Return Volatility, Market Microstructure Noise, and Institutional Investors: Evidence from High Frequency Market

Yuting Tan, Lan Zhang

R/Finance 2017

ytan36@uic.edu

May 19, 2017

Objective

- Examine the role of institutional investors (as well as mutual funds) in the high frequency world.
- Investigate whether institutional investors increase the stock return volatility and whether institutional investors affect the market microstructure noise of stocks.

Challenge

Many estimators and measures developed from low frequency data are no longer applicable in the high frequency world.

- Realized Volatility estimator *vs.* TSRV estimator

Integrated Volatility

The old estimator: Realized Volatility

$[X, X]_T \triangleq \sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2$

where X_{t_i}'s are log stock prices observations in $[0, T]$.

- X_t is assumed to follow an Itô process
- This model is justified theoretically
- The realized volatility computed from the highest-frequency data should provide the best possible estimate for the integrated volatility $\int_0^T \sigma_t^2 dt$
- However it fails in real application due to the existence of market microstructure noise

Integrated Volatility Cont'd

The true model

$Y_{t_i} = X_{t_i} + \epsilon_{t_i}$
Y_t is the observed log price process
X_t is the latent true log price process
and ϵ_{t_i}'s are noises around the latent prices

This is a model in the real world capturing market microstructure noise. If we use the old estimator, it'd become

The noise problem ($[Y, Y]_T^{(all)}$)

$$\sum_{t_i, t_{i+1} \in [0,T]} (Y_{t_{i+1}} - Y_{t_i})^2 = 2nE\epsilon^2 + O_p(n^{1/2})$$

where n is the number of sampling intervals over $[0, T]$.

Thus the realized volatility estimator gives us the variance of noise $E\epsilon^2$ rather than the true integrated volatility $\langle X, X \rangle_T$.

Integrated Volatility Cont'd

The construction of the "two-scales estimator" (also known as TSRV) is not complicated.

TSRV estimator

$\widehat{\langle X, X \rangle}_T = [Y, Y]_T^{(avg)} - \frac{\bar{n}}{n}[Y, Y]_T^{(all)}$ where T =1 day

We construct the estimator $[Y, Y]_T^{(avg)}$ by subsampling every 5 minutes (or other time intervals according to the amount of observations in original data set). We start subsampling from the first observation, then start with the second observation, and so on. The average of the results obtained from those subsamples is $[Y, Y]_T^{(avg)}$.

Market Microstructure Noise

We easily get the estimator of variance of the noise term from the old realized volatility estimator

The noise estimator

$$\widehat{E\epsilon^2} = \frac{1}{2n}[Y, Y]_T^{(all)}$$

The Data

We collect data from stocks of Dow 30 companies during the period Jan 1, 2003 - Dec 31, 2012.

- **TAQ** database for high frequency trades: millisecond consolidated trades occurred during the regular trading hours.
- **CRSP** database for stock basic information: daily close price, best bid/ask, daily return, market capitalization, shares outstanding, daily trading volume, etc.
- **Thomson Reuters** database for ownership information: institutional ownership, mutual fund ownership.

The Data: Summary Statistics

Table : Descriptive statistics of variables

	Mean	Std Dev	Median	Min	Max
Volatility	0.0173	0.049	0.0099	0.000	2.128
Noise	0.0008	0.004	0.0001	0.000	0.711
InstOwn	0.6511	0.116	0.6507	0.159	0.947
MfOwn	0.2386	0.060	0.2362	0.082	0.424
MktCap	11.4296	0.761	11.4499	8.888	13.397
1/*Price*	0.0233	0.013	0.0194	0.001	0.150
Amihud	0.0410	0.365	0.0000	0.000	14.200
Spread	0.0004	0.001	0.0003	-0.009	0.081

Observations: 68,276

Results: Institutional ownership

Table : Regression of volatility and noise on institutional ownership

	Volatility	
	Estimate	**t value**
Constant	−0.006	−1.247
InstOwn	0.030	14.408***
MktCap	0.001	1.646
1/*Price*	−0.134	−9.472***
Amihud	−0.001	−1.983*
Spread	1.699	8.040***

	Noise	
	Estimate	**t value**
Constant	2.335×10^{-3}	6.244***
InstOwn	7.080×10^{-4}	4.216***
MktCap	-1.366×10^{-4}	−5.300***
1/*Price*	-1.830×10^{-2}	−15.944***
Amihud	-2.857×10^{-4}	−6.756***
Spread	-6.098×10^{-3}	−0.356

Observations: 68,276

Results: Mutual Fund ownership

Table : Regression of volatility and noise on mutual fund ownership

	Volatility	
	Estimate	**t value**
Constant	-0.019	-4.882***
MfOwn	0.091	24.729***
MktCap	0.002	5.383***
1/*Price*	-0.191	-13.428***
Amihud	-0.001	-1.977*
Spread	1.901	9.019***

	Noise	
	Estimate	**t value**
Constant	1.018×10^{-3}	3.216**
MfOwn	3.567×10^{-3}	11.898***
MktCap	-5.155×10^{-5}	-2.152*
1/*Price*	-2.041×10^{-2}	-17.638***
Amihud	-2.754×10^{-4}	-6.532***
Spread	1.129×10^{-3}	0.066

Observations: 68,276

Results: The Amihud Ratio

- It is well known that illiquidity measures are positively related to transaction costs as well as the volatility and noise in return series.
- However, from the sections above, we notice that the coefficients on the Amihud variable are very unusual.
- The Amihud Ratios in our study are negatively related with the volatility and the noise.
- One possible reason is that daily Amihud measure does not work well with high frequency data.

Results: The Amihud Ratio Cont'd

Table : Comparison of *Amihud* measures in different paper

	Amihud(2002)	Ait-Sahalia and Yu(2009)	Our Research
Stocks	NYSE stocks	NYSE stocks	Dow 30 stocks
Time period	1963-1997	1995-2005	2003-2012
Avg. *Amihud*	0.337		0.041
Avg. *Amihud_sqrt*		0.94-1.43	0.029

This table shows the difference in average *Amihud* value. If we exclude Visa,Inc from the list, the average *Amihud* value of the remaining 29 stocks becomes 2.57×10^{-5}, and the average *Amihud_sqrt* becomes 0.004.

It seems that Amihud measure may not be a sensitive measure for highly liquid stocks.

Conclusion

- In this paper we related stock return volatility estimations and market microstructure noise to its level of ownership by institutional investors and mutual funds.
- We apply a TSRV estimator in estimating the high frequency return volatility and noise.
- The results suggest that institutional /mutual funds ownership are positively correlated with both return volatility and noise.
- This work is robust to firm fixed effect, quarter fixed effect, and data from other period of time.