QUANTITATIVE METHODS IN ECONOMICS
Vol. XV, No. 2, 2014, pp. 135 –143

ON LOW-FREQUENCY ESTIMATION OF BID-ASK SPREAD IN THE STOCK MARKET

Marek Andrzej Kociński
Department of Applied Mathematics
Warsaw University of Life Sciences – SGGW
e-mail: marek_kocinski@sggw.pl

Abstract: In the article two popular low-frequency methods od bid-ask spread estimation are presented and applied to the stocks quoted on the Warsaw Stock Exchange (WSE): the Roll method [Roll 1984] and Corwin-Schultz method [Corwin and Schultz 2012]. The widely available data on average spreads published by WSE are used as benchmark and proxy of information, usually received from difficult to access and limited high-frequency financial data

Keywords: high-frequency data, low-frequency data, Roll estimator, Corwin-Schultz estimator

INTRODUCTION

Transaction costs are any payments for an opportunity of exchange of the good from one party of the transaction to the other. They are inherently part of a market microstructure. Estimation and analysis of transaction costs are important for portfolio managers, regulators and scholars. The costs of transaction executions are significant determinant of the investment performance since the net profits can be substantially reduced by these costs. Though a trading cost is often a small fraction of the value of a single transaction, in long-time horizon such expenses can considerably lower the return attained by the implementation of the investment strategy, especially when many purchases or sales are required. This is particularly important in the area of algorithmic and high-frequency trading which has been developing fast in recent years. Whether transaction costs are handled effectively or not by the asset manager can thus make all difference in possibility to outperform a given benchmark, for example a stock index. Moreover the transaction costs are in practice the measure of liquidity which is the essential concept in financial theory and practice, since it represents the ease with which financial instruments can be traded. Monitoring liquidity risk, which means monitoring the amount of transaction costs, is one of the main tasks of financial regulators, particularly after the crisis of 2008. It was the time that due to the shortfalls of liquidity, some funds were not in able to redeem their units. The funds operating in the money market were particularly affected, because they methods of investing required high degree of liquidity.

Typically, in investment management, three major sources of transaction costs are taken into consideration: commissions and other direct fees, bid-ask spreads and market impact [Elton et al. 2010], [Sharpe et al. 1999]. The expenses such as broker commissions, taxes, and other fees are direct and predictable therefore they usually are not included in the quantitative analysis of transaction costs. Market impact refers to the effect of the order on the price of the traded security. This consists in inducing by the order the movement of the price against the order maker that is upward when buying and downward when selling. Thus, it is a source of transaction costs. The market impact is a complex phenomenon which, despite numerous articles on this subject, is not explained and quantified in a satisfactory manner. However, in practice it influences only very large transactions and can be neutralized by splitting the order and executing it incrementally over a longer period. The most important source of the transaction costs appears to be the bid-ask spread and estimation of this cost based on easy available financial data is the subject of this study. The bid-ask spread is the difference between the highest bid price and the lowest ask price for the stock; it measures the loss from buying a share of stock and then immediately selling it. It is a common measure of market liquidity and costs incurred by investors [Anand and Karagozoglu 2006]. Some authors include the bid-ask spread in market impact, given that it is a market phenomenon, in opposition to commissions and taxes [Grinold and Kahn 2000]. Bid–ask spreads are so important in market microstructure that researchers must have reliable estimators for spreads if they are to have substantial research on financial markets. Additionally, with the development of electronic trading in stock markets, investors, exchanges, and regulators are likely to be interested in precise bid–ask spread estimation methods. An accurate bid–ask estimator can help to compare the transaction costs in two competing markets. Moreover, the reliable spread estimators could help to optimize the investment strategies. The aim of this paper is to examine the possibility of bid-ask spread estimation based on low-frequency data from Warsaw Stock Exchange. The transaction costs in the polish stock market were consider before in the literature, see, for example [Olbryś 2012] or [Gniadkowska 2012] but it seems that there is no study on estimation of bid-ask spread based on easy available data for polish stocks, though the methods of estimation are present in financial theory at least from the work of [Roll 1984]. Thus, in my opinion, this article may be considered as the contribution to fill the gap in studies on the structure of the polish stock market.

