MARKET ILLIQUIDITY AND MARKET EXCESS RETURN: 
CROSS-SECTION AND TIME-SERIES EFFECTS

A Study of the Shanghai Stock Exchange

Authors: Hong Xi
Li Weitian

Master Degree Thesis
Supervisor: Lars Pettersson
Co–Supervisor: Anatoliy Malyarenko
Examiner: Sergei Silvestrov

29th November, 2013
Acknowledgments

This thesis is written specifically for all of the colleagues in Financial Engineering at MDH (Mälardalen Högskola) who we have worked with over the years who have stimulated us to understand and become better students.

We sincerely thank our Supervisor Lars Pettersson who has inspired our interest in Portfolio Theory who has given his expertise and best effort in contributing to this work and helping us develop this thesis.

We would like to thank our Co-Supervisor Anatoliy Malyarenko who have inspired our interest in Mathematics and who have given specific guidance on improving the math content of the current thesis.
Abstract
The purpose of the current paper is to explore the cross-sectional relationship between market illiquidity and market excess return on stocks traded in the Shanghai Stock Exchange (SSE) over-time; using data from monthly and yearly databases of CSMAR (China Securities Market and Accounting Research) and statistics annual Shanghai Stock Exchange from 2001.1-2012.12.

We believe that the empirical tests on the stocks traded in the New York Stock Exchange (NYSE) of the well-established paper by Amihud (2002) would be potentially useful to be tested in a different setting, the SSE; in doing so, we apply the same illiquidity measure and estimating models to examine the hypotheses of the current study. In consideration of the aim of the current study, an illiquidity measure proposed by a Chinese scholar Huang (2009) is also applied in the empirical tests.

Due to that Chinese stock market is still young and under development, any outcomes from the current study that are dissimilar to the ones appeared in Amihud (2002) in the sense of the effectiveness of market illiquidity have nothing to do with the utility of illiquidity theory; rather, different market characteristics should be taken into account, such as the unpredictability of frequent policy interventions on a Chinese stock market, following Wang Fang, Han Dong and Jiang Xianglin (2002).

Keywords: Market Illiquidity, Amihud-measure, Huang-measure, China Stock Market
# Table of Contents

1. INTRODUCTION .............................................................................................................. 4  
   1.1 What is Illiquidity ........................................................................................................ 4  
   1.2 Some Measures of Illiquidity and Their Effect on Stock Return ......................... 5  
2. THEORETICAL SECTION .................................................................................................. 6  
   2.1 Theoretical and Mathematical Background .................................................................. 6  
      2.1.1 Law of Large Numbers (LLN) – Strong Law .................................................. 7  
      2.1.2 Central Limit Theorem (CLT) ........................................................................... 7  
      2.1.3 The t-statistic .................................................................................................... 8  
      2.1.4 The p-value ....................................................................................................... 8  
      2.1.5 Autocorrelation Function (ACF) ...................................................................... 9  
      2.1.6 Sample Autocorrelation Function (SAF).............................................................. 9  
      2.1.7 Partial Autocorrelation Function (PACF)............................................................ 9  
      2.1.8 The Autoregressive Model and the AR (1) Model ............................................. 10  
      2.1.9 The ARIMA Model .......................................................................................... 11  
      2.1.10 Cross-sectional regression model and Least squares method ......................... 11  
      2.1.11 R-squared and Adjusted R-squared .................................................................. 12  
      2.1.12 F-statistic ........................................................................................................ 12  
   2.2 Illiquidity Modeling ....................................................................................................... 13  
      2.2.1 Amihud’s Illiquidity Measure ............................................................................ 13  
      2.2.2 Huang’s Illiquidity Measure .............................................................................. 14  
   2.3 Liquidity Proxies .......................................................................................................... 15  
   2.4 A Concise Review of Amihud (2002)’s Cross-sectional Estimation ......................... 16  
   2.5 Data and Data Application ......................................................................................... 19  
   2.6 The Key Constructs .................................................................................................... 22  
3. MAIN SECTION – EMPIRICAL TESTS ............................................................................ 23  
   3.1 The Correlation between Illiquidity and Liquidity Proxies ........................................ 23  
      3.2 Results of the Correlation Test .............................................................................. 24  
      3.2.1 Amihud’s measure ........................................................................................... 24  
      3.2.2 Huang’s measure ............................................................................................. 24  
   3.3 The Cross-Section Effects of Illiquidity on Market Excess Return ......................... 25  
      3.3.1 The Annual Test ................................................................................................ 25  
      3.3.2 The Monthly Test ........................................................................................... 29  
4. RESULTS SECTION .......................................................................................................... 32  
   4.1 Amihud’s measure ....................................................................................................... 32  
   4.2 Huang’s measure ....................................................................................................... 35  
5. CONCLUSION .................................................................................................................. 38  
   6.1 Objective 1 – Knowledge and understanding .............................................................. 40  
   6.2 Objective 2 – Methodological knowledge .................................................................. 40  
   6.3 Objective 3 – Critically and systematically integrate knowledge .............................. 41  
   6.4 Objective 4 – Independently and creatively identify and carry out advanced tasks .... 41
6.5 Objective 5 – Present and discuss conclusions and knowledge ........................................42
6.6 Objective 6 – Scientific, social and ethical aspects ...........................................................42
7. REFERENCES ..................................................................................................................43
8. APPENDIX .....................................................................................................................46
   8.1 Appendix 1: Fama and MacBeth (1973) method ..............................................................46
   8.2 Appendix 2: R-programming analysis of generating the PACF figure .........................47
   8.3 Appendix 3: R-programming analysis of the Correlation Test ...................................49
   8.4 Appendix 4: R-programming analysis of the Empirical Test in Section 3.3 ..............50
1. INTRODUCTION

1.1 What is Illiquidity

The general concept of stock liquidity or, illiquidity, has already been mentioned in many previously published papers. Stated briefly, a stock’s liquidity/illiquidity refers to the ease/difficulty of buying or selling its shares (generally a large amount) without having a significant movement in the stock price or a substantial trading cost, which in turn implies less/more illiquidity risk for both buyers and sellers. And the market illiquidity tells about that overall condition across stocks traded in a stock market.

Good liquidity is one of the key elements for a steady stock market; for investors, a more liquid stock is generally preferred over a less liquid one.

Illiquidity has received certain discussion in financial academia mainly due to its hypothetical positive relationship with stock return. Some scholars within the area (e.g., Amihud et al., 1986) have examined that relationship by using different measures of illiquidity in their previous studies -- overall, a positive relationship between illiquidity and stock return has become evident, suggesting that the higher the illiquidity cost, the more the compensation will be required on stock return. However, according to our literature search, it appears that most published studies focus exclusively on the illiquidity-return relationship within a U.S. stock market; consequently, this paper extends a line of knowledge of illiquidity-return relationship to a Chinese stock market (the Shanghai Stock Exchange – SEE).

Perhaps the most influential study within Illiquidity is conducted by Amihud (2002): Illiquidity and Stock Return, Cross-Section and Time-Series effects. In that paper Amihud initially employed an illiquidity measure, denoted by \( \text{ILLIQ}_t \), into a cross-sectional model to examine the over-time relationship between market illiquidity (ILLIQ across stocks) and market excess return – the risk premium. His study has confirmed the following phenomena on the New York Stock Exchange (NYSE): (i) over time, market expected illiquidity positively affects expected market

---

1. In either way, influences on trade prices from equilibrium prices can help or hurt the investor’s return, however, these differences increase the variability of return the investors will receive and thus are a cost in the sense they increase the investor’s risk (illiquidity).

2. Please refer to Section 1.2.

3. Amihud (2002)’s regression procedure applies the well-known Fama and MacBeth (1973) method; please refer to Section 8.1 – Appendix 1 for the description and history of this method.

4. The risk premium of market return in excess of U.S. Treasury bill rates is used throughout in Amihud (2002). U.S. Treasury bills are considered to have no risk of default, have very short-term maturities (less and equal than a year), and have a known return; they are the most liquid of all money market instruments that traded in active markets, and are the closest approximations available to a riskless investment.
excess return and, thus, suggesting that the expected risk premium partly represents an illiquidity premium; (ii) over time, stock returns are negatively related to contemporaneous unexpected illiquidity. However, another study has shown that the above phenomena cannot be consistently replicated in an Asian Stock Market. Following ‘Illiquidity, illiquidity risk and stock returns: evidence from Japan’ by Jing Fang et al. (2006), market illiquidity has a positive impact on stock returns in Japan in general but not in the second sub-sample period of 1990-1999. They failed to find a significant positive relationship between illiquidity and stock returns in the second sub-sample period of their investigation, while unexpected illiquidity does have a significant negative impact on contemporaneous stock return in the whole period.


Both authors took on good effort for the present thesis. Hong Xi has written and is responsible for Sections 1, 2, and 3. Li Weitian has written and is responsible for Sections 4 and 5. Sections 6, 7 and 8 are done by the two authors. Li Weitian is more active on the programming part than Hong Xi due to reality factors.

1.2 Some Measures of Illiquidity and Their Effect on Stock Return

In this section we do a brief literature review on several measures of illiquidity that have been used in some previously published studies.

According to Amihud and Mendelson (1980) and Amihud (2002), illiquidity reflects the impact of order flow or transaction volume on stock price; this is referred to as the price impact. The higher the price impact, the higher the trading costs or illiquidity costs. There are three major sources of trading costs. First are the direct costs, commission to the brokers plus a tax on the trade. The second is the bid-ask spread, the difference in the bid and ask is a cost to the investor buying and then selling the stock (called round-trip). Third is the potential price impact of a large sale or purchase, which may cause an adverse change in the bid and asks.

There can be specific illiquidity measures depending on different extents of price impact. For example, the price impact of a standard-size transaction is the bid-ask spread, in that case, illiquidity can be directly measured by the bid-ask spread6; in a study of the cross-sectional effect of illiquidity on expected stock return by Amihud and Mendelson (1986), a significant positive relationship between the quoted bid-ask

---

5 The authors of the current paper did not replicate material from other research papers and all citations used in the current study are prerequisite.

6 \( bid - ask \) spread = \( \frac{ask \ price - bid \ price}{bid \ price} \times 100\% \)
spread and stock return has been found. Another example, sometimes groups of informed traders could easily lead to imbalances of order flow\(^7\), which could induce a greater impact on price and a higher illiquidity risk (see Kyle, 1985; Kraus and Stoll, 1972; Keim and Madhavan, 1996). Then *the probability of information-based trading, PIN\(^8\)* is a suited measure for illiquidity, its effect on expected stock return has been found positive and significant both cross-sectionally and across stocks (see Eleswarapu, 1997; Easley et al., 1999). *The Amortized effective spread\(^9\)* was employed as an illiquidity measure by Chalmers and Kadlec (1998), they used quotes and subsequent transactions to generate it and have found that stock return is an increasing function of the amortized effective spread. Brennan and Subrahmanyam (1996) used transactions and quotes from intra-day continuous data to discover *the fixed-cost component of trading* and used it as a measure of illiquidity, and a positive effect of it on expected stock return has been confirmed. *The risk of stock price deviation from its full-information value and the price response to signed order flow* etc. were also found to be positively related to stock return (see Kyle, 1985; Brennan and Subrahmanyam, 1996; Glosten and Harris, 1988; Easley et al., 1999).

### 2. THEORETICAL SECTION

#### 2.1 Theoretical and Mathematical Background

In this Section essential mathematical theorems and models will be illustrated, this is in order to let the readers know what statistical knowledge will be used in the Main Section or the Empirical Tests of the current thesis. At the end of reading this paper readers would understand how the authors are refining the characteristics from the sample of the SSE by using statistical and time-series models. *R*-programming and Excel are applied in estimating data construct and testing all the models.

When we talk about statistics, one should realize that the sample mean or other estimated parameters should be regarded as random variables. In this case, the convergence problem should be considered.

---

\(^7\) The greater the proportion and the higher the quality of the superior information for information-based traders, the more market makers will have to make on the bid-ask spread and the higher the trading cost must be.

\(^8\) \(PIN = \frac{\alpha \mu}{\alpha \mu + 2 \varepsilon}\) where \(\alpha\) is the probability of an information event that taking place, \(\mu\) is the arrival rates of the news to the informed trader, \(\varepsilon\) is the arrival rates of the news to the uninformed trader.

\(^9\) \(AS_T = \frac{\sum_{t=1}^{T} |P_t - M_t| \times V_t}{P_T \times Shares\; Outstanding_T}\) where \(P_t\) is transaction price, \(M_t\) is midpoint of the prevailing bid-ask quote, \(V_t\) is the number of shares traded, \(P_T \times Shares\; Outstanding_T\) is firm’s market value of equity at the end of day \(T\).
2.1.1 Law of Large Numbers (LLN) – Strong Law

In probability theory, the Law of Large Numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

Let $X_1, X_2, \ldots$ be a sequence of independent identical random variables (i.i.d.).

Let $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 < \infty$. Define the sample average $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Then,

$$P \left( \lim_{n \to \infty} \bar{X}_n = \mu \right) = 1. \quad (1)$$

That is, under general conditions, the sample mean approaches the population mean as $n \to \infty$ with probability 1.

