1

Statistical Analysis of Count Time Series Models: A
GLM Perspective

Konstantinos Fokianos

CONTENTS

1.1 Introduction ...................................................................................... 3
1.2 Modeling .......................................................................................... 4
  1.2.1 The Linear Model........................................................................ 6
  1.2.2 Log-Linear Models for Count Time Series........................................ 7
  1.2.3 Nonlinear Models for Count Time Series........................................ 8
1.3 Inference .......................................................................................... 9
  1.3.1 Transactions Data...................................................................... 11
1.4 Diagnostics ...................................................................................... 12
  1.4.1 Residuals ................................................................................ 12
  1.4.2 Goodness-of-Fit Test................................................................... 13
1.5 Prediction ........................................................................................ 16
  1.5.1 Assessment of Probabilistic Calibration............................................ 16
  1.5.2 Assessment of Marginal Calibration................................................ 18
  1.5.3 Assessment of Sharpness............................................................. 18
1.6 Other Topics ..................................................................................... 20
  1.6.1 Testing for Linearity................................................................... 20
  1.6.2 Intervention Analysis................................................................. 21
  1.6.3 Robust Estimation..................................................................... 22
  1.6.4 Multivariate Count Time Series Models....................................... 22
  1.6.5 Parameter-Driven Models............................................................ 22
1.7 Other Extensions ............................................................................... 23
References ............................................................................................. 24

1.1 Introduction

In this chapter we discuss statistical models and methods that have been developed over the past few years for the analysis of count time series which occur in diverse application areas, like finance, biostatistics, environmetrics, and others. The analysis of count time series data has attracted considerable attention, see Kedem and Fokianos (2002, Secs. 4 & 5) for several references and Fokianos (2012) for a more recent review of this research area. In what follows, I will present the generalized linear methodology (GLM) advanced
Handbook of Discrete-Valued Time Series

by Nelder and Wedderburn (1972) and McCullagh and Nelder (1989). This framework naturally generalizes the traditional ARMA methodology and includes several complicated data generating processes besides count data such as binary and categorical data. In addition, the fitting of such models can be carried out by likelihood methods; therefore testing, diagnostics, and all type of likelihood arguments are available to the data analyst. The Bayesian point of view towards the analysis of such models will be discussed later in this handbook; see chapters by Gamerman et al. (2015; Chapter 8 in this volume) and Soyer et al. (2015; Chapter 11 in this volume).

However, a theoretical study of GLM-based models poses challenging problems some of which are still open for further research; see the review article by Tjøstheim (2012) and Section 1.4. The main problem posed by count time series in theoretical analysis is that the observed response is discrete valued and therefore it might not be strong mixing (see Andrews, 1984). The theoretical analysis of count time series has been based so far on a perturbation argument (see Fokianos et al. 2009), weak dependence (Doukhan et al. 2012, 2013), and Markov chain theory without irreducibility assumptions (see Woodard et al. 2011, Douc et al. 2013). On the other hand, Neumann (2011) has proved a number of mixing results by employing coupling techniques. Since most of the theoretical results will be presented in other chapters of this handbook, we will confine our attention to data analysis tools which are useful for estimation, testing goodness of fit, and prediction.

In Section 1.2, we present several statistical models that mimic the usual AR models for count time series analysis. In Section 1.3, we discuss quasi-maximum likelihood estimation (QMLE) of the unknown parameters. In Sections 1.4 and 1.5, we develop goodness-of-fit methodology and prediction for count time series, respectively. Finally, we conclude the chapter with a discussion of possible research topics. The present contribution can be thought of as a companion paper to Fokianos (2012), where detailed properties for GLM-based count time series models were discussed.

As a final introductory remark, we note that there are other alternative classes of regression models for count time series; the most prominent being the integer autoregressive models. These models are based on the notion of thinning. Accordingly, integer autoregressive models imitate the structure of the usual autoregressive processes, in the sense that thinning is applied instead of scalar multiplication. These models will be discussed in other chapters of this handbook, and therefore, they will not be included in this presentation.

1.2 Modeling

To motivate our discussion, consider the transactions data used by Fokianos et al. (2009). This data set consists of the number of transactions per minute for the stock Ericsson B over a single day. To economize on space, we do not show plots of the data or their autocorrelation since these can be easily obtained from the reference cited earlier. However, we report that the mean of the data is 9.909 and the variance is 32.837; this is a clear case of overdispersion where the mean is less than the variance. The minimum value of this series is 0 and the maximum is 37. The sample autocorrelation function is positive; at lag 1 (respectively, lag 2) it is equal to 0.405 (respectively, 0.340) and decays slowly toward zero; for related plots see Fokianos et al. (2009). These data are counts and therefore a reasonable model for their analysis can be based on a counting process. This is the point of view taken by Rydberg and Shephard (2000) who argue that the number of transactions influence the
price of an asset. Hence they study models of the form (1.5)—and more generally of the form (1.7)—to analyze transactions within a small time interval. These models are natural analogues of real-valued time series models and their development can be based on their similarities with other econometric models; see Bollerslev (1986) for the GARCH family of models. We will be more specific in what follows.

Let \( \{Y_t, t \geq 1\} \) denote a count time series, and let \( \{\lambda_t, t \geq 1\} \) be a latent mean process. Let \( \mathcal{F}_{t-1}^{Y, \lambda} = \sigma(Y_s, s \leq t, \lambda_0) \), where \( \lambda_0 \) denotes a starting value, denote the past of the process up to and including time \( t \). We will assume that the observed count time series conditionally on its past is Poisson distributed with mean \( Z_t \lambda_t \), where \( \{Z_t\} \) denotes an i.i.d. sequence of positive random variables with mean 1, which is independent of \( Y_s \), for all \( s < t \). That is

\[
Y_t \mid \mathcal{F}_{t-1}^{Y, \lambda} \sim \text{Po}(Z_t \lambda_t). \tag{1.1}
\]

In the context of transactions data, (1.1) implies that the total number of transactions at time \( t \) is a realization of a Poisson random variable with a time varying stochastic intensity. The random variables \( \{Z_t\} \) correspond to random fluctuations (different trading strategies, delays, and other). Model (1.1) is called a mixed Poisson model (Mikosch, 2009) and its usefulness is not confined to the analysis of econometric data, but can also be employed for modeling medical, environmental, and other data. In fact, (1.1) can be viewed as a stochastic intensity model, similar to the stochastic volatility model.

