Ann Inst Stat Math (2015) 67:129–156
DOI 10.1007/s10463-013-0442-9

Generalized duration models and optimal estimation using estimating functions

Aerambamoorthy Thavaneswaran · Nalini Ravishanker · You Liang

Received: 27 June 2013 / Revised: 27 June 2013 / Published online: 3 January 2014

Abstract This article introduces a class of generalized duration models and shows that the autoregressive conditional duration (ACD) models and stochastic conditional duration (SCD) models discussed in the literature are special cases. The martingale estimating functions approach, which provides a convenient framework for deriving optimal inference for nonlinear time series models, is described. It is shown that when the first two conditional moments are functions of the same parameter, and information about higher order conditional moments of the observed duration process become available, combined estimating functions are optimal and are more informative than component estimating functions. The combined estimating functions approach is illustrated on three classes of generalized duration models, viz., multiplicative random coefficient ACD models, random coefficient models with ACD errors, and log-SCD models. Recursive estimation of model parameters based on combined estimating functions provides a mechanism for fast estimation in the general case, and is illustrated using simulated data sets.

Keywords ACD models · Combined estimating functions · Generalized martingale differences · Quadratic log-SCD models · Random coefficients · Recursive estimates

A. Thavaneswaran · Y. Liang
Department of Statistics, University of Manitoba, 338 Machray Hall, Winnipeg, MB R3T 2N2, Canada
e-mail: Aerambamoorthy.Thavaneswaran@ad.umanitoba.ca

Y. Liang
e-mail: umlian33@cc.umanitoba.ca

N. Ravishanker (✉)
Department of Statistics, University of Connecticut, 215 Glenbrook Road, Storrs, CT 06269, USA
e-mail: nalini.ravishanker@uconn.edu

1 Introduction

There has been growing interest in the statistical analysis of random durations between events. Let t_i be the time until the ith event with t_0 being the starting time. The ith duration, defined as the time interval between two consecutive events, is denoted by

$$x_i = t_i - t_{i-1}, \quad i = 1, 2, \ldots$$

and is a non-negative random variable for each positive integer i. Statistical models used to analyze durations between market events, such as trades, have become extremely popular in finance, for the analysis of intra-day financial data such as transaction and quote data which are increasingly often being provided by several stock exchanges. There is also an increasing number of applications of such models for understanding patterns in durations between economic events, health-policy events, etc. Inference for duration analysis has developed along several directions, for example, using likelihood and quasi-likelihood approaches.

Engle and Russell (1998) proposed the autoregressive conditional duration (ACD) model to study the dynamic structure of durations in the context of irregularly spaced financial transactions data. This model shares similarities with the generalized autoregressive conditional heteroscedastic (GARCH) models. The crucial assumption underlying the ACD model is that the time dependence in the durations is described through a function ψ_i, which is the conditional expectation of the duration x_i between the $(i-1)$th and ith events. The ACD model has the form

$$\begin{aligned} x_i &= \psi_i \varepsilon_i, \\ \psi_i &= E[x_i | \mathcal{F}_{i-1}^x] = \omega + \alpha x_{i-1} + \beta \psi_{i-1}, \end{aligned} \tag{1}$$

where $\omega > 0, \alpha \geq 0, \beta \geq 0, \alpha + \beta < 1$, and ε_i are i.i.d. non-negative random variables with $E(\varepsilon_i) = 1$ and density function $f_\varepsilon(.)$, while $\mathcal{F}_{i-1}^x$ denotes the σ-field generated by $x_1, x_2, \ldots, x_{i-1}$, assumed to be independent of ε_i. Different specifications of $f_\varepsilon(\cdot)$ with unit mean for the non-negative random variables ε_i yield different ACD models. Under the more general ACD(p, q) model, ψ_i takes the form

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}, \tag{2}$$

where the conditions on the model parameters, viz., $\omega > 0$, $\alpha_j \geq 0$ for $j = 1, \ldots, p$, $\beta_j \geq 0$ for $j = 1, \ldots, q$ and $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j < 1$, ensure positive conditional durations which are weakly stationary and the existence of the marginal mean of x_i.

The model (1) has been extended along several directions. Jasiak (1998) analyzed the persistence of inter-trade durations using the fractionally integrated ACD (FIACD) model and showed that the autocorrelation function of the durations can show a slow, hyperbolic rate of decay which is typical of long memory processes. Bauwens and Giot (2000) developed a logarithmic ACD model which avoids positivity restrictions on the parameters and is, therefore, more flexible in terms of accommodating exogenous variables. Bauwens and Giot (2003) considered an asymmetric ACD model where the dynamics of the duration process depend on the state of the price process. Pacurar (2008) has given an excellent survey of the theoretical and empirical literature on ACD models.

The stochastic conditional duration (SCD) model has also received considerable attention for modeling durations. Consider the SCD model

$$x_i = \exp(\psi_i)\varepsilon_i,$$
$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + z_i, \tag{3}$$

where $z_i|\mathcal{F}_{i-1}^x$ are i.i.d. $N(0, \sigma_z^2)$ variables, $\varepsilon_i|\mathcal{F}_{i-1}^x$ follows a distribution with positive support, and z_i's are distributed independently of $\varepsilon_j|\mathcal{F}_{i-1}^x$ for all i, j, and $\mathcal{F}_{i-1}^c$ denotes the information set at the end of the $(i-1)$th duration, including past values of x_i and ψ_i. This model is a generalization of the model proposed in Bauwens and Veredas (2004); Bauwens et al. (2008), and is a special case of the model proposed by Thavaneswaran and Ghahramani (2011). While the SCD model has a multiplicative specification similar to the ACD model, it differs from the latter because it is a doubly stochastic process. The conditional expected duration, which was a fixed function of unknown parameters under the ACD model, is assumed to be a random variable under the SCD model, as in state space models. Thavaneswaran and Ghahramani (2011) have also proposed quadratic SCD and long memory SCD models, and studied their moment properties. For recent references on applications to modeling durations, see Pacurar (2008) and references therein, as well as Allen et al. (2008).

One of the main difficulties in estimation, especially with SCD models, lies in the evaluation of the likelihood function for carrying out parametric inference, because the latent variable must be integrated out. This can of course be performed using computer-intensive simulation methods, or naturally using a Bayesian framework. For fast estimation, especially with long time series, it is useful to employ other methods that are less demanding in terms of computing time and do not need to evaluate the exact likelihood function. In the literature, the quasi-maximum likelihood (QML) approach and the generalized method of moments (GMM) approach have been studied in the context of stochastic volatility models (Ruiz 1994; Jacquier et al. 1994). Overall, the estimation approaches used in the literature are based on a transformation of the nonlinear time series model into a linear state space representation and application of the Gaussian Kalman filter (see Broto and Ruiz 2004 for a review). One major drawback is that these approaches neither address the information associated with the corresponding estimating functions nor the efficiency of the resulting estimates. Nevertheless, there are some interesting duration models, such as the log-SCD models with non-normal errors which are proposed in Sect. 2, for which the approach of using the Kalman filter on the linearized model cannot be applied to construct the estimates. The combined estimating functions approach based on generalized martingale estimating functions is discussed in this article and provides a viable approach for optimal estimation in such problems, also providing the information associated with the corresponding optimal estimating functions.

Godambe (1985) first studied inference for discrete-time stochastic processes using the estimating functions approach. Thavaneswaran and Abraham (1988) described estimation for nonlinear time series models using linear estimating functions. Naik-Nimbalkar and Rajarshi (1995) and Thavaneswaran and Heyde (1999) studied problems in filtering and prediction using linear estimating functions in the Bayesian context. Merkouris (2007), Ghahramani and Thavaneswaran (2009), Ghahramani and Thavaneswaran (2012), and Thavaneswaran et al. (2012), among others, have studied estimation problems for time series via the estimating functions approach. Bera et al. (2006) give a review of the historical use of estimating equations in economic applications. Here, we describe combined estimating functions based on linear and *generalized* martingale differences for the duration models and show that the combined estimating functions are more informative when the conditional mean and variance of the observed process depend on the same parameter of interest.

We then describe the use of optimal combined estimating functions for carrying out parameter estimation for generalized duration models. Inference for the parametric basic ACD model of Engle and Russell is shown to be a special case. We also provide *generalized* recursive estimates based on combined estimating functions. We study the estimation for log-SCD models using the extended version of the prefiltered estimation method discussed in Thavaneswaran and Abraham (1988).

This paper is organized as follows: Section 2 presents the generalized durations model and shows that ACD and SCD models discussed in the literature are special cases. Section 3 describes optimal combined estimation for the class of generalized duration models and provides detailed illustrations for three models, viz., multiplicative random coefficient ACD models, random coefficient models with ACD errors, and log-SCD models. Section 4 presents simulation studies to illustrate the recursive estimation approach for the duration models. Section 5 provides a discussion and summary. Proofs of theorems from Sect. 3 are given in the Appendix.

2 Generalized duration models

We introduce a new class of generalized duration models which includes the various ACD and SCD models proposed in the literature as special cases. The generalized duration model has the form

$$x_i = \widetilde{h}(\mathcal{F}_{i-1}^x, \psi_i, z_i)\varepsilon_i, \tag{4}$$

where $\{\psi_i\}$ is the conditional mean of x_i given the information set $\mathcal{F}_{i-1}^x$ and $\{z_i\}$ is a random process independent of the history $\mathcal{F}_{i-1}^x$. The ACD models in (1) and (2), and the SCD model in (3) are seen to be special cases of (4) by assigning $\widetilde{h}(\mathcal{F}_{i-1}^x, \psi_i, z_i)$ to be, respectively ψ_i and $\exp(z_i)$. We discuss below other popular duration models proposed in the literature, and in each case, we provide the forms of the first four conditional moments of the process, which enable construction of the estimating equations.

2.1 Long memory ACD (p, d, q) and SCD (p, d, q) models

The long memory ACD(p, d, q) model, which was referred to as the FIACD model by Jasiak (1998), has the form

$$
\begin{aligned}
(1-B)^d x_i &= \psi_i \varepsilon_i, \\
\psi_i &= \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j},
\end{aligned}
$$

where $\omega > 0, \alpha_j > 0$ for $j = 1, \ldots, p, \beta_j > 0$ $j = 1, \ldots, q, \sum_{j=1}^{\max(p,\,q)}(\alpha_j + \beta_j) < 1$, B denotes the backshift operator, and $d \in (-0.5, 0.5)$ is the fractional differencing parameter such that

$$
(1-B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} (-1)^k B^k.
$$

Let $\boldsymbol{\theta} = (d, \omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$ be the vector of unknown model parameters. Here, and in the rest of this paper unless stated otherwise, we assume that ε_i are i.i.d. non-negative random variables with mean μ_ε, variance σ_ε^2, third central moment γ_ε, and fourth central moment κ_ε. This model is seen to be a special case of the general model (4) by assigning $\widetilde{h}(\mathcal{F}_{i-1}^x, \psi_i, z_i) = (1-B)^{-d}\psi_i$. The central moments of $\{x_i\}$ conditional on $\mathcal{F}_{i-1}^x$ are $\mu_i = \mu_\varepsilon(1-B)^{-d}\psi_i$, $\sigma_i^2 = \sigma_\varepsilon^2(1-B)^{-2d}\psi_i^2$, $\gamma_i = \gamma_\varepsilon(1-B)^{-3d}\psi_i^3$, and $\kappa_i = \kappa_\varepsilon(1-B)^{-4d}\psi_i^4$. The ACD$(p, q)$ model belongs to the class of long memory ACD(p, d, q) models when $d = 0$. The corresponding class of long memory SCD (p, d, q) models, with $\widetilde{h}(\mathcal{F}_{i-1}^x, \psi_i, z_i) = \exp(\psi_i)$, is defined as

$$
\begin{aligned}
(1-B)^d x_i &= \exp(\psi_i)\varepsilon_i, \\
\psi_i &= \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + z_i,
\end{aligned}
$$

where the latent process $\{z_i\}$ was defined under (3), and when $d = 0$, this model corresponds to (3).

2.2 Log-ACD and log-SCD models

The log-ACD$_1$ model has the form discussed in Pacurar (2008) as an extension of the model proposed in Bauwens and Giot (2000):

$$
\begin{aligned}
x_i &= \exp(\psi_i)\varepsilon_i, \\
\psi_i &= \omega + \sum_{j=1}^{p} \alpha_j \log x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j},
\end{aligned}
$$

where $\sum_{j=1}^{\max(p,\, q)} (\alpha_j + \beta_j) < 1$. For the log-ACD$_2$ model, ψ_i takes the form

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \frac{x_{i-j}}{\exp(\psi_{i-j})} + \sum_{j=1}^{q} \beta_j \psi_{i-j},$$

where $\sum_{j=1}^{q} \beta_j < 1$. This model is seen to be a special case of the general model (4) by assigning $\widetilde{h}(\mathcal{F}_{i-1}^x, \psi_i, z_i) = \exp(\psi_i)$. The conditional central moments of $\{x_i\}$ are derived as $\mu_i = \mu_\varepsilon \exp(\psi_i)$, $\sigma_i^2 = \sigma_\varepsilon^2 \exp(2\psi_i)$, $\gamma_i = \gamma_\varepsilon \exp(3\psi_i)$, and $\kappa_i = \kappa_\varepsilon \exp(4\psi_i)$, with the ψ_i corresponding to each of the two model forms. The unknown parameters are $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$.

For the corresponding log-SCD$_1$ and log-SCD$_2$ models, ψ_i takes the respective forms

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \log x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + z_i,$$

and

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \frac{x_{i-j}}{\exp(\psi_{i-j})} + \sum_{j=1}^{q} \beta_j \psi_{i-j} + z_i.$$

Here, z_i are i.i.d. random variables with mean 0 and variance σ_z^2, and are assumed to be mutually uncorrelated with ε_i. The log-SCD model is again a special case of (4) by defining $\widetilde{h}(\mathcal{F}_{i-1}^x, \psi_i, z_i) = \exp(\psi_i)$.

Moreover, if we define $x_i = \exp(\alpha\psi_i + \beta\psi_i^2)\varepsilon_i$, we obtain the quadratic log-SCD model (see Thavaneswaran and Ghahramani 2011), which is seen to be a special case of the general model (4) by defining $\widetilde{h}(\mathcal{F}_{i-1}^x, \psi_i, z_i) = \exp(\alpha z_i + \beta z_i^2)$.