LOW-FREQUENCY VERSUS HIGH-FREQUENCY DATA

In recent years the growing popularity of high-frequency data has been observed. It is driven by the rapid development of computer and information technologies which enable to effectively handle with enormous sets of data. The average daily number of number of observations of an actively traded NYSE stock can be higher than 20000 [Zivot 2005]. Transaction or quotes data observed in time periods shorter than once a day are called high-frequency data, in practice the length of time interval between consecutive moments of observations is very short and measured in seconds. The full information on market prices contains tick-by-tick data where tick refers to the change in the price of a security from trade to trade. The tick-by-tick data are the highest possible frequency data. In the extreme case, when the analyst has access to full record of characteristics of each transaction then one can talk about ultra-high-frequency data. The high-frequency financial data sets have been widely used to study various aspects of market microstructure. At first sight, the higher the number of independently measured observations, the higher is the precisions of estimation and the better model of market behavior can be obtained, however the situation is not as simple as it appears. The very detailed financial data can cause many problems. Some of them may result from human input errors, such as typing errors leading to data outliers. Other errors may be computer system errors, such as transmission failures which lead to data gaps, and database bugs causing mis-ordered time series observations [Zivot 2005]. Cleaning and correcting the data can be cumbersome. Furthermore tick-by-tick data are very noisy, because of discrete prices and nonsynchronous reporting which may deform inferences based on standard statistical models. The enormous number of data enforces the need of time-consuming data handling and filtering techniques. The sets of high-frequency data are materially limited. In the US markets transaction data are only available since 1983 and in many countries they are not available at all [Goyenko et al. 2009]. The presented problems with high-frequency data are sufficient to justify the need of the reliable bid-ask spread estimator based on low-frequency data by which I mean easy accessible daily characteristics of stocks such as opening price, closing price, low price, high price and trading volume. Researchers have an access to such data over a long price history and in many markets. The US daily stock returns and volume data are available from the Center for Research in Security Prices (CRSP) for NYSE/AMEX firms from 1926 to the present and for NASDAQ firms from 1983 to the present. A wide variety of services provide daily stock returns and volume data for international equity markets. For example, Thomson Financial’s Data-stream provides the considered data for more than 60 countries from 1994 to the present and daily stock returns for several developed markets from the early 1970s [Goyenko et al. 2009]. Thus, reliable low-frequency spread proxies enable research of the bid-ask spread for a long time period and many markets. High-performing low-frequency spread measures would be very helpful to the theory of market efficiency and in corporate finance. Such estimators are needed to verify whether described in literature trading strategies that appear to generate significant abnormal returns are truly profitable net of a relatively precise measure of transaction costs. Moreover, spread estimators across countries would greatly extend the potential diversity of international corporate finance environments to analyze [Holden 2009]. Recent articles provide indications that there is a chance of useful estimation of high-frequency benchmarks by means of low-frequency spread proxies in stock market [Goyenko et al. 2009], foreign exchange market [Karnaukh et al. 2014] and it seems that there is a hope for a success of low-frequency spread estimation in other types of markets where bid and ask prices appear. In fact, most studies published in financial literature deal with low frequency, regularly spaced data.

ESTIMATORS AND THEIR EVALUATION

In the financial literature there exist quite a number of methods of low-frequency spread estimation, a good overview of such estimators one can find, for example, in [Anand and Karagozoglu 2006], [Corwin and Schultz 2012] and [Goyenko et al. 2009]. In this paper I test the applicability of two of them on polish stock market: Roll estimator [Roll 1984] and Corwin-Schultz estimator[1] [Corwin and Schultz 2012]. The Roll method appeared as the first tool for measuring the bid-ask spread by means of easily available financial daily data. Although nowadays intraday data are widely available, researchers still frequently use this estimator or its extensions in applications such as, for example, asset pricing and testing market efficiency [Corwin and Schultz 2012]. The popularity of the Roll estimator may result from the fact that it enables the rapid measurement of transaction costs solely on the basis of the observed prices [Doman 2011]. As describes [Roll 1984], the method requires only the prices, so it is very cheap. It requires two major assumptions [Roll 1984]:

1. The asset is traded in an informationally efficient market.
2. The probability distribution of observed price changes is stationary (at least for short intervals).