To study (1.1) in detail consider the following representation for the process \( \{Y_t, t \geq 1\} \),

\[
Y_t = N_t(0, Z_t \lambda_t), \quad t \geq 1. \tag{1.2}
\]

In (1.2), \( \{N_t, t \geq 1\} \) is an i.i.d. sequence from a standard homogeneous Poisson process (that is, a Poisson process with rate equal to 1) and \( \{Z_t\} \) denotes an i.i.d. sequence of positive random variables with mean 1, which is independent of \( N_t \), for all \( t \). In addition, we assume that \( \{Z_t\} \) is independent of \( Y_s \), for all \( s < t \). This class of models is large and includes two important distributions routinely employed for the analysis of count time series, viz. the Poisson distribution

\[
P[Y_t = y \mid \mathcal{F}_{t-1}^{Y, \lambda}] = \frac{\exp(-\lambda_t) \lambda_t^y}{y!}, \quad y = 0, 1, 2, \ldots \tag{1.3}
\]

and the negative binomial distribution

\[
P[Y_t = y \mid \mathcal{F}_{t-1}^{Y, \lambda}] = \frac{\Gamma(\nu + y)}{\Gamma(y + 1)\Gamma(\nu)} \left( \frac{\nu}{\nu + \lambda_t} \right)^\nu \left( \frac{\lambda_t}{\nu + \lambda_t} \right)^y, \quad y = 0, 1, 2, \ldots \tag{1.4}
\]

where the dispersion parameter \( \nu \) is positive. It is obvious that (1.3) is a special case of (1.2) when \( \{Z_t\} \) is a sequence of degenerate random variables with mean 1, and (1.4) is a special case of (1.2) when \( \{Z_t\} \) are i.i.d. Gamma variables with mean 1 and variance \( 1/\nu \). Regardless of the choice of \( Z \)'s in (1.2), the conditional mean of \( Y_t \) is always equal to \( \lambda_t \). Furthermore, the conditional variance of \( Y_t \) is given by \( \lambda_t + \sigma_Z^2 \lambda_t^2 \), with \( \sigma_Z^2 = \text{Var}(Z_t) \). The conditional variance of the Poisson distribution is equal to \( \lambda_t \), whereas the conditional variance of the negative binomial variable is equal to \( \lambda_t + \lambda_t^2/\nu \). Hence, although a Poisson-based conditional model is capable of accommodating overdispersion (i.e., the mean of the data is less than the variance), we anticipate that the negative binomial distribution will be a more suitable model for accommodating this fact.
In what follows, we review some standard models that have been suggested over the past few decades for modeling count time series data. Since most of them have been already discussed in the recent survey by Fokianos (2012) and Tjøstheim (2012), we only give their definitions and some motivation. The interested reader is referred to the earlier two articles for further details.

1.2.1 The Linear Model

The linear model for the analysis of count time series is specified by

$$\lambda_t = d + a_1 \lambda_{t-1} + b_1 Y_{t-1}, \quad t \geq 1,$$

where the parameters $d, a_1, b_1$ are assumed to be nonnegative and to satisfy $0 < a_1 + b_1 < 1$. If $a_1 = b_1 = 0$, then we obtain a model for i.i.d data with mean $d$. When $a_1 = 0$, (1.5) reduces to an ordinary AR(1) model; in this case the mean of the process is influenced by its past. In the context of transactions data, (1.5), with $a_1 = 0$, implies that the mean number of transactions at time $t$ depends on the number of transactions at time $t-1$. This is a sensible way to analyze the data and generalizes naturally the framework of AR models. When $a_1 > 0$, (1.5) can be viewed as a parsimonious way to analyze count time series whose sample autocorrelation function decays slowly towards zero; this is the case of the transactions data. In fact, when both $a_1$ and $b_1$ are positive, (1.5) has the same second-order properties as those of an ARMA(1,1) model. For more details about (1.5), see Rydberg and Shephard (2000), Streett (2000), Heinen (2003), Ferland et al. (2006), and more recently Fokianos et al. (2009) for the case of Poisson distributed data. The case of negative binomial distributed data has been studied by Zhu (2011), Davis and Liu (2015), and Christou and Fokianos (2014).

The mean parametrization based on (1.4) yields a stationary region which is independent of the additional dispersion parameter $\nu$. This implies that fitting count time series models, like (1.5), to data is implemented by constrained optimization of the log-likelihood function with respect to the regression parameters ($a_1, b_1$ in the case of (1.5)). The parameter $\nu$ (equivalently $\sigma^2_Z$ in the general framework) is estimated separately. We emphasize this point further by comparing the proposed parametrization (1.4) with that of Zhu (2011) given by

$$P[Y_t = y | Y_{t-1}, \lambda_t] = \frac{\binom{y - r - 1}{r - 1} p_t^r (1 - p_t)^{y-r}}{\binom{y - r - 1}{r - 1} p_t^r (1 - p_t)^{y-r}}, \quad y = 0, 1, 2, \ldots$$

In (1.6), $r$ is a positive integer and $p_t$ denotes the conditional probability of success. Zhu (2011) models the odds ratio, denoted by $\lambda_t = p_t/(1 - p_t)$, as a function of past values of itself and past values of the response. Stationarity conditions for the model studied by Zhu (2011) depend upon the parameter $r$ in (1.6). Therefore, it is challenging to maximize the log-likelihood function since such an optimization problem imposes restrictions on the regression parameters and $r$. Furthermore, it is well known that the estimation of the odds ratio might be problematic when probabilities are either very small or very large. For related work on negative binomial–based models, see also Davis and Liu (2015).
Ferland et al. (2006) enlarge the class of models (1.5) to the general model of order \((p, q)\):

\[
Y_t | \mathcal{F}_{t-1}^Y \sim \text{Poisson}(\lambda_t), \quad \lambda_t = d + \sum_{i=1}^{p} a_i \lambda_{t-i} + \sum_{j=1}^{q} b_j Y_{t-j}, \quad t \geq \max(p, q),
\]

(1.7)

and show that it is second-order stationary provided that \(0 < \sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1\) under the Poisson assumption. These conditions are true for the mixed Poisson process (1.2) provided that \(E(Z_t) = 1\) (see Christou and Fokianos 2014).

The properties of (1.5) have been studied in detail by Fokianos (2012). Here, we only mention that by repeated substitution in (1.5), the mean process \(\lambda_t\) is given by

\[
\lambda_t = d + a_1 \lambda_{t-1} + b_1 \sum_{i=0}^{t-1} a_i Y_{t-i-1}.
\]

In other words, the hidden process \(\{\lambda_t\}\) is determined by past functions of lagged responses and the initial value \(\lambda_0\). Therefore (1.5) belongs to the class of observation-driven models in the sense of Cox (1981).

### 1.2.2 Log-Linear Models for Count Time Series

As empirical experience has shown, (1.5) and its variants can successfully accommodate dependent count data, especially when there exists positive autocorrelation and there are no covariates. However, in applications, we observe time series data that might be negatively correlated, and possibly with covariates. In this case, the logarithmic function is the most popular link function for modeling count data. In fact, this choice corresponds to the canonical link of generalized linear models. Log-linear models for dependent count data have been considered by Zeger and Qaqish (1988), Li (1994), MacDonald and Zucchini (1997), Brumback et al. (2000), Kedem and Fokianos (2002), Benjamin et al. (2003), Davis et al. (2003), Fokianos and Kedem (2004), Jung et al. (2006), Creal et al. (2008), Fokianos and Tjøstheim (2011), and Douc et al. (2013), among others.

