Microstructure Noise Estimates and Inference Regarding Stock Returns: a Nonparametric Approach

Jalal Seifoddini
Department of Financial Management, Science and Research Branch, Islamic Azad University, Tehran, Iran

*Fraydoon Rahnamay Roodposhti
Department of Financial Management, Science and Research Branch, Islamic Azad University, Tehran, Iran
*Corresponding Author: f-rahnamayroudpshiti@srbiau.ac.ir

Hashem Nikoomaram
Department of Financial Management, Science and Research Branch, Islamic Azad University, Tehran, Iran

Abstract

Numerous studies have demonstrated that noise is crucial for the existence of a liquid financial market. But the existence of noise in prices indicates a temporary deviation in prices from their fundamental values. If the noise level is high in a stock or a portfolio for a period of time, it can be identified as a risk factor. In this paper, after estimating the high frequency microstructure noise of prices through a nonparametric approach, we use a portfolio switching approach, to compare the performance of portfolios having a high level of noise with the performance of portfolios having a lower level of noise. We found out that that the risk of a high level of noise in prices, presents itself as a risk premium in the future return.

Keywords: Microstructure Noise, High Frequency Data, Two Scales Realized Volatility.
1) Introduction

Market microstructure is a branch of financial economics that investigates trading and the structure of markets (Harris, 2003), or in other words, investigates the process of financial price formation at the market (Jong & Rindi, 2009). In general, it can be said that market microstructure, studies the securities transaction mechanism (Vishwanath & Krishnamurti, 2009). In the present paper we study noise in prices as one aspect of financial markets microstructure. Generally, noise in the field of economies is the opposite of information. This noise is sometimes caused by wrong perceptions, and sometimes by false data. Noise can exist anywhere in economy (Black, 1986) but in financial literature it mostly refers to microstructure noise in prices. Noise in prices is a result of structural frictions such as changes in supply and demand. It also appears due to behavioral factors. In general, any form of temporary deviation in price from its fundamental value is called noise. The role of noise in financial markets is both positive and negative. Fisher Black in an article he published in 1986, was one of the first researchers who investigated how noise affects the financial world. He showed that without noise, no financial market would exist because noise creates liquidity in financial markets. In this area, Hu, Pan and Wang believe that temporary price deviations involve important information about the level of liquidity in the overall market (Hu, Pan, & Wang, 2013).

