Minimum Distance Estimation of Long-Memory Stochastic Duration Models

Mauricio Zevallos *

Abstract

This paper proposes a minimum distance estimator for long-memory stochastic duration models which satisfies a central limit theorem. Distinctive features of the proposed method are: it is easy to calculate and implement, allows fast estimation even for huge data sets, and provides asymptotic standard errors for the estimators. Monte Carlo experiments indicate that the proposed estimator performs very well. The proposed method is illustrated with the estimation of a real-life time series of nearly a million observations.

Keywords: Autocovariance differences, high frequency, intertrade durations.
MSC classification: 62M10
JEL classifications: C22

1 Introduction

The topic of modelling intertrade durations of financial assets, pioneered by the study of Engle and Russell (1998), has attracted substantial attention in the financial econometric literature; see the reviews of Pacurar (2008) and Bhogal and Variyam (2018).

Duration models can be classified in two classes: observation-driven and parameter-driven. In the first class, the Autoregressive Conditional Duration (ACD) model proposed by Engle and Russell (1998)-and its extensions, consider that the conditional expected duration process given the past is a function of past durations, and maximum likelihood estimates are easily obtained. In the second class, the process is driven by an unobserved latent variable and the likelihood is expressed as a multiple integral with dimension equal to the (durations) time series size. A very popular model in this class is the Stochastic Conditional Duration (SCD) model of Bauwens and Veredas (2004).

The cited ACD and SCD models have short-memory, and when applied to financial durations they tipically find that the degree of persistence is very close to one, indicating that long-memory

*Affiliation: Department of Statistics, University of Campinas, Brazil. Correspondence to: Mauricio Zevallos, IMECC-UNICAMP, Rua Sérgio Buarque de Holanda 651, Cidade Universitária, Barão Geraldo, CEP 13083-859, Campinas, São Paulo, Brazil. E-mail: amadeus@unicamp.br

behaviour can exist. Empirical evidence of long-memory has been found by Deo et al. (2010) and Zikes et al. (2017), among others.

Deo et al. (2010) proposed the Long-Memory Stochastic Duration (LMSD) model, in which the latent variable follows an autoregressive fractionally integrated moving average (ARFIMA) model (proposed by Granger and Joyeux (1980) and Hosking (1981)). Here, the estimation is performed through a frequency-domain quasi-maximum likelihood estimator using the Whittle approximation. This is a very appealing estimation approach given that nowadays time series of intertrade durations have hundreds of thousands of observations.

This paper proposes a minimum distance estimation method for the parameters of the LMSD model through autocovariance differences. This Generalized Method of Moments (GMM) type-based method allows dealing with the most challenging aspects of the problem: huge sample sizes and latent variables. Distinctive features of the proposed method are: it is easy to calculate and implement, and it provides asymptotic standard errors for the estimators[1]. Moreover, the estimates can be obtained very quickly.

The proposed method has roots in two previous GMM contributions in which the key is the existence of a central limit theorem (CLT) for the moments. In the context of stochastic volatility models (very close in nature to duration models), Wright (1999) proposed a minimum distance estimator based on autocovariances. However, the CLT only holds for values of the long memory parameter $d \in (-1/2, 1/4)$, so the method has limited application because empirical evidence supports very long memory, i.e., $d \in (1/4, 1/2)$. To overcome this problem, Chen and Deo (2005) proposed a set of moment conditions based on autocovariance differences that are asymptotically normal, but left the matter of which moments to use for future research.

The remainder of this paper is organised as follows. The proposed MDE is described in Section 2 and the evaluation of its performance through Monte Carlo experiments is presented in Section 3. The application of the method to a real-life time series is presented in Section 4 and conclusions are given in Section 5.

2 Minimum Distance Estimation

Let $\{y_1, \ldots, y_n\}$ be a sequence of observations from the LMSD process defined by[2]:

$$y_t = \beta \exp(h_t)\varepsilon_t, \tag{1}$$

$$\phi(B)(1-B)^d h_t = \theta(B)\eta_t, \tag{2}$$

where B is the backshift operator $By_t = y_{t-1}$, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$, and $(1-B)^d$ is the fractional difference operator defined by

[1] On the contrary, to the best of our knowledge, precision estimators for the method proposed by Deo et al. (2010) are unknown.

[2] Unlike Deo et al. (2010), we consider the scale parameter (β) in the observation equation instead of the ARFIMA equation.

$(1-B)^d = \sum_{j=0}^{\infty} B^j \Gamma(j-d)\Gamma(j+1)/\Gamma(-d)$. It is assumed that $\{\eta_t\}$ is a Gaussian zero-mean independent identically distributed (IID) sequence with $E(\eta_t^2) = \sigma^2$. In addition, $\{\varepsilon_t\}$ is an IID sequence of random variables with density $f(\varepsilon)$ having positive support (in applications, usual choices are the exponential, gamma and Weibull densities). Moreover, η_t and ε_t are independent for all t.

Applying the log transformation in Equation (1) we obtain the ARFIMA plus noise model:

$$x_t = c + h_t + \xi_t, \tag{3}$$

where $x_t = \log(y_t)$, $c = \log(\beta) + E[\log(\varepsilon_t)]$ and $\xi_t = \log(\varepsilon_t) - E[\log(\varepsilon_t)]$. Let σ_ξ^2 and κ_ξ be the variance and fourth cumulant of ξ_t, respectively.

Here we assume that $\{h_t\}$ is stationary with $d < \frac{1}{2}$ and that ε_t has density completely specified or indexed by a parameter ν such that κ_ξ is finite.

