UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

CENTER FOR INTERNATIONAL ECONOMICS

Working Paper Series

Working Paper No. 2019-08

A Box-Cox semiparametric multiplicative error model

Xuehai Zhang

August 2019

CENTER FOR INTERNATIONAL ECONOMICS

A Box-Cox semiparametric multiplicative error model

Xuehai Zhang*

Faculty of Business Administration and Economics, Paderborn University, Germany

Abstract

A general class of SemiMEM (semiparametric multiplicative error) models is proposed by introducing a scale function into a MEM (multiplicative error) class model to analyze the non-negative observations. The estimation of the scale function is not limited by any parametric models specification and the moments condition is also reduced via the Box-Cox transformation. For the purpose, an equivalent scale function is applied in a local linear approach and converted to the scale function under weak moment conditions. The equivalent scale function estimation and the bandwidth, the constant factor in the asymptotic variance and the power transformation parameters estimation are proposed based on the iterative plug-in (IPI) algorithms. In the power transformation estimation, the maximum likelihood estimation (MLE), the normality test and the the quantile-quantile regression (QQr) are employed and simulation algorithms for the confidence interval of estimated power transformation parameter are also developed by the block bootstrap method. The algorithms fit the selected real data well.

1 Introduction

The MEM model was built up by Engle (2002) to model the common nonnegative financial data in practice, such as the mean duration (MD), the absolute returns (AR), the trading volume (VO) and the trading number (TR). The MEM models were first developed as autoregressive conditional duration (ACD, Engle and Russell, 1998) model for featuring the stochastic process of the time between events. Manganelli (2005) applied the MEM model to the nonnegative

*E-mail: xuehai@mail.uni-paderborn.de; Tel.: +49-5251-603225.

expected trading volume series and proposed the autoregressive conditional volume (ACV) model. Besides, the MEM framework was extended to a more general functional form by Box-Cox transformation. The Box-Cox ACD model (Dufour and Engle, 2000; Hautsch, 2002; Fernandes and Grammig, 2006) were proposed as a flexible model to analyze the process of the conditional mean to recent durations based on Box-Cox transformation. If the Box-Cox parameter reduces to zero, the duration can be modeled by the Log-ACD model (Bauwens and Giot, 2000) or the EACD model (Karanasos, 2008) and the log-data follows an ARMA process. Recently, the MEM model is also applied widely. Taylor and Xu (2017) proposed a log-vMEM model, discussing the cross-dependent error terms and the nonnegative conditional mean without any restriction. The multiplicative error model with volatility jumps (MEM-J, Caporin et al., 2017) was developed to investigate the probability and density of the extreme values in the daily volatility. The MEM model is also feasible under the semiparametric framework, such as the SemiGARCH model (Feng, 2004), the Spline-GARCH model (Engle and Rangel, 2008), the general Box-Cox SemiGARCH model (Zhang et al., 2017), etc.

In this paper, we develop a general class of semiparametric MEM model, applying a time-varying scale function with the Box-Cox transformation into the MEM model to analyze the nonnegative financial time series. Nonparametric estimation of the scale function is studied in detail. It is shown that the scale function can be estimated using nonparametric regression based on the Box-Cox transformation of the data and it is closely related to its equivalent scale function, which is obtained in the local linear regression. Note that the difference between the scale function and the equivalent scale function is only a constant parameter, depending on the power transformation of the data under a weak moment condition, while, the parametric model can be also estimated under the weak moment condition after removing the long term trend component. For another, a combined iterative plug-in (IPI, Gasser et al. 1991) algorithm is developed for the bandwidth, the power parameter and the lag-window estimator c_f selection. Following Beran and Feng (2002), the initial bandwidth is selected by exponential inflation method (EIM) as $b_0 = T^{-5/7}$ instead of that with the power minus one of the sample size (Gasser et al., 1991) to minimize the rate of mean square error (MSE). In the c_f estimation, the spectral density at the origin is of great interest and the data-driven algorithm (Bühlmann, 1996; Feng and Gries, 2017; Feng et al., 2019) are employed according to the Bartlett-window, leading to an optimal c_f choice rather than manual input. Meanwhile, the power transformation parameter is also estimated in the data-driven algorithm with various criteria, such as the maximum likelihood estimation (MLE, Box and Cox, 1964), the normality test (Jarque-Bera test, JB; Shapiro-Wilk test, SW) and the quantile-quantile regression (QQr). A block bootstrap method is put forward to the power transformation parameter confidence interval (CI) estimation without sample distribution assumptions due to its dependence. The simulation results show that the absolute values of the power parameter are far smaller than one, satisfying the requirement of the weak moments conditions in real financial markets and the properties of the power parameter depend on the considered sample classes, e.g. for trading volume, the power parameter is about zero, indicating a possible logarithmic data transformation. Finally, the selection of the above parameters are implemented in the same IPI procedure and all the parameters can achieve a convergence value in a few IPI steps. The applications to the nonnegative data samples show that the algorithm is feasible.

The paper is organized as follows. In Section 2, the model is interpreted. Section 3 proposes the semiparametric estimation and the properties of the estimators. The data-driven algorithms are raise in Section 4. Data examples in Section 5 show the proposal works well in real financial markets. Section 6 concludes the paper. The proof of some results is provided in the appendix.

2 The model

2.1 The varying scale MEM model

Let $x_1, ..., x_n$, denote the observed non-negative observations. The MEM model was introduced by Engle (2002) for modeling positive time series. The MEM class is a wide class of models including the ARCH and GARCH family as a special case. Numerous parameterizations for the expected variables are proposed and studied in the literature. In this section, the MEM will be generalized to a semiparametric class by introducing a smooth mean function into the parametric MEM model so that slowly changing dynamics caused by the economic environment can be modeled. This leads to the conditional distribution defined by

$$X_t = m(\tau_t)\psi_t\eta_t, \tag{1}$$

where $\tau_t = t/n$ is the rescaled time, $m(\cdot) > 0$ is a smooth trend, which is the localized unconditional mean function or the scale function in X_t, $\psi_t \geq 0$ denote the conditional mean, $\eta_t \geq 0$ are i.i.d. innovations with unit mean and

$$x_t = \psi_t\eta_t \tag{2}$$

is the descaled stationary process. Throughout this paper, the notation of the condition, i.e. the past information set, $\mathcal{F}_{t-1}$ will be omitted for simplicity. The use of re-scaled time $\tau_t \in [0,1]$ is a standard technique in nonparametric regression with time series errors. Due to $E(x_t|\mathcal{F}_{t-1}) = \psi_t E(\eta_t|\mathcal{F}_{t-1}) = \psi_t$, it is obvious that ψ_t is the conditional mean. To ensure that model (1) is uniquely defined, it is assumed that $E(\psi_t) = 1$. However, it is not necessary and the quantities $E(\eta_t)$ and $E(\psi_t)$ may also be determined by the situation under consideration or by the estimation procedure used. This model will be called a varying scale MEM model (VSMEM), which is also indeed the SemiMEM model. Equation (1) defines indeed a sequence of models. The process $\{X_t\}$ is nonstationary unless $m(\cdot)$ is constant. But it can be shown that the SemiMEM model is locally stationary following Dahlhaus (1997). Our purpose is to develop consistent data-driven estimators of $m(\cdot)$ under certain moment condition $E(X_t^k) < \infty$, where $k > 0$ is a real number, e.g. $k = 2$. Then an approximation of the descaled stationary process x_t can be achieved which can be further studied following some known proposals in the literature. We will see that the use of some suitable power transformation X_t^λ will do the job well, where with $|\lambda| \leq k/4$ is an exponent. Note that X_t^λ belongs to the SemiMEM class.

A closely related class of models is the general semiparametric GARCH framework proposed by Feng (2004). The definition of such a model in the second order sense is given by

$$r_t = s(\tau_t)\sqrt{h_t}\varepsilon_t, \tag{3}$$

where $s^2(\cdot) > 0$ is the variance of r_t, h_t is the conditional variance and ε_t are i.i.d. $N(0,1)$ innovations. The descaled stationary process $\xi_t = \sqrt{h_t}\varepsilon_t$ stands for an ARCH type model. It is assumed that $E(h_t) = 1$ to ensure that the model is well defined. Here $s^2(\cdot)$ is the scale function in r_t^2 and $s(\cdot)$ is the scale function in r_t. Due to the assumptions $E(\varepsilon_t^2) = 1$ and $E(h_t) = 1$, we have $E(r_t^2) = s^2(\tau_t)$. But $s(\tau_t)$ is not the mean of $|r_t|$. We will see that the difference between $s(\tau_t)$ and $E(|r_t|)$ is a constant factor depending on the distribution of ξ_t. Model (3) will be called a varying scale GARCH model (VSGARCH, also SemiGARCH). A nonparametric trend in the mean or a parametric regression for the mean based on some exogenous variables can also be included in (3). Assume again that $E(r_t^k) < \infty$ with e.g. $k = 4$. The scale function $s(\cdot)$, up to a constant factor, can be estimated consistently from $|r_t|^\lambda$ by some data-driven nonparametric regression algorithm, provided $0 < \lambda \leq k/4$ is used. Again, $|r_t|^\lambda$ also belongs to the SemiMEM class.

We see, $m(\cdot)$ in (1) and $s(\cdot)$ in (3) can be estimated in the same way. And the conditional mean ψ_t in (1) can also be modelled following the idea of ARCH and GARCH models for the conditional variance h_t in (3) (Engle, 2002). But the SemiGARCH and SemiMEM classes do not coincide with each other. On the one hand, the signs of the observations of a SemiGARCH process may play an important role theoretically and in practice. Hence an original SemiGARCH process is not a member of the SemiMEM class. On the other hand, if the original process is non-negative, e.g. a duration process, it is of course not a member of the SemiGARCH class.

The descaled x_t in (2) process can be modelled following any suitable parameterization. Different parameterizations will lead to different semiparametric models. For instance, if X_t is a duration process, ψ_t can then be modelled by the well known ACD model introduced by Engle and Russell (1998) with

$$\psi_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j \psi_{t-j}, \tag{4}$$

where p and q are the orders, $\alpha_0 > 0$, $\alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q \geq 0$ are unknown parameters such that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. Equations (1) and (4) together define a semiparametric varying scale ACD model (VSACD, also SemiACD).

The overall mean in the SemiACD model can be thought of as a weighted sum of the unconditional local mean $m(\tau_t)$, the last q conditional means and the last p observations, which reflect long run, middle term and short term dynamics in the mean of X_t. Their weights are $1 - \sum \alpha_i - \sum \beta_j$; $\beta_1, ..., \beta_q$; and $\alpha_1, ..., \alpha_p$, respectively. Due to the restriction of the conditional mean, the scale parameter, α_0, is no more a free parameter, because it holds $\alpha_0 = 1 - \sum \alpha_i - \sum \beta_j$. That is α_0 in the conditional mean of the resclaed process is itself the weight for the unconditional local mean. The difference between the ACD and the SemiACD is just that the unconditional mean in the former is a constant but it is a smooth nonparametric function in the latter. Moreover, let $m_t^* = m(\tau_t)\psi_t$, we have

$$m_t^* = \alpha_0(\tau_t) + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j m_{t-j}^* + O(n^{-1}), \tag{5}$$

where $\alpha_0(\tau_t) = \alpha_0 m(\tau_t)$ is a time-varying parameter. We see, the SemiACD model is approximately an ACD model with a time-varying scale parameter.