2.1.2 Central Limit Theorem (CLT)

The Central Limit Theorem states that the probability distribution function for mean values of a number of samples is approximately a normal distribution regardless of the actual population distribution, if sample size is large.

Let $X_1, X_2, \ldots$ be a sequence of independent identical random variables (i.d.d.).

Let $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 > 0$. Define $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Let $G_n(x)$ denote the cumulative density function of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any $x, -\infty < x < +\infty$,

$$\lim_{n \to \infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy. \quad (2)$$

That is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution $N(0,1)$.

By the Central Limit Theorem, the sample covariances are asymptotically normal, and using least squares to fit the parameters to these covariances we obtain estimates that are also asymptotically normal. Note: In real world situation, sometimes the sample size is small compare to the whole population and population standard deviation is not used. Then the sample statistics from the sampling distribution does not conform to the normal distribution but to the t-distribution.

In modeling and analysis, it is important to use statistics to exam the significance of the estimated parameter after fitting a model.

2.1.3 The t-statistic

In statistics, the t-statistic is a ratio of the departure of an estimated parameter from its notional value and its standard error. Let $\hat{\beta}$ be an estimator of parameter $\beta$ in some statistical model. Then a t-statistic for this parameter is any quantity of the form:
\[ t_{\hat{\beta}_n} = \frac{\hat{\beta}_n - \beta_0}{\text{s.e.} (\hat{\beta}_n)} \]  

(3)

where \( \beta_0 \) is a non-random, known constant, and s.e. \((\hat{\beta}_n)\) is the standard error of the estimator \( \hat{\beta}_n \).

By default, statistical packages report t-statistic with \( \beta_0 = 0 \) (these t-statistics are used to test the significance of the estimated coefficient). To check the significance of corresponding repressors, one should check the table of t-distribution with respect to \( t_{\hat{\beta}_n} \). By using LLN and CLT, and as we consider \( \hat{\beta}_n \) and \( t_{\hat{\beta}_n} \) as random variables, when \( n \to \infty \), \( t_{\hat{\beta}_n} \) will converge to \( N(0,1) \) weakly. In that case, we are able to draw some conclusion of the whole population based on the information from our sample.

2.1.4 The p-value

A p-value \( p(X) \) is a test statistic satisfying \( 0 \leq p(X) \leq 1 \) for every sample point \( x \). Small values of \( p(X) \) give evidence that \( H_1 \) is true. A p-value is valid if, for every \( \theta \in \Theta_0 \) and every \( 0 \leq \alpha \leq 1 \),

\[ P_{\theta}(p(X) \leq \alpha) \leq \alpha. \]

If \( p(X) \) is a valid p-value, it is easy to construct a level \( \alpha \) test based on \( p(X) \). The test will reject \( H_0 \) if and only if \( p(X) \leq \alpha \). The p-value can be found on the t-distribution table or normal distribution table with regard to the sample size.

The following example shows the t-statistic and p-value of estimated coefficients by using \( R \):

```
> ar1 intercept
    coef   4.66846e-01  -0.4328928
    s.e.  7.330069e-02  0.2144431
    t ratio 6.366715e+00  -2.0172848
    p-value 1.931191e-10  0.0436658
```

To check the significance of the estimated parameter or coefficient, \( R \) will extract the standard error (s.e.) from the diagonal of variance-covariance matrix, i.e. \( 7.330069 \times 10^{-2} \). Then by using equation (3), the t-statistic of a coefficient will be calculated as \( \frac{4.66846 \times 10^{-1}}{7.330069 \times 10^{-2}} = 6.366715 \). As shown, \( R \) will find the p-value for that coefficient with respect to the t-statistic of 6.366715.

We should also comprehend and use the functions of ACF, SAF, and the very important PACF figure, they are defined below.
2.1.5 Autocorrelation Function (ACF)

When the linear dependence between \( r_t \) and its past values \( r_{t-1} \) is of interest, the concept of the correlation is generalized to autocorrelation (i.e., the correlation of \( r_t \) with its previous value \( r_{t-1} \)). The correlation coefficient between \( r_t \) and \( r_{t-1} \) is called the lag-1 autocorrelation of \( r_t \) and is commonly denoted by \( \rho_l \):

\[
\rho_l = \frac{\text{Cov}(r_t, r_{t-1})}{\sqrt{\text{Var}(r_t)\text{Var}(r_{t-1})}} = \frac{\gamma_l}{\gamma_0}.
\]  

(4)

2.1.6 Sample Autocorrelation Function (SAF)

For a given sample of returns \( \{r_t\}_{t=1}^T \), let \( \bar{r}_t \) be the sample mean (i.e., \( \bar{r} = \frac{\sum_{t=1}^T r_t}{T} \)). Then the lag-1 sample autocorrelation of \( r_t \) is:

\[
\hat{\rho}_l = \frac{\sum_{t=2}^T (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}.
\]  

(5)

2.1.7 Partial Autocorrelation Function (PACF)

The partial autocorrelation function (PACF) plays an important role in time series analysis. By plotting the PACF figures one could determine the order \( p \) of an AR or the extended ARIMA (\( p, d, \) and \( q \)) model.

Given a time series \( r_t \) and denote \( \alpha(l) \) as the partial autocorrelation of lag \( l \),

\[
\alpha(1) = \text{Cor}(r_t, r_{t+1}),
\]

\[
\alpha(l) = \text{Cor}\left(r_{t+l} - P_{t,l}(r_{t+l}), r_t - P_{t,l}(r_t)\right), \text{ for } l \geq 2,
\]  

(6)

where \( P_{t,l}(x) \) denotes the projection of \( x \) onto the space spanned by: \( r_{t+1}, \ldots, r_{t+l-1} \).

An approximate test that a given partial autocorrelation is significant at 5% significant level is given by comparing the sample partial autocorrelations against the critical region (dotted line on the following figure) with upper and lower limits given by \( \pm 1.96/\sqrt{n} \), where \( n \) is the number of the points of the time series being analyzed.
Figure 1 and Figure 2 are PACF figures generated by R, in the Empirical Tests (the Main Section) we will point out that the above PACF figures strongly indicate our data (market illiquidity measures) match the AR (1) model. For this detail please refer to Section 3.3.2.

Next we will introduce the AR Model, the ARIMA Model as well as the relevant lag-1 auto regression model – the AR (1) model.

2.1.8 The Autoregressive Model and the AR (1) Model

In statistics, an autoregressive (AR) model is a representation of a type of random process; as such, it describes certain time-varying processes in nature, economics, etc. The autoregressive model specifies that the output variable depends linearly on its own previous values.

The notation AR(p) indicates an autoregressive model of order p. The AR(p) model is defined as:

$$ r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \alpha_t, $$

where $\phi_1 \ldots \phi_p$ are the parameters of the model, $\phi_0$ is a constant, $\alpha_t$ is the white noise.

A linear time series $r_t$ follows the autoregressive model of order 1 (or just AR (1) model) if $\phi_2 = \phi_3 = \cdots = 0$. The coefficients $\psi$ is called the impulse responses of $r_t$, which describes the reaction of dynamic system as a function of time or possible as a function of some other independent variable that parameterizes the dynamic behavior of the system. It is common to denote $\phi_0 = \mu, \phi_1 = \psi_1$ and write this model as:

$$ r_t = \phi_0 + \phi_1 r_{t-1} + \alpha_t $$

10 Please refer to Section 8.2 – Appendix 2 for the R-programming code analysis of the PACF figure – Figure 1.
If you observe model (27) in our Main Section (Empirical Tests) you would find that it is the same form of an AR (1) model. If one assumes a set of data follows an AR (1) model, one could generate a PACF figure to observe whether that set of data fits the autoregressive model of order 1 before fitting the data into the AR (1) Model.

2.1.9 The ARIMA Model

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. The model is generally referred to as an ARIMA(p, d, q) model where parameters p, d, and q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

Given a time series of data \( r_t \), where \( t \) is an integer index and \( r_t \) are real numbers, then an ARMA(p, q) model is given by:

\[
(1 - \sum_{i=1}^{p} \phi_i L^i) r_t = (1 + \sum_{i=1}^{q} \theta_i L^i) \alpha_t,
\]

where \( L \) is the lag operator, \( \phi_i \) are the parameters of autoregressive part of the model, \( \theta_i \) are the parameters of moving average part of the model, and \( \alpha_t \) is the white noise.

Since the aim of this paper is to estimate the cross-sectional relationship between market illiquidity and market excess return, we shall fully present this model as well as impactful statistical methods according to our empirical tests both here and in Sections 3 and 4 (the Main Section and the Results Section).

2.1.10 Cross-sectional regression model and Least squares method

Suppose that there are \( k \) assets and \( T \) time periods. Let \( r_{it} \) be the return of asset \( i \) in the time period \( t \). A general form for the factor model is:

\[
r_{it} = \alpha_i + \beta_{ij}f_{jt} + \cdots + \beta_{im}f_{mt} + \epsilon_{it}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, k,
\]

where \( \alpha_i \) is a constant representing the intercept, \( \{f_{ij} | j = 1, \ldots, m\} \) are \( m \) common factors, \( \beta_{ij} \) is the factor loading for asset \( i \) on the \( j \)th factor, and \( \epsilon_{it} \) is the specific factor of asset \( i \).

Readers can later observe that models (31), (32), (39) and (40) in Section 3.3 are the final cross-sectional and estimating models of testing the relationship between illiquidity factors and market excess return over time.

In matrix form, the factor model above can be written as:

\[
r_{it} = \alpha_i + \beta_i f_t + \epsilon_{it},
\]

where \( \beta_i = (\beta_{i1}, \ldots, \beta_{im}) \) is a row vector of loadings, and the joint model for the \( k \) assets (market excess return of all the stocks on the SSE) at time \( t \) is \( r_t = \alpha + \beta f_t + \epsilon_t \), \( t = 1, \ldots, T \). Here \( r_t = (r_{1t}, \ldots, r_{kt}) \), \( \alpha = (\alpha_1, \ldots, \alpha_k) \), \( \beta = [\beta_{ij}] \) is a \( k \times m \) loading matrix, and \( \epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{kt}) \) is the error vector.
The model presenting above is in a cross-sectional regression form if the factors $f_{jt}, j = 1, \ldots, m$ are observed. One can fit a cross-section model in $R$ with its build-in function ‘lm’. Function ‘lm’ is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance. In this case, one could use the method of least squares, which is a standard approach to the approximate solution of a regression model. As defined before, $\epsilon_{it}$ stands for the residual of the regression model and it is the difference between an observed value and the fitted value provided by the model:

$$\epsilon_{it} = r_{it} - (\alpha_i + \beta_i f_t).$$

(12)

Define $S$ as the sum of the squared residuals:

$$S = \sum_{i=1}^{n} \epsilon_{it}^2.$$  

(13)

The least squares method will find the optimum solution of the coefficient of the regression model when $S$ is a minimum.\(^{11}\)

2.1.11 R-squared and Adjusted R-squared

In statistics, the coefficient of determination, denoted by $R^2$, indicates how well data points fit a line or curve. It is a statistic used in the context of statistical models whose main purpose is either the prediction of future outcomes or the testing of hypotheses on the basis of other related information,

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}},$$

(14a)

where $SS_{res}$ is the sum of squares of residuals, and $SS_{tot}$ is the total sum of squares.

The use of an adjusted $R^2$ is an attempt to take account of the phenomenon of the $R^2$ automatically and spuriously increasing when extra explanatory variables are added to the model,

$$R^2 = R^2 - (1 - R^2) \frac{p}{n - p - 1},$$

(14b)

where $p$ is the total number of regressors in the model, and $n$ is the sample size.

2.1.12 F-statistic

An F-statistic test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled. Exact F-tests mainly arise when the models have been fitted to the data using least squares,

\(^{11}\) Currently, $R$ software will use QR decomposition to solve the least squares problems. See:  
\[ F = \frac{\text{explained variable}}{\text{Unexplained variable}} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}, \]  

where \( \hat{y}_i \) is the predicted value, \( \bar{y} \) is the sample mean, \( \hat{\varepsilon}_i \) is the residual.

Again, it is worth mentioning that all of the essential statistical knowledge that is used in our empirical tests will be explained more demonstrably and intuitively in the Empirical Tests and Results Sections of the current paper.

2.2 Illiquidity Modeling

All aforementioned illiquidity measures in the preceding section (Section 1.2) are, in essence, fine illiquidity measures, for each of them represents a particular aspect of illiquidity. Yet, constructing them requires a great amount of microstructure data that are not readily available over long periods of time among most databases (including the CSMAR databases). Due to the key purpose of Amihud (2002) and the current paper as well – exploring the relationship between market illiquidity and market excess return over time, an illiquidity measure generated by data that are available and accessible over time would preferably be the option. Thereupon we are going to discuss two measures that meet with the purpose.

2.2.1 Amihud’s Illiquidity Measure

Below we introduce the key construct in Amihud (2002).

Amihud (2002) initially applied the following most often quoted illiquidity measure:

\[ ILLIQ_i = \frac{|R_{iyd}|}{VOLD_{iyd}}, \]  

\( ILLIQ_i \) is reasonable to be used for empirical tests because of the following reasons.

