Trading costs - The Roll Spread Estimator

Bernt Arne Ødegaard

21 December 2023

Contents

1 Looking at Roll [1984], estimating bid ask spread from time series of transactions. **1**

2 Actual calculation **3**
2.1 What data to apply the calculation to? 3

3 Later developments **9**

4 Summary **9**

5 Literature **9**
5.1 Textbook discussions 9

1 Looking at Roll [1984], estimating bid ask spread from time series of transactions.

Roll (1984) not really a microstructure paper, takes the bid/ask spread as given, but a very neat idea.

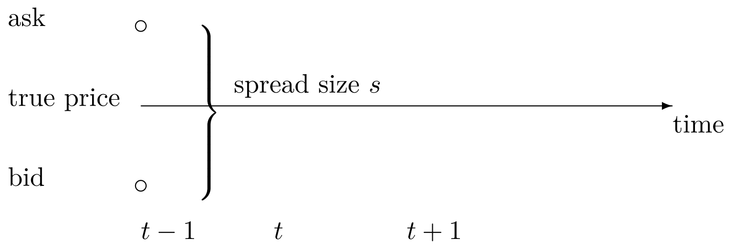
Problem: How can one measure bid/ask spread from observed trade prices?

Well, if we take as given the existence of spread, implication for successive price movements: Tend to bounce back and forth, depending on whether trades are buyer or seller initiated.

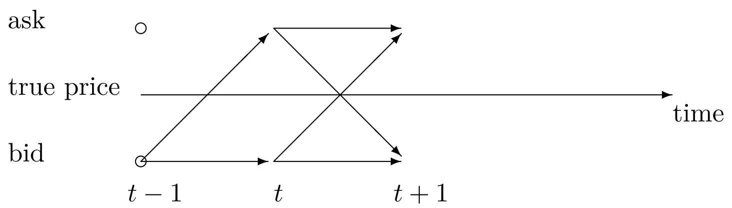
If no new information hits the market, underlying "true" price does not move. Impossible to have two succesive positive or negative price movements.

This enough to generate interesting time series implications.

Setup: Suppose last observed trade (at time $t-1$) was at bid.



How can prices change over next two periods, if no new information moves the price?



Only two possible prices at times t and $t+1$: P_{t-1} and $P_{t-1}+s$.

Probabilities

Assume each order arriving to the market has equal probabilities of being a buyer and a seller initiated order.

Then: Each branch has probability $\frac{1}{2}$.

$$P(P_t = P_{t-1} | P_{t-1} \text{ was at bid }) = \frac{1}{2}$$

$$P(P_t = P_{t-1} + s | P_{t-1} \text{ was at bid }) = \frac{1}{2}$$

What implications has this for succesive price changes?

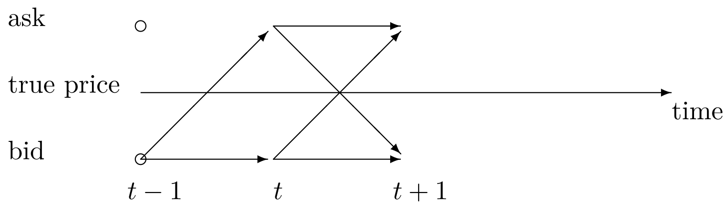
Calculate joint (unconditional) probabilities.

Define

$$\Delta P_t = P_t - P_{t-1}$$

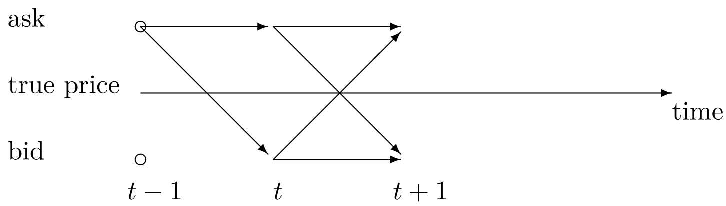
Calculate probabilities conditional on being at the bid

		$\Delta P_t = P_t - P_{t-1}$		
		$-s$	0	$+s$
	Prob.	0	$\frac{1}{2}$	$\frac{1}{2}$
ΔP_{t+1}	$-s$	0	0	$\frac{1}{2}\frac{1}{2} = \frac{1}{4}$
	0	0	$\frac{1}{2}\frac{1}{2} = \frac{1}{4}$	$\frac{1}{2}\frac{1}{2} = \frac{1}{4}$
	$+s$	0	$\frac{1}{2}\frac{1}{2} = \frac{1}{4}$	0



