On Fréchet autoregressive conditional duration models

Yao Zheng, Yang Li and Guodong Li
Department of Statistics and Actuarial Science,
University of Hong Kong, Pokfulam Road, Hong Kong
yaozheng@connect.hku.hk; snliyang@connect.hku.hk; gdli@hku.hk

Abstract

Some durations such as those of block trades may have the properties of both heavy tails and extreme values. To model such type of data, we suggest the Fréchet distribution for the innovations of the autoregressive conditional duration (ACD) model, and hence the Fréchet ACD model. Some statistical inference tools including the maximum likelihood estimation and diagnostic tools for model adequacy are derived, and their finite-sample performance is evaluated by Monte Carlo simulation experiments. The usefulness of the new model is demonstrated by analyzing the durations of block trades on two stock exchanges.

Keywords: ACD models, Extreme values, Fréchet distribution, Heavy tails.

1. Introduction

Consider the autoregressive conditional duration (ACD) model,

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

where $t_0 < t_1 < \cdots < t_n < \cdots$ are arrival times, $x_i = t_i - t_{i-1}$ is an interval, $\omega > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$, and the innovations $\{\varepsilon_i\}$ are identically and independently distributed (*i.i.d.*) nonnegative random variables with mean one (Engle and Russell, 1998). This model has been widely applied to high-frequency and ultra-high-frequency data, which usually have unequally spaced time intervals, and have become common in financial modeling due to the great improvement of information technology and the popularity of electronic trading (Engle, 2000). For the innovation ε_i, Engle and Russell (1998) considered an exponential distribution and a Weibull distribution, and the corresponding maximum likelihood estimations (MLE) were also discussed. Note that the hazard rate is a constant for the exponential distribution, and is monotonic for the Weibull distribution. Grammig and Maurer (2000) introduced a Burr distribution for ε_i to make the conditional hazards of the durations $\{x_i\}$ more flexible.

In the meanwhile, many financial time series are heavy-tailed and, when the commonly used generalized autoregressive conditional heteroscedastic (GARCH) model (Engle, 1982; Bollerslev, 1986) is applied to these sequences, Gaussian innovations usually produce tails which are thinner than those of the real data; see Mikosch and Starica (2000) and Li and Li (2005). To improve the efficiency of the Gaussian quasi-MLE for these heavy-tailed time series, some robust approaches have been discussed for GARCH models, e.g. the least absolute deviation estimation in Peng and Yao (2003) and Li and Li (2008a). Bollerslev (1987) alternatively considered a GARCH model with Student's t innovations, and the heavy-tailed Student's t distribution can help to explain the excess dispersion to some extent as well as to improve the efficiency of the resulting estimation. For the ACD model, Engle and Russell (1998) found that, after accounting for the temporal dependence, both the exponential and the Weibull distributions failed to explain the excess dispersion in the IBM transaction duration data, and Zhang et al. (2001) considered a generalized gamma distribution to account for the heavy tails in this dataset. Note that the ACD model for durations is analogous to the GARCH model for returns (Engle and Russell, 1998), and the Fréchet distribution has a relatively heavier right tail compared with other nonnegative distributions including the aforementioned four distributions. Along the line of Bollerslev (1987), this paper considers the ACD model with ε_i having the Fréchet distribution, which we call the Fréchet ACD model for simplicity. This new model is supposed to provide a more robust estimation for heavy-tailed durations.

The Fréchet distribution is one of the three types of extreme value distributions, and we may frequently encounter extreme value problems in modeling durations. As an illustrative example, consider the block trades in a stock market, and suppose that there are L stocks in this market. For the lth stock with $1 \leq l \leq L$, denote by $0 =$

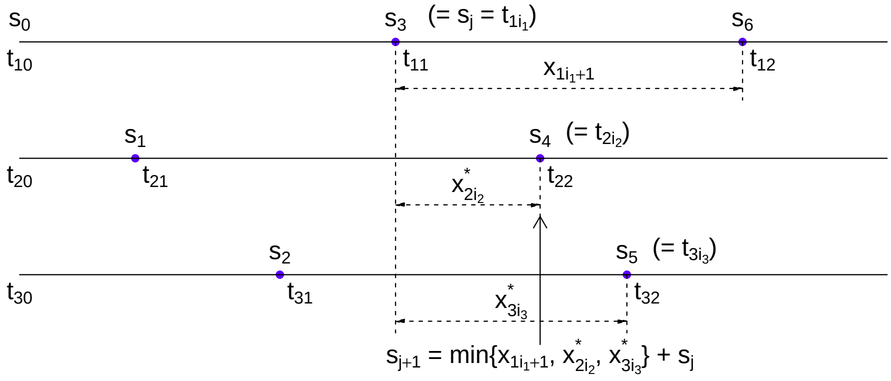

Figure 1: Illustration on arrival times of block trades with $L = 3$.

$t_{l0} < t_{l1} < \cdots < t_{ln} < \cdots$ the arrival times, by $\{x_{li}\}$ the durations, and by $N_l(t)$ the associated marked point process. Let $0 = s_0 < s_1 < \cdots < s_n < \cdots$ be the arrival times of the block trades, and $\{y_i\}$ be their durations. See Figure 1 for an illustration with $L = 3$. Note that $\{s_j\}$ is the order statistics of $\{t_{li}, 1 \le l \le L, i = 0, 1, ...\}$, and $y_1 = s_1 - s_0 = \min\{x_{l1}, 1 \le l \le L\}$. For $j \ge 1$, without loss of generality, we assume that $s_j = t_{1i_1}$, i.e. the block trade happens on the first stock; see the example of $j = 3$ in Figure 1. Let $i_l = \min\{i : t_{li} > s_j\}$ and $x^*_{li_l} = t_{li_l} - s_j$ for $l \ge 2$. As in Engle and Russell (1998), we assume that the marked point processes $N_l(t)$ evolves without after-effects and are conditionally orderly. Then the conditional intensity of point process $N_l(t)$ remains unchanged after s_j, implying that the random variables $x^*_{li_l}$ and $x_{li_l} = t_{l,i_l} - t_{l,i_l-1}$ have the same marginal distribution as well as the same dependence structure on other durations. Hence, $y_{j+1} = s_{j+1} - s_j = \min\{x_{1,i_1+1}, x^*_{2i_2}, ..., x^*_{Li_L}\}$ has the same distribution as $\min\{x_{1,i_1+1}, x_{2i_2}, ..., x_{Li_L}\}$, and then it is natural to consider involving an extreme value distribution for the innovation ε_i of model (1). Note that among the three extreme value distributions, the Gumbel distribution is two-sided, while the Weibull distribution and the Fréchet distribution are one-sided (Embrechts et al., 1997). Therefore, with its right tail heavier than the Weibull distribution, the Fréchet distribution may be of particular interest in modeling the durations of block trades.

The rest of the paper is structured as follows. Section 2 discusses the Fréchet ACD model and derives some statistical inference tools including the MLE and diagnostic tools for model adequacy. Section 3 conducts several Monte Carlo simulations to study the finite-sample performance of these inference tools. Section 4 demonstrates the usefulness of the Fréchet ACD model by analyzing the durations of block trades on two stock exchanges: the Hong Kong Stock Exchange (SEHK) and the London Stock Exchange (LSE). The proofs of Theorems 1 and 2 are relegated to the Appendix.

2. Fréchet ACD models

For the autoregressive conditional duration (ACD) model at (1), we consider the Fréchet distribution for the innovation ε_i, which has density function of the form

$$f(x;\gamma,s,m)=\frac{\gamma}{s}\left(\frac{x-m}{s}\right)^{-1-\gamma}\exp\left\{-\left(\frac{x-m}{s}\right)^{-\gamma}\right\},\quad x\geq m,$$

where $\gamma>0$ is the shape parameter, $s>0$ is the scale parameter, and $m\in\mathbb{R}$ is the location parameter. Due to the non-negativity of the observed durations $\{x_i\}$, we need to restrict m to zero. Additionally, to ensure the identifiability of model (1), the constraint $E(\varepsilon_t)=1$ is imposed; see, e.g., Engle and Russell (1998). Hence the innovation ε_i follows the standardized Fréchet distribution with shape parameter γ, which has mean one and density function of the form

$$f_\gamma(x)=\gamma c_\gamma x^{-1-\gamma}\exp\{-c_\gamma x^{-\gamma}\},\quad x\geq 0,$$

where $c_\gamma=[\Gamma(1-\gamma^{-1})]^{-\gamma}$ with $\Gamma(\cdot)$ being the Gamma function. Analogous to the Student's t distribution, the standardized Fréchet distribution has a heavier right tail as the shape parameter γ is smaller, and it has finite mth moment $E(\varepsilon_i^m)$ as long as $m<\gamma$. We denote this model by FACD(p,q) for simplicity.

2.1. Maximum likelihood estimation

Let $\boldsymbol{\alpha}=(\alpha_1,...,\alpha_p)'$, $\boldsymbol{\beta}=(\beta_1,...,\beta_q)'$ and $\boldsymbol{\theta}=(\omega,\boldsymbol{\alpha}',\boldsymbol{\beta}')'$. Denote by $\boldsymbol{\lambda}=(\gamma,\boldsymbol{\theta}')'$ the parameter vector of the Fréchet ACD model, and by $\Lambda=\mathbb{R}^+\times\Theta$ the parameter space, where $\Theta\subset\mathbb{R}^{p+q+1}$ is a compact set. The true parameter vector $\boldsymbol{\lambda}_0=(\gamma_0,\boldsymbol{\theta}_0')'$ is an interior point of Λ, and the following conditions hold for each $\lambda\in\Lambda$.

Assumption 1. *$\gamma > 1$, $\omega > 0$, $\alpha_j \geq 0$ for $1 \leq j \leq p$, $\beta_j \geq 0$ for $1 \leq j \leq q$, $\alpha_p\beta_q > 0$, $\sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j < 1$, and the polynomials $\sum_{j=1}^p \alpha_j x^j$ and $1 - \sum_{j=1}^q \beta_j x^j$ have no common root.*

Given nonnegative observations $x_1, ..., x_n$, we can iteratively define the functions

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

based on equation (1), and then the log-likelihood function of the Fréchet ACD model is

$$\begin{aligned} L_n(\boldsymbol{\lambda}) &= \sum_{i=1}^n l_i(\boldsymbol{\lambda}) \\ &= \sum_{i=1}^n \left[\log f_\gamma\left(\frac{x_i}{\psi_i(\boldsymbol{\theta})}\right) - \log \psi_i(\boldsymbol{\theta})\right] \\ &= \sum_{i=1}^n \left[\gamma \log \psi_i(\boldsymbol{\theta}) - c_\gamma \left(\frac{x_i}{\psi_i(\boldsymbol{\theta})}\right)^{-\gamma}\right] - (1+\gamma)\sum_{i=1}^n \log(x_i) + n\log(\gamma \cdot c_\gamma). \end{aligned}$$

Note that the above functions all depend on unobservable values of x_i with $i \leq 0$, and some initial values are hence needed for $x_0, x_{-1}, ..., x_{1-p}$ and $\psi_0(\boldsymbol{\theta}), \psi_{-1}(\boldsymbol{\theta}), ..., \psi_{1-q}(\boldsymbol{\theta})$. We simply set them to be $\bar{x} = n^{-1}\sum_{i=1}^n x_i$, and denote the corresponding functions $\psi_i(\boldsymbol{\theta})$, $l_i(\boldsymbol{\lambda})$ and $L_n(\boldsymbol{\lambda})$ respectively by $\widetilde{\psi}_i(\boldsymbol{\theta})$, $\widetilde{l}_i(\boldsymbol{\lambda})$ and $\widetilde{L}_n(\boldsymbol{\lambda})$.

