+1)\Delta-]} Z_i. \end{aligned}$$

where $\Delta > 0$ is a specified length of time. Note that p_t models the rate of return on the investment over a fixed time period.

From the definition of p_t, it is obvious that the process $N_t := N[(t+1)\Delta-] - N(t\Delta)$, the number of transactions occuring in the interval $[n\Delta, (n+1)\Delta)$, plays an important role in the modeling framework. It is this process, $\{N_t\}$, on which we focus our attention in Section 2.1. Note that if Δ is small, little or no information will be lost by studying the process $\{N_t\}$ rather than $\{N(t)\}$. The idea behind this modeling approach is that we "bin" our time interval into sections of the form $[t\Delta, (t+1)\Delta)$ and count the number of trades in each (given by N_t). Therefore, Rydberg and Shephard refer to their proposed models for $\{N_t\}$ as BIN(p,q) models. These models take the following form:

$$\begin{aligned} N_t \mid \mathcal{F}_{t\Delta-} &\sim \text{Poisson}(\lambda_t), \\ \lambda_t &= \alpha + \sum_{j=1}^{p} \gamma_j N_{t-j} + \sum_{j=1}^{q} \delta_j \lambda_{t-j}, \end{aligned}$$

where $\alpha, \gamma_j, \delta_j > 0$, $1 \leq j \leq \max(p,q)$ and $\mathcal{F}_{t\Delta-}$ is the information available up to an infinitesimal time before $t\Delta$. The BIN(1):=BIN(1,0) model and the BIN(1,1) model are discussed in greater detail in Section 2.1.

1.5.5 Models for Count Data

Observation driven models are also very useful for modeling time series of count data. There are many models which have been specified in this case, several of these are discussed below.

Wong (1986) suggests the following model for analyzing count data:

$$Y_t|W_t \sim \text{Poisson}(\mu_t),$$

where

$$\mu_t = W_t = \mu[1 + e^{-\theta_0 - \theta_1 Y_{t-1}}], \ \ \theta_i > 0. \tag{1.4}$$

One of the drawbacks to this model, as discussed in Zeger and Qaqish (1988), is that the previous outcome can affect the present conditional expectation by at most a factor of 2μ. However, this constraint also allows ergodicity of the chain to be established without much difficulty.

Zeger and Qaqish (1988) propose two alternatives to Wong's model. For both, they use the canonical link for the Poisson distribution, the logarithmic function. The first model they suggest is as follows:

$$\log(\mu_t) = W_t = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{q} \theta_i[\log(\max(Y_{t-i}, c)) - \mathbf{x}_{t-i}'\boldsymbol{\beta}]$$

where c, satisfying $0 < c < 1$, prevents 0 from becoming an absorbing state. In the case where $q = 1$, c determines the probability that $y_t > 0$ given $y_{t-1} = 0$. The other alternative they propose is

$$\log(\mu_t) = W_t = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{q} \theta_i\{\log(Y_{t-i} + c) - \log[e^{\mathbf{x}_{t-i}'\boldsymbol{\beta}} + c]\}.$$

When $q = 1$,

$$\mu_t = e^{\mathbf{x}_t'\boldsymbol{\beta}} \left[\frac{Y_{t-1} + c}{e^{\mathbf{x}_{t-1}'\boldsymbol{\beta}} + c}\right]^{\theta_1}$$

and c may be thought of as an "immigration" rate. Ergodicity for $\{W_t\}$ can be established in both cases when $\mathbf{x}_t'\boldsymbol{\beta} = \mu > 0$, $q = 1$ and $0 < \theta_1 < 1$ using results for

Markov chains found in Meyn and Tweedie (1993). For both models, estimation of the parameter c could be difficult. Also, aside from the case $q = 1$, the meaning of c is not clear.

Davis, Wang and Dunsmuir (1999) consider yet another variation. They propose the following model for Poisson counts:

$$\log(\mu_t) = W_t = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{\infty} \tau_i e_{t-i} \tag{1.5}$$

where

$$e_t = \frac{Y_t - \mu_t}{\mu_t^{\lambda}}, \quad \lambda \geq 0 \text{ fixed} \tag{1.6}$$

and

$$\begin{aligned} \sum_{i=1}^{\infty} \tau_i &= \left(1 - \sum_{i=1}^{p} \phi_i z^i\right)^{-1} \left(1 + \sum_{i=1}^{q} \theta_i z^i\right) - 1 \\ &= \phi(z)^{-1}\theta(z) - 1. \end{aligned} \tag{1.7}$$

Observe that $\sum_{i=1}^{\infty} \tau_i e_{t-i}$ is the one-step ahead predictor of e_t based on an ARMA(p,q) model. As this model is quite similar to the one proposed by Shepard (Section 1.5.3), we will also refer to it as a GLARMA model.

Note that $\{e_t : t \leq s-1\}$, given initial conditions $\{\mu_t : -\max(p,q)+1 \leq t \leq 0\}$ is equivalent to $\{Y_t : t \leq s-1\}$ or $\{\mu_t : t \leq s-1\}$. It then follows that the e_t form a martingale difference sequence since

$$E(e_s|\mathcal{F}_{s-1}^e) = 0,$$

where $\mathcal{F}_{s-1}^e$ is the σ-algebra generated by $\{e_t : t \leq s-1\}$. Hence, the e_t have zero mean and variance

$$E(e_t^2) = E[E(e_t^2|\mu_t)] = E[\mu_t^{1-2\lambda}],$$

which, for $\lambda = 0.5$, is unity. Also from the martingale difference property we have that the e_t are uncorrelated.

From the above properties we have, for any λ,

$$E(W_t) = x_t'\beta$$

which is a desirable property for the log mean. Also,

$$\text{Var}(W_t) = \sum_{i=1}^{\infty} \tau_i^2 E[\mu_{t-i}^{1-2\lambda}]$$

and for $s = t + l$, $l > 0$,

$$\text{Cov}(W_t, W_s) = \sum_{i=1}^{\infty} \tau_i \tau_{i+l} E[\mu_{t-i}^{1-2\lambda}].$$

Again, if $\lambda = 0.5$, the covariances do not depend on time t.

Conditions for stationarity and ergodicity for models of this type will be discussed in Chapter 2. Realizations of the process (1.5) with $p = 0$, $q = 1$, $x_t'\boldsymbol{\beta} = 0$, $\lambda \in \{0.5, 1\}$ and $\theta_1 \in \{0.25, 0.75\}$ are given in Figure 1.3. Note that as θ_1 increases, the dispersion increases. For the Poisson distribution, the estimated dispersion should be close to 1. Here, however, for $\theta_1 = 0.75$, the dispersion is estimated to be 13.27. This gives a simple illustration of the advantage that these types of models have in analyzing data which is over-dispersed.

1.6 Overview

The remainder of the thesis focuses on developing asymptotic properties for the models discussed in Sections 1.5.3-1.5.5. In Chapter 2, we consider models where the observation process has a Poisson distribution whose mean is a function of the state process. Specifically, we study the models discussed in Sections 1.5.4 and 1.5.5. Using results from Meyn and Tweedie (1993), we show that a unique stationary distribution exists for the BIN(1) and BIN(1,1) models (Section 1.5.4). We also derive stationarity properties for two variations of the GLARMA models for Poisson counts defined in (1.5). For one of these variations, we are able to establish

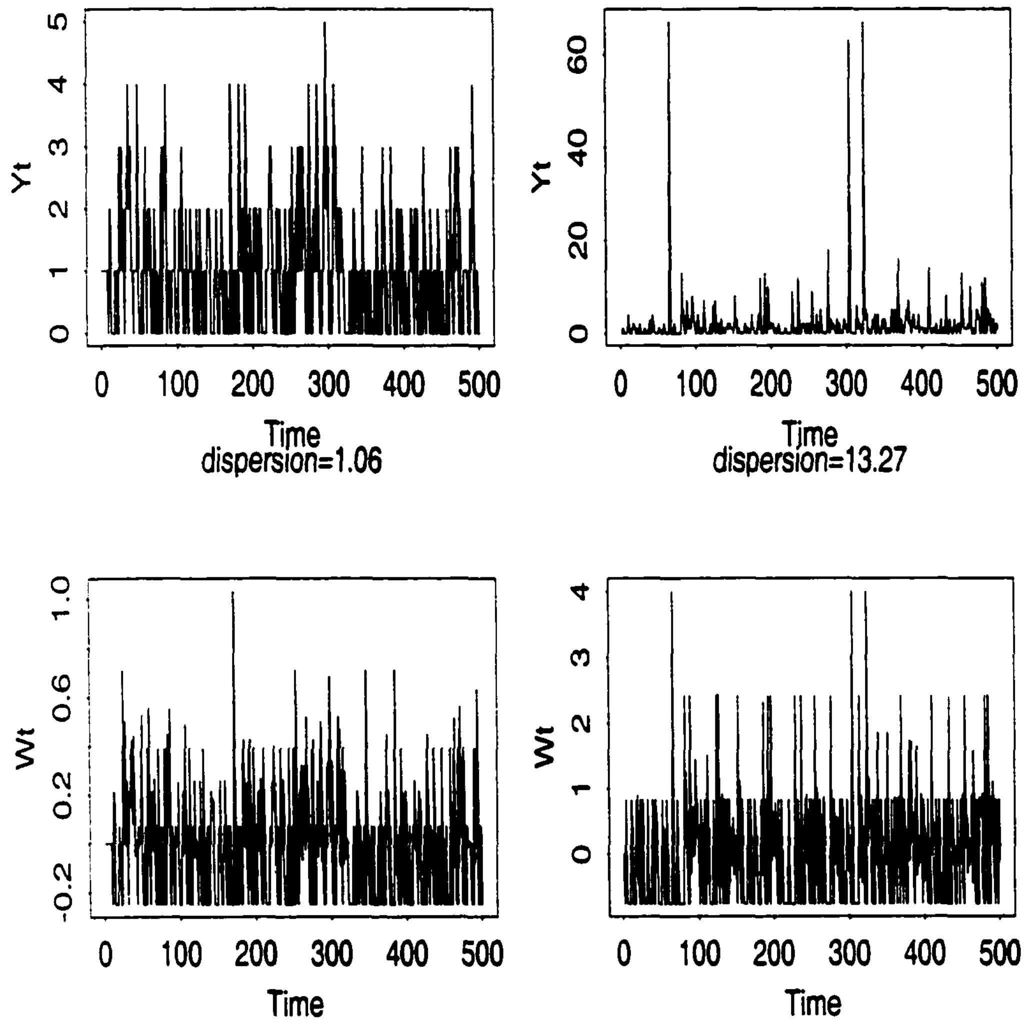

Figure 1.3: Realizations from a model for count data. Equation 1.5 with $p = 0$, $q = 1$, $x_t'\beta = 0$ and $\lambda = 0.5$, $\theta_1 = 0.25$ (top, left), $\theta_1 = 0.75$ (top, right), $\lambda = 1$, $\theta_1 = 0.25$ (bottom, left), $\theta_1 = 0.75$ (bottom, right).

that the process is uniformly ergodic. This enables us to use a procedure created by Murdoch and Green (1998) to sample exactly from the stationary distribution.

In Chapter 3, we consider Rydberg and Shephard's GLARMA models (1998) for modeling price activity. Again, we are able to establish the existence of a unique stationary distribution. Here, the key element of the argument is to show that the process has transition probabilities which are equicontinuous.

In the last chapter, we return to the models for count data given by (1.5). We consider maximum likelihood estimates for the model parameters and derive their asymptotic distribution. Simulations are then considered to support the derived theory. Lastly, we fit these models to a data set consisting of asthma cases recorded at a hospital outside of Sydney, Australia. The data is presented in Figure 1.4. Upon inspection, it is easily seen that this is a good candidate for these models due to the overdispersion present in the data. One might also note the presence of a seasonal trend. This data set is discussed further in Chapter 4.

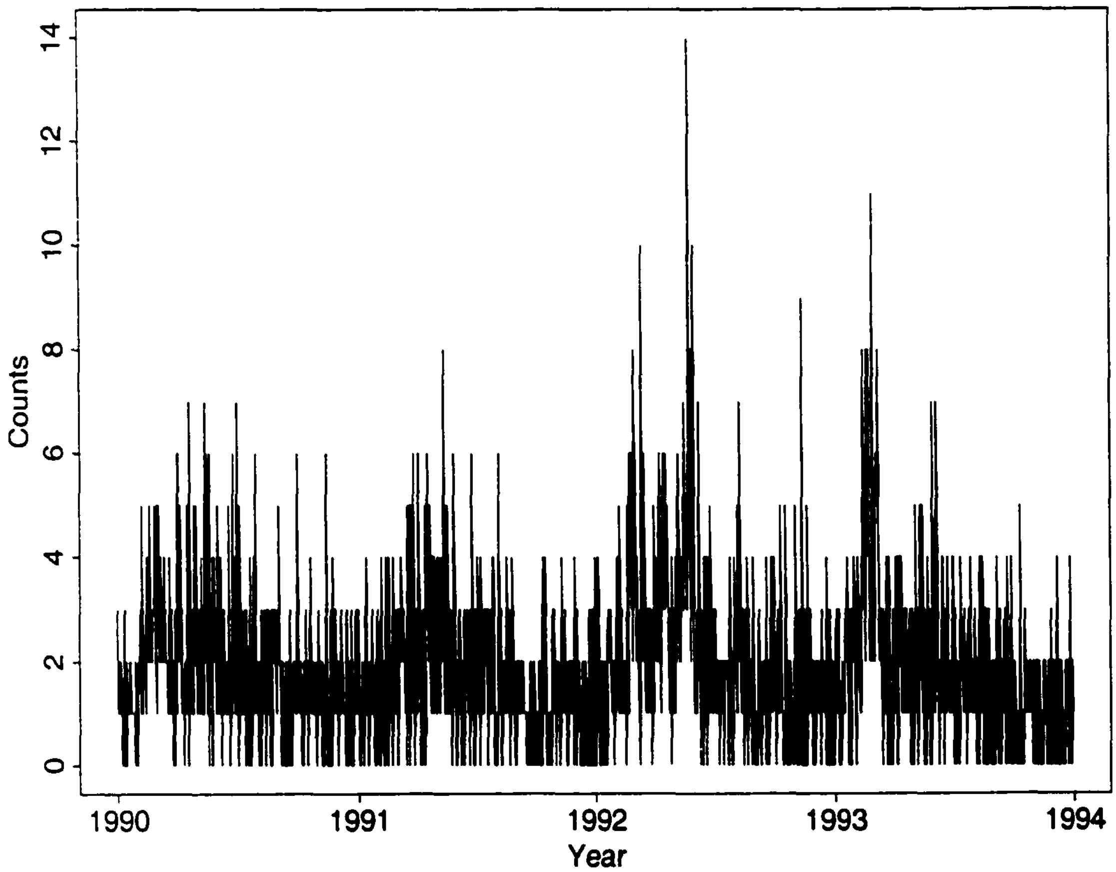

Figure 1.4: Asthma Data

Chapter 2

THE POISSON PROCESS

In this chapter, we derive stationarity results for two types of observation driven models for time series of counts. The first family of models we consider are BIN models which were introduced in Section 1.5.4. The second model family, which we call GLARMA models after Rydberg and Shephard (1998), were described in Section 1.5.5. As well as establishing stationarity properties for these models, we show how to sample from the stationary distribution exactly for the case of the existence of a unique stationary distribution using an algorithm called the "mutigamma" coupler (Murdoch and Green, 1998).

2.1 BIN Models

As mentioned in Section 1.5.4, the task of modeling the rate of return on an investment can be modeled as

$$p_n = \sum_{i=N(n\Delta-)+1}^{N[(n+1)\Delta-]} Z_i,$$

where Z_i is the price change associated with the i^{th} transaction, $N(n)$ gives the number of trades recorded up to time t and Δ is a fixed unit of time. The process on which we focus our attention in this section is $N_n := N\big[(n+1)\Delta-\big] - N(n\Delta)$, the number of trades occurring in the interval $\big[n\Delta,\ (n+1)\Delta\big)$. Because their modeling procedure involves partitioning the time interval into "bins", Rydberg and Shephard (1999) refer to these models as BIN(p,q) models. Recall the BIN modeling framework:

$$N_n \mid \mathcal{F}_{n\Delta-} \sim \text{Poisson}(\lambda_n), \tag{2.1}$$
$$\lambda_n = \alpha + \sum_{j=1}^{p} \gamma_j N_{n-j} + \sum_{j=1}^{q} \delta_j \lambda_{n-j},$$

where $\alpha, \gamma_j, \delta_j \geq 0$, $1 \leq j \leq \max(p,q)$ and $\mathcal{F}_{n\Delta-}$ is the information available up to an infinitesimal time before $n\Delta$.

2.1.1 BIN(1) Model

We first consider the BIN(1) model:

$$\lambda_n = \alpha + \beta N_{n-1}. \tag{2.2}$$

α and $\beta > 0$, $\beta < 1$. Replacing N_n by $\lambda_n + u_n$, $u_n := N_n - \lambda_n$ and noting that $\{u_n\}$ is a Martingale difference sequence, $E(u_n \mid \mathcal{F}_{n\Delta-}) = 0$, it follows that

$$\begin{aligned} N_n &= \lambda_n + u_n \\ &= \alpha + \beta N_{n-1} + u_n, \end{aligned}$$

is a weakly stationary AR(1) type process with

$$\begin{aligned} E(N_n) &= E\big[E(N_n \mid \mathcal{F}_{n\Delta-})\big] \\ &= E\big[E(\alpha + \beta N_{n-1} + u_n \mid \mathcal{F}_{n\Delta-})\big] \\ &\vdots \\ &= E\left[E\left(\alpha \sum_{i=0}^{\infty} \beta^i + \sum_{i=0}^{\infty} \beta^i u_{n-i} \,\Big|\, \mathcal{F}_{n\Delta-}\right)\right] \\ &= \frac{\alpha}{1-\beta}, \end{aligned}$$

and

$$\text{Cov}(N_n,\ N_{n+h}) = \frac{\alpha \beta^{|h|}}{(1-\beta)(1-\beta^2)},$$

since

$$
\begin{aligned}
\mathrm{Var}(u_n) &= E\big[\mathrm{Var}(u_n \mid \mathcal{F}_{n\Delta-})\big] + 0 \\
&= E\big[E((N_n - \lambda_n)^2 \mid \mathcal{F}_{n\Delta-})\big] \\
&= E(\lambda_n) \\
&= E(N_n - u_n) \\
&= \frac{\alpha}{1-\beta}.
\end{aligned}
$$

Thus, the process $\{N_n\}$ defined by (2.1) with λ_n as defined in (2.2) is weakly stationary. In the next section, we show that the BIN(1,1) model, which includes the BIN(1) model, is strictly stationary.

2.1.2 BIN(1,1) Model

Here we define the state process $\{\lambda_n\}$ as follows:

$$\lambda_n = \alpha + \beta N_{n-1} + \gamma \lambda_{n-1}, \tag{2.3}$$

$\alpha, \beta, \gamma \geq 0$, $\beta + \gamma < 1$. It is easily seen that $\lambda_n \geq \frac{\alpha}{1-\gamma}$ by considering the following form for λ_n :

$$
\begin{aligned}
\lambda_n &= \alpha + \beta Y_{n-1} + \gamma(\alpha + \beta Y_{n-2} + \lambda_{n-2}) \\
&\vdots \\
&= \alpha \sum_{i=0}^{\infty} \gamma^i + \beta \sum_{i=1}^{\infty} \gamma^{i-1} Y_{n-i} \\
&= \frac{\alpha}{1-\gamma} + \beta \sum_{i=1}^{\infty} \gamma^{i-1} Y_{n-i}.
\end{aligned}
$$

Since $\beta, \gamma \geq 0$, it follows that $\lambda_n \geq \frac{\alpha}{1-\gamma}$. We use results from Meyn and Tweedie (1993) to establish the existence of a unique stationary distribution. The condition of ψ-irreducibility (Definition A.2.2) is an important one for establishing the existence of a unique invariant probability measure. However, when it is not known whether a chain possesses this property, one can show the existence of at least one stationary distribution by verifying that the chain is bounded in probability on average (Definition A.2.12). In addition, if the chain is an e-chain (Definition A.2.17) and a reachable state exists, the stationary measure is unique (Theorem A.2.18). We will begin by establishing that the mean process $\{\lambda_n\}$ has equicontinuous transition probabilities and is therefore an e-chain.

Proposition 2.1.1 *The Markov chain $\{\lambda_n\}$ is an e-chain.*

Proof: It suffices to show that for any continuous function f with compact support and $\delta > 0$, there exists an $\epsilon > 0$ such that $|P_x^k f - P_z^k f| < \delta$ for $|x - z| < \epsilon$ and $k = 1, 2, \ldots$. For $\delta > 0$ given, choose $\delta', \epsilon > 0$ sufficiently small such that $\delta' + \frac{2\epsilon}{1-\gamma} < \delta$ and $\left|f(x) - f(z)\right| < \delta'$ if $\left|x - z\right| < \epsilon$. Note that such a δ' and ϵ exist since f is uniformly continuous. Without loss of generality, we may assume $|f| \leq 1$, since otherwise, divide f by its maximum value. Let $\left|x - z\right| < \epsilon$ and recall that $p(n|x) = \frac{e^{-x}x^n}{n!}$. Then, for the case $k = 1$,

$$
\begin{aligned}
|P_x f - P_z f| &= \left|\sum_{n=0}^{\infty}\left[p(n|x)f(\alpha+\beta n+\gamma x) - p(n|z)f(\alpha+\beta n+\gamma z)\right]\right| \\
&\leq \sum_{n=0}^{\infty} p(n|x)\left|f(\alpha+\beta n+\gamma x) - f(\alpha+\beta n+\gamma z)\right| \\
&\quad + \sum_{n=0}^{\infty}\left|p(n|x) - p(n|z)\right|\left|f(\alpha+\beta n+\gamma z)\right| \\
&\leq \sum_{n=0}^{\infty}\frac{e^{-x}x^n}{n!}\left|f(\alpha+\beta n+\gamma x) - f(\alpha+\beta n+\gamma z)\right| \\
&\quad + \sum_{n=0}^{\infty}\left|f(\alpha+\beta n+\gamma z)\right|\left|\frac{e^{-x}x^n - e^{-z}z^n}{n!}\right| \\
&= \mathrm{I} + \mathrm{II}.
\end{aligned}
$$

Since $\left|(\alpha + \beta n + \gamma x) - (\alpha + \beta n + \gamma z)\right| = \gamma\left|x - z\right| < \epsilon$, we have $\mathrm{I} \leq \delta'$.

Now, for any $x > z$, we have

$$\begin{aligned}
\text{II} &\le \sum_{n=0}^{\infty} \frac{|e^{-x}x^n - e^{-z}z^n + e^{-x}z^n - e^{-x}z^n|}{n!} \\
&\le \sum_{n=0}^{\infty} \left[\frac{|e^{-x} - e^{-z}|z^n}{n!} + \frac{e^{-x}|x^n - z^n|}{n!} \right] \\
&= (e^{-z} - e^{-x})e^{z} + e^{-x}(e^{x} - e^{z}) = 2(1 - e^{-|x-z|}).
\end{aligned}$$

The same bound is also valid for the case $z \ge x$. Combining the inequalities given for I and II, it follows that

$$|P_x f - P_z f| \le \delta' + 2(1 - e^{-|x-z|}). \tag{2.4}$$

Now consider the case $k = 2$:

$$|P_x^2 f - P_z^2 f| = |P_x(Pf) - P_z(Pf)| = \sum_{n=0}^{\infty} |p(n|x)P_{x'}f - p(n|z)P_{z'}f|,$$

where $x' = \alpha + \beta n + \gamma x$ and $z' = \alpha + \beta n + \gamma z$,

$$\begin{aligned}
&= \sum_{n=0}^{\infty} \Big(p(n|x)\, |P_{x'}f - P_{z'}f| + P_{z'}f\, |p(n|x) - p(n|z)| \Big) \\
&\le \delta' + 2(1 - e^{-|x'-z'|}) + 2(1 - e^{-|x-z|}) \quad \text{(from 2.4)} \\
&= \delta' + 2(1 - e^{-|x-z|}) + 2(1 - e^{-\gamma|x-z|}).
\end{aligned}$$

Inductively, we obtain

$$\begin{aligned}
|P_x^k f - P_z^k f| &= |P_x(P^{k-1}f) - P_z(P^{k-1}f)| \\
&\le \sum_{n=0}^{\infty} \Big[p(n|x)\, |P_{x'}^{k-1}f - P_{z'}^{k-1}f| + P_{z'}^{k-1}f\ |p(n|x) - p(n|z)| \Big] \\
&\le \delta' + 2\sum_{i=0}^{n-1} (1 - e^{-\gamma^i|x-z|}) \\
&\le \delta' + 2\sum_{i=0}^{n-1} (1 - e^{-\gamma^i\epsilon}) \\
&\le \delta' + 2\sum_{i=0}^{n-1} \Big[1 - (1 - \gamma^i\epsilon + o(\gamma^{2i})) \Big] \\
&\le \delta' + 2\epsilon\sum_{i=0}^{n-1} \gamma^i \\
&\le \delta' + \frac{2\epsilon}{1-\gamma} < \delta,
\end{aligned}$$

which completes the proof. □

The next step toward obtaining our result is to show that the process $\{\lambda_n\}$ is bounded in probability on average. To do so, results from Glynn and Meyn (1997) will be used. The idea behind this theorem, stated as Theorem A.2.15 in the appendix, is that uniform integrability of the return time to a properly chosen set $A \subseteq X$ is sufficient for establishing tightness of the averaged transition probabilities assuming certain restrictions on the set A. The third condition of the theorem places the necessary structure on A while the first and second conditions verify the uniform integrability condition.

Proposition 2.1.2 *The chain* $\{\lambda_n\}$ *is bounded in probability on average.*

Proof: It will suffice to show that the conditions of Theorem A.2.15 hold.

Condition 1: There exists a non-negative function V such that $\Delta V(x) := EV(x) - V(x) \leq -1 + bI_A(x)$, where $b = 1 + \alpha$ and $A = \left[\frac{\alpha}{1-\gamma}, \frac{\alpha+1}{1-(\beta+\gamma)}\right]$.

To see this, define $V(x) = x$ and recall that the state space is $\left[\frac{\alpha}{1-\gamma}, \infty\right)$. Then,

$$\begin{aligned}
\Delta V(x) &= E\left[\lambda_n \mid \lambda_{n-1} = x\right] - x \\
&= E\left[\alpha + \beta Y_{n-1} + \gamma\lambda_{n-1} \mid \lambda_{n-1} = x\right] - x \\
&= \alpha + x(\beta + \gamma - 1) \\
&\leq -1,
\end{aligned}$$

for $x \geq \frac{\alpha+1}{1-(\beta+\gamma)}$.

Condition 2: $\lim_{n\to\infty} \sup_{a\in A} E_a[|\lambda_n| I_{[\tau_A>n]}] = 0$, where $A = \left[\frac{\alpha}{1-\gamma}, \frac{\alpha+1}{1-(\beta-\gamma)}\right]$.

