7

Estimating Equation Approaches for Integer-Valued Time Series Models

Aerambamoorthy Thavaneswaran and Nalini Ravishanker

CONTENTS

7.1 Introduction ...................................................................................145
7.2 A Review of Estimating Functions (EFs)..................................................147
7.3 Models and Moment Properties for Count Time Series .........................148
  7.3.1 Models for Nominally Dispersed Counts.........................................149
  7.3.2 Models for Counts with Excess Zeros.............................................150
  7.3.3 Models in the GAS Framework.....................................................152
7.4 Parametric Inference via EFs..............................................................152
  7.4.1 Linear EFs..............................................................................152
  7.4.2 Combined EFs.........................................................................154
7.5 Hypothesis Testing and Model Choice....................................................160
7.6 Discussion and Summary...................................................................161
References............................................................................................162

7.1 Introduction

There is considerable current interest in the study of integer-valued time series models, and for time series of counts, in particular. Applications abound in biometrics, ecology, economics, engineering, finance, public health, etc. Given the increase in stochastic complexity and data sizes, there is a need for developing fast and optimal approaches for model inference and prediction. Several observation-driven and parameter-driven (Cox, 1981) modeling frameworks for count time series have been discussed over the past few decades. Further, although there is a large literature for count time series without zero-inflation, including both observation-driven and parameter-driven models, very few papers have been published for modeling time series with excess zeros.

In parameter-driven models, temporal association is modeled indirectly by specifying the parameters in the conditional distribution of the count random variable to be a function of a correlated latent stochastic process (West and Harrison, 1997). In observation-driven models, temporal association is modeled directly via lagged values of the count variable, adopting strategies such as binomial thinning to preserve the integer nature of the data (Al-Osh and Alzaid, 1987; McKenzie, 2003). Davis et al. (2003), Jung and Tremayne (2006), and Neal and Subba Rao (2007), among others, have discussed estimation and inference...
for these models. Heinen (2003) and Ghahramani and Thavaneswaran (2009b) described autoregressive conditional Poisson (ACP) models. Ferland et al. (2006) and Zhu (2011, 2012a,b) defined classes of integer-valued time series models following different conditional distributions, which they called INGARCH models, and studied the first two process moments. Although these are called INGARCH models, only the conditional mean of the count variable is modeled, and not its conditional variance. In a recent paper, Creal et al. (2013) described generalized autoregressive score (GAS) models to study time-varying parameters in an observation-driven modeling framework, while MacDonald and Zucchini (2015; Chapter 12 in this volume) discussed a hidden Markov modeling framework.

Estimating functions (EFs) have a long history in statistical inference. For instance, Fisher (1924) showed that maximum likelihood and minimum chi-squared methods are asymptotically equivalent by comparing the first order conditions of the two estimation procedures, that is, by analyzing properties of estimators by focusing on the corresponding EFs rather than on the objective functions or estimators themselves. Godambe (1960) and Durbin (1960) gave a fundamental optimality result for EFs for the scalar parameter case. Following Godambe (1985), who first studied inference based on the EF approach for discrete-time stochastic processes, Thavaneswaran and Abraham (1988) described estimation for nonlinear time series models using linear EFs. Naik-Nimbalkar and Rajarshi (1995) and Thavaneswaran and Heyde (1999) studied problems in filtering and prediction using linear EFs in the Bayesian context. Merkouris (2007), Ghahramani and Thavaneswaran (2009a, 2012), and Thavaneswaran et al. (2015), among others, studied estimation for time series via the combined EF approach. Bera et al. (2006) gave an excellent survey on the historical development of this topic.

Except for a few papers, (Dean, 1991), who discussed estimating equations for mixed Poisson models given independent observations, application of the EF approach to count time series is still largely unexplored. In the following sections, we extend this approach for count time series models. For some recently proposed integer-valued time series models (such as the Poisson, generalized Poisson (GP), zero-inflated Poisson, or negative binomial models), the conditional mean and variance are functions of the same parameter. This motivates considering more informative quadratic EFs for joint estimation of the conditional mean and variance parameters, rather than only using linear EFs. It is also possible to derive closed form expressions for the information gain (Thavaneswaran et al., 2015).

In this chapter, we describe a framework for optimal estimation of parameters in integer-valued time series models via martingale EFs and illustrate the approach for some interesting count time series models. The EF approach only relies on a specification of the first few moments of the random variable at each time conditional on its history, and does not require specification of the form of the conditional probability distribution. We start with a brief review of the general theory of EFs in Section 7.2. In Section 7.3, we describe the conditional moment properties for some recently proposed classes of generalized integer-valued models, such as those discussed in Ferland et al. (2006). Specifically, we derive the first four conditional moments, which are typically required for carrying out inference on model parameters using the theory of combined martingale EFs (Liang et al., 2011). Section 7.4 describes the optimal EFs that enable joint parameter estimation for such models. We also derive fast, recursive, on-line estimation techniques for parameters of interest and provide examples. In Section 7.5, we describe how hypothesis testing based on optimal estimation facilitates model choice. Section 7.6 concludes with a summary and a brief discussion of parameter-driven doubly stochastic models for count time series.
7.2 A Review of Estimating Functions (EFs)

Godambe (1985) first described an EF approach for stochastic process inference. Suppose that \( \{y_t, t = 1, \ldots, n\} \) is a realization of a discrete time stochastic process, and suppose its conditional distribution depends on a vector parameter \( \theta \) belonging to an open subset \( \Theta \) of the \( p \)-dimensional Euclidean space, with \( p \ll n \). Let \( (\Omega, \mathcal{F}, P_\theta) \) denote the underlying probability space, and let \( \mathcal{F}_t \) be the \( \sigma \)-field generated by \( \{y_1, \ldots, y_t, t \geq 1\} \). Let \( m_t = m_t(y_1, \ldots, y_t, \theta), 1 \leq t \leq n \), be specified \( q \)-dimensional martingale difference vectors. Consider the class \( \mathcal{M} \) of zero-mean, square integrable \( p \)-dimensional martingale EFs, viz.,

\[
\mathcal{M} = \left\{ g_n(\theta) : g_n(\theta) = \sum_{t=1}^{n} a_{t-1}(\theta)m_t \right\}, \tag{7.1}
\]

where \( a_{t-1}(\theta) \) are \( p \times q \) matrices that are functions of \( \theta \) and \( y_1, \ldots, y_{t-1}, 1 \leq t \leq n \). It is further assumed that \( g_n(\theta) \) are almost surely differentiable with respect to the components of \( \theta \), and are such that for each \( n \geq 1 \), \( E \left( \frac{\partial g_n(\theta)}{\partial \theta} \bigg| \mathcal{F}_{n-1} \right) \) and \( E(g_n(\theta)g_n(\theta)'| \mathcal{F}_{n-1}) \) are non-singular for all \( \theta \in \Theta \), where all expectations are taken with respect to \( P_\theta \). An estimator of \( \theta \) is obtained by solving the estimating equation \( g_n(\theta) = 0 \). Furthermore, the \( p \times p \) matrix \( E(g_n(\theta)g_n(\theta)'| \mathcal{F}_{n-1}) \) is assumed to be positive definite for all \( \theta \in \Theta \). Then, in the class of all zero-mean and square integrable martingale EFs \( \mathcal{M} \), the optimal EF \( g^*_n(\theta) \) that maximizes, in the partial order of nonnegative definite matrices, the information

\[
I_{g_n}(\theta) = \left[ E \left( \frac{\partial g_n(\theta)}{\partial \theta} \bigg| \mathcal{F}_{n-1} \right) \right]' \left[ E(g_n(\theta)g_n(\theta)'| \mathcal{F}_{n-1}) \right]^{-1} \left[ E \left( \frac{\partial g_n(\theta)}{\partial \theta} \bigg| \mathcal{F}_{n-1} \right) \right],
\]

is given by

\[
g^*_n(\theta) = \sum_{t=1}^{n} a^*_{t-1}(\theta)m_t = \sum_{t=1}^{n} E \left( \frac{\partial m_t}{\partial \theta} \bigg| \mathcal{F}_{t-1} \right)' \left[ E(m_t m_t'| \mathcal{F}_{t-1}) \right]^{-1} m_t, \tag{7.2}
\]

and the corresponding optimal information reduces to

\[
I_{g_n}(\theta) = E(g^*_n(\theta)g^*_n(\theta)'| \mathcal{F}_{n-1}). \tag{7.3}
\]

