International Research Journal of Applied Finance ISSN 2229 – 6891
Vol. X Issue – 9 September, 2019 www.irjaf.com

# The Logarithmic Multiplicative Error Model: Moments and Applications

**Li Li Wanbo Lu* Rui Ke**

**Abstract**

The multiplicative error model (MEM) proposed by Engle (2002) is suitable for all non-negative financial time series. In this paper, the logarithmic multiplicative error model (Log-MEM) is constructed based on the classical MEM. Meanwhile, we inferred the conditions of the existence and analytical expressions of the unconditional moments of Log-MEM. The dispersion index and the autocorrelation function of the Log-MEM are also obtained. We select the high frequency trading data of five minutes of Shanghai Stock Exchange Composite Index and Shanghai and Shenzhen 300 Index to illustrate how to select an appropriate model for the bid-ask spread in application.

**Key words:** Log-MEM; Unconditional Moments; Dispersion Index; ACF; Bid-ask Spread
**Mathematics Subject Classification (2010): 90C39, 49L20**

## 1. Introduction

More and more researches focus on the dynamic analysis of the nonnegative financial time series in financial markets, such as volume, volatility, range, duration, bid-ask spread, number of trades and so on. Generalizing the GARCH (Bollerslev (1986)) and autoregressive conditional duration (ACD) (Engle and Russell (1998)) approaches, Engle (2002) extends to give the Multiplicative Error Model (MEM). They are particularly suited to model such nonnegative time series by setting the dynamics of the nonnegative variable as the product of two terms, namely a scale factor evolving in a conditionally autoregressive way and an independent and identically distributed (i.i.d.) error term with unit mean. It is a suitable representation of nonnegative valued processes, which have dynamic interactions with one another. Another example is the conditional autoregressive range (CARR) model proposed by Chou (2005). The GARCH, ACD and CARR model are special cases of MEM.

Gallo and Velucchi (2005) used MEMs to study different measures of volatility based on 5 minutes returns. Brownless and Gallo (2005) studied the hourly return, volumes and number of trades. Engle and Gallo (2006) used MEM to model three different indicators of volatility (absolute returns, daily range and realized volatility). The empirical result indicated that the three indicators have significant correlation and the MEM has the good forecast effect. Their estimation method has been widely used, the examples may be found in Engle, Gallo and Velucchi (2005, 2008) for the volatility spillover effects among financial markets, Giovannetti and Velucchi (2011) for the relationships between developed markets (US, UK and China) and some Sub-Saharian African (SSA) emerging markets (Kenya, Nigeria and South Africa) in the period 2004~2009.

At the same time, many researches also concentrate on extension of MEMs. Luca and Gallo (2004) applied the simple mixture exponential distribution to the MEM and gained better prediction effect, compared with simple exponential distribution. Lanne (2006) proposed a new MEM with time-varying parameters and an error term following a mixture of gamma distributions and the forecasting performance of the new model is superior to that of the standard set. Ahoniemi and Lanne (2009) proposed a new bivariate mixture multiplicative error model and show that it is a good fit to Nikkei 225 index call and put option implied volatility. Using Engle's MEM formulation, Lam and Ng (2009) formulated MEM-GARCH modeling procedures for the embedding of intra-daily information using range-based volatilities. Brownlees, Cipollini and Gallo (2011) discussed the specification of the conditional expectation and the error term and provided a general framework, allowing for richer specifications of the conditional mean. The outcome is a novel MEM called Composite MEM. Engle, Fleming, Ghysels and Nguyen also proposed a new class of dynamic order book model based the MEM framework and their new model overcome the main challenges faced by the standard linear framework in modeling non-negative valued process. Otranto (2011) proposed to classify the level of the unconditional volatility obtained from MEM with the possibility of changes in the parameters of the model in terms of regime switching or time varying smoothed coefficients and use 15 series of realized kernel volatility relative to the main financial indices to test their method. Brownlees and Gallo (2011) investigated a shrinkage estimator semiparametric MEM. Gallo and Otranto (2012) insert a Markovian dynamics in a MEM to represent the conditional expectation of the realized volatility, allowing us to address the issues of a slow moving average level of volatility and of a different dynamics across regime. Han, Park and Zhang (2015) introduced a MEM that incorporates heterogeneous components: weekly and monthly realized volatility measures and applied it to model 34 different assets. The result showed that the new model has good performance in both in-sample and out-of-sample forecasts.

To sum up, most of works on MEM used the classical linear conditional expectation equation in MEM due to the non-negative limitation of the parameters which is violate in including exogenous variables in practice. Also many works are done for the volatility, duration and price range. For bid-ask spread, researchers mainly discussed its linear influence capability on other financial variables and omit the dynamic conditional correlation in itself. So far, we extend the classical MEM to the Log-MEM and obtain its analytical expressions for the unconditional moments, dispersion index and autocorrelation function (ACF). We also provide an empirical application in which we compute the unconditional moments and ACF for the MEM and Log-MEM estimated on bid-ask spread for the two main indices in China stock market.

The paper is organized as follows. In Section 2, we define the class of Log-MEMs. In Section 3, we provide the conditions of existence and the general formulae of the moments, study the properties of the dispersion index and the ACF. Section 4 presents the model selection using real data. Section 5 concludes the paper.

## 2. Definition of Log-MEMs

Assume that $\{x_t\}$ is a stationary process defined on $[0,+\infty)$ and adapted to the filtration $\{F_t; t \in Z\}$ with $F_t = \sigma\{x_s; s \le t\}$ and having the form

$$x_t = \mu_t \varepsilon_t , \quad (1)$$

where $\mu_t$ is a determinant or predictable positive function conditionally on $F_{t-1}$ according to a parameter vector $\vartheta$, e.g. $\mu_t(\vartheta; F_{t-1})$; $\{\varepsilon_t\}$ is an i.i.d. nonnegative innovation process with unit mean and finite second moment. The general baseline MEM(*r*,*q*) relies on a linear parameterization of the available information set $F_{t-1}$ in (1):

$$\mu_t = \omega + \sum_{i=1}^{r} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j \mu_{t-j}, \tag{2}$$

where $r, q \geq 0, \omega > 0, \alpha = (\alpha_1, \ldots, \alpha_r) \geq 0, \beta = (\beta_1, \ldots, \beta_q) \geq 0$ are sufficient to ensure positive for all possible realizations of $\{x_t\}$. The distribution of the error term $\varepsilon_t$ has positive support. For the usually used exponential, Weibull and gamma distribution, we call (1) and (2) as EMEM(*r*,*q*), WMEM(*r*,*q*) and GMEM(*r*,*q*).

