Finance Research Letters xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Estimation accuracy of high–low spread estimator

Chien-Chih Lin *

Department of Banking and Finance, Tamkang University, 151, Ying-Chuan Rd., Administration Building, B820, Tamsui, New Taipei City 25137, Taiwan

ARTICLE INFO

Article history:
Received 30 March 2013
Accepted 14 May 2013
Available online xxxx

JEL classification:
C02
C13
G14

Keywords:
High–low spread estimator
Estimation accuracy
Effective spread
Measurement error

ABSTRACT

In this paper we analyze the estimation accuracy of high–low spread estimator. It is found that the performance of high–low spread estimator depends on the size of the true spread, the level of transaction frequency, and the degree of volatility. Analyzing the probability of measurement error, it is shown that the high–low spread estimators have better performance when the size of the spread is even wider, when the level of transaction frequency is even higher, or when the degree of volatility is relatively lower.

1. Introduction

Given that the first-order serial covariance of price changes is due only to the covariance induced by the spread, Roll (1984) developed a spread estimator that has widely been applied in financial markets. One of the main advantages of Roll's measure is the data merely requiring transaction prices, which is the most basic information in financial market. Another advantage is the ease of computation; Roll's estimator can be easily measured by $2\sqrt{-Cov}$, where "*Cov*" is the first-order serial covariance of price changes.

Although Roll's estimator has the advantages referred above, its performance is unstable when spreads are estimated with low frequency data. With a dataset of stocks returns over 1963–1982, Roll found a high proportion of daily and weekly spread estimates are negative. In addition, the weekly spread estimates were significantly greater than the daily spread estimates.

* Tel.: +886 22621 5656 2859; fax: +886 22621 4755.
E-mail address: 139714@mail.tku.edu.tw

1544-6123/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.frl.2013.05.004

Thereafter, a series of papers explored Roll's problematic empirical results. According to Garbade and Lieber (1977), transaction types tend to cluster over short intervals of time, Choi et al. (1988) extended Roll's formula for the effective bid-ask spread by incorporating the possibility of serial correlation in the transaction type, and showed that if there is positive serial correlation in transaction type, then Roll's estimator would be downward-biased. On the other hand, incorporating the effect of asymmetric information, Glosten (1987) argued that the bid-ask spread can be decomposed into two parts: one due to asymmetric information, and the other due to other factors which affect the properties of the transaction-price process differently; the spread proposed by Roll estimates the total spread only when there is no adverse-selection spread. Considering the effect of price reversal, Stoll (1989) developed a more general model to estimate three components of spread: adverse information cost, order processing cost, and inventory cost; showing that Roll's measure only reflects order processing cost (other important studies include George et al., 1991; Holden, 2009; Hasbrouck, 2009).

In contrast to the covariance model which implies a bid-ask spread from the serial dependence of observed prices, Corwin and Schultz (2012) proposed a new approach to develop a spread estimator from daily high and low prices. Their high–low spread estimator is based on a simple insight: Daily high (low) prices are almost always buy (sell) trades, and hence, the high–low ratio reflects bid-ask spreads. Their simulations reveal that the high–low spread estimator has a number of advantages over the daily estimators used in prior research. This paper expands upon this study to explore the influence of the size of the true spread, the level of transaction frequency, and the degree of volatility on the performance of the high–low spread estimator. Analyzing the probability of measurement error, it is shown that the high–low spread estimators have better performance when the size of the spread is even wider, when the level of transaction frequency is even higher, or when the degree of volatility is relatively lower.

The remainder of this paper is organized as follows. Section 2 derives the high–low spread estimators. Sections 3 and 4 demonstrate the simulation procedures and results, respectively. Some explanations and discussions of the simulation results are provided in Section 5. Section 6 presents the study conclusions and implications for future research.