The minimum distance estimation (MDE) method proposed here is intended to estimate the parameters $\lambda = (d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma, \nu)$, based on the minimisation of the distance between sample and theoretical autocovariance differences. Thus, let $\delta_x(k)$ be the k-th theoretical autocovariance difference of process $\{x_t\}$, defined as:

$$\delta_x(k) = \gamma_x(0) - \gamma_x(k), \tag{4}$$

for $k = 1, \ldots, M$, where $\gamma_x(0), \ldots, \gamma_x(M)$ are the autocovariances up to lag M. Moreover, sample autocovariance differences are calculated by:

$$\hat{\delta}_x(k) = \hat{\gamma}_x(0) - \hat{\gamma}_x(k), \quad k = 1, \ldots, M, \tag{5}$$

where $\hat{\gamma}_x(k) = n^{-1} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$ for $k = 0, 1, \ldots, M$. Assuming that sample autocovariance differences satisfy the central limit theorem:

$$\sqrt{n}\,(\widehat{\delta} - \delta) \to N(0, \Omega(\lambda)) \quad \text{as} \quad n \to \infty \tag{6}$$

for fixed M, where $\delta = (\delta_x(1), \ldots, \delta_x(M))'$ and $\hat{\delta} = (\hat{\delta}_x(1), \ldots, \hat{\delta}_x(M))'$, the MDE of λ is the value that minimises the criterion function:

$$S(\lambda) = (\widehat{\delta} - \delta\,)' \Omega(\lambda)^{-1} (\widehat{\delta} - \delta\,). \tag{7}$$

For instance, let $\widehat{\lambda}$ be the value that minimizes $S(\lambda)$ and let λ_0 be the true parameter. The MDE method intends to find conditions so that:

$$\sqrt{n}(\widehat{\lambda} - \lambda_0) \to N(0, \Lambda(\lambda)) \quad \text{as} \quad n \to \infty, \tag{8}$$

where the variance-covariance matrix $\Lambda(\cdot)$ is given by:

$$\Lambda(\lambda) = \left[D(\lambda)' \Omega(\lambda)^{-1} D(\lambda) \right]^{-1}, \tag{9}$$

with matrix $D(\lambda) = \partial\delta/\partial\lambda$.

Thus, the success of the MDE method crucially depends on the existence of the CLT (6) and the computation of the asymptotic covariance matrix of the autocovariance differences $\Omega(\lambda)$. Based on results derived by Hosking (1996), Chen and Deo (2005) showed that (6) is valid. However, the expression of $\Omega_{i,j}(\lambda)$, the (i,j)-th element of $\Omega(\lambda)$, provided by these authors depends on a term written as an integral. In this respect, applying the results of Zevallos (2023) we can compute the elements of $\Omega(\lambda)$ in terms of analytic functions. The next proposition provides a procedure to calculate these elements.

Proposition 2.1. The elements of matrix $\Omega(\lambda)$ defined in (6), $\Omega_{i,j}$ for $i,j = 1,\ldots,M$, can be calculated recursively as:

$$\begin{aligned}
\Omega_{1,1} &= [3\tau(1) - \tau(2)] + 2\sigma_\xi^2[3a(1) - a(2)] + 3\sigma_\xi^4 + \kappa_\xi, && (10)\\
\Omega_{i+1,i+1} &= \Omega_{i,i} + [4\tau(i+1) - \tau(2i+1) - \tau(2i+2)] \\
&+ 2\sigma_\xi^2[4a(i+1) - a(2i+1) - a(2i+2)], \quad i = 1,\ldots,M-1 && (11)\\
\Omega_{i,i+1} &= \Omega_{i,i} + [2\tau(i+1) - \tau(1) - \tau(2i+1)] \\
&+ 2\sigma_\xi^2[2a(i+1) - a(1) - a(2i+1)] - \sigma_\xi^4, \quad i = 1,\ldots,M-1 && (12)\\
\Omega_{i,i+k} &= \Omega_{i,i+k-1} + [2\tau(i+k) - \tau(k) - \tau(2i+k)] \\
&+ 2\sigma_\xi^2[2a(i+k) - a(k) - a(2i+k)], \quad i = 1,\ldots,M-2, \quad k = 2,\ldots,M-i, && (13)
\end{aligned}$$

where functions $\tau(\cdot)$ and $a(\cdot)$ are defined as

$$\begin{aligned}
\tau(s) &= \vartheta(s-1) - \vartheta(s), && (14)\\
a(s) &= \gamma_h(s-1) - \gamma_h(s), && (15)
\end{aligned}$$

for $s = 1,2,\ldots$, $\vartheta(s) = \sum_{j=-\infty}^{\infty} \gamma_h(j)\gamma_h(j+s)$ and $\gamma_h(\cdot)$ are the autocovariances of the process $\{h_t\}$ defined in (2). Expressions for $\tau(s)$ in ARFIMA models were derived in Zevallos (2023); see some examples in Section 2.1 below.

Proof. Chen and Deo (2005) showed that (6) holds where the (i,j)-th element of $\Omega(\lambda)$, denoted by Ω_{ij} $i,j = 1,\ldots,M$, is equal to $\Omega_{ij} = \Psi_{i,j} + \Xi_{i,j}$, with $\Psi_{i,j} = 4\pi\int_{-\pi}^{\pi} \cos(i\omega)\cos(j\omega)|1 - \exp(\imath\omega)|^4 f_h^2(\omega)d\omega$, $\Xi_{i,j} = E(\xi_t^4) - \sigma_\xi^4 + \sigma_\xi^4 I_{[i=j]} + \sigma_\xi^2 Cov(2h_t - h_{t-i} - h_{t+i}, 2h_t - h_{t-j} - h_{t+j})$, and $I_{[.]}$ is the indicator function[3].