2.3 Augmented ACD and SCD models

Augmented ACD models were introduced by Fernandes and Grammig (2006) to allow asymmetric responses to small and large shocks:

$$\begin{aligned} x_i &= \psi_i \varepsilon_i, \\ \psi_i^\lambda &= \omega + \alpha \psi_{i-1}^\lambda [|\varepsilon_{i-1} - b| + c(\varepsilon_i - b)]^\nu + \beta \psi_{i-1}^\lambda, \end{aligned}$$

where $\omega > 0$, $\alpha > 0$ and $\beta > 0$. The parameter λ determines whether the Box–Cox transformation is concave ($\lambda \leq 1$) or convex ($\lambda \leq 1$). The shift parameter b enables identification of the asymmetric response implied by the shocks impact curve. The parameter c determines whether the rotation is clockwise ($c < 0$) or counterclockwise ($c > 0$). The parameter ν induces concavity ($\nu \leq 1$) or convexity ($\nu \geq 1$) to the shock impact curve. This model is seen to be a special case of the general model (4) by assigning $\widetilde{h}(\mathcal{F}_{i-1}^x, \psi_i, z_i) = \psi_i$, and the unknown parameters are specified by $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$. The forms of the first four conditional central moments are given by

$\mu_i = \mu_\varepsilon(\psi_i^\lambda)^{1/\lambda}$, $\sigma_i^2 = \sigma_\varepsilon^2(\psi_i^\lambda)^{2/\lambda}$, $\gamma_i = \gamma_\varepsilon(\psi_i^\lambda)^{3/\lambda}$, and $\kappa_i = \kappa_\varepsilon(\psi_i^\lambda)^{4/\lambda}$. We define the corresponding augmented SCD models as

$$\begin{aligned} x_i &= \exp(\psi_i)\varepsilon_i, \\ \psi_i^\lambda &= \omega + \alpha\psi_{i-1}^\lambda[|\varepsilon_{i-1} - b| + c(\varepsilon_i - b)]^\nu + \beta\psi_{i-1}^\lambda + z_i. \end{aligned}$$

2.4 Multiplicative random coefficient ACD models

In analogy with random coefficient autoregressive (RCA) models, we introduce a new class of models, called the multiplicative random coefficient ACD (p, q) models of the form

$$\begin{aligned} x_i &= (z_i + \psi_i)\varepsilon_i, \\ \psi_i &= \omega + \sum_{j=1}^{p}\alpha_j x_{i-j} + \sum_{j=1}^{q}\beta_j\psi_{i-j}, \end{aligned} \tag{5}$$

where $\omega > 0$, $\alpha_j > 0$, $\beta_j > 0$, and $\sum_{j=1}^{\max(p,q)}(\alpha_j + \beta_j) < 1$. We assume that z_i are i.i.d. non-negative random variables with mean $\mu_z(\boldsymbol{\phi})$, variance $\sigma_z^2(\boldsymbol{\phi})$, third central moment $\gamma_z(\boldsymbol{\phi})$, and fourth central moment $\kappa_z(\boldsymbol{\phi})$, where $\boldsymbol{\phi} \in \mathcal{R}^l$. We also assume that ε_i and z_i are mutually independent. This model is a special case of model (4) by letting $\widetilde{h}(\mathcal{F}_{i-1}^x, \psi_i, z_i) = z_i + \psi_i$. For this model, the conditional moments of $\{x_i\}$ can be calculated as

$$\begin{aligned} \mu_i &= \mu_\varepsilon(\mu_z + \psi_i), \\ \sigma_i^2 &= \sigma_\varepsilon^2(\mu_z + \psi_i)^2 + \sigma_z^2(\mu_\varepsilon^2 + \sigma_\varepsilon^2), \\ \gamma_i &= \gamma_\varepsilon(\mu_z + \psi_i)^3 + (3\sigma_z^2\gamma_\varepsilon + 6\mu_\varepsilon\sigma_z^2\sigma_\varepsilon^2)(\mu_z + \psi_i) + \gamma_z(\mu_\varepsilon^3 + 3\mu_\varepsilon\sigma_\varepsilon^2 + \gamma_\varepsilon), \\ \kappa_i &= \kappa_\varepsilon(\mu_z + \psi_i)^4 + 6\sigma_z^2(\mu_\varepsilon^2\sigma_\varepsilon^2 + 2\mu_\varepsilon\gamma_\varepsilon + \kappa_\varepsilon)(\mu_z + \psi_i)^2 \\ &\quad + 4\gamma_z(3\mu_\varepsilon^2\sigma_z^2 + 3\mu_\varepsilon\gamma_\varepsilon + \kappa_\varepsilon)(\mu_z + \psi_i) + \kappa_z(\mu_\varepsilon^4 + 4\mu_\varepsilon\gamma_\varepsilon + 6\mu_\varepsilon^2\sigma_\varepsilon^2 + \kappa_\varepsilon). \end{aligned}$$

In Sect. 3.1, we describe the estimating functions approach for estimation of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\phi}', \omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$.

2.5 Random coefficient autoregressive models with ACD errors

We define the RCA model with ACD errors as

$$\begin{aligned} x_i &= (\theta + z_i)x_{i-1} + \psi_i\varepsilon_i \\ \psi_i &= \omega + \sum_{j=1}^{p}\alpha_j x_{i-j} + \sum_{j=1}^{q}\beta_j\psi_{i-j}, \end{aligned} \tag{6}$$

where θ is a non-negative unknown parameter, $\{z_i\}$ is a sequence of i.i.d. non-negative random variables with four central moments $\mu_z(\boldsymbol{\phi})$, $\sigma_z^2(\boldsymbol{\phi})$, $\gamma_z(\boldsymbol{\phi})$, and $\kappa_z(\boldsymbol{\phi})$ which depend on an unknown real parameter vector $\boldsymbol{\phi} \in \mathcal{R}^l$. We assume that $\{\varepsilon_i\}$ is a non-negative process with first four central moments $\mu_\varepsilon(\theta)$, $\sigma_\varepsilon^2(\theta)$, $\gamma_\varepsilon(\theta)$, and $\kappa_\varepsilon(\theta)$. Moreover, $\{b_i\}$ and $\{\varepsilon_i\}$ are assumed to be mutually uncorrelated. This model is a special case of the vector case of the general model (4) by letting $\widetilde{\mathbf{h}}(\mathcal{F}_{i-1}^x, \psi_i, z_i) = ((\theta + z_i)x_{i-1}, \psi_i)'$ and $\varepsilon_i = (1, \varepsilon_i)'$, where $\boldsymbol{\theta} = (\theta, \boldsymbol{\phi}', \omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$ is the vector of model parameters. In this model, the first four conditional central moments of the process $\{x_i\}$ are

$$
\begin{aligned}
\mu_i &= (\theta + \mu_z(\phi))x_{i-1} + \mu_\varepsilon(\theta)\psi_i,\\
\sigma_i^2 &= \sigma_z^2(\phi)x_{i-1}^2 + \sigma_\varepsilon^2(\theta)\psi_i^2,\\
\gamma_i &= \gamma_z(\phi)x_{i-1}^3 + \gamma_\varepsilon(\theta)\psi_i^3,\\
\kappa_i &= \kappa_z(\phi)x_{i-1}^4 + \kappa_\varepsilon(\theta)\psi_i^4 + 6\sigma_z^2(\phi)\sigma_\varepsilon^2(\theta)x_{i-1}^2\psi_i^2.
\end{aligned}
$$

3 Combined estimating functions

Suppose that $\{x_i, i = 1, \ldots, n\}$ is a realization of a discrete-time stochastic process and its distribution depends on a vector parameter $\boldsymbol{\theta}$ belonging to an open subset $\boldsymbol{\Theta}$ of the P-dimensional Euclidean space. Let $(\Omega, \mathcal{F}, P_{\boldsymbol{\theta}})$ denote the underlying probability space, and let $\mathcal{F}_i^x$ be the σ-field generated by $\{x_1, \ldots, x_i, i \geq 1\}$. Let $\mathbf{h}_i(\boldsymbol{\theta}) = \mathbf{h}_i(x_1, \ldots, x_i, \boldsymbol{\theta})$, $1 \leq i \leq n$ be specified Q-dimensional vectors which are martingale differences. We consider the class $\mathcal{M}$ of zero mean and square integrable P-dimensional martingale estimating functions of the form

$$
\mathcal{M} = \left\{ \mathbf{g}_n(\boldsymbol{\theta}) : \mathbf{g}_n(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{a}_{i-1}(\boldsymbol{\theta})\mathbf{h}_i(\boldsymbol{\theta}) \right\},
$$

where $\mathbf{a}_{i-1}(\boldsymbol{\theta})$ are $P \times Q$ matrices depending on $x_1, \ldots, x_{i-1}$, $1 \leq i \leq n$ and the parameter $\boldsymbol{\theta}$, and $P \leq Q$. The estimating functions $\mathbf{g}_n(\boldsymbol{\theta})$ are further assumed to be almost surely differentiable with respect to the components of $\boldsymbol{\theta}$ and are such that $\mathrm{E}\left[\frac{\partial \mathbf{g}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{n-1}^x \right]$ and $\mathrm{E}[\mathbf{g}_n(\boldsymbol{\theta})\mathbf{g}_n(\boldsymbol{\theta})' | \mathcal{F}_{n-1}^x]$ are nonsingular for all $\boldsymbol{\theta} \in \Theta$ and for each $n \geq 1$. The expectations are always taken with respect to $P_{\boldsymbol{\theta}}$. Estimators of $\boldsymbol{\theta}$ can be obtained by solving the P estimating equations $\mathbf{g}_n(\boldsymbol{\theta}) = \mathbf{0}$ (Godambe 1985). Furthermore, the $P \times P$ matrix $E[\mathbf{g}_n(\boldsymbol{\theta})\mathbf{g}_n(\boldsymbol{\theta})' | \mathcal{F}_{n-1}^x]$ is assumed to be positive definite for all $\boldsymbol{\theta} \in \Theta$. Then, in the class of all zero mean and square integrable martingale estimating functions $\mathcal{M}$, the optimal estimating function $\mathbf{g}_n^*(\boldsymbol{\theta})$ which maximizes, in the partial order of nonnegative definite matrices, the information matrix

$$
\begin{aligned}
\mathbf{I}_{\mathbf{g}_n}(\boldsymbol{\theta}) &= \left(\sum_{i=1}^n \mathbf{a}_{i-1}(\boldsymbol{\theta}) E\left[\frac{\partial \mathbf{h}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{i-1}^x \right] \right)' \\
&\quad \times \left(\sum_{i=1}^n E[(\mathbf{a}_{i-1}(\boldsymbol{\theta})\mathbf{h}_i(\boldsymbol{\theta}))(\mathbf{a}_{i-1}(\boldsymbol{\theta})\mathbf{h}_i(\boldsymbol{\theta}))' | \mathcal{F}_{i-1}^x] \right)^{-1} \\
&\quad \left(\sum_{i=1}^n \mathbf{a}_{i-1}(\boldsymbol{\theta}) E\left[\frac{\partial \mathbf{h}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{i-1}^x \right] \right)
\end{aligned}
$$

is given by

$$\mathbf{g}_n^*(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{a}_{i-1}^*(\boldsymbol{\theta})\mathbf{h}_i(\boldsymbol{\theta}) = \sum_{i=1}^n \left(E\left[\frac{\partial \mathbf{h}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{i-1}^x \right] \right)' (E[\mathbf{h}_i(\boldsymbol{\theta})\mathbf{h}_i(\boldsymbol{\theta})' | \mathcal{F}_{i-1}^x])^{-1}\mathbf{h}_i,$$

and the corresponding optimal information reduces to

$$\mathbf{I}_{\mathbf{g}_n^*}(\boldsymbol{\theta}) = \sum_{i=1}^n \left(E\left[\frac{\partial \mathbf{h}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{i-1}^x \right] \right)' (E[\mathbf{h}_i(\boldsymbol{\theta})\mathbf{h}_i(\boldsymbol{\theta})' | \mathcal{F}_{i-1}^x])^{-1} \left(E\left[\frac{\partial \mathbf{h}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{i-1}^x \right] \right).$$

Now, consider a real-valued discrete-time stochastic process $\{x_i, i = 1, 2, \ldots\}$ with conditional moments

$$\mu_i(\boldsymbol{\theta}) = E[x_i | \mathcal{F}_{i-1}^x], \tag{7}$$

$$\sigma_i^2(\boldsymbol{\theta}) = \mathrm{Var}(x_i | \mathcal{F}_{i-1}^x). \tag{8}$$

In order to estimate the parameter $\boldsymbol{\theta}$ based on the observations $x_1, \ldots, x_n$, we consider the usual martingale differences $\{m_i(\boldsymbol{\theta}) = x_i - \mu_i(\boldsymbol{\theta})\}$ and the generalized martingale differences $\{M_i(\boldsymbol{\theta}) = \widetilde{q}(m_i(\boldsymbol{\theta})) - E[\widetilde{q}(m_i(\boldsymbol{\theta}))|\mathcal{F}_{i-1}^x]\}$, for $i = 1, \ldots, n$. The generalized martingale difference corresponds to the least absolute deviations (LAD) case when $\widetilde{q}(m_i(\boldsymbol{\theta})) = sgn(m_i(\boldsymbol{\theta}))$, to the quadratic estimating function case when $\widetilde{q}(m_i(\boldsymbol{\theta})) = m_i(\boldsymbol{\theta})^2$, and to the transformed estimating function (Merkouris 2007) when $\widetilde{q}(m_i(\boldsymbol{\theta})) = \exp(\iota u m_i(\boldsymbol{\theta}))$ for real u and for $\iota = \sqrt{-1}$. The quadratic variations of $m_i(\boldsymbol{\theta})$, $M_i(\boldsymbol{\theta})$, and the quadratic covariation of $m_i(\boldsymbol{\theta})$ and $M_i(\boldsymbol{\theta})$ are, respectively,

$$\langle m \rangle_i = E[m_i^2(\boldsymbol{\theta})|\mathcal{F}_{i-1}^x] = \sigma_i^2(\boldsymbol{\theta}),$$
$$\langle M \rangle_i = E[\widetilde{q}^2(m_i(\boldsymbol{\theta})|\mathcal{F}_{i-1}^x] - (E[\widetilde{q}(m_i(\boldsymbol{\theta}))|\mathcal{F}_{i-1}^x])^2,$$

and

$$\langle m, M \rangle_i = E[m_i(\boldsymbol{\theta})\widetilde{q}(m_i(\boldsymbol{\theta}))|\mathcal{F}_{i-1}^x].$$

For notational convenience, we denote $\widetilde{q}(m_i(\boldsymbol{\theta}))$ by $\widetilde{q}_i$ in the rest of the paper. In general, $\widetilde{q}_i$ is any differentiable function with respect to $\boldsymbol{\theta}$ chosen in a way such that $\langle M \rangle_i$ and $\langle m, M \rangle_i$ exist. However, although it is not differentiable at 0 in the LAD case, it admits positive and negative derivatives (see Thavaneswaran and Heyde 1999 for details). The optimal estimating functions based on the martingale differences $m_i(\boldsymbol{\theta})$ and $M_i(\boldsymbol{\theta})$ are, respectively,

$$g_m^*(\boldsymbol{\theta}) = -\sum_{i=1}^n \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{m_i}{\langle m \rangle_i}$$

and

$$g_M^*(\boldsymbol{\theta}) = \sum_{i=1}^n \left(E\left[\frac{\partial \widetilde{q}_i}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{i-1}^x \right] - \frac{\partial E[\widetilde{q}_i | \mathcal{F}_{i-1}^x]}{\partial \boldsymbol{\theta}} \right) \frac{M_i}{\langle M \rangle_i}.$$

The information associated with $g_m^*(\boldsymbol{\theta})$ and $g_M^*(\boldsymbol{\theta})$ is, respectively,

$$\mathbf{I}_{g_m^*}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \frac{1}{\langle m \rangle_i},$$

$$\mathbf{I}_{g_M^*}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(E\left[\frac{\partial \widetilde{q}_i}{\partial \boldsymbol{\theta}} \frac{\partial \widetilde{q}_i}{\partial \boldsymbol{\theta}'} \middle| \mathcal{F}_{i-1}^x \right] - \frac{\partial E[\widetilde{q}_i \,|\, \mathcal{F}_{i-1}^x]}{\partial \boldsymbol{\theta}} \frac{\partial E[\widetilde{q}_i \,|\, \mathcal{F}_{i-1}^x]}{\partial \boldsymbol{\theta}'} \right) \frac{1}{\langle M \rangle_i}.$$

The following theorem first extends the results for *quadratic* estimating functions in Liang et al. (2011) to combined estimating functions based on the martingale differences $m_i(\boldsymbol{\theta})$ and the *generalized* martingale differences $M_i(\boldsymbol{\theta})$. Next, the theorem provides the form of *recursive* estimates based on the generalized combined estimating functions. Neither the combined estimation nor the recursive estimation has been previously discussed to this level of generality in the literature, and in particular for duration models.