The majority of studies in the field of market microstructure noise have used high frequency data. High frequency financial data, usually refer to data sampled at a time horizon smaller than a trading day. Although, in practice, this definition is not strictly applied, and in some papers, data with daily intervals, have been considered as high frequency data. However, the meaning of "high frequency", has changed over the years following the availability of more and more detailed information on the trading process (Lillo & MiccichÈ, 2010), and also because advances in information processing technologies in recent years have made it possible to process and analyze financial data at scales and frequencies previously unthinkable. This trend is quite obvious in market microstructure data analysis. Now, unlike before when average of prices or transactions values over a time period were used in low frequency analysis, the details of all the transactions can be accessible by researchers, depending on the market under study. This has eliminated the possibility to take advantages of using averages instead of all the data. Statistically, it is clear that data averaging reduces the impact of outlier data in analysis, so the noise in prices would be mitigated, but when all data are used, outlier data are incorporated into analysis which can affect the results or in other words, the results would be noisy. Hence, in this paper we try to determine the magnitude of the noise in the volatility estimates from high frequency data and separate it from the price process. Now, according to some researchers, the existence of the noise in a market, makes the market not be completely efficient. Fisher Black is the first researcher to introduce this hypothesis. He believes that noise causes the market to be inefficient. Noise-based trades, make prices deviate from their fundamental values. Therefore, the more the amount of noise-based trades increase, the more profitable information-based trades will be; but this only happens because prices have more noises (Black, 1986). Other researchers such as Delong et al. (1990) also argue that noise trading can lead to a large divergence between market prices and fundamental values. On the other hand, some researchers such as Friedman (1953), Fama (1965) and Benos (1998) believe that the existence of noise traders, does not necessarily impact price efficiency, rather, with the assumption that the nature of noise is a temporary deviation and not a permanent one, with increased market liquidity, market
efficiency increases. They point out that noise traders are met in the market by rational arbitrageurs who trade against them and in the process drive prices close to fundamental values (Delang et al, 1990). Of course, both groups of researchers consider the existence of noise traders and the noise which can be caused by the activities of this group of traders, to be the vital condition of a liquid market. Because if there are only information traders present in the market, since either side of a deal believe that the other side also trades on information, therefore they are worried that the other side's information may be more accurate than their own, and that’s why the other side has taken an opposite position and consequently they are reluctant to make the deal. In other words, the existence of traders who trade for reasons other than information, provides the required diversity to the market. Thus, noise-based trades are essential for market liquidity (Morawski, 2009). As a result, in this regard there are two hypotheses concerning the noise in the market: “The efficient market hypothesis” and “the noisy market hypothesis” (Bodie, 2009). Therefore, the issue that this research focuses on is that whether market microstructure noise can be explained by asset pricing models like CAPM which is based on the efficient market hypothesis or its existence can generate excess returns which indicate that the market is not efficient. So far, researchers who use high-frequency data, often looked for ways to eliminate these noises in their studies and accordingly, used methods such as simulation and data filtering but, in this study, in order to examine our hypothesis, instead of eliminating the market microstructure noise, we look for a way to disentangle from high frequency observations on the stocks transaction prices a fundamental component and a microstructure noise component. Therefore, based on the research purposes and questions, after estimating the market microstructure noise in prices, the main hypothesis which will be examined is that "whether the high-level of noise in high-frequency price data is priced as a risk premium in stock returns and that this return cannot be explained by efficient market asset pricing models".

There are two approaches for estimating noise. Studies performed in quote-driven markets usually use market makers quotes. They argue that it is common practice in the realized variance literature to use midpoints of bid-ask quotes as measures of the true prices. While these measures are affected by residual noise, they are generally less noisy measures of the efficient prices than are transaction prices since they do not suffer from bid-ask bounce effects. Thus, these studies use midpoints of bid-ask quotes to measure prices, and they generate models of mid-quote determination based on efficient price and residual microstructure noise (see e.g. Bandi & Russell (2006, 2008), Mancino & Sanfelici (2008), Griffin & Oomen (2011). On the other hand, studies performed in order-driven markets use transactions prices. In studies performed to estimate noise by using transactions prices, two types of estimators have been used. The first one is a parametric estimator and the second one is a non-parametric estimator (Ait-Sahalia & Xiu, 2012). Maximum Likelihood Estimation (MLE) is the parametric estimator provided by Ait-Sahalia, Mykland, & Zhang (2005) and the non-parametric estimator is called Two Scales Realized Volatility (TSRV) which is provided by Zhang, Mykland & Ait-Sahalia (2005). To determine that which one is a better estimator of noise in stock prices, Ait-Sahalia & Yu (2009) performed a Monte Carlo simulation. Based on their investigations, in all cases the MLE and TSRV estimators of noise are robust to various types of departures from their model’s basic assumptions under a wide range of simulation design values, including properties of the volatility and the sampling mechanism. However, because TSRV does not depend on the probability distribution, in the present paper, we use TSRV to estimate noise.
After estimating noise, now we ask whether a high level of noise in prices, as a risk factor, is priced in the market, that is, stocks that co-vary with our high-frequency measure of noise tend to get compensated in the form of higher returns. We examine this question through a portfolio switching approach. In the financial field, switching can occur at two levels: the securities and the portfolio level. Switching, at the securities level, means closing a position on a security and taking a position on a more promising security. Switching at the portfolio level, usually in the family of funds, refers to the transition of an investment from a portfolio to another, or in general, the transition from one portfolio to another, is called portfolio switching. This may happen in one or more markets. For example, Gooptu (1994), focuses on the issue of the possibility of portfolio switching between emerging markets and considers it as one of the risks of emerging markets. Piasecki (2004), focuses on the executive costs of transition from one portfolio to another in the process of portfolio switching. Grant's view focuses on the matter that in market timing and portfolio management, what is the number of optimal or ideal times for portfolio switching (Grant, 1978). We use this approach to construct portfolios based on the level of noise in stock prices, and at the end of each month we switch to the portfolio with the highest level of noise.