Now, $\Xi_{i,j}$ can be written as $\Xi_{i,j} = \kappa_\xi + 2\sigma_\xi^4 + \sigma_\xi^4 I_{[i=j]} + 2\sigma_\xi^2 B_{i,j}$ with

$$B_{i,j} = 2\gamma_h(0) - 2\gamma_h(i) - 2\gamma_h(j) + \gamma_h(j-i) + \gamma_h(j+i), \quad (16)$$

[3]Note that we adapted the result for the j-th autocovariance defined as a ratio with the denominator $1/n$ instead of $1/(n-j)$. Additionally, there is a typo in Chen and Deo (2005): the term $\sigma_\xi^4 I_{[i=j]}$ is missing in the expression (denoted here by) $\Xi_{i,j}$.

and the terms $\Psi_{i,j}$ can be evaluated using the results given in Zevallos (2023) where

$$\Psi_{i,j} = 2\vartheta(0) - 2\vartheta(i) - 2\vartheta(j) + \vartheta(j-i) + \vartheta(j+i), \tag{17}$$

with $\vartheta(s) = \sum_{j=-\infty}^{\infty} \gamma_h(j)\gamma_h(j+s)$. Therefore, we work with

$$\Omega_{i,j} = \Psi_{i,j} + 2\sigma_\xi^2 B_{i,j} + (\kappa_\xi + 2\sigma_\xi^4) + \sigma_\xi^4 I_{[i=j]}, \tag{18}$$

and it only remains to write these elements in terms of functions $\tau(\cdot)$ and $a(\cdot)$ given in (14)-(15). Note that $\Psi_{i,j}$ and $B_{i,j}$ have similar structures; thus by replacing $\vartheta(\cdot)$ with $\gamma_h(\cdot)$ in $\Psi_{i,j}$ we obtain $B_{i,j}$ and vice versa. The same occurs for $\tau(\cdot)$ and $a(\cdot)$.

First, consider $i = j = 1$ in (18). Thus, $\Omega_{1,1} = \Psi_{1,1} + 2\sigma_\xi^2 B_{1,1} + (\kappa_\xi + 3\sigma_\xi^4)$ where by (17) $\Psi_{1,1} = 3\vartheta(0) - 4\vartheta(1) + \vartheta(2)$, which in turn is equal to $3\tau(1) - \tau(2)$ using (14). Similarly, from (16) and (15), $B_{1,1} = 3\tau(1) - \tau(2)$, so we obtain (10). Second, since $\Omega_{i,i} = \Psi_{i,i} + 2\sigma_\xi^2 B_{i,i} + (\kappa_\xi + 3\sigma_\xi^4)$, then $\Omega_{i+1,i+1} - \Omega_{i,i} = (\Psi_{i+1,i+1} - \Psi_{i,i}) + 2\sigma_\xi^2(B_{i+1,i+1} - B_{i,i})$. Now, $\Psi_{i+1,i+1} - \Psi_{i,i} = 4\vartheta(i) - 4\vartheta(i+1) - \vartheta(2i) + \vartheta(2i+2) = 4[\vartheta(i) - \vartheta(i+1)] - [\vartheta(2i) - \vartheta(2i+i)] - [\vartheta(2i+i) - \vartheta(2i+2)] = 4\tau(i+1) - \tau(2i+1) - \tau(2i+2)$. In the same way, we obtain $B_{i+1,i+1} - B_{i,i} = 4a(i+1) - a(2i+1) - a(2i+2)$ and (11) follows. Third, we obtain (12) from $\Omega_{i,i+1} - \Omega_{i,i} = (\Psi_{i,i+1} - \Psi_{i,i}) + 2\sigma_\xi^2(B_{i,i+1} - B_{i,i}) - \sigma_\xi^4$ for $i = 1, \ldots, M-1$, where $\Psi_{i,i+1} - \Psi_{i,i} = 2\tau(i+1) - \tau(1) - \tau(2i+1)$ and $B_{i,i+1} - B_{i,i} = 2a(i+1) - a(1) - a(2i+1)$. Fourth, let $i = 1, \ldots, M-2$ and $k = 2, \ldots, M-i$. Then (13) follows, noting that $\Omega_{i,i+k} - \Omega_{i,i+k-1} = (\Psi_{i,i+k} - \Psi_{i,i+k-1}) + 2\sigma_\xi^2(B_{i,i+k} - B_{i,i+k-1})$, where $\Psi_{i,i+k} - \Psi_{i,i+k-1} = 2\tau(i+k) - \tau(k) - \tau(2i+k)$ and $B_{i,i+k} - B_{i,i+k-1} = 2a(i+k) - a(k) - a(2i+k)$. □

Therefore, to calculate the asymptotic covariance matrix $\Omega(\lambda)$, the elements of the first two diagonals are calculated first and then the rest of the elements are evaluated by rows. Note that to compute all these elements, we need to calculate $\tau(s)$ and $a(s)$ for $s = 1, \ldots, 2M$.

As a consequence of (6) and by standard GMM techniques we have:

Theorem 2.1. Let $\{y_t\}$ be a LMSD process defined in Equations (1) and (2) with parameter vector λ_0. The MDE $\widehat{\lambda}$ is a consistent estimator of λ_0 and satisfies (8)-(9).

Proof. The proof is very similar to that given in Baillie and Chung (2001) so we omit it. We must replace the theoretical and sample autocorrelations with the theoretical and sampled autocovariance differences, respectively. See also Wright (1999). □

2.1 Implementation of the MDE method

Given the sample $y_1, \ldots, y_n$, the procedure to obtain the minimum distance estimates $\hat{\lambda}$ can be summarised in the next steps:

1. Choose M, the number of autocovariance differences.

2. Calculate the sampled autocovariance differences defined in (5), where $x_t = \log(y_t)$.

3. Find the MDE value $\hat{\lambda}$ by minimising the function $S(\lambda)$ in (7).

4. The precision of the estimates are obtained by evaluating (9) in $\hat{\lambda}$.

Regarding step 1, it is expected that the theoretical precision of the estimators increases (closer to the maximum likelihood precision) as M increases. However, in practice sampled autocovariances at high lags can exhibit substantial uncertainty, and in addition, the optimisation procedure becomes slow when using very large values of M. Monte Carlo experiments presented in the next section show that values of $M \in \{50, 75\}$ are a good choice.

