ESSAYS IN FINANCIAL ECONOMETRICS AND TIME SERIES ANALYSIS

Thesis submitted in accordance with the requirements of

The University of Liverpool

for the degree of

Doctor in Philosophy

by

Ruijun Bu

August 2006

Abstract

This Ph.D thesis is divided into two parts which deal with two self-contained studies in the field of financial econometrics and time series analysis.

Part I is concerned with estimating option implied risk-neutral probability density functions (RNDs), where we examine the ability of two recent methods — the smoothed implied volatility smile method (SML) and the density functionals based on confluent hypergeometric functions (DFCH) — for estimating RNDs from European options. Two complementary Monte Carlo experiments are conducted and the performance of the two methods is evaluated by the Root Mean Integrated Squared Error (RMISE) criterion. Results from both experiments indicate that the DFCH method dominates the SML method for the overall quality of the estimated RNDs concerning both accuracy and stability. In our application to real option data, the DFCH performs consistently well, whereas the SML has problems with the choice of the smoothing parameter.

Part II of the thesis considers maximum likelihood (ML) estimation of higher-order integer-valued autoregressive (INAR(p)) processes. A recursive representation of the transition probability function is proposed, based on which we derive the score and Fisher information in terms of conditional expectations. These new expressions enhance the interpretation of these quantities and lead to new definitions of residuals. Using the INAR(2) specification with Poisson innovations, we investigate both the asymptotic efficiency and the finite sample performance of the ML estimator (MLE).

Our results confirm that the MLE is asymptotically more efficient than the Yule-Walker estimator (YWE) and the conditional least squares estimator (CLSE), and that there is also a potential gain in implementing the MLE in small samples in terms of bias and mean squared error (MSE). A computationally efficient approach based on Markov chain techniques for producing probability forecasts for INAR(p) models is also proposed. In our empirical analysis of the Westgren (1916) data, likelihood based inferences reveal new evidence for improving model specification and forecasts produced by the new approach suggest substantial benefit from the enriched information and improved efficiency.

Acknowledgements

I would like to express my deepest and everlasting gratitude to my supervisors, Professor Kaddour Hadri and Professor Brendan McCabe, without whom my thesis would not have been successful. I am also extremely grateful to Professor David Sapsford, Head of Economics Division at the University of Liverpool Management School, for provision of resources and helpful advices. I would also like to thank Mr. Simon Blackman, Mrs. Paula Ferrington and all other staff members of the school for their continuous help and encouragement. I wish to give my special thanks to my parents, Guolin Bu and Xinfeng Chen, my brother, Ruiyi Bu and my girlfriend, Yao Rao, for their endless love, patience and support. I finally want to thank Universities UK for the ORS Award and the University of Liverpool for the International Research Scholarship.

To my parents, my brother and my love

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

This Ph.D thesis consists of two major parts that address two self-contained studies in financial econometrics and time series analysis which can be unified in the common field of econometrics.

Part I is concerned with estimating option implied risk-neutral probability density functions (RNDs), which is a topic that sits in the financial econometrics literature. We examine the ability of two recent methods for estimating RNDs from European-style options. One is the smoothed implied volatility smile method (SML) developed by Bliss and Panigirtzoglou (2002) and the other is the density functionals based on confluent hypergeometric functions (DFCH) proposed by Abadir and Rockinger (2003). To compare the two methods, we carry out two complementary Monte Carlo experiments based on the pseudo-prices methodology. The performance of alternative methods is evaluated by the criterion of Root Mean Integrated Squared Error

(RMISE). Results from both experiments indicate that the DFCH method outperforms the SML method in terms of the overall quality of the RND estimates concerning both accuracy and stability.

As an illustration of the two methods in real settings, we apply both methods to OTC currency option data. We find that the DFCH method behaves uniformly well in terms of good fitting as well as maintaining proper shapes of the estimated RNDs. For the SML method, however, the choice of the smoothing parameter remains an arbitrary factor.

Part I of the thesis is organized as follows: Chapter 2 is the introduction. Chapter 3 sets out the technical details of the two estimation methods. In Chapter 4, we present the two Monte Carlo experiments and discuss the results. The application is presented in Chapter 5. Chapter 6 concludes. Bu and Hadri (2005) is based on the main results of this study.

In Part II of the thesis, we consider a special type of time series models of count data, the integer-valued autoregressive (INAR) models. Our attention is focused on the maximum likelihood (ML) estimation of higher-order integer-valued autoregressive (INAR(p)) models. We propose a recursive representation of the transition probability function for the INAR(p) model, which not only simplifies the computation of the likelihood but also streamlines the derivation of the score functions and the Fisher information matrix for the INAR(p) model. We show that if the density functions of unobserved model components belong to a special class, the score functions

and the Fisher information matrix can be neatly expressed in terms of conditional expectations. These new expressions not only enhance the interpretation of these quantities but also lead to new definitions of the residuals of the model.

Using the INAR(2) specification with Poisson innovations, we investigate both the asymptotic efficiency and the finite sample performance of the ML estimator (MLE) in comparison with the widely used Yule-Walker estimator (YWE) and the conditional least squares estimator (CLSE). Our results confirm that the MLE is asymptotically more efficient than the YWE and the CLSE, and that there is also a potential gain in implementing the MLE in small samples in terms of bias and mean squared error (MSE). In particular, we find that the magnitude of efficiency gain in implementing MLE is positively related to the degree of persistence of the underlying process.

An efficient approach built on Markov chain theory for producing probability forecasts for INAR(p) models is also proposed. Since this approach is based on the transition matrix method, it is computationally attractive. A method for incorporating parameter uncertainties into the probability forecasts is also suggested.

We carry out an empirical analysis of the Westgren (1916) gold particle data under the ML framework developed in this study. The emphasis is on the issue of model adequacy. Both residual analysis and specification testing are discussed. We show that in the light of likelihood estimation it is possible to unveil previously hidden evidence suggesting possible improvements of model specification. Also in this application, we illustrate the effectiveness of the newly developed forecasting tools by

producing distribution forecasts for the Westgren data based on the fitted model. We find that in terms of the enriched information and the advanced efficiency the benefit of implementing the approach is considerable.

The structure of Part II of the thesis is as follows: Chapter 7 gives an introduction. In Chapter 8, we review the INAR(p) model and existing estimation methods. Chapter 9 looks at the maximum likelihood estimation of a generalized INAR(p) model. In Chapter 10, we investigate the relative performance of the three estimators. Chapter 11 considers forecasting with INAR(p) models. The application is presented in Chapter 12. In Chapter 13, we conclude. Bu et al. (2006a,b,c) are based on the main results of this part of the thesis.

Part I

Estimating Option Implied Risk-Neutral Densities using Spline and Hypergeometric Functions

Chapter 2

Introduction

Cross sections of observed option prices have long been used to estimate implied risk-neutral probability density functions (RNDs). These RNDs represent forward-looking forecast of the distributions of the prices of the underlying asset. They have proved to be particularly useful for various applications. They are used for pricing complex derivatives; estimating parameters of the underlying stochastic processes — Bates (1996); testing market rationality — Bondarenko (1997); estimating risk preferences — Ait-Sahalia and Lo (2000), Jackwerth (2000), Rosenberg and Engle (2002). In particular, option implied RNDs have found an extensive use for monetary policy purposes by an increasing number of Central Banks. These applications include, Söderlind and Svensson (1997) who discuss the extraction of interest rate expectations for monetary purposes; Bahra (1997) illustrates how RNDs are used by policy-makers for assessing monetary conditions, monetary credibility, the timing and

effectiveness of monetary operations and identifying anomalous market prices; McManus (1999) uses Eurodollar options to examine the evolution of market sentiment over the possible future values of Eurodollar rates; Jondeau and Rockinger (2000) apply the method to exchange rate options for two dates (calm and agitated market) to find out how the market participants expectation is affected; Söderlind (2000) employs daily option prices to estimate how the market's probability distribution of the future mark-pound exchange rate and UK and German interest rates changed before and after the ERM crisis. Other applications of the method include Melick and Thomas (1997), who use American options to estimate the market participants' expected distribution of oil prices during the Gulf crisis.

A large number of methods have been developed for recovering the implied RNDs. Generally, these methods can be divided into parametric and nonparametric ones. Parametric methods rely on specific assumptions on the data generating process. Examples that have been used include: generalized distribution methods of Aparicio and Hodges (1998), Rosenberg (1998) and Lim et al. (2005); expansion methods of Jarrow and Rudd (1982) and Rubinstein (1998); the lognormal mixture model of Bahra (1997) and Melick and Thomas (1997); and models for stochastic process of Heston (1993), Bates (1996) and Wu and Huang (2004). Nonparametric methods are flexible, data-driven methods. Examples of nonparametric methods include: implied trees of Rubinstein (1994); kernel estimation methods of Ait-Sahalia and Lo (1998, 2000) and Ait-Sahalia et al. (2001); smoothing techniques of Campa et al. (1998)

and Bliss and Panigirtzoglou (2002); maximum entropy methods of Buchen and Kelly (1996) and Stutzer (1996); and neural network approach of Garcia and Gencay (2000) and Gottschling et al. (2000). For surveys of existing methods, see Jackwerth (1999), Jondeau and Rockinger (2000) and Bliss and Panigirtzoglou (2002) among others.

While many papers have estimated and interpreted the option implied RNDs, relatively few have considered the reliability of these methods for estimating implied RNDs. Among the latter are Söderlind and Svensson (1997), Melick and Thomas (1998), Cooper (1999), Söderlind (2000), Bliss and Panigirtzoglou (2002), and Bondarenko (2003). Söderlind and Svensson (1997) and Melick and Thomas (1998) both worked with the parameter variance-covariance matrix. Relying on the assumption of the distribution of the estimated parameters, the confidence intervals of the estimated RNDs were examined. Cooper (1999), Söderlind (2000) and Bondarenko (2003), on the other hand, used the pseudo-prices method. The pseudo-prices method begins with known RNDs which are used to generate fitted prices. These fitted prices are then randomly perturbed to generate pseudo-prices. These pseudo-prices are finally used to estimate the implied RNDs, based on which the performance of an estimation method is assessed. Bliss and Panigirtzoglou (2002), however, focused only on the stability of the estimated implied RNDs. Therefore, they chose to perturb real option prices. Both Cooper (1999) and Bliss and Panigirtzoglou (2002) examined the two most commonly used methods: the double lognormal approximating function method (DLN) and the smoothed implied volatility smile method (SML). Both

concluded that the SML method dominates the DLN method as a technique for estimating option implied RNDs. Many authors have since used the SML method in various studies. Bliss and Panigirtzoglou (2004) and Panigirtzoglou and Skiadopoulos (2004) are among the examples.

Abadir and Rockinger (2003) recently proposed an alternative method for estimating option implied RNDs. We call it the density functionals based on the confluent hypergeometric functions (DFCH). This method is solidly founded in the theory of statistical density functionals and is particularly appealing for its semi-nonparametric nature. It is more efficient than fully nonparametric estimation and more flexible than purely parametric methods. It encompasses a large class of traditional densities, such as normal, gamma, inverse gamma, Weibull, Pareto and mixtures thereof. Thus, the possibility of misspecification is believed to be small. They showed that their method performed uniformly well in their two applications. Although the DFCH method appears to be an appealing alternative, surprisingly it did not attract any noticeable follow-up, at least to our best knowledge. The main contribution of this part of the thesis the comparison of the SML method and the DFCH method for estimating option implied RNDs.

To compare the two methods, we conduct two Monte Carlo experiments. Both experiments are based on the pseudo-prices methodology. In the first experiment, the true RNDs are generated by implementing the Heston (1993) stochastic volatility model. Different sets of parameters are selected for this model so that our true RNDs

incorporate various market conditions. In the second experiment, the true RND is specified as a mixture of three lognormals. In order to generate the true RND that is representative to the observed world, we calibrate the model using observed prices of a typical cross section of S&P 500 Index options traded at Chicago Board Option Exchange (CBOE). The two experiments can be regarded as complementary to each other in the sense that when combined they represent a broader setting for making comparison. In both experiments, we examine the ability of the two methods for recovering the true RNDs in the presence of small pricing errors. We evaluate the performance of the two RND estimators by focusing on the criterion of Root Mean Integrated Squared Error (RMISE). Results from both experiments indicate that the DFCH method dominates the SML method for the overall quality of the estimated RNDs concerning both the accuracy and the stability defined in this study.

The remainder of Part I of the thesis is organized as follows. Chapter 3 sets out the technical details of the two estimation methods. In particular, we improve the SML method by providing an analytic expression for the RND estimator, which constitutes another contribution of this study. In Chapter 4, we present the two Monte Carlo simulation experiments and discuss the results. As an illustration of the two methods, we present an application to OTC currency option data in Chapter 5. Some concluding remarks are given in Chapter 6. The main results of this part of the thesis appear in Bu and Hadri (2005).

Chapter 3

Methods for Estimating Implied RNDs

We begin this chapter by a brief review of the economics underlying the methods for estimating option implied RNDs in Section 3.1. Technical details of the SML method is presented in Section 3.2. In particular, we refine the SML method by providing an analytic expression for the RND estimator, which improves the computational efficiency of this method. An introduction to the hypergeometric functions and details of the DFCH method are discussed in Section 3.3.

3.1 Option Prices and Risk-Neutral Densities

Prices of European call options at time zero on the underlying asset S with expiration at T and strike price K are related to the risk-neutral probability density

function (RND), $f(\cdot)$, through the following expression:

$$C(K) = e^{-rT} \int_K^\infty (S_T - K) f(S_T) \, dS_T$$

where r is the continuously compounded risk-free interest rate. As noticed by Breeden and Litzenberger (1978), differentiating the integral with respect to strike price K gives

$$\frac{\partial C(K)}{\partial K} = -e^{-rT} \int_K^\infty f(S_T) \, dS_T = -e^{-rT} [1 - F(K)] \tag{3.1}$$

where $F(\cdot)$ is the cumulative distribution function (CDF) corresponding to the probability density function (PDF), $f(\cdot)$. The second derivative is given by

$$\left. \frac{\partial^2 C(K)}{\partial K^2} \right|_{K=S_T} = e^{-rT} f(S_T) \tag{3.2}$$

which reveals the required RND, $f(S_T)$. It follows that the implied RND can be recovered by calculating the compounded second partial derivative of the call pricing function with respect to the strike price. In practice, however, some approximating or smoothing method has to be used to construct such a function due to the limited number of observed call prices in a cross section.

3.2 Smoothed Implied Volatility Smile

3.2.1 General Procedure

The smoothed implied volatility smile (SML) method was originally developed by Shimko (1993). It is an approximating function method applied to the implied

volatility smile. Option prices are first converted to implied volatilities using Black-Scholes option pricing formula. A continuous smoothing function is then fitted to the implied volatilities and the associated strike prices. The reason for smoothing the volatility smile instead of interpolating the call pricing function directly is that it is technically difficult to fit accurately the shape of the latter and small fitted price errors tend to have large affects on the resulting RNDs, particularly in the tails. It is important to note that the use of the Black-Scholes formula is solely to convert data from one space to another, where smoothing can be done more efficaciously. It does not assume that the underlying price process is lognormal. Shimko (1993) used a quadratic functional form to interpolate across the implied volatilities. The continuum of fitted implied volatilities were then converted back to a continuum of fitted option prices. The implied RNDs can be obtained by applying equation (3.2). Malz (1997a,b) also used a low-order polynomial as the smoothing function, but fitted the implied volatility against the Black-Scholes option delta ($\delta = \partial C/\partial S$). Campa et al. (1998) introduced the use of a smoothing spline for fitting implied volatility curves. They also applied this to smoothing the implied volatility/strike function.

The SML estimation method used in this study was developed by Bliss and Panigirtzoglou (2002). It follows Malz (1997a,b) in smoothing in implied volatility/delta space and Campa et al. (1998) in using a natural spline to smooth the function. The natural spline minimizes the following objective function:

$$\min_{\Theta}(1-\lambda)\sum_{i=1}^{N} w_i\left(IV_i-\widehat{IV}_i(\delta_i,\Theta)\right)^2+\lambda\int_0^{e^{-r^*T}} g''(x;\Theta)^2\,dx \qquad (3.3)$$

where Θ is the matrix of polynomial parameters of the cubic spline; $g(\Theta)$ is the cubic spline function; and $\widehat{IV}_i(\delta_i,\Theta)$ is the fitted implied volatility at δ_i given the spline parameters Θ. Relative weights[1] to each observation are determined by the values of w_i. The smoothness of the spline is controlled by the smoothing parameter, λ, which multiplies a measure of the degree of curvature in the function — the integral of the squared second derivative of the function over its range. It should be recalled that $0 \leq \delta_i \leq e^{-r^*T}$, where r^* is the dividend rate of the underlying asset.

A natural spline is superior to a low-order polynomial because it allows for more flexibility in the shape of the fitted volatility smile and it also permits the user to control the smoothness of the fitted function. Using the option delta rather than the strike price as the function argument groups away-from-the-money implied volatilities more closely together than near-the-money implied volatilities, permitting greater flexibility in the shape of the approximating function near the center of the distribution (where data is more reliable) without having to use a variable smoothing parameter[2]. In addition, since possible values in the delta space always range from 0 to e^{-r^*T}, the extrapolation area becomes relatively smaller.

Once the natural spline is constructed, the fitted volatility smile is then converted

[1]Bliss and Panigirtzoglou (2002) discussed different types of weighting schemes and how the weighting can account for different sources of pricing error.

[2]See Waggoner (1997) for more discussions on variable smoothness penalties in spline regression.

back to the fitted call pricing function. As before, the implied RNDs are obtained by applying equation (3.2).

3.2.2 Smile Conversion *vs* Point Conversion

The construction of the SML method proposed by Bliss and Panigirtzoglou (2002) requires that the implied volatility smile be smoothed in delta space. Two different ways of converting a strike into its delta have been suggested in the literature, differing in their choice of the volatility in the delta function. The original one, proposed by Malz (1997a,b), is to use the implied volatility that corresponds to the strike price. This is achieved by converting strike prices into deltas using the Black-Scholes delta given by the following equation:

$$\delta_K = e^{-r^*T}\Phi\left(\frac{\ln S_0 - \ln K + \left(r - r^* + \frac{\sigma_K^2}{2}\right)T}{\sigma_K\sqrt{T}}\right)$$

The subscript of σ_K emphasizes that a particular strike price K in a given cross section is converted into δ_K through its corresponding implied volatility σ_K on the volatility smile. We call it "smile conversion". Bliss and Panigirtzoglou (2004) suggested an alternative way. That is to use a single at-the-money implied volatility to convert all strike prices in a given cross section. We define it as "point conversion". This is in fact accomplished through the following equation:

$$\delta_K = e^{-r^*T}\Phi\left(\frac{\ln S_0 - \ln K + \left(r - r^* + \frac{\sigma_A^2}{2}\right)T}{\sigma_A\sqrt{T}}\right) \tag{3.4}$$

where σ_A is the at-the-money volatility. Note that transforming each strike into a delta using the at-the-money implied volatility has the advantage that the ordering of deltas is always the same as that of the strikes. Panigirtzoglou and Skiadopoulos (2004) pointed out that using the implied volatilities that correspond to each strike could change the ordering in the delta space, in cases where steep volatility skews are observed. This would result in generating volatility smiles with artificially created kinks. As a result, they applied equation (3.4) to convert strikes in their study. In this study, we do the same.

3.2.3 Analytic CDF and PDF

In previous studies, once the natural spline function is fitted, a large number of equally δ-spaced points on the function are computed. These are then converted to equally K-spaced values in price/strike space. These in turn are used to compute the implied CDF or PDF numerically. See Bliss and Panigirtzoglou (2004) and Panigirtzoglou and Skiadopoulos (2004) for more details.

However, we note that in the case of "point conversion" the implied CDF and PDF can be evaluated analytically. Therefore, in what follows we improve the computational efficiency of the SML method by providing the analytic expression of the estimated option implied CDF and PDF for "point conversion". This constitutes another contribution of this study to the literature of RND estimation.

Let $\widehat{g}(\delta, \Theta)$ denote the fitted natural spline function in the implied volatility/delta

space. Note that by the construction of the natural spline, $\widehat{g}(\delta, \Theta)$ is a piecewise function of the following form:

$$\widehat{g}(\delta, \Theta) = \begin{cases} \widehat{g}_0(\delta) & \text{if} \quad \delta < \delta_1 \\ \widehat{g}_1(\delta) & \text{if} \quad \delta_1 \leq \delta < \delta_2 \\ \widehat{g}_2(\delta) & \text{if} \quad \delta_2 \leq \delta < \delta_3 \\ \vdots & \\ \widehat{g}_{n-1}(\delta) & \text{if} \quad \delta_{n-1} \leq \delta < \delta_n \\ \widehat{g}_n(\delta) & \text{if} \quad \delta \geq \delta_n \end{cases} \tag{3.5}$$

where $\widehat{g}_i(\delta)$ is a third degree polynomial defined by

$$\widehat{g}_i(\delta) = a_i(\delta - \delta_i)^3 + b_i(\delta - \delta_i)^2 + c_i(\delta - \delta_i) + d_i$$

for $i = 0, 1, 2, \ldots n$, where n is the number of strikes in a particular cross section. The first and second derivatives of these n equations are

$$\widehat{g}_i'(\delta) = 3a_i(\delta - \delta_i)^2 + 2b_i(\delta - \delta_i) + c_i$$

$$\widehat{g}_i''(\delta) = 6a_i(\delta - \delta_i) + 2b_i$$

For any terminal asset price $S_T = K$, at which the option implied CDF or PDF is to be evaluated, the corresponding delta point, δ_K, in the delta space can be directly calculated through equation (3.4). The fitted implied volatility σ_K is then evaluated at δ_K by the spline function $\widehat{g}_i(\delta)$ in (3.5) for $\delta_i \leq \delta_K < \delta_{i+1}$. Finally, the fitted European call option price, $\widehat{C}(K)$, can be calculated by substituting both K and σ_K into the Black-Scholes call option pricing formula. Under this framework, the fitted

European call option pricing function implied by "point conversion" can be written as

$$\widehat{C}(K) = e^{-r^*T} S_0 \Phi(d_1) - e^{-rT} K \Phi(d_2)$$

where

$$d_1 = \frac{\ln S_0 - \ln K + \left(r - r^* + \frac{\sigma_K^2}{2}\right) T}{\sigma_K \sqrt{T}}$$

$$d_2 = \frac{\ln S_0 - \ln K + \left(r - r^* - \frac{\sigma_K^2}{2}\right) T}{\sigma_K \sqrt{T}}$$

$$\sigma_K = \widehat{g}_i(\delta_K)$$

$$\delta_K = e^{-r^*T} \Phi(d_A)$$

$$d_A = \frac{\ln S_0 - \ln K + \left(r - r^* + \frac{\sigma_A^2}{2}\right) T}{\sigma_A \sqrt{T}}$$

The cumulative distribution function (CDF) is obtained by differentiating $\widehat{C}(K)$ once with respect to K, which gives

$$\frac{\partial \widehat{C}(K)}{\partial K} = e^{-rT} \left\{ S_0 e^{(r-r^*)T} \sqrt{T} \Phi'(d_1) \frac{\partial \sigma_K}{\partial \delta_K} \frac{\partial \delta_K}{\partial K} - \Phi(d_2) \right\} \tag{3.6}$$

where

$$\frac{\partial \sigma_K}{\partial \delta_K} = \widehat{g}_i'(\delta_K)$$

$$\frac{\partial \delta_K}{\partial K} = -\frac{e^{-r^*T} \Phi'(d_A)}{K \sigma_A \sqrt{T}}$$

with $\Phi'(\cdot)$ being the standard normal probability density function. It follows from equation (3.1) that the implied CDF is given by

$$F(S_T) = -\frac{S_0 e^{(r-2r^*)T} \widehat{g}_i'(\delta_K) \Phi'(d_1) \Phi'(d_A)}{K \sigma_K} - \Phi(d_2) + 1 \Big|_{K=S_T}$$

which can be directly evaluated at any terminal asset price S_T.

The probability density function (PDF) is obtained as follows. To simplify the notations, we first let

$$\begin{aligned} A &= \Phi'(d_1) \\ B &= \frac{\partial \sigma_K}{\partial \delta_K} \\ C &= \frac{\partial \delta_K}{\partial K} \end{aligned}$$

and rewrite equation (3.6) as

$$\frac{\partial \widehat{C}(K)}{\partial K} = e^{-rT}\left\{S_0 e^{(r-r^*)T}\sqrt{T}ABC - \Phi(d_2)\right\} \tag{3.7}$$

Note that A, B, C and $\Phi(d_2)$ are all functions of K. Differentiating (3.7) once more gives

$$\frac{\partial^2 \widehat{C}(K)}{\partial K^2} = e^{-rT}\left\{S_0 e^{(r-r^*)T}\sqrt{T}\left(A'BC + AB'C + ABC'\right) - \Phi'(d_2)\frac{\partial d_2}{\partial K}\right\}$$

It can be easily verified that

$$A' = -Ad_1\frac{\partial d_1}{\partial K}$$

$$B' = \widetilde{g}_i''(\delta_K)C$$

$$C' = e^{-r^*T}\Phi'(d_A)\frac{\sigma_A\sqrt{T} - d_A}{K^2\sigma_A^2 T}$$

$$\frac{\partial d_1}{\partial K} = -\frac{1}{K\sigma_K\sqrt{T}} + \left(\sqrt{T} - \frac{d_1}{\sigma_K}\right)BC$$

$$\frac{\partial d_2}{\partial K} = \frac{\partial d_1}{\partial K} - \sqrt{T}BC$$

It follows from equation (3.2) that the implied PDF, $f(S_T)$, is given by

$$f(S_T) = S_0 e^{(r-r^*)T}\sqrt{T}\left(A'BC + AB'C + ABC'\right) - \Phi'(d_2)\left.\frac{\partial d_2}{\partial K}\right|_{K=S_T} \tag{3.8}$$

We apply this analytic formula in our study.

3.3 Density Functionals Based on Confluent Hypergeometric Functions

3.3.1 The Hypergeometric Functions

Abadir and Rockinger (2003) proposed a semi-nonparametric approach for estimating density related functionals without prior knowledge of the density's functional form. It is based on a special class of transcendental functions, known as hypergeometric functions. Let $\mathbb{N}$ denote a set of natural numbers and $\mathbb{R}$ a set of real numbers. The indicator function is written as $1_{\mathcal{K}}$, returning 1 if condition $\mathcal{K}$ is satisfied and 0 otherwise. The (complete) gamma function is denoted by $\Gamma(\nu)$ for $\nu \in \mathbb{R}$, and defining

$$(a)_j \equiv (a)(a+1)\ldots(a+j-1) = \frac{\Gamma(a+j)}{\Gamma(a)}$$

leads to the generalized hypergeometric function

$${}_pF_q\left(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;z\right) \equiv \sum_{j=0}^{\infty} \frac{\prod_{k=1}^{p}(\alpha_k)_j}{\prod_{k=1}^{q}(\beta_k)_j}\frac{z^j}{j!} \tag{3.9}$$

where $-\beta \notin \mathbb{N} \cup \{0\}$. A special case of particular interest is obtained when $p = q = 1$ in (3.9). It is called Kummer's function,

$$ {}_1F_1 \equiv \sum_{j=0}^{\infty} \frac{(\alpha)_j}{(\beta)_j} \frac{z^j}{j!} \equiv 1 + \frac{\alpha}{\beta} z + \frac{\alpha(\alpha+1)}{\beta(\beta+1)} \frac{z^2}{2} + \cdots , $$

also known as a confluent hypergeometric function. The ${}_1F_1$ can be used to represent a variety of density-related functions[3]. Special cases of interest are

$$ \gamma(\nu, z) \equiv \int_0^z e^{-x} x^{\nu-1} dx \equiv \frac{z^\nu}{\nu} {}_1F_1(\nu; \nu+1; -z), \qquad -\nu \notin \mathbb{N} \cup \{0\} $$

$$ \Phi(z) \equiv \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \equiv \frac{1}{2} + \frac{z}{\sqrt{2\pi}} {}_1F_1\left(\frac{1}{2}; \frac{3}{2}; -\frac{z^2}{2}\right) \equiv \frac{1}{2} + \frac{\operatorname{sgn}(z)}{\sqrt{2\pi}} \gamma\left(\frac{1}{2}, \frac{z^2}{2}\right) $$

where $\gamma(\cdot,\cdot)$ is the incomplete-gamma function, $\Phi(z)$ is the standard normal cumulative distribution function, and $\operatorname{sgn}(z)$ is the sign function.

3.3.2 Density Functionals for Option Pricing

The functionals in the context of option pricing are based on a couple of confluent hypergeometric functions. We call it the density functionals based on the confluent hypergeometric functions (DFCH). In contrast to the SML technique, the DFCH method is an approximating function method applied to the option prices. It specifies the European call pricing function as a mixture of two confluent hypergeometric functions:

$$ \begin{aligned} C(K) \equiv\ & c_1 + c_2 K + 1_{K>m_1} a_1 (K - m_1)^{b_1} {}_1F_1\left(a_2; a_3; b_2 (K - m_1)^{b_3}\right) \\ & + (a_4) {}_1F_1\left(a_5; a_6; b_2 (K - m_2)^2\right) \end{aligned} \tag{3.10} $$

[3]The usefulness of the Kummer's function in econometrics and dynamic economics is elaborated in Abadir (1999).

where $-a_3, -a_6 \notin \mathbb{N} \cup \{0\}, b_2, b_4 \in \mathbb{R}_-$. The indicator function is required to represent a component of the density with bounded support. It is also sufficient for keeping the function real-valued for general b_1 and b_3. It can be shown that the first ${}_1F_1$ function in $C(\cdot)$ covers the double integrals of the gamma and other asymmetric generalizations and the second covers the double integrals of a family of symmetric quadratic exponential densities such as the normal. Thus, the DFCH approach encompasses many known distributions in statistics and their mixtures. Examples of special cases giving integrals of known density functions include

$$\begin{aligned}
\text{Gamma:} \quad & a_1 = \frac{(-b_2)^{b_1-1}}{\Gamma(b_1+1)}, \quad a_2 = b_1 - 1, \quad a_3 = b_1 + 1, \quad a_4 = 0, \quad b_3 = 1; \\
\text{Inverse gamma:} \quad & a_1 = \frac{-(-b_2)^{1-b_1}}{b_1 \Gamma(2-b_1)}, \quad a_2 = -b_1, \quad a_3 = 2 - b_1, \quad a_4 = 0, \quad b_3 = -1; \\
\text{Weibull:} \quad & a_1 = -1, \quad a_2 = \frac{1}{b_3}, \quad a_3 = \frac{1}{b_3} + 1, \quad a_4 = 0, \quad b_1 = 1; \\
\text{Normal:} \quad & a_1 = 0, \quad a_4 = \frac{1}{2\sqrt{-b_4 \pi}}, \quad a_5 = -\frac{1}{2}, \quad a_6 = \frac{1}{2}; \\
\text{Pareto:} \quad & a_1 = \frac{m_1}{a_2}, \quad a_3 = -m_1 b_2 \text{ with } b_2 \to \infty, \quad a_4 = 0, \quad b_1 = 0, \quad b_3 = 1;
\end{aligned}$$

where standardization (e.g., centering around zero) is not imposed and the constants of integration c_1 and c_2 are to be determined by the problem at hand.

Differentiating (3.10) once[4] gives the implied cumulative distribution function:

$$\begin{aligned}
F(S_T) &= e^{rT} \left. \frac{\partial C(K)}{\partial K} \right|_{K=S_T} + 1 \\
&= e^{rT} \Big\{ c_2 + 1_{K>m_1} a_1 (K - m_1)^{b_1 - 1}
\end{aligned}$$

[4]An important feature of the ${}_1F_1$ function is that iterated integrals and derivatives of ${}_1F_1$ gives mixtures of ${}_1F_1$, which makes it a natural tool to model option prices and, more generally, functionals of densities.

$$
\begin{aligned}
&\times \left[(b_1)\ {}_1F_1\left(a_2; a_3; b_2 (K-m_1)^{b_3}\right) \right. \\
&\quad + \frac{a_2}{a_3} b_2 b_3 (K-m_1)^{b_3} \\
&\quad \left. \times\ {}_1F_1\left(a_2+1; a_3+1; b_2 (K-m_1)^{b_3}\right) \right] \\
&+2a_4 \frac{a_5}{a_6} b_4 (K-m_2) \\
&\times {}_1F_1\left(a_5+1; a_6+1; b_4 (K-m_2)^2\right) \Bigg\} \Bigg|_{K=S_T} + 1
\end{aligned}
$$

and the implied risk neutral probability density function is given by

$$
\begin{aligned}
f(S_T) &\equiv e^{rT} \left. \frac{\partial^2 C(K)}{\partial K^2} \right|_{K=S_T} \\
&= e^{rT} \left\{ 1_{K>m_1} a_1 (K-m_1)^{b_1-2} \left[b_1 (b_1-1)\, {}_1F_1\left(a_2; a_3; b_2 (K-m_1)^{b_3}\right) \right. \right. \\
&\qquad + \frac{a_2}{a_3} b_2 b_3 (2b_1 + b_3 - 1) (K-m_1)^{b_3} \\
&\qquad \times {}_1F_1\left(a_2+1; a_3+1; b_2 (K-m_1)^{b_3}\right) \\
&\qquad \times \frac{a_2 (a_2+1)}{a_3 (a_3+1)} b_2^2 b_3^2 (K-m_1)^{2b_3} \\
&\qquad \left. \left. \times {}_1F_1\left(a_2+2; a_3+2; b_2 (K-m_1)^{b_3}\right) \right] \right. \\
&\quad +2a_4 \frac{a_5}{a_6} b_4 \left[{}_1F_1\left(a_5+1; a_6+1; b_4 (K-m_2)^2\right) \right. \\
&\qquad + 2\frac{a_5+1}{a_6+1} b (K-m_2)^2 \\
&\qquad \left. \left. \times {}_1F_1\left(a_5+2; a_6+2; b_4 (K-m_2)^2\right) \right] \right\} \Bigg|_{K=S_T}
\end{aligned}
$$

Abadir and Rockinger (2003) showed that subject to these functions being nondegenerate (i.e., the existence condition for the moments), the explicit characterization

of the moments of the implied RND is given by

$$\begin{aligned} E(K^n) &= \int_{K_l}^{K_u} K^n d(F(K)) \\ &= \int_{K_l}^{K_u} K^n d\frac{\partial C(K)}{\partial K} \\ &= \left(K^n \frac{\partial C(K)}{\partial K} - nK^{n-1} C(K) \right) \Big|_{K_l}^{K_u} + n(n-1) \int_{K_l}^{K_u} K^{n-2} C(K)\, dK \end{aligned}$$

which can be obtained by integrating by parts two times.

Given observations of call and put prices, the parameters of the implied PDF can be estimated using several different methods which include maximum likelihood (ML), generalized least squares (GLS), generalized methods of moments (GMM), and so on. In this study, we use non-linear optimization methods[5] to minimize the sum of squared fitted pricing errors.

$$\min_{\Theta} \left\{ \sum_{i=1}^{N} w_i \left[C(K_i) - \widehat{C}(K_i|\Theta) \right] \right\}^2 \tag{3.11}$$

where as before w_i represents the relative weights placed on each observation.

It is important to note that not all the parameters in (3.10) are free to vary unrelatedly. For the function to be the integral of a CDF, at least three restrictions in general and possibly seven in the problem at hand can be imposed on the parameter space[6]. The restrictions imposed by Abadir and Rockinger (2003) are given by (3.12)-(3.16) which include the martingale condition in (3.16). As a result, the actual number

[5] Abadir and Rockinger (2003) proved that the nonlinear LS estimators are consistent and asymptotically normal for any of the parameters that have a nonzero impact on the function.

[6] See Abadir and Rockinger (2003) for derivations.

of parameters to be estimated is reduced to seven.

$$a_5 = -\frac{1}{2}, a_6 = \frac{1}{2}, b_1 = 1 + a_2 b_3 \tag{3.12}$$

$$c_1 = -c_2 m_2 \tag{3.13}$$

$$c_2 = -e^{-rT} + a_4 \sqrt{-b_4 \pi} \tag{3.14}$$

$$a_4 = \frac{1}{2\sqrt{-b_4 \pi}} \left[e^{-rT} - a_1 (-b_2)^{-a_2} \frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)} \right] \tag{3.15}$$

$$E(z) = e^{rT} a_1 \frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)} (-b_2)^{-a_2} (m_1 - m_2) + m_2 \tag{3.16}$$

Chapter 4

Monte Carlo Experiments

In this chapter, we conduct two complementary Monte Carlo experiments based on pseudo-prices method to examine the ability of the two methods for recovering option implied RNDs. In the first experiment, the true RNDs are generated by the Heston (1993) stochastic volatility model and in the second experiment we specify the true RND as a mixture of three lognormals. Details of the two experiments including the simulation results are presented in Section 4.1 and Section 4.2, respectively. Section 4.3 gives a summary of this chapter.

4.1 Monte Carlo Experiment Based on the Heston Model

4.1.1 The Heston Stochastic Volatility Model

As pointed out previously, a good RND estimation technique should be able to recover the true RNDs whatever the complexity of their shapes. Therefore, for comparison purposes the choice of the true RNDs in a simulation should itself be able to take on a wide range of different shapes reflecting various empirical features of asset distributions[7]: high or low volatility, positive or negative skewness, excess kurtosis; and cater for the full range of maturities that are encountered in practice. To generate risk-neutral densities that incorporate these features, we follow Cooper (1999) and use Heston (1993) stochastic volatility model to generate true RNDs and fitted option prices. These fitted prices are to be used by alternative methods to recover the true RNDs. Under the Heston model, the underlying asset price dynamics are described by the following stochastic differential equations:

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dz_1 \\ dv_t &= \kappa\left(\theta - v_t\right) d_t + \sigma_v \sqrt{v_t} dz_2 \end{aligned} \tag{4.1}$$

Here the volatility of the underlying asset $\sqrt{v_t}$ is also stochastic. The conditional variance v_t follows a mean reverting process such that the volatility mean-reverts to a

[7]Note that option implied RNDs are risk-neutral and thus different from the empirical asset distributions. But one can justifiably suppose a rough similarity between the risk-neutral and the objective distributions. See Rubinstein (1994).

long run of $\sqrt{\theta}$ at a rate dictated by κ. The term σ_v sets the volatility of the volatility. The two Wiener process dz_1 and dz_2 have a correlation given by ρ. By changing the correlation parameter we can generate skewness in asset returns[8]. Heston shows that the European call option price on an asset that behaves according to (4.1) has a closed form solution, which is given by

$$C\left(S_0, v_0, K, T\right) = S_0 e^{-r^* T} P_1 - K e^{-rT} P_2 \tag{4.2}$$

where for $j = 1, 2$

$$P_j\left(\ln\left(S_0\right), v_0, T; \ln\left(K\right)\right) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{e^{-i\phi \ln(K)} f_j\left(\ln\left(S_0\right), v_0, T; \phi\right)}{i\phi}\right] d\phi$$

$$f_i\left(\ln\left(S_0\right), v_0, T; \phi\right) = e^{C(T;\phi) + D(T;\phi) v_0 + i\phi \ln(S_0)}$$

$$C\left(T; \phi\right) = \left(r - r^*\right) \phi i T + \frac{a}{\sigma_v^2}\left\{\left(b_j - \rho\sigma_v \phi i + d\right) T - 2 \ln\left[\frac{1 - g e^{dT}}{1 - g}\right]\right\}$$

$$D\left(T; \phi\right) = \frac{b_j - \rho\sigma_v \phi i + d}{\sigma_v^2}\left[\frac{1 - e^{dT}}{1 - g e^{dT}}\right]$$

$$g = \frac{b_j - \rho\sigma_v \phi i + d}{b_j - \rho\sigma_v \phi i - d}$$

$$d = \sqrt{\left(\rho\sigma_v \phi i - b_j\right)^2 - \sigma_v^2\left(2u_j \phi i - \phi^2\right)}$$

$$u_1 = \frac{1}{2}, u_2 = -\frac{1}{2}, a = \kappa\theta, b_1 = \kappa + \lambda - \rho\sigma_v, b_2 = \kappa + \lambda$$

As in Cooper (1999), we test performance across a range of six scenarios, which correspond to combinations of low and high volatility and three levels of skewness. We then generate European-style call and put option prices with 4 different contract maturities from 2 weeks up to 6 months. Thus, a total of 24 different pairs of scenario

[8]See Cooper (1999) for an illustration of the effect of ρ on the implied RNDs.

Table 4.1: Model Parameters under Each Scenario

	Strong Negative Skew		Strong Positive Skew
	Scenario 1	Scenario 2	Scenario 3
Low Volatility	$\kappa = 2, \sqrt{\theta} = 0.1$	$\kappa = 2, \sqrt{\theta} = 0.1$	$\kappa = 2, \sqrt{\theta} = 0.1$
	$\sigma_v = 0.1, \rho = -0.9$	$\sigma_v = 0.1, \rho = 0$	$\sigma_v = 0.1, \rho = 0.9$
	Scenario 4	Scenario 5	Scenario 6
High Volatility	$\kappa = 2, \sqrt{\theta} = 0.3$	$\kappa = 2, \sqrt{\theta} = 0.3$	$\kappa = 2, \sqrt{\theta} = 0.3$
	$\sigma_v = 0.4, \rho = -0.9$	$\sigma_v = 0.4, \rho = 0$	$\sigma_v = 0.4, \rho = 0.9$

and maturity are generated. The Heston model parameters used for each scenario are set out in Table 4.1. These are chosen to generate true RNDs that correspond to situations of negative skewness, and weak and strong positive skewness in the terminal asset price and also conditions of low and high volatility. To generate these levels of skewness in the terminal asset price distributions, we use three different values for the correlation parameter -0.9, 0 and 0.9. The long run volatilities for the high volatility scenarios are chosen on the basis of the levels of implied volatility typically observed within equity markets. The low volatility scenarios are used to mimic data from Stock Index, FX and interest rate markets.