2. The high–low spread estimator

In this study, the model proposed by Corwin and Schultz's (2012) is modified to analyze the estimation accuracy of high–low spread estimator. Suppose the bid-ask spread is $S\%$ of the true value of the asset price; hence the bid price is lower than the true value by $S/2\%$, while the ask price is higher than the true value by $S/2\%$. A trading day is equally divided into n sub-periods: $[t_0, t_1], [t_1, t_2], \ldots, [t_{n-1}, t_n]$, where t_0 and t_n are the opening time and closing time of the market respectively. Hence each sub-period is equal to $1/n$ trading day. In addition, it is assumed that each relatively high price[1] is a buy trade, while each relatively low price is a sell trade. Therefore, in each sub-period $[t_{k-1}, t_k]$ the highest possible ask price $H^o_{t,k}$ is higher than the true value by $S/2\%$, while the lowest possible bid price $L^o_{t,k}$ is lower than the true value by $S/2\%$. With $H^A_{t,k}$ ($L^A_{t,k}$) denoting the highest (lowest) true value over the sub-period $[t_{k-1}, t_k]$, yields

$$\left[\ln(H^o_{t,k}/L^o_{t,k})\right]^2 = \left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right]^2 + 2\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right]\left[\ln\left(\frac{2+S}{2-S}\right)\right] + \left[\ln\left(\frac{2+S}{2-S}\right)\right]^2. \tag{1}$$

Taking expectation of (1) yields

$$E\left[\ln\left(H^o_{t,k}/L^o_{t,k}\right)\right]^2 = E\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right]^2 + 2\alpha E\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right] + \alpha^2 \tag{2}$$

where

[1] A relatively high (low) price is a price higher (lower) than the prices immediately before and after it.

Please cite this article in press as: Lin, C.-C. Estimation accuracy of high–low spread estimator. Finance Research Letters (2013), http://dx.doi.org/10.1016/j.frl.2013.05.004

$$\alpha = \left[\ln\left(\frac{2+S}{2-S}\right)\right]. \tag{3}$$

In order to compute the expectation in the right hand side of (2), the true asset price is assumed following a geometric Brownian motion. Appling the results of Parkinson (1980) and Garman and Klass (1980) yields

$$E\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right]^2 = \frac{k_1}{n}\sigma^2 \tag{4}$$

and

$$E\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right] = \frac{k_2}{\sqrt{n}}\sigma, \tag{5}$$

where $k_1 = 4\ln 2$, $k_2 = \sqrt{8/\pi}$, and σ denotes the volatility of the true price. Substituting (4) and (5) into (2) yields

$$\frac{1}{n}k_1\sigma^2 + \frac{2}{\sqrt{n}}k_2\sigma\alpha + \alpha^2 - \beta_n = 0 \tag{6}$$

where $\beta_n = E[\ln(H^o_{t,k}/L^o_{t,k})]^2$. Note that β_n can be estimated from sample by taking the highest observed price as the estimate of the highest possible ask price and lowest observed price as the estimate of the lowest possible bid price. Appendix A provides a law of large numbers for $E[\ln(H^o_{t,k}/L^o_{t,k})]^2$ to yield consistency.

To solve the other two unobserved parameters σ and α, another equation is required. Squaring the log high–low ratio over a 2-intraday sub-period yields

$$\left[\ln\left(H^o_{t,k,k+1}/L^o_{t,k,k+1}\right)\right]^2 = \left[\ln\left(\frac{H^A_{t,k,k+1}}{L^A_{t,k,k+1}}\right)\right]^2 + 2\left[\ln\left(\frac{H^A_{t,k,k+1}}{L^A_{t,k,k+1}}\right)\right]\left[\ln\left(\frac{2+S}{2-S}\right)\right] + \left[\ln\left(\frac{2+S}{2-S}\right)\right]^2 \tag{7}$$

where $H^o_{t,k,k+1}(L^o_{t,k,k+1})$ is the highest (lowest) possible ask (bid) price in two sub-periods (i.e. the sub-periods $[t_{k-1}, t_k]$ and $[t_k, t_{k+1}]$), while $H^A_{t,k,k+1}(L^A_{t,k,k+1})$ denotes the highest (lowest) true price over two sub-periods. Considering expectations in (7) yields

$$\frac{2}{n}k_1\sigma^2 + 2\sqrt{\frac{2}{n}}k_2\sigma\alpha + \alpha^2 - \gamma_n = 0 \tag{8}$$

where $\gamma_n = E\{[\ln(H^o_{t,k,k+1}/L^o_{t,k,k+1})]^2\}$.