There are some restrictions on parameters in (2) in order to guarantee the non-negativity of the process. But if we have some exogenous variable in (2), the restrictions will be violated. Like the EGARCH model proposed by Nelson (1991) and the Log-ACD model proposed by Bauwens and Giot (2000), a Log-MEM specification of the observed nonnegative process is

$$\begin{cases} x_t = e^{\mu_t} \varepsilon_t \\ \mu_t = \omega + \sum_{i=1}^{r} \alpha_i g(\varepsilon_{t-i}) + \sum_{j=1}^{q} \beta_j \mu_{t-j} \end{cases} \tag{3}$$

where $g(\varepsilon_{t-i}) = \ln(\varepsilon_{t-i})$ or $g(\varepsilon_{t-i}) = \varepsilon_{t-i}$. Since $e^{\mu_t}$ is always positive, $\mu_t$ can take any value in real number set, the restrictions on the parameters disappear. We call (3) as Log-MEM$_1$ when $g(\varepsilon_{t-i}) = \ln(\varepsilon_{t-i})$, otherwise as Log-MEM$_2$ when $g(\varepsilon_{t-i}) = \varepsilon_{t-i}$. Several choices are available for the distribution of $\varepsilon_t$: exponential, gamma, generalized gamma, Weibull, lognormal, Pareto …, in practice any distribution with positive support. But it is a key and difficult step in practice to select an ideal distribution based on the real data.

In this paper, we try to infer the ideal distribution using the moment, the dispersion index and ACF of $x_t$. We first infer the sufficient and necessary condition of the existence of the moments and the analytical expressions for the unconditional moments. Then the MLE for each model, distribution and data series are used to compute the analytical expressions for the unconditional moments and dispersion indices. The results based on the analytical expressions will be compared with the empirical unconditional moments and dispersion indices.

To infer the moments of Log-MEM, we transfer (3) into its general form. Let *p*=max{*r*, *q*}, (3) is written as

$$\begin{cases} x_t = e^{\mu_t} \varepsilon_t \\ \mu_t = \omega + \sum_{i=1}^{p} \alpha_i g(\varepsilon_{t-i}) + \sum_{j=1}^{p} \beta_j \mu_{t-j} \end{cases}. \tag{4}$$

So in Log-MEM$_1$, the equation of $\mu_t$ becomes

$$\mu_t = \omega + \sum_{j=1}^{p} \alpha_j \ln \varepsilon_{t-j} + \sum_{j=1}^{p} \beta_j \mu_{t-j} = \omega + \sum_{j=1}^{p} \alpha_j \ln x_{t-j} + \sum_{j=1}^{p} \left(\beta_j - \alpha_j\right) \mu_{t-j}. \tag{5}$$

In Log-MEM$_2$, the equation of $\mu_t$ becomes

$$\mu_t = \omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j} + \sum_{j=1}^{p} \beta_j \mu_{t-j} = \omega + \sum_{j=1}^{p} \alpha_j \frac{x_{t-j}}{e^{\mu_{t-j}}} + \sum_{j=1}^{p} \beta_j \mu_{t-j} . \tag{6}$$

### 3. Moment, Dispersion Index and Autocorrelation Function

The nonnegative valued time series are often characterized by two features. One important fact is overdispersion, i.e., the standard deviation is lower than the mean, for example, price duration or underdispersion, i.e., the standard deviation is larger than the mean, for example, volume duration. Another important stylized fact is the shape of the ACF, which usually decreases slowly from a relatively low first-order autocorrelation. It is therefore essential that Log-MEM be able to fit such stylized facts. Before the analytical expressions for the dispersion index and ACF are given, we first give the results of the unconditional moments.

Refer to the idea in Hamilton (1994) to solve $p$ order difference equation, let us define the matrix

$$\Omega = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_{p-1} & \beta_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{7}$$

and the coefficients

$$\phi_k = \beta' \Omega^{k-p-1} \phi, \quad k > p , \tag{8}$$

where $\beta = (\beta_1, \cdots, \beta_p)'$ and $\phi = (\phi_1, \cdots, \phi_p)'$ such that

$$\begin{aligned} &\phi_0 = 1, \\ &\phi_1 = \beta_1, \\ &\phi_s = \sum\nolimits_{j=1}^{s} \beta_j \phi_{s-j}, s = 1, 2, \cdots, p, \\ &\phi_s = \sum\nolimits_{j=1}^{p} \beta_j \phi_{s-j}, s > p. \end{aligned} \tag{9}$$

Let $\lambda(\Omega)$ be the absolute value of the maximum eigenvalue of the matrix $\Omega$. The unconditional moments of $x_t$ exist and are independent of $t$ as $k \to \infty$ if and only if $\lambda(\Omega) < 1$. Therefore, $\sum_{j=0}^{\infty} \phi_j = \sum_{j=0}^{p} \phi_j + \beta'(I - \Omega)^{-1} \phi = (1 - \sum_{j=1}^{p} \beta_j)^{-1}$, the sequence $\{\phi_i\}$ converges to a finite value due to $\Omega^k \to 0$, $\sum\nolimits_{j=0}^{k} \Omega^j \to (I - \Omega)^{-1}$ as $k \to \infty$.

**Proposition 1.** *Assume that* $\mathrm{E}\exp\left[m\theta_j g(\varepsilon_t)\right]$ *and* $\mu_m = \mathrm{E}|\varepsilon_t|^m$ *exist for an arbitrary* $m \in R^+$. *For the Log-MEM process defined in (3), the condition* $\lambda(\Omega) < 1$ *is necessary and sufficient for the existence of the m-th moment* $\mathrm{E}x_t^m$. *We have*

$$\mathrm{E}x_t^m = \mu_m \exp\left[ m\omega \left( 1 - \sum_{j=1}^{p} \beta_j \right)^{-1} \right] \prod_{j=1}^{\infty} \mathrm{E}\exp\left[ m\theta_j g(\varepsilon_t) \right], \tag{10}$$

*where* $\theta_s = \begin{cases} \sum_{j=1}^{s} \alpha_j \phi_{s-j}, s = 1, 2, \cdots, p, \\ \sum_{j=1}^{p} \alpha_j \phi_{s-j}, s > p. \end{cases}$ *For the Log-MEM(1,1), if* $\mathrm{E}\exp\left[m\alpha_1 \beta_1^{j-1} g(\varepsilon_t)\right]$ *and* $\mu_m < \infty$

*for an arbitrary* $m \in R^+$ *and* $\lambda(\Omega) = |\beta_1| < 1$,

$$\mathrm{E}x_t^m = \mu_m \exp\left[\frac{m\omega}{1-\beta_1}\right]\prod_{j=1}^{\infty}\mathrm{E}\exp\left[m\alpha_1\beta_1^{j-1}g(\varepsilon_i)\right]. \tag{11}$$

In practical computation of (10), the infinite product that appears in the moment expression can be truncated after a sufficiently large number of terms since $\beta^{j-1}$ tends to 0, for example, we can obtain high accuracy for first and second moments by fixing $j = 1001$.