It is important to note that the Heston model is used here simply as a convenient tool to generate underlying RNDs that incorporate the empirical features discussed above, namely, different levels of spread, skewness and excess kurtosis of the implied RNDs. Doing so does not presume that equation (4.1) correctly describes the asset price dynamics in the real world. For this reason, it is innocuous to assume for simplicity that the market price of volatility risk is zero and that the time zero

conditional volatility is equal to the long run volatility, as long as the resulting RNDs serve our purpose.

To obtain the true RND for each scenario and maturity pair, we generate a large number of theoretical prices of Heston call options using the pricing formula in equation (4.2) and calculate the RND numerically by applying equation (3.2). Table 4.2 presents some descriptive statistics for the true RNDs used in this experiment. For simplicity, the mean of the true RNDs are set equal to 100. As can be seen from the table, our true RNDs take on a wide range of different shapes, with the standard deviations ranging from 2.038 (small spread) to 23.060 (large spread), skewness from -0.474 (large negative skew) to 1.964 (large positive skew), and kurtosis from 2.770 (thin tails) to 10.847 (fat tails). Specifically, scenario 1, 2, and 3 are low volatility cases and scenario 4, 5, and 6 are high volatility cases; Scenario 1 and 4 represent strong negative skewness, while others are positive cases. As we would expect, all the three measures including the kurtosis increase with time-to-expiry.

For each of the 24 cases generated above, we compute theoretical option prices at a number of different strikes. We assume that we observe enough strikes which span a sufficient range of the true RND[9]. We construct strike interval equal to 1 for low volatility scenarios and 5 for high volatility scenarios, respectively. The final number of strikes for each cross section is set out in Table 4.3. It can be seen that the numbers of strikes used in this study reflect the real world situation in the following

[9]Specifically, the strike range is constructed to just cover the area between the 1st percentile and the 99th percentile.

Table 4.2: Descriptive Statistics of the True RNDs

	Scenario	2 weeks	1 month	3 months	6 months
Mean	1	100.000	100.000	100.000	100.000
	2	100.000	100.000	100.000	100.000
	3	100.000	100.000	100.000	100.000
	4	100.000	100.000	100.000	100.000
	5	100.000	100.000	100.000	100.000
	6	100.000	100.000	100.000	100.000
	Scenario	2 weeks	1 month	3 months	6 months
Std Dev	1	2.038	2.877	4.956	6.965
	2	2.041	2.887	5.003	7.081
	3	2.045	2.898	5.052	7.200
	4	6.085	8.555	14.529	20.127
	5	6.130	8.677	15.094	21.491
	6	6.175	8.802	15.702	23.060
	Scenario	2 weeks	1 month	3 months	6 months
Skewness	1	-0.206	-0.281	-0.418	-0.474
	2	0.062	0.089	0.159	0.231
	3	0.331	0.459	0.743	0.956
	4	-0.172	-0.229	-0.304	-0.275
	5	0.188	0.273	0.505	0.762
	6	0.551	0.781	1.362	1.964
	Scenario	2 weeks	1 month	3 months	6 months
Kurtosis	1	3.045	3.082	3.180	3.222
	2	3.046	3.088	3.223	3.356
	3	3.178	3.346	3.931	4.602
	4	2.983	2.966	2.888	2.770
	5	3.135	3.270	3.821	4.678
	6	3.532	4.081	6.487	10.847

Table 4.3: Number of Strikes under Each Scenario

Scenario	2 Weeks	1 Month	3 Months	6 Months
1	12	16	24	34
2	11	15	26	36
3	12	15	26	36
4	7	10	15	20
5	8	10	17	23
6	8	11	17	25

two respects. Firstly, they are increasing in time-to-expiry. Secondly, they are close to the actual numbers of strikes one may observe in reality for corresponding maturities.

4.1.2 Error Specification

As discussed earlier, accuracy and stability are both desirable properties of a good RND estimator. To test the robustness of alternative methods to small errors embedded in option prices, we add noise ε_i to the theoretical prices computed above. Noise ε_i is introduced to model observational errors that arise from market imperfections such as nonsynchronicity, bid-ask spread, and discreteness, etc. See Bliss and Panigirtzoglou (2002) for a discussion on potential sources of errors in option prices.

The error specification used in this study was first introduced by Bondarenko (2003). It focuses on pricing errors resulting from the bid-ask spread and is constructed based on the following reasoning. An empiricist observes both bid and ask quotes, q_i^b and q_i^a, for put options with strikes $\{K_i\}$. This empiricist uses the midpoint quote $P_i = 0.5(q_i^b + q_i^a)$ as an approximation to the true price. As a result, the introduced measurement error ε_i is assumed to be uniformly distributed on $[-0.5s_i, 0.5s_i]$, where s_i is the bid-ask spread, i.e., $s_i = q_i^b - q_i^a$. The value s_i depends on the strike K_i, which is larger for in-the-money options and smaller for out-of-the-money options. For example, the CBOE rules state that the maximum bid-ask spread is $\frac{1}{4}$ for options with bid quote q^b below than \$2, $\frac{3}{8}$ for bid quotes between \$2 and \$5, $\frac{1}{2}$ for bid quotes between \$5 and \$10, $\frac{3}{4}$ for bid quotes between \$10 and \$20, and 1 for bid quotes

above \$20. Such information can be used to construct a function $M(q)$ to represent the maximum spread for the quote q. Specifically, let

$$M(0) = \frac{1}{8}, M(2) = \frac{1}{4}, M(5) = \frac{3}{8}, M(10) = \frac{1}{2}, M(20) = \frac{3}{4}, M(q) = 1, q \geq 50$$

and $M(q)$ is linearly interpolated for all other $q \in [0, 50]$.

On the other hand, the empiricist also observes quotes for call options. Since in practice out-of-the-money options are more liquid, they have smaller spreads than in-the-money options. It is assumed that the empiricist would use more accurate options for estimating RNDs. In particular, for large strikes the put-call parity relation is used to convert more accurate out-of-the-money call prices into the corresponding put prices. In other words, for the strike K_i, the relevant spread is the minimum of the spreads for put P_i and call $C_i = S_t + P_i - K_i$. In addition, if we assume that the actual spread s_i is proportional to the maximum spread permitted by the exchange, we can eventually write

$$s_i = c \min(M(P_i), M(S_t + P_i - K_i)),$$

with constant $c \in [0, 1]$.

The advantages of such specification for ε_i are that noise is smaller in the absolute terms but larger in the relative terms for far-from-the-money strikes. The presence of the scale constant c allows us to proportionally increase or decrease the level of noise across all strikes. This is particularly important in the present study. Note that such error specification does not guarantee nonnegative option prices after perturba-

tion for some deep out-of-the-money options. Recall that the SML method requires option prices to be first converted to the implied volatilities, which are not defined for negative option prices. Therefore, failure in monitoring the nonnegativity of the option prices will lead to failure of the SML method. Since the scale parameter c allows us to control the size of the noises across strikes, an obvious solution to this problem is to chose an arbitrarily small value of c. Nevertheless, too small noises may invalidate the test for the robustness. In order to reconcile these two, we select the maximum possible value of c that still guarantees nonnegativity of the option prices after perturbation.

4.1.3 The Root Mean Integrated Squared Error

It is difficult to present and compare more than a few RNDs in the same graph. Therefore, we analyze the perturbed-price RNDs by examining certain summary statistics. Note that both the accuracy and the stability are important properties of a good estimator. To compare these two different RND estimators under both accuracy and stability considerations, we focus on the criterion of root mean integrated squared error (RMISE). If $\widehat{f}(S_T)$ is the RND estimator of the true RND $f(S_T)$, then the RMISE is defined as

$$\mu\left(\widehat{f}\right) := \text{RMISE}\left(\widehat{f}\right) = \sqrt{E\left[\int_{-\infty}^{\infty}\left(\widehat{f}(S_T) - f(S_T)\right)^2 dS_T\right]} \tag{4.3}$$

It is convenient to represent RMISE as

$$\mu^2\left(\widehat{f}\right) = \mu_1^2\left(\widehat{f}\right) + \mu_2^2\left(\widehat{f}\right)$$

$$\mu_1\left(\widehat{f}\right) := \text{RISB}\left(\widehat{f}\right) = \sqrt{\int_{-\infty}^{\infty}\left(E\left[\widehat{f}\left(S_T\right)\right] - f\left(S_T\right)\right)^2 dS_T} \tag{4.4}$$

$$\mu_2\left(\widehat{f}\right) := \text{RIV}\left(\widehat{f}\right) = \sqrt{\int_{-\infty}^{\infty} E\left[\left(\widehat{f}\left(S_T\right) - E\left[\widehat{f}\left(S_T\right)\right]\right)^2\right] dS_T} \tag{4.5}$$

where RISB is the root integrated squared bias and RIV is the root integrated variance. We define the RMISE as our measure of the overall quality of the estimator, RISB as our measure of the accuracy, and RIV as our measure of the stability. The representation allows us to study the relative contributions of the bias μ_1 and the variability μ_2 to the RMISE of different methods. For each cross section, we repeat the procedure of shocking the prices and then fitting the RND for 100 times. The RMISE, RISB and RIV are then obtained by applying equation (4.3) to (4.5).

4.1.4 Results

To examine the impact of the choice of weighting schemes on the relative performance of the two methods, we apply three weighting schemes in the estimation. With decreasing relative weights on near-the-money options, these three weighting schemes are option vega weighting, equal weighing, and inverse variance weighting. The RMISE, RISB and RIV results for both methods under the three weighting schemes are presented in Table 4.4, 4.5 and 4.6, respectively. The results for the SML method are displayed on the left panel. Recall from (3.3) that for the SML method

the smoothing parameter λ is a free parameter which allows the user to control the trade-off between the smoothness and the goodness of fit. In this study, we search for the optimal parameter λ that minimizes the RMISE. It is important to note that this is only possible in simulation studies where the true RND is known. In real world where the true RND is unknown the smoothing parameter has to be selected by the user. Different values of λ will result in different RNDs. In the following comparison, we use the minimum RMISE for the SML[10]. For the DFCH method, however, the RMISE result is unique for each cross section.

Examining the results in Table 4.4, which corresponds to option vega weighting, we find that in 18 out of 24 cases the DFCH provides lower RMISE than that of the SML, indicating better overall quality of the DFCH as an RND estimator. Specifically, for scenario 1 and 4, which represent those important negative skewness cases[11], the DFCH dominates the SML method across all maturities by a substantial margin: The RMISE for SML are from 138% to 453% larger than that for DFCH.

For the 6 cases where the DFCH underperforms the SML, the differences in RMISE are relatively small with the largest being 74% for the scenario 6 - 6 month maturity case. As far as the shape of the true RND is concerned, scenario 2 and 5 represent very small skewness, especially for short maturities. On the other hand, for the two long maturity cases in scenario 6, it may not be too unrealistic to assume that such large positive skewness in asset prices distribution are not as often observed as the

[10]We believe this is biased in favour of the SML method, because the RMISE's would have been larger if we chose the λ as if we did not know the actual RNDs.

[11]Negative skewness is an important empirical feature in financial asset return distributions.

Table 4.4: RMISE, RISB and RIV Results Based on the Heston Model (Vega Weighting)

	SML					DFCH				
	Scenario	2 weeks	1 month	3 months	6 months	Scenario	2 weeks	1 month	3 months	6 months
	1	0.0065	0.0067	0.0090	0.0085	1	0.0019	0.0013	0.0038	0.0029
	2	0.0022	0.0023	0.0028	0.0033	2	0.0034	0.0029	0.0023	0.0030
RMISE	3	0.0058	0.0078	0.0082	0.0085	3	0.0031	0.0050	0.0044	0.0063
	4	0.0045	0.0056	0.0074	0.0069	4	0.0018	0.0017	0.0031	0.0012
	5	0.0011	0.0024	0.0025	0.0029	5	0.0012	0.0038	0.0021	0.0028
	6	0.0047	0.0052	0.0076	0.0038	6	0.0022	0.0038	0.0091	0.0148
	Scenario	2 weeks	1 month	3 months	6 months	Scenario	2 weeks	1 month	3 months	6 months
	1	0.0057	0.0063	0.0082	0.0080	1	0.0010	0.0010	0.0026	0.0025
	2	0.0013	0.0017	0.0023	0.0027	2	0.0009	0.0010	0.0006	0.0013
RISB	3	0.0055	0.0070	0.0077	0.0080	3	0.0029	0.0036	0.0043	0.0063
	4	0.0036	0.0050	0.0066	0.0066	4	0.0004	0.0008	0.0010	0.0008
	5	0.0009	0.0018	0.0022	0.0027	5	0.0008	0.0009	0.0015	0.0018
	6	0.0046	0.0052	0.0068	0.0029	6	0.0022	0.0038	0.0090	0.0148
	Scenario	2 weeks	1 month	3 months	6 months	Scenario	2 weeks	1 month	3 months	6 months
	1	0.0031	0.0023	0.0037	0.0029	1	0.0017	0.0008	0.0027	0.0015
	2	0.0018	0.0016	0.0016	0.0019	2	0.0033	0.0028	0.0022	0.0027
RIV	3	0.0017	0.0035	0.0028	0.0031	3	0.0010	0.0035	0.0007	0.0009
	4	0.0028	0.0026	0.0035	0.0020	4	0.0017	0.0015	0.0030	0.0010
	5	0.0007	0.0015	0.0011	0.0012	5	0.0009	0.0037	0.0014	0.0022
	6	0.0004	0.0002	0.0032	0.0024	6	0.0001	0.0000	0.0013	0.0001

Table 4.5: RMISE, RISB and RIV Results Based on the Heston Model (Equal Weighting)

	SML					DFCH				
	Scenario	2 weeks	1 month	3 months	6 months	Scenario	2 weeks	1 month	3 months	6 months
	1	0.0064	0.0067	0.0089	0.0086	1	0.0016	0.0012	0.0033	0.0026
	2	0.0021	0.0021	0.0026	0.0030	2	0.0029	0.0023	0.0018	0.0022
RMISE	3	0.0060	0.0077	0.0081	0.0083	RMISE 3	0.0035	0.0052	0.0045	0.0071
	4	0.0046	0.0057	0.0073	0.0070	4	0.0013	0.0011	0.0024	0.0012
	5	0.0011	0.0022	0.0023	0.0028	5	0.0012	0.0034	0.0021	0.0025
	6	0.0044	0.0055	0.0073	0.0038	6	0.0023	0.0037	0.0102	0.0173
	Scenario	2 weeks	1 month	3 months	6 months	Scenario	2 weeks	1 month	3 months	6 months
	1	0.0057	0.0065	0.0082	0.0082	1	0.0009	0.0010	0.0025	0.0023
	2	0.0012	0.0015	0.0022	0.0025	2	0.0012	0.0008	0.0006	0.0007
RISB	3	0.0058	0.0070	0.0074	0.0076	RISB 3	0.0033	0.0042	0.0043	0.0070
	4	0.0039	0.0052	0.0066	0.0069	4	0.0004	0.0004	0.0008	0.0008
	5	0.0009	0.0016	0.0022	0.0026	5	0.0008	0.0010	0.0018	0.0020
	6	0.0044	0.0055	0.0066	0.0031	6	0.0023	0.0037	0.0101	0.0172
	Scenario	2 weeks	1 month	3 months	6 months	Scenario	2 weeks	1 month	3 months	6 months
	1	0.0028	0.0018	0.0033	0.0023	1	0.0013	0.0007	0.0022	0.0012
	2	0.0017	0.0015	0.0014	0.0017	2	0.0026	0.0021	0.0017	0.0021
RIV	3	0.0015	0.0034	0.0033	0.0034	RIV 3	0.0011	0.0031	0.0011	0.0012
	4	0.0025	0.0022	0.0032	0.0016	4	0.0013	0.0011	0.0023	0.0008
	5	0.0006	0.0015	0.0008	0.0009	5	0.0008	0.0033	0.0012	0.0015
	6	0.0003	0.0001	0.0032	0.0022	6	0.0004	0.0002	0.0016	0.0015

Table 4.6: RMISE, RISB and RIV Results Based on the Heston Model (Inverse Variance Weighting)

	SML					DFCH				
	Scenario	2 weeks	1 month	3 months	6 months	Scenario	2 weeks	1 month	3 months	6 months
	1	0.0063	0.0067	0.0087	0.0085	1	0.0016	0.0012	0.0033	0.0026
	2	0.0021	0.0021	0.0025	0.0029	2	0.0029	0.0025	0.0018	0.0022
RMISE	3	0.0060	0.0076	0.0084	0.0088	RMISE 3	0.0035	0.0053	0.0055	0.0079
	4	0.0046	0.0057	0.0073	0.0075	4	0.0013	0.0012	0.0022	0.0012
	5	0.0010	0.0020	0.0024	0.0028	5	0.0011	0.0033	0.0023	0.0029
	6	0.0044	0.0057	0.0069	0.0037	6	0.0024	0.0040	0.0122	0.0213
	Scenario	2 weeks	1 month	3 months	6 months	Scenario	2 weeks	1 month	3 months	6 months
	1	0.0057	0.0065	0.0082	0.0083	1	0.0009	0.0010	0.0024	0.0023
	2	0.0012	0.0014	0.0022	0.0025	2	0.0013	0.0008	0.0006	0.0005
RISB	3	0.0058	0.0069	0.0083	0.0087	RISB 3	0.0033	0.0043	0.0054	0.0078
	4	0.0041	0.0054	0.0068	0.0074	4	0.0005	0.0004	0.0008	0.0009
	5	0.0009	0.0015	0.0023	0.0028	5	0.0009	0.0010	0.0020	0.0026
	6	0.0043	0.0057	0.0061	0.0032	6	0.0024	0.0040	0.0121	0.0212
	Scenario	2 weeks	1 month	3 months	6 months	Scenario	2 weeks	1 month	3 months	6 months
	1	0.0027	0.0016	0.0029	0.0019	1	0.0013	0.0007	0.0022	0.0012
	2	0.0017	0.0015	0.0013	0.0014	2	0.0026	0.0024	0.0017	0.0021
RIV	3	0.0015	0.0032	0.0010	0.0010	RIV 3	0.0011	0.0030	0.0012	0.0011
	4	0.0021	0.0019	0.0026	0.0011	4	0.0012	0.0011	0.0020	0.0008
	5	0.0006	0.0014	0.0006	0.0006	5	0.0007	0.0031	0.0012	0.0013
	6	0.0003	0.0001	0.0031	0.0018	6	0.0004	0.0005	0.0020	0.0019

rest.

Investigation of the RISB reveals that the DFCH is often less biased than the SML, suggesting the flexibility of the confluent hypergeometric functions. The RIV show that the SML is relatively stable for cases where the true RND is not strongly skewed.

Figure 4.1 gives an example of the differences between the two methods. For each method, the 100 RND estimates are plotted against the true RND. It can be seen that the SML is significantly biased, particularly on the left tail of the distribution. It also shows relatively larger variations in the center of the distribution. In contrast, the DFCH fits the true RND fairly well and exhibits relatively smaller variations.

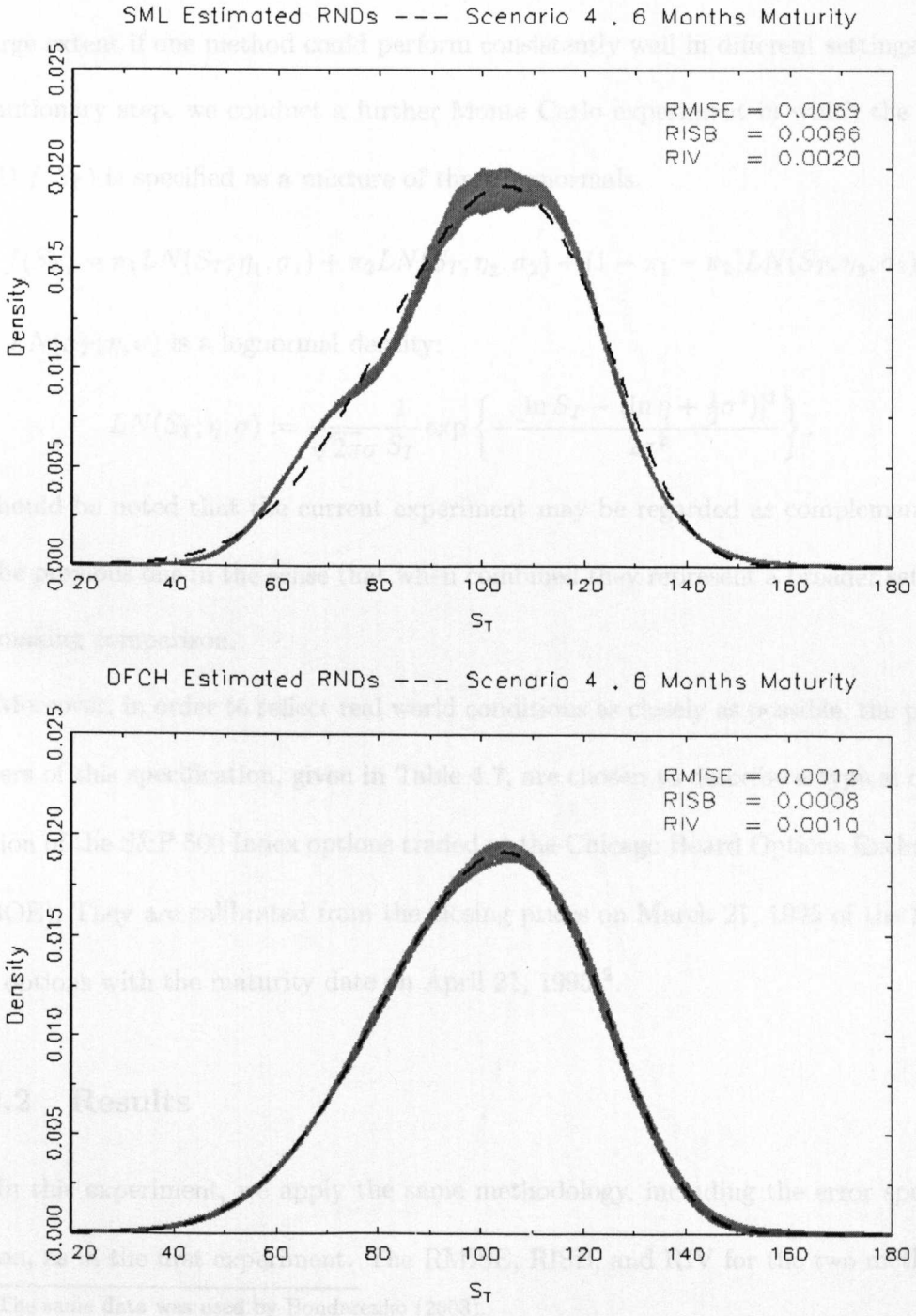
Table 4.5 and 4.6 provide the simulation results from the other two weighting schemes. As expected, because the fitted price errors are generally small, the weights used to multiply them have little impact on the estimation.

4.2 Monte Carlo Experiment Based on Mixture of Lognormals

4.2.1 The Mixture of Three Lognormals

A criticism over such simulation approaches as above is that the performance of a particular RND estimating method may be related to the choice of the true RNDs.

Figure 4.1: True RND and Estimated RNDs Based on the Heston Model



Results obtained from estimating RNDs of some particular functional form may not be generalized to RNDs outside the set examined. Such concern would be eased to a large extent if one method could perform consistently well in different settings. As a cautionary step, we conduct a further Monte Carlo experiment in which the true RND $f(S_T)$ is specified as a mixture of three lognormals.

$$f(S_T) = \pi_1 LN(S_T; \eta_1, \sigma_1) + \pi_2 LN(S_T; \eta_2, \sigma_2) + (1 - \pi_1 - \pi_2) LN(S_T; \eta_3, \sigma_3)$$

and $LN(S_T; \eta, \sigma)$ is a lognormal density:

$$LN(S_T; \eta, \sigma) := \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{S_T} \exp\left\{ -\frac{[\ln S_T - (\ln \eta + \frac{1}{2}\sigma^2)]^2}{2\sigma^2} \right\}.$$

It should be noted that the current experiment may be regarded as complementary to the previous one in the sense that when combined they represent a broader setting for making comparison.

Moreover, in order to reflect real world conditions as closely as possible, the parameters of this specification, given in Table 4.7, are chosen to describe a typical cross section of the S&P 500 Index options traded at the Chicago Board Options Exchange (CBOE). They are calibrated from the closing prices on March 21, 1995 of the S&P 500 options with the maturity date on April 21, 1995[12].

4.2.2 Results

In this experiment, we apply the same methodology, including the error specification, as in the first experiment. The RMISE, RISB, and RIV for the two methods

[12]The same data was used by Bondarenko (2003).

Figure 4.2: True RND and Estimated RNDs Based on Three Lognormals

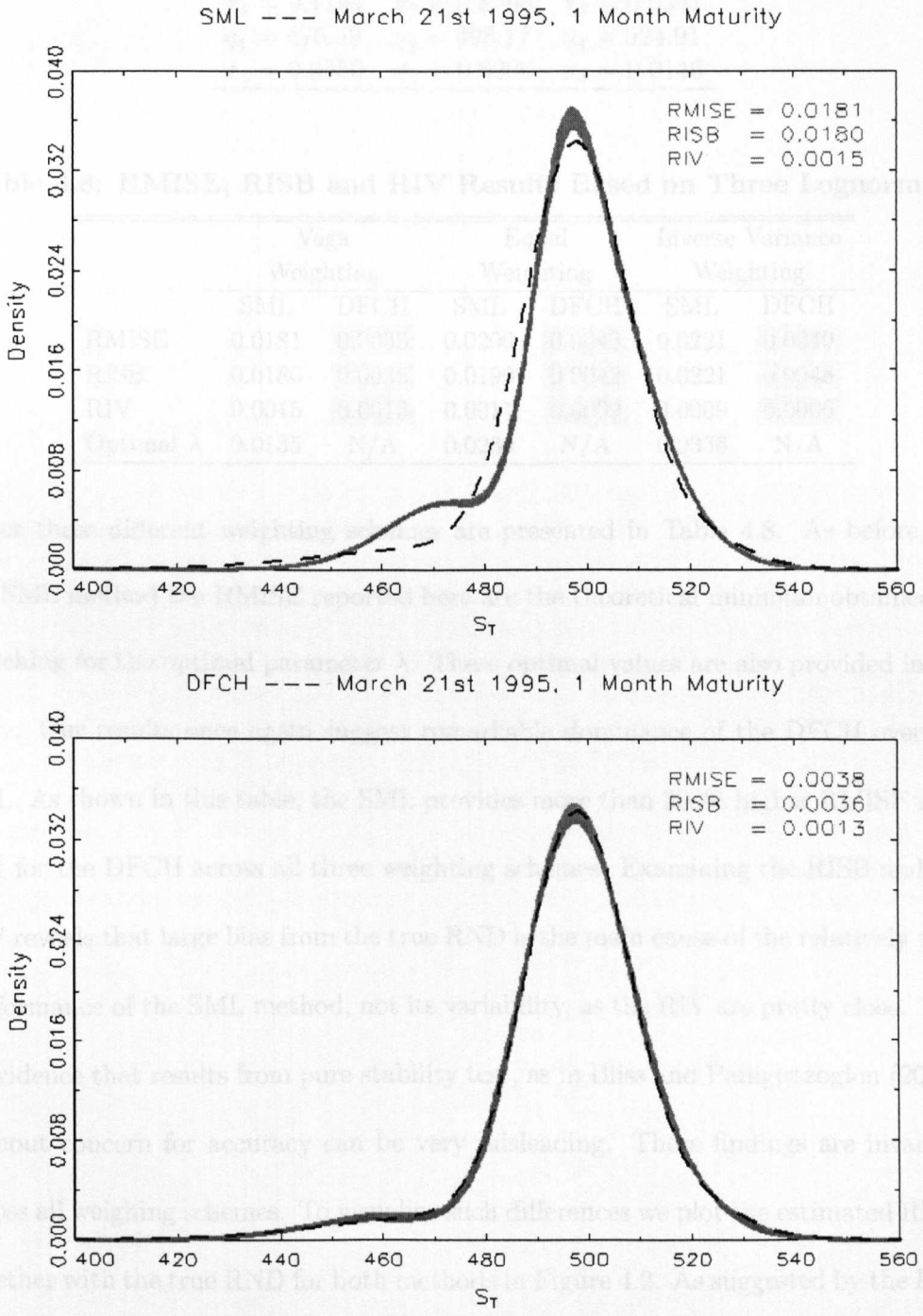


Table 4.7: The Fitted Parameters

$\pi_1 = 0.1194$	$\pi_2 = 0.8505$	$\pi_3 = 0.0301$
$\eta_1 = 475.59$	$\eta_2 = 498.17$	$\eta_3 = 524.91$
$\sigma_1 = 0.0550$	$\sigma_2 = 0.0206$	$\sigma_3 = 0.0146$

Table 4.8: RMISE, RISB and RIV Results Based on Three Lognormals

	Vega Weighting		Equal Weighting		Inverse Variance Weighting	
	SML	DFCH	SML	DFCH	SML	DFCH
RMISE	0.0181	0.0038	0.0200	0.0043	0.0221	0.0049
RISB	0.0180	0.0036	0.0199	0.0042	0.0221	0.0048
RIV	0.0015	0.0013	0.0011	0.0009	0.0009	0.0006
Optimal λ	0.0135	N/A	0.0236	N/A	0.0336	N/A

under three different weighting schemes are presented in Table 4.8. As before, for the SML method the RMISE reported here are the theoretical minimum obtained by searching for the optimal parameter λ. These optimal values are also provided in the table. Our results once again suggest remarkable dominance of the DFCH over the SML. As shown in this table, the SML provides more than 300% higher RMISE than that for the DFCH across all three weighting schemes. Examining the RISB and the RIV reveals that large bias from the true RND is the main cause of the relatively poor performance of the SML method, not its variability, as the RIV are pretty close. This is evidence that results from pure stability test, as in Bliss and Panigirtzoglou (2002), without concern for accuracy can be very misleading. These findings are invariant across all weighing schemes. To visualize such differences we plot the estimated RNDs together with the true RND for both methods in Figure 4.2. As suggested by the RIV,

both methods show similar level of variability. But whereas the SML is significantly biased, the DFCH recovers the true RND with superior precision, further evidence of the flexibility of the confluent hypergeometric functions.

4.3 Conclusion

In this chapter, we examined the ability of two methods for estimating option implied RNDs. Two complementary Monte Carlo experiments based on the pseudo-prices methodology have been conducted. In the first experiment, the Heston's stochastic volatility model were used as the benchmark model to generate true RNDs that represent various empirical features of asset distributions. The second experiment considered a mixture of three lognormals fitted from a typical cross section of S&P 500 Index option data as an alternative specification of the true RND. We compared the two methods by focusing on the RMISE criterion. Results from both experiments suggest strong evidence of the superiority of the DFCH method over the SML method under both accuracy and stability considerations. In particular, we found that the DFCH almost always more closely recovers the implied RNDs.

Chapter 5

Application to OTC Option Data

This chapter presents an application of the two RND estimation methods to a set of real option data. Section 5.1 gives some description of the data. In Section 5.2, we consider three optimization criteria for selecting the smoothing parameter for the SML method. Estimation results are discussed in Section 5.3. Section 5.4 summarizes this chapter.

5.1 The data

As an illustration of the two RND estimating methods in real settings, we apply both methods to OTC data[13] of European French franc/Deutsche mark (FF/DM) rate options of the two dates: 17 May 1996, a day when the exchange rate markets were known to be calm, and on 25 April 1997, a few days after the French President

[13]We thank Professor Karim Abadir and Professor Michael Rockinger for this data.

Chirac announced dissolution of the National Assembly, which implied nation-wide elections. This type of option is quoted in terms of delta. For both dates, we have at least information for options with delta taking the values 10, 15, 20, 30, 40, 50 (corresponding to the at the money option), 60, 70, 80, 85, 90. For the first date, we also have information for the 5 and 95 delta options. In this study we use data for all possible deltas. By using a numerical procedure we extracted for each option of a given maturity the corresponding strike price. The difference between the actual delta and the delta obtained for the optimal strike price is in all cases negligible. We have bid and ask prices for in-the-money put and call options. Following the literature, we decide to work with the average between the bid and ask prices. Even though we obtained all results for options with 1, 2, 3, 6, 9, and 12 month to maturity, we report the results for 1, 3 and 12 month maturity, representing short, medium and long horizons, respectively. The domestic (French) and foreign (German) Eurocurrency interest rates are chosen to match the expiration of the options. We transform these rates into their continuously compound equivalents. The spot exchange rate is easily available.

5.2 Selecting the Smoothing Parameter

As discussed earlier, for the SML method the presence of the smoothing parameter λ allows the user to control the trade-off between the goodness of fit and smoothness of the estimated RNDs. In the two Monte Carlo experiments conducted in the previous

chapter, because the true RND is known we were able to search for the optimal λ that minimizes the RMISE for each case. When dealing with real option data, the choice of λ has to be decided in some ad-hoc ways. A suitable smoothing parameter can be obtained by simply plotting the distribution for different smoothing parameters and choose the one which yields the "best" result. The main disadvantage of this method is that the shape of the estimated RND relies on one's subjective judgement. Therefore, two researchers may come up with different RNDs for the same data. Besides, this is a cumbersome method for studies where a large number of distributions need to be estimated.

Several procedures for automatically choosing an optimal smoothing parameter have been proposed in the spline regression literature. The most popular class of these methods is based on cross validation (CV) proposed by Craven and Wahba (1979). The basic principle of cross validation is to leave out the data points one at a time and to choose the smoothing parameter for which the missing data points are best predicted by the remainder of the data. More precisely, for a given smoothing parameter the observations are deleted one by one and a spline function is estimated in each case from the remaining observations. The sum of the squared errors between the deleted observations and the values generated by the spline functions is then calculated. The "optimal" smoothing parameter is the one that yields the smallest sum of squared errors. The simple CV criterion has been generalized so that it allows the user to reweight the contribution of deletion residuals to the total score. The

generalized cross validation method (GCV) can be written as:

$$\min_{\lambda} \sum_{i}^{N} w_i \left[\sigma_i - g_{\lambda}^{-i}\left(\delta_i, \Theta\right)\right]^2$$

where $g^{-i}(\cdot)$ is the fitted spline for a given λ with data point i omitted. Hence, this method finds an optimal λ by lowering the influence of outlying data points on the curve.

Another criterion of interest was suggested by Bliss and Panigirtzoglou (2002). It is to select the λ such that the maximum fitted price error is approximately equal to one half of a tick size[14]. They argued that by doing this one can effectively fit the data within the precision of option price measurement.

In addition, since the SML method is always capable of providing a perfect fit to the data, for comparison purposes, it is useful to see whether or not the SML method can lead to reasonable RNDs and at the same time gives as good a fit as the DFCH method does. Because the residual sum of squares is a monotonically decreasing function of the smoothing parameter λ, we are able to find the value of λ such that the two goodness of fit from the two methods are exactly equal.

We apply both RND estimating methods to the OTC currency data and the performance of the SML method is examined under all three smoothing parameter selection criteria discussed above.

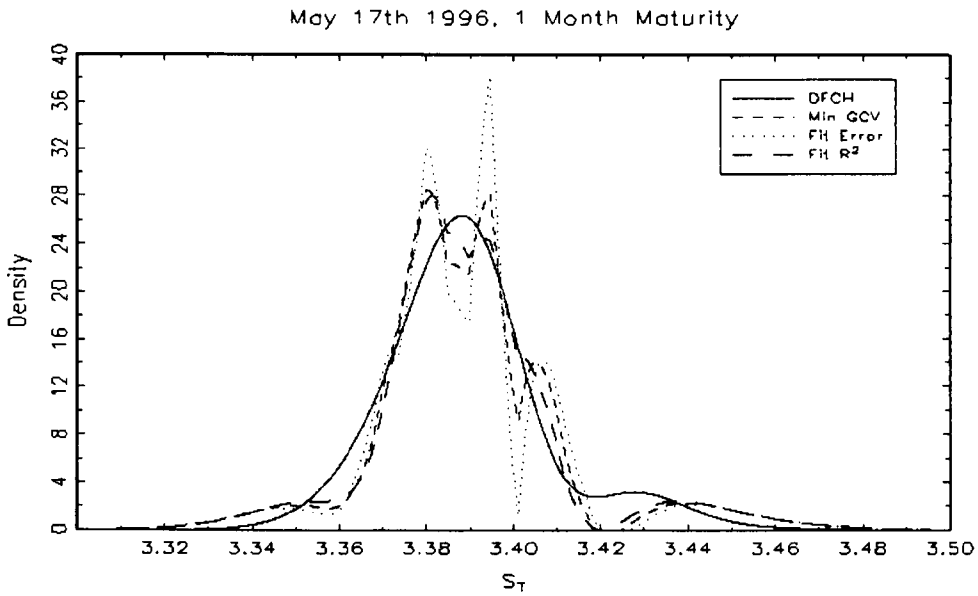
[14]Since the OTC currency options used in this study are quoted in volatilities, such criterion is implemented on volatility quotes. Accordingly, we use equal weighting to account for measurement errors in volatility quotes.

5.3 Results

Our results suggest that all three criteria provide too loose smoothing parameters and the resulting RNDs exhibit unreasonably large fluctuations across all maturities for both dates. As an example, Figure 5.1 shows the RNDs estimated from the DFCH method and the SML method with three different choices of λ for the 1 month maturity options on date 17 May 1996. The values of the λ selected by the three criteria are 1.0696×10^{-4}, 2.2451×10^{-5}, and 3.3606×10^{-4}, respectively.

The inconsistency of the GCV criterion indicates that even if the spline function is optimal according to the GCV procedure in the implied volatility/delta space, it is not necessary "optimal" after the transformation required to obtain the RND. The fact that the second criterion failed to provide reasonable result suggests that the real size of the pricing errors, at least in this market is far greater than that of those imposed by the discreteness of option quotes. In fact, implementing such criterion relies on the assumption on the size of the pricing error. As the real size of pricing errors is unknown, this criterion is arbitrary. Moreover, the smoothing parameter that provides the same goodness of fit is still too loose to generate reasonable RNDs. A much tighter λ is required to generate a plausible RND. This indicates that the SML method could not provide as good fit as the DFCH. In contrast, the nonparametric nature of the DFCH method enables it to give a high goodness of fit and at the same time the parametric property ensures that the estimated RNDs are proper density functions.

Figure 5.1: DFCH RND *vs* SML RNDs under Three λ's



The above results show that at least within the samples examined, we have found no objective ways for the choice of the smoothing parameter. In the end, we decide to select the value of λ relying on our visual inspection so that the RNDs estimated by the SML method are smooth enough[15] but with the best possible fit. The plot of estimated RNDs for the two dates and three maturities is reported in Figure 5.2 and Figure 5.3.

The implied RNDs can be described by computing a range of summary statistics. These measures are useful when analyzing changes in the shape of the implied RNDs. They are also used for comparing different estimation techniques[16]. To study the

[15] A tighter choice of λ will provide smoother RNDs, but this is at the cost of the goodness of fit.
[16] Examples are Cooper (1999) and Bliss and Panigirtzoglou (2002).