Similarly, calculate probabilities conditional on being at the ask

		ΔP_t		
		$-s$	0	$+s$
	Prob.	$\frac{1}{2}$	$\frac{1}{2}$	0
ΔP_{t+1}	$-s$	0	$\frac{1}{4}$	0
	0	$\frac{1}{4}$	$\frac{1}{4}$	0
	$+s$	$\frac{1}{4}$	0	0



Unconditional, multiply with probability of each conditional state occurring, here equal to $\frac{1}{2}$.

		ΔP_t		
		$-s$	0	$+s$
ΔP_{t+1}	$-s$	0	$\frac{1}{4}\frac{1}{2} = \frac{1}{8}$	$\frac{1}{8}$
	0	$\frac{1}{8}$	$\frac{1}{4}\frac{1}{2} + \frac{1}{4}\frac{1}{2} = \frac{1}{4}$	$\frac{1}{8}$
	$+s$	$\frac{1}{8}$	$\frac{1}{8}$	0

Calculate the theoretical covariance

$$\text{cov}(\Delta p_t, \Delta p_{t+1}) = E[(\Delta p_t - E[\Delta p_t])(\Delta p_{t+1} - E[\Delta p_{t+1}])]$$

What is $E[\Delta p_t]$?

Since each state equally likely to occur,

$$E[\Delta p_t] = E[\Delta p_{t+1}] = 0$$

Thus,

$$\text{cov}(\Delta p_t, \Delta p_{t+1}) = E[\Delta p_t \Delta p_{t+1}] = \frac{1}{8}(-s)(+s) + \frac{1}{8}(+s)(-s) = \frac{1}{4}(-s^2) = -\frac{1}{4}s^2$$

Thus, the bid ask bounce causes an induced serial correlation in returns, even if no news hits the market.

But this induced serial correlation allows us to estimate s!

Solving the above equation for s:

$$s = 2\sqrt{-\text{cov}(\Delta p_t, \Delta p_{t+1})}$$

To normalize, work with returns rather than price changes

$$\hat{s} = 200\sqrt{-\text{cov}(r_t, r_{t+1})}$$

Thus, get a nice, empirical estimate of the implied bid/ask spread.

But this is developed the assumption of constant "true" price. To develop the Roll measure, e.g. need assumptions about what happens when "true" prices moves.

2 Actual calculation

If we do the calculation from prices, we get an estimate of the kroner spread

First calculate

$$Scov = cov(\Delta p_t, \Delta p_{t-1}).$$

If we do the calculation from returns, we get an estimate of the percentage spread

$$Scov = cov(r_t, r_{t-1})$$

Calculate the spread as

$$\hat{s} = \begin{cases} 2\sqrt{-Scov} & \text{if} \quad Scov < 0 \\ ? & \text{if} \quad Scov > 0 \end{cases}$$

What to do if $Scov > 0$?

Suggestions:

- $s = -\sqrt{Scov}$ (negative spread) (Harris, 1990)
- $s = 0$
- $s = NA$

2.1 What data to apply the calculation to?

Ideally: Actual transaction data.

However: Also applied to data sampled at lower frequencies, such as daily.

Here making assumptions that the daily closing price as likely to be at the bid as at the ask.

In practice: Daily sampling gives reasonable estimates for transaction costs, but many cases where the autocovariance is positive.

The daily autocovariance is more likely to be positive for heavily traded stocks.

Makes sense: With less heavily traded stocks get closer to transaction by transaction, even at daily frequency.

Exercise 1.

The Roll (1984) estimator provides an estimate of the trading costs for a given security.

Using daily returns data for Odfjell (ODF), estimate the trading costs for the years 2008 to 2012.

Solution to Exercise 1.