Theorem 1 *For the general model defined by* (7) *and* (8)*, in the class of all combined estimating functions of the form*

$$\mathcal{G}_C = \left\{ g_C(\boldsymbol{\theta}) : g_C(\boldsymbol{\theta}) = \sum_{i=1}^{n} (\mathbf{a}_{i-1}(\boldsymbol{\theta}) m_i(\boldsymbol{\theta}) + \mathbf{b}_{i-1}(\boldsymbol{\theta}) M_i(\boldsymbol{\theta})) \right\},$$

(a) *the optimal estimating function is given by*

$$g_C^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} (\mathbf{a}_{i-1}^*(\boldsymbol{\theta}) m_i(\boldsymbol{\theta}) + \mathbf{b}_{i-1}^*(\boldsymbol{\theta}) M_i(\boldsymbol{\theta})),$$

where

$$\mathbf{a}_{i-1}^*(\boldsymbol{\theta}) = \rho_i^2 \left(-\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{\langle m \rangle_i} - \left(E\left[\frac{\partial \widetilde{q}_i}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{i-1}^x \right] - \frac{\partial E[\widetilde{q}_i \,|\, \mathcal{F}_{i-1}^x]}{\partial \boldsymbol{\theta}} \right) \eta_i \right) \tag{9}$$

and

$$\mathbf{b}_{i-1}^*(\boldsymbol{\theta}) = \rho_i^2 \left(\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \eta_i + \left(E\left[\frac{\partial \widetilde{q}_i}{\partial \boldsymbol{\theta}} \middle| \mathcal{F}_{i-1}^x \right] - \frac{\partial E[\widetilde{q}_i \,|\, \mathcal{F}_{i-1}^x]}{\partial \boldsymbol{\theta}} \right) \frac{1}{\langle M \rangle_i} \right), \tag{10}$$

where $\rho_i^2 = \left(1 - \frac{\langle m, M \rangle_i^2}{\langle m \rangle_i \langle M \rangle_i}\right)^{-1}$, *and* $\eta_i = \frac{\langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i}$;

(b) *the information* $\mathbf{I}_{g_C^*}(\boldsymbol{\theta})$ *is given by*

$$\begin{aligned}\mathbf{I}_{\mathbf{g}_C^*}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \rho_i^2 \left(\frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \frac{1}{\langle m \rangle_i} \right. \\ &\quad + \left(E\left[\frac{\partial \widetilde{q}_i}{\partial \boldsymbol{\theta}} \frac{\partial \widetilde{q}_i}{\partial \boldsymbol{\theta}'} \middle| \mathcal{F}_{i-1}^x \right] - \frac{\partial E[\widetilde{q}_i \,|\, \mathcal{F}_{i-1}^x]}{\partial \boldsymbol{\theta}} \frac{\partial E[\widetilde{q}_i \,|\, \mathcal{F}_{i-1}^x]}{\partial \boldsymbol{\theta}'} \right) \frac{1}{\langle M \rangle_i}\end{aligned}$$

$$+\left(\frac{\partial\mu_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left(E\left[\frac{\partial\widetilde{q}_i}{\partial\boldsymbol{\theta}'}\middle|\mathcal{F}^x_{i-1}\right]-\frac{\partial E[\widetilde{q}_i\,|\,\mathcal{F}^x_{i-1}]}{\partial\boldsymbol{\theta}'}\right)\right.$$
$$\left.+\left(E\left[\frac{\partial\widetilde{q}(m_i(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}}\middle|\mathcal{F}^x_{i-1}\right]-\frac{\partial E[\widetilde{q}_i\,|\,\mathcal{F}^x_{i-1}]}{\partial\boldsymbol{\theta}}\right)\frac{\partial\mu_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)\eta_i\Bigg);$$

(c) *the optimal estimating function is equal to* $g^*_m(\boldsymbol{\theta})$ *if for each* i

$$\frac{\partial\mu_i}{\partial\boldsymbol{\theta}}\eta_i=-\left(E\left[\frac{\partial\widetilde{q}_i}{\partial\boldsymbol{\theta}}\middle|\mathcal{F}^x_{i-1}\right]-\frac{\partial E[\widetilde{q}_i\,|\,\mathcal{F}^x_{i-1}]}{\partial\boldsymbol{\theta}}\right);$$

(d) *the recursive estimate for* $\boldsymbol{\theta}$ *is given by*

$$\widehat{\boldsymbol{\theta}}_i=\widehat{\boldsymbol{\theta}}_{i-1}+\mathbf{K}_i(\mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})m_i(\widehat{\boldsymbol{\theta}}_{i-1})+\mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})M_i(\widehat{\boldsymbol{\theta}}_{i-1})),\tag{11}$$

$$\mathbf{K}_i=\mathbf{K}_{i-1}(\mathbf{I}_p-(\mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})\frac{\partial m_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial\boldsymbol{\theta}'}+\frac{\partial\mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial\boldsymbol{\theta}}m_i(\widehat{\boldsymbol{\theta}}_{i-1})$$
$$+\mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})\frac{\partial M_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial\boldsymbol{\theta}'}+\frac{\partial\mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial\boldsymbol{\theta}}M_i(\widehat{\boldsymbol{\theta}}_{i-1}))\mathbf{K}_{i-1})^{-1},\tag{12}$$

where $\mathbf{I}_P$ *is* P*-dimensional identity matrix, and* $\mathbf{a}^*_{i-1}$ *and* $\mathbf{b}^*_{i-1}$ *can be calculated by substituting* $\widehat{\boldsymbol{\theta}}_{i-1}$ *in* (9) *and* (10)*, respectively.*

(e) *for the scalar parameter case, the recursive estimate of* θ *is given by*

$$\widehat{\theta}_i=\widehat{\theta}_{i-1}+K_i(a^*_{i-1}(\widehat{\theta}_{i-1})m_i(\widehat{\theta}_{i-1})+b^*_{i-1}(\widehat{\theta}_{i-1})M_i(\widehat{\theta}_{i-1})),$$
$$K_i-K_{i-1}=A_iK_{i-1}K_i,$$

where $A_i=a^*_{i-1}(\widehat{\theta}_{i-1})\frac{\partial m_i(\widehat{\theta}_{i-1})}{\partial\theta}+\frac{\partial a^*_{i-1}(\widehat{\theta}_{i-1})}{\partial\theta}m_i(\widehat{\theta}_{i-1})+b^*_{i-1}(\widehat{\theta}_{i-1})\frac{\partial M_i(\widehat{\theta}_{i-1})}{\partial\theta}+\frac{\partial b^*_{i-1}(\widehat{\theta}_{i-1})}{\partial\theta}M_i(\widehat{\theta}_{i-1})$.

The proof of the theorem is given in the Appendix.

Note 1 The optimal information matrix based on the first i observations is given by $-E\left[\frac{\partial g^*_C(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\middle|\mathcal{F}^x_{i-1}\right]$, and hence, $K_i^{-1}=-\sum_{s=1}^{i}\frac{\partial g^*_C(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial\boldsymbol{\theta}}$ can be interpreted as the observed information matrix associated with the optimal combined estimating function $g^*_C(\boldsymbol{\theta})$. Equations (11) and (12) update the recursive estimate as well as the associated information matrix. For more details and an interpretation of these terms, see Thavaneswaran and Abraham (1988).

Note 2 When ε_i follows an exponential distribution with scale parameter λ, $\mu_\varepsilon=1/\lambda$, $\sigma^2_\varepsilon=1/\lambda^2$, $\gamma_\varepsilon=2/\lambda^3$, and $\kappa_\varepsilon=9/\lambda^4$. Then, $\mathbf{I}_{g^*_m}(\boldsymbol{\theta})=\sum_{i=1}^n\frac{1}{\psi_i^2}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}$, $\mathbf{I}_{g^*_M}(\boldsymbol{\theta})=\frac{1}{2}\sum_{i=1}^n\frac{1}{\psi_i^2}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}$, and $\mathbf{I}_{g^*_C}(\boldsymbol{\theta})=\sum_{i=1}^n\frac{1}{\psi_i^2}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}$. Moreover, $\frac{\partial\mu_i}{\partial\boldsymbol{\theta}}\frac{\langle m,M\rangle_i}{\langle m\rangle_i}=\frac{\partial\sigma_i^2}{\partial\boldsymbol{\theta}}$, and hence $g^*_C(\boldsymbol{\theta})=g^*_m(\boldsymbol{\theta})$ is optimal.

Note 3 Suppose that ε_i follows a lognormal distribution with parameters μ and σ^2. The moments of ε_i are given by $\mu_\varepsilon = \exp(\mu + \frac{\sigma^2}{2})$, $\sigma_\varepsilon^2 = (\exp(\sigma^2) - 1)\exp(2\mu+\sigma^2)$, $\gamma_\varepsilon = \exp(\sigma^2+2)\sqrt{\exp(\sigma^2-1}$, and $\kappa_\varepsilon = \exp(4\sigma^2)+2\exp(3\sigma^2)+3\exp(2\sigma^2)-6$. Then, it can be shown that $\mathbf{I}_{g_m^*}(\boldsymbol{\theta}) = \frac{1}{e^{\sigma^2}-1}\sum_{i=1}^n \frac{1}{\psi_i^2}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}$, $\mathbf{I}_{g_M^*}(\boldsymbol{\theta}) = \frac{4}{e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-4}\sum_{i=1}^n \frac{1}{\psi_i^2}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}$, and $\mathbf{I}_{g_C^*}(\boldsymbol{\theta}) = \frac{e^{2\sigma^2}+2e^{\sigma^2}-1}{e^{3\sigma^2}-e^{\sigma^2}}\sum_{i=1}^n \frac{1}{\psi_i^2}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}$, and hence, $\mathbf{I}_{g_C^*}(\boldsymbol{\theta}) > \mathbf{I}_{g_m^*}(\boldsymbol{\theta})$ and $\mathbf{I}_{g_C^*}(\boldsymbol{\theta}) > \mathbf{I}_{g_M^*}(\boldsymbol{\theta})$. Moreover, when ε_i follows a double exponential distribution with density function $f(\varepsilon;\mu,b) = \frac{1}{2b}e^{-|\varepsilon-\mu|/b}$, where μ is a known location parameter and $b > 0$ is a known scale parameter, the moments of ε_i are given by $\mu_\varepsilon = \mu$, $\sigma_\varepsilon^2 = 2b^2$, $\gamma_\varepsilon = 0$, and $\kappa_\varepsilon = 3$. Then $\mathbf{I}_{g_m^*}(\boldsymbol{\theta}) = \frac{\mu^2}{2b^2}\sum_{i=1}^n \frac{1}{\psi_i^2}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}$, $\mathbf{I}_{g_M^*}(\boldsymbol{\theta}) = \frac{2}{3}\sum_{i=1}^n \frac{1}{\psi_i^2}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}$, and $\mathbf{I}_{g_C^*}(\boldsymbol{\theta}) = \frac{5\mu^2+8b^2}{10b^2}\sum_{i=1}^n \frac{1}{\psi_i^2}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}$, and hence, $\mathbf{I}_{g_C^*}(\boldsymbol{\theta}) > \mathbf{I}_{g_m^*}(\boldsymbol{\theta})$ and $\mathbf{I}_{g_C^*}(\boldsymbol{\theta}) > \mathbf{I}_{g_M^*}(\boldsymbol{\theta})$. It follows from Lindsay (1985) that the estimate obtained by solving the estimating equation $g_C^*(\boldsymbol{\theta}) = 0$ is more efficient than that obtained by solving the estimating equations $g_m^*(\boldsymbol{\theta}) = 0$, or $g_M^*(\boldsymbol{\theta}) = 0$.