Figure 5.2: Estimated RNDs for the First Date

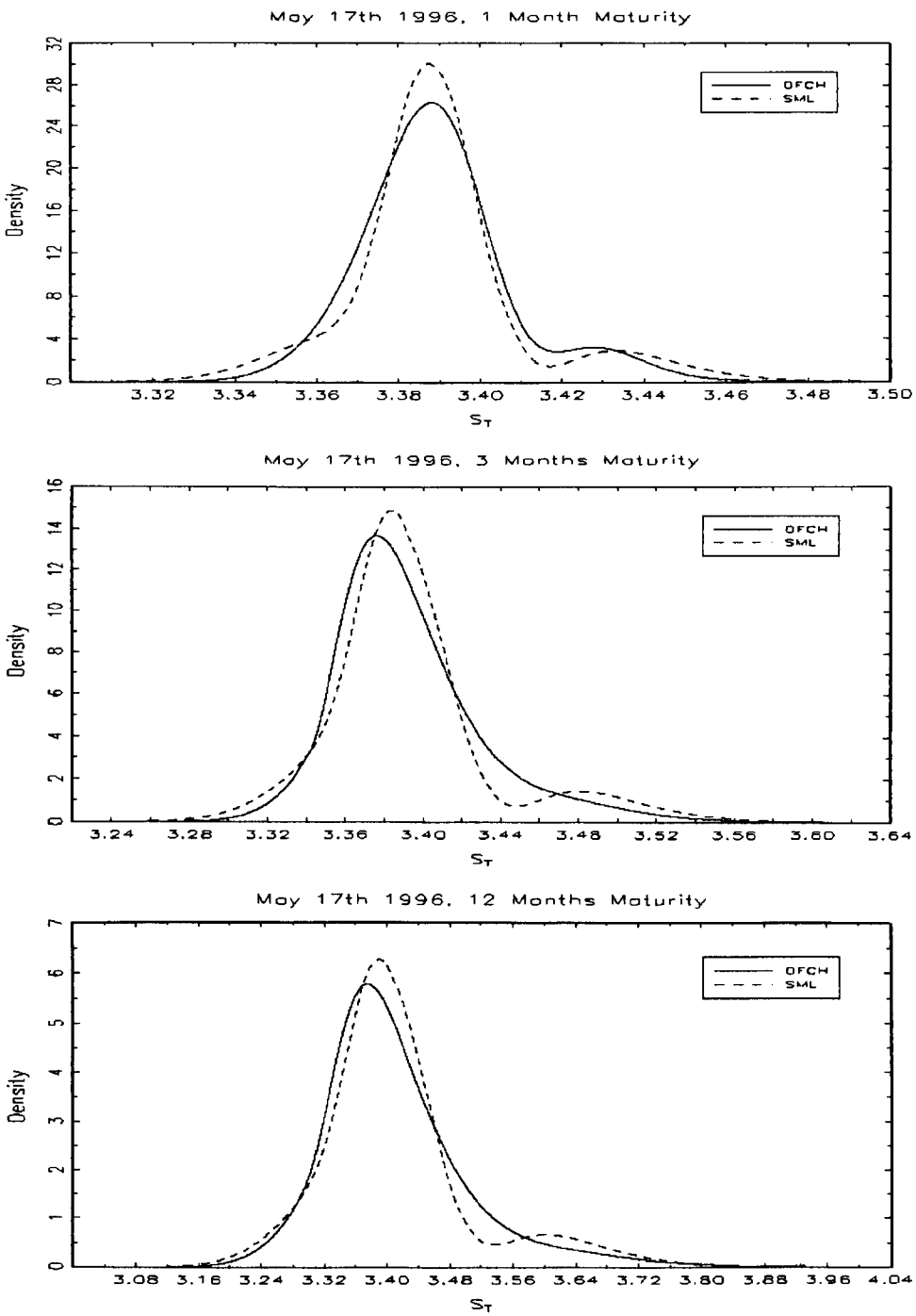


Figure 5.3: Estimated RNDs for the Second Date

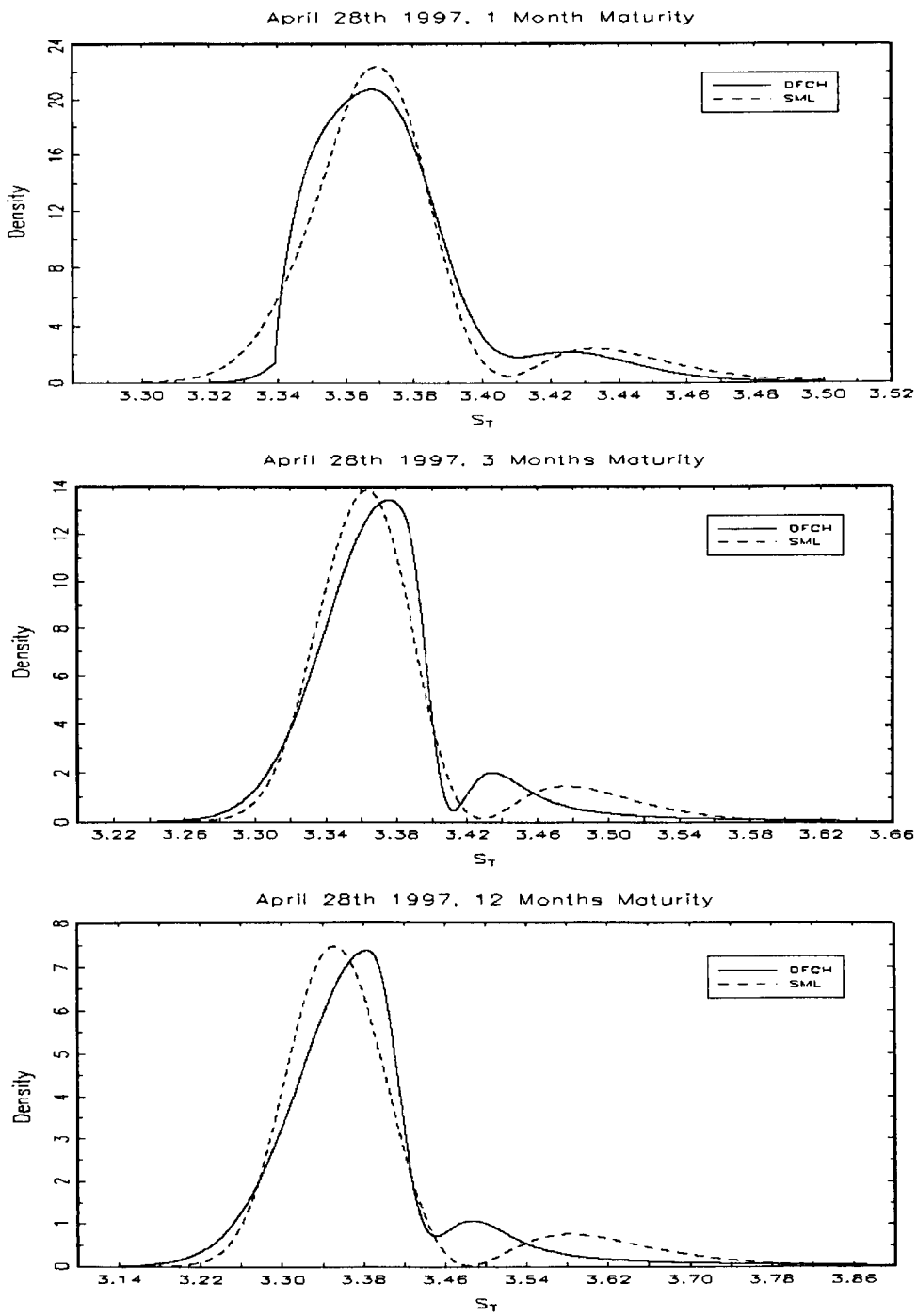


Table 5.1: RND Summary Statistics for OTC Currency Option Data

		17.05.96			28.04.97		
		1 M	3 M	12 M	1 M	3 M	12 M
$\widehat{\mu}$	SML	3.3896	3.3929	3.4121	3.3743	3.3761	3.3852
	DFCH	3.3896	3.3929	3.4121	3.3743	3.3761	3.3852
$\widehat{\sigma}$	SML	0.0222	0.0437	0.1012	0.0291	0.0507	0.1012
	DFCH	0.0193	0.0393	0.0951	0.0250	0.0444	0.0843
Skew	SML	0.8335	1.2072	1.1710	1.5190	1.7126	1.8315
	DFCH	0.6502	1.1511	1.1468	1.6009	1.4277	1.4264
Kurt	SML	5.2820	5.6405	5.3788	5.9971	6.1997	6.3860
	DFCH	4.2705	5.6707	5.5896	7.0460	8.1539	8.2997
R^2	SML	0.9967	0.9842	0.9365	0.9972	0.9883	0.9479
	DFCH	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
λ	SML	0.0035	0.0045	0.0045	0.0035	0.0150	0.0250

implied RNDs estimated by the two methods. We examine the following distribution summary statistics:

1. $\widehat{\mu}$: The Mean.

2. $\widehat{\sigma}$: The Standard Deviation.

3. *Skew*: The skewness coefficient defined as the third central moment normalized by the cube of the standard deviation.

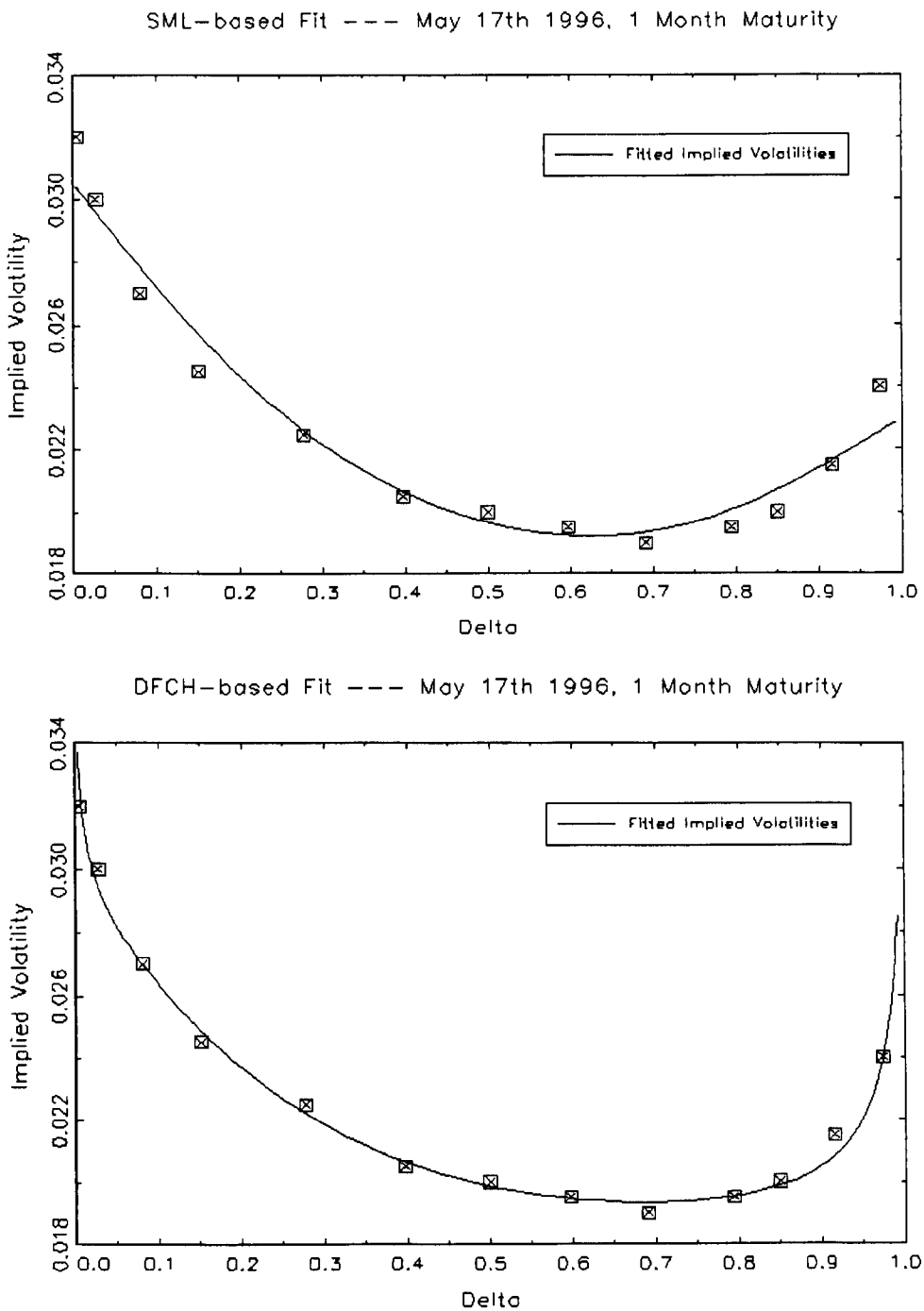
4. *Kurt*: The kurtosis coefficient defined as the fourth central moment normalized by the square of the variance.

The estimated distribution summary statistics, R^2 and the values of λ chosen in SML estimation are all presented in Table 5.1. The DFCH method provides very high values of R^2 across all estimations, indicating great flexibility of the confluent

hypergeometric functions. As shown earlier, a λ that provides the same goodness of fit is too loose to generate non-oscillating RNDs. Thus, the value of λ is manually selected to provide the best possible fit while ensuring minimum acceptable smoothness in the resulting RND. It turned out that the R^2 for the SML method under such a λ are lower by a sizable magnitude than that of the DFCH. Figure 5.4 displays the plot of the original volatility quotes and the fitted volatility function for the two methods, which is representative of the fit across all estimations we tried. We notice an excellent fit from the DFCH method as compared to the SML method, particularly for away-from-the-money options.

By construction the SML method will always fit the mean of the implied RND to the forward price. Thus, calculating the mean of the implied RND does not provide additional information. To make other distribution statistics more comparable, we impose the mean-forward equality as a constraint in the DFCH procedure. Therefore, for a particular date and maturity we obtain the same values of the mean from both methods. We notice that the differences in the standard deviations between the two methods are less remarkable than those of the two higher moments. The reason is that higher moments such as skewness and kurtosis are fairly sensitive to the tails of the distribution where observed option data provide little information. As discussed in Melick and Thomas (1998), there is an infinite variety of probability masses outside the strike range that can be consistent with the observed option prices. The allocation of tail probability mass is specific to each estimating technique. In this

Figure 5.4: Original Volatility Quotes and Fitted Volatility Functions



respect, previous results of stability test based on higher moments are questionable. Nevertheless, it is still worth mentioning that the DFCH provides much better fit for away-from-the-money options than that of the SML, as has been shown in Figure 5.4.

Interesting conclusion can be drawn from these estimated summary statistics. We find these statistics reflect the influence of major events upon this market. Comparing the mean for the two dates reveals an overall shift to the left for all maturities. This has come from the fact that the FF had appreciated against the DM. When we compare the standard deviations for the two dates, there is a larger spread for the implied RNDs at all maturities for the second date. It is an indication that for the second date there is a greater uncertainty among the markets participants about how the exchange rate will evolve towards maturity, following President Chirac's announcement of the dissolution of the National Assembly. Specifically, across all maturities the right tail of the RND decays more slowly for the second date, suggesting that the market is contemplating a nonnegligible probability of subsequent depreciation of the FF. Market participants' uncertainty about the exchange rate movement and fear of large price changes are also reflected by the large values of Kurtosis from both methods on both dates, as they are willing to pay a high premium for protection against such large price changes.

5.4 Conclusion

In this chapter, we carried out an application of the two methods to a set of OTC currency option data. For the SML method, we found that in the absence of the knowledge of the true RND, all three objective optimal λ selection criteria do not lead to reasonable RND estimates. Arbitrary choices of λ have to be used. In contrast, the DFCH method not only produced proper RND estimates but also provided remarkable goodness of fit. Based on the RNDs estimated from two different dates, we show how RNDs can be used to analyze the effects of major events on the market's expectations of future exchange rate movements.

Chapter 6

Conclusion

In this study, we compared the widely known SML method with a new semi-nonparametric DFCH method to estimate option implied RNDs. We conducted two Monte Carlo experiments based on the pseudo-prices methodology. This methodology consists of re-estimating implied RNDs from randomly perturbed cross sections of fitted theoretical option data based on presumed true RNDs. In the first experiment, the true RNDs were generated by the Heston (1993) stochastic volatility model. These RNDs were selected to represent various empirical features of asset distributions. In the second experiment, an alternative specification of the true RND was considered. It was based on a mixture of three lognormals, and the parameters of this specification were calibrated from a typical cross section of S&P 500 Index option data.

To compare the two RND estimating methods, we focused on the RMISE criterion, which is a measure of average distance between the true RND and the estimated ones.

Results from both experiments provide strong evidence of the superiority of the DFCH method over the SML method under both accuracy and stability considerations. In particular, the DFCH almost always more closely recovers the implied RNDs. We also found that these results are insensitive to the weighting schemes applied in the estimation.

We also applied the two methods to OTC currency option data. The statistical analysis conducted in our Monte Carlo experiments ignores the problem relative to the choice of the smoothing parameter in the SML method and assumes that the theoretically optimal λ is applied. In this empirical study, as the true RND is unknown, we tried three different objective choices of λ. We found that all three criteria failed to generate RNDs with reasonable shape. In particular, we found that in our examples the SML could not provide as good fit to the data as the DFCH while still generating non-oscillating RNDs. An arbitrary choice of λ was then used. The main disadvantage of this is that two researchers may come up with different RNDs for the same data! Based on the fitted RNDs for two distinctive dates, we showed how RNDs can be used to analyze the effects of major events on the market's expectations of future exchange rate movements.

The SML method is attractive as it is a practically very efficient way of extracting implied RNDs from option data, particularly with its analytic form of the PDF and CDF provided in this study. Moreover, previous research also suggested that it is both absolutely and relatively more robust to errors embedded in option prices than

the widely used mixture of lognormals technique. However, our study shows that the SML is outperformed by the semi-nonparametric DFCH method. The DFCH is a theoretically well-founded statistical density functional model. Because of its semi-nonparametric nature, the DFCH technique is less data-intensive than those fully nonparametric methods, and more flexible than purely parametric methods in generating abundant potential probability density shapes. Another potential advantage of the DFCH over the SML method is that we can test to which known distribution or mixture of distributions the estimated RND is not significantly different from.

References

[1] Abadir, K.M. (1999). An Introduction to Hypergeometric Functions for Economists. *Econometric Reviews* 18 (3), 287-330.

[2] Abadir, K.M. and Rockinger, M. (2003). Density Functionals, with an Option-Pricing Application. *Econometric Theory* 19, 778-811.

[3] Ait-Sahalia, Y. and Lo, A.W. (1998). Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices. *Journal of Finance* 53, 499-547.

[4] Ait-Sahalia, Y. and Lo, A.W. (2000). Nonparametric Risk Management and Implied Risk Aversion. *Journal of Econometrics* 94, 9-51.

[5] Ait-Sahalia, Y., Wang, Y. and Yared, F. (2001). Do Option Markets Correctly Price the Probabilities of Movement of the Underlying Asset? *Journal of Econometrics* 102, 67–110.

[6] Aparicio, S. and Hodges, S. (1998). Implied Risk-Neutral Distribution: A Comparison of Estimation Methods. Working Paper, Warwick University.

[7] Bahra, B. (1997). Implied Risk-Neutral Probability Density Functions from Option prices: Theory and Application. Working Paper, Bank of England.

[8] Bates, D. (1996). Jumps and Stochastic Volatility: Exchange Rate Process Implicit in Deutsche Mark Options. *Review of Financial Studies* 9, 69–107.

[9] Bliss, R. and Panigirtzoglou, N. (2002). Testing the Stability of Implied Probability Density Functions. *Journal of Banking and Finance* 26, 381-422.

[10] Bliss, R. and Panigirtzoglou, N. (2004). Option Implied Risk Aversion Estimates. *Journal of Finance* 59 (1), 407-446.

[11] Breeden, D.T. and Litzenberger, R.H. (1978). Prices of State-Contingent Claims Implicit in Option Prices. *Journal of Business* 51, 621-651.

[12] Bondarenko, O. (1997). Testing Rationality of Financial Markets: An Application to S&P 500 Index Options. Ph.D Thesis, California Institute of Technology.

[13] Bondarenko, O. (2003). Estimation of Risk-Neutral Densities Using Positive Convolution Approximation. *Journal of Econometrics* 116, 85-112.

[14] Bu, R. and Hadri, K. (2005). Estimating Option Implied Risk-Neutral Densities using Spline and Hypergeometric Functions. Working Paper, University of Liverpool.

[15] Buchen, P.W. and Kelly, M. (1996). The Maximum Entropy Distribution of an

Asset Inferred from Option Prices. *Journal of Finance and Quantitative Analysis* 31, 143-159.

[16] Campa, J.M., Chang, P.H.K. and Reider, R.L. (1998). Implied Exchange Rate Distributions: Evidence from OTC Option Markets. *Journal of International Money and Finance* 17, 117-160.

[17] Cooper, N. (1999). Testing Techniques for Estimating Implied RNDs from the Prices of European and American Options. Working Paper, Bank of England.

[18] Craven, P. and Wahba, G. (1979). Smoothing Noisy Data with Spline Functions. *Numerische Mathemaik* 31, 377-403.

[19] Garcia, R. and Gencay, R. (2000). Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint. *Journal of Econometrics* 94, 93–115.

[20] Gottschling, A., Haefke, C. and White, H. (2000). Closed Form Integration of Artificial Neural Networks with Some Applications to Finance. Working Paper, University of California at San Diego.

[21] Heston, S. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies* 6, 327-343.

[22] Jackwerth, J.C. (1999). Implied Binomial Trees: A Literature Review. *Journal of Derivatives* 7, 66-82.

[23] Jackwerth, J.C. (2000). Recovering Risk Aversion from Option Prices and Realized Returns. *Review of Financial Studies* 13, 433–451.

[24] Jarrow, R. and Rudd, A. (1982). Approximate Option Valuation for Arbitrary Stochastic Processes. *Journal of Financial Economics* 10, 347–369.

[25] Jondeau, E. and Rockinger, M. (2000). Reading the Smile: The Message Conveyed by Methods Which Infer Risk Neutral Densities. *Journal of International Money and Finance* 19, 885-915.

[26] Lim, G.C., Martin, G.M. and Martin, V.L. (2005). Parametric Pricing of Higher Order Moments in S&P 500 Options. *Journal of Applied Econometrics* 20, 377-404.

[27] Malz, A.M. (1997a). Option-Implied Probability Distributions and Currency Excess Returns. Staff Reports, Federal Reserve Bank of New York.

[28] Malz, A.M. (1997b). Estimating the Probability Distribution of the Future Exchange Rate from Options Prices. *Journal of Derivatives* 5 (2) 18-36.

[29] McManus, D.J. (1999). The Information Content of Interest Rate Futures Options. Working Paper, Bank of Canada.

[30] Melick, W.R. and Thomas, C.P. (1997). Recovering an Asset's Implied PDF from Option Prices: An Application to Crude Oil during the Gulf Crisis. *Journal of Financial and Quantitative Analysis* 32, 91-115.

[31] Melick, W.R. and Thomas, C.P. (1998). Confidence Intervals and Constant Maturity Series for Probability Measures Extracted from Option Prices. Paper Presented at the Bank of Canada Conference on Information Contained in Prices of Financial Assets.

[32] Panigirtzoglou, N. and Skiadopoulos, G. (2004). A New Approach to Modeling the Dynamics of Implied Distribution: Theory and Evidence from the S&P 500 Options. *Journal of Banking and Finance* 28, 1499-1520.

[33] Rosenberg, J. (1998). Pricing Multivariate Contingent Claims using Estimated Risk-Neutral Density Functions. *Journal of International Money and Finance* 17, 229-247.

[34] Rosenberg, J. and Engle, R. (2002). Empirical Pricing Kernels. *Journal of Financial Economics* 64, 341-372.

[35] Rubinstein, M. (1994). Implied Binomial Trees. *Journal of Finance* 49, 771-818.

[36] Rubinstein, M. (1998). Edgeworth Binomial Trees. *Journal of Derivatives* 5 (3), 20 27.

[37] Shimko, D.C. (1993). Bounds of Probability. *Risk* 6, 33-37.

[38] Söderlind, P. (2000). Market Expectations in the UK before and after the ERM Crisis. *Economica* 67,1-18.

[39] Söderlind, P. and Svensson, L.E.O. (1997). New Techniques to Extract Market Expectations from Financial Instruments. *Journal of Monetary Economics* 40, 383-429.

[40] Stutzer, M. (1996). A Simple Nonparametric Approach to Derivative Security Valuation. *Journal of Finance* 51, 1633–1652.

[41] Waggoner, D.F. (1997). Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices. Working Paper, Federal Reserve Bank of Atlanta.

[42] Wu, L. and Huang, J. (2004). Specification Analysis of Option Pricing Models Based on Time-Changed Levy Processes. *Journal of Finance* 59, 1405-1439.

Part II

Maximum Likelihood Estimation of Higher-Order Integer-Valued Autoregressive Processes

Chapter 7

Introduction

Recently there has been growing interest in modelling time series of small counts that arise in various fields of statistics. Examples include the number of customers waiting to be served at a counter recorded at discrete points in time; the monthly cases of rare infectious diseases in a specified area; the monthly number of claimants collecting wage loss benefit for injuries in the workplace, and so on. Typically, such time series take on only small non-negative integer values and exhibit short-range dependence. Traditional continuous variable models are apparently inappropriate in that they would invariably produce non-integer forecast values. As a result, some specific class of time series models has to be entertained to explicitly account for the discreteness. This part of the thesis is concerned with a special class of observation-driven models called integer-valued autoregressive (INAR(p)) models introduced independently by Al-Osh and Alzaid (1987) and McKenzie (1988). The INAR(p) model

not only specifies the dependence structure of the observations but also allows for a wide class of discrete marginal distributions, which are jointly determined by the distributions of the innovation sequence and the thinning operators.

Estimation of the INAR(p) process can be carried out in a variety of ways. These include the moments based Yule-Walker (YW) estimation method and the conditional least squares (CLS) estimation method of Klimko and Nelson (1978). The implementation of both approaches is relatively simple and they are asymptotically equivalent. A recent study which involves the estimation of the INAR model is provided by Jung and Tremayne (2006), where they considered the estimation of an INAR(2) model using the method of moments. Al-Osh and Alzaid (1987) showed that maximum likelihood (ML) can be implemented for estimating the parameters of the INAR(1) model when Binomial thinning is used and the innovation sequence is assumed to be Poisson. They compared the finite sample properties of the three estimation methods and concluded that the ML is worth the extra calculation because of the gain in terms of the bias and the mean squared error (MSE). Freeland and McCabe (2004a) considered the conditional maximum likelihood (CML) estimation of the INAR(1) model and derived new expressions for the score and information matrix. A general test for model specification based on information matrix equality was also proposed. However, both works of Al-Osh and Alzaid (1987) and Freeland and McCabe (2004a) are confined to the first-order model and assume only Binomial thinning operator and Poisson innovations. The main objective of this study is to extend previous works

and develop a general framework for ML estimation of higher-order INAR models with general thinning operators and innovation distributions.

On the other hand, one of the main objectives of modelling time series data is to produce forecasts for various purposes. Freeland and McCabe (2004b) suggested that for count data models forecasts be provided for each point mass of the distribution, using the median as coherent point forecast, and that the probabilities associated with each point mass be modified to reflect the variation in parameter estimation. The reason for doing this is that the estimated point mass forecasts are more informative than those supplied by single statistics, such as mean, median and mode, of the forecast distributions. Following the same ideas, McCabe and Martin (2005) explored the issue of coherent forecasting for a class of INAR models under the Bayesian framework. The disadvantage is that only first-order INAR models are concerned and their method is based on computer intensive numerical evaluation. Jung and Tremayne (2006) recently considered coherent forecasting for higher-order INAR models, but their method is based on Monte Carlo experiments which also requires considerable computational work. The second objective of this study is to extend the ideas of Freeland and McCabe (2004b) and develop an efficient procedure for producing coherent forecasts with higher-order INAR models.

The main contributions of this study can be summarized as follows:

A generalized version of the INAR(p) model is suggested, which encompasses the widely used Binomial (thinning)-Poisson (Innovations) specification. A recursive

representation of the conditional (transition) probability function (which serves as the basis of ML estimation) of the generalized INAR(p) model is also proposed, which greatly improves the efficiency of computation of these probabilities and substantially facilitates the derivation of the score functions and the Fisher information matrix of the model. We show that when certain conditions on the distributions of the thinning process and innovation sequence are satisfied, all elements of the score and Fisher information matrix can be represented in terms of conditional expectations, which enhance the interpretation of these quantities and lead to new definitions for the residuals of the INAR(p) model. More specific details on ML estimation of the Binomial-Poisson INAR(p) model are provided, including the asymptotic distribution of the ML estimator.

The second contribution of this study is the comparison of alternative estimation methods. Asymptotic relative efficiency (ARE) of the ML estimator (MLE) in relation to the YW estimator (YWE) and the CLS estimator (CLSE) are examined. Our results confirm that the newly proposed MLE is asymptotically more efficient than the YWE and the CLSE. We also find that the magnitude of efficiency gain is most substantial when the values of the thinning coefficients are large. Finite sample performance of the three estimators are also investigated. Results from our Monte Carlo experiments indicate that there is also a potential gain in implementing the ML in small samples in terms of bias and mean squared error (MSE), especially when the processes considered have high degrees of persistence.

Another contribution of this study is to the field of forecasting count data. Built on the theory of stationary higher-order Markov Chain processes, a new approach for producing distribution forecasts for higher-order INAR models is developed and a procedure based on the δ-method for calculating confidence intervals for these forecast probabilities is also suggested. Since the new method is based on the transition matrix method, no restrictions on the innovation distributions are needed. Most importantly, it is computationally efficient.

Also in this study, we carry out an empirical analysis of the Westgren (1916) gold particle data under the ML framework developed in this study. We show that in the light of the likelihood method, new weapons of statistical inferences can be used, and as a result new evidence has emerged regarding the suitability of the Binomial assumption of the thinning process in the fitted model. Forecasts are also produced for the Westgren data based on the new method. We find that the benefit of implementing the method, in terms of the enriched information and the improved efficiency, is substantial.

The rest of Part II of the thesis is organized as follows: Chapter 8 reviews the INAR(p) model and estimation methods. Chapter 9 looks at the likelihood estimation of a generalized INAR(p) model. In Chapter 10, we examine the relative performance of alternative estimators. Chapter 11 is concerned with forecasting with INAR(p) models. The empirical study is presented in Chapter 12. Chapter 13 concludes. Proofs and other details are contained in a set of appendices.

Chapter 8

Review of the INAR(p) Model and Estimation Methods

In this chapter, we present a brief review of the INAR(p) model of Du and Li (1991) and revisit three estimation methods that have been used in the literature. Section 8.1 sets out the details of the Du and Li (1991) type INAR(p) model and briefly discusses its main statistical properties. In Section 8.2, we review the three methods for estimating INAR models. The first is the moments based Yule-Walker estimation method and the second is the conditional least squares estimation method of Klimko and Nelson (1978). Both methods can be used for estimating higher-order INAR models. The maximum likelihood method proposed by Al-Osh and Alzaid (1987) for estimating the first-order INAR model is also discussed.

8.1 The INAR(p) Model

Du and Li (1991) define the INAR(p) model to be

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \cdots + \alpha_p \circ X_{t-p} + \varepsilon_t \tag{8.1}$$

where the innovation process $\{\varepsilon_t\}$ is an i.i.d process which is assumed to be independent of all thinning operations $\alpha_k \circ X_{t-k}$ for $k = 1, 2, \ldots, p$, which are in turn conditionally independent. The "$\circ$" is the thinning operator. Conditional on X_{t-k} it is defined as

$$\alpha_k \circ X_{t-k} = \sum_{i=1}^{X_{t-k}} \mathrm{B}_{i,k}$$

where each collection $\{\mathrm{B}_{i,k}, i = 1, 2, \ldots, X_{t-k}\}$ consists of independently distributed random variables (taken here to be Bernoulli) with parameter α_k and the collections are mutually independent for $k = 1, 2, \ldots, p$. Intuitively, $\alpha_k \circ X_{t-k}$ is the number of individuals that would independently survive a Binomial experiment in a given period, where each of the X_{t-k} individuals has identical surviving probability α_k. The case where $p = 1$, $\{\varepsilon_t\}$ is Poisson and $\mathrm{B}_{i,1}$ is Bernoulli is known as Poisson autoregression, henceforth denoted as PoINAR(1), since in this case the marginal distribution of X_t is also Poisson. When $p > 1$ and $\{\varepsilon_t\}$ is Poisson, it can be shown that the unconditional mean of X_t and the unconditional variance of X_t are generally not equal and so that the marginal distribution of X_t is no longer Poisson even though the innovations are. Dion et al. (1995) are able to show that the INAR(p) process may be generally viewed as a special multitype branching process with immigration.

Du and Li (1991) show that, for $\alpha_k \in [0,1)$, the INAR(p) process is asymptotically stationary as long as $\sum_{k=1}^{p} a_k < 1$ and that the correlation properties of this process are identical to the linear Gaussian AR(p) process.

The Alzaid and Al-Osh (1990) specification of the INAR(p) process differs from that of Du and Li (1991) in that it employs the alternative assumption that the conditional distribution of $(\alpha_1 \circ X_{t-p}, \alpha_2 \circ X_{t-p}, \ldots, \alpha_p \circ X_{t-p})'$ given X_{t-p} is multinomial with parameters $(\alpha_1, \alpha_2, \ldots, \alpha_p, X_{t-p})$. The statistical properties of the Alzaid and Al-Osh (1990) model are very different from that of Du and Li (1991) and the model is much less tractable. In this study, we confine ourselves to the case where the thinning operators are conditionally independent.

8.2 Estimation Methods

It is clearly the case that the estimation problem connected with the INAR process is more complicated than that of the Gaussian AR process. The complication arises from the fact that the process is nonlinear due to the thinning operator and the conditional distribution of X_t given its lags is the convolution of the distribution of ε_t and the distributions of p Binomials determined by $\alpha_k \circ X_{t-k}$ for $k = 1, 2, \ldots, p$. Several estimation methods have been proposed in the literature. These include the moments based Yule-Walker estimation method and the conditional least squares estimation method. Al-Osh and Alzaid (1987), assuming Poisson arrivals, presented a maximum likelihood procedure for estimating the PoINAR(1) process. The objective

of this section is to provide a brief review of the three estimation methods.

8.2.1 Yule-Walker Estimation

As discussed earlier, the autocorrelation structure of the INAR(p) process is identical to the linear Gaussian AR(p) process. If we define

$$\gamma_k = Cov(X_t, X_{t-k}),$$

it can be easily verified that for the process in (8.1)

$$\gamma_k = \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{k-p}$$

or

$$\rho_k = \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \cdots + \alpha_p\rho_{k-p} \tag{8.2}$$

It follows that a simple way to get an estimator for the model parameters is to replace the autocorrelation coefficient ρ_k with the sample autocorrelation coefficient $\widehat{\rho}_k$ in the Yule-Walker equations. For $k = 1, 2, \ldots, p$ in (8.2), we have the Yule-Walker equations

$$\boldsymbol{\Gamma\alpha} = \boldsymbol{\rho} \tag{8.3}$$

where

$$\boldsymbol{\Gamma} = [\rho_{|i-j|}]_{p\times p}$$

$$\begin{aligned} \boldsymbol{\alpha} &= (\alpha_1, \alpha_2, \ldots, \alpha_p) \\ \boldsymbol{\rho} &= (\rho_1, \rho_2, \ldots, \rho_p)\,. \end{aligned}$$

Replacing $\boldsymbol{\rho}$ with its sample estimate $\widehat{\boldsymbol{\rho}}$ in (8.3) yields the Yule-Walker estimate $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$, which satisfies

$$\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\alpha}}=\widehat{\boldsymbol{\rho}}. \tag{8.4}$$

Let λ denote the mean of ε_t, then the estimate of λ for the INAR(p) model is given by

$$\widehat{\lambda} = (1 - \widehat{\alpha}_1 - \widehat{\alpha}_2 - \cdots - \widehat{\alpha}_p)\overline{X} \tag{8.5}$$

where $\overline{X}$ is the sample mean given by $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$.

Let $\boldsymbol{\theta}$ denote the parameter vector for the INAR(p) process $(\alpha_1, \alpha_2, \ldots, \alpha_p, \lambda)$. Du and Li (1991) show that the Yule-Walker estimator $\widehat{\boldsymbol{\theta}}_{YW}$ is strongly consistent.

8.2.2 Conditional Least Squares Estimation

For the INAR(p) process in (8.1), the conditional expectation of X_t is given by

$$E(X_t|\Im_{t-1}) = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \lambda \equiv g_t(\boldsymbol{\theta}, \Im_{t-1})$$

where $\Im_t$ is the standard filtration, that is $\Im_t = \sigma(X_1, X_2, \ldots, X_t)$. The conditional least squares (CLS) minimizes the following function over the parameter space

$$\begin{aligned} Q(\boldsymbol{\theta}) &= \sum_{t=p+1}^{T} [X_t - g_t(\boldsymbol{\theta}, \Im_{t-1})]^2 \\ &= \sum_{t=p+1}^{T} [X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \cdots - \alpha_p X_{t-p} - \lambda]^2 \end{aligned} \tag{8.6}$$

i.e.

$$Q(\widehat{\boldsymbol{\theta}}_{CLS}) = \min Q(\boldsymbol{\theta})$$

where $\widehat{\boldsymbol{\theta}}_{CLS}$ can be solved from the first order condition

$$\frac{\partial Q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

which admits

$$\widehat{\boldsymbol{\Gamma}}^* \widehat{\boldsymbol{\alpha}}^* = \widehat{\boldsymbol{\rho}}^* \tag{8.7}$$

$$\widehat{\lambda}^* = \overline{X}^{(0)} - \sum_{j=1}^{p} \widehat{\alpha}_j^* \overline{X}^{(j)}, \tag{8.8}$$

where

$$\overline{X}^{(j)} = \frac{1}{T-p} \sum_{t=p+1}^{T} X_{t-j}$$

$$\widehat{\gamma}_{k-j}^* = \frac{1}{T-p} \sum_{t=p+1}^{T} \left(X_{t-j} - \overline{X}^{(j)} \right) \left(X_{t-k} - \overline{X}^{(k)} \right)$$

$$\widehat{\rho}_{k-j}^* = \frac{\widehat{\gamma}_{k-j}^*}{\widehat{\gamma}_0^*}$$

$$\widehat{\boldsymbol{\Gamma}}^* = [\widehat{\rho}_{|i-j|}^*]_{p \times p}$$

$$\widehat{\boldsymbol{\rho}}^* = \left(\widehat{\rho}_1^*, \widehat{\rho}_2^*, \ldots, \widehat{\rho}_p^* \right).$$

It is easily seen that when T is large enough, $\widehat{\Gamma}^* - \widehat{\Gamma}$, $\widehat{\boldsymbol{\rho}}^* - \widehat{\boldsymbol{\rho}}$, and $\overline{X}^{(j)} - \overline{X}$ are nearly zero. Therefore, one would expect that the CLS estimator in (8.7) and (8.8)[17] are

[17]Alternatively, note that the objective function in (8.6) has exactly the same expression as in the Gaussian AR(p) case. It follows that we could simply apply the ordinary least squares estimator to the INAR(p) case. Thus we can rewrite the CLS estimator in the following matrix form.

$$\widehat{\boldsymbol{\theta}}_{CLS} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

where

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}_t^{(0)} \\ \mathbf{X} &= (\mathbf{X}_t^{(1)}, \mathbf{X}_t^{(2)}, \ldots, \mathbf{X}_t^{(p)}) \end{aligned}$$

with $\mathbf{X}_t^{(j)} = (X_{p-j+1}, X_{p-j+2}, \ldots, X_{T-j})'$.

very close to those in (8.4) and (8.5). In fact, it can be shown that the two methods are asymptotically equivalent.

It can easily be verified that g_t, $\partial g_t/\partial\boldsymbol{\theta}$ and $\partial^2 g_t/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ satisfy all the regularity conditions proposed by Klimko and Nelson (1978). It follows that the CLS estimator $\widehat{\boldsymbol{\theta}}_{CLS}$ is strongly consistent and has the following asymptotic distribution

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_{CLS}-\boldsymbol{\theta}\right)\xrightarrow{d} N(\mathbf{0},\mathbf{j}^{-1})$$

where $\mathbf{j}$ is the Godambe information matrix given by

$$\mathbf{j}=\mathbf{S}\mathbf{V}^{-1}\mathbf{S}$$

where

$$\mathbf{S}=E\left[\frac{\partial g_t(\boldsymbol{\theta},\Im_{t-1})}{\partial\boldsymbol{\theta}}\frac{\partial g_t(\boldsymbol{\theta},\Im_{t-1})}{\partial\boldsymbol{\theta}'}\right] \tag{8.9}$$

$$\mathbf{V}=E\left[u_t^2(\boldsymbol{\theta})\frac{\partial g_t(\boldsymbol{\theta},\Im_{t-1})}{\partial\boldsymbol{\theta}}\frac{\partial g_t(\boldsymbol{\theta},\Im_{t-1})}{\partial\boldsymbol{\theta}'}\right] \tag{8.10}$$

with

$$u_t(\boldsymbol{\theta})=X_t-g_t(\boldsymbol{\theta},\Im_{t-1}).$$

8.2.3 Maximum Likelihood Estimation of the INAR(1) Model

Assuming Poisson innovations, Al-Osh and Alzaid (1987) proposed that the first-order INAR process be estimated by maximum likelihood. For the PoINAR(1) process, the likelihood function of a sample of T observations can be written as

$$L(\alpha,\lambda)=P(X_1)\prod_{t=2}^{T}P(X_t|X_{t-1})$$

where $P(X_t|X_{t-1})$ is the conditional probability of X_t given X_{t-1}, which is also known as the transition probability for the Markov chain implied by the process. It is well known that this transition probability is given by

$$P(X_t|X_{t-1}) = \sum_{i=0}^{\min(X_{t-1},X_t)} \binom{X_{t-1}}{i} \alpha^i (1-\alpha)^{X_{t-1}-i} \frac{e^{-\lambda}\lambda^{X_t-i}}{(X_t-i)!}.$$

Since the marginal distribution of the PoINAR(1) process is also Poisson with mean equal to $\lambda/(1-\alpha)$, the unconditional likelihood function can be written as

$$L(\alpha,\lambda) = \frac{e^{-\lambda/(1-\alpha)}\left[\lambda/(1-\alpha)\right]^{X_1}}{(X_1)!} \prod_{t=2}^{T} \left\{ \sum_{i=0}^{\min(X_{t-1},X_t)} \binom{X_{t-1}}{i} \alpha^i (1-\alpha)^{X_{t-1}-i} \frac{e^{-\lambda}\lambda^{X_t-i}}{(X_t-i)!} \right\}. \tag{8.11}$$

In the case of conditional maximum likelihood (CML) estimation, X_1 is treated as given and the conditional likelihood function is of the form:

$$L(\alpha,\lambda) = \prod_{t=2}^{T} \left\{ \sum_{i=0}^{\min(X_{t-1},X_t)} \binom{X_{t-1}}{i} \alpha^i (1-\alpha)^{X_{t-1}-i} \frac{e^{-\lambda}\lambda^{X_t-i}}{(X_t-i)!} \right\}. \tag{8.12}$$

The ML estimator and the CML estimator can be obtained by maximizing the log of the likelihood function in (8.11) and (8.12), respectively. Al-Osh and Alzaid (1987) showed that the procedure of Sprott (1983) can be used in the current case to eliminate one of the parameters in the derivatives of the log-likelihood function, and that the ML estimates $\widehat{\alpha}$ and $\widehat{\lambda}$ have the following asymptotic distribution:

$$\sqrt{T}\begin{pmatrix} \widehat{\alpha}-\alpha \\ \widehat{\lambda}-\lambda \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{i}^{-1})$$

where the matrix $\mathbf{i}$ is the Fisher information, i.e. the expectation of the negative second derivatives of the log-likelihood function.