Defining the Roll estimator

```
> Roll <- function(inp){
+   a <- acf(as.matrix(inp),plot=FALSE,lag=1,type="covariance")
+   r <- 0
+   Scov <- a$acf[2]
+   if (Scov<0){ return (2*sqrt(-Scov)) }
+   else { return (NA) }
+ }
```

Reading the data

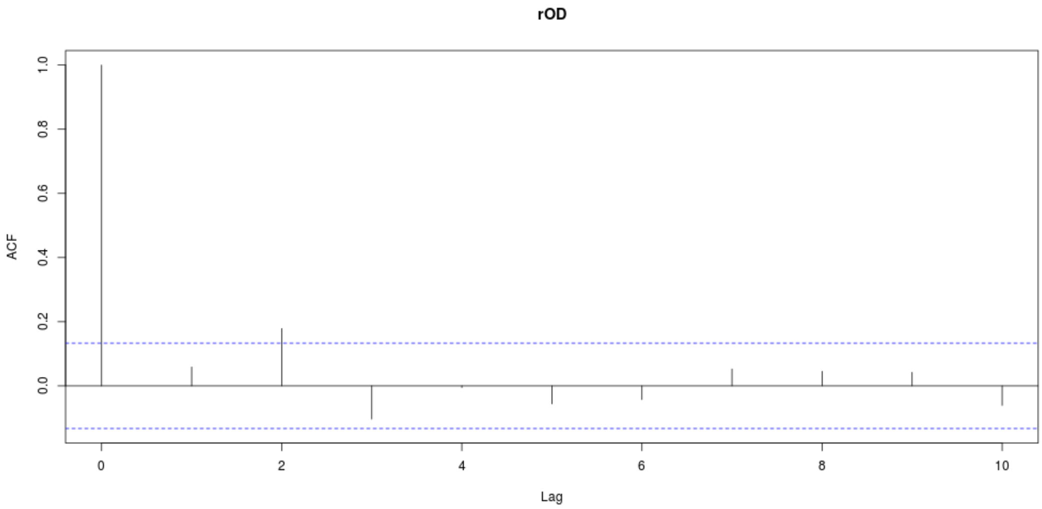
```
pOD  <- read.zoo("../data/odfjell.csv",
                  header=TRUE,sep=";",format="%d.%m.%y")
pOD <- pOD$Siste
rOD  <- diff(pOD)/pOD[-length(pOD)]
names(rOD) <- "rOD"
```

```
> summary(rOD)
     Index                 rOD
 Min.   :2008-02-07   Min.   :-0.16631
 1st Qu.:2009-05-12   1st Qu.:-0.01575
 Median :2010-08-11   Median : 0.00000
 Mean   :2010-08-08   Mean   :-0.00125
 3rd Qu.:2011-11-06   3rd Qu.: 0.01250
 Max.   :2013-02-05   Max.   : 0.13031
                      NA's   :229
```

First doing the estimation for 2008:

```
> r <- na.omit( window(rOD, start=as.Date("2008-01-01"),end=as.Date("2008-12-31")))
> Roll(r)
[1] NA
```

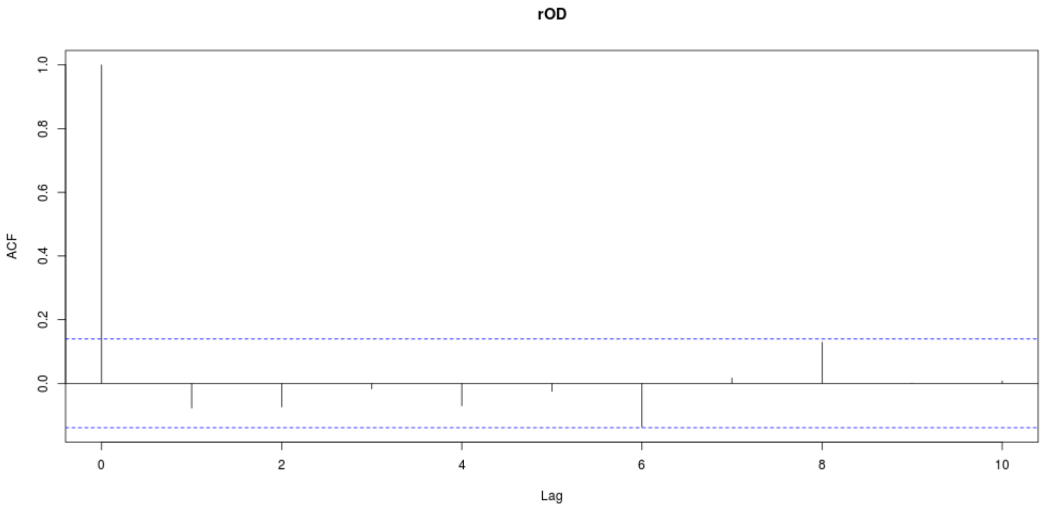
Problem, which we see from the acf function