In the remaining part of the paper, only the estimation of $m(\cdot)$ in (1) and the estimation of the unknown parameters in (4) based on the descaled data will be discussed in detail and the estimation of $s(\cdot)$ in (3) and of the unknown parametric parameters can be done following the same procedure. The scale function $m(\cdot)$ will be first estimated without using any further parametric assumption. It allows further parameter estimation under different, not only the ACD, specifications.

2.2 The Box-Cox SemiMEM model

The discussion of the scale function $m(\cdot)$ and the conditional mean ψ_t in the SemiMEM model are always considered. In the section, we study the estimation of the slow scale change based on the Box-Cox power transformation x_t^λ for some $|\lambda| < 1$. Here a nonparametric estimator of the scale function in x_t will be estimated first. Then the back-transformed estimator will be used. We will see that the resulting estimator is usually not a consistent estimator of $m(\cdot)$ but of some equivalent scale function.

In this section, any function of the form $\tilde{m}(\tau_t) = C \cdot m(\tau_t)$ with $C > 0$ will be called an equivalent scale function. Because $\tilde{x}_t = X_t/\tilde{m}(\tau_t)$, then $\tilde{x}_t = C^{-1}x_t$ is a stationary process having the same properties as x_t but with a different scale parameter. Hence the resulting estimator based on x_t can be used to remove the effect of the slowly changing scale in X_t. There are different further transformations, which can be used for estimating an equivalent scale function. The power transformations (or equivalently the Box-Cox transformations with $|\lambda| < 1$) are just the most simple examples. For a more general description on this point, we refer the reader to Eagleson and Müller (1997).

We see, the development of a consistent data-driven estimator is always possible based some power transformation. If it is assumed that $E(X_t) < \infty$, the maximal allowed λ is 1/4. For conducting maximum likelihood estimators of the unknown parameters, the condition $E(X_t^2) < \infty$ was indicated by Lee and Hansen (1994) and Engle and Russell (1998). Now, $|\lambda| \leq 1/2$ can be used. Assume that $E(X_t^k) < \infty$. Let $\delta = \min(k/4, 1)$ for some $|\lambda| \leq \delta$. We have the Box-Cox SemiMEM model as

$$X_t = m_\lambda(\tau_t)x_{t,\lambda}, \tag{6}$$

where $x_{t,\lambda}$ is stationary with $E(x_{t,\lambda}^\lambda) = 1$. If $x_{t,\lambda}$ satisfies a specific parameterization, e.g. an ACD process, model (6) becomes a Box-Cox SemiACD (or a Box-Cox VSACD) process. Obviously, the stationary process $x_{t,\lambda} = c_\lambda^{-1/\lambda}x_t$ and the equivalent scale function $\tilde{m}(\tau_t) = m_\lambda(\tau_t) = C_\lambda^{1/\lambda} \cdot m(\tau_t)$. The value of C_λ is determined by λ and the marginal distribution of x_t. Under the assumption $E(x_t) = 1$ as used in (2) we have $C_1 \equiv 1$. This shows again why commonly proposed estimators of $m(\cdot)$ is based on x_t. However, model (1) can also be defined according to the λ_0-th moments of the process for some $|\lambda_0| \leq k/4$. That is, we can assume that $E(x_t^{\lambda_0}) = 1$ and $E(\psi_t^{\lambda_0}) = 1$. This implies that $E(X_t^{\lambda_0}) =: m_{\lambda_0}(\tau_t)$ is the scale function in $X_t^{\lambda_0}$. Under this definition we have $C_{\lambda_0} \equiv 1$. The Box-Cox SemiACD model in the section has a close relation with the parametric Box-Cox ACD model, i.e. if the power parameter is nonzero, a varying scale Box-Cox ACD is considered and if the power parameter is zero, then the model reduces to a logarithmic form.

3 The model estimation and properties

3.1 Estimation of $m_\lambda(\tau_t)$

In Efromovich (1999), the scale function can be estimated by a general nonparametric regression process. Following the proposal, Equation (6) can be written as a nonparametric regression model

$$X_t^\lambda = g_\lambda(\tau_t) + g_\lambda(\tau_t)\zeta_{t,\lambda}, \tag{7}$$

where $g_\lambda(\tau_t) = c_\lambda m^\lambda(\tau_t) = [m_\lambda(\tau_t)]^\lambda$ and $\zeta_{t,\lambda} = x_{t,\lambda}^\lambda - 1$ with $E(\zeta_{t,\lambda}) = 0$ and $\operatorname{var}(\zeta_{t,\lambda}) = 1$. Hence $g_\lambda(\cdot)$ can be estimated for instance by kernel (Feng and Heiler 1998 and Feng, 2004) or local liner regression (Fan and Yao, 1998).

Since by definition the observations x_t are non-negative, kernel estimates of $g_\lambda(\tau_t)$ based on non-negative kernels are always positive. However, at boundary points (i.e. $\tau_t < b$ or $\tau_t > 1 - b$), the rate of convergence of kernel estimators is lower than in the interior unless boundary kernels are used. Alternatively, one may use local linear estimates which have the same rate of convergence for all $\tau_t \in [0, 1]$. However, boundary kernels or local linear regression lead to estimates that may be negative at boundary points. We therefore propose to use the estimator $\hat{g}_\lambda(\tau_t) = |\tilde{g}_\lambda(\tau_t)|$ where $\tilde{g}_\lambda(\tau_t) = \hat{a}_0(x)$ is a local linear estimate obtained by minimizing

$$Q(a_0, a_1) = \sum_{t=1}^{T} \left\{X_t^\lambda - a_0(\tau) - a_1(\tau)(\tau_t - \tau)\right\}^2 K\left(\frac{\tau_t - \tau}{b}\right) \tag{8}$$

with K being a symmetric non-negative kernel function and $0 < b < \frac{1}{2}$ the bandwidth. Using $\hat{g}_\lambda(\tau_t)$ instead of $\tilde{g}_\lambda(\tau_t)$ can be justified as follows. First, note that $\hat{g}_\lambda(\tau_t) = \tilde{g}_\lambda(\tau_t)$, if $\tilde{g}_\lambda(\tau_t) \geq 0$. If $\tilde{g}_\lambda(\tau_t) < 0$, we have $|\tilde{g}_\lambda(\tau_t) - g_\lambda(\tau_t)| = g_\lambda(\tau_t) + |\tilde{g}_\lambda(\tau_t)|$ and $|\hat{g}_\lambda(\tau_t) - g_\lambda(\tau_t)| = ||\tilde{g}_\lambda(\tau_t)| - g_\lambda(\tau_t)| < g_\lambda(\tau_t) + |\tilde{g}_\lambda(\tau_t)|$. Hence $E[(\hat{g}_\lambda(\tau_t) - g_\lambda(\tau_t))^2] \leq E[(\tilde{g}_\lambda(\tau_t) - g_\lambda(\tau_t))^2]$. That is the MSE of $\hat{g}_\lambda(\tau_t)$

is no larger than that of $\tilde{g}_\lambda(\tau_t)$. Moreover, the possibility of negative values is a finite sample problem and can only occur at boundary points, because in the interior the weights of a local linear estimate are exactly equal to the kernel weights and are hence non-negative. Finally, even at a boundary point, $\hat{g}_\lambda(\tau_t)$ and $\tilde{g}_\lambda(\tau_t)$ coincide with asymptotic probability one, as shown in Zhang et al. (2017).

Let that $\hat{m}_\lambda(\cdot)$ be a consistent estimator of $m_\lambda(\cdot)$, we can obtain $\hat{x}^*_{t,\lambda} = x_{t,\lambda}/[\hat{m}_\lambda(\tau_t)]^{1/\lambda}$ is an approximation of the stationary process $x^*_{t,\lambda}$. Define $Z_{t,\lambda} = X_t^\lambda$, then $\hat{z}^*_{t,\lambda} = Z_{t,\lambda}/\hat{m}_\lambda(\tau_t)$ is its approximate estimation of $x_{t,\lambda}$ with Box-Cox transformation. A suitable parametric model can be fitted to either $\hat{x}^*_{t,\lambda}$ or $\hat{z}^*_{t,\lambda}$.

Assume that the moments condition $E(X_t^k) < \infty$ and $E(x_t^k) = 1$ hold, for $k \neq 0$. The scale function $m(\tau_t)$ in the Box-Cox SemiMEM model is estimated by its equivalent scale $g_\lambda(\tau_t)$ as $\hat{m}(\tau_t) = \hat{C}_\lambda^{-1/\lambda} \cdot \hat{g}_\lambda(\tau_t)$, where $\hat{C}_\lambda^{-1/\lambda} = [\frac{1}{n}\sum_{t=1}^{n} \hat{x}^k_{t,\lambda}]^{1/k}$. Zhang et al. (2017) discussed the estimation under a Box-Cox SemiGARCH framework and the constant parameter of the equivalent scale function is regarded related to the mean of the stationary second order stationary process. Following the definition of Box-Cox SemiMEM model, for $\lambda \neq 0$, we can conclude that the scale function is estimated by its equivalent scale function as $\hat{m}_\lambda(\tau_t) = c_\lambda^{1/\lambda}\hat{m}(\tau_t)$. Then,

$$\begin{aligned} X_t &= m_\lambda(\tau_t) \cdot x_{t,\lambda} \\ &= c_\lambda^{1/\lambda} \cdot m(\tau_t) \cdot x_{t,\lambda} \\ &= m(\tau_t) \cdot x_t, \end{aligned} \tag{9}$$

it is obvious that $x_{t,\lambda} = c_\lambda^{-1/\lambda} x_t$. Under the assumptions of Proposition, if $E(x_t^k) = 1$ and $k \neq 0$ hold, the power k-th expectation of the transformed non-negative variable is

$$\begin{aligned} E(x^k_{t,\lambda}) &= E[(c_\lambda^{-1/\lambda} \cdot x_t)^k] \\ &= c_\lambda^{-k/\lambda} \cdot E(x_t^k) \\ &= c_\lambda^{-k/\lambda}. \end{aligned} \tag{10}$$

Also, the constant for the equivalent scale function is

$$c_\lambda^{-1/\lambda} = [E(x^k_{t,\lambda})]^{1/k}. \tag{11}$$

Obviously, if the stationary process follows a GARCH process and a second order transformation, the constant of the equivalent scale function is

$$c_\lambda^{-1/\lambda} = \sqrt{E(\hat{x}_{t,\lambda}^2)}$$
$$= \sqrt{\frac{1}{n}\sum_{t=1}^{n}\hat{x}_{t,\lambda}^2}, \tag{12}$$

and for the ACD model, under the assumption $E(x_t) = 1$, the constant reads as

$$c_\lambda^{-1/\lambda} = E(\hat{x}_{t,\lambda})$$
$$= \frac{1}{n}\sum_{t=1}^{n}\hat{x}_{t,\lambda}, \tag{13}$$

which is indeed the mean of the stationary process.

3.2 Properties of $\hat{g}_\lambda(\tau_t)$

The asymptotic properties of $\hat{g}_\lambda(\tau_t)$ are similar to those of a local linear estimator of the mean function in the presence of heteroskedastic time series errors and are closely related to known results on nonparametric regression with dependent errors (see e.g. Altman, 1990, Hart, 1991 and Beran and Feng, 2002). The results given in this section do not depend on any parametric specification of $x_{t,\lambda}$.