Together with (6) and (8) α can be solved. Neglecting Jensen's inequality (see Corwin and Schultz's, 2012), a simple transformation of α in (3) then provides a spread estimator, as follows

$$S = \frac{2(e^\alpha - 1)}{1 + e^\alpha}, \tag{9}$$

where

$$\alpha = \frac{\sqrt{4\beta_n} - \sqrt{2\beta_n}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma_n}{3 - 2\sqrt{2}}}. \tag{10}$$

3. Simulation procedures

To see how well the high–low spread estimator works under different scenarios, T days of prices data were simulated. Each day contains 300 min and each minute uniformly generates m prices; hence the transaction price can be monitored every $60/m$ s. At the beginning of the first trading day, the asset price is arbitrarily set to \$100. Then at each transaction time s of trading day t, the true price of the asset $P^A_{t,s}$ is simulated as

Please cite this article in press as: Lin, C.-C. Estimation accuracy of high–low spread estimator. Finance Research Letters (2013), http://dx.doi.org/10.1016/j.frl.2013.05.004

$$P^A_{t,s} = P^A_{t,s-\Delta s}\exp(\mu\Delta s + \sigma Z_{t,s}\sqrt{\Delta s}), \quad (11)$$

where μ is the daily expected return of the asset, σ denotes its daily volatility, $\Delta s = 1/(300m)$ day, and $\{Z_{t,s}\}$ are i.i.d. standard normal random variables. The bid (ask) price is then obtained by multiplying $P^A_{t,s}$ by one minus (plus) half the bid-ask spread. It is assumed that there is a 50% chance that the observed price is a bid, and a 50% chance that it is an ask. Hence the observed price $P^o_{t,s}$ is simulated as

$$P^o_{t,s} = (1+\delta_{t,s})P^A_{t,s} \quad (12)$$

where $\{\delta_{t,s}\}$ are independent random variables, such that

$$\Pr\left(\delta_{t,s} = \frac{+S}{2}\right) = \Pr\left(\delta_{t,s} = \frac{-S}{2}\right) = \frac{1}{2}. \quad (13)$$

Each trading day t was split into n sub-periods: $[0,1/n],[1/n,2/n],\ldots,[(n-1)/n,1]$. In each sub-period, the observed highest prices and lowest prices are computed as follows:

$$H^o_{t,k} = \max\{P^o_{t,s}|s \in [(k-1)/n, k/n]\}, \quad (14)$$

$$L^o_{t,k} = \min\{P^o_{t,s}|s \in [(k-1)/n, k/n]\}, \quad (15)$$

where $k = 1, 2, \ldots, n$. Using the data generated from (14) and (15), β_n and γ_n can be estimated by $\widehat{\beta_{t,n}}$ and $\widehat{\gamma_{t,n}}$ as follows:

$$\widehat{\beta}_{t,n} = \frac{1}{n}\sum_{k=1}^{n}[\ln(H^o_{t,k}/L^o_{t,k})]^2. \quad (16)$$

$$\widehat{\gamma_{t,n}} = \frac{1}{n-1}\sum_{k=1}^{n-1}[\ln(H^o_{t,k,k+1}/L^o_{t,k,k+1})]^2 = \frac{1}{(n-1)}\sum_{k=1}^{n-1}\left[\ln\left(\frac{\max(H^0_{t,k}, H^0_{t,k+1})}{\min(L^0_{t,k}, L^0_{t,k+1})}\right)\right]^2. \quad (17)$$