The autocovariance is not negative.
Doing 2009

```
> r <-na.omit( window(rOD, start=as.Date("2009-01-01"),end=as.Date("2009-12-31")))
> Roll(r)
[1] 0.01367632
```

So here we estimate the trading cost at 1.4%.



Doing this for the other three years

```
> r <- na.omit( window(rOD, start=as.Date("2010-01-01"),end=as.Date("2010-12-31")))
> Roll(r)
[1] 0.01846564
> r <- na.omit( window(rOD, start=as.Date("2011-01-01"),end=as.Date("2011-12-31")))
> Roll(r)
[1] NA
> r <- na.omit( window(rOD, start=as.Date("2012-01-01"),end=as.Date("2012-12-31")))
> Roll(r)
[1] 0.0153559
```

Summarizing

Year	Cost estimate
2008	NA
2009	0.01367632
2010	0.01846564
2011	NA
2012	0.0153559

Exercise 2.

The simple Roll estimator of trading costs (from returns) is

$$TC = 2\sqrt{-Scov}$$

where $Scov$ is the autocovariance of stock returns.

We will use the Roll estimator to estimate trading costs for NHY for a number of different datasets.

- Using daily returns data for OSE, estimate the trading costs of Norsk Hydro for 1995, 1997, 1999, 2001, 2005 and 2012.
- Using actual trades of NHY for september 1997, estimate the trading costs.
- Using the actual trades of NHY for 3 dec 2012, estimate the trading costs using the Roll estimator.

Solution to Exercise 2.

Let us start by writing the estimation of the Roll model as a R function

```
# estimating the roll model
Roll <- function(inp){
  a <- acf(as.matrix(inp),plot=FALSE,lag=1,type="covariance")
  r <- 0
  Scov <- a$acf[2]
  if (Scov<0){ return (2*sqrt(-Scov)) }
  else { return (NA) }
}
```

and then we show the output of reading in the data and running the Roll estimation

```
# read files pulled from OSE homepage
library(zoo)
NHYPric   <- read.zoo("../data/nhy.csv",
                      header=TRUE,format="%d.%m.%y",sep="\t",skip=1)
NHYPric   <- NHYPric[,1]
names(NHYPric)[1] <- "NHYClose"
```

Summarizing the data

```
> summary(DretNHY)
     Index               DretNHY
 Min.   :1980-01-03   Min.   :-0.2227000
 1st Qu.:1988-04-08   1st Qu.:-0.0101000
 Median :1996-07-05   Median : 0.0000000
 Mean   :1996-07-05   Mean   : 0.0005897
 3rd Qu.:2004-10-11   3rd Qu.: 0.0108000
 Max.   :2012-12-28   Max.   : 0.2064000

> NHY <- window(DretNHY, start=as.Date("1995-01-01"),end=as.Date("1995-12-31"))
> Roll(NHY)
[1] NA
> NHY <- window(DretNHY, start=as.Date("1997-01-01"),end=as.Date("1997-12-31"))
> Roll(NHY)
[1] NA
> NHY <- window(DretNHY, start=as.Date("1999-01-01"),end=as.Date("1999-12-31"))
> Roll(NHY)
[1] NA
> NHY <- window(DretNHY, start=as.Date("2001-01-01"),end=as.Date("2001-12-31"))
> Roll(NHY)
[1] NA
> NHY <- window(DretNHY, start=as.Date("2005-01-01"),end=as.Date("2005-12-31"))
> Roll(NHY)
[1] 0.009454619
> NHY <- window(DretNHY, start=as.Date("2012-01-01"),end=as.Date("2012-12-31"))
> Roll(NHY)
[1] 0.002750503
```

The estimates for NHY show an issue with using the Roll estimator. It may at times be undefined, since the calculation

$$\sqrt{-Scov}$$

is only defined when $Scov < 0$. Otherwise the estimator is not defined. So the Roll estimator may at times not actually give you an estimate.