First note that by Cauchy-Schwartz,

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{a\in A} E\left[|\lambda_n|\, I_{[\tau_A>n]} \mid \lambda_0 = a\right] \\
&\qquad\leq \lim_{n\to\infty} \sup_{a\in A} E^{1/2}\left[|\lambda_n|^2 \mid \lambda_0 = a\right] P^{1/2}(\tau_A > n \mid \lambda_0 = a).
\end{aligned}$$

Focusing on the first term, we have

$$\begin{aligned}
&E\big[\lambda_n^2 \mid \lambda_0 = a\big] \\
&\quad = E\Big[\big(\alpha + \beta Y_{n-1} + \gamma\lambda_{n-1}\big)^2 \mid \lambda_0 = a\Big] \\
&\quad = E\Big[E\big(\alpha^2 + 2\alpha\beta Y_{n-1} + 2\alpha\gamma\lambda_{n-1} + \beta^2 Y_{n-1}^2 + 2\beta\gamma Y_{n-1}\lambda_{n-1} \\
&\qquad + \gamma^2\lambda_{n-1}^2 \mid \lambda_{n-1}\big) \mid \lambda_0 = a\Big] \\
&\quad = E\Big[\alpha^2 + \lambda_{n-1}\big(2\alpha(\beta+\gamma) + \beta^2\big) + \lambda_{n-1}^2(\beta+\gamma)^2 \mid \lambda_{n-1} = a\Big] \\
&\quad = \alpha^2 + \big(2\alpha(\beta+\gamma) + \beta^2\big)E\big[\lambda_{n-1} \mid \lambda_0 = a\big] + (\beta+\gamma)^2 E\big[\lambda_{n-1}^2 \mid \lambda_0 = a\big].
\end{aligned}$$

Now,

$$\begin{aligned}
&E\big[\lambda_{n-1} \mid \lambda_0 = a\big] \\
&\quad = E\big[E(\alpha + \beta Y_{n-2} + \gamma\lambda_{n-2} \mid \lambda_{n-2}) \mid \lambda_0 = a\big] \\
&\quad = E\big[\alpha + (\beta+\gamma)\lambda_{n-2} \mid \lambda_0 = a\big] \\
&\quad = \alpha + (\beta+\gamma)E\big[\lambda_{n-2} \mid \lambda_0 = a\big] \\
&\quad\ \ \vdots \\
&\quad = \alpha + (\beta+\gamma)\big[\alpha + (\beta+\gamma)\big[\alpha + (\beta+\gamma)\cdots(\beta+\gamma)\big[\alpha + (\beta+\gamma)a\big]\cdots\big]\big] \\
&\quad = \sum_{i=0}^{n-2} \alpha(\beta+\gamma)^i + (\beta+\gamma)^{n-1}a \\
&\quad \le \frac{\alpha}{1-(\beta+\gamma)} + a.
\end{aligned}$$

Therefore,

$$\begin{aligned}
E\big[\lambda_n^2 \mid \lambda_0 = a\big] &\le c_1 + (\beta+\gamma)^2 E(\lambda_{n-1}^2 \mid \lambda_0 = a) \\
&= c_1 \sum_{i=0}^{n-1} (\beta+\gamma)^{2i} + (\beta+\gamma)^{2n}a^2 \\
&\le \frac{c_1}{1-(\beta+\gamma)^2} + a^2 := c_2,
\end{aligned}$$

where $c_1 = \alpha^2 + \big(2\alpha(\beta+\gamma) + \beta^2\big)\left(\frac{\alpha}{1-(\beta+\gamma)} + a\right)$. It then follows that

$$\lim_{n\to\infty} \sup_{a\in A} E\big[|\lambda_n| I_{[\tau_A > n]} \mid \lambda_0 = a\big] \le c_2^{1/2} \lim_{n\to\infty} \sup_{a\in A} P^{1/2}(\tau_A > n \mid \lambda_0 = a).$$

Applying Theorem A.2.16, we have

$$P(\tau_A > n \mid \lambda_0 = a) \le \frac{E_a(\tau_A)}{n+1} \le \frac{V(a) + (1+\alpha)}{n+1}, \; a \in A$$

from which we obtain the desired result:

$$\begin{aligned} \lim_{n\to\infty} \sup_{a\in A} E\left[|\lambda_n| I_{[\tau_A > n]} \mid \lambda_0 = a\right] &\le c_2^{1/2} \lim_{n\to\infty} \sup_{a\in A} \left(\frac{a + (1+\alpha)}{n+1}\right)^{1/2} \\ &\le c_2^{1/2} \lim_{n\to\infty} \left(\frac{\frac{\alpha+1}{1-(\beta+\gamma)} + (1+\alpha)}{n+1}\right)^{1/2} = 0. \end{aligned}$$

Condition 3: The family of probability measures $\{1/m \sum_{k=1}^{m} P^k(a,\cdot) : a \in A\}$ is tight for each $m \ge 1$.

Since $\{\lambda_n\}$ is weak Feller (Definition A.2.8) and $A = \left[\frac{\alpha}{1-\gamma}, \frac{\alpha+1}{1-(\beta+\gamma)}\right]$ is a compact set, Condition 3 follows from Theorem A.2.14. Therefore, the chain is bounded in probability on average.□

Lastly, it remains to show that the chain possesses a reachable state. The proof is straightforward if the process is recursed backward until it depends solely on the previous counts and initial condition.

Proposition 2.1.3 *The point* $\{\frac{\alpha}{1-\gamma}\}$ *is a reachable state of the process* $\{\lambda_n\}$.

Proof: First note the following form for λ_n :

$$\lambda_n = \alpha + \beta Y_{n-1} + \gamma\lambda_{n-1} = \alpha \sum_{i=1}^{n} \gamma^{i-1} + \beta \sum_{i=1}^{n} \gamma^{i-1} Y_{n-i} + \gamma^n \lambda_0.$$

Now, if $Y_i = 0$ for all $i \le n$, $\lambda_n = \alpha\frac{1-\gamma^n}{1-\gamma} + \gamma^n\lambda_0 \overset{n\to\infty}{\to} \frac{\alpha}{1-\gamma}$. Thus, if O is an open set containing $\frac{\alpha}{1-\gamma}$, we may choose M such that $x_n := \alpha\sum_{i=1}^{n}\gamma^{i-1} + \gamma^n\lambda_0 \in O$ for all $n \ge M$. Then,

$$\begin{aligned} P^M(y, O) &= P(\lambda_M \in O \mid \lambda_0 = y) \\ &\ge P(\lambda_1 = x_1, \lambda_2 = x_2, \ldots, \lambda_M = x_M \mid \lambda_0 = y) \end{aligned}$$

$$
\begin{aligned}
&= P(\lambda_M = x_M \mid \lambda_{M-1} = x_{M-1}) \cdots P(\lambda_1 = x_1 \mid \lambda_0 = y) \\
&= P(Y_{M-1} = 0 \mid \lambda_{M-1} = x_{M-1}) \cdots P(Y_0 = 0 \mid \lambda_0 = y) \\
&> 0.
\end{aligned}
$$

Therefore, $\{\frac{\alpha}{1-\gamma}\}$ is a reachable point.

Having shown that the process $\{\lambda_n\}$ is an e-chain, is bounded in probability on average and possesses a reachable point, we obtain the desired conclusion that a unique stationary distribution exists for the chain $\{\lambda_n\}$.

Theorem 2.1.4 *By Theorem A.2.18, the process $\{\lambda_n\}$ has a unique stationary distribution.*

2.2 GLARMA Models

Let $\{Y_t\}$ be the observation driven process defined by the following equation:

$$
p(y_t \mid w_t) = p(y_t \mid \mathbf{w}^{(t)}, \mathbf{y}^{(t-1)}) = \frac{e^{-\mu_t}\mu_t^{y_t}}{y_t!} I_{\{0,1,\ldots\}}(y_t).
$$

Recall the model for the process $\{W_t\}$ defined by equations (1.5), (1.6) and (1.7):

$$
W_t = \log(\mu_t) = \mathbf{x}_t'\boldsymbol{\beta} + \sum_{i=1}^{\infty} \tau_i e_{t-i}
$$

where

$$
e_t = \frac{Y_t - \mu_t}{\mu_t^{\lambda}}, \quad \lambda \geq 0 \text{ fixed}
$$

and

$$
\begin{aligned}
\sum_{i=1}^{\infty} \tau_i &= \left(1 - \sum_{i=1}^{p} \phi_i z^i\right)^{-1} \left(1 + \sum_{i=1}^{q} \theta_i z^i\right) - 1 \\
&= \phi(z)^{-1}\theta(z) - 1.
\end{aligned}
$$

In this section, we establish the existence of a stationary distribution for a subset of these models with $\mathbf{x}_t'\boldsymbol{\beta} = \beta$ and $p = 0$. Thus, the models we consider in this section have state equations of the following form:

$$
W_t = \log(\mu_t) = \beta + \sum_{i=1}^{q} \gamma_i e_{t-i}, \tag{2.5}
$$

where

$$e_t = \frac{y_t - \mu_t}{\mu_t^\lambda}, \ \lambda \geq 0 \text{ fixed}, \tag{2.6}$$

and W_0 is a given initial condition.

For $q = 1$, $\{W_t\}$ is a Markov process with

$$\begin{aligned} E(W_t) &= \beta + E\big[E(W_t \mid W_{t-1})\big] \\ &= \beta + E\big[\gamma_1 e^{-\lambda W_{t-1}} E(Y_{t-1} - e^{W_{t-1}} \mid W_{t-1})\big] \\ &= \beta. \end{aligned} \tag{2.7}$$

and

$$\begin{aligned} \text{Var}(W_t) &= \text{Var}\big[E(W_t \mid W_{t-1})\big] + E\big[\text{Var}(W_t \mid W_{t-1})\big] \\ &= \gamma_1^2 E[\text{Var}((Y_{t-1} - e^{W_{t-1}})e^{-\lambda W_{t-1}} \mid W_{t-1})] \\ &= \gamma_1^2 E\big[e^{-2\lambda W_{t-1}} e^{W_{t-1}}\big] \\ &= \gamma_1^2 E\big[e^{W_{t-1}(1-2\lambda)}\big]. \end{aligned} \tag{2.8}$$

Note that the variance of W_t is constant for $\lambda = 1/2$. Additionally,

$$\mathbf{W_t} = [W_t, \ldots, W_{t-q+1}]' \tag{2.9}$$

is a q^{th} order Markov chain with components

$$W_t = \beta + \sum_{i=1}^{q} \gamma_i (Y_{t-i} - e^{W_{t-i}}) e^{-\lambda W_{t-i}} \tag{2.10}$$

satisfying equations (2.7) and

$$\text{Var}(W_t) = \sum_{i=1}^{q} \gamma_i^2 E\big[e^{W_{t-i}(1-2\lambda)}\big].$$

2.2.1 Model 1

We begin by letting $\lambda = 1$ in (2.6) which corresponds to standardization of the mean corrected Poisson counts by their conditional variance. For these models, we are able to establish that the process $\{W_t\}$ defined by (2.5) is uniformly ergodic. We first show this to be true for the 1^{st} order Markov chain, $q = 1$, and then extend this result to the q^{th} order process.

2.2.1.1 1^{st} Order Markov

When $q = 1$, the state process W_t is bounded in one direction. This is easily seen when the process is written in the following form:

$$W_t = \beta - \gamma + \gamma Y_{t-1} e^{-W_{t-1}} \tag{2.11}$$
$$\begin{cases} \leq \beta - \gamma, & \text{for } \gamma < 0, \\ \geq \beta - \gamma, & \text{for } \gamma > 0. \end{cases}$$

Under these constraints, we are able to establish that the process is uniformly ergodic (Definition A.2.21). This is done by showing that the chain satisfies Doeblin's condition (Definition A.2.22) and is strongly aperiodic (Definition A.2.4). Uniform ergodicity then follows from Theorem A.2.23. This property is vital for developing asymptotic results for the parameter estimates. These estimates are derived in Chapter 4.

Theorem 2.2.1 *The process $\{W_t\}$ as defined by equation (2.11) satisfies Doeblin's Condition and is strongly aperiodic; hence, the process is uniformly ergodic.*

Proof: In order to establish Doeblin's Condition, we must show that there exists a probability measure ν satisfying the property that for some $m \geq 1$, $\epsilon < 1$ and $\delta > 0$,

$$\nu(A) > \epsilon \Rightarrow P^m(x, A) \geq \delta, \quad \text{for all } x \in X.$$

We consider the two cases $\gamma < 0$ and $\gamma > 0$.

Case 1: $\gamma < 0$

From (2.11), it is easily seen that W_t has an upper bound of $\beta - \gamma$ when $\gamma < 0$. Define the measure ν to have unit point mass at $\{\beta - \gamma\}$. It then suffices to only consider Borel sets B with $\beta - \gamma \in B$. Then, for all $x \leq \beta - \gamma$,

$$P(x, B) = P(W_t \in B \mid W_{t-1} = x)$$

$$
\begin{aligned}
&\geq P(W_t = \beta - \gamma \mid W_{t-1} = x) \\
&= P(Y_{t-1} = 0 \mid W_{t-1} = x) \\
&= e^{-e^x} \\
&\geq e^{-e^{\beta-\gamma}}.
\end{aligned}
$$

Hence, Doeblin's condition is satisfied for this case.

Case 2: $\gamma > 0$

For $\gamma > 0$, recall that W_t has a lower bound of $\beta - \gamma$. As in Case 1, we will take the measure ν to have unit mass at $\{\beta - \gamma\}$. Let $C = [\beta - \gamma, \max(\epsilon, \beta + \gamma)]$, where $\epsilon > 0$. Then, for all $x \in C$ and Borel sets B containing $\beta - \gamma$,

$$
\begin{aligned}
P(x, B) &= P(W_t \in B \mid W_{t-1} = x) \\
&\geq P(W_t = \beta - \gamma \mid W_{t-1} = x) \\
&= P(Y_{t-1} = 0 \mid W_{t-1} = x) \\
&= e^{-e^x} \\
&\geq e^{-e^{\max(\epsilon,\beta+\gamma)}} := \delta_1, \qquad (2.12)
\end{aligned}
$$

and

$$
\begin{aligned}
P^2(x, B) &\geq P(W_{t+1} = \beta - \gamma, W_t = \beta - \gamma \mid W_{t-1} = x) \\
&= P(W_{t+1} = \beta - \gamma \mid W_t = \beta - \gamma)(W_t = \beta - \gamma \mid W_{t-1} = x) \\
&\geq \delta_1^2.
\end{aligned}
$$

On the other hand if $x \notin C$, then $x > \max(\epsilon, \beta + \gamma)$ and we have

$$
\begin{aligned}
P(x, C) &= P(W_t \in C \mid W_{t-1} = x) \\
&\geq P(\beta - \gamma \leq W_t \leq \beta + \gamma \mid W_{t-1} = x) \\
&= P(|W_t - \beta| \leq \gamma \mid W_{t-1} = x) \\
&\geq 1 - \gamma^{-2}\mathrm{Var}(W_t \mid W_{t-1} = x) \text{ (by Chebyshev's Inequality)} \\
&= 1 - \gamma^{-2}\gamma^2 e^{-x} \\
&\geq 1 - e^{-\max(\epsilon,\beta+\gamma)} := \delta_2. \qquad (2.13)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
P^2(x,B) &= P(W_t \in B \mid W_{t-2} = x) \\
&\geq P(W_t \in B, W_{t-1} \in C \mid W_{t-2} = x) \\
&= \sum_{y \in C} P(W_t \in B, W_{t-1} = y \mid W_{t-2} = x) \\
&= \sum_{y \in C} P(W_t \in B \mid W_{t-1} = y) P(W_{t-1} = y \mid W_{t-2} = x) \\
&\geq \delta_1 \sum_{y \in C} P(W_{t-1} = y \mid W_{t-2} = x) \\
&= \delta_1 P(W_{t-1} \in C \mid W_{t-2} = x) \\
&\geq \delta_1 \delta_2.
\end{aligned}
$$

Thus, Doeblin's condition is also satisfied for the case $\gamma > 0$.

For either case the chain $\{W_t\}$ is strongly aperiodic since

$$
\begin{aligned}
P(\beta - \gamma, \beta - \gamma) &= P(W_t = \beta - \gamma \mid W_{t-1} = \beta - \gamma) \\
&= P(Y_{t-1} = 0 \mid W_{t-1} = \beta - \gamma) \\
&= \epsilon^{-e^{\beta - \gamma}} \\
&> 0.
\end{aligned}
$$

We conclude that $\{W_t\}$ must be uniformly ergodic.□

It follows from Theorem A.2.25 that $\{W_t\}$ is geometrically mixing.

Having shown that the chain is uniformly ergodic and geometrically mixing, we will be able to develop asymptotic results for the parameter estimates in this case. This is considered in detail in Chapter 4

2.2.1.2 q^{th} Order Markov

Using similar steps, it can be shown that the q^{th} order Markov chain defined by (2.9) and (2.10) with $\lambda = 1$, $\gamma_i \geq 0$, $|\beta| < \sum_{i=1}^{q} \gamma_i$ is uniformly ergodic. Thus, in this section we will consider the existence of a unique stationary distribution for the process $\mathbf{W}_t$ defined below.

$$\mathbf{W}_t = [W_t, \ldots, W_{t-q+1}]', \tag{2.14}$$

where

$$W_t := \beta + \sum_{i=1}^{q} \gamma_i (Y_{t-i} - e^{W_{t-i}}) e^{-W_{t-i}}. \tag{2.15}$$

Theorem 2.2.2 *The process $\{\mathbf{W}_t\}$ as defined by (2.14) and (2.15) satisfies Doeblin's Condition and is strongly aperiodic; hence, the process is uniformly ergodic.*

Proof: Recall from Section 2.2.1.1 that in order to establish Doeblin's Condition, we must show that there exists a probability measure ν satisfying the property that for some $m \geq 1$, $\epsilon < 1$ and $\delta > 0$,

$$\nu(A) > \epsilon \Rightarrow P^m(\mathbf{x}, A) \geq \delta, \quad \text{for all } \mathbf{x} \in X.$$

Define $\nu(\cdot)$ to have point mass at $(\beta - \eta)_q$ where $\eta = \sum_{i=1}^{q} \gamma_i$. Also define $\mathcal{C} = C^q = [\beta - \eta,\ \beta + \eta]^q$. Note that $W_t \geq \beta - \eta$; thus, if $W_t \notin C$, $W_t > \beta + \eta$. To show Doeblin's condition, we only need to consider sets $B \in \mathcal{B}(X)$ satisfying $(\beta - \eta)_q \in B$. First, we will consider the case $\mathbf{x} \in \mathcal{C}$. Here,

$$\begin{aligned}
&P^q(\mathbf{x}, B) \\
&= P(\mathbf{W}_t \in B \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&\geq P(\mathbf{W}_t = (\beta - \eta)_q \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&= P(W_t = \ldots = W_{t-q+1} = \beta - \eta \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&= P(W_t = \beta - \eta \mid W_{t-1} = \ldots = W_{t-q+1} = \beta - \eta, \mathbf{W}_{t-q} = \mathbf{x}) \\
&\quad \cdot P(W_{t-1} = \ldots = W_{t-q+1} = \beta - \eta \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&\vdots \\
&= \left(\prod_{i=0}^{q-2} P(W_{t-i} = \beta - \eta \mid W_{t-i-1} = \ldots = W_{t-q+1} = \beta - \eta, \mathbf{W}_{t-q} = \mathbf{x}) \right)
\end{aligned}$$

$$
\begin{aligned}
&\cdot P(W_{t-q+1} = \beta - \eta \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&= \left(\prod_{i=1}^{q-1} P(Y_{t-i} = \ldots = Y_{t-i-q+1} = 0 \mid \; W_{t-i} = \ldots = W_{t-q+1} = \beta - \eta, \mathbf{W}_{t-q} = \mathbf{x}) \right) \\
&\quad \cdot P(Y_{t-q} = \ldots = Y_{t-2q+1} = 0 \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&\geq \left(e^{-(q-1)e^{\beta-\eta}} e^{-e^{\beta+\eta}} \right) \left(e^{-(q-2)e^{\beta-\eta}} e^{-2e^{\beta+\eta}} \right) \cdots \left(e^{-qe^{\beta+\eta}} \right) \\
&= e^{-e^{\beta-\eta} \sum_{i=1}^{q-1} i} e^{-e^{\beta+\eta} \sum_{i=1}^{q} i} \\
&= e^{-\epsilon^{-\eta} q(q-1)/2} e^{-e^{\beta+\eta} q(q+1)/2} := \delta_1
\end{aligned}
$$

and

$$
\begin{aligned}
P^{2q}(\mathbf{x}, B) &\geq P(\mathbf{W}_t = (\beta - \eta)_q, \; \mathbf{W}_{t-q} = (\beta - \eta)_q \mid \mathbf{W}_{t-2q} = \mathbf{x}) \\
&= P(\mathbf{W}_t = (\beta - \eta)_q \mid \mathbf{W}_{t-q} = (\beta - \eta)_q) \\
&\quad \cdot P(\mathbf{W}_{t-q} = (\beta - \eta)_q \mid \mathbf{W}_{t-2q} = (\beta - \eta)_q) \\
&\geq \delta_1^2.
\end{aligned}
$$

Now, suppose $\mathbf{x} \notin C$. Then $x_i > \beta + \eta$ for at least one $i = 1, \ldots, q$. It follows that

$$
\begin{aligned}
P^q(\mathbf{x}, C) &= P(\mathbf{W}_t \in C \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&= P(W_t \in C, \ldots, W_{t-q+1} \in C \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&= \sum_{y_{q-1} \in C} P(W_t \in C, \ldots, W_{t-q+2} \in C, W_{t-q+1} = y_{q-1} \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&= \sum_{y_{q-1} \in C} P(W_t \in C, \ldots, W_{t-q+2} \in C \mid W_{t-q+1} = y_{q-1}, \mathbf{W}_{t-q} = \mathbf{x}) \\
&\quad \cdot P(W_{t-q+1} = y_{q-1} \mid \mathbf{W}_{t-q} = \mathbf{x}) \\
&\;\;\vdots \\
&= \sum_{y_{q-1} \in C} \cdots \sum_{y_1 \in C} \Big[P(W_t \in C \mid W_{t-1} = y_1, \ldots, W_{t-q+1} = y_{q-1}, \mathbf{W}_{t-q} = \mathbf{x}) \\
&\quad \cdot \prod_{i=1}^{q-2} \left(P(W_{t-i} = y_i \mid W_{t-i-1} = y_{i+1}, \ldots, W_{t-q+1} = y_{q-1}, \mathbf{W}_{t-q} = \mathbf{x}) \right) \\
&\quad \cdot P(W_{t-q+1} = y_{q-1} \mid \mathbf{W}_{t-q} = \mathbf{x}) \Big].
\end{aligned}
$$

Consider $P(W_n \in C | \mathbf{W}_{n-1} = \mathbf{w}_{n-1})$. Define $\mathbf{W}'_{n-1}$ and $\mathbf{w}'_{n-1}$ to be a reordering of $\mathbf{W}_{n-1}$ and $\mathbf{w}_{n-1}$, respectively, such that the first r elements of $\mathbf{w}'_{n-1}$ have values in C and the last $q-r$ elements of $\mathbf{w}'_{n-1}$ have values in C^c, $0 \le r \le q$. We may then write $W_n = \beta + \sum_{i=1}^{q} \gamma'_i (Y'_{n-i} - e^{W'_{i-1}}) e^{-W'_{i-1}}$, where the elements of $\gamma = (\gamma_i, \dots, \gamma_q)^t$ and $\mathbf{Y}_{n-1} = (Y_{n-1}, \dots, Y_{n-q})^t$ have been re-arranged according to the ordering of $\mathbf{W}'_{n-1}$ described above to obtain γ' and $\mathbf{Y}'_{n-1}$. Thus,

$$
\begin{aligned}
&P(W_n \in C \mid \mathbf{W}_{n-1} = \mathbf{w}_{n-1}) \\
&\qquad = P\Big(\beta + \sum_{i=1}^{q} \gamma'_i (Y'_{n-i} - e^{W'_{n-i}}) e^{-W'_{n-i}} \in C \mid \mathbf{W}'_{n-1} = \mathbf{w}'_{n-1}\Big).
\end{aligned}
$$

Now, for $r = q$, each element of $\mathbf{w}'_{n-1} \in C$. Hence,

$$
\begin{aligned}
&P(W_n \in C \mid \mathbf{W}_{n-1} = \mathbf{w}_{n-1}) \\
&\quad \ge P\Big(\beta + \sum_{i=1}^{q} \gamma'_i (Y'_{n-i} - e^{W'_{n-i}}) e^{-W'_{n-i}} = \beta - \eta \mid \mathbf{W}'_{n-1} = \mathbf{w}'_{n-1}\Big) \\
&\quad = P(Y'_{n-1} = \dots = Y'_{n-q} = 0 \mid \mathbf{W}'_{n-1} = \mathbf{w}'_{n-1}) \\
&\quad \ge e^{-qe^{\beta+\eta}}.
\end{aligned}
$$

For $r < q$,

$$
\begin{aligned}
&P(W_n \in C \mid \mathbf{W}_{n-1} = \mathbf{w}_{n-1}) \\
&\ge P\Big(Y'_{n-1} = \dots = Y'_{n-r} = 0, \\
&\qquad \beta + \sum_{i=r+1}^{q} \gamma'_i (Y'_{n-i} - e^{W'_{n-i}}) e^{-W'_{n-i}} - \sum_{i=1}^{r} \gamma'_i \in [\beta - \gamma, \beta + \gamma] \,\Big|\, \mathbf{W}'_{n-1} = \mathbf{w}'_{n-1}\Big) \\
&\ge e^{-re^{\beta+\eta}} P\Big(\sum_{i=r+1}^{q} \gamma'_i (Y'_{n-i} - e^{W'_{n-i}}) e^{-W'_{n-i}} \\
&\qquad \in \Big[-\sum_{i=r+1}^{q} \gamma'_i, \ 2\sum_{i=1}^{r} \gamma'_i + \sum_{i=r+1}^{q} \gamma'_i \Big] \,\Big|\, \mathbf{W}'_{n-1} = \mathbf{w}'_{n-1}\Big) \\
&\ge e^{-re^{\beta+\eta}} P\Big(\sum_{i=r+1}^{q} \gamma'_i (Y'_{n-i} - e^{W'_{n-i}}) e^{-W'_{n-i}} \in \Big[-\sum_{i=r+1}^{q} \gamma'_i, \ \sum_{i=r+1}^{q} \gamma'_i \Big] \,\Big|\, \mathbf{W}'_{n-1} = \mathbf{w}'_{n-1}\Big)
\end{aligned}
$$

$$\geq e^{-re^{\beta+\eta}}\left(1-\frac{\sum_{i=r+1}^{q}\gamma_i'^2 e^{-w_{n-i}'}}{\left(\sum_{i=r+1}^{q}\gamma_i'\right)^2}\right), \text{ (by Chebyshev's inequality)}$$

$$\geq e^{-re^{\beta+\eta}}\left(1-\frac{e^{-(\beta+\eta)}\sum_{i=r+1}^{q}\gamma_i'^2}{\left(\sum_{i=r+1}^{q}\gamma_i'\right)^2}\right),$$

since $w_{n-i}' > \beta+\eta$ for $r+1 \leq i \leq q$. Thus, if we define r_i to be the number of elements of $\mathbf{W}_{t-i}$ with values in C, $i = 1, \ldots, q$, and

$$\delta_{r_i} := \begin{cases} e^{-r_i e^{\beta+\eta}}\left(1-\frac{e^{-(\beta+\eta)}\sum_{j=r_i+1}^{q}(\gamma_j')^2}{\left(\sum_{j=r_i+1}^{q}\gamma_j'\right)^2}\right), & r_i < q, \\ e^{-qe^{\beta+\eta}}, & r_i = q, \end{cases}$$

(note that $\delta_{r_i} > 0$ for all $i = 1, \ldots, q$), we obtain

$$\begin{aligned} P^q(x, \mathcal{C}) \geq & \sum_{y_{q-1}\in C} \cdots \sum_{y_1 \in C}\Bigg[\delta_{r_1} \\ & \cdot \prod_{i=1}^{q-2}\Big(P(W_{t-i}=y_i \mid W_{t-i-1}=y_{i+1},\ldots, W_{t-q+1}=y_{q-1}, \mathbf{W}_{t-q}=\mathbf{x})\Big) \\ & \cdot P(W_{t-q+1}=y_{q-1} \mid \mathbf{W}_{t-q}=\mathbf{x})\Bigg] \\ \geq & \ \delta_{r_1}\sum_{y_{q-1}\in C}\cdots\sum_{y_2\in C}\Bigg[P(W_{t-1}\in C \mid W_{t-2}=y_2, \ldots W_{t-q+1}=y_{q-1}, \mathbf{W}_{t-q}=\mathbf{x}) \\ & \cdot \prod_{i=2}^{q-2}\Big(P(W_{t-i}=y_i \mid W_{t-i-1}=y_{i+1},\ldots, W_{t-q+1}=y_{q-1}, \mathbf{W}_{t-q}=\mathbf{x})\Big) \\ & \cdot P(W_{t-q+1}=y_{q-1} \mid \mathbf{W}_{t-q}=\mathbf{x})\Bigg] \\ & \vdots \\ \geq & \prod_{i=1}^{q}\delta_{r_i}. \end{aligned}$$

Thus, for $\mathbf{x} \notin \mathcal{C}$,

$$\begin{aligned} P^{2q}(\mathbf{x}, \mathcal{C}) &= P(\mathbf{W}_t \in B \mid \mathbf{W}_{t-2q}=\mathbf{x}) \\ &\geq P(\mathbf{W}_t\in B,\ \mathbf{W}_{t-q}\in\mathcal{C} \mid \mathbf{W}_{t-2q}=\mathbf{x}) \\ &= \sum_{\mathbf{y}\in\mathcal{C}} P(\mathbf{W}_t\in B \mid \mathbf{W}_{t-q}=\mathbf{y})\ P(\mathbf{W}_{t-q}=\mathbf{y} \mid \mathbf{W}_{t-2q}=\mathbf{x}) \\ &\geq \delta_1 \sum_{\mathbf{y}\in\mathcal{C}} P(\mathbf{W}_{t-q}\in\mathcal{C} \mid \mathbf{W}_{t-2q}=\mathbf{x}) \end{aligned}$$

$$\geq \delta_1 \prod_{i=1}^{q} \delta_{r_i} > 0.$$

Therefore, Doeblin's condition is satisfied. The chain is strongly aperiodic since

$$\begin{aligned} &P((\beta-\eta)_q, (\beta-\eta)_q) \\ &\quad = P(\mathbf{W}_t = (\beta-\eta)_q \mid \mathbf{W}_{t-1} = (\beta-\eta)_q) \\ &\quad = P(W_t = \beta-\eta, \ldots, W_{t-q+1} = \beta-\eta \mid W_{t-1} = \beta-\eta, \ldots, W_{t-q} = \beta-\eta) \\ &\quad = P(W_t = \beta-\eta \mid W_{t-1} = \beta-\eta, \ldots, W_{t-q} = \beta-\eta) \\ &\quad = P(Y_{t-1} = \ldots = Y_{t-q} = 0 \mid W_{t-1} = \beta-\eta, \ldots, W_{t-q} = \beta-\eta) \\ &\quad = e^{-qe^{\beta-\eta}} \\ &\quad > 0. \end{aligned}$$

Thus, it follows from Theorem A.2.23 that $\{\mathbf{W}_t\}$ is uniformly ergodic.□
By Theorem A.2.25, we also obtain that $\{\mathbf{W}_t\}$ is geometrically mixing, i.e., strong mixing at a geometric rate.