To implement step 3, we need the expressions for the autocovariance function and the asymptotic variance-covariance matrix $\Omega(\lambda)$. In addition, to implement step 4 we need the derivatives $D(\lambda)$. Next we describe how to calculate these expressions efficiently for the fractional noise ARFIMA(0,d,0) model and the ARFIMA$(1, d, 0)$ model. The parameter vector is $\lambda = (\lambda_h, \nu)$, where $\lambda_h = (d, \sigma)$ for the fractional noise model and $\lambda_h = (d, \phi, \sigma)$ for the ARFIMA$(1, d, 0)$ model. We illustrate the method assuming that ε_t follows an exponential or Weibull density. Thus, when ε_t follows a Weibull$(\nu, 1)$ distribution with density

$$f(\varepsilon_t) = \nu \varepsilon_t^{\nu-1} \exp(-\varepsilon_t^{\nu}), \quad \varepsilon_t > 0, \quad \nu > 0, \tag{19}$$

it is easily shown that

$$\sigma_\xi^2 = \psi'(1)/\nu^2, \quad \kappa_\xi = \psi'''(1)/\nu^4, \tag{20}$$

where $\psi'(\cdot)$ and $\psi'''(\cdot)$ are the first and third derivatives of the digamma function $\psi(\cdot)$. When ε_t follows an exponential distribution with rate one, we replace ν by one in expressions (19)-(20).

(a) *Autocovariances and autocovariance differences.* From Equation (3), given that h_t and ξ_t are independent processes, then $\gamma_x(k) = \gamma_h(k) + \sigma_\xi^2 I_{[k=0]}$ where $I_{[\cdot]}$ is the indicator function. Therefore

$$\delta_x(k) = \delta_h(k) + \sigma_\xi^2, \quad k = 1, 2, \ldots. \tag{21}$$

where σ_ξ^2 is given in (20) and $\delta_h(k) = \gamma_h(0) - \gamma_h(k)$. The autocovariances $\gamma_h(k)$ are calculated as follows. Note that in order to compute $\tau(k)$ and $a(k)$ given in (14)-(15), we have to consider $k = 0, \ldots, 2M$.

For the fractional noise model, the autocovariances, denoted by $\gamma_0(k)$, can be easily calculated recursively by $\gamma_0(k+1) = \gamma_0(k)(k+d)/(1+k-d)$ for $k = 0, 1, \ldots, 2M-1$, with $\gamma_0(0) = \sigma^2\Gamma(1-2d)/\Gamma^2(1-d)$. For the ARFIMA$(1, d, 0)$ model, the method suggested in Section 3.1 of Doornik and Ooms (2003) is followed. First, calculate $\gamma_0(k)$ for $k = 0, \ldots, 2M$ as described above. Second, let function G defined as $G(a; c; \rho) = \sum_{i=0}^{\infty} \rho^i (a)_{i+1}/(c)_{i+1}$ with $(b)_i = b(b+1)\ldots(b+i-1)$, $(b)_0 = 1$. Define $g(k) = G(d+k; 1-d+k; \phi)$ and compute $g(2M-1)$ using a finite sum. Then calculate the rest of the values of $g(\cdot)$ recursively backwards by $g(k-1) = [1+\phi g(k)](d+k-1)/(k-d)$ for $k = 2M-1, 2M-2, \ldots, 2-2M$. Third, evaluate $C(d, k, \phi) =$

$\gamma_0(k)[\phi^2 g(k) + \phi + g(-k)]$ for $k = 1-2M, \ldots, 1$ and then $\gamma_h(k) = C(d, 1-k, \phi)/(1-\phi^2)$ for $k = 0, \ldots, 2M$.

(b) *Calculation of* $\Omega(\lambda)$. The elements of matrix $\Omega(\lambda)$ are calculated using Proposition 2.1 with σ_ξ^2 and κ_ξ given by (20), and $\tau(\cdot)$ values are obtained from Zevallos (2023). Thus, for the ARFIMA(0,d,0) process, $\tau(s) = \sigma^4 \Delta(s)$, $s = 1, 2, \ldots, 2M$, where the terms $\Delta(s)$ can be easily computed recursively as follows: $\Delta(s+1) = \Delta(s)(s-1+2d)/(s+1-2d)$, for $s = 1, 2, \ldots, 2M-1$, with $\Delta(1) = 0.5\Gamma(3-4d)/\Gamma^2(2-2d)$. For the ARFIMA(1,$d$,0) model, $\tau(s) = \sigma^4(1-\phi^2)^{-2}\Delta(2+s)\Upsilon(d, 2+s, \phi)$, where $\Delta(s)$ values are given above and $\Upsilon(\cdot)$ is calculated by means of Corollary 3 in Zevallos (2023).

(c) *Computation of* $D(\lambda)$. We use results obtained in (a). For each parameter $\lambda_i \in \lambda$ we have to compute $D_{\lambda_i}(k) = \partial \delta_x(k)/\partial \lambda_i$ for $k = 1, \ldots, M$. These values constitute a row in matrix $D(\lambda)$. From (21) and (20), we obtain $D_\nu(k) = \partial \sigma_\xi^2/\partial \nu = -2\psi'(1)/\nu^3$ and $D_{\lambda_i}(k) = \partial \delta_h(k)/\partial \lambda_i = \partial \gamma_h(0)/\partial \lambda_i - \partial \gamma_h(k)/\partial \lambda_i$ when $\lambda_i \in \lambda_h$. For instance, when $\lambda_i = \sigma$, $D_\sigma(k) = 2\delta_h(k)/\sigma$ for $k = 1, \ldots, M$.

