Statistical inference for negative binomial processes with applications to market research

Vippal Savani
School of Mathematics, Cardiff University, Wales

August 2006

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed

Date 30-8-06

Statement 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed

Date 30-8-06

Statement 2

This thesis is the result of my own independent investigations, except where otherwise stated. Other sources are acknowledged by references and a bibliography at the end of the thesis.

Signed

Date 30-8-06

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed

Date 30-8-06

Acknowledgments

I would like to dedicate this thesis to my parents Bhanu and Harilal Savani who have provided me with enormous amounts of love and support throughout my education and upbringing. Without their support, I am certain that I would not have been able to progress this far into my education.

I would like to express enormous gratitude to my supervisors Prof. A. Zhigljavsky and Prof. N. Leonenko for having faith in my abilities and for guiding me throughout my thesis. I would like to show my appreciation to Prof. A. Zhigljavsky for providing me with a variety of research opportunities and also for providing me with a tremendous amount of support during my years of study. I would also like to thank Mike Kennard for computing support and Maureen Mills for her help and guidance on lots of different aspects. Similarly my gratitude extends to many members of staff at Cardiff University who have supported and helped develop my education.

In terms of research opportunities I wish to thank Prof. L. Pronzato for giving me the opportunity to conduct research with him at the CNRS in Nice. I wish also to thank Phil Parker of ACNielsen BASES for giving me the opportunity to conduct research for ACNielsen BASES and to have access to enormous amounts of precious data which, without him, would have been very difficult to obtain. I wish to also acknowledge the EPSRC for the funding of my PhD.

Finally, I would like to thank all my friends both within the university and outside of the university for all their support and for making my life enjoyable during my undergraduate and postgraduate years at university. Special thanks go to, in no particular order: Hinesh, the three Matt's, Sajid, Reena, Rachel, Anthony, Neil, Gill, Arti, Kalps, Bahader, Taz, Mans, Amit, Sheeps, Kishen, Ravi, Eric, Sachin, Rupen and all the postgraduates studying maths at Cardiff.

Abstract

The negative binomial distribution (NBD) and negative binomial processes have been used as natural models for events occurring in fields such as accident proneness; accidents and sickness; market research; insurance and risk theory. The fitting of negative binomial processes in practice has mainly focussed on fitting the one-dimensional distribution, namely the NBD, to data. In practice, the parameters of the NBD are usually estimated by using inefficient moment based estimation methods due to the ease in estimating moment based estimators in comparison to maximum likelihood estimators.

This thesis develops efficient moment based estimation methods for estimating parameters of the NBD that can be easily implemented in practice. These estimators, called power method estimators, are almost as efficient as maximum likelihood estimators when the sample is independent and identically distributed. For dependent NBD samples, the power method estimators are more efficient than the commonly used method of moments and zero term method estimators.

Fitting the one-dimensional marginal distribution of negative binomial processes to data gives partial information as to the adequacy of the process being fitted. This thesis further develops methods of statistical inference for data generated by negative binomial processes by comparing the dynamical properties of the process to the dynamical properties of data. For negative binomial autoregressive processes, the dynamical properties may be checked by using the autocorrelation function. The dynamical properties of the gamma Poisson process are considered by deriving the asymptotic covariance and correlation structures of estimators and functionals of the gamma Poisson process and verifying these structures against data.

The adequacy of two negative binomial processes, namely the gamma Poisson process and the negative binomial first-order autoregressive process, as models for consumer buying behavior are considered. The models are fitted to market research data kindly provided by ACNielsen BASES.

Contents

Chapter 1

Introduction

Negative binomial processes have been used as a natural model for events occurring in continuous or discrete time in many fields. Negative binomial processes have been successfully applied in the modeling of, for example: accident proneness (Greenwood and Yule (1920)); accidents and sickness (Lundberg (1964)); market research (Ehrenberg (1988)); risk theory (Grandell (1997)) and more recently in clinical trials (Cook and Wei (2003)).

The fitting of negative binomial processes in practice has mainly focussed on the fitting of the corresponding one-dimensional marginal distribution of the process, i.e. the negative binomial distribution (NBD), to data. Parameter estimation for the NBD using maximum likelihood has been considered independently by Fisher (1941) and Haldane (1941) and moment based estimators for the NBD have been considered by Anscombe (1950). Moment based estimators were considered due to the computational difficulties of maximum likelihood estimators. With the computational power available today, the computation of maximum likelihood estimators is no longer an issue. In many practices, however, the use of moment based estimators is still predominant (see e.g. Ehrenberg (1988)) even though maximum likelihood estimators are asymptotically the most efficient in the class of all asymptotically normal estimators.

A number of negative binomial processes have been presented in literature (see e.g. Barndorff-Nielsen and Yeo (1969); McKenzie (1986); Grandell (1997)). The fitting of the NBD to data over a fixed time interval therefore provides partial indication as to the adequacy of the theoretical process being fitted. Ehrenberg (1988) somewhat addressed the problem of assessing the goodness of fit of a particular negative binomial process, known as the gamma Poisson process, by comparing observed and expected frequencies as well as comparing the fit of numerous statistical measures computed in two different time intervals. No statistical tests were, however, presented to test whether the statistical measures computed in the two time intervals were (statistically) significantly different. The comparison of statistical measures computed in two different time intervals by Ehrenberg (1988) was mainly empirical.

Aim

The aim of this thesis is to further develop methods of statistical inference for data generated by negative binomial processes. This thesis will concentrate on methods of statistical inference that are efficient and methods that can be practically applied in the field of market research and other similar fields of practice.

The negative binomial processes considered will be restricted to the gamma Poisson process and the negative binomial first-order autoregressive process. Using empirical evidence, Ehrenberg (1988) has shown that the gamma Poisson process is suitable for modeling the number of purchases of various products by households within a population. The negative binomial first-order autoregressive process is a simple process in the family of autoregressive processes and will be used as a source of comparison against the gamma Poisson process.

Overview of the thesis

Chapter 2 provides a detailed background to the thesis. The chapter begins with a description of the numerous ways in which the NBD may be derived and parameterized. This will present the primary set of parameters upon which inference is to be made. The chapter then reviews methods of parameter estimation in the form of maximum likelihood and general moment based estimators.

The δ-method of obtaining the asymptotic normal distribution of various functionals of asymptotically normal statistics is described. By checking the covariance structure of functionals of data to the covariance structure of corresponding functionals of the proposed statistical distribution or process, it is possible to verify whether data could be generated from that distribution or process.

A concise description of two negative binomial processes, namely the gamma Poisson process and the negative binomial first-order autoregressive process, follows. The derivation of the processes are important when studying the statistical properties of estimators. The chapter finishes with some methods that are currently used in literature to fit and assess goodness of fit of negative binomial processes.

Chapters 3 and 4 further develop methods of statistical inference for data generated by negative binomial processes with application to market research data in mind. Chapter 3 investigates the problem of efficiently fitting the NBD using moment based estimators. Chapter 3 works on the basis that maximum likelihood can be difficult to implement in practice. Chapter 4 analyzes the dynamical behavior of negative binomial processes by considering the covariance of statistics computed in different time intervals. Checking the covariance structure of functionals of the data to the covariance structure of functionals of the theoretical model gives a method for testing goodness of fit.

Chapter 5 applies the results of Chapters 3 and 4 to market research data kindly provided by ACNielsen BASES. The data comprises of raw transaction data obtained from the scanning of individual items by a panel of 34,647 households representative of the United States for the duration of the year 2000. Since the data has been collected by the use of scanners, the database contains a comprehensive list of products purchased by each household. This list includes the epochs when a product is purchased and the number of products purchased at each epoch, thus allowing the NBD and negative binomial processes to be fit to the data.

Finally, Chapter 6 draws conclusions on statistical inference for the NBD and negative binomial processes with particular emphasis made on fitting these models to market research data. The standard methodology described in literature of fitting the NBD is compared to the methodology suggested in Chapters 3 and 4. A discussion is then presented on further possible research stemming from the research conducted in this thesis.

Chapter 2

Background

This chapter reviews methods of fitting the negative binomial distribution (NBD) and negative binomial processes. Section 2.1 introduces the NBD and presents ways in which the distribution may be parameterized. The derivations of the NBD that are presented provide indication of the many settings in which the NBD may be used. The natural settings of the NBD often allow natural interpretations for the numerous parameters of the NBD. Various well known methods of estimating negative binomial parameters are presented and the efficiency of these estimation methods are given.

Section 2.2 modifies the approach of the δ-method to derive the asymptotic normal distribution of a general class of moment based estimators, and also of various functionals of data, computed using data from a specified distribution. Testing goodness of fit of the NBD or negative binomial processes can be consequently achieved by verifying covariance structures of functionals of raw data to covariance structures of functionals of the model being fitted.

Section 2.3 introduces two types of negative binomial processes: the gamma Poisson process and the negative binomial first-order autoregressive processes (or simply the NBD INAR(1) process) and finally Section 2.4 reviews methods of fitting these processes to observed data.

2.1 The negative binomial distribution

The NBD is a two parameter distribution that has been used in the modeling of various types of events. For example, the NBD has been used to model: accident proneness and sickness (see e.g. Yule (1910); Lundberg (1964)); the frequency of accidents (see e.g. Greenwood and Yule (1920); Arbous and Kerrich (1951)); animal populations (see e.g. Kendall (1948); Anscombe (1949)), market research (see e.g. Goodhardt, Ehrenberg, and Chatfield (1984); Ehrenberg (1988)) and risk theory (see e.g. Grandell (1997)).

The ability of the NBD to model a diverse range of events arises from the fact that the NBD can be derived, using natural assumptions, in a number of different ways. The various derivations of the NBD leads to numerous ways in which the NBD may be parameterized and these are presented in Section 2.1.1. Methods of estimating parameters of the NBD have also varied according to the field in which the NBD is applied. Natural methods of estimating the NBD parameters include using the standard method of moments, the zero term method and the maximum likelihood method. Common methods of estimating NBD parameters are discussed in Section 2.1.2. Finally, the efficiency of these estimation methods are discussed in Section 2.1.3

2.1.1 Derivations and parameter representations

Inverse binomial sampling. Yule (1910) derived the NBD as a waiting time distribution. He considered a model for the time, more specifically the age in years, at which deaths occur within a population. Suppose that death per individual occurs at the exposure of k fatal accidents and that the event of a fatal accident occurring at discrete time points of fixed length is independent and identically Bernoulli distributed with the probability of a fatal accident given by p. The probability of death occurring at discrete time points x ($x = 0, 1, 2, \ldots$) beyond the k'th time point from time zero is then given by the NBD.

Let X be a random variable from the NBD then the probabilities of the NBD are

$$p_x = \mathbb{P}(X = x) = \frac{\Gamma(k+x)}{x!\Gamma(k)} p^k (1-p)^x, \qquad \begin{array}{l} x = 0, 1, 2, \ldots \\ k = 1, 2, 3, \ldots, \; p > 0. \end{array}$$

The NBD, in the case where k is integer, is known as the Pascal distribution.

Heterogenous Poisson sampling. Greenwood and Yule (1920) later showed, using entirely different arguments to inverse binomial sampling, that the distribution of the number of accidents encountered by individuals may also be modeled by the NBD. Suppose that the number of accidents follow a Poisson distribution with mean λ_j for individual j. Assume that these means λ_j, within the population of individuals, follow the gamma distribution with probability density function given by

$$f(y) = \frac{1}{a^k \Gamma(k)} y^{k-1} \mathrm{e}^{-y/a}, \quad a > 0,\, k > 0, \quad y > 0,$$

then the distribution of the number of accidents registered by different individuals chosen at random follows the NBD with

$$p_x = \int_{0-}^{\infty} \frac{y^x \mathrm{e}^{-y}}{x!} f(y)dy = \frac{\Gamma(k+x)}{x!\Gamma(k)} \left(\frac{1}{1+a}\right)^k \left(\frac{a}{1+a}\right)^x, \qquad \begin{array}{l} x = 0, 1, 2, \ldots \\ k > 0, \; a > 0. \end{array}$$

The NBD parametrization in this setting was also used by Fisher (1941) who thoroughly investigated estimation properties of these parameters using maximum likelihood and moment based estimators and applied the model to the number of ticks found in sheep.

Urn models. Eggenberger and Pólya (1923) considered the probability of choosing white balls, in a sequence of trials, from a single urn containing black and white balls. Suppose that there are initially Np white balls and $N(1-p)$ black balls in an urn containing a total of N balls. Additionally, each time a ball is chosen, assume that the ball is replaced together with $N\nu$ balls of the same color. Then the probability of obtaining x white balls in a sequence of n trials is given by the Pólya-Eggenberger distribution with probabilities

$$p_x = \frac{\binom{n}{x}\prod_{j=0}^{x-1}(p+j\nu)\prod_{j=0}^{n-x-1}(1-p+j\nu)}{\prod_{j=0}^{n-1}(1+j\nu)}, \qquad \begin{array}{l} x=0,1,2,\ldots \\ 0<p<1,\ \nu\geqslant 0,\ n=0,1,2,\ldots, \end{array}$$

where ν is such that $N\nu$ is a non-negative integer. Note that the probabilities do not depend on the total number of balls N in the urn at the first trial. Assume that $\lim_{n\to\infty} np = m$ and that $\lim_{n\to\infty} n\nu = m/k$, then the probability of obtaining x white balls in an infinite number of trials is NBD with

$$p_x = \frac{\Gamma(k+x)}{x!\Gamma(k)}\left(1+\frac{m}{k}\right)^{-k}\left(\frac{m}{m+k}\right)^{x}, \qquad \begin{array}{l} x=0,1,2,\ldots \\ k>0,\ m>0. \end{array}$$

This distribution is sometimes known as the Pólya distribution. If $\nu = 0$ then a sequence of i.i.d. trials is obtained. The probability of obtaining x white balls in a finite number of trials is then binomially distributed with mean np and variance $np(1-p)$. Additionally, if $\lim_{n\to\infty} np = m$ then the probability of obtaining x white balls in an infinite number of trials is Poisson distribution with mean m.

Consumer buying behavior. In the case of market research, where the NBD is used to model the frequency of consumer purchases, the NBD is often parameterized by two alternative, but highly interpretable, 'repeat-buying' measures called the penetration and mean purchase frequency. Let p_x $(x = 0, 1, \ldots)$ denote the probabilities of the NBD and let X be a NBD random variable then the penetration, b, and the purchase frequency, w, are defined by

$$b = 1 - p_0 \qquad \text{and} \qquad w = \mathbb{E}(X|X \geqslant 1) \qquad 0 \leqslant b \leqslant 1,\ w > 1.$$

The NBD probabilities cannot be explicitly presented in terms of the parameters b and w. To obtain the NBD probabilities, the equations above for b and w must first be solved in terms of (m, k) (see Eq. (2.1.1)). Note that for the NBD to be a valid distribution, it must be the case that $w > -\log(1-b)/b$.

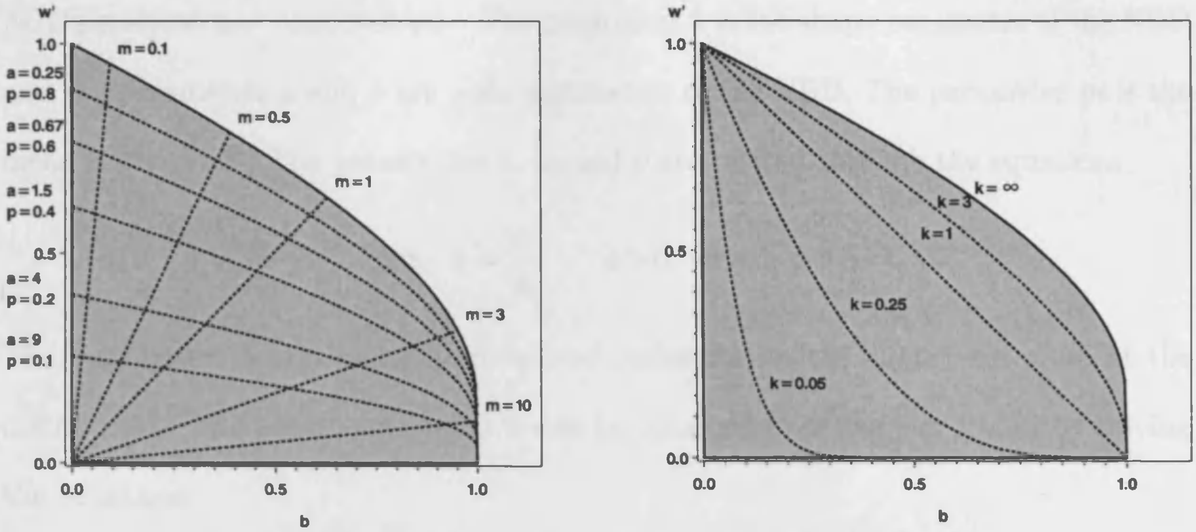


Figure 2.1: a, m, p and k versus (b, w').

A closed NBD parameter space. In this thesis an alternative parametrization denoted by (b, w') with $w' = 1/w$ is considered. Its appeal lies in the fact that the corresponding parameter space is within the unit square $(b, w') \in [0, 1]^2$, which makes it easier to make a visual comparison of different characteristics of NBD parameters for all NBD parameter values. Examples of characteristics include: plotting the efficiency of estimators; plotting the coefficient of variation of estimators or, more generally, plotting the covariances of estimators with respect to other estimators.

Fig. 2.1 shows the contour levels of a, m, p and k within the (b, w')-parameter space. The NBD is only defined for the parameter pairs $(b, w') \in (0, 1) \times (0, 1)$ such that $w' < -b/\log(1-b)$ (shaded region in Fig. 2.1). The relationship $w' = -b/\log(1-b)$ represents the limiting case of the distribution as $k \to \infty$, when the NBD converges to the Poisson distribution with mean m. The NBD is not defined on the axis $w' = 0$ (where $m = \infty$) and is degenerate on the axis $b = 0$ (as $p_0 = 1$).

It is clear from Fig. 2.1 that the parameter pairs (a, k), (m, k), (p, k), (b, w) and (b, w') all have a one-to-one relationship. This simplifies the comparison of the estimators for NBD parameters since only one of the parameter pairs needs to be estimated.

NBD parameter pair relationships. The parameter k is the shape parameter of the NBD and the parameters a and p are scale parameters of the NBD. The parameter m is the mean of the NBD. The parameters a, m and p are related through the equations

$$p = \frac{k}{m+k} \quad \text{and} \quad a = \frac{m}{k} \qquad a>0,\ m>0,\ p>0,\ k>0.$$

The parameters b and w have an indirect influence on the shape and scale of the distribution. The parameters b and w can be obtained from the pair (m, k) by solving the equations

$$b = 1 - \left(1 + \frac{m}{k}\right)^{-k} \qquad \text{and} \qquad w = \frac{m}{b}. \tag{2.1.1}$$

To avoid confusion with the NBD parameterizations in this thesis, the notation described in Table 2.1 will be used. The use of multiple notations will allow simplifications in formulae used later in the thesis. For example, it is much simpler to compare efficiency of estimators of different estimation methods using the parametrization (a, k). The NBD(m, k) notation, where the first parameter m refers to the mean and the second parameter k refers to the shape of the distribution, will be used throughout the thesis. If there is ambiguity in the NBD(m, k) notation, the parameterization NBD$_m(m, k)$ will be used.

Parameterization	Probabilities	Parameter constraints
NBD(m, k)	$\frac{\Gamma(k+x)}{x!\Gamma(k)}\left(1+\frac{m}{k}\right)^{-k}\left(\frac{m}{m+k}\right)^{x}$	$m>0,\ k>0$
NBD$_a(a, k)$	$\frac{\Gamma(k+x)}{x!\Gamma(k)}\left(\frac{1}{1+a}\right)^{k}\left(\frac{a}{1+a}\right)^{x}$	$a \geqslant 0,\ k>0$
NBD$_p(p, k)$	$\frac{\Gamma(k+x)}{x!\Gamma(k)}p^{k}(1-p)^{x}$	$0 \leqslant p \leqslant 1,\ k>0$
NBD$_w(b, w)$	re-parameterize	$0 \leqslant b \leqslant 1,\ w>1$
NBD$_{w'}(b, w')$	re-parameterize	$0 \leqslant b \leqslant 1,\ 0<w'<1$

Table 2.1: Table of NBD probabilities distributions

2.1.2 Parameter estimation for i.i.d. NBD samples

The estimation of NBD parameters given an i.i.d. sample has been considered independently by Fisher (1941) and Haldane (1941) who used maximum likelihood (ML) and by Anscombe (1950) who used general moment based methods. This section reviews estimation methods for the parameter pair (m, k) given an i.i.d. NBD sample of size N with observations $(x_1,\ x_2, \ldots,\ x_N)$. The parameter pair (m, k) is statistically convenient since the maximum likelihood estimator and natural moment based estimators for the pair (m, k) are asymptotically uncorrelated given an i.i.d. NBD sample.

Exponential families. Note that in general the NBD does not fit into the exponential family. If the NBD was in the family of exponential distributions then it would be possible to find complete sufficient statistics as estimators for m and k. For fixed k, however, the NBD does fit into the exponential family and the statistic $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is a complete minimal sufficient statistic for m.

Willson, Folks, and Young (1986) have shown using the result of Lehmann and Scheffé (1950, Theorem 6.3) that if k is unknown, then the set of all order statistics of the sample is minimal sufficient. Willson et al. (1986) have, however, also shown that the set of all order statistics of the sample is not complete, so that the search for a minimum variance unbiased estimator for k is not straightforward. In fact, Wang (1996) has shown that an unbiased estimator for k does not exist.

The log-likelihood function. The log-likelihood function for a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, where each x_i $(i = 1, 2, \ldots, N)$ are i.i.d. NBD(m, k), is

$$\begin{aligned} L_N(\boldsymbol{x}; m, k) &= \log\left(\prod_{i=1}^{N} \frac{\Gamma(k+x_i)}{x_i!\Gamma(k)}\left(1+\frac{m}{k}\right)^{-k}\left(\frac{m}{m+k}\right)^{x_i}\right) \\ &= \sum_{i=1}^{N}\left(\log\Gamma(k+x_i) - \log(x_i!) + x_i\log\left(\frac{m}{m+k}\right)\right) - N\log\left(\Gamma(k)\left(1+\frac{m}{k}\right)^{k}\right). \end{aligned}$$

Maximum likelihood estimators. The ML estimator for m is given by the sample mean

$$\hat{m} = \bar{x} = \frac{1}{N}\sum_{l=1}^{N} x_l,$$

however there is no closed form solution for $\hat{k}_{ML}$, the ML estimator of k. The estimator $\hat{k}_{ML}$ is defined as the solution, in z, to the equation

$$\log\left(1+\frac{\bar{x}}{z}\right) = \sum_{i=1}^{\infty}\frac{n_i}{N}\sum_{j=0}^{i-1}\frac{1}{z+j}, \tag{2.1.2}$$

where N is the sample size and n_i denotes the observed frequency of $i = 0, 1, 2, \ldots$ within the sample. Equation (2.1.2) can be solved using numerical methods.

Note that the maximum likelihood estimator for the parameter k requires knowledge of the frequencies n_i. In market research it is difficult to obtain these frequencies either due to difficulties in collecting data or due to problems such as ownership of raw data. Instead, it is often the case that market research companies are easily able to request and obtain statistics associated with consumer purchases. Moment based estimators are, therefore, an important alternative to estimating the NBD parameters.

Moreover, in market research, analyzing consumer purchase data often requires investigating data over different time periods of varying lengths (see e.g. Ehrenberg (1988)). Fitting the NBD to such data using the ML approach will require calculation of the n_i from raw transaction data for each analysis period. Since it is very uneconomical to store and very difficult to obtain such raw transaction data, ML estimation is hardly ever used in the practice of market research.

In Section 2.4.2 the problem of estimating NBD parameters from a dependent sample is investigated. The dependency in the observations makes it extremely difficult to analytically solve the ML equations in order to obtain ML estimators. Moment based estimators in this situation provide a simple alternative to ML estimators and may be preferred even when all the frequencies are available.

Generalized moment based estimators. Moment based estimation methods were considered by Anscombe (1950), as an alternative to ML estimators, on the basis that the ML estimator for k required the solution of the 'tedious' equation given by Eq. (2.1.2). The increase in computational power today makes the difficulty of solving an equation such as Eq. (2.1.2) obsolete. However, as discussed above, there are situations in which moment based estimators may still be preferred to ML estimators.

The estimation of the parameter pair (m, k) requires the choice of two sample moments. A natural choice for the first moment is the sample mean $\bar{x}$ which is both an efficient and an unbiased estimator for the parameter m. An additional moment is then required to estimate k. Denote this moment by $\bar{f} = \frac{1}{N}\sum_{l=1}^{N} f(x_l)$. The estimator for k is obtained by equating the sample moment $\bar{f}$ to its expected value $\mathbb{E}f(X)$, with m replaced by $\hat{m} = \bar{x}$, and solving the corresponding equation $\bar{f} = \mathbb{E}f(X)$ for k.

Anscombe (1950) considered various statistics $\bar{f}_j = \frac{1}{N}\sum_{l=1}^{N} f_j(x_l)$ for the estimation of k and these functions are shown in Table 2.2. In Table 2.2 the function $I_{[x=0]}$ denotes the indicator function of the event $x = 0$ so that $I_{[x=0]} = 1$ if $x = 0$ and $I_{[x=0]} = 0$ otherwise. Note that $\bar{f}_4 = \frac{1}{N}\sum_{l=1}^{N} c^{x_l}$ depends on an additional parameter c $(c > 0, c \neq 1)$. If $c = 0$, then defining $c^{x_l} = 1$ if $x_l = 0$ and $c^{x_l} = 0$ if $x_l \neq 0$, it is clear that $\bar{f}_4 = \bar{f}_2$ and the two moment based estimation methods become equivalent.

$f(x)$	$\mathbb{E}f(X)$	$\widehat{\mathbb{E}f(X)} = \bar{f}$
$f_1(x) = x^2$	$m(m+1) + \frac{m^2}{k}$	$\overline{x^2} = \frac{1}{N}\sum_{l=1}^{N} x_l^2$
$f_2(x) = I_{[x=0]}$	$\left(1 + \frac{m}{k}\right)^{-k}$	$\hat{p}_0 = \frac{n_0}{N}$
$f_3(x) = \frac{1}{x+1}$	$\frac{k}{m(k-1)}\left[1 - \left(\frac{k}{m+k}\right)^{k-1}\right]$	$\widehat{x_{(-1)}} = \frac{1}{N}\sum_{l=1}^{N} \frac{1}{1+x_l}$
$f_4(x) = c^x \quad (c > 0, c \neq 1)$	$\left(1 + \frac{m(1-c)}{k}\right)^{-k}$	$\widehat{c^x} = \frac{1}{N}\sum_{l=1}^{N} c_l^x$
x	m	$\bar{x} = \frac{1}{N}\sum_{l=1}^{N} x_l$

Table 2.2: Moments and moment estimators for the NBD

Method name		$f_j(x)$	$\hat{k}$	Estimator or equation for $\hat{k}$
Method of moments	(MOM)	$f_1(x)$	$\hat{k}_{MOM}$	$\frac{\bar{x}^2}{\overline{x^2}-\bar{x}^2-\bar{x}}$
Zero term method	(ZTM)	$f_2(x)$	$\hat{k}_{ZTM}$	$\widehat{p_0} = \left(1+\frac{\bar{x}}{z}\right)^{-z}$
Factorial method	(FM)	$f_3(x)$	$\hat{k}_{FM}$	$\widehat{x_{(-1)}} = \frac{z}{\bar{x}(z-1)}\left[1-\left(\frac{z}{\bar{x}+z}\right)^{z-1}\right]$
Power method	(PM)	$f_4(x)$	$\hat{k}_{PM(c)}$	$\widehat{c^X} = \left(1+\frac{\bar{x}(1-c)}{z}\right)^{-z}$

Table 2.3: Moment based estimators for the NBD parameter k

The estimator for m is always $\bar{x}$ irrespective of the additional function $f_j(x)$ chosen for estimating the parameter pair (m, k). Anscombe (1950) proved that if $f(x)$ is any integrable convex or concave function on the non-negative integers then $\mathbb{E}f(X)$ with m substituted by $\bar{x}$ is a monotone function in k. Estimating the parameter k by solving the equation $\bar{f} = \mathbb{E}f(X)$ will therefore have at most one solution.

Table 2.3 shows the moment based estimators for k, denoted by $\hat{k}$, for the different functions $f_j(x)$ presented in Table 2.2. Although an explicit formula exists for $\hat{k}_{MOM}$, no analytical solution exists for $\hat{k}_{ZTM}$, $\hat{k}_{FM}$ or $\hat{k}_{PM(c)}$. Since there is at most one solution for $\hat{k}_{ZTM}$, $\hat{k}_{FM}$ and $\hat{k}_{PM(c)}$, these estimators may be obtained by using numerical algorithms to solve the corresponding equations given in Table 2.3 for z. Note that the PM estimator for k is equal to the ZTM estimator if the additional PM parameter $c = 0$ and tends to the MOM estimator as $c \to 1$.

For each estimation method in Table 2.3 and the ML method there is, for any $m > 0$ and $k > 0$, a positive but small probability that the estimator for k will be negative even though the sample may be NBD. For the MOM it is clear that $\hat{k}_{MOM}$ is negative when $\overline{x^2} - \bar{x}^2 < \bar{x}$. For the PM, the estimator $\hat{k}_{PM(c)}$ is negative when $\widehat{c^X} < \exp(-\bar{x}(1-c))$. In literature (see e.g. Anscombe (1950); Ehrenberg (1988)) it is common to set $\hat{k} = \infty$ whenever a negative estimate for k is obtained; however the setting of $\hat{k} = \infty$ is not fully justified in the literature. This topic is investigated further in Section 3.3.2.

2.1.3 Efficiency of estimators

The variances of the ML estimators are the minimum possible asymptotic (as $N \to \infty$) variances attainable in the class of all asymptotically normal estimators and therefore provide a lower bound for the asymptotic variance of moment based estimators. Fisher (1941) and Haldane (1941) independently derived an expression for the asymptotic covariance matrix for the ML estimators by taking the inverse of the Fisher information matrix. The asymptotic variance of the moment based estimators and asymptotic covariances between moment based estimators for (m, k) were derived by Anscombe (1950) using the so-called δ-method (see e.g. Serfling (1980), Chapter 3).

Maximum likelihood estimators. The asymptotic normalized variances of $\hat{m}$ and $\hat{k}_{ML}$ are

$$\begin{aligned} \lim_{N\to\infty} N \operatorname{Var}(\hat{m}) &= ka(1+a), \qquad (2.1.3) \\ v_{ML} = \lim_{N\to\infty} N \operatorname{Var}\left(\hat{k}_{ML}\right) &= \frac{2k(k+1)(a+1)^2}{a^2\left(1+2\sum_{j=2}^{\infty}\left(\frac{a}{a+1}\right)^{j-1}\frac{j!\Gamma(k+2)}{(j+1)\Gamma(k+j+1)}\right)}, \end{aligned}$$

where $a = m/k$. Using the inverse of the Fisher information matrix, the asymptotic normalized covariance between the estimators is $\lim_{N\to\infty} N \operatorname{Cov}(\hat{m}, \hat{k}_{ML}) = 0$ and hence the ML estimators are asymptotically uncorrelated.

Generalized moment based estimators. The asymptotic normalized variance of $\hat{m} = \bar{x}$ is given by Eq. (2.1.3). The asymptotic normalized variance for general moment based estimators of k for a given function $f(\cdot)$ is

$$\lim_{N\to\infty} N \operatorname{Var}\left(\hat{k}\right) = \frac{\mathbb{E}f^2(X) - [\mathbb{E}f(X)]^2 - \left(m+\frac{m^2}{k}\right)\left[\frac{\partial}{\partial m}\mathbb{E}f(X)\right]^2}{\left[\frac{\partial}{\partial k}\mathbb{E}f(X)\right]^2}.$$

Using the δ-method the asymptotic normalized covariance between moment based estimators $\hat{m}$ and $\hat{k}$ is $\lim_{N\to\infty} N \operatorname{Cov}(\bar{x}, \hat{k}) = 0$.

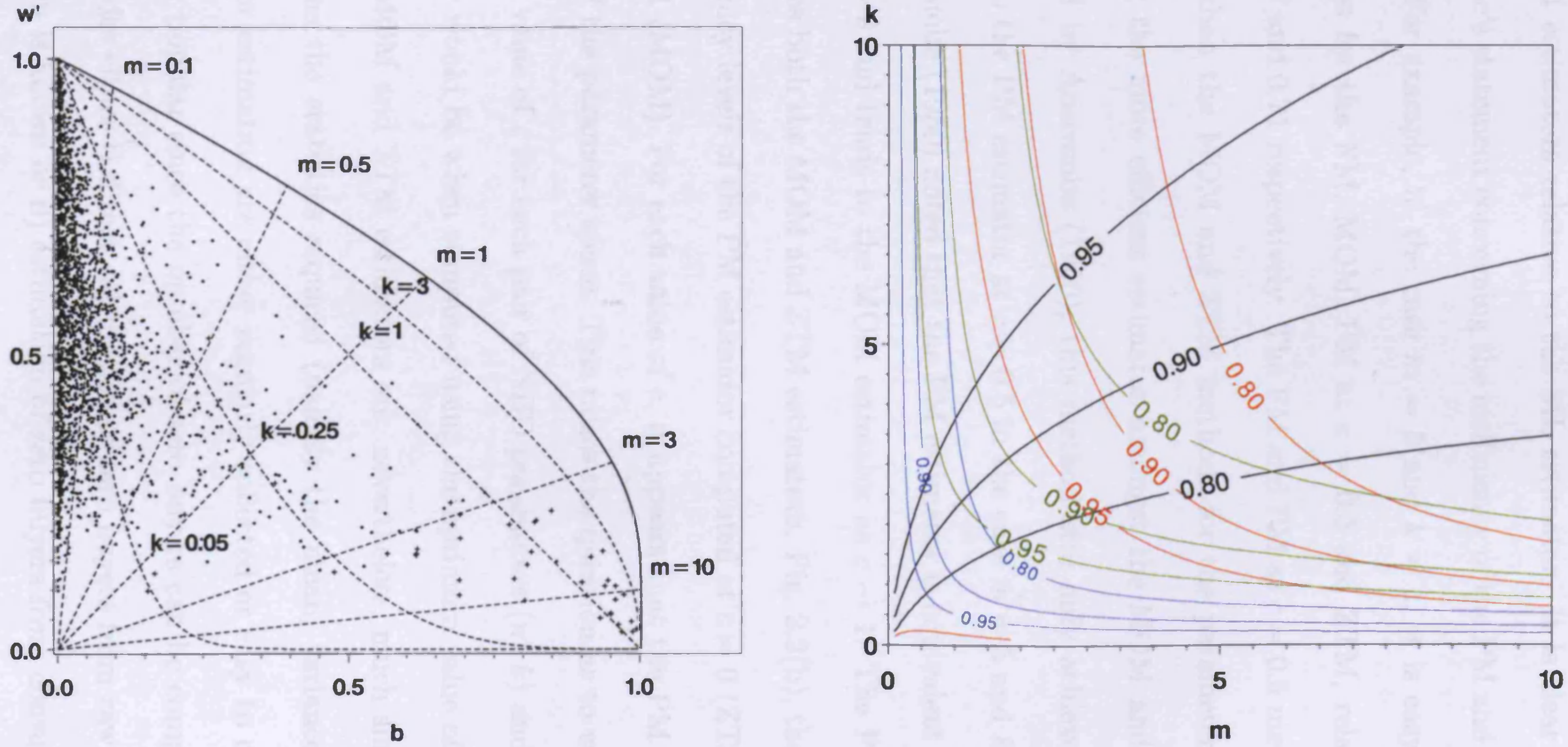
The asymptotic normalized variances of $\hat{k}_{MOM}$, $\hat{k}_{ZTM}$ and $\hat{k}_{PM}$ are

$$\begin{aligned} v_{MOM} &= \lim_{N\to\infty} N\text{Var}\left(\hat{k}_{MOM}\right) = \frac{2k(k+1)(a+1)^2}{a^2}, \\ v_{ZTM} &= \lim_{N\to\infty} N\text{Var}\left(\hat{k}_{ZTM}\right) = \frac{(a+1)^{k+2}-(a+1)^2-ka(a+1)}{[(a+1)\log(a+1)-a]^2}, \\ v_{PM}(c) &= \lim_{N\to\infty} N\text{Var}\left(\hat{k}_{PM(c)}\right) = \frac{(1+a-ac^2)^{-k}r^{2k+2}-r^2-ka(a+1)(1-c)^2}{[r\log(r)-r+1]^2}, \end{aligned}$$

where $r = 1 + a - ac$. The asymptotic normalized variance of $\hat{k}_{FM}$ is difficult to express explicitly and for an expression of the variance see Anscombe (1950, p. 369). Since, amongst the class of moment based estimators considered, the estimator for m is the same and the asymptotic covariance between the estimators of k and m is zero, the most efficient estimation method is determined by the method that minimizes the variance of $\hat{k}$.

The efficiency of estimating k, relative to ML, using the MOM and ZTM was plotted by Anscombe (1950) over the parameter space $0.04 \leqslant m \leqslant 400$ and $0.1 \leqslant k \leqslant 100$. A comparison of the efficiencies of the MOM, ZTM, PM and FM estimators was made, although no contours of the efficiency of the PM and FM estimators were plotted. Anscombe (1950) noted that the PM and FM estimators are nowhere uniformly more efficient than the more efficient of the MOM and ZTM estimators.

Fig. 2.2(a) shows ZTM estimates for NBD parameters when fitting the NBD to the number of purchases made by households for 46 different categories and the top 50 brands within each category. The estimator $\hat{k}_{ZTM} < 3$ for all the products considered. For large values of k the Poisson distribution serves as a very good approximation to the NBD. Since this thesis is primarily concerned with market research data, and the Poisson distribution serves as a good approximation for the NBD for large values of k, this thesis will be primarily concerned with estimation of NBD parameters in areas of the parameter space which is of practical importance in market research.



(a) (b)

Figure 2.2: (a) ZTM estimators for NBD parameters when fitting the NBD to 46 categories and the 50 top brands within each category for consumer purchases. Data courtesy of ACNielsen BASES. (b) Contour levels of the efficiency of FM (v_{ML}/v_{FM}), MOM (v_{ML}/v_{MOM}), PM at $c = 0.5$ ($v_{ML}/v_{PM}(0.5)$) and ZTM (v_{ML}/v_{ZTM}) estimators relative to ML.

Fig. 2.2(b) shows contour levels of the efficiency for the FM , MOM, PM at $c = 0.5$ and ZTM estimators relative to the ML estimator. It is clear from Fig. 2.2(b) that Anscombe's statement concerning the inefficiency of the PM and FM methods is clearly untrue. For example, in the case $m = 5$ and $k = 1$, it is easy to compute that the efficiencies for the FM, MOM, PM at $c = 0.5$ and ZTM, relative to ML, are 0.96, 0.56, 0.97 and 0.71 respectively. The FM and PM at $c = 0.5$ methods are clearly more efficient than the MOM and ZTM methods for the parameters $m = 5$ and $k = 1$. Choosing the more efficient estimator amongst the MOM and ZTM estimators was suggested by Anscombe (1950); this method still only achieves 73% efficiency with respect to the PM estimator at $c = 0.5$ in the case $m = 5$ and $k = 1$.

Anscombe (1950) noted that the PM estimator is equivalent to the ZTM estimator when $c = 0$ and tends to the MOM estimator as $c \to 1$. The PM estimator therefore generalizes both the MOM and ZTM estimators. Fig. 2.2(b), therefore, in effect shows the efficiency levels of the PM estimator computed at $c = 0$ (ZTM), $c = 0.5$ (PM(0.5)) and $c = 1$ (MOM). For each value of c, it appears that the PM is efficient in different regions of the parameter space. This raises the question as to whether there exists an optimum value of c for each pair of NBD parameters (m, k) and how efficient the PM estimator would be when computed using the optimum value of c.

The MOM and ZTM estimators are, nevertheless, much simpler to implement in practice as the statistics required (namely the mean, variance and number of zero events) for estimation are either regularly collected or easy to compute. The ZTM is especially popular since the number of zero buyers can be computed in various ways, this includes either i) direct calculation of zero buyers from raw data if the size of the population is known or ii) estimation of zero buyers from consumer surveys or by the use of supermarket retail data.

2.2 Asymptotic properties of a general estimator

Section 2.1 considered the asymptotic distribution of maximum likelihood estimators and a class of moment based estimators called power method estimators. The power method estimators include the case of method of moments and zero term method estimators. This section considers the asymptotic distribution of a general class of estimators for a vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^T$ where the estimators satisfy the equation $G_i(\boldsymbol{\theta}, \overline{f_i}) = 0$ $(i = 1, \ldots, d)$, using d statistics $\overline{f_i}$, with $G_i(\boldsymbol{\theta}, \overline{f_i}) = \mathbb{E}f_i(\zeta; \boldsymbol{\theta}) - \overline{f_i}$. The covariance matrix of the limiting normal distribution of the estimators is derived. The construction of the limiting normal distribution of the estimator of $\boldsymbol{\theta}$ satisfying the general equation $G_i(\boldsymbol{\theta}, \overline{f_i}) = 0$ with $G_i(\boldsymbol{\theta}, \overline{f_i}) = \mathbb{E}f_i(\zeta; \boldsymbol{\theta}) - \overline{f_i}$ is useful in that the limiting distribution for estimators of any combination of parameters can be derived.

In the case of the NBD, for example, the joint distribution of the vector of parameters $(\hat{m}, \hat{k}_{MOM}, \hat{k}_{PM}(c), \hat{k}_{ZTM})^T$ may be derived. When considering the estimation of parameters from a process, the general scheme of estimation allows the joint distribution of estimators computed in different time intervals to be derived. Take, for example, a negative binomial process where the distribution of events over different time intervals is NBD. Using the general methodology discussed in this section, it is possible to derive the limiting normal distribution of the vector of parameters $(\hat{k}^{(1)}_{PM}(c), \hat{k}^{(2)}_{PM}(c), \ldots, \hat{k}^{(t)}_{PM}(c))^T$, where $\hat{k}^{(i)}_{PM}(c)$ is the PM(c) estimator for k computed in the i'th time interval.

The results of this section are a particular case of the results on M- and Z-estimators as noted in van der Vaart (1998, Chapters 3-5). This section considers the possibility of applying these results in the case where the distribution is discrete and in particular negative binomially distributed.

2.2.1 General method of estimation

Let ζ be a random variable taking values in some set $\mathcal{Z}$ and let ζ have probability mass function $p(z;\boldsymbol{\theta})$, $z \in \mathcal{Z}$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^T$ $(d \geq 1)$ is a vector of parameters taking values in some set $\Theta \subseteq \mathbb{R}^d$ with non-empty interior $\mathrm{int}(\Theta)$. Define the vector $\boldsymbol{f} = (f_1, \ldots, f_d)^T \in \mathbb{R}^d$ such that $f_i : \mathcal{Z} \times \Theta \to \mathbb{R}$ $(i = 1, \ldots, d)$ are some functions which are smooth enough and possibly depend on $\boldsymbol{\theta}$. Let $\{z_1, \ldots, z_N\}$ be a sample of values of ζ and set $\bar{\boldsymbol{f}} = (\bar{f}_1, \ldots, \bar{f}_d)^T \in \mathbb{R}^d$ with $\overline{f_i} = \frac{1}{N}\sum_{l=1}^{N} f_i(z_l;\boldsymbol{\theta})$. Finally, assume that $\mathbb{E}\overline{f_i} = \mathbb{E}f_i(\zeta;\boldsymbol{\theta})$; which is indeed true in the case where the sample is i.i.d. and also true in the case where $\{z_1, \ldots, z_N\}$ are observed values indexed by time obtained from an ergodic time series. A general estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_d)^T$ for $\boldsymbol{\theta}^* = (\theta_1^*, \ldots, \theta_d^*)^T \in \mathrm{int}(\Theta)$, the true parameter values of the sampling distribution, is then defined to be the solution to the equations

$$G_i(\boldsymbol{\theta}, \overline{f_i}) = 0 \qquad i = 1, \ldots, d\,, \tag{2.2.1}$$

where $G_i(\boldsymbol{\theta}, \overline{f_i}) = \mathbb{E}f_i(\zeta;\boldsymbol{\theta}) - \overline{f_i}$.

Example 2.2.1. $f_i(z;\boldsymbol{\theta}) = \partial \log(p(z;\boldsymbol{\theta}))/\partial\theta_i$ implying $\mathbb{E}f_i(\zeta;\boldsymbol{\theta}) = 0$,

Example 2.2.2. $f_i(z;\boldsymbol{\theta}) = f_i(z)$ so that the functions f_i do not depend on $\boldsymbol{\theta}$,

Example 2.2.3. $f_i(z;\boldsymbol{\theta}) = z^i$ implying $\mathbb{E}f_i(\zeta;\boldsymbol{\theta}) = \mathbb{E}\zeta^i$,

Note that the system of Eqs. (2.2.1) may be represented in vector form as

$$\boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}}) = (G_1(\boldsymbol{\theta}, \overline{f_1}), \ldots, G_d(\boldsymbol{\theta}, \overline{f_d}))^T = \mathbf{0}. \tag{2.2.2}$$

For each i, the $G_i(\boldsymbol{\theta}, \overline{f_i})$ may be represented as $G_i(\boldsymbol{\theta}, \overline{f_i}) = \frac{1}{N}\sum_{l=1}^{N} g_i(z_l, \boldsymbol{\theta}) = \bar{g}_i$, where $g_i(z, \boldsymbol{\theta}) = \mathbb{E}f_i(\zeta;\boldsymbol{\theta}) - f_i(z;\boldsymbol{\theta})$.

2.2.2 Asymptotic normality of estimators

The following theorem summarizes the multivariate version of the so-called δ-method (see e.g. Serfling (1980, Chapter 3)) and the implicit function theorem (see e.g. Schwartz (1967)). The results can also be found in van der Vaart (1998, Chapters 3-5).

Theorem 2.2.1. *Assume that the function $\boldsymbol{G}$ is invertible as a function of $\boldsymbol{\theta}$ in some neighbourhood of $(\boldsymbol{\theta}^*, \mathbb{E}\boldsymbol{f})$ and let $\hat{\boldsymbol{\theta}}$ be the solution of $\boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}}) = \mathbf{0}$. Assume that $\mathbb{E}\left|\partial g_i(\zeta, \boldsymbol{\theta})/\partial\theta_j\right| < \infty$ for all i, j. Additionally, assume that the estimator $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$ and $\sqrt{N}(\bar{\boldsymbol{f}} - \mathbb{E}\boldsymbol{f})$ is asymptotically normally distributed $\mathcal{N}(0, \mathbb{D}\boldsymbol{f})$, where $\mathbb{D}\boldsymbol{f} = \mathbb{E}(\boldsymbol{f} - \mathbb{E}\boldsymbol{f})(\boldsymbol{f} - \mathbb{E}\boldsymbol{f})^T = \|\mathrm{Cov}(f_i(\zeta; \boldsymbol{\theta}), f_j(\zeta; \boldsymbol{\theta}))\|_{i,j=1}^d$. Then as $N \to \infty$,*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \boldsymbol{V}(\mathbb{D}\boldsymbol{f})\boldsymbol{V}^T) \tag{2.2.3}$$

where $\overset{\mathcal{D}}{\Rightarrow}$ is convergent in distribution and

$$\boldsymbol{V} = \left[\lim_{N\to\infty} \left.\frac{\partial \boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]^{-1}. \tag{2.2.4}$$

Proof. According to the weak law of large numbers as $N \to \infty$, $\bar{\boldsymbol{f}} \to \mathbb{E}\boldsymbol{f}$ in probability and for any $\boldsymbol{\theta}$ there exists the weak limit

$$\lim_{N\to\infty} \left\|\frac{\partial \boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})}{\partial \boldsymbol{\theta}}\right\| = \lim_{N\to\infty} \left\|\frac{1}{N}\sum_{l=1}^{N} \frac{\partial g_i(z_l, \boldsymbol{\theta})}{\partial \theta_j}\right\|$$

which is a non-random matrix.

Since $\boldsymbol{G}$ is invertible as a function of $\boldsymbol{\theta}$ in the neighbourhood of $(\boldsymbol{\theta}^*, \mathbb{E}\boldsymbol{f})$, for N large enough the inverse $\left(\frac{\partial \boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})}{\partial \boldsymbol{\theta}}\right)^{-1}$ exists in the neighbourhood of $\boldsymbol{\theta}^*$. Using the first order Taylor expansion Eq. (2.2.2) is approximated by

$$\boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}}) \simeq \boldsymbol{G}(\boldsymbol{\theta}^*, \bar{\boldsymbol{f}}) + \left.\frac{\partial \boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\boldsymbol{\theta} - \boldsymbol{\theta}^*) = 0. \tag{2.2.5}$$

According to the well known δ-method (see e.g. Serfling (1980)) the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is the same as the asymptotic distribution of $\tilde{\boldsymbol{\theta}}$, which is the solution to

Eq. (2.2.5). Solving Eq. (2.2.5) we obtain

$$\boldsymbol{\theta}^* - \tilde{\boldsymbol{\theta}} = \left(\frac{\partial \boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})}{\partial \boldsymbol{\theta}}\right)^{-1}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \boldsymbol{G}(\boldsymbol{\theta}^*, \bar{\boldsymbol{f}}).$$

The asymptotic distribution of $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ can be related to the asymptotic distribution of $\sqrt{N}\boldsymbol{G}(\boldsymbol{\theta}^*, \bar{\boldsymbol{f}})$ using Slutsky's theorem which allows the replacement of

$$\left(\frac{\partial \boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})}{\partial \boldsymbol{\theta}}\right)^{-1}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \quad \text{with} \quad \boldsymbol{V} = \left[\lim_{N\to\infty} \frac{\partial \boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})}{\partial \boldsymbol{\theta}}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]^{-1}$$

and we obtain that the asymptotic distributions of $\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)$ and $\boldsymbol{V}\left(\sqrt{N}\boldsymbol{G}(\boldsymbol{\theta}^*, \bar{\boldsymbol{f}})\right)$ coincide. Note that $\sqrt{N}(\bar{\boldsymbol{f}} - \mathbb{E}\boldsymbol{f}) = \sqrt{N}\boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})$ and therefore $\sqrt{N}\boldsymbol{G}(\boldsymbol{\theta}, \bar{\boldsymbol{f}})$ is asymptotically normally distributed $\mathcal{N}(0, \mathbb{D}\boldsymbol{f})$. This implies that $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ is asymptotically normally distributed $\mathcal{N}(0, \boldsymbol{V}(\mathbb{D}\boldsymbol{f})\boldsymbol{V}^T)$. □

2.2.3 Examples of estimation methods

Example 2.2.4. Maximum likelihood. The functions f_i are of the form $f_i(z;\boldsymbol{\theta}) = \partial \log(p(z;\boldsymbol{\theta}))/\partial\theta_i$ $(i = 1, \ldots, d)$ so that $\mathbb{E}\overline{f_i} = \mathbb{E}f_i(\zeta;\boldsymbol{\theta}) = 0$ and

$$\mathbb{D}\boldsymbol{f} = \left\|\mathbb{E}\frac{\partial}{\partial\theta_i}\log p(\zeta;\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\log p(\zeta;\boldsymbol{\theta})\right\| = I(\boldsymbol{\theta}),$$

$$V^{-1} = -\lim_{N\to\infty}\left\|\frac{\partial}{\partial\theta_j}\frac{1}{N}\sum_{l=1}^{N}\frac{\partial}{\partial\theta_i}\log p(z_l;\boldsymbol{\theta})\right\| = -\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{l=1}^{N}\frac{\partial^2}{\partial\theta_j\partial\theta_i}\log p(z_l;\boldsymbol{\theta})\right\| = I(\boldsymbol{\theta}),$$

where $I(\boldsymbol{\theta})$ is the Fisher information matrix. The covariance matrix of the maximum likelihood estimators is therefore $\mathbb{D}\hat{\theta} = I(\boldsymbol{\theta})^{-1}I(\boldsymbol{\theta})I(\boldsymbol{\theta})^{-1} = I(\boldsymbol{\theta})^{-1}$.

Example 2.2.5. General method of moments. The functions f_i are of the form $f_i(z,\boldsymbol{\theta}) = f_i(z)$ $(i = 1, \ldots, d)$ so that the functions f_i do not depend on the unknown parameters $\boldsymbol{\theta}$. This implies

$$\mathbb{D}\boldsymbol{f} = \|\mathrm{Cov}(f_i(\zeta), f_j(\zeta))\| \quad \text{and} \quad V^{-1} = \left\|\frac{\partial\mathbb{E}\boldsymbol{f}(\zeta)}{\partial\boldsymbol{\theta}}\right\|.$$

Example 2.2.6. Standard method of moments: $f_i(z,\boldsymbol{\theta}) = z^i$ $(i = 1, \ldots, d)$ implying $\mathbb{D}\boldsymbol{f} = ||\mathbb{E}\zeta^{i+j} - \mathbb{E}\zeta^i\mathbb{E}\zeta^j||$.

2.3 Negative binomial processes

The aim of this section is to provide a concise description of two negative binomial processes, namely the gamma Poisson process and the negative binomial first-order autoregressive process. The derivation of these processes are important when studying the statistical properties of estimators computed from data generated by these processes. Practical examples where these processes have been used in literature are also presented.

This section begins with some definitions important in the studying of stochastic processes. Section 2.3.2 then defines the gamma Poisson process while Section 2.3.3 defines negative binomial first-order autoregressive processes or in short the NBD INAR(1) processes.

2.3.1 Definitions and notation

A stochastic process $\{X(t) : t \in \mathcal{T}\}$ is a set of random variables indexed by time t. In this thesis, only stochastic processes where $X(t)$ takes values on the non-negative integers and the set $\mathcal{T} = [0, \infty)$ or $\mathcal{T} = \{0, 1, 2, \ldots\}$ will be considered. In the case $\mathcal{T} = [0, \infty)$, $\{X(t) : t \in \mathcal{T}\}$ is a continuous time stochastic process and in the case $\mathcal{T} = \{0, 1, 2, \ldots\}$, $\{X(t) : t \in \mathcal{T}\}$ is a discrete time stochastic process.

Let $\boldsymbol{X} = (X(t_1), \ldots, X(t_n))$ be a vector of time indexed random variables and let $\boldsymbol{x} = (x_1, \ldots, x_n)$ be a vector of non-negative integers then the distributional properties of a non-negative integer-valued stochastic process $\{X(t) : t \in \mathcal{T}\}$, with $\mathcal{T} = \{t_1, \ldots, t_n\}$, are defined by its finite-dimensional distributions (f.d.d.'s)

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \mathbb{P}\left(X(t_1) = x_1, \ldots,\ X(t_n) = x_n\right) \qquad n = 1, 2, \ldots .$$

Homogeneity and stationarity

A process is called homogenous in space if the f.d.d.'s are invariant under shifts in the state space. A process is called homogenous in time or strictly stationary if the f.d.d.'s are invariant under shifts in time i.e. if the process $\{X(t) : t \in \mathcal{T}\}$, with $\mathcal{T} = \{t_1, \ldots, t_n\}$ satisfies

$$\mathbb{P}(X(t_1)=x_1, \ldots, X(t_n)=x_n) = \mathbb{P}(X(t_1+h)=x_1, \ldots, X(t_n+h)=x_n)$$

for all $n=1,2,3,\ldots$ and $h>0$. A process is called weakly stationary if $\mathbb{E}X(t_1) = \mathbb{E}X(t_2)$ and $\mathrm{Cov}(X(t_1), X(t_2)) = \mathrm{Cov}(X(t_1+h), X(t_2+h))$ for all t_1, t_2 and $h > 0$. A process homogenous in both space and time is simply known as a homogenous process.

Types of processes

Renewal processes. Let $T_0 = 0$, $T_n = W_1 + W_2 + \ldots + W_n$ $(n \geqslant 1)$ and let W_i $(i = 1, \ldots, n)$ be i.i.d. non-negative random variables, then a renewal process $Z = \{Z(t) : t \in \mathcal{T}\}$ is the process defined by $Z(t) = \max\{n : T_n \leqslant t\}$.

Autoregressive processes. A process $X(t)$ is said to be an autoregressive process of the order r if the process satisfies

$$X(t) = \sum_{i=1}^{r} \alpha_i X(t-i) + \varepsilon_t$$

where $\alpha_i, i = 1, \ldots, r$ are constants and ε_t forms a sequence of uncorrelated random variables.

Markov processes. A process is called a Markov process if, given $t_i \leqslant t_{i+1}$ $(i = 1, \ldots, n)$,

$$\mathbb{P}(X(t_n) = x_n | X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}) = \mathbb{P}(X(t_n) = x_n | X(t_{n-1}) = x_{n-1}).$$

2.3.2 The gamma Poisson process

The gamma Poisson process is a count process that falls into the class of immigration, birth and death processes or mixed Poisson processes. The class of mixed Poisson processes has been thoroughly studied by Lundberg (1964) and Grandell (1997). Some important results on mixed Poisson processes will now be presented.

Mixed Poisson processes

Let $\boldsymbol{Z} = (Z(t_1), Z(t_2), \ldots, Z(t_n))$ be a random vector with $0 = t_0 \leqslant t_1 \leqslant \ldots \leqslant t_n$ representing an increasing sequence of time points, let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be a vector of non-negative integers with $0 = x_0 \leqslant x_1 \leqslant \ldots x_n$ and let $\lambda > 0$ be the intensity of a process, then given the multivariate Poisson distribution

$$\mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{x} | \Lambda = \lambda\right) = \prod_{i=0}^{n-1} \frac{[\lambda(t_{i+1} - t_i)]^{x_{i+1} - x_i}}{(x_{i+1} - x_i)!} \exp\left(-\lambda(t_{i+1} - t_i)\right), \tag{2.3.1}$$

the mixed Poisson process is consequently defined as a process $\{Z(t) : t \in \{t_1, t_2, \ldots, t_n\}\}$ whose f.d.d.' s are

$$\mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{x}\right) = \int_{0-}^{\infty} \mathbb{P}\left(\boldsymbol{Z} = \boldsymbol{x} | \Lambda = \lambda\right) dU_{\Lambda}(\lambda; \boldsymbol{\theta}). \tag{2.3.2}$$

Here $U_{\Lambda}(\lambda; \boldsymbol{\theta})$, commonly known as the structure distribution, is the distribution function for the random variable Λ with support $(0, \infty)$ and $\boldsymbol{\theta}$ is a vector of unknown parameters. Grandell (1997, p. 27) noted that any distribution for Λ, with support on the interval $(0, \infty)$, that is infinitely divisible may be used for $U_{\Lambda}(\lambda; \boldsymbol{\theta})$. [A random variable Λ is said to be infinitely divisible (see e.g. Feller (1966, p. 176)) if and only if Λ can be represented as the sum of n independent random variables with identical distribution function U_n for every n.] Note that the mixed Poisson process conditioned upon $\Lambda = \lambda$, so that the value of λ is fixed, is simply a pure Poisson process with stationary and independent increments whose f.d.d.'s are given by Eq. (2.3.1).

The gamma Poisson process. The gamma Poisson process is a mixed Poisson process whose structure distribution $U_\Lambda(\lambda;\boldsymbol{\theta})$ is the gamma distribution with probability density function

$$g(\lambda;a,k)=\frac{1}{a^k\Gamma(k)}\lambda^{k-1}\mathrm{e}^{-\lambda/a},\quad a>0,\, k>0,\quad \lambda>0.$$

The f.d.d. of the gamma Poisson process is the multivariate NBD with probabilities

$$\begin{aligned}\mathbb{P}(\boldsymbol{Z}=\boldsymbol{x})&=\int_{0-}^{\infty}\left(\prod_{i=0}^{n-1}\frac{[\lambda(t_{i+1}-t_i)]^{x_{i+1}-x_i}}{(x_{i+1}-x_i)!}\exp\left(-\lambda(t_{i+1}-t_i)\right)\right)g(\lambda;a,k)\,d\lambda\\&=\frac{\Gamma(k+x_n)}{\Gamma(k)}\left(\prod_{i=0}^{n-1}\frac{(t_{i+1}-t_i)^{x_{i+1}-x_i}}{(x_{i+1}-x_i)!}\right)\frac{a^{x_n}}{(1+at_n)^{x_n+k}}.\end{aligned}\tag{2.3.3}$$

The one dimensional distribution of the process is the NBD with probabilities

$$\mathbb{P}(Z(t_1)=x)=\frac{\Gamma(k+x)}{\Gamma(k)x!}\left(\frac{1}{1+at_1}\right)^k\left(\frac{at_1}{1+at_1}\right)^x\qquad x=0,1,2,\ldots.\tag{2.3.4}$$

Note that the gamma Poisson process is homogenous neither in time nor space since the f.d.d. of the process may not be represented as a function of $t_{j+1}-t_j$ $(j=0,\ldots,n-1)$ nor $x_{j+1}-x_j$ $(j=0,\ldots,n-1)$ respectively.

Birth and immigration processes

For any $t\geqslant 0$ and $h>0$, let x_t be a non-negative integer with $x_{t+h}\geqslant x_t$ and let $\lambda_{x_t}(t)\geqslant 0$ for any x_t, then an immigration, birth and death process $\{Z(t):t\in[0,\infty)\}$ is a process such that

1) $Z(0)=0$ and $Z(t)\leqslant Z(t+h)$;

2) $\mathbb{P}(Z(t+h)-Z(t)=x_{t+h}-x_t|Z(s)=x_s,Z(t)=x_t)$
$$=\mathbb{P}(Z(t+h)-Z(t)=x_{t+h}-x_t|Z(t)=x_t)\quad\text{for } 0<s<t;$$

3) $$\mathbb{P}(Z(t+h)-Z(t)=x_{t+h}-x_t|Z(t)=x_t)=\begin{cases}1-\lambda_{x_t}(t)h+o(h), & \text{if } x_{t+h}-x_t=0\\ \lambda_{x_t}(t)h+o(h), & \text{if } x_{t+h}-x_t=1\,.\\ o(h), & \text{if } x_{t+h}-x_t>1\end{cases}$$

It is clear from these properties that the birth and immigration process is also a Markov process. Property 3) is called the transition probability and defines the distributional behavior of the count process $Z(t)$.

The gamma Poisson process. The gamma Poisson process may be characterized within the class of birth and death processes (see e.g. Grandell (1997, p. 62)) as having the intensity

$$\lambda_{x_t}(t) = \mathbb{E}(\Lambda_{Z(t)}(t)|Z(s) = x_t) = \frac{\int_{0-}^{\infty} \lambda^{x_t+1} e^{-\lambda t} f(\lambda) d\lambda}{\int_{0-}^{\infty} \lambda^{x_t} e^{-\lambda t} f(\lambda) d\lambda} = \frac{a(k + x_t)}{1 + at}.$$

The mixed Poisson process for consumer buying behavior.

The analysis of modeling consumer buying behavior using the gamma Poisson process was originally considered by Ehrenberg (1988). Consumer purchase occasions represent the rate of recurrence with which households purchase products. Let $\{z_l(t_1), \ldots, z_l(t_n)\}$ represent the number of purchase occasions for household l up to times $\{t_1, t_2, \ldots, t_n\}$ and let $z_l(0) = 0$ (i.e. there are zero purchases at time zero for household l). Assume that the purchasing process of a household follows a Poisson process with mean λ_l over a unit time interval. The distribution of purchases for a fixed household is then given by Eq. (2.3.1).

If the intensity λ_l varies between individuals so that λ_l has the distribution function $U_\Lambda(\lambda; \boldsymbol{\theta})$ then, for fixed time points $\{t_1, \ldots, t_n\}$, the number of purchase occasions $\{z_l(t_1), \ldots, z_l(t_n)\}$ for a random household follows the mixed Poisson distribution given by Eq. (2.3.2). It is assumed that purchasing across households are independent events. The mixed Poisson process, when λ_l is gamma distributed, was applied to consumer buying behavior by many authors (see e.g. Goodhardt et al. (1984); Ehrenberg (1988)).

2.3.3 Negative binomial first-order autoregressive processes

First-order autoregressive integer-valued processes, or INAR(1) processes, were independently constructed by McKenzie (1986, 1988) and Al-Osh and Alzaid (1987) in an effort to introduce a discrete-valued process analogous to the continuous valued first-order autoregressive, or AR(1), process with a stationary marginal distribution.

The AR(1) process and self-decomposability. The AR(1) process is a Markov process. Let π denote the marginal distribution of the process $\{X_t; t \in \mathbb{Z}\}$ and let X_π be a random variable with distribution π, then the AR(1) process is a discrete time process X_t that satisfies

$$X_t \overset{\mathcal{D}}{=} \alpha X_{t-1} + \varepsilon_t.$$

Here X_{t-1} and ε_t are mutually independent random variables from a continuous distribution, ε_t is a sequence of uncorrelated random variables for $t \in \mathbb{Z}$ and the value of α satisfies $\alpha \in (0,1)$. Here $\overset{\mathcal{D}}{=}$ means equivalence in distribution. The existence of an AR(1) process with a stationary marginal distribution requires self-decomposability of the marginal distribution such that its characteristic function $\phi_{X_\pi}(t) = Ee^{itX_\pi}$ satisfies the equation

$$\phi_{X_\pi}(t) = \phi_{X_\pi}(\alpha t)\phi_\varepsilon(t;\alpha) \qquad \alpha \in (0,1),\ t \in \mathbb{R}, \tag{2.3.5}$$

where $\phi_\varepsilon(t;\alpha) = Ee^{it\varepsilon}$ is the characteristic function of ε depending on α. Steutel and van Harn (1979) noted that the construction of discrete valued AR(1) processes is made difficult by the fact no non-degenerate discrete random variable satisfies Eq. (2.3.5). In response, they presented a discrete analogue of self-decomposability that allows the construction of discrete-valued process that resembles the AR(1) processes.

Discrete self-decomposability. A non-negative integer-valued random variable X is said to be discrete self-decomposable (see Steutel and van Harn (1979)) if for every $\alpha \in (0,1)$ the random variable X can be written as

$$X \stackrel{\mathcal{D}}{=} \alpha \circ X + X_\alpha. \tag{2.3.6}$$

Here the random variables $\alpha \circ X$ and X_α are independent. The '$\circ$' operator is called the thinning operator and $\alpha \circ X$ is defined as

$$\alpha \circ X \stackrel{\mathcal{D}}{=} \sum_{j=1}^{X} U_j, \tag{2.3.7}$$

where the U_j are i.i.d. Bernoulli random variables with $P(U_j = 1) = \alpha$ and $P(U_j = 0) = 1-\alpha$. Note that the random variable $\alpha \circ X$ conditioned upon $X = x$ follows a binomial distribution with mean $x\alpha$ and variance $x\alpha(1-\alpha)$. Using this definition of $\alpha \circ X$ the probability generating function (PGF) of the random variable $\alpha \circ X$ is $G_{\alpha \circ X}(c) = G_X(1-\alpha+\alpha c)$ and Eq. (2.3.6) may therefore be expressed as

$$G_X(c) = G_X(1-\alpha+\alpha c)G_{X_\alpha}(c) \tag{2.3.8}$$

where $G_{X_\alpha}(s)$ is the PGF of X_α.

The INAR(1) process. Al-Osh and Alzaid (1987) and McKenzie (1988) defined a non-negative integer-valued process $\{X_t; t \in \mathbb{Z}\}$ to be an INAR(1) process if the process satisfies the equation

$$X_t \stackrel{\mathcal{D}}{=} \alpha \circ X_{t-1} + \varepsilon_t, \tag{2.3.9}$$

where $\alpha \circ X_{t-1}$ and ε_t are mutually independent discrete random variables and the ε_t form a sequence of uncorrelated random variables for $t \in \mathbb{Z}$. The value of α must satisfy $\alpha \in (0,1)$ for the process to be stationary. It is assumed that the X_t and ε_t have finite means and variances.

The INAR(1) process X_t with marginal distribution π will have a stationary marginal distribution, i.e. $X_t \overset{\mathcal{D}}{=} X_{t-1} \overset{\mathcal{D}}{=} X_\pi$ for all $t \in \mathbb{Z}$, if and only if the random variable X_π is discrete self-decomposable and satisfies Eq. (2.3.8) so that

$$G_{X_\pi}(s) = G_{X_\pi}(1 - \alpha + \alpha s)G_\varepsilon(s; \alpha) \qquad \alpha \in (0, 1). \tag{2.3.10}$$

The autocorrelation function of an INAR(1) process was derived by both Al-Osh and Alzaid (1987) and McKenzie (1988). Due to the discrete self-decomposability of the INAR(1) process, implying stationarity, the autocorrelation function only depends on the time interval between events and is in fact identical to the autocorrelation function of an AR(1) process. Let X_t be an INAR(1) process with finite first and second moments then the autocorrelation function at lag u is given by

$$\rho(X_t, X_{t+u}) = \frac{E(X_t - EX_t)(X_{t+u} - EX_{t+u})}{\sqrt{\mathrm{Var}(X_t)\mathrm{Var}(X_{t+u})}} = \rho(u) = \alpha^{|u|}, \quad u \in \mathbb{Z}. \tag{2.3.11}$$

Further developments of discrete valued processes. McKenzie (1986, 1988) has considered integer-valued autoregressive (INAR) and moving-average (INMA) processes with Poisson and NBD marginal distributions. The INAR(1) process has been generalized to the INAR(p) process by Al-Osh and Alzaid (1990) and Du and Li (1991). Both authors discuss similarities and differences between the INAR(p) and AR(p) processes. Du and Li (1991), in particular, show that the INAR(p) process is ergodic. Latour (1997, 1998) generalizes the INAR(p) process by allowing a general thinning operator, similar to Eq. (2.3.7), where the Bernoulli random variables in the thinning operation are substituted by any non-negative integer-valued random variables with finite mean and variance. Methods of estimation have so far only utilized the methods described by Al-Osh and Alzaid (1987). The problem of estimation will be discussed in more detail in Section 2.4.2. In this thesis only the INAR(1) process will be considered.

Negative binomial first-order autoregressive processes

Two different negative binomial first-order autoregressive processes will now be introduced. These processes were constructed by McKenzie (1986) as discrete analogues to gamma autoregressive processes. The processes differ in that a NBD INAR(1) process can be constructed by considering α, the thinning parameter, as either deterministic or stochastic.

The NBD INAR(1) process with deterministic thinning

The first NBD INAR(1) process obtained by McKenzie (1986) was derived by considering the standard form of the INAR(1) process defined by Eq. (2.3.9) with α fixed. Note that if the process X_t has a NBD(m,k) marginal distribution then X_π is discrete self-decomposable since the PGF of X_π can be written in the form of Eq. (2.3.10) with

$$\underbrace{\left(1+\frac{m(1-c)}{k}\right)^{-k}}_{G_{X_\pi}(c)} = \underbrace{\left(1+\frac{m\alpha(1-c)}{k}\right)^{-k}}_{G_{X_\pi}(1-\alpha+\alpha c)} \underbrace{\left(\frac{k+m(1-c)}{k+m\alpha(1-c)}\right)^{-k}}_{G_\varepsilon(c;\ \alpha)}. \tag{2.3.12}$$

The generating function of the error distribution, $G_\varepsilon(c;\alpha)$, presented in this equation is indeed a well defined PGF. McKenzie (1986) noted that the PGF of the ε_t was of an obscure form and did not specify the distribution of the errors. A method for generating a random variable from the distribution of ε_t, however, was presented, since it was shown that the errors could be represented in the form of a compound Poisson process given by

$$\varepsilon_t = \sum_{j=1}^{P}(\alpha^{U_i}) \circ Y_i \quad \alpha \in (0,1). \tag{2.3.13}$$

Here P is Poisson distributed with mean $-k\log\alpha$, the U_i are uniformly distributed on $(0,1)$ and the Y_i are NBD$(m/k,1)$ random variables. The random variables N, U_i and Y_i, $i=1,2,\ldots,N$ are all independent of each other.

The NBD INAR(1) process with stochastic thinning

As an alternative to the NBD INAR(1) process with deterministic thinning, McKenzie (1986) proposed a process whereby the errors also have a NBD distribution. Assume that there exists a non-negative integer-valued autoregressive process X_t with i.i.d. stochastic thinning parameters A_t supported on the interval $(0, 1)$, then the INAR(1) process with stochastic thinning is defined by

$$X_t \stackrel{\mathcal{D}}{=} A_t \circ X_{t-1} + \varepsilon_t, \tag{2.3.14}$$

where for fixed t the A_t, X_{t-1} and ε_t are independent random variables. If the process X_t defined by Eq. (2.3.14) is to be a stationary process then the PGF of X_π must satisfy

$$G_{X_\pi}(c) = \int_0^1 G_{X_\pi}(1 - y + yc)dF_A(y)G_\varepsilon(c;\ \alpha), \tag{2.3.15}$$

where $F_A(y)$ is the cumulative distribution function (c.d.f.) of A_t.

McKenzie (1986) derived a stationary NBD INAR(1) process with stochastic thinning by letting X_t be NBD(m, k) and letting A_t follow a Beta distribution defined by

$$f_{A_t}(y) = \frac{y^{l-1}(1-y)^{k-l-1}}{B(l, k-l)}, \quad l > 0,\ k - l > 0,\ 0 < y < 1,$$

where $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ is the beta function. The NBD INAR(1) process with stochastic thinning can be represented in terms of Eq. (2.3.15) by

$$\underbrace{\left(1 + \frac{m(1-c)}{k}\right)^{-k}}_{G_{X_\pi}(c)} = \underbrace{\left(1 + \frac{m(1-c)}{k}\right)^{-l}}_{\int G_{X_\pi}(1-y+yc)dF_A(y)} \underbrace{\left(1 + \frac{m(1-c)}{k}\right)^{-(k-l)}}_{G_\varepsilon(c;\ \alpha)}.$$

The generating function of the error distribution may be represented in the form

$$G_\varepsilon(c;\ \alpha) = \left(1 + \frac{m(1-l/k)(1-c)}{k-l}\right)^{-(k-l)}$$

from which it becomes clear that the errors are NBD$(m(1 - l/k), k - l)$.

Long-range dependent processes

A process is often said to be long-range dependent or have long-memory if the process has non-summable correlations or if the spectral density has a pole at the origin. There are various statistical definitions of long-memory and they are not all equivalent. A thorough review on long-range dependence has been made by Beran (1994) and more recently by Doukhan, Oppenheim, and Taqqu (2003).

Barndorff-Nielsen (1998) constructed a stationary long-memory normal-inverse Gaussian (NIG) process in continuous time by the superposition (or aggregation) of short-memory Ornstein-Uhlenbeck type processes with NIG marginal distributions. For suitable parameters of the individual short-memory NIG processes, each with the same autocovariance function, the aggregated process was shown to have long-memory with autocovariance function of the form

$$R(u) \simeq L(u)u^{-2(1-H)}, \quad H \in (0.5,1)\,, \; u \in \mathbb{R} \quad \text{as} \quad u \to \infty, \tag{2.3.16}$$

where H is the long-memory (or Hurst) parameter and $L(u)$ is a slowly varying function.

In particular, a stationary process X_t has long memory, if there exist constants H $\in (0.5, 1)$ and $c_\rho > 0$ such that the correlation function $\rho(u)$ of the process X_t satisfies

$$\lim_{u\to\infty} \rho(u) \,/\, \left[c_\rho\, u^{2H-2}\right] = 1.$$

If the above condition is satisfied then H is called the Hurst parameter. Alternatively, a stationary process X_t has long memory, if for some $\kappa \in (0,1)$ and $c_f > 0$, the spectral density $f(\lambda)$ of X_t satisfies

$$\lim_{\lambda\to\infty} f(\lambda)/ \left[c_f|\lambda|^{-\kappa}\right] = 1.$$

2.4 Fitting negative binomial processes

This section reviews well known methods of fitting the negative binomial processes considered in Section 2.3. Methods of fitting these processes have often relied on fitting the marginal distribution to data. Since there are different processes with negative binomial marginal distributions, an adequate fit of the NBD to data provides only partial indication about the adequacy of the process. Methods of fitting the gamma Poisson process and INAR(1) processes are discussed in sections 2.4.1 and 2.4.2 respectively.