Substituting $\widehat{\beta_{t,n}}$ and $\widehat{\gamma_{t,n}}$ into Eq. (10) for β_n and γ_n, respectively, provides an estimate of α:

$$\widehat{\alpha_t} = \frac{\sqrt{4\widehat{\beta_{t,n}}} - \sqrt{\widehat{\beta_{t,n}}}}{3-2\sqrt{2}} - \sqrt{\frac{\widehat{\gamma_{t,n}}}{3-2\sqrt{2}}}. \quad (18)$$

Substituting $\widehat{\alpha_t}$ into (9), yields an spread estimate of trading day t

$$\widehat{S_t} = \frac{2(e^{\widehat{\alpha_t}}-1)}{1+e^{\widehat{\alpha_t}}}. \quad (19)$$

Calculating the average value of spread estimates yields a spread estimator

$$\widehat{S} = \frac{1}{T}\sum_{t=1}^{T}\widehat{S_t}. \quad (20)$$

4. Simulation results

The above described simulation procedures are repeated to create N spread estimates: $\widehat{S(1)}, \widehat{S(2)}, \ldots, \widehat{S(N)}$. In order to study the estimation accuracy of the high–low spread estimator, the percentage error is calculated as

$$\text{Error} = \frac{1}{N}\sum_{j=1}^{N}|\widehat{S}(j) - S|/S. \quad (21)$$

Fig. 1 is constructed to analyze the impact of the size of true spread on the estimation accuracy of high–low spread estimators. The dotted line depicts a situation with a narrow true spread ($S = 0.5\%$); the dashed line is a case of wider spread ($S = 1.0\%$); and the solid line is a situation of even

Please cite this article in press as: Lin, C.-C. Estimation accuracy of high–low spread estimator. Finance Research Letters (2013), http://dx.doi.org/10.1016/j.frl.2013.05.004

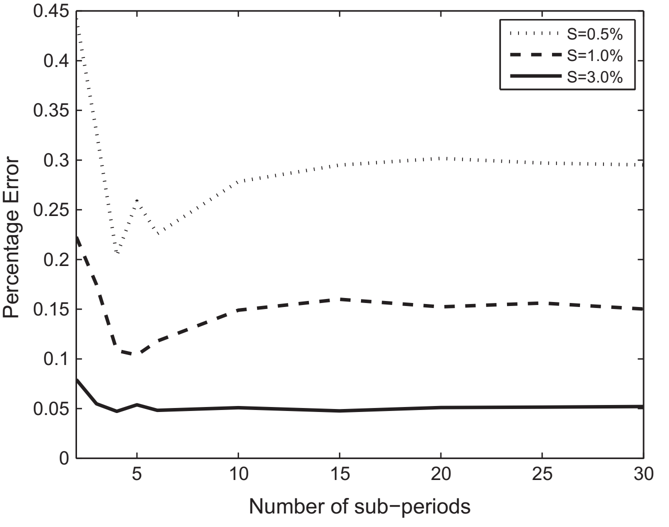

Fig. 1. The estimation error of high–low spread estimators for different sizes of spreads. (The values of parameters: $N = 30$, $m = 6$, $\mu = 0$, $\sigma = 0.03$, $T = 48$.)

wider spread ($S = 3.0\%$). It is clear that the dashed line is right below the dotted line, while the solid line is right below the dashed line, indicating that a wider true spread leads to a lower estimation error, and hence higher estimation accuracy.