As usual way, let us define the dispersion index $\delta_x = \frac{\sigma_x}{\mu_x}$ of the random variable $x$, where $\sigma_x$, $\mu_x$ are the standard deviation and mean of $x$, respectively. The ratio is larger than one in the case of overdispersion. Since for the dispersion index of the Log-MEM process,

$$1+\delta_x^2 = 1+\frac{\sigma_x^2}{\mu_x^2} = 1+\frac{\mathrm{E}x_t^2-(\mathrm{E}x_t)^2}{(\mathrm{E}x_t)^2} = \frac{\mathrm{E}x_t^2}{(\mathrm{E}x_t)^2}, \tag{12}$$

we take *m*=1 and 2 in Proposition 1 to infer following result:

**Proposition 2.** *For the Log-MEM process defined by (3), assume that the hypotheses of Proposition 1 hold for m=1, 2. We have*

$$1+\delta_x^2 = (1+\delta^2)\frac{\prod_{j=1}^{\infty}\mathrm{E}\exp\left[2\theta_j g(\varepsilon_t)\right]}{\left\{\prod_{j=1}^{\infty}\mathrm{E}\exp\left[\theta_j g(\varepsilon_t)\right]\right\}^2} \ge 1+\delta^2, \tag{13}$$

*where* $\delta = \sigma/\mu$ *is the dispersion index of* $\varepsilon_t$, $\mu = \mathrm{E}[\varepsilon_t]$ *and* $\sigma = \sqrt{\mathrm{E}\varepsilon_t^2 - (\mathrm{E}\varepsilon_t)^2}$.

The dispersion index of $x_t$ cannot be smaller than that of $\varepsilon_t$ in (13). Thus, it suffices that $\varepsilon_t$ be equidispersed for $x_t$ to be overdispersed, as long as $\alpha \ne 0$. Table 1 illustrated the variation of $\delta_x$ as a change of $\alpha$ (from 0.05 to 0.2) and $\beta$ (from 0.8 to 0.98) when $\varepsilon_t$ is exponential (so that $\delta = 1$) and the models are Log-MEM$_1$ (1,1) and Log-MEM$_2$ (1,1). From Table 1, we find that the dispersion index of Log-MEM$_1$ (1,1) is slightly larger than that of Log-MEM$_2$ (1,1) (except for the case $\alpha = 0.2, \beta = 0.98$). In general, the influence on the dispersion index of $\alpha$ is larger than that of $\beta$, one can obtain a larger dispersion index by taking larger $\alpha$.

Table 1 Dispersion index of Log-MEM$_1$ (1,1) and Log-MEM$_2$ (1,1)

| Parameters | Log-MEM$_1$ (1,1) | Log-MEM$_2$ (1,1) |
|---|---|---|
| $\alpha = 0.05, \beta = 0.8$ | 1.013 | 1.009 |
| $\alpha = 0.05, \beta = 0.9$ | 1.025 | 1.018 |
| $\alpha = 0.05, \beta = 0.98$ | 1.122 | 1.084 |
| $\alpha = 0.1, \beta = 0.8$ | 1.041 | 1.033 |
| $\alpha = 0.1, \beta = 0.9$ | 1.079 | 1.061 |
| $\alpha = 0.1, \beta = 0.98$ | 1.386 | 1.295 |
| $\alpha = 0.2, \beta = 0.8$ | 1.152 | 1.158 |
| $\alpha = 0.2, \beta = 0.9$ | 1.293 | 1.296 |
| $\alpha = 0.2, \beta = 0.98$ | 2.661 | 2.640 |

Let us define the matrix $\Gamma = \begin{bmatrix} \beta^*_{1n} & \beta^*_{2n} & \beta^*_{3n} & \cdots & \beta^*_{p-1,n} & \beta^*_{pn} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$, where $n$ is the order of ACF

and

$$\beta^*_{1n} = \phi_n + 1, n \ge 1,$$

$$\beta^*_{hn} = \sum_{j=1}^{n} \beta_{h+j-1}\phi_{n-j}, 2 \le h \le p-n+1, 2 \le j \le p-1,$$

$$\beta^*_{hn} = \sum_{j=1}^{h} \beta_{p-n-1+h+j}\phi_{n-j}, p-n+2 \le h \le p-1, 2 \le j \le p-1,$$

$$\beta^*_{pn} = \beta_p \phi_{n-1}, 1 \le n \le p,$$

$$\beta^*_{jn} = \sum_{h=1}^{p+1-j} \beta_{j+h-1}\phi_{n-h}, n \ge p, 2 \le j \le p.$$

The next proposition provides the ACF.

**Proposition 3.** *For the Log-MEM process defined by (3), assume that* $\mu < \infty$, $\mu_2 = \sigma^2 + \mu^2 < \infty, \lambda(\Gamma) < 1$ , $\mathrm{E}[e^{\delta g(\varepsilon_t)}] < \infty$ *for any* $\delta \in R$ , $\mathrm{E}[\varepsilon_{t-n}\exp(\theta_n g(\varepsilon_{t-n}))] < \infty$ *for any* $n \in \mathrm{N}^+$ *, and* $\mathrm{E}\left[\exp\left(\left(\phi_{n-j}\alpha_{j+h} + \theta^*_{hn}\right) g\left(\varepsilon_{t-n-h}\right)\right)\right] < \infty$ *for j and h such that* $n \ge 1$*. Then, for* $n \ge 1$*, the n-th order ACF of* $x_t$ *has the form*

$$\rho_n = \frac{\mu \mathrm{E}[\varepsilon_t e^{\theta_n g(\varepsilon_t)}] \prod_{j=1}^{n-1} \mathrm{E}[e^{\theta_j g(\varepsilon_t)}] \prod_{j=p}^{\infty} \mathrm{E}\left(e^{\theta^*_{jn} g(\varepsilon_t)}\right) H_{n,p} - \mu^2 \left[\prod_{j=1}^{\infty} \mathrm{E}[e^{\theta_j g(\varepsilon_t)}]\right]^2}{\mu_2 \prod_{j=1}^{\infty} \mathrm{E}[e^{2\theta_j g(\varepsilon_t)}] - \mu^2 \left[\prod_{j=1}^{\infty} \mathrm{E}[e^{\theta_j g(\varepsilon_t)}]\right]^2}, \qquad (14)$$

where

$$H_{n,p} = \begin{cases} \prod_{h=1}^{p-n} \mathrm{E}\left\{\exp\left[(\sum_{j=1}^{n} \phi_{n-j}\alpha_{h+j} + \theta^*_{hn}) g\left(\varepsilon_{t-n-h}\right)\right]\right\} \\ \cdot \prod_{h=1}^{n-1} \mathrm{E}\left\{\exp\left[(\sum_{j=1}^{h} \phi_{n-j}\alpha_{p-h+j} + \theta^*_{p-h,n}) g\left(\varepsilon_{t-n-h}\right)\right]\right\}, 1 \le n \le p \\ \prod_{h=1}^{p-1} \mathrm{E}\left\{\exp\left[(\sum_{j=1}^{p-h} \phi_{n-j}\alpha_{h+j} + \theta^*_{hn}) g\left(\varepsilon_{t-n-h}\right)\right]\right\}, n > p \end{cases}$$