2.2.1.3 Exact Sampling

Having shown that the chain $\{W_t\}$ defined by (2.5) and (2.6) with $\lambda = 1$ is uniformly ergodic, it is possible to sample directly from its stationary distribution using an algorithm developed by Murdoch and Green (1998). In this section, we give an overview of their algorithm and then apply it to the first order Markov chain, $q = 1$, with no regressors present, $\beta = 0$, and with $\gamma > 0$, discussed in Section 2.2.1.1.

The algorithm we employ is based on a construction created by Murdoch and Green called the "mutigamma" coupler. The idea behind the algorithm is to start the chain in the infinite past and consider all possible paths. From time to time, two of the paths will couple. If all possible paths have joined by time 0, the draw at time 0 will be from the stationary distribution (see Propp and Wilson (1996)).

In practice, one must find a backward coupling time, t_c, such that if the chain is started from any possible state, $x \in X$, at time $-t_c$, all possible sample paths will have joined by time 0, giving a draw from the stationary distribution.

Recall that when a Markov chain is uniformly ergodic, there exists a probability measure ν such that for some $k \geq 1$, $\rho > 0$ the chain satisfies

$$P^k(x, \cdot) \geq \rho\nu(\cdot)$$

for every $x \in X$ (A.2.23). This condition meets the assumptions necessary to employ the multigamma coupler algorithm. For the case $k = 1$, we simulate from the chain by first drawing a sequence $\{U_n\}$ of iid $U(0,1)$ random variables and a sequence $\{V_n\}$ of random variables with distribution ν. Then, if $U_{n+1} < \rho$, we define $W_{n+1} := V_{n+1}$. If $U_{n+1} \geq \rho$, we define W_{n+1} to be the value obtained by a draw from $R(X_n, \cdot) := [P(X_n, \cdot) - \rho\nu(\cdot)]/(1-\rho)$, the "residual" distribution. Since

$$\begin{aligned} P(W_{n+1} \leq x \mid W_n = w) &= \rho\nu(x) + (1-\rho)R(w, x) \\ &= \rho\nu(x) + P(w, x) - \rho\nu(x) \\ &= P(w, x), \end{aligned}$$

it is clear that W_{n+1} has the appropriate distribution. When applying this algorithm, it is imperative that the set of uniform random variables which were used to find the backward coupling time are the same ones that are used in determining from which distribution to draw the next value.

For the process $\{W_t\}$ defined by (2.5) and (2.6), $k = 2$. Thus, we must modify the sampling algorithm described above. From (2.12) and (2.13), it is clear that for any x,

$$P(x, C) \geq \min(1 - e^{-\gamma}, e^{-e^{\gamma}}).$$

It is also easily seen that once the chain is in the set C, the chain takes the value $-\gamma$ at the next step with probability $e^{-e^{\gamma}}$. This allows us to find a backward coupling time, t_c, for which we know the value of the chain at time $-(t_c - 2)$. Thus, for this case, it is not necessary to first compare the uniform random variable to the value ρ before deciding from which distribution to draw. Here, the next value of the chain, W_i is obtained from the previous value as follows:

1. $Y_{i-1} = \text{Pois}^{\leftarrow}(U_{i-1}, e^{W_{i-1}})$,
   the smallest integer n, $n \geq 0$, such that $P(Y \leq n) \geq U_{i-1}$, $Y \sim \text{Pois}(\text{mean} = e^{W_{i-1}})$.

2. $W_i = \gamma(Y_{i-1} - e^{W_{i-1}})e^{-W_{i-1}}$.

Figure 2.1 displays a histogram of the stationary distribution of $\{W_t\}$ for a sample of size 10,000 for $\gamma = 0.25$, and $\gamma = 0.75$. Also shown are corresponding histograms of the values obtained by running a simulation of the process with a chain length of 10,000. For the simulations, varying values were given as the starting point: however, there were no detectable changes. As can be seen by comparing the simulations to the "exact" sample, the process converges to its stationary distribution quite rapidly, there is negligible difference between the exact sample and the simulation.

The time it took before a backward coupling time, t_c, could be guaranteed ($W_{-(t_c-1)} \in C$, $W_{-(t_c-2)} = -\gamma$) grew larger as γ increased. Table 2.2.1.3 gives the mean and standard deviation (rounded to two decimals) of the backward coupling times for $\gamma = 0.25$, 0.5 and 0.75.

Table 2.1: Backward Coupling Time Statistics

γ	mean	std. deviation
0.25	20.24	18.71
0.5	32.06	30.50
0.75	76.95	75.39

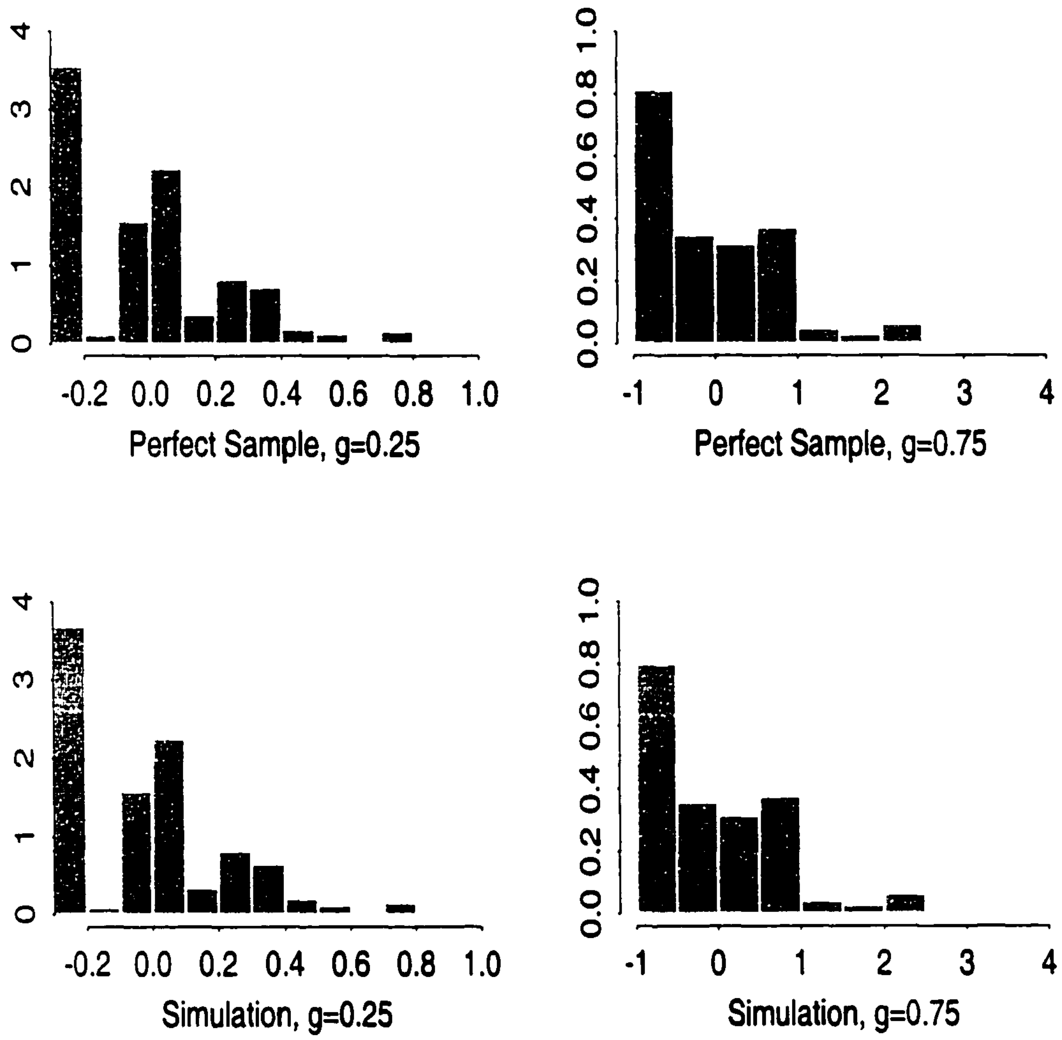

Figure 2.1: Exact Sampling. Top left: exact sampling with $\gamma = 0.25$; top right: exact sampling with $\gamma = 0.75$; bottom left: simulation with $\gamma = 0.25$: bottom right: simulation with $\gamma = 0.75$.

In addition to the point $\{-\gamma\}$, the process $\{W_t\}$ possesses many other reachable states (Definition A.2.9). It is interesting to consider the sequence of Poisson counts through which the chain arrives at these states. although it is not surprising. A table listing the states to which the chain returned on more than 50 occasions (rounded to 4 decimals), the number of visits made to the state and the sequence of Poisson counts leading the chain to the state are given in Table 2.2.1.3 for the

Table 2.2: Reachable States

$\gamma = 0.25$			$\gamma = 0.75$		
State	# Visits	Preceding Poisson Counts	State	# Visits	Preceding Poisson Counts
-0.25	3560	$\{\ldots,0\}$	-0.75	4033	$\{\ldots,0\}$
0.0710	1342	$\{\ldots,0,1\}$	0.8378	1176	$\{\ldots,0,1\}$
0.3920	572	$\{\ldots,0,2\}$	-0.1010	295	$\{\ldots,0,1,2\}$
-0.0171	490	$\{\ldots,0,1,1\}$	2.4255	275	$\{\ldots,0,2\}$
0.2157	237	$\{\ldots,0,1,2\}$	-0.4255	240	$\{\ldots,0,1,1\}$
0.0043	189	$\{\ldots,0,1,1,1\}$	0.2235	236	$\{\ldots,0,1,3\}$
-0.0811	154	$\{\ldots,0,2,1\}$	0.5480	128	$\{\ldots,0,1,4\}$
0.0878	144	$\{\ldots,0,2,2\}$	0.0797	104	$\{\ldots,0,1,2,1\}$
0.7130	132	$\{\ldots,0,3\}$	0.3978	95	$\{\ldots,0,1,1,1\}$
-0.0485	97	$\{\ldots,0,1,2,1\}$	-0.1502	74	$\{\ldots,0,1,3,1\}$
0.2586	80	$\{\ldots,0,1,1,2\}$	0.8726	58	$\{\ldots,0,1,5\}$
0.4486	77	$\{\ldots,0,1,3\}$	0.4495	56	$\{\ldots,0,1,3,2\}$
0.2568	69	$\{\ldots,0,2,3\}$	0.9094	53	$\{\ldots,0,1,2,2\}$
-0.0011	67	$\{\ldots,0,1,1,1,1\}$			
0.0211	56	$\{\ldots,0,2,1,1\}$			
0.1530	52	$\{\ldots,0,1,2,2\}$			

two cases $\gamma = 0.25$ and 0.75. For example, when $\gamma = 0.25$, the value $W_0 = 0.2157$ was "drawn" 237 out of 10,000 times. The corresponding Poisson counts which led to this state were as follows: $Y_{-3} = 0$, $Y_{-2} = 1$ and $Y_{-1} = 2$. The table illustrates the importance of our ability to bound the chain $\{W_t\}$ in at least one direction. When $Y_{t-1} = 0$ we know that the next value of the chain will be $W_t = -\gamma$. It is this result which enabled us to establish that the process is uniformly ergodic. When we are not able to bound the chain, establishing stationarity properties for the process becomes much more difficult as will be seen in Section 2.2.2.

2.2.2 Model 2

For the case $\lambda = 1/2$, $\beta = 0$, $q = 1$ and $\gamma > 0$ the process defined by (2.5) and (2.6) becomes

$$W_t = \gamma(Y_{t-1} - e^{W_{t-1}})e^{-1/2W_{t-1}}, \qquad (2.16)$$

and the situation is much more complicated. Here, we can only establish the existence of at least one stationary solution. We give two proofs of this result using vastly different methods.

2.2.2.1 Direct Approach

Here the process will be approached directly and results from Appendix A will be used to guarantee the existence of at least one stationary solution. First note that the chain satisfies the weak Feller property (Definition A.2.8). Using Theorem A.2.15 it will be shown that the process defined by (2.16) is bounded in probability on average. and hence, by Theorem A.2.13, the chain has at least one stationary solution.

Theorem 2.2.3 *The chain $\{W_t\}$ is bounded in probability on average; therefore, there exists at least one invariant measure.*

Proof: In order to verify that the chain is bounded in probability on average. it suffices to verify the three conditions of Theorem A.2.15.

Condition 1: There exists a non-negative function V such that $\Delta V(x) := EV(x) - V(x) \leq -1 + bI_A(x)$ where $b = |1+2\gamma|$ and $A = \{a : a \in [-b, b]\}$.

Defining $V(x) = |x|$, we have

$$\begin{aligned}
\Delta V(x) &= E\big[|W_t| \mid W_{t-1} = x\big] - |x| \\
&= E\big[W_t I_{[W_t \geq 0]} \mid W_{t-1} = x\big] + E\big[-W_t I_{[W_t<0]} \mid W_{t-1} = x\big] - |x| \\
&= E\big[\gamma(Y_t - e^x)e^{-x/2} I_{[Y_t \geq e^x]} \mid W_{t-1} = x\big] \\
&\quad + E\big[\gamma(e^x - Y_t)e^{-x/2} I_{[Y_t < e^x]} \mid W_{t-1} = x\big] - |x|. \qquad (2.17)
\end{aligned}$$

We will now consider two cases: i) $x \geq 0$ and ii) $x < 0$.

case i): $x \geq 0$

$$(2.17) = \gamma e^{-x/2} E\big[(Y_t - e^x)I_{[Y_t \geq e^x]} + (e^x - Y_t)(1 - I_{[Y_t \geq e^x]}) \mid W_{t-1} = x\big] - |x|$$

$$\begin{aligned}
&= \gamma e^{-x/2} E\big[2(Y_t - e^x) I_{[Y_t \geq e^x]} \mid W_{t-1} = x\big] + e^x - E\big[Y_t \mid W_{t-1} = x\big] - |x| \\
&= 2\gamma e^{-x/2} E\big[(Y_t - e^x) I_{[Y_t \geq e^x]} \mid W_{t-1} = x\big] - |x| \\
&\leq 2\gamma e^{-x/2} E^{1/3}\Big[(Y_t - e^x)^3 \mid W_{t-1} = x\Big] \; E^{2/3}\Big[I^{3/2}_{[Y_t \geq e^x]} \mid W_{t-1} = x\Big] - |x| \\
&\quad \text{(by Hölder's inequality)} \\
&\leq 2\gamma e^{-x/2} E^{1/3}\Big[(Y_t - e^x)^3 \mid W_{t-1} = x\Big] - |x| \\
&= 2\gamma e^{-x/2} e^{x/3} - |x| \\
&\leq 2\gamma - |x| \text{ for } x \geq 0.
\end{aligned}$$

case ii): $x < 0$

$$\begin{aligned}
(2.17) &= E\big[\gamma(Y_t - e^x) e^{-x/2} I_{[Y_t > 0]} \mid W_{t-1} = x\big] \\
&\quad + E\big[\gamma(e^x - Y_t) e^{-x/2} I_{[Y_t = 0]} \mid W_{t-1} = x\big] - |x| \\
&= \gamma e^{-x/2} E\big[(Y_t - e^x)(1 - I_{[Y_t = 0]}) - (Y_t - e^x) I_{[Y_t = 0]} \mid W_{t-1} = x\big] - |x| \\
&= -2\gamma e^{-x/2} E\big[(Y_t - e^x) I_{[Y_t = 0]} \mid W_{t-1} = x\big] - |x| \\
&= 2\gamma e^{x/2} e^{-e^x} - |x| \\
&< 2\gamma - |x| \text{ for } x < 0.
\end{aligned}$$

Therefore, $\Delta V(x) \leq -1 + b I_A(x)$.

Condition 2: $\lim_{n\to\infty} \sup_{a \in A} E_a\big[|W_n| I_{[\tau_A > n]}\big] = 0$, where $A = \{a : a \in [-b, b]\}$.

By the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{a \in A} E\big[|W_n| I_{[\tau_A > n]} \mid W_0 = a\big] \\
&\qquad \leq \lim_{n\to\infty} \sup_{a \in A} E^{1/2}[|W_n|^2 \mid W_0 = a] P^{1/2}(\tau_A > n \mid W_0 = a).
\end{aligned}$$

Now,

$$\begin{aligned}
E\big[|W_n|^2 \mid W_0 = a\big] &= E\Big[E\big[|W_n|^2 \mid W_{n-1}\big] \mid W_0 = a\Big] \\
&= E\Big[E\big[\gamma^2 (Y_n - e^{W_{n-1}})^2 e^{-W_{n-1}} \mid W_{n-1}\big] \mid W_0 = a\Big] \\
&= \gamma^2.
\end{aligned}$$

Whence,

$$\limsup_{n\to\infty}\sup_{a\in A} E^{1/2}\Big[|W_n|^2 \mid W_0 = a\Big] P^{1/2}\Big(\tau_A > n|W_0 = a\Big) = \limsup_{n\to\infty}\sup_{a\in A} \gamma P^{1/2}\Big(\tau_A > n\Big).$$

Using Markov's inequality and Theorem A.2.16, it follows that

$$P_a(\tau_A > n) \leq \frac{E_a(\tau_A)}{n+1} \leq \frac{V(a)+b}{n+1} \text{ for } a \in A.$$

and hence,

$$\begin{aligned}
\limsup_{n\to\infty}\sup_{a\in A} E_a\left[|W_n| I_{[\tau_A>n]}\right] &\leq \gamma \limsup_{n\to\infty}\sup_{a\in A}\left(\frac{V(a)+b}{n+1}\right)^{1/2} \\
&\leq \gamma \lim_{n\to\infty}\left(\frac{2b}{n+1}\right)^{1/2} \\
&= 0.
\end{aligned}$$

Condition 3: The family of probability measures $\{1/m \sum_{k=1}^{m} P^k(a,\cdot) : a \in A\}$ is tight for each $m \geq 1$.

By Theorem A.2.14, Condition 3 holds. Therefore, we conclude that there exists at least one invariant measure[1].□

2.2.2.2 Addition of Noise Approach

Here, we will establish the existence of a stationary distribution by adding independent Gaussian noise with variance σ^2 to the original series. Then, as σ^2 decreases to 0, properties of the original series can be recovered.

We will use the following set-up for this analysis: Let $\{W_{t,k}\}$ follow the Markov recursions given by

$$W_{t,k} = \gamma(Y_{t-1} - e^{W_{t-1,k}})e^{-1/2W_{t-1,k}} + Z_t, \tag{2.18}$$

with

[1]Through a personal correspondence with William Dunsmuir and Ben Goldys, it was shown that the process satisfies a condition stronger than bounded in probability on average. Inspired by their observation, a simpler proof of the existence of a stationary distribution is included as an addendum to this chapter.

$$
\begin{aligned}
Y_t | W_{t-1,k} &\sim \text{Poisson}(e^{W_{t-1,k}}), \\
Z_{t,k} &\sim \text{N}(0, \sigma_k^2),
\end{aligned}
$$

and Y_t, Z_t are independent. Since the value of σ_k^2 is irrelevant in showing that the chain is uniformly ergodic, we set $\sigma_k^2 = 1$ in order to establish this property. For this case, we will denote $\{W_{t,k}\}$ as $\{W_t\}$ and $\{Z_{t,k}\}$ as $\{Z_t\}$.

To establish the existence of at least one stationary measure, we will show that for any choice of σ_k^2, the process defined by (2.18) is uniformly ergodic. This will be accomplished using Theorem A.2.23. It is necessary to show that the chain is aperiodic and that Doeblin's condition is met. Once this has been established, we will show that the sequence of stationary measures $\{\mu_{\sigma_k}\}$ corresponding to a sequence of variances $\{\sigma_k\}$ decreasing to 0 is tight (Definition A.2.11). By tightness, it then follows that the limit of these measures is a stationary solution to the process defined in (2.16). However, it does not give us the existence of a unique stationary solution.

Proposition 2.2.4 *For any $\delta > 0, C := [-\delta, \delta]$ is a petite set (Definition A.2.6).*

Proof: Note that $x \in C$ if and only if $-\gamma e^{x/2} \in [-\gamma e^{\delta/2}, -\gamma e^{-\delta/2}]$. Define $f_\delta(\cdot)$ to be the probability density function of a normally distributed random variable with mean δ and variance 1. Set $f(x) = \min\left(f_{\delta_*}(x), f_{\delta^*}(x)\right)$ with $\delta_* = -\gamma e^{\delta/2}$ and $\delta^* = -\gamma e^{-\delta/2}$. Now define $\nu(dx) = e^{-e^\delta} f(x)dx$. Then, for all $x \in C$

$$
\begin{aligned}
P(x, B) &= P(W_t \in B \mid W_{t-1} = x) \\
&= P\big(\gamma(Y_{t-1} - e^{W_{t-1}})e^{-W_{t-1}/2} + Z_t \in B \mid W_{t-1} = x\big) \\
&\geq P(-\gamma e^{x/2} + Z_t \in B)e^{-e^x} \\
&\geq e^{-e^x} \int_B f(z)dz \\
&\geq e^{-e^\delta} \int_B f(z)dz \\
&= \nu(B).
\end{aligned}
$$

Therefore, with

$$K_a(x,B) \;:=\; \sum_{n=1}^{\infty} P^n(x,B)a_n, \quad \text{where } a_n = \begin{cases} 1, & \text{for } n = 1, \\ 0, & \text{for } n \geq 2, \end{cases}$$

we have

$$K_a(x,B) \;\geq\; \nu(B).$$

Hence, C is a petite set.□

Proposition 2.2.5 *The process defined in (2.18) satisfies Doeblin's condition.*

Proof: Choose a such that $a^2 > \gamma^2 + \sigma^2$. Let $A = [-a, a]$. Define $f_a(\cdot)$ and $\nu(dx)$ as in the proof of Proposition 2.2.4: define $f_a(\cdot)$ to be the probability density function of a normally distributed random variable with mean $= a$ and variance $= 1$, set $f(x) = \min(f_{a_*}(x), f_{a^*}(x))$ with $a_* = -\gamma e^{a/2}$ and $a^* = -\gamma e^{-a/2}$, and define $\nu(dx) = e^{-e^a} f(x)dx$. By Proposition 2.2.4, we have that $P(x,B) \geq \nu(B)$ for all $x \in A$. Now consider the case when $x \notin A$.

$$\begin{aligned}
P(x,A) &= P(W_t \in A | W_{t-1} = x) \\
&= P(|W_t| < a | W_{t-1} = x) \\
&\geq 1 - 1/a^2 E(W_t^2 | W_{t-1} = x) \quad \text{(by Chebyshev's inequality)} \\
&= 1 - 1/a^2 \left[E(\gamma^2 e^{-x}(Y_{t-1} - e^x)^2) + E(Z_t^2) \right] \\
&= 1 - 1/a^2(\gamma^2 + \sigma^2) \\
&= \delta_1 > 0.
\end{aligned}$$

Hence,

$$\begin{aligned}
P^2(x,B) &= P(W_t \in B | W_{t-2} = x) \\
&\geq P(W_t \in B, W_{t-1} \in A | W_{t-2} = x) \\
&= P(W_t \in B, W_{t-1}(x) \in A | W_{t-2} = x),
\end{aligned}$$

where $W_{t-1}(x)$ represents the random variable W_{t-1} given $W_{t-2} = x$.

$$\begin{aligned} &= \int_A P(W_t \in B | W_{t-1}(x) = y)\ P(W_{t-1}(x) \in dy) \\ &\geq \nu(B) \int_A P(W_{t-1}(x) \in dy) \\ &= \nu(B) P(W_{t-1} \in A | W_{t-2} = x) \\ &\geq \nu(B)\delta_1 \text{ for every } x \in X. \end{aligned}$$

Thus, the chain $\{W_t\}$ satisfies Doeblin's condition.□

The chain is strongly aperiodic since the set A defined in the previous proof is a small set with positive ν measure (Definition A.2.7).

Theorem 2.2.6 *The process defined by equation (2.18) is uniformly ergodic.*

Proof: Since the chain is strongly aperiodic and satisfies Doeblin's condition, by Theorem A.2.23, the chain is uniformly ergodic.□

To establish that the process given by (2.16) has at least one invariant measure, it remains to show that the sequence of measures $\{\mu - \sigma_k\}$ is tight.