The function \( g^*_n(\theta) \) is also called the “quasi-score” and has properties similar to those of a score function: \( E(g^*_n(\theta)) = 0 \) and \( E(g^*_n(\theta)g^*_n(\theta)'| = -E(\partial g^*_n(\theta)/\partial \theta') \). This is a general result in that we do not need to assume that the true underlying conditional distribution belongs to an exponential family of distributions. The maximum correlation between the optimal EF and the true unknown score justifies the terminology “quasi-score” for \( g^*_n(\theta) \). It is useful to note that the same procedure for derivation of optimal estimating equations may be used when the time series is stationary or nonstationary. Moreover, the finite sample
properties of the EFs remain the same, although asymptotic properties will differ. In Chapter 12 of his book, Heyde (1997) discussed general consistency and asymptotic distributional results.

Consider an integer-valued discrete-time scalar stochastic process \( \{y_t, t = 1, 2, \ldots\} \) with conditional mean, variance, skewness, and kurtosis given by

\[
\begin{align*}
\mu_t(\theta) &= \mathbb{E}(y_t|F_{t-1}), \\
\sigma_t^2(\theta) &= \text{Var}(y_t|F_{t-1}), \\
\gamma_t(\theta) &= \frac{1}{\sigma_t^3(\theta)}\mathbb{E}\left((y_t - \mu_t(\theta))^3|F_{t-1}\right), \text{ and} \\
\kappa_t(\theta) &= \frac{1}{\sigma_t^4(\theta)}\mathbb{E}\left((y_t - \mu_t(\theta))^4|F_{t-1}\right).
\end{align*}
\]

(7.4)

To jointly estimate the conditional mean and variance, which are both functions of \( \theta \), Liang et al. (2011) defined optimal combined EFs. We assume that \( \mu_t(\theta) \) and \( \sigma_t^2(\theta) \) are differentiable with respect to \( \theta \), and that the skewness and kurtosis of the standardized \( y_t \) do not depend on additional parameters beyond \( \theta \). For each data/model combination, our estimation approach for \( \theta \) requires (1) computation of the first four moments of \( y_t \) conditional on the process history, (2) selection of suitable linear and/or quadratic martingale differences, (3) construction of optimal combined EFs, and (4) derivation of recursive estimators of \( \theta \) when possible. In Section 7.4, we describe optimal estimating equations for \( \theta \) for some of the integer-valued models discussed in Section 7.3.

### 7.3 Models and Moment Properties for Count Time Series

Several models have been discussed in the literature for count time series, where parameter estimation using maximum likelihood or Bayesian approaches have been described. For the estimating equations framework described in this chapter, we start from the conditional moments of the process \( \{y_t\} \) given the history \( F_{t-1} \). The conditional moments are assumed to be functions of an unknown parameter vector \( \theta \) and form the basis for constructing the optimal estimating equation. For simplicity, we suppress \( \theta \) in the notation for the conditional moments and other derived quantities in the following examples. Consider the discrete-time model for \( \mu_t \) with \( P + Q + 1 \) parameters defined by

\[
\mu_t = \delta + \sum_{i=1}^{P} \alpha_i y_{t-i} + \sum_{j=1}^{Q} \beta_j \mu_{t-j},
\]

(7.5)

where \( \delta > 0 \), \( \alpha_i \geq 0 \) for \( i = 1, \ldots, P \) and \( \beta_j \geq 0 \) for \( j = 1, \ldots, Q \). Let \( \theta = (\delta, \alpha', \beta')' \) where \( \alpha = (\alpha_1, \ldots, \alpha_P)' \) and \( \beta = (\beta_1, \ldots, \beta_Q)' \). We assume that the conditional variance \( \sigma_t^2 \) as well as \( \mu_t \) depend on \( \theta \), and that the conditional skewness \( \gamma_t \) and conditional kurtosis \( \kappa_t \) are available and do not depend on any additional parameters. The higher order conditional moment properties for the models described in Sections 7.3.1 and 7.3.2, especially for the
zero-inflated case, are obtained using Mathematica. Section 7.3.3 proposes a model in the framework of the GAS models of Creal et al. (2013).

Equation (7.5) posits an ARMA model for \{y_t\}. This ARMA representation is useful for obtaining unconditional moments such as skewness and kurtosis under the stationarity assumption and is often useful in model identification in data analysis. We consider the martingale difference \( m_t = y_t - \mu_t \), with conditional mean 0 and conditional variance \( \sigma_t^2 \). Then (7.5) can be written as

\[
y_t - m_t = \delta + \sum_{i=1}^{P} \alpha_i y_{t-i} + \sum_{j=1}^{Q} \beta_j (y_{t-j} - m_{t-j}).
\]

Rearranging terms and simplifying, we can write

\[
\left( 1 - \sum_{i=1}^{\max(P,Q)} (\alpha_i + \beta_i) B^i \right) y_t = \delta + \left( 1 - \sum_{j=1}^{Q} \beta_j B^j \right) m_t, \text{ or } \phi(B)y_t = \delta + \beta(B)m_t,
\]

where \( B \) denotes the backshift operator. That is, (7.5) can be written as an ARMA model for \( \{y_t\} \) with \( \phi(B) = 1 - \sum_{i=1}^{\max(P,Q)} \phi_i B^i, \phi_i = \alpha_i + \beta_i, \beta(B) = 1 - \sum_{i=1}^{Q} \beta_i B^i, \) and \( \psi(B)\phi(B) = \beta(B) \psi(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i \). Similar to the continuous-valued case (Gourieroux, 1997), this model has the same second-order properties as an INARMA(\( P, Q, Q \)) model. When all solutions to \( \phi(z) = 0 \) lie outside the unit circle, we may write the moving average representation of the causal process as \( y_t = \mu + \psi(B)m_t \), where \( \psi(B) = \beta(B)/\phi(B) \) and \( \mu = \delta/(1 - \phi_1 - \ldots - \phi_{\max(P,Q)}) \) is the marginal mean of \( y_t \). The lag \( k \) autocovariance and autocorrelation of the process are, respectively, \( \gamma_k = E(\sigma_t^2) \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \) and \( \rho_k = \gamma_k / \gamma_0 = \sum_{j=0}^{\infty} \psi_j \psi_{j+k} / \sum_{j=0}^{\infty} \psi_j^2 \), where \( E(\sigma_t^2) \) is the unconditional variance of \( \{y_t\} \).