Note that a local linear estimator at point x generates an equivalent kernel $K_\tau(u)$ which is the same as $K(u)$ for $b \leq \tau_t \leq 1-b$ and equal to a boundary kernel at boundary points. To reduce the variance, we will use a varying bandwidth b_τ at boundary points, such that the length of the window is always $2b$. For kernel functions K and K_τ, define $R(K) = \int K^2(u)du$, $I(K) = \int u^2K(u)du$, $R(K_\tau) = \int K_\tau^2(u)du$ and $I(K_\tau) = \int u^2K_\tau(u)du$. Furthermore, let $\gamma(k) = cov(x_{t,\lambda}, x_{t+k,\lambda})$. Assume that $\gamma(k)$ are absolutely summable and c_f denotes the value of the spectral density of $x_{t,\lambda}$ at the origin $\lambda = 0$. Under the assumptions A1 through A5 stated in the appendix the following holds, as $T \to \infty$, $b \to 0$ and $nT \to \infty$:

The bias and variance of $\hat{g}_\lambda(\tau_t)$ are

$$B[\hat{g}_\lambda(\tau_t)] = E(\hat{g}_\lambda(\tau_t)) - g_\lambda(\tau_t) = \frac{1}{2}b_\tau^2 g_\lambda''(\tau_t)I(K_\tau) + o(b_\tau^2), \tag{14}$$

and

$$var(\hat{g}_\lambda(\tau_t)) = \frac{2\pi c_f g_\lambda^2(\tau_t)R(K_\tau)}{b_\tau T} + o\left(\frac{1}{b_\tau T}\right) = \frac{V}{b_\tau T} + o\left(\frac{1}{b_\tau T}\right), \tag{15}$$

where $V = 2\pi c_f g_\lambda^2(\tau_t)R(K_\tau)/(b_\tau T)$.

Based on the bias and variance of the estimated equivalent scale function, the mean integrated squared error of $\hat{g}_\lambda$, MISE($\hat{g}_\lambda$)=$\int_0^1 \{\hat{g}_\lambda(\tau_t) - g_\lambda(\tau_t)\}^2 d\tau$, is

$$\text{MISE}(\hat{g}_\lambda) = b^4 \frac{\int [g_\lambda''(\tau)]^2 d\tau I^2(K_\tau)}{4} + \frac{2\pi c_f \int g_\lambda^2(\tau) d\tau R(K_\tau)}{bT} + \max\left\{o(b^4), o\left(\frac{1}{bT}\right)\right\}.$$

Then, the asymptotically optimal bandwidth, which minimizes the dominating part of the MISE is given by

$$b_A = \left(2\pi c_f \frac{R(K_\tau)}{I^2(K_\tau)} \frac{\int g_\lambda^2(\tau)d\tau}{\int [g_\lambda''(\tau)]^2 d\tau}\right)^{1/5} T^{-1/5} = \left(2\pi c_f \frac{R(K_\tau)}{I^2(K_\tau)} \frac{I(g_\lambda^2)}{I[(g_\lambda'')^2]}\right)^{1/5} T^{-1/5}, \tag{16}$$

where $I(g_\lambda^2) = \int g_\lambda^2(\tau)d\tau$ and $I([g_\lambda'']^2) = \int [g_\lambda''(\tau)]^2 d\tau$.

Further, if a bandwidth $b = o(b_A)$ is used, the bias is asymptotically negligible and

$$\sqrt{Tb}[\hat{g}_\lambda(\tau_t) - g_\lambda(\tau_t)] \xrightarrow{\mathcal{D}} N(0, V), \tag{17}$$

where V is as defined in (15).

Note that both the kernel $K_\tau(u)$ and the bandwidth b_τ in the bias and variance of $\hat{g}(\tau_t)$ depend on τ_t. But the effect of boundary points on the MISE of $\hat{g_\lambda}$ is asymptotically negligible. Thus, the asymptotically optimal global bandwidth can be calculated using the MISE over the whole interval $\tau_t \in [0, 1]$. Eq. (16) provides the basis for developing a plug-in bandwidth selector. The difference between the formula of the optimal bandwidth b_A here, compared to nonparametric regression with i.i.d. errors is that the two unknown constants c_f and $I(g_\lambda)$ are different. Here, the factor c_f measures the effect of the stationary time series errors on b_A. The constant $I(g_\lambda)$ is determined by the (deterministic) heteroskedasticity characterized by the scale function in (7). The result of Eq. (17) shows that $\hat{g}_\lambda$ is asymptotically unbiased and asymptotically normal if a bandwidth of a smaller order than b_A is used.

4 The data-driven algorithms

4.1 The bandwidth selection

In this section, we develop a data-adaptive bandwidth selector for the SemiACD model following the iterative plug-in method (IPI, see Gasser et al. 1991). For simplicity, global bandwidth selection will be considered. The plug-in bandwidth selector is based on an iteration algorithm where estimates of c_f, $I(g_\lambda^2)$ and $I([g_\lambda'']^2)$ are plugged into (16). The key point here is the estimation of $I([g_\lambda'']^2)$, because the estimation of c_f and $I(g_\lambda^2)$ are relatively easy. The IPI method has been extended successfully to nonparametric regression with time series errors (see e.g. Herrmann et al. 1992, Brockmann et al. 1993, Beran and Feng 2002 and Ghosh and Draghicescu 2002). We will, therefore, use this approach here.

Let b_{j-1} denote the bandwidth in the $(j-1)$th iteration. The IPI algorithm estimates $I([g_\lambda'']^2)$ in the jth iteration by $\hat{I}_j([g_\lambda'']^2) = T^{-1}\sum_{t=1}^{T}[\hat{g}_{\lambda_j}''(\tau_t)]^2$, where $\hat{g}_{\lambda_j}''(\tau_t)$ is estimated using a bandwidth b_{2j} obtained from b_{j-1} by a so-called inflation method. The word "inflation" comes from the fact that b_{2j} is much larger than that the bandwidth for estimating m itself. Once the estimate $\hat{I}_j([g_\lambda'']^2)$ is calculated, it is inserted into (16) to calculate b_j in the jth iteration. This procedure is repeated until the selected bandwidth converges.

Two choices have to be made to define the algorithm. One needs to fix an initial bandwidth b_0, and a concrete inflation method has to be specified. Gasser et al. (1991) propose to set $b_0 = T^{-1}$. However, this bandwidth cannot be used for estimating $I(m)$ in the first iteration. Beran and Feng (2002) therefore suggest $b_0 = T^{-5/7}$ which is a very small bandwidth but satisfies the assumptions in Theorem 6.1. As it turns out, the final bandwidth is not sensitive to the choice of b_0. However, the number of required iterations depends heavily on b_0. A data-driven possibility that reduces the number of iterations is $b_0 = \hat{b}_{\mathrm{CV}}$ where $\hat{b}_{\mathrm{CV}}$ is the bandwidth selected by the cross-validation criterion (Wahba and Wold 1975).

With respect to the inflation method, the so-called exponential inflation method proposed (EIM) in Beran and Feng (2002) with $b_{2j} = (b_{j-1})^\alpha$ (where $0 < \alpha < 1$) is well suited in the current context. We propose to choose $\alpha = 5/7$ which means that the rate of the MSE of $\hat{I}([g_\lambda'']^2)$ is minimized. Note that, in comparison, the original multiplicative inflation method (MIM) of Gasser et al. (1991) defined by $b_{2j} = b_{j-1}T^{1/10}$ does not work well in the current context. The main reason is that the rate of convergence of the estimated bandwidth is $O(T^{-1/5})$ which is clearly lower than the rate of convergence $O(T^{-2/7})$ achieved by the EIM method (see the theorem below). An additional problem with the MIM is that for large sample sizes T the inflation factor is too strong which often leads to poor final bandwidth. In each iteration, $I(g_\lambda^2)$ can be estimated from $\hat{g}_\lambda$ which is based on the previous bandwidth b_{j-1}. Since $\hat{b}_{\mathrm{CV}}$ is already a consistent estimator of b_A, using $b_0 = \hat{b}_{\mathrm{CV}}$ is likely to provide a good initial estimate of $I(g_\lambda^2)$.

The asymptotic performance of $\hat{b}_A$ is characterized by the following theorem if the following assumptions are satisfied.

A1. The scale function $g_\lambda(\tau_t)$ is strictly positive, bounded, and at least twice continuously differentiable on [0, 1].

A2. The kernel $K(u)$ is a symmetric density with compact support [-1, 1].

A3. $\{x_t\}$ is a stationary process with unit mean and absolutely summable autocovariance.

Theorem 1. *Under Assumptions A1 to A3 and the additional assumption that* $E(x_t^4) < \infty$, *we have*

$$\hat{b}_A = b_A[1 + O_p(T^{-2/7}) + O(T^{-1/3})]. \tag{18}$$

A sketched proof of Theorem 6.1 is given in the appendix. The $O_p(T^{-2/7})$ term in (18) is caused by the error in $\hat{I}([g''_\lambda]^2)$ and whereas the $O(T^{-1/3})$ term is due to the error in $\hat{c}_f$. If the parametric specification in (2) and (4) is used, then the (relative) effect of the error in $\hat{c}_f$ will be of the order $O(T^{-2/5})$ (see e.g. Feng, 2004), which is smaller than $O(T^{-1/3})$. However, the $O(T^{-1/3})$ term is still asymptotically negligible compared to $O_p(T^{-2/7})$. Hence the asymptotic properties and in particular the rate of convergence of $\hat{b}_A$ are determined by those of $\hat{I}([g''_\lambda]^2)$. Note in particular that, although the bandwidth selection problem under model (7) is more complex, the rate of convergence of the selected bandwidth is the same as for the DPI (direct plug-in) bandwidth selector proposed by Ruppert et al. (1997) in the context of nonparametric regression with i.i.d. errors. Furthermore, b_A is not well defined, if $g_\lambda(\tau_t) \equiv g_0$, because $I([g''_\lambda]^2) = 0$. Nevertheless, the SemiACD model and the proposed algorithm are still applicable in this case. For instance, if y_t follows an MEM model, it can be shown that b_j converges to a nonzero constant as $j \to \infty$. Besides, it can be shown that $\hat{\theta}$ (obtained from $\hat{x}_t$) has the same asymptotic properties as $\tilde{\theta}$ (obtained from the MEM observations x_t), because $\hat{b} >> O_p(T^{-1/2})$. Moreover, suppose that no maximal number of iterations is fixed. Then $(Tb_j)^{-1}$ is asymptotically of the order T^{-1} though $b_j < 1$. Therefore $\hat{g}_\lambda(\tau_t)$ is $\sqrt{T}$-consistent, with some loss in efficiency compared to a parametric estimator.

4.2 The c_f estimation

The remaining unknown quantity c_f can be estimated using any nonparametric estimator of the spectral density. The spectral density of $x_{t,\lambda}$ is

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_j(k) e^{ik\lambda}, -\pi \leq \lambda \leq \pi. \tag{19}$$

If $\lambda = 0$, the same as we discussed in Section 4.1, we can obtain that $c_f = f(0)$. In the following, we will use the lag-window estimator with the Bartlett-window (see e.g. Priestley, 1981)

$$\hat{c}_f^j = \frac{1}{2\pi} \sum_{k=-K}^{K} w_k \hat{\gamma}_j(k) \tag{20}$$

to estimate the c_f. In the formula, $\hat{\gamma}_j(k)$ denotes the sample autocovariance at lag k calculated from the residuals in the $(j-1)$th iteration, $w_k = 1 - k/(K+1)$ and $K << T$ is the bandwidth of the lag-window, included in $\hat{c}_f$.