and

$$\theta^*_{jn} = \begin{cases} \sum_{h=1}^{j} \alpha_h \phi^*_{j+1-h,n}, & h = 1, \dots, p, \\ \sum_{h=1}^{p} \alpha_h \phi^*_{j+1-h,n}, & h > p, \end{cases}$$

in which $\phi^*_{0n} = 1$ , $\phi^*_{kn} = \sum_{j=1}^{k} \beta_j \phi^*_{k-j,n}$ for $k = 1, 2, \dots, p$ , $\phi^*_{kn} = \beta' \Gamma^{k-p-1} \phi$ for $k > p$ . *For the Log-MEM(1,1) process, we have*

$$\rho_n = \frac{\mu \mathrm{E}\left(\varepsilon_t e^{\alpha\beta^{n-1} g(\varepsilon_t)}\right)\prod_{j=1}^{n-1}\mathrm{E}\left(e^{\alpha\beta^{j-1} g(\varepsilon_t)}\right)\prod_{j=1}^{\infty}\mathrm{E}\left(e^{\alpha\beta^{j-1}\left(\beta^n+1\right) g(\varepsilon_t)}\right) - \mu^2\left[\prod_{j=1}^{\infty}\mathrm{E}\left(e^{\alpha\beta^{j-1} g(\varepsilon_t)}\right)\right]^2}{\mu_2\prod_{j=1}^{\infty}\mathrm{E}\left(e^{2\alpha\beta^{j-1} g(\varepsilon_t)}\right) - \mu^2\left[\prod_{j=1}^{\infty}\mathrm{E}\left(e^{\alpha\beta^{j-1} g(\varepsilon_t)}\right)\right]^2}. \quad (15)$$

The proofs of Proposition 1 and Proposition 3 are easily obtained since the similar structure between the MEM process and ACD process, the detail proof of the latter model can be seen in Bauwens, Galli and Giot (2003). The result of Proposition 2 is the direct application of Proposition 1 by taking the first and second unconditional moment. On the features of the ACF provided by Proposition 3, it is firstly worthwhile to notice that $\lim_{n\to\infty}\rho_n = 0$. This can be easily seen, in the Log-MEM(*p*, *p*) by considering that, as $n \to \infty$,

$$\mathrm{E}\varepsilon_t e^{\theta_n g(\varepsilon_{t-n})} \to \mu,$$

$$\prod_{j=1}^{n-1}\mathrm{E}e^{\theta_j g(\varepsilon_t)}\prod_{j=p}^{\infty}\mathrm{E}\left(e^{\theta_{jn}^{*} g(\varepsilon_t)}\right) \to \left[\prod_{j=1}^{n-1}\mathrm{E}e^{\theta_j g(\varepsilon_t)}\right]^2, \text{ and} \quad (16)$$

$$H_{n,p} \to 1.$$

Secondly, the shape of $\rho_n$ as a function of *n* in Proposition 3 is determined by the absolute value of the maximum eigenvalue of the $\Omega$ matrix. The closer $\lambda(\Omega)$ is to 1, the more persistent the autocorrelation.

## 4. Empirical Application

### *4.1. Data description*

We select the high frequency trading data of five minutes of Shanghai Stock Exchange Composite Index (SSEI) and Shanghai and Shenzhen 300 Index (CSI 300) as the research objects. They are both compiled by the China Securities Index Company, Ltd. The former has strong representation for the Shanghai stock market. The latter has strong representation for the China stock market since it consists of 300 stocks with the largest market capitalization and liquidity from the entire universe of listed A Share companies in China. The total sample period is from December 11, 2013 to March 30, 2014. We only model the bid-ask spread for these two series, which is defined as

$$Spread = \frac{P_a - P_b}{\left(P_a + P_b\right)/2} \times 100\%, \quad (17)$$

where $P_b$ and $P_a$ are the highest bid price and lowest ask price in five minutes trading interval. The total sample size is 2,784.

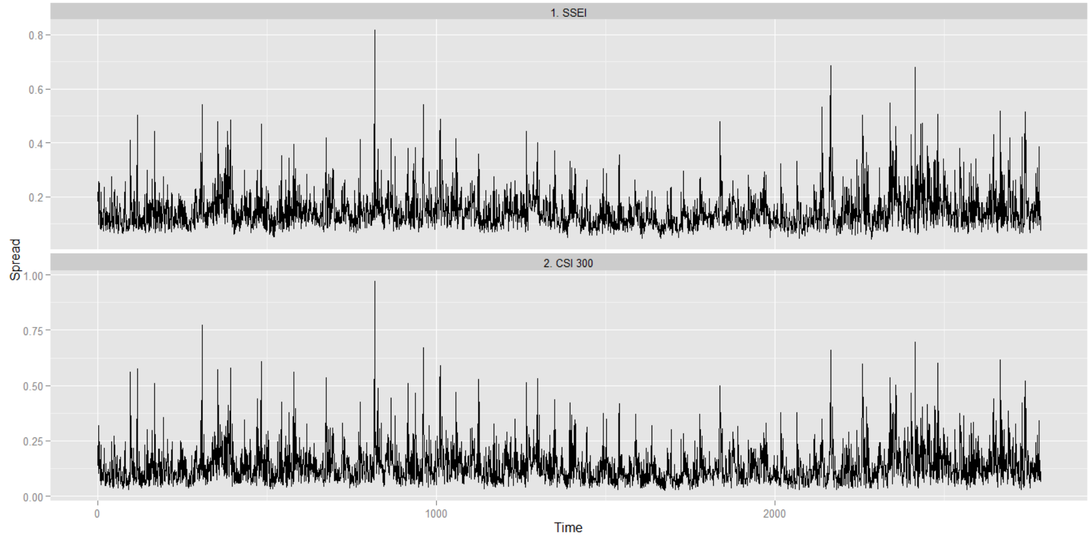


Figure 1. The bid-ask spreads of SSEI and CSI 300

Figure 1 gives the plot of the bid-ask spread of SSEI and CSI 300. We observe the bid-ask spread fluctuates around its mean. The bid-ask spread of CSI 300 has more changing range than that of SSEI and displays more volatility. There is volatility clustering in the bid-ask spread for both indices. Table 2 gives the descriptive statistics for the lowest ask price, the highest bid price and the bid-ask spread of SSEI and CSI 300.