3.1 Multiplicative random coefficient ACD models

Let $m_i = x_i - \mu_i$ and $M_i = m_i^2 - \sigma_i^2$ be the sequences of martingale differences such that

$$\begin{aligned}
\langle m\rangle_i &= \sigma_\varepsilon^2(\mu_z+\psi_i)^2 + \sigma_z^2(\mu_\varepsilon^2+\sigma_\varepsilon^2)\\
\langle M\rangle_i &= (\kappa_\varepsilon-\sigma_\varepsilon^4)(\mu_z+\psi_i)^4 + 2\sigma_z^2(2\mu_\varepsilon^2\sigma_\varepsilon^2+6\mu_\varepsilon\gamma_\varepsilon+3\kappa_\varepsilon-\sigma_\varepsilon^4)(\mu_z+\psi_i)^2\\
&\quad + 4\gamma_z(3\mu_\varepsilon^2\sigma_z^2+3\mu_\varepsilon\gamma_\varepsilon+\kappa_\varepsilon)(\mu_z+\psi_i) + \kappa_z(\mu_\varepsilon^4+4\mu_\varepsilon\gamma_\varepsilon+6\mu_\varepsilon^2\sigma_\varepsilon^2+\kappa_\varepsilon)\\
&\quad - \sigma_z^4(\mu_\varepsilon^4+2\mu_\varepsilon^2\sigma_\varepsilon^2+\sigma_\varepsilon^4),\\
\langle m, M\rangle_i &= \gamma_\varepsilon(\mu_z+\psi_i)^3 + (3\sigma_z^2\gamma_\varepsilon+6\mu_\varepsilon\sigma_z^2\sigma_\varepsilon^2)(\mu_z+\psi_i) + \gamma_z(\mu_\varepsilon^3+3\mu_\varepsilon\sigma_\varepsilon^2+\gamma_\varepsilon).
\end{aligned}$$

Theorem 2 *For the model (5), in the class of all quadratic estimating functions of the form* $\mathcal{G}_C = \{g_C(\boldsymbol{\theta}) : g_C(\boldsymbol{\theta}) = \sum_{i=1}^n(\mathbf{a}_{i-1}m_i + \mathbf{b}_{i-1}M_i)\}$,

(a) *the optimal estimating function is given by* $g_C^*(\boldsymbol{\theta}) = \sum_{i=1}^n(\mathbf{a}_{i-1}^*m_i + \mathbf{b}_{i-1}^*M_i)$, *where*

$$\mathbf{a}_{i-1}^* = \rho_i^2(r_i, R_i, R_ix_{i-1},\ldots,R_ix_{i-p}, R_i\psi_{i-1},\ldots,R_i\psi_{i-q})'$$

where

$$\langle m\rangle_i\langle M\rangle_i r_i = -\mu_\varepsilon\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\langle M\rangle_i + \left(2\sigma_\varepsilon^2(\mu_z+\psi_i)\frac{\partial\mu_z}{\partial\boldsymbol{\phi}} + (\mu_\varepsilon^2+\sigma_\varepsilon^2)\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}}\right)\langle m, M\rangle_i,$$

$$\langle m\rangle_i\langle M\rangle_i R_i = -\mu_\varepsilon\langle M\rangle_i + 2\sigma_\varepsilon^2(\mu_z+\psi_i)\langle m, M\rangle_i$$

and

$$\mathbf{b}^*_{i-1} = \rho_i^2 (s_i, S_i, S_i x_{i-1}, \ldots, S_i x_{i-p}, S_i \psi_{i-1}, \ldots, S_i \psi_{i-q})'$$

where

$$\langle m\rangle_i \langle M\rangle_i s_i = \mu_\varepsilon \frac{\partial \mu_z}{\partial \boldsymbol{\phi}} \langle m, M\rangle_i - \left(2\sigma_\varepsilon^2 (\mu_z + \psi_i) \frac{\partial \mu_z}{\partial \boldsymbol{\phi}} + (\mu_\varepsilon^2 + \sigma_\varepsilon^2) \frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}}\right) \langle m\rangle_i,$$

$$\langle m\rangle_i \langle M\rangle_i S_i = \mu_\varepsilon \langle m, M\rangle_i - 2\sigma_\varepsilon^2 (\mu_z + \psi_i) \langle m\rangle_i$$

(b) *the recursive estimate for* $\boldsymbol{\theta} = (\boldsymbol{\phi}', \omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$ *is given by*

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_i &= \widehat{\boldsymbol{\theta}}_{i-1} + \mathbf{K}_i (\mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) m_i(\widehat{\boldsymbol{\theta}}_{i-1}) + \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) M_i(\widehat{\boldsymbol{\theta}}_{i-1})), \\
\mathbf{K}_i &= \mathbf{K}_{i-1} \left(\mathbf{I}_{p+q+l+1} - \left(\mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) \frac{\partial m_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}'} + \frac{\partial \mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} m_i(\widehat{\boldsymbol{\theta}}_{i-1}) \right.\right. \\
&\qquad \left.\left. + \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) \frac{\partial M_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}'} + \frac{\partial \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} M_i(\widehat{\boldsymbol{\theta}}_{i-1}) \right) \mathbf{K}_{i-1} \right)^{-1}.
\end{aligned}$$

Explicit expressions for $g^*_C(\boldsymbol{\theta})$, the associated information matrix $\mathbf{I}_{g^*_C}(\boldsymbol{\theta})$, as well as the proof of the theorem are given in the Appendix.

3.2 RCA models with ACD errors

Corresponding to the model (6), let $m_i = x_i - \mu_i$ and $M_i = m_i^2 - \sigma_i^2$ such that $\langle m\rangle_i = \sigma_z^2(\phi) x_{i-1}^2 + \sigma_\varepsilon^2(\boldsymbol{\theta}) \psi_i$, $\langle M\rangle_i = (\kappa_z(\phi) - \sigma_z^4(\phi)) x_{i-1}^4 + (\kappa_\varepsilon(\boldsymbol{\theta}) - \sigma_\varepsilon^4(\boldsymbol{\theta})) \psi_i^4 + 4\sigma_z^2(\phi)\sigma_\varepsilon^2(\boldsymbol{\theta}) x_{i-1}^2 \psi_i^2$, $\langle m, M\rangle_i = \gamma_z(\phi) x_{i-1}^3 + \gamma_\varepsilon(\boldsymbol{\theta}) \psi_i^3$.

Theorem 3 *For the model* (6)*, in the class of all quadratic estimating functions of the form* $\mathcal{G}_C = \{g_C(\boldsymbol{\theta}) : g_C(\boldsymbol{\theta}) = \sum_{i=1}^n (\mathbf{a}_{i-1} m_i + \mathbf{b}_{i-1} M_i)\}$,

(a) *the optimal estimating function is given by* $g^*_C(\boldsymbol{\theta}) = \sum_{i=1}^n (\mathbf{a}^*_{i-1} m_i + \mathbf{b}^*_{i-1} M_i)$, *where*

$$\mathbf{a}^*_{i-1} = \rho_i^2 (v_{i1}, v_{i2}, V_i, V_i x_{i-1}, \ldots, V_i x_{i-p}, V_i \psi_{i-1}, \ldots, V_i \psi_{i-q})'$$

where

$$\langle m\rangle_i \langle M\rangle_i v_{i1} = -\left(x_{i-1} + \frac{\partial \mu_\varepsilon(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \psi_i\right) \langle M\rangle_i + \frac{\partial \sigma_\varepsilon^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \psi_i^2 \langle m, M\rangle_i$$

$$\langle m\rangle_i \langle M\rangle_i v_{i2} = -\frac{\partial \mu_z(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} x_{i-1} \langle M\rangle_i + \frac{\partial \sigma_z^2(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} x_{i-1}^2 \langle m, M\rangle_i,$$

$$\langle m\rangle_i \langle M\rangle_i V_i = -\mu_\varepsilon \langle M\rangle_i + 2\sigma_\varepsilon^2 \psi_i \langle m, M\rangle_i$$

and

$$\mathbf{b}^*_{i-1} = \rho_i^2(w_{i1}, w_{i2}, W_i, W_i x_{i-1}, \ldots, W_i x_{i-p}, W_i \psi_{i-1}, \ldots, W_i \psi_{i-q})'$$

where

$$\begin{aligned}
\langle m\rangle_i \langle M\rangle_i w_{i1} &= \left(x_{i-1} + \frac{\partial \mu_\varepsilon(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\psi_i\right)\langle m, M\rangle_i - \frac{\partial \sigma_\varepsilon^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\psi_i^2 \langle m\rangle_i,\\
\langle m\rangle_i \langle M\rangle_i w_{i2} &= \frac{\partial \mu_z(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} x_{i-1}\langle m, M\rangle_i - \frac{\partial \sigma_z^2(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} x_{i-1}^2 \langle m\rangle_i,\\
\langle m\rangle_i \langle M\rangle_i W_i &= \mu_\varepsilon \langle m, M\rangle_i - 2\sigma_\varepsilon^2 \psi_i \langle m\rangle_i
\end{aligned}$$

(b) *the recursive estimate for* $\boldsymbol{\theta} = (\boldsymbol{\phi}', \omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$ *is given by*

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_i &= \widehat{\boldsymbol{\theta}}_{i-1} + \mathbf{K}_i(\mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) m_i(\widehat{\boldsymbol{\theta}}_{i-1}) + \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1}) M_i(\widehat{\boldsymbol{\theta}}_{i-1})),\\
\mathbf{K}_i &= \mathbf{K}_{i-1}\Bigg(\mathbf{I}_{p+q+l+2} - \Bigg(\mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})\frac{\partial m_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}'} + \frac{\partial \mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} m_i(\widehat{\boldsymbol{\theta}}_{i-1})\\
&\qquad + \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})\frac{\partial M_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}'} + \frac{\partial \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} M_i(\widehat{\boldsymbol{\theta}}_{i-1})\Bigg)\mathbf{K}_{i-1}\Bigg)^{-1}.
\end{aligned}$$

Explicit expressions for $g^*_C(\boldsymbol{\theta})$, the associated information matrix $\mathbf{I}_{g^*_C}(\boldsymbol{\theta})$, as well as the proof of the theorem are shown in the Appendix.

3.2.1 Recursive estimation for RCA(1) model with ACD(1,1) errors

Consider the RCA(1) model with ACD(1,1) errors of the form

$$\begin{aligned}
x_i &= (\theta + z_i)x_{i-1} + \psi_i \varepsilon_i\\
\psi_i &= \omega + \alpha x_{i-1} + \beta \psi_{i-1},
\end{aligned}$$

where z_i follows a centered log-normal distribution with mean 0 and variance ϕ, and $\varepsilon_i \sim Gamma(1, \theta)$, and $\{z_i\}$ and $\{\varepsilon_i\}$ are assumed to be mutually uncorrelated. We discuss the estimation of the parameter vector $\boldsymbol{\theta} = (\theta, \phi, \omega, \alpha, \beta)'$. In this model, the first four conditional central moments are $\mu_i = \theta x_{i-1} + \theta \psi_i$, $\sigma_i^2 = \phi x_{i-1}^2 + \theta^2 \psi_i^2$, $\gamma_i = 2\theta^3 \psi_i^3$ and $\kappa_i = 3\phi^2 x_{i-1}^4 + 6\phi\theta^2 x_{i-1}^2 \psi_i^2 + 9\theta^4 \psi_i^4$. Let $m_i = x_i - \theta(x_{i-1} + \psi_i)$ and $M_i = m_i^2 - \sigma_i^2 = x_{i-1}(x_{i-1} + 2\psi_i)\theta^2 - 2x_i(x_{i-1} + \psi_i)\theta + x_i^2 - \phi x_{i-1}^2$ be the sequences of martingale differences; then $\langle m\rangle_i = \phi x_{i-1}^2 + \theta^2 \psi_i^2$, $\langle M\rangle_i = 2\phi^2 x_{i-1}^4 + 4\phi\theta^2 x_{i-2}^2 \psi_i^2 + 8\theta^4 \psi_i^4$, and $\langle m, M\rangle_i = 2\theta^3 \psi_i^3$. We describe recursive estimation of θ for known scale ϕ, ω, α, and β. The optimal estimate is obtained by solving the estimating equation

$$g^*_C(\theta) = \sum_{i=1}^n a^*_{i-1}(\theta) m_i + b^*_{i-1}(\theta) M_i = 0,$$

where

$$a^*_{i-1}(\theta) = -\frac{2\psi_i^4(2x_{i-1}+\psi_i)\theta^4 + 2\phi x_{i-1}^2\psi_i^2(x_{i-1}+\psi_i)\theta^2 + \phi^2 x_{i-1}^4(x_{i-1}+\psi_i)}{2\psi_i^6\theta^6 + 6\phi x_{i-1}^2\psi_i^4 + 3\phi^2 x_{i-1}^4\psi_i^2\theta^2 + \phi^3 x_{i-1}^6},$$

$$b^*_{i-1}(\theta) = \frac{x_{i-1}\psi_i^3\theta^3 - \phi x_{i-1}^2\psi_i^2\theta}{2\psi_i^6\theta^6 + 6\phi x_{i-1}^2\psi_i^4\theta^4 + 3\phi^2 x_{i-1}^4\psi_i^2\theta^2 + \phi^3 x_{i-1}^6}.$$

Therefore, the recursive estimate of θ is given by

$$\begin{aligned}\widehat{\theta_i} &= \widehat{\theta}_{i-1} + K_i(a^*_{i-1}(\widehat{\theta}_{i-1})m_i(\widehat{\theta}_{i-1}) + b^*_{i-1}(\widehat{\theta}_{i-1})M_i(\widehat{\theta}_{i-1})),\\ K_i - K_{i-1} &= C_i K_{i-1} K_i,\end{aligned}$$

where $C_i = (x_{i-1}+\psi_i)a^*_{i-1}(\widehat{\theta}_{i-1}) + \frac{\partial a^*_{i-1}(\widehat{\theta}_{i-1})}{\partial\theta} m_i(\widehat{\theta}_{i-1}) + 2(x_{i-1}(x_{i-1}+2\psi_i)\widehat{\theta}_{i-1} - x_i(x_{i-1}+\psi_i))b^*_{i-1}(\widehat{\theta}_{i-1}) + \frac{\partial b^*_{i-1}(\widehat{\theta}_{i-1})}{\partial\theta} M_i(\widehat{\theta}_{i-1})$.

3.3 Log-SCD models

For the log-SCD model defined by

$$\begin{aligned}x_i &= \exp(\psi_i)\varepsilon_i,\\ \psi_i &= \omega + \sum_{j=1}^{p}\alpha_j \log x_{i-j} + \sum_{j=1}^{q}\beta_j\psi_{i-j} + z_i,\end{aligned} \tag{13}$$

neither the linear nor the combined estimating functions approach has been discussed in the literature. This model is similar to the doubly stochastic time series model discussed in Thavaneswaran and Abraham (1988). In order to estimate the model parameters, we express the log-transformed model in non-Gaussian state space form, which implies that the usual Kalman filtering is not appropriate. In Theorem 4, we consider a class of linear filtered estimates and find the optimal estimate which minimizes the conditional MSE. Then, we use Theorem 1 and construct the estimating functions based on the martingale differences, m_i and $M_i = m_i^2 - \sigma_i^2$.

Specifically, using the logarithmic transformation for x_i, and setting $p = 1$ and $q = 1$, we have

$$\log x_i = \omega + \alpha \log x_{i-1} + \beta\psi_{i-1} + \mu_{\log\varepsilon} + \xi_i,$$

where $\xi_i = z_i + \log\varepsilon_i - \mu_{\log\varepsilon}$ is uncorrelated with ψ_{i-1} and $\mu_{\log\varepsilon} = E(\log\varepsilon_i)$, and $\mathrm{Var}(\log\varepsilon_i) = \sigma^2_{\log\varepsilon}$. For instance, if ε_i follows the lognormal (μ, σ^2) distribution, then $\mu_{\log\varepsilon} = \mu$ and $\sigma^2_{\log\varepsilon} = \sigma^2$. If ε_i follows a Weibull $(\alpha, 1)$ distribution, then $\log\varepsilon_i$ follows the extreme value distribution $(0, \pi^2/6\alpha^2)$ with $\mu_{\log\varepsilon} = -0.5772/\alpha$ and $\sigma^2_{\log\varepsilon} = \pi^2/6\alpha^2$. The mean and variance of ξ_i are given by $E(\xi_i) = 0$ and $\mathrm{Var}(\xi_i) = \sigma_z^2 + \sigma^2_{\log\varepsilon}$, while the covariance between ξ_i and z_i is $E(\xi_i z_i) = \sigma_z^2$.