2) Methodology

We estimate noise through a nonparametric approach, where volatility is left unspecified, stochastic, and we now explain the TSRV approach to separating the fundamental and noise volatilities in this case. In the following, we assume that the process of log prices follows the Itô process as follows (Itô, 2006):

$$dX_t = \mu(X_t; \theta)dt + \sigma dW_t$$

(1)

Where $X_0=0$, $W_t$ represents a Brownian motion, $\mu(.,.)$ is a drift function, $\theta$ represents drift parameter and $\sigma$ is an instantaneous volatility or diffusion coefficient which $\sigma>0$.

Here in the high frequencies context, the drift component is negligible. This is validated empirically: including a drift actually deteriorates the performance of variance estimates from high frequency data since the drift is estimated with a large standard error. So we simplify the analysis one step further by setting $\mu=0$. In their paper published in 2005, Ait-Sahalia and Mykland showed that the removal of these conditions does not eventually change the results (Ait-Sahalia, Mykland, & Zhang, 2005). Therefore, we have:

$$dX_t = \sigma dW_t$$

(2)

The object of interest is now the quadratic variation $\langle X, X \rangle_T = \int_0^T \sigma_t^2 dt$ over a fixed time period $[0,T]$. The usual estimator of $\langle X, X \rangle_T$ is the realized volatility (RV):

$$[Y, Y]_T = \sum_{i=1}^n (Y_{t_i+1} - Y_{t_i})^2$$

(3)

In the absence of noise, $[Y, Y]_T$ consistently estimates $\langle X, X \rangle_T$. However, ignoring market microstructure noise leads to a dangerous situation when $T \to \infty$. After suitable scaling, $RV$ based on the observed log-returns is a consistent and asymptotically normal estimator—but of the quantity $2nE[\epsilon^2]$ rather than of the object of interest, $[Y, Y]_T$. This is of course already visible in the special case of constant volatility. Since the expressions above are exact small-sample ones, they can, in particular, be specialized to analyze the situation where one samples at increasingly higher frequency over a fixed time period. With $N = T/\Delta$, we have
\[ E[\hat{\sigma}^2] = \frac{2n\sigma^2}{\tau} + o(n) = \frac{2nE[\varepsilon^2]}{\tau} + o(n) \]

(6)

\[ \text{Var}[\hat{\sigma}^2] = \frac{2n(6\sigma^4 + 2\text{Cum}_4[\varepsilon])}{\tau^2} + o(n) = \frac{4nE[\varepsilon^4]}{\tau^2} + o(n) \]

(7)

So \((T/2n)[\hat{\sigma}^2]\) becomes an estimator of \(E[\varepsilon^2] = \sigma^2\) whose asymptotic variance is \(E[\varepsilon^4]\). Note, in particular, that \(\hat{\sigma}^2\) estimates the variance of the noise, which is essentially unrelated to the object of interest \(\sigma^2\).

It has long been known that sampling as prescribed by \(\tau\) is not a good idea. The recommendation in the literature has then been to sample sparsely at some lower frequency, by using a realized volatility estimator \([Y, Y]_T^{(sparse)}\) constructed by summing squared log-returns at some lower frequency. Reducing the value of \(n\), has the advantage of reducing the magnitude of the bias term \(2nE[\varepsilon^2]\). However, one of the most basic lessons of statistics is that discarding data is, in general, not advisable.