The Roll estimate of the spread understood as the relative proportion of the actual or theoretical price is given as follows:

$$S = 2\sqrt{-\operatorname{cov}(R_t, R_{t+1})} \tag{1}$$

where R_t and R_{t+1} are rate of returns over a day t and $t+1$ respectively.

[1] also known as high-low estimator

If the returns are arithmetic then the formula holds approximately [Roll 1984], it holds exactly for logarithmic returns [Doman 2011]. The elegant derivation of the Roll formula can be found in [Zhang and Hodges 2012].

The method of Roll was later modified and extended. [Bleaney and Li 2014] even classify the bid-ask estimators as the Roll family of estimators and the other ones. Although simple and based on reasonable assumptions, the Roll method is strongly impaired by the fact that in about half the cases the autocovariance is positive The common ad hoc way to deal with this problem is to set Roll spreads to zero in these cases [Corwin and Schultz 2012].

[Corwin and Schultz 2012] spread estimator uses the daily high and low prices to estimate the spread. It assumes that:

1. the daily high price is a buyer-initiated trade and the daily low price is a seller-initiated trade
2. the percentage spread is constant over the 2-day estimation period
3. the true, unobserved price follows a geometric Brownian motion with zero drift

Especially the first assumption is a brilliant idea which seems to be very well adjusted to market reality.

Let H_t^0 and L_t^0 denote the observed daily high and low price, respectively, on day t. Moreover, let $H_{t,t+1}^0$ and $L_{t,t+1}^0$ be observed high and low, respectively, over the 2 days t and $t+1$. Then, the parameters β and γ are defined as follows:

$$\beta = E\left(\left[\ln\left(\frac{H_t^0}{L_t^0}\right)\right]^2 + \left[\ln\left(\frac{H_{t+1}^0}{L_{t+1}^0}\right)\right]^2\right) \text{ and } \gamma = \left[\ln\left(\frac{H_{t,t+1}^0}{L_{t,t+1}^0}\right)\right]^2. \tag{2}$$

The value S of the Corwin-Schultz spread estimator in its original form is determined by the equations [Corwin and Schultz 2012]:

$$S = \frac{2\left(e^{\alpha}-1\right)}{1+e^{\alpha}} \tag{3}$$

$$\alpha = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}\,. \tag{4}$$

The last equation can be, by means of elementary transformations, significantly simplify and present the formula for α in the following, more appealing form [Karnaukh et al. 2014]:

$$\alpha = \left(1+\sqrt{2}\right)\left(\sqrt{\beta}-\sqrt{\gamma}\right). \tag{5}$$

In practice, β can be calculated as the sum over two days of the squared daily natural logarithms of the proper high/low ratios [Karnaukh et al. 2014]. The results of [Corwin and Schultz 2012] suggest that such approach produces more accurate estimates of average monthly spreads than approximation of the expected value by averaging β over the month.

There are a number of implicit assumptions underlying the high–low spread estimator which are usually not true in practice. In order to benefit effectively from the considered method it is good to make the following adjustments which are described in detail in [Corwin and Schultz 2012]:

- adjustment for overnight price changes
- adjustment for the same price for all daily trades
- adjustment for the negative values of the estimates

To evaluate the performance of bid–ask spread estimators it is natural to examine two issues:

1. how much the estimates differ from the real bid-ask spread
2. to what extent they exhibit the same behavior as the spread in different trading environments and time periods.

With regard to the first issue, one uses mean absolute error (MAE) or root mean squared error. Both measures capture both the bias in the estimate and its variability. Lower values of the considered errors indicate that the estimated bid-ask spreads are closer to the actual spread. Second issue can be studied by calculation the correlations between the estimated and actual spreads. The higher the correlation the closer similarity of behavior od estimates and true spreads. [Corwin and Schultz 2012] carefully study the second issue and find that the accuracy of their model in terms of correlation is very good. Interestingly, they devote very little attention to the question of error of estimation. They write in the footnote that calculated mean absolute errors based on the difference between monthly spread estimates for each of considered by them estimators and monthly effective spreads from TAQ (trades and quotes database) across all sample months, and present the results of these calculations. For their high-low method, mean absolute error was 0,9%. Taking into account that the simple average effective spread from TAQ across all stock-months was 2.38%, one can easily compute that the ratio of these values which is the relative error is about 38%. It is not dramatically bad but the simulations of [Bleaney and Li 2014] reveal that the Corwin-Schultz estimator may be yet seriously biased. When it comes to the Roll method, it is rather commonly known that is far from satisfactory with respect to both aspects of performance assessment.