Next we present how to compute $\partial \gamma_h(k)/\partial \lambda_i$ when $\lambda_i \in \{d, \phi\}$ for $k = 0, \ldots, M$, in each model. For fractional noise models, let $\gamma'_{0d}(k) = \partial \gamma_0(k)/\partial d$ and $\omega(k) = \gamma'_{0d}(k)/\gamma_0(k)$. Since $\omega(0) = 2[\psi(1-d) - \psi(1-2d)]$ where $\psi(\cdot)$ is the digamma function, values of $\omega(k)$ for $k = 1, \ldots, M$ can be calculated recursively by $\omega(k) = \omega(k-1) + (2k-1)/((k-1+d)(k-d))$, and therefore $\gamma'_{0d}(k) = \omega(k)\gamma_0(k)$ for $k = 0, \ldots, M$.

For ARFIMA(1,d,0) models, let $g'_d(k) = \partial g(k)/\partial d$ and $g'_\phi(k) = \partial g(k)/\partial \phi$. First compute $g'_d(M-1)$ and $g'_\phi(M-1)$, and then calculate $g'_d(k)$ and $g'_\phi(k)$ recursively backwards by $g'_d(k-1) = \phi g'_d(k)(d+k-1)/(k-d) + [1+\phi g(k)](2k-1)/(k-d)^2$ and $g'_\phi(k-1) = [g(k) + \phi g'_\phi(k)](d+k-1)/(k-d)$ for $k = M-1, M-2, \ldots, 2-M$. Then, evaluate $C'_d(d, k, \phi) = \partial C(d, k, \phi)/\partial d$ and $C'_\phi(d, k, \phi) = \partial C(d, k, \phi)/\partial \phi$ through $C'_d(d, k, \phi) = C(d, k, \phi)\omega(k) + \gamma_0(k)[\phi^2 g'_d(k) + g'_d(-k)]$, $C'_\phi(d, k, \phi) = \gamma_0(k)[1 + 2\phi g(k) + \phi^2 g'_\phi(k) + g'_\phi(-k)]$, for $k = 1-M, \ldots, 0, 1$. Next, $\gamma'_d(k) = \partial \gamma_h(k)/\partial d$ and $\gamma'_\phi(k) = \partial \gamma_h(k)/\partial \phi$ are calculated by $\gamma'_d(k) = C'_d(d, 1-k, \phi)/(1-\phi^2)$ and $\gamma'_\phi(k) = C'_\phi(d, 1-k, \phi)/(1-\phi^2) + 2\phi C(d, 1-k, \phi)/(1-\phi^2)^2$ for $k = 0, 1, \ldots, M$.

2.2 Estimation of β

For completeness we discuss the estimation of the scale parameter β in (1). For instance, from Equation (3) we have $E(x_t) = \log(\beta) + E[\log(\varepsilon_t)]$. Then a moment estimator of β is given by $\hat{\beta} = \exp(\bar{x} - e(\hat{\nu}))$, where $\bar{x} = n^{-1}\sum_{t=1}^n x_t$ and $e(\nu) = E[\log(\varepsilon_t)]$. Thus, when ε_t follows the Weibull distribution given in (19), it can be shown that $e(\nu) = -\nu^{-1}C$, where C is the Euler constant ($C = 0.577215$). Accordingly, when ε_t follows the exponential distribution with rate one, $e(\nu) = -C$.

3 Simulations

The performance of the proposed estimator is assessed through Monte Carlo experiments. We considered LMSD models where the latent process h_t follows an ARFIMA(1,d,0) model and the perturbations ε_t follow Weibull and exponential distributions. For the Weibull case, the parameters were chosen based on some estimates obtained by Deo et al. (2010) in real-life duration data, and for the exponential case we selected from the latter reference some cases with values of ν near one. Additionally, in all cases $\beta = 1$. For each parameter combination, $1,000$ replications of time series of sizes n were generated, where $n = 50,000$ for the exponential case and $n \in \{50,000; 250,000\}$ for the Weibull case.

We report MDE results using $M \in \{25, 50\}$ and $M \in \{25, 50, 75\}$ for the exponential and Weibull cases, respectively. In order to obtain these estimates, we used a program written in the R package (R Core Team, 2017) with a Fortran subroutine. The optimisation was performed using the method of Nelder and Mead (1965) implemented in the `optim` library. Since this method performs unrestricted optimisation, we used transformations to account for the following restrictions: $d \in (-0.5, 0.5)$, $\phi \in (-1, 1)$, $\sigma > 0$ and $\nu > 0$. In all setups the initial values in the optimisation process were: $d_0 = 0.25$, $\phi_0 = -0.25$, and σ_0 equal to the standard deviation of the $\{x_t\}$ series. In turn, for the Weibull case, $\nu_0 = 1$.

The quantities reported in this simulation study are: the bias, the standard deviation and the root mean squared error (RMSE). These results are exhibited in tables 1, 2 and 3. In addition, we report the theoretical standard errors of the MDE method.

For the exponential distribution case, the numbers reported in Table 1 indicate that the MDE method is very good in terms of bias and the standard deviations are almost equal to their theoretical counterparts. Moreover, for each parameter combination, by comparing the results in terms of RMSE, we obtained slightly better results using $M = 50$ than using $M = 25$. However, even though (as expected) the theoretical standard deviation with $M = 50$ is lower than with $M = 25$, this difference is small. Therefore, in practice regarding the bias and precision, selecting $M = 25$ for calculating minimum distance estimates suffices.