To gain some insight into the estimation accuracy of the effect of transaction frequency, Fig. 2 depicts the estimation error for different levels of transaction frequency. The dotted line depicts a situation when the transaction frequency is slightly high ($m = 6$), the dashed line a situation with higher frequency ($m = 12$), and the solid line an even more extreme situation ($m = 30$). As one can see, the

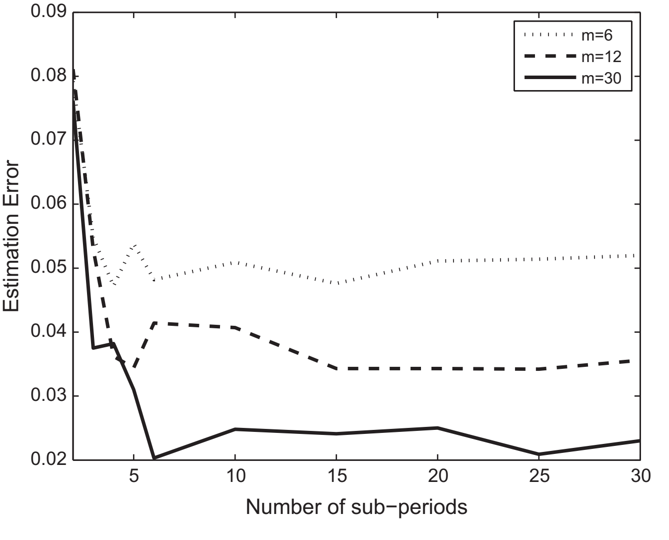

Fig. 2. The estimation error of high–low spread estimators for different levels of transaction frequency. (The values of parameters: $N = 30$, $m = 6$, $\mu = 0$, $\sigma = 0.03$, $T = 48$, $S = 3\%$.)

Please cite this article in press as: Lin, C.-C. Estimation accuracy of high–low spread estimator. Finance Research Letters (2013), http://dx.doi.org/10.1016/j.frl.2013.05.004

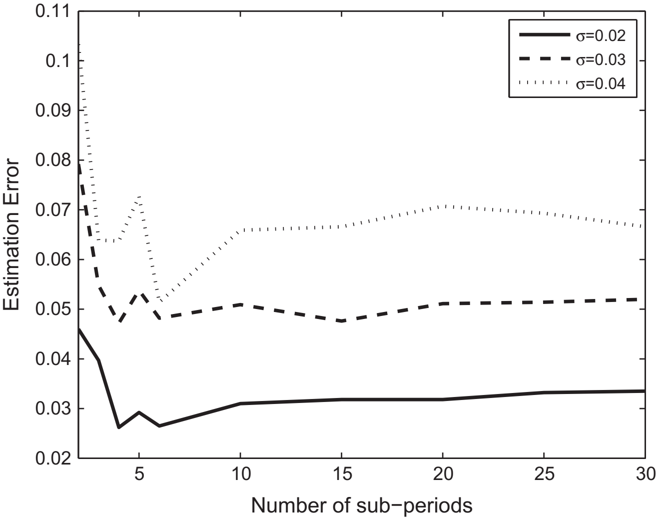

Fig. 3. The estimation error of high–low spread estimators for different degrees of volatility. (The values of parameters: $N = 30$, $m = 6$, $\mu = 0$, $\sigma = 0.03$, $T = 48$, $S = 3\%$.)

dashed line generally lies below the dotted line, while the solid line lies below the dashed line, indicating that, as the level of transaction frequency increases, the estimation error of the high–low spread estimators is reduced. As a consequence, the estimation accuracy is improved.

In order to examine the influence on the estimation accuracy of the high–low spread estimator for different degrees of price volatility, Fig. 3 depicts the estimation error for different degrees of volatility. The solid line presents a situation when the degree of volatility is slightly high ($\sigma = 0.02$), the dashed line a case with higher level of volatility ($\sigma = 0.03$), and the dotted line an even more extreme situation ($\sigma = 0.04$). It is obvious to see that the dashed line lies above the solid line, while the dotted line lies above the dashed line, indicating that a higher level of volatility leads to a worse estimation error. Consequently, the lower level of price volatility leads to the higher estimation accuracy.

5. Explanations of the effect of parameters

This section sheds light on why the high–low spread estimators are more accurate when the true spread is even wider, when the level of transaction frequency is even higher, or when the degree of volatility is relatively lower.