Table 2 Descriptive statistics of the lowest ask price, highest bid price and bid-ask spread of SSEI and CSI 300

| | SSEI | | | CSI 300 | | |
|---|---|---|---|---|---|---|
| | $P_a$ | $P_b$ | *Spread*(%) | $P_a$ | $P_b$ | *Spread*(%) |
| Mean | 2079.5792 | 2076.5461 | 0.1461 | 2255.0256 | 2251.9974 | 0.1347 |
| Min | 1987.3430 | 1984.8240 | 0.0411 | 2096.5760 | 2095.0670 | 0.0262 |
| Max | 2228.6330 | 2224.6070 | 0.8173 | 2439.9440 | 2436.6710 | 0.9725 |
| S.D. | 51.9183 | 51.9494 | 0.0725 | 69.8821 | 69.9892 | 0.0884 |
| Skewness | 0.6635 | 0.6660 | 2.3643 | 0.4622 | 0.4603 | 2.3318 |
| Kurtosis | -0.1271 | -0.1253 | 9.4176 | -0.3317 | -0.3269 | 9.1331 |

From Table 2, we find that the average spread of the lowest ask price between the highest bid price for SSEI and CSI 300 are about 3 CNY and 3.2 CNY, respectively. Meanwhile, the mean of the bid-ask spread for SSEI and CSI 300 are 0.146% and 0.135%, respectively. There is not so much difference. Angel (2000) found the mean of the bid-ask spread for the 30 stocks of Dow Jones Industrial Average is 0.32%. Ahn and Bae (2000) found the mean of the bid-ask spread for the 33 stocks in Hang Seng Index is 0.47%. Therefore, the liquidity of China's stock market is in a relatively high level compared with these two markets. As a kind of mechanism with high efficiency and low cost, the current order driven market is in line with the specific operating environment in China stock market. We also find that the positive skewness and large kurtosis compared with normal density. The standard deviation of the bid-ask spread is lower than the mean, which displays the feature of overdispersion.

*4.2. Model selection and application*

In this subsection, we consider an application to bid-ask spreads for above SSEI and CSI 300. The objective of this empirical application is to provide an illustrative example of the use of the results derived in the previous section. The possibility of calculating the moments that are implied by the estimated parameters allows us also to compare various specifications (MEM, Log-MEM$_1$ and Log-MEM$_2$) and distributions for the baseline bid-ask spread in their ability to fit the sample moments of the data. Thus we first obtain the maximum likelihood estimation (MLE) of parameters, then calculate the moment and dispersion index, last compare these results with the sample moment and dispersion index from the data to select the appropriate model.

Cubic spline method suggested by Engle and Russell (1998) is used to take care of the deterministic component of daily trading activities and to display the intraday seasonal effect. The nodes were set on each hour from 9:30 am to 15:00 pm. Figure 2 gives the plot of the diurnally adjusted bid-ask spread of SSEI and CSI 300. Table 3 gives the basic descriptive statistics for the deseasonalized bid-ask spread of SSEI and CSI 300. For SSEI, the smallest bid-ask spread is 0.2963, the largest bid-ask spread is 5.0590, and the average bid-ask spread is 1.0860. For CSI 300, the smallest bid-ask spread is 0.2085, the largest bid-ask spread is 7.4050, and the average bid-ask spread is 1.1540. We also find that the positive skewness and large kurtosis compared with normal density. The JB test rejects the normal distribution for both bid-ask spreads. The standard deviation of is also lower than the mean, which displays the feature of overdispersion for the deseasonalized bid-ask spreads.

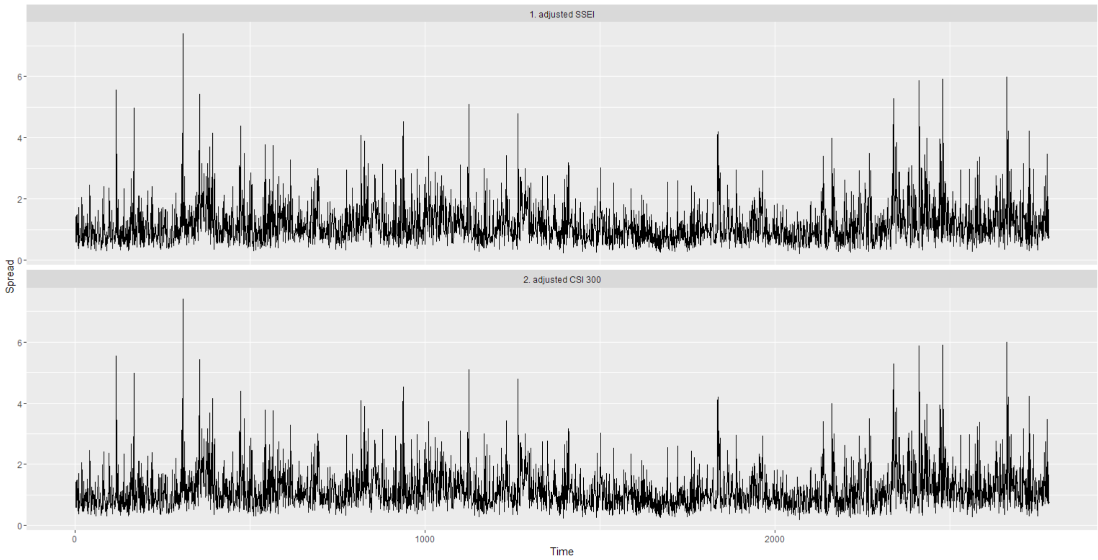

Figure 2. The adjusted bid-ask spreads of SSEI and CSI 300

After removing the deterministic part of intraday bid-ask spread, each deseasonalized bid-ask spread has been estimated by MEM(1,1), Log-MEM$_1$(1,1) and Log-MEM$_2$(1,1), and for each one of these models we can consider a series of distributions for the conditional bid-ask spread, namely: exponential, gamma, generalized gamma, Weibull, Burr, lognormal, Pareto …, in principle any distribution with positive support. The choice of a particular distribution should be guided by the desire to have a “correct” specification, and perhaps by its convenience for estimation. Among these distributions, the Burr and the Pareto do not necessarily have finite moments. The generalized gamma includes the gamma and the Weibull. In order to demonstrate the model selection process more clearly, we just use the exponential, Weibull and gamma distributions. All these distributions depend on a scale parameter that we normalize at 1. Therefore, we have following options on $\varepsilon_t$: $\varepsilon_t \sim \exp(1)$ $\varepsilon_t \sim \text{Weibull}(1, shape)$ and $\varepsilon_t \sim \text{gamma}(shape, 1)$. There are totally nine models to fit each deseasonalized bid-ask spread.

Table 3 Descriptive statistics of the deseasonalized bid-ask spread of SSEI and CSI 300

| | Mean | Min | Max | S.D. | Skewness | Kurtosis | JB test |
|---|---|---|---|---|---|---|---|
| SSEI | 1.0860 | 0.2963 | 5.0590 | 0.4968 | 2.2123 | 8.2795 | 10241 (0.0000) |
| CSI 300 | 1.1540 | 0.2085 | 7.4050 | 0.6911 | 2.2371 | 8.8130 | 11352 (0.0000) |

Table 4 reports the MLE for the deseasonalized bid-ask spread of SSEI and CSI 300. We found the most of parameters are significant at 1% level and a few of parameters are significant at 5% level. It seems that any one model can describe the bid-ask spread. But it is well-known that a significant model can fit data poorly. Which one is the best? We can make decision based on the moment and dispersion index. The MLE for each model, distribution and data series were used to compute the analytical expressions for the unconditional moments, dispersion indices and ACF. The results based on the analytical expressions were then compared with the empirical unconditional moments, dispersion indices and ACF. In order to confirm the feasibility of above simple selection process, the usually used model selection criteria such as AIC and BIC are reported in Table 4 for model comparison. In time series context, it is also important to check if the residuals violate the assumptions imposed on the innovation process. We use Ljung-Box test to show this in Table 4.