Next, we discuss the results for the Weibull distribution. For sake of clarity, we denote the chosen parameter combinations as Case I to Case IV (see tables 2 and 3). For the Weibull distribution with $n = 50,000$, the results of Table 2 reveal that the MDE method is very good in terms of bias. In addition, for $(\hat{d}, \hat{\phi}, \hat{\sigma})$ the standard deviations are very close to their theoretical counterparts for cases II and III and moderately close for cases I and IV. Moreover, when comparing the standard deviation of $\hat{\nu}$ with its theoretical counterpart, in general we did not find good results for $M = 25$. On the contrary, for $M \in \{50, 75\}$, overall we found good results except for Case I (however, even here the sampled and theoretical standard deviations had the same order of magnitude). Furthermore, when comparing the RMSE values in terms of M for a fixed estimator and case, we found smaller values using $M = 75$, but in many cases these values

Table 1: Monte Carlo experiments. 1,000 replications of LMSD-(1,d,0) exponential processes with $\beta = 1$ of size $n = 50,000$. Parameters d, ϕ and σ are estimated by the MDE method with $M \in \{25, 50\}$. For each combination, ^{a} is the sample bias, ^{b} is the standard deviation and ^{c} is the RMSE. In addition, theoretical asymptotic standard deviations for the MDE method are given in ^{d}.

			M=25			M=50		
d	ϕ	σ	$\widehat{d}$	$\widehat{\phi}$	$\widehat{\sigma}$	$\widehat{d}$	$\widehat{\phi}$	$\widehat{\sigma}$
0.45	-0.5	0.7	-0.001^a	0.001	0.002	-0.002	0.003	0.005
			0.016^b	0.019	0.014	0.013	0.018	0.014
			0.016^c	0.019	0.014	0.013	0.018	0.015
			0.017^d	0.019	0.014	0.014	0.018	0.014
0.45	-0.6	0.9	-0.000	0.000	0.002	-0.001	0.001	0.004
			0.011	0.011	0.012	0.010	0.010	0.012
			0.011	0.011	0.012	0.010	0.010	0.012
			0.011	0.011	0.012	0.010	0.010	0.012
0.40	-0.4	0.9	-0.001	0.001	0.002	-0.001	0.002	0.004
			0.013	0.016	0.011	0.011	0.015	0.011
			0.013	0.016	0.012	0.011	0.015	0.012
			0.013	0.016	0.011	0.011	0.015	0.011
0.30	-0.5	0.8	-0.002	0.002	0.003	-0.002	0.003	0.005
			0.016	0.014	0.012	0.013	0.014	0.012
			0.016	0.015	0.013	0.013	0.014	0.013
			0.017	0.015	0.012	0.014	0.015	0.012

were very close to those obtained with $M = 50$.

Table 3 reports the results for the Weibull distribution with $n = 250,000$. It can be seen that the results are very good in terms of bias. Additionally, for $M \in \{50, 75\}$, the standard deviations are close or moderately close to their theoretical counterparts, except for $\hat{\nu}$ in Case III. Moreover, when comparing the RMSE values in terms of M for a fixed estimator and case, we found smaller values using $M = 75$, except in cases I and III, where we observed slightly greater values using $M = 75$ compared to $M = 50$.

For a given parameter combination, estimator and M value, when comparing tables 2 and 3, we observe that both the bias and RMSE values decrease as the sample size increases, except for Case III, where the biases are almost the same, and the RMSE of $\hat{\nu}$ with $M = 25$ is slightly smaller for $n = 50,000$.

For completeness, we report a summary of the performance of the estimator of β presented in Section 2.2[4]. For the Weibull distribution when $n = 250,000$ and using $M = 75$, the figures of (bias, RMSE) are: (-0.005,0.052), (-0.003,0.028), (-0.005,0.027) and (-0.003,0.030) for cases 1 to 4, respectively. These values are very close to those obtained when $M = 50$. Therefore, the results are very good.

[4]Full results are available upon request.

Table 2: Monte Carlo experiments. 1,000 replications of LMSD-(1,d,0) Weibull processes with $\beta = 1$ of size $n = 50,000$. Parameters d, ϕ, σ and ν are estimated by the MDE method with $M \in \{25, 50, 75\}$. For each combination, ^{a} is the sample bias, ^{b} is the standard deviation and ^{c} is the RMSE. In addition, theoretical asymptotic standard deviations for the MDE method are given in ^{d}.

					M=25				M=50				M=75			
Case	d	ϕ	σ	ν	$\widehat{d}$	$\widehat{\phi}$	$\widehat{\sigma}$	$\widehat{\nu}$	$\widehat{d}$	$\widehat{\phi}$	$\widehat{\sigma}$	$\widehat{\nu}$	$\widehat{d}$	$\widehat{\phi}$	$\widehat{\sigma}$	$\widehat{\nu}$
I	0.40	-0.5	0.8	0.9	-0.014^a	0.019	0.036	0.020	-0.007	0.010	0.018	0.007	-0.006	0.009	0.018	0.006
					0.053^b	0.067	0.116	0.066	0.038	0.050	0.079	0.032	0.036	0.048	0.077	0.031
					0.055^c	0.069	0.122	0.069	0.039	0.051	0.081	0.033	0.036	0.049	0.079	0.032
					0.037^d	0.045	0.065	0.021	0.033	0.043	0.064	0.021	0.031	0.041	0.062	0.020
II	0.35	-0.4	0.5	1.3	-0.010	0.011	0.022	0.025	-0.004	0.004	0.014	0.015	-0.002	0.002	0.010	0.010
					0.073	0.086	0.091	0.073	0.061	0.076	0.081	0.063	0.054	0.069	0.074	0.055
					0.074	0.087	0.094	0.077	0.061	0.076	0.083	0.065	0.054	0.069	0.074	0.055
					0.073	0.087	0.083	0.052	0.061	0.077	0.076	0.047	0.053	0.069	0.069	0.044
III	0.25	-0.4	0.5	1.3	-0.007	0.009	0.017	0.019	-0.006	0.008	0.016	0.017	-0.006	0.008	0.016	0.016
					0.049	0.077	0.081	0.065	0.048	0.078	0.081	0.063	0.047	0.077	0.081	0.067
					0.050	0.078	0.083	0.068	0.049	0.078	0.082	0.066	0.048	0.078	0.083	0.069
					0.046	0.072	0.069	0.044	0.045	0.071	0.068	0.043	0.044	0.071	0.068	0.043
IV	0.25	-0.2	0.5	2.0	0.016	-0.015	-0.009	0.020	0.009	-0.008	-0.002	0.033	0.004	-0.004	0.002	0.039
					0.065	0.058	0.080	0.203	0.055	0.049	0.076	0.208	0.048	0.043	0.073	0.206
					0.067	0.060	0.080	0.204	0.055	0.050	0.076	0.210	0.048	0.043	0.073	0.210
					0.118	0.100	0.160	0.379	0.073	0.063	0.107	0.252	0.059	0.052	0.090	0.212