Chapter 9

Maximum Likelihood Estimation of the INAR(p) Model

This chapter looks at likelihood based estimation of a generalized INAR(p) process. Section 9.1 sets out the specification of the generalized model and Section 9.2 considers the maximum likelihood estimation of the model. A recursive representation of the transition probability function for the INAR(p) model is proposed, based on which we derive the expressions for the score function and the Fisher information matrix with respect to the conditional likelihood. We show that under certain conditions these quantities can be neatly represented as conditional expectations. The unconditional likelihood, however, is slightly complicated by the lack of knowledge of the analytic expression for the joint distribution of the first p observations. Nevertheless, we overcome this by proposing a simple numerical procedure for transforming

the conditional probability into the joint probability. This joint probability, when added to the results from the conditional case, produces the exact likelihood as well as the corresponding score and information quantities. In Section 9.3, we consider the special case of INAR(p) model with Binomial thinning and Poisson Innovations, for which more specific details on ML estimation are provided, including the asymptotic distribution of the ML estimator. The main results of this chapter appear in Bu et al. (2006a).

9.1 The Generalized INAR(p) Model

We consider a generalization of the model in (8.1), i.e.

$$X_t = \alpha_1 \cdot X_{t-1} + \alpha_2 \cdot X_{t-2} + \cdots + \alpha_p \cdot X_{t-p} + \varepsilon_t \tag{9.1}$$

where $\alpha_k \cdot X_{t-k}$ is, conditional on X_{t-k}, a real-valued random variable (operator) with parameter α_k. The "$\cdot$" is denoted as a general thinning operator. The variables $\alpha_k \cdot X_{t-k}$ for $k = 1, 2, \ldots, p$ are conditionally mutually independent. The operator thus delivers a random value and the dependence in $\{X_t\}$ is induced via the conditioning variables X_t. For a general treatment of such operators, see Joe (1996). The probability (density) function of $\alpha_k \cdot X_{t-k}$ conditioned on X_{t-k}, with respect to some measure υ, is written as

$$f(s_k|x_{t-k}) = f(s_k|x_{t-k}; \alpha_k) \tag{9.2}$$

while that of ε_t is

$$g(\varepsilon) = g(\varepsilon; \boldsymbol{\lambda}). \tag{9.3}$$

for some parameter vector $\boldsymbol{\lambda}$.

In integer models of principal concern here, υ is regarded as a counting measure and the model in (9.1) is henceforth referred to as the generalized INAR(p) model. Therefore, in many occasions (9.2) and (9.3) are treated as probability functions, e.g. $f(s_k|X_{t-k}; \alpha_k)$ and $g(\varepsilon_t; \boldsymbol{\lambda})$. Obvious special cases include distributions such as Binomial and Poisson.

The main task of this chapter is to develop a framework for maximum likelihood estimation of the generalized INAR(p) model. For notation convenience, the parameters α_k and the vector $\boldsymbol{\lambda}$ are often suppressed in the following exposition.

9.2 Likelihood Calculations

In this section we consider the maximum likelihood estimation of the generalized INAR(p) model. As mentioned earlier, the exact likelihood of the INAR(p) model is complicated by the joint distribution of the first p observations. For ease of exposition, we begin our exploration with the conditional likelihood estimation and derive the corresponding score functions and Fisher information matrix. We then show how these results for the conditional case can be extended to the unconditional case with the aid of a numerical procedure for computing the joint distribution of

$(X_1, X_2, \ldots X_p)$.

9.2.1 The Conditional Likelihood

Conditioning on the first p observations leads to a simple form of the likelihood, viz.

$$L(\alpha_1, \ldots, \alpha_p, \lambda) = \prod_{t=p+1}^{T} P(X_t | X_{t-1}, \ldots, X_{t-p}). \tag{9.4}$$

Obviously, the knowledge of the transition probability function $P(X_t|X_{t-1}, \ldots, X_{t-p})$ is sufficient for the construction of the conditional likelihood.

9.2.1.1 The Transition Probability Function

The primary difficulty of implementing ML estimation lies in the fact that the transition probability of the INAR(p) model is a $(p+1)$-fold convolution and thus difficult to calculate efficiently. Theorem 9.1 below shows how these transition probabilities may be calculated by a simple recursive mechanism. Notice that the recursion is defined on the set of conditioning arguments.

Theorem 9.1 *For the generalized INAR(p) model in (9.1), the transition probability function can be expressed in the recursive form as*

$$\begin{aligned} & P(X_t | X_{t-1}, \ldots, X_{t-p}) \\ = & \sum\nolimits_{s_1} f(s_1 | X_{t-1}; \alpha_1) P(X_t - s_1 | X_{t-2}, \ldots, X_{t-p}) \end{aligned} \tag{9.5}$$

where the starting value is given by

$$P\left(X_t - \sum_{k=1}^{p-1} s_k \middle| X_{t-p}\right) = \sum\nolimits_{s_p} f\left(s_p | X_{t-p}\right) g\left(X_t - \sum_{k=1}^{p} s_k\right). \tag{9.6}$$

Proof. Given in Appendix 1. ■

The recursive representation in (9.5) has several advantages. Firstly, since it is derived from a general specification of thinning operator and innovation sequence, it is valid for any probability distribution functions that satisfy the conditions set out in Section 9.1. Secondly, the recursive mechanism substantially enhances the computation of the required transition probability by sequentially lowering the order of the convolution. The main advantage of this representation, however, is that it provides a succinct expression for the conditional probability function of higher-order INAR models, which greatly facilitates the derivation of the score and information quantities. See Appendix 3 for an example of its effectiveness.

Using this transition probability function, the conditional likelihood of the INAR(p) model can be easily evaluated via (9.4).

9.2.1.2 The Score and Information Matrix

As in Freeland and McCabe (2004a), it proves convenient to express the score function in terms of certain conditional expectations. The following theorems extend the INAR(1) results of Freeland (1998) to the vector parameter case.

Theorem 9.2 *Let $\dot{\ell}_{\alpha_k}$ denote the score with respect to α_k for $k \in [1, p]$ and $\dot{\ell}_{\boldsymbol{\lambda}}$ the score with respect to the vector $\boldsymbol{\lambda}$. Denote by $E_t[\cdot]$ the conditional expectation with*

respect to the sigma field, $\Im_t = \sigma\left(X_t, X_{t-1}, ..., X_{t-p}\right)$. *Assume the density functions* f *and* g *in (9.2) and (9.3) satisfy*

$$\begin{aligned} \frac{\partial f(s_k|x_{t-k};\alpha_k)}{\partial \alpha_k} &= \tau(s_k) f(s_k|x_{t-k};\alpha_k) \\ \frac{\partial g(\varepsilon;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} &= \boldsymbol{\gamma}(\varepsilon) g(\varepsilon;\boldsymbol{\lambda}) \end{aligned}$$

for some scalar function $\tau(\cdot)$ *and vector function* $\boldsymbol{\gamma}(\cdot)$. *Then for the model (9.1)*

$$\dot{\ell}_{\alpha_k} = \sum_{t=p+1}^{T} E_t\left[\tau\left(\alpha_k \cdot X_{t-k}\right)\right]$$

and

$$\dot{\ell}_{\boldsymbol{\lambda}} = \sum_{t=p+1}^{T} E_t[\boldsymbol{\gamma}\left(\varepsilon_t\right)].$$

Proof. Given in Appendix 1. ■

The information matrix can also be expressed in a similar way in terms of conditional expectations.

Theorem 9.3 *Let* $\ddot{\ell}_{ab}$ *denote the second derivatives of the log-likelihood with respect to* a *and* b *and let* τ_{α_k} *denote the derivative of the function* τ *with respect to* α_k. *The matrix* $\boldsymbol{\gamma}_{\boldsymbol{\lambda}}$ *is defined as the derivative of the vector function* $\boldsymbol{\gamma}$ *with respect to the vector* $\boldsymbol{\lambda}$. *Under the conditions of Theorem 9.2 the following results hold for the model (9.1):*

$$\ddot{\ell}_{\alpha_k\alpha_k} = \sum_{t=p+1}^{T} \left\{E_t\left[\tau_{\alpha_k}\left(\alpha_k \cdot X_{t-k}\right)\right] + Var_t\left[\tau\left(\alpha_k \cdot X_{t-k}\right)\right]\right\},$$

$$\ddot{\ell}_{\alpha_m\alpha_n} = \sum_{t=p+1}^{T} Cov_t\left[\tau\left(\alpha_m \cdot X_{t-m}\right), \tau\left(\alpha_n \cdot X_{t-n}\right)\right],$$

$$\ddot{\ell}_{\alpha_k \lambda} = \sum_{t=p+1}^{T} Cov_t \left[\tau \left(\alpha_k \cdot X_{t-k} \right), \gamma \left(\varepsilon_t \right) \right]$$

and

$$\ddot{\ell}_{\lambda\lambda'} = \sum_{t=p+1}^{T} \left\{ E_t \left[\gamma_\lambda \left(\varepsilon_t \right) \right] + Var_t \left[\gamma \left(\varepsilon_t \right) \right] \right\}.$$

where k, m *and* $n \in [1, p]$ *and* $m \neq n$.

Proof. Given in Appendix 1. ■

It is worth mentioning that to represent the score functions and information matrix in terms of conditional expectation is not just a matter of convenience. It also provides new interpretations to these quantities. For example, it can be seen that $\ddot{\ell}_{\alpha_k \alpha_k}$ and $\ddot{\ell}_{\lambda\lambda'}$ reflect the mean-variance relations of each individual component of the model, and $\ddot{\ell}_{\alpha_m \alpha_n}$ and $\ddot{\ell}_{\alpha_k \lambda}$ reflect the conditional independence assumption between model components. The conditions on derivatives of the densities of the unobserved components will be satisfied by members of the exponential family which includes, of course, the Poisson and Binomial distributions. It will be shown in Section 9.3 that in the Binomial-Poisson case these new expressions also lead to new definitions of the residuals of the model.

9.2.2 The Unconditional Likelihood

Let $P(X_1, \ldots, X_p)$ denote the joint probability distribution of the first p observations. Then the unconditional likelihood function of the INAR(p) model can be

written as

$$L(\alpha_1, \ldots, \alpha_p, \lambda) = P(X_1, \ldots, X_p) \prod_{t=p+1}^{T} P(X_t | X_{t-1}, \ldots, X_{t-p}).$$

It can be seen that under the conditions of Theorem 9.2 and 9.3 the score functions can be written as

$$\dot{\ell}^0_{\alpha_k} = \frac{\partial P(X_1, \ldots, X_p)}{\partial \alpha_k} + \sum_{t=p+1}^{T} E_t \left[\tau(\alpha_k \cdot X_{t-k}) \right],$$

$$\dot{\ell}^0_{\boldsymbol{\lambda}} = \frac{\partial P(X_1, \ldots, X_p)}{\partial \boldsymbol{\lambda}} + \sum_{t=p+1}^{T} E_t[\boldsymbol{\gamma}(\varepsilon_t)],$$

and the information matrix are given by

$$\ddot{\ell}^0_{\alpha_k \alpha_k} = \frac{\partial^2 P(X_1, \ldots, X_p)}{\partial \alpha_k^2} + \sum_{t=p+1}^{T} \left\{ E_t \left[\tau_{\alpha_k}(\alpha_k \cdot X_{t-k}) \right] + Var_t \left[\tau(\alpha_k \cdot X_{t-k}) \right] \right\},$$

$$\ddot{\ell}^0_{\alpha_m \alpha_n} = \frac{\partial^2 P(X_1, \ldots, X_p)}{\partial \alpha_m \partial \alpha_n} + \sum_{t=p+1}^{T} Cov_t \left[\tau(\alpha_m \cdot X_{t-m}), \tau(\alpha_n \cdot X_{t-n}) \right],$$

$$\ddot{\ell}^0_{\alpha_k \boldsymbol{\lambda}} = \frac{\partial^2 P(X_1, \ldots, X_p)}{\partial \alpha_k \partial \boldsymbol{\lambda}} + \sum_{t=p+1}^{T} Cov_t \left[\tau(\alpha_k \cdot X_{t-k}), \boldsymbol{\gamma}(\varepsilon_t) \right]$$

and

$$\ddot{\ell}^0_{\boldsymbol{\lambda}\boldsymbol{\lambda}'} = \frac{\partial^2 P(X_1, \ldots, X_p)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} + \sum_{t=p+1}^{T} \left\{ E_t \left[\boldsymbol{\gamma}_{\boldsymbol{\lambda}}(\varepsilon_t) \right] + Var_t \left[\boldsymbol{\gamma}(\varepsilon_t) \right] \right\}.$$

In the relatively simple case addressed by Al-Osh and Alzaid (1987), it can be shown that the marginal distribution of X_t is also Poisson. As a result, both $P(X_1)$ and its derivatives can be evaluated analytically. However, in the case of higher-order INAR processes, analytic expressions for the joint probability function $P(X_1, \ldots, X_p)$ and its derivatives are usually not available.

Nevertheless, it can be noted that the joint probability $P(X_{t-1}, \ldots, X_{t-p})$ can be uniquely determined by the conditional probability function $P(X_t|X_{t-1}, \ldots, X_{t-p})$ providing that the process is strictly stationary[18]. This intrinsic connection between these two quantities can be exploited for calculating the exact likelihood. To this end, we propose a simple procedure for transforming the conditional probability $P(X_t|X_{t-1}, \ldots, X_{t-p})$ into the joint probability $P(X_{t-1}, \ldots, X_{t-p})$ for stationary processes. The details of the procedure are given in Appendix 2.

Based on this procedure, we are able to evaluate both $P(X_1, \ldots, X_p)$ and its derivatives numerically. It follows that the unconditional maximum likelihood estimates of the INAR(p) model can be obtained by setting $\dot{\ell}^0_{\alpha_k} = 0$ and $\dot{\ell}^0_{\lambda} = 0$. The asymptotic covariance matrix can be obtained by calculating the Fisher information given above.

9.3 The Binomial-Poisson Specification

For the theorems proposed in the Section (9.2) to be useful in practice, both the conditional probability distribution of $\alpha_k \cdot X_{t-k}|X_{t-k}$ and the distribution of ε_t need to be explicitly specified. The most widely adopted assumption in the literature is that $\alpha_k \cdot X_{t-k}|X_{t-k}$ follows a Binomial distribution[19] and the innovations are Poisson. In this section, we provide precise details on the ML estimation for the Binomial-Poisson

[18]In fact, for a stationary process the joint distribution of any set of observations is uniquely determined by the conditional probability function.

[19]For this reason, all INAR processes considered hereafter assume Binomial thinning operations.

specification. The asymptotic distribution of the ML estimator is also presented.

9.3.1 Maximum Likelihood Estimation

Under the Binomial thinning assumption, the conditional probability distribution function of $\alpha_k \cdot X_{t-k}|X_{t-k}$ is given by

$$f\left(s_k|X_{t-k};\alpha_k\right) = \binom{X_{t-k}}{s_k}\alpha_k^{s_k}(1-\alpha_k)^{X_{t-k}-s_k} \tag{9.7}$$

for $k = 1, 2, \ldots, p$. For Poisson innovations, the probability function of ε_t is given by

$$g\left(\varepsilon_t|\lambda\right) = \frac{e^{-\lambda}\lambda^{\varepsilon_t}}{\varepsilon_t!} \tag{9.8}$$

The following corollary gives the conditional probability function for the INAR(p) model with Poisson arrivals

Corollary 9.1 *For the INAR(p) model with Poisson innovations, the recursive representation of the conditional probability function can be written as*

$$\begin{aligned} & P(X_t|X_{t-1},\ldots,X_{t-p}) \\ = & \sum_{i_1=0}^{\min(X_{t-1},X_t)} \binom{X_{t-1}}{i_1}\alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1}P(X_t - i_1|X_{t-2},\ldots,X_{t-p}) \end{aligned} \tag{9.9}$$

where the starting value is given by

$$\begin{aligned} & P\left(X_t - \sum_{k=1}^{p-1} i_k \middle| X_{t-p}\right) \\ = & \sum_{i_p=0}^{\min[X_{t-p},X_t-(i_1+\cdots+i_{p-1})]} \binom{X_{t-p}}{i_p}\alpha_p^{i_p}(1-\alpha_p)^{X_{t-p}-i_p}\frac{e^{-\lambda}\lambda^{X_t-(i_1+\cdots+i_p)}}{[X_t-(i_1+\cdots+i_p)]!} \end{aligned}$$

and the complete expression of the conditional probability can be obtained by repeated substitution and is given by

$$
\begin{aligned}
& P(X_t|X_{t-1},\ldots,X_{t-p}) \\
= & \sum_{i_1=0}^{\min(X_{t-1},X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} \sum_{i_2=0}^{\min(X_{t-2},X_t-i_1)} \binom{X_{t-2}}{i_2} \alpha_2^{i_2}(1-\alpha_2)^{X_{t-2}-i_2} \\
& \cdots \sum_{i_p=0}^{\min[X_{t-p},X_t-(i_1+\cdots+i_{p-1})]} \binom{X_{t-p}}{i_p} \alpha_p^{i_p}(1-\alpha_p)^{X_{t-p}-i_p} \frac{e^{-\lambda}\lambda^{X_t-(i_1+\cdots+i_p)}}{[X_t-(i_1+\cdots+i_p)]!}.
\end{aligned}
\tag{9.10}
$$

Proof. Given in Appendix 1. ■

Clearly, the expression in (9.10) is very cumbersome and therefore of little use. The recursions of Theorem 9.1 not only facilitate the computation of the likelihood but they are also extremely useful in computing derivatives and hence the score and information quantities. The effectiveness of the recursions is exemplified in the derivation of the conditional expectations in Appendix 3.

It can be easily verified that the conditions on the derivatives of the densities of the unobserved components set out in Theorem 9.2 and 9.3 are satisfied by the Poisson and Binomial distributions. It follows that the score functions and Fisher information matrix of the process can be represented in terms of conditional expectations. In particular, it can be shown that for the Binomial-Poisson case

$$
\tau(s_k;\alpha_k) = \frac{s_k}{\alpha_k(1-\alpha_k)} - \frac{\alpha_k X_{t-k}}{\alpha_k(1-\alpha_k)} \tag{9.11}
$$

and

$$
\gamma(\varepsilon_t;\lambda) = \frac{\varepsilon_t}{\lambda} - 1. \tag{9.12}
$$

The following two corollaries give explicit forms for the score and information matrix.

Corollary 9.2 *Under the conditions of Theorem 9.2 the following results hold for the INAR(p) model with Poisson innovations:*

$$\dot{\ell}_{\alpha_k} = \frac{1}{\alpha_k(1-\alpha_k)} \sum_{t=p+1}^{T} \{E_t[\alpha_k \circ X_{t-k}] - E_{t-1}[\alpha_k \circ X_{t-k}]\}$$

and

$$\dot{\ell}_{\lambda} = \frac{1}{\lambda} \sum_{t=p+1}^{T} \{E_t[\varepsilon_t] - E_{t-1}[\varepsilon_t]\}.$$

Proof. Given in Appendix 1. ■

Corollary 9.3 *Under the conditions of Theorem 9.2 the following results hold for the INAR(p) model with Poisson innovations:*

$$\begin{aligned}\ddot{\ell}_{\alpha_k\alpha_k} &= \frac{1}{\alpha_k^2(1-\alpha_k)^2} \sum_{t=p+1}^{T} \{(2\alpha_k - 1)E_t[\alpha_k \circ X_{t-k}] \\ &\quad + Var_t[\alpha_k \circ X_{t-k}] - \alpha_k E_{t-1}[\alpha_k \circ X_{t-k}]\},\end{aligned}$$

$$\ddot{\ell}_{\alpha_m\alpha_n} = \frac{1}{\alpha_m\alpha_n(1-\alpha_m)(1-\alpha_n)} \sum_{t=p+1}^{T} Cov_t[\alpha_m \circ X_{t-m}, \alpha_n \circ X_{t-n}],$$

$$\ddot{\ell}_{\alpha_k\lambda} = \frac{1}{\lambda\alpha_k(1-\alpha_k)} \sum_{t=p+1}^{T} Cov_t[\alpha_k \circ X_{t-k}, \varepsilon_t]$$

and

$$\ddot{\ell}_{\lambda\lambda} = \frac{1}{\lambda^2} \sum_{t=p+1}^{T} \{Var_t[\varepsilon_t] - E_t[\varepsilon_t]\}.$$

Proof. Given in Appendix 1. ■

These representations clearly show that the scores and information implied by the INAR(p) model with Poisson innovations can be decomposed into quantities associated with each component of the model. This is analogous to the results for the PoINAR(1) case given in Freeland and McCabe (2004a). Specifically, the terms in the score functions measure the incremental contribution of the information arriving at time t for each process and the second derivatives characterize the properties of the component processes. For example, the expression $\ddot{\ell}_{\lambda\lambda}$ reflects the Poisson mean-variance relationship given the additional information available at time t.

In addition to enhancing the interpretation of the model, these conditional expectations are also an important computational tool. We show in Appendix 3 that all of the conditional expectations required can be expressed as functions of the transition probability. For example,

$$E_t\left[\alpha_k \circ X_{t-k}\right] = \frac{\alpha_k X_{t-k} P(X_t - 1 | X_{t-1}, \ldots, X_{t-k} - 1, \ldots, X_{t-p})}{P(X_t | X_{t-1}, \ldots, X_{t-p})} \tag{9.13}$$

$$E_t[\varepsilon_t] = \frac{\lambda P(X_t - 1 | X_{t-1}, \ldots, X_{t-p})}{P(X_t | X_{t-1}, \ldots, X_{t-p})} \tag{9.14}$$

and the conditional probabilities required may be computed either by (9.10) or, more efficiently, by the recursive representation of (9.9). It should be mentioned that the conditional probability would be zero if negative values appear in the probability function, e.g.

$$P(-1|0,0) = P(0|-1,0) = P(0|0,-1) = 0.$$

where the conditional probabilities are written in an obvious short notation[20]. It is

[20]For example, $P(X_t = 0 | X_{t-1} = 0, X_{t-2} = 0)$ is given by $P(0|0,0)$.

also important to note that the time t expectations are different from those calculated at time $t-1$. For instance, we have $E_{t-1}[\alpha_k \circ X_{t-k}] = \alpha_k X_{t-k}$ and $E_{t-1}[\varepsilon_t] = \lambda$.

As in Freeland and McCabe (2004a) the new representation of the score functions leads to new definitions of the residuals in the model (9.1). In particular for the INAR(p) model with Poisson innovations, there are residuals for the thinning components

$$r_{kt}^* = E_t[\alpha_k \circ X_{t-k}] - E_{t-1}[\alpha_k \circ X_{t-k}]$$

for $k = 1, 2, \ldots, p$, and for the arrival component

$$r_{\varepsilon t}^* = E_t[\varepsilon_t] - E_{t-1}[\varepsilon_t]$$

It can be easily verified that adding the new sets of residuals together gives the usual definition of residuals, i.e.

$$\hat{\varepsilon}_t = X_t - \sum_{k=1}^{p} \hat{\alpha}_k X_{t-k}.$$

These residuals may be used in the usual ways to assess the adequacy of the model and may suggest improved specifications. A discussion on testing for model specification under this new framework is presented in Chapter 12.

9.3.2 Asymptotic Distribution of the Maximum Likelihood Estimator

Al-Osh and Alzaid (1987) show that for the PoINAR(1) model the ML estimator is consistent and asymptotically normal. It can be verified that the likelihood function

for the INAR(p) with Poisson innovations satisfies all the regularity conditions for the consistency and asymptotic normality of ML estimators. We thus have the following result.

Theorem 9.4 *Let $\boldsymbol{\theta}$ denote the parameter vector for the INAR(p) model with Poisson innovations. The maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{ML}$ has the following asymptotic distribution:*

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{i}^{-1})$$

where the matrix $\mathbf{i}$ is the Fisher information per observation, i.e. the expectation of the negative second derivatives as given in Corollary 9.3.

Proof. Given in Appendix 4. ■

The parameter estimates for the model can be found using Newton-Raphson type iterative procedures. Standard errors of the estimates are readily available from the observed Fisher information matrix. Alternatively, if the time series is comprised of low counts, the expected Fisher Information can also be calculated numerically using the results in Corollary 9.3. See Section 10.1.1 for more details.

9.4 Conclusion

In this chapter, we extended earlier work of Freeland and McCabe (2004a) and proposed a framework for ML estimation of a general INAR(p) process. The likelihood function as well as the score and information matrix are derived based on a recursive

representation of the transition probability function for the INAR(p) model. These quantities form the basis for maximum likelihood estimation and inferences. We show that under certain conditions the score function and the Fisher information matrix can be neatly represented as conditional expectations. These new expressions enhance the interpretation of these quantities and lead naturally to new definitions of the residuals of the INAR(p) model. Our expositions are elaborated on the Binomial-Poisson specification, for which the asymptotic distribution of the ML estimator is also provided.

Chapter 10

Comparison of Methods

This chapter investigates the performance of the ML estimator (MLE) in comparison with the Yule-Walker estimator (YWE) and the CLS estimator (CLSE). Both asymptotic and finite sample properties are examined. Section 10.1 examines the asymptotic relative efficiency (ARE) of the MLE in relation to the CLSE. Our results confirm that the proposed MLE is asymptotically more efficient than the CLSE. In Section 10.2 we compare the performance of alternative estimators in small samples. Monte Carlo experiments are conducted and our results suggest that there is a potential gain in using the MLE over the YWE and the CLSE in terms of bias and mean squared error (MSE). In both studies, we found that the efficiency gain of implementing ML is most substantial for persistent processes. The main results of this chapter appear in Bu et al. (2006b).

10.1 Asymptotic Relative Efficiency

In this section, we examine the asymptotic efficiency of the MLE in relation to the CLSE and the YWE. That is, we consider what happens as the sample size goes to infinity. Since the CLSE and the YWE are asymptotically equivalent, it is sufficient to focus on MLE and CLSE only. We compare the two estimators by evaluating the asymptotic relative efficiency (ARE) between the two estimators. The ARE between estimators is defined as the ratio of their asymptotic variances (see Cox and Hinkley (1974)). Let $\widehat{\boldsymbol{\theta}}$ be estimate of $\boldsymbol{\theta}$ and denote by i_{kk}^{-1} the (k,k) element of $\mathbf{i}^{-1}$, the inverse of the Fisher information matrix. Similarly, let j_{kk}^{-1} be the (k,k) element of $\mathbf{j}^{-1}$, which is the inverse of the Godambe information matrix. The ARE for the k^{th} component of $\widehat{\boldsymbol{\theta}}$ is then defined as

$$ARE(\widehat{\theta}_{kk}) = \frac{i_{kk}^{-1}}{j_{kk}^{-1}} \tag{10.1}$$

Clearly, in this setup, an ARE less than unity would suggest better efficiency for the MLE. Notice that there are no simulations involved in this comparison and the sample size is infinitely large. Furthermore, the comparison is between ML and CLS, i.e. conditioning on the initial observations has a negligible asymptotic effect.

10.1.1 The INAR(2) Specification and Information Matrix

In our comparison, we entertain the INAR(2) specification

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \varepsilon_t$$

where ε_t has a Poisson distribution with mean equal to λ. For the MLE, the expected Fisher information matrix can be written as

$$\mathbf{i} = \left(-E \left[\frac{\partial^2 \ln P(X_t|X_{t-1}, X_{t-2})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \right)^{-1} \tag{10.2}$$

where $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \lambda)$ and $P(X_t|X_{t-1}, X_{t-2})$ is the probability of X_t conditioned on X_{t-1} and X_{t-2}. Following Corollary 9.1, this conditional probability is given by

$$\begin{aligned} & P(X_t|X_{t-1}, X_{t-2}) \\ = & \sum_{i_1=0}^{\min(X_{t-1}, X_t)} \left\{ \binom{X_{t-1}}{i_1} \alpha_1^{i_1} (1-\alpha_1)^{X_{t-1}-i_1} \right. \\ & \left. \sum_{i_2=0}^{\min(X_{t-2}, X_t - i_1)} \binom{X_{t-2}}{i_2} \alpha_2^{i_2} (1-\alpha_2)^{X_{t-2}-i_2} \frac{e^{-\lambda} \lambda^{X_t - i_1 - i_2}}{(X_t - i_1 - i_2)!} \right\} \end{aligned} \tag{10.3}$$

By Corollary 9.3

$$\frac{\partial^2 \ln P(X_t|X_{t-1}, X_{t-2})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} \ddot{\ell}_{\alpha_1\alpha_1} & \ddot{\ell}_{\alpha_1\alpha_2} & \ddot{\ell}_{\alpha_1\lambda} \\ \ddot{\ell}_{\alpha_1\alpha_2} & \ddot{\ell}_{\alpha_2\alpha_2} & \ddot{\ell}_{\alpha_2\lambda} \\ \ddot{\ell}_{\alpha_1\lambda} & \ddot{\ell}_{\alpha_2\lambda} & \ddot{\ell}_{\lambda\lambda} \end{bmatrix}$$

where each element in this information matrix can be calculated as specified in Appendix 3. The expectation in (10.2) is calculated numerically. Specifically, we select a large enough positive integer value M such that the probability of a count larger than M is zero. Then, for the INAR(2) model, there are $(M+1)^3$ possible outcomes of the joint observation of $\{X_t, X_{t-1}, X_{t-2}\}$ to sum over for each element of the Fisher information[21]. For example, summing over all $(M+1)^3$ possible values of

[21] If $M = 6$, for instance, there are 343 possible outcomes of joint observation of $\{X_t, X_{t-1}, X_{t-2}\}$. They are $\{0,0,0\}$, $\{0,0,1\}$,..., and $\{6,6,6\}$.

$\{X_t, X_{t-1}, X_{t-2}\}$,

$$E[\ddot{\ell}_{\lambda\lambda}] = \sum_{all\ \{X_t, X_{t-1}, X_{t-2}\}} P(X_t, X_{t-1}, X_{t-2}) \times \left\{ \frac{P(X_t - 2|X_{t-1}, X_{t-2})}{P(X_t|X_{t-1}, X_{t-2})} - \left[\frac{P(X_t - 1|X_{t-1}, X_{t-2})}{P(X_t|X_{t-1}, X_{t-2})} \right]^2 \right\}$$

where $P(X_t, X_{t-1}, X_{t-2})$ is the joint probability of $\{X_t, X_{t-1}, X_{t-2}\}$, which is also calculated numerically based on the conditional probability function in (10.3). Details of transforming the conditional probability into the required joint probability for stationary processes are given in Appendix 2.

For the CLSE, the Godambe information matrix is defined in Section 8.2.2. Specifically, for the INAR(2) model with Poisson innovations we have

$$g_t(\boldsymbol{\theta}, \Im_{t-1}) = E(X_t|\Im_{t-1}) = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \lambda$$

$$u_t(\boldsymbol{\theta}) = X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \lambda$$

The expectations in (8.9) and (8.10) are also evaluated numerically in the same way as in the MLE case.

10.1.2 Results

We calculate and examine the ARE of the two estimators for a range of different parameter values. To ensure that the processes examined are stationary and nondegenerate, the sum of the two thinning parameters, α_1 and α_2, is confined within the range of $[0.1, 0.9]$ and for each of the two thinning parameters, a sequence of different values on a grid of 0.05, ranging from 0.05 to 0.85, are entertained. All possible

combinations of α_1 and α_2, a total of 153 cases, are examined. In addition, in order to reflect varied arrival rates we also consider a sequence of different values of λ.

Since our conclusions are not to be affected by the choice of λ, we select to discuss the results for the typical case where $\lambda = 1$. Tables 10.1, 10.2, and 10.3 show the ARE ratios for the three parameters, $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, and $\widehat{\lambda}$, respectively. As expected, our results confirm that the MLE is asymptotically more efficient than the CLSE for all three parameters, since all the ARE ratios are less than unity. It is generally true that more substantial efficiency gains can be obtained from using the ML as the process becomes more persistent (higher values of α_1 or α_2, or both). Specifically, it can be seen from Table 10.1 that for a given value of α_2 the ARE of $\widehat{\alpha}_1$ decreases as the value of α_1 increases. In particular, when α_2 is low it approaches zero very quickly, indicating substantial advantage of the MLE in these situations; Moreover, for a given α_1 the ARE of $\widehat{\alpha}_1$ decreases as α_2 increases. But we can see that such decrease is not as fast as in the previous case, suggesting that the efficiency gain on $\widehat{\alpha}_1$ is most substantial, particularly when the value of α_1 itself is high. Similarly, Table 10.2 shows that for a fixed α_1 the ARE of $\widehat{\alpha}_2$ decreases rapidly as the α_2 itself increases. However, it is also observed that the ARE of $\widehat{\alpha}_2$ decreases at as high a rate as α_1 increases when α_2 is fixed. This is in contrast to the ARE of $\widehat{\alpha}_1$ which decreases at a much slower pace as α_2 grows than as α_1 itself increases. These observations reflect the dominant role of α_1 in the INAR(2) process in terms of efficient estimation. Table 10.3 once again confirms that more substantial gains are obtained from persistent processes on

Table 10.1: ARE of $\widehat{\alpha}_1$ for the INAR(2) Model ($\lambda = 1$)

	α_2																
α_1	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
0.05	0.9877	0.9849	0.9816	0.9756	0.9690	0.9617	0.9532	0.9458	0.9382	0.9301	0.9218	0.9127	0.9028	0.8918	0.8782	0.8586	0.8226
0.10	0.9631	0.9622	0.9606	0.9573	0.9530	0.9476	0.9422	0.9362	0.9299	0.9228	0.9139	0.9030	0.8886	0.8680	0.8353	0.7750	
0.15	0.9298	0.9315	0.9317	0.9312	0.9299	0.9280	0.9251	0.9215	0.9164	0.9095	0.8995	0.8847	0.8619	0.8243	0.7564		
0.20	0.8897	0.8945	0.8983	0.9013	0.9038	0.9051	0.9057	0.9047	0.9016	0.8953	0.8837	0.8635	0.8280	0.7627			
0.25	0.8446	0.8531	0.8606	0.8680	0.8744	0.8802	0.8848	0.8875	0.8874	0.8825	0.8692	0.8415	0.7863				
0.30	0.7952	0.8071	0.8189	0.8305	0.8420	0.8531	0.8629	0.8709	0.8751	0.8724	0.8574	0.8182					
0.35	0.7406	0.7561	0.7721	0.7886	0.8057	0.8229	0.8395	0.8543	0.8645	0.8657	0.8480						
0.40	0.6806	0.6995	0.7195	0.7411	0.7642	0.7885	0.8131	0.8364	0.8551	0.8620							
0.45	0.6153	0.6367	0.6603	0.6866	0.7157	0.7474	0.7810	0.8149	0.8448								
0.50	0.5449	0.5678	0.5940	0.6240	0.6583	0.6969	0.7397	0.7854									
0.55	0.4705	0.4936	0.5207	0.5527	0.5901	0.6337	0.6837										
0.60	0.3935	0.4153	0.4414	0.4728	0.5101	0.5541											
0.65	0.3164	0.3352	0.3581	0.3858	0.4181												
0.70	0.2418	0.2564	0.2741	0.2945													
0.75	0.1728	0.1824	0.1933														
0.80	0.1123	0.1169															
0.85	0.0629																

Table 10.2: ARE of $\widehat{\alpha}_2$ for the INAR(2) Model ($\lambda = 1$)

	α_2																
α_1	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
0.05	0.9877	0.9626	0.9298	0.8902	0.8472	0.8014	0.7530	0.7039	0.6532	0.6010	0.5474	0.4921	0.4354	0.3772	0.3174	0.2556	0.1917
0.10	0.9852	0.9622	0.9310	0.8933	0.8516	0.8067	0.7599	0.7110	0.6608	0.6088	0.5549	0.4994	0.4420	0.3827	0.3213	0.2572	
0.15	0.9808	0.9610	0.9318	0.8968	0.8576	0.8152	0.7702	0.7231	0.6738	0.6226	0.5693	0.5138	0.4560	0.3956	0.3318		
0.20	0.9734	0.9579	0.9328	0.9015	0.8659	0.8264	0.7844	0.7397	0.6926	0.6432	0.5911	0.5362	0.4781	0.4157			
0.25	0.9632	0.9531	0.9328	0.9064	0.8749	0.8398	0.8016	0.7606	0.7170	0.6705	0.6207	0.5671	0.5082				
0.30	0.9499	0.9452	0.9305	0.9091	0.8832	0.8538	0.8209	0.7853	0.7468	0.7047	0.6585	0.6066					
0.35	0.9312	0.9322	0.9235	0.9083	0.8889	0.8660	0.8403	0.8119	0.7804	0.7451	0.7044						
0.40	0.9055	0.9122	0.9094	0.9008	0.8884	0.8734	0.8563	0.8370	0.8151	0.7894							
0.45	0.8713	0.8826	0.8852	0.8828	0.8777	0.8710	0.8634	0.8549	0.8450								
0.50	0.8268	0.8412	0.8477	0.8503	0.8511	0.8517	0.8528	0.8544									
0.55	0.7712	0.7862	0.7939	0.7984	0.8020	0.8059	0.8102										
0.60	0.7037	0.7159	0.7211	0.7229	0.7229	0.7205											
0.65	0.6247	0.6303	0.6279	0.6200	0.6056												
0.70	0.5353	0.5300	0.5142	0.4876													
0.75	0.4375	0.4170	0.3816														
0.80	0.3336	0.2943															
0.85	0.2262																

Table 10.3: ARE of $\widehat{\lambda}$ for the INAR(2) Model ($\lambda = 1$)

	α_2																
α_1	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
0.05	0.9868	0.9721	0.9543	0.9304	0.9039	0.8751	0.8432	0.8126	0.7815	0.7498	0.7183	0.6872	0.6576	0.6307	0.6082	0.5936	0.5954
0.10	0.9728	0.9546	0.9333	0.9074	0.8790	0.8479	0.8168	0.7850	0.7540	0.7239	0.6943	0.6668	0.6427	0.6239	0.6144	0.6232	
0.15	0.9550	0.9336	0.9077	0.8795	0.8493	0.8179	0.7858	0.7540	0.7228	0.6932	0.6657	0.6415	0.6231	0.6140	0.6231		
0.20	0.9333	0.9091	0.8811	0.8507	0.8194	0.7859	0.7532	0.7206	0.6895	0.6607	0.6353	0.6154	0.6049	0.6120			
0.25	0.9106	0.8842	0.8537	0.8220	0.7880	0.7535	0.7192	0.6860	0.6552	0.6278	0.6056	0.5925	0.5966				
0.30	0.8891	0.8597	0.8272	0.7923	0.7566	0.7211	0.6855	0.6521	0.6219	0.5966	0.5801	0.5804					
0.35	0.8677	0.8360	0.8010	0.7640	0.7265	0.6889	0.6528	0.6194	0.5907	0.5702	0.5658						
0.40	0.8472	0.8135	0.7762	0.7371	0.6976	0.6590	0.6224	0.5898	0.5649	0.5552							
0.45	0.8284	0.7923	0.7529	0.7121	0.6712	0.6318	0.5956	0.5663	0.5512								
0.50	0.8110	0.7733	0.7323	0.6901	0.6484	0.6092	0.5759	0.5557									
0.55	0.7965	0.7572	0.7150	0.6720	0.6305	0.5941	0.5696										
0.60	0.7841	0.7444	0.7016	0.6589	0.6201	0.5921											
0.65	0.7752	0.7354	0.6929	0.6525	0.6215												
0.70	0.7700	0.7307	0.6900	0.6562													
0.75	0.7688	0.7311	0.6955														
0.80	0.7723	0.7385															
0.85	0.7819																

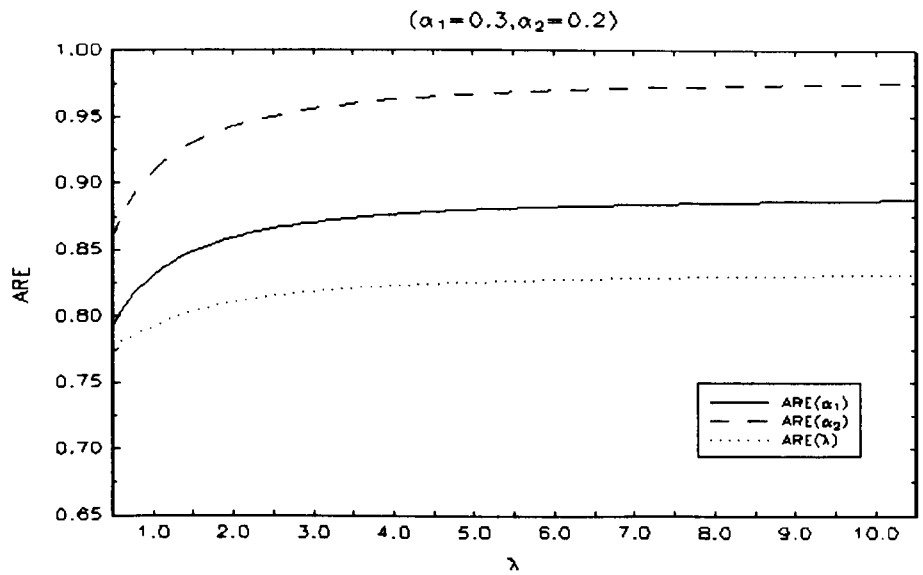
estimating λ, especially if either $\widehat{\alpha}_1$ or $\widehat{\alpha}_2$ approaches unity. But it is interesting to note that, unlike the previous two cases, the ARE of $\widehat{\lambda}$ is slightly more sensitive to the scale of α_2.