Zhang, Mykland and Ait-Sahalia (2005) proposed a methodology which makes use of the full data sample, however delivers consistent estimators of both \(\langle X, X \rangle_T\) and \(\alpha^2\). The estimator, Two Scales Realized Volatility (TSRV), is based on subsampling, averaging, and bias-correction. By evaluating the quadratic variation at two different frequencies, averaging the results over the entire sampling, and taking a suitable linear combination of the result at the two frequencies, one obtains a consistent and asymptotically unbiased estimator of \(\langle X, X \rangle_T\).

TSRV’s construction is quite simple: first, partition the original grid of observation times, \(G = \{t_0, ..., t_n\}\) into subsamples, \(G^{(k)}, k = 1, ..., K\), where \(n/K \to \infty\) as \(n \to \infty\). For example, for \(G^{(1)}\) start at the first observation and take an observation every 5 minutes; for \(G^{(2)}\) start at the second observation and take an observation every 5 minutes, etc. Then we average the estimators obtained on the subsamples. So the benefit of subsampling can be retained, whereas the variation of the estimator will be lessened by the averaging. Averaging over the subsamples gives rise to the estimator

\[ \hat{\sigma}^{(avg)} = \frac{1}{K} \sum_{k=1}^{K} [Y, Y]_T^{(k)} \]

constructed by averaging the estimators \([Y, Y]_T^{(k)}\) obtained on \(K\) grids of average size \(\bar{n} = n/K\). While a better estimator than \([Y, Y]_T^{(all)}\), \([Y, Y]_T^{(avg)}\) remains biased. The bias of \([Y, Y]_T^{(avg)}\) is \(2\bar{n}E[\varepsilon^2]\); of course, \(\bar{n} < n\), so progress is being made. But one can go one step further. Indeed, \(E[\varepsilon^2]\) can be consistently approximated using \(RV\) computed with all the observations:

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

(9)

Hence, the bias of \([Y, Y]_T^{(avg)}\) can be consistently estimated by \(\frac{\bar{n}}{n} [Y, Y]_T^{(all)}\). TSRV is the bias-adjusted estimator for \(\langle X, X \rangle\) constructed as

\[ \langle X, X \rangle^{(tsrv)}_T = \frac{\bar{n}}{n} [Y, Y]_T^{(all)} \]

If the number of subsamples is optimally selected as \(K^* = cn^{2/3}\), then TSRV has the following distribution:
Unlike all the previously considered ones, this estimator is now correctly centered. Finally, the noise-to-signal ratio (NSR) which shows the share of noise in the variance, will be as follows.

\[
NSR = \frac{\text{Var}(\text{Noise})}{\text{Var}(\text{Signal})}
\]

After estimating the noise via high frequency price data, we guess that if in a company, the value of noise is high for a period of time, then a premium should have been considered for it, and consequently, portfolios with higher noise, compared to portfolios formed by stocks having a low noise, have a higher return. Also, this extra return should not be of systematic type and explainable by the CAPM. For investigation of this issue, first, we conduct a test for noise effect on return, and if this effect does exist, we sort the sample companies based on the average noise of their last month prices and then categorize them into portfolios sorted based on high to low noise. Next, we calculate the return of the mentioned portfolios in the next month. We repeat this action at the beginning of each month. We calculate the average return for the sorted portfolios and investigate whether by switching from low-noise portfolios to higher noise portfolios, the return level monotonically increases or not. In order to make sure that the portfolio return was not caused by systematic factors and that we can consider noise as its cause, instead of the return of each portfolio, we use the Treynor ratio of each portfolio. The method for calculation of the Treynor ratio is as follows:

\[
T = \frac{r_i - rf}{\beta_i}
\]

Where, \(r_i\) is the monthly return of the \(i^{th}\) portfolio (here the \(i^{th}\) quartile), \(rf\) is the monthly risk-free return and \(\beta_i\) is the beta of the \(i^{th}\) portfolio which will be calculated from the weighted average of betas of the shares forming the portfolio.