Proposition 2.2.7 *The sequence of stationary probability measures $\{\mu_{\sigma_k} : k \in \mathbb{Z}_+\}$ for the set of processes $\{W_{t,k}\}$ defined by (2.18) with corresponding noise variance $\{\sigma_k\}$ satisfying $\sigma_1 = 1, \sigma_k \to 0$ as $k \to \infty$, is tight; i.e. for each $\epsilon > 0$, there exists a compact subset $C \subset X$ such that*

$$\mu_{\sigma_k}(C) \geq 1 - \epsilon \text{ for all } k.$$

Proof: Let $C = [-c, c]$. Given $\epsilon > 0$, set $\epsilon_1 = \epsilon/2$ and $\epsilon_2 = \epsilon^2/4$. Choose N, M and c large such that $\sum_{i=0}^{N} \frac{e^{-1}}{i!} \geq 1 - \epsilon_1$, $\sum_{i=1}^{N} \frac{e^{-e^{-M}} e^{-Mi}}{i!} < \epsilon_2$, and

$$P\left[-c + \gamma(\epsilon_1)^{-1/2} \leq Z \leq \min\left(c - \gamma(\epsilon_1)^{-1/2},\ c - \gamma(N - e^{-M})e^{M/2}\right)\right] \geq 1 - \epsilon_1,$$

where Z is a normally distributed random variable with mean$=0$ and variance$=1$. Then,

$$\mu_{\sigma_k}(C) = \int P(W_t \in C \mid W_{t-1} = x)\mu_{\sigma_k}(dx)$$

$$
\begin{aligned}
&= \int P\left(\gamma(Y_{t-1} - e^x)e^{-x/2} + Z_t \in C\right) \mu_{\sigma_k}(dx) \\
&= \int_{x>0} \sum_{i=0}^{\infty} P\left(\gamma(i - e^x)e^{-x/2} + Z_t \in C\right) P(Y_{t-1} = i)\mu_{\sigma_k}(dx) \\
&+ \int_{x\leq 0} \sum_{i=0}^{\infty} P\left(\gamma(i - e^x)e^{-x/2} + Z_t \in C\right) P(Y_{t-1} = i)\mu_{\sigma_k}(dx) \\
&= \mathrm{I} + \mathrm{II}.
\end{aligned}
$$

Now, defining $S_x := \{i : e^x - e^{x/2}\epsilon_1^{-1/2} \leq i \leq e^x + e^{x/2}\epsilon_1^{-1/2}\}$,

$$
\begin{aligned}
\mathrm{I} &\geq \int_{x>0} \sum_{i\in S_x} P\left(\gamma(i - e^x)e^{-x/2} + Z_t \in C\right) P(Y_{t-1} = i)\mu_{\sigma_k}(dx) \\
&\geq \int_{x>0} \sum_{i\in S_x} P\left(\frac{-c + \gamma\epsilon_1^{-1/2}}{\sigma_k} \leq Z \leq \frac{c - \gamma\epsilon_1^{-1/2}}{\sigma_k}\right) P(Y_{t-1} = i)\mu_{\sigma_k}(dx) \\
&\geq \int_{x>0} (1 - \epsilon_1)P(-e^{x/2}\epsilon_1^{-1/2} \leq Y_{t-1} - e^x \leq e^{x/2}\epsilon_1^{-1/2})\mu_{\sigma_k}(dx) \\
&\geq \int_{x>0} (1 - \epsilon_1)(1 - (\epsilon_1/e^x)e^x)\mu_{\sigma_k}(dx) \;\; \text{(by Chebyshev's inequality)} \\
&= (1 - \epsilon_1)^2\mu_{\sigma_k}(0, \infty),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{II} &= \int_{x\leq 0} P(-\gamma e^{x/2} + Z_t \in C)P(Y_{t-1} = 0)\mu_{\sigma_k}(dx) \\
&\quad + \int_{x\leq 0} \sum_{i=1}^{\infty} P(Y_{t-1} = i)P(\gamma(i - e^x)e^{-x/2} + Z_t \in C)\mu_{\sigma_k}(dx) \\
&\geq \int_{x\leq 0} P\left(\frac{-c + \gamma}{\sigma_k} \leq Z \leq \frac{c}{\sigma_k}\right) P(Y_{t-1} = 0)\mu_{\sigma_k}(dx) \\
&\quad + \int_{x\leq 0} \sum_{i=1}^{\infty} P(Y_{t-1} = i)P\left(\frac{-c}{\sigma_k} \leq Z \leq \frac{c - \gamma(i - e^x)e^{-x/2}}{\sigma_k}\right) \mu_{\sigma_k}(dx) \\
&\geq \int_{x\leq 0} (1 - \epsilon_1)P(Y_{t-1} = 0)\mu_{\sigma_k}(dx) \\
&\quad + \int_{x\leq 0} \sum_{i=1}^{N} P(Y_{t-1} = i)P\left(\frac{-c}{\sigma_k} \leq Z \leq \frac{c - \gamma(N - e^x)e^{-x/2}}{\sigma_k}\right) \mu_{\sigma_k}(dx) \\
&\geq \int_{x\leq 0} (1 - \epsilon_1)P(Y_{t-1} = 0)\mu_{\sigma_k}(dx) \\
&\quad + \int_{-M\leq x\leq 0} \sum_{i=1}^{N} P(Y_{t-1} = i)P\left(\frac{-c}{\sigma_k} \leq Z \leq \frac{c - \gamma(N - e^{-M})e^{M/2}}{\sigma_k}\right) \mu_{\sigma_k}(dx)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \int_{x\leq 0}(1-\epsilon_1)P(Y_{t-1}=0)\mu_{\sigma_k}(dx)+\int_{-M\leq x\leq 0}\sum_{i=1}^{N}P(Y_{t-1}=i)(1-\epsilon_1)\mu_{\sigma_k}(dx)\\
&\geq \int_{x\leq 0}(1-\epsilon_1)\sum_{i=0}^{N}P(Y_{t-1}=i)\mu_{\sigma_k}(dx)-\int_{x\leq -M}\sum_{i=1}^{N}P(Y_{t-1}=i)\mu_{\sigma_k}(dx).
\end{aligned}
$$

Now, if $Y \sim$ Poisson(λ_1) and $X \sim$ Poisson(1), $0 \leq \lambda_1 \leq 1$, then

$$
\begin{aligned}
P(Y > N) &= \sum_{i=N+1}^{\infty}\frac{e^{-\lambda_1}\lambda_1^i}{i!} = e^{-\lambda_1}\lambda_1^{N+1}\sum_{i=N+1}^{\infty}\frac{\lambda_1^{i-N-1}}{i!}\\
&\leq e^{-\lambda_1}\lambda_1\sum_{i=N+1}^{\infty}\frac{1}{i!} \leq e^{-1}\sum_{i=N+1}^{\infty}\frac{1}{i!} = P(X > N).
\end{aligned}
$$

since $e^{-x}x$ is an increasing function with a maximum at $x = 1$ for $0 \leq x \leq 1$. Note that this implies that $P(Y \leq N) \geq e^{-1}\sum_{i=0}^{N}\frac{1}{i!}$. Now, if $Y \sim$ Poisson(λ_1) and $X \sim$ Poisson(λ_2), $0 \leq \lambda_1 \leq \lambda_2 \leq 1$, then

$$
\begin{aligned}
P(1 \leq Y \leq N) &= \sum_{i=1}^{N}\frac{e^{-\lambda_1}\lambda_1^i}{i!} = e^{-\lambda_1}\lambda_1\sum_{i=1}^{N}\frac{\lambda_1^{i-1}}{i!}\\
&\leq e^{-\lambda_2}\lambda_2\sum_{i=1}^{N}\frac{\lambda_2^{i-1}}{i!} = P(1 \leq X \leq N).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\text{II} &\geq \int_{x\leq 0}(1-\epsilon_1)\sum_{i=0}^{N}\frac{e^{-1}}{i!}\mu_{\sigma_k}(dx)-\int_{x\leq -M}\sum_{i=1}^{N}\frac{e^{-e^{-M}}e^{-Mi}}{i!}\mu_{\sigma_k}(dx)\\
&\geq (1-\epsilon_1)^2\mu_{\sigma_k}(-\infty,0]-\epsilon_2.
\end{aligned}
$$

Therefore,

$$
\mu_{\sigma_k}(C) \geq (1-\epsilon_1)^2-\epsilon_2 = 1-\epsilon.
$$

Hence, the sequence of probability measures $\{\mu_{\sigma_k} : k \in \mathbb{Z}_+\}$ is tight for any sequence $\{\sigma_k\}$ satisfying $\sigma_1 = 1, \sigma_k \to 0$ as $k \to \infty$.□

Theorem 2.2.8 *The process $\{W_t\}$ defined by (2.16) has at least one invariant measure.*

Proof: Follows by Theorem 2.2.6 and Proposition 2.2.7.□

2.2.2.3 Reachable Point

One approach we have considered for establishing the uniqueness of a stationary solution concerns the theory of e-chains. Although we have not been able to establish that the sequence of transition probabilities is equicontinuous, we have been able to show that the chain is bounded in probability on average (Section 2.2.2.1, Theorem 2.2.3) and that the chain possesses a reachable point. The proof of the latter is the focus of this section.

Proposition 2.2.9 *The process defined by (2.16) has a reachable point (Definition A.2.9), x^*, where x^* is defined as the unique solution to the following equation:*

$$x = -\gamma e^{x/2}. \tag{2.19}$$

Proof: We will first show that the n^{th} iterate of $g(x_0) := -\gamma e^{x_0/2}$ converges to x^* by showing that the sequences $\{x_{2n}\}$ and $\{x_{2n+1}\}$ converge to x^*, where x_n is defined to be the n^{th} iterate of $g(x_0)$. Without loss of generality, we will assume $x_0 < x^*$. It is easily seen that under this assumption, $x_{2n} < x^* < x_{2n+1}$ for all n.

Define $h(\cdot)$ to be the inverse function of $g(\cdot)$ (i.e., the reflection of $g(\cdot)$ about the line $y = x$). Then $h(x) = 2\ln\left(\frac{-x}{\gamma}\right)$. Since $g(\cdot)$ is strictly decreasing, we have

1. for $x < x^*$, $h(x) > g(x)$
2. for $x > x^*$, $h(x) < g(x)$
3. if $g(x) = h(y)$ then
   (a) $x < y$ if $x < x^*$,
   (b) $x > y$ if $x > x^*$,
   (c) $x = y$ if $x = x^*$.

Combining these results we obtain $x_{2n} < x_{2n+2}$ and $x_{2n+1} < x_{2n-1}$ for all n since $h(x_{2n+2}) = x_{2n+1} = g(x_{2n})$ and $h(x_{2n+1}) = x_{2n} = g(x_{2n-1})$. A graph depicting $g(x)$, $h(x)$ and $y = x$ is given in Figure 2.2. The above relationships are seen more easily upon examination of the graph.

To show that $\{x_{2n}\}$ and $\{x_{2n+1}\}$ converge to x^*, suppose $\lim_{n\to\infty} x_{2n} = b$ and $\lim_{n\to\infty} x_{2n+1} = a$. Now, $a = g(b) = h(b)$ and $b = g(a) = h(a)$. From 3c. this implies that $a = b = x^*$. Thus, $\{x_{2n}\}$ and $\{x_{2n+1}\}$ converge to x^* from which it follows that $\{x_n\}$ converges to x^*.

Now, let O be an open set containing x^*. Since $\{x_n\}$ converges to x^*, we may choose M satisfying $x_M \in O$. Then

$$\begin{aligned} P^M(y, O) &= P(W_M \in O \mid W_0 = y) \\ &\geq P(W_1 = x_1, W_2 = x_2, \ldots, W_M = x_M \mid W_0 = y) \\ &= P(W_M = x_M \mid W_{M-1} = x_{M-1}) \cdots P(W_1 = x_1 \mid W_0 = y) \\ &= P(Y_{M-1} = 0 \mid W_{M-1} = x_{M-1}) \cdots P(Y_0 = 0 \mid W_0 = y) \\ &> 0. \end{aligned}$$

Therefore, $\sum_n P^n(y, O) > 0$; hence, x^* is reachable.□

2.3 Addendum

Here we show that a stronger "tightness" condition holds for the GLARMA models of Section 2.2 with $\mathbf{x}_t'\boldsymbol{\beta} = 0$, $p = 0$, $q = 1$ and $1/2 \leq \lambda \leq 1$ using Chebyshev's inequality. This condition provides a more concise means of establishing the existence of a stationary distribution and will hopefully aid in proving that the process is an e-chain. We begin by stating the results for a general Markov chain and conclude by showing that the conditions are met for the GLARMA models of Section 2.2 under certain assumptions.

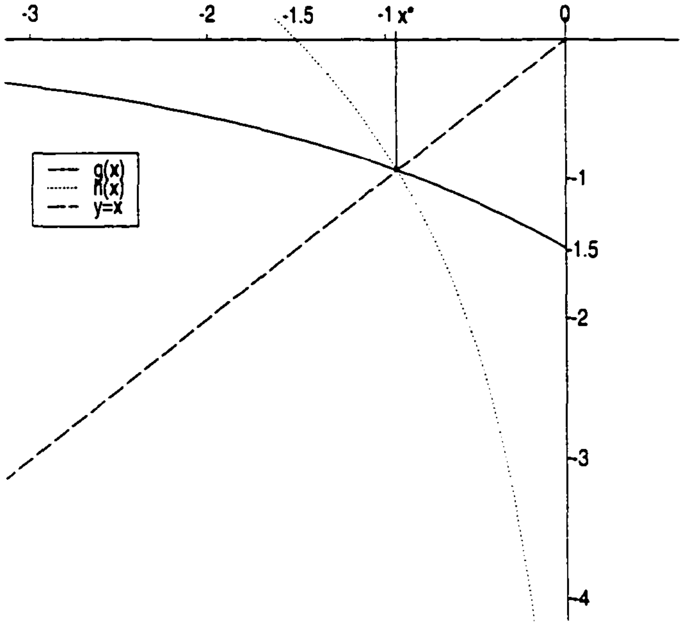

Figure 2.2: Reachable point: Poisson Process. $g(x) = -1.5e^{x/2}$, $h(x) = 2\ln\left(\frac{-x}{1.5}\right)$

Proposition 2.3.1 *If $\{\mathbf{X}_n\}$ is a weak Feller chain and if for any $\epsilon > 0$ there exists a compact set $C \subset X$ such that*

$$P(x, C^c) < \epsilon, \text{ for all } x \in X,$$

then $\{\mathbf{X}_n\}$ is bounded in probability; thus, there exists at least one stationary distribution for the chain.

Proof: Assume that for any $\epsilon > 0$ there exists a compact set $C \subset X$ such that $P(x, C^c) < \epsilon$ for all $x \in X$. If $P^k(x, \cdot)$ denotes the k-step transition probability of the chain starting from state x then,

$$\begin{aligned} P^k(x, C^c) &= \int P(y, C^c) P^{k-1}(x, dy) \\ &< \epsilon. \end{aligned}$$

Thus, the chain is bounded in probability. In fact, the tightness of the k-step transition probabilities holds uniformly in x. It follows that the chain is bounded in probability on average and hence, by Theorem A.2.13, there exists a stationary distribution.□

Proposition 2.3.2 *Let*

$$Y_t \sim \text{Poisson}(e^{W_t}),$$

where

$$W_t = \gamma(Y_{t-1} - e^{W_{t-1}})e^{-\lambda W_{t-1}}.$$

$1/2 \leq \lambda \leq 1$. *Then the chain is bounded in probability, and therefore, admits an invariant measure.*

Proof: First note that the chain is weak Feller. Define $C := [-c, c]$. Then,

$$\begin{aligned} P(x, C^c) &= P(W_t \in C \mid W_{t-1} = x) \\ &= P\big[\gamma(Y_{t-1} - e^{W_{t-1}})e^{-\lambda W_{t-1}} \in [-c, c]^c \mid W_{t-1} = x\big], \end{aligned}$$

which, by Markov's inequality,

$$\begin{aligned} &\leq \begin{cases} (\gamma/c)^2 e^{-2\lambda x}\text{Var}\big[Y_t \mid W_{t-1} = x\big], & x \geq 0 \\ (\gamma/c)e^{-\lambda x}E\big[\lvert Y_{t-1} - e^x\rvert \mid W_{t-1} = x\big], & x < 0 \end{cases} \\ &\leq \begin{cases} (\gamma/c)^2 e^{(1-2\lambda)x}, & x \geq 0 \\ 2(\gamma/c)e^{(1-\lambda)x}, & x < 0 \end{cases} \\ &\leq \begin{cases} (\gamma/c)^2, & x \geq 0 \\ 2(\gamma/c), & x < 0. \end{cases} \end{aligned}$$

Thus, given $\epsilon > 0$ choose c large such that $\max\left(2(\gamma/c), (\gamma/c)^2\right) < \epsilon$. The result follows from Proposition 2.2.9.□

Chapter 3

THE BERNOULLI PROCESS

3.1 Background

The focus of this chapter will be on a component used in the modeling framework of Rydberg and Shephard (1998) for price activity, previously mentioned in Section 1.5.3. Recall that they suggest modeling the price change of a given stock by considering three separate processes: one which models whether or not the price of the stock has moved (activity); one modeling the direction of the price change (direction); and one modeling the magnitude of the change (size). The model they favor for analyzing the price activity is a GLARMA model. Recall that $N_t \in \{0, 1\}$ is the event of a price change at time t. If $\mathcal{F}_{t-1}$ represents the σ-field generated by $\{N_x, x \leq t\}$, the conditional probability of N_t given $\mathcal{F}_{t-1}$ is given by

$$p(N_t = n \mid \mathcal{F}_{t-1}) = p_t^n (1 - p_t)^{1-n} I_{\{0,1\}}(n),$$

where $p_t = e^{\theta_t}(1 + e^{\theta_t})^{-1}$, $\theta_t = \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t$, $\mathbf{x}_t$ is a vector of exogenous variables, and

$$\epsilon_t = \sum_{j=1}^{p} \gamma_j \epsilon_{t-j} + \sigma U_t + \sigma \sum_{j=1}^{q} \delta_j U_{t-j}, \quad \sigma > 0,$$

with

$$U_t = \frac{N_{t-1} - p_{t-1}}{\sqrt{p_{t-1}(1 - p_{t-1})}}, \; t \geq 1. \tag{3.1}$$

Note that $\{U_t\}$ is a martingale difference sequence:

$$\begin{aligned} E[U_{t+1} \mid \mathcal{F}_t] &= E\left[\frac{N_t - p_t}{\sqrt{p_t(1-p_t)}} \,\Big|\, U_{t-1}\right] \\ &= \left(p_t(1-p_t)\right)^{-1/2}\left(E[N_t \mid U_{t-1}] - p_t\right) \\ &= 0, \end{aligned}$$

with

$$\begin{aligned} \mathrm{Var}[U_{t+1} \mid \mathcal{F}_t] &= E\left[\left(p_t(1-p_t)\right)^{-1}(N_t - p_t)^2 \mid \mathcal{F}_t\right] \\ &= \left(p_t(1-p_t)\right)^{-1} p_t(1-p_t) = 1. \end{aligned}$$

In this chapter, we show that the process $\{U_t\}$ has a unique stationary distribution for the special case of the GLARMA model where $\beta = p = q = 0$. In this case,

$$p_t = e^{\sigma U_t}\left(1 + e^{\sigma U_t}\right)^{-1}, \tag{3.2}$$

where U_t is given by (3.1). Results from Meyn and Tweedie (1993) will be applied to prove the above.

3.2 Asymptotic Results

In order to establish the existence of a unique stationary distribution for the process $\{U_t\}$, we will show that the conditions of Theorem A.2.18 are met. In other words, we need to show the following:

1. The process $\{U_t\}$ is an e-chain (Definition A.2.17), that is, its transition probabilities are equicontinuous.

2. The chain possesses a reachable point (Definition A.2.9).

3. The chain is bounded in probability on average (Definition A.2.12). Alone, this condition provides a means for establishing the existence of a stationary

distribution for weak Feller chains without the requirement of irreducibility. Coupled with the conditions given by (1) and (2) above, boundedness in probability on average provides us with a tool for verifying the existence of a unique stationary distribution.

3.2.1 E-chain

Theorem 3.2.1 *The process $\{U_t\}$ defined by (3.1) and (3.2) is an e-chain.*

Proof:

In order to establish equicontinuity of the transition probabilities, it is necessary to show that for any continuous function f with compact support, the functions $P_x^k f, k = 1, 2, \ldots$ are equicontinuous on every compact set C. We will first derive a more convenient expression for $|P_x^k f - P_z^k f|$. Now,

$$\begin{aligned}
&|P_{x_0} f - P_{z_0} f| \\
&:= |E_{x_0}[f(U_1)] - E_{z_0}[f(U_1)]| \\
&= \left| \sum_{n_0=0}^{1} \left\{ p(n_0 \mid x_0) f(x_1) - p(n_0 \mid z_0) f(z_1) \right\} \right| \\
&\leq \sum_{n_0=0}^{1} \left\{ |p(n_0 \mid x_0) - p(n_0 \mid z_0)| \, |f(z_1)| + p(n_0 \mid x_0) |f(x_1) - f(z_1)| \right\}. \qquad (3.3)
\end{aligned}$$

where

$$x_i := n_{i-1}(1 + e^{\sigma x_{i-1}}) e^{-\sigma x_{i-1}/2} - e^{\sigma x_{i-1}/2}$$

and

$$z_i := n_{i-1}(1 + e^{\sigma z_{i-1}}) e^{-\sigma z_{i-1}/2} - e^{\sigma z_{i-1}/2}.$$

Note that

$$x_i = \begin{cases} -e^{\sigma x_{i-1}/2}, & \text{if } n_{i-1} = 0, \\ e^{-\sigma x_{i-1}/2}, & \text{if } n_{i-1} = 1, \end{cases} \qquad (3.4)$$

$$p(n_i = 1 \mid x_i) = \frac{e^{\sigma x_i}}{(1 + e^{\sigma x_i})} = (1 + e^{-\sigma x_i})^{-1}, \qquad (3.5)$$

and

$$p(n_i = 0 \mid x_i) \;=\; (1 + e^{\sigma x_i})^{-1}, \tag{3.6}$$

with similar relations holding with x_i replaced by z_i. Throughout the remainder of the proof, we shall assume x_i and z_i evolve according to (3.4) for the same set of $n_0, n_1, \ldots$. Replacing f with Pf in (3.3), we have

$$\begin{aligned}
&\left|P^2_{x_0} f - P^2_{z_0} f\right| \\
&\quad \le \sum_{n_0=0}^{1} \left\{ \left|p(n_0 \mid x_0) - p(n_0 \mid z_0)\right| \left|P_{z_1} f\right| + p(n_0 \mid x_0) \left|P_{x_1} f - P_{z_1} f\right| \right\} \\
&\quad \le \sum_{n_0, n_1 = 0}^{1} \Big\{ \left|p(n_0 \mid x_0) - p(n_0 \mid z_0)\right| p(n_1 \mid z_1) \left|f(z_2)\right| \\
&\qquad + p(n_0 \mid x_0) \Big(\left|p(n_1 \mid x_1) - p(n_1 \mid z_1)\right| \left|f(z_2)\right| + p(n_1 \mid x_1) \left|f(x_2) - f(z_2)\right| \Big) \Big\} \\
&\quad = \sum_{n_0, n_1 = 0}^{1} \Big\{ \Big(\left|p(n_0 \mid x_0) - p(n_0 \mid z_0)\right| p(n_1 \mid z_1) \\
&\qquad + p(n_0 \mid x_0) \left|p(n_1 \mid x_1) - p(n_1 \mid z_1)\right| \Big) \left|f(z_2)\right| + p(n_0 \mid x_0) p(n_1 \mid x_1) \left|f(x_2) - f(z_2)\right| \Big\}.
\end{aligned}$$

Iterating, we obtain

$$\begin{aligned}
&\left|P^k_{x_0} f - P^k_{z_0} f\right| \\
&\quad \le \sum_{n_0=0}^{1} \left\{ \left|p(n_0 \mid x_0) - p(n_0 \mid z_0)\right| \left|P^{k-1}_{z_1} f\right| + p(n_0 \mid x_0) \left|P^{k-1}_{x_1} f - P^{k-1}_{z_1} f\right| \right\} \\
&\quad \le \sum_{n_0, n_1 = 0}^{1} \Big\{ \left|p(n_0 \mid x_0) - p(n_0 \mid z_0)\right| p(n_1 \mid z_1) \left|P^{k-2}_{z_2} f\right| \\
&\qquad + p(n_0 \mid x_0) \Big(\left|p(n_1 \mid x_1) - p(n_1 \mid z_1)\right| \left|P^{k-2}_{z_2} f\right| + p(n_1 \mid x_1) \left|P^{k-2}_{x_2} f - P^{k-2}_{z_2} f\right| \Big) \Big\} \\
&\quad \vdots \\
&\quad \le \sum_{n_0, \ldots, n_{k-1} = 0}^{1} \Big\{ \Big(\left|p(n_0 \mid x_0) - p(n_0 \mid z_0)\right| p(n_1 \mid z_1) \cdots p(n_{k-1} \mid z_{k-1}) \\
&\qquad + p(n_0 \mid x_0) \left|p(n_1 \mid x_1) - p(n_1 \mid z_1)\right| p(n_2 \mid z_2) \cdots p(n_{k-1} \mid z_{k-1}) \\
&\qquad + \ldots + p(n_0 \mid x_0) \cdots p(n_{k-2} \mid x_{k-2}) \left|p(n_{k-1} \mid x_{k-1}) - p(n_{k-1} \mid z_{k-1})\right| \Big) \left|f(z_k)\right| \\
&\qquad + p(n_0 \mid x_0) \cdots p(n_{k-1} \mid x_{k-1}) \left|f(x_k) - f(z_k)\right| \Big\}
\end{aligned}$$

$$= \sum_{i=0}^{k-1} \Bigg\{ \sum_{n_0,\ldots,n_{k-1}=0}^{1} \big|f(z_k)\big|\, p(n_0 \,|\, x_0) \cdots p(n_{i-1} \,|\, x_{i-1})$$

$$\cdot \big|p(n_i \,|\, x_i) - p(n_i \,|\, z_i)\big|\, p(n_{i+1} | z_{i+1}) \cdots p(n_{k-1} \,|\, z_{k-1}) \Bigg\}$$

$$+ \sum_{n_0,\ldots,n_{k-1}=0}^{1} p(n_0 \,|\, x_0) \cdots p(n_{k-1} \,|\, x_{k-1}) \big|f(x_k) - f(z_k)\big|, \qquad (3.7)$$

where $p(n_i \,|\, x_i) \cdots p(n_{i-1} \,|\, x_{i-1}) := 1$. In order to extend this bound, it will be necessary to bound $\big|p(n_i \,|\, x_i) - p(n_i \,|\, z_i)\big|$ and $\big|x_{i+1} - z_{i+1}\big|$. An application of the Mean Value Theorem gives us the following bounds:

$$\begin{aligned} \big|p(n_i \,|\, x_i) - p(n_i \,|\, z_i)\big| &= \begin{cases} \big|(1 + e^{\sigma x_i})^{-1} - (1 + e^{\sigma z_i})^{-1}\big|, & \text{if } n_i = 0, \\ \big|(1 + e^{-\sigma x_i})^{-1} - (1 + e^{-\sigma z_i})^{-1}\big|, & \text{if } n_i = 1, \end{cases} \\ &= \begin{cases} \frac{\sigma e^{\sigma c_1}}{(1+e^{\sigma c_1})^2}\big|x_i - z_i\big|, & \text{if } n_i = 0, \\ \frac{\sigma e^{-\sigma c_2}}{(1+e^{-\sigma c_2})^2}\big|x_i - z_i\big|, & \text{if } n_i = 1, \end{cases} \\ &= \begin{cases} \frac{\sigma}{(e^{-\sigma c_1/2}+e^{\sigma c_1/2})^2}\big|x_i - z_i\big|, & \text{if } n_i = 0, \\ \frac{\sigma}{(e^{\sigma c_2/2}+e^{-\sigma c_2/2})^2}\big|x_i - z_i\big|, & \text{if } n_i = 1, \end{cases} \\ &\le (\sigma/4)\big|x_i - z_i\big|, \end{aligned} \qquad (3.8)$$

since $e^{-x} + e^{x} \ge 2$, and

$$\begin{aligned} \big|x_{i+1} - z_{i+1}\big| &= \begin{cases} \big|e^{\sigma x_i/2} - e^{\sigma z_i/2}\big|, & n_i = 0, \\ \big|e^{-\sigma x_i/2} - e^{-\sigma z_i/2}\big|, & n_i = 1, \end{cases} \\ &= \begin{cases} \frac{\sigma}{2} e^{\sigma c_3/2}\big|x_i - z_i\big|, & n_i = 0, \\ \frac{\sigma}{2} e^{-\sigma c_4/2}\big|x_i - z_i\big|, & n_i = 1, \end{cases} \end{aligned} \qquad (3.9)$$

where c_i, $i = 1, \ldots, 4$, are constants between x_i and z_i.