Thus, the MLE can be defined as

$$\widetilde{\boldsymbol{\lambda}}_n = (\widetilde{\gamma}_n, \widetilde{\boldsymbol{\theta}}_n')' = \underset{\boldsymbol{\lambda}\in\Lambda}{\operatorname{argmax}}\, \widetilde{L}_n(\boldsymbol{\lambda}).$$

Let

$$c_1(x, \gamma) = -\frac{\partial \log f_\gamma(x)}{\partial x}x - 1 = \gamma(1 - c_\gamma x^{-\gamma})$$

and

$$c_2(x, \gamma) = \frac{\partial \log f_\gamma(x)}{\partial \gamma} = c_\gamma x^{-\gamma}\log(x) - \log(x) - c_\gamma' x^{-\gamma} + \gamma^{-1} + c_\gamma'/c_\gamma,$$

where $c_\gamma' = \partial c_\gamma / \partial\gamma$. It can be verified that $E[c_1(\varepsilon_i, \gamma_0)] = 0$ and $E[c_2(\varepsilon_i, \gamma_0)] = 0$. Denote $\kappa_1 = \text{var}[c_1(\varepsilon_i, \gamma_0)]$, $\kappa_2 = \text{var}[c_2(\varepsilon_i, \gamma_0)]$, $\kappa_3 = \text{cov}[c_1(\varepsilon_i, \gamma_0), c_2(\varepsilon_i, \gamma_0)]$ and

$$\Sigma = \begin{pmatrix} \kappa_2 & \kappa_3 E[\psi_i^{-1}(\boldsymbol{\theta}_0)\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}'] \\ \kappa_3 E[\psi_i^{-1}(\boldsymbol{\theta}_0)\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}] & \kappa_1 E\{\psi_i^{-2}(\boldsymbol{\theta}_0)[\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}][\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}']\} \end{pmatrix}.$$

Theorem 1. *Under Assumption 1, it holds that $\widetilde{\boldsymbol{\lambda}}_n$ converges to $\boldsymbol{\lambda}_0$ in almost surely sense as $n \to \infty$.*

If we further assume that $\gamma > 2$, then the matrix Σ is positive definite and $\sqrt{n}(\widetilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) \to_d N(0, \Sigma^{-1})$ as $n \to \infty$.

Denote by $\{\widetilde{\varepsilon}_i\}$ the residual sequence from the fitted Fréchet ACD model, where $\widetilde{\varepsilon}_i = x_i/\widetilde{\psi}_i(\widetilde{\boldsymbol{\theta}}_n)$. For the quantities in the information matrix Σ, κ_1, κ_2, κ_3, $E[\psi_i^{-1}(\boldsymbol{\theta}_0)\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}]$, and $E[\psi_i^{-2}(\boldsymbol{\theta}_0)(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta})(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}')]$, we can estimate them respectively by

$$\widetilde{\kappa}_1 = \frac{1}{n}\sum_{i=1}^{n}[c_1(\widetilde{\varepsilon}_i, \widetilde{\gamma}_n)]^2, \quad \widetilde{\kappa}_2 = \frac{1}{n}\sum_{i=1}^{n}[c_2(\widetilde{\varepsilon}_i, \widetilde{\gamma}_n)]^2, \quad \widetilde{\kappa}_3 = \frac{1}{n}\sum_{i=1}^{n} c_1(\widetilde{\varepsilon}_i, \widetilde{\gamma}_n)c_2(\widetilde{\varepsilon}_i, \widetilde{\gamma}_n),$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widetilde{\psi}_i(\widetilde{\boldsymbol{\theta}}_n)}\frac{\partial\widetilde{\psi}_i(\widetilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widetilde{\psi}_i^2(\widetilde{\boldsymbol{\theta}}_n)}\frac{\partial\widetilde{\psi}_i(\widetilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\frac{\partial\widetilde{\psi}_i(\widetilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}'}.$$

From the proof of Theorem 1, the above estimators are all consistent, and hence constitute a consistent estimator of the information matrix Σ. We may sometimes be interested in the parameter vector $\boldsymbol{\theta}$ only, and it is implied by Theorem 1 that

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to_d N(0, \Sigma_1^{-1})$$

as $n \to \infty$, where

$$\Sigma_1 = \kappa_1 \cdot E\left[\frac{1}{\psi_i^2(\boldsymbol{\theta}_0)}\frac{\partial\psi_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\psi_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right] - \frac{\kappa_3^2}{\kappa_2} \cdot E\left[\frac{1}{\psi_i(\boldsymbol{\theta}_0)}\frac{\partial\psi_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right] E\left[\frac{1}{\psi_i(\boldsymbol{\theta}_0)}\frac{\partial\psi_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right].$$

2.2. Diagnostic tools

Residuals from a fitted time series model play an important role in checking the adequacy of the model. In particular, residual autocorrelations, which are autocorrelations of the residual sequence, were first employed in Box and Pierce (1970) and Ljung and Box (1978). However, portmanteau tests based on residual autocorrelations usually have no power in detecting the possible misspecifications of the conditional variance (Li and Li, 2008a). Some improved diagnostic tools include those based on the squared residual autocorrelations (McLeod and Li, 1983) and those based on the absolute residual autocorrelations (Li and Li, 2005). This subsection derives the asymptotic distribution of the residual autocorrelations from the fitted Fréchet ACD model, and hence a portmanteau test for checking the adequacy of this model. It is worth pointing out that the residuals are nonnegative, and that therefore, residual autocorrelation and absolute residual autocorrelation coincide.

Without confusion, we denote $\widetilde{\psi}_i(\widetilde{\boldsymbol{\theta}}_n)$ and $\psi_i(\boldsymbol{\theta}_0)$ respectively by $\widetilde{\psi}_i$ and ψ_i for simplicity. Consider the residual sequence $\{\widetilde{\varepsilon}_i\}$ with $\widetilde{\varepsilon}_i = x_i/\widetilde{\psi}_i$. Note that $n^{-1}\sum_{i=1}^{n}\widetilde{\varepsilon}_i = 1+o_p(1)$. Hence, for a positive integer k, the lag-k residual autocorrelation can be defined as

$$\widetilde{r}_k = \frac{\sum_{i=k+1}^{n}(\widetilde{\varepsilon}_i - 1)(\widetilde{\varepsilon}_{i-k} - 1)}{\sum_{i=1}^{n}(\widetilde{\varepsilon}_i - 1)^2}.$$

We next consider the asymptotic distributions of the first K residual autocorrelations, $\widetilde{R} = (\widetilde{r}_1, ..., \widetilde{r}_K)'$, where K is a predetermined positive integer.

Theorem 2. *Under the conditions of Theorem 1, it holds that*

$$\sqrt{n}\widetilde{R} \to_d N(0, \Omega),$$

as $n \to \infty$, *where* $\Omega = \mathbf{I} - \sigma_{\gamma_0}^{-4} H'\Sigma_1^{-1} H$, $\sigma_{\gamma_0}^2 = \mathrm{var}(\varepsilon_i)$, $H = (H_1, ..., H_K)$ *with* $H_k = -E[\psi_i^{-1}(\varepsilon_{i-k} - 1)\partial\psi_i/\partial\boldsymbol{\theta}]$, *and* Σ_1 *is as defined in Section 2.1.*

Let $\widetilde{\sigma}_{\gamma_0}^2 = n^{-1}\sum_{i=1}^{n}(\widetilde{\varepsilon}_i - 1)^2$, $\widetilde{H} = (\widetilde{H}_1, ..., \widetilde{H}_K)$ with $\widetilde{H}_k = -n^{-1}\sum_{i=1}^{n}\widetilde{\psi}_i^{-1}(\widetilde{\varepsilon}_{i-k} - 1)\partial\widetilde{\psi}_i/\partial\boldsymbol{\theta}$, and $\widetilde{\Omega} = \mathbf{I} - \widetilde{\sigma}_{\gamma_0}^{-4}\widetilde{H}'\widetilde{\Sigma}_1^{-1}\widetilde{H}$, where $\widetilde{\Sigma}_1$ is as defined in the previous subsection. From the proofs of Theorems 1 and 2, we can show that $\widetilde{\Omega}$ is a consistent estimator of Ω. Denote the diagonal elements of $\widetilde{\Omega}$ by $\widetilde{\Omega}_{kk}$ with $1 \leq k \leq K$. We then can check the significance of $\widetilde{r}_k$ by comparing its absolute value with $1.96\sqrt{\widetilde{\Omega}_{kk}/n}$, where the significance level is 5%.

To check the significance of $\widetilde{R} = (\widetilde{r}_1, ..., \widetilde{r}_K)'$ jointly, we can construct a portmanteau test statistic,

$$Q(K) = n\widetilde{R}'\widetilde{\Omega}^{-1}\widetilde{R},$$

and it will be asymptotically distributed as χ_K^2, the chi-squared distribution with K degrees of freedom.

3. Simulation experiments

In this section, we conduct three Monte Carlo simulation experiments to evaluate the finite-sample performance of the proposed inference tools in the previous section.

The first experiment is for the MLE $\widetilde{\boldsymbol{\lambda}}_n$ in Theorem 1, and the following three data generating processes are employed:

Fréchet ACD(1,1) model: $x_i = \psi_i \varepsilon_i, \quad \psi_i = 0.1 + 0.2x_{i-1} + 0.6\psi_{i-1}$;

Fréchet ACD(2,1) model: $x_i = \psi_i \varepsilon_i, \quad \psi_i = 0.1 + 0.1x_{i-1} + 0.3x_{i-2} + 0.5\psi_{i-1}$; and

Fréchet ACD(1,2) model: $x_i = \psi_i \varepsilon_i, \quad \psi_i = 0.1 + 0.2x_{i-1} + 0.5\psi_{i-1} + 0.1\psi_{i-2}$.

We consider the shape parameter $\gamma = 1.6$ and 5 for the associated standardized Fréchet distribution of ε_i, corresponding to a heavy-tailed distribution and a lighter-tailed one respectively, where the case with $\gamma = 1.6$ is employed to evaluate the robustness of the estimating procedure as the asymptotic normality in Theorem 1 requires γ to be greater than two. The sample size is set to $n = 200$, 500 or 1000, and there are 1000 replications for each sample size. Tables 1-3 list the bias, empirical standard deviations (ESDs) and asymptotic standard deviations (ASDs) of the MLE $\widetilde{\boldsymbol{\lambda}}_n = (\widetilde{\gamma}_n, \widetilde{\boldsymbol{\theta}}'_n)'$ for the three data generating processes respectively. It can be seen that almost all biases become smaller when the sample size n increases, and the biases of $\widetilde{\boldsymbol{\theta}}_n$ for the Fréchet ACD(1,1) model tend to be smaller than those of the other two models. For the estimator $\widetilde{\boldsymbol{\theta}}_n$, the ASDs are close to their corresponding ESDs when the sample size is as small as $n = 200$, except the estimators of β_1 and β_2 for the Fréchet ACD(1,2) model, in which case the ASDs and ESDs both have large values. For the estimator $\widetilde{\gamma}_n$, the discrepancies between its ASDs and ESDs are larger for the Fréchet ACD(1,2) model with larger γ. However, all ASDs are generally closer to their corresponding ESDs with an increasing sample size n. Note that as implied by the iterative functions (2), $\widetilde{\psi}_i(\boldsymbol{\theta})$ is a polynomial with respect to the β_j, while it is linear with respect to the α_j. Hence, it is not surprising that the estimating procedure will become less stable numerically when there are more parameters, especially more β_j's, in the model.

The second experiment is for the proposed diagnostic tools in section 2.2. We first evaluate the sample approximation for the asymptotic variance of residual autocorrelations Ω, and the data generating process is

$$x_i = \psi_i \varepsilon_i, \quad \psi_i = 0.1 + \alpha x_{i-1} + \beta \psi_{i-1},$$

with shape parameter $\gamma = 1.6$ or 5 for the associated Fréchet distribution, and $(\alpha, \beta)' = (0.2, 0.6)'$ or $(0.4, 0.5)'$ which corresponds to a stronger or weaker persistence of shocks respectively. As in the first experiment, the sample size is set to $n = 200$, 500 or 1000, and there are 1000 replications for each sample size. As shown in Table 4, the ASDs of the residual autocorrelations at lags 2, 4 and 6 are close to their corresponding ESDs when the sample size is as small as $n = 200$. Moreover, the discrepancies between ASDs and their corresponding ESDs are smaller when the generated sequence is lighter-tailed (i.e. $\gamma = 5$).