First, it is a relative general measure as compared to the ones described in Section 1.2, as it estimates the average ratio of daily stock return (\(|R_{iyd}|\): the absolute return of stock \( i \) in day \( d \) of year \( y \)) to each dollar of their trading volume (\( VOLD_{iyd} \)), which can be seen as the everyday average price impact of a stock on its trading volume. We will explain this description in more detail in Section 2.6. Second, it has been found that \( ILLIQ_i \) is positively and significantly related to two of the aforementioned illiquidity measures – the price response to signed order flow and the bid-ask spread – \( ILLIQ_i \) is positively and strongly related to microstructure estimates of illiquidity. Brennan and Subrahmanyam (1996) used two measures of illiquidity, obtained from data on intraday transactions and quotes: Kyle’s \( \lambda \), the price impact measure, and \( \psi \), the fixed-cost component related to the bid-ask spread. The estimates are done by the method of Glosten and Harris (1988). The table 5 in their paper shows a regression
result that the coefficient of corresponding parameters is positive and significant. Mathematically, for a fitted model, statistical significance refers to whether the outcome of an observation is an effective one or simply due to chance. If an outcome meets the standard for statistical significance, we say that this outcome is an effective one which is unlikely occurred by chance. The last but not the least, $ILLIQ_i$ can be generated by data on returns and volumes that are readily obtainable in databases of most stock markets over time, including the SSE.

### 2.2.2 Huang’s Illiquidity Measure

In consideration of the aim and interest of the current paper, we will recommend a new measure of illiquidity proposed by a Chinese scholar.

In Huang et al. (2009): *A Discussion about Chinese Stock Market and Liquidity*, there is an argument about Amihud-ratio and a new measure based on the structure of $ILLIQ_i$ has been suggested. The new measure considers the effect of the turnover rate of a stock on its price amplitude:

$$
ILLIQ_i^H = \frac{S_{i,y,d}}{Turnover_{i,y,d}} = \frac{High_{i,y,d} - Low_{i,y,d}}{Vol_{i,y,d} / OUTSIZE_{i,y,d}},
$$

(17)

In the above equation $S_{i,y,d}$ is the price amplitude of stock $i$ in day $d$ of year $y$, price amplitude measures the ratio of the highest price minus the lowest price of stock $i$ in day $d$ of year $y$ to its *closing* price on the day before $d$, $d-1$. $Turnover_{i,y,d}$ measures the turnover rate of stock $i$ in day $d$ of year $y$. Huang’s illiquidity measure virtually implies that if a stock is lack of liquidity then even a low turnover rate could induce high fluctuations in its price amplitude, hence a higher illiquidity.

Huang (2009) argues that the essence of measuring illiquidity is to look at how trading activities affect stock price, whereas Amihud’s ratio uses the daily price return to indicate such influence, which in itself has noticeable estimation error. The reason is that using daily price return as a measure of price changing also involves price changes caused by non-trading factors, for instance, during the non-trading hours of a stock (i.e. the time interval between the closing and the following day's opening)\(^1\). In fact, a number of factors could lead to a stock’s opening price to be different from its yesterday’s closing price (either higher, lower, or a price gap\(^1\)), such as a news of issuing a positive earnings announcement, or a news regarding the fundamental variation in the value of a corporation, etc. Yet using price amplitude to measure price movements would allow us to observe price changes due to factors before the

---

\(^1\) The daily price return of stock $i$ is the ratio of the highest price minus the lowest price of stock $i$ in day $d$ to its *opening* price on the same day.

\(^2\) For instance, a stock might shoot up from a closing price of $20 a share, marking the high point of an $18–$20 trading range for that day, and begin trading in a $22–$24 range the next day on the news of a takeover bid. Or a company that reports lower than expected earnings might drop from the $18–$20 range to the $13–$15 range without ever trading at intervening prices.
non-trading hours of a stock and after the opening of a stock, which subtly avoids price movements accounted for by non-trading factors.

Also in view of the higher discrepancy between outstanding shares and issued shares on Chinese stock market, Huang aptly suggests that using stock turnover rate instead of trading volume could be more suitable in determining the market illiquidity (the overall illiquidity condition across stocks) on a Chinese stock market. By this time, it is good to know that both turnover rate and price amplitude are readily available in databases of the SSE over time. Moreover, although trading volume may be a major determinant for illiquidity or liquidity, its close relationship with stock market value could make the measuring of stock illiquidity/liquidity be partially accounted for by the size of stocks rather than simply the order flow. Later in Section 2.4, readers would see that, size, being a proxy of liquidity has been found negatively correlated with Amihud’s illiquidity measure.

Due to these viewpoints we will additionally employ Huang’s illiquidity measure into the same empirical tests for Amihud-ratio in the current paper, that is, from measuring illiquidity construct to testing its correlations with other liquidity proxies, and then the examination of its over-time cross-sectional relationship with market excess return.

By this time we have presented a number of illiquidity measures. However, due to the inherent complexity of market mechanics as well as various external factors from time to time (there are too many unpredictable external factors in a Chinese financial environment), none of an existing illiquidity measure can, at the same time (i) capture every aspect of price impact, and (ii) be estimated by data that are available over long periods of time. In short, the actual illiquidity of a stock or a market is impossible to be interpreted in a single cogent construct.

2.3 Liquidity Proxies

Now we are going to introduce three widely used liquidity proxies. Seeing that these liquidity proxies are often used as the benchmark of liquidity in many stock markets, introducing them and testing their correlations with illiquidity measures could help us exploring how these proxies interact with illiquidity on a Chinese stock market. Moreover, it is reasonable to assume that stock return is negatively related to stock liquidity.

First, size, it is the market value or capitalization of a stock (stock market price multiplied by the number of outstanding shares). Large stocks with lower bid-ask spreads are often thought to be less risky than small stocks, by seeing that many larger stock issues have smaller price impact on order flow, Banz (1981), Reinganum (1981), and Fama- French (1992) have found that stock expected return is negatively related to size. Another proxy of liquidity that uses data on outstanding shares is stock

\[ \text{Capital}(\text{Size}) = \text{shares price} \times \text{shares outstanding} \]
turnover rate\(^{15}\), which is the ratio of trading volume to the number of shares outstanding. Stock turnover rate reflects transaction efficiency, and a number of studies have discovered that a stock’s return is decreasing in its turnover rate (See Haugen and Baker, 1996; Datar et al., 1998; Hu, 1997a; Rouwenhorst, 1998; Chordia et al., 2001). Third, trading volume\(^{16}\), Brennan et al. (1998) have found that stock dollar volume has the same cross-sectional negative effect as turnover rate on stock return.\(^{17}\)

In Amihud (2002), these three liquidity proxies or stocks characteristics were found negatively related with his illiquidity measure for stocks traded on the NYSE. Therefore, similar correlation tests will be carried out for stocks traded on the SSE using both Amihud’s and Huang’s measures, and the mathematical tool will be illustrated in Section 3.1.1.

2.4 A Concise Review of Amihud (2002)’s Cross-sectional Estimation

Even though the main purpose of Amihud (2002) is to examine the time-series and cross-sectional effects of illiquidity on stock return, however, a test for the relationship between illiquidity and stock return when other stock characteristics (including stock illiquidity) are entered into the cross-section model was, made before. This test will be included for review consideration and the test results contained in Table 1 will be discussed both in this section and in Section 2.6.

This test involves the cross-sectional regressions of monthly stock return on seven stock characteristics where stock illiquidity is included. The data source is CRSP (Center for Research of Securities Prices of the University of Chicago) databases for the NYSE traded stocks over the period of 1963-1997. Stocks need to comply with the following criteria to exclude penny stocks and outstanding loans for less bias and errors in a statistical sense:\(^{18}\)

(a) Those stocks have data on returns and volumes for more than 200 days during each the year of the estimation period and should have listing records at the end of the year.
(b) Those stocks have prices higher than $5 at the end of the years of the estimation period.
(c) Those stocks that have market capitalization on CRSP at the end of each year during the estimation period.
(d) Eliminating stocks that have illiquidity in the upper and lower 1% tails among the

---

\(^{15}\) turnover rate = \(\frac{\text{trading volume}}{\text{shares outstanding}} \times 100\%

\(^{16}\) trading volume (in dollar or other currency) = trading price \(\times\) trading volume

\(^{17}\) Mathematically, a negative (or positive) relation or negative (or positive) effect means that the regression test of the cross-sectional model gave negative (or positive) coefficients for the corresponding liquidity proxies or parameters.

\(^{18}\) As a result, 1061 – 2291 NYSE stocks matched the criteria for the cross-sectional estimation.
distribution of the individual values after meeting (a), (b), and (c).

Below are the seven stock characteristics (annual values) and their assumed relationships with stock return:

1. **ILLIQ**\(_{iy}\) (multiplied by \(10^6\)). Stock illiquidity is assumed to be positively related with stock return. The yearly illiquidity for each stock \(i\) is (recall (16)):

\[
ILLIQ_{iy} = \frac{1}{D_{iy}} \sum_{t=1}^{D_{iy}} \frac{|R_{iyd}|}{VOLD_{iyd}},
\]

where \(D_{iy}\) is the number of days that data are available for stock \(i\) in year \(y\).

2. **Beta**\(_{iy}\). Since beta measures the systematic risk of stocks, it is assumed to have a positive effect on stock return.

3. **SDRET**\(_{iy}\) (multiplied by \(10^2\)). Standard deviation will not be well diversified if one’s portfolio is constrained, thus is assumed to be positively related with stock return.\(^{19}\)

4. **DIVYLD**\(_{iy}\). Its relationship with stock return can be either way.\(^{20}\)

5. **Size**\(_{iy}\). As mentioned in the section of Liquidity Proxies, size, being a lieu of liquidity has been found negatively related with stock return.

6. Stock past return, **R**\(_{100iy}\). Stock \(i\)'s average return during the last 100 days of year \(y\).

7. **R**\(_{100YR}_{iy}\). Stock \(i\)'s average return over the rest of the period, between the beginning of the year and 100 days before its ends. Past stock return is assumed to have a positive effect on stock return.

The regression procedure applies the well-known Fama and MacBeth (1973) method. A cross-sectional model is regressed on monthly stock returns (over the period of January 1964 – December 1997, 34 years, a total of 408 months) as a function of the seven stock characteristics \((J = 7)\) for stock \(i\):

\[
R_{imy} = k_{omy} + \sum_{j=1}^{7} K_{jimy} X_{ji,y-1} + U_{imy}, \tag{19}
\]

where,

- \(R_{imy}\) — The return of stock \(i\) in month \(m\) of year \(y\) (there will be 408 monthly returns for each stock).
- \(X_{ji,y-1}\) — The seven characteristic \(j, j = 1, 2, ..., 7\) of stock \(i\) (annual values), they are estimated in year \(y-1\), and known to investors at the beginning of year \(y\) so that they can make investment decision from that time.
- \(K_{jimy}\) — The coefficients \(K_{jimy} = 1, 2, ..., 7\), they measure the effects of stock characteristics on monthly stock returns (408 monthly regressions performed for each stock characteristic).
- \(U_{imy}\) — Residuals.

\(^{19}\) It is also possible that standard deviation has a negative effect on stock return. As tax trading option suggests, higher volatile stocks should generate lower expected return.

\(^{20}\) Higher dividends usually impose higher tax than capital gains do, in this case compensation are required for higher stock returns; on the other hand, dividend yield may has a negative effect on returns if it is negatively related with an unobserved risk factor.
After generating the monthly estimates of each coefficient \( K_{jimy} = 1, 2, ..., 7 \) for stock \( i \), the means of these estimates can be found simply by averaging the individual values, then the average values of these means for all chosen stocks will be calculated. As a result, we will obtain seven ‘market coefficients’. The cross-sectional test is also performed for the means that exclude the January coefficients, following Keim et al. (1983 and 1986) that excluding January could make the effects of beta, size and bid-ask spread insignificant. Table 1 illustrates the results.

<table>
<thead>
<tr>
<th>Variable</th>
<th>All months</th>
<th>Excl. January</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.922</td>
<td>1.568</td>
</tr>
<tr>
<td></td>
<td>(4.06)</td>
<td>(3.32)</td>
</tr>
<tr>
<td>BETA</td>
<td>0.217</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td>(0.64)</td>
<td>(0.79)</td>
</tr>
<tr>
<td>LnSIZE</td>
<td>–0.134</td>
<td>–0.073</td>
</tr>
<tr>
<td></td>
<td>(3.50)</td>
<td>(2.00)</td>
</tr>
<tr>
<td>SDRET</td>
<td>–0.179</td>
<td>–0.274</td>
</tr>
<tr>
<td></td>
<td>(1.90)</td>
<td>(2.89)</td>
</tr>
<tr>
<td>DIVYLD</td>
<td>–0.048</td>
<td>–0.063</td>
</tr>
<tr>
<td></td>
<td>(3.36)</td>
<td>(4.28)</td>
</tr>
<tr>
<td>R100</td>
<td>0.888</td>
<td>1.335</td>
</tr>
<tr>
<td></td>
<td>(3.70)</td>
<td>(6.19)</td>
</tr>
<tr>
<td>R100YR</td>
<td>0.359</td>
<td>0.439</td>
</tr>
<tr>
<td></td>
<td>(3.40)</td>
<td>(4.27)</td>
</tr>
<tr>
<td>ILLIQ</td>
<td>0.112</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td>(5.39)</td>
<td>(4.91)</td>
</tr>
</tbody>
</table>

Examining Table 1 reveals that stock illiquidity is a positive price of risk, which means that Amihud’s \( ILLIQ \) is positive related with stock return. Its coefficient remains positive and significant (0.112 with t-statistic of 5.39) while other stock characteristics are involved. Excluding January didn’t make illiquidity insignificant, yet, once again with an acceptable t-statistic. Furthermore, the relationships between the seven stock characteristics and stock return were as anticipated.