Recall that \(\{Y_t\}\) denotes a count time series and following the notation introduced in (1.2), let \(\nu_t \equiv \log \lambda_t\). A log-linear model with feedback for the analysis of count time series (Fokianos and Tjøstheim 2011) is defined as

\[
\nu_t = d + a_1 \nu_{t-1} + b_1 \log(Y_{t-1} + 1), \quad t \geq 1.
\]

(1.8)

In general, the parameters \(d, a_1, b_1\) can be positive or negative but they need to satisfy certain conditions so that we obtain a stationary time series. Note that the lagged observations of the response \(Y_t\) are fed into the autoregressive equation for \(\nu_t\) via the term \(\log(Y_{t-1} + 1)\). This is a one-to-one transformation of \(Y_{t-1}\) which avoids taking logarithm of zero data values. Moreover, both \(\lambda_t\) and \(Y_t\) are transformed into the same scale. Covariates can be easily accommodated by model (1.8). When \(a_1 = 0\), we obtain an AR(1) type model in terms of \(\log(Y_{t-1} + 1)\). In addition, the log-intensity process of (1.8) can be rewritten as

\[
\nu_t = d + a_1 \nu_{t-1} + b_1 \sum_{i=0}^{t-1} a_i \log(1 + Y_{t-i-1} + 1),
\]
after repeated substitution. Hence, we obtain again that the hidden process \( \{\nu_t\} \) is determined by past functions of lagged responses. Equivalently, the log-linear model (1.8) belongs to the class of observation-driven models and possesses similar properties to the linear model (1.5). For more details, see Fokianos (2012).

### 1.2.3 Nonlinear Models for Count Time Series

A large class of models for the analysis of count time series is given by the following nonlinear mean specification

\[
\lambda_t = f(\lambda_{t-1}, Y_{t-1}), \quad t \geq 1,
\]

where \( f(\cdot) \) is known up to an unknown finite dimensional parameter vector such that \( f : (0, \infty) \times \mathbb{N} \to (0, \infty) \). The function \( f(\cdot) \) is assumed to satisfy the contraction condition

\[
|f(\lambda, y) - f(\lambda', y')| \leq \alpha_1 |\lambda - \lambda'| + \gamma_1 |y - y'|,
\]

for \((\lambda, y)\) and \((\lambda', y')\) in \((0, \infty) \times \mathbb{N}\), where \(\alpha_1 + \gamma_1 < 1\); see Fokianos et al. (2009), Neumann (2011), Doukhan et al. (2012, 2013), and Fokianos and Tjøstheim (2012).

An interesting example of a nonlinear regression model for count time series is given by

\[
f(\lambda, y) = d + (a_1 + c_1 \exp(-\gamma y^2))\lambda + b_1 y,
\]

where \(d, a_1, c_1, b_1, \gamma\) are positive parameters, which is similar to the exponential autoregressive model (Haggan and Ozaki 1981). In Fokianos et al. (2009), (1.11) was studied for the case \(d = 0\). The parameter \(\gamma\) introduces a perturbation of the linear model (1.5), in the sense that when \(\gamma\) tends either to 0 or infinity, then (1.11) approaches two distinct linear models. It turns out that the regression coefficients of model (1.11) must satisfy the condition \(a_1 + b_1 + c_1 < 1\) to guarantee ergodicity and stationarity of the joint process \((Y_t, \lambda_t)\); see Doukhan et al. (2012) for more precise conditions. Model (1.11) shows that there is a smooth transition between two linear models for count time series in terms of the unobserved process. This transition might be difficult to estimate because of the nonlinear parameter \(\gamma\) and the fact that \(\lambda_t\) is not directly observed. An alternative method for introducing the smooth transition is by employing the observed data. In other words, instead of (1.11), we can consider

\[
f(\lambda, y) = d + a_1 \lambda + \left(b_1 + c_1 \exp\left(-\gamma y^2\right)\right) y.
\]

The previous model is interpreted in a similar manner to (1.11) and must satisfy the condition \(a_1 + b_1 + c_1 < 1\) to obtain ergodicity and stationarity of the joint process \((Y_t, \lambda_t)\).

Another nonlinear model studied by Fokianos and Tjøstheim (2012) is given by

\[
f(\lambda, y) = \frac{d}{(1 + \lambda_{t-1})^\gamma} + a_1 \lambda_{t-1} + b_1 Y_{t-1},
\]

where all regression parameters are assumed positive. Here, the inclusion of \(\gamma\) introduces a nonlinear perturbation, in the sense that small values of \(\gamma\) cause (1.12) to approach (1.5).
Moderate values of $\gamma$ introduce a stronger perturbation. Models of the form (1.9) have also been studied in the context of the negative binomial distribution by Christou and Fokianos (2014). The condition $\max\{a_1, d\gamma - a_1\} + b_1 < 1$ guarantees ergodicity and stationarity of the joint process $(Y_t, \lambda_t)$. Following the arguments made earlier in connection with model (1.11), we can alternatively consider the following modification of (1.12):

$$f(\lambda, y) = \frac{d}{(1 + Y_{t-1})^\gamma} + a_1 \lambda_{t-1} + b_1 Y_{t-1},$$

with the required stationarity condition $\max\{b_1, d\gamma - b_1\} + a_1 < 1$.

An obvious generalization of model (1.9) is given by the following specification of the mean process (see Franke 2010 and Liu 2012):

$$\lambda_t = f(\lambda_{t-1}, \ldots, \lambda_{t-p}, Y_{t-1}, \ldots, Y_{t-q}), \quad (1.13)$$

where $f(.)$ is a function such that $f : (0, \infty)^p \times \mathbb{N}^q \rightarrow (0, \infty)$. It should be clear that models (1.11) and (1.12) can be extended according to (1.13). Such examples are provided by the class of smooth transition autoregressive models of which the exponential autoregressive model is a special case (cf. Teräsvirta 1994, Teräsvirta et al. 2010). Further examples of nonlinear time series models can be found in Tong (1990) and Fan and Yao (2003). These models have not been considered earlier in the literature in the context of generalized linear models for count time series, and they provide a flexible framework for studying dependent count data. For instance, nonlinear models can be quite useful when testing departures from linearity; this topic is partially addressed in Section 1.6.1. A more general approach would have been to estimate the function $f$ of (1.13) by employing nonparametric methods. However, such an approach is missing from the literature.

### 1.3 Inference

Maximum likelihood inference for the Poisson model (1.3) and the negative binomial model (1.4) has been developed by Fokianos et al. (2009), Fokianos and Tjøstheim (2012), and Christou and Fokianos (2014). They develop estimation procedures based on the Poisson likelihood function, which for the Poisson model (1.3) is obviously the true likelihood. However, for the negative binomial model (1.4), and more generally for mixed Poisson models, this method resembles the QMLE method for GARCH models which employs the Gaussian likelihood function irrespective of the assumed error distribution. The QMLE method, in the context of GARCH models, has been studied in detail by Berkes et al. (2003), Francq and Zakoian (2004), Mikosch and Straumann (2006), Bardet and Wintenberger (2010), and Meitz and Saikkonen (2011), among others. This approach yields consistent estimators of regression parameters under a correct mean specification and it bypasses complicated likelihood functions (Godambe and Heyde 1987, Zeger and Qaqish 1988, Heyde 1997).