Let us show the ACF for 1995, to illustrate this

```
\includegraphics[scale=0.4]{../R_plots/acf_daily_nhy_rets_1995}
```

The first order autocovariance is positive.

Let us next look at the trade by trade record for trading in NHY in september of 1997.

We consider two inputs: The trade by trade record, and daily returns, constructed from closing prices (last trade price of the day), and do it for the same time period, september of 2007.

```
> summary(prices97)
     Index                        Price
 Min.   :1997-09-01 09:51:00   Min.   :395.0
 1st Qu.:1997-09-04 16:59:00   1st Qu.:419.0
 Median :1997-09-11 15:51:00   Median :426.0
 Mean   :1997-09-13 13:13:15   Mean   :425.0
 3rd Qu.:1997-09-21 01:25:30   3rd Qu.:431.0
 Max.   :1997-09-30 17:01:00   Max.   :440.5
```

Let us first look at the trade by trade Roll estimate

```
> R <- diff(log(P))
> Roll(R)
[1] 0.002604668
```

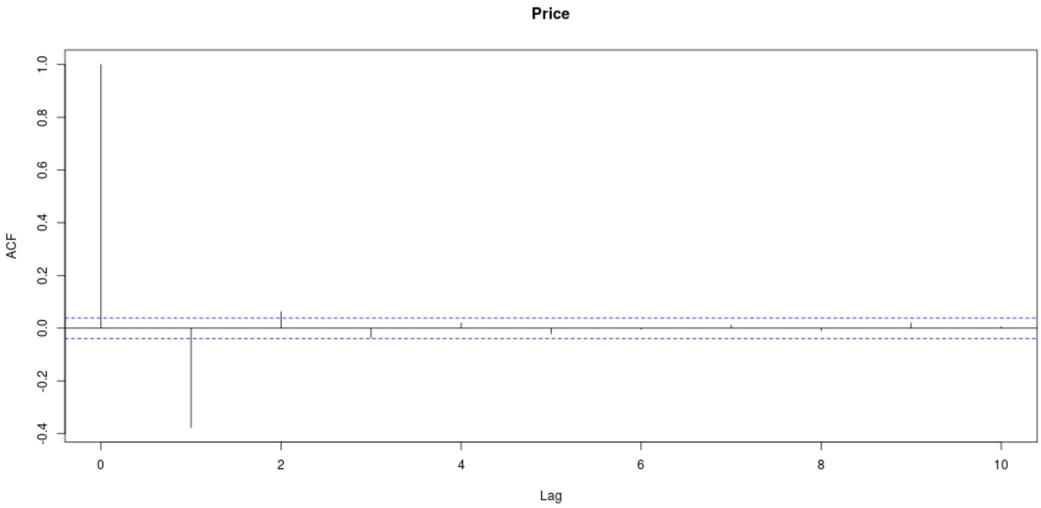
and compare it to the estimate from taking the daily prices over the same time period.

```
> SRets <- read.zoo("../data/daily_rets.csv",
+                    format="%Y%m%d", header=TRUE,skip=2,sep=",");
> NHY <- window(na.omit(SRets$Norsk.Hydro), start=as.Date("1997-09-01"),end=as.Date("1997-10-01"))
> Roll(NHY)
[1] NA
```