The optimal choice of K is of the order $O(T^{1/3})$. To focus on the main results, we do not address the issue of data-driven algorithms for selecting K. In the applications we will use $K = [c_f T^{1/3}]$ with c_f selected in the same IPI algorithm. The proposed IPI data-driven algorithm, including the bandwidth selection with power transformation and the c_f estimation, reads as follows:

1. Select b_0 using the CV criterion ignoring correlations and changes in the scale. Set $j = 1$.
2. Select bandwidth correlations and changes in the scale, using J_1 IPI-iterations.
3. In the j-th iteration for $j \geq J_1$ carry out the following calculations:
    a) Estimate $\hat{g}_\lambda(\tau_t)$ with b_{j-1} and let $\hat{x}_{t,\lambda} = X_t/[\hat{g}_\lambda(\tau_t)]^{1/\lambda}$.
    b) Estimate $\hat{c}_f$ with the Bartlett-lag-window estimator.
        i) Set the starting Bartlett-window as $M_0 = [n/2]$, where $[\cdot]$ is the integer part.
        ii) Global estimation. Following Bühlmann (1996), estimate the integration of the first derivative $\int f^{(1)}(\lambda)d\lambda$ with the Bartlett-window width $K'_j = K_{j-1}/n^{2/21}$. Insert the estimates into the optimal global window width equation (Bühlmann, 1996, Eq. 5), then calculate K_j. Increase j by one and repeat the above procedures until the selected K converges or reach the maximal 20 iterations. Denote the selected optimal global window width as K_{Gl}.

iii) Local estimation. Calculate $\int f^{(1)}(\lambda)d\lambda$ with the optimal local window width $K_{Lo} = K_{Gl}/n^{2/21}$ and insert the estimates into optimal local window width equation (Bühlmann, 1996) with $\lambda = 0$

$$K_{opt} = n^{-1/3}(\frac{\int_{-\pi}^{\pi}(f(\lambda)^2)d\lambda}{3\int_{-\pi}^{\pi} f^{(1)}(\lambda)^2 d\lambda})^{1/3}. \tag{21}$$

The finally selected Bartlett-window width is denoted as $\hat{K} = K_{opt}$.

c) Denote by $\hat{I}(g_\lambda)$ the estimate of $I(g_\lambda)$ obtained using b_{j-1} and by $\hat{I}(g''_\lambda)$ the estimate of $I(g''_\lambda)$ based on bandwidth $b_{2j} = b_{j-1}^{5/7}$. Consider the power transformation parameter in the range $[-1, 1]$ and let the power transformation parameter increase interval be 0.01. In the k-th iteration with $k > 1$, estimate the trend $\hat{g}_{\lambda_k}(\tau_t)$ using $\hat{\lambda}_{k-1}$. Remove the trend and obtain the optimal transformation parameter λ_k by the MLE, JB, SW and QQr criteria. Increase k by one and repeat the previous steps until reach the convergence or the maximal number of iterations.

d) Improve b_{j-1} by

$$b_j = \left(2\pi\hat{c}_f \frac{R(K)}{I^2(K)} \frac{\hat{I}(g_\lambda)}{\hat{I}(g''_\lambda)}\right)^{1/5} T^{-1/5}. \tag{22}$$

e) Increase j by one and repeatedly carry out a) to c) until convergence or a given maximal number of iterations has been reached.

The finally selected bandwidth $\hat{b}_A$ is obtained in the last iteration of Step 3. In the IPI procedures of the bandwidth, c_f and the power parameter estimation, we find that if the selected statistics converge, the values usually are not affected by the initial inputs. Hence, the proposed IPI algorithm can be carried out without starting restrictions. In the section for simplicity, the Bartlett-window is applied in the c_f selection algorithm. Besides, Bühlmann (1996) also discussed a C^2-window for the optimal window width. Finally, Feng and Gries (2017) introduced that the estimation quality of c_f can be improved if the optimal bandwidth for calculating the $x_{t,\lambda}$ is considered and vice versa.

4.3 The confidence interval simulation of λ

In the previous section, we obtain the stationary time series $x_{1,\lambda}, \ldots, x_{t,\lambda}$ by removing the time-varying scale function. For estimating the power transformation parameter λ, we propose to construct an estimator $\hat{\lambda}$ based on the descaled time series via the MLE, JB, SW and QQr criteria. The confidence interval of λ at a given confidence level with MLE has been well discussed. However, the confidence intervals depending on the other criteria, are unknown in practice and often very complicated. Bootstrap methods provide a general solution for estimating the confidence interval of λ without the consideration of the time series model assumptions.

The bootstrap was published by Efron (1979) for i.i.d. random variables as a simulation method using the resampling technique to estimate the variables distributions and the statistic inference, such as the bias, the variance, the confidence intervals, the reject probability in a hypothesis test, etc. The advantage of the bootstrap method is that the simulation algorithm requires no distribution or parametric assumption of the under analysis data set or process. The bootstrap provides information about the whole sampling distribution and performs computational efficient rather than the other resampling technique– jackknife (Tukey, 1958), recognized as the 'leave-one-out' method.

For the time series, however, the dependent and correlated data are not suitable to directly apply the bootstrap algorithm. As discussed by Künsch (1989), Hall (1992), Liu and Singh (1992) and Politis and Romano (1993), for the dependent time series data, the block bootstrap method is the potential solution to estimate the unknown distribution by dividing the data into several blocks, to hold the original time series dependence structure within blocks. Because the asymptotic properties of the estimator may be affected by the block selection, i.e. the dependence of the resampling time series is always regarding to the randomly selected blocks. Thus, a modification bootstrap– moving block bootstrap (MBB) was proposed by selecting the optimal block length, also recognized as the overlapping block bootstrap, which preserves the data structure of the original series in each formed block. In the MBB, the length of the block is

$$l = n/N_{bl}, \tag{23}$$

where n is the length of the original time series and N_{bl} is the number of the resampling blocks and according to the MBB, the independence of the l subsampling is for sure. If dependent data is considered, then an unnecessary requirement is

$$l \to \infty \text{ and } N_{bl}^{-1} l \to 0 \text{ as } N_{bl} \to \infty. \tag{24}$$

Obviously, the selection of the block length is the most concerned problem in the block bootstrap

(Efron and Tibshirani, 1993). As indicated by Bühlmann and Künsch (1999) and Bühlmann (2002), they proposed the bootstrap variance is equivalent to a lag weight estimator of the spectral density of the bootstrapped variable's influence function (IF, Hampel et al., 1986) at the interesting origin and the block length is obvious obtained as inverse of the bandwidth. Furthermore, they also developed a data-driven algorithm, suggesting a two-step procedure selecting the optimal block length in the blockwise bootstrap and in this algorithm, the IF is estimated first, then optimal block length is obtained as the estimated value of the lag weight spectral density with certain lag window (e.g. Bartlett window) at frequency zero. Following the algorithm, the optimal block length is approximate to the cubic root of the sample size. Politis and White (2004) discussed a data-based block length selection algorithm and found that the optimal block length order for the stationary bootstrap is also one third.

Besides, we still consider a model-based bootstrap method (such as autoregressive bootstrap (ARB), Efron, 1979), by removing the nonstationary trend and building up the i.i.d. error terms under the MEM model assumption. Due to the model-based bootstrap depending on not only the parameters in models but also the identified structure with the original data. Obviously, the model-based resampling is greatly affected by both the model and its structure and the asymptotic properties of original data can not be correctly revealed if the model misspecification occurs, leading to the inconsistent between the built-up i.i.d data and the original ones. In the section, the MBB methodology works very well in practice, however, the model-based idea seems not.

There are several procedures to calculate the confidence intervals of the resampling data sets, such as the percentile method, the bootstrap-t method, the bias-corrected (BC) method, the bias-corrected and accelerated (BCa) method, the approximate bootstrap confidence (ABC) method and calibration. For simplicity, we consider only the ordinary percentile method.

In the percentile approach, we consider the $\alpha/2$ and $(1-\alpha)/2$ percentiles of the bootstrap distribution of $\hat{\lambda}$. Let the number of the bootstrap replications of $\hat{\lambda}$ be N_B. Then, the order statistic of λ

$$\hat{\lambda}_1^* \leq \cdots \leq \hat{\lambda}_{k_L}^* \leq \cdots \leq \hat{\lambda}_{k_U}^* \leq \cdots \leq \hat{\lambda}_{N_B}^*, \tag{25}$$

and the bootstrap confidence interval (BCI) at the $100(1-\alpha)\%$ is

$$\lambda \in [\hat{\lambda}_{k_L}^*, \hat{\lambda}_{k_U}^*], \tag{26}$$

where $k_L = [\frac{\alpha}{2}(N_B+1)]$ and $k_U = [(N_B+1) - k_L]$

The percentile method gets rid of the assumption limitation of the bootstrap variable in theory, however, if the resampling replication number is too small, the simulation may not perform well, requiring the percentiles corrections, new resampling histogram assumptions, bias-correction factors and acceleration parameters.

5 The empirical examples

The selection of the $\hat{\lambda}$ is an important issue regarding to the model selection in the risk management. In the section, we apply the proposal to the nonnegative transaction data (MD, AR, VO and TR) of Siemens (SIE) and Deutsche Bank (DBK) from 2000 to 2013. The MD, VO and TR data are organized as equidistant daily data from the intraday high-frequency transaction records, meanwhile, the AR data is calculated as the absolute returns of the daily stock close price.

From Fig. 1 to Fig. 8, the histograms of the selected $\hat{\lambda}$ with the MLE, JB, SW and QQr are displayed respectively. Because the sample sizes of the considered data sets are about 3500, we apply the length of 16 (the approximate cubic root of the size of the observation) as the optimal block length in the bootstrap procedures.