EVALUATION OF SPREADS OF THE STOCKS LISTED ON THE WARSAW STOCK EXCHANGE

The estimation concerns the stocks quoted on the Warsaw Stock Exchange (WSE) in 2013. From 460 stocks of which average annual spreads have been published in WSE Statistic Bulletin for the year 2013 I took a random sample of 30 stocks. The average annual spread from the bulletin is a benchmark by which the accuracy of low-frequency estimation is measured. The requirements from a share to be taken into account in drawing were very low. They were the following:

1. the full record of daily data (247 days) and non-ambiguous name, in the quotations archive in the portal http://www.gpwinfostrefa.pl
2. non-zero spread in 2013 WSE Statistic Bulletin

This approach quite strongly differs from standard methods of stock selection described in academic literature where usually the criteria are much more demanding. But in my opinion such high requirements are designed for the comfort of researchers rather than for managers and analysts of real financial markets. Of course, applying the given estimator to some set of data may be theoretically unjustified, though numerically possible. But also then the obtained result is the indicator of the method accuracy. The complete results are presented in table 1.

Table 1. Annual average spreads and Roll and Corwin-Schultz estimations

Stock name	Average spread	Roll estimation	Corwin-Schultz estimation
ALCHEMIA	0,87%	0,78%	0,94%
AVIASG	1,59%	0,00%	0,77%
BSCDRUK	3,56%	2,09%	1,45%
BUDIMEX	0,58%	0,00%	0,64%
BYTOM	2,26%	3,95%	1,65%
CASHFLOW	3,41%	0,00%	1,40%
COALENERG	1,58%	0,00%	2,11%
ERBUD	1,61%	0,00%	1,26%
FERRUM	1,82%	2,09%	1,05%
GETINOBLE	0,60%	1,40%	0,72%
INPRO	3,59%	1,57%	2,59%
INVISTA	3,35%	2,48%	2,28%
IQP	2,72%	2,10%	1,55%
IVMX	2,87%	2,03%	1,95%
KDMSHIPNG	4,18%	0,46%	1,90%
KOMPAP	0,91%	3,35%	1,35%
LENTEX	0,72%	0,00%	0,42%
MAKRUM	2,15%	0,00%	1,31%
MILLENNIUM	0,41%	0,00%	0,81%
MIT	1,92%	1,32%	2,01%
PCCINTER	2,91%	1,04%	0,89%
MOSTALZAB	1,06%	0,00%	1,10%
PLAZACNTR	2,53%	4,81%	2,90%
PRAGMAFA	1,25%	1,74%	0,29%
SANWIL	4,45%	3,77%	3,23%
SKOTAN	0,81%	0,00%	0,97%
TESGAS	1,93%	0,00%	1,14%
TRITON	4,01%	0,00%	1,82%
VARIANT	2,68%	0,00%	1,72%
VINDEXUS	1,89%	0,00%	1,03%

Source: WSE Statistic Bulletin 2013 and own calculations based on data from www.gpw.infostrefa

In 14 cases the Roll estimates is 0 which means that the autocovariance is positive and it is consistent with common knowledge that positive autocovariance happens in about half cases. also calculated the correlation coefficients r_{Roll} and r_{CS} between the averages spreads from 2013 WSE Bulletin and the Roll and the Corwin-Schultz estimators, respectively and verified their significance by the standard significance test with test statistics $t = r\sqrt{\frac{n-2}{1-r^2}}$ where r is the correlation coefficient and n is a number of observations. The following results were obtained (p is the p-value):

$$r_{Roll} = 0{,}290; p = 0{,}12006$$

$$r_{CS} = 0{,}790, p = 0{,}00001$$

It is apparent from the carried out calculations that the Corwin-Schultz estimator is much better than the Roll one with respect to the similarity if behavior to the actual spread. The results do not even confirm the relationship between the Roll estimates and actual spreads. One may suspect that this is due to small sample but on the other hand it was large enough to prove strong interdependence between the Corwin-Schultz estimator and true values of spread.