Table 4 MLE for the deseasonalized bid-ask spread of SSEI and CSI 300

| | SSEI | | | CSI 300 | | |
|---|---|---|---|---|---|---|
| | exponential | Weibull | gamma | exponential | Weibull | gamma |
| MEM | | | | | | |
| $\omega$ | 0.04014 (0.02762) | 0.03506*** (0.01041) | 0.04014*** (0.01043) | 0.04224* (0.02332) | 0.04959*** (0.01318) | 0.04224*** (0.01170) |
| $\alpha_1$ | 0.08967*** (0.02944) | 0.09906*** (0.01177) | 0.08967*** (0.01111) | 0.07510*** (0.02073) | 0.08256*** (0.01109) | 0.07510*** (0.01040) |
| $\beta_1$ | 0.87340*** (0.04841) | 0.86870*** (0.01809) | 0.87340*** (0.01829) | 0.88840*** (0.03633) | 0.87530*** (0.01947) | 0.88840*** (0.01823) |
| shape | - | 2.38970*** (0.03008) | 7.01370*** (0.1837) | - | 1.88710*** (0.02468) | 3.97250*** (0.1023) |
| AIC | 5978.6320 | 4281.8080 | 3617.4160 | 6302.9150 | 4728.3890 | 4300.6300 |
| BIC | 5996.4270 | 4305.5350 | 3641.1430 | 6320.7100 | 4752.1160 | 4324.3570 |
| LB(10) | 14.0092 | 14.4694 | 14.0092 | 14.8484 | 14.1531 | 14.8495 |
| LB(15) | 18.8864 | 20.7139 | 18.8866 | 16.8720 | 15.7927 | 16.8721 |
| LB(20) | 26.9566 | 28.9647 | 26.9568 | 23.9059 | 22.9144 | 23.9060 |
| Log-MEM$_1$ | | | | | | |
| $\omega$ | 0.01136*** (0.00474) | 0.01062*** (0.00184) | -0.08884*** (0.00906) | 0.01668*** (0.00524) | 0.01788*** | -0.06589*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | (0.00294) | (0.00765) |
| $\alpha_1$ | 0.11190*** | 0.11410*** | 0.09241*** | 0.08582*** | 0.08708*** | 0.07278*** |
| | (0.03026) | (0.01189) | (0.00935) | (0.01994) | (0.01059) | (0.00821) |
| $\beta_1$ | 0.95600*** | 0.96550*** | 0.84170*** | 0.95880*** | 0.95650*** | 0.86170*** |
| | (0.02920) | (0.01057) | (0.01931) | (0.02128) | (0.01157) | (0.01895) |
| shape | - | 2.37150*** | 6.95300*** | - | 1.87630*** | 3.94910*** |
| | | (0.02983) | (0.18210) | | (0.02455) | (0.10170) |
| AIC | 5983.7890 | 3321.8400 | 2642.7470 | 6310.6560 | 3760.0910 | 3318.4320 |
| BIC | 6001.5840 | 3345.5660 | 2666.4730 | 6328.4510 | 3783.8180 | 3342.1580 |
| LB(10) | 14.1336 | 13.7982 | 85.2887 | 16.6118 | 16.3718 | 62.4965 |
| LB(15) | 17.0645 | 17.6303 | 107.2012 | 17.5385 | 17.1889 | 74.0603 |
| LB(20) | 25.2460 | 25.4610 | 136.3525 | 23.9927 | 23.6711 | 91.3439 |
| Log-MEM$_2$ | | | | | | |
| $\omega$ | -0.08450*** | -0.09501*** | -0.08450*** | -0.06776*** | -0.07404*** | -0.06776*** |
| | (0.02674) | (0.01088) | (0.01009) | (0.01825) | (0.00970) | (0.00915) |
| $\alpha_1$ | 0.08735*** | 0.09739*** | 0.08735** | 0.07276*** | 0.08061*** | 0.07276*** |
| | (0.02779) | (0.01118) | (0.01049) | (0.01977) | (0.01054) | (0.00992) |
| $\beta_1$ | 0.96210*** | 0.96610*** | 0.96210*** | 0.96250*** | 0.95590*** | 0.96250*** |
| | (0.02510) | (0.00944) | (0.009473) | (0.02035) | (0.01136) | (0.01021) |
| shape | - | 2.39210*** | 7.02066** | - | 1.88840*** | 3.97500*** |
| | | (0.03012) | (0.1839) | | (0.02470) | (0.10240) |
| AIC | 5978.2200 | 4277.2890 | 3614.5250 | 6302.4390 | 4725.3070 | 4298.7410 |
| BIC | 5996.0150 | 4301.0160 | 3638.2510 | 6320.234 | 4749.0330 | 4322.4670 |
| LB(10) | 13.7354 | 14.6247 | 13.7355 | 14.5182 | 13.9751 | 14.5180 |
| LB(15) | 18.5996 | 20.8357 | 18.5996 | 16.4768 | 15.5350 | 16.4765 |
| LB(20) | 26.7798 | 29.2124 | 26.7798 | 23.5564 | 22.7514 | 23.5562 |

Note: Estimates of the parameters of the MEM(1,1), Log-MEM$_1$(1,1) and Log-MEM$_2$(1,1), assuming various distributions for Log-MEM$_2$ for $\varepsilon_t$. "-" denotes that there are no estimation, standard errors are in parentheses. "**", "***" denote that the parameters are not significantly zero at the 5%, 1% level. The 5% critical values of a Chi-squared distribution with 10, 15 and 20 degrees of freedom are 18.3070, 24.9958 and 31.4104, respectively.

Figure 3 reports the graph of the empirical ACF of the bid-ask spread and the ACF computed from the estimated parameters of various models for SSEI. We observe that the ACF computed from MEM and Log-MEM$_1$ fits the empirical ACF for all distributions considered here. The ACF computed from exponential and gamma distribution almost the same for both models. But for Log-MEM$_2$, there is much difference in the computed ACF from exponential, Weibull and gamma distribution. It is obvious that the computed ACF from exponential and Weibull distribution fits well than that from gamma distribution for Log-MEM$_2$. In order to further select the best model, we compute the theoretical first two moments and dispersion indices resulting from the analytical expressions and then compared with the empirical moments and dispersion index computed from the data for the bid-ask spread.