The empirical evidence in Table 1 strengthened the *illiquidity theory* and the role of illiquidity risk in asset pricing model. Perhaps the most obvious opportunity for further research is to conduct a study of a more straightforward cross-sectional analysis between stock illiquidity and stock return while utilizing the knowledge of *time-series*. This is exactly the proposition of Amihud (2002) and the current study as well.

---

2.5 Data and Data Application

The purpose of this section is not only to indicate who will be the target sample data of the current investigation but also to analyze why they were chosen, their quality, shortage etc.

Amihud (2002) used 1061-2291 chosen stocks based on the four criteria to determine the ‘market illiquidity’ (the key construct) for his empirical tests, i.e. he calculates the illiquidity value for each stock among the 1061-2291, adding up all the values and averaging them to obtain a ‘market illiquidity’ that explains the overall illiquidity condition among the chosen stocks for a certain period. Since the authors of this paper do not have the access to advanced databases that allow searching out each qualified stock according to the four criteria in Amihud (2002), please refer to Section 2.4. Therefore, picking out each stock from the whole sample would become a manual screening with quite possibilities of inaccuracies that could be made, then we would be better off just using the data based on the whole sample. Such as the market price index, the market volume index, the market turnover ratio, in estimating the market illiquidity and its cross-sectional over-time effects on market excess return. All required market data in this study are obtained from the monthly and yearly CSMAR and statistics annual Shanghai Stock Exchange databases.22

In doing so we are in an attempt to see how the ‘real market illiquidity’ affects the expected market excess return on a stock market. The advantage is that we will be able to explore the ‘real cross-section and over-time effects’ on the whole sample of the stocks, which could in a sense increase the generalizability of the estimation results. The disadvantage is that using the whole sample cannot avoid possible statistical bias and errors, hence, theoretically, may weaken the effectiveness of illiquidity measure during empirical tests.

Our investigation takes the time from January 2001 to December 2012, a time period of 12 years, a total of 144 months. The reason for investigating from 2001 is interpreted below.

Today the most representative quantitative index of Chinese stock market is the SSE Composite Index.23 It began to release to the public not until the early 90s. At the meantime, China’s economy has experienced a successful transformation from central planning economy to market economy. Even so, China kept holding an ambiguous attitude to the reforms of all industries for years, and the features of the

22 CSMAR website: (http://www.gtarsc.com/)

statistics annual Shanghai Stock Exchange website:
(http://www.sse.com.cn/researchpublications/publication/yearly/)

23 The SSE Composite Index is a market weighted index of all stocks (A shares and B shares) that are traded in the SSE. It’s a price index that does not include dividends.
reforms were ‘reforming without announcements’. Chinese government held the same attitude to the establishment of stock exchanges, and even the mainstream media was not allowed to broadcast any information about the market to the public. In nearly ten years since the stock market was first opened, many brokers made full use of asymmetric information and even gain the first chances in illegal profiteering. These issues have been gradually suppressed with the supervision of the CSRC (China Securities Regulatory Commission) as well as the discipline of the market itself, laws and regulations etc. In the 21st century, after China's accession to the WTO, state-owned shares initially began to circulate in the stock market.

The Shanghai Stock Exchange was established in 1990 and began its operation in December at the same year. Through years of operation it has become the most preeminent stock market in Mainland China. We believe that after 10 years of its development as well as the overall market condition from that time and on, the trading mechanism would be better, the market data provided by the monthly and yearly CSMAR databases and statistics annual Shanghai Stock Exchange databases should be more reliable to be used for empirical tests. Given that the Chinese stock market is still young and still under development, we must learn and use these knowledge more carefully.

Table 2 illustrates the number of listed stocks on the SSE of each year of 2001-2012.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stocks</td>
<td>744</td>
<td>828</td>
<td>913</td>
<td>996</td>
<td>886</td>
<td>878</td>
<td>904</td>
<td>908</td>
<td>914</td>
<td>935</td>
<td>978</td>
<td>998</td>
</tr>
</tbody>
</table>

(Source: China Securities Market and Accounting Research databases – CSMAR)

In Figure 3 below we have plotted the monthly behavior of the SSE Composite Index from January 2001 to December 2012 and Figure 4 illustrates the monthly market illiquidity estimated by Amihud-ratio over the same period. Both figures are generated in R. (Data Source: Databases of CSMAR and statistics annual Shanghai Stock Exchange)
As shown in Figure 3, between 2001 and 2005, due to the aftermath effect of the Asian Financial Crisis, the SSE Composite Index continuously presented a downward trend and reached its lowest point in 2005 (998 points). In 2006 China started a ‘share - merger reform’, stock index began a sharp increase and peaked in 2007 (6124 points). However, with the outbreak of Global Financial Crisis, stock index have been pounded and quickly fell to 1664 points in 2008. In order to protect the overall economy Chinese government launched an ‘Economic Stimulus’ scheme with four trillion RMB, stock market also benefited from it, the SSE Composite Index rose to 3000 points in the year of 2009.

From Figure 4 we can observe that the ‘ups and downs’ of monthly market illiquidity is approximately consistent with the ‘downs and ups’ of monthly price index in Figure 3, that is, over time, market performance is consistent with illiquidity risk. The monthly market illiquidity reached its highest point in 2005 with the stock index reached its lowest point at the same time, and it reached very low point in 2007 while the stock index reached its record high. Some economic factors behind the ‘ups and downs’ of the market illiquidity will be briefly discussed.

Between 2001 and 2006, as Chinese capital market and banking system gradually opened themselves to the outside world and as the reformed stock market became more and more close to international standards, the number of domestic and foreign investments in Chinese stock market ushered in ongoing growth. Even though market illiquidity kept high for years since 2001 but it declined rapidly after 2005, it then fluctuated over the 2006 – 2012 period, however, low values in general. At the end of 2007, the U.S. Subprime Mortgage Crisis gave rise to a global economic downturn; however, the market illiquidity on the SSE seemed not to be greatly affected.
2.6 The Key Constructs

In Sections 2.2.1 and 2.2.2 we have introduced Amihud’s and Huang’s daily illiquidity measures for stock $i$, see (16) and (17). The current section presents these two measures in terms of market data (averaged data on the whole sample of stocks).

Before continuing we want to point out an ambiguously defined variable in Amihud-ratio.

Amihud (2002) noted that $illiq_i = \frac{|r_{iyd}|}{vold_{iyd}}$ indicates the absolute rate of price change per dollar of trading volume, where $|r_{iyd}|$ represents the absolute stock return and is used as the absolute rate of price change of each stock. As known, a stock’s return is normally regarded as the sum of its rate of price change and its rate of dividends, for dividends are any income received over a stock’s holding period. If the numerator of Amihud-ratio contained information of dividends, then how could it be used as a measure of the absolute rate of price change? The authors of the current paper received a response from Professor Yakov Amihud – he had tried to observe whether there was any information in the price that includes dividends, but there was not. Then the next thing to consider is that whether the ‘absolute stock return’ should contain any information of dividends. The answer can be found if we take a look at the cross-sectional results in Section 2.4. As shown, the coefficient of the dividend yield is a negative number with t-statistic of 3.36, suggesting that dividend is negatively and significantly related with stock return. Thus using an illiquidity measure that includes dividends would influence the true estimating relationship between that measure and stock return. Note that now we should be aware of using a market price index (the SSE Composite Index) that does not include stock return from dividends in calculating the price change rate on stocks, since the so-called ‘absolute stock return’ shouldn’t contain dividends.

As has been mentioned in Section 2.5, due to not having the access to advanced databases, we will be using market or summary data in calculating our key constructs – market illiquidity.

Since we will be examining the yearly and monthly relationships between market illiquidity and market excess return, only the monthly and annual market illiquidity will be generated. Therefore the numerator of Amihud-ratio would become the absolute rate of the market price change in month $m$ of year $y$ and in year $y$, or $|R_{My}|$ and $|R_{My}|$. And the denominator of Amihud-ratio would become the market volume in month $m$ of year $y$ and in year $y$, or $vold_{My}$ and $vold_{My}$. All these market data are obtainable from the monthly and yearly CSMAR databases over the period of 2001 January to 2012 December and we should ‘absolute’ each market price change rate afterwards. To compute the rate of the market price change rate, CSMAR takes the change in the SEE Composite Index over a day, a month, or a year and divide the value of the index at the beginning of that day, that month, or that year.
The monthly and yearly illiquidity constructs of Amihud-ratio are presented below:

\[ \text{MILLIQ}_{Mym} = \frac{|R_{Mym}|}{VOLD_{Mym}} \text{ (Monthly)}, \]  
\[ \text{AILLIQ}_{My} = \frac{|R_{My}|}{VOLD_{My}} \text{ (Yearly)}. \]  

(20) \hspace{1cm} (21)

In constructing the monthly and yearly Huang-ratio, we apply the monthly and yearly data of the ‘market turnover rate’ obtained directly from statistics annual Shanghai Stock Exchange databases over the period of 2001 January to 2012 December. According to our work, the monthly and yearly values of ‘market price amplitude’ are calculated by Excel.

The monthly and yearly illiquidity constructs of Huang-ratio should be:

\[ \text{MILLIQ}_{Mym}^{H} = \frac{S_{Mym}}{Turnover_{Mym}} \text{ (Monthly)}, \]  
\[ \text{AILLIQ}_{My}^{H} = \frac{S_{My}}{Turnover_{My}} \text{ (Yearly)}. \]  

(22) \hspace{1cm} (23)

All the estimated monthly and yearly illiquidity values will be multiplied by \(10^9\) for empirical tests.

3. MAIN SECTION-EMPIRICAL TESTS

3.1 The Correlation between Illiquidity and Liquidity Proxies

In Section 2.3 we have mentioned three widely used liquidity proxies; they are size, turnover rate, and trading volume. In this section we take correlation tests between illiquidity measures and liquidity proxies for stocks traded on the SSE utilizing both Amihud and Huang’s ratios. Pearson product-moment correlation coefficient is the tool that we used in testing the correlation coefficient.

The correlation coefficient between two random variables \(X\) and \(Y\) is defined as:

\[ \rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{E(X-\mu_X)^2E(Y-\mu_Y)^2}}. \]  

(24)

It measures the strength of dependence between \(X\) and \(Y\), where \(\mu_X\) and \(\mu_Y\) are the means of \(X\) and \(Y\), and \(-1 \leq \rho_{X,Y} \leq 1\).
Moreover, \( R \) will estimate a p-value for each of the testing result to perform the check of the hypothesis ‘the correlation coefficient is negative’ against the alternative hypothesis ‘the correlation coefficient is zero’.

The annual market data of size, turnover, and trading volume will be directly applied for the correlation test. Readers can refer to the preceding section for the annual Amihud-ratio and Huang-ratio constructs as they are presented in (21) and (23) respectively.

According to the estimation, \( E[(X - \mu_x)(Y - \mu_y)] \) is the expected value of the product of ‘two deviations’: (i) the deviation of annual market size, turnover, and trading volume from their means, and (ii) the deviation of annual Amihud’s and Huang’s ratios from their means. Due to China’s stock market is still young and under development we prefer 0.1(p-value) as the significance level for all of our empirical tests in hopes of increasing the effectiveness of illiquidity measure.

### 3.2 Results of the Correlation Test

#### 3.2.1 Amihud’s measure

The average correlations between the three liquidity proxies and Amihud’s measure as well as their inter-correlations are contained in Table 3.

<table>
<thead>
<tr>
<th>Table 3&lt;sup&gt;24&lt;/sup&gt;</th>
<th>The inter-correlation matrix: Amihud’s measure and liquidity proxies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amihud ILLIQ</strong></td>
<td>Turnover</td>
</tr>
<tr>
<td>Turnover</td>
<td>-0.1399</td>
</tr>
<tr>
<td>p-value</td>
<td>(0.6644)</td>
</tr>
<tr>
<td>Trading Volume</td>
<td>-0.9143</td>
</tr>
<tr>
<td>p-value</td>
<td>(3.155 \times 10^{-5})</td>
</tr>
<tr>
<td>Capitalization</td>
<td>-0.8805</td>
</tr>
<tr>
<td>p-value</td>
<td>(1.563 \times 10^{-4})</td>
</tr>
</tbody>
</table>

Examining Table 3 reveals that the correlations between Amihud’s ratio and the three liquidity proxies are all negative values, but the average correlation coefficient between Amihud ILLIQ and turnover rate is not significant negative for the stock traded on the SSE.