We next check the power of the proposed portmanteau test $Q(K)$ using the data generating process,

$$x_i = \psi_i \varepsilon_i, \quad \psi_i = 0.1 + 0.1 x_{i-1} + \alpha_2 x_{i-2} + 0.3 \psi_{i-1},$$

where $\alpha_2 = 0$, 0.2 or 0.4, and ε_i follows the standardized Fréchet distribution with $\gamma = 1.5$, 2 or 2.5. All the other settings are preserved from the first two experiments. We fit the model of orders $(1, 1)$ to the generated data; hence, the case with $\alpha_2 = 0$ corresponds to the size and those with $\alpha_2 > 0$ to the power. The rejection rates of test statistic $Q(K)$ with $K = 6$ are given in Table 5, where the critical value is the upper 5th percentile of the χ^2_6 distribution. The test is slightly sensitive, and the sizes are close to the nominal value of 0.05 when the sample size is $n = 1000$. While unsurprisingly the powers are larger as the sample size is larger, they are interestingly observed to have smaller values when the generated data are more heavy-tailed.

The last experiment compares the MLEs of the exponential ACD model, the Weibull ACD model and the Fréchet ACD model. The innovations associated with the Weibull ACD model follows the standardized Weibull distribution with mean of one and density function of the form,

$$f_\gamma(x) = \gamma b_\gamma x^{\gamma-1} \exp\{-b_\gamma x^\gamma\}, \quad x \geq 0, \tag{3}$$

where $\gamma > 0$ is the shape parameter, $b_\gamma = [\Gamma(1+\gamma^{-1})]^\gamma$, and $\Gamma(\cdot)$ is the Gamma function. The data generating process is the Fréchet ACD model,

$$x_i = \psi_i \varepsilon_i, \quad \psi_i = 0.1 + 0.2x_{i-1} + 0.6\psi_{i-1},$$

with $\gamma = 1.6$ or 2.4 for the associated Fréchet distribution. We set the sample size to $n = 1000$, and generated 1000 replications. Each generated sequence is estimated by the MLEs of the aforementioned three models. Boxplots for the estimators of $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$ are presented in Figure 2. While it is not surprising that the Fréchet ACD model has the best performance, the other two models even have inconsistent estimators for the parameters α and β. This further justifies the necessity of considering the Fréchet ACD model in real applications.

4. Two empirical examples

4.1. Durations of block trades on the SEHK

In the first empirical example, we consider the durations of block trades on the Hong Kong Stock Exchange (SEHK). The 50 stocks comprising the Hang Seng Index (HSI) are taken into account, with trades of 0.5 million Hong Kong dollars or greater sampled as block trades. The stocks are traded in two regular trading sessions on the SEHK: the morning session from 9:30 to 12:00 and the afternoon session from 13:00 to 16:00. We discarded the observations in the first 30 minutes of the morning session and the last 30 minutes of the afternoon session since they consist of extremely short durations even for block trades. Moreover, we treat multiple block trades within a second as a single trade; i.e., we ignore zero durations. Finally, each week is analyzed separately, and intersession durations and overnight durations are ignored.

As is well known in the literature, intraday duration series typically contain strong diurnal patterns. Specifically, the frequency of transactions is higher near the open and close of the market. A common practice is to first assume a deterministic function of time of day for the diurnal pattern, and then estimate this function by a semi- or non-parametric approach; see the cubic spline in Engle and Russell (1998) and Grammig and Maurer (2000), and the local regression smoothing in Zhang et al. (2001). Then the time-of-day detrended duration is calculated by $x_i = z_i/\phi(t_i)$, where t_i is the arrival time of the ith trade, $z_i = t_i - t_{i-1}$ is the observed duration, and $\phi(t_i)$ is the estimated diurnal pattern. We tried the detrending method in Engle and Russell (1998) on both the morning and afternoon sessions, but it has a poor performance for our dataset. For simplicity, we estimate $\phi(t_i)$ by fitting two cubic smoothing splines respectively for the two trading sessions with the R function *smooth.spline*. The knots are evenly spaced, and the number of knots are set to be 5 for the morning session and 6 for the afternoon session so that the intervals between any two consecutive knots are around 30 minutes.

We consider the durations of block trades on the SEHK in the following four weeks of 2014: January 6 to 10 (Week 1), January 13 to 17 (Week 2), January 20 to 24 (Week 3), and February 10 to 14 (Week 4). Note that the two weeks spanning from January 27 to February 7 are not included since they each contain a weekday on which SEHK is closed (i.e., January 31 and February 3). During this period, block trades account for around 4% of all trades on a normal trading day.

We first fit the Weibull ACD model of orders (p, q)=(1,2) to the four diurnally adjusted sequences, respectively. The distribution of the Weibull innovations are as specified in (3) in the previous section, and this specification applies to all the Weibull ACD models in the sequel. As shown in the upper panel of Figure 3, outliers above the reference line at the upper-right corner can be observed in the QQ plots of the standardized residuals from the four fitted Weibull ACD models, suggesting the right tails of the fitted Weibull distributions are not heavy enough for these datasets. This finding is indeed consistent with the observation in Engle and Russell (1998) where transaction durations data are fitted with the Weibull ACD model.

To reveal more details, we next concentrate on the durations of block trades in Week 1, and apply the proposed Fréchet ACD model as well as the Weibull ACD model to the diurnally adjusted sequence.

We first fit the data with the Weibull ACD model,

$$x_i = \psi_i \varepsilon_i, \;\; \psi_i = 0.0294_{0.0000} + 0.0919_{0.0001} x_{i-1} + 0.6614_{0.0113} \psi_{i-1} + 0.2185_{0.0096} \psi_{i-2}, \quad (4)$$

where ε_i follows the standardized Weibull distribution with parameter $\widetilde{\gamma}_n = 1.0573_{0.0001}$, and the subscripts of the parameter estimates are their associated standard errors. The portmanteau test $Q(K)$ has p-values of 0.2723, 0.5486 and 0.4225 for $K = 6$, 12 and 18, respectively, and the adequacy of the fitted model is further confirmed by the plot of residual autocorrelations in Figure 4 since the autocorrelations slightly exceed the reference values at lag 16 only.

We next consider the Fréchet ACD model. Note that both the right and left tails of the standardized Fréchet distribution become lighter when the shape parameter γ increases. As a result, this distribution has a very small probability near the origin (see Figure 6 for an illustration). This feature may affect the accuracy of its corresponding estimates. To overcome this problem, this section considers a slightly different standardization of Fréchet distributions. Specifically, we impose the following two conditions upon a 3-parameter Fréchet distribution: (i) $E(\varepsilon_i) = 1$; and (ii) $F_\gamma(0) = \delta$ for a predetermined small positive number δ, which is set to be 0.05 in the following. Moreover, in order to obtain a stable estimate, we fix the shape parameter γ in the estimating procedure.

With the value of γ chosen to be 3.5, the fitted Fréchet ACD model has the form of

$$x_i = \psi_i \varepsilon_i, \;\; \psi_i = 0.0354_{0.0000} + 0.1009_{0.0001} x_{i-1} + 0.6351_{0.0088} \psi_{i-1} + 0.2161_{0.0072} \psi_{i-2}, \quad (5)$$

where the portmanteau test statistic $Q(K)$ has p-values of 0.0789, 0.1143 and 0.1253 for $K = 6$, 12 and 18, respectively. Residual autocorrelations are all within the boundaries (see Figure 4), supporting the adequacy of the fitted model.

Lastly, the lower panel of Figure 3 presents the QQ plots of the standardized residuals from the fitted Fréchet ACD models for Weeks 1-4. It can be seen that the points fall approximately along the reference line, indicating that the Fréchet distribution fits the right tails better than the Weibull distribution.

4.2. Durations of block trades on the LSE

To further demonstrate the usefulness of the proposed model, in the second example we explore the behavior of the durations of block trades on the London Stock Exchange (LSE). The block trades of the FTSE 100 index components in the first five-day trading week of 2015 (i.e., January 5 to 9, 2015) are considered. Trades of 2 million pounds or greater are sampled as block trades, which account for around 2.2% of all trade during this period. Considering that the normal trading session of the LSE is from 8:00 to 16:30 without breaks, we discarded the observations in the first and the last 30 minutes of this period to avoid extremely short durations. The adjustments to the preliminary data are the same as those in the first empirical example, except that the number of knots is set to be 16 for the whole session. All the tick-by-tick transactions data used in this and the previous subsection are downloaded from Bloomberg.

For the durations of block trades on the LSE, we choose the value of γ to be 4, and the fitted Fréchet ACD model has the form of

$$x_i = \psi_i \varepsilon_i, \;\; \psi_i = 0.0511_{0.0000} + 0.0818_{0.0000} x_{i-1} + 0.4560_{0.0041} \psi_{i-1} + 0.3967_{0.0035} \psi_{i-2}.$$

The portmanteau test $Q(K)$ has p-values of 0.8728, 0.8539 and 0.3034 for $K = 6$, 12 and 18, respectively, and the adequacy of the fitted model is further confirmed by the plot of residual autocorrelations in Figure 5.

As a comparison, the fitted Weibull ACD(1,2) model to this dataset is

$$x_i = \psi_i \varepsilon_i, \;\; \psi_i = 0.0475_{0.0000} + 0.0848_{0.0000} x_{i-1} + 0.4359_{0.0038} \psi_{i-1} + 0.4326_{0.0033} \psi_{i-2},$$

with the estimated shape parameter $\widetilde{\gamma}_n = 1.0932_{0.0000}$. The portmanteau test $Q(K)$ has p-values of 0.8374, 0.8482 and 0.2668 for $K = 6$, 12 and 18, respectively, and the adequacy of the fitted model is further suggested by the plot of residual autocorrelations in Figure 5.

Similar to the first empirical example, the QQ plots of the standardized residuals from the two fitted models indicate that the Fréchet ACD model is preferable to the Weibull ACD model for this dataset (see the left panel of Figure 5).

Acknowledgment

We are grateful to the editor, an associate editor, and three anonymous referees for their valuable comments and constructive suggestions that led to the substantial improvement of this paper.

Appendix: Proofs of Theorems 1 and 2

Proof of Theorem 1. Our proof is split into three parts: some preliminary results; strong consistency; and asymptotic normality.

Part I. Some Preliminary Results

This part attempts to establish some preliminary results below:

$$\sup_{\boldsymbol{\theta}\in\Theta} |\widetilde{\psi}_i(\boldsymbol{\theta}) - \psi_i(\boldsymbol{\theta})| \leq \zeta\rho^i, \tag{6}$$

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial\widetilde{\psi}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} - \frac{\partial\psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \right\| \leq \zeta\rho^i, \quad \text{and} \quad \sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{\partial^2\widetilde{\psi}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2\psi_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \right\| \leq \zeta\rho^i, \tag{7}$$

where $0 < \rho < 1$, the random variable ζ is independent of i with $E|\zeta| < \infty$, and $\widetilde{\psi}_i(\boldsymbol{\theta})$ is defined based on the initial values of $x_0, \ldots, x_{1-p}, \psi_0(\boldsymbol{\theta}), \ldots, \psi_{1-q}(\boldsymbol{\theta})$.