In the case of mixed Poisson models (1.2), it is impossible, in general, to have a readily available likelihood function since the distribution of $Z$’s is generally unknown. Hence, we resort to QMLE methodology and, for defining properly the QMLE, we consider the
Poisson log-likelihood function, as in Fokianos et al. (2009), where \( \theta \) denotes the unknown parameter vector, 

\[
l(Y; \theta) = \sum_{i=1}^{n} l_i(\theta) = \sum_{i=1}^{n} (Y_i \log \lambda_i(\theta) - \lambda_i(\theta)). \tag{1.14}
\]

The quasi-score function is defined by

\[
S_n(\theta) = \frac{\partial l(Y; \theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial l_i(\theta)}{\partial \theta} = \sum_{i=1}^{n} \left( \frac{Y_i}{\lambda_i(\theta)} - 1 \right) \frac{\partial \lambda_i(\theta)}{\partial \theta}, \tag{1.15}
\]

where the vector \( \frac{\partial \lambda_i(\theta)}{\partial \theta} \) is defined recursively depending upon the mean specification employed. The solution of the system of nonlinear equations \( S_n(\theta) = 0 \), if it exists, yields the QMLE of \( \theta \) which we denote by \( \hat{\theta} \). The conditional information matrix is defined by

\[
G_n(\theta) = \sum_{i=1}^{n} \text{Var} \left[ \frac{\partial l_i(\theta)}{\partial \theta} \bigg| \mathcal{F}_{t-1} \right] = \sum_{i=1}^{n} \left( \frac{1}{\lambda_i(\theta)} + \sigma^2 \right) \left( \frac{\partial \lambda_i(\theta)}{\partial \theta} \right)^2 \left( \frac{\partial \lambda_i(\theta)}{\partial \theta} \right)^2, \tag{1.16}
\]

The asymptotic properties of \( \hat{\theta} \) have been studied in detail by Fokianos et al. (2009) and Fokianos and Tjøstheim (2012) for the case of the Poisson model (1.3). For the case of the negative binomial distribution, see Christou and Fokianos (2014). In both cases, consistency and asymptotic normality of \( \hat{\theta} \) is achieved. In fact, it can be shown that the QMLE is asymptotically normally distributed, i.e.,

\[
\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}(0, \mathbf{G}^{-1}(\theta_0) \mathbf{G}_1(\theta_0) \mathbf{G}^{-1}(\theta_0)),
\]

where the matrices \( \mathbf{G} \) and \( \mathbf{G}_1 \) are given by the following:

\[
\mathbf{G}(\theta) = E \left( \frac{1}{\lambda_i(\theta)} \left( \frac{\partial \lambda_i(\theta)}{\partial \theta} \right)^2 \right), \quad \mathbf{G}_1(\theta) = E \left( \left( \frac{1}{\lambda_i(\theta)} + \sigma^2 \right) \left( \frac{\partial \lambda_i(\theta)}{\partial \theta} \right)^2 \right). \tag{1.16}
\]

For the case of Poisson distribution (1.3), we obtain that

\[
\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}(0, \mathbf{G}^{-1}(\theta_0)),
\]

since \( \sigma^2 = 0 \). If \( \sigma^2 > 0 \), then we need to estimate this parameter for maximum likelihood inference. We propose to estimate \( \sigma^2 \) by solving the equation

\[
\sum_{i=1}^{n} \frac{(Y_i - \hat{\lambda}_t)^2}{\hat{\lambda}_t(1 + \hat{\lambda}_t \sigma^2)} = n - m, \tag{1.17}
\]

where \( m \) denotes the dimension of \( \theta \) and \( \hat{\lambda}_t = \lambda_i(\hat{\theta}) \); see Lawless 1987. In particular, we recognize that for the negative binomial case, \( \sigma^2 = 1/\nu \). Therefore, the previous estimation method is standard in applications; see Cameron and Trivedi (1998, Ch. 3) among others.
Although the earlier mentioned formulas are given for the linear model (1.5), they can be modified suitably for the log-linear model (1.8) and the nonlinear model (1.9).

### 1.3.1 Transactions Data

Recall the transactions data discussed in Section 1.2. To fit model (1.5) to those data we proceed as follows: we set \( \lambda_0 = 0 \) and \( \partial \lambda_0 / \partial \theta = 0 \) for initialization of the recursions; recall (1.15). Starting values for the parameter vector \( \theta \) are obtained after observing that (1.5) can be expressed as an ARMA(1,1) model of the form

\[
(Y_t - \frac{d}{1 - (a_1 + b_1)}) = (a_1 + b_1)(Y_{t-1} - \frac{d}{1 - (a_1 + b_1)}) + \epsilon_t - a_1 \epsilon_{t-1}
\]

(1.18)

where \( \epsilon_t = Y_t - \lambda_t \) is a white noise process.

The first three lines of Table 1.1 report estimation results after fitting model (1.5) to these data. The first line reports estimates of the regression parameters (previously reported in Fokianos 2012). Regardless of the assumed distribution, the estimators are identical because we maximize the log-likelihood function (1.14). If the Poisson distribution is assumed to be the true data generating process, then these estimators are MLE; otherwise they are QMLE. If the true data generating process is assumed to be the negative binomial distribution, then we can also estimate the dispersion parameter \( \nu \) by means of (1.17) and the fact that \( \sigma_Z^2 = 1/\nu \). The next two rows report the standard errors of the estimators. For the regression parameters, these are calculated either by using the Poisson assumption and the matrix \( G \) from (1.16) (second row of Table 1.1) or by using the sandwich matrix obtained by \( G \) and \( G_1 \) from (1.16) (third row of Table 1.1). We note that the standard errors obtained from the sandwich matrix are somewhat larger than those obtained from \( G \). These are robust standard errors in the sense that we are using a working likelihood—namely the Poisson likelihood—to carry out inference (see White 1982). The standard error of \( \hat{\nu} \) has been computed by parametric bootstrap. In other words, given \( \hat{\theta} \) and \( \hat{\nu} \), generate a large number

| TABLE 1.1 |
| QMLE and their standards errors (in parentheses) for the Linear Model (1.5), the Nonlinear Model (1.12) with \( \gamma = 0.5 \) and the Loglinear Model (1.8), for the total number of transactions per minute for the Stock Ericsson B for the time period between July 2 and July 22, 2002 |

<table>
<thead>
<tr>
<th></th>
<th>Maximum Likelihood Estimates</th>
<th>Estimates of ( \nu )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{d} ) ( \hat{a}_1 ) ( \hat{b}_1 )</td>
<td>( \hat{\nu} )</td>
</tr>
<tr>
<td>Linear model (1.5)</td>
<td>0.581 0.745 0.199</td>
<td>7.158</td>
</tr>
<tr>
<td></td>
<td>(0.148) (0.030) (0.022)</td>
<td>(0.907)</td>
</tr>
<tr>
<td></td>
<td>(0.236) (0.047) (0.035)</td>
<td></td>
</tr>
<tr>
<td>Nonlinear model (1.12)</td>
<td>1.327 0.774 0.186</td>
<td>7.229</td>
</tr>
<tr>
<td></td>
<td>(0.324) (0.026) (0.021)</td>
<td>(1.028)</td>
</tr>
<tr>
<td></td>
<td>(0.506) (0.041) (0.034)</td>
<td></td>
</tr>
<tr>
<td>Log-linear model (1.8)</td>
<td>0.105 0.746 0.207</td>
<td>6.860</td>
</tr>
<tr>
<td></td>
<td>(0.032) (0.028) (0.022)</td>
<td>(1.067)</td>
</tr>
<tr>
<td></td>
<td>(0.125) (0.081) (0.070)</td>
<td></td>
</tr>
</tbody>
</table>

Note: The total number of observations is 460.
of count time series models by means of (1.5) and using (1.4). For each of the simulated count time series, carry out QML estimation and get an estimator of \( \nu \) by using (1.17). The standard error of these replicates is reported in Table 1.1, underneath \( \hat{\nu} \).