Table 3: Monte Carlo experiments. 1,000 replications of LMSD-(1,d,0) Weibull processes with $\beta = 1$ of size $n = 250,000$. Parameters d, ϕ, σ and ν are estimated by the MDE method with $M \in \{25, 50, 75\}$. For each combination, ^{a} is the sample bias, ^{b} is the standard deviation and ^{c} is the RMSE. In addition, theoretical asymptotic standard deviations for the MDE method are given in ^{d}.

					M=25				M=50				M=75			
Case	d	ϕ	σ	ν	$\widehat{d}$	$\widehat{\phi}$	$\widehat{\sigma}$	$\widehat{\nu}$	$\widehat{d}$	$\widehat{\phi}$	$\widehat{\sigma}$	$\widehat{\nu}$	$\widehat{d}$	$\widehat{\phi}$	$\widehat{\sigma}$	$\widehat{\nu}$
I	0.40	-0.5	0.8	0.9	-0.007^a	0.010	0.018	0.010	-0.001	0.001	0.002	0.001	-0.001	0.001	0.002	0.000
					0.036^b	0.046	0.082	0.050	0.015	0.020	0.030	0.010	0.015	0.020	0.030	0.010
					0.036^c	0.047	0.084	0.051	0.015	0.020	0.030	0.010	0.015	0.020	0.030	0.010
					0.017^d	0.020	0.029	0.010	0.015	0.019	0.028	0.009	0.014	0.018	0.028	0.009
II	0.35	-0.4	0.5	1.3	-0.010	0.011	0.014	0.012	-0.003	0.004	0.006	0.005	-0.001	0.002	0.003	0.002
					0.040	0.048	0.051	0.037	0.031	0.039	0.041	0.029	0.025	0.033	0.033	0.022
					0.042	0.050	0.052	0.039	0.031	0.040	0.041	0.030	0.025	0.033	0.034	0.022
					0.033	0.039	0.037	0.023	0.027	0.034	0.034	0.021	0.024	0.031	0.031	0.020
III	0.25	-0.4	0.5	1.3	-0.011	0.018	0.022	0.020	-0.010	0.016	0.020	0.018	-0.010	0.017	0.022	0.019
					0.032	0.053	0.060	0.053	0.031	0.050	0.058	0.053	0.031	0.052	0.062	0.058
					0.034	0.056	0.064	0.056	0.032	0.053	0.062	0.056	0.033	0.055	0.066	0.061
					0.021	0.032	0.031	0.019	0.020	0.032	0.031	0.019	0.020	0.032	0.031	0.019
IV	0.25	-0.2	0.5	2.0	0.012	-0.010	-0.012	-0.011	0.002	-0.002	-0.001	0.009	0.002	-0.002	-0.001	0.008
					0.037	0.032	0.049	0.122	0.028	0.025	0.042	0.104	0.025	0.023	0.039	0.098
					0.039	0.034	0.050	0.123	0.028	0.025	0.042	0.105	0.025	0.023	0.039	0.098
					0.053	0.045	0.072	0.169	0.033	0.028	0.048	0.113	0.026	0.023	0.040	0.095

4 Illustration

In this section we briefly illustrate the application of the proposed method to a real time series. We use the time series of nonzero durations (in seconds) adjusted for daily seasonality of the Caterpillar stock in the period from January 4, 2010 to May 28, 2010[5]. Daily seasonal adjustment was performed using the method described in Deo et al. (2010). The time series consists of $n = 985,365$ observations and has the following statistics: 1.350 (mean), 1.404 (standard deviation), 0.459 (minimum), 0.572 (Q_1), 0.720 (Q_2), 1.455 (Q_3) and 48.880 (maximum). Figure 1 shows the autocorrelation function and the histogram based on the durations up to the 99% quantile. Here, it can be seen that the series exhibits long-memory and that a Weibull distribution instead an exponential distribution is more suitable for modelling purposes.

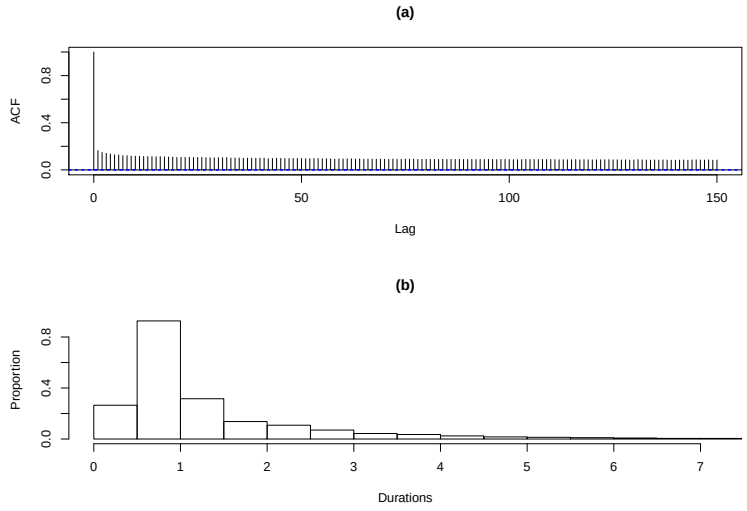

Figure 1: *(a) ACF of the durations; (b) Histogram of the durations up to quantile 99%.*