The spread estimator derived here provides a method for inferring the true spread from the high and low prices of each sub-period. This method requires one major assumption: *Relatively high prices are buy trades, while relatively low prices are sell trades.* However, in practice it is possible that an observed relatively high price is a seller-initiated trade, and is therefore discounted by half of the spread, while an observed relatively low price is a buyer-initiated trade and is therefore grossed up by half of the spread.

In order to analyze the influences of the size of the true spread, the level of transaction frequency, and the degree of volatility on the accuracy of the high–low spread estimator, the probability of one type of measurement error occurring from a relatively high price lower than the actual values by half of the spread is computed, which is referred to as the following proposition.

Proposition. *Suppose that $P^o_{t,s}$ is a relatively high price, then the probability of $P^o_{t,s} = P^A_{t,s}(1 - S/2)$ is given by*

$$P(P^o_{t,s} = P^A_{t,s}(1 - S/2)|P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}) = \frac{(1/2 + N(-d))^2}{(1/2 + N(d))^2 + (1/2 + N(-d))^2} \quad (22)$$

where $N(\cdot)$ *is the cumulative standard normal distribution function and* $d = \frac{1}{\sigma\sqrt{\Delta s}} \ln\left[\frac{2+S}{2-S}\right]$.

Proof. See Appendix B □.

Remark 1. Similar to the proof in Appendix B, the probability of the other type of measurement error occurred from a relatively low price, which is higher than the actual values by half of the spread, can be shown to be exactly equal to the probability of a relatively high price lower than the actual values by half of the spread.

Remark 2. Rewriting (22), one has

$$P(P^o_{t,s} = P^A_{t,s}(1 - S/2)|P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}) = \frac{1}{\left(\frac{1/2+N(d)}{1/2+N(-d)}\right)^2 + 1}. \quad (23)$$

As the size of S increases, or the degree of σ decreases, both lead to a higher value of $N(d)$ (lower value of $N(-d)$), and hence a lower probability of a relatively high price grossed down by half of the spread. Consequently, the probability of the occurrence of measurement error becomes lower. As a result, the high–low spread estimator is more accurate when the true spread is even wider, or when price volatility is relatively lower. Similarly, as the level of transaction frequency is increased, the value of Δs is decreased. As a result, the estimation accuracy is improved.

6. Conclusions

In this study the estimation accuracy of high–low spread estimator is analyzed. We found the performance of high–low spread estimator depending on the size of the true spread, the level of transaction frequency, and the degree of volatility. Analyzing the probability of measurement error of high–low spread estimator, it is shown that the estimation error is reduced when the size of the spread is even wider, when the level of transaction frequency is even higher, or when the degree of volatility is relatively lowers.

Although the accuracy of the high–low spread estimator has been discussed in detail, empirical research is still needed to provide further evidence to support the analysis results described herein. The difficulty of the empirical work would be to construct a proxy for the unobservable effective spread. A possible proxy is the difference between the transaction price and the average of the bid and ask prices which are observed immediately before the transaction price. A more detailed empirical analysis can be resolved by future work.

Acknowledgments

I highly appreciate the helpful comments from the Editor-in-Chief and the anonymous referee. In addition, many thanks for the financial support provided from the National Science Council (NSC 101-2410-H-032-024).

Appendix A

To obtain a law of large numbers for $E[\ln(H^o_{t,k}/L^o_{t,k})]^2$ to yield consistency, it is sufficient to show (see Eq. (2))

$$\frac{1}{M}\sum_{t,k}\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right] \xrightarrow{P} E\left[\ln(H^A_{t,k}/L^A_{t,k})\right] \quad (A1)$$

$$\frac{1}{M}\sum_{t,k}\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right]^2 \xrightarrow{P} E\left[\ln\left(H^A_{t,k}/L^A_{t,k}\right)\right]^2 \tag{A2}$$

where M is the number of observations. Together with (4), (5), and $\{\ln(H^A_{t,k}/L^A_{t,k})\}$ being independent identically distributed, it is straightforward to obtain Eq. (A1).