We now fit model (1.12) to these data with \( \gamma = 0.5 \). Although we have assumed for simplicity that \( \gamma \) is known, it can be estimated along with the regression parameters as outlined in Fokianos and Tjøstheim (2012). In principle, additional nonlinear parameters can be estimated by QMLE, but the sample size should be sufficiently large. For fitting this model, we set again \( \lambda_0 = 0 \) and \( \partial \lambda_0 / \partial \theta = 0 \). Starting values for the parameter vector \( \theta \) are obtained by initially fitting a linear model (1.5). Table 1.1 shows the results of this exercise. The estimators of \( a_1 \) and \( b_1 \) from model (1.12) are close to those obtained from fitting model (1.5). The same observation holds for \( \hat{\nu} \) and the standard errors of these coefficients which are computed in an analogous manner to that of the linear model. It is worth pointing out that the sum of \( \hat{a}_1 \) and \( \hat{b}_1 \) is close to unity. This fact provides some evidence of nonstationarity when we fit these types of models to transactions data. These observations are repeated when the log-linear model (1.8) is fitted to those data; see the last three lines of Table 1.1. Some empirical experience with these models has shown that when there is positive correlation among the data, then all models will produce similar output. In general, the log-linear model will be more useful when some covariate information is available.

## 1.4 Diagnostics

A detailed discussion concerning diagnostic methods for count time series models has been given by Kedem and Fokianos (2002, Sec. 1.6); also see the chapter by Jung et al. (2015; Chapter 9 in this volume). Various quantities, like Pearson and deviance residuals, for example, have been suggested and it was shown that they can be easily calculated using standard software. In what follows we focus our attention on the so-called Pearson residuals and on a new test statistic proposed recently by Fokianos and Neumann (2013) for testing the goodness of fit of the model under a conditional Poisson model. However, properties of different types of residuals (raw and deviance residuals for instance) have not been investigated in the literature. A slightly different definition of the residuals than the one which we will be using has been given recently by Zhu and Wang (2010). They also study the large sample behavior of the autocorrelation function of the residuals that they propose, but for a model of the form (1.5) without feedback.

### 1.4.1 Residuals

To examine the adequacy of the fit, we consider the so-called Pearson residuals. Set

\[
e_{t} = \frac{Y_t - E\left(Y_t \mid F_{t-1}^{Y, \lambda}ight)}{\sqrt{\text{Var}\left(Y_t \mid F_{t-1}^{Y, \lambda}\right)}} = \frac{Y_t - \lambda_t}{\sqrt{\lambda_t + \lambda_t^2 \sigma_Z^2}}, \quad t \geq 1, \tag{1.19}
\]

where the first equality is the general definition of the Pearson residuals and the second equality follows from (1.2). Under the true model, the sequence \( \{e_t, t \geq 1\} \) is a white noise
sequence with \( \text{Var}(e_t) = 1 \). It is straightforward to see that under the Poisson assumption, (1.19) becomes

\[
e_t = \frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}, \quad t \geq 1,
\]

while for the case (1.4), we obtain

\[
e_t = \frac{Y_t - \lambda_t}{\sqrt{\lambda_t + \lambda_t^2/\nu}}, \quad t \geq 1.
\]

To compute the Pearson residuals in either case, substitute \( \lambda_t \) by \( \hat{\lambda}_t \equiv \lambda_t(\hat{\theta}) \) and \( \sigma_Z^2 \) by \( \hat{\sigma}_Z^2 \). Construction of their autocorrelation function and cumulative periodogram plots (see Brockwell and Davis 1991, Sec. 10.2) give some clue about the whiteness of the sequence \( \{e_t, t \geq 1\} \).

Figure 1.1 shows the plots of the autocorrelation function and the cumulative periodogram of the Pearson residuals obtained after fitting (1.5) to the transactions data. The upper panel corresponds to the case of Poisson distribution and the lower panel is constructed using the negative binomial assumption. We observe that both models fit the data quite adequately. We have also computed the Pearson residuals for models (1.8) and (1.11). The corresponding plots are not shown because the results are quite analogous to the case of the simple linear model.

### 1.4.2 Goodness-of-Fit Test

A goodness-of-fit test for model (1.5), and more generally of (1.9), was recently proposed by Fokianos and Neumann (2013), by considering two forms of hypotheses. The first of these refers to the simple hypothesis

\[
H_0^{(s)}: \quad f = f_0 \quad \text{against} \quad H_1^{(s)}: \quad f \neq f_0,
\]

for some completely specified function \( f_0 \) which satisfies (1.10). However, in applications, the most interesting testing problem is given by the following composite hypotheses

\[
H_0: \quad f \in \{f_0: \Theta \in \Theta\} \quad \text{against} \quad H_1: \quad f \notin \{f_0: \Theta \in \Theta\}, \quad (1.20)
\]

where \( \Theta \subseteq \mathbb{R}^m \) and the function \( f_0 \) is known up to a parameter \( \theta \) and again satisfies (1.10).

The methodology for testing (1.20) is quite general and can be applied to all models considered so far. Recall that \( \Theta \) denotes the QMLE and \( \hat{\lambda}_t = \lambda_t(\hat{\theta}) \). If \( \hat{e}_t \) are the Pearson residuals (1.19), the statistic for testing (1.20) is given by

\[
\hat{T}_n = \sup_{x \in \Pi} |\hat{G}_n(x)|, \quad \hat{G}_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{e}_t w(x - \hat{l}_{t-1}),
\]

where \( x \in \Pi := [0, \infty)^2, \hat{l}_t = (\hat{\lambda}_t, Y_t), \) and \( w(\cdot) \) is some suitably defined weight function. In the applications, we can consider the weight function to be of the form \( w(x) = w(x_1, x_2) = K(x_1)K(x_2) \) where \( K(\cdot) \) is a univariate kernel and \( x = (x_1, x_2) \in \Pi \). We can employ the
FIGURE 1.1
Diagnostic plots for the Pearson residuals (1.19) after fitting model (1.5) to the transactions data. (a) Autocorrelation function using the Poisson assumption. (b) Cumulative periodogram plot using the Poisson assumption. (c) Autocorrelation function using the negative binomial assumption. (d) Cumulative periodogram plot using the negative binomial assumption.

uniform, Gaussian and the Epanechnikov kernels. For instance, when the uniform kernel is employed, compute the test statistic (1.21) by using the weights

\[ w(x - I_{t-1}) = K(x_1 - \lambda_{t-1})K(x_2 - Y_{t-1}) = \frac{1}{4} 1(|x_1 - \lambda_{t-1}| \leq 1) 1(|x_2 - Y_{t-1}| \leq 1), \]

where \(1(A)\) is the indicator function of a set \(A\). Then, the test statistic (1.21) becomes

\[ \hat{T}_n = \sup_{x \in \Pi} |\hat{G}_n(x)|, \]

where

\[ \hat{G}_n(x_1, x_2) = \frac{1}{4\sqrt{n}} \sum_{t=1}^n \hat{e}_t 1(|x_1 - \hat{\lambda}_{t-1}| \leq 1) 1(|x_2 - Y_{t-1}| \leq 1). \]
Obvious formulas hold for other kernel functions. It turns out that (1.21) yields a consistent procedure when testing against Pitman’s local alternatives. It converges weakly with the usual parametric rate under some regularity conditions on the kernel function; see Fokianos and Neumann (2013) for more details.