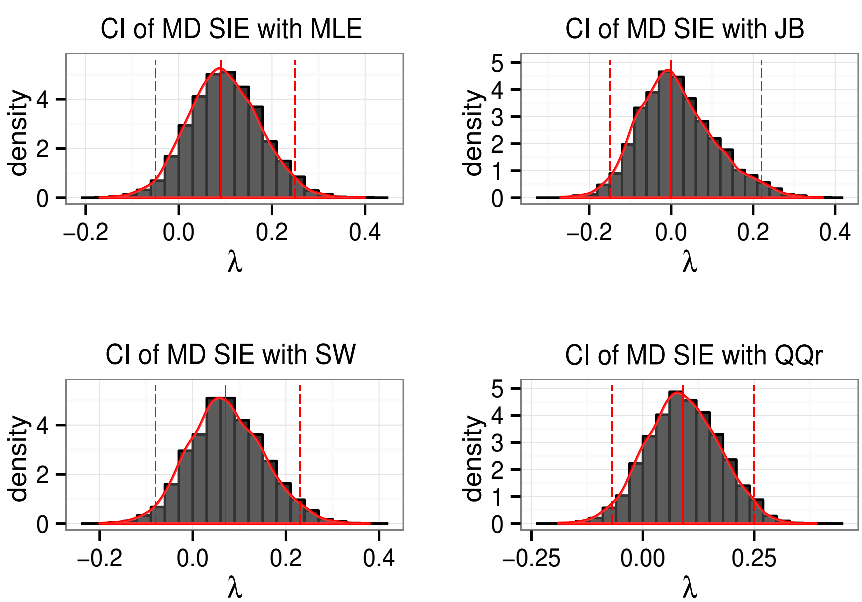

Figure 1: Simulated λ confidence interval of the SIE's mean duration

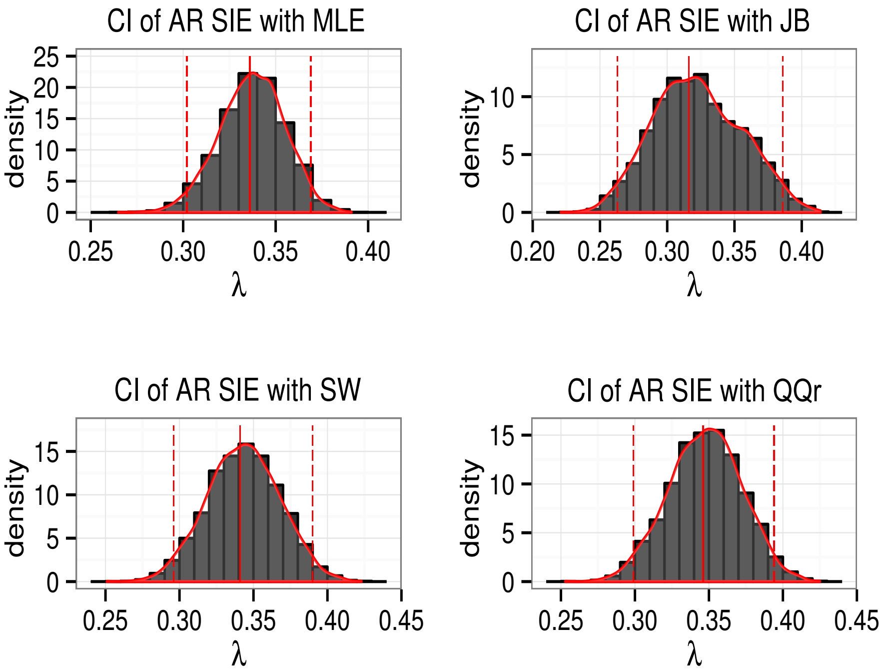

Figure 2: Simulated λ confidence interval of the SIE's absolute returns

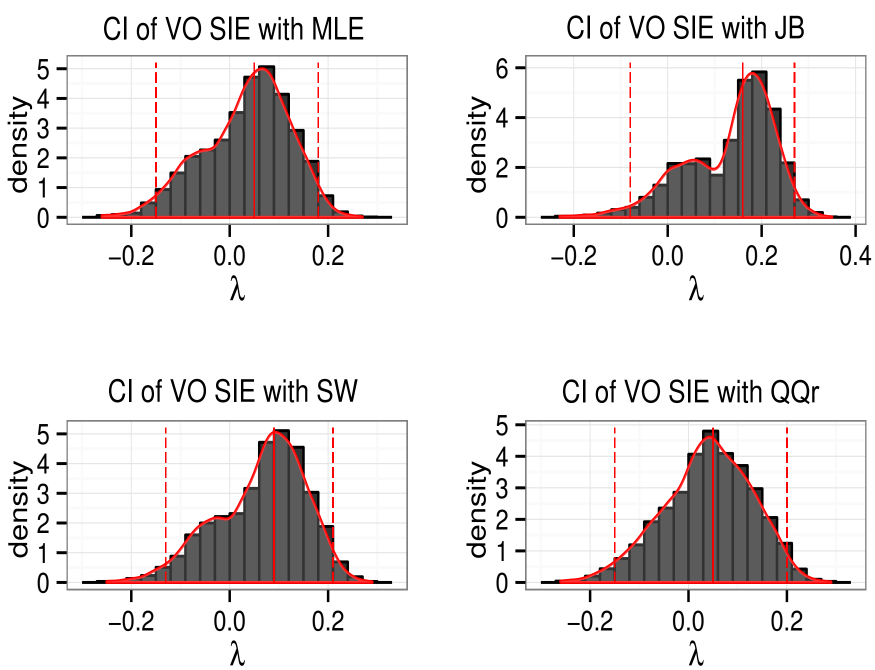

Figure 3: Simulated λ confidence interval of the SIE's trading volume

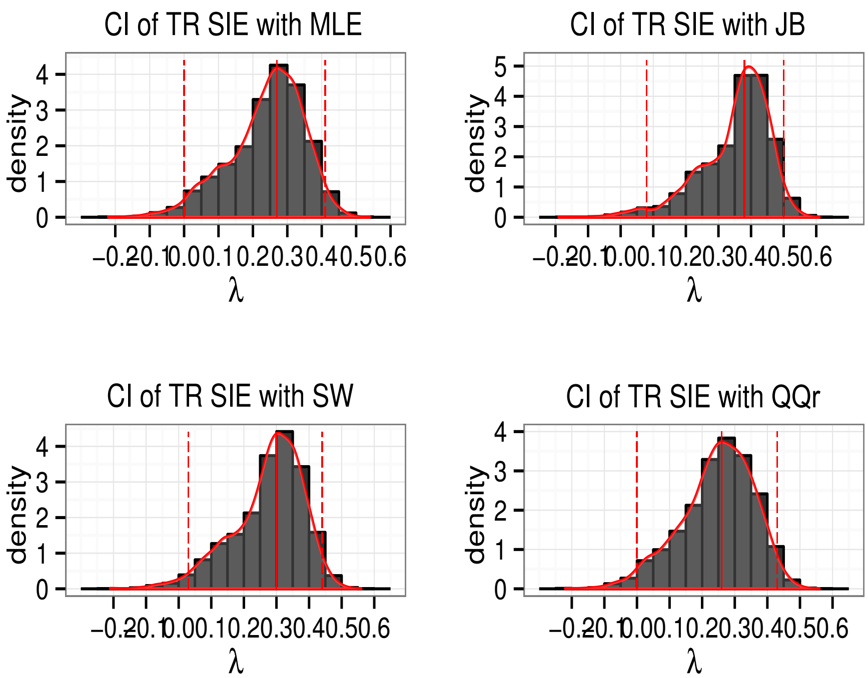

Figure 4: Simulated λ confidence interval of the SIE's trading numbers

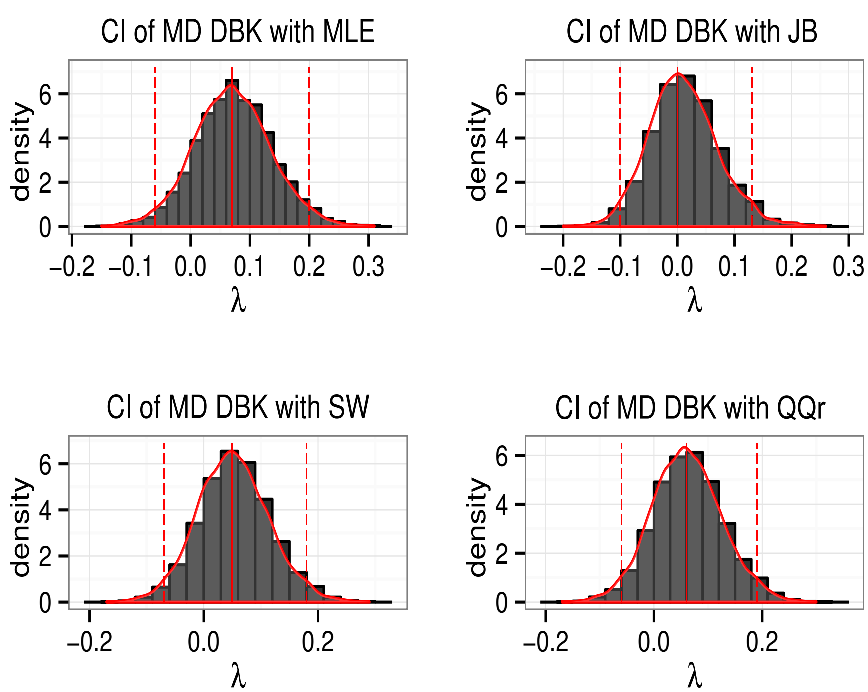

Figure 5: Simulated λ confidence interval of the DBK's mean duration

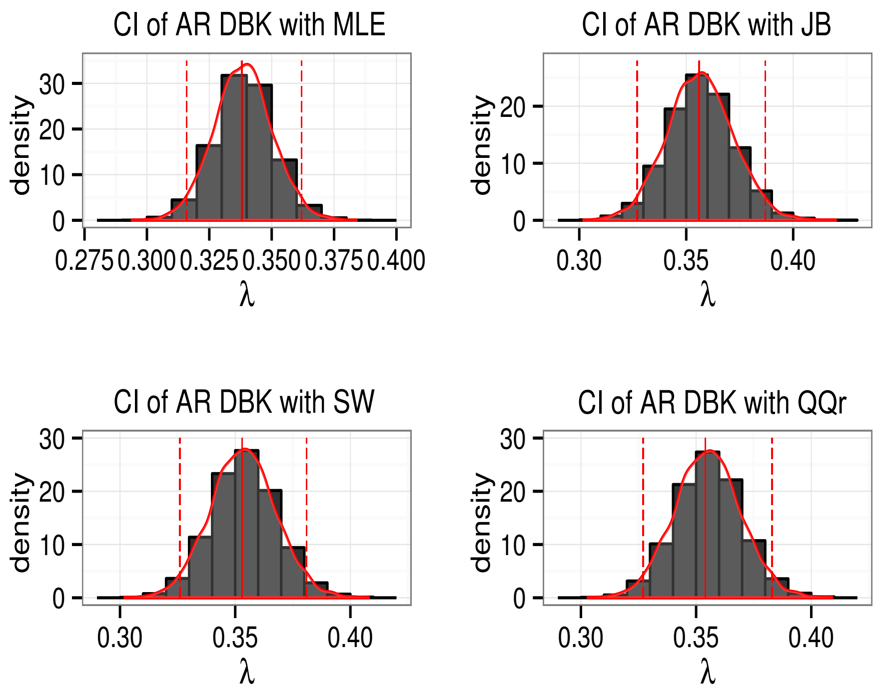

Figure 6: Simulated λ confidence interval of the DBK's absolute returns

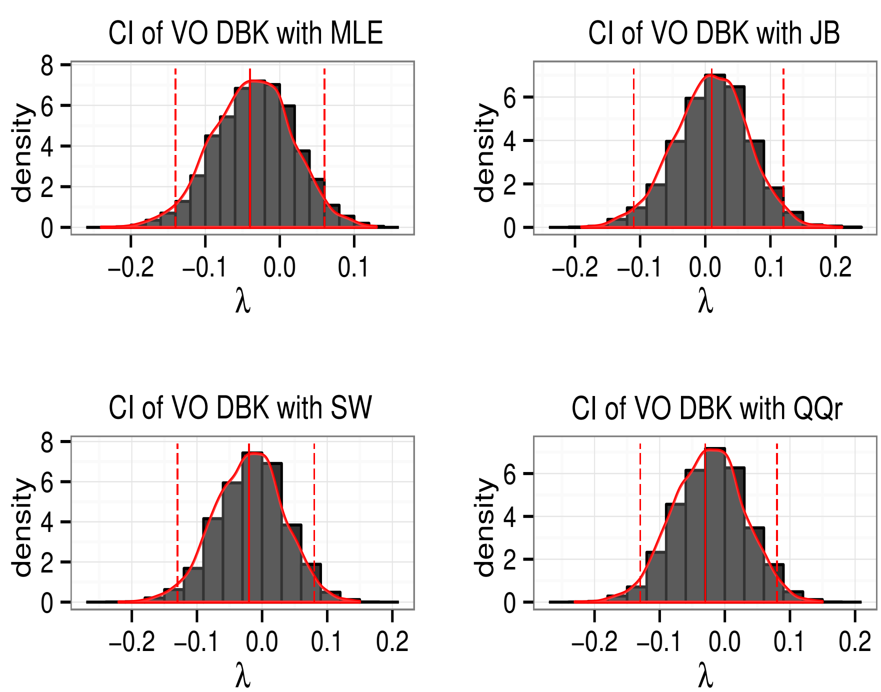

Figure 7: Simulated λ confidence interval of the DBK's trading volume

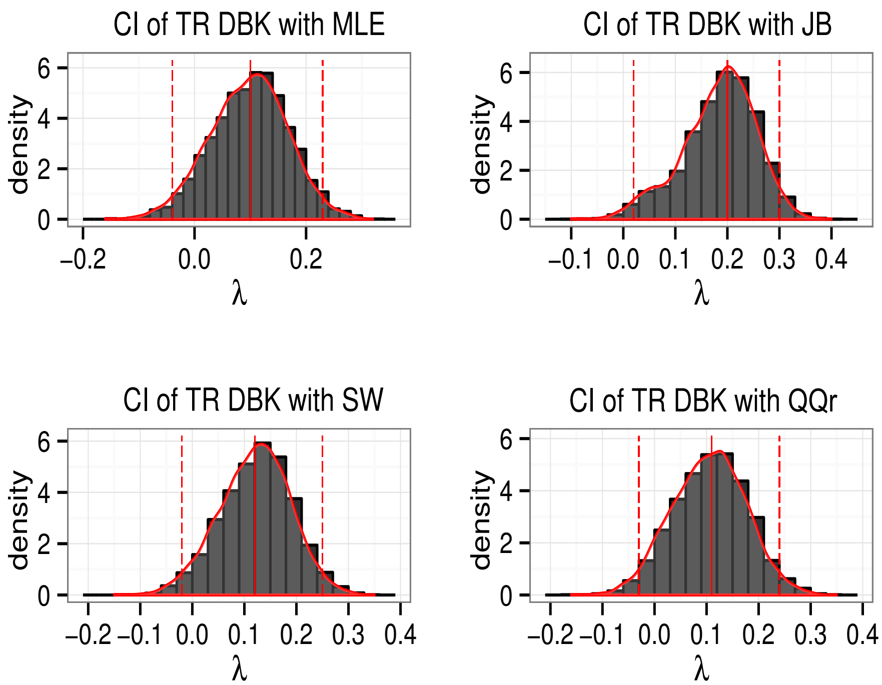

Figure 8: Simulated λ confidence interval of the DBK's trading numbers

In the figures, we find that the $\hat{\lambda}$ of MD is very close to 0 in all cases and the confidence interval also covers the origin. In Table 1, the CIs of SIE's MD are $[-0.05, 0.25]$ with the MLE, $[-0.08, 0.23]$ with the SW, $[-0.15, 0.22]$ with the JB and $[-0.07, 0.25]$ with the QQr and he CIs of DBK's MD are $[-0.06, 0.20]$ with the MLE, $[-0.07, 0.18]$ with the SW, $[-0.10, 0.13]$ with the JB and $[-0.06, 0.19]$ with the QQr. Due to the small value $\hat{\lambda}$ and its CI, it seems that a logarithmic transformation is suitable in the power transformation process. In Fig. 9 and 10, it is obvious that after the Box-Cox power transformation with $\hat{\lambda}$, both the histogram and the Q-Q plot of SIE and DBK perform better than before. Therefore, in the semiparametric analysis, a Semi-Log-MEM process is preferred in this case.