#### 3.2.2 Huang’s measure

The average correlations between the three liquidity proxies and Huang’s measure

<sup>24</sup> Please refer to Section 8.3 – Appendix 3 for the R-programming code analysis of the correlation test for Table 3.
as well as their inter-correlations are illustrated in Table 4.

<table>
<thead>
<tr>
<th>Table 4&lt;sup&gt;25&lt;/sup&gt;</th>
</tr>
</thead>
</table>

The inter-correlation matrix: Huang’s measure and liquidity proxies

<table>
<thead>
<tr>
<th></th>
<th>Huang ILLIQ</th>
<th>Turnover</th>
<th>Trading Vol.</th>
<th>Capitalization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Huang ILLIQ</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Turnover</strong></td>
<td>-0.7639</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>(0.077)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trading volume</strong></td>
<td>-0.9230</td>
<td>0.4813</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>(0.0087)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Capitalization</strong></td>
<td>-0.7221</td>
<td>0.7155</td>
<td>0.6479</td>
<td>1</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>(0.10)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown, the correlations between Huang’s measure and the three liquidity proxies are all negative and statistically significant. The negative relationships are even stronger between Huang ILLIQ and turnover rate, and between Huang ILLIQ and trading volume than those in Table 3. In that sense Huang’s ratio appears to be more effective for stocks traded on the SSE.

With regard to Table 3 and Table 4, the correlation results are consistent with the common sense that the larger the size, the larger the trading volume or the higher the turnover rate of a market, the better the market liquidity.

### 3.3 The Cross-Section Effects of Illiquidity on Market Excess Return

#### 3.3.1 The Annual Test

The claim of the current paper is that over time, expected market illiquidity positively affects expected market excess return (market return in excess of the risk-free interest rate<sup>26</sup>). As the aggregate of investors have anticipated higher risk of market illiquidity, they will price stocks to generate higher expected return for compensation, in other words, that higher expected return is an illiquidity premium.

We will make a hypothesis that the expected market excess return is an increasing function of the expected market illiquidity, since it stands to reason that the expected return on stocks in excess of the risk-free interest rate should be considered as compensation for illiquidity risk in addition to its standard interpretation as compensation for risk – illiquidity is also a prevailing risk among stocks and stock markets.

Our final cross-section estimating model will follow the methodology of French et al. (1987) who tested the effect of risk on stock excess return and the well-known

<sup>25</sup> Please refer to Section 8.3 – Appendix 3 for R-programming code analysis of the Correlation Test.

<sup>26</sup> This paper uses the one-year interest rate as the risk-free interest rate.
The expected effect of market illiquidity on expected market excess return is described by the model:

\[
E(RM_y - Rf_y) = f_0 + f_1 \ln \text{AILLIQ}_{My}^E,
\]

where

- \( RM_y \) — The annual market return of year \( y \).
- \( Rf_y \) — The one-year interest rate of year \( y \).
- \( E(RM_y - Rf_y) \) — The expected market excess return in year \( y \).
- \( \ln \text{AILLIQ}_{My}^E \) — The expected market illiquidity in year \( y \) (Amihud’s ratio) based on information in year \( y-1 \).

Our tests will use the logarithmic transformation of market illiquidity:

\[
\ln \text{AILLIQ}_{My}^E \text{ (multiplied by } 10^9).\]

Investors are assumed to predict the market illiquidity for year \( y \) based on information available in year \( y-1 \) and use the prediction to set prices that will generate the desired expected return for year \( y \). The higher the expectation on stock return, the more it will be in setting the price. The model of the expected market illiquidity is presented in (26):

\[
\ln \text{AILLIQ}_{My}^E = c_0 + c_1 \ln \text{AILLIQ}_{My-1}^E.\]

It is reasonable to assume that the coefficient \( c_1 > 0 \), that is, the expected market illiquidity in year \( y \) is an increasing function of the actual market illiquidity in year \( y-1 \). \( c_1 > 0 \) also implies that the higher the unexpected illiquidity appeared in year \( y-1 \), the higher the expectation on market illiquidity for the following year. The underlying rationale is that once the market exists unexpected illiquidity in year \( y-1 \) that is beyond the former prediction of illiquidity in year \( y-2 \), that 'shock' could lead to predicting the market illiquidity for year \( y \) more carefully — investors would rather believe that there will be a higher market illiquidity in year \( y \) than in year \( y-1 \), or they would rather over-predict the market illiquidity for year \( y \) since they don’t want to make any loss due to under-predicting illiquidity risk.

The market illiquidity of each year is assumed to follow a time-series autoregressive model. It is reasonable to assume \( c_1 > 0 \) – higher illiquidity in one year should be associated with higher illiquidity in the following year:

\[
\ln \text{AILLIQ}_{My} = c_0 + c_1 \ln \text{AILLIQ}_{My-1} + v_{My},
\]

where \( v_{My} \) is the residual of the unexpected market illiquidity in year \( y \).

In Section 2.1.10 we have mentioned that model (27) matches the form of an AR (1) model. Since we assume that the market illiquidity follows such a model, we then
need to check weather our data (market illiquidity over the investigation period 2001.1-2012.12) fit model (27), or mathematically, the AR (1) model. In doing so we should generate and observe the PACF figure of our data. For brevity we will give illustration of two PACF figures for both ratios (Amihud’s and Huang’s) using the data from the monthly illiquidity - a greater sample of data of 144 months, readers can refer to the next Section of the Monthly Test.

Recall model (25) of the expected market excess return for year \( y \) and if we substitute (26) into (25), we will find the following:

\[
E(RM_y - Rf_y) = f_0 + f_1 c_0 + f_1 c_1 \ln AILLIQ_{MY-1}.
\] (28)

Next follows the assumed model for the actual market excess return in year \( y \):

\[
(RM_y - Rf_y) = f_0 + f_1 \ln AILLIQ_{MY}^E + u_{MY} = g_0 + g_1 \ln AILLIQ_{MY-1} + u_{MY},
\] (29)

where

\[
g_0 = f_0 + f_1 c_0 \quad \text{and} \quad g_1 = f_1 c_1.
\]

\( u_{MY} \) – Unexpected market excess return in year \( y \). Parts of unexpected excess return in year \( y \) are due to the contemporaneous unexpected market illiquidity, the \( v_{MY} \) in (27).

Next we will make another hypothesis that the effect of unexpected market illiquidity in year \( y \) on contemporaneous unexpected market excess return of year \( y \) is negative – unexpected market illiquidity in one year would lead to a decrease in market return over the same period. The argument is that the unexpected illiquidity appeared in year \( y \) raises the expected illiquidity and thus the expected return in year \( y+1 \); in doing so the aggregate of investors in year \( y \) will price stocks down to make the expected return for the following year a higher one.\(^{27}\) Thus results in an unexpected negative excess return (a decrease in the stock price) in year \( y \). The actual unexpected market excess return in year \( y \) is interpreted below:

\[
u_{MY} = g_2 \ln AILLIQ^U_{MY} + w_{MY},
\] (30)

where

\[
\ln AILLIQ^U_{MY} \quad - \quad \text{Unexpected illiquidity in year } y, \text{ the } v_{MY} \text{ in (27)}.
\]

\( w_{MY} \) – Unexpected market excess return in year \( y \) caused by other unexpected factors over the same period.

If we take (30) into (29) and we will obtain our final cross-section estimating model using Amihud-ratio (Annual Test):

\[
(RM_y - Rf_y) = g_0 + g_1 \ln AILLIQ_{MY-1}^U + g_2 \ln AILLIQ^U_{MY} + w_{MY}.
\] (31)

\(^{27}\) Assuming there is no relation between corporate cash flows and market liquidity.
Likewise, the final cross-section estimating model using Huang-ratio (Annual Test) is:

\[
(RM_y - Rf_y) = g_0 + g_1 \ln AILLIQ^H_{MY-1} + g_2 \ln AILLIQ^{HU}_{MY} + w_{MY}.
\]  

(32)

In the above equations coefficients \(g_1\) and \(g_2\) suggest two hypotheses as we have discussed:

H – 1: \(g_1 > 0\), and
H – 2: \(g_2 < 0\).

Hypothesis 1: The coefficient \(g_1\) should be positive (significant) – expected market excess return is an increasing function of expected market illiquidity.
Hypothesis 2: The coefficient \(g_2\) should be negative (significant) – unexpected market illiquidity has a negative effect on contemporaneous unexpected market excess return.28

Moreover, with reference to previous studies, the estimated coefficient \(c_1\) was biased downward in estimating the time-series autoregressive model (model (27)). Kendall’s (1954) suggested a bias correction approximation procedure by which the estimated coefficient \(c_1\) is enhanced by the term \((1 + 3c_1)/T\), where T is the sample size. The same method is applied in the current paper to adjust \(c_1\). Hence the procedure is to first calculates the residual value \(v_{MY}\) in (27) after its coefficients are adjusted by \((1 + 3c_1)/T\) and then to use the adjusted residual value into model (31) as \(\ln AILLIQ^{MY}_{MY}\) to estimate the coefficients \(g_1\) and \(g_2\) in testing the two hypotheses. The same procedure will be done for Huang-ratio as well.

The estimation of model (27) after adjusting \(c_1\) provides the following result for Amihud-ratio:29

\[
\ln AILLIQ_{MY} = 0.094 + 1.081 \ln AILLIQ_{MY-1} + v_{MY}.
\]  

(33)

For Huang-ratio:

\[
\ln AILLIQ^{H}_{MY} = 0.092 + 0.056 \ln AILLIQ^{H}_{MY-1} + v_{MY}.
\]  

(34)

By applying Kendall’s (1954) bias correction method, the bias-corrected estimated slope coefficients \(c_1\) are 1.081 and 0.056 respectively in above models, suggesting that higher illiquidity in one year raises expected illiquidity for the following year. It is therefore reasonable to proceed with the adjusted coefficients that are estimated

28 This hypothesis is consistent with the findings of Amihud et al. (1990).
29 For brevity we directly give the estimation results of after adjusting \(c_1\) for both ratios of their annual data, but we will give more specific estimation information and explanation for both ratios of their monthly data in the Section of the Monthly Test, Section 3.3.2.
using the entire data.

The results of estimating (31) and (32) for stocks traded on the SSE are presented in the Result Section - Section 4.

3.3.2 The Monthly Test

The same methodology and procedure in the above section is replicated here using monthly data.

The monthly market illiquidity data of both ratios is also assumed to follow a time-series autoregressive model.

We have employed R to generate the PACF figures for the monthly illiquidity data of both ratios over the investigation period of January 2001 – December 2012. As mentioned before, this is in order to check whether our data match such a model before fitting them into the model, and the two figures below indicate that, for both Amihud and Huang ratios, the line representing lag 1 is significant. Then we are confident with fitting our market illiquidity data estimated by Amihud’s and Huang’s measures into the AR (1) model. Readers can refer to Sections 2.1.7 and 2.1.8.

![PACF of Amihud-ratio](image1)
![PACF of Huang-ratio](image2)

The monthly version of model (27) is as follows using Amihud-ratio:

\[
\ln MILLIQ_{Mym} = -0.433 + 0.467 \ln MILLIQ_{M(y_{m-1})} + \nu_{Mym}. \tag{35}
\]

\[
(t = -2.017, \quad R^2 = 0.22, \quad p = 0.0437, \quad 1.931 \times 10^{-10}).
\]

As shown here are three statistics: t-statistic, p-value and R-squared. As we have explained in the Theoretical Section, t-statistic measures how many standard errors

\[30\] Please refer to Section 8.2 – Appendix 2 for the R-programming code analysis of this PACF figure.

29
the estimated coefficient is away from zero. Generally, any t-value greater than +2 or less than -2 is acceptable, and any t-value between -2 to +2 indicates the low reliability of the predictive power of the coefficient.

p-value measures the probability of the estimated coefficient happened by chance. Ordinarily, a p-value of 0.05 is considered to be meeting the standard statistical significance, it means that the outcome of an observation or estimated coefficient is 5% likely by chance and 95% likely resulted by no chance. An advantage of reporting a test result via p-value is that each reader can choose the significance level he or she considers appropriate and then compare the p-value to that level and know whether a fitted model lead to acceptance or rejection of Hypothesis H₀ (happened by chance) after observing its estimated coefficients.

R-squared, the coefficient of determination, is defined as the ratio of the regression sum of squares to the total sum of squares. The coefficient of determination measures the proportion of the total variation of the data that is explained by the fitted line.

As supported by the p-value and t-statistic, the coefficient 0.467 in equation (35) is a positive and strongly significant value, which indicates that investors in the SSE predict market illiquidity for month m as an increasing function of the market illiquidity in month m-1. Indeed, the p-value of this coefficient with respect of the t-statistic is \(1.931 \times 10^{-10}\), which indicates that this coefficient we got is significantly not happened by chance. And the t-value also indicates that this coefficient is acceptable. The R-squared shows that the above AR (1) model can explain 22 percent of the data point.

By applying Kendall’s (1954) bias correction method the bias-corrected estimated model of (35) looks like:

\[
\ln \text{MILLIQ}_{\text{Mym}} = -0.483 + 0.435 \ln \text{MILLIQ}_{\text{M(ym-1)}} + v_{\text{Mym}}.
\]  (36)

As shown the adjusted coefficient still presents a positive number in (36).