When $p = q = 1$, it can be deduced that for $i \geq 2$,

$$\psi_i(\boldsymbol{\theta}) = \frac{\omega}{1-\beta_1} + \alpha_1 \sum_{j=0}^{\infty} \beta_1^j x_{i-j-1} = \sum_{j=0}^{i-2} \beta_1^j(\omega + \alpha_1 x_{i-j-1}) + \beta_1^{i-1}\psi_1(\boldsymbol{\theta}), \tag{8}$$

and similarly, $\widetilde{\psi}_i(\boldsymbol{\theta}) = \sum_{j=0}^{i-2} \beta_1^j(\omega + \alpha_1 x_{i-j-1}) + \beta_1^{i-1}\widetilde{\psi}_1(\boldsymbol{\theta})$. Therefore

$$\widetilde{\psi}_i(\boldsymbol{\theta}) - \psi_i(\boldsymbol{\theta}) = \beta_1^{i-1}[\widetilde{\psi}_1(\boldsymbol{\theta}) - \psi_1(\boldsymbol{\theta})] = \beta_1^{i-1}[\alpha_1(c - x_0) + \beta_1(c - \psi_0(\boldsymbol{\theta}))],$$

where c is the arbitrary starting value for x_0 and $\psi_0(\boldsymbol{\theta})$.

From Assumption 1, there exist $0 < \underline{\omega} < \overline{\omega} < \infty$ and $0 < \rho < 1$ such that $\underline{\omega} \leq \omega \leq \bar{\omega}$, $\alpha_1 \leq \rho$ and $\beta_1 \leq \rho$. Then by (8) we have $\sup_{\boldsymbol{\theta}\in\Theta} |\psi_0(\boldsymbol{\theta})| = \overline{\omega}(1-\rho)^{-1} + \sum_{j=1}^{\infty} \rho^j x_{-j}$, and hence also the result at (6),

$$\sup_{\boldsymbol{\theta}\in\Theta} |\widetilde{\psi}_i(\boldsymbol{\theta}) - \psi_i(\boldsymbol{\theta})| \leq \rho^i \sup_{\boldsymbol{\theta}\in\Theta}[|c - x_0| + |c - \psi_0(\boldsymbol{\theta})|] \leq \zeta\rho^i.$$

Hence, $E\sup_{\boldsymbol{\theta}\in\Theta} |\psi_0(\boldsymbol{\theta})| = \overline{\omega}(1-\rho)^{-1} + \sum_{j=1}^{\infty} \rho^j E(x_{-j}) = O(1)$. Using (8), similarly we can obtain the results at (7).

Analogous to (8), for general p and q, we have

$$\psi_i(\boldsymbol{\theta}) = \frac{\omega}{1 - \sum_{j=1}^{p} \beta_j} + \sum_{j=0}^{\infty} \iota_q' B^j \iota_q \sum_{l=1}^{p} \alpha_l x_{i-j-l}, \tag{9}$$

where $\iota_q = (1, 0, ..., 0)'$ and

$$B = \begin{pmatrix} \beta_1 & \cdots & \beta_{q-1} & \beta_q \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}.$$

By Lemma A.1 of Li and Li (2008b), we have $|\iota_q' B^j \iota_q| \leq M\rho^j$ for some $M > 0$ and $0 < \rho < 1$. Then the results at (6) and (7) can also be verified for general p and q.

Part II. Strong Consistency

To show the strong consistency of $\widetilde{\boldsymbol{\lambda}}_n$, as in, e.g., Huber (1967), Kukush et al. (2004) and Francq and Zakoian (2004), it is sufficient to establish the following intermediate results:

(i) For any $\overline{\gamma} > 0$, $\sup_{1<\gamma\leq\overline{\gamma},\boldsymbol{\theta}\in\Theta} n^{-1}|\widetilde{L}_n(\boldsymbol{\lambda}) - L_n(\boldsymbol{\lambda})| \to 0$ almost surely as $n \to \infty$.

(ii) The estimator of the shape parameter $\widetilde{\gamma}_n$ is stochastically bounded, i.e. there exist $1 < \underline{\gamma} \leq \gamma_0 \leq \overline{\gamma}$ such that $\widetilde{\gamma}_n \in [\underline{\gamma}, \overline{\gamma}]$ with probability one when n is large enough.

(iii) $E[l_1(\boldsymbol{\lambda})] \leq E[l_1(\boldsymbol{\lambda}_0)]$ for all $\boldsymbol{\lambda} \in \Lambda$, and the equality holds if and only if $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$.

(iv) Any $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0$ has a neighborhood $V(\boldsymbol{\lambda})$ such that

$$\limsup_{n\to\infty} \sup_{\boldsymbol{\lambda}^* \in V(\boldsymbol{\lambda}), \gamma\in[\underline{\gamma},\overline{\gamma}], \boldsymbol{\theta}\in\Theta} n^{-1}\widetilde{L}_n(\boldsymbol{\lambda}^*) < E[l_1(\boldsymbol{\lambda}_0)]$$

almost surely.

We first prove (i). It holds that $0 < c_\gamma < 1$ when $\gamma > 1$ and, by (6) and the Taylor series expansion of L_n as a function of $\psi_i(\boldsymbol{\theta})$, we have

$$\begin{aligned} \frac{1}{n}\left|\widetilde{L}_n(\boldsymbol{\lambda}) - L_n(\boldsymbol{\lambda})\right| &= \frac{1}{n}\left|\sum_{i=1}^n \frac{\gamma}{\psi_i^*(\boldsymbol{\theta})}\left\{1 - c_\gamma\left(\frac{x_i}{\psi_i^*(\boldsymbol{\theta})}\right)^{-\gamma}\right\}(\widetilde{\psi}_i(\boldsymbol{\theta}) - \psi_i(\boldsymbol{\theta}))\right| \\ &\leq \frac{\overline{\gamma}\zeta}{\underline{\omega}}\frac{1}{n}\sum_{i=1}^n \rho^i \left\{1 + c_\gamma\left(\frac{\psi_i^*(\boldsymbol{\theta})}{x_i}\right)^\gamma\right\} \\ &\leq \frac{1}{n}\frac{\rho(1-\rho^n)\overline{\gamma}\zeta}{(1-\rho)\underline{\omega}} + \frac{\overline{\gamma}\zeta}{\underline{\omega}}\frac{1}{n}\left[\sum_{i=1}^n \rho^{i/\gamma}\frac{\zeta\rho^i + \psi_i(\boldsymbol{\theta})}{x_i}\right]^\gamma, \end{aligned} \quad (10)$$

where $\psi_i^*(\boldsymbol{\theta})$ is between $\widetilde{\psi}_i(\boldsymbol{\theta})$ and $\psi_i(\boldsymbol{\theta})$, and then $\psi_i^*(\boldsymbol{\theta}) \leq |\widetilde{\psi}_i(\boldsymbol{\theta}) - \psi_i(\boldsymbol{\theta})| + \psi_i(\boldsymbol{\theta}) \leq \zeta\rho^i + \psi_i(\boldsymbol{\theta})$ with probability one. Note that $\rho^{i/\gamma} < \rho^i$,

$$\sum_{i=1}^{n} \rho^{2i}/x_i \leq \underline{\omega}^{-1} \sum_{i=1}^{\infty} \rho^{2i}\varepsilon_i^{-1} \quad \text{and} \quad \sum_{i=1}^{n} \rho^i \psi_i(\boldsymbol{\theta})/x_i \leq \underline{\omega}^{-1} \sum_{i=1}^{\infty} \rho^i \varepsilon_i^{-1} \sup_{\boldsymbol{\theta}\in\Theta} \psi_i(\boldsymbol{\theta}),$$

where $E|\sum_{i=1}^{\infty} \rho^{2i}\varepsilon_i^{-1}| < \infty$ and $E|\sum_{i=1}^{\infty} \rho^i\varepsilon_i^{-1} \sup_{\boldsymbol{\theta}\in\Theta} \psi_i(\boldsymbol{\theta})| < \infty$. As a result, the second term at the last line of (10) converges to zero almost surely as $n \to \infty$. Thus, we accomplish the proof for (i).

Next we prove (ii). We first show that $\widetilde{\gamma}_n$ is stochastically bounded from above. Since $\gamma > 1$ and $0 < c_\gamma < 1$, by elementary algebra, we can show

$$-\frac{c_\gamma}{2} x^{-\gamma} - (1+\gamma)\log x < 2\log c_\gamma^{-1}$$

and hence, for $x \neq 1$,

$$\begin{aligned}
\log f_\gamma(x) &= \log(\gamma c_\gamma) + I(x > 1)\{-(1+\gamma)\log x - c_\gamma x^{-\gamma}\} \\
&\quad + I(0 < x < 1)\{-\frac{c_\gamma}{2} x^{-\gamma} + [-\frac{c_\gamma}{2} x^{-\gamma} - (1+\gamma)\log x]\} \\
&\leq \log\gamma + 2\log c_\gamma^{-1} - (1+\gamma) I(x > 1)\log x - \frac{c_\gamma}{2} I(0 < x < 1) x^{-\gamma}.
\end{aligned}$$

As a result, we have

$$\begin{aligned}
\frac{1}{n}\widetilde{L}_n(\boldsymbol{\lambda}) &= \frac{1}{n}\sum_{i=1}^{n} \log f_\gamma\left(\frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})}\right) - \frac{1}{n}\sum_{i=1}^{n}\log\widetilde{\psi}_i(\boldsymbol{\theta}) \\
&\leq -(1+\gamma)\frac{1}{n}\sum_{i=1}^{n} I\left(\frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})} > 1\right)\log\left(\frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})}\right) \\
&\quad - \frac{c_\gamma}{2}\frac{1}{n}\sum_{i=1}^{n} I\left(0 < \frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})} < 1\right)\left(\frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})}\right)^{-\gamma} \\
&\quad + \log\gamma + 2\log c_\gamma^{-1} - \frac{1}{n}\sum_{i=1}^{n}\log\widetilde{\psi}_i(\boldsymbol{\theta}),
\end{aligned} \tag{11}$$

where

$$\frac{1}{n}\sum_{i=1}^{n} I\left(0 < \frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})} < 1\right)\left(\frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})}\right)^{-\gamma} \geq \left[\frac{1}{n}\sum_{i=1}^{n} I\left(\frac{\widetilde{\psi}_i(\boldsymbol{\theta})}{x_i} > 1\right)\frac{\widetilde{\psi}_i(\boldsymbol{\theta})}{x_i}\right]^{\gamma}. \tag{12}$$

Define a function $g_1(x) = \log(x) I(x > 1)$, and then it holds that $|g_1(x) - g_1(y)| \leq |x - y|$. By (6), we have

$$\frac{1}{n}\sum_{i=1}^{n} |g_1(x_i/\widetilde{\psi}_i(\boldsymbol{\theta})) - g_1(x_i/\psi_i(\boldsymbol{\theta}))| \leq \frac{\zeta}{\underline{\omega}^2}\frac{1}{n}\sum_{i=1}^{n}\rho^i x_i \to 0$$

with probability one. Moreover, by the ergodic theorem, we have

$$\frac{1}{n}\sum_{i=1}^{n} g_1(x_i/\psi_i(\boldsymbol{\theta})) \geq \frac{1}{n}\sum_{i=1}^{n} \inf_{\boldsymbol{\theta}\in\Theta} g_1(x_i/\psi_i(\boldsymbol{\theta})) \to K_1$$

with probability one, where $K_1 = E\inf_{\boldsymbol{\theta}\in\Theta} g_1(x_i/\psi_i(\boldsymbol{\theta})) > 0$. Thus, there exists an N_1 such that

$$n^{-1}\sum_{i=1}^{n} g_1(x_i/\widetilde{\psi}_i(\boldsymbol{\theta})) > 0.5K_1 \text{ as } n > N_1. \tag{13}$$