Without loss of generality, we will assume $\sup_{x \in C} |f(x)| \le 1$ since f, being continuous with compact support, is bounded. Upon combining these assumptions with (3.8), the bound on $\big|P^k_{x_0} f - P^k_{z_0} f\big|$ may be extended as follows:

$$\big|P^k_{x_0} f - P^k_{z_0} f\big| \le (\sigma/4) \sum_{i=0}^{k-1} \Bigg\{ \sum_{n_0,\ldots,n_{k-1}=0}^{1} \Big(p(n_0 \,|\, x_0) \cdots p(n_{i-1} \,|\, x_{i-1}) \big|x_i - z_i\big|$$

$$\cdot p(n_{i+1} | z_{i+1}) \cdots p(n_{k-1} \,|\, z_{k-1}) \Big) \Bigg\}$$

$$+ \sum_{n_0,\ldots,n_{k-1}=0}^{1} p(n_0 \,|\, x_0) \cdots p(n_{k-1} \,|\, x_{k-1}) \big|f(x_k) - f(z_k)\big|. \qquad (3.10)$$

For the remainder of the argument, we require that σ be such that there exist two solutions, $y_l < y_u$, to the equation $e^{\sigma|x|/2} = |x|$. Note that for $\sigma = 2/e$, there exists exactly one solution to the equation, namely, $x = e$. Therefore, it is required that $\sigma < 2/e$. Define $y_m := \text{soln}(\frac{d}{dx}(e^{\sigma|x|/2}) = 1)$. Since $g(x) := e^{\sigma|x|/2}$ is convex, $x > g(x)$ for $y_l < x < y_u$, and $x < g(x)$ for $x \in (0, y_l) \cup (y_u, \infty)$. It is useful to indicate the following:

$$y_l > 1 \text{ for } \sigma > 0, \tag{3.11}$$

and

$$(\sigma/2)y_l = (\sigma/2)e^{(\sigma/2)y_l} < (\sigma/2)e^{(\sigma/2)y_m} = 1. \tag{3.12}$$

One might also note that $-y_l > -y_u$ are the two solutions to $-g(-x) = -x$ and $-y_m = \text{soln}(\frac{d}{dx}(-e^{-\sigma|x|/2}) = 1)$. The functions $g(x), -g(-x)$ and $h(x) := x$ along with $\pm y_l, \pm y_m$ and $\pm y_u$ are plotted in Figure 3.1 for the case $\sigma = 1/2$. There are four separate cases of initial values to consider: $|x_0|, |z_0| \leq y_l$, $y_l \leq |x_0|, |z_0| \leq y_m$, $y_m \leq |x_0|, |z_0| \leq y_u$ and $|x_0|, |z_0| \geq y_u$. Note that because the distance between x_0 and z_0 can be chosen to be arbitrarily small, if $|x_0|, |z_0|$ are not in the same case, they must be in adjacent ones. Thus, since each case includes its boundaries, it is not necessary to consider any additional cases.

Before proceeding with the proof of the four cases, it is helpful to point out two key characteristics of the realizations $\{x_k\}$.

c1. First, since the sequences $\{x_k\}$ and $\{z_k\}$ are driven by the same sequence of observed values of the Bernoulli random variables, $\{n_k\}$, it is easily seen from (3.4) that x_i and z_i have the same sign for $i \geq 1$. Thus, we will assume without loss of generality that $x_0, z_0 > 0$ throughout.

c2. Now, if we define $g_n(x)$ to be the n^{th} iterate of $g(x) := e^{\sigma|x|/2}$, with $g_0(x) := |x|$, it can also be seen from (3.4) that $|x_n| = g_n(x_0)$ (which is ≥ 1) if and only

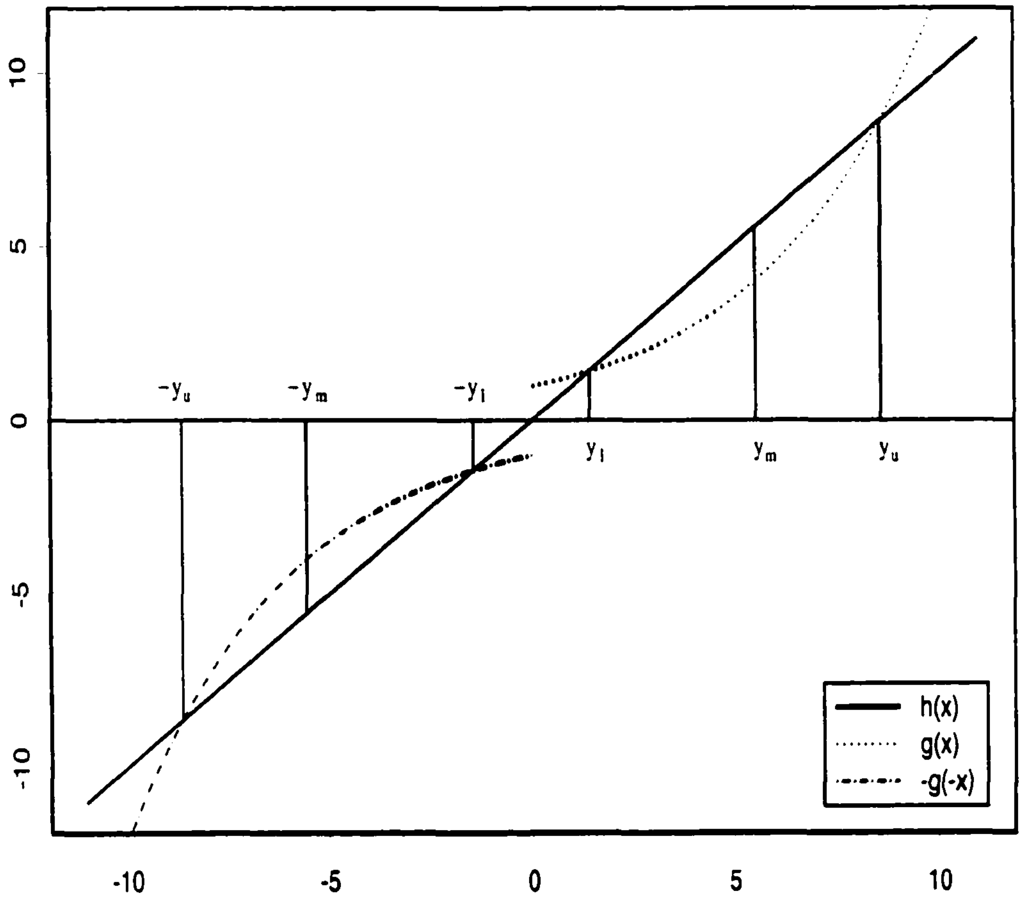

Figure 3.1: E-chain cases. $h(x) = x, g(x) = e^{x/4}$

if $\text{sign}(x_i) \neq \text{sign}(x_{i-1})$ for all $i \leq n$. If $\text{sign}(x_i) = \text{sign}(x_{i-1})$ for some $i \leq n$, say $i = s$, and $\text{sign}(x_i) \neq \text{sign}(x_{i-1})$ for all $i < s$, then $|x_s| = e^{-(\sigma/2)|x_{s-1}|} \leq e^{(\sigma/2)|x_{s-1}|} = e^{(\sigma/2)g_{s-1}(x_0)} = g_s(x_0)$. Therefore, by (3.4), $|x_{s+1}| \leq e^{(\sigma/2)|x_s|} = e^{(\sigma/2)g_s(x_0)} = g_{s+1}(x_0)$, from which it follows that $|x_n| \leq g_n(x_0)$, for all $n \geq 0$.

Throughout the proof, we will denote the first $i \geq 1$ such that $\text{sign}(x_i) = \text{sign}(x_{i-1})$ by τ_A.

case 1: $|x_0|, |z_0| \leq y_l$

We begin by noting that $|x_i|, |z_i| \leq y_l$ for all $i \geq 1$ when $x_0,\ z_0 \leq y_l$ (recall from cl that $x_0,\ z_0 > 0$ by assumption). To see this, recall (3.4):

$$x_i = \begin{cases} -e^{\sigma x_{i-1}/2}, & n_{i-1} = 0, \\ e^{-\sigma x_{i-1}/2}, & n_{i-1} = 1. \end{cases}$$

It follows that for $0 \leq x_{i-1} \leq y_l$,

$$x_i \in \begin{cases} [-y_l, -1], & n_{i-1} = 0, \\ [e^{-\sigma y_l/2}, 1], & n_{i-1} = 1. \end{cases}$$

Likewise, for $-y_l \leq x_{i-1} \leq 0$,

$$x_i \in \begin{cases} [-1, -e^{-\sigma y_l/2}], & n_{i-1} = 0, \\ [\ 1, y_l], & n_{i-1} = 1. \end{cases}$$

Thus, $|x_i| \leq y_l$ for all $i \geq 1$ if $x_0 \leq y_l$.

Now, since $e^{\sigma c/2} \leq e^{\sigma y_l/2} = y_l$ for $c \leq y_l$, and $y_l > 1$, we see from (3.9) that

$$\left|x_{i+1} - z_{i+1}\right| \ \leq\ (\sigma/2) y_l \left|x_i - z_i\right|. \tag{3.13}$$

Combining this result with the bound for $\left|P^k_{x_0} f - P^k_{z_0} f\right|$ given by (3.10), we obtain:

$$\begin{aligned} &\left|P^k_{x_0} f - P^k_{z_0} f\right| \\ &\quad \leq (\sigma/4) \sum_{i=0}^{k-1} \Bigg\{ \sum_{n_0,\dots,n_{k-1}=0}^{1} \Big(p(n_0 \mid x_0) \cdots p(n_{i-1} \mid x_{i-1}) (\sigma y_l/2)^i |x_0 - z_0| \\ &\qquad \cdot p(n_{i+1} | z_{i+1}) \cdots p(n_{k-1} \mid z_{k-1}) \Big) \Bigg\} \\ &\qquad + \sum_{n_0,\dots,n_{k-1}=0}^{1} p(n_0 \mid x_0) \cdots p(n_{k-1} \mid x_{k-1}) \left| f(x_k) - f(z_k) \right|. \end{aligned}$$

Since $\sum_{n_0,\dots,n_{k-1}=0}^{1} p(n_0 \mid x_0) \cdots p(n_{i-1} \mid x_{i-1}) p(n_{i+1} | z_{i+1}) \cdots p(n_{k-1} \mid z_{k-1}) \ = \ 2$, we have

$$\begin{aligned} &\left|P^k_{x_0} f - P^k_{z_0} f\right| \\ &\quad \leq (\sigma/2) \left|x_0 - z_0\right| \sum_{i=0}^{k-1} (\sigma y_l/2)^i + \sum_{n_0,\dots,n_{k-1}=0}^{1} p(n_0 \mid x_0) \cdots p(n_{k-1} \mid x_{k-1}) \left| f(x_k) - f(z_k) \right| \\ &\quad \leq \frac{\sigma \left|x_0 - z_0\right|}{2 - \sigma y_l} + \sum_{n_0,\dots,n_{k-1}=0}^{1} p(n_0 \mid x_0) \cdots p(n_{k-1} \mid x_{k-1}) \left| f(x_k) - f(z_k) \right|. \end{aligned} \tag{3.14}$$

Recall from (3.12) that $(\sigma/2)y_l < 1$ which implies $2-\sigma y_l > 0$. Therefore, given $\delta > 0$, we may choose ϵ such that for $|x_0 - z_0| < \epsilon$, $\frac{\sigma|x_0-z_0|}{2-\sigma y_l} < \delta/2$ and $|f(x_0) - f(z_0)| < \delta/2$. Since $(\sigma/2)y_l < 1$, it follows by (3.13) that $|x_k - z_k| \leq |x_0 - z_0| < \epsilon$, and hence,

$$|P^k_{x_0}f - P^k_{z_0}f| < \delta/2 + \delta/2 = \delta.$$

Thus, for $|x_0|, |z_0| \leq y_l$, the transition probabilities are equicontinuous.

case 2: $y_l \leq |x_0|, |z_0| \leq y_m$

Here we will use the fact that $\frac{d}{dx}(e^{\sigma x/2}) = (\sigma/2)e^{\sigma x/2} \leq 1$ for $y_l \leq x \leq y_m$ and $\frac{d}{dx}(-e^{-\sigma x/2}) = (\sigma/2)e^{-\sigma x/2} \leq 1$ for $-y_m \leq x \leq -y_l$ (see Figure 3.1) to show that the successive distances between the sequences $\{x_k\}$ and $\{z_k\}$ contract. From (3.9) and the preceeding statement, we have

$$\begin{aligned} |x_1 - z_1| &= \begin{cases} \frac{\sigma}{2}e^{\sigma c_1/2}|x_0 - z_0|, & n_i = 0 \\ \frac{\sigma}{2}e^{-\sigma c_2/2}|x_0 - z_0|, & n_i = 1 \end{cases} \\ &\leq |x_0 - z_0|, \end{aligned} \tag{3.15}$$

where again, c_1, c_2 are constants between x_0 and z_0. In this case, it is possible for $|c_i| = y_m$ which would imply $(\sigma/2)e^{\sigma|c_i|/2} = 1$. Thus, it is necessary to carry this bound a step further. Similar to case 1, it is easily shown that $|x_i| \leq e^{(\sigma/2)y_m}$ for all $i \geq 1$ if $|x_0| \leq y_m$. From (3.4), we have that for $y_l \leq x_{i-1} \leq y_m$:

$$x_i \in \begin{cases} [-e^{(\sigma/2)y_m}, -y_l], & n_{i-1} = 0, \\ [e^{-(\sigma/2)y_m}, e^{-(\sigma/2)y_l}], & n_{i-1} = 1, \end{cases}$$

and for $-y_m \leq x_{i-1} \leq -y_l$:

$$x_i \in \begin{cases} [-e^{-(\sigma/2)y_l}, -e^{-(\sigma/2)y_m}], & n_{i-1} = 0, \\ [y_l, e^{(\sigma/2)y_m}], & n_{i-1} = 1. \end{cases}$$

Hence, $|x_i| \leq e^{(\sigma/2)y_m} = g_1(y_m)$ for all $i \geq 1$ if $|x_0| \leq y_m$. Thus, for c between $|x_1|$ and $|z_1|$, we obtain the following:

$$\begin{aligned} |x_2 - z_2| &\leq (\sigma/2)e^{\sigma c/2}|x_1 - z_1| \\ &\leq (\sigma/2)e^{\sigma g_1(y_m)/2}|x_1 - z_1| \\ &\leq (\sigma/2)e^{\sigma g_1(y_m)/2}|x_0 - z_0| \text{ (by 3.15)}. \end{aligned}$$

Note that since $g_1(y_m) < y_m$ (see Figure 3.1), $(\sigma/2)e^{\sigma g_1(y_m)/2} < 1$. This implies $|x_2 - z_2| \le |x_0 - z_0|$. It then follows from (3.9) that for all $i \ge 1$,

$$|x_i - z_i| \le \left((\sigma/2)e^{\sigma g_1(y_m)/2}\right)^{i-1} \big|x_0 - z_0\big| := r^{i-1}\big|x_0 - z_0\big|.$$

Thus, from (3.10), we have

$$\begin{aligned}
\big|P^k_{x_0} f - P^k_{z_0} f\big| \le (\sigma/4)\Big\{ &\sum_{n_0,\dots n_{k-1}=0}^{1} \big|x_0 - z_0\big|\, p(n_1 \,|\, z_1)\cdots p(n_{k-1} \,|\, z_{k-1}) \\
&+ \sum_{i=1}^{k-1} \sum_{n_0,\dots,n_{k-1}=0}^{1} \Big(p(n_0 \,|\, x_0)\cdots p(n_{i-1} \,|\, x_{i-1})\, r^{i-1}\, \big|x_0 - z_0\big| \\
&\cdot p(n_{i+1} | z_{i+1})\cdots p(n_{k-1} \,|\, z_{k-1})\Big)\Big\} \\
&+ \sum_{n_0,\dots,n_{k-1}=0}^{1} p(n_0 \,|\, x_0)\cdots p(n_{k-1} \,|\, x_{k-1})\, \big|f(x_k) - f(z_k)\big| \\
\le (\sigma/2)\,&\big|x_0 - z_0\big| + \frac{(\sigma/2)|x_0 - z_0|}{1 - r} \\
&+ \sum_{n_0,\dots,n_{k-1}=0}^{1} p(n_0 \,|\, x_0)\cdots p(n_{k-1} \,|\, x_{k-1})\, \big|f(x_k) - f(z_k)\big|.
\end{aligned}$$

Hence, given $\delta > 0$, we may choose $\epsilon > 0$ such that for $\big|x_0 - z_0\big| < \epsilon$, $\big|f(x_0) - f(z_0)\big| < \delta/2$, and $(\sigma/2)\big|x_0 - z_0\big|\,[1 + (1-r)^{-1}] < \delta/2$. Then, since $\big|x_k - z_k\big| \le \big|x_0 - z_0\big|$, $\big|P^k_{x_0} f - P^k_{z_0} f\big| < \delta$, for $|x_0 - z_0| < \epsilon$.

case 3: $y_m \le |x_0|, |z_0| \le y_u$

The remaining two cases are quite similar. In this case, given $\delta > 0$, we will choose k^* such that $\frac{2}{(1+e^{\sigma y_l})^{k^*}} < \delta/4$. Next select ϵ_1 satisfying $\big|f(x) - f(z)\big| < \delta/4$ for $|x - z| < \epsilon_1$ and choose ϵ_2 such that for $|x_0 - z_0| < \epsilon_2$, $\max_{0 \le i \le k^*-1} \big|g_i(x_0) - g_i(z_0)\big| < \min\left(\frac{\delta}{2\sigma k^*}, \frac{\delta(2-\sigma y_l)}{2\sigma^2}\right)$, $\max_{0 \le i \le k^*} \big|g_i(x_0) - g_i(z_0)\big| < \epsilon_1$ and $\frac{2\sigma}{2-\sigma}\big|x_0 - z_0\big| < \delta/4$.

We will first consider the case $k \le k^*$. Recall that $\tau_A = j$ if and only if $\text{sign}(x_j) = \text{sign}(x_{j-1})$ and $\text{sign}(x_i) \ne \text{sign}(x_{i-1})$ for all $i < j$. Assume $\tau_A = j \le k^*$. Then, since x_i

and z_i have the same sign (recall c1), it follows from c2 that $|x_i - z_i| = |g_i(x_0) - g_i(z_0)|$ for all $i < j$, and

$$\begin{aligned} |x_j - z_j| &= (\sigma/2)e^{-\sigma|c|/2}|x_{j-1} - z_{j-1}| \\ &= (\sigma/2)e^{-\sigma|c|/2}|g_{j-1}(x_0) - g_{j-1}(z_0)| \\ &\le |g_{j-1}(x_0) - g_{j-1}(z_0)|, \end{aligned}$$

where c is a constant between x_{j-1} and z_{j-1}. Also, note that $|x_j|$ and $|z_j| \le 1$, and hence, $< y_l$. It follows by case 1, for all $i > j$,

$$\begin{aligned} |x_i - z_i| &\le (\sigma y_l/2)|x_{i-1} - z_{i-1}| \\ &\le |x_{i-1} - z_{i-1}| \\ &\vdots \\ &\le |x_j - z_j| \\ &< |g_{j-1}(x_0) - g_{j-1}(z_0)|. \end{aligned}$$

Thus, for all $i \le k \le k^*$, $|x_i - z_i| \le \max_{0\le n\le k^*-1} |g_n(x_0) - g_n(z_0)| \le \max_{0\le n\le k^*} |g_n(x_0) - g_n(z_0)|$ for $\tau_A \le k^*$. Now suppose $\tau_A = j > k^*$. Then $|x_i - z_i| = |g_i(x_0) - g_i(z_0)|$ for all $i \le k \le k^*$. It follows that $|x_i - z_i| \le \max_{0\le n\le k^*-1} |g_n(x_0) - g_n(z_0)|$ for $i < k^*$ and $|x_k - z_k| \le \max_{0\le n\le k^*} |g_n(x_0) - g_n(z_0)|$. Hence, for $k \le k^*$, it follows from (3.10) that

$$\begin{aligned} |P_{x_0}^k f - P_{z_0}^k f| &\le (\sigma/2)\sum_{i=0}^{k-1} \max_{0\le n\le k^*-1} |g_n(x_0) - g_n(z_0)| \\ &\quad + \sum_{n_0,\dots,n_{k-1}=0}^{1} p(n_0 \,|\, x_0)\cdots p(n_{k-1} \,|\, x_{k-1})\, |f(x_k) - f(z_k)| \\ &\le (\sigma/2)k^* \max_{0\le n\le k^*-1} |g_n(x_0) - g_n(z_0)| + \delta/4 \\ &< \delta/2. \end{aligned}$$

For $k > k^*$, a different approach will be taken. Note that for $x_0 > 0$, $\tau_A = j$ if $n_0 = 0$, $n_i \ne n_{i-1}$ for $i \le j-1$ and $n_j = n_{j-1}$. Thus, given $x_0 > 0$, $\tau_A = j$, it follows from (3.4) that $x_i = (-1)^i e^{\sigma|x_{i-1}|/2}$ for $1 \le i < j$ and $x_j = (-1)^{j-1} e^{-\sigma|x_{j-1}|/2}$. Now, consider the following expression for $\left|P^k_{x_0} f - P^k_{z_0} f\right|$:

$$\begin{aligned}
\left|P^k_{x_0} f - P^k_{z_0} f\right| &= \Bigg|\sum_{j=1}^{\infty} E_{x_0}\left[f(U_k)\,|\,\tau_A = j\right] P_{x_0}(\tau_A = j) \\
&\quad - \sum_{j=1}^{\infty} E_{z_0}\left[f(U_k)\,|\,\tau_A = j\right] P_{z_0}(\tau_A = j)\Bigg| \\
&= \Bigg|\sum_{j=1}^{\infty}\left\{E_{x_0}\left[f(U_k)\,|\,\tau_A = j\right] - E_{z_0}\left[f(U_k)\,|\,\tau_A = j\right]\right\} P_{x_0}(\tau_A = j) \\
&\quad - \sum_{j=1}^{\infty}\left\{P_{z_0}(\tau_A = j) - P_{x_0}(\tau_A = j)\right\} E_{z_0}\left[f(U_k)\,|\,\tau_A = j\right]\Bigg| \\
&\le \sum_{j=1}^{\infty} P_{x_0}(\tau_A = j)\left|E_{x_0}\left[f(U_k)\,|\,\tau_A = j\right] - E_{z_0}\left[f(U_k)\,|\,\tau_A = j\right]\right| \\
&\quad + \sum_{j=1}^{\infty}\left|P_{z_0}(\tau_A = j) - P_{x_0}(\tau_A = j)\right| \\
&= \mathrm{I} + \mathrm{II}. \qquad (3.16)
\end{aligned}$$

First we will consider I:

$$\begin{aligned}
\mathrm{I} \le\ & \sum_{j=1}^{k^*} P_{x_0}(\tau_A = j)\left|E_{x_0}\left[f(U_k)\,|\,\tau_A = j\right] - E_{z_0}\left[f(U_k)\,|\,\tau_A = j\right]\right| \\
&+ \sum_{j=k^*+1}^{\infty} 2P_{x_0}(\tau_A = j).
\end{aligned}$$

Note that the sets $\{U_0 = x_0, \tau_A = j\}$ and $\{U_0 = x_0, U_1 = x_1, \ldots, U_j = x_j\}$ are the same, where $x_0 > 0$ by assumption, $x_i = (-1)^i e^{\sigma|x_{i-1}|/2}$ for $1 \le i < j$ and $x_j = (-1)^{j-1} e^{-\sigma|x_{j-1}|/2}$ (note that this implies $|x_j| \le 1$). It then follows by the Markov property of the $U_i^{\prime}s$ that

$$\begin{aligned}
\mathrm{I} &\le \sum_{j=1}^{k^*} P_{x_0}(\tau_A = j)\left|E_{x_j}\left[f(U_{k-j})\right] - E_{z_j}\left[f(U_{k-j})\right]\right| + \sum_{j=k^*+1}^{\infty} 2P_{x_0}(\tau_A = j) \\
&= \sum_{j=1}^{k^*} P_{x_0}(\tau_A = j)\left|P^{k-j}_{x_j} f - P^{k-j}_{z_j} f\right| + \sum_{j=k^*+1}^{\infty} 2P_{x_0}(\tau_A = j) \\
&:= \sum_{j=1}^{k^*} P_{x_0}(\tau_A = j)a + b. \qquad (3.17)
\end{aligned}$$

Since $|x_j|, |z_j| \leq 1 < y_l$ if $\tau_A = j$, to obtain a bound for a, we can use the same idea as in case 1. From case 1, (3.14), we have

$$\begin{aligned} a &:= \left|P_{x_j}^{k-j} f - P_{z_j}^{k-j} f\right| \\ &\leq \frac{\sigma|x_j - z_j|}{2 - \sigma y_l} + \sum_{n_j, \ldots, n_{k-1}=0}^{1} p(n_j \mid x_j) \cdots p(n_{k-1} \mid x_{k-1}) \left|f(x_k) - f(z_k)\right|. \end{aligned}$$

Now, for c between x_{j-1} and z_{j-1}, where recall, x_{j-1} and z_{j-1} have the same sign, we have

$$\begin{aligned} |x_j - z_j| &= \left|e^{-(\sigma/2)|x_{j-1}|} - e^{-(\sigma/2)|z_{j-1}|}\right| \\ &= (\sigma/2) e^{-\sigma|c|/2} |x_{j-1} - z_{j-1}| \\ &\leq (\sigma/2)|x_{j-1} - z_{j-1}| \\ &= (\sigma/2)|g_{j-1}(x_0) - g_{j-1}(z_0)| \\ &\leq (\sigma/2) \max_{0 \leq i \leq k^* - 1} \left|g_i(x_0) - g_i(z_0)\right|. \end{aligned}$$

Thus,

$$\begin{aligned} a &\leq \frac{(\sigma^2/2)}{2 - \sigma y_l} \max_{0 \leq i \leq k^* - 1} \left|g_i(x_0) - g_i(z_0)\right| \\ &\quad + \sum_{n_j, \ldots, n_{k-1}=0}^{1} p(n_j \mid x_j) \cdots p(n_{k-1} \mid x_{k-1}) \left|f(x_k) - f(z_k)\right| \\ &< \delta/4 + \delta/4 = \delta/2. \end{aligned}$$

since $|x_k - z_k| \leq |x_j - z_j| \leq \max_{0 \leq i \leq k^*} \left|g_i(x_0) - g_i(z_0)\right| < \epsilon_1$.