Note that the temporal correlation \( \rho_k \) depends only on the model parameters in (7.5) and not on the conditional distribution of the observed process \( \{y_t\} \). Also, the kurtosis of \( \{y_t\} \) is given by

\[
K^{(y)} = 3 + \frac{(K^{(m)} - 3) \sum_{j=0}^{\infty} \psi_j^4}{\left( \sum_{j=0}^{\infty} \psi_j^2 \right)^2},
\]

where \( K^{(m)} = E(m_t^4)/[E(m_t^2)]^2 \). These results follow directly from properties of stationary ARMA processes and often provide guidance in model order choice. By substituting suitable values of \( \psi_j \), we can derive the kurtosis for the integer-valued processes discussed in the following sections.

### 7.3.1 Models for Nominally Dispersed Counts

Considerable attention has been paid in the literature for modeling count time series via observation-driven models (Zeger and Qaqish, 1988; Davis et al., 2003) and parameter-driven models (Chan and Ledolter, 1995; West and Harrison, 1997). We consider three examples.
Example 7.1

Suppose that the conditional mean, variance, skewness, and kurtosis of $y_t$ are specified as $\mu_t = \lambda_t$, $\sigma_t^2 = \lambda_t$, $\gamma_t = \lambda_t^{-1/2}$, and $\kappa_t = \lambda_t^{-1}$, and suppose that $\mu$ is modeled by (7.5). These moments match the first four moments of $y_t$ generated by what Ferland et al. (2006) referred to as a Poisson INGARCH process, which assumes that $y_t|F_{t-1} \sim$ Poisson($\lambda_t$), so that $\lambda_t$ is the conditional variance as well as the conditional mean. While it seems a misnomer to use the term INGARCH for modeling the conditional mean and not the conditional variance as GARCH models do, the form of (7.5) is similar to the normal-GARCH model (Bollerslev, 1986), where $y_t|F_{t-1} \sim N(0, \sigma_t^2)$ for all $t$, and the model for the conditional variance $\sigma_t^2$ follows the right side of (7.5), subject to the same conditions on the parameters. For conformity with the literature, we use the term INGARCH in this chapter. The moments of the Poisson INGARCH random variable $y_t$ are easily derived from the probability generating function $G_y(s) = E(exp(s y_t)|F_{t-1}) = \exp(\lambda_t(s - 1))$. Implementation of the EF approach does not require that at each time $t$, $y_t$ has a conditional Poisson distribution, but only requires specification of the conditional moments of $y_t|F_{t-1}$ for each $t$. Such moment specifications are also sufficient for the other INGARCH models described in this chapter. □

Example 7.2

Suppose the conditional mean, variance, skewness, and kurtosis of $y_t$ given $F_{t-1}$ are $\mu_t = \lambda_t^*/(1 - \tau) = \lambda_t$, $\sigma_t^2 = \lambda_t^*/(1 - \tau)^3 = \tau^2(1 - \tau)$, $\gamma_t = (1 + 2\tau)/\sqrt{\lambda_t^*/(1 - \tau)}$, and $\kappa_t = (1 + 8\tau + 6\tau^2)/[\lambda_t^*/(1 - \tau)]$, corresponding to moments from the GP INGARCH process (Zhu, 2012a), where $\tau^* = 1/(1 - \tau)$. This process is defined as $y_t|F_{t-1} \sim GP(\lambda_t^*, \tau)$, $\lambda_t^* = (1 - \tau)\lambda_t$, $\max(-1, -\lambda_t^*/4) < \tau < 1$, and the conditional mean is again modeled by (7.5). The GP distribution for $y_t$ conditional on $F_{t-1}$ is

$$P(y_t = k|F_{t-1}) = \begin{cases} \lambda_t(\lambda_t + \tau k)^{k-1} \exp[-(\lambda_t + \tau k)]/k!, & k = 0, 1, 2, \ldots \, , \\ 0, & k > m \text{ if } \tau < 0 \end{cases}$$

(7.7)

where $m$ is the largest positive integer for which $\lambda_t + \tau m > 0$ when $\tau < 0$. To derive the conditional moments of the GP distribution shown above, we can use the recursive relation for the $r$th raw moment $\mu(r)$, that is, $(1 - \tau)\mu(r) = \lambda_t(\mu(r-1) + \lambda_t \partial \mu(r)/\partial \lambda_t + \tau^2 (\partial^2 \mu(r)/\partial \tau^2)$, where $\lambda_t > 0$ and $\max(-1, -\lambda_t/m) < \tau < 1$. □

Example 7.3

For $p_t = 1/(1 + \lambda_t)$ and $q_t = 1 - p_t$, suppose the conditional mean, variance, skewness, and kurtosis of $y_t$ given $F_{t-1}$ are $\mu_t = \eta_t/p_t = r\lambda_t$, $\sigma_t^2 = \eta_t/q_t^2$, $\gamma_t = (r - p_t)/\eta_t$, and $\kappa_t = (r^2 - 6p_t + 6)/\eta_t$, which correspond to the moments of a negative binomial INGARCH process, where $y_t|F_{t-1} \sim NB(r, \lambda_t)$, the conditional mean is modeled by (7.5) as before, and the probability generating function is given by $G_y(s) = p_t^e/(1 - q_t s)^r$, and the conditional probability mass function (pmf) of $y_t$ has the form

$$P(y_t = k|F_{t-1}) = \binom{k + r - 1}{r - 1} p_t^e q_t^k, \, k = 0, 1, 2, \ldots \,.$$  

(7.8)

7.3.2 Models for Counts with Excess Zeros

In several applications, observed counts over time may show an excess of zeros, and the usual Poisson or negative binomial models are inadequate. One example in the area of public health could involve surveillance of a rare disease over time, where the observed
counts typically show zero inflation. Yang (2012) studied statistical modeling for time series with excess zeros. We consider two examples.

Example 7.4

When count time series are observed with an excess of zeros in applications, we may assume a specification of the conditional mean, variance, skewness, and kurtosis of $y_t$ given $F_{t-1}$ as $\mu_t = (1 - \omega)t$, $\sigma_t^2 = (1 - \omega)(1 + \omega \lambda_t)$, $\gamma_t = ((1 - \omega)\lambda_t)^{-1/2}(1 + \omega \lambda_t)^{-3/2}(1 + \omega \lambda_t(3 + 2\omega \lambda_t + \lambda_t))$, and $\kappa_t = [(1 - \omega)(1 + \omega \lambda_t)^2 \lambda_t]^{-1} [1 + \omega \lambda_t (7 + \lambda_t (-6 + 12 \omega + \lambda_t (1 - 6(1 - \omega) \omega))) ]$, where $0 < \omega < 1$, and we assume that the conditional mean is modeled by (7.5). The moments correspond to the first four moments of a zero-inflated Poisson INGARCH process (Zhu, 2012b) given by $y_t | F_{t-1} \sim ZIP(\lambda_t, \omega)$ defined by

$$P(y_t = k | F_{t-1}) = \omega \Delta_{k,0} + (1 - \omega)\lambda_t^k \exp(-\lambda_t)/k!,$$

(7.9)

for $k = 0, 1, 2, \ldots$. When $\omega = 0$, the ZIP INGARCH model reduces to the Poisson INGARCH model. The probability generating function of a ZIP random variable is given by $G_y(s) = \omega + (1 - \omega) \exp[\lambda_t(s - 1)]$. The conditional variance exceeds the conditional mean, so the ZIP-INGARCH model can handle overdispersion. As mentioned earlier, in this and the following models, we only require specification of the conditional moments of $y_t | F_{t-1}$ for each $t$ (and no marginal distributional assumptions).