Similar to the MD, the VO series also possesses an almost zero $\hat{\lambda}$ and a through-origin CI, which also indicates a log-transform consideration. However, the AR set performs a little different. The $\hat{\lambda}$ with all methods are definitely positive and the values are between 0.32 and 0.36. Like the returns in the SemiGARCH model with Box-Cox transformation (Zhang et al., 2017), a third-root power transformation based on a Semi-MEM process seems to be applicable to the AR series. Besides, the results of the TR series are between the above two circumstances. The simulated results of SIE are close to the AR-type, however, those of DBK are in the form

Table 1: The selected λ, confidence interval and bandwidth

		MLE			SW			JB			QQr		
		λ	CI	b	λ	CI	b	λ	CI	b	λ	CI	b
SIE	MD	0.09	[-0.05, 0.25]	0.07	0.07	[-0.08, 0.23]	0.07	0.00	[-0.15, 0.22]	0.08	0.09	[-0.07, 0.25]	0.07
	AR	0.34	[0.30, 0.37]	0.12	0.34	[0.30, 0.39]	0.12	0.32	[0.26, 0.39]	0.12	0.35	[0.30, 0.39]	0.12
	VO	0.05	[-0.15, 0.18]	0.07	0.09	[-0.13, 0.21]	0.07	0.16	[-0.08, 0.27]	0.07	0.05	[-0.15, 0.20]	0.07
	TR	0.27	[0.00, 0.41]	0.08	0.30	[0.03, 0.44]	0.08	0.38	[0.08, 0.50]	0.08	0.26	[0.00, 0.43]	0.08
DBK	MD	0.07	[-0.06, 0.20]	0.07	0.05	[-0.07, 0.18]	0.07	0.00	[-0.10, 0.13]	0.07	0.06	[-0.06, 0.19]	0.07
	AR	0.34	[0.32, 0.36]	0.07	0.35	[0.33, 0.38]	0.07	0.36	[0.33, 0.39]	0.07	0.35	[0.33, 0.38]	0.07
	VO	-0.04	[-0.14, 0.06]	0.08	-0.02	[-0.13, 0.08]	0.08	0.01	[-0.11, 0.12]	0.08	-0.03	[-0.13, 0.08]	0.08
	TR	0.10	[-0.04, 0.23]	0.07	0.12	[-0.02, 0.25]	0.07	0.20	[0.02, 0.30]	0.07	0.11	[-0.03, 0.24]	0.07

of the MD-VO-type. Thus, for the TR series, both the Semi-MEM process and the Semi-Log-MEM process are the possible power transformation models, depending on the exact $\hat{\lambda}$ selection.