2.4.1 Fitting the gamma Poisson process

The gamma Poisson process has the feature that the one-dimensional marginal distribution of the process is NBD whose parameter parameter k remains constant in time and whose mean m increases linearly with time. Here m and k are parameters of the NBD for a unit time interval. In literature the fit of the gamma Poisson process has mainly focussed on fitting the one-dimensional marginal distribution to the data (see Greenwood and Yule (1920); Lundberg (1964); Grandell (1997)). In the work of Ehrenberg (1988), however, a more detailed investigation into the adequacy of the gamma Poisson process as a model for consumer buying behavior is presented.

Using household panel data Ehrenberg (1988) verified that consumer purchase occasions could be successfully modeled by the gamma Poisson process. Consumer purchase occasions represent the rate of recurrence with which households purchase products. An advantage of panel data is that multiple realizations of the gamma Poisson process are observed. In the case of market research, when collecting household panel data, the number of purchases are recorded for many customers over a specific time period. Each customer, therefore, has their own realization and this information can be used to test adequacy of the gamma Poisson process.

Ehrenberg (1988) tested the adequacy of the gamma Poisson process by fitting the NBD to observed data over time intervals of different length and comparing the observed and expected frequencies of consumer purchase occasions. In addition Ehrenberg (1988) considered the behavior of various repeat buying measures commonly used in market research. Ehrenberg (1988) used the multivariate NBD (Eq. (2.3.3)) to consider how repeat buying measures associate between two non-overlapping time intervals and compared the theoretical and observed patterns.

A description of repeat buying measures follows. The repeat buying measures considered by Ehrenberg (1988) are then described under the heading of single-period repeat buying theory and multi-period repeat buying theory. Single-period repeat buying theory considers how repeat buying measures develop as time increases whereas multi-period repeat buying theory considers the relationship between repeat buying measures in different time intervals.

Repeat buying measures

The single-period repeat buying measures are functionals of the one-dimensional marginal distribution that have a natural interpretation in the field of market research. Assume, for simplicity, that the marginal distribution is NBD. These measures are often estimated in practice by using intuitive methods where probabilities are replaced by observed proportions. Such estimators may, however, be biased. Let X be a random variable from the NBD and let p_x denote the probabilities of the NBD.

Penetration. The simplest measure of consumer buying behavior is the penetration of a product, which represents the probability that an individual makes at least one purchase in a given time period. The penetration is defined by

$$b = 1 - p_0, \qquad 0 \leqslant b \leqslant 1. \tag{2.4.1}$$

When estimating the NBD parameters by the zero term method, popular in the field of market research, the penetration is estimated by the frequency of non-zero buyers. In practice, estimation of penetration using the zero term method can cause problems due to the ambiguity in the definition of a zero buyer; indeed it is difficult to distinguish between zero buyers who are potential buyers and zero buyers who will never purchase the product in their lifetime.

Purchase frequency. The purchase frequency of an item represents the mean number of purchase occasions of the population who purchase an item at least once in the analysis period. The purchase frequency w is

$$w = \mathbb{E}(X|X \geqslant 1) = \frac{m}{b}, \qquad w \geqslant 1. \tag{2.4.2}$$

Measured repeat. The r-th ($r = 1, 2, 3, \ldots$) measured repeat of a product represents the proportion of households who bought a product at least $r+1$ times out of those households who bought the product at least r times. Theoretically, the r-th measured repeat is

$$\beta_r = \mathbb{P}(X \geqslant r+1|X \geqslant r) = \frac{1 - \sum_{j=0}^{r} \mathbb{P}(X = j)}{1 - \sum_{j=0}^{r-1} \mathbb{P}(X = j)}. \tag{2.4.3}$$

Repeats per repeater. The r-th ($r = 1, 2, 3, \ldots$) repeats per repeater of a product represents the mean purchase frequency of the households who bought a product at least $r+1$ times. The mean purchase frequency is usually shifted by the value r so that the minimum possible purchase frequency is always one. The r-th theoretical repeats per repeater is

$$\omega_r = \mathbb{E}(X - r|X \geqslant r+1) = \frac{m - \sum_{j=0}^{r} j\,\mathbb{P}(X = j)}{1 - \sum_{j=0}^{r} \mathbb{P}(X = j)} - r. \tag{2.4.4}$$

Single-period repeat buying

In single-period repeat buying analysis the length of time, $t > 0$, over which data is analyzed may be taken to be variable. Equation (2.3.4) showed the one-dimensional distribution of the gamma Poisson process as a function of time. The repeat buying measures as a function of time can therefore be easily obtained. Let $X(t)$ be a random variable from the gamma Poisson process observed over a time interval of length t and let p_x denote the probabilities of the NBD as given by Eq. (2.3.4).

Penetration. The penetration $b(t)$ as a function of time is

$$b(t) = 1 - \mathbb{P}(X(t) = 0) = 1 - (1 + at)^{-k} \quad 0 \leqslant b(t) \leqslant 1. \tag{2.4.5}$$

The penetration is a non-linear non-decreasing function of t as t increases. Since no purchases may be made at time intervals of length $t = 0$ units, we have $b(0) = 0$. At time $t = \infty$ the penetration $b(\infty) = 1$ and the model presumes that given an infinite amount of time the whole population will make at least one purchase of the item.

Purchase frequency. The purchase frequency $w(t)$ is

$$w(t) = \mathbb{E}(X(t)|X(t) \geqslant 1) = \frac{akt}{b(t)} \qquad w(t) \geqslant 1. \tag{2.4.6}$$

As a function of time $w(t)$ is a strictly increasing function.

Measured repeat. The measured repeat is

$$\mathbb{P}(X(t) \geqslant r + 1|X(t) \geqslant r) = \frac{1 - \sum_{j=0}^{r} \mathbb{P}(X(t) = j)}{1 - \sum_{j=0}^{r-1} \mathbb{P}(X(t) = j)}. \tag{2.4.7}$$

Repeats per repeater. The theoretical repeats per repeater is

$$\mathbb{E}(X(t) - r|X(t) \geqslant r + 1) = \frac{m - \sum_{j=0}^{r} j\,\mathbb{P}(X(t) = j)}{1 - \sum_{j=0}^{r} \mathbb{P}(X(t) = j)} - r. \tag{2.4.8}$$

Multi-period repeat buying

In the analysis of multi-period repeat buying, Ehrenberg (1988) considered the association between market measures in two different non-overlapping time intervals. The association is made simply by using the two-dimensional distribution of the gamma Poisson process which can be obtained from Eq. (2.3.3). Although the time intervals do not necessarily have to be of equal length, Ehrenberg (1988) mainly considered equal length time periods due to the simplification in theoretical formulae.

Ehrenberg (1988) considered how market measures for purchases in the combination of two non-overlapping intervals relate to market measures for purchases in two different time-periods of equal length. Since the time-periods are of equal length, the NBD parameters are identical in each of the two individual time periods. Therefore, without loss of generality, it may be assumed that purchases follow the $\text{NBD}_a(a, k)$ distribution in each of the individual time periods.

Penetration in two equal length time-periods. Let b_r denote the probability that a consumer buys in both periods and let consumer buying behavior follow a gamma Poisson process such that purchases are $\text{NBD}_a(a, k)$ in the two individual time-periods of equal length, then

$$b_r = 1 - 2(1 + a)^{-k} + (1 + 2a)^{-k}.$$

Note that b_r is not equivalent to the penetration in the two combined periods and the representation used by Ehrenberg (1988) can be therefore be misleading. The probability that a consumer buys in only one of the two intervals is $b_n = b - b_r$ where $b = 1 - (1 + a)^{-k}$ and b is the penetration in an individual time period. Ehrenberg (1988) used b_n to check frequencies of new buyers that did not purchase in the first time period but did purchase in the second time period.

Purchase frequency in two equal length time-periods. The mean purchase frequency, w_r, of consumers that purchase in both periods is

$$w_r = b_r/m_r, \quad \text{where} \quad m_r = m\left(1 - (1+a)^{-k-1}\right)$$

and the mean purchase frequency, w_n of consumers that purchase in only one period is

$$w_n = b_n/m_n, \quad \text{where} \quad m_n = m(1+a)^{-k-1}.$$

Conditional trend analysis

To consider a more detailed fit of the gamma Poisson process to data Ehrenberg (1988) considered the use of "conditional trend analysis". Here, two consecutive periods are taken and the distribution of purchases in the second period are analyzed conditional upon the number of purchases observed in the first time period.

Let the distribution of purchases for a random individual in the first time interval be $\mathrm{NBD}_a(a,k)$. Ehrenberg (1988) noted that given y purchases are made in the first time interval, the probability mass function of purchases made in the second time interval, on the assumption of time intervals of equal length, is

$$\mathbb{P}\left(Z(t_2) = x | Z(t_1) = y\right) = \frac{\Gamma(k+y+x)}{\Gamma(k+y)x!}\left(\frac{1+2a}{1+a}\right)^{-(k+y)}\left(\frac{a}{1+2a}\right)^{x} \qquad x = 0, 1, 2, \dots,$$

which is the $\mathrm{NBD}_a(a/(1+a), k+y)$ distribution.

The observed and expected market measures in the second period can be compared conditional upon the observed frequency of purchases in the first period. The market measures in the second period, conditional upon the fact that y purchases are made in the first period are given by Eqs. (2.4.2)-(2.4.4) with the parameter pair (a,k) in these formulae replaced by the parameter pair $(a/(1+a), k+y)$. For example, let b_y and w_y represent the conditional measures then

$$b_y = 1 - \left(\frac{1+2a}{1+a}\right)^{-(k+y)} \quad \text{and} \quad w_y = \frac{m_y}{b_y} \quad \text{where} \quad m_y = \frac{a(k+y)}{(1+a)}.$$

2.4.2 Fitting the INAR(1) process

The fitting of the INAR(1) process was first considered by Al-Osh and Alzaid (1987) in the case when the marginal distribution of the process is Poisson distributed. The model was fit to data simulated from an INAR(1) process. In practice, the Poisson INAR(1) process has been applied by Franke and Seligmann (1993) and Silva and Oliveira (2005) in the case of epileptic seizure counts employing the methods described by Al-Osh and Alzaid (1987). The INAR model has also been applied by Gourieroux and Jasiak (2004) in the case of car insurance claims.

Al-Osh and Alzaid (1987) proposed to determine the adequacy of the INAR(1) process by verifying that the empirical autocorrelation function has the equivalent theoretical form

$$\rho(X_t, X_{t+u}) = \frac{E(X_t - EX_t)(X_{t+u} - EX_{t+u})}{\sqrt{\mathrm{Var}(X_t)\mathrm{Var}(X_{t+u})}} = \rho(u) = \alpha^{|u|}, \quad u \in \mathbb{Z}. \tag{2.4.9}$$

Since the autocorrelation function of the INAR(1) process is identical to the autocorrelation function of the AR(1) process, the problem of estimating α has been well documented (see e.g. Brockwell and Davis (2002)). On estimating the parameter α, the problem is then reduced to that of estimating the parameters of the marginal distribution of the process.

In addition, Al-Osh and Alzaid (1987) considered the problem of estimating the mean parameter λ of the Poisson INAR(1) process using three different types of estimators. The first two types of estimators, called the Yule-Walker estimator and the conditional least squares estimator, use moment based methods and are asymptotically equivalent (see Freeland and McCabe (2005)). The third method of estimation uses the maximum likelihood approach. It will be assumed that $(x_1, x_2, \ldots, x_N)$ is a sample of size N from an INAR(1) process X_t.

Yule-Walker estimators The Yule-Walker method estimates parameters of a time series by equating the theoretical autocorrelation function to the empirical autocorrelation function. The autocorrelation function of an INAR(1) process is given by $\rho(u) = \alpha^{|u|}$, $u \in \mathbb{Z}$. The parameter α of the INAR(1) process may therefore be estimated by the equation

$$\widehat{\rho(u)} = \hat{\alpha}^u = \frac{\sum_{t=1}^{N-u}(x_t - \bar{x})(x_{t+u} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \qquad u \in \mathbb{Z}, \tag{2.4.10}$$

where $\bar{x} = \sum_{t=0}^{N} x_t$. Note that multiple estimates for α may be obtained using different values of the lag u. Using the properties of the thinning operator, the expected value of the errors is $\mathbb{E}[\varepsilon_t] = \mathbb{E}[X_t] - \alpha\mathbb{E}[X_{t-1}]$. The estimated value of α, denoted by $\hat{\alpha}$, may therefore be used to obtain estimates for the observations of the uncorrelated errors by computing $\widehat{\varepsilon_t} = x_t - \hat{\alpha}x_{t-1}$ for $t = 1, 2, \ldots, N$. The distribution of the errors may then be used to estimate the distributional parameters of the process. In the case of the Poisson INAR(1) process with thinning parameter α and X_π having mean λ, the value of λ is estimated by $\hat{\lambda} = \frac{(1-\hat{\alpha})}{N} \sum_{t=1}^{N} \widehat{\varepsilon_t}$.

Conditional least squares estimators The conditional least squares estimators are derived by minimizing the sum of squares of X_t conditioned upon the value of X_{t-1}. The estimators for an INAR(1) process are therefore derived by minimizing the function

$$\sum_{t=2}^{N}(X_t - E[X_t|X_{t-1}])^2 = \sum_{t=2}^{N}(X_t - \alpha X_{t-1} - E[\varepsilon_t])^2 = \sum_{t=2}^{N}(X_t - \alpha X_{t-1} - (1-\alpha)E[X_t])^2.$$

with respect to α and the distributional parameters. The parameter estimates for the Poisson INAR(1) process are

$$\hat{\alpha} = \frac{\sum_{t=2}^{N} x_t x_{t-1} - \frac{1}{N-1}\left(\sum_{t=2}^{N} x_t \sum_{t=2}^{N} x_{t-1}\right)}{\sum_{t=2}^{N} x_{t-1}^2 - \frac{1}{N-1}\left(\sum_{t=2}^{N} x_{t-1}\right)^2} \quad \text{and} \quad \hat{\lambda} = \frac{1}{N}\left(\sum_{t=2}^{N} x_t - \hat{\alpha}\sum_{t=2}^{N} x_{t-1}\right).$$

The conditional least squares estimators and the Yule-Walker estimators, at $u = 1$, are asymptotically equivalent for the Poisson INAR(1) process.

Conditional maximum likelihood estimators The conditional maximum likelihood estimators are obtained by maximizing the likelihood function given an initial value x_1 for the sample. Let $\mathbf{x} = (x_1, x_2, \dots, x_N)$ be an observed sample from an INAR(1) process and let Θ be the set of parameters for the INAR(1) process then the likelihood function is

$$\begin{aligned} L(\mathbf{x};\ \Theta) &= \mathbb{P}(X_1 = x_1)\prod_{t=2}^{N}\mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}) \\ &= \mathbb{P}(X_1 = x_1)\prod_{t=2}^{N}\mathbb{P}\left(\alpha \circ x_{t-1} + \varepsilon_t = x_t\right) \\ &= \mathbb{P}(X_1 = x_1)\prod_{t=2}^{N}\sum_{r=0}^{\min(x_t,x_{t-1})}\binom{x_{t-1}}{r}\alpha^r(1-\alpha)^{x_{t-1}-r}\mathbb{P}(\varepsilon_t = x_t - r). \end{aligned} \tag{2.4.11}$$

Denote $\mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}) = P_{x_{t-1},x_t}(t-1,t)$, then Al-Osh and Alzaid (1987) defined the conditional maximum likelihood function, given the value of x_1, to be

$$L(\mathbf{x}|x_1;\ \Theta) = \prod_{t=2}^{N} P_{x_{t-1},x_t}(t-1,t) = \prod_{t=2}^{N}\sum_{r=0}^{\min(x_t,x_{t-1})}\binom{x_{t-1}}{r}\alpha^r(1-\alpha)^{x_{t-1}-r}\mathbb{P}(\varepsilon_t = x_t - r). \tag{2.4.12}$$

Al-Osh and Alzaid (1987) used a method described by Sprott (1983) to maximize the conditional maximum likelihood function for a sample generated by a Poisson INAR(1) process. The conditional maximum likelihood function for a Poisson INAR(1) process with thinning parameter α and $\mathbb{E}[X_\pi] = \lambda$ is

$$L(\mathbf{x}|x_1;\ \Theta) = \prod_{t=2}^{N}\sum_{r=0}^{\min(x_t,x_{t-1})}\frac{e^{-\lambda(1-\alpha)}\left(\lambda(1-\alpha)\right)^{x_t-r}}{(x_t-r)!}\binom{x_{t-1}}{r}\alpha^r(1-\alpha)^{x_{t-1}-r}. \tag{2.4.13}$$

Sprott (1983) noted that with the computational power available it is possible to numerically maximize the likelihood function. Brännäss (1994) noted that maximum likelihood estimation for the NBD INAR(1) model is difficult due to the complex form of the maximum likelihood equations even in the case of an i.i.d. NBD sample (i.e. in the case $\alpha = 0$).

Chapter summary and discussion

Negative binomial processes have been used as models in fields such as accident proneness, accidents and sickness, market research, risk theory and more recently in clinical trials. This thesis concentrates on the efficient fitting of the NBD and negative binomial processes to observed data, with application to market research data in mind.

Fitting the NBD. The NBD can be parameterized in numerous ways and the first problem in the estimation of NBD parameters is the choice of parameters to be estimated. The NBD parameters all have a one-to-one relationship and it therefore suffices to estimate just one of the parameter pairs. Since parameter estimates of m and k are asymptotically uncorrelated for natural moment based estimators and maximum likelihood estimators (m is the mean and k is the shape parameter of the NBD), estimation in literature has justifiably focussed on the estimation of (m, k).

In practice, maximum likelihood estimators are difficult to implement and, depending upon the NBD parameter values, the standard moment based estimators currently used can be inefficient. Chapter 3 will investigate problems related to the efficient estimation of NBD parameters using moment based estimators for i.i.d. NBD samples.

Parameter estimation in the case when the sample follows the INAR(1) processes has been considered by Al-Osh and Alzaid (1987). The methods suggested by Al-Osh and Alzaid (1987) require estimating the moments of the error distribution. The error distribution in the case of the NBD INAR(1) process is complex. The INAR(1) process is an ergodic process. It should therefore be possible to equate moments of a single observed realization to the moments of the stationary distribution in order to estimate parameters of the process. Chapter 3 will also consider efficient estimation of NBD parameters for NBD INAR(1) samples.

Negative binomial processes. Numerous negative binomial processes have been considered in literature. These include the gamma Poisson process, which falls into the class of mixed Poisson processes, and the negative binomial INAR(1) process, which falls into the class of integer-valued first-order autoregressive processes. The two classes of processes mentioned have the common feature that the marginal distribution of the process is negative binomial. Chapter 4 will deal with statistical inference for the gamma Poisson process and the negative binomial INAR(1) process. The two processes will be considered separately as they both come from different families of processes (the families of renewal processes and autoregressive processes respectively).

Ehrenberg (1988) has considered the goodness of fit of the gamma Poisson process to consumer buying data by empirically comparing various market research measures computed over varying time intervals. The gamma Poisson process is not an ergodic process. In the case of market research, however, multiple realizations of the process are observed thus enabling valid statistical inference to be made from data generated by the gamma Poisson process. Chapter 4 will consider statistical inference of the gamma Poisson process by investigating the joint asymptotic distribution of various statistics or estimators computed from data generated in different time intervals. The methodology discussed in Section 2.2 will be used to compute that asymptotic distribution of the statistical pairs.

The most common method of statistical inference of autoregressive processes is to consider the autocorrelation function of the time series (time domain analysis) or to consider the spectral density of the process (frequency domain analysis). Long-range dependence has been of great interest in literature (see e.g. Beran (1994); Doukhan et al. (2003)). Chapter 4 will consider developing the NBD INAR(1) models by extending the NBD INAR(1) models to NBD INAR(1) models with long-range dependence.

Developing discrete-valued time series models with long-range dependence will provide possible models for integer-valued data observing long-range dependence.

Application to market research data. Chapter 5 will analyze market research data and assess suitability of the gamma Poisson process and the NBD INAR(1) process to the market research data kindly provided by ACNielsen BASES. Ehrenberg (1988) has empirically verified goodness of fit of the gamma Poisson process to consumer buying behavior through the use of the Chi-squared test statistic and empirically compared observed and expected values of numerous market research measures. In this chapter the goodness of fit of the gamma Poisson process will be verified by using the traditional Chi-squared technique as well as the statistical inference procedures developed in Chapter 3 and Chapter 4.

Chapter 3

The power method for estimating parameters of the NBD

This chapter investigates the efficiency of the power method (PM) estimator for independent and dependent INAR(1) samples and considers the implementation of PM estimators in practice. Section 3.1 investigates the efficiency of PM estimators for i.i.d. NBD samples. Section 3.2 investigates the PM for estimating NBD parameters from NBD INAR(1) samples. Finally, Section 3.3 considers the implementation of PM estimators in practice.

Only estimation of the NBD parameter pair (m, k) will be considered since maximum likelihood (ML) estimators and all natural moment based estimators for (m, k) are asymptotically uncorrelated given an i.i.d. NBD sample. For dependent samples, the parameter pair (m, k) is no longer uncorrelated due to dependence in sample observations. Nevertheless, the parameter pair (m, k), as opposed to other parameter pairs, will be considered for simplicity.

The choice of an optimum estimation method in this thesis is determined by the method whose estimators minimize the determinant of the covariance matrix. The covariance matrix of the ML estimators provides a lower bound for the covariance matrix of all asymptotically normal, and hence moment based, estimators.

3.1 Efficient moment estimators for i.i.d. samples

Anscombe (1950) proved that the ML estimator and all natural moment based estimators for the parameter pair (m, k), when estimating from i.i.d. NBD samples, are asymptotically uncorrelated so that $\lim_{N\to\infty} \mathrm{Cov}(\hat{m}, \hat{k}) = 0$. The parameter m is always efficiently estimated by $\hat{m} = \bar{x}$ for both ML and moment based estimation methods. The most efficient moment based method of estimation for parameters of an i.i.d. NBD sample is therefore determined by the method whose estimator for k achieves the lowest asymptotic variance. The asymptotic normalized variance for the ML estimator of k is

$$v_{ML} = \lim_{N\to\infty} N \,\mathrm{Var}\left(\hat{k}_{ML}\right) = \frac{2k(k+1)(a+1)^2}{a^2\left(1 + 2\sum_{j=2}^{\infty}\left(\frac{a}{a+1}\right)^{j-1}\frac{j!\Gamma(k+2)}{(j+1)\Gamma(k+j+1)}\right)}, \quad (3.1.1)$$

where $a = m/k$. This variance v_{ML} is a lower bound for the asymptotic normalized variance of all asymptotically normal estimators.

The asymptotic normalized variances for the method of moments (MOM), zero term method (ZTM) and power method (PM) estimators of k are

$$v_{MOM} = \lim_{N\to\infty} N\mathrm{Var}\left(\hat{k}_{MOM}\right) = \frac{2k(k+1)(a+1)^2}{a^2}, \quad (3.1.2)$$

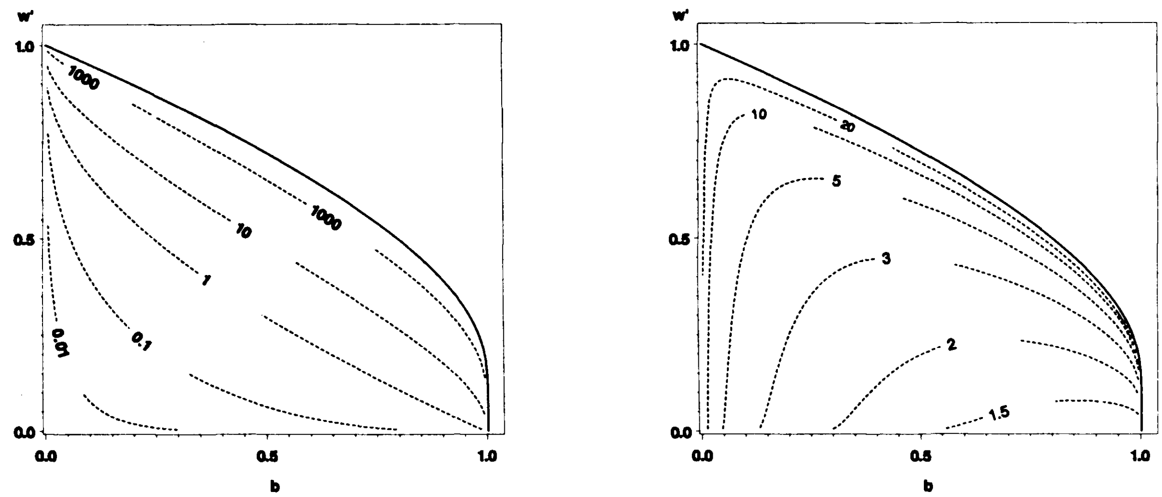
$$v_{ZTM} = \lim_{N\to\infty} N\mathrm{Var}\left(\hat{k}_{ZTM}\right) = \frac{(a+1)^{k+2} - (a+1)^2 - ka(a+1)}{[(a+1)\log(a+1) - a]^2}, \quad (3.1.3)$$

$$v_{PM}(c) = \lim_{N\to\infty} N\mathrm{Var}\left(\hat{k}_{PM(c)}\right) = \frac{(1+a-ac^2)^{-k} r^{2k+2} - r^2 - ka(a+1)(1-c)^2}{[r\log(r) - r + 1]^2}, \quad (3.1.4)$$

where $r = 1 + a - ac$. The variance of $\hat{k}$ for the factorial method is difficult to express explicitly; an expression of the variance is given in Anscombe (1950, p. 369).

Behavior of the ML estimator for k

Before investigating the efficiency of the PM estimator, the limits of the efficiency levels of moment based estimators is considered by considering the behavior of the asymptotic normalized variance of the ML estimator for k.



(a) (b)

Figure 3.1: Contour levels of (a) v_{ML} and (b) $\sqrt{v_{ML}}/k$.

Fig. 3.1(a) shows contour levels of the asymptotic normalized variance, v_{ML}, for $\hat{k}_{ML}$. It is clear that v_{ML} increases as $k \to \infty$ and the NBD converges to the Poisson distribution. For large values of k, the Poisson distribution is a good approximation to the NBD and the probabilities of the NBD will therefore be dominated by the mean m of the distribution; in such cases, the probabilities of the NBD will be insensitive to changes in the value of k.

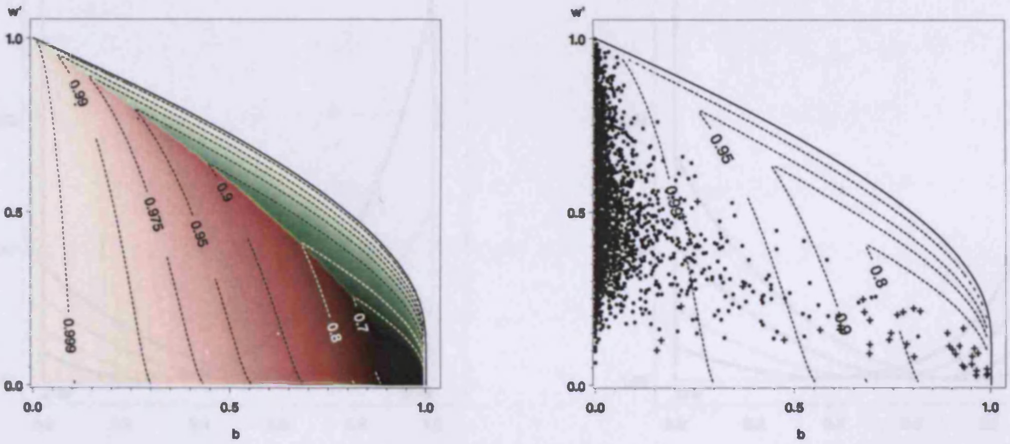
Fig. 3.1(b) shows contour levels of the coefficient of variation $\sqrt{v_{ML}}/k$, thereby indicating areas of the NBD parameter space where estimation of k is difficult even for ML. Assume that k is fixed but m is allowed to vary, then Fig. 3.1(b) indicates that a smaller m would require larger sample sizes in comparison to a large m in order to obtain a fixed precision of the ML estimator for k.

In consideration of the results shown in Fig. 3.1 and the fact that in the practice of market research large values of k appear to be rarely observed (see e.g. Fig. 2.2(a) where no products with $\hat{k}_{ZTM} > 3$ were observed), it seems sensible to concentrate on efficient estimation of parameters in areas of the parameter space where $k < 3$.

3.1.1 Efficiency of the MOM/ZTM estimator

The MOM and ZTM estimators are commonly used in practice for the estimation of NBD parameters given i.i.d. NBD samples. These estimators, however, achieve low efficiency levels in certain regions of the NBD parameter space when compared to ML. Fig. 3.2(a) shows the efficiency of the MOM/ZTM estimator, the more efficient method amongst the MOM and ZTM estimators relative to the ML estimator, given by $v_{ML}/\min\{v_{MOM}, v_{ZTM}\}$. The green and red shading in the figure respectively represents areas where the MOM and ZTM are the more efficient in comparison to each another.

Fig. 3.2(b) shows ZTM estimates of the NBD parameters for 46 different categories and the top 50 brands within each category with data courteously provided by AC-Nielsen BASES. Fig. 3.2(b) indicates that the NBD parameters for numerous products are inefficiently estimated, with efficiency levels sometimes reaching below 70% when compared to ML. In the practice of market research, where MOM and ZTM estimators are commonly used, parameter estimates for large values of m and small values of k may be inefficient.



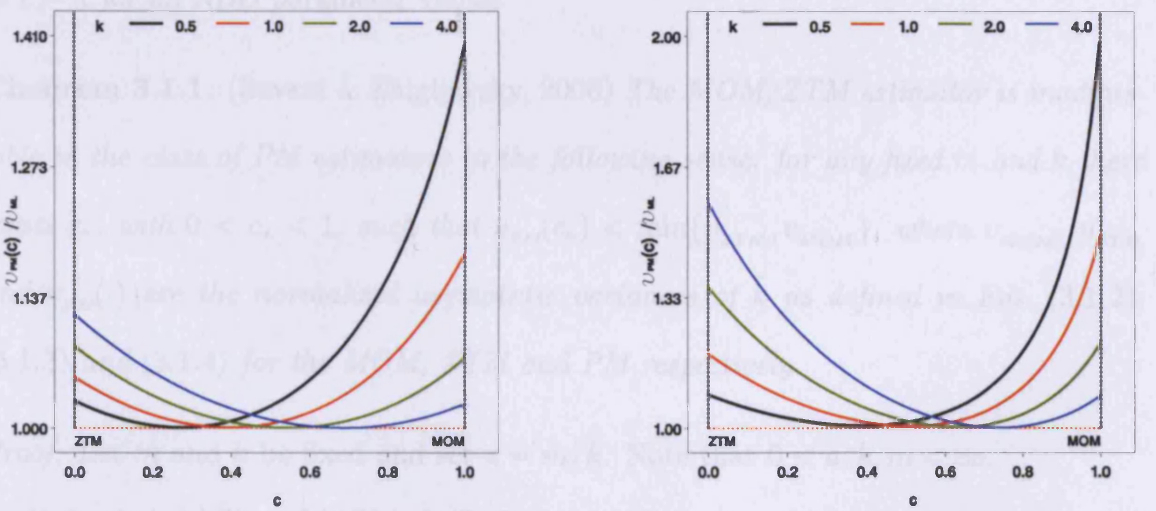
(a) (b)

Figure 3.2: (a) Efficiency of the more efficient amongst the MOM and ZTM estimator ($v_{ML}/\min\{v_{MOM}, v_{ZTM}\}$) (b) Contour levels of $v_{ML}/\min\{v_{MOM}, v_{ZTM}\}$ together with ZTM estimators for NBD parameters when fitting the NBD to the top 50 brands in each category and 46 categories in consumer buying behavior. Data courtesy of ACNielsen BASES.

3.1.2 Efficiency of the power method estimator

The computation of the PM estimator for k depends on an additional parameter c; the PM estimator is equal to the ZTM estimator if $c = 0$ and tends to the MOM estimator as $c \to 1$. Fig. 2.2(a) showed contour levels of the efficiency of the MOM, ZTM and PM(0.5) estimators in the (m, k) parameter space. All three methods of estimation achieved high levels of efficiency in different regions of the NBD parameter space. This raises the question as to whether efficient estimators can be obtained by choosing an appropriate value of c depending upon the parameters (m, k).

Denote the PM estimator for k computed at c as the PM(c) estimator. Fig. 3.3 shows the relative asymptotic efficiency, $v_{PM}(c)/v_{ML}$, of the PM(c) estimators for k with respect to the ML estimator for k for different parameter values (m, k). The values of the efficiency are plotted against the power method parameter $c \in (0, 1)$. Note that $v_{PM}(0) = v_{ZTM}$ and $v_{PM}(1) = v_{MOM}$. Fig. 3.3 shows that there exists a range of values of c_* such that $v_{PM}(c_*) < \min\{v_{ZTM}, v_{MOM}\}$.



$m = 0.5$ $m = 2$

Figure 3.3: $v_{PM}(c)/v_{ML}$ versus c for different parameter values (m, k).

Note that for these cases there exists a single optimum c,

$$c_o = \text{argmin}_{c\in(0,1)} v_{PM}(c), \tag{3.1.5}$$

where $v_{PM}(c_o)/v_{ML} \approx 1$, so that the PM($c_o$) estimator is almost as efficient as the ML estimator for k. The proof that a single optimum c exists for all NBD parameter values requires proving the convexity of the function $v_{PM}(c)$ in c. The complex form of the function $v_{PM}(c)$, however, makes it difficult to prove that the function is indeed a convex function. An attempt to prove the convexity of the function $v_{PM}(c)$ was unfortunately unsuccessful.

Inadmissability of the MOM/ZTM

The inadmissability of the MOM and ZTM estimators is now proven in that there always exists a $c \in (0,1)$ such that $v_{PM}(c) < \min\{v_{ZTM}, v_{MOM}\}$ for all NBD parameter values. The proof basically relies on the fact that $v_{PM}(c)$ is a continuous function for $c \in [0,1]$ and that the gradient of $v_{PM}(c) < 0$ at $c = 0$ and the gradient of $v_{PM}(c) > 0$ at $c = 1$ for all NBD parameter values.

Theorem 3.1.1. (Savani & Zhigljavsky, 2006) *The MOM/ZTM estimator is inadmissible in the class of PM estimators in the following sense: for any fixed m and k there exists c_*, with $0 < c_* < 1$, such that $v_{PM}(c_*) < \min\{v_{ZTM}, v_{MOM}\}$, where v_{MOM}, v_{ZTM} and $v_{PM}(\cdot)$ are the normalized asymptotic variances of $\hat{k}$ as defined in Eqs.* (3.1.2), (3.1.3) *and* (3.1.4) *for the MOM, ZTM and PM respectively.*

Proof. Let m and k be fixed and set $a = m/k$. Note that $0 < a, k, m < \infty$.

i) Inadmissability of MOM. A Taylor expansion of $v_{PM}(c)$ in the neighborhood of $c = 1$ gives

$$v_{PM}(c) = \frac{2k(k+1)(1+a)^2}{a^2} - \frac{8k(k+1)(1+a)^2}{3a}(1-c) + O\left((1-c)^2\right), \quad c \to 1.$$

In view of (3.1.2) this implies $v_{PM}(1) = v_{MOM}$. Additionally, the derivative of $v_{PM}(c)$ at $c = 1$ is

$$\left.\frac{\partial v_{PM}(c)}{\partial c}\right|_{c=1} = \frac{8k(k+1)(1+a)^2}{3a},$$

which is strictly positive for all m and k. Hence, there always exists c' such that $0 < c' < 1$ and $v_{PM}(c') < v_{PM}(1) = v_{MOM}$.

ii) Inadmissability of ZTM. A Taylor expansion of $v_{PM}(c)$ in the neighborhood of $c = 0$ gives

$$v_{PM}(c) = v_{PM}(0) + c\left.\frac{\partial v_{PM}(c)}{\partial c}\right|_{c=0} + O\left(c^2\right), \quad c \to 0. \tag{3.1.6}$$

Equation (3.1.3) and (3.1.4) directly imply that $v_{PM}(0) = v_{ZTM}$. The derivative of $v_{PM}(c)$ at $c = 0$ can be written as

$$\begin{aligned}\left.\frac{\partial v_{PM}(c)}{\partial c}\right|_{c=0} &= -2a(a+1)\frac{(1+a)^k\left[k(1+a)\log(1+a)-a(k+1)\right]+a(k+1)-k\log(1+a)}{((1+a)\log(1+a)-a)^3}\\ &= -\frac{2a(a+1)}{[h(a)]^3\log(1+a)}\sum_{j=2}^{\infty}\frac{[k\log(1+a)]^j}{j!}h_j(a),\end{aligned} \tag{3.1.7}$$

where $h(a) = (1+a)\log(1+a) - a$ and $h_j(a) = [(j-1)a + j]\log(1+a) - aj$. The infinite series in (3.1.7) is derived by a Taylor expansion of $(1+a)^k$ (at $k = 0$) in the numerator. Lemma 3.1.1 implies that $h(a) > 0$ and $h_j(a) > 0$ for all $a > 0$ and all $j \geqslant 2$. All the terms in the infinite series in (3.1.7) are therefore positive for all k and a. This implies first, that the series is absolutely convergent for all k and a and second, that the derivative (3.1.7) is negative for all k and a. Hence, there always exists c'' such that $0 < c'' < 1$ and $v_{PM}(c'') < v_{PM}(0) = v_{ZTM}$.

Let c' and c'' be particular values as above. Define

$$c_* = \begin{cases} c' & \text{if } v_{ZTM} \geq v_{MOM} \\ c'' & \text{if } v_{ZTM} < v_{MOM} \end{cases} \tag{3.1.8}$$

then we obviously have $v_{PM}(c_*) < \min\{v_{ZTM}, v_{MOM}\}$. □

Lemma 3.1.1. *The functions*

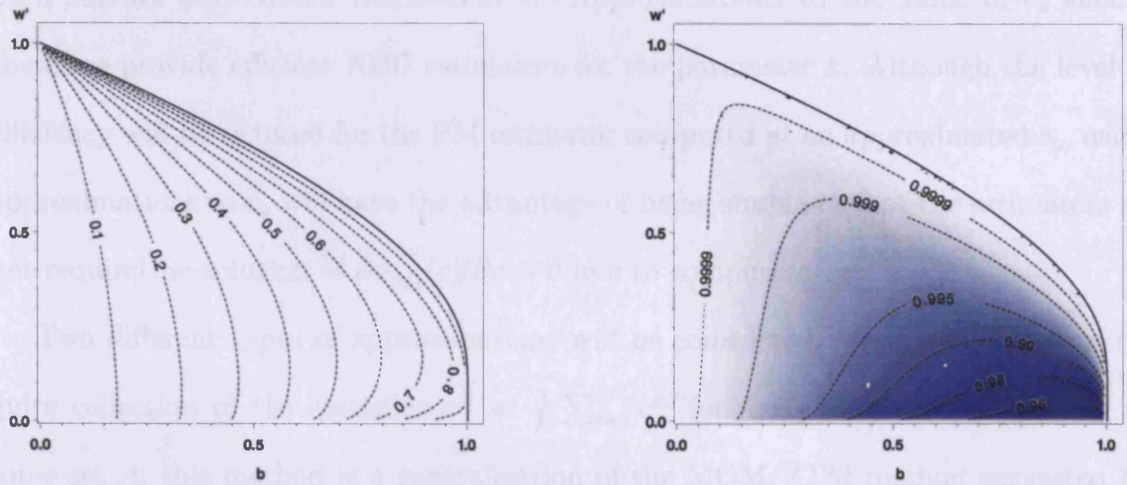
$$h(a) = (1+a)\log(1+a) - a \quad \text{and} \quad h_j(a) = [(j-1)a + j]\log(1+a) - aj$$

are positive for all $a > 0$ and $j \geqslant 2$.

Proof. We have $h(0) = 0$ and $h'(a) = \log(1+a) > 0$ for all $a > 0$, implying that $h(a) > 0$ for all $a > 0$. Similarly, for all $j \geqslant 2$ we have $h_j(0) = 0$ and $h'_j(a) = (j-2)\log(1+a) + h(a)/(1+a) > 0$ for all $a > 0$, implying that $h_j(a) > 0$ for all $a > 0$ and $j \geqslant 2$. □

Efficiency of the PM(c_o) estimator

Consider the PM estimator computed at $c_o = \operatorname{argmin}_{c\in(0,1)} v_{PM}(c)$. It is difficult to express c_o analytically since the solution, with respect to c, of the equation $\partial v_{PM}(c)/\partial c = 0$ is intractable. Fig. 3.3 showed, for various values of the NBD parameters (m, k), that the function $v_{PM}(c)$ is a convex function in c. If the function $v_{PM}(c)$ is a convex function in c for all parameter values, then the equation $\partial v_{PM}(c)/\partial c = 0$ may be solved numerically to compute the optimum value of c.



(a) Contour levels for c_o (b) $v_{ML}/v_{PM}(c_o)$

Figure 3.4: Optimum values of the power method parameter c and efficiency of the power method estimator computed at optimum c for all admissible values of NBD parameters.

Fig. 3.4(a) shows contour levels of c_o within the NBD parameter space. The contour levels are plotted from values of c_o obtained by numerical minimization of $v_{PM}(c)$ for a fine grid of values of b and $w' < -b/\log(1-b)$. The fact that the contour lines vary smoothly over the parameter space indicate no erratic jumps in the value of c_o and therefore that the function $v_{PM}(c)$ may well have only one minimum.

Note that $c \to 0$ as $b \to 0$, in this case the probability of observing a zero event tends to one; the ZTM is therefore asymptotically efficient when the NBD is degenerate. Furthermore, $c \to 1$ as $k \to \infty$; the MOM is therefore asymptotically efficient when the NBD converges to the Poisson distribution. The asymptotic efficiency $v_{ML}/v_{PM}(c_o)$ is shown in Fig. 3.4(b). The PM(c_o) estimator achieves an efficiency of greater than 0.96 for the majority of the (b, w')-parameter space.

3.1.3 Approximating optimum c

Fig. 3.3 showed that there is a range of values c_* such that $v_{PM}(c_*) < \min\{v_{ZTM}, v_{MOM}\}$. Moreover, for the parameter values shown in Fig. 3.3, the function $v_{PM}(c)$ appears to be a smooth and convex function in c. Approximations to the value of c_o should therefore provide efficient NBD estimators for the parameter k. Although the level of efficiency will be reduced for the PM estimator computed at an approximated c_o, using approximations to c_o will have the advantage of being simple in that the estimators do not require the solution of $\partial v_{PM}(c)/\partial c = 0$ in c to compute c_o.

Two different types of approximations will be considered. Set approximations require collection of the statistics $\widehat{c^X} = \frac{1}{N}\sum_{i=1}^{N} c^{x_i}$ for fixed values of c belonging to some set A; this method is a generalization of the MOM/ZTM method suggested by Anscombe (1950). Alternatively, approximations of c_o can be obtained using regression methods.

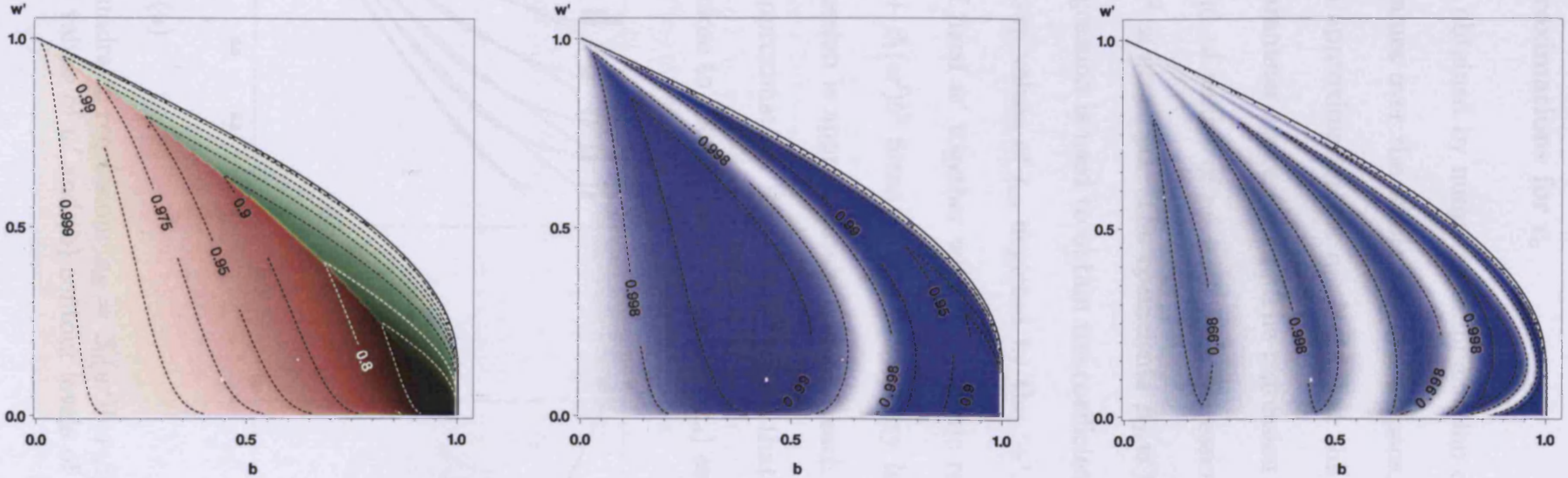
Set approximations for c_o

In the computation of the combined MOM/ZTM estimator for k, the more efficient estimator amongst the MOM and ZTM estimator is chosen. The MOM/ZTM estimator can therefore be thought of as a PM estimator computed at an approximated c_o by choosing the value of $c \in \{0, 1\}$ such that $v_{PM}(c) = \min\{v_{PM}(0), v_{PM}(1)\}$.

A generalization of the MOM/ZTM estimator is achieved by extending the set of possible values of c such that $v_{PM}(c)$ is minimized. Denote by A the set of values of c that will be used to approximate c_o, then the value of c_o is approximated by c_A such that $c_A = \text{argmin}_{c\in A} v_{PM}(c)$. Fig. 3.5 shows the asymptotic efficiency of $v_{PM}(c_o)/v_{PM}(c_A)$, where $c_A = \text{argmin}_{c\in A} v_{PM}(c)$, for the combined MOM/ZTM estimator ($A = \{0, 1\}$) and two different sets $A = \{0, \frac{1}{2}, 1\}$ and $A = \{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$.

Fig. 3.5(a) shows that the asymptotic normalized efficiency of the MOM/ZTM estimator relative to the PM(c_o) estimator lies in the interval $(0, 1)$ for all parameter values of the NBD. This indicates that the PM(c_o) estimator is always more efficient than the MOM/ZTM estimator. Areas of red shading represents regions where $v_{ZTM} < v_{MOM}$ and areas of green shading represents regions where $v_{MOM} < v_{ZTM}$. Asymptotically the MOM/ZTM estimator becomes efficient in the following sense: if either m or $a = m/k$ is fixed then $v_{ML}/v_{MOM} \to 1$ as $k \to \infty$ and $v_{ML}/v_{ZTM} \to 1$ as $k \to 0$. The combined MOM/ZTM estimator can have efficiency levels as low as 0.7 and below when compared to the PM(c_o) estimator.

Fig. 3.5(b) and Fig. 3.5(c) show that extending the set A improves the efficiency of the estimation method. These estimators are just as simple as the combined MOM/ZTM estimator, apart from the fact that the collection of extra statistics $\widehat{c^X} = \frac{1}{N}\sum_{i=1}^{N} c^{x_i}$ for all $c \in A$ is required. It is clear that greater efficiency levels can be obtained by further extending the set A at the expense of computing additional statistics.



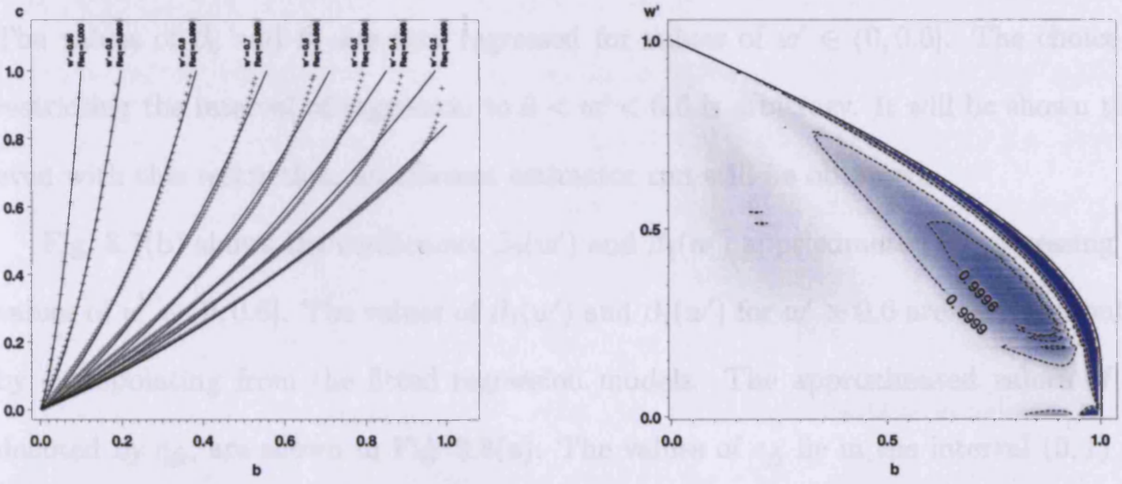
(a) (b) (c)

Figure 3.5: Contour levels for $v_{PM}(c_o)/v_{PM}(c_A)$ where $c_A = \text{argmin}_{c \in A} v_{PM}(c)$ for the sets (a) $A = \{0, 1\}$, (b) $A = \left\{0, \frac{1}{2}, 1\right\}$ and (c) $A = \left\{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\right\}$.

Regression approximations for c_o

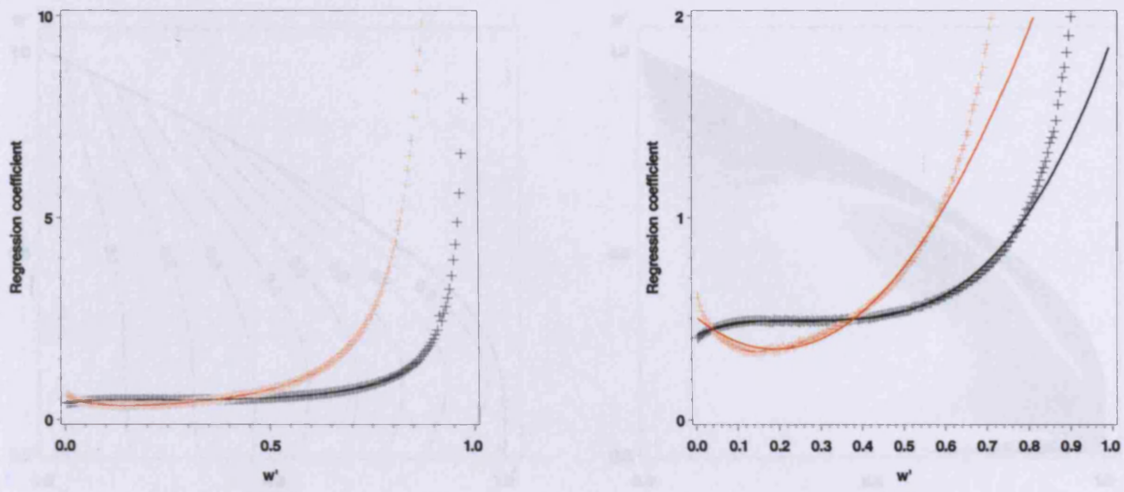
The value of c_o is obtained by numerical minimization of $v_{PM}(c)$. Using knowledge of these numerical values over the whole parameter space, regression techniques can be used to obtain an approximation for c_o. Note that the values of c_o are a continuous function of the parameter space (b, w'). The regression approximation is obtained as follows. For a grid of values of $w' \in (0, 1)$ a regression equation of the form $c_B = \beta_0(w')b + \beta_1(w')b^2$ is obtained. The coefficients $\beta_0(w')$ and $\beta_1(w')$ are then plotted against w' and regression is used to obtain the coefficients $\beta_0(w')$ and $\beta_1(w')$.

Fig. 3.6(a) shows values of c_o, depicted by the '+' symbol, plotted against b for different values of fixed w' together with a quadratic regression approximation given by $c_B = \beta_0(w')b + \beta_1(w')b^2$ fitted using the ordinary least squares method. The R^2 value for the regression is approximately 0.998 for each fixed value of w' indicating a good fit for the approximation. Fig. 3.6(b) shows that the efficiency of the PM(c_B) estimator is very close to the efficiency of the PM(c_o) estimator.



(a) (b)

Figure 3.6: (a) Quadratic regression: $c_B = \beta_0(w')b + \beta_1(w')b^2$ to approximate values of c_o (+) for fixed values of w' and (b) contour levels of $v_{PM}(c_o)/v_{PM}(c_B)$.



(a) (b)

$+++\ \beta_0$ $+++\ \beta_1$ —— $\hat{\beta}_0$ —— $\hat{\beta}_1$

Figure 3.7: (a) β_0 and β_1 versus w' where β_0 and β_1 are the regression coefficients in $c_B = \beta_0(w')b + \beta_1(w')b^2$. (b) Values of β_0 and β_1 approximated by $\hat{\beta}_0 = 0.4206 + 0.8065w' - 2.9790w'^2 + 3.644w'^3$ and $\hat{\beta}_1 = 0.509 - 1.6594w' + 4.3075w'^2$ respectively.

Fig. 3.7 shows the regression coefficients $\beta_0(w')$ and $\beta_1(w')$ plotted against w'. Both regression coefficients are continuous functions of w' for $w' \in (0, 1)$. For $0 < w' < 0.6$ the values of β_0 and β_1 behave like cubic and quadratic functions of w' respectively. For $0.6 < w' < 1$ the values of β_0 and β_1 increase at an exponential rate as w' increases. The values of β_0 and β_1 are now regressed for values of $w' \in (0, 0.6]$. The choice of restricting the interval of regression to $0 < w' < 0.6$ is arbitrary. It will be shown that even with this restriction, an efficient estimator can still be obtained.

Fig. 3.7(b) shows the coefficients $\beta_0(w')$ and $\beta_1(w')$ approximated by regressing on values of $w' \in (0, 0.6]$. The values of $\beta_0(w')$ and $\beta_1(w')$ for $w' > 0.6$ are then computed by extrapolating from the fitted regression models. The approximated values of c_o, denoted by $c_{\tilde{B}}$, are shown in Fig. 3.8(a). The values of $c_{\tilde{B}}$ lie in the interval $(0, 1)$ for all parameter values of the NBD. The efficiency of the PM($c_{\tilde{B}}$) estimator for k with respect to the PM(c_o) estimator is shown in Fig. 3.8(b). The efficiency is at least 0.99 (white contour) for the majority of the (b, w')-parameter space.

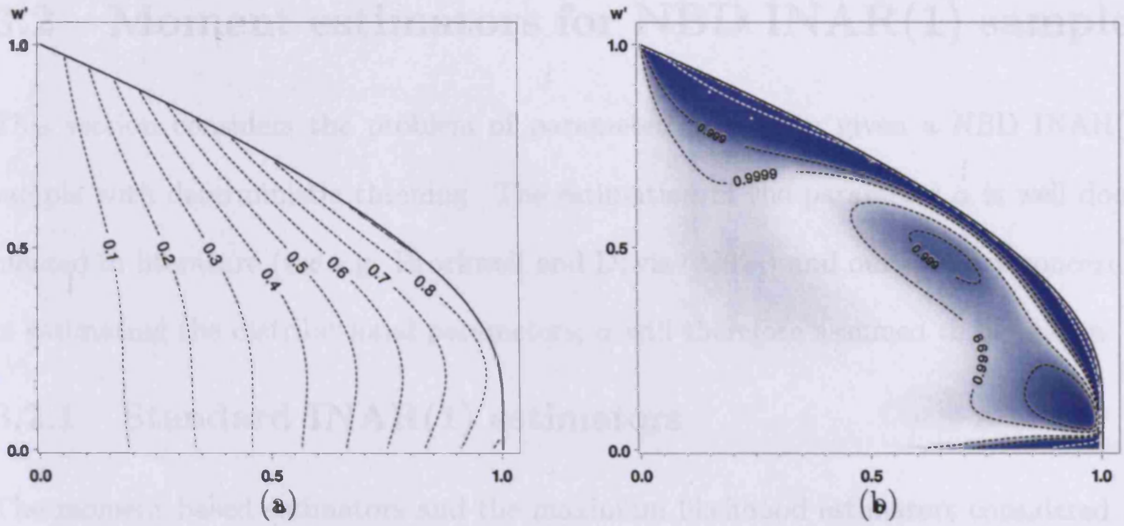


Figure 3.8: (a) $c_{\tilde{B}}$ and (b) $v_{PM}(c_o)/v_{PM}(c_{\tilde{B}})$ where $c_{\tilde{B}} = \hat{\beta}_0 b + \hat{\beta}_1 b^2$ with $\hat{\beta}_0 = 0.4206 + 0.8065w' - 2.9790w'^2 + 3.644w'^3$ and $\hat{\beta}_1 = 0.509 - 1.6594w' + 4.3075w'^2$.

Sensitivity of efficiency to changes in c

All of the results shown in Section 3.1.3 indicate that efficient NBD estimators for the parameter k can be obtained by using the PM estimator computed at a suitable value of c. The PM(c_o) estimator, where $c_o = \text{argmin}_{c\in(0,1)} v_{PM}(c)$, is required to obtain the most efficient PM estimator. The value of c_o can be approximated extremely well by using regression techniques. Using the regression technique mentioned in Section 3.1.3 a negligible loss of efficiency is seen for the majority of NBD parameter values. Finally, a much simpler approximation can be made by the use of set approximations to c_o. The loss of efficiency for set approximations depends on the set A.

The approximations have the advantage that they do not require the solution of $\partial v_{PM}(c)/\partial c = 0$ in c to compute c_o. The fact that different approximations of c_o exist, to give highly efficient PM estimators relative to the PM(c_o) estimator, show the insensitive nature of the PM(c) estimator to small changes in c. This insensitive nature is important when implementing the PM estimator to efficiently estimate k in the case, as in practice, when the NBD parameters are unknown (see Section 3.3.1).

3.2 Moment estimators for NBD INAR(1) samples

This section considers the problem of parameter estimation given a NBD INAR(1) sample with deterministic thinning. The estimation of the parameter α is well documented in literature (see e.g. Brockwell and Davis (2002)) and our primary concern is in estimating the distributional parameters; α will therefore assumed to be known.

3.2.1 Standard INAR(1) estimators

The moment based estimators and the maximum likelihood estimators considered by Al-Osh and Alzaid (1987) use the distribution of the errors for the INAR(1) process. Although the probability generating function of the errors for the NBD INAR(1) process is known (see McKenzie (1986)), the distribution was never written down explicitly.

Proposition 3.2.1. *Let X_t be a NBD INAR(1) process with thinning parameter α and marginal distribution NBD(m,k), then ε_t has a negative-binomial geometric distribution, NBD-G$(k, k/(k+m\alpha), \alpha)$ with probability mass function*

$$\mathbb{P}\{\varepsilon_t = x\} = \sum_{j=0}^{\infty} \binom{j+x-1}{x} \left(\frac{k}{k+m\alpha}\right)^j \left(\frac{m\alpha}{k+m\alpha}\right)^x \binom{k+j-1}{j} \alpha^k (1-\alpha)^j, \qquad x = 0, 1 \dots .$$

Proof. Note that the generating function of the errors can be written as

$$\begin{aligned} G_\varepsilon(c) &= \left(\frac{k+m(1-c)}{k+m\alpha(1-c)}\right)^{-k} = \alpha^k \left(\frac{k\alpha+m\alpha(1-c)}{k+m\alpha(1-c)}\right)^{-k} = \alpha^k \left(1 - \frac{k(1-\alpha)}{k+m\alpha(1-c)}\right)^{-k} \\ &= \alpha^k \left(1 - (1-\alpha)\left(\frac{k}{k+m\alpha}\right)\left(1 - \frac{m\alpha}{k+m\alpha}c\right)^{-1}\right)^{-k}. \end{aligned} \tag{3.2.1}$$

This is of the form of the generating function of the negative-binomial geometric NBD-G(μ, ν, θ) distribution (see Wimmer and Altmann (1990, pp. 459–460)) given by

$$G(c) = \theta^\mu \left(1 - (1-\theta)\nu(1-(1-\nu)c)^{-1}\right)^{-\mu}.$$

Here $0 < \nu \leqslant 1$, $0 < \theta \leqslant 1$ and $\mu > 0$. □

The negative-binomial geometric distribution is a compound Poisson distribution in direct agreement with the result by McKenzie (1986) (see Eq. 2.3.13). Note that the generating function of ε_t can be written in the form $G_\varepsilon(c) = \exp\{\lambda\,(G_\phi(c)-1)\}$ with

$$\lambda = -k\log\alpha \quad \text{and} \quad G_\phi(c) = (\log\alpha)^{-1}\log\left(1-(1-\alpha)\left(\frac{k}{k+m\alpha}\right)\left(1-\frac{m\alpha}{k+m\alpha}c\right)^{-1}\right).$$

Here $G_\phi(c)$ is the generating function of the logarithmic-geometric distribution (see Wimmer and Altmann (1990, pp. 388–389)) with

$$\mathbb{P}\{\phi = x\} = \sum_{j=1}^{\infty}\binom{j+x-1}{x}\left(\frac{k}{k+m\alpha}\right)^j\left(\frac{m\alpha}{k+m\alpha}\right)^x\frac{(1-\alpha)^k}{-k\log\alpha}.$$

Generating random variables from the negative-binomial geometric NBD-G(μ,ν,θ) is made simple by the fact that the distribution is equivalent to the NBD$_p(\nu, r)$ distribution where r is itself a random variable with a NBD$_p(\theta, m)$ distribution. The errors therefore have the distribution

$$\varepsilon_t \stackrel{\mathcal{D}}{=} \text{NBD-G}\left(k, \frac{k}{k+m\alpha}, \alpha\right) \stackrel{\mathcal{D}}{=} \text{NBD}_p\left(\frac{k}{k+m\alpha}, r\right)\bigwedge_r \text{NBD}_p(\alpha, k). \tag{3.2.2}$$

The maximum likelihood estimator. Let $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ be an observed sample from an INAR(1) process and let Θ be the set of parameters for the INAR(1) process then the likelihood function is

$$L(\mathbf{x};\ \Theta) = \mathbb{P}(X_1 = x_1)\prod_{t=2}^{N}\sum_{r=0}^{\min(x_t,x_{t-1})}\binom{x_{t-1}}{r}\alpha^r(1-\alpha)^{x_{t-1}-r}\mathbb{P}(\varepsilon_t = x_t - r). \tag{3.2.3}$$

The likelihood function for the NBD INAR(1) process requires the distribution of errors, i.e. the negative binomial geometric distribution. Since the likelihood function is complex, it is difficult to obtain the Fisher information matrix and hence analytically analyze the efficiency of moment based estimators with respect to maximum likelihood estimators.

Nevertheless, parameter estimates from an NBD INAR(1) sample can be obtained by maximizing the likelihood function using global optimization algorithms. Simulation results will therefore provide some indication as to the efficiency of moment based methods of estimation with respect to maximum likelihood estimation. These results are shown in Section 3.3.

Standard INAR(1) moment based estimators. For the standard INAR(1) moment based estimators considered by Al-Osh and Alzaid (1987), the thinning parameter α is estimated from the autocorrelation function of the INAR(1) process. (The same method is applied in the case of estimating the corresponding parameter for continuous AR(1) processes.) Since the autocorrelation function of the INAR(1) process is identical to that of the AR(1) process, the problem of estimating α is well documented in many textbooks (see e.g. (Brockwell & Davis, 2002)). Once α is estimated by $\hat{\alpha}$, Al-Osh and Alzaid obtain a sequence of estimators $\hat{\varepsilon}_t$ using the equation $\hat{\varepsilon}_t = x_t - \hat{\alpha}x_{t-1}$. Note that the $\hat{\varepsilon}_t$ are no longer integer-valued although the distribution of ε_t is discrete. Standard moment based estimation methods are then used to estimate the parameters of the marginal distribution of the error process. The distribution of the errors for the NBD INAR(1) process is not simple and this makes inference about the estimators of (m, k) difficult.

Since the INAR(1) process is a stationary and ergodic process (see e.g. Du and Li (1991)), the expected value of the sample moments for an observed realization are equivalent to the expected value of the sample moments of the stationary distribution. It therefore seems reasonable to use the moments of the observed realization to estimate the distributional parameters of the process; in this way the method will not need to use the complex structure of the distribution of the innovation process. Moreover, the power method estimators may be used to obtain efficient estimators.

3.2.2 Efficiency of the power method estimator

Since the INAR(1) process is a stationary and ergodic process, the expected value of the sample moments for an observed realization are equivalent to the expected value of the sample moments of the stationary distribution. An unbiased estimator for m is therefore $\hat{m} = \bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t$. The PM estimator $\hat{k}_{PM}$ for the shape parameter of the NBD distribution is computed by solving, in z, the equation $\widehat{c^X} = \frac{1}{N}\sum_{t=1}^{N} c^{x_t} = \left(1 + \frac{\bar{x}(1-c)}{z}\right)^{-z}$.

Although computing moment based estimators for a NBD INAR(1) process and an i.i.d. NBD sample are identical, the fact that the values of x_t are correlated for INAR(1) samples implies that the covariance matrices of the estimators of (m, k) are different. The asymptotic distribution of the estimators $(\hat{m}, \hat{k}_{PM})$ can be derived by using a multivariate version of the so-called δ-method (see e.g. Serfling (1980)). Using the δ-method if $\left(\bar{x}, \widehat{c^X}\right)$ is asymptotically normally distributed then the estimators $\left(\hat{m}, \hat{k}_{PM}\right)$ are also asymptotically normally distributed. In this section the asymptotic covariance matrix for the statistics $\left(\bar{x}, \widehat{c^X}\right)$ is derived. The covariance matrix of $\left(\hat{m}, \hat{k}_{PM}\right)$ is consequently obtained.

Theorem 3.2.2. (Savani & Zhigljavsky, 2006) *Let $\{x_t;\ t = 1, 2, \ldots, N\}$ be a sample realization from an INAR(1) process X_t with stationary distribution π. Let $\boldsymbol{f} = \left(x, c^X\right)^T$, $\bar{\boldsymbol{f}} = \left(\bar{x}, \widehat{c^X}\right)^T$ with $\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t$ and $\widehat{c^X} = \frac{1}{N}\sum_{t=1}^{N} c^{x_t}$, with $c > 0$ and $c \neq 1$. Then $\bar{\boldsymbol{f}}$ has an asymptotic normal distribution given by $\lim_{N\to\infty} \sqrt{N}(\bar{\boldsymbol{f}} - \mathbb{E}\boldsymbol{f}) \sim \mathcal{N}(0, \mathbb{D}\boldsymbol{f})$ with covariance matrix*

$$\mathbb{D}\boldsymbol{f} = \mathbb{E}(\boldsymbol{f} - \mathbb{E}\boldsymbol{f})(\boldsymbol{f} - \mathbb{E}\boldsymbol{f})^T = \begin{pmatrix} V_{\bar{X}} & C_{\bar{X},\widehat{c^X}} \\ C_{\bar{X},\widehat{c^X}} & V_{\widehat{c^X}} \end{pmatrix}. \tag{3.2.4}$$

Here

$$V_{\bar{X}} = \lim_{N\to\infty} N \text{ Var}\left(\bar{X}\right) = \left(\frac{1+\alpha}{1-\alpha}\right) \text{Var}[X_\pi], \tag{3.2.5}$$

$$\begin{aligned} V_{\widehat{c^X}} &= \lim_{N\to\infty} N \text{ Var}\left(\widehat{c^X}\right) \\ &= \text{Var}\left(c^{X_\pi}\right) + 2\lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{G_{X_\pi}\left(c\left[1-\alpha^r+\alpha^r c\right]\right)G_\varepsilon(c;\alpha^r) - G^2_{X_\pi}(c)\right\}, \end{aligned} \tag{3.2.6}$$

$$\begin{aligned} C_{\bar{X},\widehat{c^X}} &= \lim_{N\to\infty} N \text{ Cov}\left(\bar{X},\widehat{c^X}\right) = \text{Cov}\left(X_\pi c^{X_\pi}\right) \\ &+ \lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{E\left[X_\pi\left(1-\alpha^r+\alpha^r c\right)^{X_\pi}\right]G_\varepsilon(c;\alpha^r) - E[X_\pi]G_{X_\pi}(c)\right\} \\ &+ \lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{G_{X_\pi}\left(c\left[1-\alpha^r+\alpha^r c\right]\right) - \alpha^r E[X_\pi]G_{X_\pi}(c)\right\}. \end{aligned} \tag{3.2.7}$$

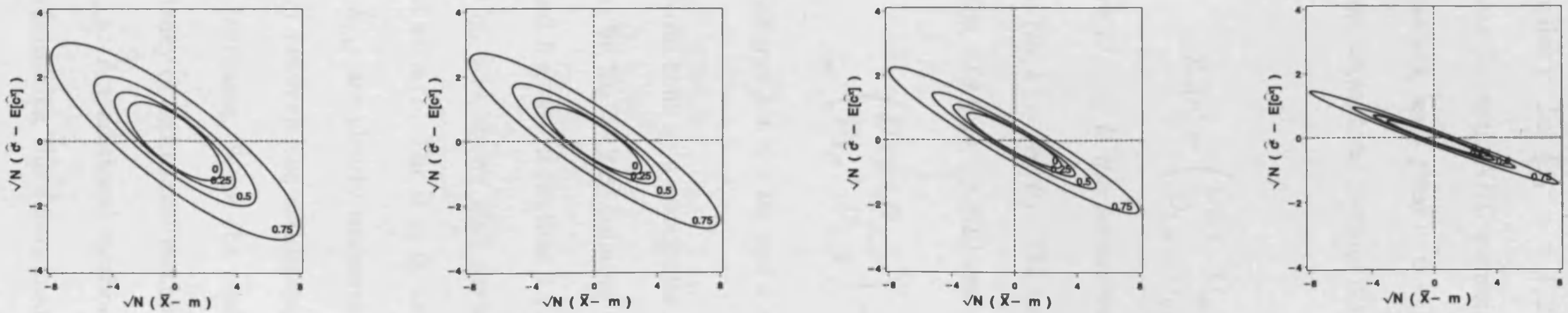
Proof. See Appendix A.1. □

Note that the asymptotic distribution of $\bar{\boldsymbol{f}} = \left(\bar{x},\widehat{c^X}\right)^T$ derived in Theorem 3.2.2 holds for any INAR(1) process and not just the NBD INAR(1) process. Fig. 3.9 shows 95% asymptotic bivariate normal confidence ellipses for $E\bar{\boldsymbol{f}}$, centered at zero, given by the equation

$$\left(\bar{\boldsymbol{f}} - E\bar{\boldsymbol{f}}\right)^T D_c^{-1}\left(\bar{\boldsymbol{f}} - E\bar{\boldsymbol{f}}\right) \leqslant \chi_{0.95}(2) \simeq 5.99\,.$$

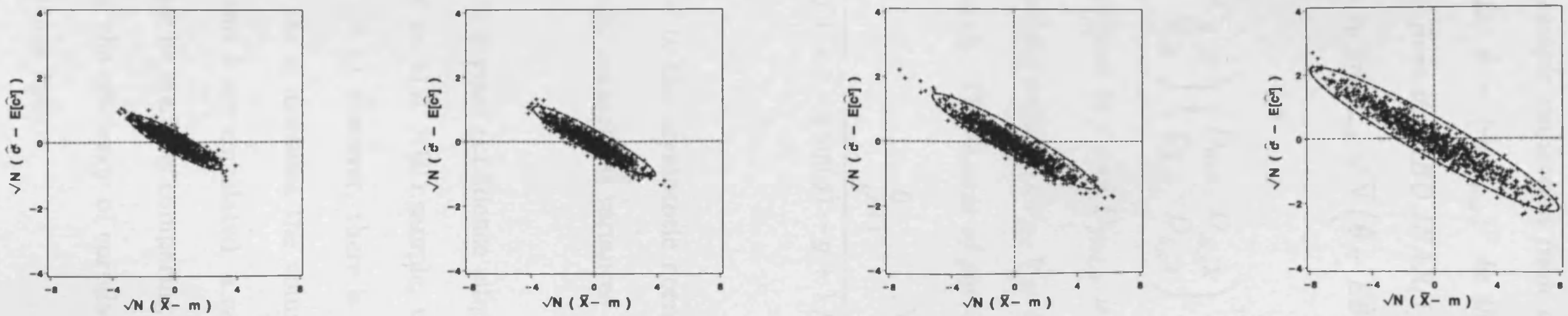
As α increases the correlation between the statistics $\bar{x}$ and $\widehat{c^X}$ clearly increases irrespective of the value of the PM parameter c.

Fig. 3.10 shows estimates $\bar{\boldsymbol{f}} - E\bar{\boldsymbol{f}}$ obtained from 1000 simulations from a NBD INAR(1) process together with corresponding 95% asymptotic bivariate normal confidence ellipses. The parameters used for the NBD INAR(1) process are $m = 1$, $k = 2$, $N = 1000$, $\alpha \in \{0, 0.25, 0.5, 0.75\}$ and the PM estimator $\widehat{c^X}$ is computed using the value $c = 0.5$.



$c = 0$ $\quad$ $c = 0.25$ $\quad$ $c = 0.5$ $\quad$ $c = 0.75$

Figure 3.9: 95% asymptotic bivariate normal confidence ellipses for $E\bar{\boldsymbol{f}}$ ($\bar{\boldsymbol{f}} = (\bar{x}, \widehat{c^X})^T$), centered at zero, for a NBD INAR(1) sample with $m = 1$, $k = 2$, $\alpha \in \{0, 0.25, 0.5, 0.75\}$ and $c \in \{0, 0.25, 0.5, 0.75\}$.



$\alpha = 0$ $\quad$ $\alpha = 0.25$ $\quad$ $\alpha = 0.5$ $\quad$ $\alpha = 0.75$

Figure 3.10: 1000 simulated $\bar{\boldsymbol{f}} - E\bar{\boldsymbol{f}}$ ($\bar{\boldsymbol{f}} = (\bar{x}, \widehat{c^X})^T$) with 95% asymptotic bivariate normal confidence ellipses for a NBD INAR(1) sample with $m = 1$, $k = 2$, $N = 1000$, $c = 0.5$ and $\alpha \in \{0, 0.25, 0.5, 0.75\}$.