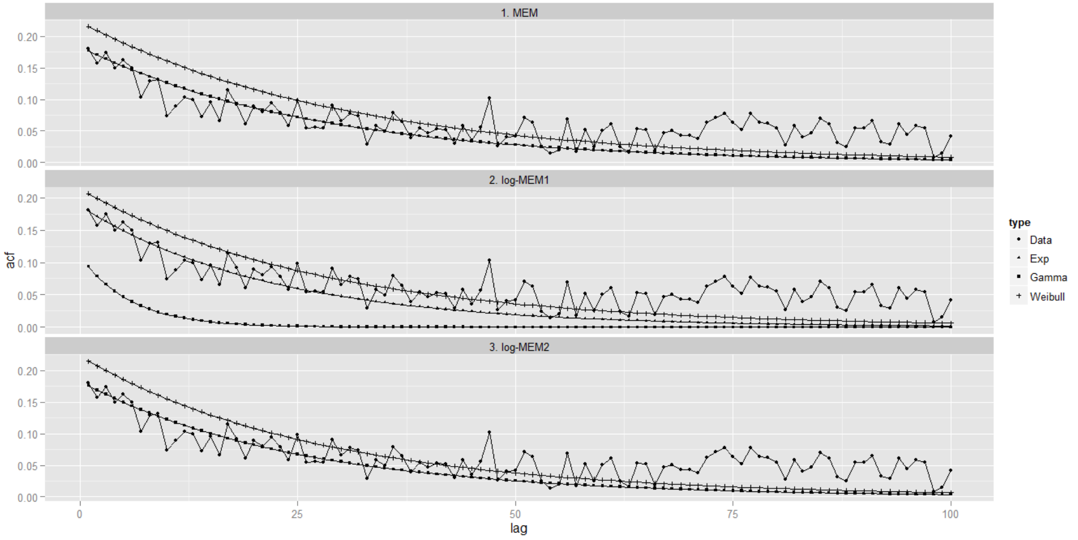


Figure 3. The ACFs of SSEI

Table 5 reports the theoretical first two moments and dispersion indices resulting from the analytical expressions for the three models, the empirical moments and dispersion index computed from the data for the bid-ask spread of SSEI. For all models, the Weibull seems to be the best distribution, followed by the gamma. The ranking of the difference and the sum of ranks seem to show the Weibull Log-MEM$_1$ is the preferable specification, which was reflected in Figure 3. Based on the AIC, BIC and the Ljung-Box test statistics in Table 4, we find that the gamma Log-MEM$_1$ has the smallest AIC and BIC for SSEI, but the Ljung-Box test rejects the model. In other eight models, all the Ljung-Box test statistics are smaller than the critical value. We arrange these models based on the value of AIC and BIC. Compare the rank of the last second columns in Table 5, our findings through moments and dispersion indices are confirmed again.

Table 5 Moments, dispersion indices and ranking results of SSEI

| model | distribution | mean | standard deviation | dispersion index | sum of ranks | rank of AIC, BIC and LB |
|---|---|---|---|---|---|---|
| MEM | exp | 1.08815 (2) | 1.21650 (9) | 1.11795 (9) | 20 | 7 |
| | Weibull | 1.08912 (6) | 0.52971 (3) | 0.48636 (3) | 12 | 5 |
| | gamma | 1.08816 (3) | 0.43657 (5) | 0.40120 (5) | 13 | 3 |
| Log-MEM$_1$ | exp | 1.08175 (7) | 0.15571 (7) | 0.14395 (7) | 21 | 8 |
| | Weibull | 1.08802 (1) | 0.52371 (1) | 0.48401 (2) | 4 | 1 |
| | gamma | 0.70252 (9) | 0.27086 (6) | 0.38555 (6) | 21 | 9 |
| Log-MEM$_2$ | exp | 1.08893 (5) | 0.15362 (8) | 0.14112 (8) | 21 | 6 |
| | Weibull | 1.09180 (8) | 0.52540 (2) | 0.48126 (1) | 11 | 4 |
| | gamma | 1.08863 (4) | 0.44245 (4) | 0.40643 (4) | 12 | 2 |
| Empirical moments | | 1.08580 | 0.49683 | 0.45758 | | |

Note: Unconditional moments and dispersion indices for the MEM(1,1), Log-MEM$_1$(1,1) and Log-MEM$_2$(1,1) computed by applying the analytical expressions with the estimated parameters. The last raw gives the empirical moments and dispersion index computed from the data. The ranking of the difference between the theoretical and the empirical moments and dispersion indices are in parentheses. The last second column gives the sum of the ranks. The last column gives the rank of the nine models based on AIC, BIC and Ljung-Box test.

Figure 4 reports the graph of the empirical ACF of the bid-ask spread and the ACF computed from the estimated parameters of various models for CSI 300. We observe similar pattern with Figure 3. The ACF computed from exponential and gamma distribution almost the same for MEM and Log-MEM$_1$. While for Log-MEM$_2$, the computed ACF from exponential and Weibull distribution fits well than that from gamma distribution.

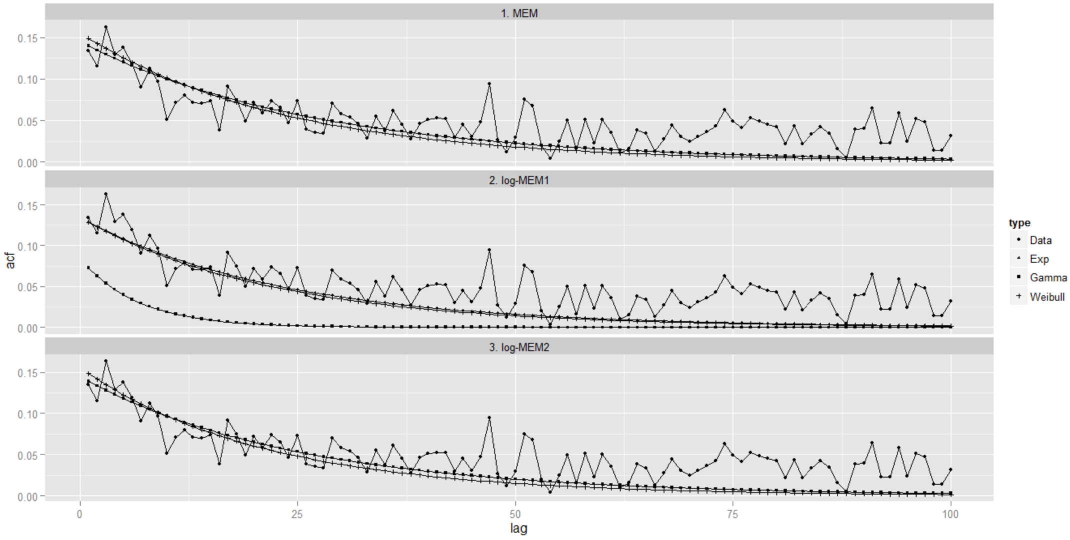


Figure 4. The ACFs of CSI 300

Table 6 reports the theoretical first two moments and dispersion indices resulting from the analytical expressions for the three models, the empirical moments and dispersion index computed from the data for the bid-ask spread of CSI 300. The Weibull Log-MEM$_1$ is also the preferable specification for the bid-ask spread of CSI 300. Based on the AIC, BIC and the Ljung-Box test statistics in Table 4, the gamma Log-MEM$_1$ also has the smallest AIC and BIC for CSI 300, but the Ljung-Box test rejects the model. In other eight models, all the Ljung-Box test statistics are smaller than the critical value. These models are sorted based on the value of AIC and BIC shown at the last column in Table 6. Compare the rank of the last second columns in Table 6, our findings through moments and dispersion indices are also confirmed again.