Example 7.5

An alternate specification for count time series with an excess of zeros corresponds to the conditional moment specifications:

$$\mu_t = \frac{(1 - \omega)\lambda_t}{p_t},$$

$$\sigma_t^2 = \frac{(1 - \omega)\lambda_t(1 + r\omega \lambda_t)}{p_t^2},$$

$$\gamma_t = \frac{r\omega \lambda_t (3 - \omega \lambda_t + 2 \omega^2 \lambda_t^2) + 2 - p_t}{[(1 - \omega)\lambda_t]^{1/2}[1 - \omega \lambda_t]^{3/2}},$$

$$\kappa_t = \frac{1}{(1 - \omega)\lambda_t(1 - \omega \lambda_t)^2} \left[ \omega \lambda_t \left( 11 - 4p_t - 6r\lambda_t + r^2 q_t^2 \right) 
+ 6\omega^2 r^2 q_t^2 (2 - r\lambda_t) + 6r^3 \omega^3 q_t^3 + 6 - 6p_t + p_t^2 \right].$$

These are the first four moments of a zero-inflated negative binomial INGARCH process (Zhu, 2012b). Here, $y_t | F_{t-1} \sim ZINB(\lambda_t, \alpha, \omega)$, the conditional mean is again given by (7.5), and $\alpha > 0$ is the dispersion parameter. The conditional ZINB distribution is defined by

$$P(y_t = k | F_{t-1}) = \omega \Delta_{k,0} + (1 - \omega) \frac{\Gamma(k + \lambda_t^1 - c / \alpha)}{k! \Gamma(\lambda_t^1 - c / \alpha)} \left( \frac{1}{1 + \alpha \lambda_t^1 c / \alpha} \right)^{\lambda_t^1 - c / \alpha} \left( \frac{\alpha \lambda_t^1 c / \alpha}{1 + \alpha \lambda_t^1 c / \alpha} \right)^k,$$

for $k = 0, 1, 2, \ldots$, and where $c$ is an index that assumes the values 0 or 1 and identifies the form of the underlying negative binomial distribution. The probability generating function is $G_y(s) = \omega + (1 - \omega) p_t^r / (1 - q_t s)^r$. In comparison to the negative binomial distribution shown in (7.8), we have $p_t = 1/\alpha + \lambda_t^1, q_t = 1 - p_t, \alpha$, and $r = \lambda_t / \alpha$. Note that in the limit as $\alpha \to \infty$, the ZINB-INGARCH model reduces to the ZIP-INGARCH model, and when $\omega = 0$, the model reduces to the NB-INGARCH model.
7.3.3 Models in the GAS Framework

Recently, Creal et al. (2013) proposed a novel observation-driven modeling strategy for time series, that is, the GAS model. Following their approach, we propose an extension of the GAS model for an integer-valued time series \( \{y_t\} \) with specified first four conditional moments, and describe the use of estimating equations. Let \( f_t = f_t(\theta) \) denote a vector-valued time-varying function of an unknown vector-valued parameter \( \theta \), and suppose that the evolution of \( f_t \) is determined by an autoregressive updating equation with an innovation \( s_t \), which is a suitably chosen martingale difference vector:

\[
f_t = \omega + \sum_{i=1}^{p} A_i s_{t-i} + \sum_{j=1}^{Q} B_j f_{t-j}.
\] (7.10)

Suppose that \( \omega = \omega(\theta) \), \( A_i = A_i(\theta) \), and \( B_j = B_j(\theta) \). For instance, \( f_t \) could represent the conditional mean \( \mu_t \) of \( y_t \) or its conditional variance \( \sigma_t^2 \).

Suppose a valid conditional probability distribution \( p(y_t|\mathcal{F}_{t-1}, f_t; \theta) \) is specified (rather than just assuming the form of the first few moments). Let \( s_t \) correspond to the standardized score function, that is, \( s_t = S_t \nabla_t \), where

\[
\nabla_t = \frac{\partial}{\partial f_t} \log p(y_t|\mathcal{F}_{t-1}, f_t; \theta) \quad \text{and} \quad S_t = -E \left[ \frac{\partial}{\partial f_t} \frac{\partial}{\partial f_t} \log p(y_t|\mathcal{F}_{t-1}, f_t; \theta) \right]^{-1}.
\]

Then (7.10) corresponds to the GAS\((P, Q)\) model discussed by Creal et al. (2013). Alternate specifications have been suggested for \( S_t \) that scale \( \nabla_t \) in addition to the inverse information given above. These include the positive square root of the inverse information or the identity matrix.

7.4 Parametric Inference via EFs

In Section 7.4.1, we describe the estimation of the parameter vector \( \theta \) in integer-valued time series models via linear estimating equations and give a recursive scheme for fast optimal estimation. In Section 7.4.2, we describe a combined EF approach based on linear and quadratic martingale differences, and show that these combined EFs are more informative when the conditional mean and variance of the observed process depend on the same parameter.

7.4.1 Linear EFs

Consider the class \( \mathcal{M} \) of all unbiased EFs \( g(m_t(\theta)) \) based on the martingale difference \( m_t(\theta) = y_t - \mu_t(\theta) \). Theorem 7.1 gives the optimal linear estimating equation with corresponding optimal information and the form of an approximate recursive estimator of \( \theta \) which is based on a first order Taylor approximation.
Theorem 7.1  The optimal linear estimating equation and corresponding information are obtained by substituting \( m_t(\theta) = y_t - \mu_t(\theta) \) into (7.2) and (7.3). The recursive estimator for \( \theta \) is given by

\[
\hat{\theta}_t = \hat{\theta}_{t-1} + K_t a^*_{t-1}(\hat{\theta}_{t-1}) g(m_t(\hat{\theta}_{t-1})),
\]

\[
K_t = K_{t-1} \left( I_p - \left[ a^*_{t-1}(\hat{\theta}_{t-1}) \frac{\partial g(m_t(\hat{\theta}_{t-1}))}{\partial \theta} + \frac{\partial a^*_{t-1}(\hat{\theta}_{t-1})}{\partial \theta} g(m_t(\hat{\theta}_{t-1})) \right] K_{t-1} \right)^{-1},
\]

where \( I_p \) is the identity matrix. If \( g(x) = x \), then

\[
a^*_{t-1}(\theta) = \frac{\partial \mu_t(\theta)/\partial \theta}{\text{Var}(g(m_t(\theta))|F_{t-1})},
\]

while for any other function \( g \) (such as the score function),

\[
a^*_{t-1}(\theta) = \frac{[\partial \mu_t(\theta)/\partial \theta][\partial g(m_t(\theta))/\partial m_t(\theta)]}{\text{Var}(g(m_t(\theta))|F_{t-1})}.
\]

The proof is similar to that in Thavaneswaran and Heyde (1999) for the scalar parameter case. \( \square \)

Corollary 7.1 is a special case for the scalar parameter case, where \( a^*_{t-1} \) does not depend on \( \theta \), while Corollary 7.2 discusses a nonlinear time series model.