Corollary. *Let $\{x_t;\ t = 1, 2, \ldots, N\}$ be a sample realization from a NBD INAR(1) process X_t with NBD parameters (m, k). Let $\hat{\boldsymbol{\theta}} = (\hat{m}, \hat{k}_{PM})^T$ be the power method estimators, with fixed c, $0 < c < 1$, obtained from the NBD INAR(1) sample, then $\hat{\boldsymbol{\theta}}$ has an asymptotic normal distribution given by $\lim_{N\to\infty} \sqrt{N}\left(\hat{\boldsymbol{\theta}} - E\hat{\boldsymbol{\theta}}\right) \sim \mathcal{N}(0, \Sigma_\alpha(c))$ with*

$$\Sigma_\alpha(c) = \begin{pmatrix} D_{\hat{m},\bar{x}} & D_{\hat{m},\widehat{c^X}} \\ D_{\hat{k},\bar{x}} & D_{\hat{k},\widehat{c^X}} \end{pmatrix} \begin{pmatrix} V_{\bar{X}} & C_{\bar{X},\widehat{c^X}} \\ C_{\bar{X},\widehat{c^X}} & V_{\widehat{c^X}} \end{pmatrix} \begin{pmatrix} D_{\hat{m},\bar{x}} & D_{\hat{m},\widehat{c^X}} \\ D_{\hat{k},\bar{x}} & D_{\hat{k},\widehat{c^X}} \end{pmatrix}^T. \tag{3.2.8}$$

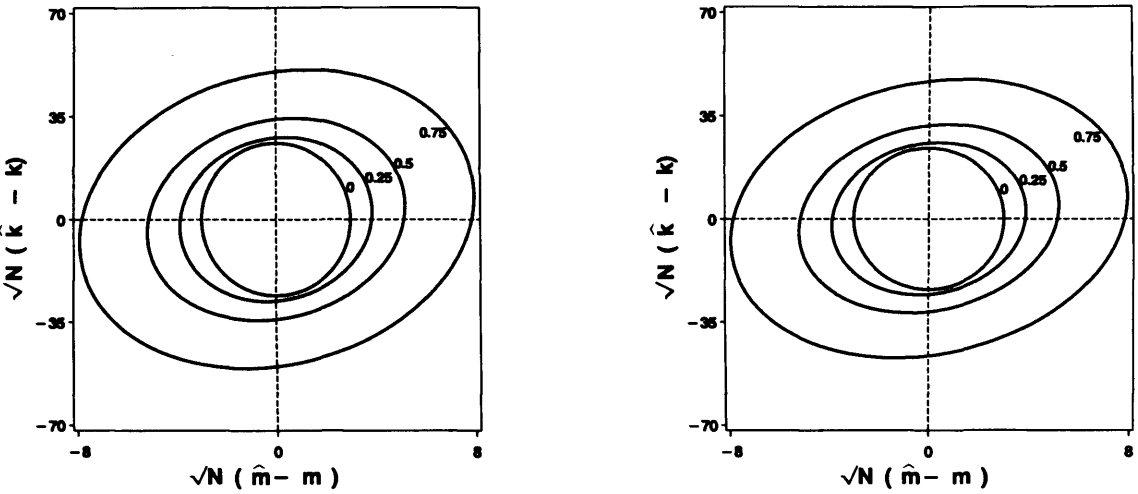
Here $D_{f(\nu),\nu}$ is the derivative of $f(\nu)$ with respect to ν and $D_{f(\nu),\nu}$ is evaluated at the point $(\hat{m}, \hat{k}) = (m, k)$. The asymptotic normalized variances $V_{\bar{X}}$, $V_{\widehat{c^X}}$ and $C_{\bar{X}}$ are given by Eqs. (3.2.5), (3.2.6) and (3.2.7) respectively. The matrix of partial derivatives is

$$\begin{pmatrix} D_{\hat{m},\bar{x}} & D_{\hat{m},\widehat{c^X}} \\ D_{\hat{k},\bar{x}} & D_{\hat{k},\widehat{c^X}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \dfrac{c-1}{g\log(g) - g + 1} & -\dfrac{g^{k+1}}{g\log(g) - g + 1} \end{pmatrix}$$

where $g = 1 + a - ac$ and $a = m/k$.

Note that as a particular case, when $\alpha = 0$, the asymptotic normalized variances given by Eq. (3.2.8) collapse to the asymptotic normalized variances of estimators for m and k given in Section 2.1.3.

Fig. 3.11 shows 95% asymptotic bivariate normal confidence ellipses for $E\hat{\boldsymbol{\theta}}$, centered at zero. For $\alpha = 0$, i.e. in the case of an i.i.d. NBD sample, the estimators $\hat{m}$ and $\hat{k}_{PM}$ are clearly uncorrelated. For $\alpha \in (0, 1)$, however, there is a positive correlation between the estimators $\hat{m}$ and $\hat{k}_{PM}$. As α increases the volume of the ellipse also increases. Since the estimators for m and k are correlated, a comparison of the efficiency of estimation methods may no longer be made by comparing just the variance of $\hat{k}_{PM}$. A traditional method for comparing the efficiency of correlated estimators is by minimizing the determinant of the covariance matrix.



$c = 0$ $c = 0.5$

Figure 3.11: 95% asymptotic bivariate normal confidence ellipses for $E\hat{\boldsymbol{\theta}}$ ($\hat{\boldsymbol{\theta}} = (\hat{m}, \hat{k}_{PM})$), centered at zero, for a NBD INAR(1) sample with $m = 1$, $k = 2$, $\alpha \in \{0, 0.25, 0.5, 0.75\}$.

In Fig. 3.11 there is no clear observable difference between the PM estimators computed at $c = 0$ and $c = 0.5$. This suggests the possibility that the volume of the ellipse may be insensitive to certain changes in the value of c. The insensitivity of the volume of the ellipse to changes in c implies the possibility of using simple approximations to c to obtain efficient PM estimators. Note that the estimator for m is identical for both PM estimators computed at $c = 0$ and $c = 0.5$ and therefore the ellipses are likely to be very similar.

As a more informative indicator of efficiency, Fig. 3.12 shows the determinant of the asymptotic normalized covariance matrix $\Sigma_\alpha(c)$ (see Eq. (3.2.8)) plotted against c for two different NBD parameter pairs $(m, k) \in \{(1, 0.5), (1, 2)\}$ and $\alpha \in \{0, 0.25, 0.5, 0.75\}$. For the NBD parameter pairs shown, the optimal values of c are never equal to 0 or 1. Moreover, there is an optimum c, denoted by c_α that minimizes $\det(\Sigma_\alpha(c))$. For the parameter pair $m = 1$ and $k = 2$ there is actually a difference between the values of the determinant for $c = 0$ and $c = 0.5$ which is not apparent in Fig. 3.11.

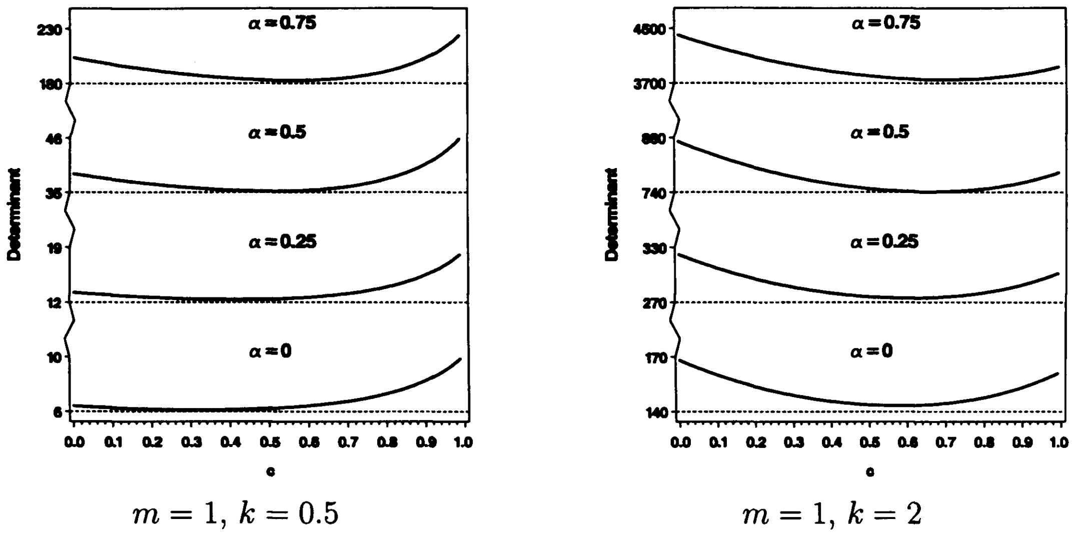


Figure 3.12: Determinant of the covariance matrix $\Sigma_\alpha(c)$, for $(\hat{m}, \hat{k}_{PM})$, versus c.

Fig. 3.13 shows contour levels of c_α, the value of $c \in (0,1)$ that minimizes $\det(\Sigma_\alpha(c))$, within the (b, w')-parameter space. The contour levels of c_α are similar as α increases highlighting the insensitive nature of c_α to changes in α. This is useful in that, if the value of α is unknown or not accounted for, then the loss in efficiency when using the standard PM(c_o) estimator as opposed to the PM(c_α) estimator will be small.

Fig. 3.13 also shows contour levels of efficiency, defined by $\det(\Sigma_0(c_0))/\det(\Sigma_\alpha(c_\alpha))$ for $\alpha \in \{0.25, 0.5, 0.75, 0.95\}$. The figure shows the loss in efficiency when using a NBD INAR(1) sample with thinning parameter α relative to an i.i.d. NBD sample. As α increases the efficiency of estimating (m, k), with respect to estimating from an i.i.d. sample using the PM(c_o) estimator, decreases.

More importantly, however, Fig. 3.13 shows the efficiency of estimating (m, k) using the PM(c_o) estimator with respect to the PM(c_α) estimator for a NBD INAR(1) sample with thinning parameter α. The efficiency is defined by $\det(\Sigma_\alpha(c_\alpha))/\det(\Sigma_\alpha(c_o))$ and shows the loss in efficiency in estimation when assuming an i.i.d. sample when in fact the sample is obtained from a NBD INAR(1) process. The loss of efficiency, even when $\alpha = 0.95$ is at most 10% for the majority of the NBD (b, w')-parameter space.

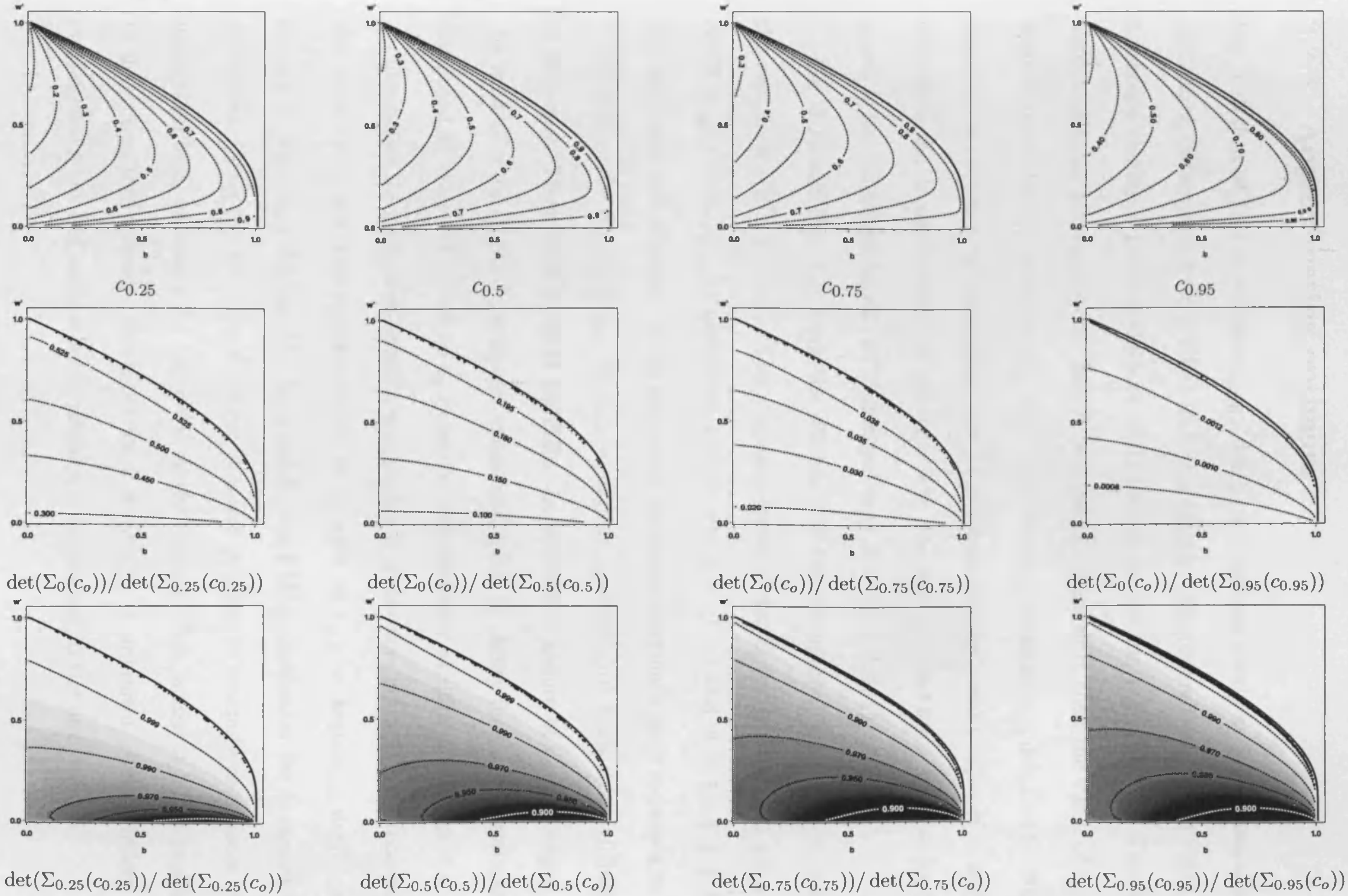


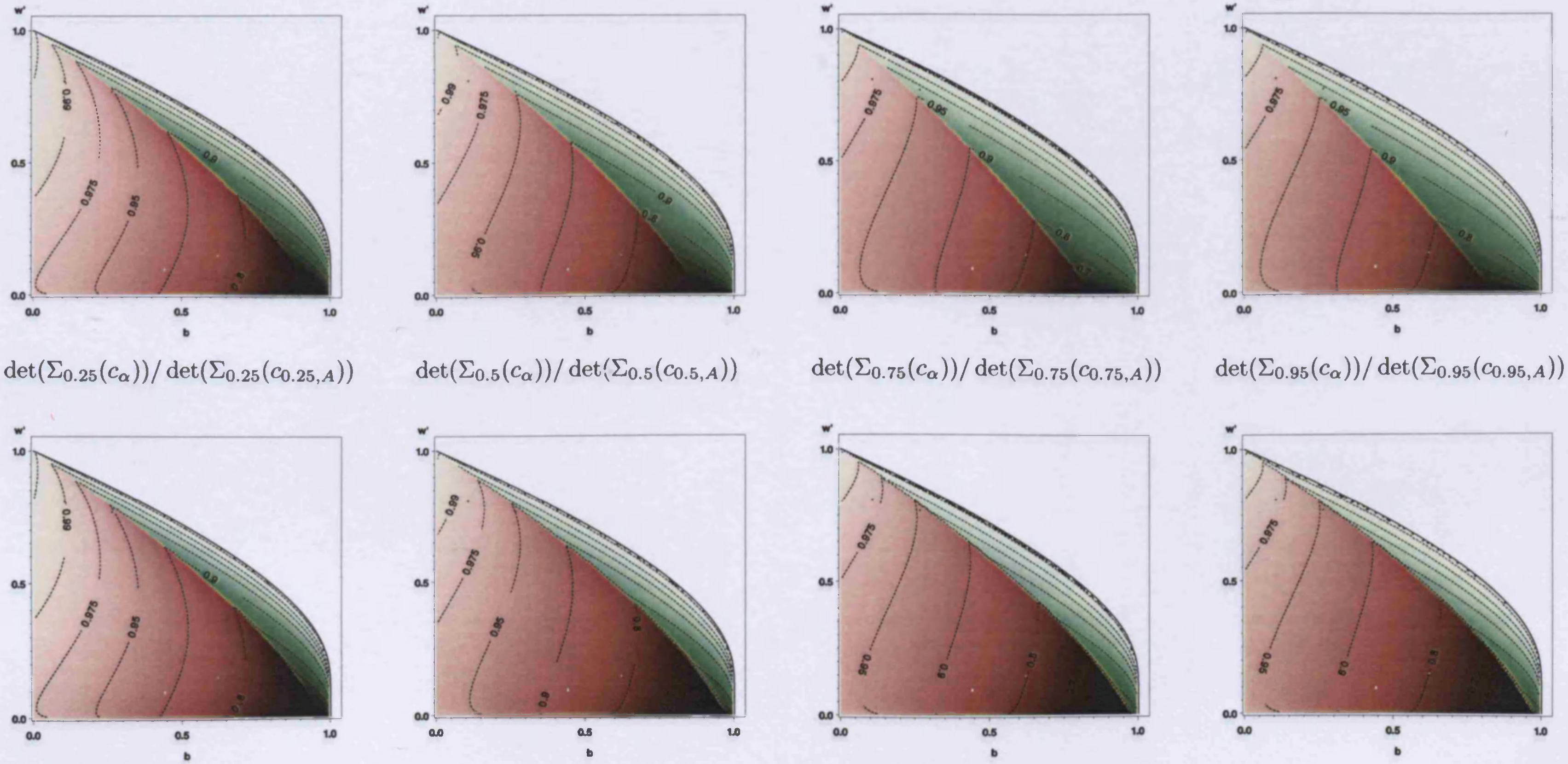
Figure 3.13: Contour levels of c_α, $\det(\Sigma_0(c_o))/\det(\Sigma_\alpha(c_\alpha))$ and $\det(\Sigma_\alpha(\alpha))/\det(\Sigma_\alpha(c_o))$

3.2.3 Approximating optimum c

Fig. 3.13 showed that $c_\alpha = \mathrm{argmin}_{\mathrm{c}\in(0,1)} \det(\Sigma_\alpha(\mathrm{c}))$ changes slowly as α increases when estimating parameters from a NBD INAR(1) sample. Approximations to the value of c_α should therefore provide efficient NBD estimators for the parameter k. Two set approximations are considered. Let A be a set of values of c, then the value of c_α is approximated by $c_{\alpha,A} = \mathrm{argmin}_{\mathrm{c}\in\mathrm{A}} \det(\Sigma_\alpha(\mathrm{c}))$ and $\tilde{c}_{\alpha,A} = \mathrm{argmin}_{\mathrm{c}\in\mathrm{A}} \det(\Sigma_0(\mathrm{c}))$. Note that $c_{\alpha,A}$ is obtained by minimizing the determinant for the correct value of α and $\tilde{c}$ minimizes the determinant in the case of $\alpha = 0$. The two sets used for A are $A = \{0, 1\}$, which is the combined MOM/ZTM method, and $A = \{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$.

Fig. 3.14 and Fig. 3.15 show the efficiency of two set approximations to c_α when the sample is NBD INAR(1). The efficiency levels of $\det(\Sigma_\alpha(c_\alpha))/\det(\Sigma_\alpha(c_{\alpha,A}))$ and $\det(\Sigma_\alpha(c_\alpha))/\det(\Sigma_\alpha(\tilde{c}_{\alpha,A}))$ are shown for the sets $A = \{0, 1\}$ and $A = \{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$. For both sets, the efficiency of the estimator decreases marginally as α increases to 1. Additionally, the choice of $c_{\alpha,A}$ by $\tilde{c}_{\alpha,A} = \mathrm{argmin}_{\mathrm{c}\in\mathrm{A}} \det(\Sigma_0(\mathrm{c}))$ leads to small losses in efficiency. These figures again highlight the insensitive nature of c_α to changes in the value of α for all NBD parameter values and thus the ability for the PM estimator computed at values of c close to c_α to retain high efficiency at different values of α.

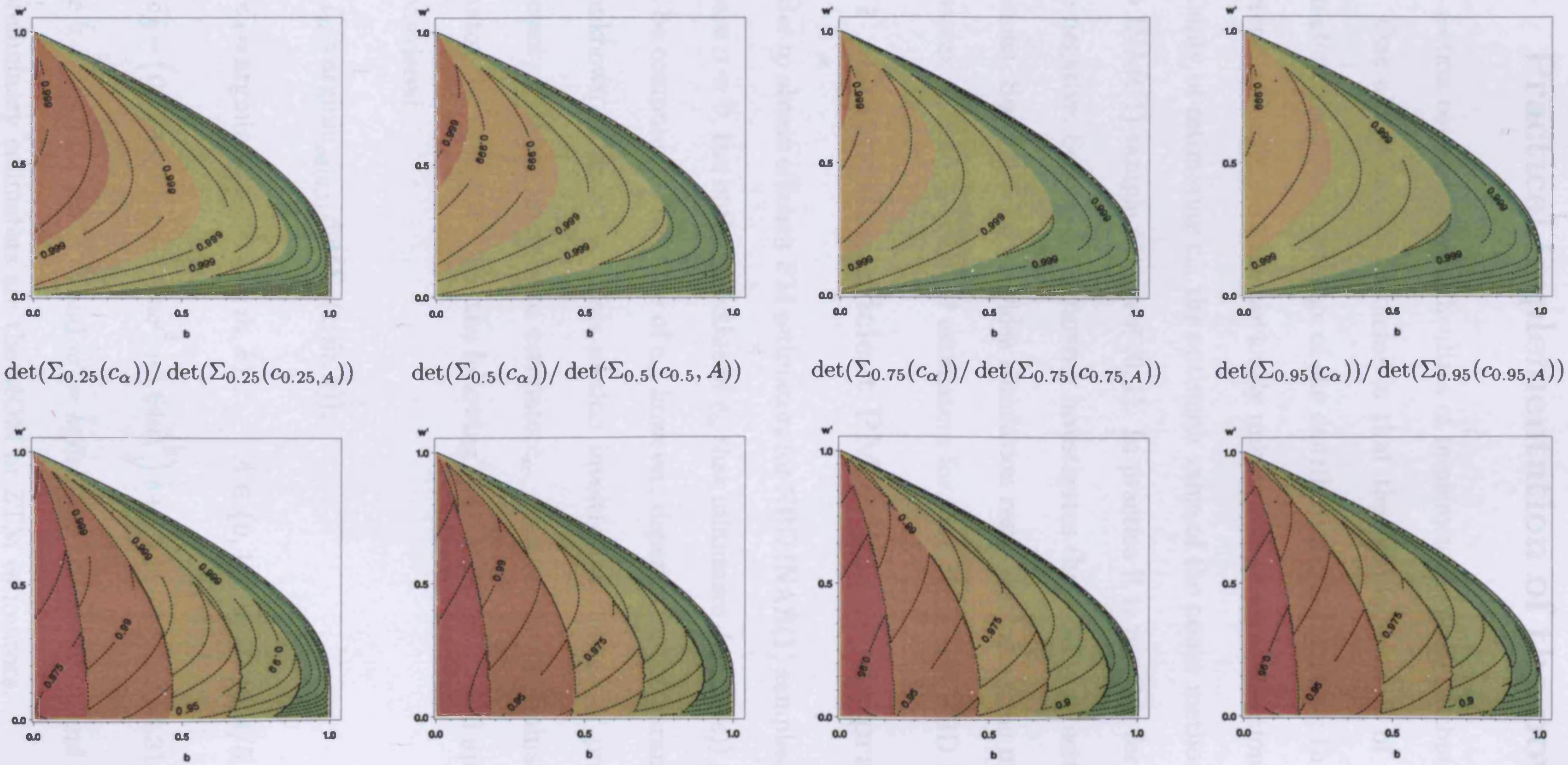
In practice the parameter values are unknown; it is therefore not possible to compute the value of c_α nor any approximation to c_α such as $c_{\alpha,A} = \mathrm{argmin}_{\mathrm{c}\in\mathrm{A}} \det(\Sigma_\alpha(\mathrm{c}))$ or $\tilde{c}_{\alpha,A} = \mathrm{argmin}_{\mathrm{c}\in\mathrm{A}} \det(\Sigma_0(\mathrm{c}))$. As a result, the PM($c_\alpha$) estimator for k cannot be computed. Using the set approximation, it is still possible to compute estimators of k using the PM(c) estimator for every $c \in A$; the only problem is that it is not possible to determine the optimum estimator for k since $c_{\alpha,A}$ is unknown. The problem of implementing the PM estimators in practice is considered in the next section.



$\det(\Sigma_{0.25}(c_\alpha))/\det(\Sigma_{0.25}(c_{0.25,A}))$ $\det(\Sigma_{0.5}(c_\alpha))/\det(\Sigma_{0.5}(c_{0.5,A}))$ $\det(\Sigma_{0.75}(c_\alpha))/\det(\Sigma_{0.75}(c_{0.75,A}))$ $\det(\Sigma_{0.95}(c_\alpha))/\det(\Sigma_{0.95}(c_{0.95,A}))$

$\det(\Sigma_{0.25}(c_\alpha))/\det(\Sigma_{0.25}(\tilde{c}_{0.25,A}))$ $\det(\Sigma_{0.5}(c_\alpha))/\det(\Sigma_{0.5}(\tilde{c}_{0.5,A}))$ $\det(\Sigma_{0.75}(c_\alpha))/\det(\Sigma_{0.75}(\tilde{c}_{0.75,A}))$ $\det(\Sigma_{0.95}(c_\alpha))/\det(\Sigma_{0.95}(\tilde{c}_{0.95,A}))$

Figure 3.14: Contour levels of $\det(\Sigma_\alpha(c_\alpha))/\det(\Sigma_\alpha(c_{\alpha,A}))$ and $\det(\Sigma_\alpha(c_o))/\det(\Sigma_\alpha(\tilde{c}_{\alpha,A}))$ for $A = \{0,1\}$. Here $c_{\alpha,A} = \mathrm{argmin}_{c\in A}\det(\Sigma_\alpha(c))$ and $\tilde{c}_{\alpha,A} = \mathrm{argmin}_{c\in A}\det(\Sigma_0(c))$. Red and green shading represents areas of the parameter space where $c_{\alpha,A}$ and $\tilde{c}_{\alpha,A}$ are 0 and 1 respectively.



$\det(\Sigma_{0.25}(c_\alpha))/\det(\Sigma_{0.25}(c_{0.25,A}))$ $\det(\Sigma_{0.5}(c_\alpha))/\det(\Sigma_{0.5}(c_{0.5},A))$ $\det(\Sigma_{0.75}(c_\alpha))/\det(\Sigma_{0.75}(c_{0.75,A}))$ $\det(\Sigma_{0.95}(c_\alpha))/\det(\Sigma_{0.95}(c_{0.95,A}))$

$\det(\Sigma_{0.25}(c_\alpha))/\det(\Sigma_{0.25}(\tilde{c}_{0.25,A}))$ $\det(\Sigma_{0.5}(c_\alpha))/\det(\Sigma_{0.5}(\tilde{c}_{0.5,A}))$ $\det(\Sigma_{0.75}(c_\alpha))/\det(\Sigma_{0.75}(\tilde{c}_{0.75,A}))$ $\det(\Sigma_{0.95}(c_o))/\det(\Sigma_{0.95}(\tilde{c}_{0.95,A}))$

Figure 3.15: Contour levels of $\det(\Sigma_\alpha(c_\alpha))/\det(\Sigma_\alpha(c_{\alpha,A}))$ and $\det(\Sigma_\alpha(c_o))/\det(\Sigma_\alpha(\tilde{c}_{\alpha,A}))$ for $A = \{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$. Here $c_{\alpha,A} = \text{argmin}_{c\in A}\det(\Sigma_\alpha(c))$ and $\tilde{c}_{\alpha,A} = \text{argmin}_{c\in A}\det(\Sigma_0(c))$. Red through to green shading represents areas of the parameter space where $c_{\alpha,A}$ and $\tilde{c}_{\alpha,A}$ are 0.2, 0.4, 0.6 and 0.8, 1 respectively.

3.3 Practical implementation of the power method

This section considers the difficulties of implementing the PM that may arise in practice. One of the major problems is that the optimum value of the power method parameter c requires knowledge of the distributional parameters; the distributional parameters and optimum c are obviously unknown in practice. Section 3.3.1 considers the possibility of estimating c_α, the optimum value of the power method parameter c for a NBD INAR(1) sample with $\alpha \in [0, 1)$. In practice it is possible for the estimator of k to be negative, Section 3.3.2 further investigates the validity of setting $\hat{k} = \infty$ in such situations. Section 3.3.3 provides simulation results on using the maximum likelihood estimators and moment based estimators for i.i.d. NBD and NBD INAR(1) samples.

3.3.1 Computing efficient PM estimators in practice

In order to obtain efficient PM estimators for NBD INAR(1) samples, or i.i.d. samples in the case $\alpha = 0$, the optimum value of c_α that minimizes $\det(\Sigma_\alpha(c))$ given in Eq. (3.2.8) must be computed. The value of c_α, however, depends on the parameters (m, k) which are unknown in practice. This section investigates the use of preliminary, possibly inefficient, NBD estimators to estimate c_α. Denote the preliminary inefficient NBD estimators by $\tilde{m}$ and $\tilde{k}$, then the following estimators for c_α and approximated c_α will be considered

$$\widehat{c_\alpha} = \text{argmin}_{c\in(0,1)} \det(\Sigma_\alpha(c; \tilde{m}, \tilde{k})), \tag{3.3.1}$$

$$\widehat{c_A} = \text{argmin}_{c\in A} \det(\Sigma_\alpha(c; \tilde{m}, \tilde{k})), \qquad A \in \{0, 1/5, 2/5, 3/5, 4/5, 1\}, \tag{3.3.2}$$

$$\widehat{c_{\tilde{B}}} = \left(0.42+0.81\tilde{w}'-2.98\tilde{w}'^2+3.64\tilde{w}'^3\right)\tilde{b}+\left(0.51-1.66\tilde{w}'+4.31\tilde{w}'^2\right)\tilde{b}^2, \tag{3.3.3}$$

where $\tilde{b} = 1 - (1 + \tilde{m}/\tilde{k})^{-\tilde{k}}$ and $\tilde{w}' = \tilde{b}/\tilde{m}$. The two simplest and natural choices for the preliminary estimators are the MOM or ZTM estimators.

Fig. 3.16 shows 95% confidence ellipses of the MOM and ZTM estimators for various values of (m, k) within the (b, w')-parameter space together with contour levels of c_α for $\alpha \in \{0, 0.25, 0.5, 0.75\}$. The ellipses have been constructed assuming a NBD INAR(1) sample of size $N = 1000$ with $\alpha \in \{0, 0.25, 0.5, 0.75\}$. The ellipses show the variability that would be expected in the estimated values of c_α given a significance level of 0.05. The confidence ellipses become larger as either N decreases, α increases or the significance level decreases.

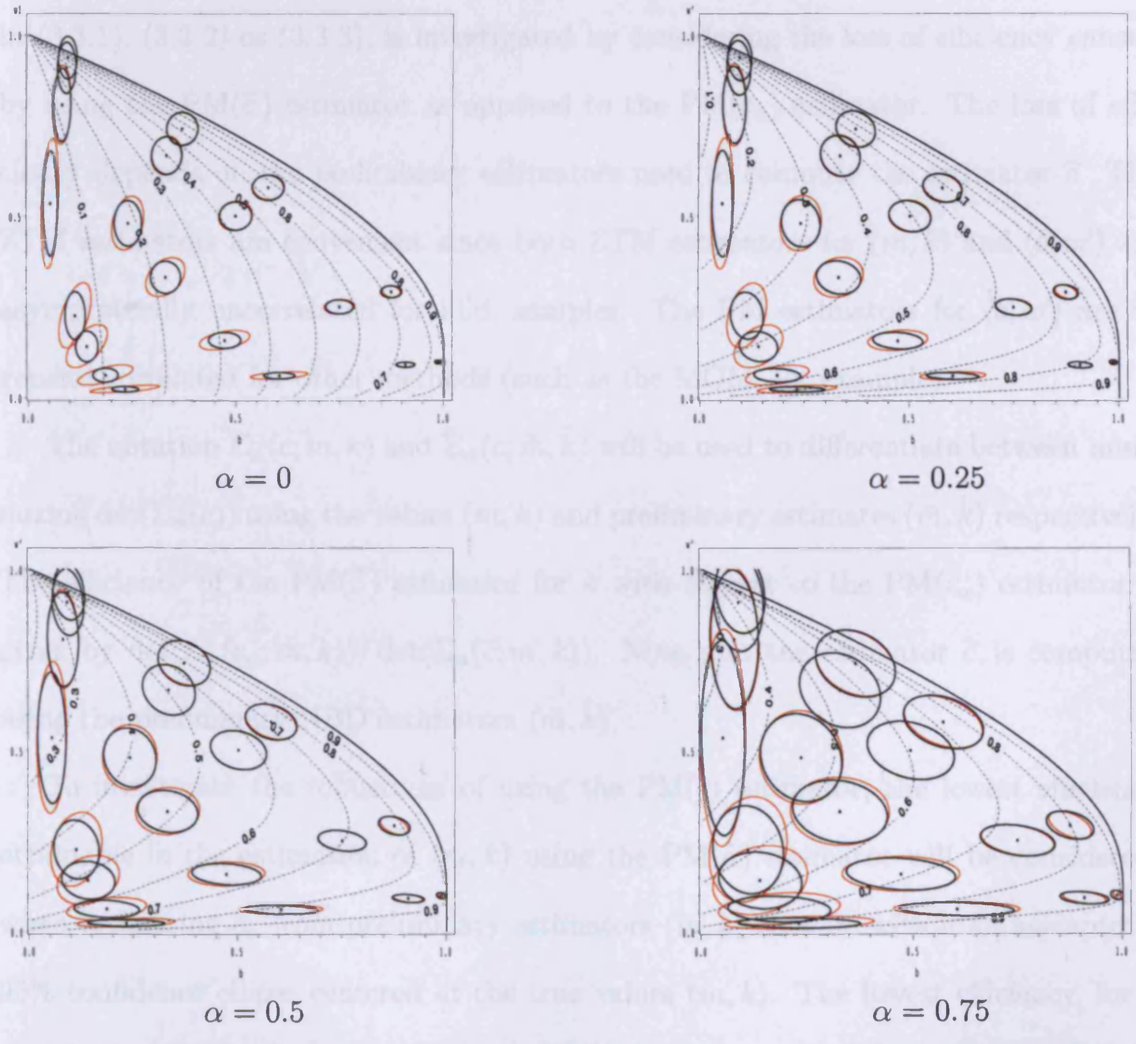


Figure 3.16: 95% confidence ellipses for NBD estimators when using (a) MOM (red) and (b) ZTM (black) using a sample size of $N = 1000$. It is assumed that N is large enough for convergence of the distribution of estimators to the normal distribution.

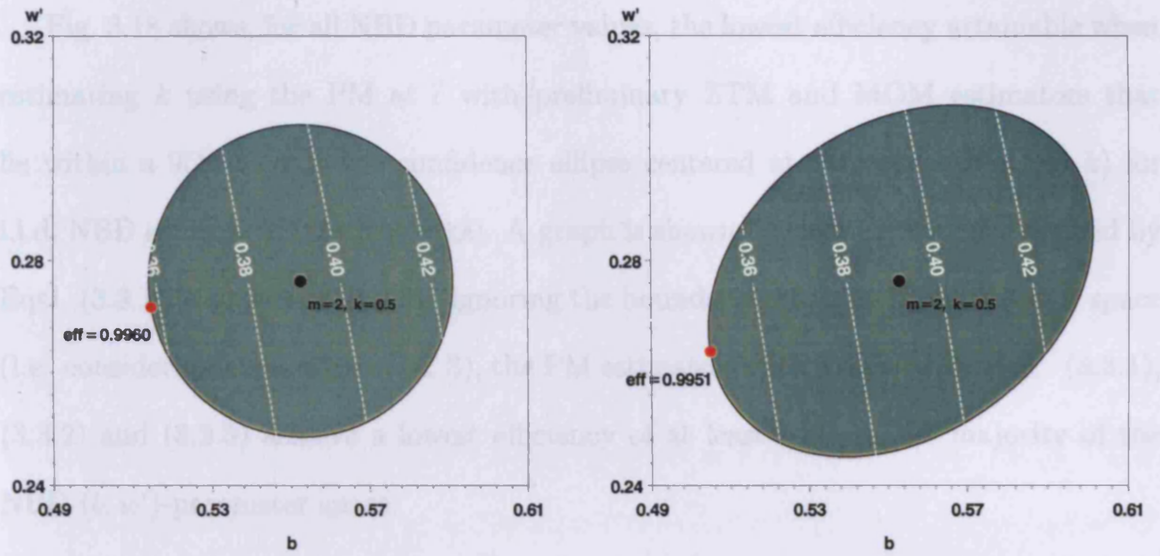
Fig. 3.16 indicates that, for the majority of NBD parameter values in the (b, w')-parameter space, the inefficient NBD estimator should provide an estimator of c_α close enough to c_α so that the PM estimator achieves a negligible loss of efficiency relative to the PM(c_α) estimator. The variability in estimated values of c_α depends on the sample size N, the value of α and the method used in estimating c_α.

Robustness

The robustness of using the PM($\widehat{c}$) estimator, where $\widehat{c}$ is one of the estimators defined in (3.3.1), (3.3.2) or (3.3.3), is investigated by considering the loss of efficiency caused by using the PM($\widehat{c}$) estimator as opposed to the PM(c_α) estimator. The loss of efficiency depends on the preliminary estimators used to compute the estimator $\widehat{c}$. The ZTM estimators are convenient since both ZTM estimators for (m, k) and (b, w') are asymptotically uncorrelated for i.i.d. samples. The PM estimators for (b, w') are in general correlated for other methods (such as the MOM, for example).

The notation $\Sigma_\alpha(c; m, k)$ and $\Sigma_\alpha(c; \tilde{m}, \tilde{k})$ will be used to differentiate between minimizing $\det(\Sigma_\alpha(c))$ using the values (m, k) and preliminary estimates $(\tilde{m}, \tilde{k})$ respectively. The efficiency of the PM($\widehat{c}$) estimator for k with respect to the PM(c_α) estimator is given by $\det(\Sigma_\alpha(c_\alpha; m, k))/\det(\Sigma_\alpha(\widehat{c}; m, k))$. Note that the estimator $\widehat{c}$ is computed using the preliminary NBD estimators $(\tilde{m}, \tilde{k})$.

To investigate the robustness of using the PM($\hat{c}$) estimator, the lowest efficiency attainable in the estimation of (m, k) using the PM($\widehat{c}$) estimator will be considered, when estimating c_α from preliminary estimators $(\tilde{m}, \tilde{k})$ that lie within an asymptotic 95% confidence ellipse centered at the true values (m, k). The lowest efficiency, for a significance level of 0.05, is given by $\det(\Sigma_\alpha(c_\alpha; m, k))/\det(\Sigma_\alpha(\widehat{c_\times}; m, k))$ where $\widehat{c_\times} = \text{argmax}_{\hat{c}\in\widehat{C}} \det(\Sigma_\alpha(c; m, k))$ and $\widehat{C}$ is the set of all possible estimators $\hat{c}$ obtained from preliminary estimators $(\tilde{m}, \tilde{k})$ that lie within the 95% confidence ellipse.



(a) (b)

Figure 3.17: Asymptotic confidence ellipses for NBD estimators $(\tilde{m}, \tilde{k})$ when using (a) ZTM and (b) MOM to estimate $(\tilde{m}, \tilde{k})$. The value 'eff' indicates the lowest value of $v_{PM}(c_o; m, k)/v_{PM}(\widehat{c_o}; m, k)$, where $\widehat{c_o} = \text{argmin}_{c\in(0,1)} v_{PM}(c; \tilde{m}, \tilde{k})$, amongst all $(\tilde{m}, \tilde{k})$ within the confidence ellipse.

Fig. 3.17(a) and (b) show examples of the robustness for the case $m = 2$ and $k = 0.5$ when using (a) ZTM and (b) MOM estimators as preliminary estimators for the estimation of $\widehat{c_o}$. A sample size of $N = 1000$ is assumed. An asymptotic 95% confidence ellipse is shown for the preliminary estimators $(\tilde{m}, \tilde{k})$ of (m, k) within the (b, w')-parameter space together with contour levels of $\widehat{c_o}$. The lowest efficiency occurs at the boundary of the ellipse and is 0.996 ($\tilde{m} = 1.90, \tilde{k} = 0.43, \widehat{c_o} = 0.36$) when using preliminary ZTM estimators and 0.995 ($\tilde{m} = 1.92, \tilde{k} = 0.40, \widehat{c_o} = 0.35$) when using the preliminary MOM estimators.

The lowest efficiency attainable depends on the significance level, the size of the sample and the preliminary estimator used. Increasing the sample size naturally reduces the size of the confidence ellipse for $(\tilde{m}, \tilde{k})$ providing a more accurate estimator for c_o and the estimation method becomes more robust. The PM$(\widehat{c})$ estimator also appears more robust when the significance level is increased.

Fig. 3.18 shows, for all NBD parameter values, the lowest efficiency attainable when estimating k using the PM at $\widehat{c}$ with preliminary ZTM and MOM estimators that lie within a 95% asymptotic confidence ellipse centered at the true values (m, k) for i.i.d. NBD samples of size $N = 1000$. A graph is shown for each estimator $\widehat{c}$ defined by Eqs. (3.3.1), (3.3.2) and (3.3.3). Ignoring the boundaries of the NBD parameter space (i.e. considering areas where $k < 3$), the PM estimators with $\hat{c}$ defined by Eqs. (3.3.1), (3.3.2) and (3.3.3) achieve a lowest efficiency of at least 0.98 for the majority of the NBD (b, w')-parameter space.

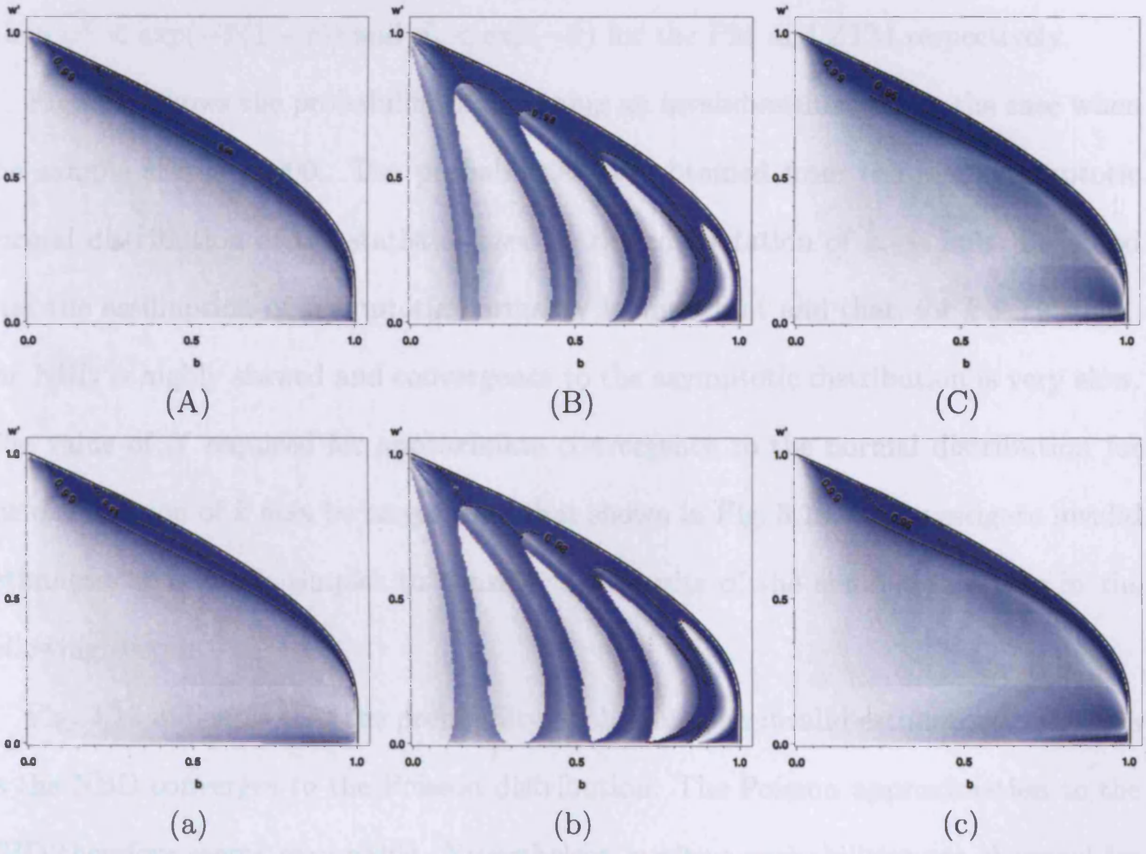


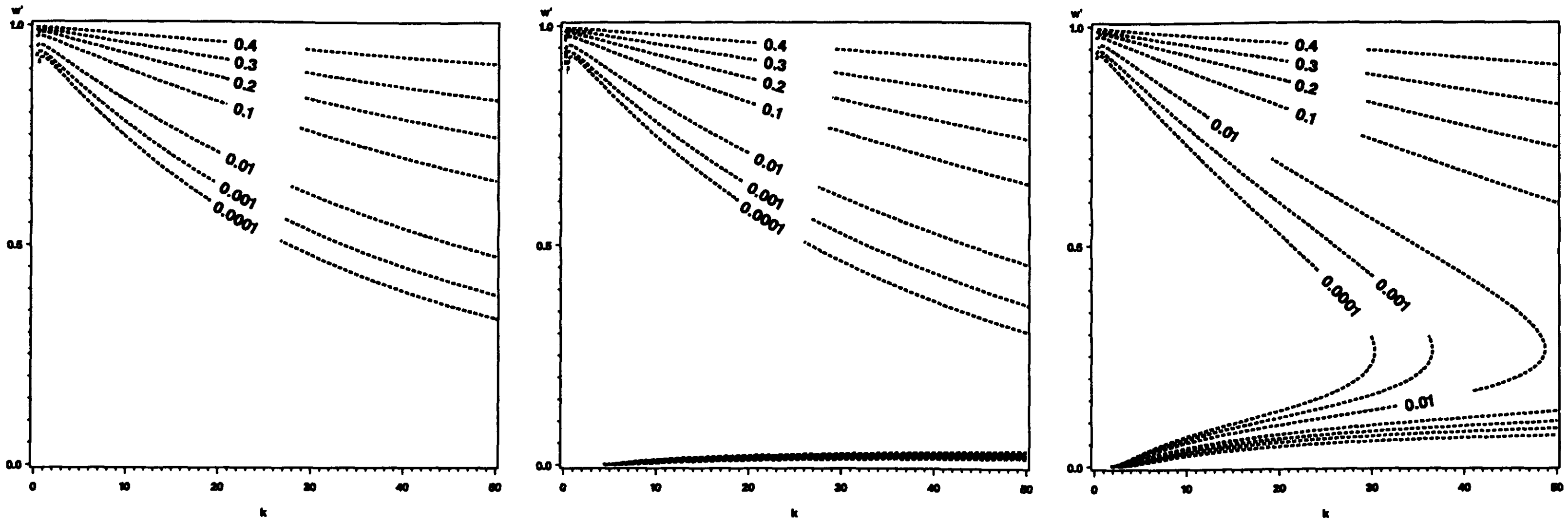
Figure 3.18: Lowest possible efficiency of the PM($\hat{c}$) estimator when using preliminary ZTM ((A),(B) and (C)) and MOM ((a), (b) and (c)) estimators for (m, k) that lie within a 95% confidence ellipse of the true values. The sample size is i.i.d. NBD ($\alpha = 0$) with $N = 1000$. The estimators $\hat{c}$ are obtained from (a) equation (3.3.1), (b) equation (3.3.2) and (c) equation (3.3.3).

3.3.2 Degenerate samples

In literature (see e.g. Anscombe (1950)) it is often assumed that, when an invalid estimator for k ($\hat{k} \leqslant 0$) is obtained, the Poisson distribution may be fitted and the estimator for k is set to $\hat{k} = \infty$. Recall that the MOM estimator for k is $\hat{k}_{MOM} = \bar{x}^2/(\overline{x^2} - \bar{x}^2 - \bar{x})$. It is clear that the MOM estimator for k will be negative when $s^2 < \bar{x}$, where $s^2 = \overline{x^2} - \bar{x}^2$. The PM estimator ($\hat{k}_{PM}$) and ZTM estimator ($\hat{k}_{ZTM}$) for k are respectively obtained by solving the equations $\widehat{c^X} = (1 + \bar{x}(1-c)/z)^{-z}$ and $\widehat{p_0} = (1 + \bar{x}/z)^{-z}$ with respect to z. Negative estimators for $\hat{k}_{PM}$ and $\hat{k}_{ZTM}$ are obtained when $\widehat{c^X} < \exp(-\bar{x}(1-c))$ and $\widehat{p_0} < \exp(-\bar{x})$ for the PM and ZTM respectively.

Fig. 3.19 shows the probability of obtaining an invalid estimator $\hat{k}$ in the case when the sample size is 10000. The probabilities are obtained from the joint asymptotic normal distribution of the statistics used in the computation of $\hat{k}$. It must be noted that the assumption of asymptotic normality is important and that, for k very small, the NBD is highly skewed and convergence to the asymptotic distribution is very slow. The value of N required for approximate convergence to the normal distribution for the distribution of $\hat{k}$ may be larger than that shown in Fig. 3.19. To investigate invalid estimators it is much simpler to consider the results of the simulation study in the following section.

Fig. 3.19 indicates that the probability of obtaining an invalid estimator $\hat{k}$ increases as the NBD converges to the Poisson distribution. The Poisson approximation to the NBD therefore seems reasonable. Nevertheless, positive probabilities are observed for all values of k shown in the picture, allbeit with very small probability. In practice, care must be taken. If estimators of market research measures (e.g. penetration or purchase frequency) are sensitive to changes in the value of k, then setting $\hat{k} = \infty$ may provide completely wrong inference on estimators for the market research measures.



(a) (b) (c)

Figure 3.19: Probability of obtaining a degenerate sample with sample size $N = 10000$ using (a) MOM, (b) PM ($c = 0.5$) and (c) ZTM.

3.3.3 Simulation results

This section considers the results of a simulation study comprising $R = 1000$ sample runs of the NBD distribution with sample size $N = 10000$ for various parameters (m, k) with $m \in \{0.1, 0.5, 1, 5, 10\}$ and $k \in \{0.01, 0.25, 0.5, 1, 3, 5\}$. The purpose of the simulations are to analytically confirm the results of the previous sections and also to investigate the behavior of the maximum likelihood estimators for NBD INAR(1) samples which are difficult to analytically analyze.

Table 3.1 shows the empirical coefficient of variation $\widehat{\kappa_k} = \sqrt{N}\sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{k}_i - k)^2}/k$ for the ML, MOM, ZTM and PM(c_o) estimators against the theoretical coefficient of variation $\kappa_k = \sqrt{v_{ML}}/k$ (see Fig. 3.1 (b)) when estimating from an i.i.d. NBD sample. A value of $\widehat{\kappa_k} = \infty$ indicates that $\hat{k}_i \leq 0$ or $\hat{k}_i = \infty$ for at least one sample. For all samples with $\widehat{\kappa_k} < \infty$ the PM(c_o) estimator has a consistently lower $\widehat{\kappa_k}$ than both the MOM and ZTM estimators. The largest percentage difference between the PM(c_o) estimator and the combined MOM/ZTM method occurs when $k = 1$ and $m = 10$ when the value of $\widehat{\kappa_k}$ is increased by a factor of 26% by using the MOM/ZTM method.

Table 3.2 shows the coefficient of variation $\widehat{\kappa_k} = \sqrt{N}\sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{k}_i - k)^2}/k$ for the ML and PM(c) estimators and Table 3.3 shows the coefficient of variation $\widehat{\kappa_m} = \sqrt{N}\sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{m}_i - m)^2}/m$ for ML estimators in the case where the sample is NBD INAR(1) with $\alpha = 0.5$. These tables compare estimators computed on the false assumption that the data is i.i.d. NBD against estimators computed using the fact that the true distribution of the sample is a NBD INAR(1) sample with $\alpha = 0.5$. For an i.i.d. NBD sample, the ML estimator for m is the sample mean and the ML estimator for k is given by Eq. (2.1.2). For an INAR(1) NBD sample, the ML estimators are computed by maximizing the likelihood function given by Eq. (3.2.3). In Table 3.2 a value of $\hat{\kappa} = \infty$ indicates that $\hat{k}_i \leq 0$ or $\hat{k}_i = \infty$ for at least one sample.

	$k=0.01$					$k=0.25$					$k=0.5$				
m	$\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM	$\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM	$\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM
0.1	**8.27**	8.78	8.78	8.78	15.21	**10.05**	10.38	10.38	10.38	11.50	**14.01**	15.34	15.34	15.34	16.25
0.5	**5.90**	5.91	5.91	5.91	13.70	**3.56**	3.62	3.63	3.62	4.76	**4.14**	4.06	4.11	4.07	4.86
1	**5.28**	5.30	5.30	5.30	13.86	**2.66**	2.65	2.67	2.65	4.00	**2.85**	2.85	2.92	2.86	3.60
5	**4.41**	4.48	4.48	4.48	13.75	**1.77**	1.80	1.83	1.81	3.34	**1.70**	1.77	1.89	1.78	2.66
10	**4.15**	4.19	4.20	4.19	13.74	**1.59**	1.61	1.66	1.61	3.30	**1.51**	1.51	1.70	1.53	2.56

	$k=1$					$k=3$					$k=5$				
m	$\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM	$\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM	$\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM
0.1	**21.55**	∞	∞	∞	∞	**50.40**	∞	∞	∞	∞	**78.86**	∞	∞	∞	∞
0.5	**5.51**	5.52	5.68	5.53	6.00	**11.21**	11.88	12.55	11.88	12.07	**16.89**	19.92	22.26	20.00	19.90
1	**3.49**	3.51	3.65	3.52	4.07	**6.30**	6.63	7.45	6.63	6.82	**9.14**	9.28	10.73	9.29	9.38
5	**1.79**	1.81	2.07	1.83	2.38	**2.36**	2.41	3.78	2.41	2.67	**2.94**	2.96	5.63	2.96	3.09
10	**1.55**	1.58	2.02	1.60	2.22	**1.86**	1.80	4.25	1.81	2.07	**2.16**	2.14	7.41	2.15	2.34

Table 3.1: Comparison of $\sqrt{N}\sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{k}_i - k)^2}/k$ against $\sqrt{v_{ML}}/k$ for the ML, ZTM, PM(c_o) and MOM estimators using $R = 1000$ i.i.d. samples of the NBD distribution with sample size $N = 10000$. A value of ∞ indicates that $\hat{k}_i \leq 0$ or $\hat{k}_i = \infty$ for at least one sample. All values are given to 2 decimal places.

	$k = 0.01$				$k = 0.25$				$k = 0.5$			
	Assuming i.i.d. sample		$\alpha = 0.5$		Assuming i.i.d. sample		$\alpha = 0.5$		Assuming i.i.d. sample		$\alpha = 0.5$	
m	MLE	PM(c_o)	MLE	PM($c_{0.5}$)	MLE	PM(c_o)	MLE	PM($c_{0.5}$)	MLE	PM(c_o)	MLE	PM($c_{0.5}$)
0.1	15.30	15.30	12.56	15.04	20.99	20.96	17.06	21.14	32.55	32.48	25.6	32.59
0.5	13.24	13.24	12.20	13.16	5.37	5.35	4.02	5.33	6.59	6.52	4.30	6.77
1	13.16	13.16	11.97	12.84	4.35	4.34	3.29	4.60	4.18	4.15	3.14	4.14
5					3.41	3.42	2.80	3.31	2.90	2.90	2.27	2.84
10					3.25	3.26	2.57	3.11	2.78	2.78	2.09	2.66

	$k = 1$				$k = 3$				$k = 5$			
	Assuming i.i.d. sample		$\alpha = 0.5$		Assuming i.i.d. sample		$\alpha = 0.5$		Assuming i.i.d. sample		$\alpha = 0.5$	
m	MLE	PM(c_o)	MLE	PM($c_{0.5}$)	MLE	PM(c_o)	MLE	PM($c_{0.5}$)	MLE	PM(c_o)	MLE	PM($c_{0.5}$)
0.1	77.97	76.71	37.48	77.63	2122.57	∞	1383.75	∞	2463.90	∞	2408.02	∞
0.5	13.66	13.60	11.38	13.92	32.65	32.70	26.84	32.66	63.03	62.30	34.86	62.30
1	5.77	5.68	3.86	6.04	21.30	21.27	18.45	21.35	30.60	30.66	25.74	30.63
5	2.79	2.76	2.16	2.75	3.78	3.72	2.80	3.91	5.04	4.99	3.97	5.16
10	2.62	2.61	1.92	2.56	2.86	2.81	2.06	2.91	3.37	3.34	2.45	3.45

Table 3.2: Comparison of $\sqrt{N}\sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{k}_i - k)^2/k}$ using the ML and PM(c_{opt}) estimators when assuming (incorrectly) that the sample is i.i.d. against the estimators obtained when the sample is NBD INAR(1) with $\alpha = 0.5$. Here $R = 1000$ samples of NBD INAR(1) realizations distribution with series length $N = 10000$. A value of ∞ indicates that $\hat{k}_i \leq 0$ or $\hat{k}_i = \infty$ for at least one sample. All values are given to 2 decimal places.

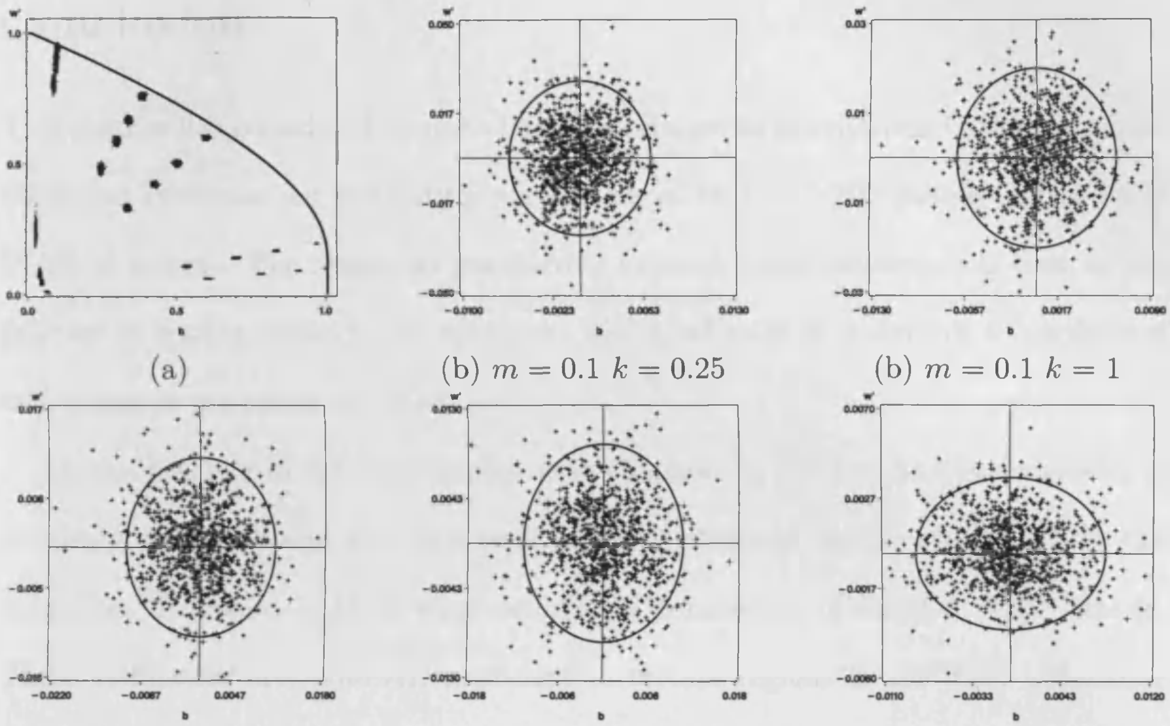
Table 3.2 shows that the coefficient of variation of the estimator for k, computed by maximizing the true likelihood equation (Eq. (3.2.3)), is clearly much lower than the coefficient of variation of the PM($c_{0.5}$) estimator as well as the ML and PM(c_o) estimators computed under the assumption that the data is i.i.d. NBD. The empirical coefficient of variation of the PM(c_o) and PM($c_{0.5}$) estimators are similar. It has already been noted in Section 3.2.2 that using the PM(c_o) estimator leads to a small loss of efficiency in comparison to using the correct PM($c_{0.5}$) estimator (see Fig. 3.13).

The numerical results in Table 3.3 show that the coefficient of variation for the estimator of m when maximizing the true likelihood function (Eq. (3.2.3)) is lower than the coefficient of variation of the maximum likelihood estimator for m of an i.i.d. NBD sample (i.e. the sample mean). This further indicates that the sample mean does not maximize the likelihood function given by Eq. (3.2.3).

	$k = 0.01$		$k = 0.25$		$k = 0.5$	
m	Assuming i.i.d. sample	$\alpha = 0.5$	Assuming i.i.d. sample	$\alpha = 0.5$	Assuming i.i.d. sample	$\alpha = 0.5$
0.1	17.45	17.17	6.30	5.64	5.97	4.96
0.5	17.86	17.68	4.28	4.04	3.49	3.29
1	16.63	16.20	3.74	3.66	2.97	2.79
5			3.59	3.57	2.61	2.58
10			3.37	3.35	2.52	2.50

	$k = 1$		$k = 3$		$k = 5$	
m	Assuming i.i.d. sample	$\alpha = 0.5$	Assuming i.i.d. sample	$\alpha = 0.5$	Assuming i.i.d. sample	$\alpha = 0.5$
0.1	6.53	5.92	7.15	6.39	6.56	5.91
0.5	2.84	2.61	2.90	2.67	2.79	2.59
1	2.51	2.37	1.97	1.83	1.86	1.75
5	1.96	1.93	1.20	1.20	1.09	1.07
10	1.83	1.81	1.12	1.11	1.00	0.98

Table 3.3: Comparison of $\sqrt{N}\sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{m}_i - m)^2}/m$ using the ML estimators when assuming (incorrectly) that the sample is i.i.d. against the estimators obtained when the sample is NBD INAR(1) with $\alpha = 0.5$. Here $R = 1000$ samples of NBD INAR(1) realizations distribution with series length $N = 10000$. All values are given to 2 decimal places.



(a) (b) $m = 0.1$ $k = 0.25$ (b) $m = 0.1$ $k = 1$

(b) $m = 1$ $k = 1$ (b) $m = 1$ $k = 5$ (b) $m = 5$ $k = 1$

Figure 3.20: (a) ZTM preliminary estimators for $m \in \{0.1, 0.5, 1, 5, 10\}$ and $k \in \{0.01, 0.25, 0.5, 1, 5\}$ in (b, w')-space. (b) ZTM preliminary estimators with corresponding 95% confidence ellipse.

Fig. 3.20 shows preliminary ZTM estimates, $(\tilde{b}, \tilde{w}')$, for different NBD parameters within the (b, w') parameter space. For each parameter pair, ZTM estimates for 1000 different NBD samples of size $N = 10000$ are shown. When comparing the ZTM estimates in Fig. 3.20 (a) to values of c_o in Fig. 3.4(b) it is clear that, even with the variation in the estimates $(\tilde{b}, \tilde{w}')$, the variation in the corresponding estimated values of c_o will be small in most regions of the (b, w')-space. The regions where c_o is sensitive to small changes in (b, w') and the corresponding maximum possible loss of efficiency in these regions was shown in Fig. 3.18. The maximum possible loss of efficiency was based on a 95% confidence ellipse. Fig. 3.20 (b) shows examples of preliminary ZTM estimates within the corresponding theoretical 95% confidence ellipses for (b, w'). These pictures are typical for each of the parameter pairs considered in Fig. 3.20 (a).

Conclusion

This chapter has considered moment based estimators as alternatives to the maximum likelihood estimator for estimating parameters of an i.i.d NBD sample and a NBD INAR(1) sample. The reason for considering moment based estimators is that, in the practice of market research, the maximum likelihood method is difficult to implement and moments are easier to obtain.

In the practice of market research it is common to use the standard method of moments estimator and the zero term method estimator as the alternative to the maximum likelihood method when estimating parameters of an i.i.d. NBD sample. These estimators are, however, inefficient in certain regions of the NBD parameter space. Importantly, this inefficient region of the parameter space includes areas where zero term method estimates of NBD parameters occur when fitting the NBD to the number of purchases of a product made by households at category level.

The power method for estimating the NBD parameters includes as particular cases the method of moments and the zero term method. The power method estimator for the NBD parameter k requires the choice of an additional parameter c. For $c = 0$, the power method estimator is equivalent to the zero term method estimator and as $c \to 1$ the power method estimator tends to the method of moments estimator.

The power method estimator is more efficient than the method of moments and zero term method estimator upon suitable choice of c except in the limiting cases as $b \to 0$ (which is when $c \to 0$) and as $k \to \infty$ (which is when $c \to 1$); in these cases the efficiency of the power method is equivalent to the efficiency of the zero term method and method of moments respectively. In the case of an i.i.d. NBD sample, it is in fact proven that there always exists a c ($c \in (0, 1)$) such that the power method estimator is more efficient than either the method of moments or zero term method estimators.

The optimum value of c depends on the NBD parameters m and k and must be computed by numerical minimization of the expression for the variance of the estimator for k given an i.i.d. NBD sample or minimization of the determinant of the covariance matrix of estimators of m and k given a NBD INAR(1) sample. In the case of an i.i.d. NBD sample, computing the power method estimator at this optimum value of c provides estimators of k almost as efficient as the maximum likelihood estimator for k. For a NBD INAR(1) sample, however, simulation results show that maximizing the likelihood function for the NBD INAR(1) model provides more efficient estimates for the NBD parameters in comparison to estimating the NBD parameters using the power method at optimum c. Nevertheless, the power method estimators at optimum c are still more efficient than the method of moments and zero term method estimators.

Simple approximations to the optimum value of c for the power method estimator have been proposed. These approximations lead to very small losses in efficiencies when estimating k relative to the power method estimator computed at the optimum value of c. Each of the approximations provide slightly different values of c; this shows the insensitive nature of the efficiency of estimating k using the power method to small changes in c.

The insensitive nature of the efficiency of the power method estimator for k to small changes in the value c in the region of the optimum value of c, enables the power method to be robustly implemented in practice. In practice, preliminary estimators that are possibly inefficient may be used to estimate the value of c; the estimated value of c can then be used to find an updated more efficient estimator for the NBD parameter k. Note that this procedure may be used iteratively until the value of estimated c or the estimates of the parameters converges. Simulation results, however, have shown that, even on the first iteration, efficient estimators for k can be obtained.

Chapter 4

Analyzing the dynamical behavior of negative binomial processes

This chapter considers the dynamical behavior of negative binomial processes by considering the correlation between statistical measures computed in varying time intervals. Analyzing the dynamical behavior of the mixed Poisson process with a negative binomial marginal distribution differs to analyzing the dynamical behavior of the negative binomial first-order autoregressive integer-valued process, which includes the sequence of i.i.d. NBD random variables.

The mixed Poisson process is not an ergodic process and therefore analyzing a single fixed realization does not represent the behavior of the process in the ensemble of realizations. In the case of panel data, many realizations are observed and it is therefore possible to check the appropriateness of fitting a mixed Poisson process by considering the covariances between statistics of the marginal distribution computed in different time intervals.

The NBD INAR(1) process is an ergodic process and the suitability of the INAR(1) process as a model for observed data can be confirmed by considering the autocorrelation function of the process. In addition to time domain analysis, where autocorrelation functions are considered, one may also consider spectral domain analysis of the INAR(1) process by considering the spectral frequencies of the process.

4.1 Mixed Poisson processes

This section considers the dynamical behavior of mixed Poisson processes by considering the covariances between statistics and estimators in two different time intervals. The gamma Poisson process is considered as a particular example in Section 4.2, where the correlation between commonly used market research measures is also considered.

Background

Recall the definition of a mixed Poisson process. Define the multivariate Poisson distribution as

$$\mathbb{P}\left(\boldsymbol{Z}=\boldsymbol{x}|\Lambda=\lambda\right)=\prod_{i=0}^{n-1}\frac{[\lambda(t_{i+1}-t_i)]^{x_{i+1}-x_i}}{(x_{i+1}-x_i)!}\exp\left(-\lambda(t_{i+1}-t_i)\right),$$

where $\lambda > 0$ is the intensity, $\boldsymbol{Z} = \{Z(t_1), Z(t_2), \ldots, Z(t_n)\}$ is a random vector, the set $\boldsymbol{x} = \{x_0, x_1, x_2, \ldots, x_n\}$ is a set of non-negative integers such that $0 = x_0 \leqslant x_1 \leqslant \ldots \leqslant x_n$ and $0 = t_0 \leqslant t_1 \leqslant \ldots \leqslant t_n$ represents an increasing sequence of time points. The mixed Poisson process is then defined as a process whose finite-dimensional distributions are

$$\mathbb{P}\left(\boldsymbol{Z}=\boldsymbol{x}\right)=\int_{0-}^{\infty}\mathbb{P}\left(\boldsymbol{Z}=\boldsymbol{x}|\Lambda=\lambda\right)dU_\Lambda(\lambda;\boldsymbol{\theta}).$$

Here $U_\Lambda(\lambda;\boldsymbol{\theta})$ is the distribution function for the random variable Λ and $\boldsymbol{\theta}$ is a vector of unknown parameters. The function $U_\Lambda(\lambda;\boldsymbol{\theta})$ is commonly known as the structure distribution of the mixed Poisson process.

In this section the asymptotic distributions of different statistics and estimators computed in two different, possibly overlapping, time intervals, using data from mixed Poisson processes, are derived. Section 4.1.1 considers the covariance of various statistics computed in two different time intervals. The joint asymptotic distribution of estimators are then derived in Section 4.1.2 using the results of Theorem 2.2.1 and Section 4.1.1.

4.1.1 Covariance of statistics

The simple case of computing the covariance of statistics computed in non-overlapping intervals is first considered. The results are then generalized to the case of statistics computed in overlapping intervals.

Non-overlapping intervals

Note that since the Poisson process is a stationary process that is homogenous in time, considering covariances of two statistics computed over the intervals $[t_1, t_2)$ and $[t_3, t_4)$ with $0 \leq t_1 < t_2 \leq t_3 < t_4$ is equivalent to considering covariances of the same statistics over the time intervals $[0, t)$ and $[t, t+s)$, so that $t_1 = 0,\ t_2 = t_3 = t$ and $t_4 = t+s$. Consider the covariance between the statistics

$$\bar{\phi}_{0,t} = \frac{1}{N}\sum_{l=1}^{N} \phi(z_l(0,t)) \quad \text{and} \quad \bar{\psi}_{t,t+s} = \frac{1}{N}\sum_{l=1}^{N} \psi(z_l(t,t+s)),$$

where $\{z_1(0,t), \dots, z_N(0,t)\}$ and $\{z_1(t,t+s), \dots, z_N(t,t+s)\}$ are i.i.d. data from a mixed Poisson process observed over two adjacent time intervals $[0,t)$ and $[t, t+s)$ respectively $(t, s > 0)$. Here ϕ and ψ are some functions possibly dependent upon the vector of parameters $\boldsymbol{\theta}$.

Note that for fixed u and v the observations $z_l(u, u+v)$ $(l = 1, \dots, N)$ are mutually independent. For fixed l, the observations $z_l(0,t)$ and $z_l(t, t+s)$ are independent Poisson distributed with means $\lambda_l t$ and $\lambda_l s$ respectively. Here λ_l is random for $l = 1, \dots, N$, but is the same for fixed l as time varies. The samples $\{z_1(0,t), \dots, z_N(0,t)\}$ and $\{z_1(t,t+s), \dots, z_N(t,t+s)\}$ are dependent since, for each l, $z_l(0,t)$ and $z_l(t,t+s)$ are Poisson distributed with a common λ_l. Let $\zeta_{u,v}$ be a random variable whose distribution is identical to the distribution of the i.i.d. random variables $z_l(u,v)$ $(l = 1, \dots, N)$, the number of events occurring in the time interval $[u, v)$. Then

$$N\mathrm{Cov}\left[\bar{\phi}_{0,t}, \bar{\psi}_{t,t+s}\right] = \mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right]. \tag{4.1.1}$$

Indeed,

$$\begin{aligned}
N\mathrm{Cov}\left[\bar{\phi}_{0,t}, \bar{\psi}_{t,t+s}\right] &= N\mathbb{E}(\bar{\phi}_{0,t} - \mathbb{E}\bar{\phi}_{0,t})(\bar{\psi}_{t,t+s} - \mathbb{E}\bar{\psi}_{t,t+s}) \\
&= N\mathbb{E}\left[\left\{\frac{1}{N}\sum_{l=1}^{N}\left[\phi(z_l(0,t)) - \mathbb{E}\phi(z_l(0,t))\right]\right\} \times \right. \\
&\qquad \left.\left\{\frac{1}{N}\sum_{l=1}^{N}\left[\psi(z_l(t,t+s)) - \mathbb{E}\psi(z_l(t,t+s))\right]\right\}\right] \\
&= \frac{1}{N}\mathbb{E}\left[\sum_{l=1}^{N}\left[\phi(z_l(0,t)) - \mathbb{E}\phi(z_l(0,t))\right]\left[\psi(z_l(t,t+s)) - \mathbb{E}\psi(z_l(t,t+s))\right]\right. \\
&\qquad \left. + \sum_{j\neq l}^{N}\left[\phi(z_j(0,t)) - \mathbb{E}\phi(z_j(0,t))\right]\left[\psi(z_l(t,t+s)) - \mathbb{E}\psi(z_l(t,t+s))\right]\right] \\
&= \frac{1}{N}\sum_{l=1}^{N}\mathbb{E}\left[\phi(z_l(0,t)) - \mathbb{E}\phi(z_l(0,t))\right]\left[\psi(z_l(t,t+s)) - \mathbb{E}\psi(z_l(t,t+s))\right] \\
&= \mathbb{E}\left[\phi(\zeta_{0,t}) - \mathbb{E}\phi(\zeta_{0,t})\right]\left[\psi(\zeta_{t,t+s}) - \mathbb{E}\psi(\zeta_{t,t+s})\right] \\
&= \mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right].
\end{aligned}$$

The covariances of pairs of statistics, commonly used in the estimation of NBD parameters, are derived below. These statistics are the method of moments statistics (z^α, z^β) for some $\alpha > 0$ and $\beta > 0$; the power method statistics (z, c^z) for some $c \neq 1$, general power method statistics (c_1^z, c_2^z) for some $c_1 \neq 1$ and $c_2 \neq 2$ and finally the functionals used in maximum likelihood $(\frac{\partial}{\partial\theta_i}\log p_{[0,t)}(z;\boldsymbol{\theta}), \frac{\partial}{\partial\theta_j}\log p_{[t,t+s)}(z;\boldsymbol{\theta}))$. Let $I_{[z=0]}$ be the indicator function such that $I_{[z=0]} = 1$ if $z = 0$ and $I_{[z=0]} = 0$ otherwise. Define $c^z = 1$ when $z = 0$ and $c^z = 0$ otherwise, then the zero term method statistics $(z, I_{[z=0]})$ are equivalent to the power method statistics (z, c^z) for $c = 0$.

Let $\mathcal{L}(c) = \mathbb{E}e^{-c\Lambda}$ be the Laplace transform of the random variable Λ with its derivative $\mathcal{L}'(c) = \frac{\partial}{\partial c}\mathbb{E}e^{-c\Lambda} = -\mathbb{E}[\Lambda e^{-c\Lambda}]$. Additionally, let $p_{[u,v)}(z;\boldsymbol{\theta})$ denote the mixed Poisson distribution over the time interval $[u,v)$. Then the covariances for the statistics discussed above are:

Case 1. $\phi(z) = z^\alpha, \psi(z) = z^\beta$:

$$\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] = \mathbb{E}\mu_\alpha(\lambda t)\mu_\beta(\lambda s) - \mathbb{E}\mu_\alpha(\lambda t)\mathbb{E}\mu_\beta(\lambda s),$$

where $\mu_\alpha(\nu) = \mathbb{E}\kappa_\nu^\alpha$ is the α-th moment of a Poisson random variable κ_ν with intensity ν.