The monthly version of model (27) is as follows using Huang-ratio:

\[
\ln \text{MILLIQ}^H_{\text{Mym}} = 0.024 + 0.624 \ln \text{MILLIQ}^H_{\text{M(ym-1)}} + v_{\text{Mym}}.
\]  (37)

\(t = \) \(6.324\) \(7.700\), \(R^2 = 0.347\).

\(p = \) \(2.545 \times 10^{-10}\) \(1.36 \times 10^{-14}\).

The results from both t-statistic and p-value indicate that model (37) is statistically significant, representing that the expected market illiquidity for month m is an increasing function of the market illiquidity appeared in m-1. The p-value of this coefficient is \(1.36 \times 10^{-14}\), which indicates that this coefficient is nearly impossible happened by chance. And the t-statistic also indicates that this coefficient is acceptable. The R-squared shows that the above AR (1) model can explain 34.7 percent of the data point.
By applying Kendall’s (1954) bias correction method the bias-corrected estimated model of (37) looks like:

\[ \ln \text{MILLIQ}_{M_{ym}}^H = 0.039 + 0.664 \ln \text{MILLIQ}_{M_{(ym-1)}}^H + v_{Mym}. \] (38)

Again the adjusted coefficient presents a positive number in (38).

After applying Kendall’s (1954) bias correction method the monthly unexpected illiquidity (the residuals) in the adjusted models (36) and (38) will be estimated and they will then be used in the monthly final cross-section estimating models for both ratios.

The monthly final cross-section estimating models of the two ratios are presented below,

Amihud-ratio:

\[ (R_{M_{ym}} - R_{f_{ym}}) = g_0 + g_1 \ln \text{MILLIQ}_{M_{(ym-1)}}^H + g_2 \ln \text{MILLIQ}_{Mym}^U + g_3 \text{JANDUM}_{My} + w_{Mym}. \] (39)

Huang-ratio:

\[ (R_{M_{ym}} - R_{f_{ym}}) = g_0 + g_1 \ln \text{MILLIQ}_{M_{(ym-1)}}^H + g_2 \ln \text{MILLIQ}_{Mym}^{HU} + g_3 \text{JANDUM}_{My} + w_{Mym}. \] (40)

Here \( g_1 \) and \( g_2 \) suggest the same hypotheses as illustrated in the preceding section, with \( g_1 > 0 \), and \( g_2 < 0 \).

As you can see we are adding a \(- g_3 \text{JANDUM}_y\) to the monthly versions, a January dummy, that accounts for the well-known January effect. And the results of estimating models (39) and (40) are presented in Section 4.
4. RESULTS SECTION

4.1 Amihud’s measure

Figure 5 reports the effect of market illiquidity on market excess return generated from the annual cross-section estimating model (31):

\[(RM_y - Rf_y) = g_0 + g_1 \ln AIIQ_{MY-1} + g_2 \ln AIIQ_{MY}^U + w_{MY}\]

Market illiquidity and Market excess return-Annual data-Amihud

<table>
<thead>
<tr>
<th>Call:</th>
<th>lm(formula = y ~ logILLIQ1 + UlogILLIQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals:</td>
<td>Min     1Q   Median     3Q   Max</td>
</tr>
<tr>
<td></td>
<td>-0.04723 -0.01868    0.01282  0.01668  0.03506</td>
</tr>
<tr>
<td>Coefficients:</td>
<td>Estimate Std. Error  t value Pr(&gt;</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-0.003587    0.008723 -0.411  0.69172</td>
</tr>
<tr>
<td>logILLIQ1</td>
<td>-0.003056    0.008832 -0.346  0.73828</td>
</tr>
<tr>
<td>UlogILLIQ</td>
<td>-0.061581    0.013862 -4.442  0.00216 **</td>
</tr>
<tr>
<td>---</td>
<td>Signif. codes: 0 ‘<em><strong>’ 0.001 ‘</strong>’ 0.01 ‘</em>’ 0.05 ‘.’ 0.1 ‘ ’ 1</td>
</tr>
<tr>
<td>Residual standard error: 0.02786 on 8 degrees of freedom</td>
<td></td>
</tr>
<tr>
<td>Multiple R-squared: 0.7378, Adjusted R-squared: 0.6722</td>
<td></td>
</tr>
<tr>
<td>F-statistic: 11.25 on 2 and 8 DF, p-value: 0.004728</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5

As shown in Figure 5, the value of the first-quantile of residual is -0.01868, which means 1/4 of the residual is less than this value. The median of the residual is 0.01282 which means that if we rank all the residual from minimum to maximum, 0.01282 is the one lies in the middle of them. The third-quantile of the residual is 0.01648, which means 1/4 of the residual is greater than this value.

The estimated intercept is -0.003587 with standard error of 0.008723. By using equation (3) we could get the corresponding t-statistic equals to \(-\frac{-0.003587}{0.008723} \approx -0.411\), using \(R\) we can also obtain the corresponding p-value equals to 0.69172 with respect to the t-statistic of -0.411.

The estimated coefficient of the market illiquidity in year \(y-1\) is -0.003056 with standard error of 0.008832. Thus the t-statistic is \(-\frac{-0.003056}{0.008832} \approx -0.346\), and the p-value is 0.73828 with respect to the t-statistic of -0.346.
However, we have the estimated coefficient of the unexpected illiquidity in year $y$ equals to -0.061581 with a very significant t-statistic of -4.442 and a very significant p-value of 0.00216 ** with respect to the t-statistic of -4.442.

The residual standard error is 0.02786, which is close to zero, indicates that the quality of the fitted model is good. The R-squared shows that the above model can explain 73.78% of the data point. However, some of the data may be explained by chance, therefore, adjusted R-squared suggests that this model can explain 67% of the data after a correction. The F-statistic is 11.25 with p-value equals to 0.004728, which indicates that $H_0: R^2 = 0$ should be rejected on the 0.01 significance level, in other words, this model gives a significant good fit to the corresponding data and has a good explaining power.

Figure 6 reports the effect of market illiquidity on market excess return generated from the monthly cross-section estimating model (39):

$$(RM_{ym} - RF_{ym}) = g_0 + g_1 \ln MILLIQ_{M(ym-1)} + g_2 \ln MILLIQ_{Mym}^U + g_3 \text{JANDUM}_{Mym} + w_{Mym}.$$ 

Marked illiquidity and Market excess return - Monthly data - Amihud

---

The value of the first-quantile of residuals is -0.05098, which means 1/4 of the residual is less than this value. The median of the residual is -0.01675 which means that if we rank all the residual from minimum to maximum, -0.01675 is the one lies in the middle of them. The third-quantile of residuals is 0.05430, which means 1/4 of the residual is greater than this value.

---

Please refer to Section 8.4 – Appendix 4 for checking the R-programming code analysis of the time-series and cross-sectional estimations for Figure 6.
The estimated intercept is $0.005760$ with standard error of $0.007944$. By using (3) we could get the corresponding t-statistic equals to $\frac{0.005760}{0.007944} \approx 0.725$ and we have also obtained the corresponding p-value equals to $0.4696$ with respect to the t-statistic of $0.725$.

The estimated coefficient of the market illiquidity in month $m-1$ is $0.003805$ with standard error of $0.004695$, thus the t-statistic equals to $\frac{0.003805}{0.004695} \approx 0.810$, and the p-value equals to $0.4191$ with respect to the t-statistic of $0.810$.

However, we have the estimated coefficient of the unexpected illiquidity in month $m$ equals to $-0.013350$ with a significant t-statistic of $-2.505$ and a significant p-value of $0.0134 \ast$ with respect to the t-statistic of $-2.505$.

The estimated coefficient of the January dummy is $0.020627$ with insignificant t-statistic of $0.745$ and p-value of $0.4577$.

The residual standard error is $0.08797$, which is close to zero and indicates the quality of the fitted model is good. The R-squared shows that the above model can explain $5\%$ of the data. However, some of the data may be explained by chance, therefore, adjusted R-squared suggests that this model can only explain $3\%$ of the data after a correction, the explaining power of this model is very low. The F-statistic is $2.467$ with p-value equals to $0.06473$ indicates that $H_0: R^2 = 0$ should be rejected on the $0.1$ significance level. In short, model (39) gives a good fit to the corresponding data but a low explaining power.

Through time-series and cross section regressions, it is shown that only the coefficients of unexpected illiquidity are significant in Figure 5 and 6, where the average coefficient $g_2$ from the annual regressions is $-0.061581$ with a p-value of $0.00216 \ast\ast$ (significant at $0.01$ level) and the average coefficient $g_2$ from the monthly regressions is $-0.013350$ with a p-value of $0.0134 \ast$ (significant at $0.05$ level). Since unexpected illiquidity and market excess return presented significant negative relationship, it thus suggesting that Hypothesis 2 (please refer to Section 3.3) of unexpected market illiquidity negatively affects contemporaneous unexpected market excess return holds in the SSE. Both estimated coefficients of market illiquidity ($g_1$) in year $y-1$ and month $m-1$ shown insignificant values and even with a negative $g_1$ in Figure 5. However it is worth noting that the reason might not be that expected illiquidity does not positively affect expected market return on the SSE, the reasons, however, could be:

1. Frequent interventions. According to Wang Fang, Han Dong and Jiang Xianglin (2002), any unpredictable information regarding policies or important events contained in data provided by CSMAR would make the correlation between market illiquidity and market excess return become unstable or even show opposite results – negative value. In that sense the effect of unexpected market illiquidity on contemporaneous market excess return would likely to become relative significant, as shown in both Figures 5 and 6.
2. According to Huang’s arguments, Amihud-ratio in itself has ineluctable
measurement defects and may not be the best choice in estimating stock or market illiquidity for stocks traded on a Chinese stock market. For more details, please refer to Section 2.2.2.

3. Using the data of market illiquidity based on the whole sample of stocks may, theoretically, weaken the effectiveness of illiquidity measure during empirical estimations; readers can refer to Section 2.5.

4.2 Huang’s measure

Figure 7 reports the effect of market illiquidity on market excess return generated from the annual cross-section estimating model (32):

$$(RM_y - Rf_y) = g_0 + g_1 \ln AILLIQ^H_{My-1} + g_2 \ln AILLIQ^{HU}_{My} + w_{My}.$$ 

**Market illiquidity and Market excess return-Annual data-Huang**

```
Call: lm(formula = yRETURN ~ yILLIQN + yUILLIQN)

Residuals:
   1       2       3       4       5
-0.008194 -0.004938  0.015042 -0.008690  0.006780

Coefficients: 
 Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.10704   0.01765   -6.065   0.0261 *
yILLIQN       0.13843   0.02222    6.230   0.0248 *
yUILLIQN     -0.08745   0.02310   -3.786   0.0632 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01402 on 2 degrees of freedom
Multiple R-squared: 0.9587,   Adjusted R-squared: 0.9174
F-statistic: 23.21 on 2 and 2 DF,  p-value: 0.04131
```

**Figure 7**

The estimated intercept is -0.10704 with standard error of 0.01765. By using equation (3) we could get the corresponding t-statistic equals to $\frac{-0.10704}{0.0165} \approx -6.065$ and then $R$ will find the corresponding p-value, which equals to 0.0261 * with respect to the t-statistic of -6.065.

The estimated coefficient of the market illiquidity in year y-1 is 0.13843 with standard error of 0.02222. Its t-statistic is $\frac{0.13843}{0.02222} \approx 6.230$ and the p-value is 0.0248 * with respect to the t-statistic of 6.230, they are all very significant.

By the same way, we have the estimated coefficient of the unexpected illiquidity
equals to \(-0.08745\) with a significant t-statistic of \(-3.786\) and a significant p-value of 0.0632 \(\cdot\) with respect to the t-statistic of \(-3.786\).

The residual standard error is 0.01482, which is close to zero, indicates that the quality of the fitted model is good. The R-squared shows that the above model can explain 95.87\% of the data points. However, some of the data may be explained by chance, therefore, adjusted R-squared suggests that this model can explain 91.74 \% of the data after a correction. The F-statistic is 23.21 with p-value equals to 0.04131 indicates that \(H_0: R^2 = 0\) should be rejected on the 0.05 significance level. Therefore, model (32) gives a significant good fit to the corresponding data with a very strong explaining power.

Figure 8 reports the effect of market illiquidity on market excess return generated from the monthly cross-section estimating model (40):

\[
(RM_{ym} - Rf_{ym}) = g_0 + g_1 \ln M\text{ILLIQ}^H_{M(ym-1)} + g_2 \ln A\text{ILLIQ}^H_{Mym} + g_3 J\text{ANDU}M_{Mym} + w_{Mym}.
\]

**Market illiquidity and Market excess return - Monthly data - Huang**

| Call: lm(formula = y ~ ILLIQ + UILLIQ + mon2) |
| Residuals: Min 1Q Median 3Q Max |
| -0.23161 -0.05159 -0.01081 0.06623 0.22886 |
| Coefficients: Estimate Std. Error t value Pr(>|t|) |
| (Intercept) 0.038213 0.021804 1.753 0.084254 |
| ILLIQ -0.035733 0.022214 -1.609 0.112405 |
| UILLIQ -0.097505 0.027610 -3.531 0.000753 *** |
| mon2 -0.009286 0.045270 -0.205 0.838096 |
| --- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 |
| Residual standard error: 0.0974 on 67 degrees of freedom |
| Multiple R-squared: 0.1777, Adjusted R-squared: 0.1409 |
| F-statistic: 4.827 on 3 and 67 DF, p-value: 0.004213 |

**Figure 8**

The estimated intercept is 0.038213 with standard error of 0.021804. By using (3) we could get the corresponding t-statistic equals to \(\frac{0.038213}{0.021804} \approx 1.753\) and \(R\) will find the p-value equals to 0.084254 \(\cdot\) with respect to the t-statistic of 1.753.