LMSD models with Weibull perturbations were fitted for two specifications of the latent ARFIMA process: (0,d,0) and (1,d,0). Minimum distance estimation results for $M \in \{50, 75, 100\}$ are presented in Table 4. Here, it can be seen that in all cases the estimates are highly significant for the (0,d,0) specification, and for each parameter the estimates are very stable (have similar values) in terms of M. Besides this, for the (1,d,0) specification, we obtain non-significant estimates for the autoregressive parameter ϕ, but the rest of estimates are highly significant and stable across M for each parameter. For both specifications, the results support the presence of long-memory, with memory parameter estimates around 0.38, and distribution parameter estimates around 2.1. In addition, based on the consistent estimator shown in Section 2.2, we obtain estimates of β around 1.3.

[5]The row data were downloaded from https://faculty.chicagobooth.edu/ruey-s-tsay/research/an-introduction-to-analysis-of-financial-data-with-r, inside the ch6data.zip file. A subset of this time series was analysed in Chapter 6 of Tsay (2013) based on ACD models.

Table 4: Minimum distance estimates of Caterpillar duration time series. Standard errors are given in parentheses.

Specification	M	$\hat{d}$	$\hat{\phi}$	$\hat{\sigma}$	$\hat{\nu}$
(0,d,0)-Weibull	50	0.3790		0.2412	2.1380
		(0.0083)		(0.0055)	(0.0075)
	75	0.3900		0.2351	2.1302
		(0.0070)		(0.0047)	(0.0064)
	100	0.3878		0.2364	2.1316
		(0.0063)		(0.0045)	(0.0062)
(1,d,0)-Weibull	50	0.3708	-0.0213	0.2534	2.1536
		(0.0140)	(0.0258)	(0.0171)	(0.0231)
	75	0.4034	-0.0318	0.2354	2.1269
		(0.0107)	(0.0263)	(0.0136)	(0.0163)
	100	0.3780	-0.0262	0.2506	2.1487
		(0.0097)	(0.0228)	(0.0133)	(0.0176)

5 Conclusions

A new estimation method based on minimum distance is proposed. The main features of the MDE estimator can be summarized as follows: First, it is easy to calculate and implement, since the computation of the estimates is based on a reduced number of sample autocovariance differences. Second, the computation is very fast even for time series with nearly a million observations. For instance, the computing time to obtain all the estimates of Table 4 was 73.1 seconds[6], while it was 17.7 and 55.4 seconds for the $(0, d, 0)$ and $(1, d, 0)$ specifications, respectively. Third, it provides precision estimates based on asymptotic standard errors.

Monte Carlo experiments with LMSD-$(1, d, 0)$ models indicate that the proposed method exhibits very small bias. Moreover, precision estimates are very accurate using $M = 50$ in time series of size 50,000 for the exponential case, and using $M \in \{50, 75\}$ in time series with size $250,000$ for the Weibull case.

To sum up, given its numerical efficiency, the MDE technique is useful for handling huge datasets, typically found in real intertrade duration time series.

Acknowledgements

This research was partially supported by the São Paulo Research Foundation (FAPESP) grant number 2018/04654-9. The author acknowledges the support of the Centre for Applied Research on Econometrics, Finance and Statistics (CAREFS).

[6]Using an Intel(C) CoreTM i7-8550U CPU @ 1.80GHz×4, 15.4 GiB of RAM memory computer.

References

R.T Baillie and H. Chung. Estimation of garch models from the autocorrelations of the squares of a process. *Journal of Time Series Analysis*, 22(6):631–650, 2001.

L. Bauwens and D. Veredas. The stochastic conditional duration model: A latent variable model for the analysis of financial durations. *Journal of Econometrics*, 119(2):381–412, 2004.

S. K. Bhogal and R. T. Variyam. Conditional duration models for high frequency data: A review on recent developments. *Journal of Economic Surveys*, 33(1):252–273, 2018.

W. Chen and R.S. Deo. GMM estimation for long memory latent variable volatility and duration Models. Technical report, Econometrics 0501006, University Library of Munich, Germany, 2005. URL `https://ideas.repec.org/p/wpa/wuwpem/0501006.html`.

R. Deo, M. Hsieh, and C.M. Hurvich. Long memory in intertrade durations, counts and realized volatility of nyse stocks. *Journal of Statistical Planning and Inference*, 140(12):3715–3733, 2010.

J.A. Doornik and M. Ooms. Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis*, 42(3):333–348, 2003.

R.F. Engle and J.R. Russell. Autoregressive conditional duration: A new model for irregularly spaced time series data. *Econometrica*, 66(5):1127–1162, 1998.

C. W. J. Granger and R. Joyeux. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29, 1980.

J.R.M. Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

J.R.M. Hosking. Asymptotic distributions of the sample mean, autocovariance and autocorrelation of long-memory time series. *Journal of Econometrics*, 73(1):261–284, 1996.

J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

M. Pacurar. Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys*, 22(4):711–751, 2008.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

R.S. Tsay. *An Introduction to Analysis of Financial Data with R*. John Wiley & Sons, New Jersey, 2013.

J.H. Wright. A new estimator of the fractionally integrated stochastic volatility model. *Economics Letters*, 63(3):295–303, 1999.

M. Zevallos. On the asymptotic distribution of sample autocovariance differences of long-memory processes. *Brazilian Journal of Probability and Statistics*, (Forthcoming), 2023.

F. Zikes, J. Baruník, and N. Shenai. Modeling and forecasting financial durations. *Econometric Reviews*, 36(10):1081–1110, 2017.