We will use the fact that the probability of observing the sequence $\{n_i : n_i = 1 - n_{i-1}, i \geq 1\}$ decreases as i increases in order to bound b. Recall,

$$b := \sum_{j=k^*+1}^{\infty} 2P_{x_0}(\tau_A = j) = 2P_{x_0}(\tau_A > k^*),$$

and $\tau_A > k^*$ if and only if $n_0 = 0$ and $n_i \neq n_{i-1}$ for $1 \leq i \leq k^*$ ($x_0 > 0$ by assumption). It follows from (3.5) and (3.6) that

$$P_{x_0}(\tau_A > k^*) = \left[(1 + e^{\sigma x_0})(1 + e^{\sigma g_1(x_0)}) \cdots (1 + e^{\sigma g_{k^*-1}(x_0)})\right]^{-1},$$

where $g_i(x_0) = |x_i| = e^{(\sigma/2)g_{i-1}(x_0)} \leq g_{i-1}(x_0)$ for $1 \leq i \leq k^* < \tau_A, x_0 \geq y_m$. Note that for $x_0 \geq y_l$, $g_i(x_0) \geq y_l$ for $1 \leq i < \tau_A$. Hence,

$$\begin{aligned} b &\leq 2(1 + e^{\sigma g_{k^*-1}(x_0)})^{-k^*} \\ &\leq \frac{2}{(1 + e^{\sigma y_l})^{k^*}} \\ &< \delta/4. \end{aligned}$$

Therefore,

$$\begin{aligned} I &< \sum_{j=1}^{k^*} P_{x_0}(\tau_A = j)\delta/2 + \delta/4 \\ &\leq 3\delta/4. \end{aligned}$$

It remains to bound II. By the Mean Value Theorem, $\left|P_{z_0}(\tau_A = j) - P_{x_0}(\tau_A = j)\right| = P_c'(\tau_A = j)|x_0 - z_0|$ where $P_c'(\tau_A = j) = \frac{d}{dx}P_x(\tau_A = j)\big|_{x=c}$, and c is a value between x_0 and z_0. Now, for $x_0 > 0$,

$$P_{x_0}(\tau_A = j) = \left[(1 + e^{\sigma x_0})(1 + e^{\sigma g_1(x_0)}) \cdots (1 + e^{\sigma g_{j-2}(x_0)})(1 + e^{-\sigma g_{j-1}(x_0)})\right]^{-1},$$

from which it follows that

$$\begin{aligned} P_{x_0}'(\tau_A = j) &= -P_{x_0}^{-2}(\tau_A = j)\left\{ \sum_{i=0}^{j-2} \frac{d}{dx_0}\left(e^{\sigma g_i(x_0)}\right) \prod_{\substack{k=0 \\ k \neq i}}^{j-2} \left(1 + e^{\sigma g_k(x_0)}\right) \right. \\ &\quad \left. + \frac{d}{dx_0}\left(e^{-\sigma g_{j-1}(x_0)}\right) \prod_{k=0}^{j-2} \left(1 + e^{\sigma g_k(x_0)}\right) \right\}. \end{aligned}$$

But,

$$\begin{aligned} \frac{d}{dx_0}\left(e^{\sigma g_i(x_0)}\right) &= e^{\sigma g_i(x_0)}\sigma \frac{d}{dx_0}\left(g_i(x_0)\right) \\ &= \sigma e^{\sigma g_i(x_0)}\sigma/2 e^{(\sigma/2)g_{i-1}(x_0)} \frac{d}{dx_0}\left(g_{i-1}(x_0)\right) \\ &\vdots \\ &= \sigma(\sigma/2)^i e^{\sigma g_i(x_0) + (\sigma/2)g_{i-1}(x_0) + \ldots + (\sigma/2)g_1(x_0) + (\sigma/2)x_0}. \end{aligned}$$

Therefore,

$$P_{x_0}^{-2}(\tau_A = j)\frac{d}{dx_0}\left(e^{\sigma g_i(x_0)}\right) \prod_{\substack{k=0\\k\neq i}}^{j-2} \left(1+e^{\sigma g_k(x_0)}\right)$$

$$= \frac{d}{dx_0}\left(e^{\sigma g_i(x_0)}\right)\Big[(1+e^{\sigma x_0})(1+e^{\sigma g_1(x_0)})\cdots(1+e^{\sigma g_{i-1}(x_0)})$$
$$\cdot(1+e^{\sigma g_i(x_0)})^2(1+e^{\sigma g_{i+1}(x_0)})\cdots(1+e^{\sigma g_{j-2}(x_0)})(1+e^{-\sigma g_{j-1}(x_0)})\Big]^{-1}$$
$$= \sigma(\sigma/2)^i\Big[(e^{-(\sigma/2)x_0}+e^{(\sigma/2)x_0})\cdots(e^{-(\sigma/2)g_{i-1}(x_0)}+e^{(\sigma/2)g_{i-1}(x_0)})$$
$$\cdot(e^{-(\sigma/2)g_i(x_0)}+e^{(\sigma/2)g_i(x_0)})^2(1+e^{\sigma g_{i+1}(x_0)})\cdots(1+e^{\sigma g_{j-2}(x_0)})(1+e^{-\sigma g_{j-1}(x_0)})\Big]^{-1}$$
$$\leq \frac{\sigma(\sigma/2)^i}{2^i\cdot 4\cdot 2^{j-i-2}\cdot 1}$$
$$= \frac{\sigma(\sigma/2)^i}{2^j}.$$

and

$$P_{x_0}^{-2}(\tau_A = j)\left|\frac{d}{dx_0}\left(e^{-\sigma g_{j-1}(x_0)}\right)\right| \prod_{k=0}^{j-2}\left(1+e^{\sigma g_k(x_0)}\right)$$
$$= \left|\frac{d}{dx_0}\left(e^{-\sigma g_{j-1}(x_0)}\right)\right|\Big[(1+e^{\sigma x_0})(1+e^{\sigma g_1(x_0)})\cdots(1+e^{\sigma g_{j-2}(x_0)})(1+e^{-\sigma g_{j-1}(x_0)})^2\Big]^{-1}$$
$$= \sigma(\sigma/2)^{j-1}\Big[(e^{-(\sigma/2)x_0}+e^{(\sigma/2)x_0})\cdots(e^{-(\sigma/2)g_{j-2}(x_0)}+e^{(\sigma/2)g_{j-2}(x_0)})$$
$$\cdot(e^{(\sigma/2)g_{j-1}(x_0)}+e^{-(\sigma/2)g_{j-1}(x_0)})^2\Big]^{-1}$$
$$\leq \frac{\sigma(\sigma/2)^{j-1}}{2^{j-1}\cdot 4}$$
$$\leq \frac{\sigma(\sigma/2)^{j-1}}{2^j}.$$

It then follows that

$$\begin{aligned}\left|P_x'(\tau_A = j)\right| &\leq \frac{\sigma}{2^j}\sum_{i=0}^{j-2}(\sigma/2)^i + \frac{\sigma(\sigma/2)^{j-1}}{2^j}\\ &= \frac{\sigma}{2^j}\sum_{i=0}^{j-1}(\sigma/2)^i\\ &\leq \frac{\sigma}{2^{j-1}(2-\sigma)}.\end{aligned}$$

These results give us a bound for II:

$$\begin{aligned}
\text{II} &= \sum_{j=1}^{\infty} \left| P_{x_0}(\tau_A = j) - P_{z_0}(\tau_A = j) \right| \\
&= \sum_{j=1}^{\infty} \left| P_c'(\tau_A = j) \right| \left| x_0 - z_0 \right| \\
&\leq \sum_{j=1}^{\infty} \frac{\sigma}{2^{j-1}(2-\sigma)} \left| x_0 - z_0 \right| \\
&= \frac{2\sigma}{2-\sigma} \left| x_0 - z_0 \right| \\
&< \delta/4.
\end{aligned}$$

Therefore, the transition probabilities are equicontinuous for $|x_0|, |z_0| \leq y_u$.

case 4: $|x_0|, |z_0| \geq y_u$

Here, given $\delta > 0$, choose k^* satisfying $\frac{2}{(1+e^{\sigma y_u})^{k^*}} < \delta/4$. Next, select ϵ_1 such that $|f(x) - f(z)| < \delta/4$ for $|x - z| < \epsilon_1$ and ϵ_2 such that for $|x_0 - z_0| < \epsilon_2$, $\left| g_{k^*-1}(x_0) - g_{k^*-1}(z_0) \right| < \min\left(\frac{\delta}{2\sigma k^*}, \frac{\delta(2-\sigma y_l)}{2\sigma^2} \right)$, $\frac{2\sigma}{2-\sigma}|x_0 - z_0| < \delta/4$ and $\left| g_{k^*}(x_0) - g_{k^*}(z_0) \right| < \epsilon_1$.

For this case, note the following:

n1. $|x_{i+1}| = \begin{cases} e^{\sigma x_i/2}, & n_i = 0 \\ e^{-\sigma x_i/2}, & n_i = 1 \end{cases} \leq e^{\sigma |x_i|/2} \leq e^{\sigma g_i(x_0)/2} := g_{i+1}(x_0)$ (by c2).
and $g_{i+1}(x_0) \geq g_i(x_0)$ since $e^{\sigma x/2} \geq x$ for $x \geq y_u$. It follows that $g_i(x_0) \geq g_0(x_0) = x_0 \geq y_u$ for all $i \geq 0$.

n2. Since $\frac{d}{dx}(e^{\sigma x/2}) \geq 1$ for $x \geq y_u \geq y_m$, it follows from (n1) and the Mean Value Theorem that for c between $g_i(x_0)$ and $g_i(z_0)$,

$$\begin{aligned}
\left| g_{i+1}(x_0) - g_{i+1}(z_0) \right| &= \left| e^{\sigma g_i(x_0)/2} - e^{\sigma g_i(z_0)/2} \right| \\
&= (\sigma/2) e^{\sigma c/2} \left| g_i(x_0) - g_i(z_0) \right| \\
&\geq \left| g_i(x_0) - g_i(z_0) \right|.
\end{aligned}$$

n3. As shown in case 3,

$$\begin{aligned} |x_i - z_i| &\le \max_{0\le n\le k^*-1} \left|g_n(x_0) - g_n(z_0)\right| \\ &= \left|g_{k^*-1}(x_0) - g_{k^*-1}(z_0)\right| \text{ (by n2)}, \end{aligned}$$

for all $i \le k-1$, and

$$\begin{aligned} |x_k - z_k| &\le \max_{0\le n\le k^*} \left|g_n(x_0) - g_n(z_0)\right| \\ &= \left|g_{k^*}(x_0) - g_{k^*}(z_0)\right| \text{ (by n2)}. \end{aligned}$$

Thus, for $k \le k^*$, it follows from (3.10), (n2) and (n3) above that

$$\begin{aligned} \left|P^k_{x_0} f - P^k_{z_0} f\right| &\le (\sigma/2)k^*\left|g_{k^*-1}(x_0) - g_{k^*-1}(z_0)\right| + \delta/4 \\ &< \delta/2, \end{aligned}$$

since $\left|g_{k^*}(x_0) - g_{k^*}(z_0)\right| < \epsilon_1$ and $(\sigma/2)k^*\left|g_{k^*-1}(x_0) - g_{k^*-1}(z_0)\right| < \delta/4$.

For $k > k^*$, the same approach will be used as in case 3. The expression $\left|P^k_{x_0} f - P^k_{z_0} f\right|$ is broken into two pieces which we will call I and II (see (3.16)). I is then further divided into a and b (see (3.17)). Here, similar to case 3, we find

$$\begin{aligned} |x_j - z_j| &\le (\sigma/2)\, |g_{j-1}(x_0) - g_{j-1}(z_0)| \\ &\le (\sigma/2)\, \left|g_{k^*-1}(x_0) - g_{k^*-1}(z_0)\right|, \end{aligned}$$

by (n3) since $j \le k^*$. Thus, from (3.14),

$$\begin{aligned} a &= \left|P^{k-j}_{x_j} f - P^{k-j}_{z_j} f\right| \\ &\le \frac{\sigma|x_j - z_j|}{2-\sigma y_l} + \sum_{n_0,\ldots,n_{k-1}=0}^{1} p(n_0 \,|\, x_0)\cdots p(n_{k-1} \,|\, x_{k-1})\, \left|f(x_k) - f(z_k)\right| \\ &\le \frac{(\sigma^2/2)|g_{k^*-1}(x_0) - g_{k^*-1}(z_0)|}{2-\sigma y_l} \\ &\quad + \sum_{n_0,\ldots,n_{k-1}=0}^{1} p(n_0 \,|\, x_0)\cdots p(n_{k-1} \,|\, x_{k-1})\, \left|f(x_k) - f(z_k)\right| \\ &< \delta/4 + \delta/4 = \delta/2, \end{aligned}$$

since $\frac{(\sigma^2/2)|g_{k^*-1}(x_0)-g_{k^*-1}(z_0)|}{2-\sigma y_l} < \delta/4$ and $|x_k - z_k| \le |x_j - z_j| \le |g_{k^*}(x_0) - g_{k^*}(z_0)| < \delta/4$.

Recall that $b := 2P_{x_0}(\tau_A > k^*)$ and $P_{x_0}(\tau_A > k^*) = \left[(1+e^{\sigma x_0})(1+e^{\sigma g_1(x_0)})\cdots(1+e^{\sigma g_{k^*-1}(x_0)})\right]^{-1}$. By (n1) it follows that $b \le 2(1+e^{\sigma x_0})^{-k^*} \le \frac{2}{(1+e^{\sigma y_u})^{k^*}} < \delta/4$. Thus,

$$\mathrm{I} < \sum_{j=1}^{k^*} P_{x_0}(\tau_A = j)\delta/2 + \delta/4 < 3\delta/4.$$

The same proof used to bound II as given in case 3 applies to this case. Therefore, for $k \ge k^*$, $|P_{x_0}^k - P_{z_0}^k| < \delta$ for $|x_0 - z_0| < \epsilon_2, |x_0|, |z_0| \ge y_u$.

Since we have established equicontinuity of the transition probabilities for all four cases, it follows that the process $\{U_t\}$ defined by (3.1) and (3.2) is an e-chain.

3.2.2 Reachable Point

Proposition 3.2.2 *U_t defined by (3.1) and (3.2) has two reachable points, x^* and x_*, where x^* and x_* are defined as the unique solutions to the following equations, respectively:*

$$x = e^{-\sigma x/2} \tag{3.18}$$

$$x = -e^{\sigma x/2} \tag{3.19}$$

Proof:

Only the proof for x^* will be given, the proof for x_* being similar. First we will show that the n^{th} iterate of $g(x_0) := e^{-\sigma x_0/2}$ converges to x^* by showing that the sequences $\{x_{2n}\}$ and $\{x_{2n+1}\}$ converge to x^*, where x_n is defined to be the n^{th} iterate of $g(x_0)$. Without loss of generality, we will assume $x_0 < x^*$. It is easily seen that under this assumption, $x_{2n} < x^* < x_{2n+1}$ for all n.

Define $h(x) := -\frac{2}{\sigma}\ln x$ which is the inverse of $g(\cdot)$. Since $g(\cdot)$ is strictly decreasing, we have the following:

1. for $x < x^*, h(x) > g(x)$

2. for $x > x^*, h(x) < g(x)$

3. if $g(x) = h(y)$ then

   (a) $x < y$ if $x < x^*$,

   (b) $x > y$ if $x > x^*$,

   (c) $x = y$ if $x = x^*$.

Combining these results we obtain $x_{2n} < x_{2n+2}$ and $x_{2n+1} < x_{2n-1}$ for all n since $h(x_{2n+2}) = x_{2n+1} = g(x_{2n})$ and $h(x_{2n+1}) = x_{2n} = g(x_{2n-1})$. Figure 3.2 depicts the functions $g(x)$ and $h(x)$ when $\sigma = 1$. The above results are seen more easily upon examination of the graph. To show that $\{x_{2n}\}$ and $\{x_{2n+1}\}$ converge to x^*, suppose $\lim_{n\to\infty} x_{2n} = b$ and $\lim_{n\to\infty} x_{2n+1} = a$. Now $a = g(b) = h(b)$ and $b = g(a) = h(a)$. From 3c, this implies that $a = b = x^*$. Thus, $\{x_{2n}\}$ and $\{x_{2n+1}\}$ converge to x^* from which it follows that $\{x_n\}$ converges to x^*.

Now, let O be an open set containing x^*. Since $\{x_n\}$ converges to x^*, we may choose M so that $x_M \in O$. Then

$$\begin{aligned} P^M(y, O) &= P(U_M \in O | U_0 = y) \\ &\geq P(U_0 = y, U_1 = x_1, U_2 = x_2, \ldots, U_M = y_m) \\ &= P(U_0 = y, N_1 = 1, N_2 = 1, \ldots, N_M = 1) \\ &> 0 \end{aligned}$$

Therefore, $\sum_n P^n(y, O) > 0$; hence, x^* is reachable.

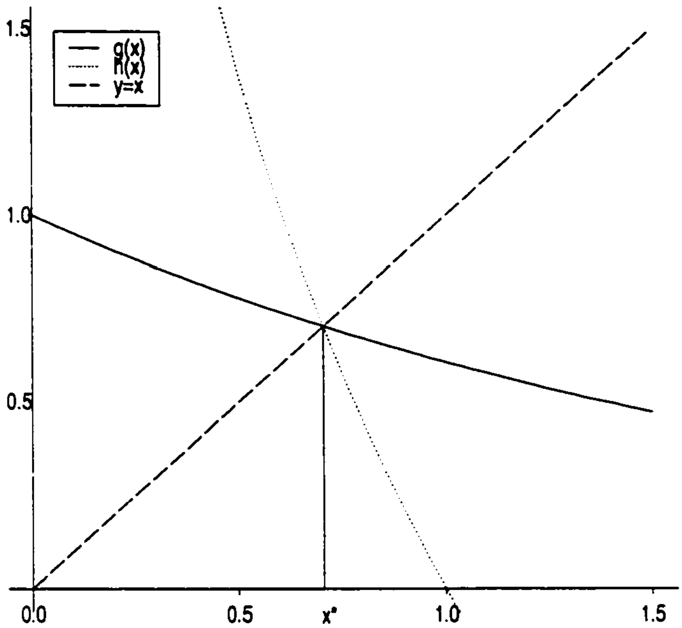

Figure 3.2: Reachable point: Bernoulli Process. $g(x) = e^{-x/2}, h(x) = -2\ln x$

3.2.3 Bounded in Probability on Average

Here we will use results from Glynn and Meyn (1997) to establish that the process $\{U_t\}$ defined by (3.1) and (3.2) is bounded in probability on average. The applicable result is reproduced as Theorem A.2.15 in the Appendix. As shown in Section 2.3, a stronger result actually holds; namely that $\{U_t\}$ is bounded in probability uniformly in x. The proof of this result is similar to the proof given in Section 2.3 for the GLARMA models of Section 2.2 and is not reproduced here.

Theorem 3.2.3 *The chain $\{U_t\}$ is bounded in probability on average.*

Proof:

The conditions of Theorem A.2.15 are shown to be satisfied.

Condition 1: $\Delta V(x) := EV(x) - V(x) \leq -1 + bI_A(x)$.

To see that this condition is met, define

$$V(x) = |x|, \tag{3.20}$$

$$b = 2, \text{ and} \tag{3.21}$$

$$A = [-2, 2]. \tag{3.22}$$

Then,

$$\begin{aligned}
\Delta V(x) &= E\big[|U_t| \mid U_{t-1} = x\big] - |x| \\
&= E\big[U_t I_{[U_t \geq 0]} \mid U_{t-1} = x\big] + E\big[-U_t I_{[U_t<0]} \mid U_{t-1} = x\big] - |x| \\
&= E\left[\left(N_t - \frac{e^{\sigma x}}{1+e^{\sigma x}}\right)(1+e^{\sigma x})e^{-\sigma x/2} I_{[N_t=1]} \mid U_{t-1} = x\right] \\
&\quad + E\left[\left(\frac{e^{\sigma x}}{1+e^{\sigma x}} - N_t\right)(1+e^{\sigma x})e^{-\sigma x/2} I_{[N_t=0]} \mid U_{t-1} = x\right] - |x| \\
&= \frac{2e^{\sigma x/2}}{1+e^{\sigma x}} - |x| \\
&\leq 1 - |x|.
\end{aligned}$$

Therefore, $\Delta V(x) \leq -1 + bI_A(x)$.

Condition 2: $\lim_{n\to\infty} \sup_{a\in A} E_a[V(U_n) I_{[\tau_A > n]}] = 0$, where A is as defined in (3.22) and $V(\cdot)$ is as defined in (3.20).

By Cauchy-Schwartz, we have

$$\lim_{n\to\infty} \sup_{a\in A} E[|U_n| I_{[\tau_A>n]} \mid U_0 = a] \leq \lim_{n\to\infty} \sup_{a\in A} E^{1/2}[|U_n|^2 \mid U_0 = a] P^{1/2}(\tau_A > n).$$

Since $E(U_n \mid U_{n-1}) = 0$ and $\mathrm{Var}(U_n \mid U_{n-1}) = 1$, it follows that

$$E[|U_n|^2 \mid U_0 = a] = 1.$$

By Markov's inequality,

$$P_a(\tau_A > n) \leq \frac{E_a(\tau_A)}{n+1},$$

and by Theorem A.2.16,

$$E_a(\tau_A) \leq V(a) + b \text{ for } a \in A.$$

Combining these results we obtain:

$$\begin{aligned}
\lim_{n\to\infty} \sup_{a\in A} E_a[V(U_n)I_{[\tau_A>n]}] &\le \lim_{n\to\infty} \sup_{a\in A} \left(\frac{E_a(\tau_A)}{n+1}\right)^{1/2} \\
&\le \lim_{n\to\infty} \sup_{a\in A} \left(\frac{V(a)+b}{n+1}\right)^{1/2} \\
&\le \lim_{n\to\infty} \left(\frac{2b}{n+1}\right)^{1/2} \\
&= 0.
\end{aligned}$$

Condition 3: The family of probability measures $\{1/m \sum_{k=1}^{m} P^k(a,\cdot) : a \in A\}$ is tight for each $m \ge 1$.

Since the chain is weak Feller (Definition A.2.8), Condition 3 holds by Theorem A.2.14. Thus, the chain is bounded in probability on average.

It has been shown that the process $\{U_t\}$ is an e-chain, possesses a reachable point and is bounded in probability on average. Applying Theorem A.2.18, we conclude that the process has a unique stationary distribution. □

Chapter 4

INFERENCE FOR GLARMA MODELS

We return to the GLARMA models discussed in Chapter 2 for analyzing time series of Poisson counts:

$$p(Y_t = k \mid W_t) = \frac{e^{-\mu_t}\mu_t^k}{k!} I_{\{0,1,\ldots\}}(k)$$

where

$$\log(\mu_t) = W_t = \mathbf{x}_t^T\boldsymbol{\beta} + \sum_{i=1}^{\infty} \tau_i e_{t-i}, \tag{4.1}$$

with

$$e_t = \frac{Y_t - \mu_t}{\mu_t^\lambda}, \quad \lambda \geq 0 \tag{4.2}$$

and

$$\begin{aligned} \sum_{i=1}^{\infty} \tau_i &= \Big(1 - \sum_{i=1}^{p} \phi_i z^i\Big)^{-1}\Big(1 + \sum_{i=1}^{q} \theta_i z^i\Big) - 1 \qquad (4.3) \\ &= \phi(z)^{-1}\theta(z) - 1. \end{aligned}$$

In this chapter, we calculate the maximum likelihood estimates for the above model parameters and derive the asymptotic theory for these estimates under specific conditions. We also consider simulations to determine how well these asymptotic properties apply. Lastly, we fit this model to a data set of asthma counts.