Corollary 7.1 (Thavaneswaran and Heyde, 1999). For the class \( \mathcal{M} \) of all unbiased EFs \( g(m_t(\theta)) \) based on the martingale difference \( m_t(\theta) = y_t - \mu_t(\theta) \), let \( \mu_t(\theta) \) be differentiable with respect to \( \theta \). The recursive estimator for \( \theta \) is given by

\[
\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{K_{t-1} a^*_{t-1}(\hat{\theta}_{t-1}) g(m_t(\hat{\theta}_{t-1}))}{1 + [\partial \mu_t(\hat{\theta}_{t-1})/\partial \theta] K_{t-1} a^*_{t-1}}, \text{ where}
\]

\[
K_t = \frac{K_{t-1}}{1 + [\partial \mu_t(\hat{\theta}_{t-1})/\partial \theta] K_{t-1} a^*_{t-1}},
\]

and \( a^*_{t-1} \) is a function of the observations. \( \square \)

Corollary 7.2 Consider nonlinear time series models of the form

\[
y_t = \theta f(F_{t-1}) + \sigma(F_{t-1}) \varepsilon_t,
\]

where \( \{\varepsilon_t\} \) is an uncorrelated sequence with zero mean and unit variance and \( f(F_{t-1}) \) denotes a nonlinear function of \( F_{t-1} \), such as \( y_{t-1}^2 \). When \( g(x) = x \) in Theorem 7.1, the recursive estimate based on the optimal linear EF \( \sum_{t=1}^{n} a^*_{t-1}(y_t - \theta f(F_{t-1})) \) is given by (Thavaneswaran and Abraham, 1988)
\[
\hat{\theta}_t = \hat{\theta}_{t-1} + K_t a^*_t[y_t - \hat{\theta}_{t-1} f(F_{t-1})], \text{ where }
K_t = \frac{K_t f(F_{t-1})}{\sigma^2(F_{t-1}) + f^2(F_{t-1}) K_{t-1}}, \tag{7.14}
\]

\[a^*_t = -f(F_{t-1})/\sigma^2(F_{t-1}) \text{ and } K^{-1}_t = \sum_{i=1}^{\infty} a^*_i f(F_{t-1}).\] After some algebra, (7.14) has the following familiar Kalman filter form:

\[
\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{K_{t-1} f(F_{t-1})}{\sigma^2(F_{t-1}) + f^2(F_{t-1}) K_{t-1}} [y_t - \hat{\theta}_{t-1} f(F_{t-1})], \text{ where }
K_t = K_{t-1} - \frac{(K_{t-1} f(F_{t-1}))^2}{\sigma^2(F_{t-1}) + f^2(F_{t-1}) K_{t-1}}. \tag{7.15}
\]

The form of the recursive estimate of the fixed parameter \(\theta\) in (7.14) motivates use of the EF approach for the model in the GAS framework discussed in Section 7.3.3. When the observation at time \(t\) comes in, we update the recursive estimate at time \(t\) as the sum of the estimate at \(t-1\) and the product of the inverse of the information \(K_{t}\) (which is the term \(S_{t}\) in the GAS formulation) and the optimal martingale difference \(a^*_t[y_t - \hat{\theta}_{t-1} f(F_{t-1})]\) (which is \(\gamma_t\) in the GAS formulation).

Estimating equation approaches for recursive estimation of a time-varying parameter have not been discussed in the literature. Consider the simple case where \(\theta_t = \phi \theta_{t-1}\) for a given \(\phi\). The following recursions have a Kalman filter form when the state equation has no error:

\[
\hat{\theta}_t = \phi \hat{\theta}_{t-1} + \frac{\phi K_{t-1} f(F_{t-1})}{\sigma^2(F_{t-1}) + f^2(F_{t-1}) K_{t-1}} [y_t - \hat{\theta}_{t-1} f(F_{t-1})], \text{ where }
K_t = \phi^2 K_{t-1} - \frac{(\phi K_{t-1} f(F_{t-1}))^2}{\sigma^2(F_{t-1}) + f^2(F_{t-1}) K_{t-1}}. \tag{7.16}
\]

### 7.4.2 Combined EFs

To estimate \(\theta\) based on the integer-valued data \(y_1, \ldots, y_n\), consider two classes of martingale differences for \(t = 1, \ldots, n\), viz., \(\{m_t(\theta) = y_t - \mu_t(\theta)\}\) and \(\{M_t(\theta) = m_t(\theta)^2 - \sigma^2_t(\theta)\}\), see Thavaneswaran et al. (2015). The quadratic variations of \(m_t(\theta)\), \(M_t(\theta)\), and the quadratic covariation of \(m_t(\theta)\) and \(M_t(\theta)\) are respectively

\[\langle m\rangle_t = E(m_t^2(\theta)|\mathcal{F}_{t-1}) = \sigma^2_t(\theta),\]
\[\langle M\rangle_t = E(M_t(\theta)^2|\mathcal{F}_{t-1}) = \sigma^4_t(\theta)(\kappa_t(\theta) - 1),\] and
\[\langle m, M\rangle_t = E(m_t(\theta)M_t(\theta)|\mathcal{F}_{t-1}) = \sigma^3_t(\theta)\gamma_t(\theta).\]
The optimal EFs based on the martingale differences $m_t(\theta)$ and $M_t(\theta)$ are respectively

$$g^*_m(\theta) = -\sum_{t=1}^{n} \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{m_t}{\langle m \rangle_t}$$

and

$$g^*_M(\theta) = -\sum_{t=1}^{n} \frac{\partial \sigma^2_t(\theta)}{\partial \theta} \frac{M_t}{\langle M \rangle_t}.$$  

The information associated with $g^*_m(\theta)$ and $g^*_M(\theta)$ are respectively

$$I_{g^*_m}(\theta) = \sum_{t=1}^{n} \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{1}{\langle m \rangle_t}$$

and

$$I_{g^*_M}(\theta) = \sum_{t=1}^{n} \frac{\partial \sigma^2_t(\theta)}{\partial \theta} \frac{\partial \sigma^2_t(\theta)}{\partial \theta} \frac{1}{\langle M \rangle_t}.$$ 

Theorem 7.2 describes the results for combined EFs based on the martingale differences $m_t(\theta)$ and $M_t(\theta)$ and provides the resulting form of the recursive estimator of $\theta$ based on a first-order Taylor approximation (Liang et al., 2011; Thavaneswaran et al., 2015.)

**Theorem 7.2** In the class of all combined EFs

$$G_C = \left\{ g_C(\theta) : g_C(\theta) = \sum_{t=1}^{n} \left[ a_{t-1}(\theta)m_t(\theta) + b_{t-1}(\theta)M_t(\theta) \right] \right\},$$

(a) The optimal EF is given by

$$g^*_C(\theta) = \sum_{t=1}^{n} \left[ a^*_{t-1}(\theta)m_t(\theta) + b^*_{t-1}(\theta)M_t(\theta) \right],$$

where

$$a^*_{t-1} = \rho^2_t \left( \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{1}{\langle m \rangle_t} - \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \sigma^2_t(\theta)}{\partial \theta} \eta_t \right)$$

and

$$b^*_{t-1} = \rho^2_t \left( \frac{\partial \sigma^2_t(\theta)}{\partial \theta} \frac{1}{\langle M \rangle_t} \right),$$

where $\rho^2_t = \left( 1 - \frac{\langle m, M \rangle_t^2}{\langle m \rangle_t^2 \langle M \rangle_t} \right)^{-1}$ and $\eta_t = \frac{\langle m, M \rangle_t}{\langle m \rangle_t \langle M \rangle_t}$. 