We define another function $g_2(x) = (x-1)I(x > 1)$, and it holds that $I(\widetilde{\psi}_i(\boldsymbol{\theta})/x_i > 1)\widetilde{\psi}_i(\boldsymbol{\theta})/x_i = g_2(\widetilde{\psi}_i(\boldsymbol{\theta})/x_i) + I(\widetilde{\psi}_i(\boldsymbol{\theta})/x_i > 1)$. By a method similar to the proof of Theorem 1 in Li et al. (2015), together with (6) and the fact that $|g_2(x)-g_2(y)| \leq |x-y|$, it can be verified that

$$\frac{1}{n}\sum_{i=1}^{n} |I(\widetilde{\psi}_i(\boldsymbol{\theta})/x_i > 1) - I(\psi_i(\boldsymbol{\theta})/x_i > 1)| \to 0$$

and

$$\frac{1}{n}\sum_{i=1}^{n} |g_2(\widetilde{\psi}_i(\boldsymbol{\theta})/x_i) - g_2(\psi_i(\boldsymbol{\theta})/x_i)| \leq \frac{\zeta}{\underline{\omega}}\frac{1}{n}\sum_{i=1}^{n} \rho^i \varepsilon_i^{-1} \to 0$$

with probability one. By the ergodic theorem again,

$$\frac{1}{n}\sum_{i=1}^{n} I(\psi_i(\boldsymbol{\theta})/x_i > 1)\psi_i(\boldsymbol{\theta})/x_i \geq \frac{1}{n}\sum_{i=1}^{n} \inf_{\theta\in\Theta} I(\psi_i(\boldsymbol{\theta})/x_i > 1)\psi_i(\boldsymbol{\theta})/x_i \to K_2$$

with probability one, where $K_2 = E\inf_{\theta\in\Theta} I(\psi_i(\boldsymbol{\theta})/x_i > 1)\psi_i(\boldsymbol{\theta})/x_i > 1$. Hence, there exists an N_2 such that

$$n^{-1}\sum_{i=1}^{n} I(\widetilde{\psi}_i(\boldsymbol{\theta})/x_i > 1)\widetilde{\psi}_i(\boldsymbol{\theta})/x_i > 1 + 0.5(K_2 - 1) \text{ as } n > N_2. \tag{14}$$

We can similarly handle the term $n^{-1}\sum_{i=1}^{n} \log \widetilde{\psi}_i(\boldsymbol{\theta})$ at (11). This, together with (11)-(14), implies that

$$\sup_{\gamma>C} \limsup_{n\to\infty} \sup_{\boldsymbol{\theta}\in\Theta} \frac{1}{n}\widetilde{L}_n(\boldsymbol{\lambda}) \to -\infty \quad \text{almost surely as } C \to +\infty. \tag{15}$$

We then show that $\widetilde{\gamma}_n$ is stochastically bounded from below. Observe that

$$\frac{1}{n}\widetilde{L}_n(\boldsymbol{\lambda}) = \log(\gamma c_\gamma) + \frac{\gamma}{n}\sum_{i=1}^{n} \log \widetilde{\psi}_i(\boldsymbol{\theta}) - \frac{1+\gamma}{n}\sum_{i=1}^{n} \log(x_i) - \frac{c_\gamma}{n}\sum_{i=1}^{n} \left(\frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})}\right)^{-\gamma}$$

and

$$-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})}\right)^{-\gamma} \leq -\left(\frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})}\right)^{-\gamma}.$$

Similarly to the proof of (15), by the ergodic theorem together with (6), we can show that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sup_{\boldsymbol{\theta}\in\Theta}\log\widetilde{\psi}_i(\boldsymbol{\theta}) = E\sup_{\boldsymbol{\theta}\in\Theta}\log\psi_i(\boldsymbol{\theta}) \leq \log\left(E\sup_{\boldsymbol{\theta}\in\Theta}\psi_i(\boldsymbol{\theta})\right)$$

and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sup_{\boldsymbol{\theta}\in\Theta}\frac{x_i}{\widetilde{\psi}_i(\boldsymbol{\theta})} = E\sup_{\boldsymbol{\theta}\in\Theta}\left(\frac{x_i}{\psi_i(\boldsymbol{\theta})}\right) \leq \frac{E(x_i)}{\underline{\omega}}$$

with probability one. Moreover, by the ergodic theorem,

$$\frac{1}{n}\sum_{i=1}^{n}\log(x_i) \to E[\log(\varepsilon_i)] - E[\log\psi_i(\boldsymbol{\theta}_0)]$$

almost surely. Hence, in view of the fact that $\gamma > 1$, $0 < c_\gamma < 1$, and $\log(c_\gamma) \to -\infty$ as $\gamma \to 1$, we have

$$\sup_{1<\gamma<\delta}\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}\in\Theta}\frac{1}{n}\widetilde{L}_n(\boldsymbol{\lambda}) \to -\infty \quad \text{almost surely as } \delta \to 1. \tag{16}$$

Furthermore,

$$\lim_{n\to\infty} n^{-1}\widetilde{L}_n(\boldsymbol{\lambda}_0) = E[l_1(\boldsymbol{\lambda}_0)] = \int_0^{+\infty} f_{\gamma_0}(x)\log f_{\gamma_0}(x)dx - E[\log\psi_1(\boldsymbol{\theta}_0)] \tag{17}$$

with probability one, where $\int_0^{+\infty} f_\gamma(x)\log f_\gamma(x)dx = \log\gamma - \gamma^{-1}\log c_\gamma - \gamma^{-1}(\gamma+1)\gamma_e - 1$ is finite for each γ, $E[\log\psi_1(\boldsymbol{\theta}_0)]$ is a constant and γ_e is Euler's constant. This, together with (15) and (16), leads to the existence of $1 < \underline{\gamma} \leq \gamma_0 \leq \overline{\gamma}$ such that $P(\{\omega : \widetilde{\gamma}_n(\omega) \in [\underline{\gamma}, \overline{\gamma}]$ as $n > n_0(\omega)\}) = 1$, where $n_0(\omega)$ is a large number, and depends on the realization ω. The proof for (ii) is accomplished.

Now we prove (iii). Denote $m(\boldsymbol{\theta}) = \psi_1(\boldsymbol{\theta}_0)/\psi_1(\boldsymbol{\theta})$, and then $m(\boldsymbol{\theta}) \in \mathcal{F}_0$ is independent of ε_1. Since

$$E\left[\frac{m(\boldsymbol{\theta})f_\gamma\left(m(\boldsymbol{\theta})\varepsilon_1\right)}{f_{\gamma_0}(\varepsilon_1)}\middle|\mathcal{F}_0\right] = \int_0^{\infty}\frac{m(\boldsymbol{\theta})f_\gamma\left(m(\boldsymbol{\theta})x\right)}{f_{\gamma_0}(x)}f_{\gamma_0}(x)\mathrm{d}x = 1,$$

we have $E\left[m(\boldsymbol{\theta})f_\gamma\left(m(\boldsymbol{\theta})\varepsilon_1\right)/f_{\gamma_0}(\varepsilon_1)\right] = 1$. Therefore by Jensen's inequality,

$$E[l_1(\boldsymbol{\lambda}) - l_1(\boldsymbol{\lambda}_0)] = E\left[\log\frac{m(\boldsymbol{\theta})f_\gamma\left(m(\boldsymbol{\theta})\varepsilon_1\right)}{f_{\gamma_0}(\varepsilon_1)}\right] \leq \log E\left[\frac{m(\boldsymbol{\theta})f_\gamma\left(m(\boldsymbol{\theta})\varepsilon_1\right)}{f_{\gamma_0}(\varepsilon_1)}\right] = 0,$$

with equality if and only if

$$f_{\gamma_0}(\varepsilon_1) = m(\boldsymbol{\theta}) f_\gamma \left(m(\boldsymbol{\theta})\varepsilon_1 \right)$$

with probability one. Evidently this holds when $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$. Conversely, if $f_{\gamma_0}(\varepsilon_1) = m(\boldsymbol{\theta}) f_\gamma \left(m(\boldsymbol{\theta})\varepsilon_1 \right)$ with probability one, then $E\left\{ \Pr\left\{ f_{\gamma_0}(\varepsilon_1) \neq m(\boldsymbol{\theta}) f_\gamma \left(m(\boldsymbol{\theta})\varepsilon_1 \right) | m(\boldsymbol{\theta}) \right\} \right\} = 0$. By Lemma 6.1.2 of Straumann (2005), we have

$$\{m(\boldsymbol{\theta}) \neq 1\} \subset \left\{ \Pr\{ f_{\gamma_0}(\varepsilon_1) \neq m(\boldsymbol{\theta}) f_\gamma \left(m(\boldsymbol{\theta})\varepsilon_1 \right) | m(\boldsymbol{\theta})\} > 0 \right\}.$$

Therefore $\Pr\left\{ m(\boldsymbol{\theta}) \neq 1 \right\} = 0$, i.e. $\psi_1(\boldsymbol{\theta}_0) = \psi_1(\boldsymbol{\theta})$ almost surely and it follows that $f_{\gamma_0}(\varepsilon_1) = f_\gamma(\varepsilon_1)$ almost surely.

On the one hand, since $\psi_1(\boldsymbol{\theta}_0) = \psi_1(\boldsymbol{\theta})$ almost surely, in view of (9), we have

$$\sum_{j=0}^{\infty} \sum_{l=1}^{p} \left[\iota_q' B^j \iota_q \alpha_l - \iota_q' B_0^j \iota_q \alpha_{0l} \right] x_{1-j-l} = \frac{\omega_0}{1 - \sum_{j=1}^{p} \beta_{0j}} - \frac{\omega}{1 - \sum_{j=1}^{p} \beta_j}$$

where $\iota_q = (1, 0, ..., 0)'$ and B_0 is the matrix B evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Suppose $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, then there exists a constant linear combination of the x_{1-j}, $j \geq 0$, and thus almost surely

$$x_1 - E(x_1 | \mathcal{F}_0) = \psi_1(\boldsymbol{\theta}_0)(\varepsilon_1 - 1) = 0,$$

which however is impossible because ε_1 is non-degenerate. Therefore we have $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

On the other hand, since $f_{\gamma_0}(\varepsilon_1) = f_\gamma(\varepsilon_1)$ almost surely and the map $(0, \infty) \times (1, \infty) \to (0, \infty) : (x, \gamma) \to f_\gamma(x)$ is continuous, we have $f_{\gamma_0}(x) = f_\gamma(x)$ for all $x \in (0, \infty)$. For the standardized Fréchet distribution, $\gamma_0 = \gamma$ necessarily follows.

Lastly we prove (iv). Let $\Lambda_1 = [\underline{\gamma}, \overline{\gamma}] \times \Theta$, where $1 < \underline{\gamma} < \overline{\gamma}$ are defined as in (ii). For any $\boldsymbol{\lambda} \in \Lambda_1$ with $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0$ and any positive integer k, let $V_k(\boldsymbol{\lambda})$ be the open ball with center $\boldsymbol{\lambda}$ and radius $1/k$, and let $U_k(\boldsymbol{\lambda}) = V_k(\boldsymbol{\lambda}) \bigcap \Lambda_1$. Owing to (i),

$$\begin{aligned}
\limsup_{n\to\infty} \sup_{\boldsymbol{\lambda}^* \in U_k(\boldsymbol{\lambda})} n^{-1} \widetilde{L}_n(\boldsymbol{\lambda}^*) &\leq \limsup_{n\to\infty} \sup_{\boldsymbol{\lambda}^* \in U_k(\boldsymbol{\lambda})} n^{-1} L_n(\boldsymbol{\lambda}^*) - \liminf_{n\to\infty} \sup_{\boldsymbol{\lambda} \in \Lambda_1} n^{-1} |\widetilde{L}_n(\boldsymbol{\lambda}) - L_n(\boldsymbol{\lambda})| \\
&\leq \limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sup_{\boldsymbol{\lambda}^* \in U_k(\boldsymbol{\lambda})} l_i(\boldsymbol{\lambda}^*).
\end{aligned}$$

Since $E[l_1^+(\boldsymbol{\lambda})] < \infty$, $E \sup_{\boldsymbol{\lambda}^* \in U_k(\boldsymbol{\lambda})} l_i(\boldsymbol{\lambda}^*) \in \mathbb{R} \bigcup \{-\infty\}$. We apply the following ergodic theorem to $\{\sup_{\boldsymbol{\lambda}^* \in U_k(\boldsymbol{\lambda})} l_i(\boldsymbol{\lambda}^*)\}_i$: if $\{X_i\}$ is a stationary and ergodic process such that

$EX_1 \in \mathbb{R} \bigcup \{-\infty\}$, then $n^{-1}\sum_{i=1}^{n} X_i$ converges almost surely to EX_1 when $n \to \infty$ (see, e.g. the proof of Theorem 2.1 of Francq and Zakoian (2004)). Therefore

$$\limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sup_{\boldsymbol{\lambda}^* \in U_k(\boldsymbol{\lambda})} l_i(\boldsymbol{\lambda}^*) = E \sup_{\boldsymbol{\lambda}^* \in U_k(\boldsymbol{\lambda})} l_1(\boldsymbol{\lambda}^*). \tag{18}$$

By the monotone convergence theorem, when k increases to ∞, $E\sup_{\boldsymbol{\lambda}^* \in U_k(\boldsymbol{\lambda})} l_1(\boldsymbol{\lambda}^*)$ decreases to $E[l_1(\boldsymbol{\lambda})]$. In view of (iii), (iv) is proved.