Case 1a. $\phi(z) = z, \psi(z) = z$:

$$\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] = \mathbb{E}\zeta_{0,t}\zeta_{t,t+s} - \mathbb{E}\zeta_{0,t}\mathbb{E}\zeta_{t,t+s} = \mathbb{E}\Lambda^2 ts - \mathbb{E}\Lambda t\mathbb{E}\Lambda s = ts\mathrm{Var}\,\Lambda.$$

Case 1b. $\phi(z) = z, \psi(z) = z^2$:

$$\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] = \mathbb{E}\zeta_{0,t}\zeta_{t,t+s}^2 - \mathbb{E}\zeta_{0,t}\mathbb{E}\zeta_{t,t+s}^2 = ts^2\mathrm{Cov}[\Lambda, \Lambda^2] + ts\mathrm{Var}\Lambda.$$

Case 2. $\phi(z) = z, \psi(z) = c^z$:

$$\begin{aligned}\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] &= \mathbb{E}\zeta_{0,t}c^{\zeta_{t,t+s}} - \mathbb{E}\zeta_{0,t}\mathbb{E}c^{\zeta_{t,t+s}} = \mathbb{E}\Lambda t e^{-\Lambda s(1-c)} - \mathbb{E}\Lambda t\,\mathbb{E}e^{-\Lambda s(1-c)}\\ &= -t\left[\mathcal{L}'(s(1-c)) + \mathbb{E}\Lambda\,\mathcal{L}(s(1-c))\right].\end{aligned}$$

Case 3. $\phi(z) = c_1^z, \psi(z) = c_2^z$:

$$\begin{aligned}\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] &= \mathbb{E}c_1^{\zeta_{0,t}}c_2^{\zeta_{t,t+s}} - \mathbb{E}c_1^{\zeta_{0,t}}\mathbb{E}c_2^{\zeta_{t,t+s}}\\ &= \mathbb{E}e^{-\Lambda t(1-c_1)}e^{-\Lambda s(1-c_2)} - \mathbb{E}e^{-\Lambda t(1-c_1)}\mathbb{E}e^{-\Lambda s(1-c_2)}\\ &= \mathcal{L}([t(1-c_1) + s(1-c_2)]) - \mathcal{L}(t(1-c_1))\mathcal{L}(s(1-c_2)).\end{aligned}$$

Case 4. $\phi(z) = \frac{\partial}{\partial\theta_i}\log p_{[0,t)}(z;\boldsymbol{\theta}), \psi(z) = \frac{\partial}{\partial\theta_j}\log p_{[t,t+s)}(z;\boldsymbol{\theta})]$:

$$\begin{aligned}
\mathrm{Cov}\left[\phi(\zeta_{0,t}),\psi(\zeta_{t,t+s})\right] &= \mathbb{E}\frac{\partial}{\partial\theta_i}\log p_{[0,t)}(\zeta_{0,t};\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\log p_{[t,t+s)}(\zeta_{t,t+s};\boldsymbol{\theta}) \\
&= \mathbb{E}\frac{1}{p_{[0,t)}(\zeta_{0,t};\boldsymbol{\theta})p_{[t,t+s)}(\zeta_{t,t+s};\boldsymbol{\theta})}\frac{\partial}{\partial\theta_i}p_{[0,t)}(\zeta_{0,t};\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}p_{[t,t+s)}(\zeta_{t,t+s};\boldsymbol{\theta}) \\
&= \sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\frac{p_{\{[0,t),[t,t+s)\}}(u,v;\boldsymbol{\theta})}{p_{[0,t)}(u;\boldsymbol{\theta})p_{[t,t+s)}(v;\boldsymbol{\theta})}\frac{\partial}{\partial\theta_i}p_{[0,t)}(u;\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}p_{[t,t+s)}(v;\boldsymbol{\theta})
\end{aligned}$$

where $p_{\{[0,t),[t,t+s)\}}$ is the joint probability mass function of the random variables $\zeta_{0,t}$ and $\zeta_{t,t+s}$. The derivative $\frac{\partial}{\partial\theta_i}p_{[u,v)}(z;\boldsymbol{\theta})$ can be computed using the formula

$$\begin{aligned}
\frac{\partial}{\partial\theta_i}p_{[u,v)}(z;\boldsymbol{\theta}) &= \frac{\partial}{\partial\theta_i}\int_{0-}^{\infty}\frac{(\lambda(v-u))^z\exp(-\lambda(v-u))}{z!}\,dU_\Lambda(\lambda;\boldsymbol{\theta}) \\
&= \frac{1}{v-u}\int_{0-}^{\infty}\frac{\lambda^z\exp(-\lambda)}{z!}\frac{\partial}{\partial\theta_i}\left\{dU_\Lambda\left(\frac{\lambda}{v-u};\boldsymbol{\theta}\right)\right\}.
\end{aligned}$$

Overlapping intervals

This section considers covariances of statistics in the most general case when the intervals are possibly overlapping. This includes the cases when the intervals do not overlap and also when the intervals coincide. Consider the covariance of the statistics

$$\bar{\phi}_{t_1,t_3} = \frac{1}{N}\sum_{l=1}^{N}\phi(z_l(t_1,t_3)) \quad\text{and}\quad \bar{\psi}_{t_2,t_4} = \frac{1}{N}\sum_{l=1}^{N}\psi(z_l(t_2,t_4)),$$

where $\{z_1(t_1,t_3),\ldots,z_N(t_1,t_3)\}$ and $\{z_1(t_2,t_4),\ldots,z_N(t_2,t_4)\}$ are data from a mixed Poisson process observed over two, possibly overlapping, intervals $[t_1,t_3)$ and $[t_2,t_4)$ with $0 \le t_1 \le t_2 \le t_3 \le t_4$. Since the $z_l(u,u+v)$ for $l = 1,\ldots,N$ are mutually independent, the covariances of the statistics can be simplified to

$$N\mathrm{Cov}\left[\bar{\phi}_{t_1,t_3},\bar{\psi}_{t_2,t_4}\right] = \mathrm{Cov}[\phi(\zeta_{t_1,t_3}),\psi(\zeta_{t_2,t_4})]. \tag{4.1.2}$$

Computing these covariances for different functions ϕ and ψ can be further simplified by using the fact that the Poisson process has stationary and independent increments.

Case 1′a. $\phi(z) = z, \psi(z) = z$:

$$\begin{aligned}\mathrm{Cov}[\phi(\zeta_{t_1,t_3}), \psi(\zeta_{t_2,t_4})] &= \mathbb{E}\zeta_{t_1,t_3}\zeta_{t_2,t_4} - \mathbb{E}\zeta_{t_1,t_3}\mathbb{E}\zeta_{t_2,t_4} \\ &= \mathbb{E}(\zeta_{t_1,t_2} + \zeta_{t_2,t_3})(\zeta_{t_2,t_3} + \zeta_{t_3,t_4}) - (\mathbb{E}\zeta_{t_1,t_2} + \mathbb{E}\zeta_{t_2,t_3})(\mathbb{E}\zeta_{t_2,t_3} + \mathbb{E}\zeta_{t_3,t_4}) \\ &= \mathrm{Cov}(\zeta_{t_1,t_2}, \zeta_{t_2,t_3}) + \mathrm{Cov}(\zeta_{t_1,t_2}, \zeta_{t_3,t_4}) + \mathrm{Cov}(\zeta_{t_2,t_3}, \zeta_{t_3,t_4}) + \mathrm{Var}(\zeta_{t_2,t_3})\end{aligned}$$

and using the results of Case 1

$$\begin{aligned}\mathrm{Cov}[\phi(\zeta_{t_1,t_3}), \psi(\zeta_{t_2,t_4})] &= (t_3 - t_2)\,\mathbb{E}\Lambda + (t_3 - t_2)^2\mathrm{Var}\Lambda \\ &\quad + [(t_2 - t_1)(t_3 - t_2) + (t_2 - t_1)(t_4 - t_3) + (t_3 - t_2)(t_4 - t_3)]\,\mathrm{Var}\Lambda \\ &= (t_4 - t_2)(t_3 - t_1)\mathrm{Var}\Lambda + (t_3 - t_2)\mathbb{E}\Lambda.\end{aligned}$$

Case 1′b. $\phi(z) = z, \psi(z) = z^2$:

$$\begin{aligned}\mathrm{Cov}[\phi(\zeta_{t_1,t_3}), \psi(\zeta_{t_2,t_4})] &= \mathbb{E}\zeta_{t_1,t_3}\zeta^2_{t_2,t_4} - \mathbb{E}\zeta_{t_1,t_3}\mathbb{E}\zeta^2_{t_2,t_4} \\ &= \mathbb{E}(\zeta_{t_1,t_2} + \zeta_{t_2,t_3})(\zeta_{t_2,t_3} + \zeta_{t_3,t_4})^2 - (\mathbb{E}\zeta_{t_1,t_2} + \mathbb{E}\zeta_{t_2,t_3})\mathbb{E}(\zeta_{t_2,t_3} + \zeta_{t_3,t_4})^2 \\ &= \mathrm{Cov}(\zeta_{t_1,t_2}, \zeta^2_{t_2,t_3}) + 2\mathrm{Cov}(\zeta_{t_1,t_2}, \zeta_{t_2,t_3}\zeta_{t_3,t_4}) + \mathrm{Cov}(\zeta_{t_1,t_2}, \zeta^2_{t_3,t_4}) \\ &\quad + \mathrm{Cov}(\zeta_{t_2,t_3}, \zeta^2_{t_2,t_3}) + 2\mathrm{Cov}(\zeta_{t_2,t_3}, \zeta_{t_2,t_3}\zeta_{t_3,t_4}) + \mathrm{Cov}(\zeta_{t_2,t_3}, \zeta^2_{t_3,t_4})\end{aligned}$$

and using the results of Case 1

$$\begin{aligned}\mathrm{Cov}[\phi(\zeta_{t_1,t_3}), \psi(\zeta_{t_2,t_4})] &= (t_4 - t_2)^2(t_3 - t_1)\mathrm{Cov}(\Lambda, \Lambda^2) + (t_4 - t_2)(t_3 - t_1)\mathrm{Var}\Lambda \\ &\quad + 2(t_4 - t_2)(t_3 - t_2)\mathbb{E}\Lambda^2 + (t_3 - t_2)\mathbb{E}\Lambda.\end{aligned}$$

Case 2′. $\phi(z) = z, \psi(z) = c^z$:

$$\begin{aligned}\mathrm{Cov}[\phi(\zeta_{t_1,t_3}), \psi(\zeta_{t_2,t_4})] &= \mathbb{E}\zeta_{t_1,t_3}c^{\zeta_{t_2,t_4}} - \mathbb{E}\zeta_{t_1,t_3}\mathbb{E}c^{\zeta_{t_2,t_4}} \\ &= \mathbb{E}(\zeta_{t_1,t_2} + \zeta_{t_2,t_3})c^{\zeta_{t_2,t_4}} - \mathbb{E}(\zeta_{t_1,t_2} + \zeta_{t_2,t_3})\mathbb{E}c^{\zeta_{t_2,t_4}} \\ &= \mathrm{Cov}\left(\zeta_{t_1,t_2}, c^{\zeta_{t_2,t_4}}\right) + \mathbb{E}\zeta_{t_2,t_3}c^{\zeta_{t_2,t_3}}c^{\zeta_{t_3,t_4}} - \mathbb{E}\zeta_{t_2,t_3}\mathbb{E}c^{\zeta_{t_2,t_4}}.\end{aligned}$$

Using the result of Case 2

$$\mathrm{Cov}\left(\zeta_{t_1,t_2}, c^{\zeta_{t_2,t_4}}\right) = -(t_2-t_1)\left[\mathcal{L}'((t_4-t_2)(1-c)) + \mathbb{E}\Lambda\mathcal{L}((t_4-t_2)(1-c))\right].$$

Similarly,

$$\begin{aligned}
\mathbb{E}\zeta_{t_2,t_3}c^{\zeta_{t_2,t_3}}c^{\zeta_{t_3,t_4}} &= \mathbb{E}_\Lambda\left[\mathbb{E}\left(\zeta_{t_2,t_3}c^{\zeta_{t_2,t_3}}|\Lambda=\lambda\right)\mathbb{E}\left(c^{\zeta_{t_3,t_4}}|\Lambda=\lambda\right)\right] \\
&= \mathbb{E}_\Lambda\left[\left(\Lambda c(t_3-t_2)e^{-\Lambda(t_3-t_2)(1-c)}\right)\left(e^{-\Lambda(t_4-t_3)(1-c)}\right)\right] \\
&= c(t_3-t_2)\mathbb{E}_\Lambda\left[\Lambda e^{-\Lambda(t_4-t_2)(1-c)}\right] \\
&= -c(t_3-t_2)\mathcal{L}'\left((t_4-t_2)(1-c)\right).
\end{aligned}$$

Noting that $\mathbb{E}\zeta_{t_2,t_3}\mathbb{E}c^{\zeta_{t_2,t_4}} = (t_3-t_2)\mathbb{E}\Lambda\mathcal{L}\left((t_4-t_2)(1-c)\right)$, the above results are combined to give

$$\begin{aligned}
&\mathrm{Cov}[\zeta_{t_1,t_3}, c^{\zeta_{t_2,t_4}}] = -(t_2-t_1)\left[\mathcal{L}'((t_4-t_2)(1-c)) + \mathbb{E}\Lambda\mathcal{L}((t_4-t_2)(1-c))\right] \\
&- c(t_3-t_2)\mathcal{L}'\left((t_4-t_2)(1-c)\right) - (t_3-t_2)\mathbb{E}\Lambda\mathcal{L}\left((t_4-t_2)(1-c)\right) \\
&= -\left[(t_2-t_1)+c(t_3-t_2)\right]\mathcal{L}'((t_4-t_2)(1-c)) - (t_3-t_1)\mathbb{E}\Lambda\mathcal{L}((t_4-t_2)(1-c)).
\end{aligned}$$

Case 3′. $\phi(z) = c_1^z, \psi(z) = c_2^z$:

$$\begin{aligned}
&\mathrm{Cov}[\phi(\zeta_{t_1,t_3}), \psi(\zeta_{t_2,t_4})] = \mathbb{E}c_1^{\zeta_{t_1,t_3}}c_2^{\zeta_{t_2,t_4}} - \mathbb{E}c_1^{\zeta_{t_1,t_3}}\mathbb{E}c_2^{\zeta_{t_2,t_4}} \\
&= \mathbb{E}c_1^{\zeta_{t_1,t_2}}(c_1c_2)^{\zeta_{t_2,t_3}}c_2^{\zeta_{t_3,t_4}} - \mathbb{E}c_1^{\zeta_{t_1,t_3}}\mathbb{E}c_2^{\zeta_{t_2,t_4}} \\
&= \mathbb{E}e^{-\Lambda(t_2-t_1)(1-c_1)}e^{-\Lambda(t_3-t_2)(1-c_1c_2)}e^{-\Lambda(t_4-t_3)(1-c_2)} - \mathbb{E}e^{-\Lambda(t_3-t_1)(1-c_1)}\mathbb{E}e^{-\Lambda(t_4-t_2)(1-c_2)} \\
&= \mathcal{L}\left((t_4-t_1)-(t_2-t_1)c_1-(t_3-t_2)c_1c_2-(t_4-t_3)c_2\right) - \mathcal{L}\left((t_3-t_1)(1-c_1)\right)\mathcal{L}\left((t_4-t_2)(1-c_2)\right).
\end{aligned}$$

If $t_1 = 0$, $t_2 = t_3 = t$ and $t_4 = t+s$ the results of Section 4.1.1 are obtained, i.e. the covariances over non-overlapping intervals, in all three cases.

Covariance between statistics in the same time interval

Consider the particular case when $t_1 = t_2 = 0$ and $t_3 = t_4 = t$ so that the statistics are computed in the same time interval. The covariances between statistics in overlapping intervals can then be simplified as follows:

Case $\tilde{1}$a. $\phi(z) = z, \psi(z) = z$:

$$\mathrm{Var}(\zeta_{0,t}) = t^2 \mathrm{Var}\Lambda + t\mathbb{E}\Lambda.$$

(This formula is given in Grandell (1997, p. 14).)

Case $\tilde{1}b$. $\phi(z) = z, \psi(z) = z^2$:

$$\mathrm{Cov}[\zeta_{0,t}, \zeta_{0,t}^2] = t^3 \mathrm{Cov}[\Lambda, \Lambda^2] + t^2 \mathrm{Var}\Lambda + 2t^2 \mathbb{E}\Lambda^2 + t\mathbb{E}\Lambda.$$

Case $\tilde{2}$. $\phi(z) = z, \psi(z) = c^z$:

$$\mathrm{Cov}[\zeta_{0,t}, c^{\zeta_{0,t}}] = -t[c\mathcal{L}'(t(1-c)) + \mathbb{E}\Lambda\, \mathcal{L}(t(1-c))].$$

Case $\tilde{3}$. $\phi(z) = c_1^z, \psi(z) = c_2^z$:

$$\mathrm{Cov}[c_1^{\zeta_{0,t}}, c_2^{\zeta_{0,t}}] = \mathcal{L}\left(t(1 - c_1 c_2)\right) - \mathcal{L}\left(t(1 - c_1)\right) \mathcal{L}\left(t(1 - c_2)\right).$$

Consider also

Case $\tilde{4}$. Any suitable $\phi(z)$ (so that the expectations below exist)

and $\psi(z) = \frac{\partial}{\partial \theta_j} \log p_{[0,t)}(z; \boldsymbol{\theta})$:

$$\begin{aligned}\mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{0,t})\right] &= \mathbb{E}\phi(\zeta_{0,t}) \frac{\partial}{\partial \theta_j} \log p_{[0,t)}(\zeta_{0,t}; \boldsymbol{\theta}) = \mathbb{E} \frac{\phi(\zeta_{0,t})}{p_{[0,t)}(\zeta_{0,t}; \boldsymbol{\theta})} \frac{\partial}{\partial \theta_j} p_{[0,t)}(\zeta_{0,t}; \boldsymbol{\theta}) \\ &= \sum_{u=0}^{\infty} \phi(u) \frac{\partial}{\partial \theta_j} p_{[0,t)}(u; \boldsymbol{\theta}).\end{aligned}$$

4.1.2 Covariances of estimators

Let $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ be estimators of $\boldsymbol{\theta}$ in the intervals $[t_1, t_3)$ and $[t_2, t_4)$ constructed using the general scheme of Section 2.2.1 with the sets of functions $\{f_i^{(1)}(z;\boldsymbol{\theta})\}_{i=1}^d$ and $\{f_i^{(2)}(z;\boldsymbol{\theta})\}_{i=1}^d$, respectively. Assume that Theorem 2.2.1 applies to $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ so that both estimators are asymptotically normal and let $\boldsymbol{V}^{(1)}$, $\boldsymbol{V}^{(2)}$, $\mathbb{D}\boldsymbol{f}^{(1)}$ and $\mathbb{D}\boldsymbol{f}^{(2)}$ be the matrices associated with $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$. Using Theorem 2.2.1, $\sqrt{N}(\bar{\boldsymbol{f}} - \mathbb{E}\boldsymbol{f})$ is asymptotically normal $\mathcal{N}(0, \mathbb{D}\boldsymbol{f})$, where

$$\boldsymbol{f}(z;\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{f}^{(1)}(z;\boldsymbol{\theta}) \\ \boldsymbol{f}^{(2)}(z;\boldsymbol{\theta}) \end{pmatrix}, \quad \bar{\boldsymbol{f}} = \begin{pmatrix} \bar{\boldsymbol{f}}^{(1)} \\ \bar{\boldsymbol{f}}^{(2)} \end{pmatrix}, \quad \mathbb{E}\boldsymbol{f} = \begin{pmatrix} \mathbb{E}\boldsymbol{f}^{(1)}(\zeta_{t_1,t_3};\boldsymbol{\theta}) \\ \mathbb{E}\boldsymbol{f}^{(2)}(\zeta_{t_2,t_4};\boldsymbol{\theta}) \end{pmatrix}$$

and

$$\mathbb{D}\boldsymbol{f} = \begin{pmatrix} \mathbb{D}\boldsymbol{f}^{(1)} & \mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}) \\ \mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)})^T & \mathbb{D}\boldsymbol{f}^{(2)} \end{pmatrix} \tag{4.1.3}$$

with

$$\mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}) = \left\| \mathrm{Cov}(f_i^{(1)}(\zeta_{t_1,t_3};\boldsymbol{\theta}), f_j^{(2)}(\zeta_{t_2,t_4};\boldsymbol{\theta})) \right\|_{i,j=1}^d .$$

The components of the matrix $\mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)})$ are computed using the results of Section 4.1.1.

Consider the problem of estimating the vector $\boldsymbol{\theta}_* = \left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right)^T$ with the estimator $\hat{\boldsymbol{\theta}}_* = \left(\hat{\boldsymbol{\theta}}^{(1)}, \hat{\boldsymbol{\theta}}^{(2)}\right)^T$, where $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ are two different copies of $\boldsymbol{\theta}$. The fact that $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ are two different copies of $\boldsymbol{\theta}$ implies that the matrix of partial derivatives $\boldsymbol{V}$, defined by Eq. (2.2.4) with $\boldsymbol{\theta}_*$ substituted for $\boldsymbol{\theta}^*$, has a block diagonal structure

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}^{(1)} & 0 \\ 0 & \boldsymbol{V}^{(2)} \end{pmatrix}. \tag{4.1.4}$$

Using Theorem 2.2.1, $\sqrt{N}(\hat{\boldsymbol{\theta}}_* - \boldsymbol{\theta}_*)$ is asymptotically normal $\mathcal{N}(0, \boldsymbol{V}(\mathbb{D}\boldsymbol{f})\boldsymbol{V}^T)$, where $\mathbb{D}\boldsymbol{f}$ and $\boldsymbol{V}$ are defined by Eqs. (4.1.3) and (4.1.4). The asymptotic covariance matrix is therefore

$$\boldsymbol{V}(\mathbb{D}\boldsymbol{f})\boldsymbol{V}^T = \begin{pmatrix} \boldsymbol{V}^{(1)}\, \mathbb{D}\boldsymbol{f}^{(1)}\, (\boldsymbol{V}^{(1)})^T & \boldsymbol{V}^{(1)}\, \mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)})\, (\boldsymbol{V}^{(2)})^T \\ \boldsymbol{V}^{(2)}\, (\mathbb{C}(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}))^T\, (\boldsymbol{V}^{(1)})^T & \boldsymbol{V}^{(2)}\, \mathbb{D}\boldsymbol{f}^{(2)}\, (\boldsymbol{V}^{(2)})^T \end{pmatrix}. \tag{4.1.5}$$

4.2 The gamma Poisson process

The gamma Poisson process is a mixed Poisson process whose structure distribution $U_\Lambda(\lambda; \boldsymbol{\theta})$ is the gamma distribution with probability density function

$$g(\lambda; a, k) = \frac{1}{a^k \Gamma(k)} \lambda^{k-1} \mathrm{e}^{-\lambda/a}, \quad a > 0,\, k > 0, \quad \lambda > 0.$$

The finite-dimensional distribution of the gamma Poisson process is

$$\begin{aligned}\mathbb{P}(\boldsymbol{Z} = \boldsymbol{x}) &= \int_{0-}^{\infty} \left(\prod_{i=0}^{n-1} \frac{[\lambda(t_{i+1}-t_i)]^{x_{i+1}-x_i}}{(x_{i+1}-x_i)!} \exp\left(-\lambda(t_{i+1}-t_i)\right) \right) g(\lambda; a, k)\, d\lambda \\ &= \frac{\Gamma(k+x_n)}{\Gamma(k)} \left(\prod_{i=0}^{n-1} \frac{(t_{i+1}-t_i)^{x_{i+1}-x_i}}{(x_{i+1}-x_i)!} \right) \frac{a^{x_n}}{(1+at_n)^{x_n+k}}.\end{aligned}$$

The one-dimensional distribution of the gamma-Poisson process is the negative binomial distribution (NBD) with probabilities

$$p_x = \mathbb{P}(Z(t_1) = x) = \frac{\Gamma(k+x)}{\Gamma(k)x!} \left(\frac{1}{1+at_1} \right)^k \left(\frac{at_1}{1+at_1} \right)^x.$$

Four methods are considered in the estimation of (m, k). The estimators $\hat{m}$ and $\hat{k}$ are obtained as the solutions to the equations $\bar{f}_1 - \mathbb{E}\bar{f}_1 = 0$ and $\bar{f}_2 - \mathbb{E}\bar{f}_2 = 0$, where $\bar{f}_1 = \frac{1}{N}\sum_{l=1}^N f_1(z_l; m, k)$ and $\bar{f}_2 = \frac{1}{N}\sum_{l=1}^N f_2(z_l; m, k)$ and $\{z_1, \ldots, z_N\}$ is an i.i.d. NBD sample. The methods are defined by the functions f_1, f_2 which are as follows:

- **Maximum likelihood (ML):** $f_1(z; m, k) = \frac{\partial \log p(z;m,k)}{\partial m}$, $f_2(z; m, k) = \frac{\partial \log p(z;m,k)}{\partial k}$;
- **Standard method of moments (MOM):** $f_1(z) = z$, $f_2(z) = z^2$;
- **Zero term method (ZTM):** $f_1(z) = z$, $f_2(z) = 1$ if $z = 0$ and 0 otherwise;
- **Power method (PM):** $f_1(z) = z$, $f_2(z) = c^z$ for some $c \neq 1$;

4.2.1 Covariance of statistics

This section considers the covariance of the statistics $\bar{\phi}$ and $\bar{\psi}$ for the following pairs of functions:

$$\begin{pmatrix} \phi(z) \\ \psi(z) \end{pmatrix} \in \left\{ \begin{pmatrix} z \\ z \end{pmatrix}, \begin{pmatrix} z \\ z^2 \end{pmatrix}, \begin{pmatrix} z \\ c^z \end{pmatrix}, \begin{pmatrix} c_1^z \\ c_2^z \end{pmatrix} \right\}.$$

For the gamma distributed random variable Λ with density $g(\lambda; a, k) = \lambda^{k-1} e^{-\lambda/a}/(a^k \Gamma(k))$ with $a > 0$, $k > 0$, $\lambda > 0$ the following moments and expectations are required:

$$\mathbb{E}\Lambda^\alpha = \frac{a^\alpha \Gamma(k+\alpha)}{\Gamma(k)} \quad (\alpha = 1, 2, 3, \ldots), \quad \operatorname{Var}\Lambda = a^2 k, \quad \mathcal{L}(c) = (1+ac)^{-k},$$
$$\mathcal{L}'(c) = -ak(1+ac)^{-k-1}, \quad \operatorname{Cov}(\Lambda, \Lambda^2) = 2a^3 k(k+1).$$

The covariances between statistics in non-overlapping and overlapping intervals follow from the results in Section 4.1.

Non-overlapping intervals $[0, t)$ and $[t, t+s)$

Case 1a. $\phi(z) = z, \psi(z) = z$:

$$\operatorname{Cov}\left[\bar{\phi}, \bar{\psi}\right] = tsa^2 k.$$

Case 1b. $\phi(z) = z, \psi(z) = z^2$:

$$\operatorname{Cov}\left[\bar{\phi}, \bar{\psi}\right] = tsa^2 k(1 + 2as(k+1)).$$

Case 2. $\phi(z) = z, \psi(z) = c^z$:

$$\operatorname{Cov}\left[\bar{\phi}, \bar{\psi}\right] = -tsa^2 k(1 + as(1-c))^{-k-1}(1-c).$$

Case 3. $\phi(z) = c_1^z, \psi(z) = c_2^z$:

$$\operatorname{Cov}\left[\bar{\phi}, \bar{\psi}\right] = (1+at(1-c_1) + as(1-c_2))^{-k} - (1+at(1-c_1))^{-k}(1+as(1-c_2))^{-k}.$$

Note that in the case of $c_1 = c_2 = c$

$$\operatorname{Cov}\left[\bar{\phi}, \bar{\psi}\right] = (1 + a(t+s)(1-c))^{-k} - (1 + at(1-c))^{-k}(1 + as(1-c))^{-k}.$$

Overlapping intervals $[t_1, t_3)$ and $[t_2, t_4)$ with $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$

Case 1a. $\phi(z) = z, \psi(z) = z$:

$$\text{Cov}\left[\bar{\phi}, \bar{\psi}\right] = (t_4 - t_2)(t_3 - t_1)a^2k + ak(t_3 - t_2).$$

Case 1b. $\phi(z) = z, \psi(z) = z^2$:

$$\begin{aligned}\text{Cov}\left[\bar{\phi}, \bar{\psi}\right] &= 2(t_4 - t_2)^2(t_3 - t_1)a^3k(k+1) + (t_4 - t_2)(t_3 - t_1)a^2k \\ &\quad + 2(t_4 - t_2)(t_3 - t_2)a^2k(k+1) + (t_3 - t_2)ak.\end{aligned}$$

Case 2. $\phi(z) = z, \psi(z) = c^z$:

$$\text{Cov}\left[\bar{\phi}, \bar{\psi}\right] = -\frac{ak(1-c)[(t_3 - t_2) + a(t_3 - t_1)(t_4 - t_2)]}{(1 + a(t_4 - t_2)(1-c))^{k+1}}.$$

Case 3. $\phi(z) = c_1^z, \psi(z) = c_2^z$:

$$\begin{aligned}\text{Cov}\left[\bar{\phi}, \bar{\psi}\right] &= (1 + a\left[(t_2 - t_1)(1 - c_1) + (t_3 - t_2)(1 - c_1c_2) + (t_4 - t_3)(1 - c_2)\right])^{-k} \\ &\quad - (1 + a(t_3 - t_1)(1 - c_1))^{-k}(1 + a(t_4 - t_2)(1 - c_2))^{-k}.\end{aligned}$$

Note that for the case $t_2 = t_1$ and $t_4 = t_3$ the covariances are obtained for statistics computed in the same time interval. Let $t_2 = t_1 = 0$ and $t_4 = t_3 = t$ then

Case 1a. $\phi(z) = z, \psi(z) = z$:

$$\text{Cov}\left[\bar{\phi}, \bar{\psi}\right] = kat(1 + at).$$

Case 1b. $\phi(z) = z, \psi(z) = z^2$:

$$\text{Cov}\left[\bar{\phi}, \bar{\psi}\right] = kat(1 + at)\left(2(k+1)at + 1\right).$$

Case 2. $\phi(z) = z, \psi(z) = c^z$:

$$\text{Cov}\left[\bar{\phi}, \bar{\psi}\right] = -akt(1-c)(1+at)\left(1 + at(1-c)\right)^{-k-1}.$$

Case 3. $\phi(z) = c_1^z, \psi(z) = c_2^z$:

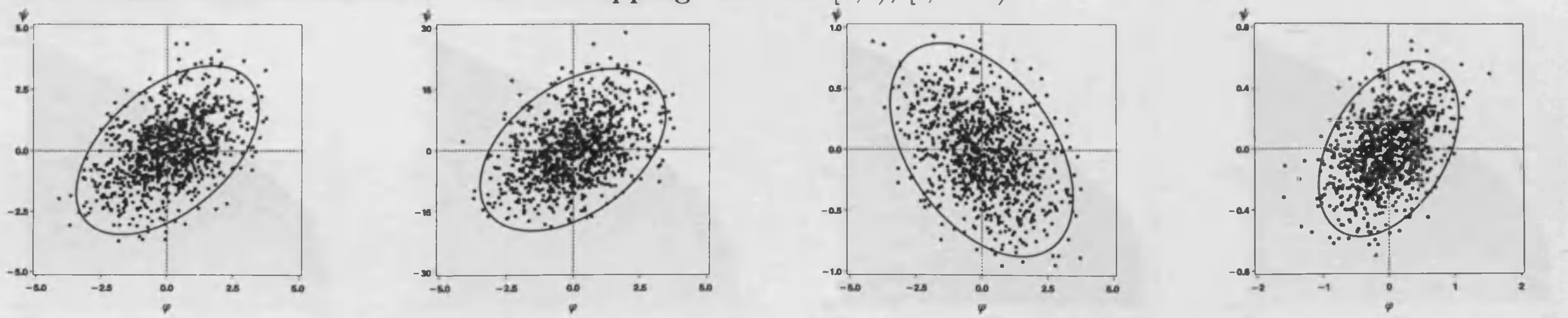
$$\mathrm{Cov}\left[\bar{\phi}, \bar{\psi}\right] = \left(1 + at(1-c_1c_2)\right)^{-k} - \left(1 + at(1-c_1)\right)^{-k}\left(1 + at(1-c_2)\right)^{-k}.$$

Fig. 4.1 and Fig. 4.2 show bivariate plots of various statistics $\bar{\phi}$ and $\bar{\psi}$ computed in different time intervals for 1000 replications of the gamma Poisson process with sample size N=1000. A 95% confidence ellipse based on the covariance matrix (4.1.3) and constructed under the assumption of asymptotic normality is also shown. Figures are shown for the two cases of overlapping and non-overlapping time intervals and confirm the results of this section.

Fig. 4.3 and Fig. 4.4 show correlations $\rho(\bar{\phi}, \bar{\psi}) = \rho(\phi, \psi)$ (follows from Eq. (4.1.1) and Eq. (4.1.2)) for various functions ϕ and ψ in the case of overlapping and non-overlapping time intervals. These correlations will be useful when computing the correlations between estimators and market research measures of the gamma Poisson process computed in different time intervals. Note that given data from the gamma Poisson process, computing the correlations between statistics in different time intervals can give some indication as to the region of the parameter space in which the parameters lie.

Fig. 4.5 and Fig. 4.6 show the correlations $\rho(\bar{\phi}, \bar{\psi}) = \rho(\phi, \psi)$ (for various functions ϕ and ψ) in the case $m = 1$ and $k = 1$ as a function of varying time for both overlapping intervals and non-overlapping intervals. In the case of overlapping intervals, for the statistics shown, the absolute value of the correlation decreases linearly as the amount of overlap decreases. As would be expected, if the statistics computed in each time interval are the same then the correlation tends to 1 as the proportion of overlap goes to 1. The adequacy of the gamma Poisson process as a model for data can be checked by comparing the empirical covariances of statistics obtained from data in varying time intervals to the expected gamma Poisson covariances of statistics.

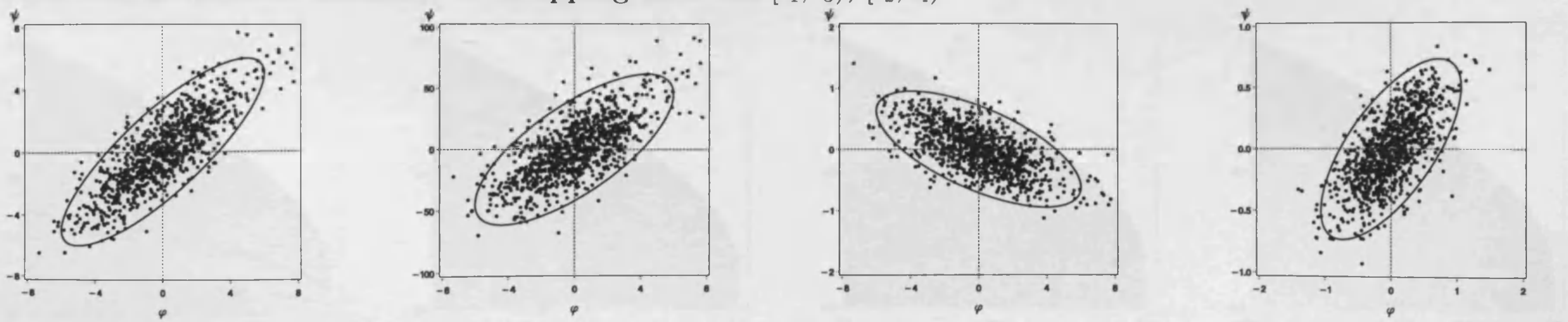
Joint distributions of statistics: Non-overlapping intervals $[0, t)$, $[t, t+s)$



(a) $\phi(z) = z,\ \psi(z) = z$ (b) $\phi(z) = z,\ \psi(z) = z^2$ (c) $\phi(z) = z,\ \psi(z) = 0.5^z$ (d) $\phi(z) = 0.25^z,\ \psi(z) = 0.75^z$

Figure 4.1: 1000 points of $\sqrt{N}(\bar{\phi} - \mathbb{E}\bar{\phi})$ versus $\sqrt{N}(\bar{\psi} - \mathbb{E}\bar{\psi})$ computed for various functions $\phi(z)$ and $\psi(z)$ when sampling from the gamma Poisson process with $m = 1, k = 1$ with samples of size $N = 1000$ in the case $t = 1, s = 1$. A 95% confidence ellipse based on the covariance matrix (4.1.3) and constructed under the assumption of asymptotic normality is also shown.

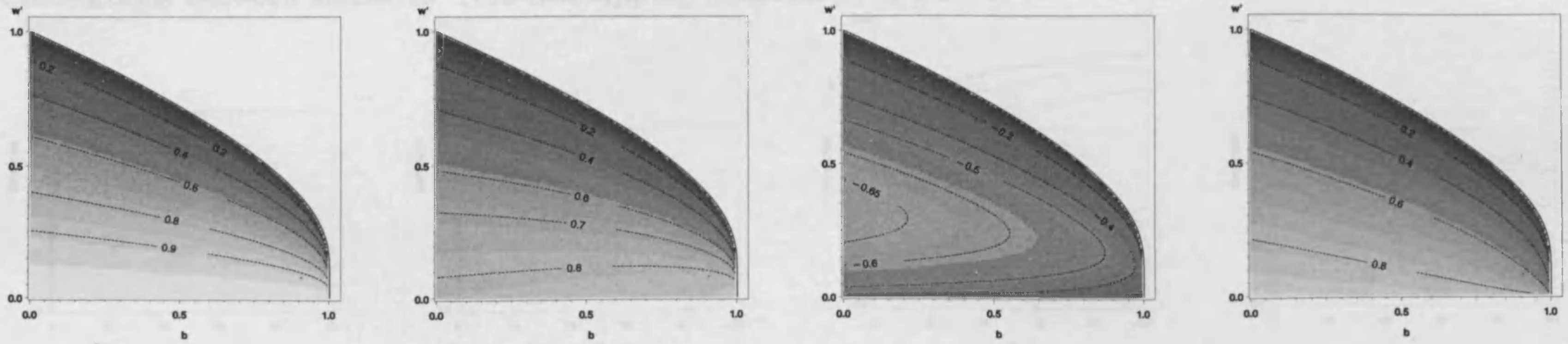
Joint distributions of statistics: Overlapping intervals $[t_1, t_3)$, $[t_2, t_4)$



(a) $\phi(z) = z,\ \psi(z) = z$ (b) $\phi(z) = z,\ \psi(z) = z^2$ (c) $\phi(z) = z,\ \psi(z) = 0.5^z$ (d) $\phi(z) = 0.25^z,\ \psi(z) = 0.75^z$

Figure 4.2: 1000 points of $\sqrt{N}(\bar{\phi} - \mathbb{E}\bar{\phi})$ versus $\sqrt{N}(\bar{\psi} - \mathbb{E}\bar{\psi})$ computed for various functions $\phi(z)$ and $\psi(z)$ when sampling from the gamma Poisson process with $m = 1, k = 1$ with samples of size $N = 1000$ in the case $t_1 = 0, t_2 = 1, t_3 = 2, t_4 = 3$. A 95% confidence ellipse based on the covariance matrix (4.1.3) and constructed under the assumption of asymptotic normality is also shown.

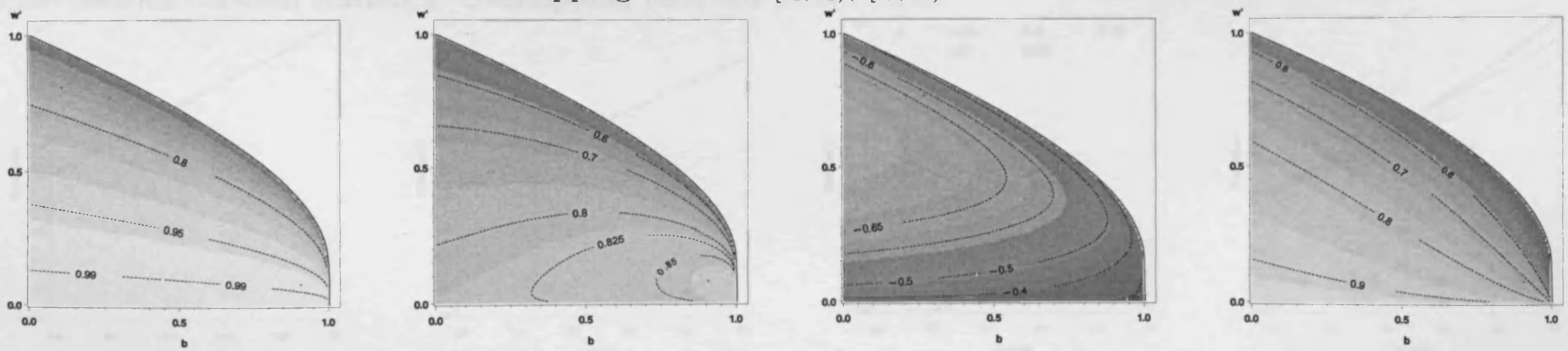
Correlations between statistics: Non-overlapping intervals $[0,t)$, $[t,t+s)$



(a) $\phi(z)=z,\ \psi(z)=z$ (b) $\phi(z)=z,\ \psi(z)=z^2$ (c) $\phi(z)=z,\ \psi(z)=0.5^z$ (d) $\phi(z)=0.25^z,\ \psi(z)=0.75^z$

Figure 4.3: Correlation $\rho(\phi(\zeta_{0,t}),\psi(\zeta_{t,t+s}))=\mathrm{Cov}\left[\phi(\zeta_{0,t}),\psi(\zeta_{t,t+s})\right]/\sqrt{\mathrm{Var}\phi(\zeta_{0,t})\mathrm{Var}\psi(\zeta_{t,t+s})}$ plotted for all NBD parameter values for various functions ϕ and ψ in the case $t=1$ and $s=1$ when sampling from the gamma Poisson process.

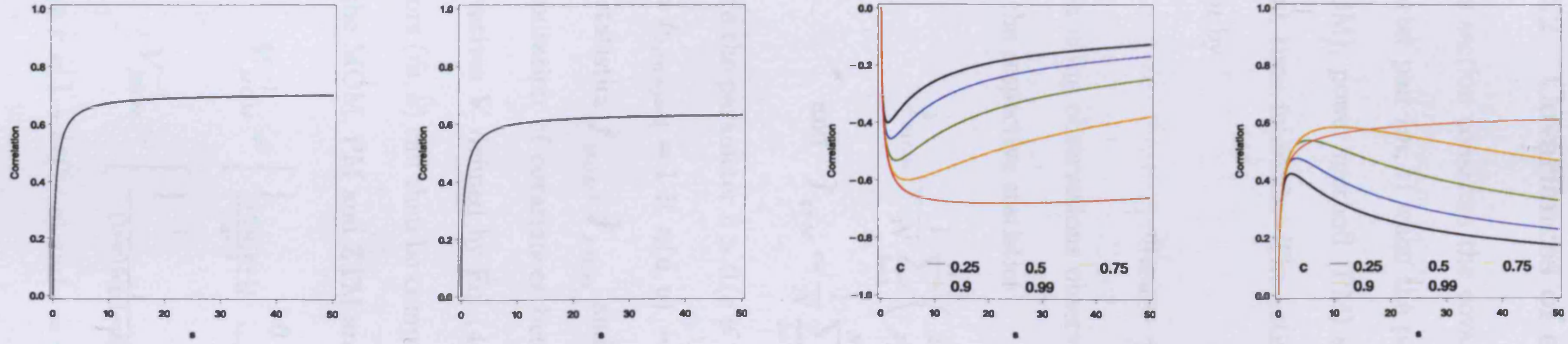
Correlations between statistics: Overlapping intervals $[t_1,t_3)$, $[t_2,t_4)$



(a) $\phi(z)=z,\ \psi(z)=z$ (b) $\phi(z)=z,\ \psi(z)=z^2$ (c) $\phi(z)=z,\ \psi(z)=0.5^z$ (d) $\phi(z)=0.25^z,\ \psi(z)=0.75^z$

Figure 4.4: Correlation $\rho(\phi(\zeta_{t_1,t_3}),\psi(\zeta_{t_2,t_4}))=\mathrm{Cov}\left[\phi(\zeta_{t_1,t_3}),\psi(\zeta_{t_2,t_4})\right]/\sqrt{\mathrm{Var}\phi(\zeta_{t_1,t_3})\mathrm{Var}\psi(\zeta_{t_2,t4})}$ plotted for all NBD parameter values for various functions ϕ and ψ in the case $t_1=0,t_2=1,t_3=2,t_4=3$ when sampling from the gamma Poisson process.

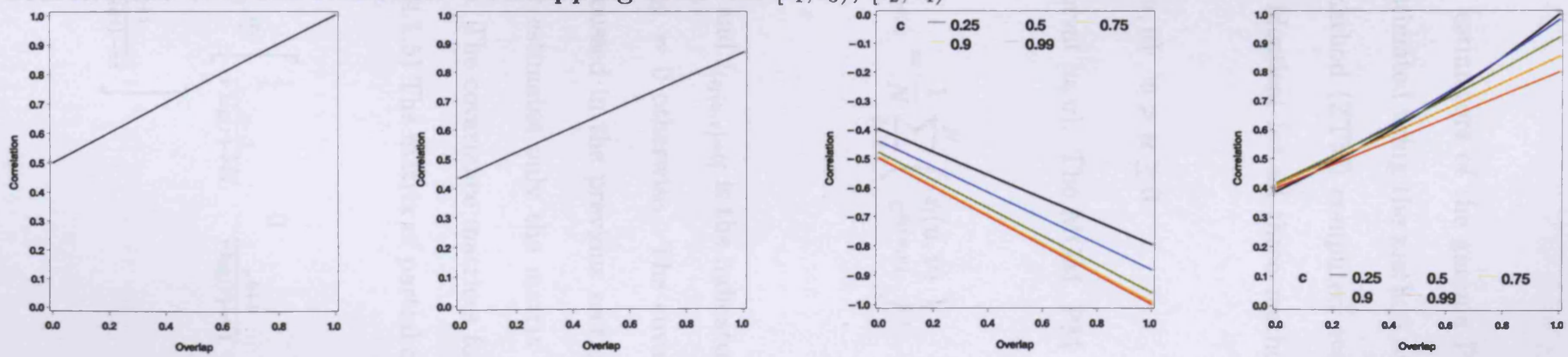
Correlations between statistics: Non-overlapping intervals $[0, t)$, $[t, t+s)$



(a) $\phi(z) = z,\ \psi(z) = z$ (b) $\phi(z) = z,\ \psi(z) = z^2$ (c) $\phi(z) = z,\ \psi(z) = c^z$ (d) $\phi(z) = 0.25^z,\ \psi(z) = c^z$

Figure 4.5: Correlation $\rho(\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})) = \mathrm{Cov}\left[\phi(\zeta_{0,t}), \psi(\zeta_{t,t+s})\right] / \sqrt{\mathrm{Var}\phi(\zeta_{0,t})\mathrm{Var}\psi(\zeta_{t,t+s})}$ plotted against s for various functions ϕ and ψ in the case $t = 1$ when sampling from the gamma Poisson process with $m = 1$ and $k = 1$.

Correlations between statistics: Overlapping intervals $[t_1, t_3)$, $[t_2, t_4)$



(a) $\phi(z) = z,\ \psi(z) = z$ (b) $\phi(z) = z,\ \psi(z) = z^2$ (c) $\phi(z) = z,\ \psi(z) = c^z$ (d) $\phi(z) = 0.25^z,\ \psi(z) = c^z$

Figure 4.6: Correlation $\rho(\phi(\zeta_{t_1,t_3}), \psi(\zeta_{t_2,t_4})) = \mathrm{Cov}\left[\phi(\zeta_{t_1,t_3}), \psi(\zeta_{t_2,t_4})\right] / \sqrt{\mathrm{Var}\phi(\zeta_{t_1,t_3})\mathrm{Var}\psi(\zeta_{t_2,t_4})}$, where $t_3 - t_1 = t_4 - t_2 = 1$ plotted against the overlap $t_3 - t_2$ for various functions ϕ and ψ when sampling from the gamma Poisson process with $m = 1$ and $k = 1$.

4.2.2 Covariances of estimators

This section considers the covariances between estimators of the gamma Poisson parameter pair (m, k) when the parameters are estimated using the method of moments (MOM), power method (PM) and zero term method (ZTM) computed over two different time intervals. The estimator for m is identical for all three methods and is given by

$$\hat{m}_{v-u} = \frac{1}{N(v-u)} \sum_{l=1}^{N} z_l(u,v) \quad v > u \geq 0$$

when using observations observed over the interval $[u, v)$. The MOM, PM and ZTM use the respective statistics

$$\bar{\boldsymbol{f}}_{MOM} = \frac{1}{N} \sum_{l=1}^{N} \begin{pmatrix} z_l(u,v) \\ z_l^2(u,v) \end{pmatrix}, \quad \bar{\boldsymbol{f}}_{PM(c)} = \frac{1}{N} \sum_{l=1}^{N} \begin{pmatrix} z_l(u,v) \\ c^{z_l(u,v)} \end{pmatrix},$$

$$\text{and} \quad \bar{\boldsymbol{f}}_{ZTM} = \frac{1}{N} \sum_{l=1}^{N} \begin{pmatrix} z_l(u,v) \\ I_{[z_l(u,v)=0]} \end{pmatrix}, \tag{4.2.1}$$

where the parameter $c > 0\,(c \neq 1)$ is a constant and $I_{[z_l(u,v)=0]}$ is the indicator function with $I_{[z_l(u,v)=0]} = 1$ if $z_l(u,v) = 0$ and $I_{[z_l(u,v)=0]} = 0$ otherwise. The covariances of the statistics $\bar{\boldsymbol{f}}_{MOM}$, $\bar{\boldsymbol{f}}_{PM(c)}$ and $\bar{\boldsymbol{f}}_{ZTM}$ were discussed in the previous section. In the computation of covariances between parameter estimates only the matrix of partial derivatives $\boldsymbol{V}$ defined by Eq. (4.1.4) is required. The covariance matrices for the estimators $(\hat{m}, \hat{k})$ can then be computed using Eq. (4.1.5) The matrix of partial derivatives for the MOM, PM and ZTM are respectively

$$\boldsymbol{V}^{-1}_{MOM} = \begin{bmatrix} \frac{1}{t} & 0 \\ \frac{1+2at(k+1)}{a^2t^2} & -\frac{1}{a^2t^2} \end{bmatrix}, \quad \boldsymbol{V}^{-1}_{PM(c)} = \begin{bmatrix} \frac{1}{t} & 0 \\ \frac{c-1}{r\log(r)-r+1} & -\frac{r^{k+1}}{r\log(r)-r+1} \end{bmatrix},$$

$$\boldsymbol{V}^{-1}_{ZTM} = \begin{bmatrix} \frac{1}{t} & 0 \\ -\frac{1}{(1+at)\log(1+at)-at} & -\frac{(1+at)^{k+1}}{(1+at)\log(1+at)-at} \end{bmatrix},$$

where $r = 1 + at(1-c)$ and $t = v - u$.

The covariance matrices for estimators $(\hat{m}, \hat{k})$ computed in different time intervals are analytically simple only in the case of non-overlapping intervals for the MOM, PM and ZTM. The covariance matrix of the MOM estimators is

$$\mathbb{D}\begin{pmatrix} \hat{m}_t \\ \hat{k}_t \\ \hat{m}_s \\ \hat{k}_s \end{pmatrix} = \begin{bmatrix} \frac{a\,k\,(1+a\,t)}{t} & 0 & a^2k & 0 \\ 0 & \frac{2k(k+1)(1+a\,t)^2}{a^2t^2} & 0 & 2\,k\,(k+1) \\ a^2k & 0 & \frac{a\,k\,(1+a\,s)}{s} & 0 \\ 0 & 2\,k\,(k+1) & 0 & \frac{2k(k+1)(1+a\,s)^2}{a^2s^2} \end{bmatrix} \tag{4.2.2}$$

and the covariance matrix for the PM and ZTM estimators is

$$\mathbb{D}\begin{pmatrix} \hat{m}_t \\ \hat{k}_t \\ \hat{m}_s \\ \hat{k}_s \end{pmatrix} = \begin{bmatrix} \frac{a\,k\,(1+a\,t)}{t} & 0 & a^2k & 0 \\ 0 & v_{PM}(c;t) & 0 & \mathbb{D}_{2,4} \\ a^2k & 0 & \frac{a\,k\,(1+a\,s)}{s} & 0 \\ 0 & \mathbb{D}_{4,2} & 0 & v_{PM}(c;s) \end{bmatrix} \tag{4.2.3}$$

$$\mathbb{D}_{2,4} = \mathbb{D}_{4,2} = \frac{{r_t}^{k+1}{r_s}^{k+1}\,(r_t + r_s - 1)^{-k} - r_t\,r_s - (1-c)^2\,a^2 t\,s\,k}{(r_t\,\log(r_t) - r_t + 1)\,(r_s\,\log(r_s) - r_s + 1)},$$

$$v_{PM}(c;u) = \frac{(1+au-auc^2)^{-k}r_u^{2k+2} - r_u^2 - kau(1+au)(1-c)^2}{[r_u\log(r_u) - r_u + 1]^2},$$

where $r_u = 1 + au(1-c)$. For the ZTM the matrix $\mathbb{D}$ can be computed using (4.2.3) with $c = 0$.

Fig. 4.7 and Fig. 4.8 show bivariate plots of various estimators $\hat{\theta}(t)$ and $\hat{\theta}(s)$ computed in different time intervals for 1000 replications of the gamma Poisson process with sample size N=1000 with $m = 1$ and $k = 1$. A 95% confidence ellipse based on the covariance matrix (4.2.3) and constructed under the assumption of asymptotic normality is also shown. Figures are shown for the two cases of overlapping and non-overlapping time intervals and confirm the results of this section.

Fig. 4.9 and Fig. 4.10 show correlations between $\rho(\hat{\theta}(t), \hat{\theta}(s))$, where $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in non-overlapping and overlapping time intervals. For fixed time intervals, the correlations for estimators of m and k increases as w increases for the MOM, PM and ZTM estimators.

Fig. 4.11 shows correlations between $\rho(\hat{\theta}(t), \hat{\theta}(s))$ against the time interval s, where $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in non-overlapping intervals in the case when $t = 1$, $m = 1$ and $k = 1$. The correlation $\rho(\hat{\theta}(t), \hat{\theta}(s))$, in the case where $\theta^* = m$ and θ^* is the MOM estimator for k, increases to a constant as the length of the second time interval increases. For any set of parameter values (a, k), where $a = m/k$, it is straightforward to show using (4.2.2) that

$$\frac{\partial}{\partial s}\rho(\hat{m}(t), \hat{m}(s)) = \left(\frac{1}{1+as}\right)^{\frac{3}{2}} \frac{at}{2\sqrt{(1+at)(1+as)ts}} \quad \text{and} \quad \lim_{s\to\infty} \rho(\hat{m}(t), \hat{m}(s)) = \sqrt{\frac{at}{1+at}},$$

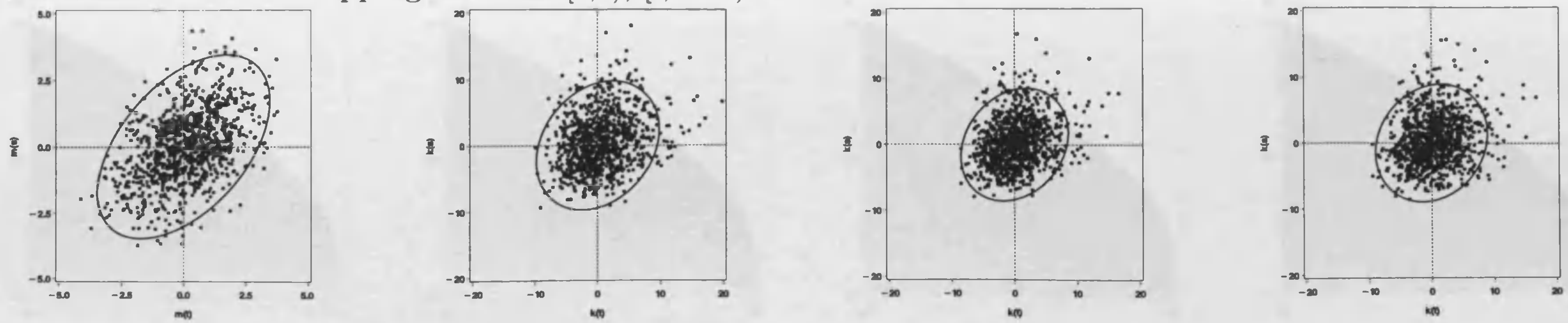
so that the derivative is positive for all $a > 0$, $k > 0$ and the correlation tends to a constant as $s \to \infty$ for the ML, MOM, PM and ZTM estimator of m. For the MOM it is straightforward to show using (4.2.2) that

$$\frac{\partial}{\partial s}\rho(\hat{k}(t), \hat{k}(s)) = \frac{a^2 t}{(1+at)(1+as)^2} \quad \text{and} \quad \lim_{s\to\infty} \rho(\hat{k}(t), \hat{k}(s)) = \frac{at}{1+at},$$

so that the derivative is positive for all $a > 0$, $k > 0$. Therefore, the correlation $\rho(\hat{k}(t), \hat{k}(s))$ for MOM estimators of k is also strictly increasing and tends to a constant. There is no simple equivalent limiting form for the covariance between the PM estimators for k.

Fig. 4.12 shows correlations between $\rho(\hat{\theta}(t), \hat{\theta}(s))$, where $t = t_3 - t_1 = 1$ and $s = t_4 - t_2 = 1$ for $t_1 \le t_2 \leqslant t_3 \le t_4$, against the length of overlap $t_3 - t_2$. Here $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in overlapping time intervals of length $t = t_3 - t_1 = 1$ and $s = t_4 - t_2 = 1$, in the case when $m = 1$ and $k = 1$. For the MOM, PM and ZTM methods, the correlation of estimators of m and k increase as the length of the overlap increases. For any fixed length of s and t and any set of parameter values (a, k), as the length of the overlap tends to zero, the correlations are equivalent to the correlations in non-overlapping intervals of the same length s and t.

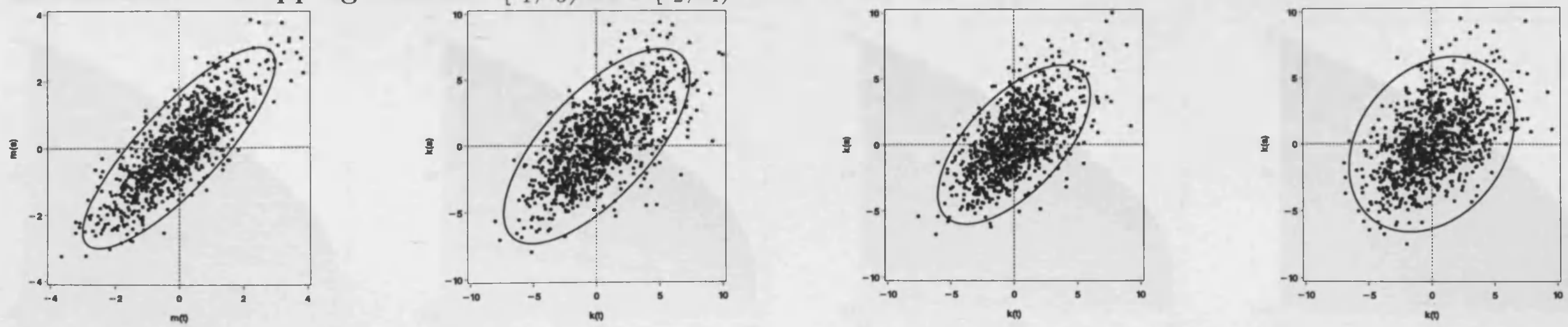
Estimators: Non-overlapping intervals $[0, t)$, $[t, t+s)$



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(0.5)) (d) $\theta^* = k$ (ZTM)

Figure 4.7: 1000 points of $\sqrt{N}(\hat{\theta}(t) - \theta^*(t))$ versus $\sqrt{N}(\hat{\theta}(s) - \theta^*(s))$ computed from data in the time intervals $[0, t)$ and $[t, t+s)$ respectively when sampling from the gamma Poisson process with $m = 1, k = 1$ and samples of size $N = 1000$ in the case $t = 1, s = 1$. A 95% confidence ellipse based on the covariance matrix (4.1.5) and constructed under the assumption of asymptotic normality is also shown.

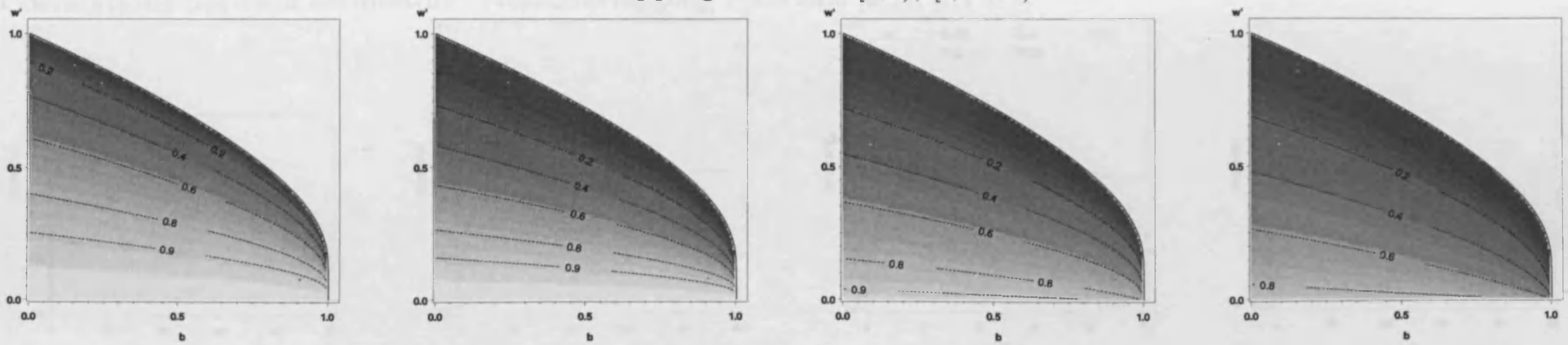
Estimators: Overlapping intervals $[t_1, t_3)$ and $[t_2, t_4)$



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(0.5)) (d) $\theta^* = k$ (ZTM)

Figure 4.8: 1000 points of $\sqrt{N}(\hat{\theta}(t) - \theta^*(t))$ versus $\sqrt{N}(\hat{\theta}(s) - \theta^*(s))$ computed from data in the time intervals $[t_1, t_3)$ and $[t_2, t_4)$ respectively when sampling from the gamma Poisson process with $m = 1, k = 1$ and samples of size $N = 1000$ in the case $t_1 = 0, t_2 = 1, t_3 = 2, t_4 = 3$. A 95% confidence ellipse based on the covariance matrix (4.1.5) and constructed under the assumption of asymptotic normality is also shown.

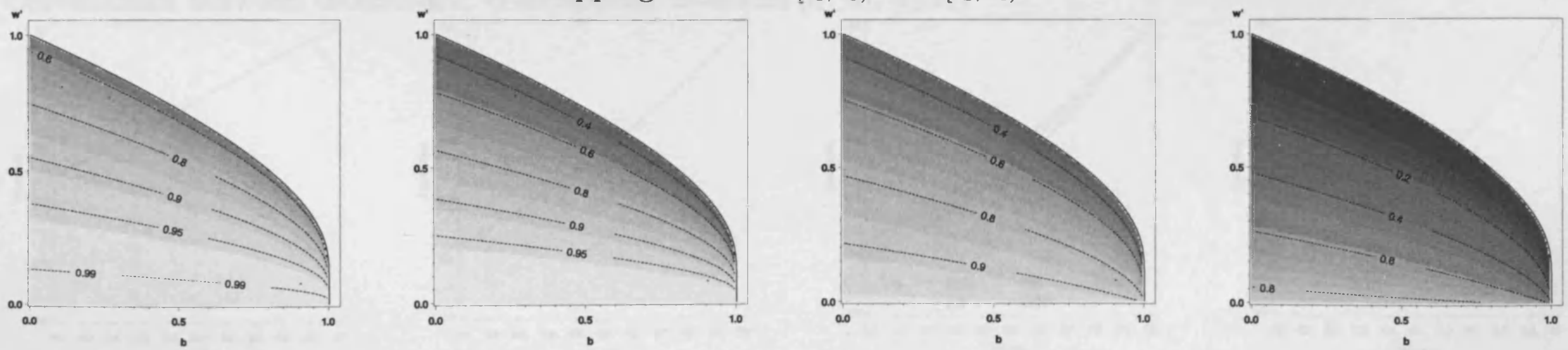
Correlations between estimators: Non-overlapping intervals $[0, t)$, $[t, t+s)$



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(0.5)) (d) $\theta^* = k$ (ZTM)

Figure 4.9: Correlation $\rho(\hat{\theta}(t), \hat{\theta}(s)) = \mathrm{Cov}[\hat{\theta}(t), \hat{\theta}(s)]/\sqrt{\mathrm{Var}[\hat{\theta}(t)]\mathrm{Var}[\hat{\theta}(s)]}$, where $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in the time intervals $[0, t)$, $[t, t+s)$ respectively. Correlations are plotted for all NBD parameter values in the case $t = 1$ and $s = 1$ when sampling from the gamma Poisson process.

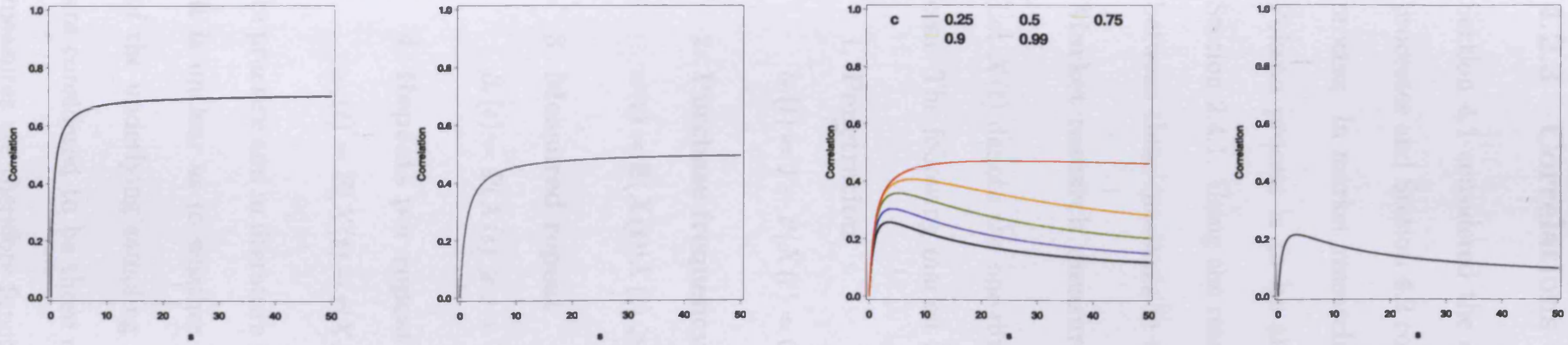
Correlations between estimators: Overlapping intervals $[t_1, t_3)$ and $[t_2, t_4)$



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(0.5)) (d) $\theta^* = k$ (ZTM)

Figure 4.10: Correlation $\rho(\hat{\theta}(t), \hat{\theta}(s)) = \mathrm{Cov}[\hat{\theta}(t), \hat{\theta}(s)]/\sqrt{\mathrm{Var}[\hat{\theta}(t)]\mathrm{Var}[\hat{\theta}(s)]}$, where $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in the time intervals $[t_1, t_3)$ and $[t_2, t_4)$ respectively. Correlations are plotted for all NBD parameter values in the case $t_1 = 0, t_2 = 1, t_3 = 2, t_4 = 3$ when sampling from the gamma Poisson process.

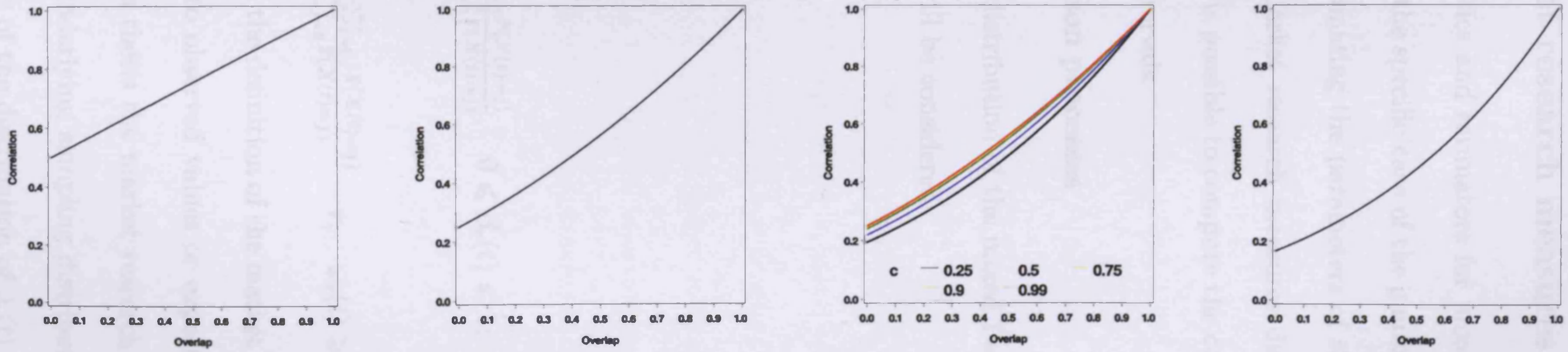
Correlations between estimators: Non-overlapping intervals $[0,t)$, $[t,t+s)$



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(c)) (d) $\theta^* = k$ (ZTM)

Figure 4.11: Correlation $\rho(\hat{\theta}(t), \hat{\theta}(s)) = \mathrm{Cov}[\hat{\theta}(t), \hat{\theta}(s)]/\sqrt{\mathrm{Var}[\hat{\theta}(t)]\mathrm{Var}[\hat{\theta}(s)]}$ versus s, where $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in the time intervals $[0,t)$, $[t,t+s)$ respectively. Correlations are plotted for $m=1$ and $k=1$ in the case $t=1$ when sampling from the gamma Poisson process.

Correlations between estimators: Overlapping intervals $[t_1,t_3)$, $[t_2,t_4)$



(a) $\theta^* = m$ (b) $\theta^* = k$ (MOM) (c) $\theta^* = k$ (PM(c)) (d) $\theta^* = k$ (ZTM)

Figure 4.12: Correlation $\rho(\hat{\theta}(t), \hat{\theta}(s)) = \mathrm{Cov}[\hat{\theta}(t), \hat{\theta}(s)]/\sqrt{\mathrm{Var}[\hat{\theta}(t)]\mathrm{Var}[\hat{\theta}(s)]}$ versus overlap $t_3 - t_2$, where $\hat{\theta}(t)$ and $\hat{\theta}(s)$ are different estimators of the same parameter θ^* computed using data in the time intervals $[t_1,t_3)$ and $[t_2,t_4)$ respectively. Correlations are plotted for $m=1$ and $k=1$ in the case $t_3 - t_1 = t_4 - t_2 = 1$ when sampling from the gamma Poisson process.

4.2.3 Correlations between market research measures

Section 4.1 considered the covariances of statistics and estimators for mixed Poisson processes and Section 4.2 considered results for the specific case of the gamma Poisson process. In market research, the purpose of estimating the parameters of the gamma Poisson process is to be able to predict the market research measures discussed in Section 2.4.1. Using the results of Section 2.2 it is possible to compute the correlations between these measures in two different time intervals.

Market research measures for mixed Poisson processes

Let $X(t)$ denote the one-dimensional marginal distribution of the mixed Poisson process. The following market research measures will be considered:

1. **Penetration**

   $b_0(t) = 1 - \mathbb{P}(X(t) = 0), \quad 0 \leqslant b(t) \leqslant 1;$

2. **Purchase frequency**

   $w(t) = \mathbb{E}(X(t)|X(t) \geqslant 1), \qquad w(t) \geqslant 1;$

3. **Measured repeat**

   $\beta_r(t) = \mathbb{P}(X(t) \geqslant r+1|X(t) \geqslant r) = \frac{1-\sum_{j=0}^{r}\mathbb{P}(X(t)=j)}{1-\sum_{j=0}^{r-1}\mathbb{P}(X(t)=j)}, \quad 0 \leqslant \beta_r(t) \leqslant 1;$

4. **Repeats per repeater**

   $\omega_r(t) = \mathbb{E}(X(t) - r|X(t) \geqslant r+1) = \frac{\mathbb{E}X(t)-\sum_{j=0}^{r} j\,\mathbb{P}(X(t)=j)}{1-\sum_{j=0}^{r}\mathbb{P}(X(t)=j)} - r, \quad \omega_r(t) \geqslant 1.$

In practice and in literature there is ambiguity in the definition of the market measures. It is unclear as to whether the measures refer to observed values or expected values of the underlying sampling distribution. In this thesis the market research measures are considered to be those obtained from the underlying sampling distribution. The measures are therefore functions of the moments of the distribution of $X(t)$.

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^T$ denote the vector of parameters for the one-dimensional marginal distribution of $X(t)$. The theoretical market research measures are straightforward to compute from knowledge of the one-dimensional marginal distribution of $X(t)$. The market measures may be estimated by using different estimators $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. The covariances of the market research measures will therefore depend on the estimation method used to estimate the parameter vector $\boldsymbol{\theta}$.

Computing the covariance matrix

The asymptotic normal distribution of different estimators of (m, k) using maximum likelihood and generalized moment based estimators given a sample of i.i.d. NBD observations was considered in Chapter 3. The asymptotic normal distribution of estimators of (m, k) computed in two different intervals was considered in Section 4.2. Using the results of these sections and the theory given in Section 2.2 it is possible to derive the covariance of market measures computed in two different time intervals when the underlying process is gamma Poisson. Let $\widehat{m_t}$ and $\widehat{k_t}$ be parameter estimates of the gamma Poisson process using data observed over a time interval of length t. The market research measures of penetration, mean purchase frequency, measured repeat and repeats per repeater are respectively given by

$$\widehat{b}(t) = 1 - \left(1 + \frac{\widehat{m_t}t}{\widehat{k_t}}\right)^{-\widehat{k_t}}, \quad \widehat{w}(t) = \frac{\widehat{m_t}t}{\widehat{b}(t)}, \quad \widehat{\beta_r}(t) = \frac{1 - \sum_{j=0}^{r} \widehat{\mathbb{P}}(X(t) = j)}{1 - \sum_{j=0}^{r-1} \widehat{\mathbb{P}}(X(t) = j)}$$

$$\text{and} \quad \widehat{\omega_r}(t) = \frac{\widehat{m_t}t - \sum_{j=0}^{r} j\, \widehat{\mathbb{P}}(X(t) = j)}{1 - \sum_{j=0}^{r} \widehat{\mathbb{P}}(X(t) = j)} - r.$$

where

$$\widehat{\mathbb{P}}(X(t) = j) = \frac{\Gamma(\widehat{k_t} + j)}{\Gamma(\widehat{k_t})j!} \left(1 + \frac{\widehat{m_t}t}{\widehat{k_t}}\right)^{-\widehat{k_t}} \left(\frac{\widehat{m_t}t}{\widehat{m_t}t + \widehat{k_t}}\right)^{j}.$$

It should be noted that for finite sample sizes $\widehat{\beta}_r(t)$ and $\widehat{\omega}_r(t)$ are biased estimators of $\beta_r(t)$ and $\omega_r(t)$ respectively. They are, however, asymptotically unbiased and the asymptotic distributions of the estimator will therefore remain unaffected.