4.1 Likelihood Calculations

Here we obtain formulas for the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma} = (\boldsymbol{\phi}^T, \boldsymbol{\theta}^T)^T$, the parameters of the model given above. Let $\delta = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ and define

$L_t(\delta) = \log f(y_t|\mathcal{F}_{t-1})$. The log-likelihood can then be written as $\sum_{t=1}^{n} L_t(\delta)$ which, upon ignoring terms which do not involve the parameters, becomes

$$L(\delta) = \sum_{t=1}^{n} \left(Y_t W_t(\delta) - e^{W_t(\delta)}\right),$$

where

$$\log(\mu_t) = W_t(\delta) = \mathbf{x}_t^T \boldsymbol{\beta} + \sum_{i=1}^{\infty} \tau_i(\boldsymbol{\gamma}) e_{t-i}(\boldsymbol{\delta})$$

and

$$e_t = (Y_t - \mu_t)/\mu_t^{\lambda}.$$

First and second derivatives are given by the following expressions

$$\frac{\partial L}{\partial \delta} = \sum_{t=1}^{n} (Y_t - \mu_t) \frac{\partial W_t}{\partial \delta} = \sum_{t=1}^{n} e_t \mu_t^{\lambda} \frac{\partial W_t}{\partial \delta}$$

and

$$\begin{aligned} \frac{\partial^2 L}{\partial \delta \partial \delta^T} &= \sum_{t=1}^{n} \left[(Y_t - \mu_t) \frac{\partial^2 W_t}{\partial \delta \partial \delta^T} - \mu_t \frac{\partial W_t}{\partial \delta} \frac{\partial W_t}{\partial \delta^T} \right] \\ &= \sum_{t=1}^{n} \left[e_t \mu_t^{\lambda} \frac{\partial^2 W_t}{\partial \delta \partial \delta^T} - \mu_t \frac{\partial W_t}{\partial \delta} \frac{\partial W_t}{\partial \delta^T} \right]. \end{aligned}$$

In order to calculate these, the following expressions are required. First note that

$$\frac{\partial e_t}{\partial \delta} = -[e^{(1-\lambda)W_t} + \lambda e_t] \frac{\partial W_t}{\partial \delta}.$$

Also

$$\frac{\partial W_t}{\partial \delta} = \frac{\partial \beta^T}{\partial \delta} x_t + \frac{\partial Z_t}{\partial \delta},$$

where

$$\begin{aligned} Z_t &= \sum_{i=1}^{\infty} \tau_i e_{t-i} \\ &= (\phi(B)^{-1} \theta(B) - 1) e_t, \end{aligned}$$

so that

$$Z_t = \sum_{i=1}^{p} \phi_i (Z_{t-i} + e_{t-i}) + \sum_{i=1}^{q} \theta_i e_{t-i}.$$

It follows that

$$\begin{aligned}\frac{\partial Z_t}{\partial \delta} &= \sum_{i=1}^{p} \frac{\partial \phi_i}{\partial \delta}(Z_{t-i} + e_{t-i}) + \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \delta} + \frac{\partial e_{t-i}}{\partial \delta} \right) \\ &+ \sum_{i=1}^{q} \frac{\partial \theta_i}{\partial \delta} e_{t-i} + \sum_{i=1}^{q} \theta_i \frac{\partial e_{t-i}}{\partial \delta}.\end{aligned}$$

In particular:

$$\frac{\partial Z_t}{\partial \beta_a} = \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \beta_a} + \frac{\partial e_{t-i}}{\partial \beta_a} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial e_{t-i}}{\partial \beta_a},$$

$$\frac{\partial Z_t}{\partial \phi_a} = Z_{t-a} + e_{t-a} + \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \phi_a} + \frac{\partial e_{t-i}}{\partial \phi_a} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial e_{t-i}}{\partial \phi_a}$$

and

$$\frac{\partial Z_t}{\partial \theta_a} = \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \theta_a} + \frac{\partial e_{t-i}}{\partial \theta_a} \right) + e_{t-a} + \sum_{i=1}^{q} \theta_i \frac{\partial e_{t-i}}{\partial \theta_a}.$$

The second derivatives are then

$$\begin{aligned}\frac{\partial^2 e_t}{\partial \delta \partial \delta^T} &= -[e^{(1-\lambda)W_t} + \lambda e_t] \frac{\partial^2 W_t}{\partial \delta \partial \delta^T} \\ &- \left[\frac{\partial W_t}{\partial \delta}(1-\lambda)e^{(1-\lambda)W_t} + \lambda \frac{\partial e_t}{\partial \delta} \right] \frac{\partial W_t}{\partial \delta^T}\end{aligned}$$

and

$$\frac{\partial^2 W_t}{\partial \delta \partial \delta^T} = \frac{\partial^2 \beta^T}{\partial \delta \partial \delta^T} x_t + \frac{\partial \delta^2 Z_t}{\partial \delta \partial \delta^T} = \frac{\partial \delta^2 Z_t}{\partial \delta \partial \delta^T},$$

in which

$$\begin{aligned}\frac{\partial^2 Z_t}{\partial \delta \partial \delta^T} &= \sum_{i=1}^{p} \left[\frac{\partial \phi_i}{\partial \delta} \left(\frac{\partial Z_{t-i}}{\partial \delta^T} + \frac{\partial e_{t-i}}{\partial \delta^T} \right) + \left(\frac{\partial Z_{t-i}}{\partial \delta} + \frac{\partial e_{t-i}}{\partial \delta} \right) \frac{\partial \phi_i}{\partial \delta^T} \right] \\ &+ \sum_{i=1}^{p} \phi_i \left(\frac{\partial^2 Z_{t-i}}{\partial \delta \partial \delta^T} + \frac{\partial^2 e_{t-i}}{\partial \delta \partial \delta^T} \right) + \sum_{i=1}^{q} \left[\frac{\partial \theta_i}{\partial \delta} \frac{\partial e_{t-i}}{\partial \delta^T} + \frac{\partial e_{t-i}}{\partial \delta} \frac{\partial \theta_i}{\partial \delta^T} \right] \\ &+ \sum_{i=1}^{q} \theta_i \frac{\partial^2 e_{t-i}}{\partial \delta \partial \delta^T}.\end{aligned}$$

Asymptotic results for these estimates are given in the next section for the first order Markov process presented in Section 2.2.1.1:

$$W_t \;=\; \beta - \gamma + \gamma Y_{t-1} e^{-W_{t-1}}.$$

We consider the following two methods of estimating the standard errors of the estimates. Let

$$\hat{\Omega}^{(1)} = -\left(\frac{\partial^2 L(\hat{\theta})}{\partial\delta\partial\delta^T}\right)^{-1}$$

and

$$\hat{\Omega}^{(2)} = \left(\sum_{t=1}^{n} \frac{\partial L_t(\hat{\theta})}{\partial\delta}\frac{\partial L_t(\hat{\theta})}{\partial\delta^T}\right)^{-1}$$

where

$$\frac{\partial L_t(\hat{\delta})}{\partial\delta} = e_t\mu_t^{\lambda}\frac{\partial W_t}{\partial\delta}.$$

The standard errors are then defined as

$$\hat{\sigma}^{(1)}_{\hat{\delta},j} = \sqrt{\Omega^{(1)}_{jj}} \tag{4.4}$$

and

$$\hat{\sigma}^{(2)}_{\hat{\delta},j} = \sqrt{\Omega^{(2)}_{jj}}. \tag{4.5}$$

4.2 Asymptotic Theory

In this section we establish asymptotic properties of the MLEs derived in the previous section for the process defined by (4.1)-(4.3) with $\lambda = 1$. $p = 0$, $q = 1$ and $\mathbf{x}_t^T\boldsymbol{\beta} = \beta$. Uniform ergodicity (as established in Theorem 2.2.1) and stationarity of $\{W_t\}$ are the key ingredients of the argument.

First replace $W_t(\delta)$ by

$$W_t^{\dagger}(\delta) = W_t(\delta_0) + (\delta - \delta_0)^T \dot{W}_t,$$

where $\dot{W}_t = \frac{\partial W_t(\delta_0)}{\partial \delta}$ and define a linearized form of the likelihood as

$$L^\dagger(\delta) = \sum_{t=1}^{n} \left(Y_t W_t^\dagger(\delta) - e^{W_t^\dagger(\delta)} \right).$$

Unless otherwise indicated, W_t and $\dot{W}_t$ are evaluated at δ_0. Now, reparameterizing with the transformation $u = n^{1/2}(\delta - \delta_0)$, we have

$$\begin{aligned} R_n^\dagger(u) := {} & L^\dagger(\delta_0) - L^\dagger(\delta_0 + un^{-1/2}) \\ = {} & - u^T n^{-1/2} \sum_{t=1}^{n} Y_t \dot{W}_t + \sum_{t=1}^{n} e^{W_t} \left(e^{u^T n^{-1/2} \dot{W}_t} - 1 \right) \\ = {} & - u^T n^{-1/2} \sum_{t=1}^{n} \left(Y_t - e^{W_t} \right) \dot{W}_t + \sum_{t=1}^{n} e^{W_t} \left(e^{u^T n^{-1/2} \dot{W}_t} - 1 - u^T n^{-1/2} \dot{W}_t \right). \quad (4.6) \end{aligned}$$

Note that $R_n^\dagger(u)$ is a convex function of u. The first term in (4.6) can be written as $-u^T H_n$ where

$$H_n := n^{-1/2} \sum_{t=1}^{n} e_t e^{W_t} \dot{W}_t$$

and $e_t = \left(Y_t - e^{W_t} \right) / e^{W_t}$. Now this is a sum of a triangular array of vector martingale differences

$$\eta_{nt} = n^{-1/2} e_t b_t,$$

where

$$b_t = \dot{W}_t e^{W_t} = \dot{W}_t \mu_t.$$

In order to apply a martingale central limit theorem, it suffices to show (see Corollary 3.1 of Hall and Heyde (1980)) that

$$\sum_{t=1}^{n} E\left(\eta_{nt}\eta_{nt}^T \mid \mathcal{F}_{t-1}\right) \xrightarrow{P} V(\delta_0), \quad (4.7)$$

where $\mathcal{F}_t = \sigma(Y_s, s \le t)$, and, for all $\epsilon > 0$,

$$\sum_{t=1}^{n} E\left(\eta_{nt}\eta_{nt}^T I\left(|\eta_{nt}| > \epsilon\right) \mid \mathcal{F}_{t-1}\right) \xrightarrow{P} 0. \quad (4.8)$$

We then have

$$H_n \xrightarrow{d} N(0, V),$$

where

$$V = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial L_t(\delta_0)}{\partial \delta} \frac{\partial L_t(\delta_0)}{\partial \delta^T} = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} e_t^2 e^{2W_t} \dot{W}_t \dot{W}_t^T.$$

The second term in (4.6) is

$$u^T \left[(2n)^{-1} \sum_{t=1}^{n} e^{W_t} \dot{W}_t \dot{W}_t^T \right] u + O_p\left(n^{-3/2} \sum_{t=1}^{n} e^{W_t} (u^T \dot{W}_t)^3 \right)$$

in which the second term converges to zero. Hence

$$R_n^{\dagger}(u) \xrightarrow{d} R(u) := -u^T N(0, V) + u^T V u/2,$$

where

$$V = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} e^{W_t} \dot{W}_t \dot{W}_t^T.$$

It then follows that $\hat{u}_n^{\dagger}$ =argmin$R_n^{\dagger}(u) \xrightarrow{d} \hat{u}$ =argmin$R(u)$. From the form of $R(u)$, we see that $\hat{u} = V^{-1}N(0, V) \sim N(0, V^{-1})$.

Next, we pass the convergence of $R_n^{\dagger}(u)$ onto $R_n(u) := L(\delta_0) - L(un^{-1/2} + \delta_0)$. Specifically, it suffices to show that $L(un^{-1/2} + \delta_0) - L^{\dagger}(un^{-1/2} + \delta_0) \xrightarrow{P} 0$ uniformly for $|u| \leq K$. Writing $\delta = un^{1/2} + \delta_0$, we have

$$\begin{aligned}
&L(\delta) - L^{\dagger}(\delta) \\
&= \sum_{t=1}^{n} Y_t W_t(\delta) - \sum_{t=1}^{n} e^{W_t(\delta)} - \sum_{t=1}^{n} Y_t \left(W_t + u^T n^{-1/2} \dot{W}_t \right) + \sum_{t=1}^{n} e^{W_t + u^T n^{-1/2} \dot{W}_t} \\
&= \sum_{t=1}^{n} Y_t \left(W_t(\delta) - W_t - u^T n^{-1/2} \dot{W}_t \right) - \sum_{t=1}^{n} \left(e^{W_t(\delta)} - e^{W_t + u^T n^{-1/2} \dot{W}_t} \right) \\
&= \sum_{t=1}^{n} (Y_t - e^{W_t}) \left(W_t(\delta) - W_t - u^T n^{-1/2} \dot{W}_t \right) \\
&\quad - \sum_{t=1}^{n} \left[e^{W_t(\delta)} - e^{W_t + u^T n^{-1/2} \dot{W}_t} - e^{W_t} \left(W_t(\delta) - W_t - u^T n^{-1/2} \dot{W}_t \right) \right]. \qquad (4.9)
\end{aligned}$$

The first term in equation (4.9) is

$$\begin{aligned} A_n &= \sum_{t=1}^{n} (Y_t - e^{W_t}) \left(W_t(\delta) - W_t - u^T n^{-1/2} \dot{W}_t \right) \\ &= u^T (2n)^{-1} \left[\sum_{t=1}^{n} (Y_t - e^{W_t}) \ddot{W}_t + \sum_{t=1}^{n} (Y_t - e^{W_t}) \left(\ddot{W}_t(\delta^*) - \ddot{W}_t \right) \right] u. \end{aligned}$$

Since $(Y_t - e^{W_t})\ddot{W}_t$ is stationary and $E[(Y_t - e^{W_t})\ddot{W}_t] = 0$, $A_n \to 0$ uniformly for $|u| \le K$ and for all $K < \infty$, where $\|\delta^* - \delta_0\| \le \|\delta - \delta_0\|$ assuming $\ddot{W}_t(\delta^*) - \ddot{W}_t \xrightarrow{P} 0$. The second term is

$$B_n = -\sum_{t=1}^{n} \left[e^{W_t(\delta)} - e^{W_t + u^T n^{-1/2} \dot{W}_t} - e^{W_t} \left(W_t(\delta) - W_t - u^T n^{-1/2} \dot{W}_t \right) \right].$$

which after expanding $e^{W_t(\delta)}, e^{u^T n^{-1/2} \dot{W}_t}$ and $W_t(\delta)$ in a Taylor series is

$$\begin{aligned} &= -\sum_{t=1}^{n} \Big[e^{W_t} + u^T n^{-1/2} e^{W_t} \dot{W}_t + u^T (2n)^{-1} e^{W_t(\delta_1^*)} \left(\dot{W}_t^2(\delta_1^*) + \ddot{W}_t(\delta_1^*) \right) u \\ &\quad - e^{W_t} \left(1 + u^T n^{-1/2} \dot{W}_t + e^c (2n)^{-1} u^T \dot{W}_t^2 u \right) \\ &\quad - e^{W_t} \left(W_t + u^T n^{-1/2} \dot{W}_t + u^T (2n)^{-1} \ddot{W}_t(\delta_2^*) u - W_t - u^T n^{-1/2} \dot{W}_t \right) \Big] \\ &= -u^T (2n)^{-1} \Bigg\{ \sum_{t=1}^{n} \left(e^{W_t(\delta_1^*)} - e^{W_t} \right) \left(\dot{W}_t^2(\delta_1^*) + \ddot{W}_t(\delta_1^*) \right) \\ &\quad + \sum_{t=1}^{n} e^{W_t} \left[\left(\dot{W}_t^2(\delta_1^*) - e^c \dot{W}_t^2 \right) + \left(\ddot{W}_t(\delta_1^*) - \ddot{W}_t(\delta_2^*) \right) \right] \Bigg\} u. \end{aligned}$$

where $0 \le c \le \frac{u^T}{2n} \dot{W}_t(\delta_0)$ and $\left\|\delta_j^* - \delta_0\right\| \le \|\delta - \delta_0\|$ for $j = 1, 2$. Assuming each average in the above expression converges to a finite quantity in probability, we have that $B_n \to 0$ uniformly on compact subsets for u. Therefore, $L(\delta) - L^\dagger(\delta) \xrightarrow{P} 0$ uniformly for $|u| \le K$, for all $K < \infty$ and we obtain the desired result:

$$R_n(u) \xrightarrow{d} R(u) := -u^T N(0, V) + u^T V u/2.$$

We now consider establishing conditions (4.7) and (4.8). From (4.4) we see that

$$\begin{aligned} \dot{W}_t &= \begin{bmatrix} \frac{\partial W_t}{\partial \gamma} \\ \frac{\partial W_t}{\partial \beta} \end{bmatrix} = \begin{bmatrix} \dot{W}_{t,1} \\ \dot{W}_{t,2} \end{bmatrix} \\ &= \begin{bmatrix} Y_{t-1} e^{-W_{t-1}} - 1 - \gamma Y_{t-1} e^{-W_{t-1}} \dot{W}_{t-1,1} \\ 1 - \gamma Y_{t-1} e^{-W_{t-1}} \dot{W}_{t-1,2} \end{bmatrix} \\ &= \begin{bmatrix} U_t + A_t \dot{W}_{t-1,1} \\ 1 + A_t \dot{W}_{t-1,2} \end{bmatrix} = \begin{bmatrix} U_t + \sum_{i=1}^{\infty} A_t \cdots A_{t-i+1} U_{t-i} \\ 1 + \sum_{i=1}^{\infty} A_t \cdots A_{t-i+1} \end{bmatrix}, \end{aligned} \tag{4.10}$$

where $U_t = Y_{t-1}e^{-W_{t-1}} - 1$ and $A_t = -\gamma Y_{t-1}e^{-W_{t-1}}$. Since $\dot{W}_t$ is a function of $\{W_s, s \le t\}$, it also is a strictly stationary ergodic process. Now,

$$\sum_{t=1}^{n} E(\eta_{nt}\eta_{nt}^T \mid \mathcal{F}_{t-1}) = \frac{1}{n}\sum_{t=1}^{n} e^{W_t}\dot{W}_t\dot{W}_t^T,$$

which is a function of two stationary ergodic processes, $\{W_t\}$ and $\{\dot{W}_t\}$. By the ergodic theorem we then have

$$\frac{1}{n}\sum_{t=1}^{n} e^{W_t}\dot{W}_t\dot{W}_t^T \xrightarrow{a.s.} V = E(e^{W_1}\dot{W}_1\dot{W}_1^T)$$

if $E|e^{W_t(\delta_0)}\dot{W}_t\dot{W}_t^T| < \infty$. Conditions under which this holds will now be derived for a particular choice of parameter values of β and γ. It suffices to show $E|e^{W_t}\dot{W}_{t,i}^2| < \infty$, $i = 1, 2$. First we will consider the case $i = 1$. Using $\|\cdot\|_2$ to denote the L_2 norm, we have from (4.10),

$$\|e^{W_t/2}\dot{W}_{t,1}\|_2 \le \|e^{W_t/2}U_t\|_2 + \sum_{i=1}^{\infty} \|e^{W_t/2}A_t\cdots A_{t-i+1}U_{t-i}\|_2.$$

Using properties of the moment generating function for a Poisson distributed random variable and the fact that the process W_t is bounded below by $\beta - \gamma$, we have

$$\begin{aligned}
\|e^{W_t/2}U_t\|_2^2 &= E\left[e^{\beta-\gamma}e^{\gamma Y_{t-1}e^{-W_{t-1}}}(Y_{t-1}e^{-W_{t-1}} - 1)^2\right] \\
&= e^{\beta-\gamma}E\left[E\left(e^{\gamma Y_{t-1}e^{-W_{t-1}}}(Y_{t-1}^2e^{-2W_{t-1}} - 2Y_{t-1}e^{-W_{t-1}} + 1) \mid W_{t-1}\right)\right] \\
&= e^{\beta-\gamma}E\Big[e^{-W_{t-1}}e^{\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)} + e^{2\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)} \\
&\qquad -2e^{\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)} + e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)}\Big] \\
&= e^{\beta-\gamma}E\left[e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)}\left[1 + e^{\gamma e^{-W_{t-1}}}\left(e^{\gamma e^{-W_{t-1}}} + e^{-W_{t-1}} - 2\right)\right]\right] \\
&\le e^{\beta-\gamma}e^{e^{\beta-\gamma}(e^{\gamma e^{-(\beta-\gamma)}}-1)}\left[1 + e^{\gamma e^{-(\beta-\gamma)}}\left(e^{\gamma e^{-(\beta-\gamma)}} + e^{-(\beta-\gamma)} - 2\right)\right] := c_1^2,
\end{aligned}$$

$$\begin{aligned}
&E\left[e^{W_t}A_t^2 \mid \mathcal{F}_{t-1}\right] \\
&\quad = E\left[\gamma^2e^{\beta-\gamma}Y_{t-1}^2e^{-2W_{t-1}}e^{\gamma Y_{t-1}e^{-W_{t-1}}} \mid W_{t-1}\right] \\
&\quad = \gamma^2e^{\beta-\gamma}\left[e^{-W_{t-1}}e^{\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)} + e^{2\gamma e^{-W_{t-1}}}e^{e^{W_{t-1}}(e^{\gamma e^{-W_{t-1}}}-1)}\right] \\
&\quad \le \gamma^2e^{\beta-\gamma}e^{e^{\beta-\gamma}(e^{\gamma e^{-(\beta-\gamma)}}-1)}e^{\gamma e^{-(\beta-\gamma)}}\left(e^{\gamma e^{-(\beta-\gamma)}} + e^{-(\beta-\gamma)}\right) := b_1^2.
\end{aligned}$$

$$
\begin{aligned}
E\left[A_t^2 \mid \mathcal{F}_{t-1}\right] &= E\left[\gamma^2 Y_{t-1}^2 e^{-2W_{t-1}} \mid W_{t-1}\right] \\
&= \gamma^2\left(1+e^{-W_{t-1}}\right) \\
&\leq \gamma^2(1+e^{-(\beta-\gamma)})
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[U_t^2 \mid \mathcal{F}_{t-1}\right] &= E\left[Y_{t-1}^2 e^{-2W_{t-1}} - 2Y_{t-1}e^{-W_{t-1}} + 1 \mid W_{t-1}\right] \\
&= e^{-W_{t-1}} \\
&\leq e^{-(\beta-\gamma)} := b_2^2.
\end{aligned}
$$

Applying these results. $\|e^{W_t/2}A_t \cdots A_{t-i+1}U_{t-i}\|_2^2$ may be calculated recursively:

$$
\begin{aligned}
\|e^{W_t/2}A_t \cdots A_{t-i+1}U_{t-i}\|_2^2 &= E\left(e^{W_t}A_t^2 \cdots A_{t-i+1}^2 U_{t-i}^2\right) \\
&= E\left[E\left(e^{W_t}A_t^2 \cdots A_{t-i+1}^2 U_{t-i}^2 \mid \mathcal{F}_{t-1}\right)\right] \\
&= E\left[A_{t-1}^2 \cdots A_{t-i+1}^2 U_{t-i}^2 E\left(e^{W_t}A_t^2 \mid \mathcal{F}_{t-1}\right)\right] \\
&\leq b_1^2 E\left[E\left(A_{t-1}^2 \cdots A_{t-i+1}^2 U_{t-i}^2 \mid \mathcal{F}_{t-2}\right)\right] \\
&= b_1^2 E\left[A_{t-2}^2 \cdots A_{t-i+1}^2 U_{t-i}^2 E\left(A_{t-1}^2 \mid \mathcal{F}_{t-2}\right)\right] \\
&\leq b_1^2\gamma^2(1+e^{-(\beta-\gamma)})E\left[A_{t-2}^2 \cdots A_{t-i+1}^2 U_{t-i}^2\right] \\
&\vdots \\
&\leq b_1^2\left(\gamma^2(1+e^{-(\beta-\gamma)})\right)^{i-1} E\left[E\left(U_{t-i} \mid \mathcal{F}_{t-i-1}\right)\right] \\
&\leq b_1^2 b_2^2\left(\gamma^2(1+e^{-(\beta-\gamma)})\right)^{i-1}.
\end{aligned}
$$

Therefore,

$$
\|e^{W_t/2}\dot{W}_{t,1}\|_2 \leq c_1 + c_2\sum_{i=1}^{\infty}\gamma^{i-1}(1+e^{\gamma-\beta})^{(i-1)/2},
$$

where $c_2 = b_1 b_2$. Likewise,

$$
\begin{aligned}
\|e^{W_t/2}\dot{W}_{t,2}\|_2 &\leq \|e^{W_t/2}\|_2 + \sum_{i=0}^{\infty}\|e^{W_t/2}A_t \cdots A_{t-i}\|_2 \\
&\leq c_3 + c_4\sum_{i=1}^{\infty}\gamma^{i-1}(1+e^{\gamma-\beta})^{(i-1)/2},
\end{aligned}
$$

where

$$c_3 = \left[e^{\beta-\gamma} e^{e^{\beta-\gamma}\left(e^{\gamma e^{-(\beta-\gamma)}}-1\right)}\right]^{1/2}$$

and

$$c_4 = \left[\gamma^2 e^{\beta-\gamma} e^{e^{\beta-\gamma}\left(e^{\gamma e^{-(\beta-\gamma)}}-1\right)} e^{\gamma e^{-(\beta-\gamma)}} \left(e^{\gamma e^{-(\beta-\gamma)}} + e^{-(\beta-\gamma)}\right)\right]^{1/2}.$$

Therefore, $E|e^{W_t}\dot{W}_t\dot{W}_t^T|$ will be finite for $\gamma(1+e^{\gamma-\beta})^{1/2} < 1$.

The convergence required in condition (4.8) is easily established using condition (4.7) and the stationarity of $\{W_t\}$. Now,

$$\begin{aligned}
&\sum_{t=1}^{n} E\Big(\eta_{nt}\eta_{nt}^T I(|\eta_{nt} > \epsilon|) \mid \mathcal{F}_{t-1}\Big) \\
&\quad = \frac{1}{n}\sum_{t=1}^{n} E\left[(Y_{t-1} - e^{W_{t-1}})^2 \dot{W}_t\dot{W}_t^T I(|(Y_{t-1} - e^{W_{t-1}})\dot{W}_t| > \epsilon\sqrt{n})|\mathcal{F}_{t-1}\right] \\
&\quad \leq \frac{1}{n}\sum_{t=1}^{n} E\left[(Y_{t-1} - e^{W_{t-1}})^2 \dot{W}_t\dot{W}_t^T I(|(Y_{t-1} - e^{W_{t-1}})\dot{W}_t| > M)|\mathcal{F}_{t-1}\right] \\
&\quad \stackrel{n\to\infty}{\longrightarrow} E\left[(Y_1 - e^{W_1})^2 \dot{W}_1\dot{W}_1^T I(|(Y_1 - e^{W_1})\dot{W}_1| > M)\right] \\
&\quad \longrightarrow 0 \text{ as } M \to \infty.
\end{aligned}$$

Therefore, the asymptotic distribution of the maximum likelihood estimates is $N(\delta_0, V^{-1})$ where

$$V = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} e^{W_t(\delta_0)}\dot{W}_t\dot{W}_t^T. \tag{4.11}$$

4.3 Simulations

To illustrate the asymptotic properties of the parameter estimates, we simulate from two models and compare the results with the theory obtained in the previous section. Both models contain an intercept term and we set $\lambda = 1$, $p = 0$ and $q = 1$ in (4.2) and (4.3). The second model also includes a simple linear trend. Thus, the two models we consider are:

$$W_t = \beta_0 + \gamma(Y_{t-1} - e^{W_{t-1}})e^{-W_{t-1}} \tag{4.12}$$

and

$$W_t = \beta_0 + \beta_1 t/50 + \gamma(Y_{t-1} - e^{W_{t-1}})e^{-W_{t-1}}. \quad (4.13)$$

Table 4.3 contains the results for the model defined by 4.12 for two choices of β_0 and γ (δ_1 and δ_2) with a sample size of $n = 500$ and $N = 5000$ replications. In this table, we define n to be the simulation sample size, N to be the number of simulations, $\hat{\mu}_{\hat{\delta}_j}$ to be the average of the N estimates of δ_j, $\hat{\sigma}_{\hat{\delta}_j}$ to be the sample standard deviation of the N estimates of δ_j, $s_{\hat{\delta}_{j,i}}$ to be the estimate of the standard error of $\delta_{j,i}$ as given by either (4.4) or (4.5), $\hat{\mu}_{s_{\hat{\delta}_j}}$ to be the average of the $s_{\hat{\delta}_{j,i}}$ and $\hat{\sigma}_{s_{\hat{\delta}_j}}$ to be the standard deviation of the $s_{\hat{\delta}_{j,i}}$, where $\boldsymbol{\delta} = (\boldsymbol{\beta}^T, \gamma)^T$.

Similarly, Table 4.3 contains the results for the model given by (4.13) for two combinations of $(\beta_0, \beta_1, \gamma) = (\delta_1, \delta_2, \delta_3)$. Other values for n and N were considered; however, significantly smaller values of n resulted in greater bias and complications in obtaining the maximum likelihood estimates.

Table 4.1: Simulations, no trend. n=500. N=5000.

parameters	$\hat{\mu}_{\hat{\delta}_j}$	$\hat{\sigma}_{\hat{\delta}_j}$	$\hat{\mu}_{\hat{\delta}_j} \pm 1.96\hat{\sigma}_{\hat{\delta}_j}/\sqrt{N}$	$\hat{\sigma}_{\hat{\delta}_j}\sqrt{1 \pm 1.96\sqrt{(2/n)}}$	$\hat{\mu}_{s_{\hat{\delta}_j}}$
$\delta_1 = 1.5$	1.4993	0.0263	(1.4985, 1.5000)	(0.0258, 0.0268)	0.0265
$\delta_2 = 0.25$	0.2489	0.0403	(0.2477, 0.2500)	(0.03945, 0.0411)	0.0408
$\delta_1 = 1.5$	1.4989	0.0366	(1.4979, 1.4999)	(0.0359, 0.0373)	0.0364
$\delta_2 = 0.75$	0.7502	0.0218	(0.7496, 0.7508)	(0.0214, 0.0222)	0.0218
$\delta_1 = 3$	3.0000	0.0125	(2.9996, 3.0003)	(0.0123, 0.0127)	0.0125
$\delta_2 = 0.25$	0.2496	0.0431	(0.2484, 0.2508)	(0.0422, 0.0439)	0.0430
$\delta_1 = 3$	2.9997	0.0175	(2.9992, 3.0002)	(0.0172, 0.0178)	0.0174
$\delta_2 = 0.75$	0.7504	0.0270	(0.7497, 0.7512)	(0.0265, 0.0275)	0.0271

Notice that the "true" parameter value, δ_j, is very close to the estimated value, $\hat{\mu}_{\hat{\delta}_j}$, in all cases and that the corresponding standard deviation of the estimates is quite small. A comparison can also be made to evaluate the accuracy of our estimates of standard error. We estimate $V_{j,j}$, as defined in (4.11), by $\hat{\sigma}_{\hat{\delta}_j}$. This value can then be compared to the average of our N estimates of standard error, $\hat{\mu}_{s_{\hat{\delta}_j}}$. Again, these values are very close, supporting the theory derived in the previous section.