An alternative test statistic for testing goodness of fit for count time series can be based on supremum-type tests of the following form (Koul and Stute 1999):

\[
\hat{H}_n = \sup_{x \in \mathbb{P}} |H_n(x)|, \quad H_n(x) = n^{-1/2} \sum_{t=1}^{n} \hat{e}_t \mathbb{I}(\hat{I}_{t-1} \leq x),
\]

(1.22)

using the same notations as before. Although the asymptotic behavior of supremum-type test statistics based on (1.22) has not been studied in the literature, it is possible to develop a theory following the arguments of Koul and Stute (1999) and utilizing the recent results on weak dependence properties obtained by Doukhan et al. (2012), at least for some classes of models.

Regardless of the chosen statistic and the distributional assumption, we can calculate critical values by using parametric bootstrap; see Fokianos and Neumann (2013) for details under the Poisson assumption. More specifically, to compute the \( p \)-value of the test statistic, (1.21) or (1.22) is recalculated for \( B \) parametric bootstrap replications of the data set. Then, if \( \hat{T}_n \) denotes the observed value of the test statistic and \( \hat{T}_{i;n} \) denotes the value of the test statistic in the \( i \)th bootstrap run, the corresponding \( p \)-value used to determine acceptance/rejection is given by

\[
p\text{-value} = \frac{\# \{ i : \hat{T}_{i;n} \geq \hat{T}_n \}}{B + 1}.
\]

A similar result holds for \( \hat{H}_n \).

Test statistics (1.21) (with the uniform and Epanechnikov kernels) and (1.22) were computed for the transactions data for testing the goodness of fit of the linear model (1.5). Under the Poisson assumption, the observed values of these test statistics were calculated to be 0.212, 0.234, and 1.390, respectively. Under the negative binomial assumption, the observed values were equal to 0.146, 0.164, and 0.818, respectively. Table 1.2 shows the bootstrap \( p \)-value of the test statistics which have been obtained by parametric bootstrap as explained earlier. We note that the test statistics formed by (1.21) yield identical conclusions; that is, the linear model can be used for fitting the transactions data regardless of the assumed distribution. However, the test statistic (1.22) raises some doubt about the linearity, under the Poisson assumption.

**TABLE 1.2**  
P-values for the transactions data when testing for the Linear Model (1.5)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>(1.22)</th>
<th>(1.21) with Uniform Kernel</th>
<th>(1.21) with Epanechnikov Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>0.024</td>
<td>0.350</td>
<td>0.279</td>
</tr>
<tr>
<td>Negative binomial</td>
<td>0.659</td>
<td>0.611</td>
<td>0.585</td>
</tr>
</tbody>
</table>

Note: Results are based on \( B = 999 \) bootstrap replications.
1.5 Prediction

Following Gneiting et al. (2007), we take the point of view that predictions should be probabilistic in nature. In addition, they should strive to maximize the sharpness of the predictive distribution subject to calibration. Calibration refers to the statistical consistency between the predictive distribution and the observations. The notion of sharpness refers to the concentration of the predictive distribution and is a property of the forecasts only. It follows that if the predictive distribution is more concentrated, then the forecasts are sharper. In this section, we provide diagnostic tools to evaluate the predictive performance. Note that calculation of all these measures requires an assumption on the conditional distribution of the process; hence general processes of the form (1.2) cannot be fitted without simulating from the mixing variables $Z_t$. Predictive performance based on the following diagnostic tools has been examined recently by Jung and Tremayne (2011) and Christou and Fokianos (2015).

1.5.1 Assessment of Probabilistic Calibration

To ascertain whether or not the negative binomial distribution is a better choice than the Poisson distribution, we use the diagnostic tool of the Probability Integral Transformation (PIT) histogram, as explained below. This tool is used for checking the statistical consistency between the predictive distribution and the distribution of the observations. If the observation is drawn from the predictive distribution, then the PIT has a standard uniform distribution. In the case of count data, the predictive distribution is discrete and therefore the PIT is no longer uniform. To remedy this, several authors have suggested a randomized PIT. However, Czado et al. (2009) recently proposed a nonrandomized uniform version of the PIT. We explain their approach in the context of count time series models. Note that the approach is quite general and can accommodate various data generating processes.

In our context, we fit any model discussed earlier to the data by using the quasi-likelihood function (1.14). After obtaining consistent estimators for the regression parameters, we estimate the mean process $\lambda_t$ by $\hat{\lambda}_t = \lambda_t(\hat{\theta})$ and the parameter $\nu$ by $\hat{\nu}$. Then, the PIT is based on the conditional cumulative distribution

$$F(u|Y_t = y) = \begin{cases} 
0 & u \leq P_{y-1}, \\
(u - P_{y-1})/(P_y - P_{y-1}) & P_{y-1} \leq u \leq P_y, \\
1 & u \geq P_y,
\end{cases}$$

where $P_y$ is equal to the conditional c.d.f. either of the Poisson distribution (1.3) evaluated at $\hat{\lambda}_t$, or of the negative binomial p.m.f. (1.4) evaluated at $\hat{\lambda}_t$ and $\hat{\nu}$. Subsequently, we form the mean PIT by

$$\bar{F}(u) = \frac{1}{n} \sum_{t=1}^{n} F^{(t)}(u|y_t), \quad 0 \leq u \leq 1.$$
\[ f_j = \bar{F} \left\{ \frac{j}{J} \right\} - \bar{F} \left\{ \frac{j-1}{J} \right\} \]

for equally spaced bins \( j = 1, \ldots, J \). Then we plot the histogram with height \( f_j \) for bin \( j \) and check for uniformity. Deviations from uniformity hint at reasons for forecasting failures and model deficiencies. U-shaped histograms point at underdispersed predictive distributions, while hump or inverse–U shaped histograms indicate overdispersion.

Figure 1.2 shows nonrandomized PIT histograms with 10 equally spaced bins for two different situations for the transactions data. The left plots show the PIT histograms when
the fit is based on the Poisson distribution, for the linear and nonlinear models. Clearly, the plots show deviations from the Poisson distribution, indicating underdispersed predictive distributions. The right plots indicate no apparent deviations from uniformity; these plots are based on the negative binomial distribution (1.4). Similar findings were obtained after fitting the log-linear model (1.8) to the transactions data.