For any $\epsilon > 0$, from (iv) and the finite covering theorem, we have that $E[l_1(\boldsymbol{\lambda}_0)] - \delta \geq \limsup_{n\to\infty} \sup_{\boldsymbol{\lambda}\in\Lambda_1, |\boldsymbol{\lambda}-\boldsymbol{\lambda}_0| \geq \epsilon} n^{-1}\widetilde{L}_n(\boldsymbol{\lambda})$ with probability one for some $\delta > 0$. Moreover, it is implied by (17) that

$$\liminf_{n\to\infty} n^{-1}\widetilde{L}_n(\boldsymbol{\lambda}_0) > E[l_1(\boldsymbol{\lambda}_0)] - 0.5\delta > E[l_1(\boldsymbol{\lambda}_0)] - \delta \geq \limsup_{n\to\infty} \sup_{\boldsymbol{\lambda}\in\Lambda_1, |\boldsymbol{\lambda}-\boldsymbol{\lambda}_0| \geq \epsilon} n^{-1}\widetilde{L}_n(\boldsymbol{\lambda})$$

with probability one. In view of (ii), as a result, there exists a large number $n_1(\omega, \epsilon)$ such that $P(\{\omega : |\widetilde{\boldsymbol{\lambda}}_n(\omega) - \boldsymbol{\lambda}_0| < \epsilon \text{ as } n > n_1(\omega, \epsilon)\}) = 1$, where $n_1(\omega, \epsilon)$ depends on ϵ and the realization ω. Thus the strong consistency follows.

Part III. Asymptotic Normality

We first gives the derivatives of the function of $l_i(\boldsymbol{\lambda})$ as follows. Write $\psi_i = \psi_i(\boldsymbol{\theta})$. The first derivatives are given by

$$\begin{aligned}
\frac{\partial l_i}{\partial \gamma} &= \left\{ c_\gamma \left(\frac{x_i}{\psi_i}\right)^{-\gamma} - 1 \right\} \log\left(\frac{x_i}{\psi_i}\right) - c'_\gamma \left(\frac{x_i}{\psi_i}\right)^{-\gamma} + \frac{1}{\gamma} + \frac{c'_\gamma}{c_\gamma}, \\
\frac{\partial l_i}{\partial \boldsymbol{\theta}} &= \gamma \left\{ 1 - c_\gamma \left(\frac{x_i}{\psi_i}\right)^{-\gamma} \right\} \left\{ \frac{1}{\psi_i} \frac{\partial \psi_i}{\partial \boldsymbol{\theta}} \right\},
\end{aligned}$$

and the second derivatives are given by,

$$\begin{aligned}
\frac{\partial^2 l_i}{\partial \gamma^2} &= \left(\frac{x_i}{\psi_i}\right)^{-\gamma} \left\{ \left[2c'_\gamma - c_\gamma \log\left(\frac{x_i}{\psi_i}\right) \right] \log\left(\frac{x_i}{\psi_i}\right) - c''_\gamma \right\} - \frac{1}{\gamma^2} + \frac{c''_\gamma}{c_\gamma} - \left(\frac{c'_\gamma}{c_\gamma}\right)^2, \\
\frac{\partial^2 l_i}{\partial \boldsymbol{\theta} \partial \gamma} &= \left\{ 1 - \left(\frac{x_i}{\psi_i}\right)^{-\gamma} \left[c_\gamma + \gamma c'_\gamma - \gamma c_\gamma \log\left(\frac{x_i}{\psi_i}\right) \right] \right\} \left\{ \frac{1}{\psi_i} \frac{\partial \psi_i}{\partial \boldsymbol{\theta}} \right\}, \\
\frac{\partial^2 l_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} &= \gamma \left\{ 1 - c_\gamma \left(\frac{x_i}{\psi_i}\right)^{-\gamma} \right\} \left\{ \frac{1}{\psi_i} \frac{\partial^2 \psi_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\} \\
&\quad - \gamma \left\{ 1 + (\gamma - 1) c_\gamma \left(\frac{x_i}{\psi_i}\right)^{-\gamma} \right\} \left\{ \frac{1}{\psi_i} \frac{\partial \psi_i}{\partial \boldsymbol{\theta}} \right\} \left\{ \frac{1}{\psi_i} \frac{\partial \psi_i}{\partial \boldsymbol{\theta}'} \right\},
\end{aligned}$$

where $c'_\gamma = \partial c_\gamma / \partial \gamma$ and $c''_\gamma = \partial^2 c_\gamma / \partial \gamma^2$. We can verifty that

$$E(\varepsilon_i^{-\gamma}) < \infty, \quad E(\log \varepsilon_i) < \infty, \quad E(\varepsilon_i^{-\gamma} \log \varepsilon_i) < \infty, \quad E(\varepsilon_i^{-\gamma} (\log \varepsilon_i)^2) < \infty.$$

Due to the consistency, there exists a compact set $\Theta_1 \subset \Theta$ such that $\boldsymbol{\theta}_0 \in \Theta_1$ and each element of Θ_1 is bounded away from zero. It then can be verified that

$$E \sup_{\boldsymbol{\theta} \in \Theta_1} \left\| \frac{1}{\psi_i(\boldsymbol{\theta})} \frac{\partial \psi_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^2 < \infty, \quad \text{and} \quad E \sup_{\boldsymbol{\theta} \in \Theta_1} \left\| \frac{1}{\psi_i(\boldsymbol{\theta})} \frac{\partial^2 \psi_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| < \infty. \tag{19}$$

Consider a Taylor expansion of the score vector around $\boldsymbol{\lambda_0}$,

$$0 = n^{-1/2} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\lambda}} \widetilde{l}_i(\widetilde{\boldsymbol{\lambda}}_n) = n^{-1/2} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\lambda}} \widetilde{l}_i(\boldsymbol{\lambda}_0) + \left[n^{-1} \sum_{i=1}^n \frac{\partial^2 \widetilde{l}_i(\boldsymbol{\lambda}^*)}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \right] \sqrt{n}(\widetilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0),$$

where the $\boldsymbol{\lambda}^*$ are between $\widetilde{\boldsymbol{\lambda}}_n$ and $\boldsymbol{\lambda}_0$. It suffices to show

(i) $n^{-1/2} \sum_{i=1}^n \partial \widetilde{l}_i(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda} \to_d N(0, \Sigma)$,

(ii) $n^{-1} \sum_{i=1} \partial^2 \widetilde{l}_i(\boldsymbol{\lambda}^*)/\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}' \to \Sigma$ in probability, and

(iii) the matrix Σ is positive definite.

We first prove (i). Since $\Sigma = E(\partial l_i(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda})(\partial l_i(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda'}) < \infty$ from (19), and $E(\partial l_i(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda} | \mathcal{F}_{i-1}) = 0$, for any $\boldsymbol{x} \in \mathbb{R}^{2+p+q}$, the sequence $\{(\partial l_i(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda}')\boldsymbol{x}, \mathcal{F}_i, i = 1, ..., n\}$ is a finite variance stationary ergodic martingale difference. By the central limit theorem and the Cramér-Wold device, we obtain $n^{-1/2} \sum_{i=1}^n \partial l_i(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda} \to_d N(0, \Sigma)$. Moreover, by a method similar to (10), $n^{-1/2} \sum_{i=1}^n |\partial \widetilde{l}_i(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda} - \partial \widetilde{l}_i(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\lambda}| \to 0$ almost surely. This accomplishes the proof for (i).

Now we show (ii). By a method similar to (10) again,

$$\frac{1}{n} \sup_{2 < \gamma \le \overline{\gamma}, \boldsymbol{\theta} \in \Theta_1} \left| \sum_{i=1}^n \frac{\partial^2 \widetilde{l}_i(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} - \frac{\partial^2 l_i(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \right| \to 0 \tag{20}$$

with probability one. From (19), we further verify that $E \sup_{2 < \gamma \le \overline{\gamma}, \boldsymbol{\theta} \in \Theta_1} |\partial^2 l_i(\boldsymbol{\lambda})/\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'| < \infty$. Note that $E\{\partial^2 l_i(\boldsymbol{\lambda})/\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda'} - E[\partial^2 l_i(\boldsymbol{\lambda})/\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}']\} = 0$. By applying Theorem 3.1 of Ling and McAleer (2003), we have

$$\frac{1}{n} \sup_{2 < \gamma \le \overline{\gamma}, \boldsymbol{\theta} \in \Theta_1} \left| \sum_{i=1}^n \frac{\partial^2 l_i(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} - E \frac{\partial^2 l_i(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \right| = o_p(1),$$

which, together with the consistency of $\widetilde{\boldsymbol{\lambda}}_n$ and (20), implies (ii).

Finally we prove (iii). Note that Σ is positive semi-definite. Assume there exists $\boldsymbol{x} = (u, \boldsymbol{v}')' \neq \mathbf{0}$ with $u \in \mathbb{R}$ and $\boldsymbol{v} \in \mathbb{R}^{1+p+q}$ such that $\boldsymbol{x}'\Sigma\boldsymbol{x} = 0$. Equivalently, $\{\partial l_i(\boldsymbol{\lambda_0})/\partial\gamma\}u + \{\partial l_i(\boldsymbol{\lambda_0})/\partial\boldsymbol{\theta}'\}\boldsymbol{v} = 0$ almost surely. The proof is split into the following three subcases:

(a) $u \neq 0$ and $\boldsymbol{v} = \mathbf{0}$. Then $\kappa_2 = \text{var}[c_2(\varepsilon_i, \gamma_0)] = 0$, where $c_2(x,\gamma) = c_\gamma x^{-\gamma}\log(x) - \log(x) - c'_\gamma x^{-\gamma} + \gamma^{-1} + c'_\gamma/c_\gamma$. This implies $c_2(\varepsilon_i, \gamma_0) = 0$ almost surely, which is however impossible because ε_i is non-degenerate.