The estimated coefficient of the market illiquidity is -0.035733 with standard error of 0.022214. Its t-statistic is \(\frac{-0.035733}{0.022214} \approx -1.609\) and the p-value is 0.112405 with respect to the t-statistic of -1.609.
As shown, we have the estimated coefficient of unexpected illiquidity equals to -0.097505 with a significant t-statistic of -3.531 and a very significant p-value of 0.000753*** with respect to the t-statistic of -3.531.

The estimated coefficient of the January dummy is -0.009286 with t-statistic of -0.205 and p-value of 0.838098.

The residual standard error is 0.0974, which is close to zero, indicates that the quality of the fitted model is good. The R-squared shows that the above model can explain 17.77 percent of the data. However, some of the data may be explained by chance, therefore, adjusted R-squared suggest that this model can explain 14.09 percent of the data after a correction. The F-statistic is 4.827 with p-value equals to 0.004213 indicates that $H_0: R^2 = 0$ should be rejected on the 0.01 significance level, in other words, model (40) gives a significant good fit to the corresponding data.

In Figure 7, the estimated coefficient of market illiquidity in year $y-1$ and unexpected illiquidity in year $y$ both presented significant values. Coefficient $g_1$ stays positive and significant with a p-value of 0.0248 * (significant at 0.05 level) and $g_2$ shows a negative and significant value of -0.08745.

Using Huang’s measure, the result in Figure 7 (annual test) is consistent with the two hypotheses in Section 3.3 and has confirmed the following phenomena on the SSE over the investigation period of 2007.1-2012.12:

1. The coefficient $g_1$ is positive and significant, which suggests that expected stock excess return is an increasing function of expected market illiquidity.
2. The coefficient $g_2$ is negative and significant, which suggests that unexpected market illiquidity has a negative effect on unexpected market excess return.

In Figure 8, the result of cross-sectional regression with respect of monthly Huang’s data shows that $g_1$ is negative and insignificant but $g_2$ stays negative and highly significant with a value of -0.097505 and a p-value of 0.000753*** (significant at 0.001 level).

At this point we are more confident with Huang’s illiquidity measure since it shows higher effectiveness in its empirical results than Amihud’s illiquidity measure does. We would suggest that Huang-ratio is a relatively more suitable illiquidity measure for the SSE or a Chinese stock market.

By observing all the estimated coefficients from Figure 5 to Figure 8, we see that all the coefficients of unexpected illiquidity in year $y$ or month $m$ caused significant negative impact on contemporaneous market excess return. By contrast, the effect of market illiquidity in year $y-1$ or month $m-1$ on expected market excess return is relatively weak. This is what we found whether by using Amihud-ratio or Huang-ratio. Overall, the results in the four figures indicate that any information of policies or important events contained in the data from CSMAR or statistics annual Shanghai Stock Exchange databases could weaken the relationship between illiquidity and
market excess return while increase the effect of unexpected illiquidity on unexpected excess return. Following Wang Fang, Han Dong and Jiang Xianglin (2002), the relationship between market illiquidity and market excess return may only present more significant values than the relationship between unexpected market illiquidity and contemporary market excess return after excluding policy influences. However that is beyond the capability of the current study.

5. CONCLUSION

This paper extends a line of knowledge of the illiquidity and market return studies conducted by Amihud (2002) to a Chinese stock market – the Shanghai Stock Exchange (SSE).

We have presented the illiquidity phenomenon of the SSE over the investigation period of 2001.01–2012.12, since there is a common phenomenon among stock markets: illiquidity premiums are brought by illiquidity risks. We have discussed two illiquidity measures: Amihud-ratio and Huang-ratio, and have carried out their time-series and cross-section analyses over the estimation time of 12 years and 144 months. It is important not only to implement time-series tests on illiquidity measures, but also to examine the utility of illiquidity constructs and to see how they would interact with a stock market. In doing so we carried out four cross-sectional analyses on the relationships between market illiquidity and market excess return and between unexpected market illiquidity and unexpected market excess return for the SSE using both ratios.

Through time-series tests we found that the time-series data of the two market illiquidity measures can be significantly fitted into the lag-1 auto regression model, the AR (1) model. After applying Kendal’s (1954) bias correction method we observed that all the coefficients of Amihud-ratio and Huang-ratio have shown positive significant values in both annual and monthly time-series autoregressive estimations (See (33), (34), (36) and (38)). Therefore, both Amihud-ratio and Huang-ratio imply such phenomena on the SSE:
1. Investors are assumed to predict illiquidity for year y based on information available in year y-1; and then use this prediction to set prices that will generate the desired expected return in year y.
2. Market illiquidity available in the previous year has a positive effect on the prediction of market illiquidity for the following year.

Through annual and monthly cross-sectional analyses we found that with the influences of policies and important events on a Chinese stock market, unexpected market excess return is significantly and negatively related with contemporary unexpected market illiquidity on the SSE, but insignificantly related with market illiquidity when illiquidity is measured by Amihud-ratio. Whereas the significance of all the coefficients estimated in the cross-sectional analyses using Huang-ratio are
comparatively higher. As we have suggested that Huang-ratio is more suitable for a Chinese stock market.

In order to clearly contrast the final cross-sectional results using both ratios in Figure 5 – Figure 8, the following tables are given.

**Table 5 - Monthly Cross-Sectional Results**

<table>
<thead>
<tr>
<th>Illiquidity measures</th>
<th>Coefficients of Market Illiquidity in m-1</th>
<th>Coefficients of Unexpected Market Illiquidity in m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amihud-ratio</td>
<td>0.003805</td>
<td>-0.013350</td>
</tr>
<tr>
<td></td>
<td>p-value (0.4191) insignificant</td>
<td>p-value (0.0134 *) significant</td>
</tr>
<tr>
<td>Huang-ratio</td>
<td>-0.035733</td>
<td>-0.097505</td>
</tr>
<tr>
<td></td>
<td>p-value (0.112405) insignificant</td>
<td>p-value (0.000753 ***) significant</td>
</tr>
</tbody>
</table>

**Table 6 - Annual Cross-Sectional Results**

<table>
<thead>
<tr>
<th>Illiquidity measures</th>
<th>Coefficients of Market Illiquidity in y-1</th>
<th>Coefficients of Unexpected Market Illiquidity in y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amihud-ratio</td>
<td>-0.003056</td>
<td>-0.061581</td>
</tr>
<tr>
<td></td>
<td>p-value (0.73828) insignificant</td>
<td>p-value (0.00216 **) significant</td>
</tr>
<tr>
<td>Huang-ratio</td>
<td>0.13843</td>
<td>-0.08745</td>
</tr>
<tr>
<td></td>
<td>p-value (0.0248 *) significant</td>
<td>p-value (0.0632 •) significant</td>
</tr>
</tbody>
</table>

Again, Examining Table 5 and 6 we see that unexpected illiquidity caused more impacts (negative impacts) on market excess return than market illiquidity whether it is by using Amihud-ratio or Huang-ratio. Therefore we reached the conclusion that the influence of unexpected illiquidity on the SSE is playing a major role – Chinese stock market is more vulnerable to the impacts of unexpected factors such as frequent policy interventions and important events. And it is worthwhile pointing out at this time, that the final test of a model is not only how reasonable the assumptions behind it appear but also how well the model describes reality.

In the light of our research, some suggestions should be presented at the end of this paper:

1. As the empirical results have shown that Huang-ratio is more effective in a Chinese stock market, investors who aim to make profit in the SSE should pay more attention to the Huang-ratio.
2. As our cross-sectional regressions performed highly significant coefficients on unexpected illiquidity, one should always take into consideration of unexpected factors when predicting market illiquidity.
We suggest that further researches should try to observe useful microstructure data and analyze meaningful data patterns for the SSE over time, which would provide a great help in developing new ideas of illiquidity measures and estimating models. Moreover, estimation results may also be improved if one can get access to advanced databases that allow searching out each data that complies with certain criteria in order to avoid bias and errors during statistical tests. Overall, Chinese stock market is still young and under development, there is still much room for improvement and research.


This section includes the summary of the reflection of objectives regarding the current thesis.

6.1 Objective 1 – Knowledge and understanding

Mainly in Sections 1 and 2, we have demonstrated as much as valuable knowledge that is in relation with illiquidity theory and the understanding of it. Overall, a positive relationship between illiquidity and stock return has become evident, which brought up to the importance of measuring illiquidity risk using time-series and cross-sectional analyses. By deeper insights we have come to that ‘Amihud illiquidity measure’ and ‘Huang illiquidity measure’ could both be generated by data that are readily accessible in the Shanghai Stock Exchange over long periods of time, which are indispensable for the cross-sectional estimations of the current study. Over the very investigations on certain Chinese literature, we realized that Huang’s illiquidity measure is practically based on the idea of the Amihud-ratio, and theoretically, more significant to be tested in a Chinese stock market as an illiquidity construct.

6.2 Objective 2 – Methodological knowledge

In the section of Theoretical and Mathematical Background we have orderly provided 12 statistical related theories that are, analytically, essential knowledge for the Empirical Tests. Together, with a complete discussion of the famous Fama and MacBeth (1973) method in Appendix 1, we should, at this point, be fully confident about the efforts on how we are employing these methodological knowledge in deriving illiquidity constructs and the final time-series and cross-sectional tests, please refer to Section 2 and 3. Given that the microstructure data illustrated in Section 1.2 do not fulfill the estimation purpose of the current study in the sense of time span, we have discussed the rationales of the two measures (Amihud-ratio and Huang-ratio) that meet with the purpose where the readers could refer to Section 2.2 and Section 2.6 for the detail. The points mentioned, of course, include the critiques,
limitations of the two measures under the nature of the developing Chinese stock market and the inherent complexity among markets.

6.3 Objective 3 – Critically and systematically integrate knowledge

Due to that Chinese stock market is still young as compared to the stock markets in many developed countries we have integrated sources and acquired knowledge more carefully. Also, as getting the access of advanced databases that allow searching out each stock that complies with certain criteria is beyond the capability of the authors, our target sample accounts the whole sample of stocks on the Shanghai Stock Exchange. As a result, it has weakened the effectiveness of illiquidity measure since more bias and errors could come along during the programming tests. Although it set a limit to the current study, however, not yet a balk. The mathematical steps illustrated in Section 3.3 took many hours of the authors’ hard work, but turned out to be an elaborative and systematical one, and is viewed as rather an intelligible way of interpreting them than the old way did in Amihud (2002).

In the light of our research for knowledge or articles regarding illiquidity theory and the Chinese stock market, we intended to avoid texts that are not necessarily credible, some are from Wikipedia, and some are unpublished short-lived information. In doing so, the plan was to focus on well-published scientific literature and famous studies within the field that are primarily acquired from searching at Mälardalen University computer system and seeking from libraries and Google webpages. The readers would see how we are refining our References Section of Section 7 into different categories on sources.

6.4 Objective 4 – Independently and creatively identify and carry out advanced tasks

Seeing that the most published studies focus exclusively on the illiquidity-return relationship within a U.S. stock market, we have extended a line of study of the illiquidity and market return relationship to a Chinese stock market. From Section 2 to Section 4, we have taken the theory of illiquidity, the knowledge of time-series and cross-sectional analyses in conjunction with a more complex market condition into analyzing the time-series and cross-sectional effect between market illiquidity and market excess return on the Shanghai Stock Exchange. In addition, an ambiguously defined factor in Amihud’s illiquidity construct has been proved beyond exception by the current authors. Based on Fama & MacBeth (1973) methodology, we have not only formulated the autoregressive model and cross-section regression model for Amihud’s measure, but also an illiquidity measure proposed by a Chinese scholar (Huang-ratio). The specific explanation and analysis on each single line of the R codes in Appendix 2, 3 and 4 of Section 8 goes deeper into the logic of programming
from the authors’ mind. As this is a thesis intended for scholars and students at advanced mathematical level, no attempt has been made to do away with all such theories and models in the study.

6.5 Objective 5 – Present and discuss conclusions and knowledge

With our continuous effort on improving the knowledge structure of this paper, and making it as professional, understandable and precise as it could, we reckon this paper might be of particular interest to the readers who are familiar with statistics, probability, time series analysis, cross-sectional regression model, econometrics, portfolio theory, and R-programming. To support the analyses of the empirical results, we have provided four result figures for each final cross-sectional estimation model with adequate analyses in the Results Section of Section 4. Later on, as the readers refer to Section 5, we have devoted this whole chapter to expound, emphasize, and analyze all the meanings and implications contained in the test results. Not only giving the discussion of both liquidity and illiquidity conditions consist in the Shanghai Stock Exchange and their relationships with market excess return over the period of 2001.01-2012.12, we have also proceed with providing useful suggestions for investors and further researchers in a Chinese stock market.