Since the estimators for m and k are asymptotically uncorrelated, the asymptotic normalized covariance matrix is of the form

$$\mathbb{D}\begin{pmatrix}\hat{m}_t\\ \hat{k}_t\\ \hat{m}_s\\ \hat{k}_s\end{pmatrix} = \lim_{N\to\infty} N \begin{bmatrix} \mathrm{Var}(\hat{m}_t) & 0 & \mathrm{Cov}(\hat{m}_t,\hat{m}_s) & 0 \\ 0 & \mathrm{Var}(\hat{k}_t) & 0 & \mathrm{Cov}(\hat{k}_t,\hat{k}_s) \\ \mathrm{Cov}(\hat{m}_t,\hat{m}_s) & 0 & \mathrm{Var}(\hat{m}_s) & 0 \\ 0 & \mathrm{Cov}(\hat{k}_t,\hat{k}_s) & 0 & \mathrm{Var}(\hat{k}_s) \end{bmatrix} \tag{4.2.4}$$

In the computation of covariances between the same market measures computed in different time intervals, only the matrix of partial derivatives $\boldsymbol{V}$ defined by Eq. (4.1.4) is required. The matrices of partial derivatives are

$$\begin{aligned} \boldsymbol{V}_b^{-1} &= \begin{bmatrix} b_{[1,t]} & b_{[2,t]} & 0 & 0 \\ 0 & 0 & b_{[1,s]} & b_{[2,s]} \end{bmatrix}, \boldsymbol{V}_w^{-1} = \begin{bmatrix} w_{[1,t]} & w_{[2,t]} & 0 & 0 \\ 0 & 0 & w_{[1,s]} & w_{[2,s]} \end{bmatrix} \\ \boldsymbol{V}_{\beta_r}^{-1} &= \begin{bmatrix} \beta_{[1,t]} & \beta_{[2,t]} & 0 & 0 \\ 0 & 0 & \beta_{[1,s]} & \beta_{[2,s]} \end{bmatrix}, \boldsymbol{V}_{\omega_r}^{-1} = \begin{bmatrix} \omega_{[1,t]} & \omega_{[2,t]} & 0 & 0 \\ 0 & 0 & \omega_{[1,s]} & \omega_{[2,s]} \end{bmatrix} \end{aligned} \tag{4.2.5}$$

where

$$\begin{aligned} b_{[1,u]} &= u\left(1+\frac{m_t u}{k_t}\right)^{-k_t-1} \\ b_{[2,u]} &= \left(1+\frac{m_t u}{k_t}\right)^{-k_t-1}\left[\log\left(1+\frac{m_t}{k_t}\right)+\frac{m_t u}{k_t}\log\left(1+\frac{m_t}{k_t}\right)+\frac{m_t u}{k_t}\right] \\ w_{[1,u]} &= -\frac{t\left(-k-mt+kb+bmt+mtbk\right)}{\left(-1+b\right)^2\left(k+mt\right)} \\ w_{[2,u]} &= -\frac{bmt\left(k\ln\left(\frac{k+mt}{k}\right)+\ln\left(\frac{k+mt}{k}\right)mt-mt\right)}{\left(-1+b\right)^2\left(k+mt\right)} \\ \beta_{[1,u]} &= \frac{\left(1-\sum_{j=0}^{r-1}\widehat{p}_j(u)\right)\frac{\partial}{\partial m}\widehat{p}_r(u)+\widehat{p}_r(u)\sum_{j=0}^{r-1}\frac{\partial}{\partial m}\widehat{p}_j(u)}{\left(1-\sum_{j=0}^{r-1}\widehat{p}_j(u)\right)^2} \\ \beta_{[2,u]} &= \frac{\left(1-\sum_{j=0}^{r-1}\widehat{p}_j(u)\right)\frac{\partial}{\partial k}\widehat{p}_r(u)+\widehat{p}_r(u)\sum_{j=0}^{r-1}\frac{\partial}{\partial k}\widehat{p}_j(u)}{\left(1-\sum_{j=0}^{r-1}\widehat{p}_j(u)\right)^2} \end{aligned}$$

$$\omega_{[1,u]} = \frac{\left(1-\sum_{j=0}^{r}\widehat{p_j}(u)\right)\left(u-\sum_{j=0}^{r} j\frac{\partial}{\partial m}\widehat{p_j}(u)\right)+\left(\hat{m}u-\sum_{j=0}^{r} j\widehat{p_j}(u)\right)\sum_{j=0}^{r}\frac{\partial}{\partial m}\widehat{p_j}(u)}{\left(1-\sum_{j=0}^{r}\widehat{p_j}(u)\right)^2}$$

$$\omega_{[2,u]} = \frac{\left(1-\sum_{j=0}^{r}\widehat{p_j}(u)\right)\left(-\sum_{j=0}^{r} j\frac{\partial}{\partial k}\widehat{p_j}(u)\right)+\left(\hat{m}u-\sum_{j=0}^{r} j\widehat{p_j}(u)\right)\sum_{j=0}^{r}\frac{\partial}{\partial k}\widehat{p_j}(u)}{\left(1-\sum_{j=0}^{r}\widehat{p_j}(u)\right)^2}$$

Here

$$\frac{\partial}{\partial m}\widehat{p_j}(u) = \frac{(j-mu)k}{m(mu+k)}\widehat{p_j}(u) \quad \text{and}$$
$$\frac{\partial}{\partial k}\widehat{p_j}(u) = \left[\Psi(k+j)-\Psi(k)-\log\left(1+\frac{mu}{k}\right)+\frac{mu-j}{mu+k}\right]\widehat{p_j}(u)$$

and $\Psi(\cdot)$ is the digamma function. Let $\mathbb{D}_{[i,j]}$ be the elements of the matrix (4.2.4) and let $\boldsymbol{V}_{[i,j]}$ denote the elements of one of the diagonal matrix of partial derivatives as given in Eq. (4.2.5). Then $\boldsymbol{V}_{[i,j]}\mathbb{D}_{[i,j]}\boldsymbol{V}^{-1}_{[i,j]}$ is of the form

$$\boldsymbol{V}^{-1}_{b} = \begin{bmatrix} b^2_{[1,t]}\mathbb{D}_{[1,1]}+b^2_{[2,t]}\mathbb{D}_{[1,2]} & b_{[1,t]}b_{[1,s]}\mathbb{D}_{[1,3]}+b_{[2,t]}b_{[2,s]}\mathbb{D}_{[2,4]} \\ b_{[1,t]}b_{[1,s]}\mathbb{D}_{[1,3]}+b_{[2,t]}b_{[2,s]}\mathbb{D}_{[2,4]} & b^2_{[1,s]}\mathbb{D}_{[3,3]}+b^2_{[2,s]}\mathbb{D}_{[4,4]} \end{bmatrix}.$$

Hypothesis testing

The construction of the joint asymptotic normal distributions of statistics and functionals of data whose underlying distribution is the gamma Poisson process, and also mixed Poisson processes in general, has the important consequence that the limiting distribution can be used in the testing of various hypotheses. For example, suppose we have a vector of estimators $\hat{\boldsymbol{\theta}} = (\widehat{\theta^{(1)}}, \widehat{\theta^{(2)}}, \ldots, \widehat{\theta^{(n)}})$ of the vector of parameters with identical entries θ, computed in time intervals $1, 2, \ldots, n$. Then it is possible to check whether the vector of estimators falls within the confidence ellipsoid of the corresponding n-dimensional asymptotic normal distribution of $\hat{\boldsymbol{\theta}}$ for a specified significance level. In performing the hypothesis test, it will be important to consider the power of the test so that the test minimizes the probability of a Type 1 error.

4.3 The NBD INAR(1) process

This section considers the dynamic properties of the INAR(1) process by considering the correlations and spectral representations of the INAR(1) process. The INAR(1) process is an ergodic process. The correlations of statistics computed in an ensemble of realizations therefore represents the correlations of statistics computed from a single realization of the process by using the autocorrelation function. As well as considering the autocorrelation function (time domain analysis), the spectral representation of the process is also considered (see e.g. Priestley (1981)). Spectral domain analysis considers the decomposition of time series into frequency components and is commonly used in the detection of long-range dependence.

Background

The INAR(1) process with deterministic thinning. Recall the definition of an INAR(1) process. A non-negative integer-valued process $\{X_t; t \in \mathbb{Z}\}$ is said to be an INAR(1) process if the process satisfies the equation

$$X_t \overset{\mathcal{D}}{=} \alpha \circ X_{t-1} + \varepsilon_t, \tag{4.3.1}$$

where $\alpha \circ X_{t-1}$ and ε_t are mutually independent random variables from a discrete distribution and the ε_t form a sequence of uncorrelated random variables for $t \in \mathbb{Z}$. The value of α must satisfy $\alpha \in (0,1)$ for the process to be stationary. It is assumed that the X_t and ε_t have finite means and variances. The INAR(1) process X_t with marginal distribution π will have a stationary marginal distribution, i.e. $X_t \overset{\mathcal{D}}{=} X_{t-1} \overset{\mathcal{D}}{=} X_\pi$ for all $t \in \mathbb{Z}$, if and only if the random variable X_π is discrete self-decomposable so that

$$G_{X_\pi}(s) = G_{X_\pi}(1-\alpha+\alpha s)G_\varepsilon(s;\alpha) \qquad \alpha \in (0,1). \tag{4.3.2}$$

The INAR(1) process with stochastic thinning. The INAR(1) process with stochastic thinning was introduced by McKenzie (1986) in the special case where the marginal distribution of the process is NBD. Assume that there exists a non-negative integer-valued autoregressive process X_t with i.i.d. stochastic thinning parameters A_t supported on the interval $(0,1)$, then the INAR(1) process with stochastic thinning is defined by

$$X_t \overset{\mathcal{D}}{=} A_t \circ X_{t-1} + \varepsilon_t, \tag{4.3.3}$$

where for fixed t the A_t, X_{t-1} and ε_t are independent random variables. If the process X_t defined by (4.3.3) is to be a stationary process then the PGF of X_π must satisfy

$$G_{X_\pi}(c) = \int_0^1 G_{X_\pi}(1-y+yc)dF_A(y)G_\varepsilon(c;\ \alpha), \tag{4.3.4}$$

where $F_A(y)$ is the cumulative distribution function (c.d.f.) of A_t.

McKenzie (1986) derived a stationary NBD INAR(1) process with stochastic thinning by letting X_t be NBD(m,k) and letting A_t follow a Beta distribution defined by

$$f_{A_t}(y) = \frac{y^{l-1}(1-y)^{k-l-1}}{B(l,k-l)}, \quad l>0,\ k-l>0,\ 0<y<1,$$

where $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ is the beta function. The NBD INAR(1) process with stochastic thinning can be represented in terms of Eq. (4.3.4) by

$$\underbrace{\left(1+\frac{m(1-c)}{k}\right)^{-k}}_{G_{X_\pi}(c)} = \underbrace{\left(1+\frac{m(1-c)}{k}\right)^{-l}}_{\int_0^1 G_{X_\pi}(1-y+yc)dF_A(y)} \underbrace{\left(1+\frac{m(1-c)}{k}\right)^{-(k-l)}}_{G_\varepsilon(c;\ \alpha)}. \tag{4.3.5}$$

The generating function of the error distribution may be represented in the form

$$G_\varepsilon(c;\ \alpha) = \left(1+\frac{m(1-l/k)(1-c)}{k-l}\right)^{-(k-l)}$$

from which it becomes clear that the errors are NBD$_m(m(1-l/k), k-l)$.

4.3.1 The NBD INAR(1) process with mixed thinning

A more general NBD INAR(1) process can be derived as a mixture of the two processes described by Eqs. (4.3.1) and (4.3.3).

Definition 4.3.1. Let X_t be a stationary non-negative integer-valued autoregressive process of the first-order with innovation process ε_t (uncorrelated for $t \neq s$), independent of X_t. Assume that both processes have finite means and variances. Additionally let $\alpha \in (0,1)$ be a deterministic thinning parameter and A_t (independent of α) be i.i.d. stochastic thinning parameters with c.d.f. F_A concentrated on the interval $(0,1)$. Then the INAR(1) process with mixed deterministic and stochastic thinning is defined by

$$X_t \quad = \quad \alpha A_t \circ X_{t-1} + \varepsilon_t. \tag{4.3.6}$$

The generating function of the process (4.3.6) is given by

$$G_X(c) = \int_0^1 G_X(1 - y\alpha + y\alpha c)\, dF_A(y)\ \ G_\varepsilon(c). \tag{4.3.7}$$

Proposition 4.3.1. *Let the process X_t have a* NBD(m,k) *marginal distribution then X_t may be represented as a process with mixed deterministic and stochastic thinning with $A_t \sim$* Beta$(\nu, k-\nu)$ *and $\varepsilon_t \sim$* NBD$(m\alpha, k-\nu) *$ NBDG$(k, k/(k+m\alpha), \alpha)$.

Proof. The proof of the proposition is obtained directly from proving Eq. (4.3.7). Suppose X_t has a NBD(m,k) distribution then

$$\begin{aligned}
&\int_0^1 G_X(1 - y\alpha + y\alpha c) dF_A(y)\, G_\varepsilon(c)\\
&= \int_0^1 \left(1 + \frac{m\alpha y(1-c)}{k}\right)^{-k} dF_A(y) \left(1 + \frac{m\alpha(1-c)}{k}\right)^{\nu-k} \left(\frac{k+m(1-c)}{k+m\alpha(1-c)}\right)^{-k}\\
&= \left(1 + \frac{m\alpha(1-c)}{k}\right)^{-\nu} \left(1 + \frac{m\alpha(1-c)}{k}\right)^{\nu-k} \left(\frac{k+m(1-c)}{k+m\alpha(1-c)}\right)^{-k}\\
&= \left(1 + \frac{m(1-c)}{k}\right)^{-k} = G_X(s).
\end{aligned}$$

□

Proposition 4.3.2. *Let X_t be an INAR(1) process with mixed deterministic and stochastic thinning with thinning parameters given by α and A_t where A_t has distribution function F_A. Assume that the process has finite first and second moments, then the autocorrelation function of the process at lag u is given by*

$$\rho(X_t, X_{t+u}) = \rho(u) = (\alpha E[A])^{|u|}, \quad u \in \mathbb{Z}. \tag{4.3.8}$$

Proof. Let A_1 and A_2 be two random variables with c.d.f. concentrated on $(0, 1)$ then it is straightforward to show that for any non-negative integer X, the thinning operation $A_1 \circ A_2 \circ X = A_1 A_2 \circ X$. Note that using an iterative technique the process X_t in (4.3.6) may be written in terms of X_{t-u} as

$$X_t = \left(\prod_{i=0}^{u-1} \alpha A_{t-i}\right) \circ X_{t-u} + \sum_{j=1}^{u-1} \left(\prod_{i=0}^{j-1} \alpha A_{t-i}\right) \circ \varepsilon_{t-j} + \varepsilon_t. \tag{4.3.9}$$

The autocovariance function at lag u is

$$\begin{aligned}
R(u) = \mathrm{Cov}[X_t, X_{t-u}] &= \mathrm{Cov}\left[\left(\prod_{i=0}^{u-1} \alpha A_{t-i}\right) \circ X_{t-u} + \sum_{j=1}^{u-1} \left(\prod_{i=0}^{j-1} \alpha A_{t-i}\right) \circ \varepsilon_{t-j} + \varepsilon_t,\ X_{t-u}\right] \\
&= \mathrm{Cov}\left[\left(\prod_{i=0}^{u-1} \alpha A_{t-i}\right) \circ X_{t-u},\ X_{t-u}\right] + \mathrm{Cov}\left[\sum_{j=1}^{u-1} \left(\prod_{i=0}^{j-1} \alpha A_{t-i}\right) \circ \varepsilon_{t-j} + \varepsilon_t,\ X_{t-u}\right] \\
&= E\left[\prod_{i=0}^{u-1} \alpha A_{t-i}\right] \mathrm{Var}\left[X_{t-u}\right] + 0 = (\alpha E[A])^u \,\mathrm{Var}\left[X_{t-u}\right], \quad u \in \mathbb{Z}^+.
\end{aligned}$$

Note that for any $t > s$ the pair (ε_t, X_s) are uncorrelated. Additionally, using the stationarity property of the process, the variances are invariant under shifts in time so that $\mathrm{Var}[X_{t-u}] = \mathrm{Var}[X_t]$. The autocorrelation function of the process then follows directly. □

Note that by taking $\alpha = 1$, the autocorrelation function of the process with random thinning (see Eq. (4.3.3)) is $E[A]^{|u|}$, $u \in \mathbb{Z}$.

4.3.2 Integer valued processes with long memory

This section derives a long-memory non-negative integer-valued process using the approach of Barndorff-Nielsen (1998) by the aggregation, $X_t = \sum_{\eta=1}^{\infty} X_t^{(\eta)}$, of a sequence of stationary and independent INAR(1) processes $X_t^{(\eta)}$ ($\eta = 0, 1, 2, \ldots$). Here $X_t^{(\eta)}$ are of the form

$$X_t^{(\eta)} = \alpha_\eta \circ X_{t-1}^{(\eta)} + \varepsilon_t^{(\eta)}, \quad \eta = 1, 2, \ldots \quad t \in \mathbb{Z}.$$

Conditions required in order to construct long-memory processes with Poisson and NBD marginal distributions are presented followed by some simulation results of the autocovariance function and spectral density.

Proposition 4.3.3. *Let $X_t = \sum_{\eta=1}^{\infty} X_t^{(\eta)}$ be the aggregation of independent INAR(1) processes with each $X_t^{(\eta)}$ having mean $\mu_{X_\eta} < \infty$ and variance $\sigma^2_{X_\eta} < \infty$ with thinning parameter α_η. If $\sigma^2_{X_\eta}$ and α_η are of the form*

$$\sigma^2_{X_\eta} = \frac{c_1}{\eta^{1+2(1-H)}}, \quad \alpha_\eta = \exp\{-c_2 / \eta\} \tag{4.3.10}$$

with some positive constants c_1, c_2 and $0.5 < H < 1$, then on the assumption that $E[X_t] = \sum_{\eta=1}^{\infty} \mu_{X_\eta} < \infty$, the limiting aggregated processes X_t is a well defined process in the L^2 sense with long-memory (or Hurst) parameter H. The autocovariance function and the spectral density of the process are given by Eqs. (4.3.12) *and* (4.3.13) *below.*

Proof. Note that the aggregated process has a finite mean (by assumption) and finite variance, which for any $H \in (0.5, 1)$ is given by

$$\mathrm{Var}[X_t] = \sum_{\eta=1}^{\infty} \sigma^2_{X_\eta} = \sum_{\eta=1}^{\infty} \frac{c_1}{\eta^{1+2(1-H)}} < \infty. \tag{4.3.11}$$

The long-memory of the process is proved by showing that the aggregated process has an autocovariance function of the form $R(u) \simeq A_1(u)u^{-\tau}$ with $\tau \in (0,1)$ as $u \to \infty$

and spectral density of the form $f(\omega) \simeq A_2(\omega)|\omega|^{-\kappa}$ with $\kappa \in (0,1)$ as $\omega \to 0$, where both A_1 and A_2 are slowly varying functions.

The autocovariance function. Let $R^{(\eta)}(u)$ represent the autocovariance function of the individual INAR(1) processes, then under the conditions of (4.3.10), the covariance of the aggregated process at lag $u = t - s$ is given by

$$\begin{aligned}
R(u) &= \sum_{\eta=1}^{\infty} R^{(\eta)}(u) = \sum_{\eta=1}^{\infty} \mathrm{Cov}(X_t^{(\eta)}, X_{t-u}^{(\eta)}) = \sum_{\eta=1}^{\infty} \sigma^2_{X_\eta}\, \alpha_\eta^{|u|} = \sum_{\eta=1}^{\infty} \frac{c_1}{\eta^{1+2(1-H)}}\, e^{-|u|c_2/\eta} \\
&\simeq \int_1^{\infty} \frac{c_1}{x^{1+2(1-H)}}\, e^{-|u|c_2/x}\, dx = \frac{c_1}{(|u|c_2)^{2(1-H)}} \int_0^{|u|c_2} z^{2(1-H)-1} e^{-z} dz \\
&\simeq \frac{c_1 \Gamma(2(1-H))}{(|u|c_2)^{2(1-H)}} = \frac{C}{|u|^{2(1-H)}} \quad \text{as} \quad |u| \to \infty
\end{aligned} \tag{4.3.12}$$

where C is a constant, $u \in \mathbb{Z}$ and $H \in (0,1)$. Note that a substitution of $z = |u|\, c_2 \,/\, x$ was made to the integral in the third line of the proof. If $H \in (0.5, 1)$ then Eq. (4.3.12) satisfies the definition of long-memory given in Eq. (2.3.16).

The spectral density. Barndorff-Nielsen (1998) constructed a long-memory process with the same autocovariance function $R(u)$ as (4.3.12) but in continuous time so that $u \in \mathbb{R}$. The corresponding spectral density $f_c(\omega)$, $\omega \in \mathbb{R}$ therefore exists and may be obtained directly from the autocovariance function (4.3.12) (see Priestley (1981, pp. 210–226)). The identity $f(\omega) = \sum_{s=-\infty}^{\infty} f_c(\omega + 2\pi s)$ where $-\pi \leq \omega < \pi$ may then be used to find the spectral density of the discrete time process with autocovariance structure of the form (4.3.12).

Let $f_c(\omega)$ denote the spectral density of a continuous time process $\{X_t;\ t \in \mathbb{R}\}$, then the spectral density for a process with autocovariance function of the form (4.3.12) under the conditions of proposition 4.3.3 is derived on re-writing $R(u)$ as

$$R(u) = \sum_{\eta=1}^{\infty} \frac{c_1}{\eta^{1+2(1-H)}} e^{-|u|c_2/\eta} = \sum_{\eta=1}^{\infty} \frac{c_1}{\eta^{1+2(1-H)}} \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\eta c_2}{c_2^2 + \eta^2 w^2} e^{i\omega u} d\omega$$

$$= R(0) \int_{-\infty}^{\infty} \left(\frac{1}{R(0)} \sum_{\eta=1}^{\infty} \frac{c_1 c_2}{\pi} \frac{1}{\eta^{2(1-H)}} \frac{1}{c_2^2 + \omega^2 \eta^2} \right) e^{i\omega u} d\omega = R(0) \int_{-\infty}^{\infty} f_c(\omega) e^{i\omega u} d\omega.$$

Hence the spectral density of the aggregated process in continuous time with autocovariance function $R(u),\ u \in \mathbb{R}$ of the form (4.3.12) has spectral density $f_c(\omega)$ given by

$$f_c(\omega) = \frac{1}{\sigma_X^2} \sum_{\eta=1}^{\infty} \frac{c_1 c_2}{\pi} \frac{1}{\eta^{2(1-H)}} \frac{1}{c_2^2 + \omega^2 \eta^2}, \quad \omega \in \mathbb{R}.$$

The equivalent spectral density for the discrete time process $f(\omega)$ is therefore

$$f(\omega) = \sum_{s=-\infty}^{\infty} f_c(\omega + 2\pi s) = \frac{1}{\sigma_X^2} \frac{c_1 c_2}{\pi} \sum_{s=-\infty}^{\infty} \left[\sum_{\eta=1}^{\infty} \frac{1}{\eta^{2(1-H)}} \frac{1}{c_2^2 + (w + 2\pi s)^2 \eta^2} \right],$$

$$-\pi \leq \omega < \pi. \qquad (4.3.13)$$

Note that the spectral density has a pole at the origin for $H \in (0.5, 1)$. Consider the individual terms in s of the spectral density (4.3.13), then the spectral density at $s = 0,\ \omega = 0$ is given by

$$f(0)|_{s=0} = \frac{1}{\sigma_X^2} \frac{c_1}{c_2 \pi} \sum_{\eta=1}^{\infty} \frac{1}{\eta^{2(1-H)}} = \infty, \quad c_1,\ c_2 > 0, \quad \text{for} \quad H \in (0.5,\ 1).$$

The spectral density can be simplified on interchanging the summation to give

$$f(\omega) = \frac{1}{\sigma_X^2} \sum_{\eta=1}^{\infty} \frac{c_1}{\eta^{1+2(1-H)}} \frac{1}{2\pi} \frac{1 - \exp\{-2c_2/\eta\}}{1 - 2\exp\{-c_2/\eta\} \cos\omega + \exp\{-2c_2/\eta\}}$$

$$= \frac{1}{\sigma_X^2} \sum_{\eta=1}^{\infty} \frac{c_1}{\eta^{1+2(1-H)}} \frac{1}{2\pi} \frac{\cosh(c_2/2\eta) \sinh(c_2/2\eta)}{\cosh(c_2/2\eta) - \cos(w/2)^2}, \quad -\pi \leq \omega < \pi.$$

□

Example: Long memory Poisson processes

This example constructs a stationary long-memory process with a Poisson marginal distribution with mean λ, autocovariance function of the form (4.3.12) and spectral density of the form (4.3.13) by the aggregation of independent Poisson INAR(1) processes.

Proposition 4.3.4. *Let $\{X_t^{(\eta)};\ \eta = 1,2,\ldots\}$ be a sequence of stationary and independent Poisson INAR(1) processes with mean λ_η and thinning parameter $\alpha_\eta =$ $\exp\{-c_2/\eta\}$ $(c_2 > 0)$ where*

$$\lambda_\eta = \frac{\lambda}{\zeta(1+2(1-H))}\frac{1}{\eta^{1+2(1-H)}}, \quad H \in (0.5, 1)$$

and $\zeta(s) = \sum_{\eta=1}^{\infty} 1/\eta^s$ is the Riemann-Zeta function. Then the aggregated process $X_t = \sum_{\eta=1}^{\infty} X_t^{(\eta)}$ has long-memory with Hurst parameter H and a Poisson marginal distribution with mean λ and autocovariance function

$$R(u) = \frac{\lambda}{\zeta(1+2(1-H))}\sum_{\eta=1}^{\infty}\frac{\exp\{-c_2|u|/\eta\}}{\eta^{1+2(1-H)}}, \quad u \in \mathbb{Z}$$

and spectral density

$$f(\omega) = \frac{c_2}{\pi\zeta(1+2(1-H))}\sum_{s=-\infty}^{\infty}\left[\sum_{\eta=1}^{\infty}\frac{1}{\eta^{2(1-H)}}\frac{1}{c_2^2 + (w+2\pi s)^2\eta^2}\right], \quad -\pi \leq \omega < \pi.$$

Proof. The proposition is easily proved by using properties of generating functions. Note that the Poisson distribution is infinitely-divisible and hence the aggregated process $X_t = \sum_{\eta=1}^{\infty} X_t^{(\eta)}$, as a sum of independent Poisson INAR(1) processes, is well defined on the assumption that X_t is L^2 convergent. Assume that the $X_t^{(\eta)}$ follow a Poisson distribution with mean λ_η then using the fact that $G_{X_t}(c) = \prod_{\eta=1}^{\infty} G_{X_t^{(\eta)}}(c)$

and the form of $\sigma^2_{X_t^{(\eta)}}$ in (4.3.10), we require for any $t \in \mathbb{Z}$

$$\exp(-\lambda(1-c)) = \exp\left(-\sum_{\eta=1}^{\infty}\lambda_\eta(1-c)\right) = \exp\left(-\left[\sum_{\eta=1}^{\infty}\frac{c_1}{\eta^{1+2(1-H)}}\right](1-c)\right).$$

This implies that the constant c_1 and the parameter λ_η must be of the form

$$c_1 = \frac{\lambda}{\sum_{\eta=1}^{\infty}\eta^{-[1+2(1-H)]}} \Rightarrow \lambda_\eta = \lambda\left(\frac{\eta^{-[1+2(1-H)]}}{\sum_{\eta=1}^{\infty}\eta^{-[1+2(1-H)]}}\right).$$

□

It is clear from the form of λ_η that the aggregated long-memory process is a sum of weighted Poisson processes whose mean and variance tend to zero in the limit as $k \to \infty$.

Simulation results. Figure 4.13 shows part of a realization of a simulated long-memory Poisson INAR(1) process of length $t = 20000$ with Hurst parameter $H = 0.8$, $\lambda = 5$ and constant $c_2 = 0.1$. Note that the constant c_1 is restricted on specifying the marginal distribution of the long-memory process. The simulations show both the short term ($t = 100$) and long term ($t = 10000$) behaviour of the process.

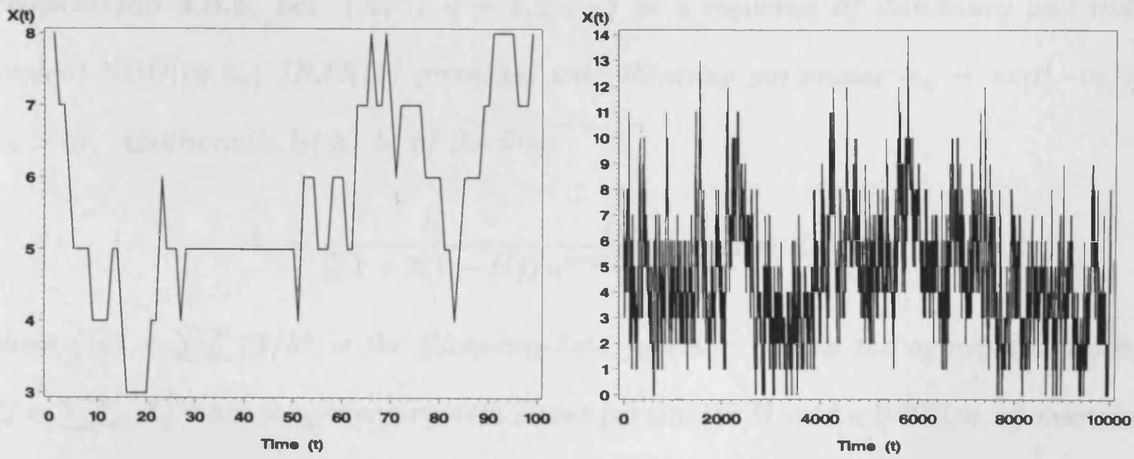


Figure 4.13: Long Memory Poisson INAR(1) Realization

Figure 4.14 shows the autocorrelation function and periodogram in logarithmic scale of the simulated long-memory process, with the solid line showing the theoretical value of the autocorrelation function and spectral density respectively.

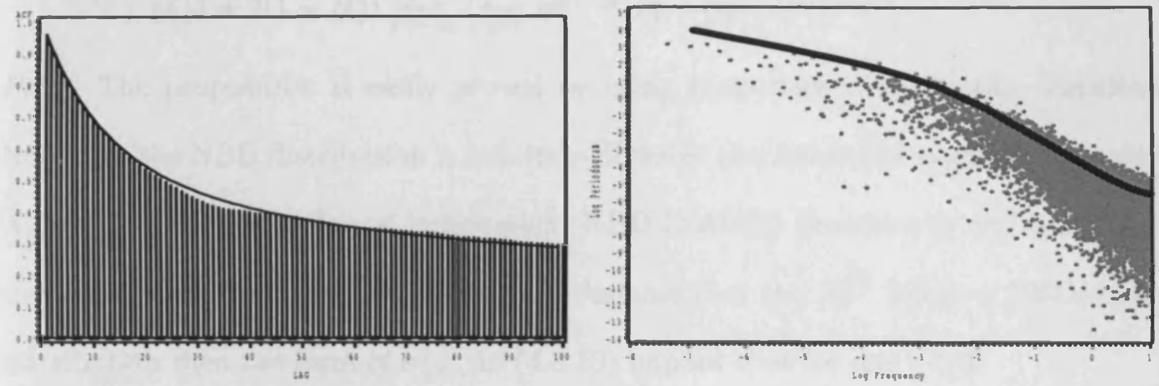


Figure 4.14: Autocorrelation function & periodogram

Example: Long memory NBD processes

This example constructs a stationary long-memory process with a negative binomial, NBD(m, k), marginal distribution with autocovariance function of the form (4.3.12) and spectral density of the form (4.3.13) by the aggregation of independent NBD INAR(1) processes.

Proposition 4.3.5. *Let* $\{X_t^{(\eta)};\ \eta = 1, 2, \ldots\}$ *be a sequence of stationary and independent* NBD(m, k_η) *INAR(1) processes with thinning parameter* $\alpha_\eta = \exp\{-c_2/\eta\}$ $(c_2 > 0)$*. Additionally let* k_η *be of the form*

$$k_\eta = \frac{k}{\zeta(1 + 2(1 - H))} \frac{1}{\eta^{1+2(1-H)}}, \quad H \in (0.5, 1)$$

where $\zeta(s) = \sum_{k=1}^{\infty} 1/k^s$ *is the Riemann-Zeta function. Then the aggregated process* $X_t = \sum_{\eta=1}^{\infty} X_t^{(\eta)}$ *has long-memory with Hurst parameter* H *and a* NBD(m, k) *marginal distribution with covariance function*

$$R(u) = \frac{m}{\zeta(1 + 2(1 - H))} \left(1 + \frac{m}{k}\right) \sum_{\eta=1}^{\infty} \frac{\exp\{-c_2|u|/\eta\}}{\eta^{1+2(1-H)}}, \quad u \in \mathbb{Z}$$

and spectral density

$$f(\omega) = \frac{c_2}{\pi\zeta(1+2(1-H))}\sum_{s=-\infty}^{\infty}\left[\sum_{\eta=1}^{\infty}\frac{1}{\eta^{2(1-H)}}\frac{1}{c_2^2+(w+2\pi s)^2\eta^2}\right], \quad -\pi \le \omega < \pi.$$

Proof. The proposition is easily proved by using properties of generating functions. Note that the NBD distribution is infinitely-divisible and hence the aggregated process $X_t = \sum_{\eta=1}^{\infty} X_t^{(\eta)}$, as a sum of independent NBD INAR(1) processes is well defined on the assumption that X_t is L^2 convergent. Assume that the $X_t^{(\eta)}$ follow a NBD(m, k_η) distribution then the form of $\sigma^2_{X^{(\eta)}}$ in (4.3.10) implies that for any $t \in \mathbb{Z}$

$$\sigma^2_{X_t^{(\eta)}} = m\left(1+\frac{m}{k_\eta}\right) = \frac{c_1}{\eta^{1+2(1-H)}} \Rightarrow k_\eta = \frac{k^2}{m(m+k)}\frac{c_1}{\eta^{1+2(1-H)}}.$$

Furthermore using the fact that $G_{X_t}(c) = \prod_{\eta=1}^{\infty} G_{X_t^{(\eta)}}(c)$, the value of k_η is obtained as

$$\begin{aligned}\left(1+\frac{m(1-c)}{k}\right)^{-k} &= \left(1+\frac{m(1-c)}{k}\right)^{-\sum_{\eta=1}^{\infty}k_\eta}\\ &\Rightarrow k = \sum_{\eta=1}^{\infty}k_\eta = \sum_{\eta=1}^{\infty}\left[\frac{k^2}{m(m+k)}\frac{c_1}{\eta^{1+2(1-H)}}\right]\\ &\Rightarrow k_\eta = k\left(\frac{\eta^{-[1+2(1-H)]}}{\sum_{\eta=1}^{\infty}\eta^{-[1+2(1-H)]}}\right).\end{aligned}$$

□

It is clear from the form of k_η that the aggregated long-memory process is also a sum of weighted NBD processes whose mean and variance tend to zero in the limit as $k \to \infty$.

Simulation results. Figure 4.15 shows part of a realization of a simulated long-memory NBD INAR(1) process of length $t = 20000$ with Hurst parameter $H = 0.8$, $m = 5$, $k = 5$ and constant $c_2 = 0.1$. Note that the constant c_1 is restricted on specifying the marginal distribution of the long-memory process. The simulations show both the short term ($t = 500$) and long term ($t = 10000$) behaviour of the process.

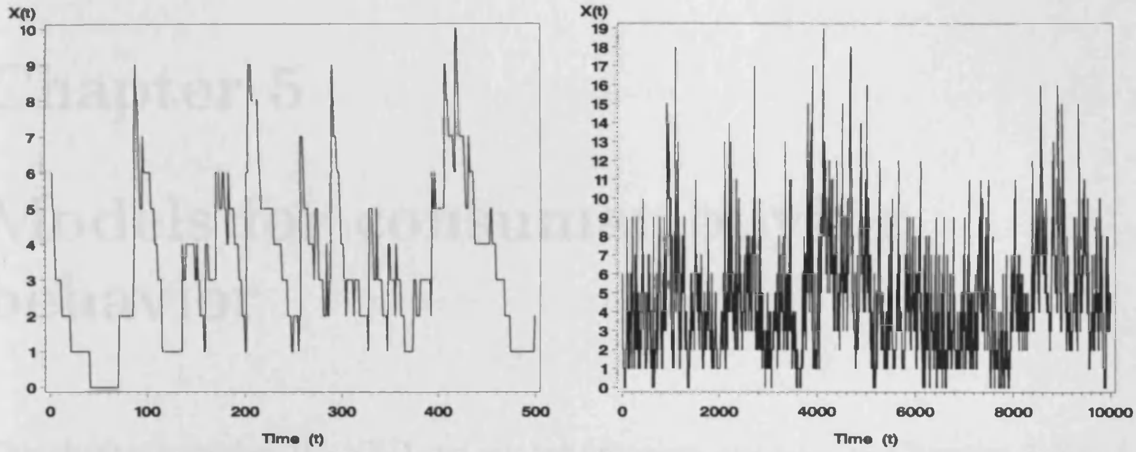


Figure 4.15: Long Memory NBD INAR(1) Realization

Figure 4.16 shows the autocorrelation function and periodogram in logarithmic scale of the simulated long-memory process, with the solid line showing the theoretical value of the autocorrelation function and spectral density respectively.

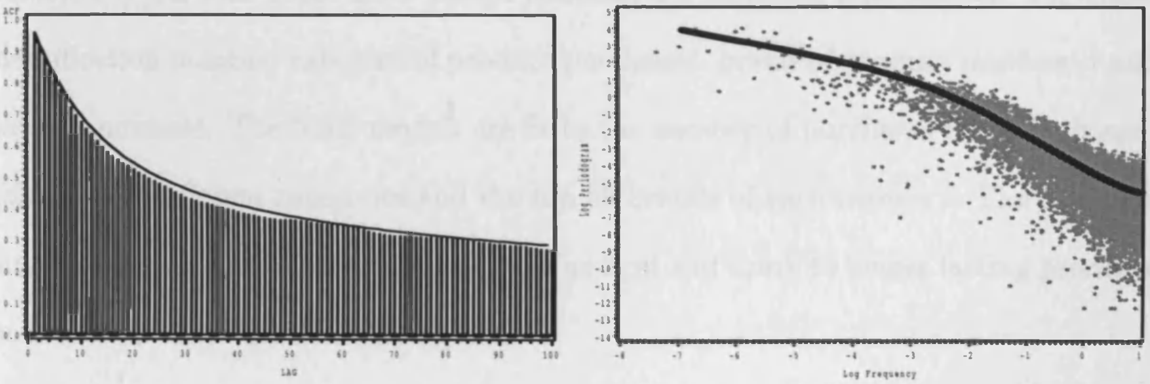


Figure 4.16: Autocorrelation function & periodogram

Chapter 5

Models for consumer buying behavior

This chapter considers the NBD and related processes discussed in Chapters 2, 3 and 4 as possible models for use in market research. The data analyzed has been courteously provided by ACNielsen BASES. The transaction data comprises a panel of 34,647 individual households representative of the United States. The database contains records of every transaction, through the scanning of individual items, of each household for the duration of the year 2000. Each record contains the following information: household identification number; category of product purchased; brand of product purchased and date of purchase. The NBD models are fit to the number of purchases made by households for 46 different categories and the top 50 brands of each category. The products range from goods purchased regularly such as food and drink to longer lasting products such as cosmetics and household goods.

The transaction data is an ideal source as the data can either be represented in the form of panel data, analyzing the number of purchases across many households, or as a single time-series of longitudinal data, analyzing the total number of purchases of a particular brand or category. The mixed Poisson processes are suitable models for panel data whereas the INAR(1) processes are suitable models for a single realization of longitudinal data with serial dependence.

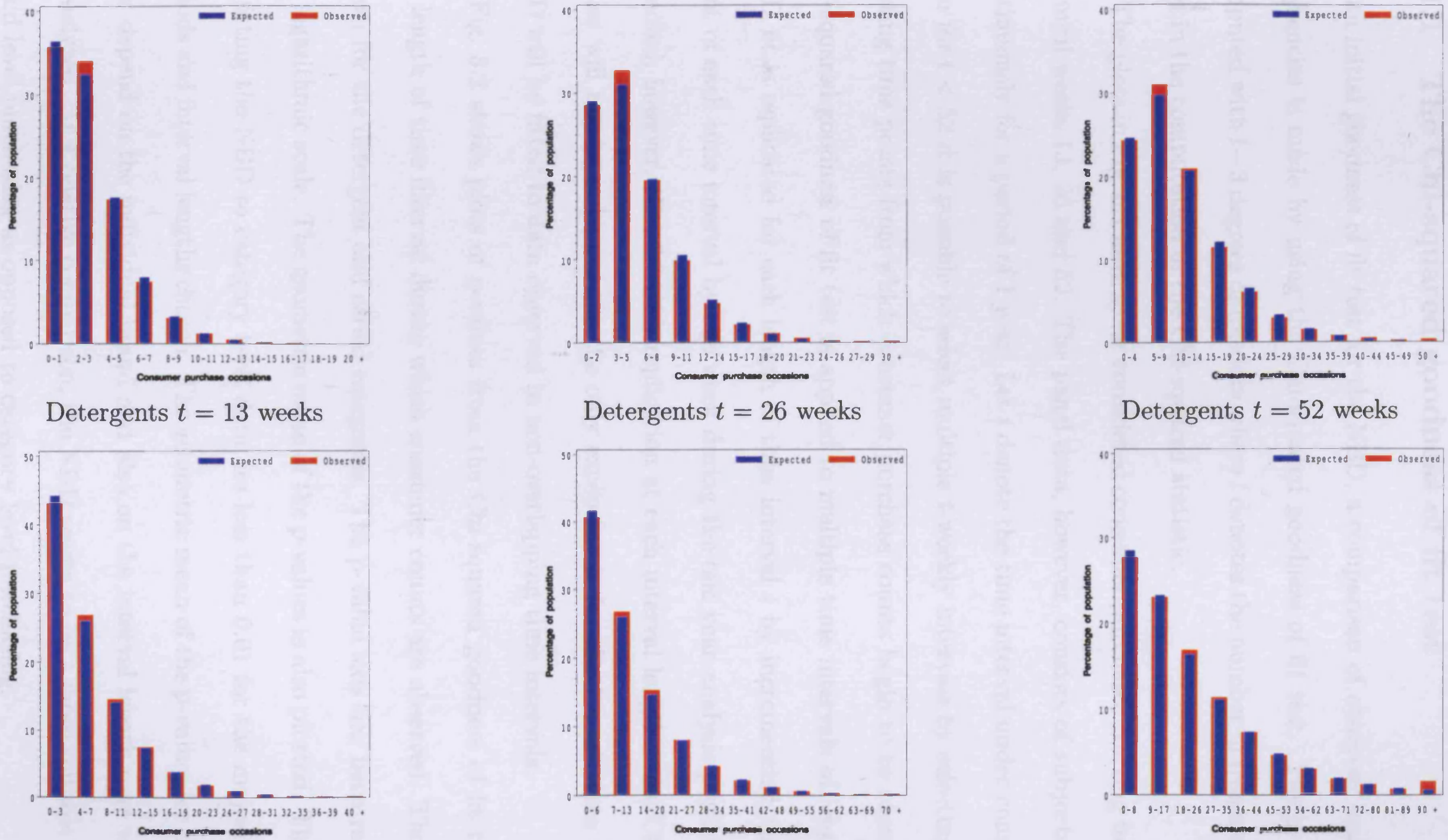
The event that will be modeled by the negative binomial processes will be consumer purchase occasions. A single consumer purchase occasion is said to occur if a household purchases a given product on any single day during the analysis period. The number of purchase occasions in a time interval therefore represents the number of days a household purchased the product within that time interval.

Section 5.1 considers the NBD as a suitable marginal distribution for consumer purchase occasions. The power method of estimation is used to estimate parameters of the NBD and the estimator is compared to the traditional method of moments and zero term method estimators. Sections 5.2 and 5.3 respectively consider the gamma Poisson process and the INAR(1) process as models for consumer purchase occasions by analyzing the covariances and correlations of functionals of the data computed in different time intervals.

5.1 The NBD model

The gamma Poisson process and the NBD INAR(1) process both have the NBD as their one-dimensional marginal distribution of the process. This is regardless of the fact that the gamma Poisson process and the NBD INAR(1) process are count and stationary processes respectively and therefore model different types of events.

Fig. 5.1 shows bar charts of observed frequencies and expected frequencies for two different types of category purchases (detergents and cereals) during time intervals of length 13, 26 and 52 weeks. The expected frequencies are computed under the assumption that the data follow the NBD. The NBD parameters are estimated by using the power method at optimum value of c using zero term method estimators as preliminary estimators. The NBD visually seems to be a good model for consumer purchase occasions for these two categories.



Detergents $t = 13$ weeks   Detergents $t = 26$ weeks   Detergents $t = 52$ weeks

Cereal $t = 13$ weeks   Cereal $t = 26$ weeks   Cereal $t = 52$ weeks

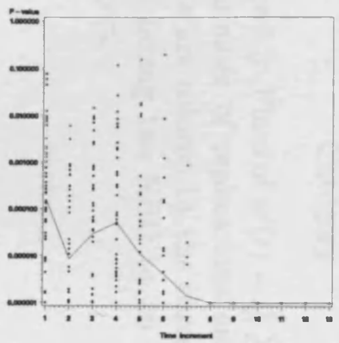
Figure 5.1: NBD fits for two different categories (detergents and cereals) for time intervals of lengths 13, 26 and 52 weeks.

5.1.1 The Chi-squared goodness of fit test

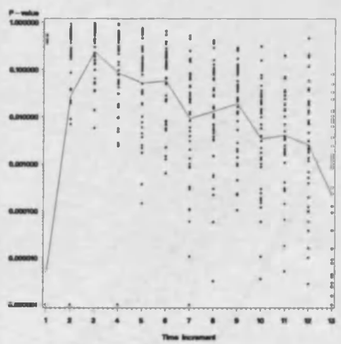
As an initial goodness of fit test for the NBD, a comparison of observed and expected frequencies is made by using the Chi-squared goodness of fit test. The p-values are computed with $l-3$ degrees of freedom where l denotes the number of frequency groups used in the computation of the Chi-squared statistic.

The plots in Fig. 5.1 have so far considered consumer purchases starting from week 1 up until weeks 13, 26 and 52. The panel data, however, consists of subjects observed continuously for a period of 1 year. Let t denote the time interval under consideration, then for $t < 52$ it is possible to select multiple t-weekly intervals by selecting different starting time points from which consumer purchase counts begin to be observed. The Chi-squared goodness of fit test is applied to multiple time intervals of length t. The NBD fit is replicated for each length of time interval t by incrementing the starting point of each time interval by one week during the one year analysis period. It must be noted, however, that for each replication at each interval length t the Chi-squared values will not be independent; the only exception is for the 1-week data where the NBD will be fitted to data observed in non-overlapping time intervals.

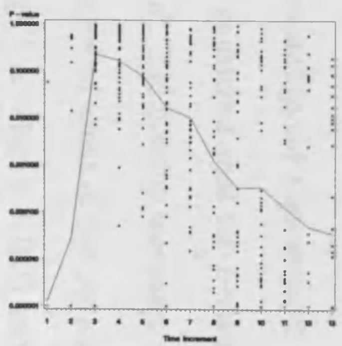
Fig. 5.2 shows plots of p-values from the Chi-squared goodness of fit test versus the length of time interval during which consumer counts are observed. The plots are shown for the detergent and cereal categories. The p-value axis has been re-scaled on the logarithmic scale. The geometric mean of the p-values is also plotted. The p-values for fitting the NBD to category level data are less than 0.01 for the majority of time periods and interval lengths chosen. The geometric mean of the p-values for brand level data depend on the individual brand and also on the interval length over which data is analyzed. As a relative comparison, the NBD seems to be a more suitable model for brand level purchasing as opposed to category level purchasing.



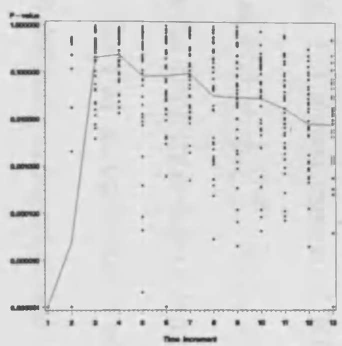
Detergent sales at category level



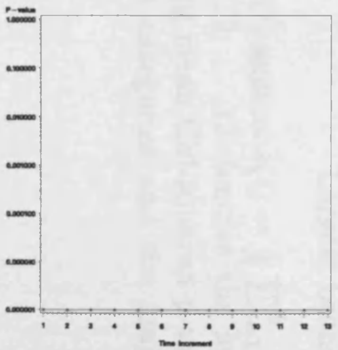
Largest brand in detergents



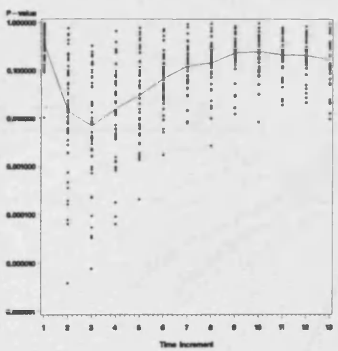
Second largest brand in detergents



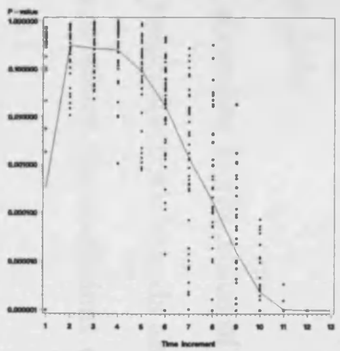
Third largest brand in detergents



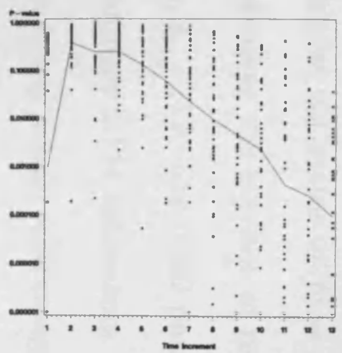
Cereal sales at category level



Largest brand in cereals



Second largest brand in cereals

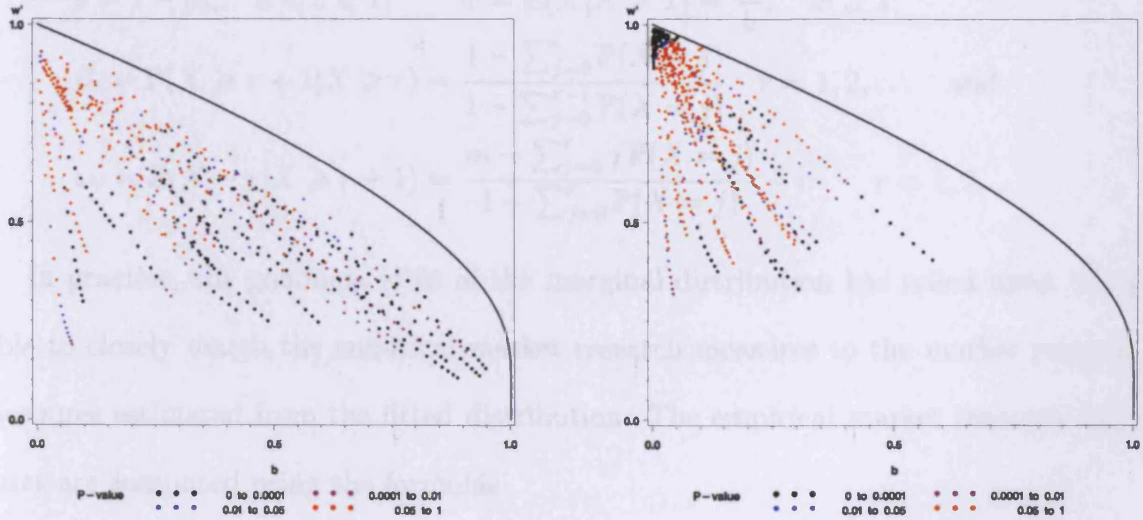


Third largest brand in cereals

Figure 5.2: Chi-squared goodness of fit test when fitting the NBD, using the PM at c-optimum with preliminary ZTM estimators, to category and brand level purchasing in two different categories (detergents and cereals) for vary length time intervals.

Fig. 5.3 shows means of NBD parameter estimates $\hat{w}'(t) = \frac{1}{R}\sum_{j=1}^{R} \hat{w}'_j(t)$ plotted against $\hat{b}(t) = \frac{1}{R}\sum_{j=1}^{R} \hat{b}_j(t)$, where R denotes the number of replications of the estimator for time intervals of length $t = 1, 2, \ldots, 13$. The NBD parameters are estimated using the PM at optimum c with preliminary ZTM estimators. Points are shown for the 46 different categories and the major brand within each category. The means of the NBD parameter estimates are taken over the replications obtained from the 1-weekly increments. The points are colored according to the corresponding mean p-value when implementing the Chi-squared goodness of fit.

It is evident that the p-value varies according to the value of the estimated parameter w'. The p-value increases as w' increases for both brands and categories, with the exception of areas of the parameter space where the coefficient of variation $(\sqrt{v_{ML}}/k)$ is large (i.e. when $b < 0.05$ and $w' > 0.95$).



Category Largest brand within category

Figure 5.3: Plots of $\hat{w}'(t) = \frac{1}{R}\sum_{j=1}^{R} \hat{w}'_j(t)$ against $\hat{b}(t) = \frac{1}{R}\sum_{j=1}^{R} \hat{b}_j(t)$, where R denotes the number of replications and $t = 1, 2, \ldots, 13$ denotes the length of time interval. Points are colored by the corresponding mean Chi-squared p-value. Points are shown when fitting the NBD to 46 different categories and the major brand within each category.

5.1.2 Single period repeat buying measures

Market measures for a general marginal distribution. In practice and in literature there is ambiguity in the definition of the market research measures. It is unclear as to whether the measures refer to observed values or values of the underlying sampling distribution. In this thesis the market research measures are considered to be those obtained from the underlying sampling distribution. Let X be a random variable from the one-dimensional distribution of purchases and let p_x denote the probabilities of purchasing $x = 0, 1, 2, \ldots$ products in the chosen time interval. The measures are then functions of the moments of the distribution of X. The penetration (b), purchase frequency (w), measured repeat (β_r) and repeats per repeater (ω_r) are defined by the equations

$$
\begin{aligned}
&b = 1 - p_0, \quad 0 \leqslant b \leqslant 1; \qquad w = \mathbb{E}(X|X \geqslant 1) = \frac{m}{b}, \quad w \geqslant 1; \\
&\beta_r = \mathbb{P}(X \geqslant r+1|X \geqslant r) = \frac{1 - \sum_{j=0}^{r} \mathbb{P}(X=j)}{1 - \sum_{j=0}^{r-1} \mathbb{P}(X=j)} \quad r = 1, 2, \ldots \quad \text{and} \\
&\omega_r = \mathbb{E}(X - r|X \geqslant r+1) = \frac{m - \sum_{j=0}^{r} j\,\mathbb{P}(X=j)}{1 - \sum_{j=0}^{r} \mathbb{P}(X=j)} - r \qquad r = 1, 2, \ldots.
\end{aligned}
$$

In practice, the goodness of fit of the marginal distribution has relied upon being able to closely match the empirical market research measures to the market research measures estimated from the fitted distribution. The empirical market research measures are computed using the formulae

$$
\tilde{b} = 1 - \frac{n_0}{N}, \quad \tilde{w} = \frac{\bar{x}}{\tilde{b}}, \quad \tilde{\beta}_r = \frac{1 - \sum_{j=0}^{r} \frac{n_j}{N}}{1 - \sum_{j=0}^{r-1} \frac{n_j}{N}} \quad \text{and} \quad \tilde{\omega}_r = \frac{\bar{x} - \sum_{j=0}^{r} j \frac{n_j}{N}}{1 - \sum_{j=0}^{r} \frac{n_j}{N}} - r. \tag{5.1.1}
$$

The empirical measured repeat and repeats per repeater are therefore computed by replacing the probability $\mathbb{P}(X = j)$ with its sample equivalent n_j/N where n_j, $(j = 1, 2, \ldots)$ are observed frequencies of j within the sample and N is the size of the sample.

The NBD as a marginal distribution for repeat buying measures. For purchases of category and the major brand within each category respectively, Fig. 5.4 and Fig. 5.5 show values of the NBD estimated market research measures ($\hat{b}$, $\hat{w}$, $\hat{\beta}_1$ and $\hat{\omega_1}$) against the empirical values of the market research measures ($\tilde{b}$, $\tilde{w}$, $\tilde{\beta}_1$ and $\tilde{\omega_1}$) defined by (5.1.1). Points are shown for each replication when fitting the NBD to 1-weekly data through to 13-weekly data in 1-week increments. The figures show data for all 46 categories. A line regressing the theoretically estimated measures on the empirical measures is shown along with a line corresponding to the 45° diagonal.

The estimated market research measures are computed using estimators for m and k obtained from the MOM/ZTM estimators and the PM estimator computed at optimum c with ZTM preliminary estimators. Note that, in the case of the MOM/ZTM estimator, if the ZTM estimator is deemed to be more efficient then $\hat{b} = \tilde{b}$ and therefore $\hat{w} = \tilde{w}$, so that the ratios for the penetration and mean purchase frequency will be equal to 1.

The NBD estimated points plotted against the empirical points for penetration ($\hat{b}$ vs. $\tilde{b}$) and mean purchase frequency ($\hat{w}$ vs. $\tilde{w}$) lie very close to the diagonal for consumer purchases of both category and the major brand within each category. There is, however, a tendency for the NBD estimates computed using the MOM/ZTM to slightly deviate from the diagonal as both penetration and purchase frequency increases; this is not the case for the NBD estimated penetration and purchase frequency obtained by using the PM at optimum c. The points for measured repeat and repeats per repeater ($r = 1$ and $r = 2$) also lie close to the diagonal for category purchases. For the purchases of the major brand within each category the fit of measured repeat and repeats per repeater become worse as r increases from $r = 1$ to $r = 2$ with outliers becoming increasingly present.

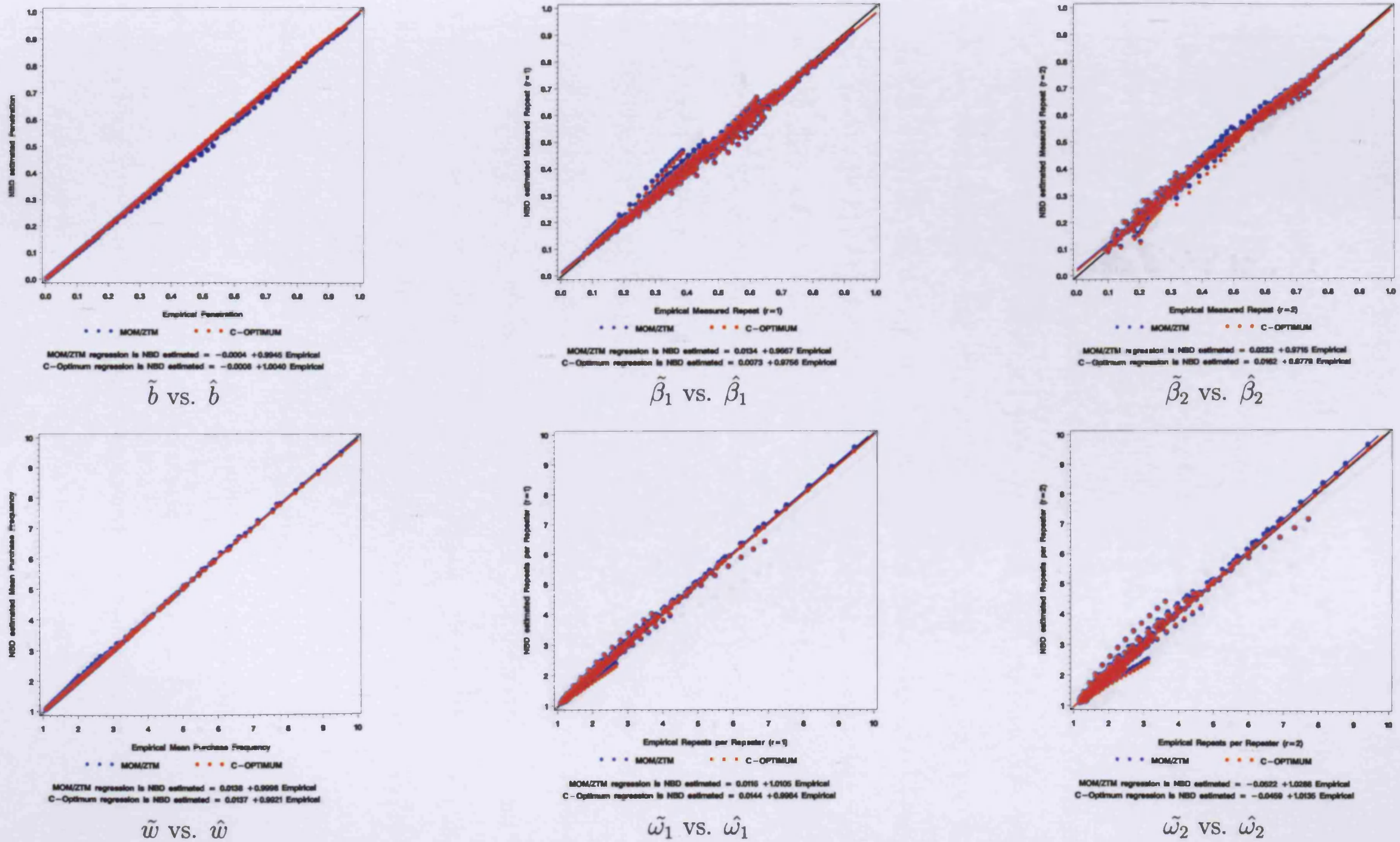


Figure 5.4: Plots of NBD estimated market research measures versus empirical market research measures for 46 different categories.

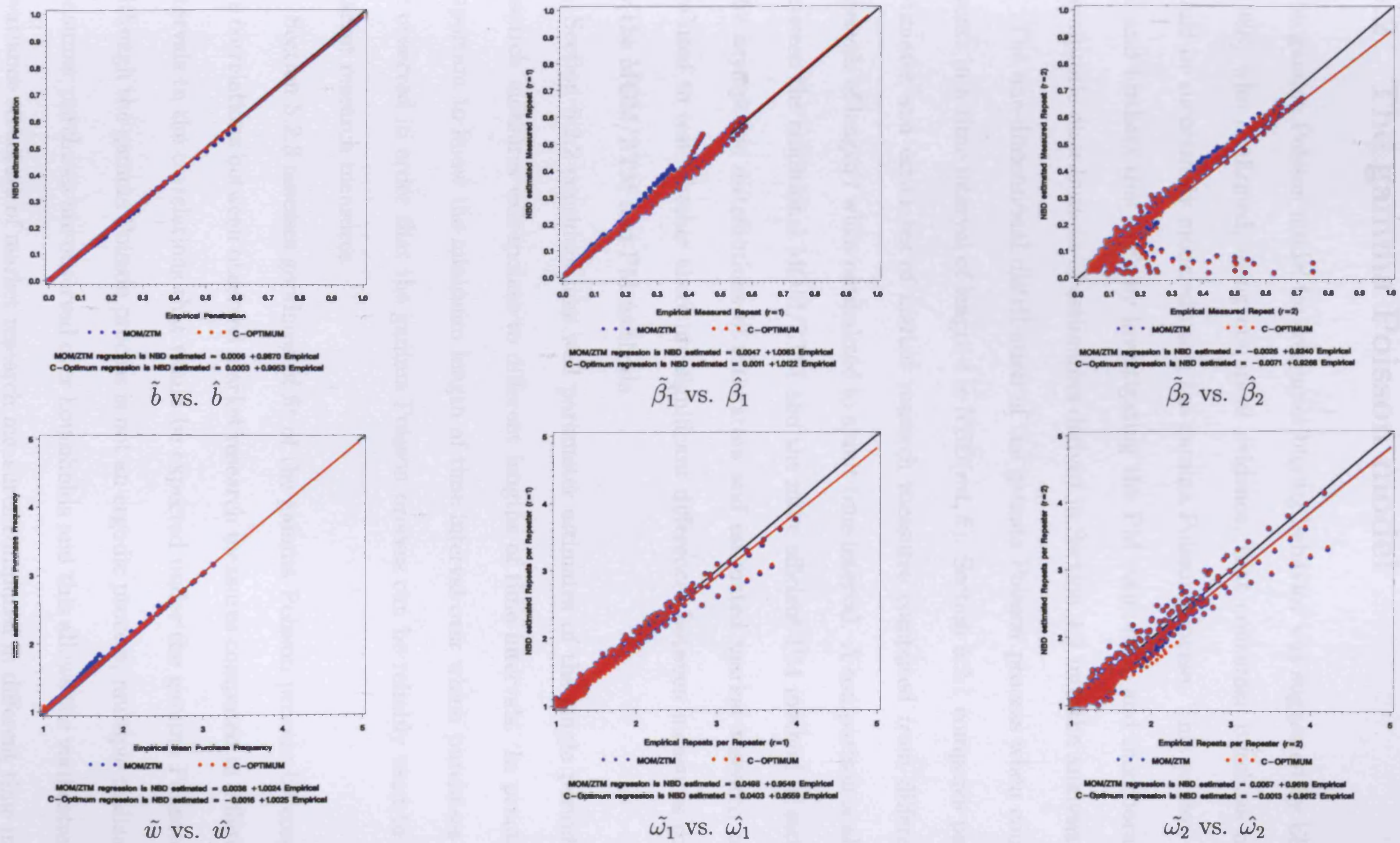


Figure 5.5: Plots of NBD estimated market research measures versus empirical market research measures for the major brand in each of 46 different categories.

5.2 The gamma Poisson model

The gamma Poisson model for consumer buying behavior was suggested by Ehrenberg (1988) who confirmed, using empirical evidence, that consumer purchase occasions could be successfully modeled using the gamma Poisson process. This section expands on and furthers this work by investigating the PM estimators and incorporating the asymptotic distributions of estimators derived in Section 4.2 into the analysis.

The one-dimensional distribution of the gamma Poisson process when considering events in a time interval of length t is NBD(mt, k). Section 5.2.1 compares parameter estimates and estimates of market research measures computed from different time intervals of length t when normalized to a unit time interval. A comparison is also made between the traditional MOM/ZTM and the more efficient PM method of estimation. The asymptotic distributions of estimators and estimated market research measures are used to test whether there are significant differences between measures computed by the MOM/ZTM and PM methods.

Section 5.2.2 considers how well parameter estimates of the single period market research measures extrapolate to different lengths of time intervals. In practice, it is important to know the minimum length of time interval over which purchases need to be observed in order that the gamma Poisson process can be reliably used to forecast market research measures.

Section 5.2.3 assesses goodness of fit of the gamma Poisson process by considering the correlations between observed market research measures computed in different time intervals to the correlations that would be expected under the gamma Poisson model. Although the gamma Poisson process is not an ergodic process, multiple realizations of consumer purchases are observed over households and this allows the verification of the covariance structure of market research measures computed in different time intervals.

5.2.1 Single period measures with varying time

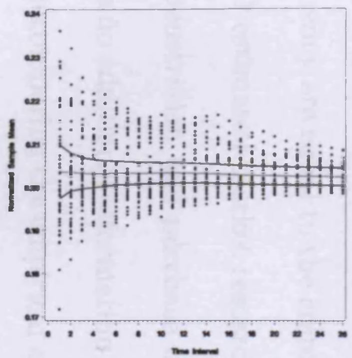
The one-dimensional distribution of the gamma Poisson process is

$$\mathbb{P}\left(Z(t)=x\right)=\frac{\Gamma(k+x)}{\Gamma(k)x!}\left(\frac{k}{k+mt}\right)^{k}\left(\frac{mt}{k+mt}\right)^{x}\qquad \begin{array}{l} x=0,1,2,\ldots \\ k>0,\ m>0. \end{array} \tag{5.2.1}$$

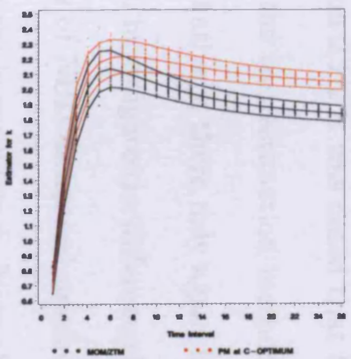
The process stipulates that the number of events within a time interval of length t is NBD(mt, k). The mean, mt, of the one-dimensional distribution increases linearly with time whereas the shape parameter k remains constant.

Fig. 5.6 shows MOM/ZTM and PM estimators (computed at optimum c using ZTM preliminary estimators) for m and k when fitting the NBD to consumer purchase occasions of cereals and detergents at different lengths of time intervals t. The estimator for m is the normalized sample mean $\hat{m}_t = \bar{x}/t$, where $\bar{x}$ is the sample mean of purchase occasions in a time interval of length t. Replications for each time interval t are obtained by incrementing the starting point of the time intervals by 1 week. A 95% lower and upper confidence bound computed using the results of Section 4.2 and the mean for each estimator is also shown by solid lines. For fixed t, the confidence bounds have been computed using the mean of the estimators for m and k over the replications.

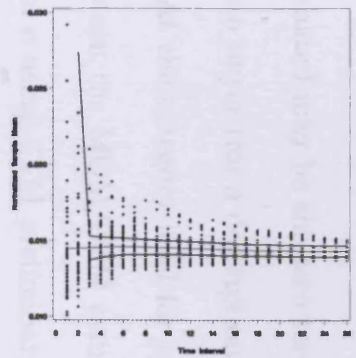
Fig. 5.6 shows the estimators for m and k converging to a constant as t increases. The variation of the estimators at each fixed t appears to decrease as t increases. It is important to note that this may be a cause of the dependence in observations and the reduction in the number of observations as the length of the time interval increases. For many of the values of t, a large number of points for $\hat{m}_t$ lie outside the 95% confidence interval for both detergents and cereals at top brand and category level. This indicates significant differences in the estimators for m over different time intervals. For $t > 4$, the estimators for k lie within the 95% confidence bounds indicating no significant differences in the shape parameter for varying time intervals.



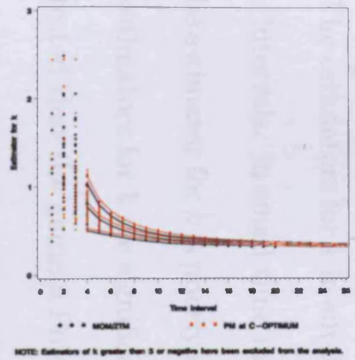
Detergents $\hat{m} = \bar{x}/t$ versus t



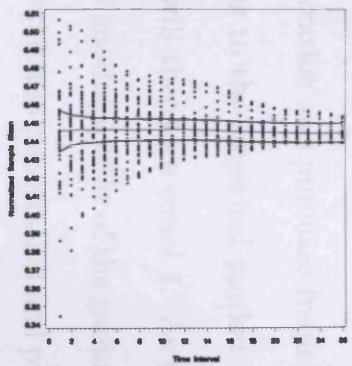
Detergents $\hat{k}$ versus t



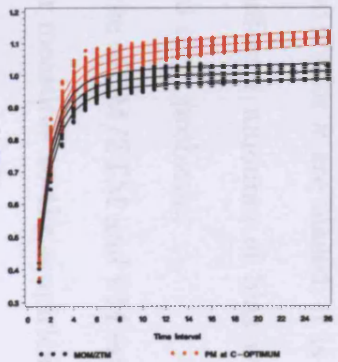
Detergents (top brand) $\hat{m} = \bar{x}/t$ versus t



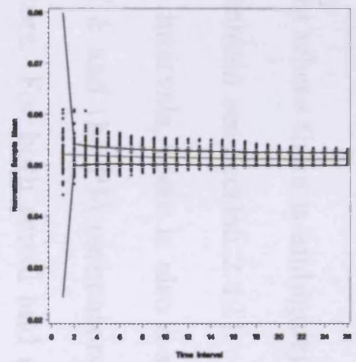
Detergents (top brand) $\hat{k}$ versus t



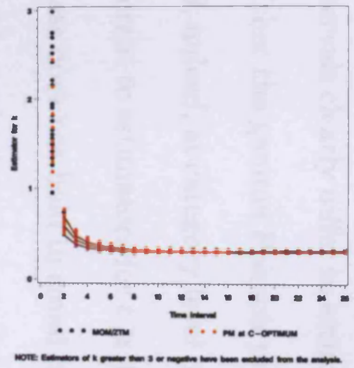
Cereal $\hat{m} = \bar{x}/t$ versus t



Cereal $\hat{k}$ versus t



Cereal (top brand) $\hat{m} = \bar{x}/t$ versus t



Cereal (top brand) $\hat{k}$ versus t

Figure 5.6: Estimators for the NBD parameters m and k using the MOM/ZTM method and the PM c-optimum method, with preliminary ZTM estimators, when estimated from different length time intervals t. The solid lines indicate the mean and corresponding 95% confidence bounds.

The estimators for k in small time intervals clearly differ to estimators for k in larger time intervals. In small time intervals either the gamma Poisson process does not hold or the estimator for k is poorly estimated; indeed, at category level purchasing, for $t < 4$ the estimators for k are significantly different to estimators for $t \geqslant 4$. Ehrenberg (1988) suggested that the gamma Poisson process does not hold in small time intervals as the Poisson process assumption of independent purchasing in consecutive time intervals by each household is unlikely to be true in practice. A possible cause of k being poorly estimated may be the zero term problem where there is ambiguity in the definition of a zero buyer (for a description of the problem see Section 2.4.1).

At the category level for large time intervals, there is also a significant difference between the MOM/ZTM estimators of k and the PM estimators computed at optimum c using ZTM preliminary estimators. For both cereal and detergent categories the MOM/ZTM estimator for k is persistently lower than the PM estimator for k. In Section 2.1.2 it was noted that the estimators for k are biased; it is therefore possible that the two estimation methods have different amounts of bias when estimating k. Alternatively, there may again be the zero term problem.

To investigate the difference between the MOM/ZTM and PM estimators for k, the ratios of NBD estimated market research measures to the empirical market research measures is considered. Note that for the MOM/ZTM method, if the ZTM method is deemed to be more efficient, then the NBD estimated penetration and purchase frequency are equal to the empirical penetration and purchase frequency. The ratios of NBD estimated market research measures to the empirical market research measures for penetration and purchase frequency will therefore equal 1. As a result, the ZTM gives no additional information in terms of goodness of fit of the gamma Poisson process when comparing the empirical and NBD estimated penetration and purchase frequency.

Fig. 5.7 shows ratios of NBD estimated market research measures to the empirical market research measures by estimation method for the detergent and cereal categories. A 95% lower and upper confidence bound computed using the results of Section 4.2 and the mean over the replications for each estimator are also shown by solid lines. For fixed t, the confidence bounds have been computed using the mean of the estimators for m and k over the replications.

Since the penetration and purchase frequency ratios equal 1, the ZTM estimator has been used for the cereal category and for $t \leqslant 3$ in the detergents category. When fitting the NBD to data in practice, it is unclear as to whether the zero counts should refer to potential buyers of the product or all non-buyers of the product. The empirical penetration used in these figures considers all buyers in the population that did not purchase a product during the time interval as zero buyers. This empirical penetration may, however, be incorrect. This problem is referred to as the zero term problem. The penetration and purchase frequency for the cereals category is therefore not considered.

In the case of detergents, the ratios for penetration and purchase frequency are closer to 1 using the PM estimators in comparison to using the MOM/ZTM estimator. The ratios are significantly closer to one for larger time intervals. The confidence intervals for penetration includes the value 1 and the confidence intervals for purchase frequency are closer to 1 than that of the MOM/ZTM method.

For longer time intervals, the PM estimator also persistently achieves a ratio closer to 1 for the ratio of NBD estimated measured repeat and repeats per repeater to the empirical measured repeat and repeats per repeater respectively. The fact that the ratios for measured repeat and repeats per repeater are closer to 1 in the cereal category, suggests that the empirical penetration in the cereal category may not be the empirical penetration required when fitting the NBD to data.

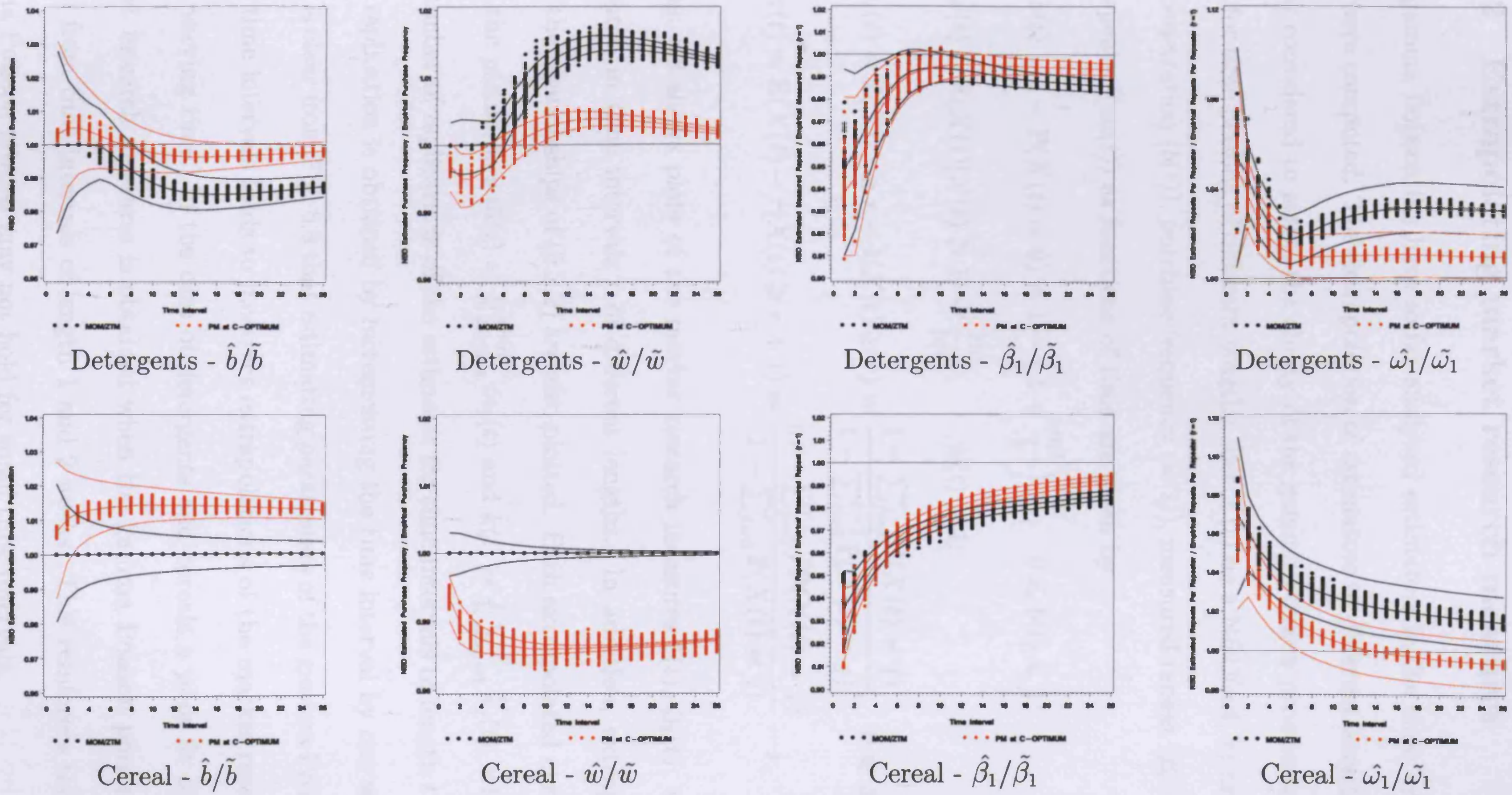


Figure 5.7: Ratios of NBD estimated market research measures to empirical market research measures by estimation method. The NBD estimated measures are computed using the MOM/ZTM method and the PM method at c-optimum with preliminary ZTM estimators. The solid lines indicate the mean and corresponding 95% confidence bounds.