In order to estimate the unknown parameters $\boldsymbol{\theta} = (\omega, \alpha, \beta)$, we first need to consider nonlinear filtering to obtain the optimal recursive estimate of ψ_i. By minimizing the prediction mean squared error (PMSE) $\Delta_i = E[(\psi_i - \widehat{\psi}_i)^2 | \mathcal{F}_i^x]$, the optimal recursive estimate $\widehat{\psi}_i$ and its PMSE Δ_i are given in the following theorem:

Theorem 4 *In the class of all estimates of the form* $\widehat{\psi}_i = \omega + \alpha \log x_{i-1} + \beta \widehat{\psi}_{i-1} + G_i(\log x_i - \omega - \beta \widehat{\psi}_{i-1} - \mu_{\log \varepsilon})$*, the optimal recursive estimate* $\widehat{\psi}_i$ *of* ψ_i *which minimizes the PMSE* Δ_i *is given by*

$$\begin{aligned}\widehat{\psi}_i = \omega + \alpha \log x_{i-1} + \beta \widehat{\psi}_{i-1} - \frac{\Delta_i}{\sigma^2_{\log \varepsilon}} (\log x_i - \omega - \alpha \log x_{i-1} \\ - \beta \widehat{\psi}_{i-1} - \mu_{\log \varepsilon}),\end{aligned} \tag{14}$$

where the optimal PMSE is given by

$$\Delta_i = \frac{(\beta^2 \Delta_{i-1} + \sigma_z^2)\sigma^2_{\log \varepsilon}}{\beta^2 \Delta_{i-1} + \sigma_z^2 + \sigma^2_{\log \varepsilon}}. \tag{15}$$

Proof Since

$$\psi_i - \widehat{\psi}_i = \beta(1 + G_i)(\psi_{i-1} - \widehat{\psi}_{i-1}) + G_i \xi_i + z_i,$$

$$\begin{aligned}\Delta_i &= E[(\psi_i - \widehat{\psi}_i)^2 | \mathcal{F}_i^x] \\ &= E[\beta^2(1 + G_i)^2(\psi_{i-1} - \widehat{\psi}_{i-1})^2 + G_i^2 \xi_i^2 + z_i^2 + 2G_i \xi_i z_i | \mathcal{F}_i^x] \\ &= \beta^2(1 + G_i)^2 \Delta_{i-1} + G_i^2(\sigma_z^2 + \sigma^2_{\log \varepsilon}) + \sigma_z^2 + 2G_i \sigma_z^2 \\ &= (\beta^2 \Delta_{i-1} + \sigma_z^2 + \sigma^2_{\log \varepsilon}) \left(G_i + \frac{\beta^2 \Delta_{i-1} + \sigma_z^2}{\beta^2 \Delta_{i-1} + \sigma_z^2 + \sigma^2_{\log \varepsilon}} \right)^2 \\ &\quad + \frac{(\beta^2 \Delta_{i-1} + \sigma_z^2)\sigma^2_{\log \varepsilon}}{\beta^2 \Delta_{i-1} + \sigma_z^2 + \sigma^2_{\log \varepsilon}}.\end{aligned}$$

Hence Δ_i is minimized by taking

$$G_i = -\frac{\beta^2 \Delta_{i-1} + \sigma_z^2}{\beta^2 \Delta_{i-1} + \sigma_z^2 + \sigma^2_{\log \varepsilon}}.$$

Now, the conditional mean and variance of $\log x_i$ are given by

$$\begin{aligned}\mu_i(\boldsymbol{\theta}) &= E[\log x_i | \mathcal{F}_{i-1}^x] = \omega + \alpha \log x_{i-1} + \beta E[\psi_{i-1} | \mathcal{F}_{i-1}^x] + \mu_{\log \varepsilon} \\ &= \omega + \alpha \log x_{i-1} + \beta \widehat{\psi}_{i-1} + \mu_{\log \varepsilon} \\ \sigma_i^2(\boldsymbol{\theta}) &= \mathrm{Var}(\log x_i | \mathcal{F}_{i-1}^x) = E[\beta^2(\psi_{i-1} - E[\psi_{i-1} | \mathcal{F}_{i-1}^x])^2 + \xi^2 | \mathcal{F}_{i-1}^x] \\ &= \beta^2 \Delta_{i-1} + \sigma_z^2 + \sigma^2_{\log, \varepsilon}\end{aligned}$$

where $\widehat{\psi}_{i-1}$ and Δ_i can be calculated recursively by (14) and (15). Then the optimal estimating function $\mathbf{g}_m^*(\boldsymbol{\theta})$ based on the martingale difference $m_i(\boldsymbol{\theta}) = \log x_i - \mu_i(\boldsymbol{\theta})$ is given by

$$-\sum_{i=1}^{n} \begin{pmatrix} 1+\beta\frac{\partial\mu_{i-1}}{\partial\omega} \\ \log x_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\alpha} \\ \mu_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\beta} \end{pmatrix} \frac{\log x_i - (\omega+\alpha\log x_{i-1}+\beta\mu_{i-1}+\mu_{\log\varepsilon})}{\beta^2\Delta_{i-1}+\sigma_z^2+\sigma_{\log\varepsilon}^2},$$

with associated information matrix $\mathbf{I}_{g_m^*}(\boldsymbol{\theta})$ given by

$$\sum_{i=1}^{n} \begin{pmatrix} 1+\beta\frac{\partial\mu_{i-1}}{\partial\omega} \\ \log x_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\alpha} \\ \mu_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\beta} \end{pmatrix} \begin{pmatrix} 1+\beta\frac{\partial\mu_{i-1}}{\partial\omega} \\ \log x_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\alpha} \\ \mu_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\beta} \end{pmatrix}' \frac{1}{\beta^2\Delta_{i-1}+\sigma_z^2+\sigma_{\log\varepsilon}^2}.$$

The recursive estimate for $\boldsymbol{\theta}$ is given by

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_i &= \widehat{\boldsymbol{\theta}}_{i-1} + \mathbf{K}_i\mathbf{a}_{i-1}^*(\widehat{\boldsymbol{\theta}}_{i-1})m_i(\widehat{\boldsymbol{\theta}}_{i-1}), \\ \mathbf{K}_i &= \mathbf{K}_{i-1}\left(\mathbf{I}_3 - \left(\mathbf{a}_{i-1}^*(\widehat{\boldsymbol{\theta}}_{i-1})\frac{\partial m_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial\boldsymbol{\theta}'} + \frac{\partial\mathbf{a}_{i-1}^*(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial\boldsymbol{\theta}} m_i(\widehat{\boldsymbol{\theta}}_{i-1})\right)\mathbf{K}_{i-1}\right)^{-1}. \end{aligned}$$

Moreover, the combined optimal estimating function $\mathbf{g}_C^*(\boldsymbol{\theta})$ based on the martingale differences $m_i(\boldsymbol{\theta}) = \log x_i - \mu_i(\boldsymbol{\theta})$ and $M_i(\boldsymbol{\theta}) = m_i^2(\boldsymbol{\theta}) - \sigma_i^2(\boldsymbol{\theta})$ and the recursive estimate of $\boldsymbol{\theta}$ can be obtained using Theorem 1. If we further assume that $\varepsilon_i \sim$ *logNormal*(μ, σ^2) and $z_i \sim N(0, \sigma_z^2)$, then the combined optimal estimating function $\mathbf{g}_C^*(\boldsymbol{\theta})$ is simplified as

$$\begin{aligned} -\sum_{i=1}^{n} \Bigg(& \begin{pmatrix} 1+\beta\frac{\partial\mu_{i-1}}{\partial\omega} \\ \log x_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\alpha} \\ \mu_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\beta} \end{pmatrix} \frac{m_i(\boldsymbol{\theta})}{\beta^2\Delta_{i-1}+\sigma_e^2+\sigma^2} \\ & + \begin{pmatrix} 0 \\ 0 \\ 2\beta\Delta_{i-1}+\beta^2\frac{\partial\Delta_{i-1}}{\partial\beta} \end{pmatrix} \frac{M_i(\boldsymbol{\theta})}{2(\beta^2\Delta_{i-1}+\sigma_e^2+\sigma^2)^2} \Bigg), \end{aligned}$$

with associated information matrix $\mathbf{I}_{g_C^*}(\boldsymbol{\theta})$ given by

$$\begin{aligned} \sum_{i=1}^{n} & \begin{pmatrix} 1+\beta\frac{\partial\mu_{i-1}}{\partial\omega} \\ \log x_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\alpha} \\ \mu_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\beta} \end{pmatrix} \begin{pmatrix} 1+\beta\frac{\partial\mu_{i-1}}{\partial\omega} \\ \log x_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\alpha} \\ \mu_{i-1}+\beta\frac{\partial\mu_{i-1}}{\partial\beta} \end{pmatrix}' \frac{1}{\beta^2\Delta_{i-1}+\sigma_e^2+\sigma_{\log\varepsilon}^2} \\ & + \begin{pmatrix} 0 \\ 0 \\ 2\beta\Delta_{i-1}+\beta^2\frac{\partial\Delta_{i-1}}{\partial\beta} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 2\beta\Delta_{i-1}+\beta^2\frac{\partial\Delta_{i-1}}{\partial\beta} \end{pmatrix}' \frac{1}{2(\beta^2\Delta_{i-1}+\sigma_e^2+\sigma^2)^2}. \end{aligned}$$

□

It follows from Theorem 1 that the combined estimating function is more informative than the linear estimating function discussed in Thavaneswaran and Abraham (1988). An approach similar to that discussed in this paper can be used to derive optimal estimates based on the quadratic estimating function for the quadratic SCD, and quadratic log-SCD models as well.

4 Monte Carlo simulations

We present simulation results to illustrate the performance of the recursive estimation for a few models discussed earlier. Tables 1 and 2 present results for the log ACD1(1,1) and log ACD2(1,1) models obtained by setting $p = q = 1$ in the respective models in Sect. 2.2. Column 1 in Tables 1 and 2 shows the true values of the parameters ω, α, and β. Using these values, we generated $L = 100$ sets of time series each of length $n = 4{,}000$. The distribution of ε_i was assumed to be exponential with mean 1. For each simulated series, we then computed the recursive estimates of the model parameters ω, α, and β using Theorem 1. We started with random initial values selected uniformly from reasonably wide intervals. Columns 2–6 in Tables 1 and 2 present selected percentiles of the empirical distribution of the estimates from the L simulated time series. The computations were easily coded in R and illustrate that the recursive estimation approach provides good estimates (close to the true values used for the simulation) of the parameters from these duration models. The quantiles enable us to understand the location, spread, symmetry, and tail behavior of the empirical distribution of the recursive estimates from the simulated data sets.

Table 3 presents results from a simulation study for the multiplicative RCACD(1,1) model. The simulation was carried out in a similar manner by setting $L = 100$ and $n = 4{,}000$. We simulated both ε_i and z_i from an exponential distribution with mean 1. In this model (see Sect. 3.1 for model details), the parameters must satisfy the

Table 1 Percentiles of the Distribution of Recursive Estimates from a Log ACD1(1,1) process: $n = 4{,}000$, $L = 100$

True parameters	Estimates				
	5th	25th	50th	75th	95th
$\omega = 0.6$	0.373	0.498	0.621	0.701	0.775
$\alpha = 0.05$	0.035	0.041	0.049	0.059	0.069
$\beta = 0.75$	0.550	0.648	0.717	0.822	0.938
$\omega = 0.6$	0.437	0.509	0.619	0.705	0.782
$\alpha = 0.15$	0.134	0.140	0.149	0.160	0.168
$\beta = 0.65$	0.455	0.532	0.601	0.717	0.827
$\omega = 2.0$	1.835	1.896	2.006	2.100	2.184
$\alpha = -0.1$	−0.116	−0.109	−0.101	−0.091	−0.083
$\beta = 0.75$	0.560	0.630	0.703	0.817	0.927
$\omega = 2.0$	1.823	1.892	2.008	2.100	2.184
$\alpha = -0.5$	−0.516	−0.510	−0.503	−0.490	−0.483
$\beta = 0.35$	0.161	0.235	0.302	0.417	0.527

Table 2 Percentiles of the distribution of recursive estimates from a log ACD2(1,1) process: $n = 4{,}000$, $L = 100$

True parameters	Estimates				
	5th	25th	50th	75th	95th
$\omega = 0.6$	0.419	0.507	0.616	0.700	0.784
$\alpha = 0.05$	0.034	0.041	0.049	0.059	0.067
$\beta = 0.75$	0.559	0.634	0.703	0.817	0.927
$\omega = 0.6$	0.424	0.505	0.621	0.701	0.785
$\alpha = 0.15$	0.133	0.140	0.149	0.159	0.168
$\beta = 0.65$	0.458	0.528	0.606	0.717	0.827
$\omega = 2.0$	1.835	1.896	2.006	2.100	2.184
$\alpha = 0.1$	0.084	0.091	0.099	0.109	0.117
$\beta = 0.45$	0.260	0.330	0.402	0.517	0.627
$\omega = 2.0$	1.826	1.903	2.011	2.100	2.185
$\alpha = -0.05$	−0.066	−0.059	−0.050	−0.041	−0.033
$\beta = 0.35$	0.159	0.231	0.298	0.417	0.527

Table 3 Percentiles of the distribution of recursive estimates from a multiplicative RCACD(1,1) process from Sect. 3.1: $n = 4{,}000$, $L = 100$, $\theta = 1, \phi = 1$

True Parameters	Estimates				
	5th	25th	50th	75th	95th
$\omega = 0.25$	0.095	0.148	0.206	0.273	0.337
$\alpha = 0.005$	−0.017	−0.002	0.005	0.010	0.016
$\beta = 0.25$	0.151	0.190	0.242	0.304	0.341
$\omega = 0.25$	0.131	0.178	0.217	0.277	0.331
$\alpha = 0.2$	0.191	0.197	0.201	0.206	0.209
$\beta = 0.35$	0.257	0.297	0.346	0.408	0.440
$\omega = 0.1$	−0.045	0.001	0.060	0.129	0.196
$\alpha = 0.01$	−0.003	0.004	0.011	0.016	0.021
$\beta = 0.45$	0.357	0.395	0.444	0.511	0.539
$\omega = 0.1$	−0.008	0.027	0.064	0.130	0.183
$\alpha = 0.2$	0.191	0.197	0.202	0.206	0.210
$\beta = 0.45$	0.360	0.397	0.448	0.509	0.539

conditions $\omega > 0, \alpha > 0, \beta > 0$ and $\alpha + \beta < 1$. From Table 3, we see that for Case 1 when the true value of α is assumed to be 0.005, the 5th and 25th percentiles of the estimates from the L simulated sets is negative. Negative values also occur in certain lower percentiles when the true values of ω and/or α are close to zero. These negative values may be ignored, or viewed as zeroes. Again, the recursive estimation approach yields estimates of the model parameters close to the true values.