Our results show that the observations discussed above are invariant to the magnitude of the arrival rate. However, it is generally the case that for fixed values of thinning coefficients the ARE ratios for all three parameters increase monotonically as a function of λ and approach a limit. These observations are graphically exemplified in Figure 10.1, which shows the ARE of $\widehat{\alpha}_1$, $\widehat{\alpha}_1$ and $\widehat{\lambda}$ as a function of λ for the case where $\alpha_2 = 0.3$ and $\alpha_2 = 0.2$. It can be seen that the ARE for all three parameters increase monotonically as λ increases and reach a limit at about 0.88, 0.98 and 0.83, respectively.

10.2 Finite Sample Performance

In the previous section, we investigated the asymptotic gain of implementing the ML method over the commonly used CLS method by calculating the ARE ratio between the two estimators. Our results confirmed that the proposed MLE is asymptotically more efficient than the CLSE. In particular, we noted that in general the magnitude of the efficiency gain increases as the process becomes persistent. In this section, we take our research into the small sample performance of the MLE in comparison to the YWE and the CLSE.

Figure 10.1: ARE of $\widehat{\alpha}_1$, $\widehat{\alpha}_2$ and $\widehat{\lambda}$ as a Function of λ



10.2.1 Estimating the INAR(2) Model

As in the previous section, we still entertain the INAR(2) specification with Poisson innovations for its relative simplicity. General procedures of the three methods for estimating the INAR(p) model have been discussed in previous chapters. In order to give the readers a better feel for the three estimation methods, in what follows we outline the specific details of the alternative estimators for the INAR(2) process.

10.2.1.1 Yule-Walker Estimator for INAR(2)

It follows from Section 8.2.1 that the autocorrelation functions (ACFs) of the INAR(2) model satisfy the second order difference equation

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2}$$

for $k \geq 2$. The fundamental idea of the YW estimation is to replace the autocorrelation coefficient ρ_k with the sample autocorrelation coefficient $\widehat{\rho}_k$ in the Yule-Walker equations. For the INAR(2) process, the first and second order sample autocorrelations can be estimated by

$$\widehat{\rho}_1 = \frac{(T-1)^{-1} \sum_{t=2}^{T} (X_t - \overline{X})(X_{t-1} - \overline{X})}{T^{-1} \sum_{t=1}^{T} (X_t - \overline{X})^2}$$

and

$$\widehat{\rho}_2 = \frac{(T-2)^{-1} \sum_{t=3}^{T} (X_t - \overline{X})(X_{t-2} - \overline{X})}{T^{-1} \sum_{t=1}^{T} (X_t - \overline{X})^2}$$

where $\overline{X}$ is the sample mean given by $\overline{X} = \frac{1}{T} \sum_{t=1}^{T} X_t$. The corresponding YWE for α_1 and α_2 are obtained by recalling that the autocorrelation structure of this process mimics that of a Gaussian AR(2) process. They are thus given by

$$\widehat{\alpha}_1 = \widehat{\rho}_1 \left[\frac{1 - \widehat{\rho}_2}{1 - \widehat{\rho}_1^2} \right]$$

and

$$\widehat{\alpha}_2 = \frac{\widehat{\rho}_2 - \widehat{\rho}_1^2}{1 - \widehat{\rho}_1^2}.$$

The associated YWE for the parameter λ can be obtained from the unconditional mean of the INAR(2) process and takes the familiar form

$$\widehat{\lambda} = (1 - \widehat{\alpha}_1 - \widehat{\alpha}_2)\overline{X}. \tag{10.4}$$

10.2.1.2 Conditional Least Squares Estimator for INAR(2)

For the INAR(2) specification, the conditional expectation is given by

$$E(X_t|\Im_{t-1}) = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \lambda$$

where $\Im_t$ is the standard filtration, $\Im_t = \sigma(X_1, X_2, \ldots, X_t)$. The CLS minimizes the following objective function over the parameter space

$$\begin{aligned} Q(\boldsymbol{\theta}) &= \sum_{t=3}^{T} [X_t - E(X_t|\Im_{t-1})]^2 \\ &= \sum_{t=3}^{T} [X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \lambda]^2 \end{aligned}$$

with respect to α_1, α_2, and λ. The first order conditions are

$$\begin{aligned} \sum_{t=3}^{T} X_{t-1} [X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \lambda] &= 0 \\ \sum_{t=3}^{T} X_{t-2} [X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \lambda] &= 0 \\ \sum_{t=3}^{T} [X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \lambda] &= 0. \end{aligned}$$

It can be easily verified that the solutions to the above conditions can be written as

$$\begin{aligned}
\widehat{\alpha}_1 &= \frac{B_2C_1 - B_1C_2}{A_1B_2 - A_2B_1} \\
\widehat{\alpha}_2 &= \frac{A_1C_2 - A_2C_1}{A_1B_2 - A_2B_1} \\
\widehat{\lambda} &= \frac{1}{(T-2)} \sum_{t=3}^{T} [X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2}]
\end{aligned} \tag{10.5}$$

where

$$\begin{aligned}
A_1 &= \sum_{t=3}^{T} X_{t-2}X_{t-1} - \frac{1}{(T-2)} \sum_{t=3}^{T} X_{t-1} \sum_{t=3}^{T} X_{t-2} \\
B_1 &= \sum_{t=3}^{T} X_{t-2}^2 - \frac{1}{(T-2)} \left(\sum_{t=3}^{T} X_{t-2} \right)^2 \\
C_1 &= \sum_{t=3}^{T} X_{t-2}X_t - \frac{1}{(T-2)} \sum_{t=3}^{T} X_t \sum_{t=3}^{T} X_{t-2} \\
A_2 &= \sum_{t=3}^{T} X_{t-1}^2 - \frac{1}{(T-2)} \left(\sum_{t=3}^{T} X_{t-1} \right)^2 \\
B_2 &= \sum_{t=3}^{T} X_{t-1}X_{t-2} - \frac{1}{(T-2)} \sum_{t=3}^{T} X_{t-2} \sum_{t=3}^{T} X_{t-1} \\
C_2 &= \sum_{t=3}^{T} X_{t-1}X_t - \frac{1}{(T-2)} \sum_{t=3}^{T} X_t \sum_{t=3}^{T} X_{t-1}.
\end{aligned}$$

10.2.1.3 Maximum Likelihood Estimator for INAR(2)

We have shown in Chapter 9 that the INAR(p) model can be estimated by either conditional or unconditional likelihood. Nevertheless, the unconditional likelihood estimation is complicated by the joint distribution of the first p observations, so that numerical methods have to be used. Since our simulation experiments require repeated estimation of a large number of replications, for computational simplicity

we focus on the conditional maximum likelihood estimation.

It follows from Section 10.1.1 that the conditional probability of X_t for the INAR(2) process is given by (10.3). Thus, conditioned on the first two observations, the likelihood function can be written as

$$L(\alpha_1, \alpha_2, \lambda) = \prod_{t=3}^{T} P(X_t|X_{t-1}, X_{t-2})$$

Using the results in Appendix 3, the score functions with respect to α_1, α_2 and λ can be written as

$$\dot{\ell}_{\alpha_1} = \sum_{t=3}^{T} \frac{X_{t-1}}{(1-\alpha_1)} \frac{[P(X_t - 1|X_{t-1} - 1, X_{t-2}) - P(X_t|X_{t-1}, X_{t-2})]}{P(X_t|X_{t-1}, X_{t-2})}$$

$$\dot{\ell}_{\alpha_2} = \sum_{t=3}^{T} \frac{X_{t-2}}{(1-\alpha_2)} \frac{[P(X_t - 1|X_{t-1}, X_{t-2} - 1) - P(X_t|X_{t-1}, X_{t-2})]}{P(X_t|X_{t-1}, X_{t-2})}$$

$$\dot{\ell}_{\lambda} = \sum_{t=3}^{T} \frac{P(X_t - 1|X_{t-1}, X_{t-2}) - P(X_t|X_{t-1}, X_{t-2})}{P(X_t|X_{t-1}, X_{t-2})}$$

respectively. The conditional maximum likelihood estimator (CMLE) for the three parameters are obtained by setting $\dot{\ell}_{\alpha_1}$, $\dot{\ell}_{\alpha_2}$, and $\dot{\ell}_{\lambda}$ equal to zero.

10.2.2 Results

In order to compare the relative performance of the three estimators in small samples, we carry out Monte Carlo experiments to examine the finite sample bias and mean squared error (MSE) of alternative estimators. To achieve this, we generate artificial time series of counts based on the INAR(2) model. As in the previous

section, 153 cases are considered, each of which represents a particular set of thinning coefficients. As mentioned earlier, values of α_1 and α_2 as well as their sum, $(\alpha_1 + \alpha_2)$, are constrained so that each case under study is ensured to be stationary and nondegenerate. For each of the 153 cases, 1000 replications are generated. For each replication, we estimate the model parameters using alternative estimators and calculate the bias and MSE of parameter estimates. Our simulation experiments are performed for samples size $T = 100$, 200, and 500.

To reflect different arrival rates, three rounds of experiments are conducted. Three different values of λ, 0.5, 1 and 2, are used. Not surprisingly, since λ is in general a factor that affects only the magnitude of the variable, our conclusions do not seem to be affected by the choice of it. We thus select to discuss the simulation results for the case $\lambda = 1$, which are typical. The bias results for the three different sample sizes are reported in Table 10.4, 10.5 and 10.6, respectively. Each table contains the results obtained from the three estimators. To save more space, we have selected to report the results for 10 typical cases from the total 153 cases. It can be noted that these presented cases represent combinations of parameter values of typical scales.

It can be seen that except for only a few cases of the CML where $\widehat{\alpha}_1$ is biased up, in all the remaining cases $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ are both biased down and $\widehat{\lambda}$ is biased up. This inverse relationship is to be expected because for a fixed marginal mean of the series X_t, decreasing α_1 and α_2 corresponds to increasing λ. For the CLS and the YW, this can be explicitly noticed from equations defining the two estimators (equation

Table 10.4: Bias Results for the INAR(2) Model ($\lambda = 1$) for Sample Size $T = 100$

	YW				CLS				CML			
		α_2				α_2				α_2		
	α_1	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70
	0.10	-0.0020	-0.0053	-0.0120	-0.0168	-0.0032	-0.0057	-0.0118	-0.0168	0.0019	-0.0018	-0.0080	-0.0177
Bias(α_1)	0.30	-0.0119	-0.0199	-0.0202		-0.0163	-0.0200	-0.0206		-0.0066	-0.0123	-0.0118	
	0.50	-0.0132	-0.0160			-0.0219	-0.0190			-0.0027	-0.0017		
	0.70	-0.0228				-0.0344				-0.0021			
		α_2				α_2				α_2			
	α_1	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70
	0.10	-0.0099	-0.0345	-0.0443	-0.0513	-0.0097	-0.0340	-0.0436	-0.0491	-0.0066	-0.0214	-0.0214	-0.0099
Bias(α_2)	0.30	-0.0139	-0.0387	-0.0513		-0.0133	-0.0377	-0.0491		-0.0104	-0.0264	-0.0231	
	0.50	-0.0145	-0.0409			-0.0135	-0.0374			-0.0142	-0.0291		
	0.70	-0.0130				-0.0111				-0.0233			
		α_2				α_2				α_2			
	α_1	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70
	0.10	0.0201	0.0573	0.1291	0.3286	0.0214	0.0562	0.1267	0.3143	0.0110	0.0283	0.0616	0.1193
Bias(λ)	0.30	0.0415	0.1283	0.3276		0.0485	0.1263	0.3206		0.0274	0.0791	0.1451	
	0.50	0.0658	0.2732			0.0832	0.2700			0.0362	0.1423		
	0.70	0.1602				0.2080				0.1068			

Table 10.5: Bias Results for the INAR(2) Model ($\lambda = 1$) for Sample Size $T = 200$

YW

Bias(α_1)

α_1 \ α_2	0.10	0.30	0.50	0.70
0.10	-0.0007	-0.0026	-0.0067	-0.0054
0.30	-0.0086	-0.0112	-0.0123	
0.50	-0.0049	-0.0096		
0.70	-0.0073			

Bias(α_2)

α_1 \ α_2	0.10	0.30	0.50	0.70
0.10	-0.0061	-0.0185	-0.0220	-0.0275
0.30	-0.0064	-0.0209	-0.0207	
0.50	-0.0107	-0.0208		
0.70	-0.0131			

Bias(λ)

α_1 \ α_2	0.10	0.30	0.50	0.70
0.10	0.0047	0.0289	0.0672	0.1588
0.30	0.0265	0.0699	0.1499	
0.50	0.0342	0.1429		
0.70	0.0983			

CLS

Bias(α_1)

α_1 \ α_2	0.10	0.30	0.50	0.70
0.10	-0.0007	-0.0025	-0.0072	-0.0056
0.30	-0.0101	-0.0113	-0.0125	
0.50	-0.0074	-0.0104		
0.70	-0.0114			

Bias(α_2)

α_1 \ α_2	0.10	0.30	0.50	0.70
0.10	-0.0061	-0.0183	-0.0218	-0.0270
0.30	-0.0065	-0.0207	-0.0201	
0.50	-0.0104	-0.0192		
0.70	-0.0122			

Bias(λ)

α_1 \ α_2	0.10	0.30	0.50	0.70
0.10	0.0047	0.0284	0.0679	0.1579
0.30	0.0288	0.0703	0.1490	
0.50	0.0398	0.1406		
0.70	0.1142			

CML

Bias(α_1)

α_1 \ α_2	0.10	0.30	0.50	0.70
0.10	0.0010	0.0001	-0.0060	-0.0068
0.30	-0.0053	-0.0042	-0.0090	
0.50	-0.0014	0.0004		
0.70	0.0006			

Bias(α_2)

α_1 \ α_2	0.10	0.30	0.50	0.70
0.10	-0.0042	-0.0121	-0.0091	-0.0058
0.30	-0.0043	-0.0156	-0.0072	
0.50	-0.0097	-0.0168		
0.70	-0.0160			

Bias(λ)

α_1 \ α_2	0.10	0.30	0.50	0.70
0.10	0.0002	0.0138	0.0328	0.0582
0.30	0.0171	0.0397	0.0681	
0.50	0.0234	0.0738		
0.70	0.0732			

Table 10.6: Bias Results for the INAR(2) Model ($\lambda = 1$) for Sample Size $T = 500$

		YW				CLS				CML			
		α_2				α_2				α_2			
	α_1	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70
	0.10	-0.0028	-0.0008	-0.0043	-0.0027	0.10	-0.0028	-0.0008	-0.0043	-0.0028	0.10	-0.0022	0.0007	-0.0038	-0.0034
Bias(α_1)	0.30	-0.0027	-0.0032	-0.0043		0.30	-0.0030	-0.0033	-0.0042		0.30	-0.0011	-0.0017	-0.0018	
	0.50	-0.0029	-0.0035			0.50	-0.0031	-0.0036			0.50	-0.0001	0.0007		
	0.70	-0.0033				0.70	-0.0035				0.70	0.0001			
		α_2				α_2				α_2			
	α_1	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70
	0.10	-0.0034	-0.0066	-0.0095	-0.0113	0.10	-0.0034	-0.0066	-0.0093	-0.0112	0.10	-0.0029	-0.0035	-0.0035	-0.0024
Bias(α_2)	0.30	-0.0043	-0.0058	-0.0089		0.30	-0.0043	-0.0058	-0.0089		0.30	-0.0033	-0.0035	-0.0042	
	0.50	-0.0069	-0.0074			0.50	-0.0069	-0.0072			0.50	-0.0062	-0.0056		
	0.70	-0.0056				0.70	-0.0055				0.70	-0.0061			
		α_2				α_2				α_2			
	α_1	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70
	0.10	0.0041	0.0132	0.0339	0.0682	0.10	0.0040	0.0131	0.0338	0.0682	0.10	0.0027	0.0052	0.0180	0.0272
Bias(λ)	0.30	0.0089	0.0225	0.0629		0.30	0.0091	0.0226	0.0632		0.30	0.0042	0.0128	0.0280	
	0.50	0.0230	0.0536			0.50	0.0233	0.0533			0.50	0.0143	0.0231		
	0.70	0.0419				0.70	0.0428				0.70	0.0275			

(10.4) and (10.5)). The relationship between the bias in $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$, however, is less evident from the tables. But a closer examination among all cases studied, including those unreported, also reveals a negative correlation, despite the fact that both are biased down. This inverse relationship is also expected for similar reason. That is, for a fixed marginal mean of X_t and λ, a large α_1 corresponds to a small α_2, and vice versa. For the CLS and the YW, this can also be noted from equation (10.4) and (10.5).

It can be seen from the tables that for all three parameters the biases in the CML estimates are in general smaller than those in the CLS and YW estimates. Consider for example when $T = 100$, only 2 cases of the YW and 3 cases of the CLS have smaller biases than the CML; when $T = 200$, the numbers of such cases are 3 and 3, respectively. For $T = 500$, each has only 2 such cases. In particular, the CML seems to dominate the YW and CLS in terms of the bias of $\widehat{\lambda}$. These results suggest that there is a gain in using the CML over the YW and the CLS in terms of the bias in estimates.

It is clear that for both the CLS and the YW the magnitude of the bias of $\widehat{\alpha}_1$ increases with the increase in α_1 and the magnitude of the bias of $\widehat{\alpha}_2$ increases with the increase in α_2, ceteris paribus. In contrast, the bias of $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ in CML estimates do not show such a tendency for increase in bias. These observations are graphically illustrated by Figure 10.2 and 10.3, which show, respectively, the bias of $\widehat{\alpha}_1$ as a function of α_1 (with $\alpha_2 = 0.3$ and $\lambda = 1$) and the bias of $\widehat{\alpha}_2$ as a function of α_2 (with

Figure 10.2: Bias of $\widehat{\alpha}_1$ as a Function of α_1 for the INAR(2) Model

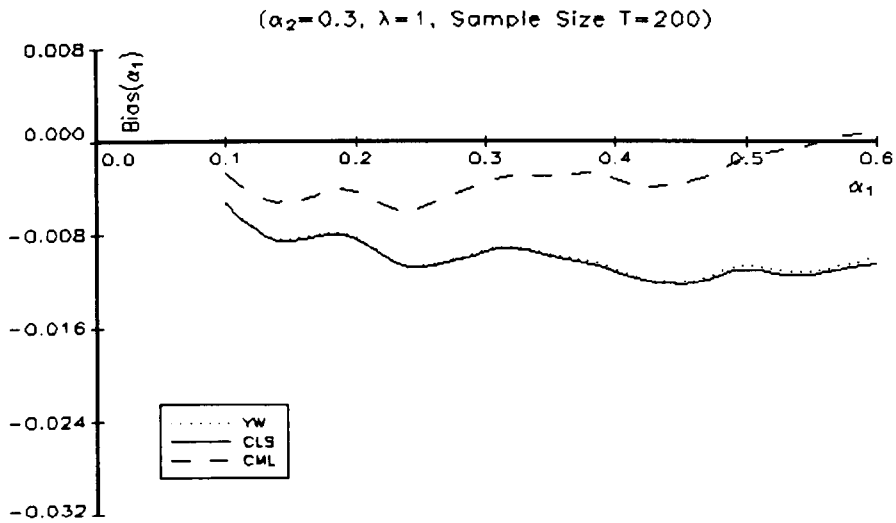
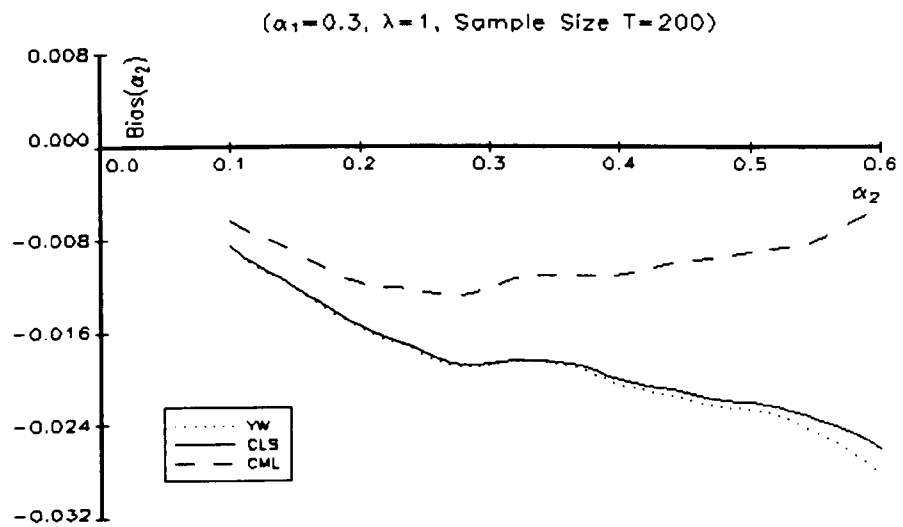


Figure 10.3: Bias of $\widehat{\alpha}_2$ as a Function of α_2 for the INAR(2) Model



$\alpha_1 = 0.3$ and $\lambda = 1$) for the three estimators based on the sample size $T = 200$. In terms of the bias of $\widehat{\lambda}$, we found that for all three methods the magnitude increases with the increase in α_1 and α_2. However, in the CML estimates the increase in the magnitude is much slower than in the YW and CLS estimates. This observation is illustrated by Figure 10.4 and 10.5, which show the biases of $\widehat{\lambda}$ as a function of α_1 (with $\alpha_2 = 0.3$ and $\lambda = 1$) and α_2 (with $\alpha_1 = 0.3$ and $\lambda = 1$), respectively.

Considering the magnitude of the biases in relation to the sample size, we found that in both the YW and the CLS estimates, and to a lesser extent, in the CML estimates, the size of bias is reciprocally related to the sample size. It can be seen that as the sample size increases from 100 to 200 and to 500 the size of bias of each of $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, and $\widehat{\lambda}$ would be reduced by roughly the same proportion. For example, consider the case $\alpha_1 = 0.7$ and $\alpha_2 = 0.1$. When $T = 100$ the bias of $\widehat{\lambda}$ for the YW, CLS and CML estimates are 0.1602, 0.2080 and 0.1068, respectively. As the sample size increases to 200, these biases reduce by approximately one-half to 0.0983, 0.1142, and 0.0732, respectively. As T increases further to 500, they reduce to 0.0419, 0.0428, and 0.0275, respectively, roughly one fifth of the initial size.

The corresponding MSE results for the three different sample sizes are given in Table 10.7, 10.8 and 10.9, respectively. Similar to the bias results, the MSE's in the CML estimates are in general smaller than their counterparts in the CLS and YW estimates for all three parameters. The dominance of the CML becomes extremely clear as the sample size increases. It can be seen from the tables that for $T = 100$,

Figure 10.4: Bias of $\widehat{\lambda}$ as a Function of α_1 for the INAR(2) Model

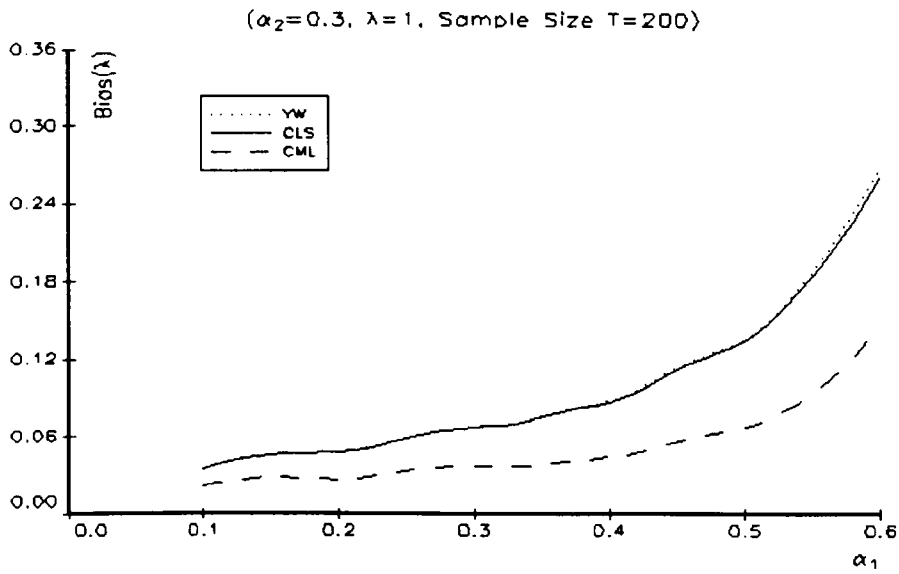


Figure 10.5: Bias of $\widehat{\lambda}$ as a Function of α_2 for the INAR(2) Model

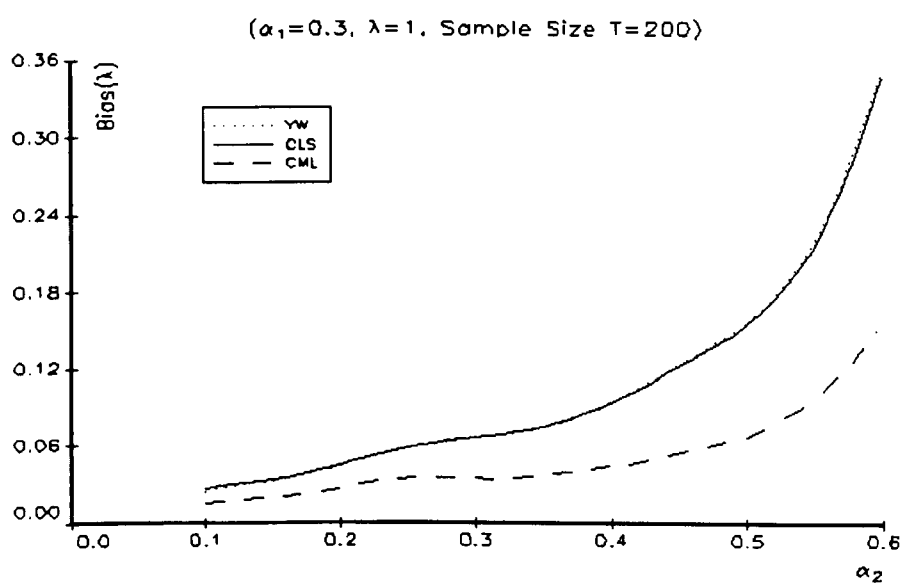


Table 10.7: MSE Results for the INAR(2) Model ($\lambda = 1$) for Sample Size $T = 100$

	YW				CLS				CML			
		α_2				α_2				α_2		
	α_1	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70
	0.10	0.0076	0.0074	0.0061	0.0053	0.0073	0.0074	0.0061	0.0053	0.0080	0.0078	0.0065	0.0051
MSE(α_1)	0.30	0.0121	0.0121	0.0094		0.0115	0.0122	0.0095		0.0103	0.0112	0.0101	
	0.50	0.0110	0.0117			0.0101	0.0119			0.0063	0.0085		
	0.70	0.0119				0.0102				0.0030			
		α_2				α_2				α_2			
	α_1	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70
	0.10	0.0074	0.0113	0.0114	0.0091	0.0074	0.0113	0.0111	0.0087	0.0077	0.0099	0.0071	0.0028
MSE(α_2)	0.30	0.0071	0.0119	0.0117		0.0071	0.0118	0.0112		0.0075	0.0113	0.0088	
	0.50	0.0068	0.0123			0.0067	0.0119			0.0068	0.0123		
	0.70	0.0070				0.0070				0.0052			
		α_2				α_2				α_2			
	α_1	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70	0.10	0.30	0.50	0.70
	0.10	0.0324	0.0481	0.1020	0.3202	0.0330	0.0487	0.1015	0.3166	0.0329	0.0417	0.0689	0.1302
MSE(λ)	0.30	0.0485	0.0991	0.3379		0.0496	0.0995	0.3388		0.0427	0.0739	0.1557	
	0.50	0.0665	0.2931			0.0708	0.3013			0.0503	0.1608		
	0.70	0.1947				0.2059				0.1100			

Table 10.8: MSE Results for the INAR(2) Model ($\lambda = 1$) for Sample Size $T = 200$

	YW				CLS				CML			
		α_2				α_2				α_2		
	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70
	0.10	0.0045	0.0042	0.0037	0.0028	0.10	0.0044	0.0043	0.0038	0.0029	0.10	0.0043	0.0043	0.0035	0.0026
MSE(α_1)	0.30	0.0057	0.0053	0.0048		0.30	0.0055	0.0054	0.0048		0.30	0.0048	0.0050	0.0046	
	0.50	0.0057	0.0052			0.50	0.0054	0.0052			0.50	0.0034	0.0039		
	0.70	0.0056				0.70	0.0050				0.70	0.0013			
		α_2					α_2					α_2			
	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70
	0.10	0.0046	0.0054	0.0049	0.0039	0.10	0.0047	0.0054	0.0049	0.0038	0.10	0.0048	0.0046	0.0030	0.0012
MSE(α_2)	0.30	0.0042	0.0057	0.0050		0.30	0.0042	0.0057	0.0050		0.30	0.0043	0.0051	0.0037	
	0.50	0.0046	0.0053			0.50	0.0046	0.0052			0.50	0.0043	0.0050		
	0.70	0.0042				0.70	0.0041				0.70	0.0028			
		α_2					α_2					α_2			
	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70
	0.10	0.0163	0.0252	0.0481	0.1353	0.10	0.0164	0.0253	0.0482	0.1346	0.10	0.0161	0.0225	0.0330	0.0657
MSE(λ)	0.30	0.0235	0.0454	0.1263		0.30	0.0240	0.0453	0.1260		0.30	0.0209	0.0328	0.0682	
	0.50	0.0335	0.1065			0.50	0.0349	0.1065			0.50	0.0269	0.0594		
	0.70	0.0804				0.70	0.0838				0.70	0.0548			

Table 10.9: MSE Results for the INAR(2) Model ($\lambda = 1$) for Sample Size $T = 500$

	YW					CLS					CML				
		α_2					α_2					α_2			
	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70
	0.10	0.0021	0.0020	0.0017	0.0011	0.10	0.0021	0.0020	0.0017	0.0012	0.10	0.0021	0.0019	0.0016	0.0010
MSE(α_1)	0.30	0.0022	0.0020	0.0017		0.30	0.0022	0.0020	0.0017		0.30	0.0019	0.0018	0.0016	
	0.50	0.0025	0.0021			0.50	0.0024	0.0021			0.50	0.0014	0.0015		
	0.70	0.0022				0.70	0.0021				0.70	0.0006			
		α_2					α_2					α_2			
	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70
	0.10	0.0022	0.0023	0.0020	0.0013	0.10	0.0022	0.0023	0.0020	0.0013	0.10	0.0022	0.0018	0.0012	0.0005
MSE(α_2)	0.30	0.0021	0.0021	0.0019		0.30	0.0021	0.0021	0.0019		0.30	0.0020	0.0019	0.0014	
	0.50	0.0021	0.0020			0.50	0.0021	0.0020			0.50	0.0018	0.0018		
	0.70	0.0021				0.70	0.0021				0.70	0.0012			
		α_2					α_2					α_2			
	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70	α_1	0.10	0.30	0.50	0.70
	0.10	0.0079	0.0121	0.0198	0.0450	0.10	0.0080	0.0121	0.0198	0.0452	0.10	0.0077	0.0098	0.0136	0.0265
MSE(λ)	0.30	0.0101	0.0156	0.0378		0.30	0.0102	0.0156	0.0379		0.30	0.0087	0.0113	0.0208	
	0.50	0.0143	0.0332			0.50	0.0145	0.0328			0.50	0.0112	0.0182		
	0.70	0.0285				0.70	0.0287				0.70	0.0203			

there are only 8 cases, respectively, for the YW and the CLS, which appear to be (very marginally) smaller than their CML counterparts. But as T increases to 200, the number of such cases reduces to only 3; For $T = 500$, the YW and the CLS are completely dominated by the CML. There is clearly a gain in implementing the CML in terms of the MSE, especially when one has a reasonably large sample size.

In contrast to the bias results above, for all three estimation methods the magnitude of MSE of $\widehat{\alpha}_1$ does not increase with the increase in α_1. Neither does the magnitude of MSE of $\widehat{\alpha}_2$ arises in α_2. However, the magnitude of the MSE of $\widehat{\lambda}$ still increases with the increase in both α_1 and α_2. These observations for $\widehat{\lambda}$ are graphically illustrated in Figure 10.6 and 10.7. It can be seen that as in the bias results the pace of increase in MSE for the CML is clearly lower than the CLS and the YW.

Similar to the bias results, for all three methods of estimation, the MSE of each parameter is reciprocally related to the sample size. The pace of decrease in the MSE as the result of increase in the sample size is also similar to the bias results.

In comparing the three methods of estimation in the light of the results of the simulation experiment, it seems that the CML method is worth the effort. Not only does the benefit from implementing the CML, in terms of bias and MSE, become substantial as sample size increases, but there is also a considerable gain when estimating relatively persistent processes. Consider only the $T = 100$ case. When $\alpha_1 = \alpha_2 = 0.1$, for instance, the bias of $\widehat{\alpha}_1$ in the CML estimate is approximately 95% of the corresponding YW bias and 60% of the CLS bias. For fixed value of α_2, this percentage

Figure 10.6: MSE of $\widehat{\lambda}$ as a Function of α_1 for the INAR(2) Model

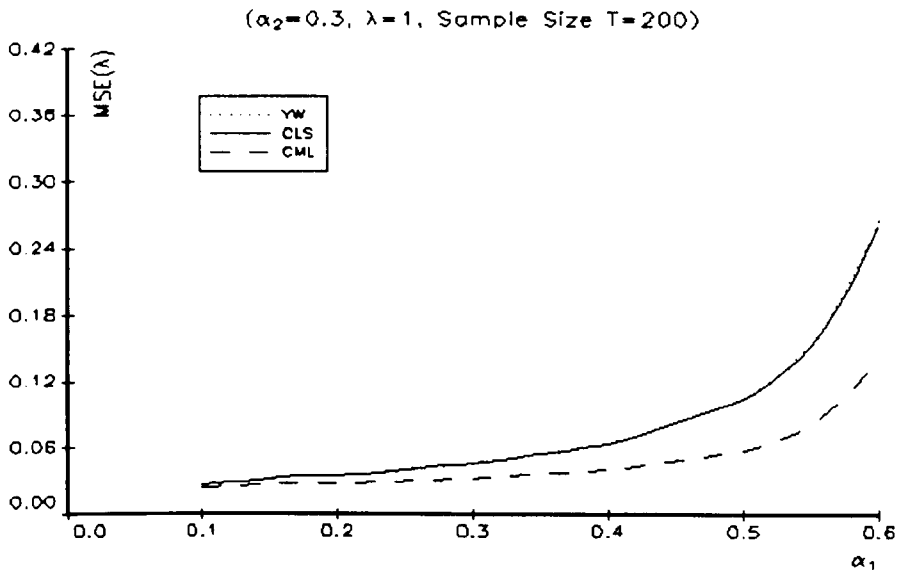
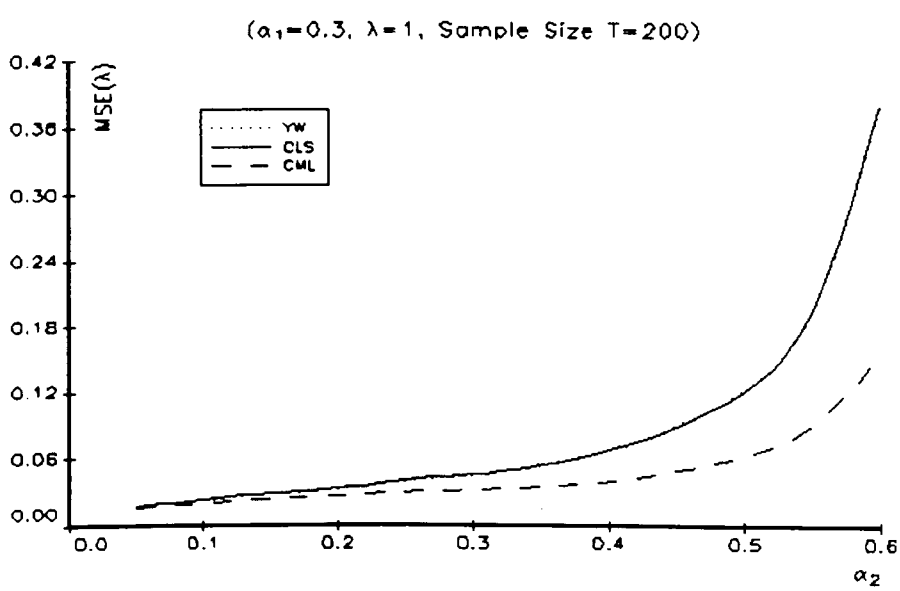


Figure 10.7: MSE of $\widehat{\lambda}$ as a Function of α_2 for the INAR(2) Model



decreases monotonically with the increase in the value of α_1, reaching only about 9% and 6%, respectively, when $\alpha_1 = 0.7$. Similarly, the bias of $\widehat{\alpha}_2$ in the CML estimate is about two thirds of the YW and CLS bias when $\alpha_1 = \alpha_2 = 0.1$. This percentage drops to around one fifth when $\alpha_2 = 0.7$. However, it should be noted that such relative gain in terms of bias of $\widehat{\lambda}$ is less sensitive to the value of α_1 and α_2.

In terms of the MSE, it can be seen that at $\alpha_1 = \alpha_2 = 0.1$, the MSE of $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, and $\widehat{\lambda}$ in the CML estimate are even slightly greater than (or very close to) their YW and CLS counterparts. But the relative efficiency gain of using the CML starts to emerge and continues to grow as the process approaches to high degree of persistency, in the direction of either α_1 or α_2 or both.

It should be mentioned in passing that for very small (close to zero) values of α_1 or α_2 there are actually some noticeable but not substantial gain, in terms of bias and MSE, of the YW and CLS estimates over the CML method. But it diminishes very rapidly with the increase in the sample size.

10.3 Conclusion

In this chapter, we examined both asymptotic and finite sample performance of the MLE. Using the INAR(2) specification with Poisson innovations, we investigated the asymptotic gain of implementing the ML method over the commonly used CLS method by calculating the ARE between the two estimators. Our results confirm that the proposed MLE is asymptotically more efficient than the CLSE and the efficiency

gain is most substantial for persistent processes. According to the results from our finite sample simulation experiments, we feel that given the potential gain, in terms of the bias and MSE, in the CML estimates compared with the YW and CLS methods, it is worth the effort to implement the CML method. This is particularly true if the size of the sample is reasonably large. The gain of implementing the CML could be substantial. Moreover, extra benefit may be achieved if the processes under study show certain degrees of persistence.

Chapter 11

Coherent Forecasting with the INAR(p) Model

One of the objectives of modelling time series data is to forecast the future values of the variables of interest. This chapter is concerned with forecasting time series of count data based on the INAR(p) model. Freeland and McCabe (2004b) suggest that for count data model forecasts be provided for each point mass of the distribution and using median as coherent point forecast. They also suggest that the probabilities associated with each point mass be modified to reflect the variation in parameter estimation. However, Freeland and McCabe (2004b) is concerned only with the PoINAR(1) model. The main contribution of this chapter is to extend their ideas to a general INAR(p) model. We begin this chapter by a brief discussion of alternative predictors based on single summary statistic of forecast distribution and

their limitations. In Section 11.2, we present a method for producing h-step ahead forecasts of the conditional probability distribution of the INAR(p) process. Meanwhile, we also consider how the model parameter uncertainty can be reflected in the confidence intervals for the probability forecasts. Section 11.3 summarizes. The main results of this chapter appear in Bu et al. (2006c).

11.1 Forecasting using Conditional Mean, Median and Mode

The most common procedure for constructing forecasts in time series models is to use conditional expectations. The reason is that this technique will yield forecasts with minimum mean squared forecast error. Consider a realization $\{X_t\}_{t=1}^{T}$ from a discrete time stochastic process. Then it can be shown that the forecast, $\tilde{X}_{T+h}$, of X_{T+h} that minimizes the expected mean squared forecast error

$$E\left[\left(X_{T+h} - \tilde{X}_{T+h}\right)^2 | \{X_t\}_{t=1}^{T}\right]$$

is the mean of the h-step ahead conditional distribution. But this method lacks data coherency when the time series under consideration has restrictions on its support. In the count data context, the process consists of only integer values and therefore in order to generate data coherent forecasts we seek a method of forecasting that produces only integer values.

For this reason, Freeland and McCabe (2004b) suggest that the h-step ahead

conditional distribution itself be used to produce coherent forecasts. One obvious idea is to use the median of the forecast distribution. The coherence of the median is given by the fact that it almost always lies in the support of the distribution when the variable is discrete and the cardinality of support is small. It can also be shown that the median has the optimality property of minimizing the expected absolute forecast error. That is, the forecast, $\tilde{X}_{T+h}$, of X_{T+h} that minimizes the expected absolute error

$$E\left[|X_{T+h} - \tilde{X}_{T+h}|| \{X_t\}_{t=1}^{T}\right]$$

is the median of the h-step ahead conditional distribution.

Freeland and McCabe (2004b) also pointed out that despite being data coherent it can be quite misleading to summarize an entire distribution by a single point. They exemplified the problem by the following two situations: in the first case, $P(X = 0) = 1 - P(X = 1) = 0.5$ while in the second $P(X = 0) = 1 - P(X = 5) = 0.9$. In both cases the median of X is 0 (the mean is 0.5), but in the second case, there is almost twice the probability of observing a zero. Since there are only 2 outcomes in these examples it would be more informative to give the probability distribution for both values in the support.

Another data coherent distribution statistic is the mode, which is the value of a random variable that occurs with the greatest probability. The mode of the forecast distribution may serve as an alternative data coherent predictor, but as in the case of the median, it also ignores the probability distribution for values other than the

mode.