Then, the mean absolute errors MAE_{Roll} and MAE_{CS} of the Roll and the Corwin-Schultz estimators were computed:

$$MAE_{Roll} = 1{,}51\%$$

$$MAE_{CS} = 0{,}85\%.$$

The results provide clear evidence that the Corwin- Schultz method outperforms the Roll estimation with regard to error of measurement. The ratio of MAE_{CS} and the average value of actual spread in the sample is below 40% which is not small value but seeming acceptable from the point of view of the bid-ask spread assessment in practice.

CONCLUSION AND PERSPECTIVE FOR RESEARCH

The paper by demonstrating the difficulties with high-frequency data argues that low-frequency estimation is important. Two popular estimators based on daily data are applied to the polish shares. The obtained results show that the Corwin-Schultz estimator is an effective tool to measure the bid-ask spread for stocks quoted on WSE on the basis of daily data and it is unambiguously better than the popular Roll method, in this task. The aim was realized by confirmation of the usefulness of the Corwin-Schultz method in the polish stock market. Moreover, the performed calculations for the polish market, confirmed the properties of both estimators known from the previous literature.. The study was done with the small sample size and concerned only one year. The natural extension of this article could be considering larger sample and longer time period. Moreover, it interesting to see how the situation changes when not annual but, as it is in [Corwin and Schultz 2012], monthly average spreads are considered.

REFERENCES

Anand A., Karagozoglu A. K. (2006) Relative performance of bid–ask spread estimators: Futures market evidence, International Financial Markets, Institutions and Money, 16, pp. 231 – 245.

Bleaney M., Li Z. (2013) The performance of bid-ask spread estimators under less than ideal condition, http://www.nottingham.ac.uk/economics/documents/discussion-papers/13-05.pdf

Bleaney M., Li Z. (2014) A new spread estimator, http://www.nottingham.ac.uk/economics/documents/discussion-papers/14-01.pdf

Corwin S. A., Schultz P. (2012) A Simple Way to Estimate Bid-Ask Spreads from daily High and Low Prices, Journal of Finance, 67, pp. 719 – 759.

Doman M. (2011) Mikrostruktura giełd papierów wartościowych, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań

Elton E. J, Gruber M. J, Brown S. J., Goetzmann W. N. (2010) Modern Portfolio Theory and Investment Analysis, John Wiley & Sons, Hoboken.

Gniadkowska A. (2012) Wpływ płynności obrotu na kształtowanie się stopy zwrotu z akcji notowanych na Giełdzie Papierów Wartościowych w Warszawie, Zarządzanie i Finanse, 10, pp. 563 – 570.

Goyenko R. Y., Holden C. W., Trzcinka C. A. (2009) Do liquidity measures measure liquidity?, Journal of Financial Economics, 92, pp. 153 – 181.

Grinold R. C., Kahn R. N. (2000) Active Portfolio Management, McGraw-Hill, New York.

Holden C. (2009) New low-frequency spread measures, Journal of Financial Markets, 12, pp. 778 – 813.

Karnaukh N., Ranaldo A., Söderlind Paul (2014) Understanding FX Liquidity, http://www1.vwa.unisg.ch/RePEc/usg/sfwpfi/WPF-1315.pdf

Olbryś J. (2012) Tarcie w procesach transakcyjnych i jego konsekwencje, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 254, pp. 181–189.

Roll R. (1984) A simple implicit measure of the effective bid-ask spread in an efficient market, Journal of Finance, 39, pp. 1127–1139.

Sharpe W. F., Alexander G., J., Bailey J. V. (1999) Investments Prentice Hall, Upper Saddle River.

WSE Statistics Bulletin, http://www.gpw.pl/pub/statystyki_roczne/2013_GPW.pdf

Zhang H., Hodges S. (2012) An Extended Model of Effective Bid-ask Spread, http://www.cass.city.ac.uk/__data/assets/pdf_file/0007/128068/H.Zhang.pdf

Zivot E. (2005) Analysis of High Frequency Financial Data: Models, Methods and Software. Part I: Descriptive Analysis of High Frequency Financial Data with S-PLUS, http://faculty.washington.edu/ezivot/research/hflectures.pdf