Table 4 presents percentiles of the distribution of recursive estimates from an RCA(1) model with ACD(1,1) errors, which was discussed in Sect. 3.2. We again set $n = 4{,}000$ and $L = 100$. We simulated ε_i from a Gamma$(1, \theta)$ distribution with $\theta = 0.1$ and simulated z_i from a centered log-normal distribution with mean 0 and

Table 4 Percentiles of the distribution of recursive estimates from a RCA(1) model with ACD(1,1) errors from Sect. 3.2: $n = 4{,}000$, $L = 100$, $\theta = 0.1$, $\phi = 0.25$

True parameters	Estimates				
	5th	25th	50th	75th	95th
$\omega = 0.6$	0.510	0.543	0.598	0.639	0.691
$\alpha = 0.005$	−0.004	0.0001	0.006	0.010	0.014
$\beta = 0.15$	0.054	0.098	0.157	0.202	0.240
$\omega = 0.6$	0.510	0.546	0.595	0.639	0.689
$\alpha = 0.005$	−0.004	0.0001	0.006	0.010	0.013
$\beta = 0.35$	0.258	0.299	0.360	0.399	0.439
$\omega = 1.2$	1.112	1.146	1.198	1.239	1.285
$\alpha = 0.01$	0.001	0.005	0.011	0.015	0.019
$\beta = 0.15$	0.059	0.100	0.153	0.197	0.240
$\omega = 1.2$	1.111	1.145	1.197	1.239	1.285
$\alpha = 0.005$	−0.004	0.0003	0.006	0.010	0.014
$\beta = 0.35$	0.259	0.300	0.353	0.397	0.440

variance $\phi = 0.25$ (starting from a $N(0, 0.188)$ distribution). From Table 4, we see that the recursive estimation approach yields good estimates of the model parameters in this case as well.

5 Discussion and summary

Duration models are now widely used in applied statistics and econometrics. Thus, it is important to investigate the behavior of these models. In this paper, we have introduced a class of generalized duration models and shown that almost all the ACD and SCD models in the literature are special cases of this generalized model. Combined estimation using generalized martingale differences yields optimal estimates of the model parameters. Further, we have obtained a recursive estimation algorithm by extending the work in Thavaneswaran and Heyde (1999) to the multi-parameter setup. This algorithm is based on the function $g_c^*(\boldsymbol{\theta})$ and leads to optimal estimates which are more informative than the estimates based on the linear estimating function $g_m^*(\boldsymbol{\theta})$. We have provided details for three classes of duration models, viz., the multiplicative random coefficient ACD models, random coefficient models with ACD errors, and the log-SCD models, for which we have provided closed form expressions for $g_c^*(\boldsymbol{\theta})$ via the quadratic estimating function and shown that it is optimal. These details can be easily derived for all the models discussed in Sect. 2. The contribution of this paper is threefold. First, this article introduces a class of generalized duration models and provides a framework to integrate a number of ACD models in the literature. Second, the martingale estimating functions approach is described for the generalized duration models. Third, it is shown that combined estimating functions are optimal and are more informative than the component estimating functions under stated conditions. We expect that the recursive estimation described in this paper will enable practitioners to apply these methods in order to carry out fast estimation for long time series.

A Proof of Theorem 1

Proof We choose two orthogonal martingale differences $m_i(\boldsymbol{\theta}) = x_i - \mu_i(\boldsymbol{\theta})$ and $\Psi_i(\boldsymbol{\theta}) = M_i(\boldsymbol{\theta}) - \frac{\langle m,M\rangle_i}{\langle m\rangle_i} m_i(\boldsymbol{\theta})$, where the conditional variance of $\Psi_i(\boldsymbol{\theta})$ is given by $\langle \Psi\rangle_i = \langle M\rangle_i - \frac{\langle m,M\rangle_i^2}{\langle m\rangle_i}$. That is, $m_i(\boldsymbol{\theta})$ and $\Psi_i(\boldsymbol{\theta})$ are uncorrelated martingale differences with conditional variances $\langle m\rangle_i$ and $\langle \Psi\rangle_i$, respectively. Moreover, the optimal martingale estimating function and associated information based on the martingale differences $\Psi_i(\boldsymbol{\theta})$ are

$$\begin{aligned}
\mathbf{g}^*_{\Psi}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} \frac{\langle m, M\rangle_i}{\langle m\rangle_i} + E\left[\left. \frac{\partial M_i}{\partial \boldsymbol{\theta}} \right| \mathcal{F}^x_{i-1} \right] \right) \frac{\Psi_i}{\langle \Psi\rangle_i} \\
&= \sum_{i=1}^{n} \rho_i^2 \left(\left(-\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} \frac{\langle m, M\rangle_i^2}{\langle m\rangle_i^2 \langle M\rangle_i} - E\left[\left. \frac{\partial M_i}{\partial \boldsymbol{\theta}} \right| \mathcal{F}^x_{i-1} \right] \eta_i \right) m_i \right. \\
&\quad \left. + \left(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} \eta_i + E\left[\left. \frac{\partial M_i}{\partial \boldsymbol{\theta}} \right| \mathcal{F}^x_{i-1} \right] \frac{1}{\langle M\rangle_i} \right) M_i \right)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{I}_{\mathbf{g}^*_{\Psi}}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}'} \frac{\langle m, M\rangle_i}{\langle m\rangle_i} + E\left[\left. \frac{\partial M_i}{\partial \boldsymbol{\theta}} \right| \mathcal{F}^x_{i-1} \right] \right) \\
&\times \left(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}'} \frac{\langle m, M\rangle_i}{\langle m\rangle_i} + E\left[\left. \frac{\partial M_i}{\partial \boldsymbol{\theta}'} \right| \mathcal{F}^x_{i-1} \right] \right) \frac{1}{\langle \Psi\rangle_i} \\
&= \sum_{i=1}^{n} \rho_i^2 \left(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} \frac{\partial \mu_i}{\partial \boldsymbol{\theta}'} \frac{\langle m, M\rangle_i^2}{\langle m\rangle_i^2 \langle M\rangle_i} + E\left[\left. \frac{\partial M_i}{\partial \boldsymbol{\theta}} \frac{\partial M_i}{\partial \boldsymbol{\theta}'} \right| \mathcal{F}^x_{i-1} \right] \frac{1}{\langle M\rangle_i} \right. \\
&\quad \left. + \left(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} \frac{\partial M_i}{\partial \boldsymbol{\theta}'} + \frac{\partial M_i}{\partial \boldsymbol{\theta}} \frac{\partial \mu_i}{\partial \boldsymbol{\theta}'} \right) \eta_i \right).
\end{aligned}$$

Then the combined estimating function based on m_i and Ψ_i becomes

$$\begin{aligned}
g^*_C(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \rho_i^2 \left(\left(-\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} \frac{1}{\langle m\rangle_i} - E\left[\left. \frac{\partial M_i}{\partial \boldsymbol{\theta}} \right| \mathcal{F}^x_{i-1} \right] \eta_i \right) m_i \right. \\
&\quad \left. + \left(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} \eta_i + E\left[\left. \frac{\partial M_i}{\partial \boldsymbol{\theta}} \right| \mathcal{F}^x_{i-1} \right] \frac{1}{\langle M\rangle_i} \right) M_i \right),
\end{aligned}$$

and satisfies the sufficient condition for optimality

$$E\left[\left. \frac{\partial g_C(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right| \mathcal{F}^x_{i-1} \right] = \mathrm{Cov}(g_C(\boldsymbol{\theta}), g^*_C(\boldsymbol{\theta}) \mid \mathcal{F}^x_{i-1})\, K, \ \forall g_C(\boldsymbol{\theta}) \in \mathcal{G}_C,$$

where K is a constant matrix. Hence, $g^*_C(\boldsymbol{\theta})$ is optimal in the class $\mathcal{G}_C$, and (a) follows. Since m_i and Ψ_i are orthogonal, the information $I_{g^*_C}(\boldsymbol{\theta}) = I_{g^*_m}(\boldsymbol{\theta}) + I_{g^*_{\Psi}}(\boldsymbol{\theta})$ and (b) follows. Hence, neither $g^*_m(\boldsymbol{\theta})$ nor $g^*_M(\boldsymbol{\theta})$ is fully informative, that is, $I_{g^*_C}(\boldsymbol{\theta}) \geq I_{g^*_m}(\boldsymbol{\theta})$

and $I_{g^*_C}(\boldsymbol{\theta}) \geq I_{g^*_M}(\boldsymbol{\theta})$. Part (c) follows immediately from the fact that if $\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} \frac{\langle m, M\rangle_i}{\langle m\rangle_i} = -E\left[\left.\frac{\partial M_i}{\partial \boldsymbol{\theta}}\right| \mathcal{F}^x_{i-1}\right]$ for each i, then $g^*_\Psi(\boldsymbol{\theta}) = \mathbf{0}$ and $I_{g^*_\Psi}(\boldsymbol{\theta}) = \mathbf{0}$.

To prove (d), we note that the optimal combined estimating function based on $m_i(\boldsymbol{\theta})$ and $M_i(\boldsymbol{\theta})$ is given by $g^*_C(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{a}^*_{i-1}(\boldsymbol{\theta}) m_i(\boldsymbol{\theta}) + \mathbf{b}^*_{i-1}(\boldsymbol{\theta}) M_i(\boldsymbol{\theta})$. Using the Taylor expansion for $\mathbf{g}^*_C(\boldsymbol{\theta})$ and substituting the recursive estimate for $\boldsymbol{\theta}$ at each step, the estimate based on the first $i-1$ observations is given by

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{i-1} &= \left\{\sum_{s=1}^{i-1}\left(\mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial}{\partial \boldsymbol{\theta}} m_s(\widehat{\boldsymbol{\theta}}_{s-1}) + \frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1}) m_s(\widehat{\boldsymbol{\theta}}_{s-1}) + \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial}{\partial \boldsymbol{\theta}} M_s(\widehat{\boldsymbol{\theta}}_{s-1})\right.\right.\\ &\quad \left.\left. + \frac{\partial \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} M_s(\widehat{\boldsymbol{\theta}}_{s-1})\right)\right\}^{-1}\left(-\sum_{s=1}^{i-1}\left(\mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1}) m_s(\widehat{\boldsymbol{\theta}}_{s-1}) + \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1}) M_s(\widehat{\boldsymbol{\theta}}_{s-1})\right)\right.\\ &\quad + \sum_{s=1}^{i-1}\left(\mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial m_s(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} m_s(\widehat{\boldsymbol{\theta}}_{s-1}) + \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial M_s(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}}\right.\\ &\quad \left.\left. + \frac{\partial \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} M_s(\widehat{\boldsymbol{\theta}}_{s-1})\right)\widehat{\boldsymbol{\theta}}_{s-1}\right).\end{aligned}$$

When the ith observation becomes available, the estimate becomes

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{i} &= \left\{\sum_{s=1}^{i}\left(\mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial m_s(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} m_s(\widehat{\boldsymbol{\theta}}_{s-1}) + \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial M_s(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}}\right.\right.\\ &\quad \left.\left. + \frac{\partial \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} M_s(\widehat{\boldsymbol{\theta}}_{s-1})\right)\right\}^{-1}\left(-\sum_{s=1}^{i}\left(\mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1}) m_s(\widehat{\boldsymbol{\theta}}_{s-1}) + \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1}) M_s(\widehat{\boldsymbol{\theta}}_{s-1})\right)\right.\\ &\quad + \sum_{s=1}^{i}\left(\mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial m_s(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} m_s(\widehat{\boldsymbol{\theta}}_{s-1}) + \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial M_s(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}}\right.\\ &\quad \left.\left. + \frac{\partial \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} M_s(\widehat{\boldsymbol{\theta}}_{s-1})\right)\widehat{\boldsymbol{\theta}}_{s-1}\right).\end{aligned}$$

Let

$$\begin{aligned}\mathbf{K}_i^{-1} &= -\sum_{s=1}^{i}\left(\mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial m_s(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{a}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} m_s(\widehat{\boldsymbol{\theta}}_{s-1})\right.\\ &\quad \left. + \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})\frac{\partial M_s(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{b}^*_{s-1}(\widehat{\boldsymbol{\theta}}_{s-1})}{\partial \boldsymbol{\theta}} M_s(\widehat{\boldsymbol{\theta}}_{s-1})\right),\end{aligned}$$

then

$$\begin{aligned}\mathbf{K}_i^{-1} &= \mathbf{K}_{i-1}^{-1} - \mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})\frac{\partial m_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{a}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} m_i(\widehat{\boldsymbol{\theta}}_{i-1})\\ &\quad + \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})\frac{\partial M_i(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{b}^*_{i-1}(\widehat{\boldsymbol{\theta}}_{i-1})}{\partial \boldsymbol{\theta}} M_i(\widehat{\boldsymbol{\theta}}_{i-1}),\end{aligned}$$

and

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_{i-1} &= \mathbf{K}_i(\mathbf{K}_i^{-1}\widehat{\boldsymbol{\theta}}_i - \mathbf{K}_i^{-1}\widehat{\boldsymbol{\theta}}_{i-1}) \\ &= \mathbf{K}_i(\mathbf{a}_{i-1}^*(\widehat{\boldsymbol{\theta}}_{i-1})m_i(\widehat{\boldsymbol{\theta}}_{i-1}) + \mathbf{b}_{i-1}^*(\widehat{\boldsymbol{\theta}}_{i-1})M_i(\widehat{\boldsymbol{\theta}}_{i-1})).\end{aligned}$$

Hence it is easy to show that the recursive equations for $\boldsymbol{\theta}$ take the form (11)–(12). Note that the proofs of (a) and (b) are somewhat similar to the proof given in Liang et al. (2011). Part (d) extends the results in Thavaneswaran and Heyde (1999) to the generalized combined estimating function $g_C^*(\boldsymbol{\theta})$ with a vector-valued parameter. □