Table 6 Moments, dispersion indices and ranking results of CSI300

| model | distribution | mean | standard deviation | dispersion index | sum of ranks | rank of AIC, BIC and LB |
|---|---|---|---|---|---|---|
| MEM | exp | 1.15613 (2) | 1.25090 (9) | 1.08197 (9) | 20 | 7 |
| | Weibull | 1.17553 (8) | 0.68230 (1) | 0.58042 (2) | 11 | 5 |
| | gamma | 1.15613 (2) | 0.60848 (5) | 0.52631 (5) | 12 | 3 |
| Log-MEM$_1$ | exp | 1.14842 (6) | 0.17667 (8) | 0.38555 (7) | 21 | 8 |
| | Weibull | 1.15504 (1) | 0.67032 (3) | 0.58034 (3) | 7 | 1 |
| | gamma | 0.73310 (9) | 0.37379 (6) | 0.50987 (6) | 21 | 9 |
| Log-MEM$_2$ | exp | 1.15699 (4) | 0.18447 (7) | 0.15944 (8) | 19 | 6 |
| | Weibull | 1.16354 (7) | 0.67599 (2) | 0.58098 (1) | 10 | 4 |
| | gamma | 1.15699 (5) | 0.61591 (4) | 0.53234 (4) | 13 | 2 |
| Empirical moments | | 1.15387 | 0.69107 | 0.59892 | | |

Note: Unconditional moments and dispersion indices for the MEM(1,1), Log-MEM$_1$(1,1) and Log-MEM$_2$(1,1) computed by applying the analytical expressions with the estimated parameters. The last line gives the empirical moments and dispersion index computed from the data. The ranking of the difference between the theoretical and the empirical moments and dispersion indices are in parentheses. The last second column gives the sum of the ranks. The last column gives the rank of the nine models based on AIC, BIC and Ljung-Box test.

## 5. Conclusions

This paper extends to give the Log-MEMs based on the classical MEM and the analytical formulae for unconditional moments, dispersion index and ACF of the Log-MEM. Using these formulae, we provide an empirical application to select the best model in which we compute the unconditional moments and ACF for the MEM and Log-MEM estimated on bid-ask spread for the high frequency trading data of five minutes of Shanghai Stock Exchange Composite Index and Shanghai and Shenzhen 300 Index in China stock market. Further study may consider constructing the former moment test statistics based on these formulae.

## Reference

Angel, J. J. Tick size, share prices, and stock splits. The Journal of Finance, 1997, 52(2): 655-681.

Ahn, H. J., Bae, K. H., Chan, K. Limit orders, depth, and volatility: Evidence from the stock exchange of Hong Kong. The Journal of Finance, 2001, 56(2): 767-788.

Ahoniemi, K., Lanne, M. Joint modeling of call and put implied volatility. International Journal of Forecasting, 2009, 25(2): 239-258.

Bauwens, L., Galli, F., Giot, P. The moments of Log-ACD models. Available at SSRN 375180, 2003.

Bauwens, L., Giot, P. The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. Annales d'Economie et de Statistique, 2000: 117-149.

Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 1986, 31(3): 307-327.

Brownlees, C. T., Gallo, G. M. Ultra-high frequency data management. Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione (ed. C. Provasi), Atti del Convegno S. Co., CLUEP, Padova, 2005.

Brownlees, C. T., Gallo, G. M. Shrinkage estimation of semiparametric multiplicative error models. International Journal of Forecasting, 2011, 27(2), 365-378.

Brownlees, C. T., Cipollini, F., Gallo, G. M. Multiplicative error models. Available at SSRN 1852285, 2011.

Chou, R. Y. Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. Journal of Money, Credit and Banking, 2005, 561-582.

De Luca, G., Gallo, G. M. Mixture processes for financial intradaily durations. Studies in Nonlinear Dynamics & Econometrics, 2004, 8(2).

Engle, R. F. New frontiers for ARCH models. Journal of Applied Econometrics, 2002, 17(5), 425-446.

Engle, R. F, Gallo, G. M. A multiple indicators model for volatility using intra-daily data. Journal of Econometrics, 2006, 131(1), 3-27.

Engle, R. F., Russell, J. R. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica, 1998, 66(5), 1127-1162.

Engle, R. F., Gallo, G. M., Velucchi, M. An MEM–based investigation of dominance and interdependence across markets. In 3rd World Conference on Computational Statistics and Data Analysis, Limassol, Cyprus, 2005.

Engle, R. F., Gallo, G. M., Velucchi, M. An MEM-based analysis of volatility spillovers in East Asian financial markets. Available at SSRN 1283254, 2008.

Engle, R. F., Fleming, M. J., Ghysels, E., et al. Liquidity and volatility in the US treasury market: evidence from a new class of dynamic order book models. Federal Reserve Bank of New York Working Paper, 2011.

Gallo, G. M., Otranto, E. The Markov switching asymmetric multiplicative error mode. WP CRENoS, 2012, 5.

Gallo, G., Velucchi, M. Ultra-high frequency measures of volatility: an mem–based investigation. Technical report, Dipartimento di Statistica 'G. Parenti', Universita di Firenze. WP 2007-01, 2007.

Giovannetti, G., Velucchi, M. A MEM analysis of African financial markets. New Perspectives in Statistical Modeling and Data Analysis. Springer Berlin Heidelberg, 2011: 319-328.

Hamilton, J. D. Time series analysis. Princeton: Princeton university press, 1994.

Han, H., Park, M. D., Zhang, S. A multiplicative error model with heterogeneous components for forecasting realized volatility. Journal of Forecasting, 2015, 34(3), 209-219.

Lam, K. P., & Ng, H. S. Intra-daily information of range-based volatility for MEM-GARCH. Mathematics and Computers in Simulation, 2009, 79(8), 2625-2632.

Lanne M. A mixture multiplicative error model for realized volatility. Journal of Financial Econometrics, 2006, 4(4): 594-616.

Nelson, D. B. Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 1991, 59(2), 347-370.

Otranto, E. Classification of volatility in presence of changes in model parameters. WP CRENoS, 2011, 13.

## Authors

**Li Li**

Chengdu Primary School Affiliated to Beijing International Studies University, Chengdu 610041, China.

**Wanbo Lu***

School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China, luwb@swufe.edu.cn

**Rui Ke**

School of Economics, Hefei University of Technology, Hefei 230009, China

*corresponding author

**Acknowledgements**

Wanbo Lu's research is sponsored by the National Science Foundation of China (71101118, 71771187) and the Program for New Century Excellent Talents in University (NCET-13-0961) and the Fundamental Research Funds for the Central Universities (JBK190602) in China.