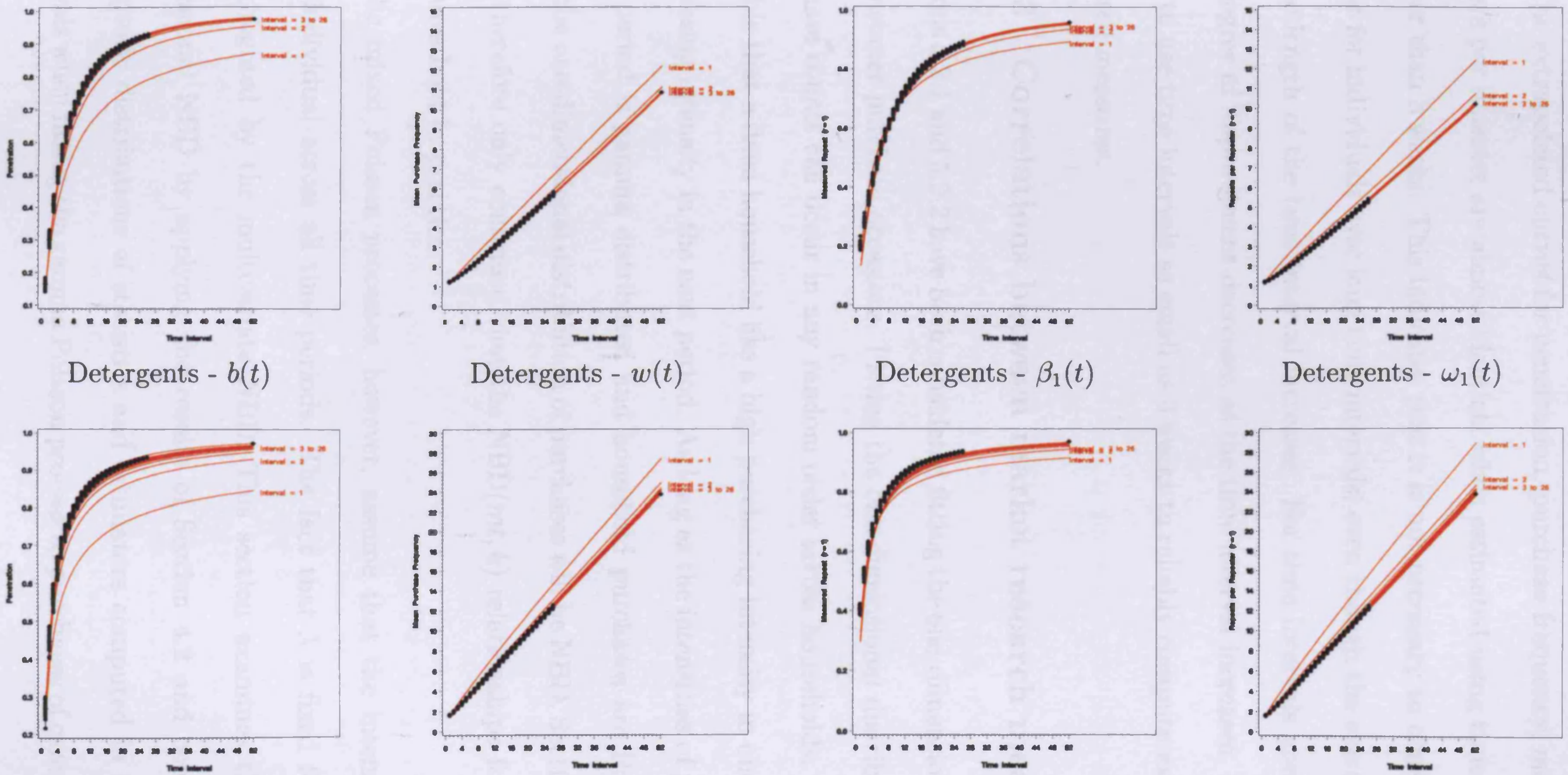
5.2.2 Extrapolating market research measures

The gamma Poisson fits have so far analyzed estimators in the time interval in which they were computed. The extrapolation of estimators to different length time intervals is now considered to assess the ability of the gamma Poisson process to forecast measures for time periods of different lengths. Let $X(t)$ be a NBD(mt, k) random variable. The penetration $(b(t))$, purchase frequency $(w(t))$, measured repeat $(\beta_r(t))$ and repeats per repeater $(\omega_r(t))$ as functions of time are given by

$$\begin{aligned}
b(t) &= 1 - \mathbb{P}(X(t) = 0) = 1 - \left(1 + \frac{mt}{k}\right)^{-k}, \quad 0 \leqslant b(t) \leqslant 1; \\
w(t) &= \mathbb{E}(X(t)|X(t) \geqslant 1) = \frac{mt}{b(t)}, \qquad w(t) \geqslant 1; \\
\beta_r(t) &= \mathbb{P}(X(t) \geqslant r+1|X(t) \geqslant r) = \frac{1 - \sum_{j=0}^{r} \mathbb{P}(X(t) = j)}{1 - \sum_{j=0}^{r-1} \mathbb{P}(X(t) = j)}, \quad 0 \leqslant \beta_r(t) \leqslant 1; \\
\omega_r(t) &= \mathbb{E}(X(t) - r|X(t) \geqslant r+1) = \frac{m - \sum_{j=0}^{r} j\,\mathbb{P}(X(t) = j)}{1 - \sum_{j=0}^{r} \mathbb{P}(X(t) = j)} - r, \quad \omega_r(t) \geqslant 1.
\end{aligned} \tag{5.2.2}$$

Fig. 5.8 shows plots of the market research measures $b(t)$, $\beta_1(t)$, $w(t)$ and $\omega_1(t)$ computed in time intervals t of different lengths. In addition, extrapolated curves using the relationships of (5.2.2) are also plotted. Each extrapolated curve is produced using the parameters $\hat{m}(t) = \frac{1}{R}\sum_{j=1}^{R} \hat{m}_j(t)$ and $\hat{k}(t) = \frac{1}{R}\sum_{j=1}^{R} \hat{k}_j(t)$, where R denotes the number of replications of the estimator for time intervals of length $t = 1, 2, \ldots, 26$. Each replication is obtained by incrementing the time interval by one week.

It is clear from Fig. 5.8 that estimating parameters of the gamma Poisson process in small time intervals leads to incorrect extrapolations of the market research measures when varying time. In the case of detergents and cereals a poor fit of the empirical market research measures is obtained when the gamma Poisson parameters are estimated from time intervals of length 1 and 2 weeks. This reinforces the fact that the gamma Poisson process may not hold for small time intervals.



Detergents - $b(t)$ | Detergents - $w(t)$ | Detergents - $\beta_1(t)$ | Detergents - $\omega_1(t)$

Cereal - $b(t)$ | Cereal - $w(t)$ | Cereal - $\beta_1(t)$ | Cereal - $\omega_1(t)$

Figure 5.8: Extrapolating estimators for market research measures using NBD parameter estimates obtained from different length time intervals. The NBD parameter estimates are computed using the power method at optimum c with ZTM preliminary estimators. Extrapolations are shown for category level purchasing only.

The extrapolated curves for penetration, purchase frequency, measured repeat and repeats per repeater are almost identical when estimated using time intervals of length greater than 3 weeks. This indicates that it is not necessary to observe purchasing behavior for individuals over long time intervals, even though the extrapolation improves as the length of the time interval increases. For time intervals greater than 3 weeks, the degree of improvement decreases as the time interval increases. It is therefore possible to use time intervals as small as 3 weeks to reliably compute extrapolated market research measures.

5.2.3 Correlations between market research measures

Sections 5.2.1 and 5.2.2 have both considered fitting the one-dimensional NBD to counts of consumer purchase occasions. Fitting the one-dimensional distribution implies that purchase counts can occur in any random order across households. For example, it is possible that a fixed household has a high purchasing intensity in one period and a low purchasing intensity in the next period. As long as the intensities of purchasing in each time period is gamma distributed and household purchases are Poisson distributed, then the one-dimensional distribution of purchases will be NBD. Sections 5.2.1 and 5.2.2 have therefore only confirmed that the NBD(mt, k) relationships for market research measures hold in practice.

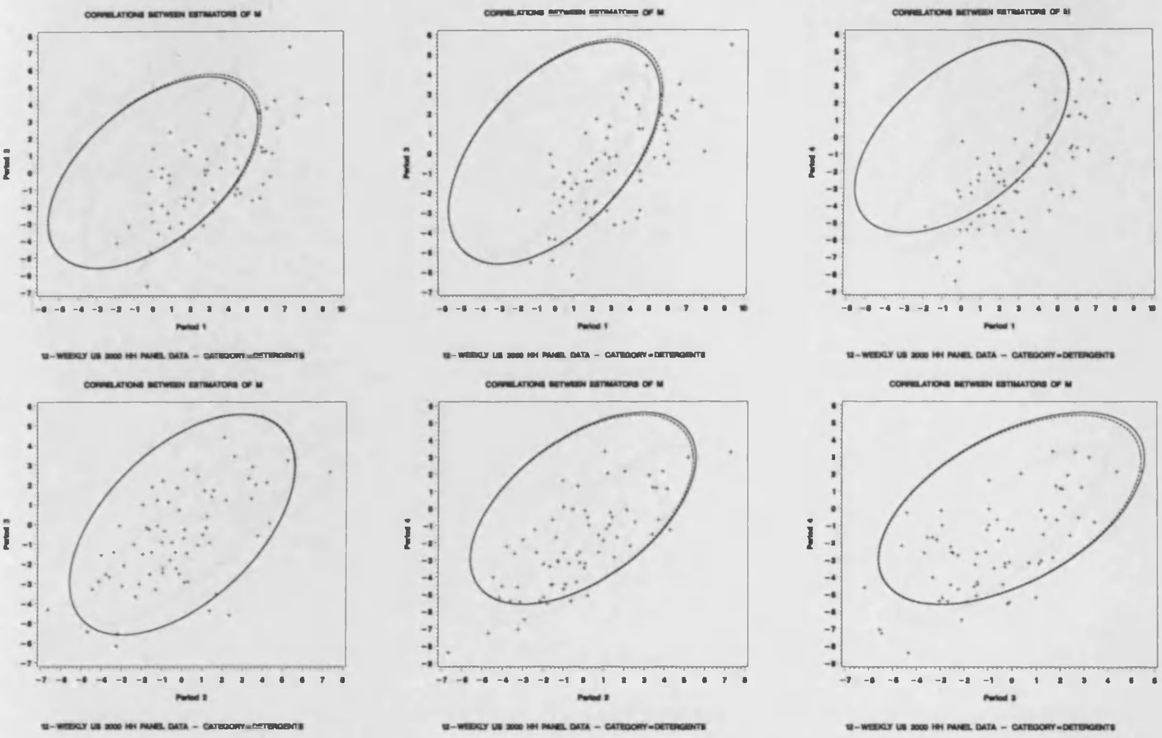
The mixed Poisson processes, however, assume that the intensity λ is fixed for each individual across all time periods. The fact that λ is fixed for each individual is highlighted by the multivariate NBD. This section examines the fit of the two-dimensional NBD by applying the results of Section 4.2 and considering the joint asymptotic distributions of statistics and estimators computed in two different time intervals when fitting the gamma Poisson process to purchases of cereals and detergents.

Computing covariances between estimators in two different time intervals requires replications of estimators. The replications cannot be obtained by incrementing the time intervals by 1 week, as in the previous section. The gamma Poisson process is not an ergodic process and therefore the correlations between estimators obtained by considering different time intervals in a single realization are not equivalent to the correlations between estimators in the ensemble of realizations.

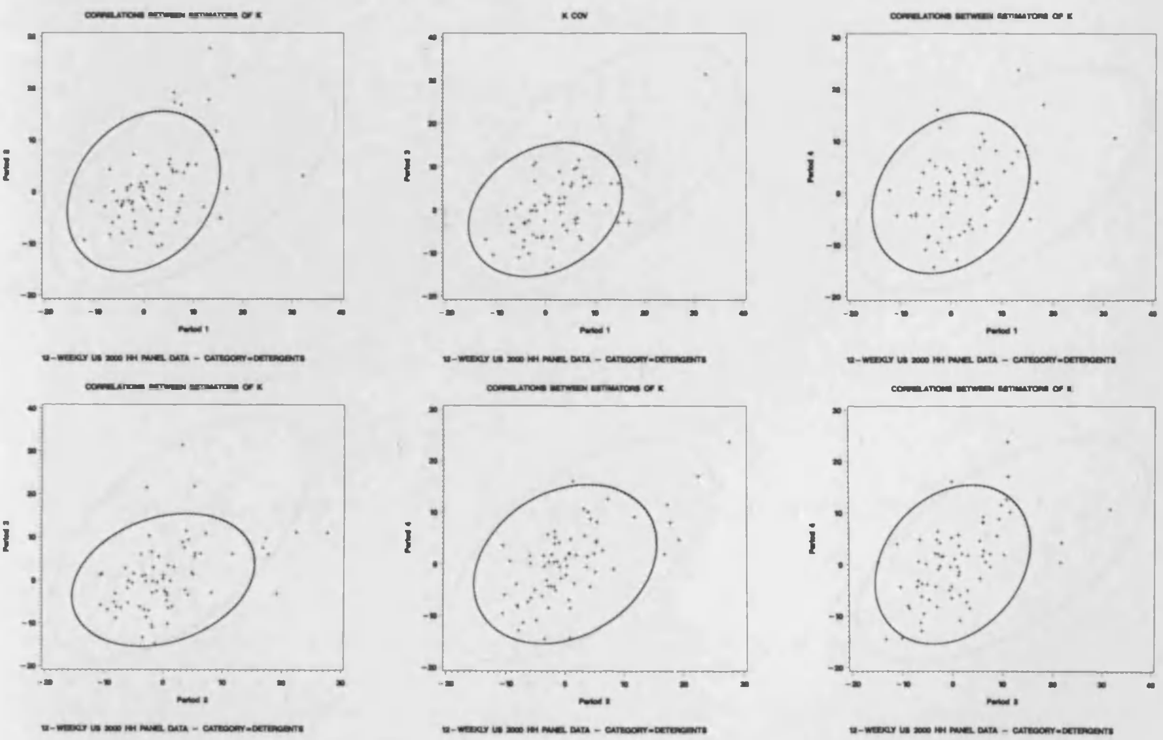
In the case of panel data, however, realizations of consumer purchase occasions are observed for each household. Replications of statistics or estimators can be obtained by taking sub-samples of the overall population and computing statistics or estimators for each sub-sample. In this thesis, the 34,467 households comprising the panel are randomly split into sub-samples of size 500 households.

Fig. 5.9 and Fig. 5.10 shows normalized estimators of the gamma Poisson parameters m and k computed in consecutive non-overlapping time intervals of length 12 weeks. In addition to the estimators, two 95% confidence ellipses constructed using the covariance matrix (4.2.2) for estimators of m and k in non-overlapping time intervals are also shown. The values m and k required to construct the ellipses are replaced by PM estimators $\hat{m}$ and $\hat{k}$ computed at optimum c using preliminary ZTM estimators. The solid confidence ellipse uses the estimators $\hat{m}$ and $\hat{k}$ obtained by fitting the NBD to the whole 52-week period, whereas the dotted confidence ellipse uses the mean of the estimators $\hat{m}$ and $\hat{k}$ obtained by fitting the NBD to each time period shown.

The estimators for k are captured well by the 95% theoretical confidence ellipses for both detergent and cereal categories. Note that in the detergent category a number of observations for estimators of k lie well outside the confidence ellipse and may be labeled as potential outliers of the model. The estimators for k in the cereal category are much more highly correlated than estimators for k in the detergents category.

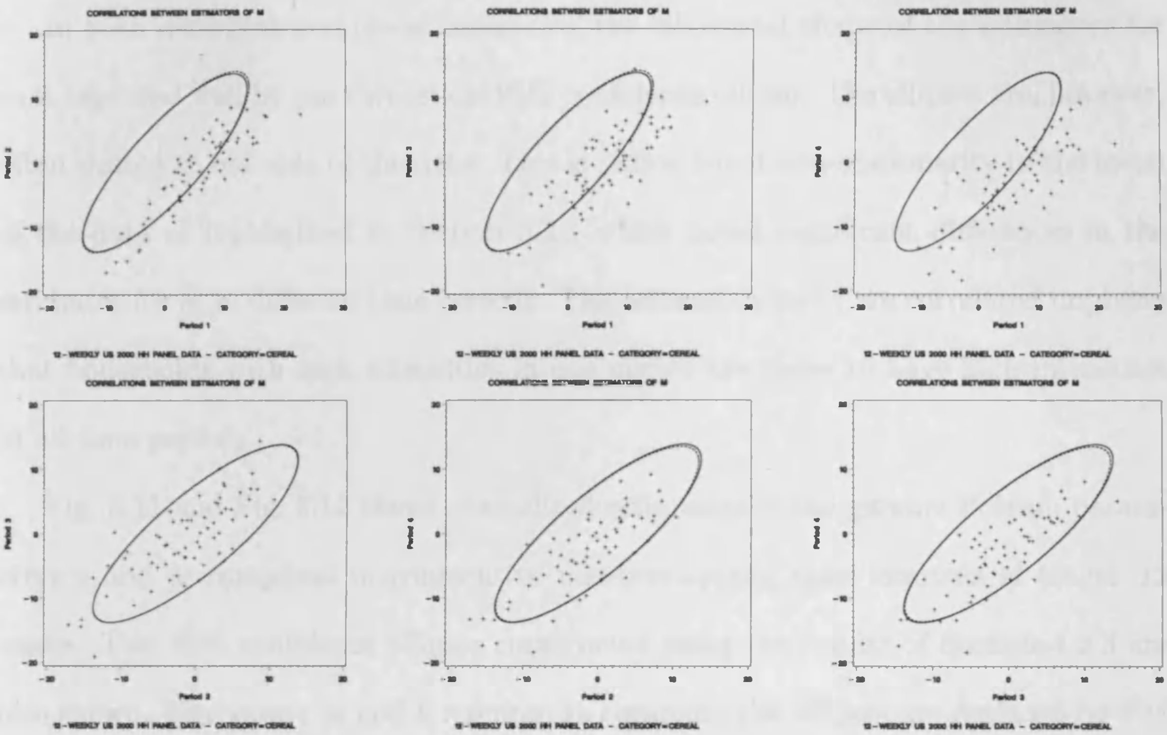


Correlations between estimators of m in different time intervals.

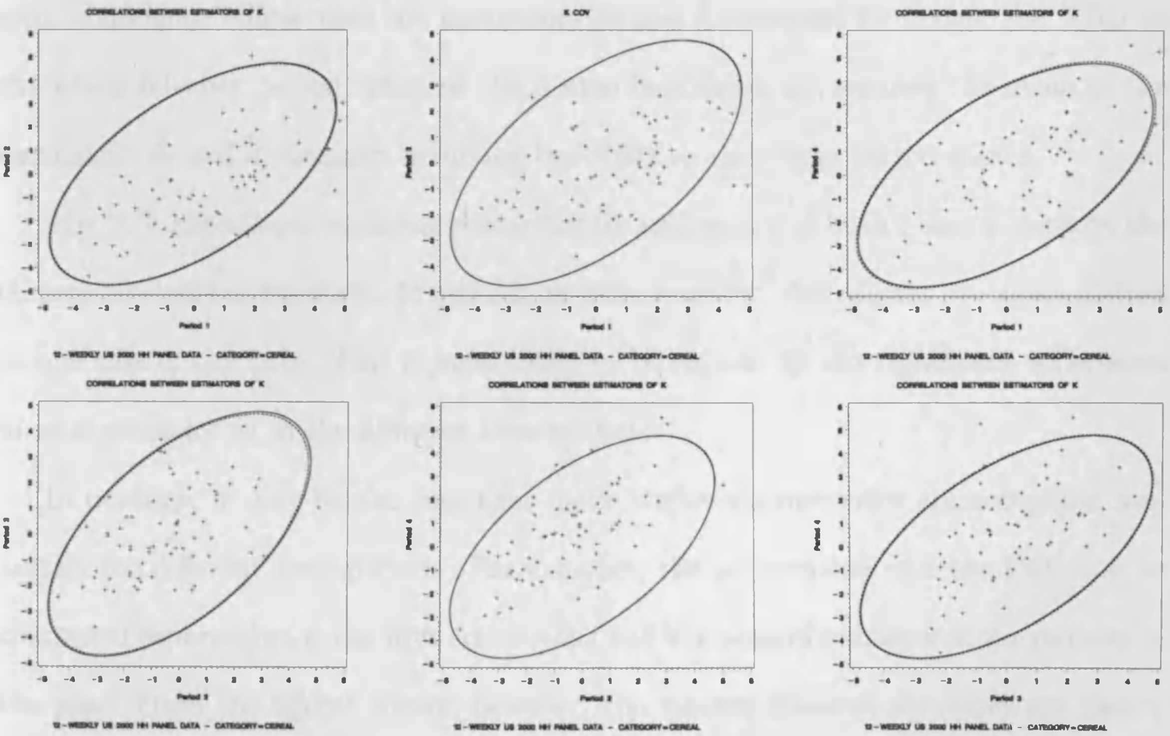


Correlations between estimators of k in different time intervals.

Figure 5.9: Correlations between estimators when fitting the gamma Poisson process to purchases of detergents at category level. Bivariate plots show estimators computed in different time periods together with corresponding 95% confidence ellipses computed under the assumption of asymptotic normality.



Correlations between estimators of m in different time intervals.



Correlations between estimators of k in different time intervals.

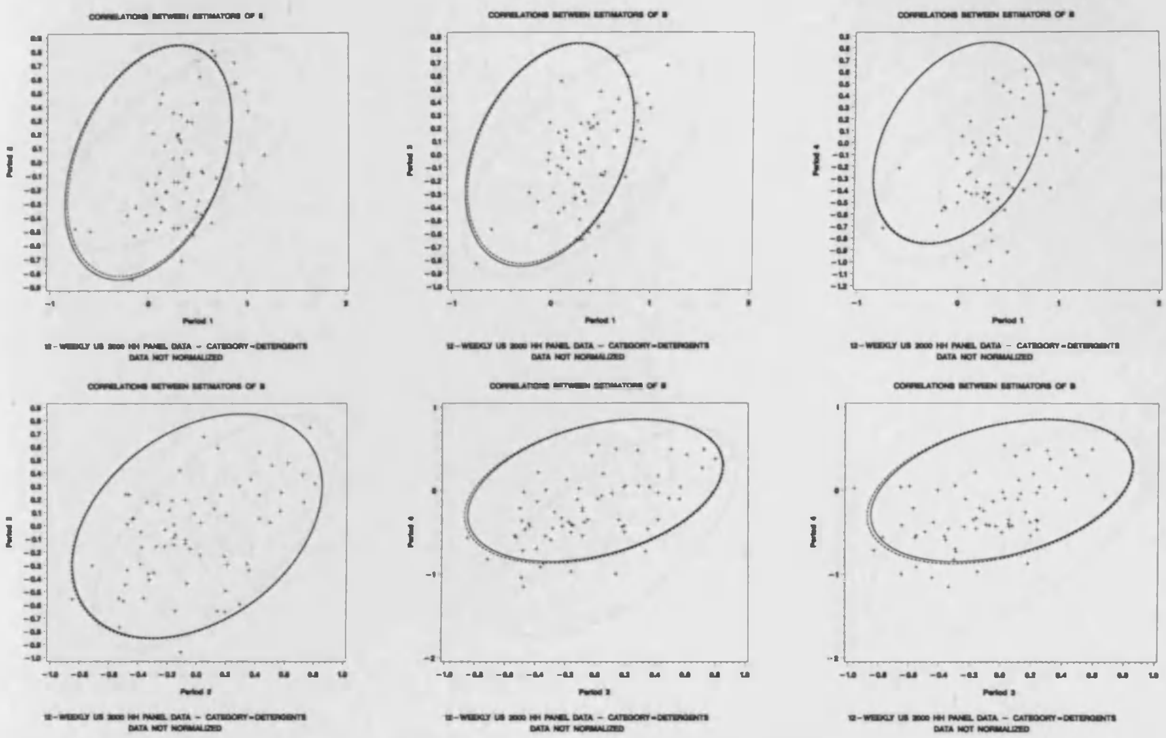
Figure 5.10: Correlations between estimators when fitting the gamma Poisson process to purchases of cereals at category level. Bivariate plots show estimators computed in different time periods together with corresponding 95% confidence ellipses computed under the assumption of asymptotic normality.

In both detergent and cereal categories, the ellipsoidal shape of the estimators for m is captured well by the theoretical 95% confidence ellipse. The ellipses are, however, often shifted to one side of the data. This is indicative of non-stationarity in the mean of the data as highlighted in Section 5.2.1 which noted significant differences in the estimator for m in different time periods. The estimators for m are correlated implying that households with high intensities in one period are likely to have high intensities in all time periods.

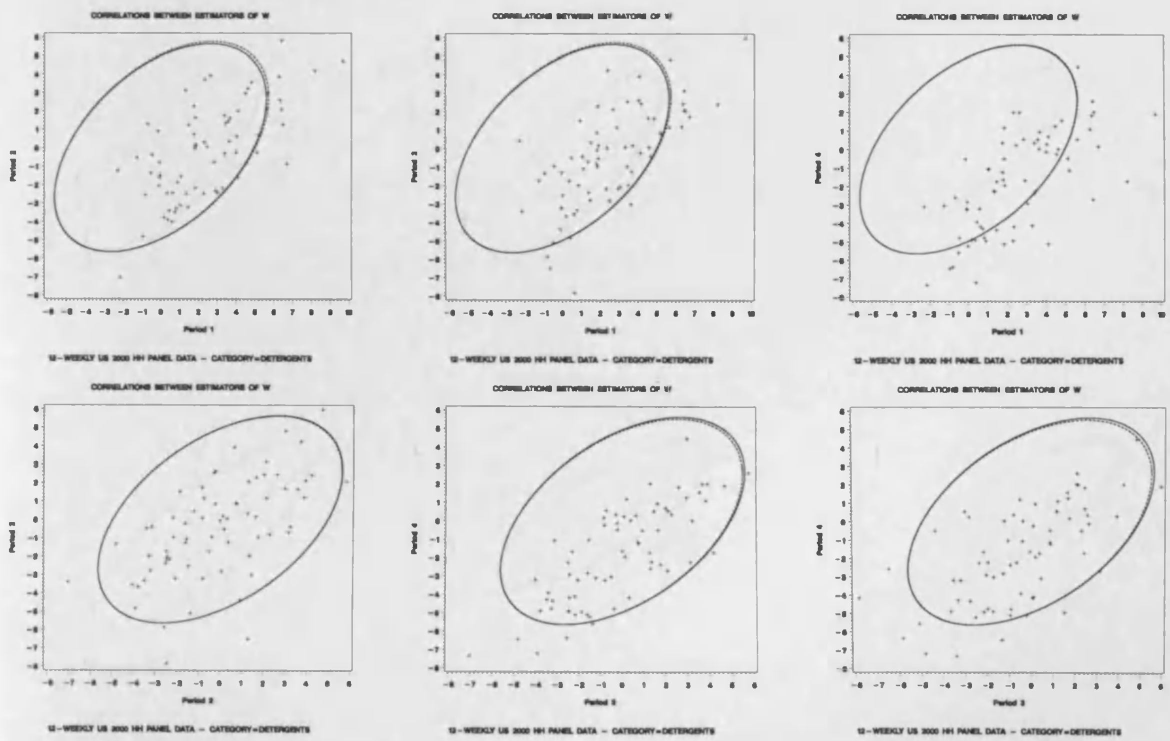
Fig. 5.11 and Fig. 5.12 shows normalized estimators of the gamma Poisson parameters b and w computed in consecutive non-overlapping time intervals of length 12 weeks. Two 95% confidence ellipses constructed using the results of Section 4.2.3 are also shown. The values m and k required to construct the ellipses are replaced by PM estimators $\hat{m}$ and $\hat{k}$ computed at optimum c using preliminary ZTM estimators. The solid confidence ellipse uses the estimators $\hat{m}$ and $\hat{k}$ obtained by fitting the NBD to the whole 52-week period, whereas the dotted confidence ellipse uses the mean of the estimators $\hat{m}$ and $\hat{k}$ obtained by fitting the NBD to each time period shown.

The 95% theoretical confidence intervals for estimators of both b and w capture the ellipsoidal shape of the data. In certain periods, however, the ellipses are again shifted to one side of the data. This is most likely to be caused by the significant differences in estimators for m in the different time periods.

In practice, it may be the case that market research measures are computed separately for different time periods. For example, the penetration of a product may be computed separately for the first six months and the second subsequent six months in the year. From the figures shown, however, the market research measures are clearly correlated. More accurate estimators may therefore be obtained by computing estimates from fitting the joint two-dimensional NBD.

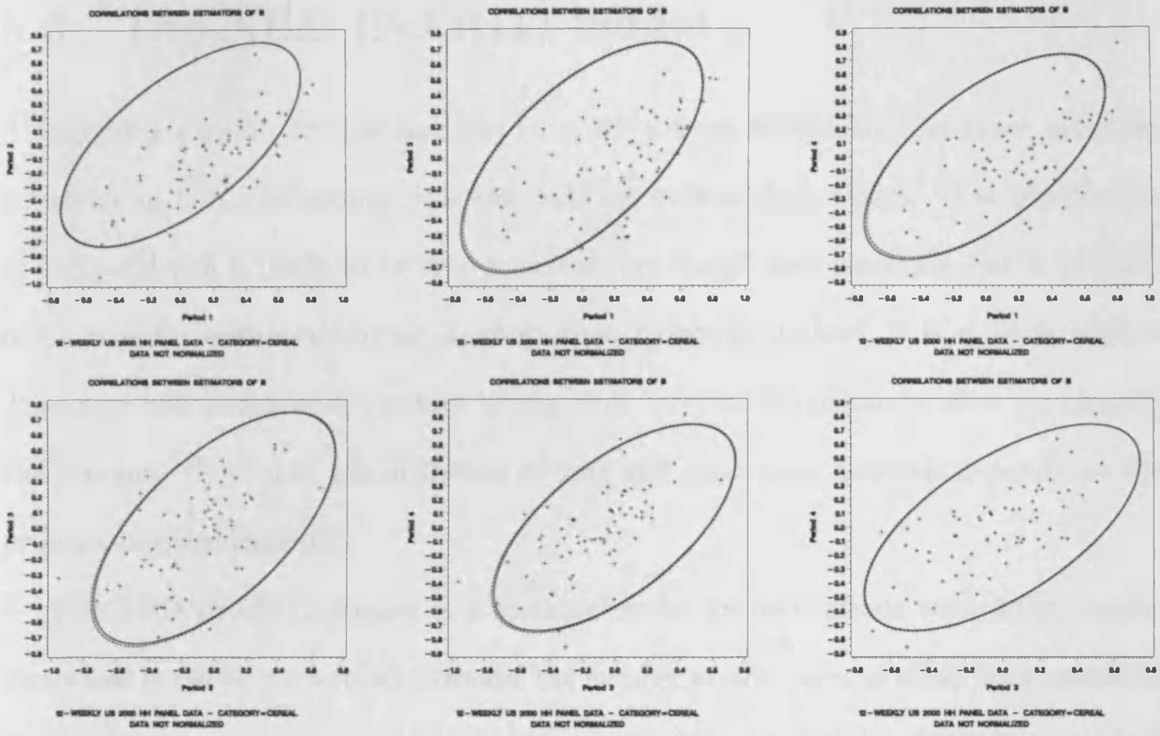


Correlations between estimators of b in different time intervals.

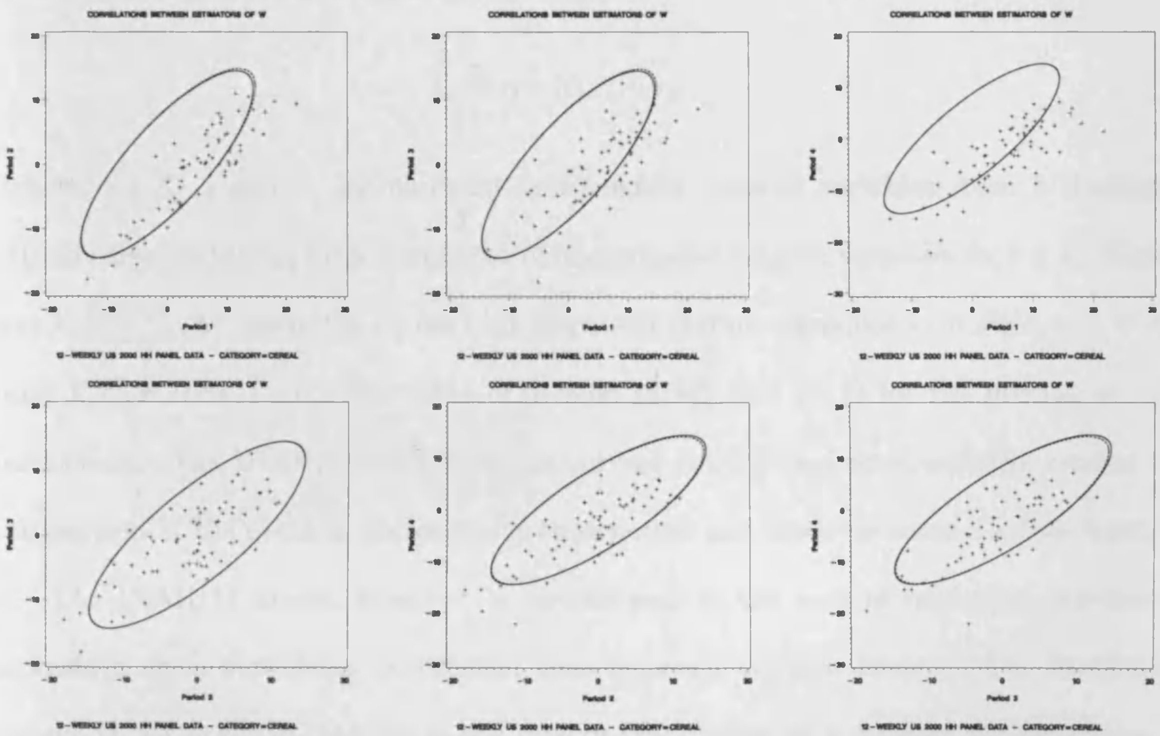


Correlations between estimators of w in different time intervals.

Figure 5.11: Correlations between estimators when fitting the gamma Poisson process to purchases of detergents at category level. Bivariate plots show estimators computed in different time periods together with corresponding 95% confidence ellipses computed under the assumption of asymptotic normality.



Correlations between estimators of b in different time intervals.



Correlations between estimators of w in different time intervals.

Figure 5.12: Correlations between estimators when fitting the gamma Poisson process to purchases of cereals at category level. Bivariate plots show estimators computed in different time periods together with corresponding 95% confidence ellipses computed under the assumption of asymptotic normality.

5.3 The NBD INAR(1) model

The gamma Poisson process assumes that, for a fixed household, consumer purchase occasions in non-overlapping time intervals are independent events. The assumption of independence is likely to be true for events in "long" time intervals, but is unlikely to be true for events occurring in short time intervals. Indeed, it is unlikely that a consumer will purchase a product in the time interval immediately after purchasing the product. Of course, the definition of long and short time intervals depends on the product in consideration.

The NBD INAR(1) process is a suitable model for realizations with serial dependence and could be introduced to model the number of purchases in short time intervals. Recall that the non-negative integer-valued process $\{X_t; t \in \mathbb{Z}\}$ is an INAR(1) process if the process satisfies the equation

$$X_t \stackrel{\mathcal{D}}{=} \alpha \circ X_{t-1} + \varepsilon_t,$$

where $\alpha \circ X_{t-1}$ and ε_t are mutually independent random variables from a discrete distribution and the ε_t form a sequence of uncorrelated random variables for $t \in \mathbb{Z}$. Here $\alpha \circ X \stackrel{\mathcal{D}}{=} \sum_{j=1}^{X} U_j$ where the U_j are i.i.d. Bernoulli random variables with $P(U_j = 1) = \alpha$ and $P(U_j = 0) = 1 - \alpha$. The value of α must satisfy $\alpha \in (0, 1)$ for the process to be stationary. The INAR(1) model for the current time period stochastically retains a proportion of the event in the previous time period and observes some random input.

The INAR(1) model, however, is not natural in the case of consumer purchase occasions since purchasing in different time intervals are new events. (The INAR(1) model is, for example, natural in the case of stock levels of a product within a store. The stock level in a time period can be represented as a retention of stock from the previous time period plus the addition of stock obtained during the current period.)

5.3.1 The INAR(1) model for the number of consumers

In the models considered so far, the event modeled has been the number of purchases made by consumers within a given time interval. In this section the analysis is concerned with the number of buyers who purchase a product. Consider the number of buyers that purchase a particular category. The number of buyers may be considered to be a retention of a proportion of the customers in the previous time period plus new customers. Such a situation could be modeled well by the INAR(1) process where X_t denotes the number of customers and α possibly denotes the level of loyalty.

Fig. 5.13 shows the autocorrelation function of the time series of the number of buyers of detergents and cereals observed in weekly increments. The shaded area indicates values of $0 \pm 1.96\sigma_l$ at each lag l where σ_l is the standard deviations of the estimated correlation. Correlation bars outside the shaded are therefore represent significant autocorrelation. Fig. 5.13 indicates that there is significant lag-1 autocorrelation of about 0.4 for both detergent and cereal categories. The remaining correlations for higher lags are insignificant. This indicates that an INAR(1) model could be appropriate for modeling the number of buyers in the detergents and cereal categories.

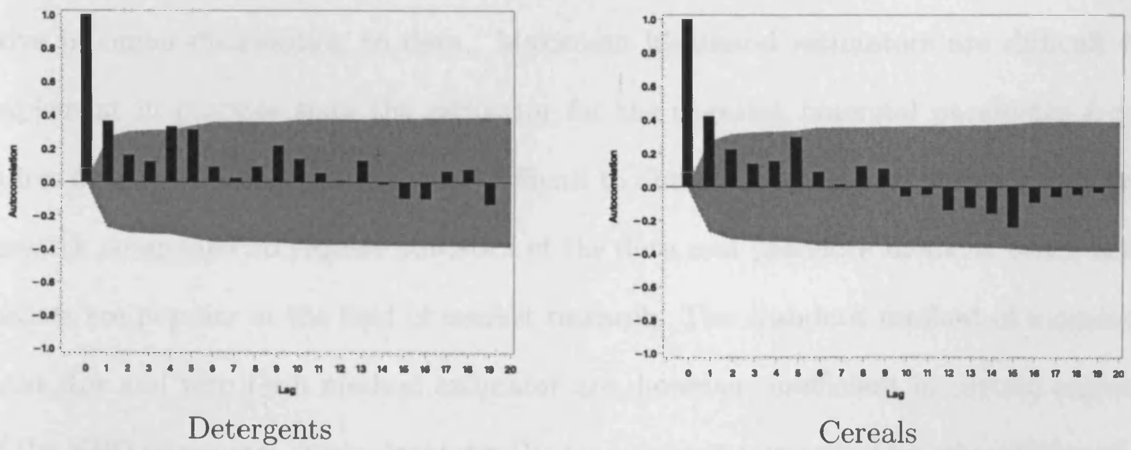


Figure 5.13: Autocorrelation functions for the number of buyers of detergents and cereal categories in different weeks.

Chapter 6

Conclusions and further work

This thesis has considered two themes in developing statistical inference for negative binomial processes. The first theme has been to consider more efficient moment based estimators for estimating parameters of the NBD than the standard method of moments and zero term method estimators. The second theme has been to assess adequacy of negative binomial processes by considering the dynamical behavior of the processes. The dynamical behavior of the processes has been assessed by verifying the correlation structure of estimators and statistics computed from data in two different time intervals to the correlation structure that would be expected given the process being fitted.

Parameters of negative binomial processes are often estimated by fitting the negative binomial distribution to data. Maximum likelihood estimators are difficult to implement in practice since the estimator for the negative binomial parameter k requires frequency counts and these are difficult to obtain. Instead, it is easier for market research companies to request statistics of the data and therefore moment based estimators are popular in the field of market research. The standard method of moments estimator and zero term method estimator are, however, inefficient in certain regions of the NBD parameter space. Importantly, many parameter estimates when fitting the NBD to purchases of a category reside in this inefficient area of the parameter space.

Fitting the negative binomial distribution only provides partial indication to the suitability of negative binomial processes for data. Ehrenberg (1988), to an empirical extent, considered assessing suitability of the gamma Poisson process for market research data, more precisely the modeling of consumer purchase occasions, by assessing relationships between market research measures computed in different time intervals and also the growth of market research measures as a function of time. These measures, however, were only assessed empirically and no method of checking the significance of the fits were presented.

6.1 Conclusion and discussion

In this thesis more efficient moment based estimation methods have been considered in the form of power method estimators. Statistically assessing the adequacy of negative binomial processes have been considered by deriving the limiting covariance matrix of estimators of the negative binomial distribution and also the limiting covariance matrix of estimators of parameters in negative binomial processes.

6.1.1 The power method estimators

The power method estimators depend on the parameter c. The power method estimator tends to the method of moments estimator as $c \to 1$ and is equivalent to the zero term method estimator when $c = 0$. Upon suitable choice of the parameter c the power method estimator can be almost as efficient as the maximum likelihood estimator when the sample is i.i.d. NBD. Moreover, upon suitable choice of c, the power method estimator is always more efficient than the method of moments estimator and zero term method estimator. The optimum choice of c that minimizes the variance of the estimator for the NBD parameter k however depends on the NBD parameter values.

In practice, since the NBD parameters are unknown, it appears as though the power method estimators may be difficult to implement. The optimum value for c, however, may be estimated using preliminary, possibly inefficient, estimators. The optimum value of c changes smoothly within the NBD (b, w')-parameter space. Estimating optimum c using preliminary NBD parameter estimates, for most NBD parameter values within the (b, w')-parameter space, will give estimates of c close enough to the value of optimum c to obtain an updated more efficient power method estimate for the parameter k.

In market research, simple estimators for the NBD parameters are required. The insensitivity of the efficiency of power method estimators to small changes in c further allows the construction of simple estimators, by approximating optimum c, that can be more easily implemented in practice. The approximations and estimations for optimum c are robust in areas of the parameter space where the coefficient of variation of the maximum likelihood estimator for k is low. The robustness of the estimators are shown in Fig. 3.18, by indicating the maximum possible loss of efficiency in estimating the NBD parameters, with respect to estimating using optimum c, with probability 0.95.

6.1.2 Fitting the NBD

The fit of the NBD at different time intervals is consistently rejected by the Chi-square test for purchases at category level. The fit of the NBD is not rejected, to the same extent as category level purchasing, for purchases of products at brand level. It is known (see e.g. Berkson (1938); Neyman (1949)) that for fixed significance level and fixed observed and expected frequencies, the power of the Chi-square test tends to one as the sample size increases. The Chi-square test is therefore not an ideal test for large sample sizes.

For both category level purchasing and brand level purchasing the NBD visually seems to be a good fit. Moreover, the observed market research measures when compared to the empirical market research measures agree extremely well, especially for category level purchasing. For brand level purchasing, more outliers are observed when comparing empirical and NBD estimated market research measures.

The δ-method has been used to construct asymptotic normal distributions of estimators of the NBD and also estimators of market research measures. The distributions have been computed as a by-product of considering the distribution of estimators of the gamma Poisson process computed in two different time intervals. The asymptotic distributions allows the construction of asymptotic confidence intervals for the estimators and therefore allows us to test whether the MOM/ZTM and the PM estimators are significantly different from each other and also if they are different from the empirical measures.

The empirical market research measures are estimated well by the NBD estimated market research measures when estimating the NBD parameters using the standard MOM/ZTM method and the PM at optimum c using ZTM preliminary estimators. The estimators for both MOM/ZTM and PM are very similar.

Using the asymptotic distribution of estimators for k and the asymptotic distribution of estimators for market research measures, the MOM/ZTM and PM methods can be shown to provide significantly different estimates when fitting the NBD to category level purchasing in large time intervals. The PM is shown to provide closer estimates for market research measures to the empirical measures than the MOM/ZTM method. The exception is in the case of penetration and purchase frequency when the ZTM of estimation is used; here the empirical and NBD estimated penetration and purchase frequencies are equal by definition of the estimator.

6.1.3 Fitting negative binomial process

This thesis has considered two negative binomial processes, namely the gamma Poisson process and the NBD INAR(1) process. The gamma Poisson process and the NBD INAR(1) process belong to different families of processes (that of renewal and autoregressive processes respectively). Assessing the adequacy of the two processes as a model for data therefore require different methods of inference.

Assessing the adequacy of the gamma Poisson process

The first method of assessing the adequacy of the gamma Poisson process extends the work of Ehrenberg (1988) by considering market research measures which, when estimated using data from a single time interval, have the ability to accurately extrapolate measures for time intervals of different lengths. The data analysis considered in this thesis showed that market research measures extrapolate well to all lengths of time periods when using parameter estimates obtained by fitting the NBD to time intervals of greater than three weeks for both cereal and detergent categories.

The second method verifies that estimators, computed using data in two different time intervals, fall within the corresponding asymptotic confidence ellipse that would be expected for estimators computed using data generated from a gamma Poisson process. The advantage of this method, over the method of assessing how well the NBD extrapolates to different lengths of time intervals, is that computing measures using the two-dimensional NBD requires individuals to retain the same Poisson (purchasing) intensity in both time intervals. Verifying the fit of the one-dimensional NBD to different length time intervals does not require the restriction that individuals must retain the same intensity in two time intervals; the only requirement is that the distribution is NBD where m increases linearly in time and k remains constant.

The estimators for the NBD parameters m, k, b and w when computed in two non-overlapping intervals all observe the ellipsoidal shape of the asymptotic confidence ellipse when fitting the gamma Poisson process to both cereal and detergent categories. However, for estimators of m and w, which are primarily location parameters, the ellipses are often shifted to one side of the data. A possible cause of this could be that there is a trend in the mean of the data.

Assessing the adequacy of the INAR(1) process

The assessment of the adequacy of the INAR(1) process considers the autocorrelation function of the process. In a similar fashion to the case of continuous valued first-order autoregressive processes, the INAR(1) process has an exponentially decaying autocorrelation function of the form $\rho(u) = \alpha^{|u|}$ at lag $u = \{0, \pm 1, \pm 2, \ldots\}$. Using the approach of Barndorff-Nielsen (1998), it is possible to construct long-memory integer valued processes by the aggregation, $X_t = \sum_{\eta=1}^{\infty} X_t^{(\eta)}$, of a sequence of stationary and independent INAR(1) processes $X_t^{(\eta)}$ $(\eta = 0, 1, 2, \ldots)$. The aggregated series has long-memory if $X_t^{(\eta)}$ has mean $\mu_{X_\eta} < \infty$ and variance $\sigma^2_{X_\eta} = c_1/(\eta^{1+2(1-H)}) < \infty$ with thinning parameter $\alpha_\eta = \exp\{-c_2 \,/\, \eta\}$ for some positive constants c_1 and c_2 with Hurst parameter $0.5 < H < 1$. As examples, a long-memory Poisson process and a long-memory NBD process were constructed.

The INAR(1) process was suggested as a possible model for the number of consumers of a product; the number of consumers in a subsequent time period can be thought of as a retention of customers from the previous time period plus the addition of new customers. The autocorrelation functions of the number of consumers in both cereal and detergent categories show that there is significant lag-1 autocorrelation. The estimate of $\hat{\alpha}$ for both categories is about 0.4 suggesting that about 40% of consumers that purchase in one time interval will also purchase in the next time interval.

6.2 Further work

This section considers further work and additional questions raised by this thesis. The topics are split into the three subsections of power method estimators, fitting the NBD and fitting negative binomial processes.

6.2.1 The power method estimators

The power method estimators have been shown to be almost as efficient as maximum likelihood estimators for i.i.d. NBD samples. For NBD INAR(1) samples it is difficult to analytically obtain the Fisher information matrix and therefore to obtain analytical expressions for the covariance matrix for maximum likelihood estimators of m and k. Simulation results, however, show that the maximum likelihood estimators are much more efficient than the power method estimators. Note that the power method estimators are still more efficient than the standard method of moments and zero term method estimators.

What is surprising, however, is that simulation studies maximizing the likelihood function for NBD INAR(1) samples show that the estimator for m is not equivalent to the sample mean of the data. The power method estimators assume that the estimator for m is efficiently estimated by the sample mean. Further study is required to check whether using a more efficient estimator for the sample mean to estimate m will provide more efficient power method estimators for the NBD parameters m and k.

6.2.2 Fitting the NBD

The zero term problem has not been fully investigated in this thesis. Since the number of zero buyers are latent, it is difficult to determine what the number of zero buyers should be when fitting the NBD. Further study is required to check goodness of fit of the NBD with varying number of zeros in the data. Note that the power method estimators may be iteratively used to estimate the number of zero buyers by updating the frequency of observed zeros with estimated zeros obtained by fitting the NBD using the power method. The process may be repeated until parameter estimates converge.

The NBD has been shown to provide a good fit for the data in terms of adequately estimating market research measures. Often, when considering frequency charts of consumer purchase occasions, long tails are observed indicating a small but significant presence of heavy buyers. Further study is required to check how much of an effect these heavy buyers have on fitting the NBD to consumer purchase occasions.

6.2.3 Fitting negative binomial process

This thesis has so far considered fitting the two-dimensional NBD to data in order to assess adequacy of the gamma Poisson process. Using the methodology used in this thesis, it should be possible to derive joint distributions of estimators computed in multiple (greater than two) time intervals and to use the joint distribution to verify whether the vector of estimators fall within the asymptotic confidence ellipsoid of estimators whose underlying process is gamma Poisson. This should provide a stronger indication of how well the assumption of constant intensity for each household holds in practice.

The asymptotic distributions of estimators computed in two different time intervals have shown that the estimators are in fact correlated. The strength of the correlation depends on the NBD parameter values. In practice, therefore, it is not sensible to compute estimators in separate time intervals as though the estimators in the two time intervals were independent. For example, it may be the case that the mean of consumer purchase occasions is estimated separately for the year 2005 and the year 2006. The moments of the two-dimensional NBD, in the case where the estimators are correlated, could be used to provide more accurate estimators for the year 2006.

Note that the joint asymptotic normal distributions of estimators computed from data generated by the gamma Poisson process allows testing of the hypothesis as to whether two estimators come from the same gamma Poisson process. This could aid in outlier detection. In this thesis the households were placed into sub-groups of size 500 households. For example, when computing estimators in two consecutive time intervals, some estimators for k in the detergent category fell far outside the confidence ellipse. This could have been a result of an outlier or outliers within the particular sub-group.

Appendices

Appendix A

A.1 Asymptotic distributions of statistics computed from INAR(1) samples

Theorem A.1.1. *Let* $\{x_t;\ t = 1, 2, \ldots, N\}$ *be a sample realization from an INAR(1) process* X_t *with stationary distribution* π. *Let* $\boldsymbol{f} = \left(x, c^X\right)^T$, $\bar{\boldsymbol{f}} = \left(\bar{x}, \widehat{c^X}\right)^T$ *with* $\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t$ *and* $\widehat{c^X} = \frac{1}{N}\sum_{t=1}^{N} c^{x_t}$, *with* $c > 0$ *and* $c \neq 1$. *Then* $\bar{\boldsymbol{f}}$ *has an asymptotic normal distribution given by* $\lim_{N\to\infty}\sqrt{N}(\bar{\boldsymbol{f}} - \mathbb{E}\boldsymbol{f}) \sim \mathcal{N}(0, \mathbb{D}\boldsymbol{f})$ *with covariance matrix*

$$\mathbb{D}\boldsymbol{f} = \mathbb{E}(\boldsymbol{f} - \mathbb{E}\boldsymbol{f})(\boldsymbol{f} - \mathbb{E}\boldsymbol{f})^T = \begin{pmatrix} V_{\bar{X}} & C_{\bar{X},\widehat{c^X}} \\ C_{\bar{X},\widehat{c^X}} & V_{\widehat{c^X}} \end{pmatrix}. \tag{A.1.1}$$

Here

$$V_{\bar{X}} = \lim_{N\to\infty} N \operatorname{Var}\left(\bar{X}\right) = \left(\frac{1+\alpha}{1-\alpha}\right)\operatorname{Var}[X_\pi], \tag{A.1.2}$$

$$\begin{aligned} V_{\widehat{c^X}} &= \lim_{N\to\infty} N \operatorname{Var}\left(\widehat{c^X}\right) \\ &= \operatorname{Var}\left(c^{X_\pi}\right) + 2\lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{G_{X_\pi}\left(c\left[1-\alpha^r+\alpha^r c\right]\right)G_\varepsilon(c;\alpha^r) - G_{X_\pi}^2(c)\right\}, \end{aligned} \tag{A.1.3}$$

$$\begin{aligned} C_{\bar{X},\widehat{c^X}} &= \lim_{N\to\infty} N \operatorname{Cov}\left(\bar{X}, \widehat{c^X}\right) = \operatorname{Cov}\left(X_\pi c^{X_\pi}\right) \\ &\quad + \lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{E\left[X_\pi\left(1-\alpha^r+\alpha^r c\right)^{X_\pi}\right]G_\varepsilon(c;\alpha^r) - E[X_\pi]G_{X_\pi}(c)\right\} \\ &\quad + \lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{G_{X_\pi}\left(c\left[1-\alpha^r+\alpha^r c\right]\right) - \alpha^r E[X_\pi]G_{X_\pi}(c)\right\}. \end{aligned} \tag{A.1.4}$$

The proof of Theorem A.1.1 uses the statistical properties of the thinning operator and the form of the INAR(1) process. Recall that the thinning operation $\alpha \circ X$ is defined as

$$\alpha \circ X \stackrel{d}{=} \sum_{j=1}^{X} Z_j \quad \alpha \in (0,1),$$

where the Z_j are i.i.d. Bernoulli random variables with $P(Z_j\!=\!1) = \alpha$ and $P(Z_j\!=\!0) = 1-\alpha$. From the definition of the thinning operator above it follows that

$$E[\alpha \circ X] = \alpha E[X] \quad \text{and} \quad E[f(X)(\alpha \circ X)] = \alpha E[X f(X)],$$

where all expectations are assumed to be finite.

Since the INAR(1) process is a stationary process we have for any $s \neq t$

$$E[f(X_s)] = E[f(X_t)] = E[f(X_\pi)].$$

From the definition of the INAR(1) process we note that the dependence between any two random variables X_t and X_s from the same INAR(1) process with $s > t$ can be written as

$$X_s = \alpha^{s-t} \circ X_t + \sum_{j=0}^{s-t-1} \alpha^j \circ \varepsilon_{s-j}\,.$$

Finally we note that since $X_t \stackrel{d}{=} \alpha \circ X_{t-1} + \varepsilon_t$ the expected value of the errors are

$$E[\varepsilon_t] = E[X_t] - E[\alpha \circ X_{t-1}] = (1-\alpha)E[X_\pi].$$

This result and many more relationships between the moments of the ε_π and the moments of X_π can be obtained using the relationship

$$G_{X_\pi}(c) = G_{X_\pi}(1-\alpha+\alpha c)G_\varepsilon(c;\ \alpha).$$

Proof of Theorem A.1.1.

Proof for Var($\bar{\mathbf{X}}$)

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{N}\sum_{t=1}^{N} X_t\right) = \frac{1}{N^2}\left\{\sum_{t=1}^{N}\mathrm{Var}(X_t) + \sum_{t\neq s}^{N}\mathrm{Cov}(X_t, X_s)\right\} \\
&= \frac{1}{N^2}\left\{N\mathrm{Var}[X_\pi] + 2\sum_{r=1}^{N-1}(N-r)\alpha^r\mathrm{Var}[X_\pi]\right\} \\
&= \frac{1}{N}\mathrm{Var}[X_\pi]\left\{1 + 2\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\alpha^r\right\}
\end{aligned}$$

$$\lim_{N\to\infty}\mathrm{Var}(\bar{X}) = \left(\frac{1+\alpha}{1-\alpha}\right)\mathrm{Var}(X_\pi). \qquad \square$$

Proof for Var($\widehat{\mathbf{c}^{\mathbf{X}}}$)

Note that

$$\begin{aligned}
E\left[c^{X_t}c^{X_s}\right] &= E\left[c^{X_t+X_s}\right] = E\left[c^{X_t+\alpha^{s-t}\circ X_t}\right]E\left[c^{\sum_{j=0}^{s-t-1}\alpha^j\circ\varepsilon_{s-j}}\right] \\
&= G_{X_\pi}\left(c(1-\alpha^{s-t}+\alpha^{s-t}c)\right)\prod_{j=0}^{s-t-1}G_\varepsilon(1-\alpha^j+\alpha^j c;\ \alpha) \\
&= G_{X_\pi}\left(c(1-\alpha^{s-t}+\alpha^{s-t}c)\right)\prod_{j=0}^{s-t-1}\frac{G_{X_\pi}(1-\alpha^j+\alpha^j c)}{G_{X_\pi}(1-\alpha+\alpha(1-\alpha^j+\alpha^j c))} \\
&= G_{X_\pi}\left(c(1-\alpha^{s-t}+\alpha^{s-t}c)\right)G_\varepsilon(c;\ \alpha^{s-t})
\end{aligned}$$

therefore

$$\begin{aligned}
\mathrm{Var}(\widehat{c^X}) &= \mathrm{Var}\left(\frac{1}{N}\sum_{t=1}^{N}c^{X_t}\right) = \frac{1}{N^2}\left\{\sum_{t=1}^{N}\mathrm{Var}(c^{X_t}) + \sum_{t\neq s}^{N}\mathrm{Cov}(c^{X_t}, c^{X_s})\right\} \\
&= \frac{1}{N^2}\left\{N\mathrm{Var}\left[c^{X_\pi}\right] + \sum_{t\neq s}^{N}\left(E\left[c^{X_t}c^{X_s}\right] - E\left[c^{X_t}\right]E\left[c^{X_s}\right]\right)\right\} \\
&= \frac{1}{N^2}\left\{N\mathrm{Var}\left[c^{X_\pi}\right] + 2\sum_{r=1}^{N-1}(N-r)\left(G_{X_\pi}\left(c(1-\alpha^r+\alpha^r c)\right)G_\varepsilon(c;\ \alpha^r) - G_{X_\pi}^2(c)\right)\right\}
\end{aligned}$$

$$\begin{aligned}
\lim_{N\to\infty} N\mathrm{Var}(\widehat{c^X}) &= \mathrm{Var}\left(c^{X_\pi}\right) \\
&\quad + 2\lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{G_{X_\pi}\left(c\left[1-\alpha^r+\alpha^r c\right]\right)G_\varepsilon(c;\alpha^r) - G_{X_\pi}^2(c)\right\}. \qquad \square
\end{aligned}$$

Proof for $\mathbf{Cov}(\bar{\mathbf{X}}, \widehat{\mathbf{c}^{\mathbf{X}}})$

Note that for $t < s$

$$\begin{aligned}
E\left[X_s c^{X_t}\right] &= E\left[\left(\alpha^{s-t} \circ X_t + \sum_{j=0}^{s-t-1} \alpha^j \circ \varepsilon_{s-j}\right) c^{X_t}\right] \\
&= E\left[c^{X_t}\left(\alpha^{s-t} \circ X_t\right)\right] + E\left[c^{X_t}\right] \sum_{j=0}^{s-t-1} E\left[\alpha^j \circ \varepsilon_{s-j}\right] \\
&= \alpha^{s-t} E\left[c^{X_t} X_t\right] + (1-\alpha^{s-t}) E\left[c^{X_t}\right] E\left[X_t\right] \\
&= \alpha^{s-t} E\left[c^{X_\pi} X_\pi\right] + (1-\alpha^{s-t}) E\left[c^{X_\pi}\right] E\left[X_\pi\right], \\
\mathrm{Cov}\left(X_s, c^{X_t}\right) &= \alpha^{s-t} \mathrm{Cov}\left(X_\pi, c^{X_\pi}\right)
\end{aligned}$$

and

$$\begin{aligned}
E\left[X_t c^{X_s}\right] &= E\left[X_t c^{\alpha^{s-t} \circ X_t}\right] E\left[c^{\sum_{j=0}^{s-t-1} \alpha^j \circ \varepsilon_{s-j}}\right] \\
&= E\left[X_t\left(1-\alpha^{s-t}+\alpha^{s-t} c\right)^{X_t}\right] \prod_{j=0}^{s-t-1} G_{\varepsilon_{s-j}}(1-\alpha^j+\alpha^j c) \\
&= E\left[X_\pi\left(1-\alpha^{s-t}+\alpha^{s-t} c\right)^{X_\pi}\right] G_\varepsilon(c;\ \alpha^{s-t}) \\
\mathrm{Cov}\left(X_t, c^{X_s}\right) &= E\left[X_\pi\left(1-\alpha^{s-t}+\alpha^{s-t} c\right)^{X_\pi}\right] G_\varepsilon(c;\ \alpha^{s-t}) - E\left[X_\pi\right] E\left[c^{X_\pi}\right],
\end{aligned}$$

therefore

$$\begin{aligned}
\mathrm{Cov}\left(\bar{X}, \widehat{c^X}\right) &= \frac{1}{N^2}\left\{\sum_{t=1}^{N} \mathrm{Cov}\left(X_t, c^{X_t}\right) + \sum_{t<s} \mathrm{Cov}(X_t, c^{X_s}) + \sum_{t>s} \mathrm{Cov}(X_t, c^{X_s})\right\} \\
&= \frac{1}{N^2}\left\{N \mathrm{Cov}\left(X_\pi, c^{X_\pi}\right) + \sum_{r=1}^{N-1} (N-r) \alpha^r \mathrm{Cov}\left(X_\pi, c^{X_\pi}\right)\right. \\
&\quad \left. + \sum_{r=1}^{N-1} (N-r)\left(E\left[X_\pi\left(1-\alpha^{s-t}+\alpha^{s-t} c\right)^{X_\pi}\right] G_\varepsilon(c;\ \alpha^{s-t}) - E\left[X_\pi\right] E\left[c^{X_\pi}\right]\right)\right\}
\end{aligned}$$

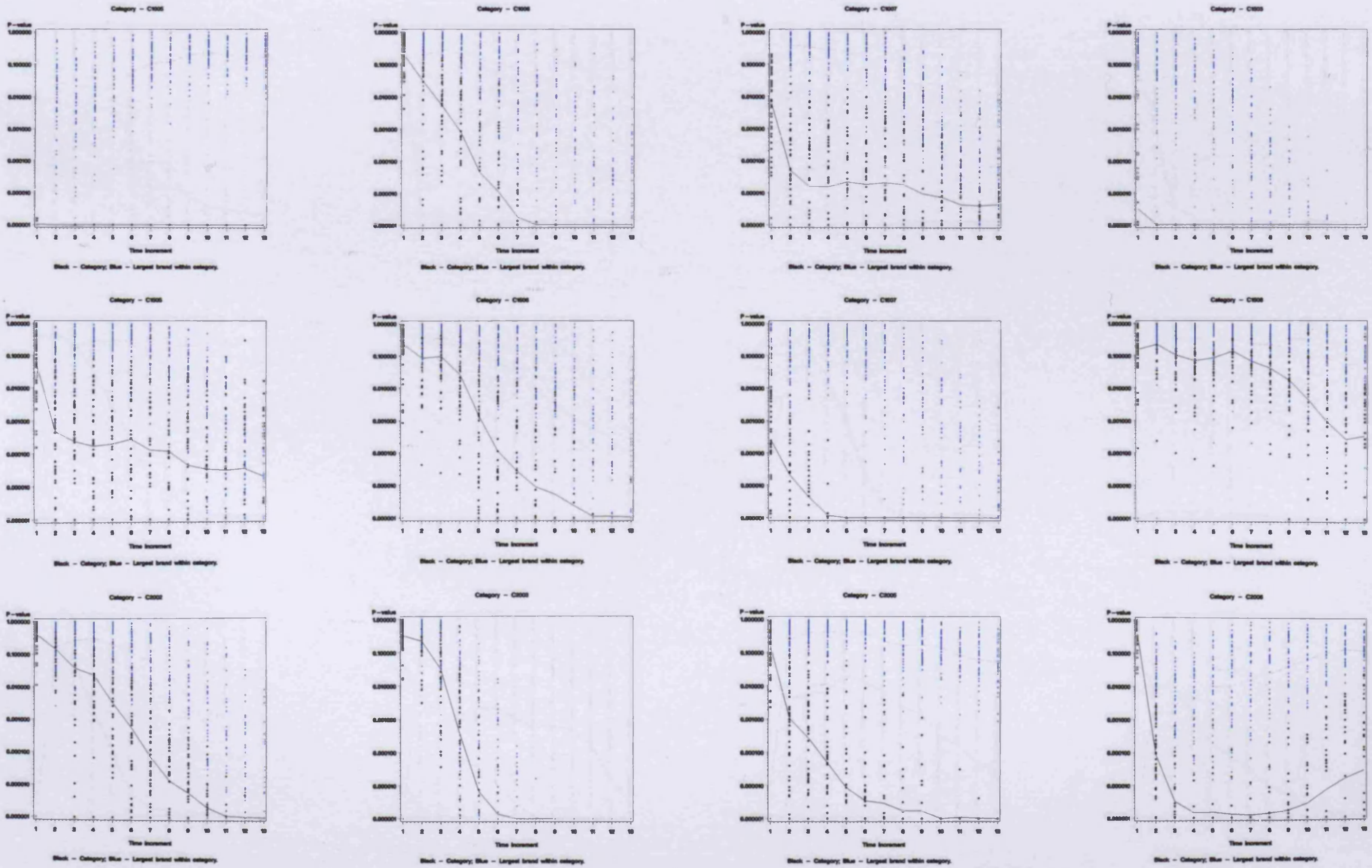
$$\begin{aligned}
\lim_{N\to\infty} N \mathrm{Cov}\left(\bar{X}, \widehat{c^X}\right) &= \frac{1}{1-\alpha} \mathrm{Cov}\left(X_\pi, c^{X_\pi}\right) \\
&\quad + \lim_{N\to\infty} \sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left(E\left[X_\pi(1-\alpha^r+\alpha^r c)^{X_\pi}\right] G_\varepsilon(c;\ \alpha^r) - E\left[X_\pi\right] E\left[c^{X_\pi}\right]\right).
\end{aligned}$$

□

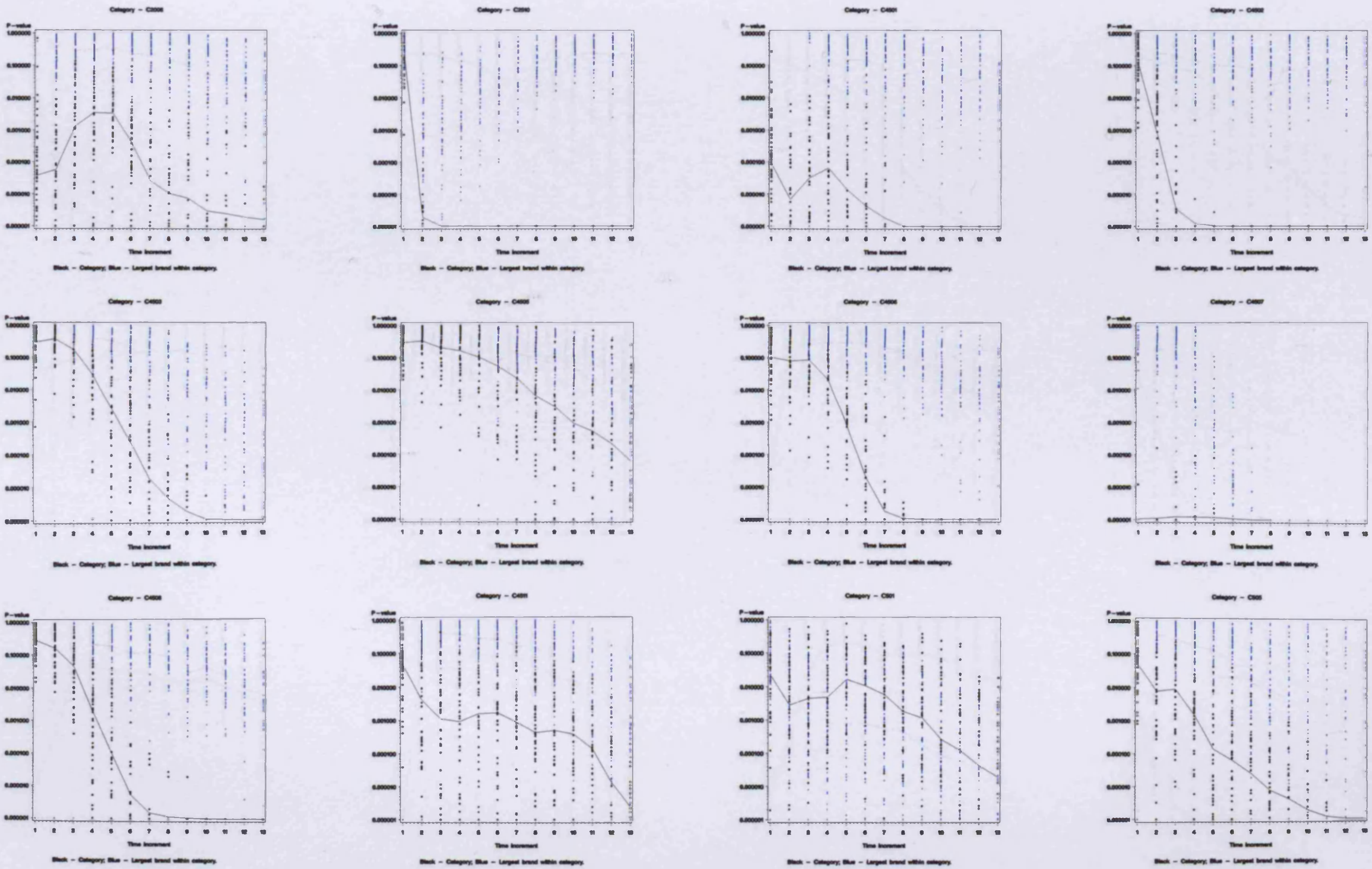
Appendix B

Fitting the gamma Poisson process to 46 categories

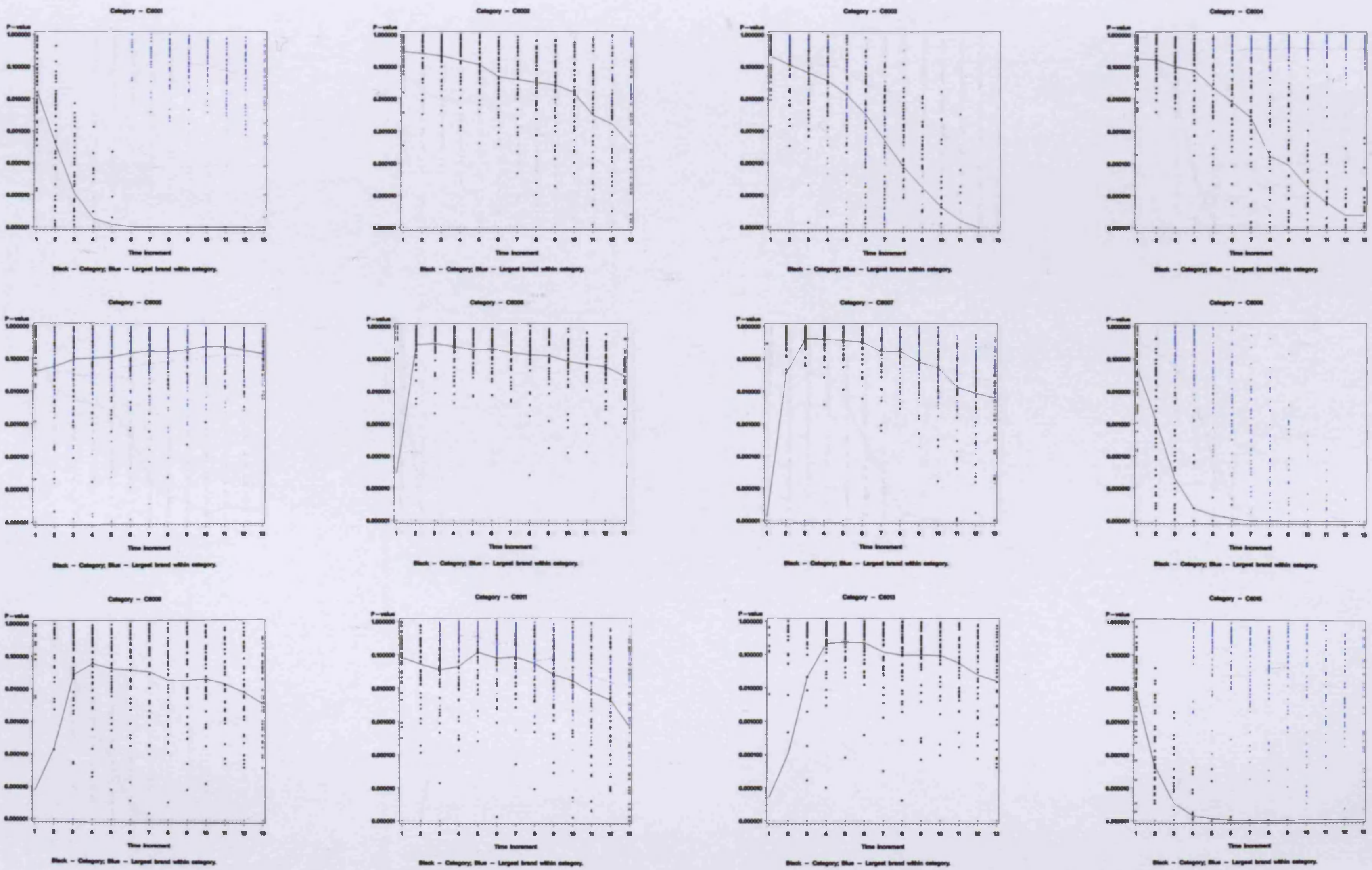
B.1 NBD Chi-square goodness of fit plots for 46 categories