A Proof of Theorem 2

Proof Since

$$\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} = \mu_\varepsilon \left(\frac{\partial \mu_z}{\partial \boldsymbol{\phi}'}, 1, x_{i-1}, \ldots, x_{i-p}, \psi_{i-1}, \ldots, \psi_{i-q}\right)'$$

and

$$\frac{\partial \sigma_i^2}{\partial \boldsymbol{\theta}} = \Bigg(2\sigma_\varepsilon^2(\mu_z + \psi_i)\frac{\partial \mu_z}{\partial \boldsymbol{\phi}'} + (\mu_\varepsilon^2 + \sigma_\varepsilon^2)\frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}'}, 2\sigma_\varepsilon^2(\mu_z + \psi_i)x_{t-1}, \ldots, \\ 2\sigma_\varepsilon^2(\mu_z + \psi_i)x_{t-p}, 2\sigma_\varepsilon^2(\mu_z + \psi_i)\psi_{i-1}, \ldots, 2\sigma_\varepsilon^2(\mu_z + \psi_i)\psi_{t-q}\Bigg)',$$

the quadratic estimating function for each component of $\boldsymbol{\theta}$ is given by

$$\begin{aligned} g_C^*(\boldsymbol{\phi}) &= \sum_{i=1}^n \rho_i^2(\mathbf{r}_i m_i + \mathbf{s}_i M_i), \\ g_C^*(\omega) &= \sum_{i=1}^n \rho_i^2(R_i m_i + S_i M_i), \\ g_C^*(\alpha_k) &= \sum_{i=1}^n \rho_i^2 x_{i-k}(R_i m_i + S_i M_i), k = 1, \ldots, p, \\ g_C^*(\beta_j) &= \sum_{i=1}^n \rho_i^2 \psi_{i-j}(R_i m_i + S_i M_i), j = 1, \ldots, q, \end{aligned}$$

where

$$\rho_i^2 = \left(1 - \frac{\langle m, M\rangle_i^2}{\langle m\rangle_i \langle M\rangle_i}\right)^{-1},$$

$$\langle m\rangle_i \langle M\rangle_i \mathbf{r}_i = -\mu_\varepsilon \frac{\partial \mu_z}{\partial \boldsymbol{\phi}}\langle M\rangle_i + \left(2\sigma_\varepsilon^2(\mu_z + \psi_i)\frac{\partial \mu_z}{\partial \boldsymbol{\phi}} + (\mu_\varepsilon^2 + \sigma_\varepsilon^2)\frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}}\right)\langle m, M\rangle_i,$$

$$
\begin{aligned}
\langle m\rangle_i\langle M\rangle_i R_i &= -\mu_\varepsilon\langle M\rangle_i + 2\sigma_\varepsilon^2(\mu_z+\psi_i)\langle m, M\rangle_i,\\
\langle m\rangle_i\langle M\rangle_i \mathbf{s}_i &= \mu_\varepsilon\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\langle m, M\rangle_i - \left(2\sigma_\varepsilon^2(\mu_z+\psi_i)\frac{\partial\mu_z}{\partial\boldsymbol{\phi}} + (\mu_\varepsilon^2+\sigma_\varepsilon^2)\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}}\right)\langle m\rangle_i,\\
\langle m\rangle_i\langle M\rangle_i S_i &= \mu_\varepsilon\langle m, M\rangle_i - 2\sigma_\varepsilon^2(\mu_z+\psi_i)\langle m\rangle_i.
\end{aligned}
$$

Moreover, the information matrix corresponding to the optimal quadratic estimating function for $\boldsymbol{\theta}$ is given by

$$
I_{g_C^*}(\boldsymbol{\theta}) = \begin{pmatrix} I_{\boldsymbol{\phi}\boldsymbol{\phi}'} & I_{\boldsymbol{\phi}\omega} & I_{\boldsymbol{\phi}\boldsymbol{\alpha}'} & I_{\boldsymbol{\phi}\boldsymbol{\beta}'} \\ I'_{\boldsymbol{\phi}\omega} & I_{\omega\omega} & I_{\omega\boldsymbol{\alpha}'} & I_{\omega\boldsymbol{\beta}'} \\ I'_{\boldsymbol{\phi}\boldsymbol{\alpha}'} & I'_{\omega\boldsymbol{\alpha}'} & I_{\boldsymbol{\alpha}\boldsymbol{\alpha}'} & I_{\boldsymbol{\alpha}\boldsymbol{\beta}'} \\ I'_{\boldsymbol{\phi}\boldsymbol{\beta}'} & I'_{\omega\boldsymbol{\beta}'} & I'_{\boldsymbol{\alpha}\boldsymbol{\beta}'} & I_{\boldsymbol{\beta}\boldsymbol{\beta}'} \end{pmatrix},
$$

where

$$
\begin{aligned}
I_{\boldsymbol{\phi}\boldsymbol{\phi}'} &= \sum_{i=1}^n \rho_i^2\left(\frac{4\sigma_\varepsilon^2(\mu_z+\psi_i)^2\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\frac{\partial\mu_z}{\partial\boldsymbol{\phi}'} + (\mu_\varepsilon^2+\sigma_\varepsilon^2)^2\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}}\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}'} + 4\sigma_\varepsilon^2(\mu_z+\psi_i)(\mu_\varepsilon^2+\sigma_\varepsilon^2)\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}'}}{\langle M\rangle_i}\right.\\
&\quad\left. - \frac{\left(4\mu_\varepsilon\sigma_\varepsilon^2(\mu_z+\psi_i)\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\frac{\partial\mu_z}{\partial\boldsymbol{\phi}'} + 2\mu_\varepsilon(\mu_\varepsilon^2+\sigma_\varepsilon^2)\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}'}\right)\langle m, M\rangle_i}{\langle m\rangle_i\langle M\rangle_i} + \frac{\mu_\varepsilon^2\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\frac{\partial\mu_z}{\partial\boldsymbol{\phi}'}}{\langle m\rangle_i}\right)
\end{aligned}
$$

$$
\begin{aligned}
I_{\boldsymbol{\phi}\omega} &= \sum_{i=1}^n \rho_i^2\left(\frac{\mu_\varepsilon^2\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}}{\langle m\rangle_i} + \frac{4\sigma_\varepsilon^4(\mu_z+\psi_i)^2\frac{\partial\mu_z}{\partial\boldsymbol{\phi}} + 2\sigma_\varepsilon^2(\mu_\varepsilon^2+\sigma_\varepsilon^2)(\mu_z+\psi_i)\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}}}{\langle M\rangle_i}\right.\\
&\quad\left. - \frac{4\mu_\varepsilon\sigma_\varepsilon^2(\mu_z+\psi_i)\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\langle m, M\rangle_i}{\langle m\rangle_i\langle M\rangle_i}\right)
\end{aligned}
$$

$$
\begin{aligned}
I_{\omega\omega} &= \sum_{i=1}^n \rho_i^2\left(\frac{\mu_\varepsilon^2}{\langle m\rangle_i} + \frac{4\sigma_\varepsilon^4(\mu_z+\psi_i)^2}{\langle M\rangle_i} - \frac{4\mu_\varepsilon\sigma_\varepsilon^2(\mu_z+\psi_i)\langle m, M\rangle_i}{\langle m\rangle_i\langle M\rangle_i}\right)\\
I_{\boldsymbol{\phi}\boldsymbol{\alpha}'} &= \left(\sum_{i=1}^n \rho_i^2 x_{i-k}\left(\frac{\mu_\varepsilon^2\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}}{\langle m\rangle_i} + \frac{4\sigma_\varepsilon^4(\mu_z+\psi_i)^2\frac{\partial\mu_z}{\partial\boldsymbol{\phi}} + 2\sigma_\varepsilon^2(\mu_\varepsilon^2+\sigma_\varepsilon^2)(\mu_z+\psi_i)\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}}}{\langle M\rangle_i}\right.\right.\\
&\quad\left.\left. - \frac{4\mu_\varepsilon\sigma_\varepsilon^2(\mu_z+\psi_i)\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\langle m, M\rangle_i}{\langle m\rangle_i\langle M\rangle_i}\right)\right)_{k=1,\ldots,p}
\end{aligned}
$$

$$
\begin{aligned}
I'_{\omega\boldsymbol{\alpha}'} &= \left(\sum_{i=1}^n \rho_i^2 x_{i-k}\left(\frac{\mu_\varepsilon^2}{\langle m\rangle_i} + \frac{4\sigma_\varepsilon^2(\mu_z+\psi_i)^2}{\langle M\rangle_i} - \frac{4\mu_\varepsilon\sigma_\varepsilon^2(\mu_z+\psi_i)\langle m, M\rangle_i}{\langle m\rangle_i\langle M\rangle_i}\right)\right)_{k=1,\ldots,p}\\
I_{\boldsymbol{\alpha}\boldsymbol{\alpha}'} &= \left(\sum_{i=1}^n \rho_i^2 x_{i-k}x_{i-j}\left(\frac{\mu_\varepsilon^2}{\langle m\rangle_i} + \frac{4\sigma_\varepsilon^2(\mu_z+\psi_i)^2}{\langle M\rangle_i} - \frac{4\mu_\varepsilon\sigma_\varepsilon^2(\mu_z+\psi_i)\langle m, M\rangle_i}{\langle m\rangle_i\langle M\rangle_i}\right)\right)_{k,j=1,\ldots,p}
\end{aligned}
$$

$$
\begin{aligned}
I_{\boldsymbol{\phi}\boldsymbol{\beta}'} &= \left(\sum_{i=1}^{n} \rho_i^2 \psi_{i-j} \left(\frac{\mu_\varepsilon^2 \frac{\partial \mu_z}{\partial \boldsymbol{\phi}}}{\langle m \rangle_i} + \frac{4\sigma_\varepsilon^4 (\mu_z + \psi_i)^2 \frac{\partial \mu_z}{\partial \boldsymbol{\phi}} + 2\sigma_\varepsilon^2 (\mu_\varepsilon^2 + \sigma_\varepsilon^2)(\mu_z + \psi_i) \frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}}}{\langle M \rangle_i} \right.\right.\\
&\quad \left.\left. - \frac{4\mu_\varepsilon \sigma_\varepsilon^2 (\mu_z + \psi_i) \frac{\partial \mu_z}{\partial \boldsymbol{\phi}} \langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i}\right)\right)_{j=1,\ldots,q}\\
I'_{\omega\boldsymbol{\beta}'} &= \left(\sum_{i=1}^{n} \rho_i^2 \psi_{i-j} \left(\frac{\mu_\varepsilon^2}{\langle m \rangle_i} + \frac{4\sigma_\varepsilon^2 (\mu_z + \psi_i)^2}{\langle M \rangle_i} - \frac{4\mu_\varepsilon \sigma_\varepsilon^2 (\mu_z + \psi_i) \langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i}\right)\right)_{j=1,\ldots,q}\\
I_{\boldsymbol{\alpha}\boldsymbol{\beta}'} &= \left(\sum_{i=1}^{n} \rho_i^2 x_{i-k} \psi_{i-j} \left(\frac{\mu_\varepsilon^2}{\langle m \rangle_i} + \frac{4\sigma_\varepsilon^2 (\mu_z + \psi_i)^2}{\langle M \rangle_i} - \frac{4\mu_\varepsilon \sigma_\varepsilon^2 (\mu_z + \psi_i) \langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i}\right)\right)_{k=1,\ldots,p,\, j=1,\ldots,q}\\
I_{\boldsymbol{\beta}\boldsymbol{\beta}'} &= \left(\sum_{i=1}^{n} \rho_i^2 \psi_{i-k} \psi_{i-j} \left(\frac{\mu_\varepsilon^2}{\langle m \rangle_i} + \frac{4\sigma_\varepsilon^2 (\mu_z + \psi_i)^2}{\langle M \rangle_i} - \frac{4\mu_\varepsilon \sigma_\varepsilon^2 (\mu_z + \psi_i) \langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i}\right)\right)_{k,j=1,\ldots,q}
\end{aligned}
$$

□

A Proof of Theorem 3

Proof Since

$$
\begin{aligned}
\frac{\partial \mu_i}{\partial \boldsymbol{\theta}} &= \left(x_{i-1} + \frac{\partial \mu_\varepsilon(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} x_{i-1}, \frac{\partial \mu_\varepsilon(\theta)}{\partial \theta} \psi_i, \mu_\varepsilon, \right.\\
&\qquad \left. \mu_\varepsilon x_{i-1}, \ldots, \mu_\varepsilon x_{t-p}, \mu_\varepsilon \psi_{i-1}, \ldots, \mu_\varepsilon \psi_{t-q}\right)',\\
\frac{\partial \sigma_i^2}{\partial \boldsymbol{\theta}} &= \left(\frac{\partial \sigma_\varepsilon^2(\theta)}{\partial \theta} \psi_i^2, \frac{\partial \sigma_\varepsilon^2(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} x_{i-1}^2, 2\sigma_\varepsilon^2 \psi_i, \right.\\
&\qquad \left. 2\sigma_\varepsilon^2 \psi_i x_{i-1}, \ldots, 2\sigma_\varepsilon^2 \psi_i x_{t-p}, 2\sigma_\varepsilon^2 \psi_i \psi_{i-1}, \ldots, 2\sigma_\varepsilon^2 \psi_i \psi_{t-q}\right)',
\end{aligned}
$$

the quadratic estimating function for each component of $\boldsymbol{\theta}$ is given by

$$
\begin{aligned}
g_C^*(\theta) &= \sum_{i=1}^{n} \rho_i^2 (v_{i1} m_i + w_{i1} M_i),\\
g_C^*(\boldsymbol{\phi}) &= \sum_{i=1}^{n} \rho_i^2 (v_{i2} m_i + w_{i2} M_i),\\
g_C^*(\omega) &= \sum_{i=1}^{n} \rho_i^2 (V_i m_i + W_i M_i),\\
g_C^*(\alpha_k) &= \sum_{i=1}^{n} \rho_i^2 x_{i-k} (V_i m_i + W_i M_i), k = 1, \ldots, p,\\
g_C^*(\beta_j) &= \sum_{i=1}^{n} \rho_i^2 \psi_{i-j} (V_i m_i + W_i M_i), j = 1, \ldots, q, \qquad (16)
\end{aligned}
$$

where

$$
\begin{aligned}
\langle m\rangle_i\langle M\rangle_i v_{i1} &= -\left(x_{i-1}+\frac{\partial\mu_\varepsilon(\theta)}{\partial\theta}\psi_i\right)\langle M\rangle_i+\frac{\partial\sigma_\varepsilon^2(\theta)}{\partial\theta}\psi_i\langle m,M\rangle_i,\\
\langle m\rangle_i\langle M\rangle_i v_{i2} &= -\frac{\partial\mu_z(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}x_{i-1}\langle M\rangle_i+\frac{\partial\sigma_z^2(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}x_{i-1}^2\langle m,M\rangle_i,\\
\langle m\rangle_i\langle M\rangle_i V_i &= -\mu_\varepsilon\langle M\rangle_i+2\sigma_\varepsilon^2\psi_i\langle m,M\rangle_i,\\
\langle m\rangle_i\langle M\rangle_i w_{i1} &= \left(x_{i-1}+\frac{\partial\mu_\varepsilon(\theta)}{\partial\theta}\psi_i\right)\langle m,M\rangle_i-\frac{\partial\sigma_\varepsilon^2(\theta)}{\partial\theta}\psi_i^2\langle m\rangle_i,\\
\langle m\rangle_i\langle M\rangle_i w_{i2} &= \frac{\partial\mu_z(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}x_{i-1}\langle m,M\rangle_i-\frac{\partial\sigma_z^2(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}x_{i-1}^2\langle m\rangle_i\\
\langle m\rangle_i\langle M\rangle_i W_i &= \mu_\varepsilon\langle m,M\rangle_i-2\sigma_\varepsilon^2\psi_i\langle m\rangle_i.
\end{aligned}
$$