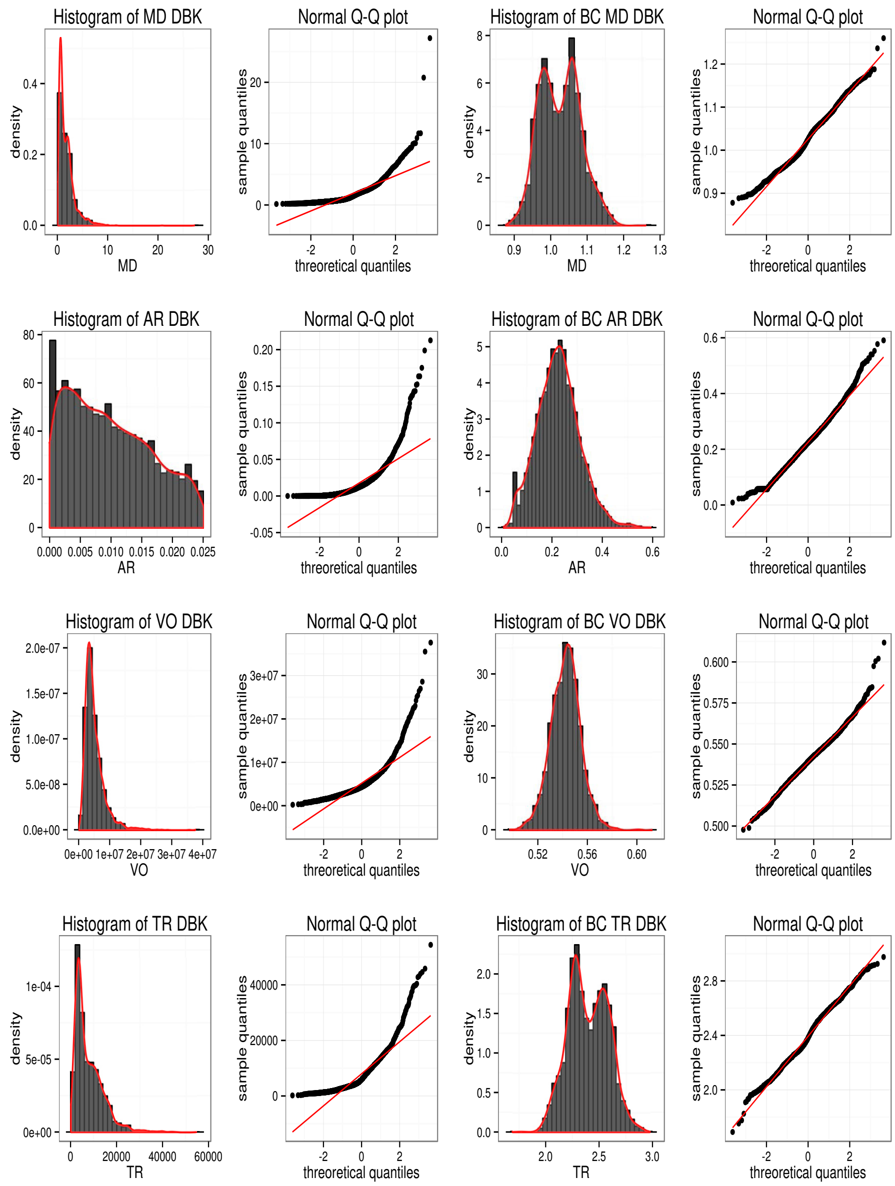

Figure 9: Histogram and QQ plot of DBK

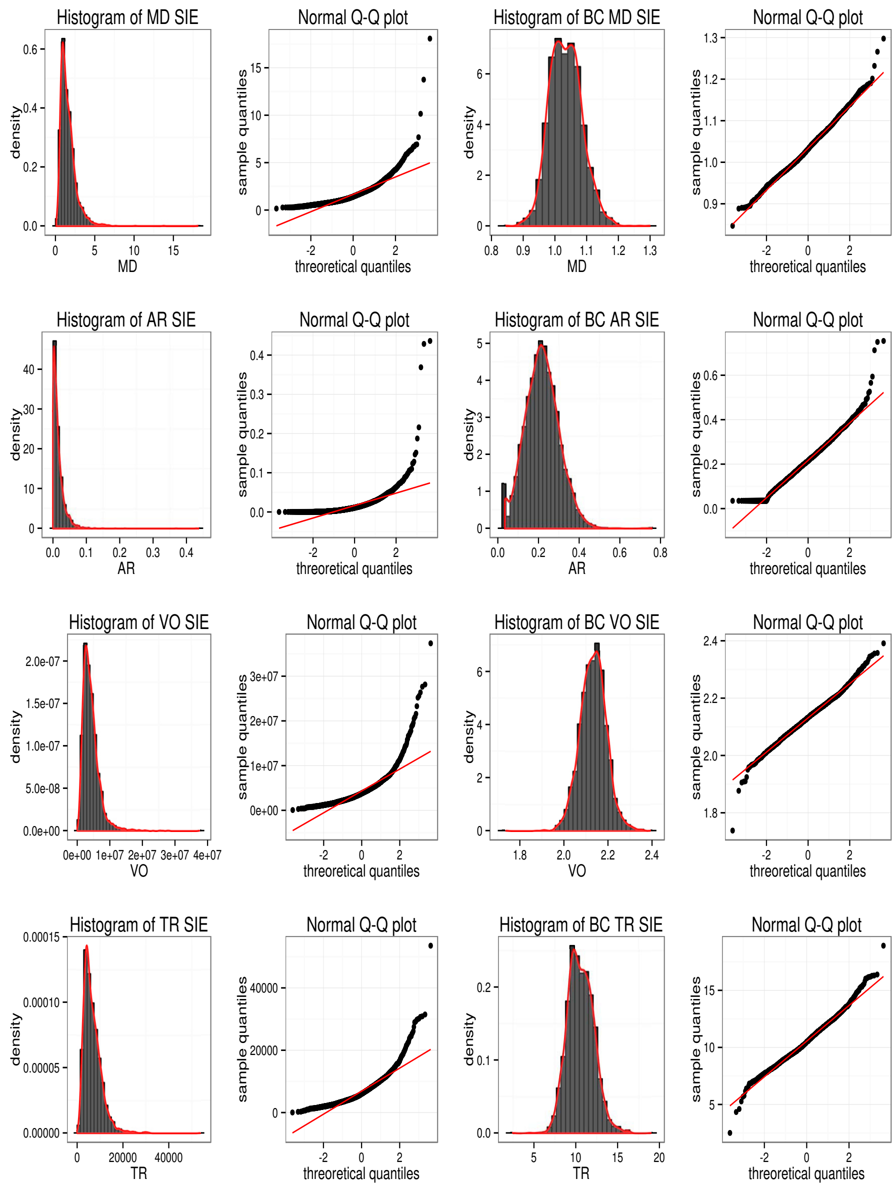

Figure 10: Histogram and QQ plot of SIE

To describe the simulation quality of the $\hat{\lambda}$'s CI, we have to bring the length (L) and shape (Δ) statistics into consideration. The two statistics are defined as

$$L = \hat{\lambda}^*_{k_U} - \hat{\lambda}^*_{k_L}, \tag{27}$$

and

$$\Delta = (\hat{\lambda}^*_{k_U} - \hat{\lambda})/(\hat{\lambda}^*_{k_L} - \hat{\lambda}). \tag{28}$$

Table 2: The length and shape of simulated CI

		MLE		SW		JB		QQr	
		L	Δ	L	Δ	L	Δ	L	Δ
SIE	MD	0.30	1.14	0.31	1.07	0.37	1.47	0.32	1.00
	AR	0.07	0.97	0.09	1.14	0.13	1.32	0.09	1.02
	VO	0.33	0.65	0.34	0.55	0.35	0.46	0.35	0.75
	TR	0.41	0.52	0.41	0.52	0.42	0.40	0.43	0.65
DBK	MD	0.26	1.00	0.25	1.08	0.23	1.30	0.25	1.08
	AR	0.04	1.09	0.05	1.29	0.06	1.07	0.05	1.07
	VO	0.20	0.92	0.21	0.91	0.23	0.92	0.21	1.10
	TR	0.27	0.91	0.27	0.93	0.28	0.56	0.27	0.93

Because we employ the percentile CI in the paper and it is indeed a first-order accurate procedure. If the coverage probabilities are identical, the two statistics can be applied to detect the CI quality. Therefore, we evaluate the simulation quality of CI through the L and Δ.

In Table 2, we can see that in most case, the lengths of the CI with MLE, SW and QQr are always more accurate than those with JB, e.g. the CI length of MD with the MLE, SW and QQr are 0.3, 0.31 and 0.32, however, the CI length of JB reach even 0.37. For another, from the performance of the shape, the shapes of the MLE and QQr are closer to one than the other two's, which means that the $\hat{\lambda}$ lies closer to the center of the CI, so that the CIs selected by the MLE and QQr are more asymmetric. In general, the CIs with the MLE and QQr perform well in practice, while the CI with JB seems to be not as good as the other ones. In addition, the CI with SW is not so stale, especially in the shape performance, e.g. the VO shape of SIE is only 0.55 and the AR shape of DBK is as high as 1.29. So, in the latter descale process, we are going to apply the $\hat{\lambda}$ by MLE (tiny difference from $\hat{\lambda}$ with QQr) as the considered power.

Note that the CIs discussed in this section only refer to the first-order accurate percentile CI. For the purpose of the CI modification, the second-order accurate methods can be introduced, such as the bootstrap-t and the BCa methods (Efron, 2003). In addition to the discussed data, we have also tried the realized volatility (RV), however the $\hat{\lambda}$ of RV is always negative, due to the extreme values. Hence, the RV is no longer considered here.

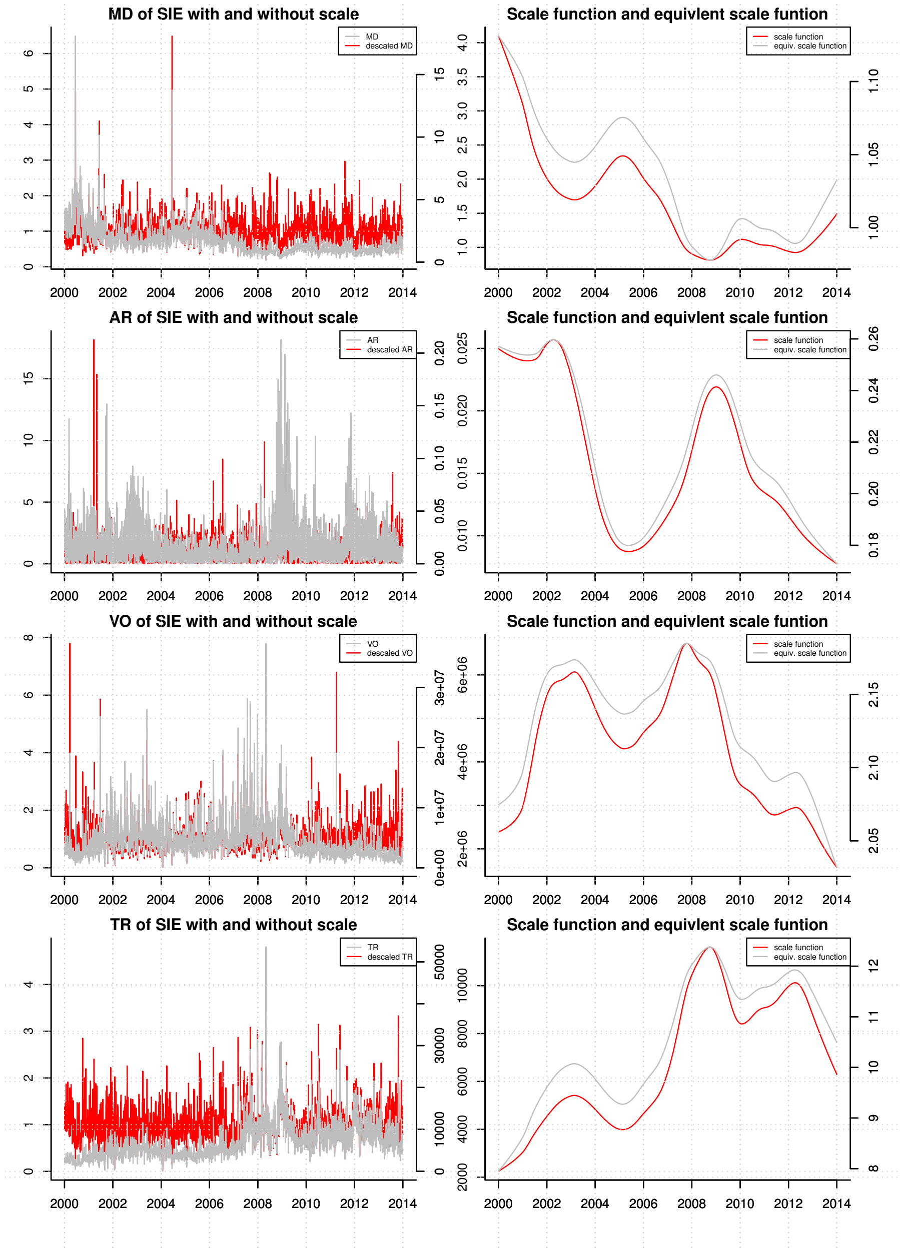

Figure 11: Estimation results of SIE from Jan 2000 to Dec 2013

Finally, in Fig. 6.11 and Fig. 6.12, the original and descaled data are displayed at the left side, while the estimation of the scale function and the equivalent scale function are at the right side.

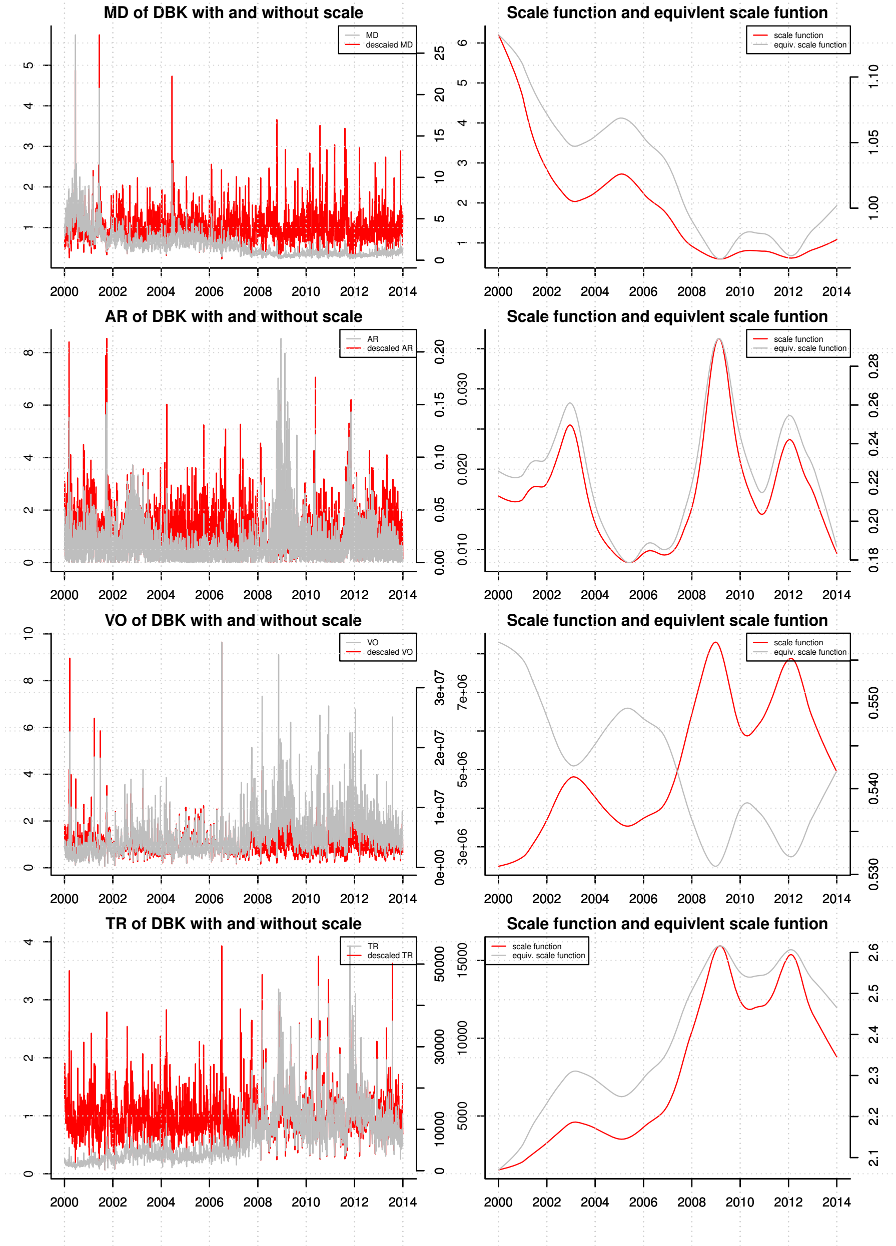

Figure 12: Estimation results of DBK from Jan 2000 to Dec 2013

Comparing the data, we can see obviously that the descaled data are more stationary than the original ones, bringing the smaller volatility after removing the trend function from the non-stationary process. Further, we have to pay attention to the change of the scale function. During the world financial crisis (2007-2008), the transaction risk reaches its peak, combining with the high transaction volume and numbers and low transaction duration. Therefore, we can see that the MD is at its minimum value, however, the AR, VO and TR all reach the peak values. Besides, the data are selected from DAX 30, the German financial crisis 'neuer Markt' (2002-2003) can also be observed by reaching the changing points in the sub-period.