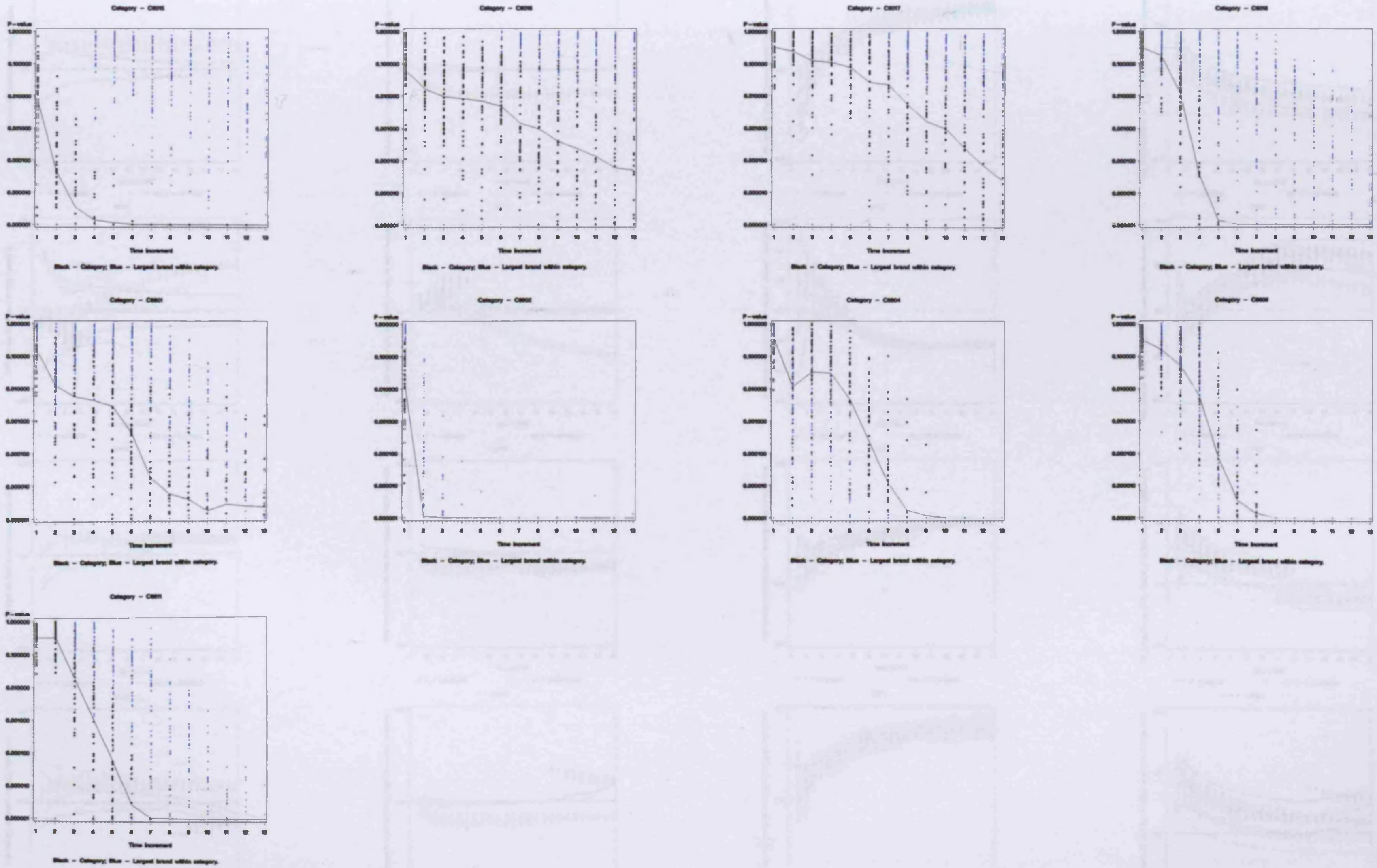
NBD Chi-square goodness of fit plots for 46 categories



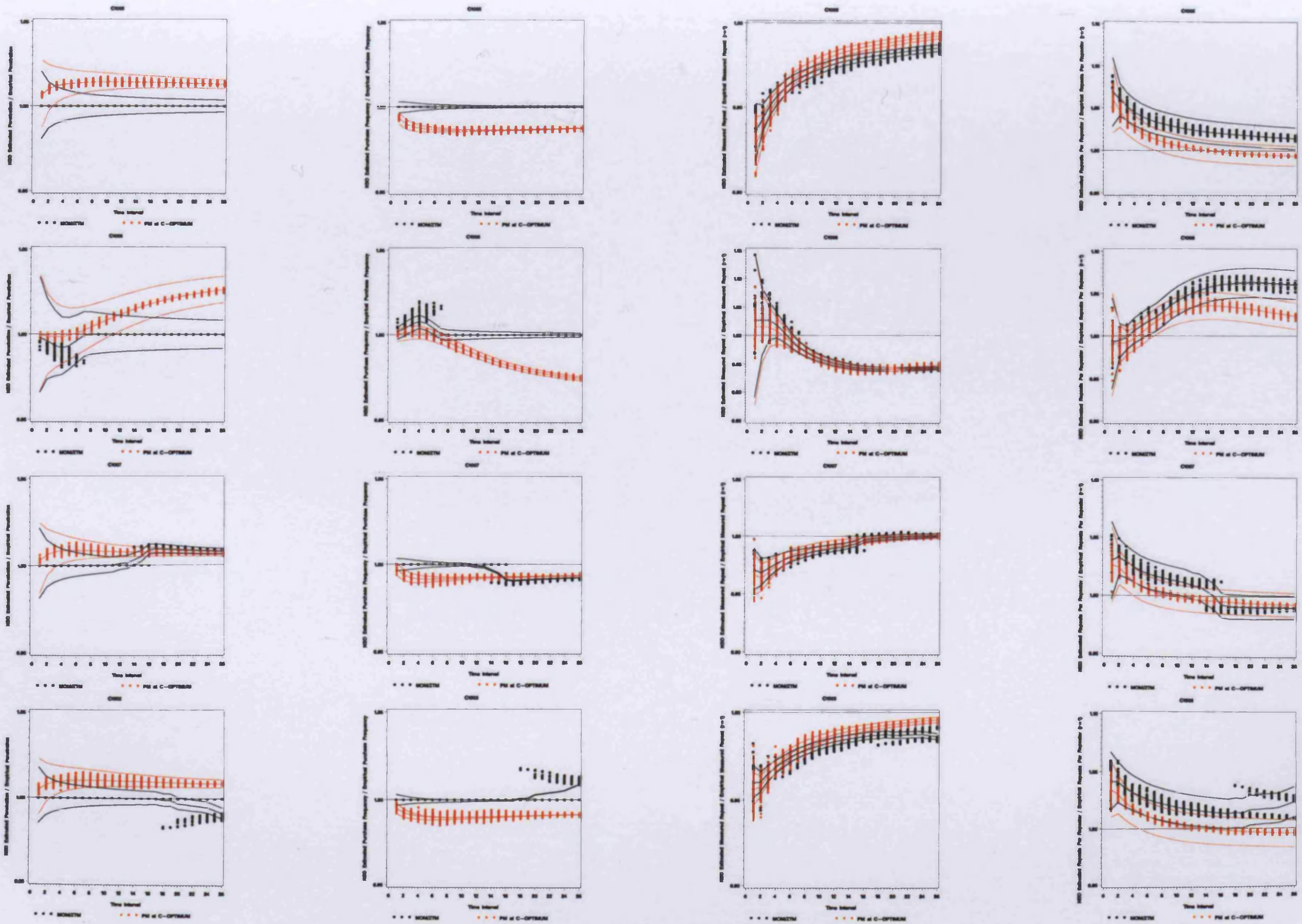
NBD Chi-square goodness of fit plots for 46 categories



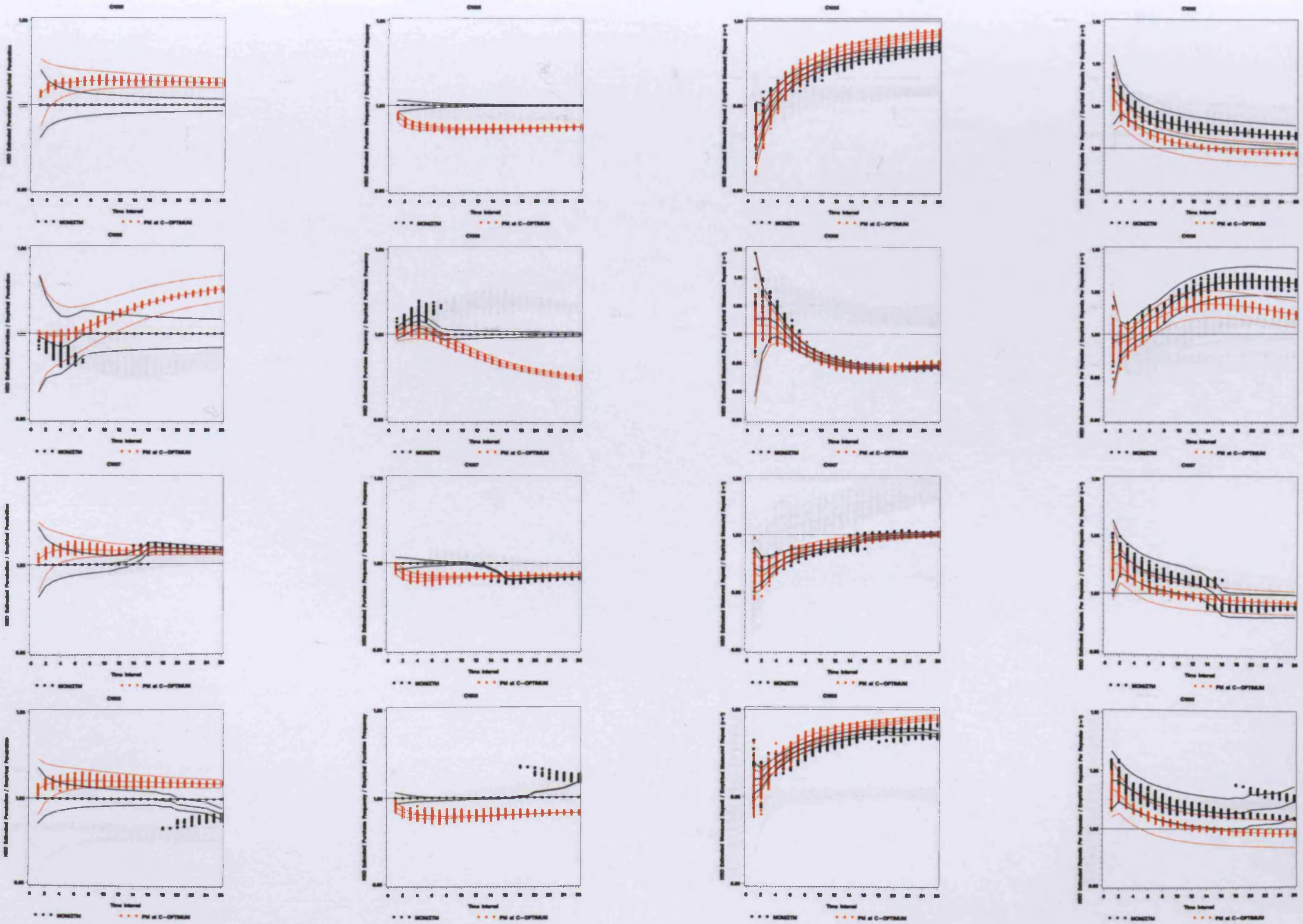
NBD Chi-square goodness of fit plots for 46 categories



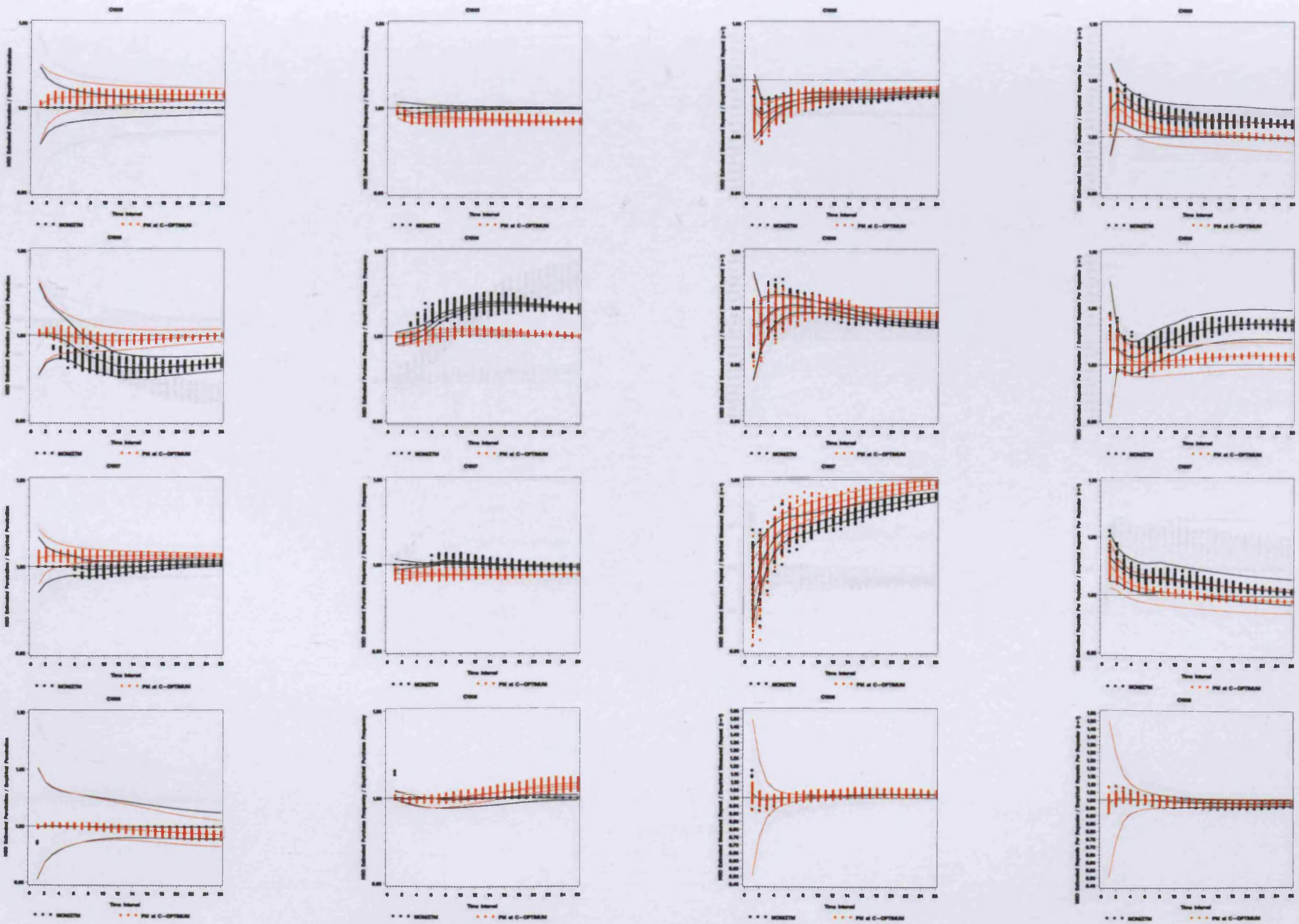
B.2 Ratio of NBD estimated market research measures to empirical market research measures



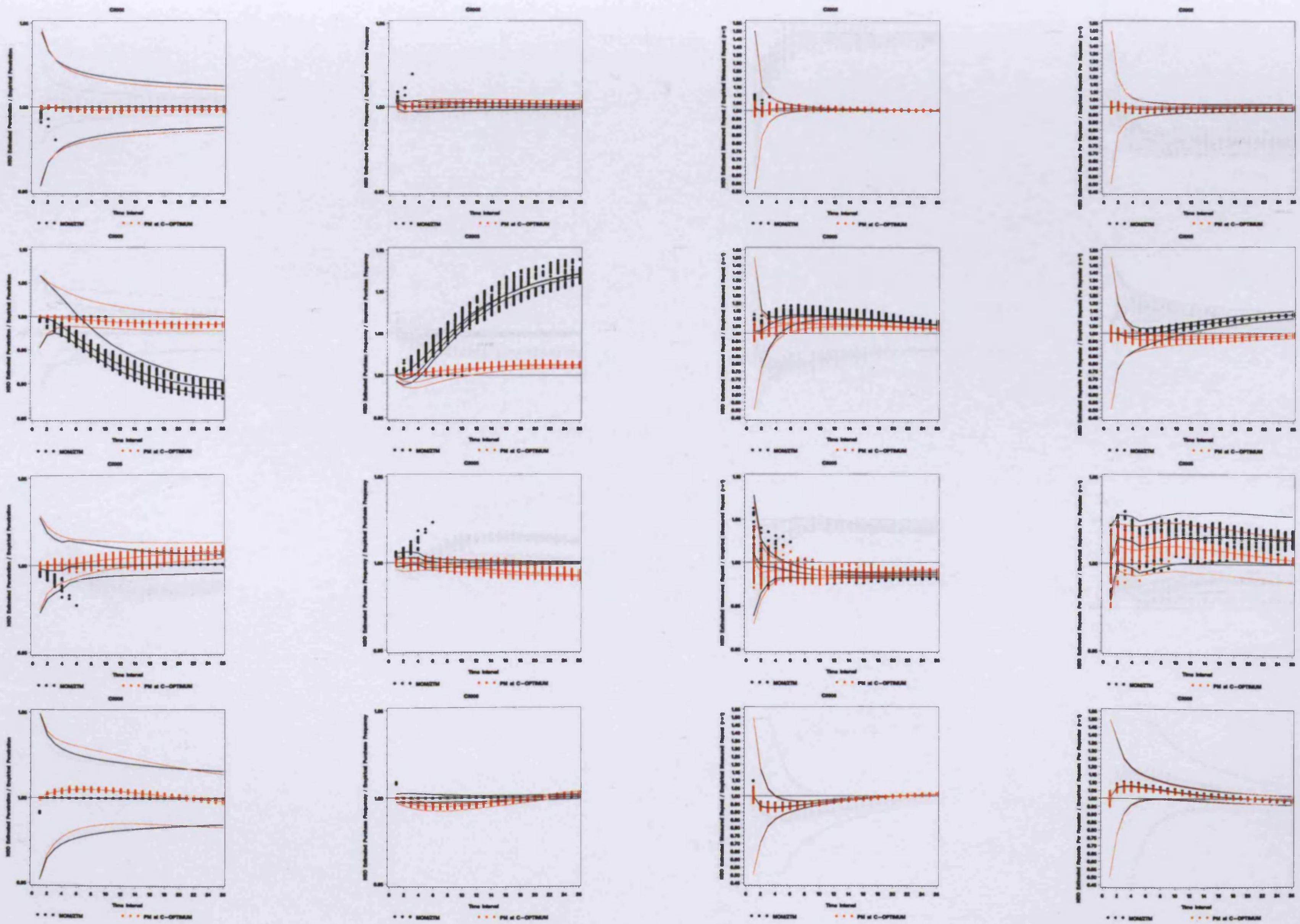
Ratio of NBD estimated market research measures to empirical market research measures



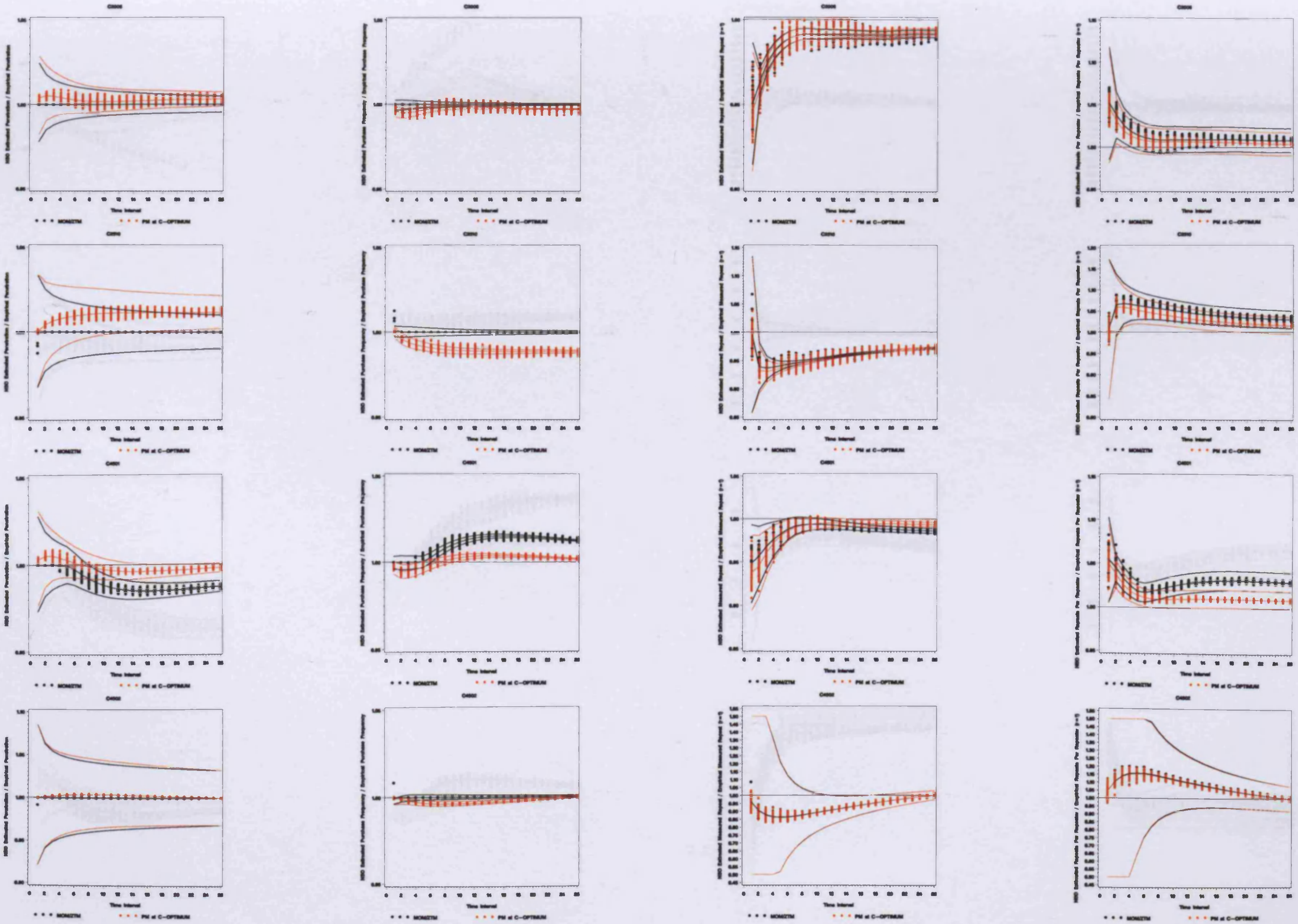
Ratio of NBD estimated market research measures to empirical market research measures



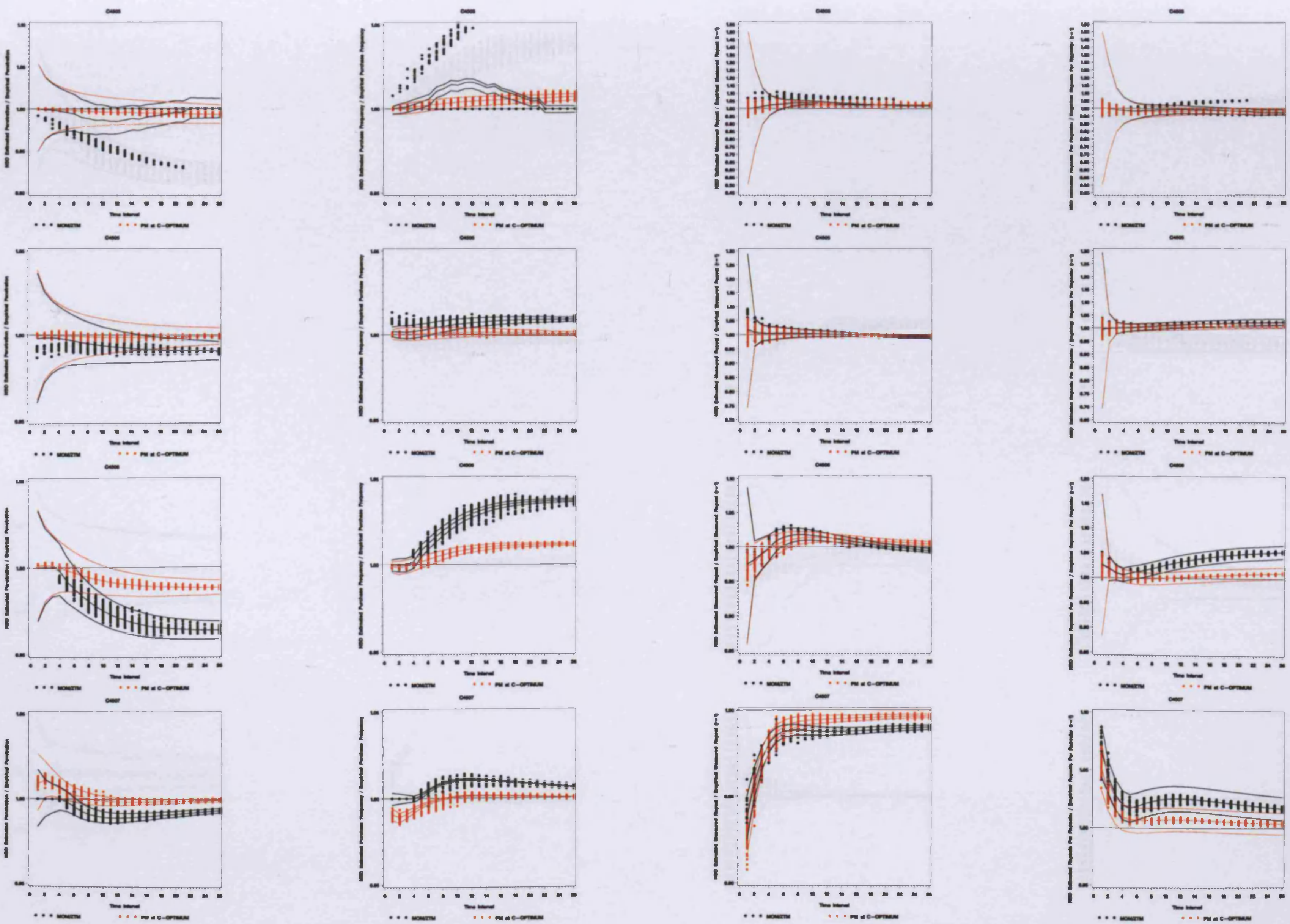
Ratio of NBD estimated market research measures to empirical market research measures



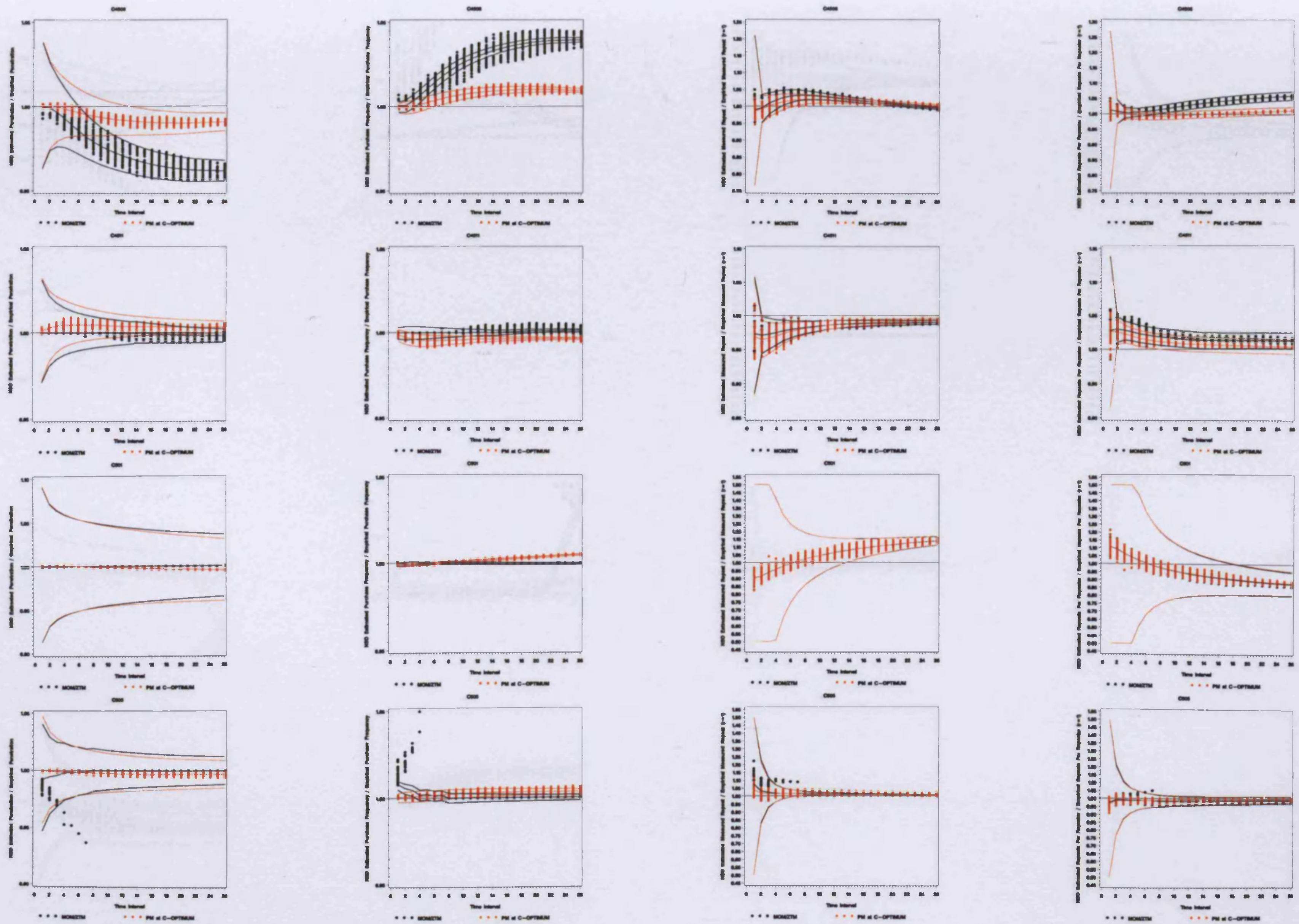
Ratio of NBD estimated market research measures to empirical market research measures



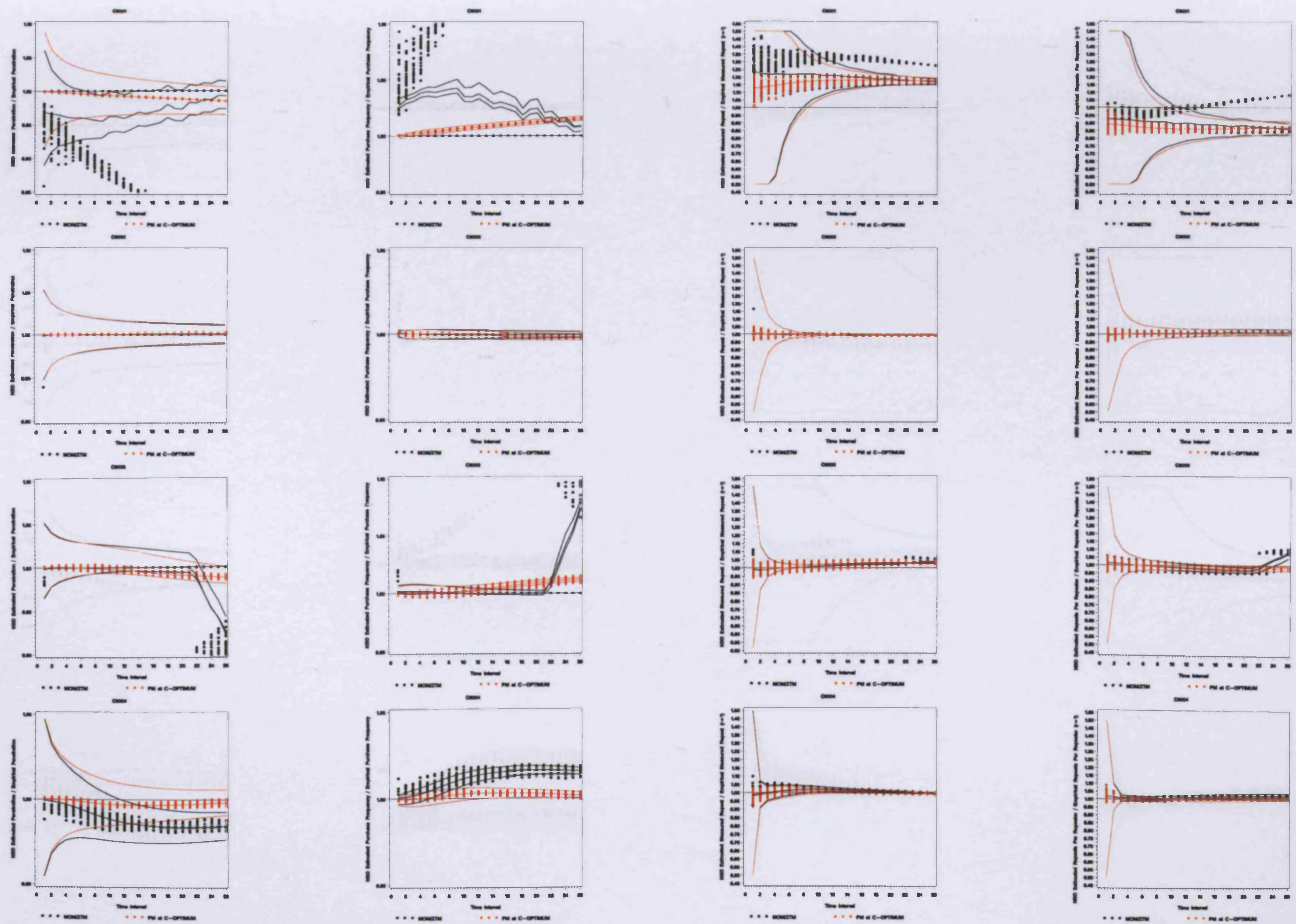
Ratio of NBD estimated market research measures to empirical market research measures



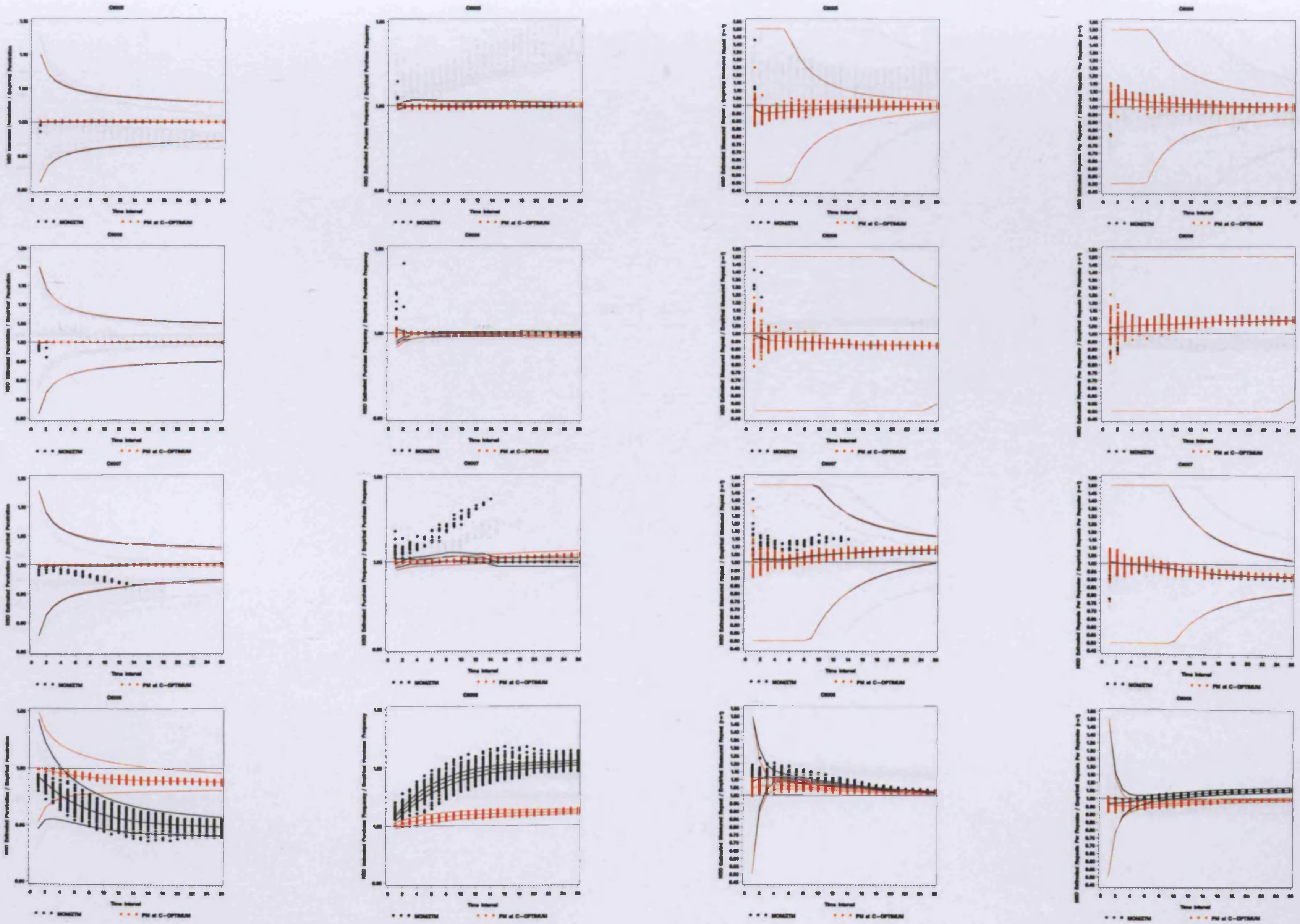
Ratio of NBD estimated market research measures to empirical market research measures



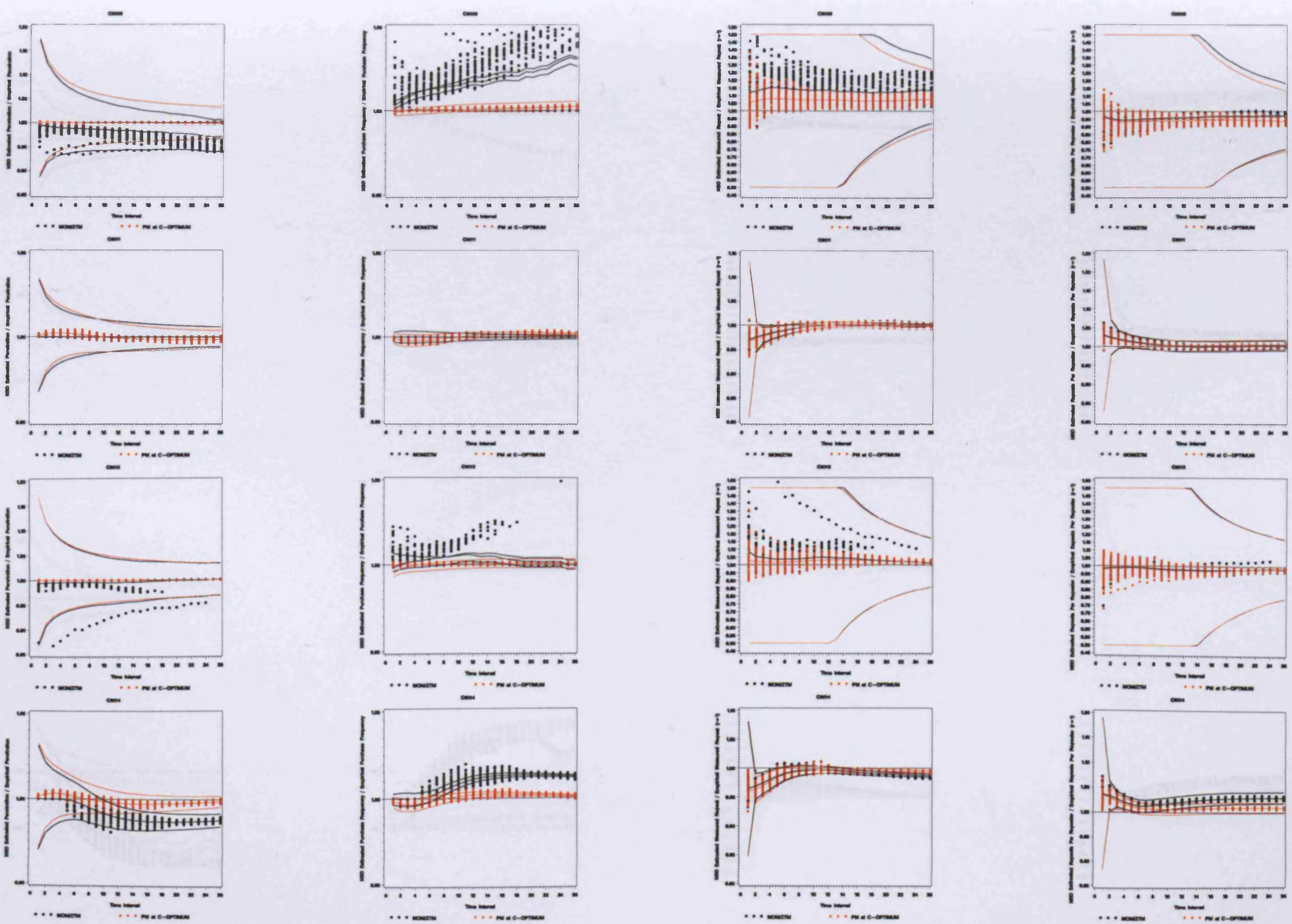
Ratio of NBD estimated market research measures to empirical market research measures



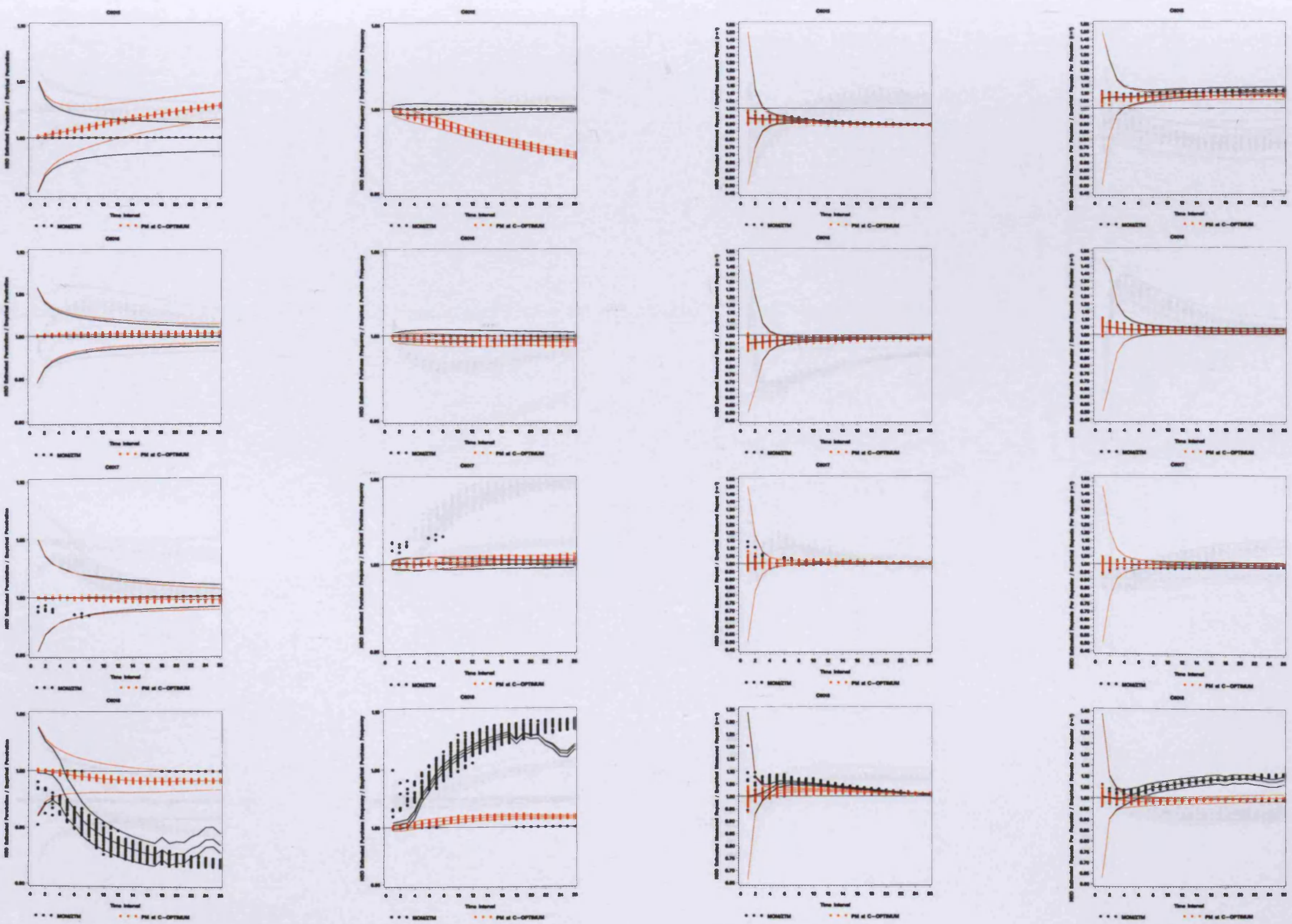
Ratio of NBD estimated market research measures to empirical market research measures



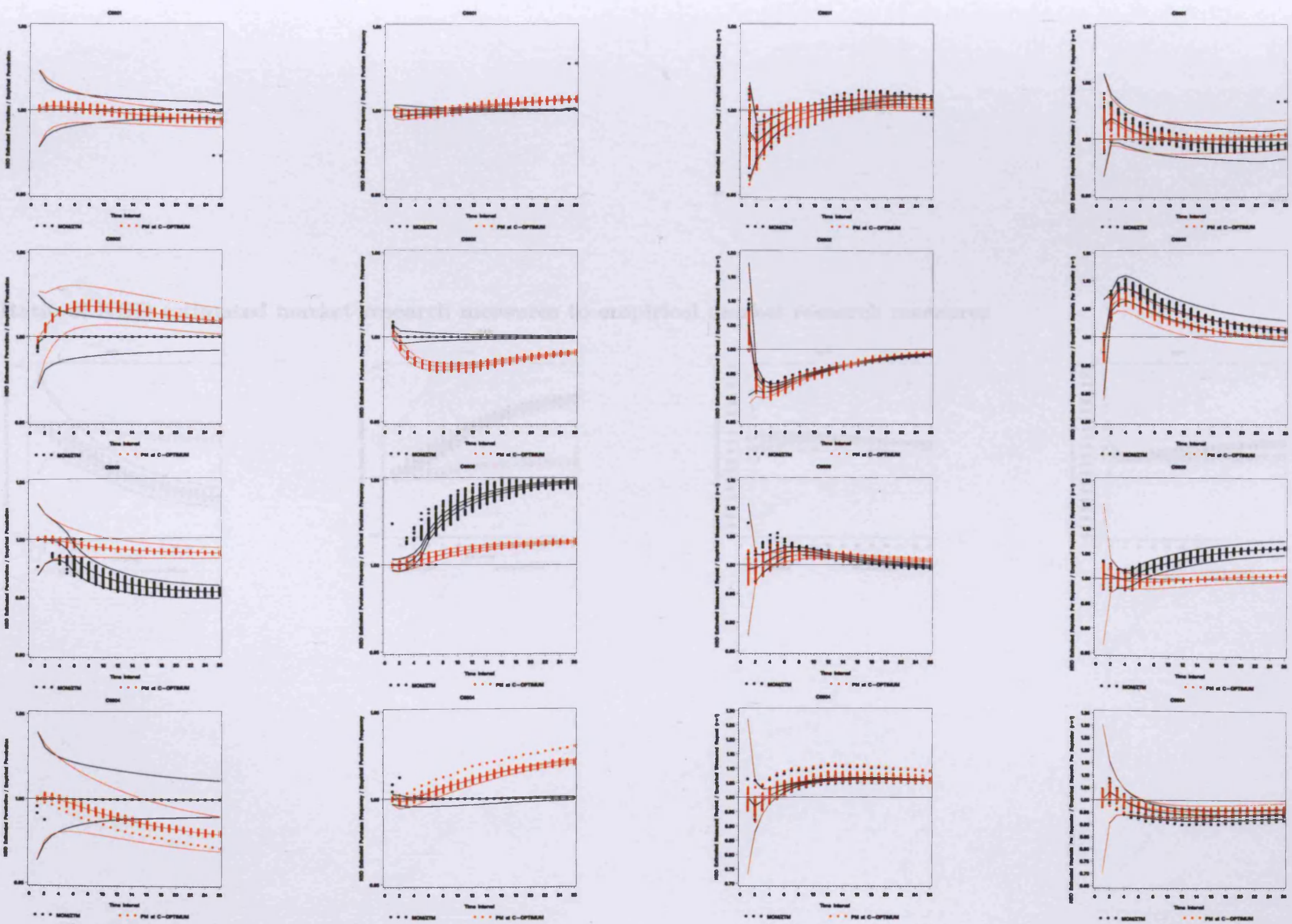
Ratio of NBD estimated market research measures to empirical market research measures



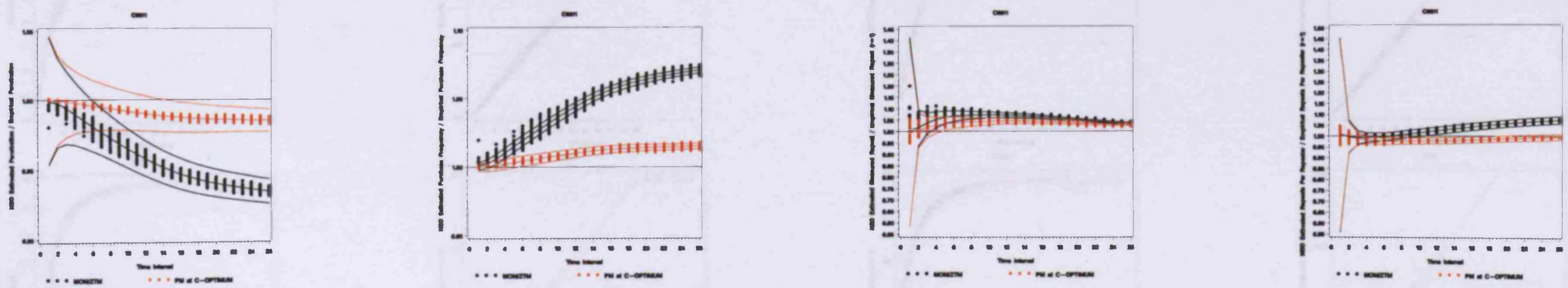
Ratio of NBD estimated market research measures to empirical market research measures



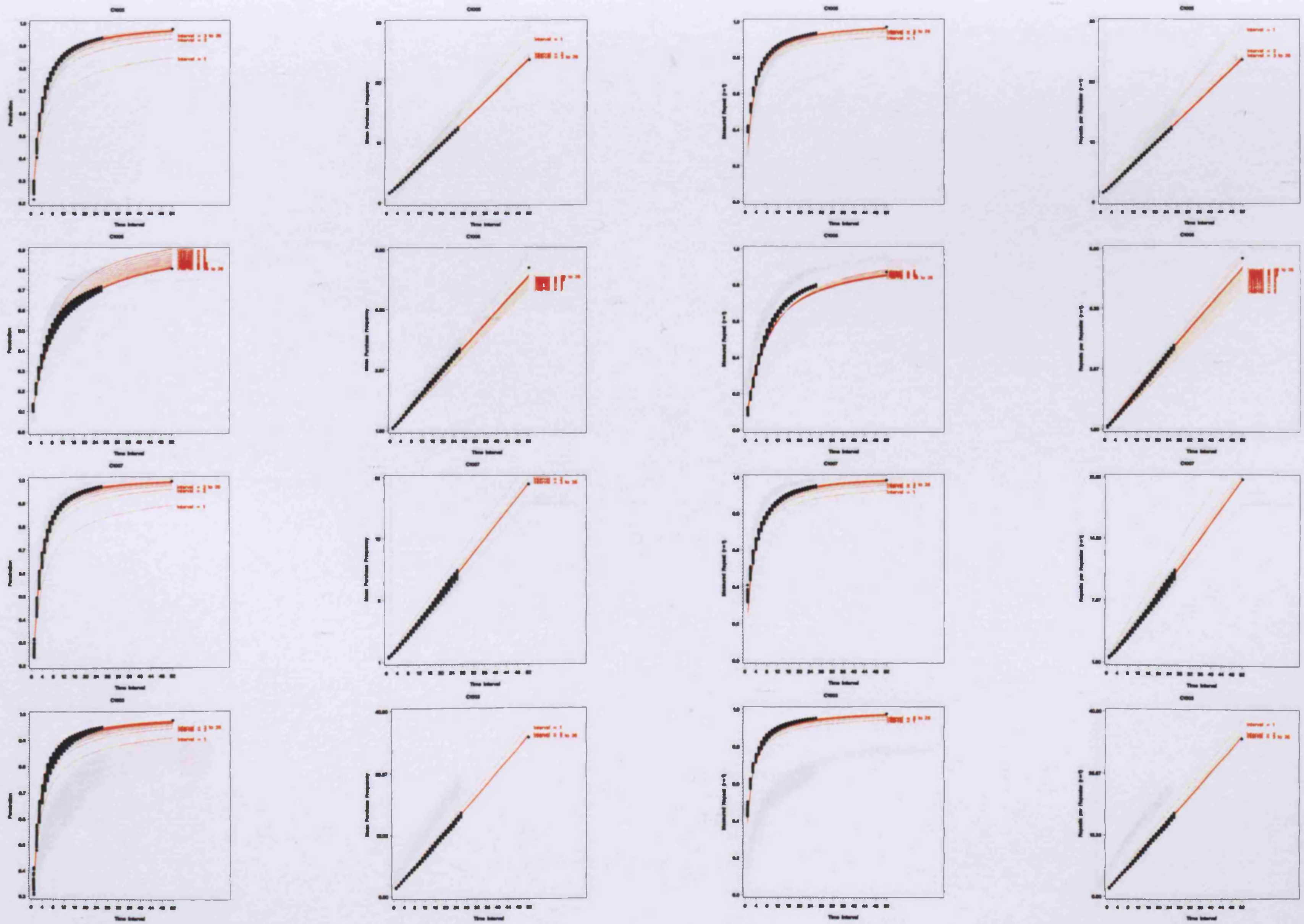
Ratio of NBD estimated market research measures to empirical market research measures

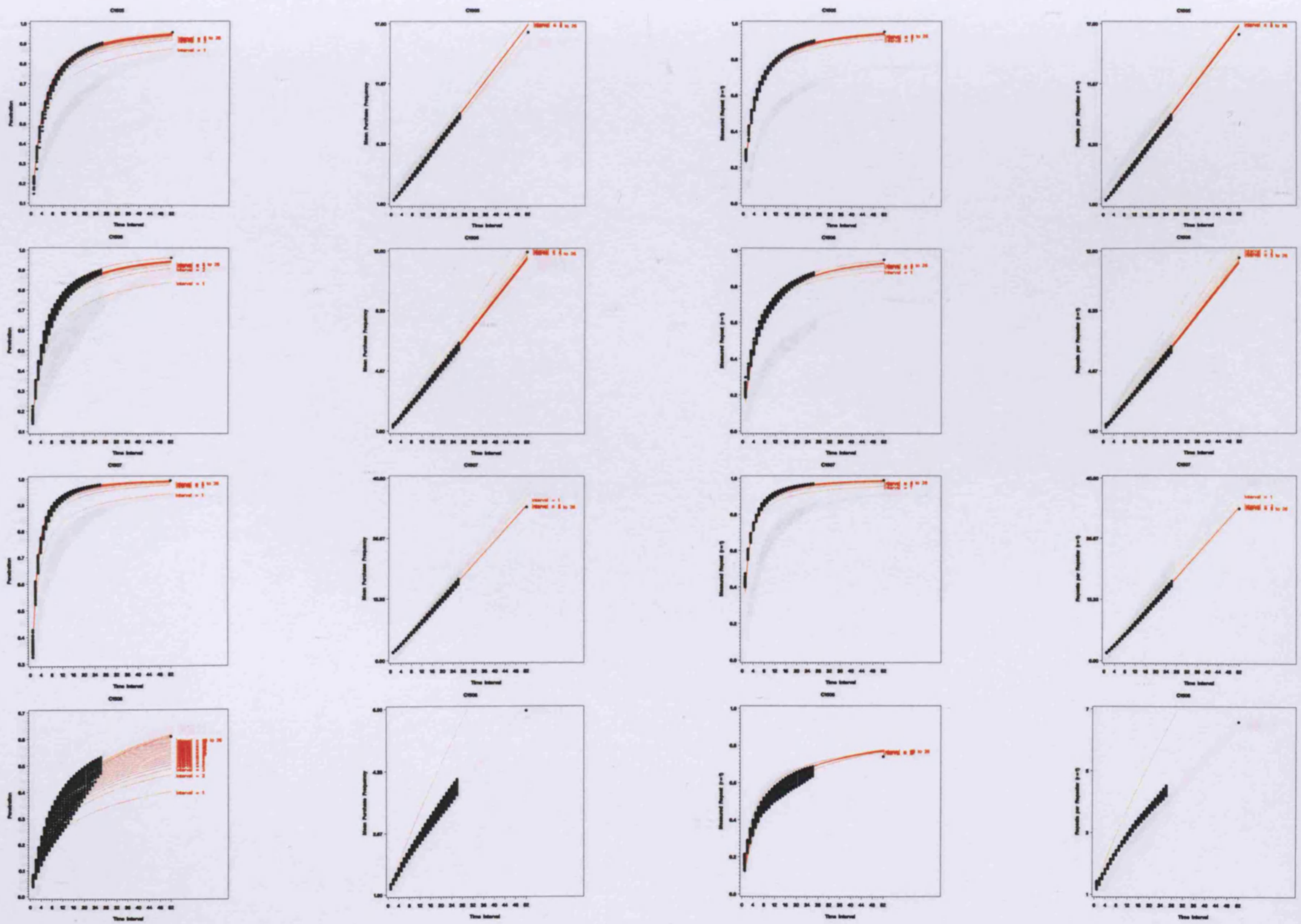


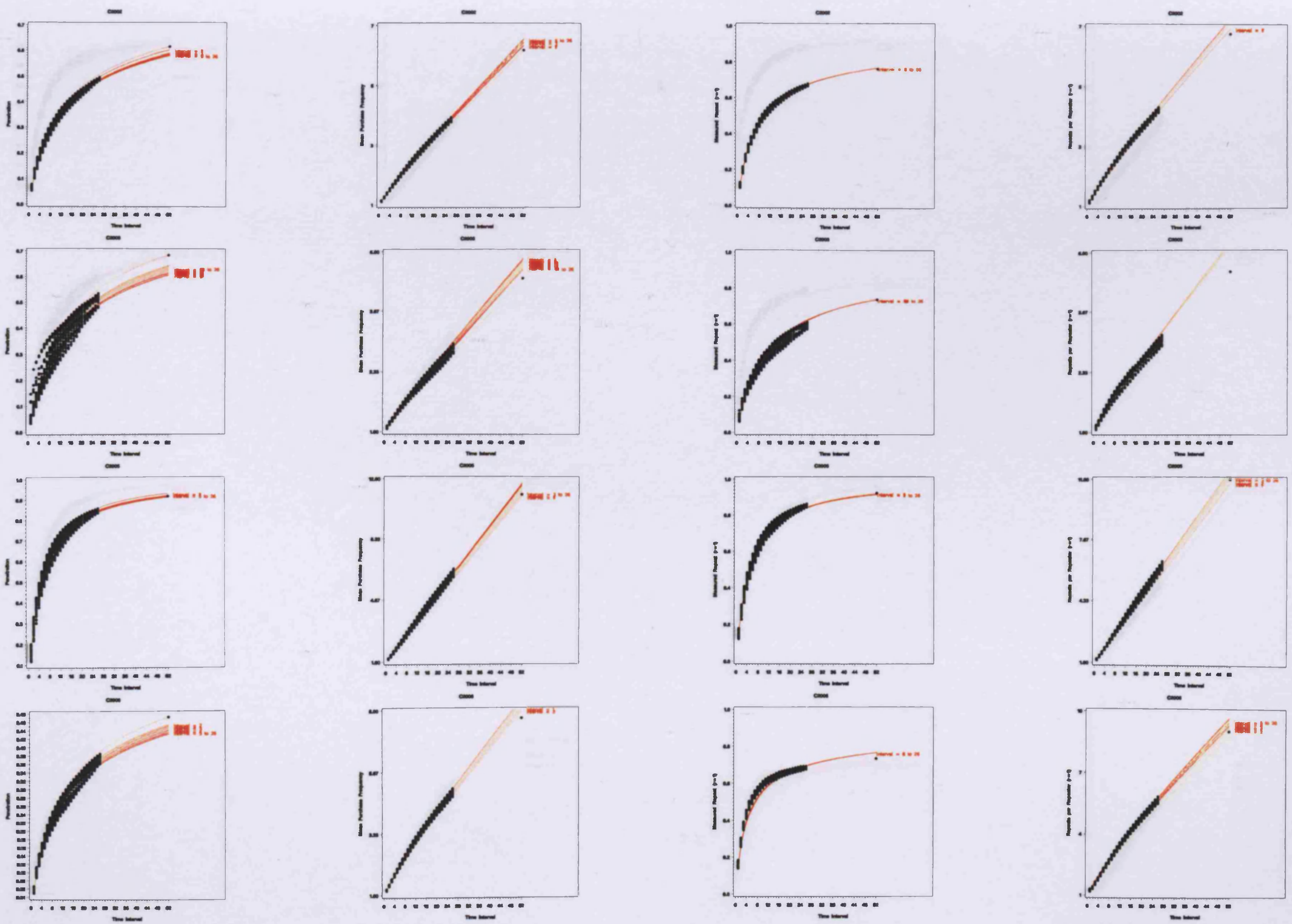
Ratio of NBD estimated market research measures to empirical market research measures



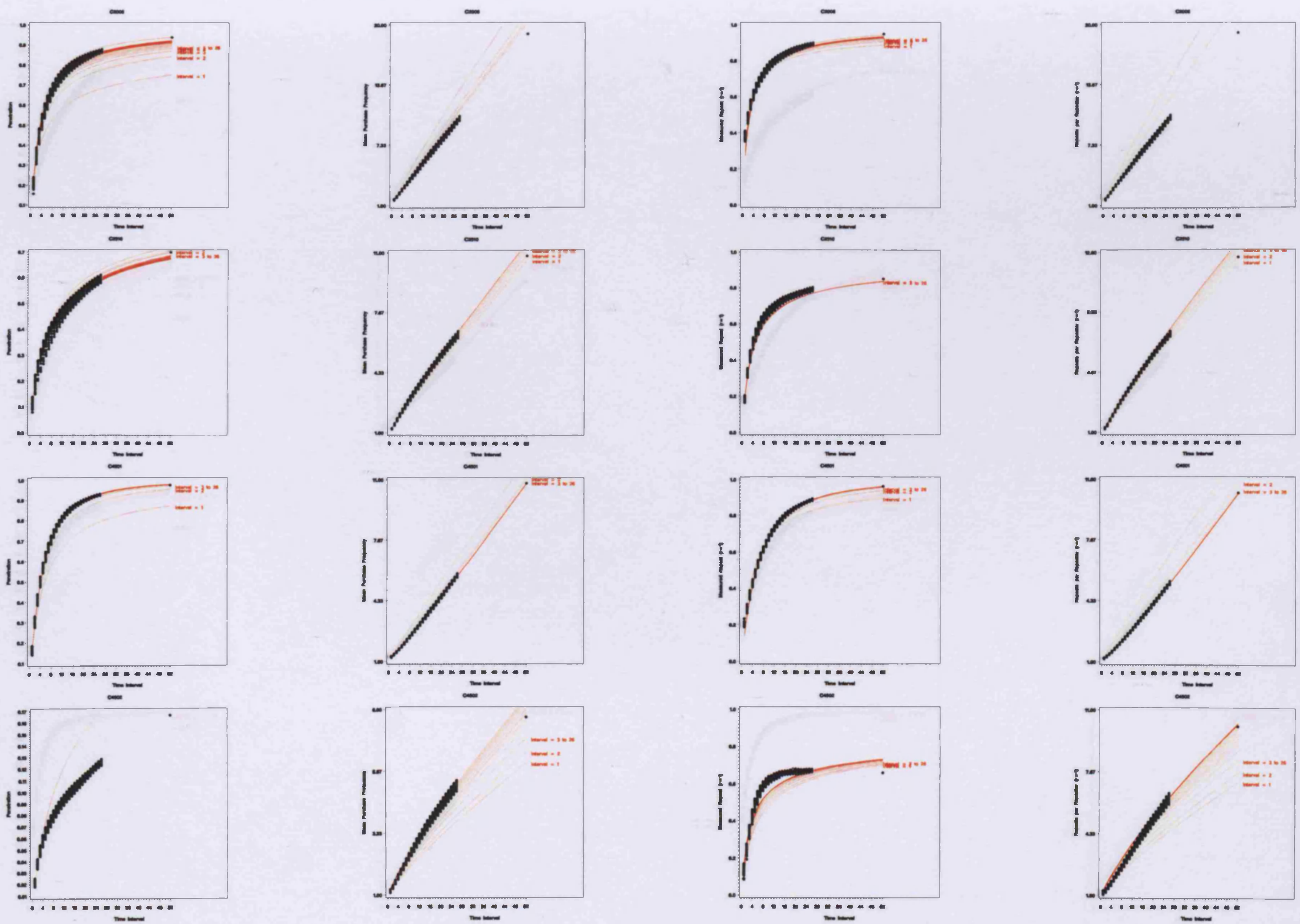
B.3 Extrapolation of market research measures to different length time intervals

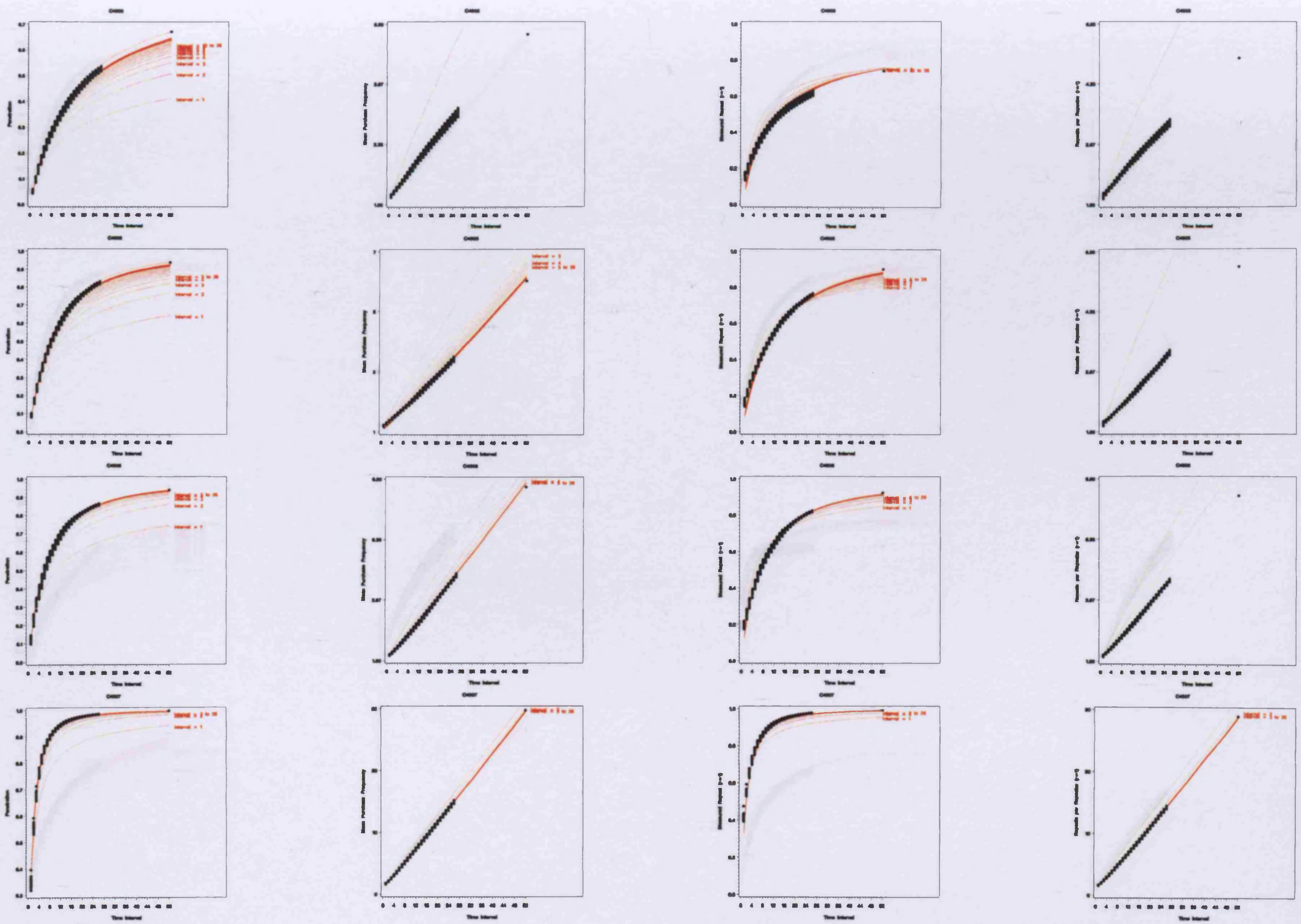




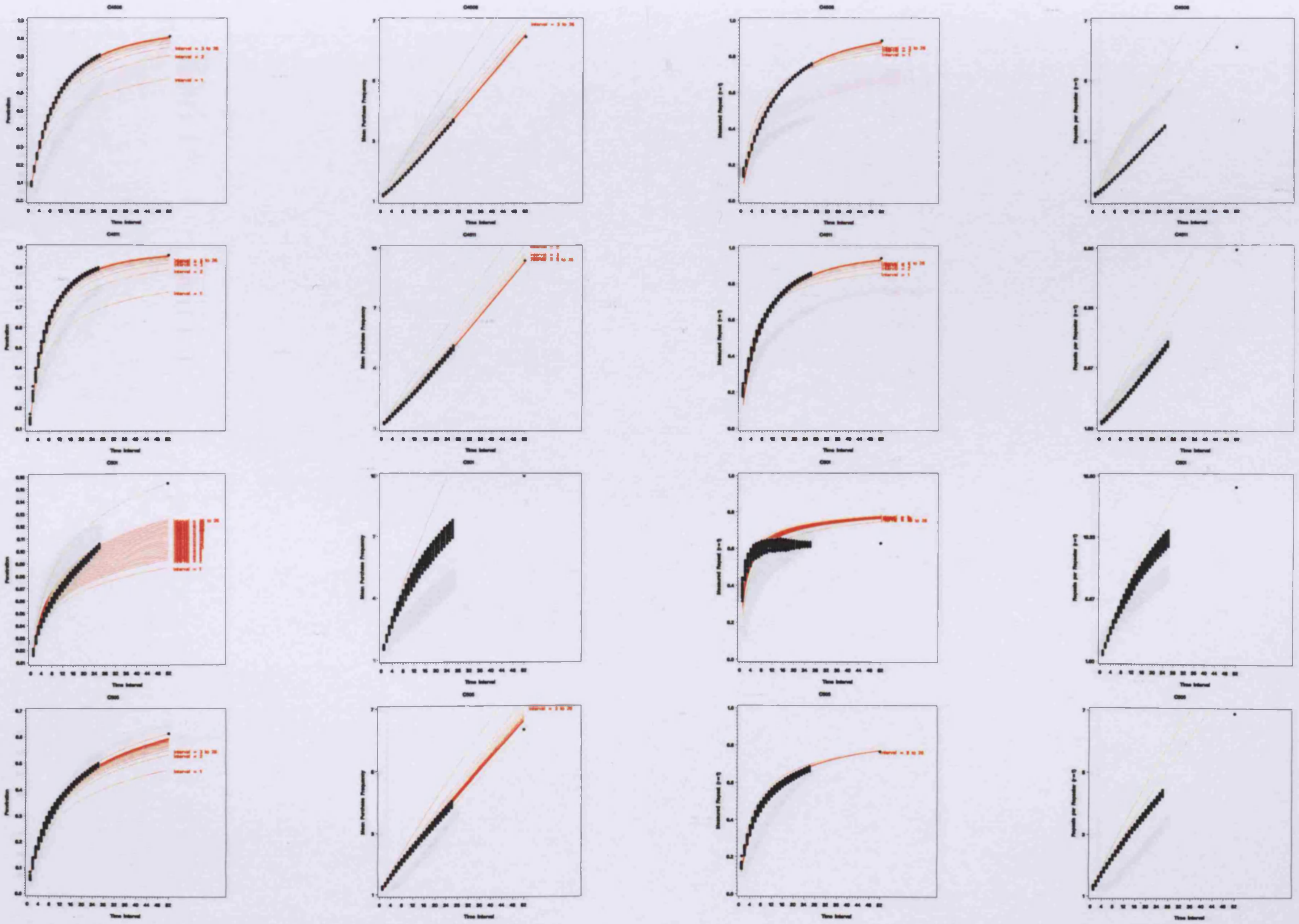


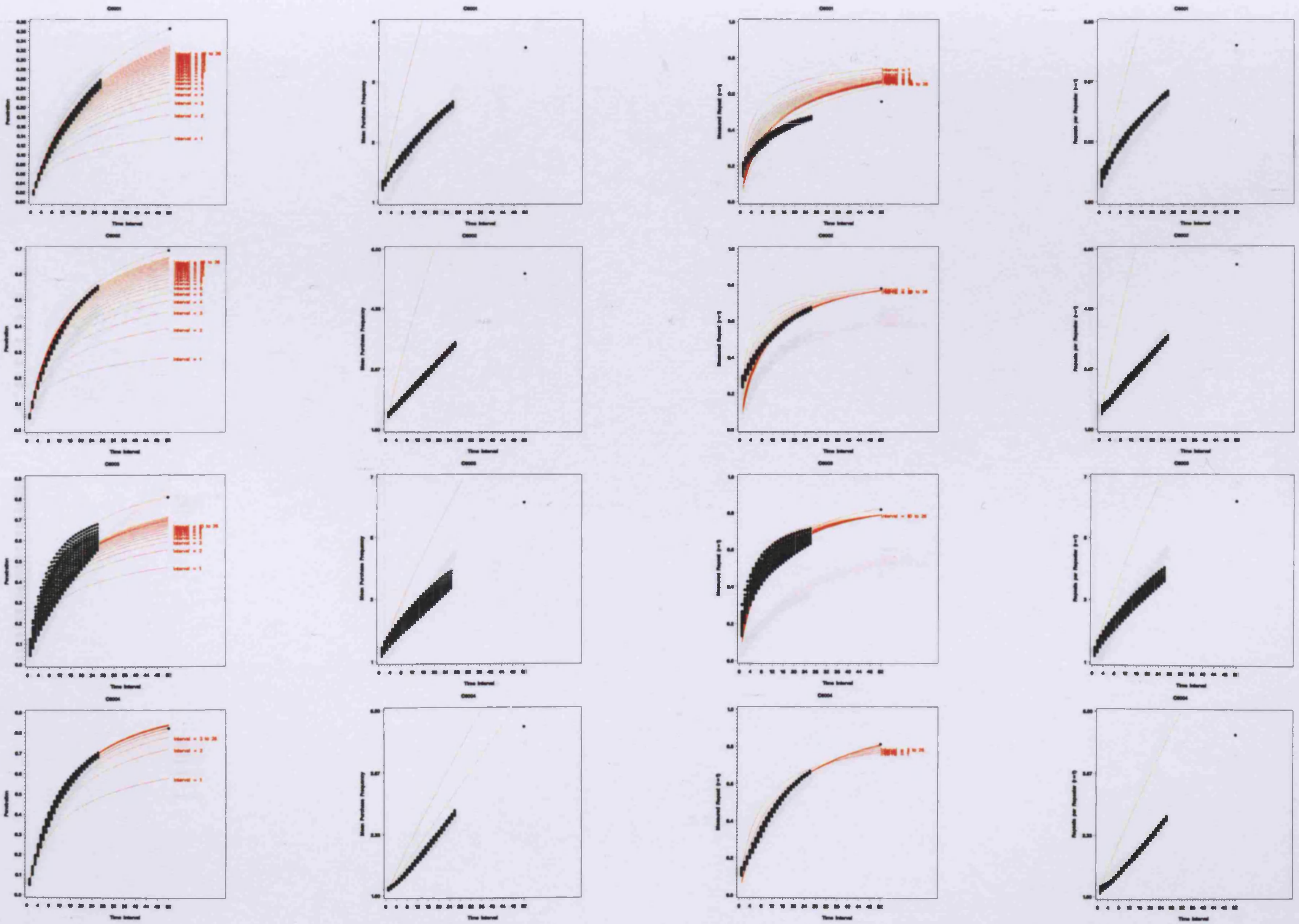
Extrapolation of market research measures to different length time intervals