© 2016 by Taylor & Francis Group, LLC
(b) The information $I_{\theta c}^*(\theta)$ is given by
\[
I_{\theta c}^*(\theta) = \sum_{t=1}^{n} \rho_t^2 \left( \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} + \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} - \left( \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} + \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} \right) \eta_t \right) ;
\]

(c) The gain in information over the linear EF is
\[
I_{\theta c}^*(\theta) - I_{\theta l}^*(\theta) = \sum_{t=1}^{n} \rho_t^2 \left( \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} + \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} - \left( \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} + \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} \right) \eta_t \right) ;
\]

(d) The gain in information over the quadratic EF is
\[
I_{\theta c}^*(\theta) - I_{\theta q}^*(\theta) = \sum_{t=1}^{n} \rho_t^2 \left( \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} + \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} - \left( \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} + \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} \right) \eta_t \right) ;
\]

(e) The recursive estimate for $\theta$ is given by
\[
\begin{aligned}
\hat{\theta}_t &= \hat{\theta}_{t-1} + K_t \left( a_{t-1}^* \hat{\theta}_{t-1} m_t(\hat{\theta}_{t-1}) + b_{t-1}^* M_t(\hat{\theta}_{t-1}) \right), \\
K_t &= K_{t-1} \left( I_p - \left( a_{t-1}^* \frac{\partial m_t(\hat{\theta}_{t-1})}{\partial \theta'} + \frac{\partial a_{t-1}^*}{\partial \theta} m_t(\hat{\theta}_{t-1}) \right) + \left( b_{t-1}^* \frac{\partial M_t(\hat{\theta}_{t-1})}{\partial \theta'} + \frac{\partial b_{t-1}^*}{\partial \theta} M_t(\hat{\theta}_{t-1}) \right) \right)^{-1},
\end{aligned}
\]

where $I_p$ is the $p \times p$ identity matrix and $a_{t-1}^*$ and $b_{t-1}^*$ can be calculated by substituting $\hat{\theta}_{t-1}$ for $\theta$ in Equations (7.17) and (7.18), respectively;

(f) For the scalar parameter case, the recursive estimate of $\theta$ is given by
\[
\hat{\theta}_t = \hat{\theta}_{t-1} + K_t [a_{t-1}^* \hat{\theta}_{t-1} m_t(\hat{\theta}_{t-1}) + b_{t-1}^* M_t(\hat{\theta}_{t-1})],
\]

where
\[ K_t = K_{t-1} \left( 1 - \left( a_{t-1}^* \frac{\partial m_t(\hat{\theta}_{t-1})}{\partial \theta} + \frac{\partial a_{t-1}^*}{\partial \theta} m_t(\hat{\theta}_{t-1}) \right) \right)^{-1} \]

(7.21)

Since \(-E\left( \frac{\partial g^*_m(\theta)}{\partial \theta} \mid F_{t-1} \right)\) denotes the optimal information matrix based on the first \(t\) observations, it follows that \(K_t^{-1} = -\sum_{s=1}^{t} \frac{\partial g^*_m(\hat{\theta}_{s-1})}{\partial \theta}\) can be interpreted as the observed information matrix associated with the optimal combined EF \(g^*_m(\theta)\). The proof of this theorem is given in Thavaneswaran et al. (2015).

In an interesting recent paper, Fokianos et al. (2009) described estimation for linear and nonlinear autoregression models for Poisson count time series, and used simulation studies to compare conditional least squares estimates with maximum likelihood estimates. Similar to Fisher (1924), we compare the information associated with the corresponding EFs, and show that the optimal EF is more informative than the conditional least squares EF.

In the class of estimating functions of the form \(g_m(\theta) = \sum_{t=1}^{n} a_{t-1}(\theta) m_t(\theta)\), the optimal EF is given by \(g^*_m(\theta) = \sum_{t=1}^{n} a_{t-1}^*(\theta) m_t(\theta)\), where \(a_{t-1}^* = \left(-\frac{\partial \mu(\theta)}{\partial \theta} \frac{1}{|m^*_t|}\right)\). The optimal EF and the conditional least squares EF belong to the class \(g\), and the optimal value of \(a_{t-1}\) is chosen to maximize the information. Hence \(I_{g_{m}^*} - I_{CLS}\) is nonnegative definite. It follows from page 919 of Lindsay (1985) that the optimal estimates are more efficient than the conditional least squares estimates for any class of count time series models.

Note that \(g^*_m = 0\) corresponds in general to a set of nonconvex, nonlinear equations. The formulas for (7.17) through (7.21) may be easily coded as functions in R. For each data/model combination, use of the EF approach in practice requires soft coding of the first four conditional moments, derivatives of the first two conditional moments with respect to model parameters, and specification of initial values to start the recursive estimation.

**Example 7.6**

Consider a zero-inflated regression model. Let \(y_t\) denote a count time series with excess zeros, and assume that the mean, variance, skewness, and kurtosis of \(y_t\) conditional on \(F_{t-1}\) are given by

\[
\begin{align*}
\mu_t(\theta) &= (1 - \omega_t)\lambda_t, \\
\sigma^2_t(\theta) &= (1 - \omega_t)\lambda_t(1 + \omega_t\lambda_t), \\
\gamma_t(\theta) &= \frac{\omega_t(1 + 2\omega_t)\lambda_t^2 + 3\omega_t\lambda_t + 1}{(1 - \omega_t)\lambda_t^{1/2}(1 + \omega_t\lambda_t)^{3/2}}, \text{ and} \\
\kappa_t(\theta) &= \frac{\omega_t(6\omega_t^2 - 6\omega_t + 1)\lambda_t^3 + 6\omega_t(2\omega_t - 1)\lambda_t^2 + 7\omega_t\lambda_t + 1}{(1 - \omega_t)\lambda_t(1 + \omega_t\lambda_t)^2},
\end{align*}
\]

therefore corresponding to the moments of a ZIP(\(\lambda_t, \omega_t\)) distribution with pmf

\[
p(y_t | F_{t-1}) = \begin{cases} 
\omega_t + (1 - \omega_t)\exp(-\lambda_t), & \text{if } y_t = 0, \\
(1 - \omega_t)\exp(-\lambda_t)\lambda_t^{y_t}/y_t!, & \text{if } y_t > 0.
\end{cases}
\]
where $\lambda_t$ is the intensity parameter of the baseline Poisson distribution and $\omega_t$ is the zero-inflation parameter. The ZIP model for count time series is an extension of the Poisson autoregression discussed in Chapter 4 of Kedem and Fokianos (2002).