Table 4.2: Simulations, linear trend. n=500, N=5000.

parameters	$\hat{\mu}_{\hat{\delta}_j}$	$\hat{\sigma}_{\hat{\delta}_j}$	$\hat{\mu}_{\hat{\delta}_j} \pm 1.96\hat{\sigma}_{\hat{\delta}_j}/\sqrt{N}$	$\hat{\sigma}_{\hat{\delta}_j}\sqrt{1 \pm 1.96\sqrt{(2/n)}}$	$\hat{\mu}_{s_{\hat{\delta}_j}}$
$\delta_1 = 1$	1.0001	0.0286	(0.9994, 1.0009)	(0.0279, 0.0291)	0.0284
$\delta_2 = 0.5$	0.5000	0.0035	(0.4999, 0.5001)	(0.0034, 0.0036)	0.0034
$\delta_3 = 0.25$	0.2477	0.0420	(0.2468, 0.2491)	(0.0411, 0.0427)	0.0426
$\delta_1 = 1$	0.9983	0.0795	(0.9960, 1.0004)	(0.0787, 0.0819)	0.0805
$\delta_2 = -0.15$	-0.1502	0.0171	(-0.1507, -0.1498)	(0.0170, 0.0176)	0.0173
$\delta_3 = 0.25$	0.2475	0.0337	(0.2465, 0.2484)	(0.0332, 0.0346)	0.0339

To further illustrate the theoretical properties of the parameter estimates derived in the previous section, Figure 4.1 contains plots of the estimated densities along with the appropriate normal density for one set of parameters from each of the two models. We have chosen the example $\beta_0 = 1.5$, $\gamma = 0.25$ for the first model (4.12) and $\beta_0 = 1$, $\beta_1 = -0.15$, $\gamma = 0.25$ for the linear trend model (4.13). The normal densities which are given have mean δ_j and variance $\hat{\sigma}_{\hat{\delta}_j}$. The densities, estimated and asymptotic, are very close in both examples.

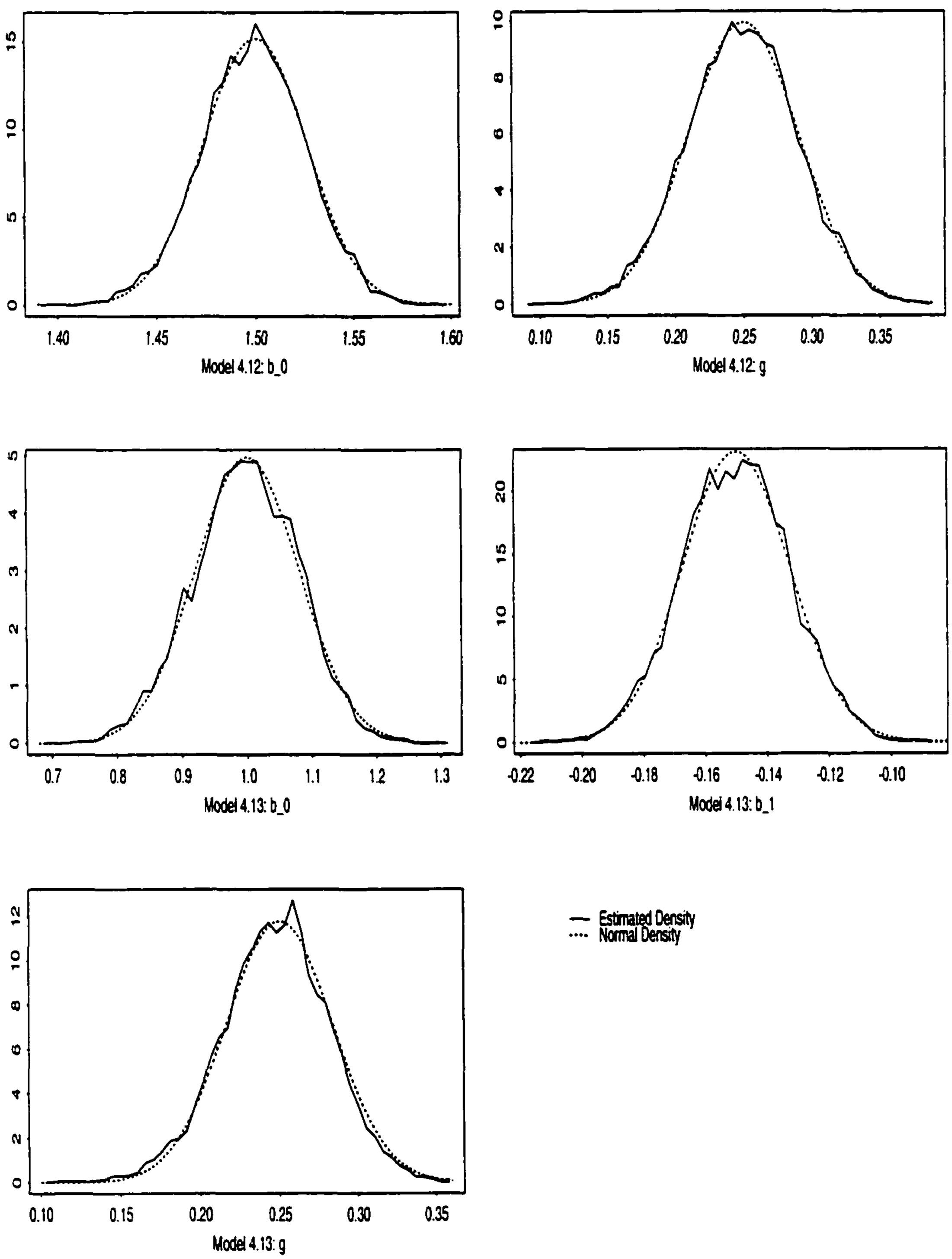

Figure 4.1: Estimated and Asymptotic Densities. Row 1: Model 4.12 with $\beta_0 = 1.5$, $\gamma = 0.25$. Rows 2 and 3: Model 4.13 with $\beta_0 = 1$, $\beta_1 = -0.15$, $\gamma = 0.25$.

4.4 Application to Asthma Data

In this last section of the thesis, we fit GLARMA models to a data set consisting of asthma counts from a hospital in Cambelltown, outside of Sydney, Australia. These daily counts are shown in Figure 4.2 for January 1, 1990 - December 31, 1993. Upon examination, it is clear that there is a seasonal aspect to the data with higher counts in the fall (March - May). It also appears as if there may be an increasing trend over time. This is most likely attributed to an increase in population and pollution in the area. Another aspect of the data which is not clear from the figure but is not difficult to explain, is the potential for a day of the week effect. As will be seen in the analysis, there is a higher incidence rate recorded on Sunday and Monday than for other days of the week. One of the more plausible explanations for this phenomena is that family physicians are typically not available on Sundays; thus, people seeking immediate medical attention must visit the hospital. Others may choose to wait a day to see their private physician, but end up in the hospital because their condition has worsened.

An exploratory analysis of this data set was performed by Davis, Dunsmuir and Wang (1999). Several of the explanatory variables they considered include a Sunday effect, a Monday effect, a linear trend term, and sinusoidal terms to model the seasonal trends in the data. Initial estimates of the parameters were found using a Poisson regression with the aforementioned terms as predictors. The trend term was not found to be significant; however, the terms $\cos(2\pi kt/365)$ and $\sin(2\pi kt/365)$, $k = 1, 2, 3, 4$ along with terms for both a Sunday and Monday effect were found to be significant. The log-likelihood under this model was -796.36. The residuals from this fit exhibited significant correlation at lags 1,3,4,7,10. The Q statistic, which is a Portmanteau style test for the existence of a latent process (see (Davis *et al.*, 1999)) was 2.39. Under the null hypothesis that there is no latent process, this statistic is normally distributed with mean = 0 and variance = 1.

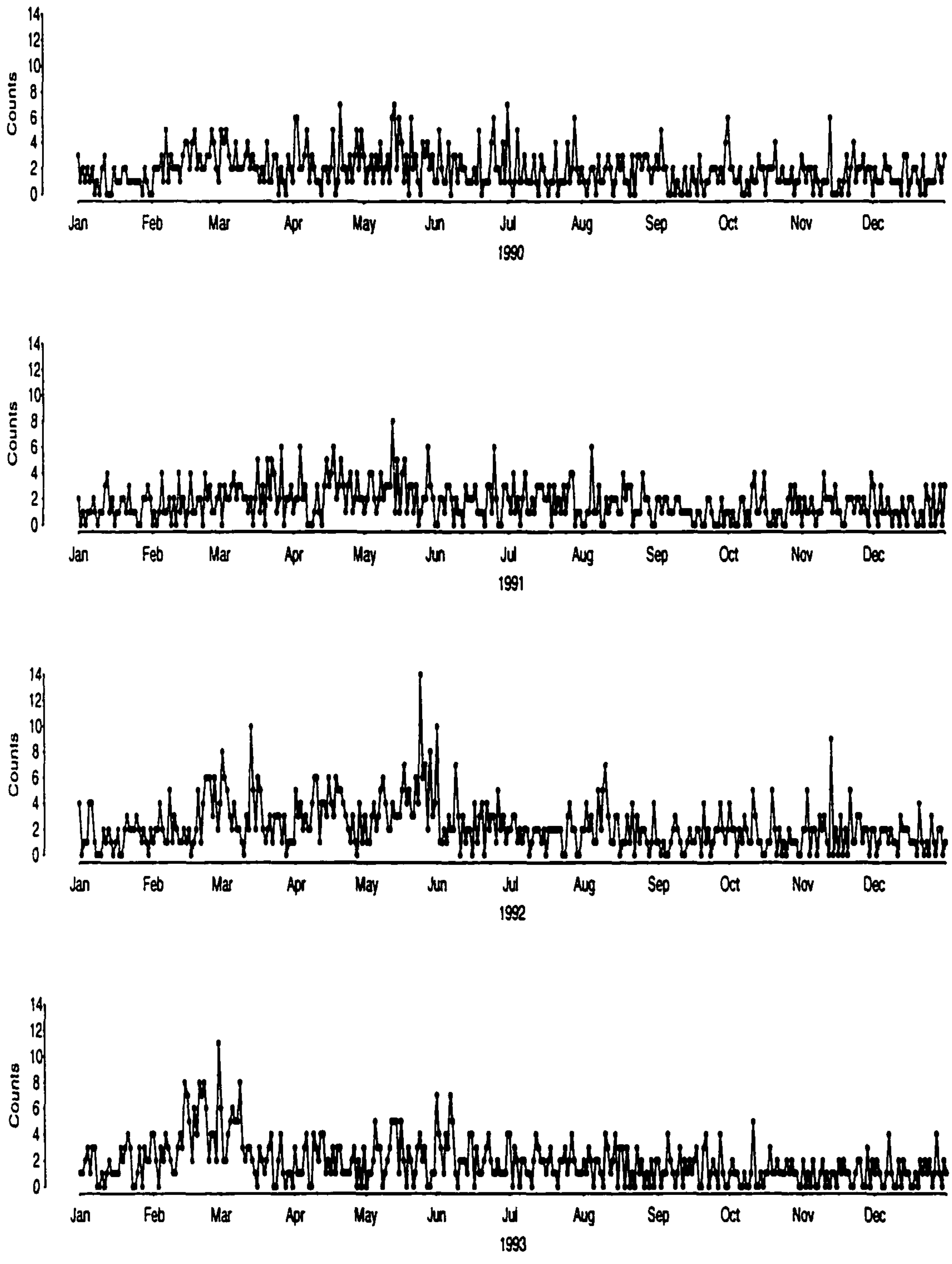

Figure 4.2: Asthma Counts

The same explanatory variables were used to fit a GLARMA model of the form (1.5) with p=0 and MA lags at 1,3,4,7,10. To initialize the recursions used to fit the model, the Poisson parameter estimates were used along with zero initial values for the ARMA terms. The resulting estimates are given in Table 4.4. The log-likelihood for this model was -779.71 which indicates a substantial improvement in the fit. Note that the MA coefficient at lag 4 is not significant. The model was refit without this lag. The results from this model fit are also given in Table 4.4. This model had a log-likelihood value of -781.03, a non-significant change from the model with all 5 lags included. The Q statistic for this model was 1.82 indicating that there is additional variance which is unaccounted for by the MA terms. However. the Q statistic is significantly lower than that obtained by the Poisson regression. Auto-correlation plots of the Pearson residuals indicate substantial improvement was made in accounting for the serial correlation present in the data. These plots are presented in Figure 4.3 for the Poisson regression model as well as the GLARMA model with MA lags 1,3,7,10.

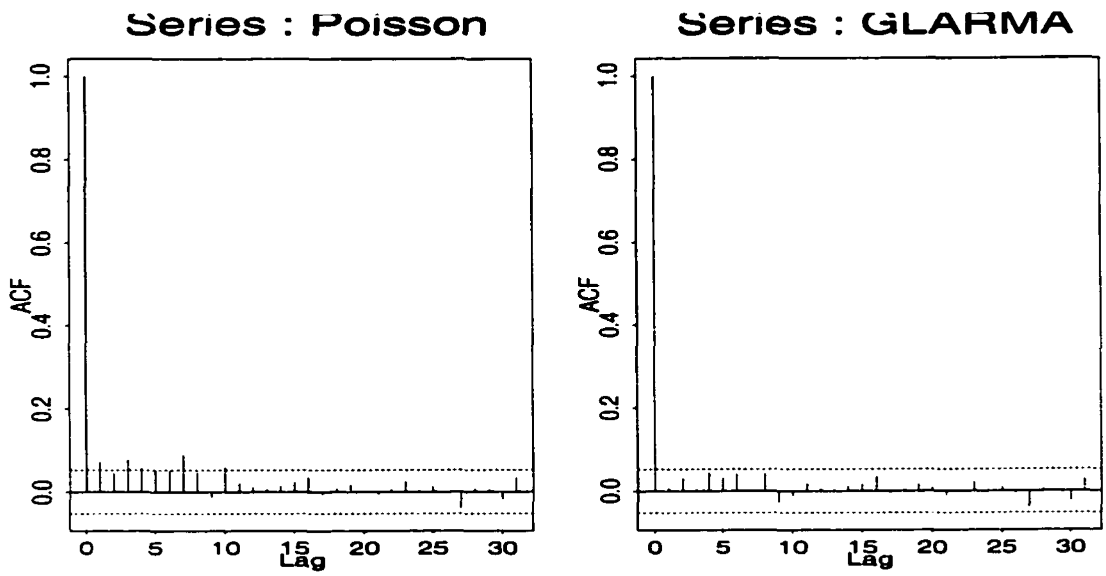

Figure 4.3: ACFs of residuals from Poisson and GLARMA fits

Table 4.3: Parameter Estimates

Term	MA(1,2,3,5,7,10) Parameter	se	MA(1,3,7,10) Parameter	se
(Intercept)	0.531	0.030	0.533	0.029
Sunday effect	0.243	0.053	0.240	0.054
Monday effect	0.253	0.053	0.249	0.054
$\cos(2\pi t/365)$	-0.164	0.037	-0.162	0.036
$\sin(2\pi t/365)$	0.362	0.036	0.362	0.035
$\cos(4\pi t/365)$	-0.065	0.038	-0.067	0.036
$\sin(4\pi t/365)$	0.024	0.035	0.023	0.034
$\cos(6\pi t/365)$	-0.082	0.036	-0.083	0.035
$\sin(6\pi t/365)$	0.011	0.036	0.009	0.035
$\cos(8\pi t/365)$	-0.156	0.035	-0.157	0.034
$\sin(8\pi t/365)$	-0.062	0.035	-0.062	0.034
θ_1	0.053	0.024	0.053	0.024
θ_3	0.064	0.024	0.061	0.024
θ_4	0.039	0.024		
θ_7	0.080	0.024	0.078	0.024
θ_{10}	0.070	0.024	0.071	0.024

The analysis of the asthma data set given in this section is by no means meant to be a comprehensive one. These results were included to indicate how the GLARMA models of Section 2.2 might be applied in practice. Further improvements on this model are the subject of ongoing research.

Bibliography

Cox, D. (1981). Statistical Analysis of Time Series: Some Recent Developments. *Scandinavian Journal of Statistics*, 8:93–115.

Davis, R. A., Wang, Y., and Dunsmuir, W. T. (1999). *Modeling Time Series of Count Data. In Asymptotics, Nonparametrics & Time Series, Ed Subir Ghosh.* New York: Marcel Dekker.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50:987–1008.

Glynn, P. and Meyn, S. (1997). Tightness for Non-irreducible Markov Chains. Submitted for publication.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Application.* New York: Academic Press.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models.* Great Britain: Chapman and Hall, second edition.

Meyn, S. and Tweedie, R. (1993). *Markov Chains and Stochastic Stability.* Great Britain: Springer-Verlag.

Murdoch, D. and Green, P. (1998). Exact sampling from a continuous state space. *Scandinavian Journal of Statistics*, 25:483–502.

Propp, J. and Wilson, D. (1996). Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252.

Rydberg, T. H. and Shephard, N. (1998). Dynamics of trade-by-trade price movements: decomposition and models. working paper.

Rydberg, T. H. and Shephard, N. (1999). BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time. working paper.

Wong, W. (1986). Theory of partial likelihood. *Annals of Statistics*, 14:88–123.

Zeger, S. L. and Qaqish, B. (1988). Markov Regression Models for Time Series: A Quasi-Likelihood Approach. *Biometrics*, 44:1019–1031.

Appendix A

PRELIMINARY RESULTS FOR MARKOV CHAINS

A.1 Introduction

This appendix provides several definitions and properties relating to Markov chains. The proofs have been omitted for the theorems given below with one exception. The omitted proofs may be found in the indicated references. Additional background material may be found in Meyn and Tweedie (1993)

A.2 Definitions and Properties

Definition A.2.1 A Markov chain $\{\mathbf{X_n}\}$ with state space X is φ *-irreducible* if there exists a measure φ on $\mathcal{B}(X)$ such that, whenever $\varphi(A) > 0$, $P_x(\tau_A < \infty) > 0$ for all $x \in X$.

Definition A.2.2 The Markov chain $\{\mathbf{X_n}\}$ is called *ψ-irreducible* if it is φ-irreducible for some φ and the measure ψ is a *maximal irreducibility* measure satisfying the conditions of the following proposition.

Proposition A.2.3 (Proposition 4.2.2 of Meyn and Tweedie (1993))
If $\{\mathbf{X_n}\}$ is φ-irreducible for some measure φ, then there exists a probability measure ψ on $\mathcal{B}(X)$ such that

(i) $\{\mathbf{X_n}\}$ *is ψ-irreducible;*

(ii) *for any other measure φ', the chain $\{\mathbf{X_n}\}$ is φ'-irreducible if and only if ψ is absolutely continuous with respect to φ' ($\psi(A) = 0$ implies $\varphi'(A) = 0$);*

(iii) *if* $\psi(A) = 0$, *then* $\psi\{y : P_y(\tau_A < \infty) > 0\} = 0$;

(iv) *the probability measure* ψ *is equivalent to*

$$\psi'(A) := \int_X \sum_{n=0}^{\infty} P^n(y, A) 2^{-(n+1)} \varphi'(dy),$$

for any finite irreducibility measure φ'.

Definition A.2.4 An irreducible chain on a countable space X is called *strongly aperiodic* if $P(x, x) > 0$ for some $x \in X$.

Definition A.2.5 A set $C \in \mathcal{B}(X)$ is called a ν_m*−small set* if there exists an $m > 0$, and a non-trivial measure ν_m on $\mathcal{B}(X)$, such that for all $x \in C$, $B \in \mathcal{B}(X)$,

$$P^m(x, B) \geq \nu_m(B).$$

Definition A.2.6 A set $C \in \mathcal{B}(X)$ is called ν_a*-petite* if the chain satisfies

$$\sum_{n=0}^{\infty} P^n(x, B) a(n) \geq \nu_a(B)$$

for all $x \in X$, $B \in \mathcal{B}(X)$, where ν_a is a non-trivial measure on $\mathcal{B}(X)$.

Definition A.2.7 A φ-irreducible Markov chain on a general state space is called *strongly aperiodic* if there exists a ν_1-small set A with $\nu_1(A) > 0$.

Definition A.2.8 A chain is said to be *weak Feller* if its transition probability kernel P maps $C(X)$ to $C(X)$, where

$$P(h(x)) \;:=\; \int P(x, dy) h(y), \quad x \in X,$$

and $C(X)$ represents the class of bounded continuous functions from X to $\mathbb{R}$ and X is the state space of the chain.

Definition A.2.9 A point $x \in X$ is called *reachable* if for every open set $O \in \mathcal{B}(X)$ containing x,

$$\sum_n P^n(y, O) > 0, \quad y \in X.$$

Definition A.2.10 A σ-finite measure π is *invariant* if

$$\pi(A) = \int_X \pi(dx) P(x, A), \quad A \in \mathcal{B}(X).$$

Definition A.2.11 A sequence of probabilities $\{\mu_k : k \in \mathbb{Z}_+\}$ is *tight* if for each $\epsilon > 0$, there exists a compact subset $C \subset X$ such that

$$\liminf_{k \to \infty} \mu_k(C) \geq 1 - \epsilon.$$

Definition A.2.12 A chain $\{\mathbf{X_n}\}$ will be called *bounded in probability on average* if for each initial condition $x \in X$ the sequence $\{\frac{1}{k} \sum_{i=1}^k P^i(x, \cdot) : k \in \mathbb{Z}_+\}$ is tight.

Theorem A.2.13 (Theorem 12.0.1 of Meyn and Tweedie (1993))
If $\{\mathbf{X_n}\}$ is a weak Feller chain which is bounded in probability on average then there exists at least one invariant probability measure.

A proof of the following theorem is outlined in Glynn and Meyn (submitted for publication) . an alternate proof relying solely on basic principles is given below.

Theorem A.2.14 *If $\{\mathbf{X_n}\}$ is a weak Feller chain, then for each $m \geq 1$ and compact A, the family of probability measures $\{1/m \sum_{k=1}^m P^k(a, \cdot) : a \in A\}$ is tight.*

Proof:

For $m \geq 1$ fixed, let $P^*(a, \cdot) = 1/m \sum_{k=1}^m P^k(a, \cdot)$. Now, by the weak Feller property, given $\epsilon > 0$ and $a \in A$, we may choose δ_a and compact C_a such that

$$\begin{aligned} P^*(a, C_a) &> 1 - \epsilon, \text{ and} \\ P^*(x, C_a) &> 1 - \epsilon \text{ for } |x - a| < \delta_a \end{aligned}$$

Since $A \subseteq \bigcup_{a \in A} B_{\delta_a}(a)$ where $B_{\delta_a}(a)$ is an open ball of radius δ_a and A is compact, $A \subseteq \bigcup_{j=1}^{n} B_{\delta_{a_j}}(a_j)$ for some finite set $\delta_{a_1}, \ldots, \delta_{a_n}$. Let $C = \bigcup_{j=1}^{n} C_{a_j}$. Then C is compact and $P^{*}(x, C) > 1 - \epsilon$ for all $x \in A$. Thus, for each $m \geq 1$, the family of probability measures $\{1/m \sum_{k=1}^{m} P^k(a, \cdot) : a \in A, A \text{ compact}\}$ is tight.

Theorem A.2.15 (Corollary 2.2 of Glynn and Meyn (submitted for publication))
Suppose that there exists a measurable function $V : X \to [0, \infty)$ satisfying

(i) For some $b < \infty$,

$$PV(x) \leq V(x) - 1 + bI_A(x), \; x \in X;$$

(ii)

$$\lim_{n \to \infty} \sup_{a \in A} E_a[V(X_n) I_{[\tau_A > n]}] = 0$$

(iii) The family of probability measures $\{1/m \sum_{k=1}^{m} P^k(a, \cdot) : a \in A\}$ is tight for each $m \geq 1$.

Then the chain, $\{\mathbf{X_n}\}$, is bounded in probability on average.

Theorem A.2.16 (Theorem 11.3.4 of Meyn and Tweedie (1993))
Suppose $C \in \mathcal{B}(X)$, and $V : X \to [0, \infty]$ satisfies

$$\Delta V(x) \leq -1 + bI_c(x), \; x \in X,$$

then

$$E_x[\tau_C] \leq V(x) + bI_C(x)$$

for all $x \in X$.

Definition A.2.17 The Markov transition function P is called equicontinuous if for each continuous function f with compact support, the sequence of functions $\{P^k f \; : \; k \in \mathbb{Z}_+\}$ is equicontinuous on compact sets. A Markov chain which possesses an equicontinuous Markov transition function will be called an *e-chain.*

Theorem A.2.18 (Theorem 18.4.4(i) of Meyn and Tweedie (1993))

If $\{\mathbf{X_n}\}$ *is an e-chain which is bounded in probability on average then a unique invariant probability measure exists if and only if a reachable state* $x^* \in X$ *exists.*

Theorem A.2.19 (Theorem 10.0.1 of Meyn and Tweedie (1993))

If the chain $\{\mathbf{X_n}\}$ *is recurrent then it admits a unique (up to constant multiples) invariant measure* π*, and the measure* π *has the representation, for any* $A \in \mathcal{B}^+(X)$

$$\pi(B) \;=\; \int_A \pi(dw)E_w\left[\sum_{n=1}^{\tau_A} I[X_n \in B]\right], \quad B \in \mathcal{B}(X).$$

Theorem A.2.20 (Theorem 13.0.1 of Meyn and Tweedie (1993))

If $\{\mathbf{X_n}\}$ *is an aperiodic Harris recurrent chain with invariant measure* π *and there exists some petite set* C*, some* $b < \infty$ *and a non-negative function* V *finite at some one* $x_0 \in X$*, satisfying*

$$\Delta V(x) \;:=\; PV(x) - V(x) \leq -1 + bI_C(x), \quad x \in X,$$

then there exists a unique invariant probability measure π *such that for every initial condition* $x \in X$,

$$\sup_{A \in \mathcal{B}(X)} |P^n(x, A) - \pi(A)| \to 0$$

as $n \to \infty$.

Definition A.2.21 A chain $\{\mathbf{X_n}\}$ is called *uniformly ergodic* if

$$\sup_{x \in X} \|P^n(x, \cdot) - \pi\| \to 0, \quad n \to \infty.$$

Definition A.2.22 (Doeblin's Condition)

Suppose there exists a probability measure ϕ with the property that for some $m \geq 1, \epsilon < 1, \delta > 0$

$$\phi(A) > \epsilon \Longrightarrow P^m(x, A) \geq \delta$$

for every $x \in X$.

Theorem A.2.23 (Theorem 16.0.2 of Meyn and Tweedie (1993))
A Markov chain $\{\mathbf{X_n}\}$ *is uniformly ergodic if and only if the chain is aperiodic and satisfies Doeblin's condition.*

Definition A.2.24 $\{\mathbf{X_n}\}$ is *geometrically mixing* if there exists $R < \infty, \rho < 1$ such that

$$\sup |E_x[g(\{X_n\}_k)h(\{X_n\}_{n+k})] - E_x[g(\{X_n\}_k)]E_x[h(\{X_{n+k})]| \leq R\rho^n, \quad n \in \mathbb{Z}_+,$$

where the supremum is taken over all $k \in \mathbb{Z}_+$, and all g and h such that $g^2(x), h^2(x) \leq 1$ for all $x \in X$.

Theorem A.2.25 (Theorem 16.1.5 of Meyn and Tweedie (1993))
If $\{\mathbf{X_n}\}$ *is uniformly ergodic then* $\{\mathbf{X_n}\}$ *is geometrically mixing.*