6 Final remarks

We discussed a general SemiMEM model with Box-Cox power transformation to analyze the nonnegative data, such as MD, AR, VO and TR by selecting the power parameter with the MLE, SW, JB and QQr criteria. In the IPI procedures, we found that, for the MD and VO, the Semi-Log-MEM model is optimal, due to the approaching zero $\hat{\lambda}$, while for the AR, the $\hat{\lambda}$ is significantly positive, leading to a SemiMEM process. Both the Semi-Log-MEM and SemiMEM are suitable for the TR, depending on the $\hat{\lambda}$ selection in cases. Furthermore, we have discovered the simulated CI of $\hat{\lambda}$ without model or distribution assumption by introducing the block bootstrap method for dependent variables. The simulation CI quality is also considered. With the accurate coverage probabilities, we introduced the length and shape parameters to detect the simulated CI quality with the four mentioned criteria and it is found that the CI qualities with the MLE and QQr methods are better. Besides, IPI algorithms are also developed, referring to the c_f selection. The c_f is selected using the algorithm in Bühlmann (1996), estimating a Bartlett-window estimator and the optimal window width is proved to be an $O(T^{-1/3})$ term. Hence, in the paper, we selected the cubic root of the sample size as the optimal block length for simplicity.

Finally, some contributions will be possibly carried out in the future. First, the long memory parameter can be introduced into the framework. The long memory models, such as the Semi-FI-MEM and Semi-FI-Log-MEM models, can be discussed with the suitable power transformation under weak moment conditions, developing the algorithm of the synchronous selection of the differencing parameter d and the power parameter λ. Then, the power transformation technique can be updated. In the paper, we applied the Box-Cox transformation and with this method, the transformation is only close to the normal distribution, leading to the application of the other distribution, such as sin-arcsin distribution, sinh-normal distribution, Birnbaum-Saunders distribution, and so on. Further, the intraday high-frequency financial data can be introduced. Set up the spatial general SemiMEM models by modeling the high-frequency data in the daily and intraday dimension.

References

Altman, N. (1990). Kernel smoothing with correlated errors. *Journal of the American Statistical Association*, 85, 749–759.

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

Beran, J. and Y. Feng (2002). Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties. *Journal of Computational and Graphical Statistics*, 11, 690–713.

Box, G. and D. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, Series B, 26, 211–252.

Brockmann, M., T. Gasser and E. Herrmann (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88, 1302–1309.

Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *Journal Time Series Analysis*, 17, 247-270.

Bühlmann, P. and H. Künsch (1999). Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis*, 31, 295–310.

Caporin, M., E. Rossi and P. Santucci de Magistris (2017). Chasing volatility. *Journal of Econometrics*, 198, 122–145.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25, 1–37.

Dufour, A and R. Engle (2000). Time and the impact of a trade. *Journal of Finance*, 55, 2467–2489.

Eagleson, G. and H. Müller (1997). Transformations for smooth regression models with multiplicative errors. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59, 173–189.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1–26.

Efron, B. and R. Tibshirani (1993). An introduction to the bootstrap. *Chapman and Hall*, New York.

Engle, R. (2002). New frontiers for arch models. *Journal of Applied Econometrics*, 17, 425–446.

Engel, R. and J. Rangel (2008). The Spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21, 1187–1222.

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

Fan, J. and Q. Yao (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85, 645–60.

Feng, Y. (2004). Simultaneously modeling conditional heteroskedasticity and scale change. *Econometric Theory*, 20, 563–596.

Feng, Y. and T. Gries (2017). Data-driven local polynomial for the trend and its derivatives in economic time series. *CIE WP*, 2017-04, Paderborn University.

Feng, Y. T. Gries and M. Fritz (2019). Data-driven local polynomial for the trend and its derivatives in economic time series. *CIE WP*, 2019-06, Paderborn University.

Feng, Y and S. Heiler (1998). Locally weighted autoregression, in: R. Galata and H. Küchenhoff (Eds.) *Econometrics in theory and practice*, Physica-Verlag, Heidelberg, 101–117.

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

Gasser, T., A. Kneip and W. Köhler (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association*, 86, 643–652.

Ghosh, S and D. Draghicescu (2002). An algorithm for optimal bandwidth selection for smooth nonparametric quantile estimation. In Y. Dodge (Ed.), *Statistics for Industry and Technology: Statistical Data Analysis Based on the L1-Norm and Related Methods*, 161–168, Basel: Birkhäuser.

Hall, P. (1992). The bootstrap and Edgeworth expansion. *Springer*, New York.

Hampel, F., E. Ronchetti, P. Rousseeuw and W. Stahel (1986). Robust statistics: The approach based on influence function. *Wiley*, New York.

Hart, J. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society*, Series B, 53, 173–188.

Hautsch, N. (2002). Modeling intraday trading activity using Box-Cox ACD models. Discussion Paper 02/05, CoFE, University of Konstanz.

Herrmann, E., T. Gasser and A. Kneip (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, 79, 783–795.

Karanasos, M. (2008). The statistical properties of exponential ACD models. *Quantitative and Qualitative Analysis in Social Sciences*, 2, 29–49.

Künsch, H (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17, 1217–1241.

Lee, S. and B. Hansen (1994). Asymptotic theory for the GARCH$(1,1)$ quasi-maximum likelihood estimator. *Econometric Theory*, 10, 29–52.

Liu, R. and K. Singh (1992). Moving blocks jackknife and bootstrap capture weak dependence. In: R. Lepage and L. Billard, Eds., Exploring the Limits of Bootstrap, *John Wiley*, New York.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, 8, 377–399.

Politis, D. and J. Romano (1993). The stationary bootstrap. *Journal of the American Statistical Association*, 89, 1301–1313.

Politis, D. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23, 53–70.

Priestley, M. (1981). *Spectral analysis and time series* (Volume 1 and 2). Academic Press, London.

Ruppert, D., M. Wand, U. Holst and O. Hössjer (1997). Local polynomial variance-function estimation. *Technometrics*, 39, 262–273.

Taylor, N. and Y. Xu (2017). The logarithmic vector multiplicative error model: an application to high frequency NYSE stock data. *Quantitative Finance*, 17, 1021–1035.

Tukey, J. (1958). Bias and confidence in not quite large samples (abstract). *The Annals of Mathematical Statistics*, 29, 614.

Wahba, G. and S. Wold (1975). A completely automatic french curve: fitting spline functions by cross validation. *Communications in Statistics*, A, 4, 1–17.

Zhang, X, Y. Feng and C. Peitz (2017). A general class of SemiGARCH models based on the Box-Cox transformation. *CIE WP*, 2017-05, Paderborn University.

Appendix: Proof of the theorem

Proof of Theorem 1

Define $b_A = C_{\mathrm{A}} T^{-1/5}$, where C_{A} is the constant in b_A. We have $\hat{b} = \hat{C}_{\mathrm{A}} T^{-1/5}$ and

$$(\hat{b} - b_{\mathrm{A}})/b_{\mathrm{A}} = C_{\mathrm{A}}^{-1}(\hat{C}_{\mathrm{A}} - C_{\mathrm{A}}). \tag{29}$$

Taylor expansion leads to

$$\hat{C}_{\mathrm{A}} - C_{\mathrm{A}} \doteq O_p(\hat{I}(g_\lambda^2) - I(g_\lambda^2)) + O_p(\hat{I}((g_\lambda'')^2) - I((g_\lambda'')^2)) + O(\hat{c}_f - c_f). \tag{30}$$

It is well known that

$$\hat{I}((g_\lambda'')^2) - I((g_\lambda'')^2) \doteq O_p(T^{-2/7}). \tag{31}$$

Following the results in chapter 6.2 of Priestley (1981), the error in the lag-window estimator of c_f using the Bartlett-window and a bandwidth $K = O(T^{1/3})$ is

$$\hat{c}_f - c_f = O_p(T^{-1/3}). \tag{32}$$

Moreover, the first term $O_p(\hat{I}(g_\lambda^2) - I(g_\lambda^2)) = O_p(n^{-1/2})$ is negligible. $\diamond$

Recent discussion papers

2019-08	Xuehai Zhang	A Box-Cox semiparametric multiplicative error model
2019-07	Angelika Endres Joachim Heinzel	The Impact of Product Qualities on Downstream Bundling in a Distribution Channel
2019-06	Xuehai Zhang	Value at Risk and Expected Shortfall under General Semi-parametric GARCH models
2019-05	Xuehai Zhang	Republished as CIE Working Paper 2019-08
2019-04	Fatma Aslan Papatya Duman Walter Trockel	Duality for General TU-games Redefined
2019-03	Joachim Heinzel	Bundling in a Distribution Channel with Retail Competition
2019-02	Joachim Heinzel	Credence Goods Markets with Fair and Opportunistic Experts
2019-01	Joachim Heinzel	Credence Goods Markets with Heterogeneous Experts
2018-07	Papatya Duman	Does Informational Equivalence Preserve Strategic Behavior? An Experimental Study on Trockel's Game
2018-06	Bernard Michael Gilroy Alexander Golderbein Christian Peitz Nico Stöckmann	Incentives for the finance sector: How the ECB affects banks' business assembling
2018-05	Marc Atkins Bernard Michael Gilroy Volker Seiler	New Dimensions of Service Offshoring in World Trade
2018-04	Vanessa Hilleringmann	The Influence of Bribery and Relative Reciprocity on a Physician's Prescription Decision – An Experiment
2018-03	Claus-Jochen Haake Cheng-Zhong Qin	On unification of solutions to the bargaining problem
2018-02	Simon Hoof	Feedback Pareto weights in cooperative NTU differential games
2018-01	Claus-Jochen Haake Nadja Stroh-Maraun	A Note on Manipulability in School Choice with Reciprocal Preferences
2017-12	Britta Hoyer Nadja Stroh-Maraun	Matching Strategies of Heterogeneous Agents under Incomplete Information in a University Clearinghouse
2017-11	Bernard Michael Gilroy Julia Günthner	The German Precariat and the Role of Fundamental Security - Is the Unconditional Basic Income a Possible Solution for the Growing Precarity in Germany?
2017-10	Achim Voß Mark Schopf	Lobbying over Exhaustible-Resource Extraction
2017-09	Walter Trockel Claus-Jochen Haake	Thoughts on Social Design