Suppose that $\lambda_t$ and $\omega_t$ are parametrized by $\lambda_t(\beta)$ and $\omega_t(\delta)$, which are flexible functions of the unknown parameters $\beta$ and $\delta$ and exogenous explanatory variables at time $t - 1$, viz., $x_{t-1}$ and $z_{t-1}$:

$$
\lambda_t(\beta) = \exp(x'_{t-1}\beta) \text{ and } \omega_t(\delta) = \frac{\exp(z'_{t-1}\delta)}{1 + \exp(z'_{t-1}\delta)}.
$$

(7.22)

Let $\theta = (\beta', \delta')'$, which appears in the conditional mean and variance of $y_t$. Let $m_t = y_t - \mu_t, M_t = m_t^2 - \sigma_t^2$, and refer to $\lambda_t(\beta)$ and $\omega_t(\delta)$ by $\lambda_t$ and $\omega_t$, respectively. Then

$$
(m)_t = \lambda_t[1 - \omega_t][1 + \omega_t\lambda_t],
$$

$$
(M)_t = (1 - \omega_t)\lambda_t(4\omega_t^2 - 4\omega_t + 1)\lambda_t^2 + 2\omega_t(4\omega_t - 1)\lambda_t^2 + 5\omega_t\lambda_t + 2),
$$

$$
(m, M)_t = (1 - \omega_t)\lambda_t(\omega_t(1 + 2\omega_t)\lambda_t^2 + 3\omega_t\lambda_t + 1).
$$

Also,

$$
\frac{\partial \mu_t}{\partial \theta} = \left(\begin{array}{c}
(1 - \omega_t) \frac{\partial \lambda_t}{\partial \beta}

- \lambda_t \frac{\partial \omega_t}{\partial \delta}
\end{array}\right)
$$

and

$$
\frac{\partial \sigma_t^2}{\partial \theta} = \left(\begin{array}{c}
(1 - \omega_t)(1 + 2\omega_t\lambda_t) \frac{\partial \lambda_t}{\partial \beta}

\lambda_t(\lambda_t(1 - 2\omega_t) - 1) \frac{\partial \omega_t}{\partial \delta}
\end{array}\right).
$$

The combined optimal EF based on $m_t$ and $M_t$ is given by

$$
g^*(\theta) = \sum_{t=1}^{n} (a^*_{t-1} m_t + b^*_{t-1} M_t),
$$

where

$$
a^*_{t-1} = \rho_t^2 \left(\begin{array}{c}
-1 - \omega_t \frac{\partial \lambda_t}{\partial \beta}

\lambda_t \frac{\partial \omega_t}{\partial \delta}
\end{array}\right) \frac{1}{m}_t + \left(\begin{array}{c}
(1 - \omega_t)(1 + 2\omega_t\lambda_t) \frac{\partial \lambda_t}{\partial \beta}

\lambda_t(\lambda_t(1 - 2\omega_t) - 1) \frac{\partial \omega_t}{\partial \delta}
\end{array}\right) \eta_t,
$$

$$
b^*_{t-1} = \rho_t^2 \left(\begin{array}{c}
-1 - \omega_t \frac{\partial \lambda_t}{\partial \beta}

\lambda_t \frac{\partial \omega_t}{\partial \delta}
\end{array}\right) \eta_t - \left(\begin{array}{c}
(1 - \omega_t)(1 + 2\omega_t\lambda_t) \frac{\partial \lambda_t}{\partial \beta}

- \lambda_t(\lambda_t(1 - 2\omega_t) - 1) \frac{\partial \omega_t}{\partial \delta}
\end{array}\right) \frac{1}{M}_t.
$$

The corresponding information matrix is given by

$$
\Gamma^*_C(\theta) = \Gamma^*_m(\theta) + \Gamma^*_M(\theta)
$$

$$
= -\sum_{t=1}^{n} \rho_t^2 \left(\begin{array}{ccc}
2(1 - \omega_t)^2(1 + 2\omega_t\lambda_t) \left(\frac{\partial \lambda_t}{\partial \beta}\right)^2

\lambda_t(1 - \omega_t)(\lambda_t(1 - 4\omega_t) - 2) \left(\frac{\partial \lambda_t}{\partial \beta}\right) \left(\frac{\partial \omega_t}{\partial \delta}\right)

-2\lambda_t^2(\lambda_t(1 - 2\omega_t) - 1) \left(\frac{\partial \omega_t}{\partial \delta}\right)^2
\end{array}\right).
$$

© 2016 by Taylor & Francis Group, LLC
The recursive estimator for $\theta$ follows from (7.19) and (7.20) as

$$\widehat{\theta}_t = \widehat{\theta}_{t-1} + K_t \left( \frac{\frac{\partial \lambda_t}{\partial \beta} \lambda_t(\widehat{\delta}_{t-1})}{\lambda_t(\lambda_t(1+\omega_t(\widehat{\delta}_{t-1})\lambda_t(\widehat{\beta}_{t-1}))} \right),$$

and

$$K_t = K_{t-1} \left( \begin{pmatrix} 1 & \frac{\lambda_t}{\lambda_t(1+\omega_t\lambda_t)} \frac{\lambda_t}{\lambda_t(1+\omega_t\lambda_t)}(1+\omega_t\lambda_t) \\ -\frac{\partial \lambda_t}{\partial \beta} \frac{\lambda_t}{\lambda_t(1+\omega_t\lambda_t)} & -\frac{\partial \lambda_t}{\partial \beta} \frac{\lambda_t}{\lambda_t(1+\omega_t\lambda_t)}(1+\omega_t\lambda_t) \end{pmatrix} \right)^{-1}.$$

For this example, it is also straightforward to show that the linear optimal EF and the corresponding optimal coefficient are given by

$$g^*_t(\theta) = \sum_{s=1}^{n} \left( -\frac{\lambda_t}{\lambda_t(1+\omega_t\lambda_t)} \right) m_t, \quad \text{and}$$

$$a^*_{t-1}(\theta) = \begin{pmatrix} -\frac{\lambda_t}{\lambda_t(1+\omega_t\lambda_t)} \\ \frac{\lambda_t}{\lambda_t(1+\omega_t\lambda_t)} \end{pmatrix}.$$

**Example 7.7**

Consider an extended GAS($P, Q$) model. As discussed in Section 7.3.3, suppose $\{y_t\}$ is an integer-valued time series, its first four conditional moments given $F_{t-1}$ are available, and $f_t$ is modeled by (7.10), $s_t$ being a suitably chosen martingale difference. Suppose the time-varying parameter $f_t$ corresponds to the conditional mean $\mu_t(\theta) = E(y_t|F_{t-1})$. Following the discussion in Theorem 7.1, it is natural to choose $s_t$ as $a^*_{t-1}(\theta)m_t(\theta)$. Suppose instead that $f_t$ corresponds to the conditional variance $\sigma^2_t(\theta) = Var(y_t|F_{t-1})$; a natural choice of $s_t$ is $b^*_{t-1}(\theta)M_t(\theta)$. When $f_t$ is modeled by (7.10), the most informative innovation is given by

$$s_t = K_t \left( a^*_{t-1}(\widehat{\theta}_{t-1})m_t(\widehat{\theta}_{t-1}) + b^*_{t-1}(\widehat{\theta}_{t-1})M_t(\widehat{\theta}_{t-1}) \right),$$

where $K_t$ is defined in (7.20) and $a^*_{t-1}$ and $b^*_{t-1}$ can be calculated by substituting $\widehat{\theta}_{t-1}$ in equations (7.17) and (7.18) respectively for the fixed parameter $\theta$.