11.2 Forecasting Conditional Distribution with the INAR(p) Model

To generate data coherent predictions, Freeland and McCabe (2004b) suggest using the h-step ahead conditional distribution and its median as a point forecast. For the PoINAR(1) model, Freeland and McCabe (2004b) presented a method of computing the conditional forecast distribution on the basis of estimated parameters and they discussed its statistical properties. Following their ideas, McCabe and Martin (2005) explored the issue of coherent forecasting with count data models under the Bayesian framework. But they are concerned only with first-order INAR models and their method is based on computer intensive numerical evaluation. More recently, Jung and Tremayne (2006) proposed a simulation based method for producing coherent forecasts for higher-order INAR models, which also requires considerable computational work. The principal intention of this section is to extend the ideas of Freeland and McCabe (2004b) and develop an efficient procedure for producing coherent forecasts with higher-order INAR models.

Coherent forecasting requires the information about the conditional forecast distribution of the count variable at subsequent periods. It can be easily noticed that the one-step ahead conditional probability is simply the transition probability of the

process. For a count series X_t which follows an INAR(p) process defined in (9.1), it follows that the probability mass function of X_{T+1} conditioned on $\{X_t\}_{t=1}^{T}$ is given by $P(X_{T+1}|X_T, \ldots, X_{T-p+1})$, which by definition is the probability of the value X_t occurring at $T+1$, according to the one-step ahead conditional distribution. Efficient procedure for computing the transition probability has been discussed in Chapter 9.

In principle, multiple-step ahead forecasting requires the information about the conditional forecast distribution of the count variable at multiple periods ahead. In the relatively simple case of PoINAR(1) model, the required distribution is a convolution of a Poisson and a Binomial random variable. Freeland and McCabe (2004b) derived the conditional probability mass function of X_{T+h} given X_T for any non-negative integer value h in an analytic form. However, for the higher-order models of principal concern here, the analytic expression for the required conditional probability function is not easily available. In what follows, we present an efficient procedure for producing distribution forecasts based on the INAR(p) model. It will become clear soon that the transition probability function forms the basis of this procedure.

11.2.1 Forecasting Count Data: A Markov Chain Approach

The Markov chain is a probabilistic model used to represent dependence between successive observations of a random variable. It is widely used in many disciplines. Comprehensive treatments of Markov chains and their applications can be found in, for example, Kemeny and Snell (1976), Kemeny et al. (1976), Karlin and Taylor

(1981) and Brémaud (1999). It is easily seen that the INAR model is a special type Markov chain and the process generated by an INAR model can be regarded as a special case of Markov system, which is by definition a system that can be in one of several (numbered) states, and can pass from one state to another at each time step according to fixed probabilities, the transition probabilities. In this section, we present a method for producing conditional probability forecasts for higher-order INAR models using Markov chain techniques.

In theory, since the support for an INAR(p) variable is from zero to infinity, there are infinite possible states in the system. But in practice, for any stationary INAR processes (at least for most of the count processes one might encounter in reality), there exists a sufficiently large positive integer M such that the probability of observing a count larger than M is negligible. Therefore, for a count series X_t which follows an INAR(p) process, we can assume that X_t takes values in the finite set $\Omega = \{0, \ldots, M\}$. Therefore, it can be easily verified that at any given period t there are $(M+1)^p$ different states, determined by $\{X_{t-p}, X_{t-p+1}, \ldots, X_t\}$, in the system generated by an INAR(p) model.

For a Markov system with finite states, the forecast distribution of each state at any time t can be obtained by means of the transition matrix method. Let $Q_{p,M}$ denote the transition matrix of an INAR(p) model with the maximum possible count M. It can be noted that $Q_{p,M}$ is a $(M+1)^p \times (M+1)^p$ probability matrix. Consider for example an INAR(2) process where $M = 1$. The corresponding transition matrix

$Q_{2,1}$ can be written as

$$Q_{2,1} = \begin{array}{cc} & X_t \\ X_{t-2} & X_{t-1} \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{array} \\ \begin{bmatrix} P(0|0,0) & 0 & P(1|0,0) & 0 \\ P(0|0,1) & 0 & P(1|0,1) & 0 \\ 0 & P(0|1,0) & 0 & P(1|1,0) \\ 0 & P(0|1,1) & 0 & P(1|1,1) \end{bmatrix} \end{array} \quad (11.1)$$

Since this is a 2nd-order Markov system, the state of the system at time t is jointly determined by the values of X_{t-1} and X_t, and can be denoted by (X_{t-1}, X_t). As shown in (11.1), there are 4 different possible states for the system. They are $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$. Each element in $Q_{2,1}$ represents the probability of the system going from one state into another. For instance, if the system is currently in state $(0,0)$, then the probability of the system going from state $(0,0)$ to state $(0,1)$ in the next period is given by $P(1|0,0)$ and the probability of the system remaining in $(0,0)$ in the next period is given by $P(0|0,0)$, etc. A Markov system is said to be homogenous if the transition matrix, $Q_{p,M}$, is time-invariant. Note that for a process with order greater than 1, as in the example, the transition matrix would generally contain several elements corresponding to transitions that cannot occur. The probability of these transitions are then 0, sometimes called "structural zero".

Recall that coherent forecasting aims to produce forecasts of the probability of each value occurring at the forecast horizon, which is in turn determined by the probability of each state occurring at the forecast horizon. To get probability forecasts for each state, we define for a system with $(M+1)^p$ states a $1 \times (M+1)^p$ probability vector,

π_t, which represents the probabilities of finding a system in each of the $(M+1)^p$ different states at a given period t. In the above example, the probability vector is a 1×4 vector which can be written as $\pi_t = \left(P_t^{0,0}, P_t^{1,0}, P_t^{0,1}, P_t^{1,1}\right)$. The elements in π_t represent the probability of the system in state $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$, respectively. It can be noted that at time t the probability of $X_t = 0$ occurring is equal to $P_t^{0,0} + P_t^{1,0}$, the sum of the probabilities of the system in state $(0,0)$ and $(1,0)$. Similarly, the probability of $X_t = 1$ occurring is given by $P_t^{0,1} + P_t^{1,1}$, the sum of the probabilities of the system in state $(0,1)$ and $(1,1)$. To formalize this principle, we define, corresponding to π_t, a $(M+1)^p \times 1$ selection vector v_i, $i = 0, 1, \ldots, M$, which has $M+1$ entries equal to one and all others equal to zero. Each of the $M+1$ entries corresponds to one of the $M+1$ states that involve $X_t = i$. In our example, $v_0 = (1,1,0,0)'$ and $v_1 = (0,0,1,1)'$. Thus, the probabilities of $X_t = 0$ and $X_t = 1$ can be written as $\pi_t v_0$ and $\pi_t v_1$, respectively.

It follows from the above reasoning that for a general INAR(p) process the conditional probability forecasts for X_{T+h} can be obtained from the forecasts of the probability vector π_{T+h}. That is

$$P\left(X_{T+h} = i | X_T, \ldots, X_{T-p+1}\right) = \pi_{T+h} v_i.$$

The following theorem thus forms the basis of forecasting conditional probability distributions for the INAR(p) process.

Theorem 11.1 *Let Q_p and $Q_p^{(h)}$ denote, respectively, the one-step transition matrix and h-step transition matrix for a homogeneous pth-order Markov system. It can be*

shown that

$$Q_p^{(h)} = Q_p^{(h-1)} Q_p = Q_p^h. \tag{11.2}$$

Let π_T *and* π_{T+h} *denote the probability vector at time* T *and* $T+h$*, respectively, we have*

$$\pi_{T+h} = \pi_{T+h-1} Q_p = \pi_T Q_p^h. \tag{11.3}$$

Equation (11.2) says that the h-step transition matrix is equal to the hth power of the one-step transition matrix and Equation (11.3) says that the h-step ahead forecast of the probability vector π_{T+h} is equal to the current probability vector π_T times the h-step transition matrix.

It should be mentioned that π_T is also known as the current state vector. Consider the system in the above example. If we observe that the last two observations of the series X_t are $X_{T-1} = X_T = 0$, for instance, it implies that the system is currently in state $(0,0)$ with probability 1. Therefore, we have $P_T^{0,0} = 1$ and $P_T^{1,0} = P_T^{0,1} = P_T^{1,1} = 0$. The probability vector at time T is thus given by $\pi_T = (1,0,0,0)$, which may be regarded as an indexing vector of the current status of the system.

To formalize the idea of forecasting conditional probability for the INAR(p) model, we have the following theorem.

Theorem 11.2 *For a general INAR(p) process with maximum possible count assumed to be* M*, the* h*-step ahead forecast of the conditional probability of* $X_{T+h|T} = i$ *is given by*

$$P\left(X_{T+h} = i | X_T, \ldots, X_{T-p+1}\right) = \pi_T Q_{p,M}^h v_i.$$

If we define a vector $\mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1})$ *such that*

$$\mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1}) = \begin{pmatrix} P(X_{T+h}=0|X_T,\ldots,X_{T-p+1}) \\ P(X_{T+h}=1|X_T,\ldots,X_{T-p+1}) \\ \vdots \\ P(X_{T+h}=M|X_T,\ldots,X_{T-p+1}) \end{pmatrix}$$

then

$$\mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1}) = \left(\pi_T Q_{p,M}^h \mathbf{v}\right)'$$

where $\mathbf{v}$ *is a selection matrix given by* $\mathbf{v} = (v_0, v_1, \ldots, v_M)$.

Consider the example in (11.1). The conditional probability forecasts of X_{T+h} given $\{X_t\}_{t=1}^T$ are given by

$$\begin{aligned} P(X_{T+h}=0|X_T,X_{T-1}) &= \pi_T Q_{2,1}^h v_0 \\ P(X_{T+h}=1|X_T,X_{T-1}) &= \pi_T Q_{2,1}^h v_1 \end{aligned}$$

or

$$\mathbf{P}(X_{T+h}|X_T,X_{T-1}) = \left(\pi_T Q_{2,1}^h \mathbf{v}\right)'$$

where

$$\mathbf{v} = (v_0, v_1) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

11.2.2 Forecasting Count Data When Parameters are Estimated

If the parameters of the model were known it would be easy to calculate the conditional probability forecasts $\mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1})$ directly using the results from Theorem 11.2. However, in almost all practical applications these parameters are unknown and have to be estimated. Therefore, it is important that this source of variation be accounted for when producing forecasts. Since forecasts are integers, Freeland and McCabe (2004b) suggest that the δ-method be used to modify the probabilities associated with each point mass to reflect the variation in parameter estimation.

As before, we denote $\boldsymbol{\theta}$ as the parameter vector of the INAR(p) model. The h-step ahead forecast of the conditional probability mass function can be written as $\mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1};\boldsymbol{\theta})$. Under standard regularity conditions, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ is asymptotically normally distributed around the true parameter value, i.e. $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \overset{a}{\sim} N(\mathbf{0},\mathbf{i}^{-1})$ where $\mathbf{i}$ is the Fisher information matrix. Let $g(\hat{\boldsymbol{\theta}})$ denote a continuous and differentiable function of the parameter estimates $\hat{\boldsymbol{\theta}}$. The δ-method is a technique for finding the asymptotic distribution of $g(\hat{\boldsymbol{\theta}})$ given that the distribution of $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ is asymptotically normal. The idea in the present context is to apply the δ-method to $g(\hat{\boldsymbol{\theta}}) = \mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1};\hat{\boldsymbol{\theta}})$.

Let $\hat{\boldsymbol{\theta}}_{ML}$ be the ML estimators of $\boldsymbol{\theta}$ in the INAR(p) model based on a sample of size T; Assuming that standard regularity conditions are satisfied, $\hat{\boldsymbol{\theta}}_{ML} \overset{a}{\sim} N[\boldsymbol{\theta}_0,$

$T^{-1}\mathbf{i}^{-1}]$. According to Theorem 11.2, the ML estimate of the h-step ahead probability mass is given by $\mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1};\hat{\boldsymbol{\theta}}_{ML})$. An application of the δ-method gives the asymptotic distribution of this quantity for values of X_{T+h}. From this we may compute a confidence interval for the probability associated with each value of X_{T+h} in the forecast distribution. Obviously these intervals may be truncated outside $[0,1]$.

Theorem 11.3 *For the INAR(p) model, ML estimates of the h-step ahead forecast,* $\mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1};\hat{\boldsymbol{\theta}}_{ML})$, *has an asymptotically normal distribution with mean vector*

$$\boldsymbol{\mu}(\boldsymbol{\theta}_0) = \mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1};\boldsymbol{\theta}_0)$$

and variance matrix

$$\mathbf{V}(\boldsymbol{\theta}_0) = T^{-1}\left(\frac{\partial \mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\mathbf{i}^{-1}\frac{\partial \mathbf{P}(X_{T+h}|X_T,\ldots,X_{T-p+1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

where $\mathbf{i}$ *is the Fisher information matrix and* $\mathbf{P}(X_{T+1}|X_T,\ldots,X_{T-p+1};\boldsymbol{\theta})$ *is given by Theorem 11.2.*

Note that since $Q^h_{P,M}$ is only a matrix of transition probabilities, the partial derivative

$$\frac{\partial \mathbf{P}(X_{T+1}|X_T,\ldots,X_{T-p+1};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial\left[\left(\pi_T Q^h_{P,M}\mathbf{v}\right)'\right]}{\partial \boldsymbol{\theta}'} \tag{11.4}$$

may be obtained analytically by applying the chain rule of differentiation if analytic derivatives of the transition probability function are available. The INAR(p) model with Binomial thinning and Poisson innovations is one of such cases.

According to Theorem 11.3, the 95% confidence intervals for the conditional probability forecast $P(X_{T+h} = i|X_T, \ldots, X_{T-p+1}; \boldsymbol{\theta}_0)$ for $i = 0, 1, \ldots, M$, can be computed based on its asymptotic distribution by means of

$$P(X_{T+1} = i|X_T, \ldots, X_{T-p+1}; \hat{\boldsymbol{\theta}}_{ML}) \pm 2\sigma_{i+1}(\hat{\boldsymbol{\theta}}_{ML})$$

where $\sigma_{i+1}(\hat{\boldsymbol{\theta}}_{ML})$ is the $(i+1, i+1)$ element of $\mathbf{V}(\hat{\boldsymbol{\theta}}_{ML})$. Moreover, it is even possible to get the joint confidence interval over the support of the forecast distribution. But what should be noted here is that in contrast to the treatment of continuous variables, it is the probability of the values that is modified to reflect the uncertainty in parameter estimation.

11.3 Conclusion

In this chapter we extended the ideas of Freeland and McCabe (2004b) and developed a method for producing data coherent forecasts for higher-order INAR models. We showed that the INAR(p) process can be regarded as a special type of Markov system and the distribution forecasts for a count series can be obtained by means of transition matrix method. A procedure based on the δ-method for calculating confidence intervals for these forecast probabilities is also suggested. Since our method is based on explicit formulation, not simulations, it is computationally efficient. An application of this procedure to the Westgren data is presented in Section 12.4 of the following chapter.

Chapter 12

Application to Westgren Data

In this chapter, we analyze the Westgren (1916) gold particle data under the maximum likelihood framework developed in Chapter 9. Some description of the Westgren data is presented in Section 12.1, where we show the presence of serial dependence. We estimate an INAR model using conditional maximum likelihood. Results are discussed in Section 12.2. Section 12.3 assesses the adequacy of the fitted model by both residual analysis and specification testing. We find that new evidence has emerged in the light of the new framework. In Section 12.4 we apply the method introduced in Chapter 11 for producing forecasts for the Westgren data. A brief summary is given in the concluding section.

12.1 The Data

The data used in this study consists of 380 counts of gold particles in a well-defined colloidal solution at equidistant points in time. The set of data was originally published in Westgren (1916). Although not a data set of economic content, it is something of a classic in the history of time series and serves admirably for the purpose at hand. The values used constitute part of a data set that has been used in both the branching process and time series literatures and analyzed by Guttorp (1991) and Grunwald et al. (2000). The most recent work of analyzing this data set is provided by Jung and Tremayne (2006), in which they propose a computer intensive method for generating coherent predictions for INAR models and the Westgren data is used as a testbed.

In the first instance we conduct some preliminary analysis to get an overall picture of the data at hand. Figure 12.1 provides the time series plot of the first $T = 370$ data points (the last 10 are used in forecasting), which shows no discernible trend or seasonal patterns. A summary of simple descriptive statistics for the data is given in Table 12.1. It can be seen that the observed counts vary from 0 to 7 with the sample mean and variance equal to 1.55 and 1.65, respectively. This suggests that there is very little or no overdispersion in the data. The marginal distribution of these data is depicted in Figure 12.2.

As a natural first step of modelling time series data, we plot the sample autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) in Figure

Figure 12.1: Time Series Plot of 370 Observations of the Westgren Data

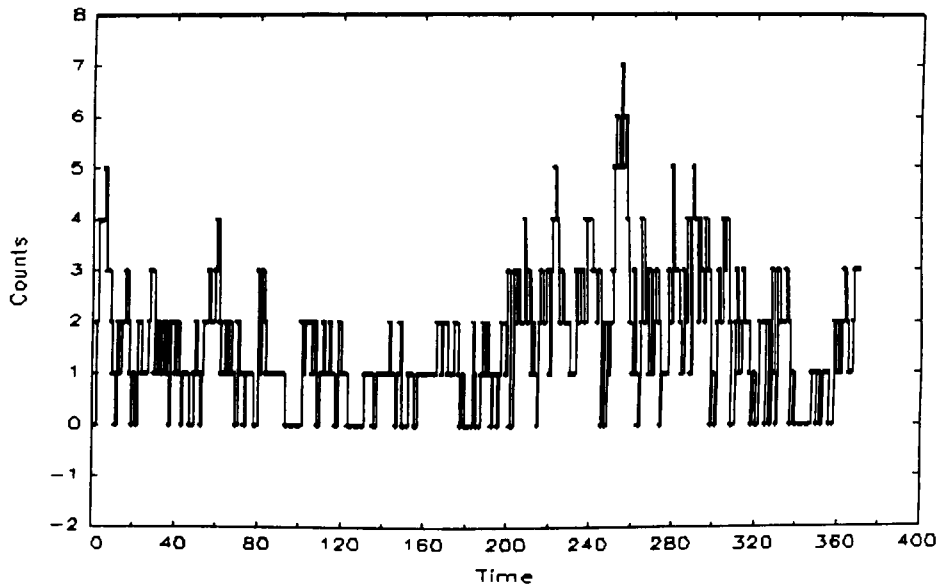
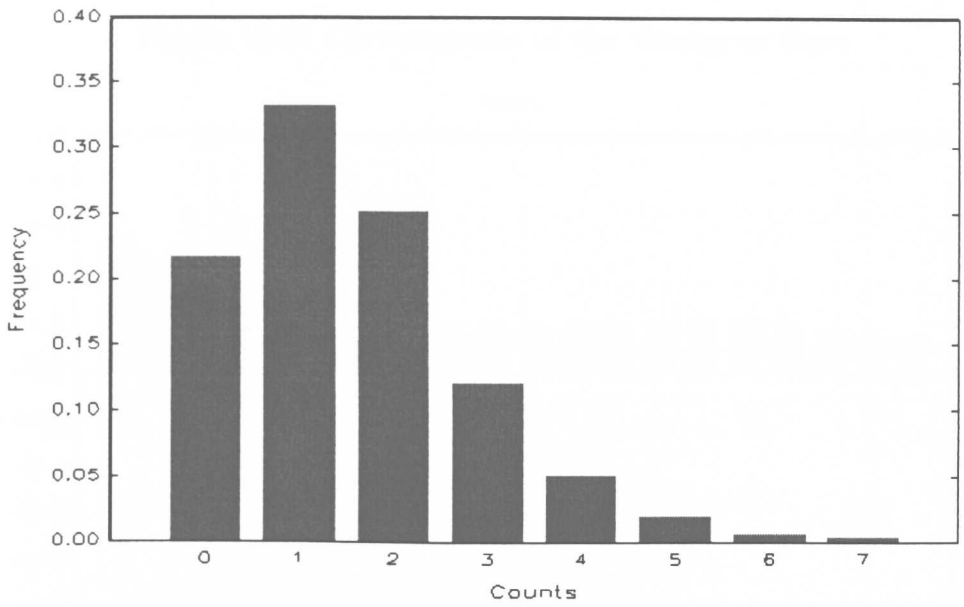


Table 12.1: Descriptive Statistics of the Westgren Data

Minimum	Maximum	Median	Mode	Mean	Variance
0	7	1	1	1.55	1.65

12.3. The sample ACFs confirm the presence of serial correlations, but little or no seasonal patterns are found. The sample PACFs suggest that an autoregressive model with dependence of order 2 is appropriate. Various tests have been proposed in the literature for detecting presence of serial dependence in time series of counts. Examples include the simple run test of Wald and Wolfowitz (1940), the modified score test of Freeland (1998) and the adapted portmanteau tests of Venkataraman (1982) and Mills and Seneta (1989). Jung and Tremayne (2003) provided an excellent survey of

Figure 12.2: Marginal Distribution of the Westgren Data

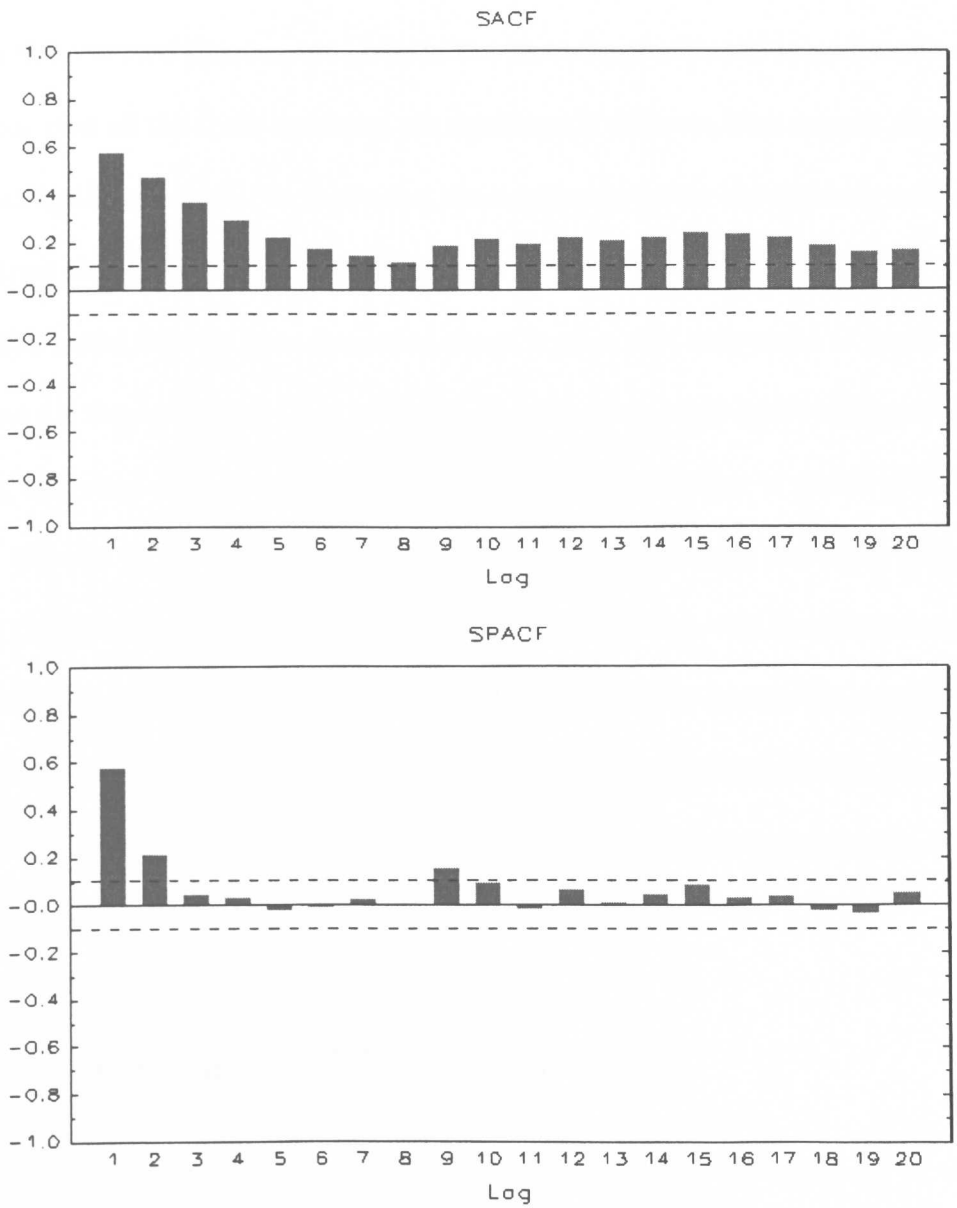


these tests. The data at hand has also been examined by Jung and Tremayne (2006) for the presence of serial dependence by means of different tests. Their results also confirmed that there is significant serial dependence in the series. Using the method of moments, they estimated both the INAR(1) model and the INAR(2) model. By means of residual analysis, they also suggested that the INAR(2) model be used.

12.2 Estimation by Maximum Likelihood

For reasons discussed above, we proceed by estimating a second order INAR model using conditional maximum likelihood. This requires the arrival process to

Figure 12.3: Correlograms of the Westgren Data



be fully specified. The Poisson distribution is a common assumption. The CML estimate of the parameters are $\widehat{\alpha}_1 = 0.4716(0.0472)$, $\widehat{\alpha}_2 = 0.1798(0.0535)$, and $\widehat{\lambda} = 0.5450(0.0724)$. The estimated asymptotic standard errors, which are obtained from the observed Hessian, are given in the parenthesis adjacent to each estimate. It is clear that all the three estimates are significantly different from zero at all conventional significance levels. In particular, the significance of the two thinning coefficients confirms that there is indeed dependence in the data.

The model that we have estimated above implies that each count at a given time period t is thinned twice: Once at time $t+1$ with an estimated probability of around 47%; and once at period $t+2$ with an estimated probability of about 18%. Note that the sum of the two thinning coefficients is relatively large, over 0.65, indicating that the series shows a relative high degree of persistence. On the other hand, the estimated arrival rate, $\widehat{\lambda}$, is around 0.5, indicating that on average in approximately every two periods there will be a new particle entering the observation area. Note that the unconditional mean of the process, the average number of particles observed in each period, implied by the parameter estimates is 1.56.

12.3 Testing for Model Adequacy

In modelling time series data, it is important to assess the adequacy of the fit. Generally, this is accomplished by means of residual analysis and specification testing. By checking the serial dependence of the residuals from the fitted models, Jung and

Tremayne (2006) suggest that the INAR(2) model is appropriate. In this section, we examine the issue of adequacy of the INAR(2) model estimated by ML. As shown in Chapter 9, the ML method enables new tools for statistical inferences to be used. As a result, new evidence emerges from the following two sections in the light of the newly developed ML framework.

12.3.1 Residual Analysis

The estimated residuals of the fitted INAR(2) model are defined as

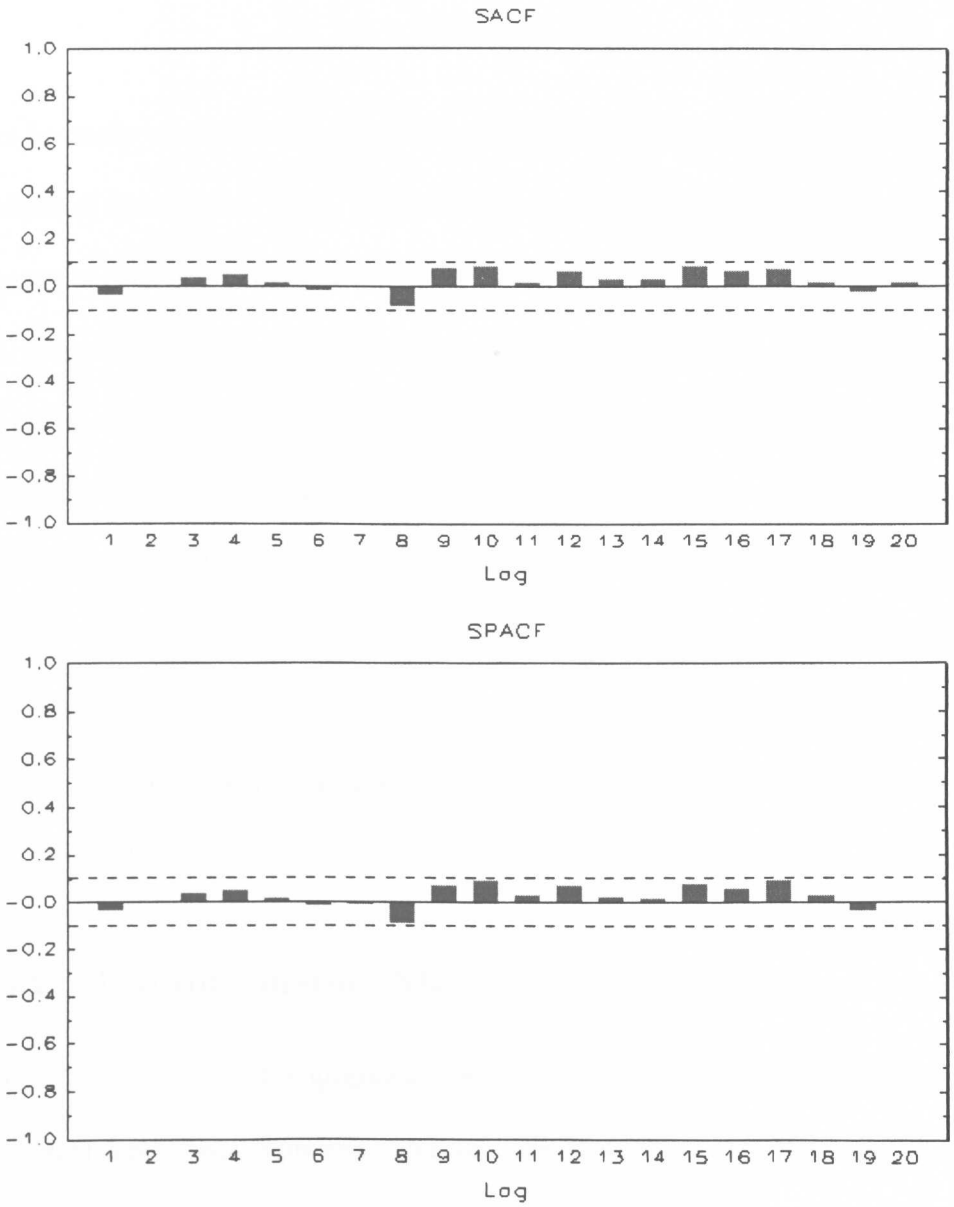
$$\widehat{\varepsilon}_t = X_t - \widehat{\alpha}_{1,ML} X_{t-1} - \widehat{\alpha}_{2,ML} X_{t-2} \tag{12.1}$$

where $\widehat{\alpha}_{1,ML}$ and $\widehat{\alpha}_{2,ML}$ are ML estimates of the thinning coefficients. In principle, the existence of any dependence structure in the residuals would suggest that a more general specification is called for. For this reason, we plot the SACFs and SPACFs of the residuals from the estimated INAR(2) model in Figure 12.4. In an informal manner, the figure indicates that there is no obvious dependence structure left in the residuals.

As we discussed earlier, the new representation of the score functions leads to new definitions of the residuals of the model. According to Section 9.3.1, there are residuals for each random component of the model. In the current case, there are residuals for three random components, i.e. for the two thinning components

$$r_{1t}^{*} = E_t[\alpha_1 \circ X_{t-1}] - E_{t-1}[\alpha_1 \circ X_{t-1}]$$

Figure 12.4: Correlograms of the Residuals $\widehat{\varepsilon}_t$ from the INAR(2) Model



$$r_{2t}^{*} = E_t[\alpha_2 \circ X_{t-2}] - E_{t-1}[\alpha_2 \circ X_{t-2}]$$

and for the arrival component

$$r_{\varepsilon t}^{*} = E_t[\varepsilon_t] - E_{t-1}[\varepsilon_t].$$

It can be easily verified that adding the new sets of residuals together gives the usual definition of residuals in (12.1). These component residuals may be used to assess the adequacy of each component of the fitted model and may suggest improved specifications. However, similar to the results from the residuals in (12.1), the correlograms for none of the three component residual series suggest significant remaining dependence structure. These observations are confirmed by the results from both the Venkataraman (1982) test and Mills and Seneta (1989) tests with various degrees of freedom applied to these residuals. Table 12.2 shows all the statistics for the tests we used, together with their corresponding p-values given in the parenthesis. Therefore, in terms of results from the residual analysis, the INAR(2) model with Poisson innovations seems to be adequate.

12.3.2 The Information Matrix Test

Nevertheless, neither the graphical method based on correlograms nor the residual serial dependence tests represent rigorous investigations of the fitted model. Serial dependence tests are only intended to provide an indication of model misspecification to motivate the application of a higher order specification. Graphical methods often

Table 12.2: Results of Tests for Serial Dependence in INAR(2) Residuals

Tests	Residual Series			
	$\hat{\varepsilon}_t$	r_{1t}^*	r_{2t}^*	$r_{\varepsilon t}^*$
$Q_{acf}(1)$	0.0007	1.1878	0.2140	0.1838
	(0.9787)	(0.2758)	(0.6437)	(0.6681)
$Q_{acf}(5)$	1.2045	4.1936	1.0649	6.6946
	(0.9444)	(0.5219)	(0.9572)	(0.2444)
$Q_{acf}(10)$	6.7416	10.2911	6.4794	14.4748
	(0.7496)	(0.4153)	(0.7735)	(0.1524)
$Q_{pacf}(1)$	0.0002	1.1086	0.2027	0.2265
	(0.9900)	(0.2924)	(0.6525)	(0.6342)
$Q_{pacf}(5)$	1.3097	4.0125	1.1488	8.0657
	(0.9339)	(0.5476)	(0.9497)	(0.1526)
$Q_{pacf}(10)$	7.3144	8.9945	6.1304	14.7222
	(0.6955)	(0.5326)	(0.8042)	(0.1425)

suggest how the model may be constructively modified, e.g. the lack of a seasonal component is usually indicated by a cyclical pattern in a residual plot. However, by comparison with graphical procedures for continuous data, those for counts are more difficult to interpret because of their discrete nature. In particular, when individual component residuals are concerned, not only are they correlated with one another within any given set but different sets of residuals from the same model are themselves correlated. In the current case, for instance, the calculated sample correlation is 0.5832 between r_{1t}^* and $r_{\varepsilon t}^*$; 0.6498 between r_{2t}^* and $r_{\varepsilon t}^*$; and 0.7173 between r_{1t}^* and r_{2t}^*. For these reasons graphical analysis often needs to be supplemented by formal statistical tests to confirm whether the specification of the components of the model is adequate.

Freeland and McCabe (2004a) suggested that for the INAR model, where ML

methods are used, the specification test based on information matrix (IM) equality can be used. They showed that the IM test could be interpreted as a test that the parameters of the model were constants against the alternative that they were random variables. They applied the IM specification test to the PoINAR(1) model and showed that the test can be decomposed into sub-tests: one for each component of the model. Thus, there is a formal test which checks the adequacy of the arrival process and one which assesses the Binomial thinning process. There is also an additional component of the overall test which checks whether these component processes are independent. The details of this specification test are elaborated in their paper.

The IM test of Freeland and McCabe (2004a) can be easily extended to higher-order INAR models. In the current situation, the test statistic for the overall adequacy of the model can be written as

$$\begin{aligned} U_n &= \sum_{t=1}^{n}\{\dot{\ell}_{\alpha_{1t}}^2 + \ddot{\ell}_{\alpha_{1t}}\} + \sum_{t=1}^{n}\{\dot{\ell}_{\alpha_{2t}}^2 + \ddot{\ell}_{\alpha_{2t}}\} + \sum_{t=1}^{n}\{\dot{\ell}_{\lambda_t}^2 + \ddot{\ell}_{\lambda_t})\} \\ &\quad -2\sum_{t=1}^{n}\{\dot{\ell}_{\alpha_{1t}}\dot{\ell}_{\lambda_t} + \ddot{\ell}_{\alpha_{1t}\lambda_t}\} - 2\sum_{t=1}^{n}\{\dot{\ell}_{\alpha_{2t}}\dot{\ell}_{\lambda_t} + \ddot{\ell}_{\alpha_{2t}\lambda_t}\} \\ &\quad -2\sum_{t=1}^{n}\{\dot{\ell}_{\alpha_{1t}}\dot{\ell}_{\alpha_{2t}} + \ddot{\ell}_{\alpha_{1t}\alpha_{2t}}\} \end{aligned} \tag{12.2}$$

where

$$\begin{aligned} \dot{\ell}_{\alpha_{kt}} &= \frac{\partial \ln P_t\left(X_t|X_{t-1}, X_{t-2}\right)}{\partial \alpha_k} \\ \dot{\ell}_{\lambda_t} &= \frac{\partial \ln P_t\left(X_t|X_{t-1}, X_{t-2}\right)}{\partial \lambda} \end{aligned}$$

and

$$\begin{aligned}
\ddot{\ell}_{\alpha_{kt}} &= \frac{\partial^2 \ln P_t\left(X_t|X_{t-1}, X_{t-2}\right)}{\partial \alpha_k^2} \\
\ddot{\ell}_{\lambda_t} &= \frac{\partial^2 \ln P_t\left(X_t|X_{t-1}, X_{t-2}\right)}{\partial \lambda^2} \\
\ddot{\ell}_{\alpha_{kt}\lambda_t} &= \frac{\partial^2 \ln P_t\left(X_t|X_{t-1}, X_{t-2}\right)}{\partial \alpha_k \partial \lambda}
\end{aligned}$$

for $k = 1$ and 2. Freeland and McCabe (2004a) showed that under the null hypothesis that the model is correctly specified, U_n is a zero mean martingale and

$$[U]_n^{-\frac{1}{2}} U_n \xrightarrow{d} N(0,1)$$

where $[U]_n$ is the quadratic variation of the martingale. Or equivalently,

$$[U]_n^{-1} U_n^2 \xrightarrow{d} \chi^2(1).$$

Although the results from the residual analysis suggest that the INAR(2) model is adequate, since the residual analysis based on graphical method and dependence tests has its limitations, we take one step further by applying the IM test to the fitted INAR(2) model. The p-value of the overall test is computed to be 0.2659, which indicates that we cannot reject the hypothesis that the INAR(2) process with Poisson arrivals sufficiently describes the variation in the data. This is consistent with the previous conclusion based on residual analysis. It is worth mentioning that although the overall test suggests that the model as a whole may be regarded as satisfactory, it does not necessarily mean that each component is satisfactory. This is because the effect of misspecification in each component may cancel each other so

that when considered together they appear to be satisfactory. It is also possible to have the opposite situation. Since the IM test can be decomposed into sub-tests. We are able to assess the adequacy of each component and the independence assumptions among them.

For the INAR(2) model at hand, it is evident from (12.2) that there can be up to 6 sub-tests. There are three components corresponding to the first and second binomial thinning operations and the Poisson arrivals. The p-values for the three sub-tests on the three components are 0.0343, 0.8259, and 0.5108, respectively. These results suggest that while we may conclude that the second Binomial operation and the Poisson arrivals are correctly specified, the first binomial thinning seems to be problematic due to the low p-value of the test. It should be noted that Binomial thinning assumes that from one period to the next the departure of individual particles from the observation area is independent and all individual particles have the same probability of staying in the area. Although to get more specific and sensible interpretations, one might have to trace back the source of the data, the result of the IM test provides statistical evidence that either of the two assumptions for the Binomial distribution or both of them are not supported by the data. One possibility to correct for such misspecification is to consider over-dispersed models such as in McKenzie (1986), Al-Osh and Aly (1992) and Joe (1996). Since this is beyond the scope of this study, we do not pursue further discussion.

The remaining three sub-tests investigate the conditional independence assump-

tions among the three components. The corresponding p-values are found to be 0.2756 for the test for independence between the two Binomial thinning operations; 0.1240 for the test between the first Binomial and the Poisson arrivals; and 0.1571 for the test between the second Binomial and Poisson arrivals. These results suggest that the assumption of conditional independence is satisfactory.

12.4 Forecasting the Westgren Data

This section aims to apply the method developed in Chapter 11 to produce forecasts for the Westgren data. Despite the evidence against the first Binomial thinning component, the result from the IM test on overall adequacy does not allow us to reject the hypothesis that the INAR(2) model with Poisson innovations is correctly specified. Therefore, we proceed to evaluate forecasts based on the fitted INAR(2) model. ML estimates for the model parameters are already given in Section 12.2. For the INAR(2) model, the conditional probability depends on two lags and can be denoted as $P(X_t|X_{t-1}, X_{t-2})$. It is observed that the last two observations of the series are $X_T = 3$ and $X_{T-1} = 3$.