Moreover, the information matrix of the optimal quadratic estimating function for $\boldsymbol{\theta}$ is given by

$$
I_{g_C^*}(\boldsymbol{\theta})=\begin{pmatrix}
I_{\theta\theta} & I_{\theta\boldsymbol{\phi}'} & I_{\theta\omega} & I_{\theta\boldsymbol{\alpha}'} & I_{\theta\boldsymbol{\beta}'}\\
I'_{\theta\boldsymbol{\phi}'} & I_{\boldsymbol{\phi}\boldsymbol{\phi}'} & I_{\boldsymbol{\phi}\omega} & I_{\boldsymbol{\phi}\boldsymbol{\alpha}'} & I_{\boldsymbol{\phi}\boldsymbol{\beta}'}\\
I'_{\theta\omega} & I'_{\boldsymbol{\phi}\omega} & I_{\omega\omega} & I_{\omega\boldsymbol{\alpha}'} & I_{\omega\boldsymbol{\beta}'}\\
I'_{\theta\boldsymbol{\alpha}'} & I'_{\boldsymbol{\phi}\boldsymbol{\beta}'} & I'_{\omega\boldsymbol{\alpha}'} & I_{\boldsymbol{\alpha}\boldsymbol{\alpha}'} & I_{\boldsymbol{\alpha}\boldsymbol{\beta}'}\\
I'_{\theta\boldsymbol{\beta}'} & I'_{\boldsymbol{\phi}\boldsymbol{\beta}'} & I'_{\omega\boldsymbol{\beta}'} & I'_{\boldsymbol{\alpha}\boldsymbol{\beta}'} & I_{\boldsymbol{\beta}\boldsymbol{\beta}'}
\end{pmatrix},
$$

where

$$
\begin{aligned}
I_{\theta\theta} &= \sum_{i=1}^n\rho_i^2\left(\frac{\left(x_{i-1}+\frac{\partial\mu_\varepsilon(\theta)}{\partial\theta}\psi_i\right)^2}{\langle m\rangle_i}+\frac{\left(\frac{\partial\sigma_\varepsilon^2(\theta)}{\partial\theta}\right)^2\psi_i^4}{\langle M\rangle_i}-\frac{2\left(x_{i-1}+\frac{\partial\mu_\varepsilon(\theta)}{\partial\theta}\psi_i\right)\frac{\partial\sigma_\varepsilon^2(\theta)}{\partial\theta}\psi_i^2\langle m,M\rangle_i}{\langle m\rangle_i\langle M\rangle_i}\right)\\
I_{\theta\boldsymbol{\phi}'} &= \sum_{i=1}^n\rho_i^2\left(\frac{\left(x_{i-1}+\frac{\partial\mu_\varepsilon(\theta)}{\partial\theta}\psi_i\right)x_{i-1}\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}}{\langle m\rangle_i}+\frac{\frac{\partial\sigma_\varepsilon^2(\theta)}{\partial\theta}\psi_i^2x_{i-1}^2\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}}}{\langle M\rangle_i}\right.\\
&\quad\left.-\frac{\left(\left(x_{i-1}+\frac{\partial\mu_\varepsilon(\theta)}{\partial\theta}\psi_i\right)x_{i-1}^2\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}}+\frac{\partial\sigma_\varepsilon^2(\theta)}{\partial\theta}\psi_i^2x_{i-1}\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\right)\langle m,M\rangle_i}{\langle m\rangle_i\langle M\rangle_i}\right)\\
I_{\boldsymbol{\phi}\boldsymbol{\phi}'} &= \sum_{i=1}^n\rho_i^2\left(\frac{x_{i-1}^2\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\frac{\partial\mu_z}{\partial\boldsymbol{\phi}'}}{\langle m\rangle_i}+\frac{x_{i-1}^4\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}}\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}'}}{\langle M\rangle_i}-\frac{2x_{i-1}^3\frac{\partial\mu_z}{\partial\boldsymbol{\phi}}\frac{\partial\sigma_z^2}{\partial\boldsymbol{\phi}'}\langle m,M\rangle_i}{\langle m\rangle_i\langle M\rangle_i}\right)\\
I_{\theta\omega} &= \sum_{i=1}^n\rho_i^2\left(\frac{\mu_\varepsilon\left(x_{i-1}+\frac{\partial\mu_\varepsilon(\theta)}{\partial\theta}\psi_i\right)}{\langle m\rangle_i}+\frac{\sigma_\varepsilon^2(\theta)\frac{\partial\sigma_\varepsilon^2(\theta)}{\partial\theta}\psi_i}{\langle M\rangle_i}\right.\\
&\quad\left.-\frac{\left(\sigma_\varepsilon^2(\theta)\left(x_{i-1}+\frac{\partial\mu_\varepsilon(\theta)}{\partial\theta}\psi_i\right)+\mu_\varepsilon(\theta)\frac{\partial\sigma_\varepsilon^2(\theta)}{\partial\theta}\psi_i\right)\langle m,M\rangle_i}{\langle m\rangle_i\langle M\rangle_i}\right)
\end{aligned}
$$

$$I_{\boldsymbol{\phi}\omega} = \sum_{i=1}^{n} \rho_i^2 \left(\frac{\mu_\varepsilon x_{i-1} \frac{\partial \mu_z}{\partial \boldsymbol{\phi}}}{\langle m \rangle_i} + \frac{2\sigma_\varepsilon^2 x_{i-1}^2 \psi_i \frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}}}{\langle M \rangle_i} - \frac{\left(2\sigma_\varepsilon^2 x_{i-1}\psi_i - \mu_\varepsilon x_{i-1}^2 \frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}}\right) \langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i} \right)$$

$$I_{\omega\omega} = \sum_{i=1}^{n} \frac{\rho_i^2 (\mu_\varepsilon^2 \langle M \rangle_i + 4\sigma_\varepsilon^4 \psi_i^2 \langle m \rangle_i - 4\mu_\varepsilon \sigma_\varepsilon^2 \psi_i \langle m, M \rangle_i)}{\langle m \rangle_i \langle M \rangle_i}$$

$$I_{\theta\boldsymbol{\alpha}'} = \left(\sum_{i=1}^{n} \rho_i^2 x_{i-k} \left(\frac{\mu_\varepsilon \left(x_{i-1} + \frac{\partial \mu_\varepsilon(\theta)}{\partial \theta} \psi_i \right)}{\langle m \rangle_i} + \frac{\sigma_\varepsilon^2(\theta) \frac{\partial \sigma_\varepsilon^2(\theta)}{\partial \theta} \psi_i}{\langle M \rangle_i} \right.\right.$$
$$\left.\left. - \frac{\left(\sigma_\varepsilon^2(\theta) \left(x_{i-1} + \frac{\partial \mu_\varepsilon(\theta)}{\partial \theta} \psi_i \right) + \mu_\varepsilon(\theta) \frac{\partial \sigma_\varepsilon^2(\theta)}{\partial \theta} \psi_i \right) \langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i} \right) \right)_{k=1,\ldots,p}$$

$$I_{\boldsymbol{\phi}\boldsymbol{\alpha}'} = \left(\sum_{i=1}^{n} \rho_i^2 x_{i-k} \left(\frac{\mu_\varepsilon x_{i-1} \frac{\partial \mu_z}{\partial \boldsymbol{\phi}}}{\langle m \rangle_i} + \frac{2\sigma_\varepsilon^2 x_{i-1}^2 \psi_i \frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}}}{\langle M \rangle_i} - \frac{\left(2\sigma_\varepsilon^2 x_{i-1}\psi_i - \mu_\varepsilon x_{i-1}^2 \frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}}\right) \langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i} \right) \right)_{k=1,\ldots,p}$$

$$I'_{\omega\boldsymbol{\alpha}'} = \left(\sum_{i=1}^{n} \frac{\rho_i^2 x_{i-k} \left(\mu_\varepsilon^2 \langle M \rangle_i + 4\sigma_\varepsilon^4 \psi_i^2 \langle m \rangle_i - 4\mu_\varepsilon \sigma_\varepsilon^2 \psi_i \langle m, M \rangle_i \right)}{\langle m \rangle_i \langle M \rangle_i} \right)_{k=1,\ldots,p}$$

$$I_{\boldsymbol{\alpha}\boldsymbol{\alpha}'} = \left(\sum_{i=1}^{n} \frac{\rho_i^2 x_{i-k} x_{i-j} \left(\mu_\varepsilon^2 \langle M \rangle_i + 4\sigma_\varepsilon^4 \psi_i^2 \langle m \rangle_i - 4\mu_\varepsilon \sigma_\varepsilon^2 \psi_i \langle m, M \rangle_i \right)}{\langle m \rangle_i \langle M \rangle_i} \right)_{k,j=1,\ldots,p}$$

$$I_{\theta\boldsymbol{\beta}'} = \left(\sum_{i=1}^{n} \rho_i^2 \psi_{i-j} \left(\frac{\mu_\varepsilon \left(x_{i-1} + \frac{\partial \mu_\varepsilon(\theta)}{\partial \theta} \psi_i \right)}{\langle m \rangle_i} + \frac{\sigma_\varepsilon^2(\theta) \frac{\partial \sigma_\varepsilon^2(\theta)}{\partial \theta} \psi_i}{\langle M \rangle_i} \right.\right.$$
$$\left.\left. - \frac{\left(\sigma_\varepsilon^2(\theta) \left(x_{i-1} + \frac{\partial \mu_\varepsilon(\theta)}{\partial \theta} \psi_i \right) + \mu_\varepsilon(\theta) \frac{\partial \sigma_\varepsilon^2(\theta)}{\partial \theta} \psi_i \right) \langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i} \right) \right)_{j=1,\ldots,q}$$

$$I_{\boldsymbol{\phi}\boldsymbol{\beta}'} = \left(\sum_{i=1}^{n} \rho_i^2 \psi_{i-j} \left(\frac{\mu_\varepsilon x_{i-1} \frac{\partial \mu_z}{\partial \boldsymbol{\phi}}}{\langle m \rangle_i} + \frac{2\sigma_\varepsilon^2 x_{i-1}^2 \psi_i \frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}}}{\langle M \rangle_i} - \frac{\left(2\sigma_\varepsilon^2 x_{i-1}\psi_i - \mu_\varepsilon x_{i-1}^2 \frac{\partial \sigma_z^2}{\partial \boldsymbol{\phi}}\right) \langle m, M \rangle_i}{\langle m \rangle_i \langle M \rangle_i} \right) \right)_{j=1,\ldots,q}$$

$$I_{\omega\boldsymbol{\beta}'} = \left(\sum_{i=1}^{n} \frac{\rho_i^2 \psi_{i-j} \left(\mu_\varepsilon^2 \langle M \rangle_i + 4\sigma_\varepsilon^4 \psi_i^2 \langle m \rangle_i - 4\mu_\varepsilon \sigma_\varepsilon^2 \psi_i \langle m, M \rangle_i \right)}{\langle m \rangle_i \langle M \rangle_i} \right)_{j=1,\ldots,q}$$

$$I_{\boldsymbol{\alpha}\boldsymbol{\beta}'} = \left(\sum_{i=1}^{n} \frac{\rho_i^2 x_{i-k} \psi_{i-j} \left(\mu_\varepsilon^2 \langle M \rangle_i + 4\sigma_\varepsilon^4 \psi_i^2 \langle m \rangle_i - 4\mu_\varepsilon \sigma_\varepsilon^2 \psi_i \langle m, M \rangle_i \right)}{\langle m \rangle_i \langle M \rangle_i} \right)_{k=1,\ldots,p,\, j=1,\ldots,q}$$

$$I_{\boldsymbol{\beta}\boldsymbol{\beta}'} = \left(\sum_{i=1}^{n} \frac{\rho_i^2 \psi_{i-k} \psi_{i-j} \left(\mu_\varepsilon^2 \langle M \rangle_i + 4\sigma_\varepsilon^4 \psi_i^2 \langle m \rangle_i - 4\mu_\varepsilon \sigma_\varepsilon^2 \psi_i \langle m, M \rangle_i \right)}{\langle m \rangle_i \langle M \rangle_i} \right)_{k,j=1,\ldots,q}$$

□

Acknowledgments We are grateful to the reviewers, associate editor, and editor for their suggestions that considerably improved the paper. The work of the first author was supported by a grant from NSERC.

References

Allen, D., Chan, F., McAleer, M., Peiris, S. (2008). Finite sample properties of QMLE for the log-ACD model: Application to Australian stocks. *Journal of Econometrics*, *147*, 163–185.

Bauwens, L., Giot, P. (2000). The logarithmic ACD model: An application to the Bid-ask quote process of three NYSE stocks. *Annales d' Économie et de Statistique*, *60*, 117–149.

Bauwens, L., Giot, P. (2003). Asymmetric ACD models: Introducing price information in ACD models. *Empirical Economics*, *28*, 709–731.

Bauwens, L., Veredas, D. (2004). The stochastic conditional duration model: A latent factor model for the analysis of financial durations. *Journal of Econometrics*, *119*, 381–412.

Bauwens, L., Galli, F., Giot, P. (2008). The moments of Log-ACD models. *Qualitative and Quantitative Analysis in Social Sciences*, *2*, 1–28.

Bera, A. K., Bilias, Y., Simlai, P. (2006). Estimating functions and equations: An essay on historical developments with applications to econometrics. In: T. C. Mills, K. Patterson (eds.), *Palgrave Handbook of Econometrics* (Vol. I, pp. 427–476). NSW: Palgrave MacMillan.

Broto, C., Ruiz, E. (2004). Estimation method for stochastic volatility models: a survey. *Journal of Economic Surveys*, *18*, 613–637.

Engle, R. F., Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, *66*, 1127–1162.

Fernandes, M., Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics*, *130*, 1–23.

Ghahramani, M., Thavaneswaran, A. (2009). Combining estimating functions for volatility. *Journal of Statistical Planning and Inference*, *139*, 1449–1461.

Ghahramani, M., Thavaneswaran, A. (2012). Nonlinear recursive estimation of the volatility via estimating functions. *Journal of Statistical Planning and Inference*, *142*, 171–180.

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic process. *Biometrika*, *72*, 319–328.

Jacquier, E., Polson, N. G., Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, *12*, 371–389.

Jasiak, J. (1998). Persistence in intratrade durations. *Finance*, *19*, 166–195.

Liang, Y., Thavaneswaran, A., Abraham, B. (2011). Joint estimation using quadratic estimating functions. *Journal of Probability and Statistics*, *2011*(2011), Article ID 372512.

Lindsay, B. G. (1985). Using empirical partially Bayes inference for increased efficiency. *Annals of Statistics*, *13*, 914–931.

Merkouris, T. (2007). Transform martingale estimating functions. *The Annals of Statistics*, *35*, 1975–2000.

Naik-Nimbalkar, U. V., Rajarshi, M. B. (1995). Filtering and smoothing via estimating functions. *Journal of the American Statistical Association*, *90*, 301–306.

Pacurar, M. (2008). Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys*, *22*, 711–751.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, *63*, 289–306.

Thavaneswaran, A., Abraham, B. (1988). Estimation of nonlinear time series models using estimating functions. *Journal of Time Series Analysis*, *9*, 99–108.

Thavaneswaran, A., Ghahramani, M. (2011). On some properties of stochastic conditional duration models. *Journal of Statistical Theory and Applications*, *4*, 571–580.

Thavaneswaran, A., Heyde, C. C. (1999). Prediction via estimating functions. *Journal of Statistical Planning and Inference*, *77*, 89–101.

Thavaneswaran, A., Liang, Y., Frank, J. (2012). Inference for random coefficient volatility models. *Statistics and Probability Letters*, *82*, 2086–2090.