Applying the results of Parkinson (1980), there exists a constant C such that

$$Var\left[\frac{1}{M}\sum_{t,k}\left[\ln(\frac{H^A_{t,k}}{L^A_{t,k}})\right]^2\right] < \frac{C}{M}. \tag{A3}$$

Using Chebyshev's inequality on $1/M\sum_{t,k}\left[\ln\left(H^A_{t,k}/L^A_{t,k}\right)\right]^2$ results in

$$P\left(\left|\frac{1}{M}\sum_{t,k}\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right]^2 - E\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right]^2\right| \geqslant \varepsilon\right)$$

$$\leqslant \frac{1}{\varepsilon^2} Var\left[\frac{1}{M}\sum_{t,k}\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right]^2\right] = \frac{C}{\varepsilon^2 M} \tag{A4}$$

Eq. (A4) can be used to obtain the following:

$$P\left(\left|\frac{1}{M}\sum_{t,k}\left[\ln\left(\frac{H^A_{t,k}}{L^A_{t,k}}\right)\right]^2 - E\left[\ln\left(\frac{H^o_{t,k}}{L^o_{t,k}}\right)\right]^2\right| < \varepsilon\right) \geqslant 1 - \frac{C}{\varepsilon^2 M}. \tag{A5}$$

As M approaches infinity, the expression approaches 1. And by definition of convergence in probability, I have obtained Eq. (A2).

Appendix B

This appendix computes the probability of a relatively high price discounted by half of the spread. By Bayes' theorem, one has

$$\begin{aligned} &P(P^o_{t,s} = P^A_{t,s}(1-S/2)|P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}) \\ &\quad = P\left(P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}\middle| P^o_{t,s} = P^A_{t,s}(1-S/2))P(P^o_{t,s} = P^A_{t,s}(1-S/2)\right) \\ &\quad\quad \div \left\{P(P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}\middle| P^o_{t,s} = P^A_{t,s}(1-S/2))P(P^o_{t,s} = P^A_{t,s}(1-S/2))\right. \\ &\left. + \left(P(P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}\middle| P^o_{t,s} = P^A_{t,s}(1+S/2))P(P^o_{t,s} = P^A_{t,s}(1+S/2)\right)\right\}. \end{aligned} \tag{B1}$$

Since $P(P^o_{t,s} = P^A_{t,s}(1-S/2)) = P(P^o_{t,s} = P^A_{t,s}(1+S/2)) = 1/2$, (B1) can be simplified as

$$\begin{aligned} &P(P^o_{t,s} = P^A_{t,s}(1-S)|P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}) \\ &= \frac{P(P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}\big| P^o_{t,s} = P^A_{t,s}(1-S/2))}{P(P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}\big| P^o_{t,s} = P^A_{t,s}(1-S/2)) + P(P^o_{t,s} > P^o_{t,s-\Delta s} \wedge P^o_{t,s} > P^o_{t,s+\Delta s}\big| P^o_{t,s} = P^A_{t,s}(1+S/2))} \end{aligned} \tag{B2}$$

$$\equiv \frac{P_I}{P_I + P_{II}}. \tag{B3}$$

To deal with P_I, first note that

$$P_I = P\left(\frac{P^A_{t,s}}{P^A_{t,s-\Delta s}} > \frac{1+\delta_{t,s-\Delta s}}{1-S/2} \wedge \frac{P^A_{t,s+\Delta s}}{P^A_{t,s}} < \frac{1-S/2}{1+\delta_{t,+\Delta s}}\right). \tag{B4}$$

Please cite this article in press as: Lin, C.-C. Estimation accuracy of high–low spread estimator. Finance Research Letters (2013), http://dx.doi.org/10.1016/j.frl.2013.05.004