1.5.2 Assessment of Marginal Calibration

We now turn to the question of examining marginal calibration. We suppose that the observed time series \( \{Y_t, t \geq 1\} \) is stationary with marginal c.d.f. \( G(x) \). In addition, we assume that we pick a probabilistic forecast in the form of a predictive c.d.f. \( P_t(x) = P(Y_t \leq x | F_{t-1}^Y) \). In our case, \( P_t(x) \) is either the c.d.f. of a Poisson random variable with mean \( \lambda_t \), or a negative binomial distribution evaluated at \( \lambda_t \) and \( \nu \). We follow Gneiting et al. (2007) to assess marginal calibration by comparing the average predictive c.d.f.

\[
\tilde{P}(x) = \frac{1}{n} \sum_{t=1}^{n} P_t(x), \quad x \in \mathbb{R},
\]

to the empirical c.d.f. of the observations given by

\[
\hat{G}(x) = \frac{1}{n} \sum_{t=1}^{n} 1(Y_t \leq x), \quad x \in \mathbb{R}.
\]

To display the marginal calibration plot, we plot the difference of the two c.d.f.,

\[
\tilde{P}(x) - \hat{G}(x), \quad x \in \mathbb{R}.
\]  

(1.23)

Figure 1.3 shows that the negative binomial assumption provides a better fit than the Poisson assumption for the transactions data. These figures were drawn by direct calculation of the average c.d.f. \( \tilde{P} \) and the empirical c.d.f., as explained earlier.

1.5.3 Assessment of Sharpness

The assessment of sharpness is accomplished via scoring rules. These rules provide numerical scores and form summary measures for the assessment of the predictive performance. In addition, scoring rules help us to rank the competing forecast models. They are negatively oriented penalties that the forecaster wishes to minimize, see also Czado et al. (2009). Table 1.3 shows a few examples of scoring rules, following Czado et al. (2009). The calculation of all these scores requires an assumption on the conditional distribution of the process. The squared error score is identical for both the Poisson and the negative binomial distributions, since the conditional means are equal. Note that the normalized square error score is formed by the Pearson residuals defined in (1.19).
Table 1.3 shows all scoring rules applied to transactions data. It is clear that the negative binomial model fits these data considerably better than the Poisson model regardless of the assumed model. We note again that all models yield almost identical scoring rules. This further supports our point that when there exists positive persistent correlation in the data, then all models will produce similar output.

For the transactions data, a simple linear model of the form (1.5) under the negative binomial assumption seems to describe the data adequately. This conclusion is a direct consequence of the earlier findings with additional evidence provided by the values of goodness-of-fit test (1.21) reported in Table 1.2.
TABLE 1.4
Scoring rules calculated for the transactions data after fitting the Linear Model (1.5), the Nonlinear Model (1.12) for $\gamma = 0.5$, and the Log-Linear Model (1.8)

<table>
<thead>
<tr>
<th>Forecaster</th>
<th>Scoring Rules</th>
<th>logs</th>
<th>qs</th>
<th>sphs</th>
<th>rps</th>
<th>dss</th>
<th>nses</th>
<th>ses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear model (1.5)</td>
<td>Poisson</td>
<td>3.126</td>
<td>−0.076</td>
<td>−0.276</td>
<td>3.633</td>
<td>4.585</td>
<td>2.326</td>
<td>23.477</td>
</tr>
<tr>
<td></td>
<td>NegBin</td>
<td>2.902</td>
<td>−0.080</td>
<td>−0.292</td>
<td>3.284</td>
<td>4.112</td>
<td>0.993</td>
<td>23.477</td>
</tr>
<tr>
<td>Nonlinear model (1.12)</td>
<td>Poisson</td>
<td>3.123</td>
<td>−0.075</td>
<td>−0.274</td>
<td>3.605</td>
<td>4.579</td>
<td>2.318</td>
<td>23.435</td>
</tr>
<tr>
<td></td>
<td>NegBin</td>
<td>2.901</td>
<td>−0.080</td>
<td>−0.289</td>
<td>3.267</td>
<td>4.107</td>
<td>0.985</td>
<td>23.435</td>
</tr>
<tr>
<td>Log-linear model (1.8)</td>
<td>Poisson</td>
<td>3.144</td>
<td>−0.081</td>
<td>−0.286</td>
<td>3.764</td>
<td>4.633</td>
<td>2.376</td>
<td>23.894</td>
</tr>
<tr>
<td></td>
<td>NegBin</td>
<td>2.910</td>
<td>−0.083</td>
<td>−0.300</td>
<td>3.334</td>
<td>4.132</td>
<td>0.993</td>
<td>23.894</td>
</tr>
</tbody>
</table>

**Note:** The two forecasters are compared by the mean logarithmic, quadratic, spherical, ranked probability, Dawid–Sebastiani, normalized squared error and squared error scores. Bold face numbers in each column indicate the minimum value obtained between the two forecasters.

1.6 Other Topics

In this section, we will discuss other interesting research topics in the context of count time series analysis. This list is not exhaustive, and several other interesting topics will be covered in the following chapters. The list below reflects our personal research interests in the framework of generalized linear models.

1.6.1 Testing for Linearity

Consider the nonlinear model (1.11) and suppose that we are interested in testing the hypothesis $H_0 : c_1 = 0$ which is equivalent to testing linearity of the model. This testing problem is not standard because under the hypothesis, the nonlinear parameter $\gamma$ is not identifiable. Furthermore, $c_1 = 0$ implies that the parameter is on the boundary of the parameter space. Similar comments can be made for model (1.12) when testing the hypothesis $H_0 : \gamma = 0$, but without the additional challenge implied by the nonidentifiability issue. These type of testing problems have been recently discussed by Christou and Fokianos (2013).

Suppose that, in general, the vector of unknown parameters can be decomposed as $\theta = (\theta^{(1)}, \theta^{(2)})$ and let $S_n = (S_n^{(1)}, S_n^{(2)})$ be the corresponding partition of the score function. Consider testing $H_0 : \theta^{(2)} = 0$ vs. $H_1 : \theta^{(2)} > 0$, componentwise. This problem is attacked by using the score test statistic which is given by

$$LM_n = S_n^{(2)'} (\tilde{\theta}_n) \Sigma^{-1}_{\theta} (\tilde{\theta}_n) S_n^{(2)} (\tilde{\theta}_n),$$

where $\tilde{\theta}_n = (\tilde{\theta}_n^{(1)}, 0)$ is the QMLE of $\theta$ under the hypothesis and $\tilde{\Sigma}$ is an appropriate estimator for the covariance matrix $\Sigma = \text{Var} \left( \frac{1}{\sqrt{n}} S_n^{(2)} (\tilde{\theta}_n) \right)$. If all the parameters are identified...
under the null hypothesis, the score statistic (1.11) follows asymptotically a \( \chi^2_{m_2} \) distribution under the null, where \( m_2 = \text{dim}(\theta^{(2)}) \) (Francq and Zakoïan 2010, Ch. 8). Model (1.12) belongs to this class.

When the parameters are not identified under the null, a supremum type test statistic resolves this problem; see for instance Davies (1987). Consider model (1.11), for example, and let \( \Gamma \) be a grid of values for the nuisance parameter, denoted by \( \gamma \). Then the sup-score test statistic is given by

\[
LM_n = \sup_{\gamma \in \Gamma} LM_n(\gamma).
\]

Critical values of the test statistics can be either based on the asymptotic chi-square approximation or by employing parametric bootstrap as in the case of the test statistic (1.21).