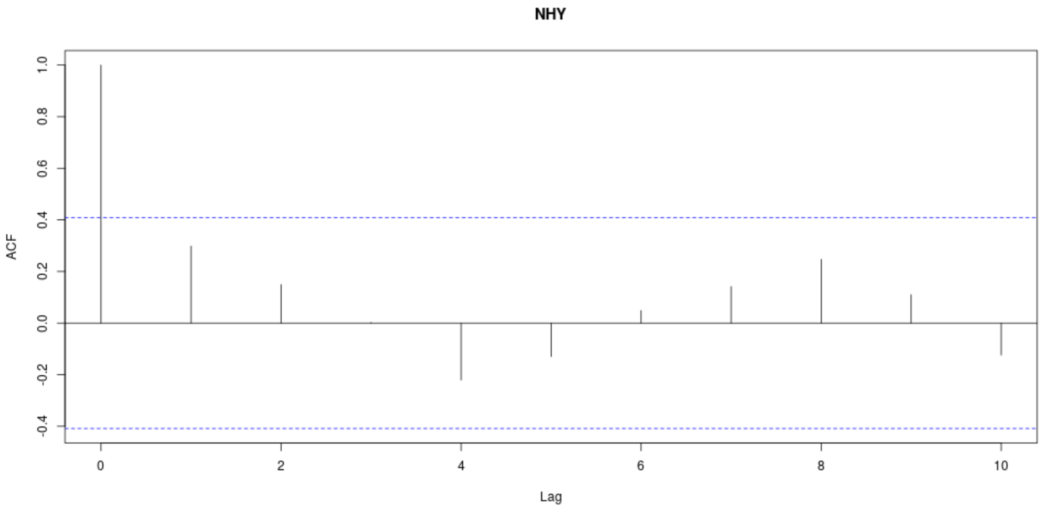
As we see the Roll estimator is not even defined for the daily data.

To see why, let us plot the ACF functions for the two cases.

ACF for trade by trade returns.



ACF for daily returns.



So, the daily returns are positively autocorrelated for lags 1-3, and the Roll estimator is not defined.

The trade by trade case is much better, at least for this example. So, it may look like the daily case is not that useful, but it does tally with other cost measures, in the sense that when the Roll measure is defined, it has positive correlation in the crossection with spreads, etc: More illiquid stocks by other measures have higher estimates of Roll. Maybe this is due to the fact that for illiquid stocks, daily close prices are much closer to trade by trade prices...

Let us look at the trade by trade case using data from december 2012.

```
> summary(p)
     Index                         Price
 Min.   :2012-12-03 08:00:22   Min.   :27.00
 1st Qu.:2012-12-03 09:37:50   1st Qu.:27.18
 Median :2012-12-03 11:43:08   Median :27.24
 Mean   :2012-12-03 11:41:58   Mean   :27.25
 3rd Qu.:2012-12-03 13:49:50   3rd Qu.:27.31
 Max.   :2012-12-03 15:25:25   Max.   :27.49
```

Running the estimates:

```
> R <- diff(log(p12))
> Roll(R)
[1] 0.0001815402
```

So the trading costs have gone down by a factor of 10.

3 Later developments

George, Kaul, and Nimalendran (1991) develop the Roll measure.

A conceptually better way to do the estimation would be to use the Bayesian framework of the Gibbs sampler in Hasbrouck (2006), which allow one to impose nonnegativety on the cost.

4 Summary

The Roll (1984) measure estimates trading costs as the effective spread implicit in the sequence of trades. If one posit the existence of a constant proportional effective spread s, Roll shows how one can back this out from the autocorrelation of successive price movements. The bouncing back and forth between bid and ask will be induced partly by the magnitude of the relative spread s, and Roll shows that this leads to a calculation as follows, where r_t is the stock return at time t:

$$Scov = cov(r_t, r_{t-1})$$

and estimate s as

$$\hat{s} = \begin{cases} 2\sqrt{-Scov} & \text{if} \quad Scov < 0 \\ \text{undefined} & \text{if} \quad Scov > 0 \end{cases}$$

5 Literature

Original article: Roll (1984)

Harris (1990) looks at the usage of the Roll estimator, in particular that one can try to take the square root of a negative number.

Extending the Roll measure: George et al. (1991)

5.1 Textbook discussions

(Foucault, Pagano, and Roell, 2013, pg 59-62)

References

Thierry Foucault, Marco Pagano, and Ailsa Roell. *Market Liquidity. Theory, Evidence and Policy.* Oxford University Press, 2013.

Thomas J George, Gautam Kaul, and M Nimalendran. Estimation of the bid–ask spread and its components: A new approach. *Review of Financial Studies*, 4:623–656, 1991.

Lawrence Harris. Statistical properties of the Roll serial covariance bid/ask spread estimator. *Journal of Finance*, 44: 579–590, 1990.

Joel Hasbrouck. Trading costs and returns for US equities: Estimating effective costs from daily data. Working Paper, Stern School of Business, New York University, August 2006.

Richard Roll. A simple implicit measure of the effective bid–ask spread in an efficient market. *Journal of Finance*, 39(4): 1127–1139, September 1984.