3) The data

We perform our study on the shares of companies listed on Tehran Stock Exchange which is an order-driven market with no specialist or market maker. As we need high frequency data, the main criteria in the selection of stocks was that they should have the highest amount of trade volume and the highest number of trading days. Because, we require stocks whose trade volume is so high that it becomes possible to obtain the required observations at the determined frequency. Also, investigating the relationship between noise and stock return, requires stocks that have the lowest number closing days Therefore, in order to select the sample stocks, we use stocks included in the list of 50 most active companies provided by Tehran Stock Exchange. The list of these companies is seasonally announced by Tehran Stock Exchange, which ranks companies based on trade volume and the number of trading days. In our research, we select companies present in this list for four subsequent seasons. The table below, provides the standard deviation, minimum and maximum number of trading
days, number of trades per day, daily trading volume and daily volume of orders of the sample in the time period under study.

<table>
<thead>
<tr>
<th>Stock characteristics</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trading days</td>
<td>221.00</td>
<td>184.00</td>
<td>201.32</td>
<td>8.41</td>
</tr>
<tr>
<td>number of trades per day</td>
<td>13,128.00</td>
<td>40.00</td>
<td>355.17</td>
<td>4.12</td>
</tr>
<tr>
<td>daily trading volume</td>
<td>226,785,635.00</td>
<td>1,940.00</td>
<td>2,533,232.03</td>
<td>7.07</td>
</tr>
<tr>
<td>daily volume of orders placed in the trading system</td>
<td>98,019.00</td>
<td>215.00</td>
<td>12,300.56</td>
<td>9.02</td>
</tr>
<tr>
<td>The average time interval between orders (s)</td>
<td>240.04</td>
<td>00.00</td>
<td>196.03</td>
<td>10.33</td>
</tr>
</tbody>
</table>

4) Findings

Based on our research methodology, we estimated the true realized volatility, noise level and noise to signal ratio, via the TSRV method. The table below shows the maximum, minimum, average and standard deviation of the noise value, RV, and noise to signal ratio of the study sample in the desired period:

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Variable symbol</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise</td>
<td>$\alpha_{i,t}$</td>
<td>0.00220461</td>
<td>0.00425563</td>
</tr>
<tr>
<td>RV</td>
<td>$\sigma_{i,t}$</td>
<td>0.37523582</td>
<td>0.39065411</td>
</tr>
<tr>
<td>noise to signal ratio</td>
<td>NSR$_{i,t}$</td>
<td>0.12156696</td>
<td>1.19247556</td>
</tr>
</tbody>
</table>

The following figure shows the density histogram of the different values of stocks daily noise of all the sample stocks during the study period.

**Figure 1: density of estimated $\alpha_{i,t}$ of the study sample in the determined time period**
The figure below shows the density histogram of various values of daily RV of all the sample stocks during the study period.

**Figure 2: density of estimated $\sigma_{jt}$ of the study sample in the determined time period**

After estimating the microstructure noise in prices, now the question remains that whether this noise is priced in the market. To answer this question, first, in order to ensure the existence of the noise effect on price return, we performed the Granger Causality Test. The table below shows the results of the Granger Causality Test regarding the effect of noise on stocks’ return.