12.4.1 One-step and h-step Forecasts

Table 12.3 gives the multiple-period ahead conditional mean, median and mode forecasts, as well as conditional probability forecasts for each value of X_t at subsequent

Table 12.3: Mean, Median, Mode and Probability Forecasts for the Westgren Data

	h									
	1	2	3	4	5	10	20	30	40	∞
Mean	2.4993	2.2632	2.0618	1.9244	1.8233	1.6143	1.5656	1.5637	1.5636	1.5636
Median	2	2	2	2	2	1	1	1	1	1
Mode	2	2	2	1	1	1	1	1	1	1
$P_h(0\|3,3)$	0.0472	0.0892	0.1314	0.1607	0.1819	0.2237	0.2329	0.2332	0.2332	0.2332
$P_h(1\|3,3)$	0.1831	0.2315	0.2617	0.2780	0.2891	0.3111	0.3162	0.3164	0.3164	0.3164
$P_h(2\|3,3)$	0.2955	0.2819	0.2664	0.2566	0.2499	0.2382	0.2360	0.2359	0.2359	0.2359
$P_h(3\|3,3)$	0.2616	0.2150	0.1836	0.1662	0.1545	0.1325	0.1277	0.1276	0.1276	0.1276
$P_k(4\|3,3)$	0.1431	0.1157	0.0961	0.0843	0.0761	0.0598	0.0560	0.0559	0.0559	0.0559
$P_h(5\|3,3)$	0.0525	0.0469	0.0405	0.0355	0.0317	0.0232	0.0211	0.0210	0.0210	0.0210
$P_h(6\|3,3)$	0.0138	0.0150	0.0144	0.0129	0.0115	0.0080	0.0071	0.0070	0.0070	0.0070
$P_h(7\|3,3)$	0.0027	0.0039	0.0044	0.0041	0.0037	0.0025	0.0022	0.0022	0.0022	0.0022
$P_h(8\|3,3)$	0.0004	0.0008	0.0012	0.0012	0.0011	0.0007	0.0006	0.0006	0.0006	0.0006
$P_h(9\|3,3)$	0.0001	0.0002	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
$P_h(10\|3,3)$	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000

periods. As expected the conditional mean forecasts are no longer integer values. It can be seen that these conditional mean forecasts converge to the mean of the marginal distribution. This is equal to the unconditional mean of the process implied by the parameter estimates. Similarly, the conditional median and mode forecasts converge to their marginal counterparts. It can be noted that the pace of convergence is relatively slow, happening after more than 30 periods. It is not surprising since as mentioned previously the data at hand exhibit relatively long distance dependence, with $\widehat{\alpha}_1 + \widehat{\alpha}_2 = 0.6514$. This observation is also consistent with what we have seen in Figure 12.3 where the SACFs remain significant even after 20 lags.

To account for parameter uncertainty in model estimation. We apply the theorem proposed in Section 11.2.2 and compute the confidence intervals for the probabilities associated with each value of the forecast distribution. These interval forecasts are presented in Table 12.4. It can be seen that, for instance, in the next period we are 95% confident that the probability of the value 0 occurring lies between 0.0295 and 0.0649; the point estimate of the probability is 0.0472. The point estimates for the values from 0 to 10 sum almost to unity and the model deems that an observation of 11 or more has zero probability of occurring in the forecast periods. In contrast, forecasts based a single summary statistic of the forecast distribution is much less informative than what we can infer from Table 12.4. The conditional mean (if rounded to the nearest integer) all indicate that 2 particles are to be expected in the following periods; The conditional median suggest 2 particles to be expected in the following 5 periods

Table 12.4: 95% Confidence Intervals for the Probability Forecasts for the Westgren Data

	h				
	1	2	3	4	5
$P_h(0\|3,3)$	(0.0295, 0.0649)	(0.0574, 0.1209)	(0.0890, 0.1738)	(0.1115, 0.2099)	(0.1283, 0.2355)
$P_h(1\|3,3)$	(0.1477, 0.2186)	(0.1881, 0.2748)	(0.2178, 0.3056)	(0.2343, 0.3218)	(0.2464, 0.3318)
$P_h(2\|3,3)$	(0.2751, 0.3160)	(0.2728, 0.2910)	(0.2629, 0.2699)	(0.2523, 0.2608)	(0.2418, 0.2580)
$P_h(3\|3,3)$	(0.2380, 0.2852)	(0.1891, 0.2408)	(0.1565, 0.2108)	(0.1373, 0.1951)	(0.1245, 0.1845)
$P_k(4\|3,3)$	(0.1164, 0.1699)	(0.0861, 0.1452)	(0.0667, 0.1255)	(0.0549, 0.1137)	(0.0472, 0.1051)
$P_h(5\|3,3)$	(0.0370, 0.0680)	(0.0291, 0.0648)	(0.0217, 0.0594)	(0.0168, 0.0543)	(0.0134, 0.0499)
$P_h(6\|3,3)$	(0.0082, 0.0193)	(0.0076, 0.0224)	(0.0054, 0.0233)	(0.0037, 0.0221)	(0.0025, 0.0205)
$P_h(7\|3,3)$	(0.0013, 0.0041)	(0.0016, 0.0062)	(0.0010, 0.0078)	(0.0004, 0.0078)	(0.0000, 0.0075)
$P_h(8\|3,3)$	(0.0002, 0.0007)	(0.0002, 0.0014)	(0.0001, 0.0023)	(0.0000, 0.0025)	(0.0000, 0.0025)
$P_h(9\|3,3)$	(0.0000, 0.0001)	(0.0000, 0.0003)	(0.0000, 0.0006)	(0.0000, 0.0007)	(0.0000, 0.0007)
$P_h(10\|3,3)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0001)	(0.0000, 0.0002)	(0.0000, 0.0002)
	10	20	30	40	∞
$P_h(0\|3,3)$	(0.1614, 0.2860)	(0.1685, 0.2972)	(0.1688 0.2976)	(0.1688 0.2976)	(0.1688 0.2976)
$P_h(1\|3,3)$	(0.2743, 0.3479)	(0.2828, 0.3496)	(0.2833 0.3496)	(0.2833 0.3496)	(0.2833 0.3496)
$P_h(2\|3,3)$	(0.2203, 0.2561)	(0.2154, 0.2566)	(0.2151 0.2566)	(0.2151 0.2566)	(0.2151 0.2566)
$P_h(3\|3,3)$	(0.1005, 0.1645)	(0.0956, 0.1599)	(0.0955 0.1597)	(0.0955 0.1597)	(0.0955 0.1597)
$P_k(4\|3,3)$	(0.0337, 0.0859)	(0.0316, 0.0804)	(0.0316 0.0802)	(0.0316 0.0802)	(0.0316 0.0802)
$P_h(5\|3,3)$	(0.0081, 0.0383)	(0.0076, 0.0346)	(0.0077 0.0344)	(0.0077 0.0344)	(0.0077 0.0344)
$P_h(6\|3,3)$	(0.0009, 0.0151)	(0.0010, 0.0132)	(0.0010 0.0131)	(0.0010 0.0131)	(0.0010 0.0131)
$P_h(7\|3,3)$	(0.0000, 0.0054)	(0.0000, 0.0045)	(0.0000 0.0045)	(0.0000 0.0045)	(0.0000 0.0045)
$P_h(8\|3,3)$	(0.0000, 0.0018)	(0.0000, 0.0014)	(0.0000 0.0014)	(0.0000 0.0014)	(0.0000 0.0014)
$P_h(9\|3,3)$	(0.0000, 0.0005)	(0.0000, 0.0004)	(0.0000 0.0004)	(0.0000 0.0004)	(0.0000 0.0004)
$P_h(10\|3,3)$	(0.0000, 0.0002)	(0.0000, 0.0001)	(0.0000 0.0001)	(0.0000 0.0001)	(0.0000 0.0001)

and 1 to be expected from the 6th period onwards; and the conditional mode predict 2 particles in the next 3 periods and 1 thereafter. Clearly, the amount of information that would have been lost as a result of using only single statistics is considerable.

12.4.2 Rolling Forecasts

The forecast distributions described above are of interest in their own right, but the forecasting scenario envisaged is somewhat different from what would likely be done in real life. This is because all the forecasts are ex post, since the actual data for $t = 371, \ldots, 380$ are available to the investigators. Moreover, in reality, as each new observation becomes available, a forecaster is likely to incorporate it into any prediction exercise. To mimic this situation, we used the idea of rolling forecasting suggested by West (1996). That is, we forecast observation 371 as before, and then forecast observation 372 based on a model fitted using a new sample consisting of observations 2 to 371, and forecast observation 373 using observations 3 to 372, and so on up to forecasting observation 380. A rolling sample with fixed size (370 observations) is used throughout in model fitting. This generates a sequence of one-step ahead rolling forecasts. The results for point and interval forecasts are reported in Table 12.5 and 12.6, respectively. West (1996) also mentions recursive forecasts based on an increasing sample size as new observations become available. We also tried this and the results are given in Table 12.7 and 12.8. But as expected these results are very close to those in Table 12.5 and 12.6. One difference to be expected from what

Table 12.5: Rolling One-Step Ahead Forecasts (Fixed Sample Size)

	Observation Number									
	371	372	373	374	375	376	377	378	379	380
Observed Value	3	2	3	2	1	2	2	1	2	1
Mean	2.4993	2.5046	2.0294	2.3023	2.0181	1.3672	1.6432	1.8342	1.3738	1.6478
Median	2	2	2	2	2	1	2	2	1	2
Mode	2	2	2	2	2	1	1	2	1	1
$P_h(0\|3,3)$	0.0472	0.0470	0.0896	0.0596	0.0913	0.2092	0.1385	0.1118	0.2074	0.1375
$P_h(1\|3,3)$	0.1831	0.1822	0.2662	0.2139	0.2684	0.3867	0.3426	0.3037	0.3856	0.3416
$P_h(2\|3,3)$	0.2955	0.2947	0.3204	0.3142	0.3199	0.2694	0.3189	0.3238	0.2707	0.3196
$P_h(3\|3,3)$	0.2616	0.2620	0.2090	0.2477	0.2072	0.1034	0.1484	0.1810	0.1045	0.1493
$P_k(4\|3,3)$	0.1431	0.1441	0.0852	0.1179	0.0841	0.0259	0.0420	0.0622	0.0262	0.0423
$P_h(5\|3,3)$	0.0525	0.0530	0.0238	0.0369	0.0234	0.0047	0.0083	0.0146	0.0048	0.0083
$P_h(6\|3,3)$	0.0138	0.0138	0.0049	0.0082	0.0048	0.0007	0.0012	0.0026	0.0007	0.0012
$P_h(7\|3,3)$	0.0027	0.0027	0.0008	0.0014	0.0008	0.0001	0.0001	0.0003	0.0001	0.0001
$P_h(8\|3,3)$	0.0004	0.0004	0.0001	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
$P_h(9\|3,3)$	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$P_h(10\|3,3)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 12.6: 95% Confidence Intervals for Rolling Probability Forecasts (Fixed Sample Size)

	Observation Number				
	371	372	373	374	375
$P_h(0\|3,3)$	(0.0295, 0.0649)	(0.0294, 0.0645)	(0.0658, 0.1134)	(0.0365, 0.0827)	(0.0670, 0.1157)
$P_h(1\|3,3)$	(0.1477, 0.2186)	(0.1470, 0.2175)	(0.2286, 0.3038)	(0.1772, 0.2506)	(0.2306, 0.3062)
$P_h(2\|3,3)$	(0.2751, 0.3160)	(0.2745, 0.3149)	(0.3030, 0.3378)	(0.3014, 0.3270)	(0.3026, 0.3373)
$P_h(3\|3,3)$	(0.2380, 0.2852)	(0.2387, 0.2854)	(0.1805, 0.2376)	(0.2182, 0.2773)	(0.1783, 0.2360)
$P_k(4\|3,3)$	(0.1164, 0.1699)	(0.1173, 0.1708)	(0.0615, 0.1090)	(0.0958, 0.1400)	(0.0604, 0.1078)
$P_h(5\|3,3)$	(0.0370, 0.0680)	(0.0375, 0.0684)	(0.0139, 0.0337)	(0.0274, 0.0465)	(0.0136, 0.0332)
$P_h(6\|3,3)$	(0.0082, 0.0193)	(0.0083, 0.0193)	(0.0022, 0.0075)	(0.0053, 0.0112)	(0.0022, 0.0074)
$P_h(7\|3,3)$	(0.0013, 0.0041)	(0.0013, 0.0041)	(0.0002, 0.0013)	(0.0007, 0.0021)	(0.0002, 0.0013)
$P_h(8\|3,3)$	(0.0002, 0.0007)	(0.0002, 0.0007)	(0.0000, 0.0002)	(0.0001, 0.0003)	(0.0000, 0.0002)
$P_h(9\|3,3)$	(0.0000, 0.0001)	(0.0000, 0.0001)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)
$P_h(10\|3,3)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)
	376	377	378	379	380
$P_h(0\|3,3)$	(0.1771, 0.2413)	(0.1008, 0.1762)	(0.0853, 0.1383)	(0.1751, 0.2398)	(0.0998, 0.1752)
$P_h(1\|3,3)$	(0.3550, 0.4185)	(0.3188, 0.3663)	(0.2776, 0.3297)	(0.3540, 0.4173)	(0.3177, 0.3655)
$P_h(2\|3,3)$	(0.2458, 0.2930)	(0.2896, 0.3483)	(0.3030, 0.3446)	(0.2469, 0.2946)	(0.2903, 0.3490)
$P_h(3\|3,3)$	(0.0795, 0.1272)	(0.1257, 0.1711)	(0.1583, 0.2037)	(0.0804, 0.1285)	(0.1265, 0.1721)
$P_k(4\|3,3)$	(0.0161, 0.0357)	(0.0306, 0.0534)	(0.0483, 0.0760)	(0.0164, 0.0361)	(0.0308, 0.0538)
$P_h(5\|3,3)$	(0.0022, 0.0072)	(0.0046, 0.0119)	(0.0096, 0.0197)	(0.0022, 0.0073)	(0.0046, 0.0120)
$P_h(6\|3,3)$	(0.0002, 0.0011)	(0.0004, 0.0020)	(0.0013, 0.0038)	(0.0002, 0.0011)	(0.0004, 0.0020)
$P_h(7\|3,3)$	(0.0000, 0.0001)	(0.0000, 0.0003)	(0.0001, 0.0006)	(0.0000, 0.0001)	(0.0000, 0.0003)
$P_h(8\|3,3)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0001)	(0.0000, 0.0000)	(0.0000, 0.0000)
$P_h(9\|3,3)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)
$P_h(10\|3,3)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)

Table 12.7: Rolling One-Step Ahead Forecasts (Increasing Sample Size)

	Observation Number									
	371	372	373	374	375	376	377	378	379	380
Observed Value	3	2	3	2	1	2	2	1	2	1
Mean	2.4993	2.5033	2.0265	2.3261	2.0315	1.3724	1.6704	1.8509	1.3717	1.6714
Median	2	2	2	2	2	1	2	2	1	2
Mode	2	2	2	2	2	1	1	2	1	1
$P_h(0\|3,3)$	0.0472	0.0468	0.0891	0.0568	0.0885	0.2065	0.1318	0.1077	0.2064	0.1313
$P_h(1\|3,3)$	0.1831	0.1824	0.2671	0.2093	0.2661	0.3879	0.3383	0.3011	0.3883	0.3382
$P_h(2\|3,3)$	0.2955	0.2954	0.3215	0.3140	0.3216	0.2702	0.3234	0.3266	0.2703	0.3239
$P_h(3\|3,3)$	0.2616	0.2622	0.2084	0.2511	0.2092	0.1037	0.1526	0.1834	0.1036	0.1528
$P_k(4\|3,3)$	0.1431	0.1436	0.0845	0.1205	0.0851	0.0261	0.0437	0.0632	0.0260	0.0437
$P_h(5\|3,3)$	0.0525	0.0527	0.0236	0.0380	0.0237	0.0048	0.0087	0.0149	0.0047	0.0087
$P_h(6\|3,3)$	0.0138	0.0138	0.0048	0.0085	0.0049	0.0007	0.0013	0.0026	0.0007	0.0013
$P_h(7\|3,3)$	0.0027	0.0027	0.0008	0.0014	0.0008	0.0001	0.0002	0.0004	0.0001	0.0002
$P_h(8\|3,3)$	0.0004	0.0004	0.0001	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
$P_h(9\|3,3)$	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$P_h(10\|3,3)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 12.8: 95% Confidence Intervals for Rolling Probability Forecasts (Increasing Sample Size)

	Observation Number				
	371	372	373	374	375
$P_h(0\|3,3)$	(0.0295, 0.0649)	(0.0292, 0.0643)	(0.0656, 0.1127)	(0.0347, 0.0789)	(0.0651, 0.1119)
$P_h(1\|3,3)$	(0.1477, 0.2186)	(0.1471, 0.2177)	(0.2296, 0.3046)	(0.1730, 0.2456)	(0.2287, 0.3035)
$P_h(2\|3,3)$	(0.2751, 0.3160)	(0.2748, 0.3159)	(0.3033, 0.3398)	(0.3008, 0.3273)	(0.3034, 0.3398)
$P_h(3\|3,3)$	(0.2380, 0.2852)	(0.2386, 0.2857)	(0.1800, 0.2369)	(0.2219, 0.2804)	(0.1809, 0.2375)
$P_k(4\|3,3)$	(0.1164, 0.1699)	(0.1169, 0.1703)	(0.0608, 0.1082)	(0.0985, 0.1425)	(0.0613, 0.1088)
$P_h(5\|3,3)$	(0.0370, 0.0680)	(0.0372, 0.0682)	(0.0137, 0.0334)	(0.0284, 0.0477)	(0.0138, 0.0336)
$P_h(6\|3,3)$	(0.0082, 0.0193)	(0.0083, 0.0193)	(0.0022, 0.0075)	(0.0055, 0.0116)	(0.0022, 0.0075)
$P_h(7\|3,3)$	(0.0013, 0.0041)	(0.0013, 0.0041)	(0.0002, 0.0013)	(0.0007, 0.0022)	(0.0002, 0.0013)
$P_h(8\|3,3)$	(0.0002, 0.0007)	(0.0002, 0.0007)	(0.0000, 0.0002)	(0.0001, 0.0003)	(0.0000, 0.0002)
$P_h(9\|3,3)$	(0.0000, 0.0001)	(0.0000, 0.0001)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)
$P_h(10\|3,3)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)
	376	377	378	379	380
$P_h(0\|3,3)$	(0.1756, 0.2374)	(0.0958, 0.1678)	(0.0822, 0.1331)	(0.1757, 0.2371)	(0.0954, 0.1672)
$P_h(1\|3,3)$	(0.3553, 0.4205)	(0.3140, 0.3625)	(0.2753, 0.3269)	(0.3556, 0.4210)	(0.3139, 0.3624)
$P_h(2\|3,3)$	(0.2473, 0.2932)	(0.2944, 0.3524)	(0.3054, 0.3478)	(0.2473, 0.2932)	(0.2949, 0.3529)
$P_h(3\|3,3)$	(0.0801, 0.1274)	(0.1302, 0.1750)	(0.1613, 0.2056)	(0.0800, 0.1272)	(0.1304, 0.1751)
$P_k(4\|3,3)$	(0.0163, 0.0359)	(0.0322, 0.0553)	(0.0493, 0.0771)	(0.0163, 0.0357)	(0.0322, 0.0552)
$P_h(5\|3,3)$	(0.0022, 0.0073)	(0.0049, 0.0125)	(0.0098, 0.0201)	(0.0022, 0.0073)	(0.0049, 0.0124)
$P_h(6\|3,3)$	(0.0002, 0.0011)	(0.0005, 0.0021)	(0.0013, 0.0039)	(0.0002, 0.0011)	(0.0005, 0.0021)
$P_h(7\|3,3)$	(0.0000, 0.0001)	(0.0000, 0.0003)	(0.0001, 0.0006)	(0.0000, 0.0001)	(0.0000, 0.0003)
$P_h(8\|3,3)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0001)	(0.0000, 0.0000)	(0.0000, 0.0000)
$P_h(9\|3,3)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)
$P_h(10\|3,3)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)

is reported in Section 12.4.1 is that no forecast distribution should tend towards the marginal distribution of the data used in the fitting.

12.5 Conclusion

In this chapter, we carried out an empirical analysis of the Westgren gold particle data under the framework of Maximum likelihood developed in Chapter 9. Estimates of parameters of the INAR(2) model were obtained by conditional maximum likelihood estimation. Issues of model adequacy were examined. Although various tests for serial dependence in the model residuals do not reject the INAR(2) model as adequate, new evidence has been found from the likelihood based IM test: While the overall test failed to reject the INAR(2) model, sub-tests on model components indicate that the first Binomial thinning operator is not supported by the data, suggesting less restricted models, i.e. overdispersed models, be used to correct for the partial misspecification.

Also in this chapter, we applied the method developed in Chapter 11 to produce distribution forecasts for the Westgren data. Our results showed that the estimated point mass forecasts are much more informative than those supplied by either the mean, median or mode of the forecast distributions. Therefore, the benefit of implementing the new method is substantial. Moreover, compared to simulation based approaches our method is computationally efficient.

Chapter 13

Conclusion

In this part of the thesis, we considered the estimation of higher-order INAR processes with general specifications for thinning operators and innovation distributions. A maximum likelihood framework for estimating the INAR(p) model has been developed. Specifically, we proposed a recursive form for the transition probability function of the INAR(p) model to facilitate the likelihood computation and the derivative calculations. Also based on this recursion, we derived the corresponding score functions and the Fisher information matrix for the INAR(p) model. In particular, we provided the conditions on the distributions of the thinning process and innovation sequence, under which the elements of both the score and Fisher information can be represented in terms of conditional expectations. We showed that these new representations not only enhance the interpretation of these quantities but also lead to new definitions for residuals of the model. For the Binomial-Poisson special case, specific

details on ML estimation have been explored, including the asymptotic distribution of the ML estimator.

We also studied the performance of the MLE in comparison to the YWE and the CLSE. Using the INAR(2) model with Poisson innovations, we investigated the asymptotic gain of implementing the ML method over the widely used CLS method by calculating the ARE between the two estimators. Our results confirm that the proposed MLE is asymptotically more efficient than the CLSE especially for high-persistence processes. The finite sample properties of the alternative estimators have been examined by means of Monte Carlo experiments. While the results from the CLSE and YWE are similar, we found that the MLE is preferable and worth the extra calculation due to potential gain in terms of bias and MSE. In particular, we found that both asymptotically and in small samples the magnitude of efficiency gain is positively related to the degree of persistence of the underlying processes.

A new approach for producing conditional probability forecasts for time series of count data based on the INAR(p) model has been proposed. We suggested that the INAR(p) process be regarded as a special type Markov system and the new method is based the transition matrix for stationary INAR processes. A procedure based on the δ-method for calculating confidence intervals has also been suggested. Compared with the simulation based method proposed by Jung and Tremayne (2006), the method developed here is computationally efficient.

An application to the Westgren Gold particle data has also been presented. We

showed that under the ML framework new tools can be used for testing model adequacy. As a result, new evidence has come out, regarding the validity of assuming Binomial distributions for the thinning processes for the Westgren data. We also applied our forecasting tools for generating distribution forecasts for the fitted model. Given the enriched information and the advanced efficiency, the benefit from the new approach is found to be substantial.

References

[1] Al-Osh, M.A. and Aly, E.A.A. (1992). First Order Autoregressive Time Series with Negative Binomial and Geometric Marginals. *Communications in Statistics* A 21, 2483-2492.

[2] Al-Osh, M.A. and Alzaid, A.A. (1987). First-Order Integer Valued Autoregressive (INAR(1)) Process. *Journal of Time Series Analysis* 8, 261-275

[3] Alzaid, A.A. and Al-Osh, M.A. (1990). An Integer-Valued *p*th-Order Autoregressive Structure (INAR(p)) Process. *Journal of Applied Probability* 27, 314-323.

[4] Brémaud, P. (1999). Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues. Springer. New York.

[5] Bu, R., Hadri, K. and McCabe, B.P.M. (2006a). Maximum Likelihood Estimation of Higher-Order Integer-Valued Autoregressive Processes. Working Paper, University of Liverpool.

[6] Bu, R., Hadri, K. and McCabe, B.P.M. (2006b). Estimating INAR(p) Processes: Finite Sample Performance. Working Paper, University of Liverpool.

[7] Bu, R., Hadri, K. and McCabe, B.P.M. (2006c). Coherent Forecasting with INAR(p) Models: A Markov Chain Approach. Working Paper, University of Liverpool.

[8] Cox, D.R. and Hinkley, D. (1974). Theoretical Statistics. Chapman and Hall, London

[9] Dion, J.-P., Gauthier, G. and Latour, A. (1995). Branching Processes with Immigration and Integer-Valued Time Series. *Serdica* 21, 123-136.

[10] Du, J.-G and Li, Y. (1991). The Integer-Valued Autoregressive (INAR(p)) Model. *Journal of Time Series Analysis* 12, 129-142.

[11] Freeland, R.K. (1998). Statistical Analysis of Discrete Time Series with Applications to the Analysis of Workers Compensation Claims Data. Ph.D Thesis. The University of British Columbia.

[12] Freeland, R.K. and McCabe, B.P.M. (2004a). Analysis of Low Count Time Series by Poisson Autoregression. *Journal of Time Series Analysis* 25, 701-722

[13] Freeland, R.K. and McCabe, B.P.M. (2004b). Forecasting Discrete Valued Low Count Time Series. *International Journal of Forecasting* 20, 427–434.

[14] Grunwald, G.K., Hyndman, R.J., Tedesco, L. and Tweedie, R.L. (2000). Non-Gaussian Conditional Linear AR(1) Models. *Australian and New Zealand Journal of Statistics* 42, 479-495.

[15] Guttorp, P. (1991). Statistical Inference for Branching Processes. Wiley, New York.

[16] Joe, H. (1996). Time Series Models with Univariate Margins in the Convolution-Closed Infinitely Divisible Class. *Journal of Applied Probability* 33, 664-677.

[17] Jung, R.C. and Tremayne, A.R. (2003). Testing for Serial Dependence in Time Series of Counts. *Journal of Time Series Analysis* 24, 65-84.

[18] Jung, R.C. and Tremayne, A.R. (2006). Coherent Forecasting in Integer Time Series Models. Working Paper, Universität Tübingen.

[19] Karlin, S. and Taylor, H.M. (1981). A Second Course in Stochastic Processes. Academic Press, New York.

[20] Kemeny, J.G. and Snell, J.L. (1976). Finite Markov Chains. Springer, New York.

[21] Kemeny, J.G., Snell, J.L. and Knapp, A.W. (1976). Denumerable Markov Chains. Springer. New York.

[22] Klimko, L.A. and Nelson, P.I. (1978). On Conditional Least Squares Estimation for Stochastic Processes. *Annals of Statistics* 6, 629-642.

[23] McCabe, B.P.M. and Martin, G.M. (2005). Bayesian Predictions of Low Count Time Series. *International Journal of Forecasting* 21, 315-330.

[24] McKenzie, E. (1986). Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions. *Advances in Applied Probability* 18, 679-705.

[25] McKenzie, E. (1988). Some ARMA Models for Dependent Sequences of Poison Counts. *Advances in Applied Probability* 20, 822-835.

[26] Mills, T.M. and Seneta, E. (1989). Goodness-of-Fit for a Branching Process with Immigration. *Stochastic Processes and Their Applications* 33, 151-161.

[27] Sprott, D.A. (1983). Estimating the Parameters of a Convolution by Maximum Likelihood. *Journal of American Statistical Association* 78, 457-460.

[28] Venkataraman, K. (1982). A Time Series Approach to the Study of the Simple Subcritical Galton-Watson Process with Immigration. *Advances in Applied Probability* 14, 1-20.

[29] Wald, A. and Wolfowitz, J. (1940). On a Test Whether Two Samples are from the Same Population. *Annals of Mathematical Statistics* 11, 147-162.

[30] West, K.D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica* 64, 1067-1084.

[31] Westgren. A. (1916). Die Veränderungsgeschwindigkeit der lokalen Teilchenkonzentration in kollioden Systemen (Erste Mitteilung). *Arkiv för Matematik, Astronomi och Fysik* 11, 1-24.

Appendices

We present now the appendices that correspond to the previous chapters of the thesis.

Appendix 1: Proofs of Theorems and Corollaries in Chapter 9

In this appendix we set out the proofs of the theorems and corollaries given in Chapter 9.

Proof. of Theorem 9.1. We regard X_t as the convolution of $\alpha_1 \cdot X_{t-1}$ and $Y = \alpha_2 \cdot X_{t-2} + \cdots + \alpha_p \cdot X_{t-p} + \varepsilon_t$, which are by definition mutually independent given the p observed lags. Thus, the transition probability density function of the INAR(p) process, namely the probability density function of X_t conditioned on $(X_{t-1}, \ldots, X_{t-p})$, can be written as

$$\begin{aligned} & h(x_t|x_{t-1}, \ldots, x_{t-p}; \alpha_1, \ldots, \alpha_p, \lambda) \\ = & \int f(s_1|x_{t-1}; \alpha_1) h_Y(x_t - s_1|x_{t-2}, \ldots, x_{t-p}; \alpha_2, \ldots, \alpha_p, \lambda) d\upsilon(s_1) \end{aligned}$$

where $h_Y(y|x_{t-2}, \ldots, x_{t-p}; \alpha_2, \ldots, \alpha_p, \lambda)$ is the conditional probability density function of Y given observations $(X_{t-2}, \ldots, X_{t-p})$ and parameters $(\alpha_2, \ldots, \alpha_p, \lambda)$. It is important to note that this quantity $h_Y(y|x_{t-2}, \ldots, x_{t-p}; \alpha_2, \ldots, \alpha_p, \lambda)$ can be evaluated using the expression of transition probability density function for an INAR(p-1) process with parameters $(\alpha_2, \ldots, \alpha_p, \lambda)$. This is purely a computational device. We thus have the following recursive representation.

$$\begin{aligned} & h^{(p)}(x_t|x_{t-1}, \ldots, x_{t-p}; \alpha_1, \ldots, \alpha_p, \lambda) \\ = & \int f(s_1|x_{t-1}; \alpha_1) h^{(p-1)}(x_t - s_1|x_{t-2}, \ldots, x_{t-p}; \alpha_2, \ldots, \alpha_p, \lambda) d\upsilon(s_1) \end{aligned}$$

The superscript denotes that the conditional probability density function has the same expression as the transition probability density function of an INAR process with corresponding order. The recursion is initialized by

$$h^{(1)}\left(x_t - \sum_{k=1}^{p-1} s_k \middle| x_{t-p}\right) = \int f(s_p|x_{t-p})\, g\left(x_t - \sum_{k=1}^{p} s_k\right) d\upsilon(s_p)$$

which is just the convolution of the INAR(1) model with arguments as specified. Since υ is a counting measure, the recursion for the transition probability function is given by

$$\begin{aligned} & P(X_t|X_{t-1},\ldots,X_{t-p}) \\ = & \sum_{s_1} f(s_1|X_{t-1};\alpha_1) P(X_t - s_1|X_{t-2},\ldots,X_{t-p}) \end{aligned}$$

with starting value

$$P\left(X_t - \sum_{k=1}^{p-1} s_k \middle| X_{t-p}\right) = \sum_{s_p} f(s_p|X_{t-p})\, g\left(X_t - \sum_{k=1}^{p} s_k\right).$$

■

The proofs of Theorem 9.2 and 9.3 are based on the following straightforward lemma.

Lemma 1 *Let $X_1,\ldots,X_p,Y$ be independent random variables and denote their densities as $f_{X_i}(x_i)$ for $i = 1,2,\ldots,p$ and $f_Y(y)$. The densities are with respect to the measure υ which may be Lebesgue measure or counting measure. Let $Z = X_1 + X_2 + \cdots + X_p + Y$ be the convolution of $X_1,\ldots,X_p$, and Y. The joint distribution of Z and $X_1,\ldots,X_p$ is given by $\prod_{i=1}^{p} f_{X_i}(x_i) f_Y(z - \sum_{i=1}^{p} x_i)$ and the density*

for Z can be found by integrating out $X_1, \ldots, X_p$,

$$f_Z(z) = \int_p \cdots \int \prod_{i=1}^{p} f_{X_i}(x_i) f_Y(z - \sum_{i=1}^{p} x_i) d\upsilon(x_1) \cdots d\upsilon(x_p)$$

The conditional joint probability density for $X_1, \ldots, X_p$ given Z is

$$f_{(X_1,\ldots,X_p)|Z}(x_1, \ldots, x_p|z) = \frac{\prod_{i=1}^{p} f_{X_i}(x_i) f_Y(z - \sum_{i=1}^{p} x_i)}{f_Z(z)}$$

Let $\phi(X_1, \ldots, X_p, Y)$ denote an arbitrary function of $X_1, \ldots, X_p$, and Y. The conditional moments for $\phi(X_1, \ldots, X_p, Y)$ given Z are then

$$\begin{aligned}
& E[\phi(X_1, \ldots, X_p, Y)|Z] \\
= \; & E\left[\phi\left(X_1, \ldots, X_p, Z - \sum_{i=1}^{p} X_i\right) \middle| Z\right] \\
= \; & \int_p \cdots \int \phi\left(x_1, \ldots, x_p, z - \sum_{i=1}^{p} x_i\right) f_{(X_1,\ldots,X_p)|Z}(x_1, \ldots, x_p|z)\, d\upsilon(x_1) \cdots d\upsilon(x_p) \\
= \; & \frac{\int_p \cdots \int \phi\left(x_1, \ldots, x_p, z - \sum_{i=1}^{p} x_i\right) \prod_{i=1}^{p} f_{X_i}(x_i) f_Y(z - \sum_{i=1}^{p} x_i) d\upsilon(x_1) \cdots d\upsilon(x_p)}{f_Z(z)}
\end{aligned}$$

For notation convenience, we also use the following additional notation. The transition probability density function is denoted by

$$h(x_t|x_{t-1}, \ldots, x_{t-p}; \alpha_1, \ldots, \alpha_p, \lambda) = h(x_t|\mathbf{x}_{-p}; \boldsymbol{\theta})$$

where $\mathbf{x}_{-p} = (x_{t-1}, \ldots, x_{t-p})$, $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_p, \boldsymbol{\lambda})'$ and $\boldsymbol{\lambda}$ may be a vector. To simplify the notation for a p-fold integration, we set $\mathbf{s} = (s_1, \ldots, s_p)$ and $d\boldsymbol{\upsilon}_s = (d\upsilon(s_1) \cdots d\upsilon(s_p))'$.

Proof. of Theorem 9.2. The conditional log-likelihood function can be written as

$$\ln L(\boldsymbol{\theta}) = \sum_{t=p+1}^{T} \ln h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})$$

where

$$h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta}) = \int \left[\prod_{k=1}^{p} f(s_k|x_{t-k};\alpha_k) g(x_t - \sum_{k=1}^{p} s_k;\boldsymbol{\lambda})\right] d\upsilon_s$$

We denote

$$f(\mathbf{s}, x_t;\boldsymbol{\theta}) = \prod_{k=1}^{p} f(s_k|x_{t-k};\alpha_k) g(x_t - \sum_{k=1}^{p} s_k;\boldsymbol{\lambda}).$$

Hence

$$h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta}) = \int f(\mathbf{s}, x_t;\boldsymbol{\theta})\, d\upsilon_s$$

It follows that the corresponding the score functions are given by

$$\dot{\ell}_{\alpha_k} = \frac{\partial \ln L(\boldsymbol{\theta})}{\partial \alpha_k} = \sum_{t=p+1}^{T} \frac{\frac{\partial}{\partial \alpha_k} h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}$$

$$\dot{\ell}_{\boldsymbol{\lambda}} = \frac{\partial \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} = \sum_{t=p+1}^{T} \frac{\frac{\partial}{\partial \boldsymbol{\lambda}} h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}$$

Under the conditions of Theorem 9.2

$$\begin{aligned} \frac{\partial f(s_k|x_{t-k};\alpha_k)}{\partial \alpha_k} &= \tau(s_k;\alpha_k) f(s_k|x_{t-k};\alpha_k) \\ \frac{\partial g(\varepsilon;\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} &= \boldsymbol{\gamma}(\varepsilon;\boldsymbol{\lambda}) g(\varepsilon;\boldsymbol{\lambda}) \end{aligned}$$

and so

$$\begin{aligned}
\frac{\frac{\partial}{\partial \alpha_k} h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} &= \frac{\frac{\partial}{\partial \alpha_k}\left[\int f(\mathbf{s},x_t;\boldsymbol{\theta})\, d\boldsymbol{\upsilon}_s\right]}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \frac{\partial}{\partial \alpha_k}\left[f(\mathbf{s},x_t;\boldsymbol{\theta})\right] d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \tau(s_k;\alpha_k)\, f(\mathbf{s},x_t;\boldsymbol{\theta})\, d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= E_t\left[\tau(\alpha_k \cdot X_{t-k})\right].
\end{aligned}$$

Note that the last equality follows from the Lemma 1 on conditional expectations.