(b) $u = 0$ and $\boldsymbol{v} \neq \mathbf{0}$. Then almost surely

$$\frac{\partial l_i(\boldsymbol{\lambda_0})}{\partial\boldsymbol{\theta}'}\boldsymbol{v} = \gamma_0\left(1 - c_{\gamma_0}\varepsilon_i^{-\gamma_0}\right)\frac{\partial\psi_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\frac{\boldsymbol{v}}{\psi_i(\boldsymbol{\theta}_0)} = 0.$$

First note that $E\{\psi_i^{-2}(\boldsymbol{\theta}_0)(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta})(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}')\}$ is positive definite. This can be proved by contradiction. Suppose there exists a nonzero vector $\boldsymbol{w} \in \mathbb{R}^{p+q+1}$ such that $\boldsymbol{w}'E\{\psi_i^{-2}(\boldsymbol{\theta}_0)(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta})(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}')\}\boldsymbol{w} = 0$, then $(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}')\boldsymbol{w} = 0$ almost surely, and hence in view of (2) and the stationarity of $\{\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\}$, we have almost surely

$$\begin{aligned} 0 = \frac{\partial\psi_i(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\boldsymbol{w} &= (1, x_{i-1}, \ldots, x_{i-p}, \psi_{i-1}(\boldsymbol{\theta}_0), \ldots, \psi_{i-q}(\boldsymbol{\theta}_0))\boldsymbol{w} + \sum_{j=1}^{q}\beta_j\frac{\partial\psi_{i-j}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\boldsymbol{w} \\ &= (1, x_{i-1}, \ldots, x_{i-p}, \psi_{i-1}(\boldsymbol{\theta}_0), \ldots, \psi_{i-q}(\boldsymbol{\theta}_0))\boldsymbol{w}. \end{aligned}$$

However, the condition that the polynomials $\sum_{j=1}^{p}\alpha_j x^j$ and $1 - \sum_{j=1}^{q}\beta_j x^j$ have no common root stated in Assumption 1 implies that the definition (1) is *minimal*: there is no (p^*, q^*) such that $p^* < p$ or $q^* < q$ and $\psi_i = \omega^* + \sum_{j=1}^{p^*}\alpha_j^* x_{i-j} + \sum_{j=1}^{q^*}\beta_j^*\psi_{i-j}$. Since $\boldsymbol{w} \neq 0$, a contradiction results. Thus, $E\{\psi_i^{-2}(\boldsymbol{\theta}_0)(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta})(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}')\}$ must be positive definite, implying $P\{\psi_i(\boldsymbol{\theta}_0)^{-1}(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}')\boldsymbol{v} \neq 0\} > 0$. Due to the independence of $\gamma_0\left(1 - c_{\gamma_0}\varepsilon_i^{-\gamma_0}\right)$ and $\psi_i(\boldsymbol{\theta}_0)^{-1}(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}')\boldsymbol{v}$, we have $\gamma_0(1 - c_{\gamma_0}\varepsilon_i^{-\gamma_0}) = 0$ with probability one, which is again impossible.

(c) $u \neq 0$ and $\boldsymbol{v} \neq \mathbf{0}$. Then $\psi_i(\boldsymbol{\theta}_0)^{-1}(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}')\boldsymbol{v} = -u\gamma_0^{-1}c_2(\varepsilon_i, \gamma_0)/\left(1 - c_{\gamma_0}\varepsilon_i^{-\gamma_0}\right)$ almost surely. However, since the right-hand side of this equation is non-degenerate, it contradicts the independence between ε_i and $\psi_i^{-1}(\boldsymbol{\theta}_0)(\partial\psi_i(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta})$. Therefore Σ is positive definite.

This completes the proof of Theorem 1. □

Proof of Theorem 2. Denote $\widetilde{\psi}_i(\widetilde{\boldsymbol{\theta}}_n)$ and $\psi_i(\boldsymbol{\theta}_0)$ respectively by $\widetilde{\psi}_i$ and ψ_i, and let $\widetilde{\varepsilon}_i = x_i/\widetilde{\psi}_i$. Let $\widetilde{C} = (\widetilde{C}_1, ..., \widetilde{C}_K)'$ and $C = (C_1, ..., C_K)'$, where

$$\widetilde{C}_k = \frac{1}{n}\sum_{i=k+1}^{n}(\widetilde{\varepsilon}_i - 1)(\widetilde{\varepsilon}_{i-k} - 1) \quad \text{and} \quad C_k = \frac{1}{n}\sum_{i=k+1}^{n}(\varepsilon_i - 1)(\varepsilon_{i-k} - 1).$$

By (6) in the proof of Theorem 1, the $\sqrt{n}$-consistency of $\widetilde{\boldsymbol{\theta}}_n$ and the ergodic theorem, it follows that

$$\frac{1}{n}\sum_{i=1}^{n}(\widetilde{\varepsilon}_i - 1)^2 = \sigma_{\gamma_0}^2 + o_p(1), \tag{21}$$

where $\sigma_{\gamma_0}^2 = \text{var}(\varepsilon_i)$, and thus it suffices to derive the asymptotic distribution of $\widetilde{C}$.

By (7) in the proof of Theorem 1 and the Taylor expansion, it holds that

$$\widetilde{C} = C + H'(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(n^{-1/2}), \tag{22}$$

where $H = (H_1, ..., H_K)$ and $H_k = -E[\psi_i^{-1}(\varepsilon_{i-k} - 1)\partial\psi_i/\partial\boldsymbol{\theta}]$. From the proof of Theorem 1, we have

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = A\Sigma^{-1} \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[c_2(\varepsilon_i, \gamma_0), \frac{c_1(\varepsilon_i, \gamma_0)}{\psi_i}\frac{\partial\psi_i}{\partial\boldsymbol{\theta}'}\right]' + o_p(1), \tag{23}$$

where the $c_j(\varepsilon_i, \gamma_0)$ is as defined in Section 2.1, and the matrix $A = (0, \mathbf{I})$ with $\mathbf{I}$ being the $(p+q+1)$-dimensional identity matrix. Note that $E[\varepsilon_i c_2(\varepsilon_i, \gamma_0)] = 0$ and $E[\varepsilon_i c_1(\varepsilon_i, \gamma_0)] = 1$. By (22), (23), the central limit theorem and the Cramér-Wold device, the theorem follows. □

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69:542–547.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association*, 65:1509–1526.

Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, Berlin.

Engle, R. F. (1982). Autoregression conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68:1–22.

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

Francq, C. and Zakoian, J. M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10:605–637.

Grammig, J. and Maurer, K.-O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal*, 3:16–38.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 221–233, Berkeley. University of California Press.

Kukush, A., Chernikov, Y., and Pfeifer, D. (2004). Maximum likelihood estimators in a statistical model of natural catastrophe claims with trend. *Extremes*, 7:309–336.

Li, G. and Li, W. K. (2005). Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika*, 92:691–701.

Li, G. and Li, W. K. (2008a). Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika*, 95:399–414.

Li, G. and Li, W. K. (2008b). Testing for threshold moving average with conditional heteroscedasticity. *Statistica Sinica*, 18:647–665.

Li, G., Li, Y., and Tsai, C.-L. (2015). Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association*, 110:246–261.

Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory*, 19:280–310.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65:297–303.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared residual autocorrelations. *Journal of Time Series Analysis*, 4:269–273.

Mikosch, T. and Starica, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *The Annals of Statistics*, 28:1427–1451.

Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika*, 90:967–975.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*. Springer, Berlin.

Zhang, M., Russell, J. R., and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104:179–207.

Table 1: Estimation results of Fréchet ACD(1,1) models with $\boldsymbol{\theta} = (\omega, \alpha, \beta)' = (0.1, 0.2, 0.6)'$ and $\gamma = 1.6$ or 5.

	n		γ	ω	α	β
$\gamma = 1.6$	200	Bias	0.0149	0.0078	0.0060	-0.0203
		ESD	0.0947	0.0406	0.0655	0.1175
		ASD	0.0892	0.0431	0.0697	0.1081
	500	Bias	0.0047	0.0031	0.0024	-0.0070
		ESD	0.0623	0.0231	0.0399	0.0674
		ASD	0.0561	0.0244	0.0425	0.0626
	1000	Bias	0.0020	0.0002	0.0001	0.0003
		ESD	0.0410	0.0146	0.0269	0.0434
		ASD	0.0396	0.0162	0.0292	0.0421
$\gamma = 5$	200	Bias	0.0820	0.0150	-0.0012	-0.0294
		ESD	0.5947	0.0481	0.0474	0.1225
		ASD	0.2867	0.0461	0.0463	0.1182
	500	Bias	0.0172	0.0049	0.0007	-0.0106
		ESD	0.1725	0.0258	0.0299	0.0700
		ASD	0.1770	0.0255	0.0289	0.0680
	1000	Bias	0.0081	0.0027	0.0009	-0.0062
		ESD	0.1279	0.0176	0.0209	0.0483
		ASD	0.1241	0.0174	0.0203	0.0467

Table 2: Estimation results of Fréchet ACD(2,1) models with $\boldsymbol{\theta} = (\omega, \alpha_1, \alpha_2, \beta)' = (0.1, 0.1, 0.3, 0.5)'$ and $\gamma = 1.6$ or 5.

	n		γ	ω	α_1	α_2	β
$\gamma = 1.6$	200	Bias	0.0338	0.0083	0.0081	-0.0194	-0.0166
		ESD	0.0858	0.0374	0.0547	0.0904	0.1101
		ASD	0.0920	0.0390	0.0551	0.1026	0.1014
	500	Bias	0.0011	0.0041	0.0030	-0.0020	-0.0039
		ESD	0.0598	0.0216	0.0329	0.0616	0.0632
		ASD	0.0566	0.0233	0.0345	0.0675	0.0598
	1000	Bias	-0.0024	0.0026	0.0024	0.0008	-0.0032
		ESD	0.0430	0.0151	0.0224	0.0417	0.0425
		ASD	0.0395	0.0158	0.0238	0.0470	0.0411
$\gamma = 5$	200	Bias	0.0324	0.0183	0.0058	0.0012	-0.0253
		ESD	0.3172	0.0439	0.0467	0.0686	0.0883
		ASD	0.2812	0.0427	0.0493	0.0684	0.0888
	500	Bias	0.0060	0.0072	0.0008	0.0007	-0.0082
		ESD	0.1903	0.0248	0.0304	0.0425	0.0535
		ASD	0.1749	0.0240	0.0308	0.0428	0.0533
	1000	Bias	0.0061	0.0032	0.0005	-0.0004	-0.0032
		ESD	0.1234	0.0161	0.0223	0.0304	0.0370
		ASD	0.1235	0.0162	0.0217	0.0303	0.0371

Table 3: Estimation results of Fréchet ACD(1,2) models with $\boldsymbol{\theta} = (\omega, \alpha, \beta_1, \beta_2)' = (0.1, 0.2, 0.5, 0.1)'$ and $\gamma = 1.6$ or 5.

	n		γ	ω	α	β_1	β_2
$\gamma = 1.6$	200	Bias	0.0307	0.0079	0.0177	-0.1495	0.1160
		ESD	0.1254	0.0425	0.0687	0.1737	0.1400
		ASD	0.0886	0.0445	0.0751	0.2663	0.2251
	500	Bias	0.0072	0.0060	0.0108	-0.0864	0.0666
		ESD	0.1020	0.0242	0.0412	0.1345	0.1073
		ASD	0.0559	0.0265	0.0472	0.1801	0.1454
	1000	Bias	0.0023	0.0036	0.0059	-0.0469	0.0348
		ESD	0.0432	0.0164	0.0280	0.1030	0.0812
		ASD	0.0396	0.0180	0.0329	0.1348	0.1069
$\gamma = 5$	200	Bias	0.0693	0.0103	0.0150	-0.1434	0.1102
		ESD	0.3865	0.0448	0.0494	0.1659	0.1464
		ASD	0.2830	0.0470	0.0506	0.2613	0.2312
	500	Bias	0.0513	0.0077	0.0083	-0.0844	0.0612
		ESD	0.4735	0.0280	0.0316	0.1312	0.1099
		ASD	0.1770	0.0282	0.0323	0.1763	0.1525
	1000	Bias	0.0109	0.0037	0.0047	-0.0480	0.0360
		ESD	0.1220	0.0189	0.0217	0.0998	0.0856
		ASD	0.1245	0.0189	0.0229	0.1291	0.1105

Table 4: Empirical standard deviations (ESD) and asymptotic standard deviations (ASD) of residual autocorrelations at lags 2, 4 and 6, for Fréchet ACD(1,1) models with $\boldsymbol{\theta} = (\omega, \alpha, \beta)' = (0.1, 0.2, 0.6)'$ or $(0.1, 0.4, 0.5)'$ and $\gamma = 1.6$ or 5.