**Table 3: Pairwise Dumitrescu Hurlin Panel Causality Tests (Lags: 1)**

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{return}$ does not homogeneously cause $\alpha$</td>
<td>0.0749</td>
</tr>
<tr>
<td>$\alpha$ does not homogeneously cause $P_{return}$</td>
<td>0.000089</td>
</tr>
</tbody>
</table>
As the results of the Granger Causality Test show, it is possible to reject the Hypothesis of the lack of noise effect on return. Therefore, it can be said that noise affects the stock return. Now that we are sure about the effect of noise on stock return, for investigation of whether stocks with higher noises can get compensated higher returns or not, we consider the return of portfolios categorized based on the noise level. For investigation of this matter at the end of each month, we categorize the stocks, based on their average noise in the previous month, into quartiles from high level of noise to low level of noise. Then, we calculate the weighted return of each portfolio for the next month. We do this every month and switch to the new categorized portfolios. Then, we compare the return of the portfolio having the highest noise with the quartile portfolio having the lowest noise.

We also categorize and investigate the returns of portfolios based on liquidity measures such as trade volume, number of trades, average volume of each trade, orders' volume, and also variables such as price levels, spread and $\sigma$. Before conclusion, first, in order to make sure that the increase in returns was not caused by systematic factors and can be related to noise, instead of portfolio return, we use portfolio Treynor ratio.

Table 4 Treynor ratio of portfolios sorted based on each measure

<table>
<thead>
<tr>
<th>Measures</th>
<th>1(minimum)</th>
<th>2</th>
<th>3</th>
<th>4(maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{j,t}$</td>
<td>0.0057</td>
<td>0.0344</td>
<td>0.0349</td>
<td>0.0527</td>
</tr>
<tr>
<td>NSR$_{j,t}$</td>
<td>0.0057</td>
<td>0.0437</td>
<td>0.6083</td>
<td>0.0007</td>
</tr>
<tr>
<td>$DV_{j,t}$</td>
<td>0.0237</td>
<td>-0.0137</td>
<td>0.0337</td>
<td>0.0507</td>
</tr>
<tr>
<td>DTN$_{j,t}$</td>
<td>0.0367</td>
<td>0.0047</td>
<td>0.0117</td>
<td>0.0397</td>
</tr>
<tr>
<td>DMTV$_{j,t}$</td>
<td>-0.0173</td>
<td>0.0397</td>
<td>0.0227</td>
<td>0.0597</td>
</tr>
<tr>
<td>DWOV$_{j,t}$</td>
<td>0.0317</td>
<td>0.0447</td>
<td>0.0037</td>
<td>0.0197</td>
</tr>
<tr>
<td>$\sigma_{j,t}$</td>
<td>0.0033</td>
<td>0.0307</td>
<td>0.0197</td>
<td>0.0547</td>
</tr>
<tr>
<td>SPREAD$_{j,t}$</td>
<td>0.0317</td>
<td>0.0447</td>
<td>0.0037</td>
<td>0.0197</td>
</tr>
<tr>
<td>PLEVEL$_{j,t}$</td>
<td>0.0384</td>
<td>0.0407</td>
<td>0.0147</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

As table (4) shows, in portfolios sorted based on noise level, by switching from low noise portfolios to high noise portfolios, the portfolio Treynor measure monotonically increases. If the risk caused by high level of noise was explainable by the stocks beta, we should not have observed a significant increase in the portfolios Treynor measures. We do not see this monotonicity in portfolios returns sorted by other measures, therefore, it cannot be considered as related to other considered variables. This shows that the noise contains information that other measures cannot justify. We also did this portfolio switching on a semimonthly and quarterly basis, and the results are similar to the monthly portfolio switching. For brevity, the results are omitted and can be obtained from the authors upon request.

5) Conclusion

In the present research, after estimating the high frequency microstructure noise of prices through a nonparametric approach, we investigated this hypothesis that the risk caused by the existence of noise in prices, due to its relationship with market liquidity and other market risks, would be priced in the market as a risk premium.
We investigated this hypothesis through switching portfolio approach and based our findings we can conclude that if the average noise in prices of a stock is high for a time period, this can be considered a risk for the stock and a premium can exist for it in future stock returns. It was also investigated that this return cannot be explained by full effectiveness market-based pricing models, therefore the hypothesis can be accepted.
References