When the form of the conditional distribution of $y_t$ given $F_{t-1}$ is available, and the score function is easy to obtain, then the optimal choice for $V_t$ is the score function. However, in situations where we do not wish to assume an explicit form for the conditional distribution, the optimal choice for $V_t$ is given by components of the optimal linear, or quadratic, or combined EFs.
Consider the nonlinear time series model in (7.13). Based on the optimal EF in the class $G$ of all unbiased EFs $g = \sum_{t=1}^{n} b_{t-1} g(y_t - \mu(t_{t-1}, \mathcal{F}_{t-1}))$, the form of the extended GAS model for the location parameter $\theta_t$ is given by

$$
\theta_t = \phi \theta_{t-1} + \frac{\phi K_{t-1} b_{t-1}^*}{1 + [\partial \mu(\hat{\theta}_{t-1}, \mathcal{F}_{t-1})/\partial \hat{\theta}] K_{t-1} b_{t-1}^*} g(y_t - \mu(\hat{\theta}_{t-1}, \mathcal{F}_{t-1})),
$$

where

$$
K_t = \frac{K_{t-1}}{1 + [\partial \mu(\hat{\theta}_{t-1}, \mathcal{F}_{t-1})/\partial \hat{\theta}] \phi^2 K_{t-1} b_{t-1}^*},
$$

where $b_{t-1}^*$ is a function of $g, \theta$, and the observations:

$$
b_{t-1}^* = \frac{[\partial \mu(\hat{\theta}_{t-1}, \mathcal{F}_{t-1})/\partial \hat{\theta}] [\partial g(\hat{\theta}_{t-1}, \mathcal{F}_{t-1})/\partial \mu]}{\text{Var}(g|\mathcal{F}_{t-1})}.
$$

### 7.5 Hypothesis Testing and Model Choice

Hypothesis testing situations in stochastic modeling are often tests of linear hypotheses about the unknown parameter $\theta \in \mathbb{R}^q$. Suppose $\theta = (\theta_1, \theta_2)$, and suppose that the optimal EF $g_n^*(\theta)$ and the corresponding unconditional information $F_n(\theta) = E[g_n^*(\theta) g_n^*(\theta)']$ have conformable partitions, that is,

$$
g_n^*(\theta) = \begin{pmatrix}
g_n^{*1}(
) \\
g_n^{*2}(
)
\end{pmatrix},
F_n(\theta) = \begin{pmatrix}
F_{n11}(\theta) & F_{n12}(\theta) \\
F_{n21}(\theta) & F_{n22}(\theta)
\end{pmatrix}.
$$

A test of $H_0 : \Theta_2 = \theta_{20}$ versus $H_1 : \Theta_2 \neq \theta_{20}$ corresponds to a comparison of a full model versus a nested model when we test $\theta_{20} = 0$. For example, in (7.5), testing $\theta_{20} = 0$ could correspond to testing $\beta = 0$, so that $H_0$ corresponds to a smaller model with only $\delta$ and $\alpha$ as parameters.

Let $\theta^*_n = (\theta_{21}^*, \theta_{22}^*)$ denote the optimal estimate of $\theta$ unrestricted by $H_0$, and let $\tilde{\theta}_n = (\tilde{\theta}_{21}, \tilde{\theta}_{22})$ denote the optimal estimate under the null hypothesis $H_0$. As in Thavaneswaran (1991), we propose two test statistics, viz., the Wald-type statistic and the score statistic, as

$$
W_n = (\theta_{21}^* - \theta_{20})' A_{n22}(\theta_{22}^*) (\theta_{22}^* - \theta_{20}),
$$

$$
Q_n = (g_n^*(\tilde{\theta}_n))' A_{n22}(\tilde{\theta}_n) g_n^*(\tilde{\theta}_n),
$$

where

$$
A_{n22}(\theta) = F_{n22}(\theta) - F_{n21}(\theta) F_{n11}(\theta)^{-1} F_{n12}(\theta)
$$

is the inverse of the second diagonal block in the inverse of the partitioned matrix $F_n(\theta)$.

The Wald and score statistics for testing a general linear hypothesis $H_0 : C\theta = c_0$ versus $H_1 : C\theta \neq c_0$, where the $r \times p$ matrix $C$ has full row rank, are

$$
\tilde{W}_n = (C \theta_n^* - c_0)' (C F_n^{-1}(\theta_n^*) C')^{-1} (C \theta_n^* - c_0) \quad \text{and}
$$

$$
\tilde{Q}_n = (g_n^*(\tilde{\theta}_n))' F_n^{-1}(\tilde{\theta}_n) g_n^*(\tilde{\theta}_n),
$$

where

$$
\tilde{\theta}_n = \hat{\theta}_n - \theta_{20}.
$$
Thavaneswaran (1991) showed that under certain regularity conditions, the test statistics in (7.26) and (7.27) are asymptotically equivalent, that is, the difference between them converges to zero in probability under $H_0$ and they have the same limiting null distributions (which is a $\chi^2_r$ distribution).

### 7.6 Discussion and Summary

Interest in developing models for integer valued time series, especially for count time series, is growing. Among these are models discussed in Ferland et al. (2006) and Zhu (2011, 2012a,b), who described classes of INGARCH models with different conditional distributional specifications for the process given its history, and primarily described likelihood based approaches for estimating model parameters, under parametric assumptions such as Poisson, negative binomial, or ZIP for the conditional distributions. Although these models are referred to as INGARCH models in the literature, they model the conditional mean of the time series and not its conditional variance. These models are similar to the ACP models discussed in Heinen (2003) and Ghahramani and Thavaneswaran (2009b). Creal et al. (2013) recently discussed GAS models, while Thavaneswaran and Ravishanker (2015) described models for circular time series.

This chapter considers modeling the conditional mean and conditional variance of an integer-valued time series $\{y_t\}$, where conditional moments are functions of $\theta$. We have described a combined EF approach for estimating $\theta$ and have also provided forms for joint recursive estimates for fixed parameters using the most informative combined martingale difference and provided its corresponding information. In Section 7.4.1, we have shown how recursive estimation extends to the case with a time-varying parameter. This approach would be valuable in a study of doubly stochastic models for integer-valued time series, which we briefly discuss below.

Similar to the well-known stochastic volatility (SV) or stochastic conditional duration (SCD) models described in the literature, a general integer-valued doubly stochastic model for $y_t$ with conditional mean $E(y_t|\mathcal{F}_{t-1}) = \mu_t = \exp(\lambda_t)$ is defined via

$$\lambda_t = \delta + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (7.28)$$

where $\delta$ is a real-valued parameter, $\sum_{j=0}^{\infty} \psi_j^2 \varepsilon_{t-j} < \infty$, $\varepsilon_t|\mathcal{F}_{t-1}$ are independent $N(0, \sigma^2_\varepsilon)$ variables, and $\varepsilon_s$ is independent of $y_t|\mathcal{F}_{t-1}$ for all $s, t$. The conditional moments of $y_t$ may match the moments of a known probability distribution for a count random variable, for example, those corresponding to a Poisson-INDS model, a GP-INDS model, or a ZIP-INDS model. In lieu of (7.28), if $\lambda_t$ is modeled by $\lambda_t - \delta = (1 - B)^{-d} \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ where $d \in (0,0.5)$ and $\psi_k = \Gamma(k + d)/\Gamma(d)\Gamma(k + 1)$, where $\Gamma(\cdot)$ is the gamma function, the model can handle long-memory behavior. We may also define an integer-valued quadratic doubly stochastic (INQDS) model by assuming the first four conditional moments of $y_t$ given $\mathcal{F}_{t-1}$ to be $\mu_t = \exp(a\lambda_t + b\lambda_t^2)$, $\sigma^2_t = \mu_t$, $\gamma_t = \mu_t^{-1/2}$ and $\kappa_t = \mu_t^{-1}$, which match the first four moments of a Poisson($\exp(a\lambda_t + b\lambda_t^2)$) process. Naik-Nimbalkar and Rajarshi (1995) and Thompson and Thavaneswaran (1999) studied filtering/estimation for state space
models and counting processes in the context of EFs. Thavaneswaran and Abraham (1988) and Thavaneswaran et al. (2015) described combining nonorthogonal EFs following prefiltered estimation.

Acknowledgments
The first author acknowledges support from an NSERC grant.

References