			$\boldsymbol{\theta} = (0.1, 0.2, 0.6)'$			$\boldsymbol{\theta} = (0.1, 0.4, 0.5)'$		
	n		2	4	6	2	4	6
$\gamma = 1.6$	200	ESD	0.0556	0.0517	0.0548	0.0588	0.0577	0.0560
		ASD	0.0671	0.0678	0.0681	0.0670	0.0674	0.0678
	500	ESD	0.0366	0.0393	0.0405	0.0393	0.0362	0.0381
		ASD	0.0436	0.0438	0.0439	0.0436	0.0438	0.0439
	1000	ESD	0.0272	0.0275	0.0282	0.0264	0.0276	0.0260
		ASD	0.0312	0.0313	0.0313	0.0313	0.0313	0.0313
$\gamma = 5$	200	ESD	0.0741	0.0755	0.0747	0.0632	0.0651	0.0659
		ASD	0.0664	0.0680	0.0690	0.0674	0.0686	0.0691
	500	ESD	0.0420	0.0427	0.0412	0.0418	0.0411	0.0431
		ASD	0.0424	0.0431	0.0436	0.0428	0.0435	0.0437
	1000	ESD	0.0294	0.0305	0.0299	0.0295	0.0306	0.0302
		ASD	0.0300	0.0305	0.0308	0.0303	0.0308	0.0309

Table 5: Rejection rates of test statistic $Q(K)$ with $K = 6$ and $\gamma = 1.5$, 2 or 2.5.

	$\alpha_2 = 0$			$\alpha_2 = 0.2$			$\alpha_2 = 0.4$		
n	1.5	2	2.5	1.5	2	2.5	1.5	2	2.5
200	0.089	0.110	0.127	0.106	0.154	0.231	0.132	0.284	0.451
500	0.077	0.074	0.091	0.118	0.233	0.333	0.185	0.508	0.761
1000	0.056	0.059	0.054	0.155	0.299	0.522	0.268	0.689	0.915

Figure 2: Boxplots for maximum likelihood estimators of $\boldsymbol{\theta}$ by exponential ACD (EACD) models, Weibull ACD (WACD) models and Fréchet ACD (FACD) models. The data generating process is the Fréchet ACD model with $\boldsymbol{\lambda} = (1.6, 0.1, 0.2, 0.6)'$ (left panel) or $\boldsymbol{\lambda} = (2.4, 0.1, 0.2, 0.6)'$ (right panel).

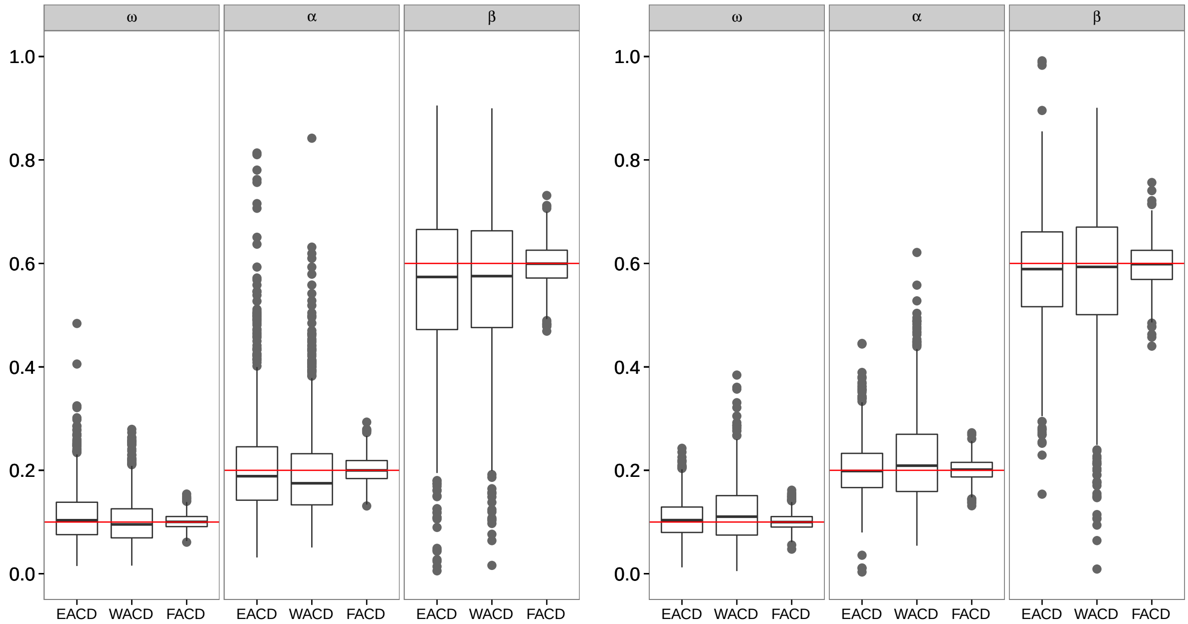

Figure 3: QQ plots of the standardized residuals from all fitted models for the SEHK block trade durations.

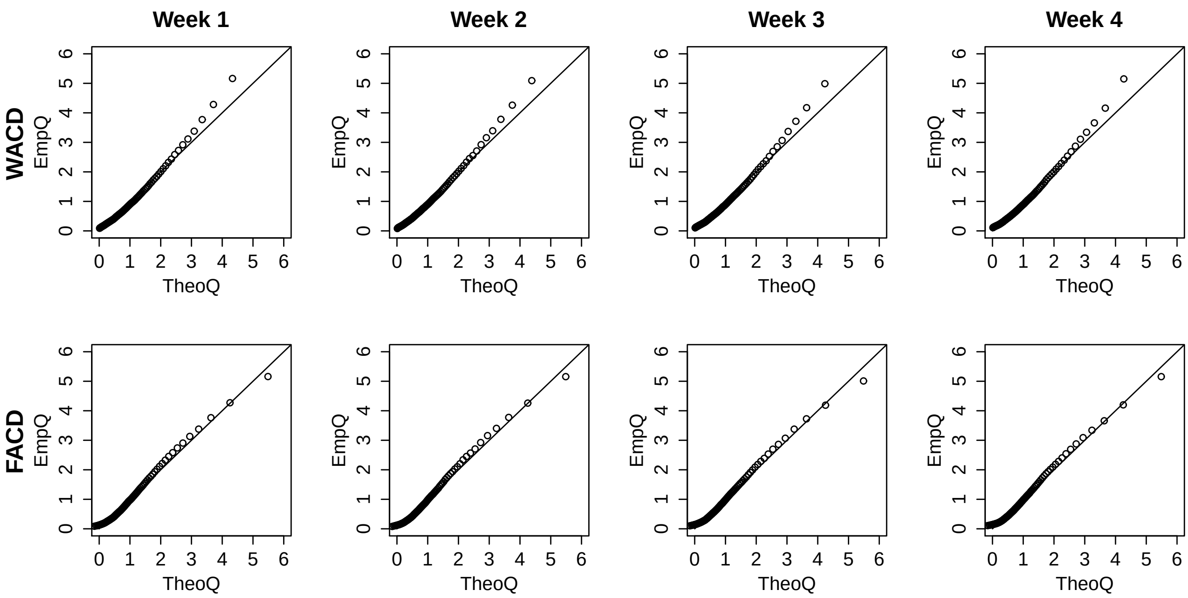

Figure 4: Estimation results of the Fréchet ACD model and the Weibull ACD model for the SEHK block trade durations (Week 1). They include QQ plots of the standardized residuals from the fitted models and the residual autocorrelations.

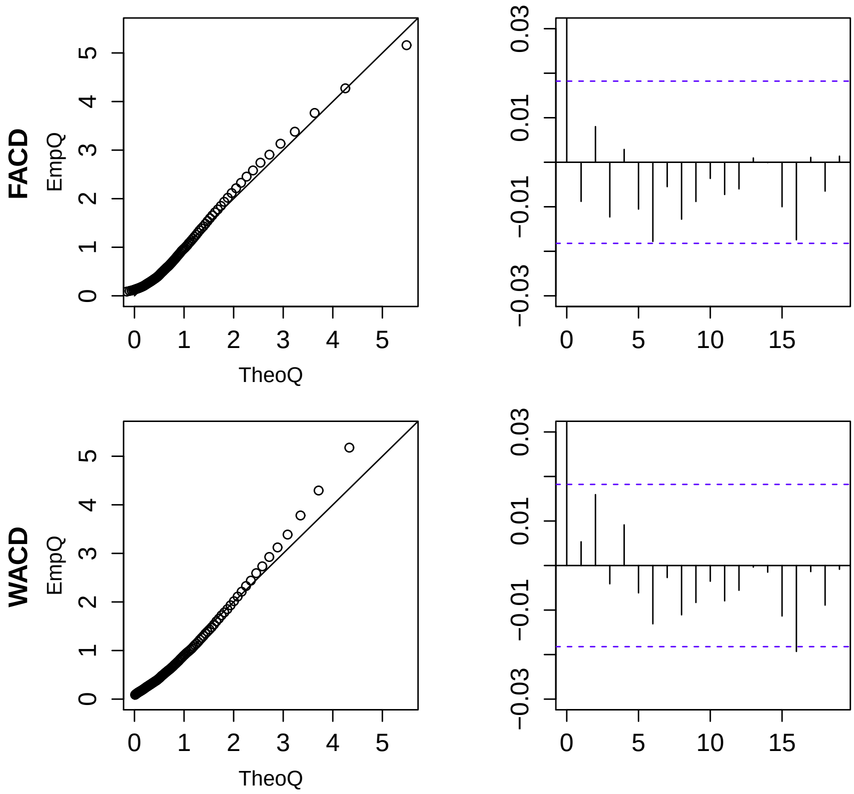

Figure 5: Estimation results of the Fréchet ACD model and the Weibull ACD model for the LSE block trade durations. They include QQ plots of the standardized residuals from the fitted models and the residual autocorrelations.

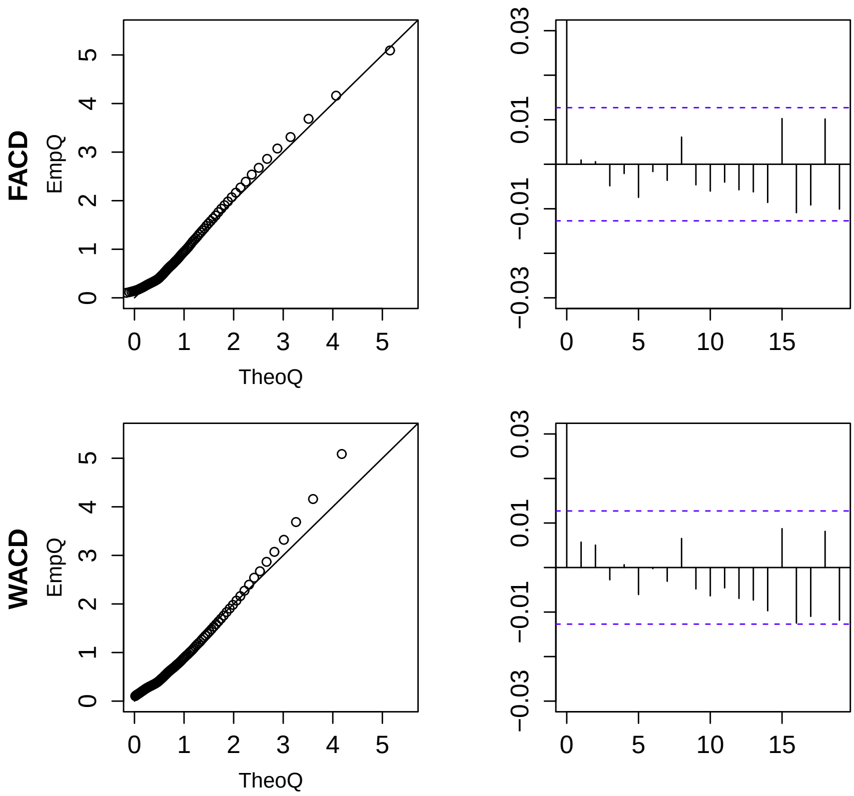

Figure 6: Density functions of the standardized Fréchet distribution with γ =1.5 (black line), 2 (gray line) and 3 (light gray line).