### 1.6.2 Intervention Analysis

Occasionally, some time series data may show that both variation and the level of the data change during some specific time interval. Additionally, there might exist outlying values (unusual values) at some time points. This is the case for the campylobacteriosis infections data reported from January 1990 to the end of October 2000 in the north of the Province of Québec, Canada; see Fokianos and Fried (2010, Fig. 1). It is natural to ask whether these fluctuations can be explained by (1.5) or whether the inclusion of some interventions will yield better results; see Box and Tiao (1975), Tsay (1986) and Chen and Liu (1993) among others.

Generally speaking, types of intervention effects on time series data are classified according to whether their impact is concentrated on a single or a few data points, or whether they affect the whole process from some specific time \( t = \tau \) on. In classical linear time series methodology, an intervention effect is included in the observation equation by employing a sequence of deterministic covariates \( \{X_t\} \) of the form

\[
X_t = \xi(B)I_t(\tau),
\]

where \( \xi(B) \) is a polynomial operator, \( B \) is the shift operator such that \( B^jX_t = X_{t-j} \), and \( I_t(\tau) \) is an indicator function, with \( I_t(\tau) = 1 \) if \( t = \tau \), and \( I_t(\tau) = 0 \) if \( t \neq \tau \). The choice of the operator \( \xi(B) \) determines the kind of intervention effect: additive outlier (AO), transient shift (TS), level shift (LS), or innovational outlier (IO). Since models of the form (1.5) are not defined in terms of innovations, we focus on the first three types of interventions (but see Fried et al. 2015 for a Bayesian point of view).

However, a model like (1.5) is determined by a latent process. Therefore, a formal linear structure, as in the case of the Gaussian linear time series model, does not hold any more and interpretation of the interventions is a more complicated issue. Hence, a method which allows the detection of interventions and estimation of their size is needed so that structural changes can be identified successfully. Important steps to achieve this goal are the following; see Chen and Liu (1993):

1. A suitable model for accommodating interventions in count time series data.
2. Derivation of test procedures for their successful detection.
3. Implementation of joint maximum likelihood estimation of model parameters and outlier sizes.
4. Correction of the observed series for the detected interventions.

All these issues and possible directions for further developments of the methodology have been addressed by Fokianos and Fried (2010, 2012) for the linear model (1.5) and the log-linear model (1.8), under the Poisson assumption.

1.6.3 Robust Estimation

The previous work on intervention analysis is complemented by developing robust estimation procedures for count time series models. The works by El Saied (2012) and El Saied and Fried (2014) address this research topic in the context of the linear model (1.5) when $a_1 = 0$. In the context of log-linear model (1.8), the work of Kitromilidou and Fokianos (2015) develops robust estimation for count time series by adopting the methods suggested by Künsch et al. (1989) and Cantoni and Ronchetti (2001). In particular, Cantoni and Ronchetti (2001) robustified the quasi-likelihood approach for estimating the regression coefficient of generalized linear models by considering robust deviances which are natural generalizations of the quasi-likelihood functions. The robustification proposed by Cantoni and Ronchetti (2001) is performed by bounding and centering the quasi-score function.

1.6.4 Multivariate Count Time Series Models

Another interesting topic of research is the analysis of multivariate count time series models; see Liu (2012), Pedeli and Karlis (2013), and Section V of this volume which contains many interesting results. The main issue for attacking the problem of multivariate count time series is that multivariate count distributions are quite complex to be analyzed by maximum likelihood methods.

Assume that $\{Y_t = (Y_{i,t}), t = 1, 2, \ldots, n\}$ denotes a $p$-dimensional count time series and suppose further that $\{\lambda_t = (\lambda_{i,t}), t = 1, 2, \ldots, n\}$ is a corresponding $p$-dimensional intensity process. Here the notation $p$ denotes dimension but not order as in (1.13). Then, a natural generalization of (1.5) is given by

$$Y_{i,t} = N_{i,t}(0, \lambda_{i,t}), \quad i = 1, 2, \ldots, p, \quad \lambda_t = d + A \lambda_{t-1} + B Y_{t-1},$$

(1.25)

where $d$ is a $p$-dimensional vector and $A, B$ are $p \times p$ matrices, all of them unknowns to be estimated. Model (1.25) is a direct extension of the linear autoregressive model (1.5) and assumes that marginally the count process is Poisson distributed. However, the statistical problem of dealing with the joint distribution of the vector process $\{Y_t\}$ requires further research; some preliminary results about ergodicity and stationarity of (1.25) have been obtained by Liu (2012). More on multivariate models for count time series is given in the chapter by Karlis (2015; Chapter 19 in this volume), and an application is discussed by Ravishanker et al. (2015; Chapter 20 in this volume).

1.6.5 Parameter-Driven Models

So far we have discussed models that fall under the framework of observation-driven models. This implies that even though the mean process $\{\lambda_t\}$ is not observed directly, it can still
be recovered explicitly as a function of the past responses. However, a different point of view has been taken by Zeger (1988), who introduced regression models for time series of counts by assuming that the observed process is driven by a latent (unobserved) process. To be more specific, suppose that, conditional on an unobserved process \( \{\xi_t, t \geq 1\}, \{Y_t, t \geq 1\} \), is a sequence of independent counts such that

\[
E[Y_t | \xi_t] = \text{Var}[Y_t | \xi_t] = \xi_t \exp(d + a_1 y_{t-1}).
\]  

In (1.26) we consider a simple model for illustration, but more complex models that include higher-order lagged values of the response and any covariates can be assumed. It can be proved that the earlier formulation, although similar to a Poisson log-linear model, reveals that the observed data are overdispersed. Estimation of all unknown parameters is discussed by Zeger (1988). A further detailed study of model (1.26) can be found in Davis et al. (2000), where the authors address the problem of existence of the latent stochastic process \( \{\xi_t\} \) and derive the asymptotic distribution of the regression coefficients when the latter exist. In the context of negative binomial regression, the latent process model (1.26) has been extended by Davis and Wu (2009). See also Harvey and Fernandes (1989) for a state-space approach with conjugate priors for the analysis of count time series and Jørgensen et al. (1999) for multivariate longitudinal count data. More generally, state space models for count time series are discussed in West and Harrison (1997), Durbin and Koopman (2001), and Cappé et al. (2005), among others.

1.7 Other Extensions

There are several other possible directions for extending the theory and methods discussed in this chapter. Threshold models have been considered recently by Woodard et al. (2011), Douc et al. (2013), and Wang et al. (2014). Several questions are posed by such models such as estimation of regression and threshold or/and delay parameters. The concept of mixture models for the analysis of count time series data is a topic closely related to that of threshold models; for the real-valued case, see Wong and Li (2001) among others. Finally, I would like to bring forward the possibility of introducing local stationarity to count time series in the sense of Dahlhaus (1997, 2000). Such models pose several problems; for instance, the question of existence of a stationary approximation and estimation of the time-varying parameters by nonparametric likelihood inference.

Acknowledgments

Many thanks to the editors for all their efforts and for inviting me to contribute to this volume. Special thanks are due to V. Christou, N. Papamichael, D. Tjostheim, and a referee for several useful comments and suggestions. Support was provided by the Cyprus Research Promotion Foundation grant TEXNOLOGIA/THEPIS/0609(BE)/02.
References


Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74:33–43.