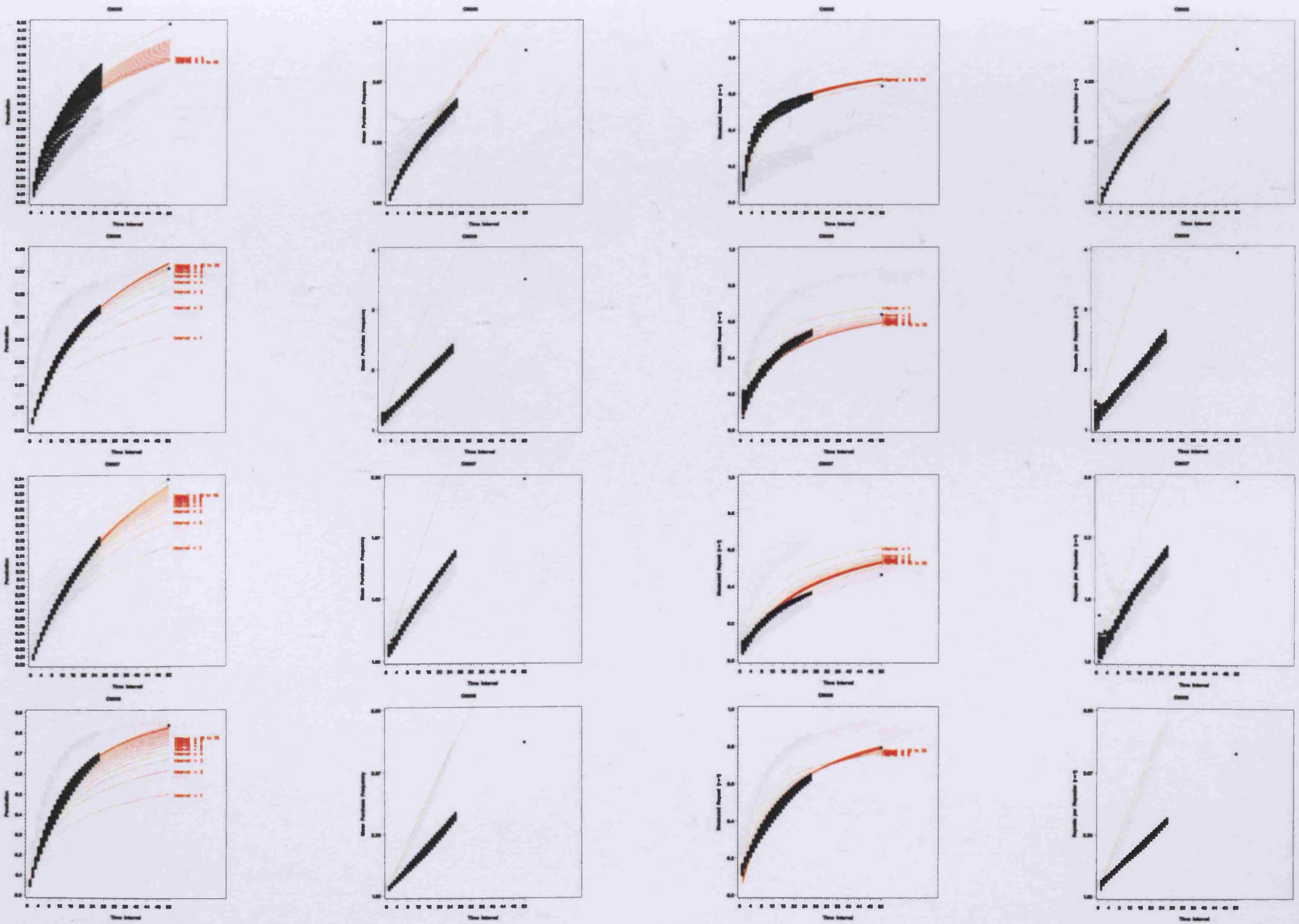


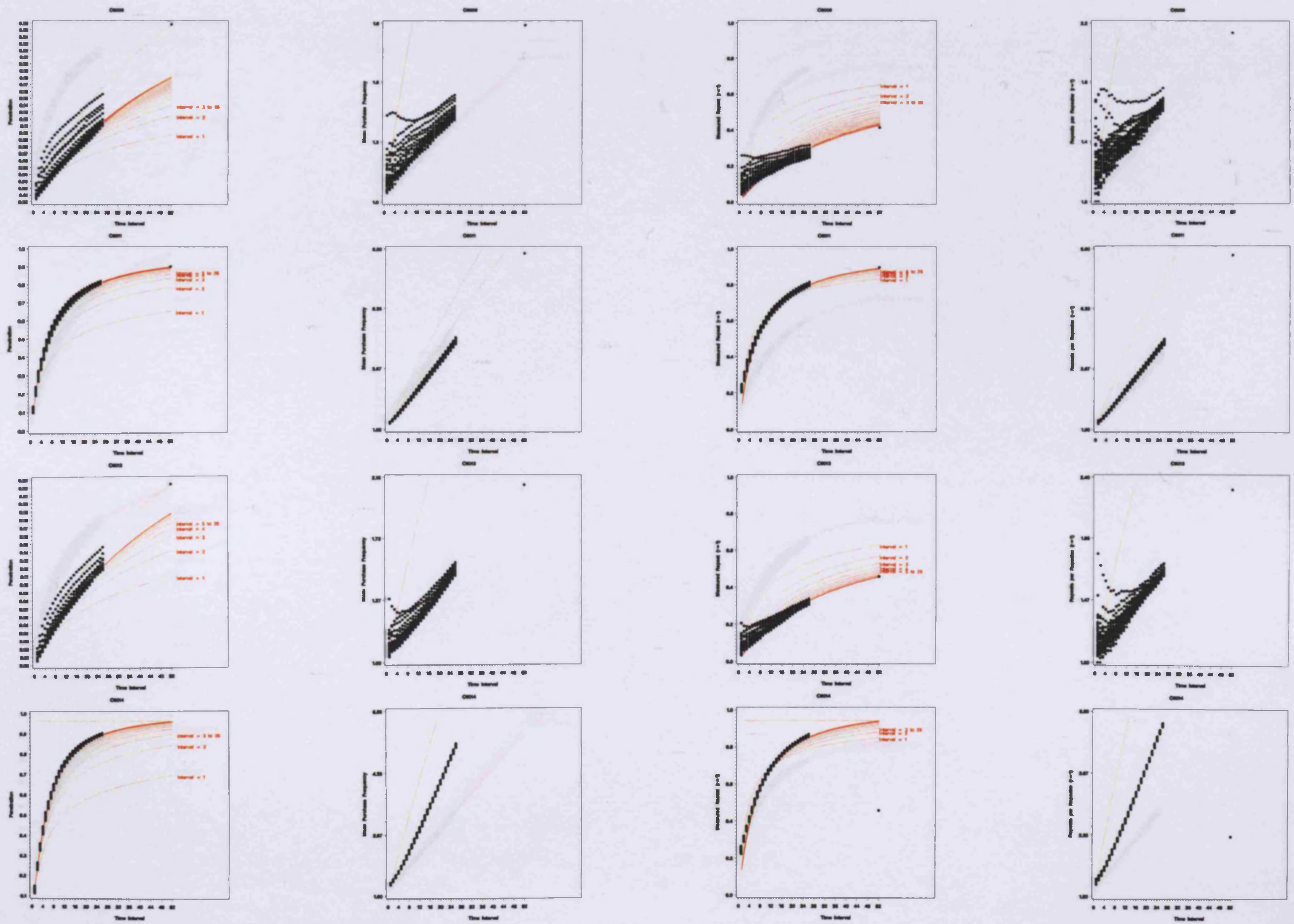
Extrapolation of market research measures to different length time intervals



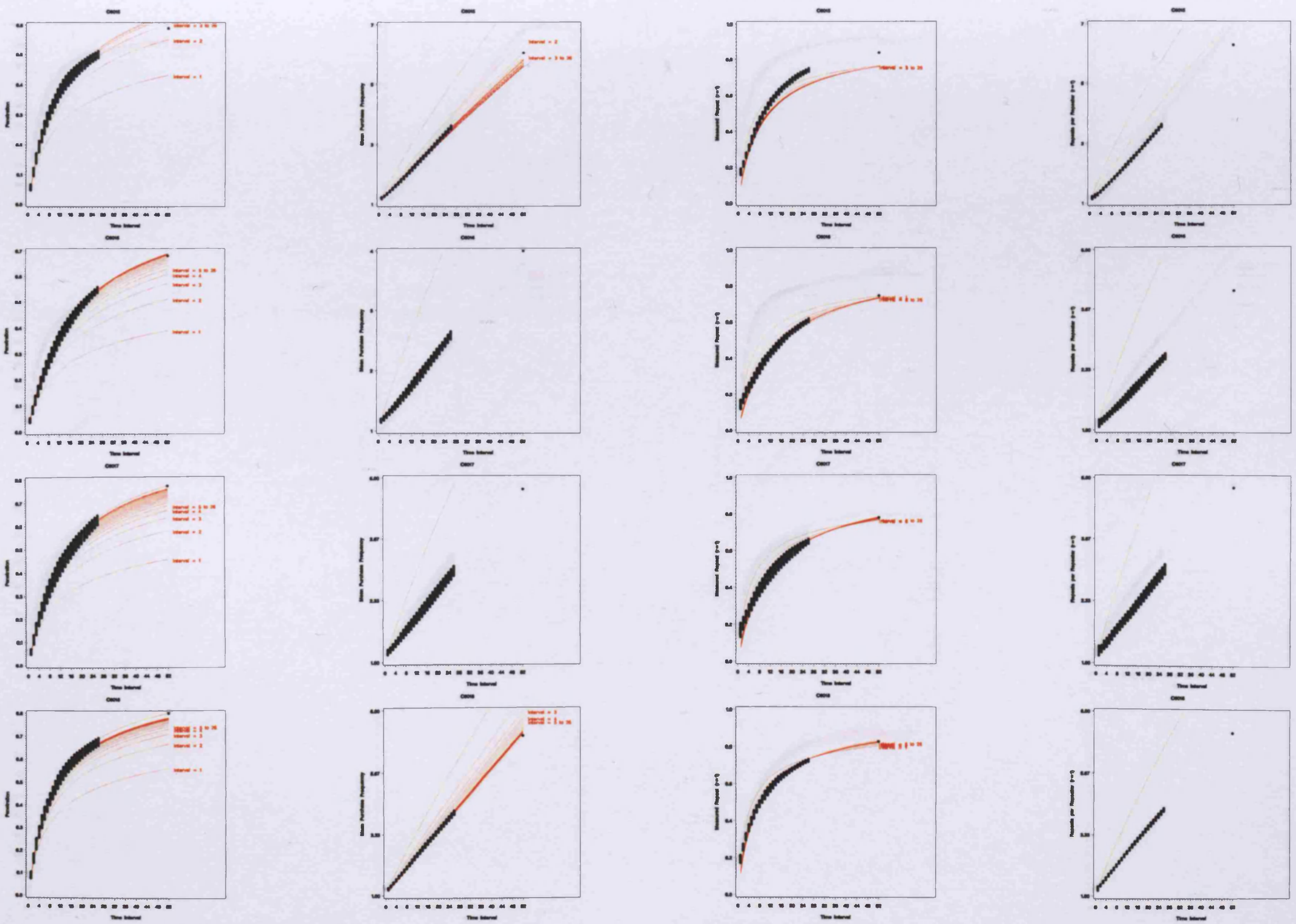


Extrapolation of market research measures to different length time intervals

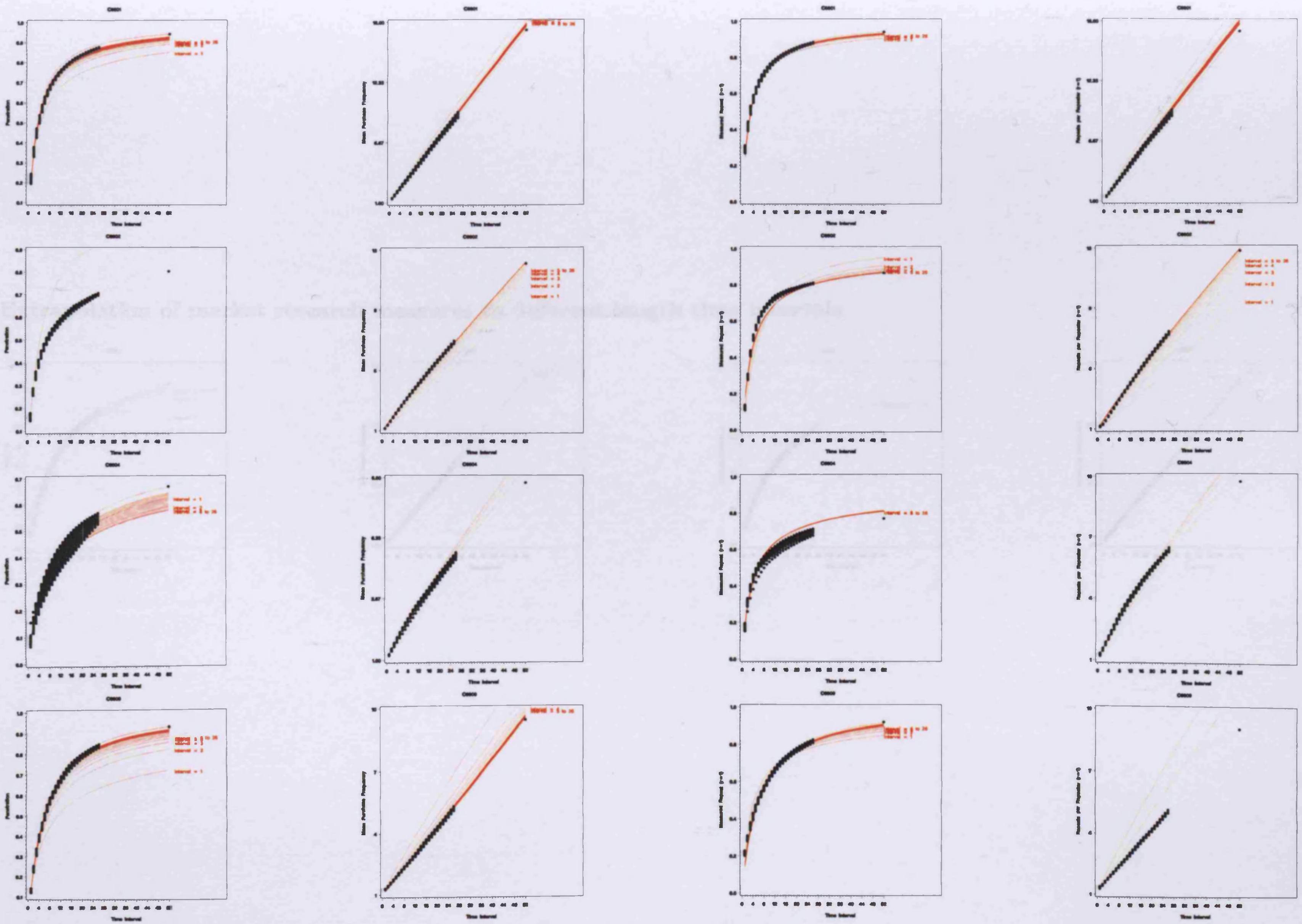




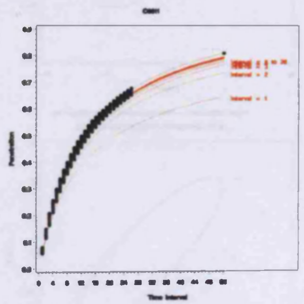
Extrapolation of market research measures to different length time intervals

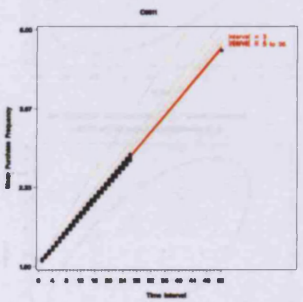


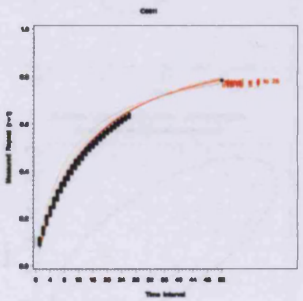
Extrapolation of market research measures to different length time intervals

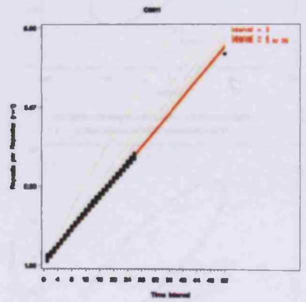


Extrapolation of market research measures to different length time intervals

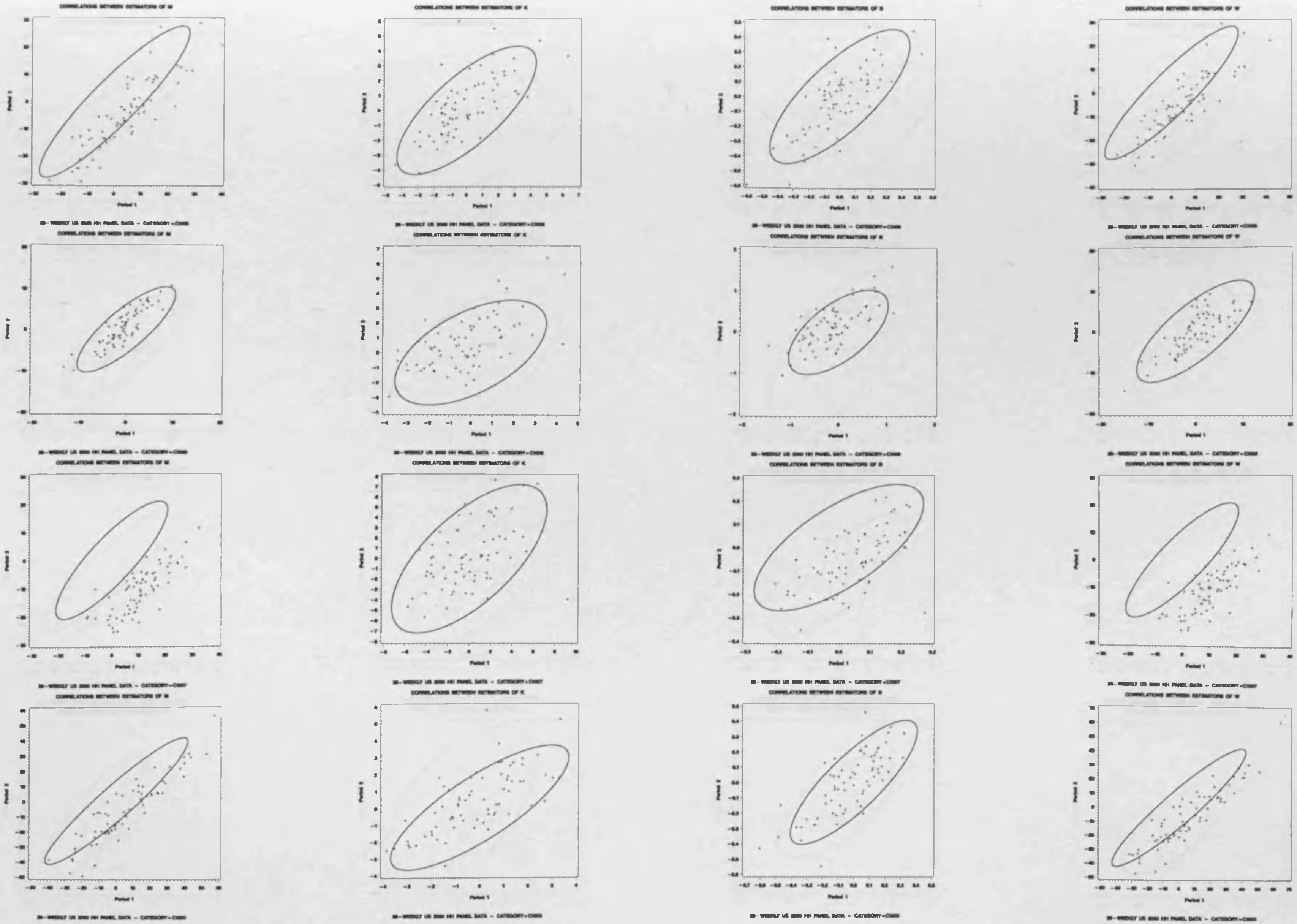




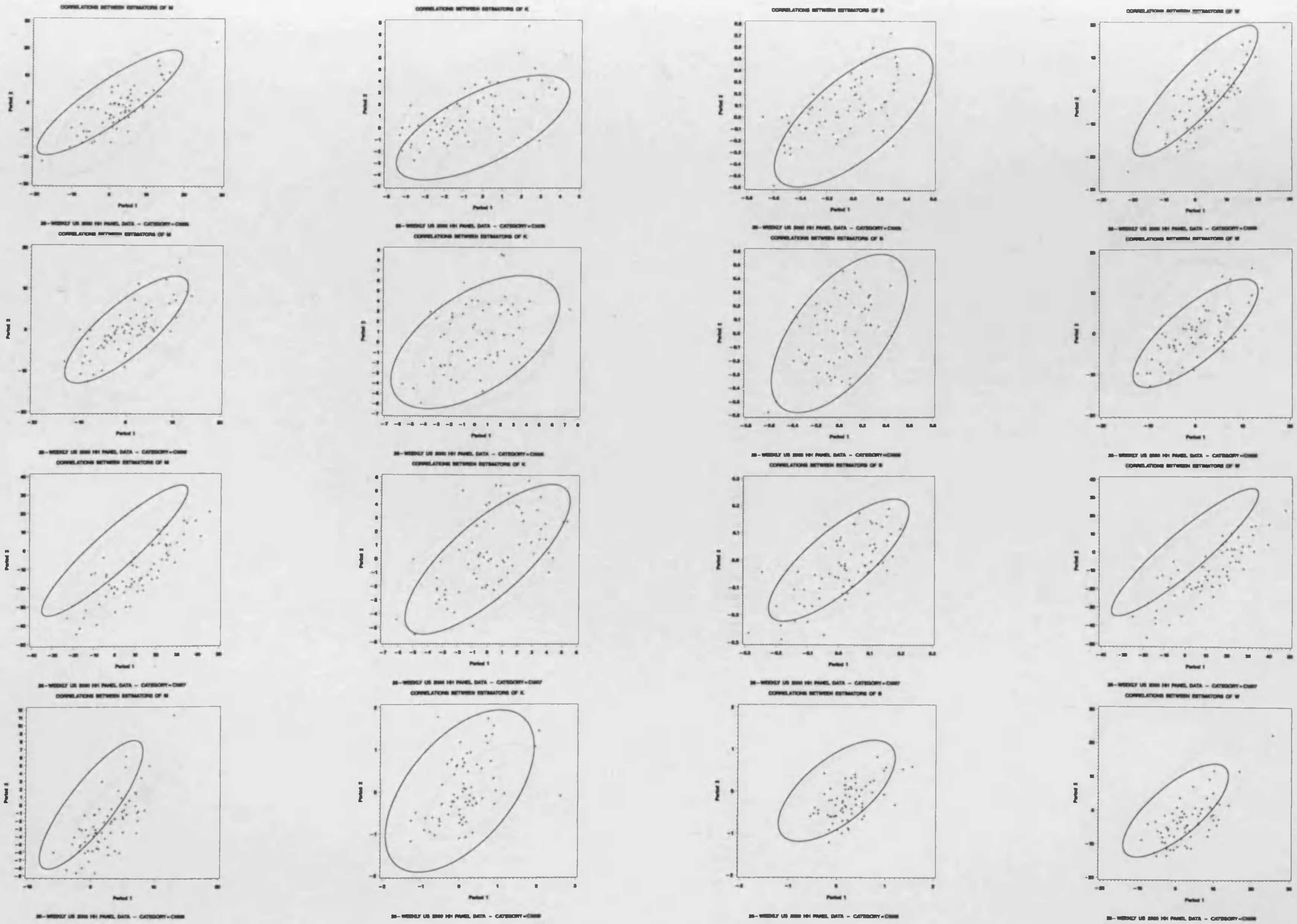




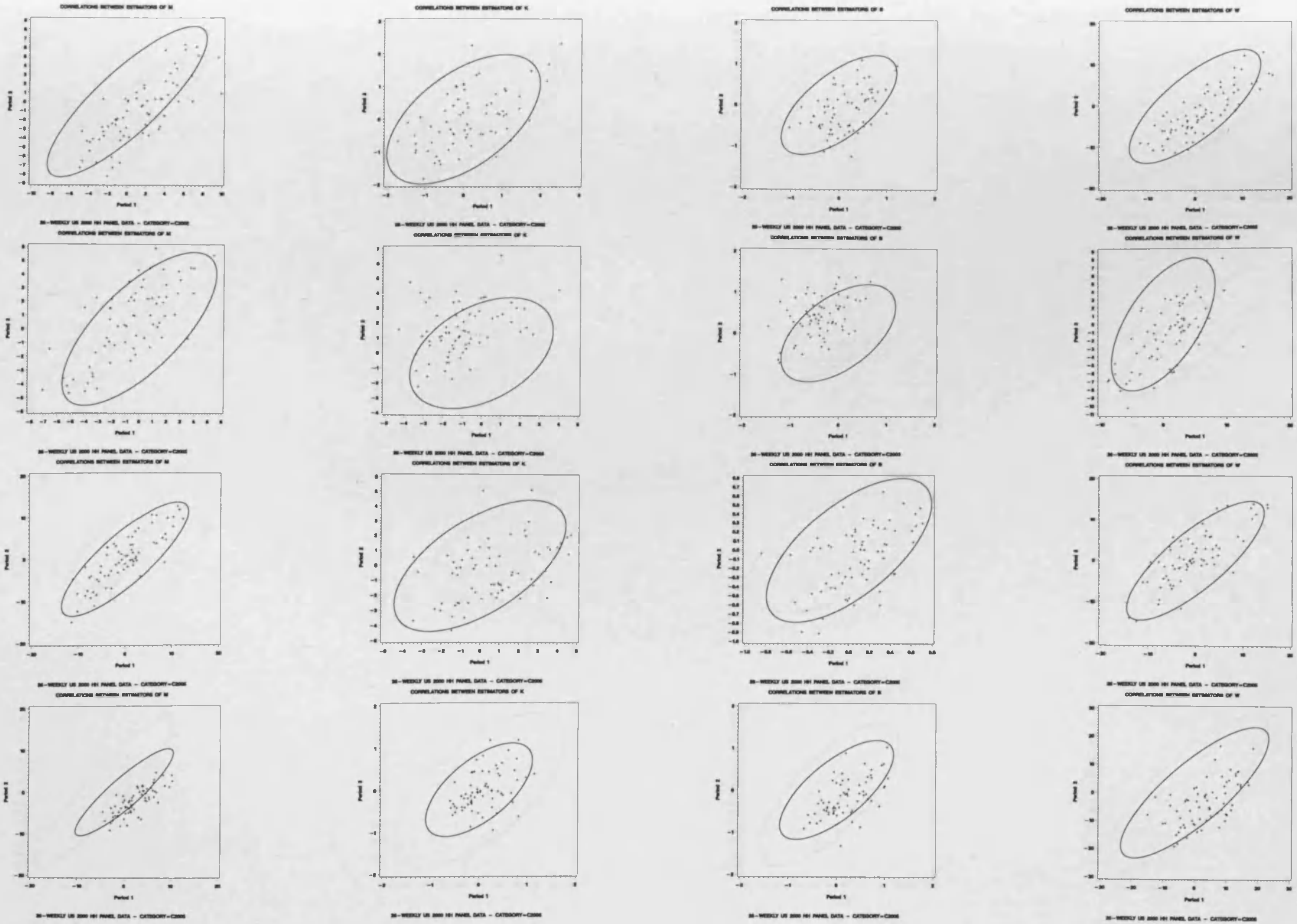
B.4 Correlations between market research measures computed in two 26-week time intervals



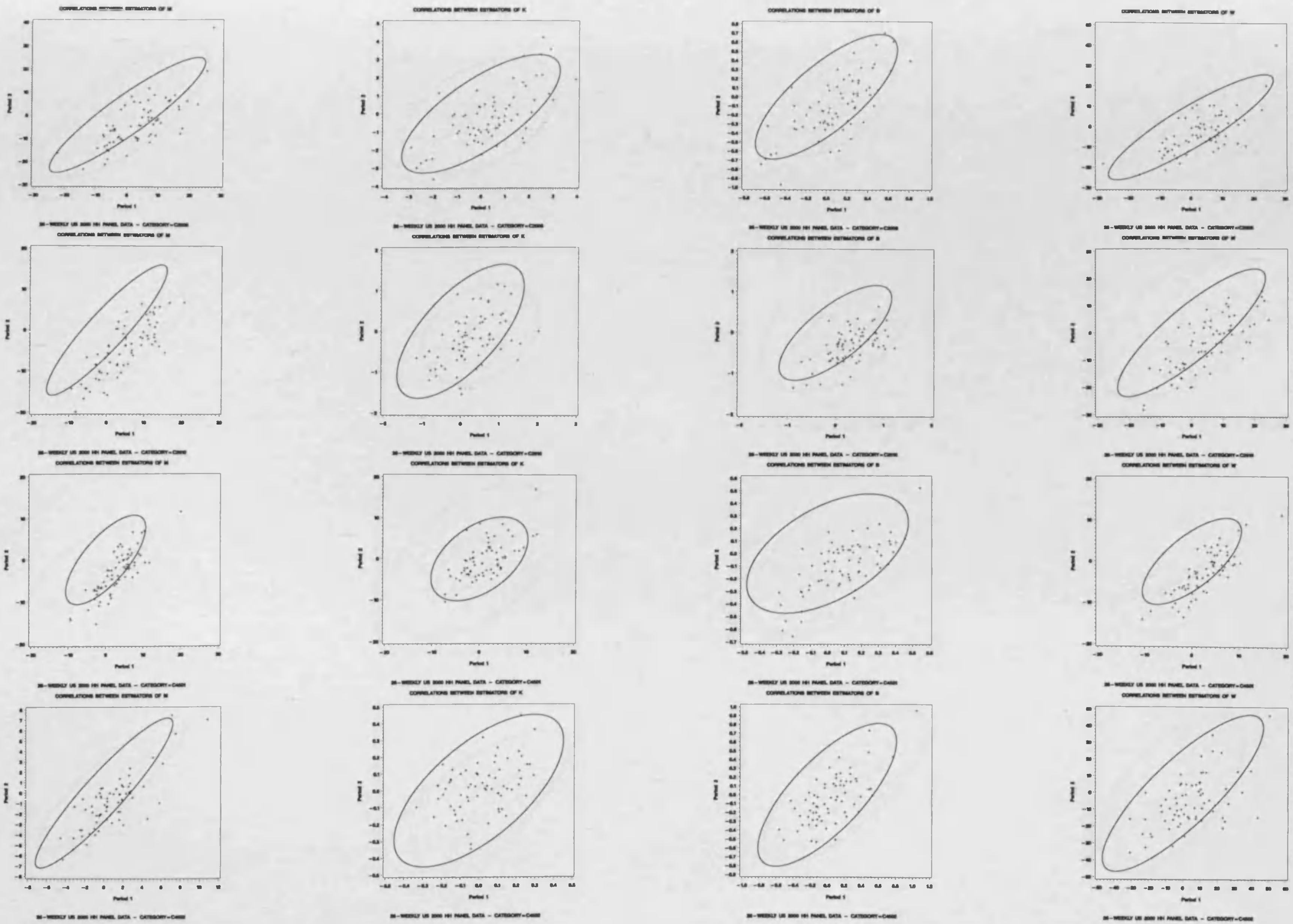
Correlations between market research measures computed in two 26-week time intervals



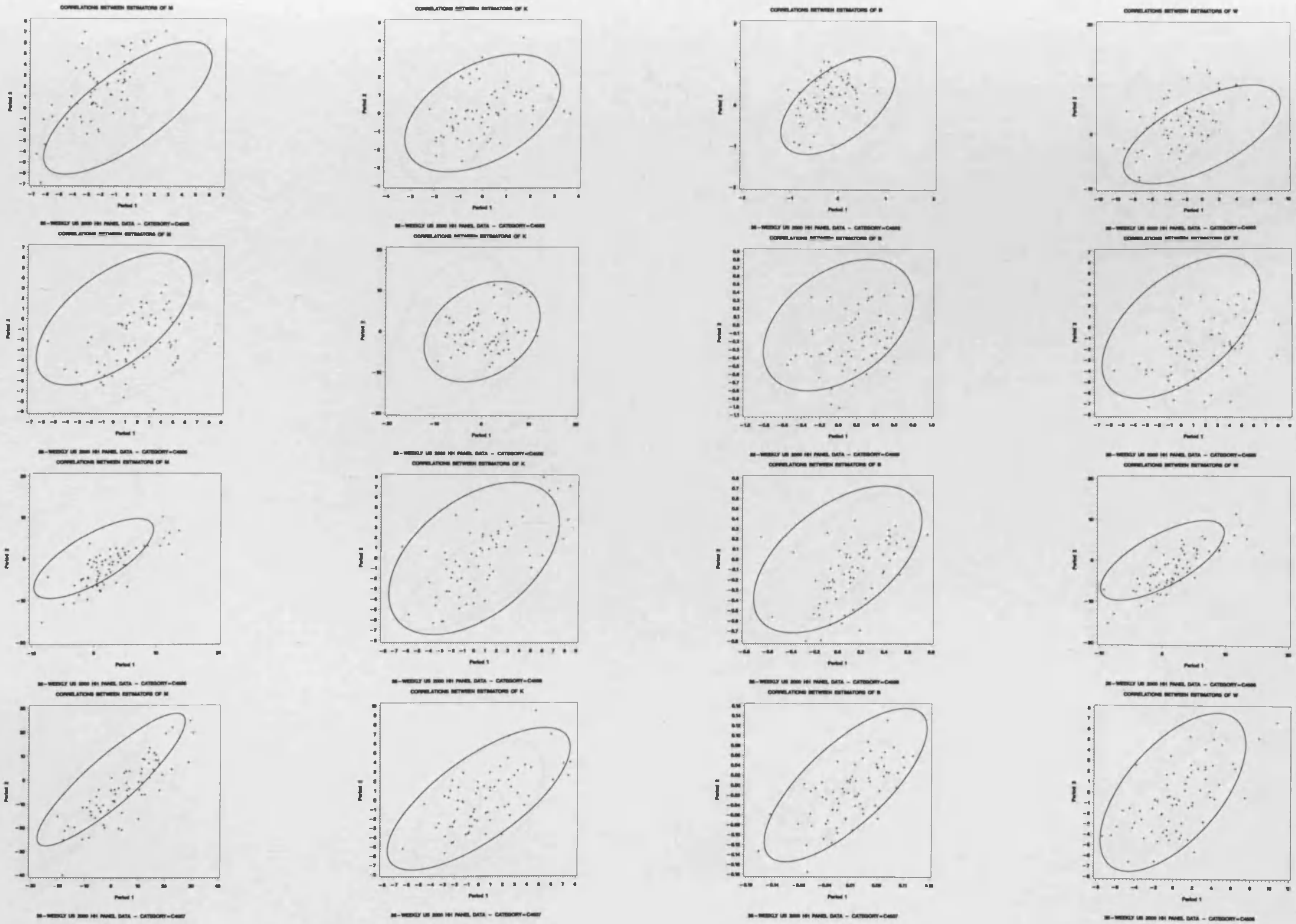
Correlations between market research measures computed in two 26-week time intervals



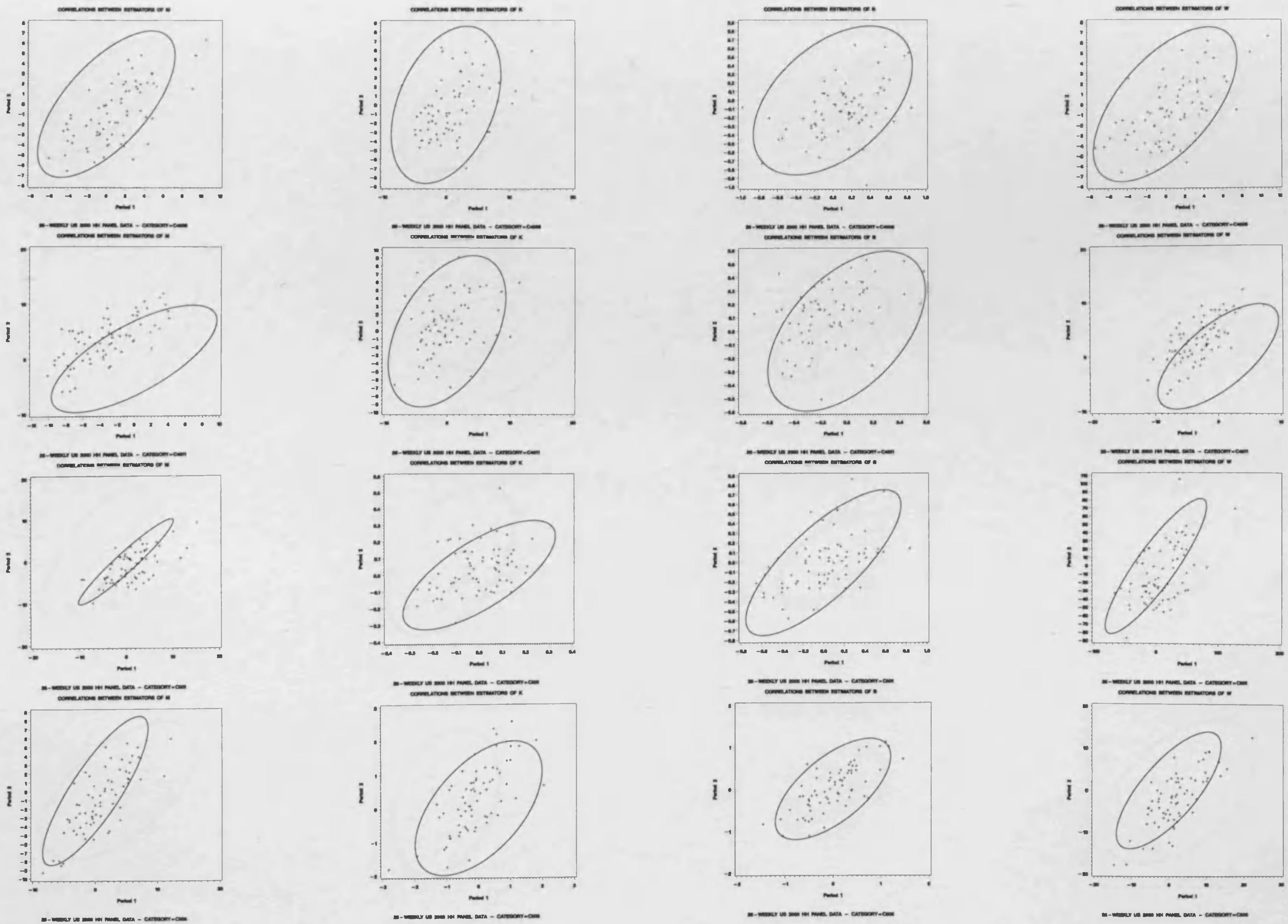
Correlations between market research measures computed in two 26-week time intervals



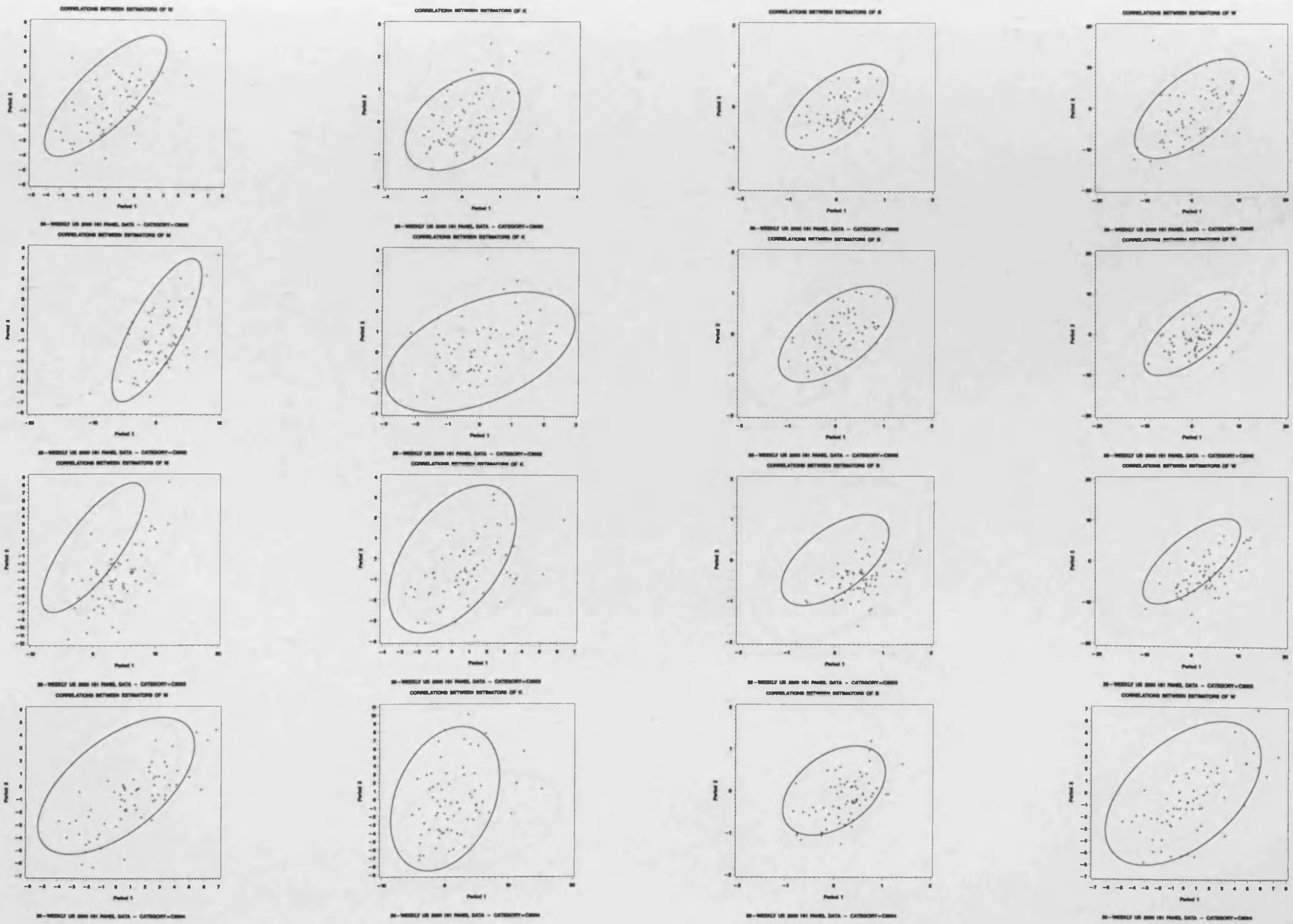
Correlations between market research measures computed in two 26-week time intervals



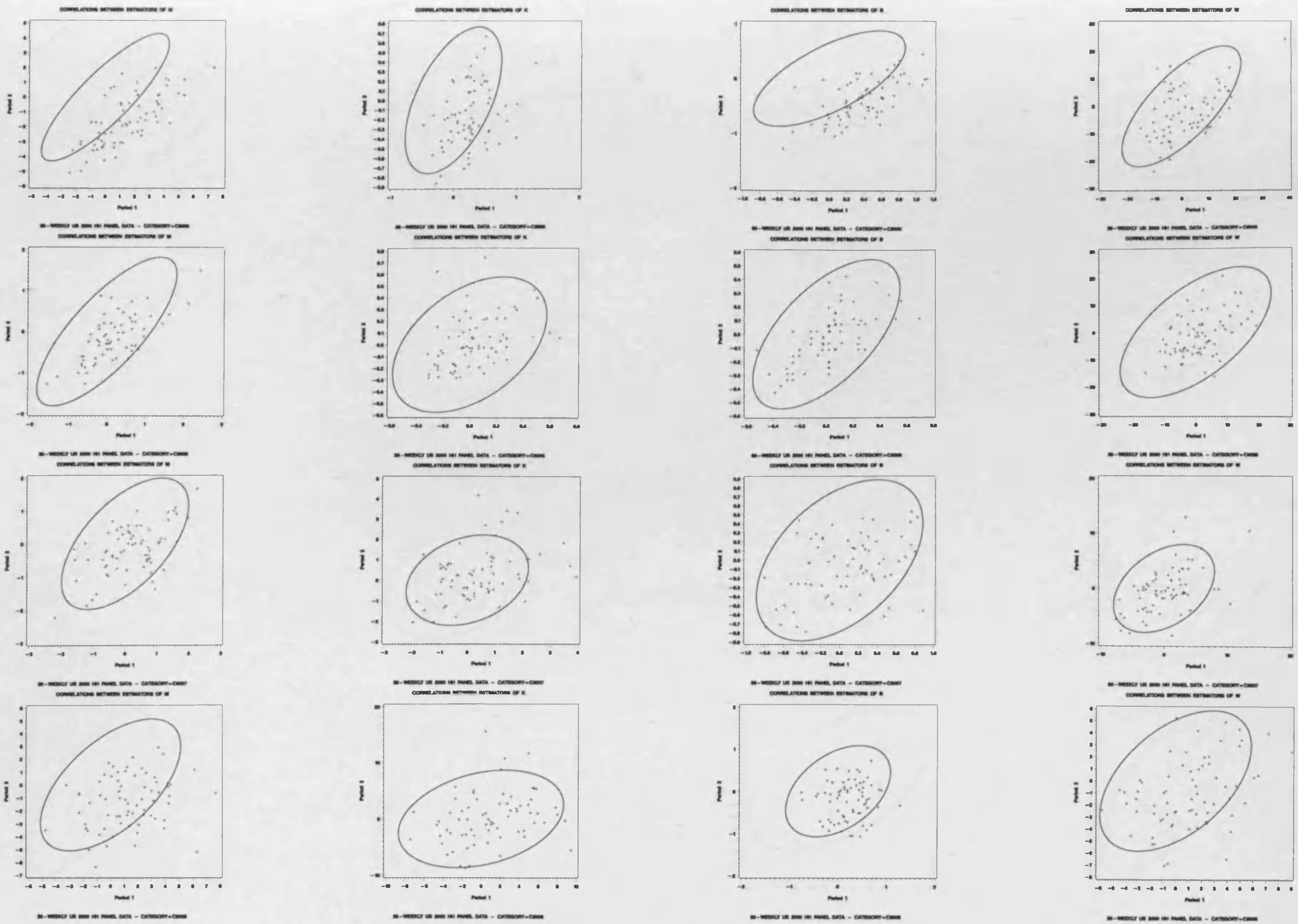
Correlations between market research measures computed in two 26-week time intervals



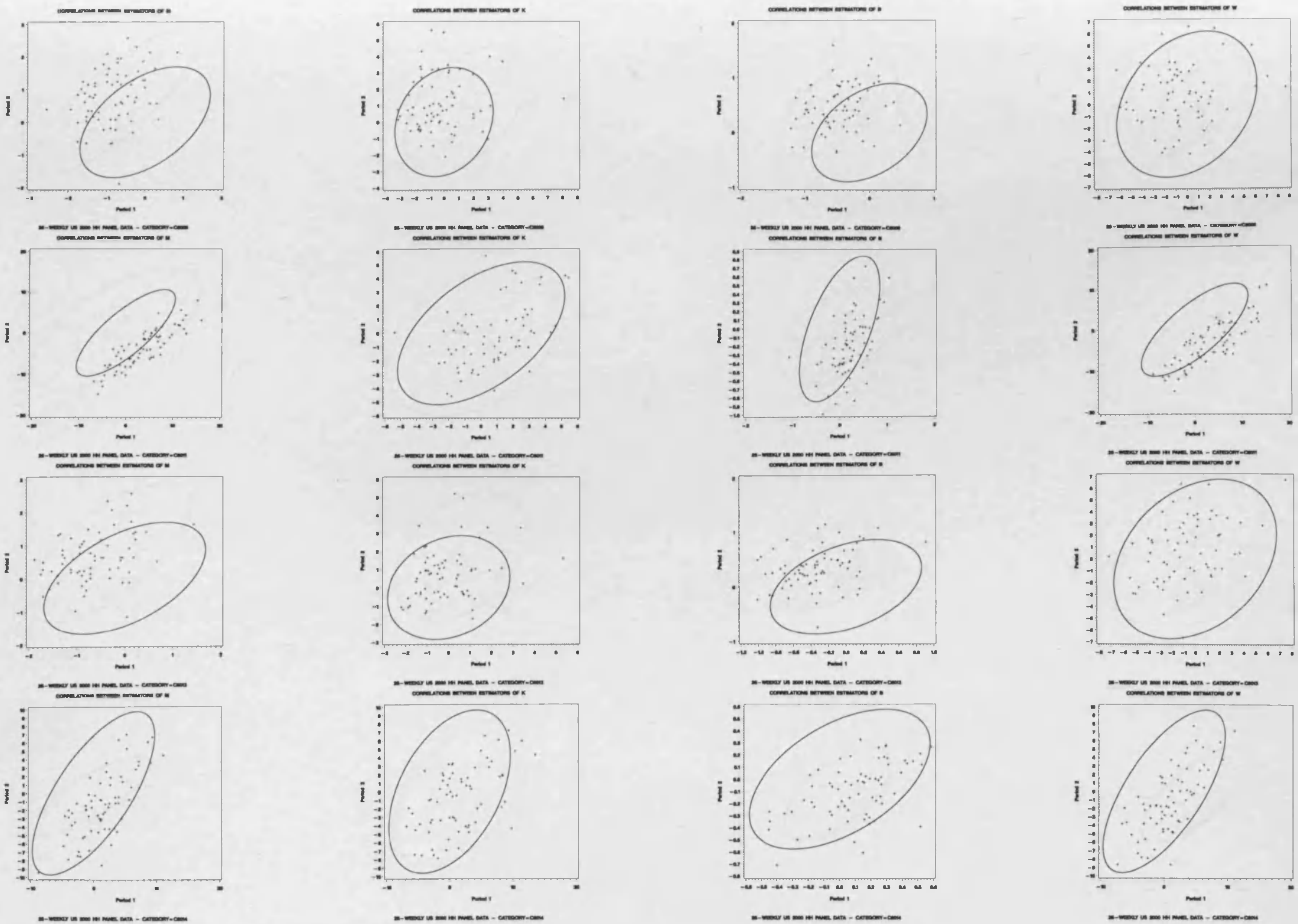
Correlations between market research measures computed in two 26-week time intervals



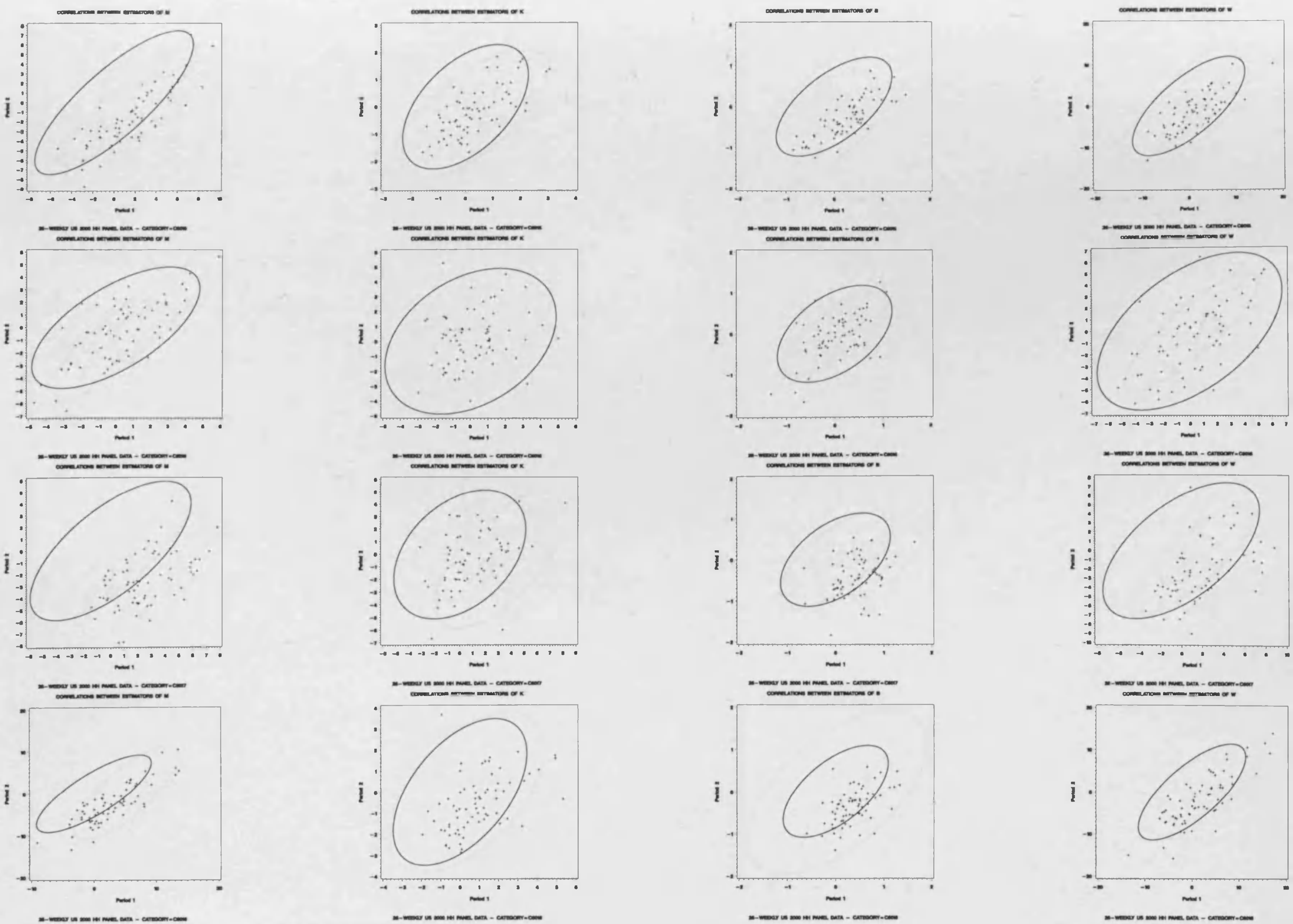
Correlations between market research measures computed in two 26-week time intervals



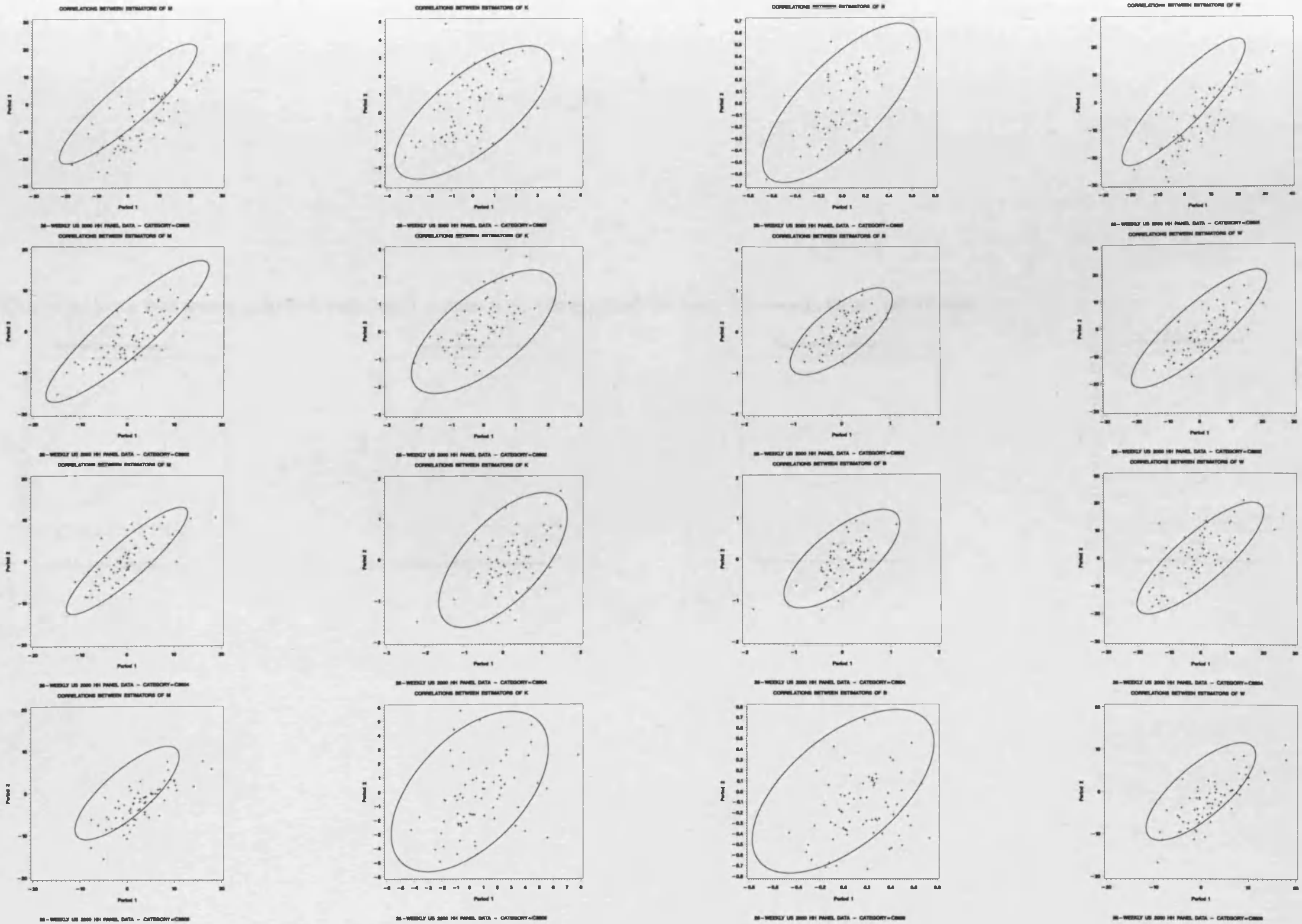
Correlations between market research measures computed in two 26-week time intervals



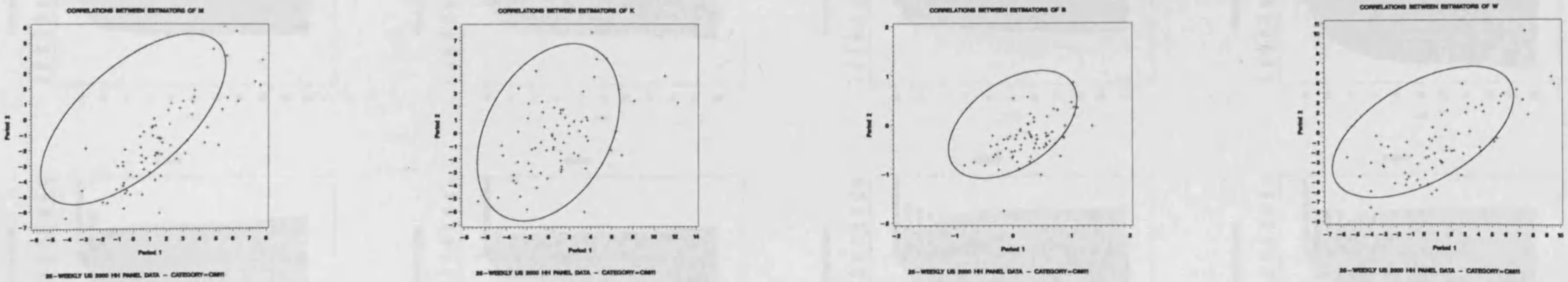
Correlations between market research measures computed in two 26-week time intervals



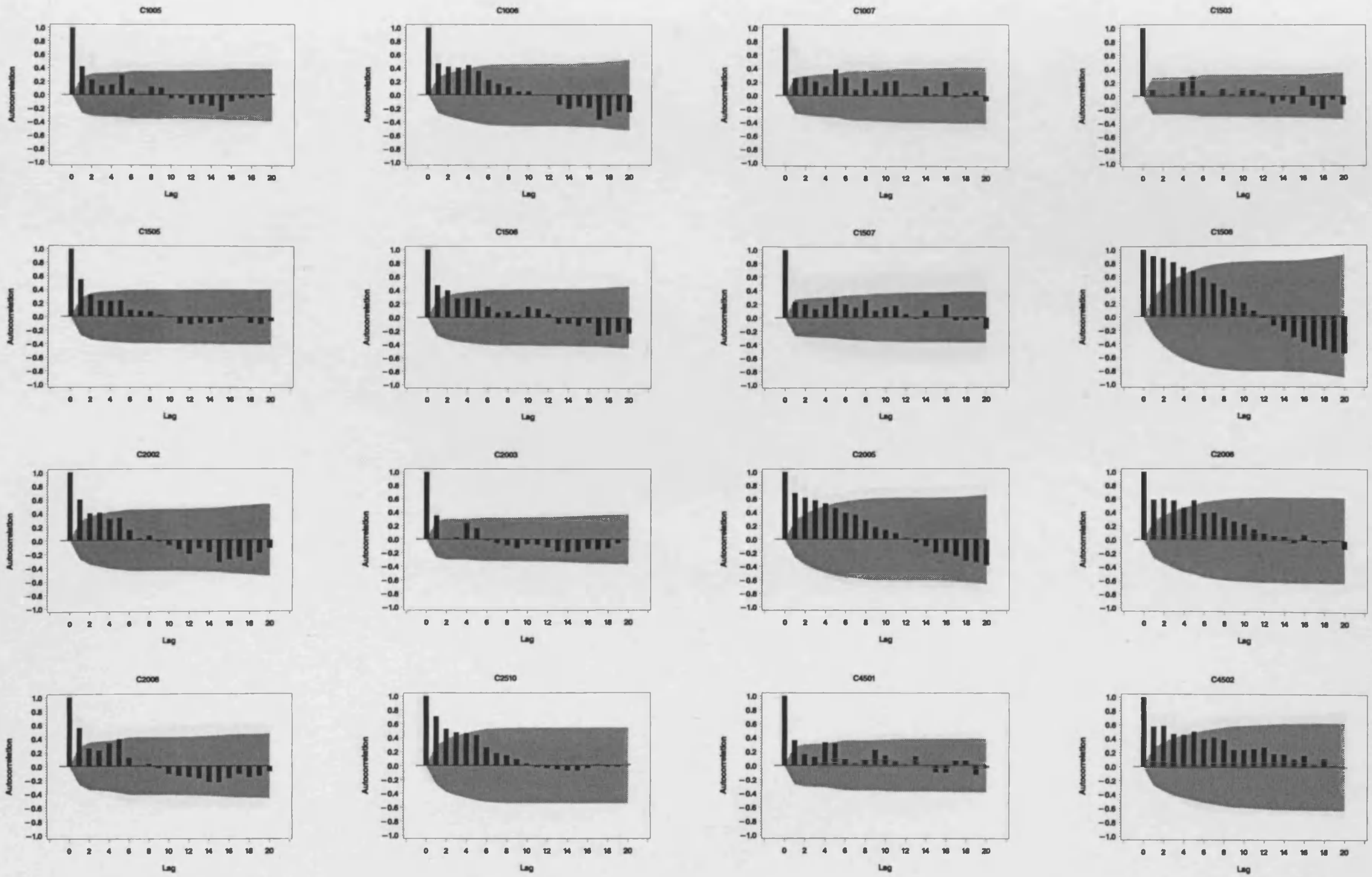
Correlations between market research measures computed in two 26-week time intervals



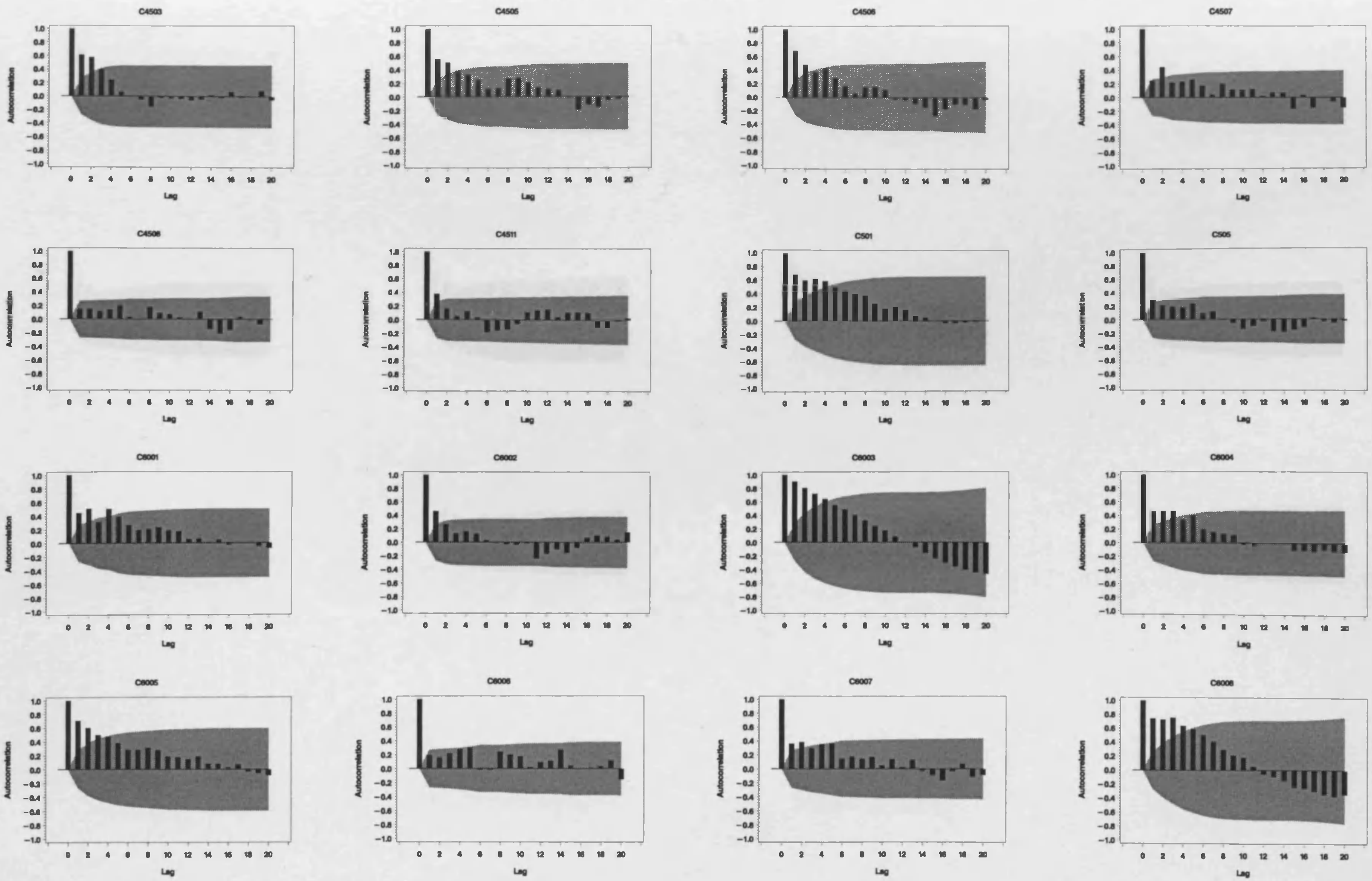
Correlations between market research measures computed in two 26-week time intervals



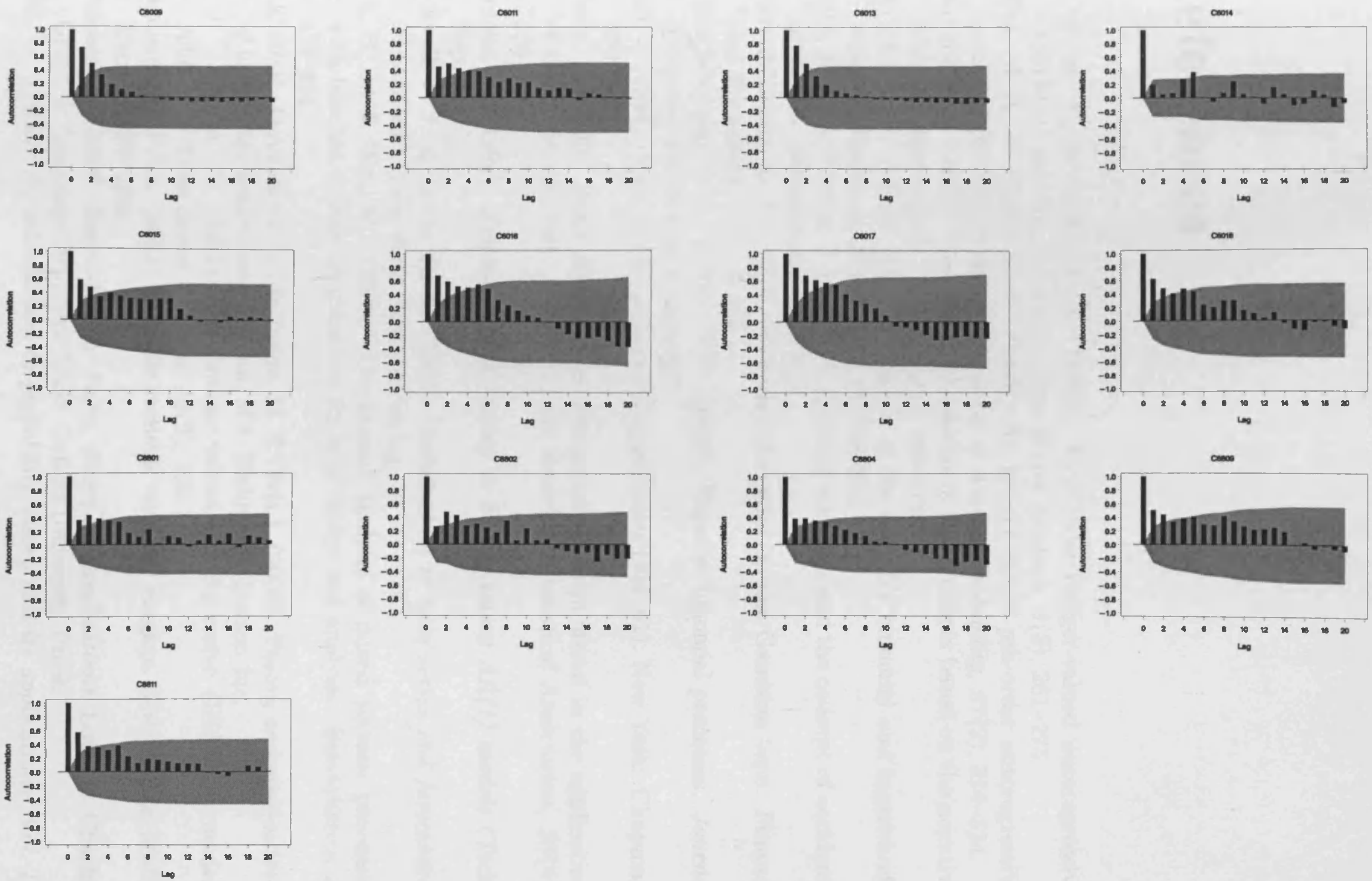
B.5 Autocorrelation function for the time series of the number of consumers in a category



Autocorrelation function for the time series of the number of consumers in a category



Autocorrelation function for the time series of the number of consumers in a category



References

Al-Osh, M. A., & Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, *8*(3), 261–275.

Al-Osh, M. A., & Alzaid, A. A. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, *27*(2), 314–324.

Anscombe, F. J. (1949). The statistical analysis of insect counts based on the negative binomial distribution. *Biometrics*, *5*, 165–173.

Anscombe, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, *37*, 358–382.

Arbous, A. G., & Kerrich, J. E. (1951). Accident statistics and the concept of accident-proneness. *Biometrics*, *7*, 340–429.

Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance and Stochastics*, *2*(1), 41–68.

Barndorff-Nielsen, O. E., & Yeo, G. F. (1969). Negative binomial processes. *Journal of Applied Probability*, *6*, 633–647.

Beran, J. (1994). *Statistics for long-memory processes* (Vol. 61). New York: Chapman and Hall.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, *33*(4), 526–536.

Brännäss, K. (1994). *Estimation and testing in integer-valued AR(1) models* (Tech. Rep.).

Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting* (Second ed.). New York: Springer-Verlag.

Cook, R. J., & Wei, W. (2003). Conditional analysis of mixed poisson processes with baseline counts: implications for trial design and analysis. *Biostatistics*, *4*, 479–494.

Doukhan, P., Oppenheim, G., & Taqqu, M. S. (Eds.). (2003). *Theory and applications of long-range dependence*. Boston, MA: Birkhäuser Boston Inc.

Du, J. G., & Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, *12*(2), 129–142.

Eggenberger, & Pólya. (1923). Uber die statistik verketter vorgage. *Zeit. Angew. Math. Mech.*, *1*, 279–289.

Ehrenberg, A. (1988). *Repeat-buying: Facts, theory and applications*. London: Charles Griffin & Company Ltd.; New York: Oxford University Press.

Feller, W. (1966). *An introduction to probability theory and its applications. Vol. II.*

New York: John Wiley & Sons Inc.

Fisher, R. A. (1941). The negative binomial distribution. *Annals of Eugenics, 11*, 182–187.

Franke, J., & Seligmann, T. (1993). Conditional maximum likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In *Developments in time series analysis* (pp. 310–330). London: Chapman & Hall.

Freeland, R. K., & McCabe, B. (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics & Probability Letters, 73*(2), 147–153.

Goodhardt, G., Ehrenberg, A., & Chatfield, C. (1984). The dirichlet: A comprehensive model of buying behaviour. *Journal of the Royal Statistical Society, A,147*, 621–655.

Gourieroux, C., & Jasiak, J. (2004). Heterogeneous INAR(1) model with application to car insurance. *Insurance: Mathematics & Economics, 34*(2), 177–192.

Grandell, J. (1997). *Mixed Poisson processes* (Vol. 77). London: Chapman & Hall.

Greenwood, M., & Yule, G. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society, 93*, 255–279.

Haldane, J. B. S. (1941). The fitting of binomial distributions. *Annals of Eugenics, 11*, 179–181.

Kendall, D. G. (1948). On the generalized "birth-and-death" process. *Annals of Mathematical Statistics, 19*, 1–15.

Latour, A. (1997). The multivariate GINAR(p) process. *Advances in Applied Probability, 29*(1), 228–248.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis, 19*(4), 439–455.

Lehmann, E. L., & Scheffé, H. (1950). Completeness, similar regions, and unbiased estimation. I. *Sankhyā, 10*, 305–340.

Lundberg, O. (1964). *On Random Processes and their Application to Sickness and Accident Statistics.* Uppsala: Almquist and Wiksells.

McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability, 18*(3), 679–705.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability, 20*(4), 822–835.

Neyman, J. (1949). Contribution to the theory of the χ^2 test. In *Proceedings of the berkeley symposium on mathematical statistics and probability, 1945, 1946* (pp. 239–273). Berkeley and Los Angeles: University of California Press.

Priestley, M. B. (1981). *Spectral analysis and time series. Vol. 1.* London: Academic Press Inc. [Harcourt Brace Jovanovich Publishers]. (Univariate series, Probability and Mathematical Statistics)

Savani, V., & Zhigljavsky, A. (2006). *Efficient parameter estimation for independent and INAR(1) negative binomial samples.* (Accepted by Metrika)

Schwartz, L. (1967). *Analyse mathématique. I.* Paris: Hermann.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* New York: John Wiley & Sons Inc.

Silva, M. E., & Oliveira, V. L. (2005). Difference equations for the higher order moments and cumulants of the INAR(p) model. *Journal of Time Series Analysis, 26*(1), 17–36.

Sprott, D. A. (1983). Estimating the parameters of a convolution by maximum likelihood. *Journal of the American Statistical Association, 78*(382), 457–460.

Steutel, F. W., & van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability, 7*(5), 893–899.

van der Vaart, A. W. (1998). *Asymptotic statistics* (Vol. 3). Cambridge: Cambridge University Press.

Wang, Y. (1996). Estimation problems for the two-parameter negative binomial distribution. *Statistics & Probability Letters, 26*(2), 113–114.

Willson, L. J., Folks, J. L., & Young, J. H. (1986). Complete sufficiency and maximum likelihood estimation for the two-parameter negative binomial distribution. *Metrika, 33*(6), 349–362.

Wimmer, G., & Altmann, G. (1990). *Thesaurus of univariate discrete probability distributions.* Stamm.

Yule, G. U. (1910). On the distribution of deaths with age when the causes of death act cumulatively, and similar frequency distributions. *Journal of the Royal Statistical Society, 73*, 26–38.