Similarly,

$$\begin{aligned}
\frac{\frac{\partial}{\partial \boldsymbol{\lambda}} h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} &= \frac{\frac{\partial}{\partial \boldsymbol{\lambda}}\left[\int f(\mathbf{s},x_t;\boldsymbol{\theta})\, d\boldsymbol{\upsilon}_s\right]}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \frac{\partial}{\partial \boldsymbol{\lambda}}\left[f(\mathbf{s},x_t;\boldsymbol{\theta})\right] d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \boldsymbol{\gamma}\left(x_t - \sum_{k=1}^{p} s_k;\boldsymbol{\lambda}\right) f(\mathbf{s},x_t;\boldsymbol{\theta})\, d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= E_t\left[\boldsymbol{\gamma}(\varepsilon_t)\right]
\end{aligned}$$

where, again, the last line follows from the Lemma 1. Finally, we have

$$\begin{aligned}
\dot{\ell}_{\alpha_k} &= \sum_{t=p+1}^{T} E_t\left[\tau(\alpha_k \cdot X_{t-k})\right] \\
\dot{\ell}_{\boldsymbol{\lambda}} &= \sum_{t=p+1}^{T} E_t\left[\boldsymbol{\gamma}(\varepsilon_t)\right].
\end{aligned}$$

■

Proof. of Theorem 9.3. Define scalar function $\tau_{\alpha_k}(s_k;\alpha_k)$ and matrix function

$\gamma_{\lambda}(\varepsilon;\boldsymbol{\lambda})$ such that

$$\begin{aligned}\tau_{\alpha_k}(s_k;\alpha_k) &= \frac{\partial\tau(s_k;\alpha_k)}{\partial\alpha_k}\\ \gamma_{\lambda}(\varepsilon;\boldsymbol{\lambda}) &= \frac{\partial\gamma(\varepsilon;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}'}.\end{aligned}$$

Under the conditions of Theorem 9.2:

$$\begin{aligned}\frac{\partial^2 f(s_k|x_{t-k};\alpha_k)}{\partial\alpha_k^2} &= \frac{\partial\left[\tau(s_k;\alpha_k)\,f(s_k|x_{t-k};\alpha_k)\right]}{\partial\alpha_k}\\ &= \frac{\partial\tau(s_k;\alpha_k)}{\partial\alpha_k}f(s_k|x_{t-k};\alpha_k)+\left[\tau(s_k;\alpha_k)\right]^2 f(s_k|x_{t-k};\alpha_k)\\ &= \left\{\tau_{\alpha_k}(s_k;\alpha_k)+\left[\tau(s_k;\alpha_k)\right]^2\right\}f(s_k|x_{t-k};\alpha_k)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 g(\varepsilon;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'} &= \frac{\partial\left[\boldsymbol{\gamma}(\varepsilon;\boldsymbol{\lambda})\,g(\varepsilon;\boldsymbol{\lambda})\right]}{\partial\boldsymbol{\lambda}'}\\ &= \frac{\partial\boldsymbol{\gamma}(\varepsilon;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}'}g(\varepsilon;\boldsymbol{\lambda})+\boldsymbol{\gamma}(\varepsilon;\boldsymbol{\lambda})\frac{\partial g(\varepsilon;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}'}\\ &= \frac{\partial\boldsymbol{\gamma}(\varepsilon;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}'}g(\varepsilon;\boldsymbol{\lambda})+\left[\boldsymbol{\gamma}(\varepsilon;\boldsymbol{\lambda})\,\boldsymbol{\gamma}(\varepsilon;\boldsymbol{\lambda})'\right]g(\varepsilon;\boldsymbol{\lambda})\\ &= \left\{\boldsymbol{\gamma}_{\lambda}(\varepsilon;\boldsymbol{\lambda})+\left[\boldsymbol{\gamma}(\varepsilon;\boldsymbol{\lambda})\,\boldsymbol{\gamma}(\varepsilon;\boldsymbol{\lambda})'\right]\right\}g(\varepsilon;\boldsymbol{\lambda}).\end{aligned}$$

It then follows that

$$\begin{aligned}\frac{\frac{\partial^2}{\partial\alpha_k^2}h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} &= \frac{\frac{\partial^2}{\partial\alpha_k^2}\left[\int f(\mathbf{s},x_t;\boldsymbol{\theta})\,d\boldsymbol{\upsilon}_s\right]}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}\\ &= \frac{\int\frac{\partial^2}{\partial\alpha_k^2}\left[f(\mathbf{s},x_t;\boldsymbol{\theta})\right]d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}\\ &= \frac{\int\left\{\tau_{\alpha_k}(s_k;\alpha_k)+\left[\tau(s_k;\alpha_k)\right]^2\right\}f(\mathbf{s},x_t;\boldsymbol{\theta})\,d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}\\ &= E_t\left[\tau_{\alpha_k}(\alpha_k\cdot X_{t-k})+\left[\tau(\alpha_k\cdot X_{t-k})\right]^2\right]\end{aligned}$$

$$\begin{aligned}
\frac{\frac{\partial^2}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} &= \frac{\frac{\partial^2}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\left[\int f(\mathbf{s},x_t;\boldsymbol{\theta})\,d\boldsymbol{\upsilon}_s\right]}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \frac{\partial^2}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\left[f(\mathbf{s},x_t;\boldsymbol{\theta})\right]d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \boldsymbol{\gamma}_{\boldsymbol{\lambda}}\left(x_t-\sum_{k=1}^{p}s_k;\boldsymbol{\lambda}\right)f(\mathbf{s},x_t;\boldsymbol{\theta})\,d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&\quad + \frac{\int \boldsymbol{\gamma}\left(x_t-\sum_{k=1}^{p}s_k;\boldsymbol{\lambda}\right)\boldsymbol{\gamma}\left(x_t-\sum_{k=1}^{p}s_k;\boldsymbol{\lambda}\right)'f(\mathbf{s},x_t;\boldsymbol{\theta})\,d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= E_t\left[\boldsymbol{\gamma}_{\boldsymbol{\lambda}}(\varepsilon_t)+\boldsymbol{\gamma}(\varepsilon_t)\boldsymbol{\gamma}(\varepsilon_t)'\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\frac{\partial^2}{\partial\alpha_k\partial\boldsymbol{\lambda}}h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} &= \frac{\frac{\partial^2}{\partial\alpha_k\partial\boldsymbol{\lambda}}\left[\int f(\mathbf{s},x_t;\boldsymbol{\theta})\,d\boldsymbol{\upsilon}_s\right]}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \frac{\partial^2}{\partial\alpha_k\partial\boldsymbol{\lambda}}\left[f(\mathbf{s},x_t;\boldsymbol{\theta})\right]d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \tau(s_k;\alpha_k)\boldsymbol{\gamma}(\varepsilon_t;\boldsymbol{\lambda})f(\mathbf{s},x_t;\boldsymbol{\theta})\,d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= E_t\left[\tau(\alpha_k\cdot X_{t-k})\boldsymbol{\gamma}(\varepsilon_t)\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\frac{\partial^2}{\partial\alpha_m\alpha_n}h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} &= \frac{\frac{\partial^2}{\partial\alpha_m\alpha_n}\left[\int f(\mathbf{s},x_t;\boldsymbol{\theta})\,d\boldsymbol{\upsilon}_s\right]}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \frac{\partial^2}{\partial\alpha_m\alpha_n}\left[f(\mathbf{s},x_t;\boldsymbol{\theta})\right]d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= \frac{\int \tau(s_m;\alpha_m)\tau(s_n;\alpha_n)f(\mathbf{s},x_t;\boldsymbol{\theta})\,d\boldsymbol{\upsilon}_s}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})} \\
&= E_t\left[\tau(\alpha_m\cdot X_{t-m})\tau(\alpha_n\cdot X_{t-n})\right].
\end{aligned}$$

Finally, the Fisher information can then be written as follows:

$$\begin{aligned}
\ddot{\ell}_{\alpha_k\alpha_k} &= \frac{\partial^2\ln L(\boldsymbol{\theta})}{\partial\alpha_k^2} \\
&= \sum_{t=p+1}^{T}\left\{\frac{\frac{\partial^2}{\partial\alpha_k^2}h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}-\left[\frac{\frac{\partial}{\partial\alpha_k}h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}{h(x_t|\mathbf{x}_{-p};\boldsymbol{\theta})}\right]^2\right\}
\end{aligned}$$

$$= \sum_{t=p+1}^{T} \left\{ E_t \left[\tau_{\alpha_k} (\alpha_k \cdot X_{t-k}) + [\tau (\alpha_k \cdot X_{t-k})]^2 \right] - (E_t [\tau (\alpha_k \cdot X_{t-k})])^2 \right\}$$

$$= \sum_{t=p+1}^{T} \{ E_t [\tau_{\alpha_k} (\alpha_k \cdot X_{t-k})] + Var_t [\tau (\alpha_k \cdot X_{t-k})] \}$$

$$\begin{aligned}
\ddot{\ell}_{\lambda\lambda'} &= \frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \\
&= \sum_{t=p+1}^{T} \left\{ \frac{\frac{\partial^2}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})}{h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})} - \frac{\frac{\partial}{\partial \boldsymbol{\lambda}} h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})}{h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})} \frac{\frac{\partial}{\partial \boldsymbol{\lambda}'} h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})}{h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})} \right\} \\
&= \sum_{t=p+1}^{T} \left\{ E_t \left[\boldsymbol{\gamma}_{\boldsymbol{\lambda}} (\varepsilon_t) + \boldsymbol{\gamma} (\varepsilon_t) \boldsymbol{\gamma} (\varepsilon_t)' \right] - E_t [\boldsymbol{\gamma} (\varepsilon_t)] E_t \left[\boldsymbol{\gamma} (\varepsilon_t)' \right] \right\} \\
&= \sum_{t=p+1}^{T} \{ E_t [\boldsymbol{\gamma}_{\boldsymbol{\lambda}} (\varepsilon_t)] + Var_t [\boldsymbol{\gamma} (\varepsilon_t)] \}
\end{aligned}$$

$$\begin{aligned}
\ddot{\ell}_{\alpha_k \boldsymbol{\lambda}} &= \frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial \alpha_k \partial \boldsymbol{\lambda}} \\
&= \sum_{t=p+1}^{T} \left\{ \frac{\frac{\partial^2}{\partial \alpha_k \partial \boldsymbol{\lambda}} h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})}{h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})} - \frac{\frac{\partial}{\partial \alpha_k} h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})}{h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})} \frac{\frac{\partial}{\partial \boldsymbol{\lambda}} h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})}{h(x_t | \mathbf{x}_{-p}; \boldsymbol{\theta})} \right\} \\
&= \sum_{t=p+1}^{T} \{ E_t [\tau (\alpha_k \cdot X_{t-k}) \boldsymbol{\gamma} (\varepsilon_t)] - E_t [\tau (\alpha_k \cdot X_{t-k})] E_t [\boldsymbol{\gamma} (\varepsilon_t)] \} \\
&= \sum_{t=p+1}^{T} Cov_t [\tau (\alpha_k \cdot X_{t-k}), \boldsymbol{\gamma} (\varepsilon_t)] .
\end{aligned}$$

■

Proof. of Corollary 9.1. Substituting (9.7) and (9.8) into (9.5) and adjusting the summation range yields

$$\begin{aligned}
& P(X_t | X_{t-1}, \ldots, X_{t-p}) \\
&= \sum_{i_1=0}^{\min(X_{t-1}, X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1} (1 - \alpha_1)^{X_{t-1} - i_1} P(X_t - i_1 | X_{t-2}, \ldots, X_{t-p}).
\end{aligned}$$

By repeated substitution we see that

$$
\begin{aligned}
& P(X_t|X_{t-1},\ldots,X_{t-p}) \\
= & \sum_{i_1=0}^{\min(X_{t-1},X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} P(X_t - i_1|X_{t-2},\ldots,X_{t-p}) \\
= & \sum_{i_1=0}^{\min(X_{t-1},X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} \left\{ \sum_{i_2=0}^{\min(X_{t-2},X_t-i_1)} \binom{X_{t-2}}{i_2} \alpha_2^{i_2}(1-\alpha_2)^{X_{t-2}-i_2} \right. \\
& \left. P(X_t - i_1 - i_2|X_{t-3},\ldots,X_{t-p}) \right\} \\
& \vdots \\
= & \sum_{i_1=0}^{\min(X_{t-1},X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} \sum_{i_2=0}^{\min(X_{t-2},X_t-i_1)} \binom{X_{t-2}}{i_2} \alpha_2^{i_2}(1-\alpha_2)^{X_{t-2}-i_2} \\
& \cdots \left\{ \sum_{i_{p-1}=0}^{\min[X_{t-(p-1)},X_t-(i_1+\cdots+i_{p-2})]} \binom{X_{t-(p-1)}}{i_{p-1}} \alpha_{p-1}^{i_{p-1}}(1-\alpha_{p-1})^{X_{t-(p-1)}-i_{p-1}} \right. \\
& \left. P\left(X_t - \sum_{k=1}^{p-1} i_k \middle| X_{t-p} \right) \right\} \\
= & \sum_{i_1=0}^{\min(X_{t-1},X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} \sum_{i_2=0}^{\min(X_{t-2},X_t-i_1)} \binom{X_{t-2}}{i_2} \alpha_2^{i_2}(1-\alpha_2)^{X_{t-2}-i_2} \\
& \cdots \sum_{i_p=0}^{\min[X_{t-p},X_t-(i_1+\cdots+i_{p-1})]} \binom{X_{t-p}}{i_p} \alpha_p^{i_p}(1-\alpha_p)^{X_{t-p}-i_p} \frac{e^{-\lambda}\lambda^{X_t-(i_1+\cdots+i_p)}}{[X_t-(i_1+\cdots+i_p)]!}.
\end{aligned}
$$

The starting value is given in the bottom line which involves the probability function of the Poisson distribution. ■

Proof. of Corollary 9.2. Using (9.11) and (9.12) we see that

$$
\dot{\ell}_{\alpha_k} = \sum_{t=p+1}^{T} E_t \left[\frac{\alpha_k \circ X_{t-k}}{\alpha_k(1-\alpha_k)} - \frac{\alpha_k X_{t-k}}{\alpha_k(1-\alpha_k)} \right]
$$

$$= \frac{1}{\alpha_k(1-\alpha_k)} \sum_{t=p+1}^{T} \{E_t [\alpha_k \circ X_{t-k}] - \alpha_k X_{t-k}\}$$

$$= \frac{1}{\alpha_k(1-\alpha_k)} \sum_{t=p+1}^{T} \{E_t [(\alpha_k \circ X_{t-k})] - E_{t-1} [(\alpha_k \circ X_{t-k})]\}.$$

By the same method,

$$\dot{\ell}_\lambda = \sum_{t=p+1}^{T} E_t \left[\frac{\varepsilon_t}{\lambda} - 1\right]$$

$$= \frac{1}{\lambda} \sum_{t=p+1}^{T} \{E_t [\varepsilon_t] - E_{t-1} [\varepsilon_t]\}.$$

■

Proof. of Corollary 9.3. By differentiating (9.11) and (9.12), we can verify that

$$\tau_{\alpha_k}(s_k; \alpha_k) = \frac{(2\alpha_k - 1)\, s_k}{\alpha_k^2(1-\alpha_k^2)} - \frac{X_{t-k}}{(1-\alpha_k)^2}$$

and

$$\gamma_\lambda(\varepsilon_t) = -\frac{\varepsilon_t}{\lambda^2}.$$

Thus, the Fisher information can be written as

$$\ddot{\ell}_{\alpha_k \alpha_k} = \sum_{t=p+1}^{T} \left\{ E_t \left[\frac{(2\alpha_k - 1)\,(\alpha_k \circ X_{t-k})}{\alpha_k^2(1-\alpha_k^2)} - \frac{X_{t-k}}{(1-\alpha_k)^2} \right] \right.$$

$$\left. -Var_t \left[-\frac{X_{t-k}}{(1-\alpha_k)} + \frac{(\alpha_k \circ X_{t-k})}{\alpha_k(1-\alpha_k)} \right] \right\}$$

$$= \frac{1}{\alpha_k^2 (1-\alpha_k)^2} \sum_{t=p+1}^{T} \{(2\alpha_k - 1) E_t[\alpha_k \circ X_{t-k}]$$

$$+ Var_t[\alpha_k \circ X_{t-k}] - \alpha_k E_{t-1}[\alpha_k \circ X_{t-k}]\},$$

$$\begin{aligned}
\ddot{\ell}_{\alpha_m\alpha_n} &= \sum_{t=p+1}^{T} Cov_t\left[\frac{\alpha_m \circ X_{t-m}}{\alpha_m(1-\alpha_m)} - \frac{X_{t-m}}{(1-\alpha_m)}, \frac{\alpha_n \circ X_{t-n}}{\alpha_n(1-\alpha_n)} - \frac{X_{t-n}}{(1-\alpha_n)}\right] \\
&= \sum_{t=p+1}^{T} Cov_t\left[\frac{\alpha_m \circ X_{t-m}}{\alpha_m(1-\alpha_m)}, \frac{\alpha_n \circ X_{t-n}}{\alpha_n(1-\alpha_n)}\right] \\
&= \frac{1}{\alpha_m\alpha_n(1-\alpha_m)(1-\alpha_n)} \sum_{t=p+1}^{T} Cov_t\left[\alpha_m \circ X_{t-m}, \alpha_n \circ X_{t-n}\right],
\end{aligned}$$

$$\begin{aligned}
\ddot{\ell}_{\alpha_k\lambda} &= \sum_{t=p+1}^{T} Cov_t\left[\frac{\alpha_k \circ X_{t-k}}{\alpha_k(1-\alpha_k)} - \frac{X_{t-k}}{(1-\alpha_k)}, \frac{\varepsilon_t}{\lambda} - 1\right] \\
&= \sum_{t=p+1}^{T} Cov_t\left[\frac{\alpha_k \circ X_{t-k}}{\alpha_k(1-\alpha_k)}, \frac{\varepsilon_t}{\lambda}\right] \\
&= \frac{1}{\lambda\alpha_k(1-\alpha_k)} \sum_{t=p+1}^{T} Cov_t\left[\alpha_k \circ X_{t-k}, \varepsilon_t\right],
\end{aligned}$$

and

$$\begin{aligned}
\ddot{\ell}_{\lambda\lambda} &= \sum_{t=p+1}^{T} \left\{E_t[-\frac{\varepsilon_t}{\lambda^2}] + Var_t\left[\frac{\varepsilon_t}{\lambda} - 1\right]\right\} \\
&= \frac{1}{\lambda^2} \sum_{t=p+1}^{T} \left\{Var_t[\varepsilon_t] - E_t[\varepsilon_t]\right\}.
\end{aligned}$$

■

Appendix 2: A Numerical Procedure for Calculating Joint Probabilities for the INAR(p) Model

In this appendix we present a numerical procedure that transforms the conditional probability $P(X_t|X_{t-1},\ldots,X_{t-p})$ of a stationary INAR process into the joint probability $P(X_{t-1},\ldots,X_{t-p})$ and $P(X_t,X_{t-1},\ldots,X_{t-p})$. $P(X_{t-1},\ldots,X_{t-p})$ can be used for calculating the unconditional likelihood of the model and $P(X_t,X_{t-1},\ldots,X_{t-p})$ is necessary for evaluating the expected Fisher information.

Without any real loss of generality, we consider a simple INAR(2) process, whose conditional probability function is given by $P(X_t|X_{t-1},X_{t-2})$. We assume that the probability of observing a count larger than $M = 1$ is negligible. Thus, for this particular Markov Chain process, the transition matrix can be written as follows:

$$\begin{array}{cc|cccc} & X_t & 0 & 0 & 1 & 1 \\ X_{t-2} & X_{t-1} & 0 & 1 & 0 & 1 \\ \hline 0 & 0 & P(0|0,0) & 0 & P(1|0,0) & 0 \\ 1 & 0 & P(0|0,1) & 0 & P(1|0,1) & 0 \\ 0 & 1 & 0 & P(0|1,0) & 0 & P(1|1,0) \\ 1 & 1 & 0 & P(0|1,1) & 0 & P(1|1,1) \end{array}$$

According to Bayes's rule, the joint probability

$$\begin{aligned} & P(X_t = 0, X_{t-1} = 0) \\ = \; & P(X_t = 0|X_{t-1} = 0, X_{t-2} = 0)P(X_{t-1} = 0, X_{t-2} = 0) \\ + \; & P(X_t = 0|X_{t-1} = 0, X_{t-2} = 1)P(X_{t-1} = 0, X_{t-2} = 1). \end{aligned}$$

Stationarity implies that

$$P(X_t = 0, X_{t-1} = 0) = P(X_{t-1} = 0, X_{t-2} = 0)$$

and thus

$$\begin{aligned} & P(X_{t-1}=0, X_{t-2}=0) \\ = \; & P(X_t=0|X_{t-1}=0, X_{t-2}=0)P(X_{t-1}=0, X_{t-2}=0) \\ + \; & P(X_t=0|X_{t-1}=0, X_{t-2}=1)P(X_{t-1}=0, X_{t-2}=1). \end{aligned} \tag{A2.1}$$

Rearranging (A2.1) yields

$$[P(0|0,0)-1]P(0,0)+P(0|0,1)P(0,1)=0$$

and a similar argument gives another two equalities

$$\begin{aligned} P(1|0,0)P(0,0)-P(1,0)+P(1|0,1)P(0,1) &= 0 \\ P(0|1,0)P(1,0)-P(0,1)+P(0|1,1)P(1,1) &= 0. \end{aligned} \tag{A2.2}$$

The unknown quantities in the above three equations are $P(0,0)$, $P(1,0)$, $P(0,1)$, and $P(1,1)$. Since the probability of observing a count other than 0 and 1 is assumed to be zero

$$P(1,1)=1-[P(0,0)+P(1,0)+P(0,1)]. \tag{A2.3}$$

Substituting (A2.3) into (A2.2) yields a system of three simultaneous equations with three unknown quantities[22]. Thus, $P(0,0)$, $P(1,0)$, and $P(0,1)$ are uniquely determined and so is $P(1,1)$. The joint probability $P(X_{t-1}, X_{t-2})$ thus obtained can then be used along with the conditional probability $P(X_t|X_{t-1}, X_{t-2})$ to get the joint probability $P(X_t, X_{t-1}, X_{t-2})$ since

$$P(X_t, X_{t-1}, X_{t-2}) = P(X_t|X_{t-1}, X_{t-2})P(X_{t-1}, X_{t-2}).$$

[22]For a general INAR(p) process, there will be a system of $(M+1)^p - 1$ simultaneous equations with the same number of unknown quantities that are the joint probabilities $P(X_t, \ldots X_{t-p})$.

Furthermore, one could also easily get the marginal probability $P(X_t)$ from the joint probability $P(X_t, X_{t-1})$ by summing over all possible values of X_{t-1}.

Appendix 3: Time t Conditional Expectations for the INAR(p) Model with Poisson Innovations

In this appendix we provide specific formulae for the time t conditional expectations required for the score functions and the Fisher information matrix for the INAR(p) model with Poisson innovations. It turns out that all these conditional expectations can be expressed as functions of the transition probability, and that the recursive representation and mechanism effectively facilitate the derivation of these results. Proofs of the results in (9.13) and (9.14) are given in details. The remaining can be obtained by similar reasoning.

Proof. of Equation (9.13). Using the recursive representation of the transition probability, it follows from the Lemma 1 that

$$
\begin{aligned}
& E_t\left[\alpha_1 \circ X_{t-1}\right] \\
= & \frac{\sum_{i_1=0}^{\min(X_{t-1},X_t)} i_1 \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} P(X_t - i_1 | X_{t-2}, \ldots, X_{t-p})}{P(X_t | X_{t-1}, \ldots, X_{t-p})} \\
= & \frac{\sum_{i_1=0}^{\min(X_{t-1},X_t)} X_{t-1} \binom{X_{t-1}-1}{i_1-1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} P(X_t - i_1 | X_{t-2}, \ldots, X_{t-p})}{P(X_t | X_{t-1}, \ldots, X_{t-p})} \\
= & \alpha_1 X_{t-1} \sum_{i_1=0}^{\min(X_{t-1},X_t)} \left\{ \binom{X_{t-1}-1}{i_1-1} \alpha_1^{i_1-1}(1-\alpha_1)^{(X_{t-1}-1)-(i_1-1)} \right. \\
& \left. \times P(X_t - i_1 | X_{t-2}, \ldots, X_{t-p}) \right\} \frac{1}{P(X_t | X_{t-1}, \ldots, X_{t-p})} \\
= & \alpha_1 X_{t-1} \sum_{i_1=0}^{\min(X_{t-1}-1,X_t-1)} \left\{ \binom{X_{t-1}-1}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{(X_{t-1}-1)-i_1} \right.
\end{aligned}
$$

$$\left.\times P(X_t - 1 - i_1 | X_{t-2}, \ldots, X_{t-p})\right\} \frac{1}{P(X_t | X_{t-1}, \ldots, X_{t-p})}$$
$$= \frac{\alpha_1 X_{t-1} P(X_t - 1 | X_{t-1} - 1, \ldots, X_{t-p})}{P(X_t | X_{t-1}, \ldots, X_{t-p})}$$

Note that the order of components is irrelevant to the result of a convolution. Thus, the above result also implies that

$$E_t[\alpha_k \circ X_{t-k}]$$
$$= \frac{\alpha_k X_{t-k} P(X_t - 1 | X_{t-k} - 1, X_{t-1}, \ldots, X_{t-k+1}, X_{t-k-1}, \ldots, X_{t-p})}{P(X_t | X_{t-1}, \ldots, X_{t-p})}$$
$$= \frac{\alpha_k X_{t-k} P(X_t - 1 | X_{t-1}, \ldots, X_{t-k} - 1, \ldots, X_{t-p})}{P(X_t | X_{t-1}, \ldots, X_{t-p})}.$$

■

Proof. of Equation (9.14). Denote $E_t^{(p)}[\varepsilon_t]$ as the time t conditional expectation of ε_t for the INAR(p) model. Then for the INAR(1) model

$$E_t^{(1)}[\varepsilon_t] P(X_t | X_{t-1})$$
$$= \sum_{i_1=0}^{\min(X_{t-1}, X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1} (1-\alpha_1)^{X_{t-1}-i_1} (X_t - i_1) \frac{e^{-\lambda} \lambda^{X_t - i_1}}{(X_t - i_1)!}$$
$$= \sum_{i_1=0}^{\min(X_{t-1}, X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1} (1-\alpha_1)^{X_{t-1}-i_1} \frac{e^{-\lambda} \lambda^{X_t - i_1}}{(X_t - 1 - i_1)!}$$
$$= \lambda \sum_{i_1=0}^{\min(X_{t-1}, X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1} (1-\alpha_1)^{X_{t-1}-i_1} \frac{e^{-\lambda} \lambda^{X_t - 1 - i_1}}{(X_t - 1 - i_1)!}$$
$$= \lambda \sum_{i_1=0}^{\min(X_{t-1}, X_t - 1)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1} (1-\alpha_1)^{X_{t-1}-i_1} \frac{e^{-\lambda} \lambda^{X_t - 1 - i_1}}{(X_t - 1 - i_1)!}$$
$$= \lambda P(X_t - 1 | X_{t-1}). \qquad \text{(A3.1)}$$

For the INAR(2) model, we have

$$
\begin{aligned}
& E_t^{(2)}\left[\varepsilon_t\right] P\left(X_t | X_{t-1}, X_{t-2}\right) \\
= & \sum_{i_1=0}^{\min\left(X_{t-1}, X_t\right)} \left\{ \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} \right. \\
& \left. \sum_{i_2=0}^{\min\left(X_{t-2}, X_t-i_1\right)} \binom{X_{t-2}}{i_2} \alpha_2^{i_2}(1-\alpha_2)^{X_{t-2}-i_2}\left(X_t-i_1-i_2\right) \frac{e^{-\lambda} \lambda^{X_t-i_1-i_2}}{\left(X_t-i_1-i_2\right)!} \right\} \\
= & \sum_{i_1=0}^{\min\left(X_{t-1}, X_t\right)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} \lambda P\left(X_t-i_1-1 | X_{t-2}\right) \\
= & \sum_{i_1=0}^{\min\left(X_{t-1}, X_t-1\right)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} \lambda P\left(X_t-1-i_1 | X_{t-2}\right) \\
= & \lambda P\left(X_t-1 | X_{t-1}, X_{t-2}\right) \qquad \text{(A3.2)}
\end{aligned}
$$

Note that the second equality used the result in (A3.1). Similarly, the result in (A3.2) can be used to obtain

$$
E_t^{(3)}\left[\varepsilon_t\right] P\left(X_t | X_{t-1}, X_{t-2}, X_{t-3}\right) = \lambda P\left(X_t-1 | X_{t-1}, X_{t-2}, X_{t-3}\right)
$$

and so on up to any given order p with the general result as

$$
\begin{aligned}
& E_t^{(p)}\left[\varepsilon_t\right] P\left(X_t | X_{t-1}, \ldots, X_{t-p}\right) \\
= & \sum_{i_1=0}^{\min\left(X_{t-1}, X_t\right)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} \lambda P\left(X_t-i_1-1 | X_{t-2}, \ldots, X_{t-p}\right) \\
= & \sum_{i_1=0}^{\min\left(X_{t-1}, X_t-1\right)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1}(1-\alpha_1)^{X_{t-1}-i_1} \lambda P\left(X_t-1-i_1 | X_{t-2}, \ldots, X_{t-p}\right) \\
= & \lambda P\left(X_t-1 | X_{t-1}, \ldots, X_{t-p}\right)
\end{aligned}
$$

or

$$
E_t\left[\varepsilon_t\right] = E_t^{(p)}\left[\varepsilon_t\right] = \frac{\lambda P\left(X_t-1 | X_{t-1}, \ldots, X_{t-p}\right)}{P\left(X_t | X_{t-1}, \ldots, X_{t-p}\right)}.
$$

■

Similar reasoning based on Lemma 1 and the recursive representation of the transition probability also yields the following results:

$$\begin{aligned} & E_t\left[(\alpha_k \circ X_{t-k})^2\right] - E_t\left[\alpha_k \circ X_{t-k}\right] \\ = & \frac{\alpha_k^2 X_{t-k}(X_{t-k}-1)P(X_t-2|X_{t-1},\ldots,X_{t-k}-2,\ldots,X_{t-p})}{P(X_t|X_{t-1},\ldots,X_{t-p})}, \end{aligned}$$

$$\begin{aligned} & E_t[\varepsilon_t^2] - E_t[\varepsilon_t] \\ = & \frac{\lambda^2 P(X_t-2|X_{t-1},\ldots,X_{t-p})}{P(X_t|X_{t-1},\ldots,X_{t-p})}, \end{aligned}$$

$$\begin{aligned} & E_t\left[(\alpha_m \circ X_{t-m})\,(\alpha_n \circ X_{t-n})\right] \\ = & \frac{\alpha_m\alpha_n X_{t-m}X_{t-n}P(X_t-2|X_{t-1},\ldots,X_{t-m}-1,\ldots,X_{t-n}-1,\ldots,X_{t-p})}{P(X_t|X_{t-1},\ldots,X_{t-p})}, \end{aligned}$$

$$\begin{aligned} & E_t[(\alpha_k \circ X_{t-k})\,\varepsilon_t] \\ = & \frac{\alpha_k\lambda X_{t-k}P(X_t-2|X_{t-1},\ldots,X_{t-k}-1,\ldots,X_{t-p})}{P(X_t|X_{t-1},\ldots,X_{t-p})}. \end{aligned}$$

Appendix 4: Asymptotic Distribution of the Maximum Likelihood Estimator for the INAR(p) Model with Poisson Innovations

In this appendix we set out the proof of the asymptotic distribution of the Maximum likelihood estimator for the INAR(p) model with Poisson innovations. Freeland (1998) provided a detailed discussion of the regularity conditions needed for the ML estimates of the INAR(p) model to be consistent and asymptotically normal. According to Freeland (1998), for ergodic processes if the Fisher information is finite and positive definite then the regularity conditions hold. Since the INAR(p) process is ergodic according to Du and Li (1991), our main task in this appendix is to show that the Fisher information is finite and positive definite.

The following theorem shows that the Fisher information is finite.

Theorem A4.1 *Let* $u_{\alpha_k,t} = U_{\alpha_k,t} - U_{\alpha_k,t-1}$, *for* $k \in [1,p]$, $u_{\lambda,t} = U_{\lambda,t} - U_{\lambda,t-1}$ *and* $u_t = \left(u_{\alpha_1,t}, \ldots, u_{\alpha_p,t}, u_{\lambda,t}\right)'$, *where* $U_{\alpha_k,t}$ *and* $U_{\lambda,t}$ *are the score functions for the INAR(p) model with Poisson innovations with respect to* α_k *and* λ, *respectively. Further, let* $\dot{u}_t$ *denote the matrix of partial derivatives of* u_t *with respect to* α_k *and* λ, *and let* $\dot{U}_t = \sum_1^n \dot{u}_t$. *For any* $p+1$ *dimensional vector* l *and any positive integer* Z, $E\left[\left(l^T u_t\right)^Z\right] < \infty$ *and* $E\left[\left(l^T \dot{u}_t l\right)\right] < \infty$.

The proof of Theorem A4.1 is based on the following lemma.

Lemma 2 *Let Z and $i_1, \ldots, i_p$ be any non-negative integers, which satisfy $i_1, \ldots, i_p \in [0, Z]$ and $(i_1 + \cdots + i_p) \leq Z$. Then, for the INAR(p) model in (8.1), the following inequality holds:*

$$X_t^Z \geq E_t[\alpha_1 \circ X_{t-1}]^{Z-(i_1+\cdots+i_p)} E_t[\alpha_2 \circ X_{t-2}]^{i_1} \cdots E_t[\alpha_p \circ X_{t-p}]^{i_{p-1}} E_t[\varepsilon_t]^{i_p}$$

Proof. of Lemma 2. For the INAR(p) model in (8.1), we can write

$$X_t = E_t[X_t] = E_t[\alpha_1 \circ X_{t-1}] + E_t[\alpha_2 \circ X_{t-2}] + \cdots + E_t[\alpha_p \circ X_{t-p}] + E_t[\varepsilon_t],$$

and

$$X_t^Z = \{E_t[\alpha_1 \circ X_{t-1}] + E_t[\alpha_2 \circ X_{t-2}] + \cdots + E_t[\alpha_p \circ X_{t-p}] + E_t[\varepsilon_t]\}^Z$$

is therefore a polynomial in $E_t[\alpha_k \circ X_{t-k}]$ and $E_t[\varepsilon_t]$ of degree Z. That is

$$\begin{aligned} X_t^Z = \sum_{\substack{i_1=0 \\ (i_1+\cdots+i_p)\leq Z}}^{Z} \cdots \sum_{i_p=0}^{Z} \{ & a_{i_1 \cdots i_p} E_t[\alpha_1 \circ X_{t-1}]^{Z-(i_1+\cdots+i_p)} E_t[\alpha_2 \circ X_{t-2}]^{i_1} \\ & \cdots E_t[\alpha_p \circ X_{t-p}]^{i_{p-1}} E_t[\varepsilon_t]^{i_p} \} \end{aligned} \tag{A4.1}$$

for some non-negative constants $a_{i_1 \cdots i_p}$. It should be noted that the p-fold summation in (A4.1) is restricted to sets of values of $i_1, \ldots, i_p$ which satisfy $(i_1 + \cdots + i_p) \leq Z$. Since all time t conditional expectations are non-negative, all summands in (A4.1) are non-negative. It follows that

$$X_t^Z \geq E_t[\alpha_1 \circ X_{t-1}]^{Z-(i_1+\cdots+i_p)} E_t[\alpha_2 \circ X_{t-2}]^{i_1} \cdots E_t[\alpha_p \circ X_{t-p}]^{i_{p-1}} E_t[\varepsilon_t]^{i_p}$$

for any non-negative integers Z and $i_1, \ldots, i_p$ which satisfy $i_1, \ldots, i_p \in [0, Z]$ and $(i_1 + \cdots + i_p) \leq Z$ ■

Proof. of Theorem A4.1. Recall from Corollary 9.2 that for the INAR(p) model with Poisson innovations

$$\begin{aligned} U_{\alpha_k,t} &= \frac{1}{\alpha_k(1-\alpha_k)} \sum_{t=p+1}^{T} \{E_t[\alpha_k \circ X_{t-k}] - E_{t-1}[\alpha_k \circ X_{t-k}]\} \\ U_{\lambda,t} &= \frac{1}{\lambda} \sum_{t=p+1}^{T} \{E_t[\varepsilon_t] - E_{t-1}[\varepsilon_t]\} \end{aligned}$$

for $k \in [1,p]$. Thus, for any $p+1$ dimensional vector $l = (l_1, \ldots, l_p, l_{p+1})'$,

$$\left\{l^T u_t\right\}^Z = \left\{l_1 u_{\alpha_1,t} + \cdots + l_p u_{\alpha_p,t} + l_{p+1} u_{\lambda,t}\right\}^Z$$

is a polynomial in $E_t[\alpha_k \circ X_{t-k}]$ and $E_t[\varepsilon_t]$ of degree Z, which can be written as

$$\begin{aligned} & \left\{l^T u_t\right\}^Z \\ = & \left\{l_1 u_{\alpha_1,t} + \cdots + l_p u_{\alpha_p,t} + l_{p+1} u_{\lambda,t}\right\}^Z \\ = & \sum_{\substack{j_1=0 \\ (j_1+\cdots+j_p)\le Z}}^{Z} \cdots \sum_{j_p=0}^{Z} \left\{a_{j_1\cdots j_p} E_t[\alpha_1 \circ X_{t-1}]^{Z-(j_1+\cdots+j_p)} E_t[\alpha_2 \circ X_{t-2}]^{j_1} \right. \\ & \left. \cdots E_t[\alpha_p \circ X_{t-p}]^{j_{p-1}} E_t[\varepsilon_t]^{j_p}\right\} \end{aligned}$$

for some constants $a_{j_1\cdots j_p}$. The expected value of $\left\{l^T u_t\right\}^Z$ is

$$\begin{aligned} & E\left[\left\{l^T u_t\right\}^Z\right] \\ = & E\left[\sum_{\substack{j_1=0 \\ (j_1+\cdots+j_p)\le Z}}^{Z} \cdots \sum_{j_p=0}^{Z} a_{j_1\cdots j_p} E_t[\alpha_1 \circ X_{t-1}]^{Z-(j_1+\cdots+j_p)} E_t[\alpha_2 \circ X_{t-2}]^{j_1} \right. \\ & \left. \cdots E_t[\alpha_p \circ X_{t-p}]^{j_{p-1}} E_t[\varepsilon_t]^{j_p}\right] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\substack{j_1=0 \\ (j_1+\cdots+j_p)\le Z}}^{Z} \cdots \sum_{j_p=0}^{Z} a_{j_1\cdots j_p} E\left[E_t[\alpha_1 \circ X_{t-1}]^{Z-(j_1+\cdots+j_p)} E_t[\alpha_2 \circ X_{t-2}]^{j_1}\right. \\
&\quad \left. \cdots E_t[\alpha_p \circ X_{t-p}]^{j_{p-1}} E_t[\varepsilon_t]^{j_p}\right] \\
&\le \sum_{\substack{j_1=0 \\ (j_1+\cdots+j_p)\le Z}}^{Z} \cdots \sum_{j_p=0}^{Z} \left|a_{j_1\cdots j_p}\right| E\left[E_t[\alpha_1 \circ X_{t-1}]^{Z-(j_1+\cdots+j_p)} E_t[\alpha_2 \circ X_{t-2}]^{j_1}\right. \\
&\quad \left. \cdots E_t[\alpha_p \circ X_{t-p}]^{j_{p-1}} E_t[\varepsilon_t]^{j_p}\right] \\
&\le \sum_{\substack{j_1=0 \\ (j_1+\cdots+j_p)\le Z}}^{Z} \cdots \sum_{j_p=0}^{Z} \left|a_{j_1\cdots j_p}\right| E\left[X_t^Z\right] \\
&< \infty
\end{aligned}
$$

Note that the second inequality follows from Lemma 2. The proof for the second part is automatic since $E\left[\left(l^T u_t\right)^2\right] = E\left[l^T u_t u_t^T l\right] = E\left[l^T \dot{u}_t l\right]$. ■

The next theorem shows that the Fisher information is positive definite.

Theorem A4.2 In the parameter space satisfying $\alpha_k \in (0,1)$, for $k \in [1,p]$, and $\lambda \in (0,\infty)$, $E\left[\left(l^T u_t\right)^2\right] = 0$ *if and only if* $l_1 = l_2 = \cdots = l_{p+1} = 0$, *where* $l = (l_1, \ldots, l_p, l_{p+1})'$.

Proof. of Theorem A4.2. It is sufficient to show that

$$
Var\left\{l_1 E_t\left[\alpha_1 \circ X_{t-1}\right] + \cdots + l_p E_t\left[\alpha_p \circ X_{t-p}\right] + l_{p+1} E_t[\varepsilon_t]\right\} \neq 0 \qquad \text{(A4.2)}
$$

Since for the INAR(p) model with Poisson innovations

$$
\begin{aligned}
E_t\left[\alpha_k \circ X_{t-k}\right] &= \frac{\alpha_k X_{t-k} P(X_t - 1 | X_{t-1}, \ldots, X_{t-k} - 1, \ldots, X_{t-p})}{P(X_t | X_{t-1}, \ldots, X_{t-p})} \\
E_t[\varepsilon_t] &= \frac{\lambda P(X_t - 1 | X_{t-1}, \ldots, X_{t-p})}{P(X_t | X_{t-1}, \ldots, X_{t-p})}
\end{aligned}
$$

for $k \in [1,p]$, (A4.2) is equivalent to

$$Var\left\{\left[\sum_{k=1}^{p} l_k\alpha_k X_{t-k}P(X_t-1|X_{t-1},\ldots,X_{t-k}-1,\ldots,X_{t-p})\right] + l_{p+1}\lambda P(X_t-1|X_{t-1},\ldots,X_{t-p})\right\} \neq 0 \tag{A4.3}$$

when l is non-zero. Note that it is assumed that the parameter space satisfies $\alpha_k \neq 0$ for $k \in [1,p]$ and $\lambda \neq 0$. This is because $\alpha_k = 0$ and $\lambda = 0$ are boundary values. Moreover, this assumption ensures that the model being considered does not degenerate. Otherwise, it can be seen that, for instance, if $\alpha_1 = 0$, there exists a non-zero vector $l = (1,0,\ldots,0)'$ such that the variance in (A4.3) is always zero.

We now prove (A4.3) by contradiction. Suppose there exists a non-zero $p+1$ dimension vector l such that

$$Var\left\{\left[\sum_{k=1}^{p} l_k\alpha_k X_{t-k}P(X_t-1|X_{t-1},\ldots,X_{t-k}-1,\ldots,X_{t-p})\right] + l_{p+1}\lambda P(X_t-1|X_{t-1},\ldots,X_{t-p})\right\} = 0$$

This implies that

$$\left[\sum_{k=1}^{p} l_k\alpha_k X_{t-k}P(X_t-1|X_{t-1},\ldots,X_{t-k}-1,\ldots,X_{t-p})\right] + l_{p+1}\lambda P(X_t-1|X_{t-1},\ldots,X_{t-p}) \equiv c \tag{A4.4}$$

almost everywhere for some constant c.

In particular, (A4.4) should hold when $X_t = 1$ and $X_{t-1} = X_{t-2} \cdots = X_{t-p} = 0$. Thus, (A4.4) becomes

$$\begin{aligned} c &\equiv l_{p+1}\lambda P(0|0,\ldots,0) \\ &= l_{p+1}\lambda e^{-\lambda} \end{aligned}$$

Because $\lambda \neq 0$, we can write

$$l_{p+1} = c\lambda^{-1}e^{\lambda}$$

Note that for any finite value of λ, $c = 0$ only if $l_{p+1} = 0$.

(A4.4) should also hold when $X_t = X_{t-1} = 1$ and $X_{t-2} = X_{t-3} \cdots = X_{t-p} = 0$. Thus

$$\begin{aligned} c &\equiv l_1\alpha_1 P(0|0,\ldots,0) + l_{p+1}\lambda P(0|1,0,\ldots,0) \\ &= l_1\alpha_1 e^{-\lambda} + l_{p+1}\lambda\,(1-\alpha_1)\,e^{-\lambda} \\ &= l_1\alpha_1 e^{-\lambda} + c\lambda^{-1}e^{\lambda}\lambda\,(1-\alpha_1)\,e^{-\lambda} \\ &= l_1\alpha_1 e^{-\lambda} + c\,(1-\alpha_1) \end{aligned}$$

which leads to

$$l_1\alpha_1 e^{-\lambda} = \alpha_1 c$$

Since $\alpha_1 \neq 0$, we get

$$l_1 = ce^{\lambda}$$

Note again that $c = 0$ only if $l_1 = 0$.

It can be shown that in general by setting $X_t = X_{t-k} = 1$ for any $k \in [1, p]$ and all other lags of X_t equal to zero, we can verify that $l_k = ce^{\lambda}$ must hold for all $k \in [1, p]$, and $c = 0$ only if $l_1 = l_2 = \cdots = l_p = 0$.

Finally, we consider when $X_t = 1$, $X_{t-1} = 2$ and $X_{t-2} = X_{t-3} = \cdots = X_{t-p} = 0$. (A4.4) then becomes

$$\begin{aligned} c &\equiv 2l_1\alpha_1 P(0|1,0,\ldots,0) + l_{p+1}\lambda P(0|2,0,\ldots,0) \\ &= 2ce^{\lambda}\alpha_1 P(0|1,0,\ldots,0) + c\lambda^{-1}e^{\lambda}\lambda P(0|2,0,\ldots,0) \\ &= 2ce^{\lambda}\alpha_1 (1-\alpha_1) e^{-\lambda} + c\lambda^{-1}e^{\lambda}\lambda (1-\alpha_1)^2 e^{-\lambda} \\ &= 2c\alpha_1 (1-\alpha_1) + c(1-\alpha_1)^2 \\ &= 2c(\alpha_1 - \alpha_1^2) + c(1 - 2\alpha_1 + \alpha_1^2) \end{aligned}$$

which leads to

$$c\alpha_1^2 = 0$$

This implies that either $\alpha_1 = 0$ or $c = 0$. In fact, by setting $X_t = 1$, $X_{t-k} = 2$ for any $k \in [1, p]$ and all other lags of X_t equal to zero, (A4.4) requires $c\alpha_k^2 = 0$, which implies that either $\alpha_k = 0$ or $c = 0$. Since $\alpha_k = 0$ is outside the parameter space, it must be the case that $c = 0$, which requires that $l_1 = \cdots = l_p = l_{p+1} = 0$. Thus, for any non-zero l, the condition in (A4.2) hold and therefore the Fisher information is positive definite. ■

Theorem A4.1 and A4.2 show that for the INAR(p) model with Poisson innovations the Fisher information matrix is positive definite. Since the process is ergodic,

all the regularity conditions for the asymptotics of the maximum likelihood estimator are satisfied. There for $\widehat{\boldsymbol{\theta}}_{ML} = \left(\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p, \widehat{\lambda}\right)$ has the following asymptotic distribution:

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{i}^{-1})$$

where the matrix $\mathbf{i}$ is the Fisher information.