Neglecting the effect of drift term ($\mu = 0$), from Eq. (11), one has

$$P_I = P\left(\sigma\sqrt{\Delta s}Z_{t,s} < \ln\left[\frac{1 - S/2}{1 + \delta_{t,s-\Delta s}}\right] \wedge \sigma\sqrt{\Delta s}Z_{t,s+\Delta s} < \ln\left[\frac{1 - S/2}{1 + \delta_{t,+\Delta s}}\right]\right). \tag{B5}$$

By the independence between $(Z_{t,s}, \delta_{t,s-\Delta s})$ and $(Z_{t,s+\Delta s}, \delta_{t,s+\Delta s})$, yields

$$\begin{aligned} P_I &= P\left(\sigma\sqrt{\Delta s}Z_{t,s} < \ln\left[\frac{1 - S/2}{1 + \delta_{t,s-\Delta s}}\right]\right)P\left(\sigma\sqrt{\Delta s}Z_{t,s+\Delta s} < \ln\left[\frac{1 - S/2}{1 + \delta_{t,s+\Delta s}}\right]\right) \\ &= \left\{P(\delta_{t,s-\Delta s} = -S/2)P(Z_{t,s} < 0) + P(\delta_{t,s-\Delta s} = +S/2)P\left(Z_{t,s} < \frac{\ln\left[(2 - S)/(2 + S)\right]}{\sigma\sqrt{\Delta s}}\right)\right\} \\ &\quad \times \left\{P(\delta_{t,s+\Delta s} = -S/2)P(Z_{t,s+\Delta s} < 0) + P(\delta_{t,s+\Delta s} = +S/2)P\left(Z_{t,s+\Delta s} < \frac{\ln[(2 - S)/(2 + S)]}{\sigma\sqrt{\Delta s}}\right)\right\} \\ &= \left\{\frac{1}{4} + \frac{1}{2}N\left(\frac{\ln[(2 - S)/(2 + S)]}{\sigma\sqrt{\Delta s}}\right)\right\}^2 \end{aligned}$$

Similarly, P_{II} can be computed as

$$P_{II} = \left\{\frac{1}{4} + \frac{1}{2}N\left(\frac{\ln[(2 + S)/(2 - S)]}{\sigma\sqrt{\Delta s}}\right)\right\}^2. \tag{B6}$$

Substituting (B5) and (B6) into (B2); consequently, Eq. (22) follows.

References

Choi, J.Y., Salandro, D., Shastri, K., 1988. On the estimation of bid-ask spreads: theory and evidence. J. Finance Quant. Anal. 23, 219–230.

Corwin, S.A., Schultz, P., 2012. A simple way to estimate bid-ask spreads from daily high and low prices. J. Finance 67, 719–760.

Garbade, K., Lieber, Z., 1977. On the independence of transactions on the New York Stock Exchange. J. Bank. Finance 1, 151–172.

Garman, M.B., Klass, M.J., 1980. On the estimation of security price volatilities from historical data. J. Business 53, 67–78.

George, T., Kaul, G., Nimalendran, M., 1991. Estimation of the bid-ask spread and its components: a new approach. Rev. Finance Stud. 4, 623–656.

Glosten, L., 1987. Components of the bid-ask spread and the statistical properties of transaction prices. J. Finance 42, 1293–1307.

Hasbrouck, J., 2009. Trading costs and returns for US equities: estimating effective costs from daily data. J. Finance 64, 1445–1477.

Holden, C., 2009. New low-frequency spread measures. J. Finance Mark. 12, 778–813.

Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. J. Business 53, 61–65.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. J. Finance 39, 1127–1140.

Stoll, H., 1989. Inferring the components of the bid-ask spread: theory and empirical tests. J. Finance 44, 115–134.

Please cite this article in press as: Lin, C.-C. Estimation accuracy of high–low spread estimator. Finance Research Letters (2013), http://dx.doi.org/10.1016/j.frl.2013.05.004