6.6 Objective 6 – Scientific, social and ethical aspects

To be straight, there have not been much sources for the scientific research of this paper, for illiquidity is not a popular theory among mathematics and finance. After all, the right ones are ample for the purpose of our investigation that they are made by chief scholars within the field as you refer to Sections 1 and 2. Extending a line of illiquidity-return relationship research to a Chinese stock market is truly bravura, never an easy one. Through our research, carried out mainly with Amihud (2002) and Huang (2009), we have finally come to plot the illiquidity phenomenon and market performance presented in the Shanghai Stock Exchange in the past ten years, and these diagrams are generated in R. On the overall market trading mechanism of a Chinese stock market and frequent policy intervention by the Chinese government, it turned out that Huang-ratio is comparatively more useful as a construct of illiquidity to the Chinese investors -- our models have described reality. Without a shadow of doubt, we ensure the legal use and correctness of all the articles, literature, data, internet webpages, and books in our study. The last paragraph of Section 1.1 also explains how we worked as a group and the responsibility of each author.
7. REFERENCES

Main References:

All required market data in the current study are obtained from the databases of:

1. SSE (Shanghai Stock Exchange)
   Homepage: (http://www.sse.com.cn/).
2. Monthly and Yearly CSMAR (China Securities Market and Accounting Research)
   Website: (http://www.gtarsc.com/).
3. Statistics annual Shanghai Stock Exchange
   Website: (http://www.sse.com.cn/researchpublications/publication/yearly/).


P-values for ARIMA coefficients in Section 2.1, Section 3, & Section 4:
In section 2.1.11 we have mentioned that currently R software uses QR decomposition to solve the least squares problem:


**English References:**


hypothesis and evidence from the Tokyo Stock Exchange. Working Paper, National Taiwan University.


Chinese References:


8. APPENDIX

8.1 Appendix 1: Fama and MacBeth (1973) method

In asset pricing and portfolio management the Fama–French three-factor model is a model designed by Eugene Fama and Kenneth French to describe stock returns. Fama and French were professors at the University of Chicago Booth School of Business.

The traditional asset pricing model, known formally as the Capital Asset Pricing Model (CAPM) uses only one variable to describe the returns of a portfolio or stock with the returns of the market as a whole:

\[ E(R_i) = R_f + \beta_i(E(R_m) - R_f), \]

where \( E(R_i) \) is the expected return on the capital asset, \( R_f \) is the risk-free rate of interest such as interest arising from government bonds, and \( \beta_i \) is the sensitivity of the expected excess asset returns to the expected excess market returns.

In contrast, the Fama–French model uses three variables. Fama and French started with the observation that two classes of stocks have tended to do better than the market as a whole: (i) small caps and (ii) stocks with a high book-to-market ratio (BtM, customarily called value stocks, contrasted with growth stocks). They then added two factors to CAPM to reflect a portfolio’s exposure to these two classes:

\[ E(R_i) = R_f + \beta_3(E(R_m) - R_f) + b_S \cdot SMB + b_H \cdot HML + \alpha. \]

In the above equation \( E(R_i) \) is the expected return on the portfolio, \( R_f \) is the risk-free rate of interest such as interest arising from government bonds, and the ‘three factor’ \( \beta_3 \) is analogous to the classical \( \beta \) but not equal to it. SMB stands for ‘Small [market capitalization] Minus Big’ and HML for ‘High [book-to-market ratio] Minus Low’. Once the SMB and HML are defined, the corresponding coefficient \( b_S \) and \( b_H \) can be determined by linear regression.

To make this model more sophisticated, Amihud and Mendelson (1986) tested their hypothesis on the positive relationship between stock or portfolio return and stock liquidity. Then the three factor model can be improved into the following form:

\[ E(R_i) = R_f + \beta_3(E(R_m) - R_f) + b_S \cdot SMB + b_H \cdot HML + b_{ILLIQ} \cdot ILLIQ + \alpha, \]

where \( ILLIQ \) stands for the Amihud illiquidity measure.

In 2002, Amihud proposed his research result based on the data from CRSP (Center for Research of Securities Prices of the University of Chicago) databases. Amihud (2002) inspired us to bring an analysis to the Chinese stock market to investigate the cross-section and time-series effect between market illiquidity and market excess return.
8.2 Appendix 2: R-programming analysis of generating the PACF figure

```r
setwd("C:/Master")
#Set the working directory.

library(fBasics)
#Load fBasics package, this program package is used for time-series analysis.

da=read.table("SSE INDEX.txt")
#Read SSE INDEX.txt in the working directory, and let program to store SSE #INDEX.txt data in variable ‘da’.

head(da)
#Display the first five rows of ’da’.

da2=read.table("CHNRf.txt")
#Read CHNRf.txt file in the working directory, and let program to store CHNRf.txt #data in ‘da2’.

head(da2)
#Display the first five rows of ‘da2’.

da3=read.table("market return.txt")
#Read market return.txt file in the working directory, and let program to store market #return.txt data in ‘da3’.

return=da[,5]
#Read the fifth column of ‘da’ and name it ‘return’. (These returns are calculated by #the SSE composite price index, a price index without returns from dividend.)

head(return)
#Display the first five rows of ‘return’.

RM=da3[,2]
#Read the second column of ‘da3’ and name it ‘RM’. (RM = total market returns, i.e. #returns including dividends.)

head(RM)
#Display the first five rows of ‘RM’.

rf=da2[,3]
#Read the third column of ‘da2’ and name it ‘rf’. (rf = risk-free interest rate, i.e. #one-year risk free interest rate.)
```
head(rf)
#Display the first five rows of ‘rf’.

volumn=da[,9]
#Read the ninth column of ‘da’ and name it ‘volumn’. (volumn = market trading volume.)

k=length(return)
#Set ‘k’ equals to the length of ‘return’.

ILLIQ=c(1:k)
#Establish a measure name it ILLIQ in which contain k numbers of grids, these grids are used for the storage of the estimated data.

#Calculate illiquidity measure in a loop:
for(i in 1:k){
#Let the program count from 1 to k.
ILLIQ[i]=(10^9)*(abs(return[i]/volumn[i]))
#Illiquidity values calculated by Amihud-ratio.
#When count to 1, fill the value in the first grid, this value is the first illiquidity value of Amihud-ratio times 10^9, and so forth, until the program count to k, the program fills the last grid with the last illiquidity value of Amihud-ratio times 10^9.

head(ILLIQ)
#Display the first five rows of ‘ILLIQ’. (Amihud’s illiquidity measure)

logILLIQ=log(ILLIQ)
#Take the logarithm of Amihud’s illiquidity measures and name it ‘logILLIQ’.

head(logILLIQ)
#Display the first five rows of ‘logILLIQ’.

par(mfrow=c(1,2))
#This command tells the program to open a window where two figures are shown horizontally in the window.

acf(logILLIQ,lag.max=10)
#Show logILLIQ’s ACF figure, at most lag-10.

pacf(logILLIQ,lag.max=10)
#Show logILLIQ’s PACF figure, at most lag-10.

par(mfrow=c(1,1))
This command tells the program to open a window where one figure is shown in the window.

Likewise, the above programming procedure applies for the illiquidity measure estimated by Huang-ratio.

### 8.3 Appendix 3: R-programming analysis of the Correlation Test

```r
setwd("C:/Master")
# Set the working directory.

library(fBasics)
# Load fBasics package, this program package is used for time-series analysis.

da=read.table("SSE INDEX.txt")
# Read SSE INDEX.txt in the working directory, and let program to store SSE INDEX.txt data in variable ‘da’.

da2=read.table("CHNRf.txt")
# Read CHNRf.txt file in the working directory, and let program to store CHNRf.txt data in ‘da2’.

da3=read.table("data1.txt")
# Read data1.txt file in the working directory, and let the program to store data1.txt data in ‘da3’.

turnover=da3[,4]
# Read the fourth column of ‘da3’ and name it ‘turnover’.

tradingvalue=da3[,5]
# Read the fifth column of ‘da3’ and name it ‘tradingvalue’.

CAP=da3[,7]
# Read the seventh column of ‘da3’ and name it ‘CAP’.

return=da[,5]
# Read the fifth column of ‘da’ and name it ‘return’. (These returns are calculated by the SSE composite price index, a price index without returns from dividend.)

volumn=da[,9]
# Read the ninth column of ‘da’ and name it ‘volumn’.

k=length(return)
# Set ‘k’ equals to the length of ‘return’.
```
ILLIQ=c(1:k)
#Establish a measure name it ILLIQ in which contain k numbers of grids, these grids
#are used for the storage of the estimated data.

#Calculate illiquidity measure in a loop:
for(i in 1:k){
#Let the program count from 1 to k.
ILLIQ[i]=(10^9)*(abs(return[i]/volumn[i]))
#Illiquidity values calculated by Amihud-ratio.
#When count to 1, fill the value in the first grid, this value is the first illiquidity value
#of Amihud-ratio times 10^9, and so forth, until the program count to k, the program
#fills the last grid with the last illiquidity value of Amihud-ratio times 10^9.
}

ILLIQy=aggregate(ILLIQ,list(da2$V1),mean)
#Take the summation of the ILLIQ data based on the first column of da2 (that is,
#based on the ‘data of date’ that extracted from the first column of da2) and calculate
#the mean value of that summation and what we will obtain is the yearly market
#illiquidity (calculated by Amihud-ratio) for the correlation test.

logILLIQy=log(ILLIQy$x)
logILLIQy
#Let logILLIQy equals to the logarithmic of ILLIQy$x, ILLIQy$x are illiquidity
#values that do not contain useless information for our estimations, which are the
#values that are extracted from ILLIQy.

cor.test(logILLIQy,turnover, method="pearson",use = "everything")
cor.test(logILLIQy,tradingvalue, method="pearson",use = "everything")
cor.test(logILLIQy,CAP, method="pearson",use = "everything")
#Use Pearson’s method calculates the correlation coefficients between the three
#illiquidity proxies and illiquidity measures calculated by Amihud-ratio. R will return
#the results of the correlation test and their corresponding t-values and p-values.

Likewise, the above programming procedure applies for the illiquidity measure
estimated by Huang-ratio.

8.4 Appendix 4: R-programming analysis of the Empirical Test in Section 3.3

Below the R-programming analysis of the time-series and cross-sectional
estimations for Figure 6 in Section 4.1 is given.

setwd("C:/Master")
#Set the working directory.
library(fBasics)
#Load fBasics package, this program package is used for time-series analysis.

da=read.table("SSE INDEX.txt")
#Read SSE INDEX.txt in the working directory, and let program to store SSE
#INDEX.txt data in variable ‘da’.

da2=read.table("CHNRf.txt")
#Read CHNRf.txt file in the working directory, and let program to store CHNRf.txt
#data in ‘da2’.

da3=read.table("market return.txt")
#Read market return.txt file in the working directory, and let program to store market
#return.txt data in ‘da3’.

return=da[,5]
#Read the fifth column of ‘da’ and name it ‘return’. (These returns are calculated by
#the SSE composite price index, a price index without returns from dividend.)

RM=da3[,2]
#Read the second column of ‘da3’ and name it ‘RM’. (RM = total market returns, i.e.
#returns including dividends.)

rf=da2[,3]
#Read the third column of ‘da2’ and name it ‘rf’. (rf = risk-free interest rate, i.e.
#one-year risk free interest rate.)

volumn=da[,9]
#Read the ninth column of ‘da’ and name it ‘volumn’. (volumn = market trading
#volume.)

mon1=as.numeric(da2[,2]==1)
#Create the January Dummy.

k=length(return)
#Set ‘k’ equals to the length of ‘return’.

ILLIQ=c(1:k)
#Establish a measure name it ILLIQ in which contain k numbers of grids, these grids
#are used for the storage of the estimated data.

#Calculate illiquidity measure in a loop:
for(i in 1:k){
#Let the program count from 1 to k.
ILLIQ[i]=(10^9)*(abs(return[i]/volumn[i]))
#Iliquidity values calculated by Amihud-ratio.
#When count to 1, fill the value in the first grid, this value is the first illiquidity value
#of Amihud-ratio times 10^9, and so forth, until the program count to k, the program
#fills the last grid with the last illiquidity value of Amihud-ratio times 10^9.

head(ILLIQ)
#Display the first five rows of ‘ILLIQ’.

logILLIQ=log(ILLIQ)
#Take the logarithm of Amihud illiquidity measures and name it ‘logILLIQ’.

head(logILLIQ)
#Display the first five rows of ‘logILLIQ’.

#fitting the AR(1) model:
m1=arima(logILLIQ,order=c(1,0,0))
#Fitting logILLIQ into AR (1) model, order = c (1, 0, 0) means we set the order of AR
#to be 1, and other factors’ orders to be zero. Since ARIMA is a three-factored model.

m1
#Show the coefficients after fitting.

UlogILLIQ=m1$resid
#Residuals are the unexpected ILLIQs after the fitting.

head(UlogILLIQ)
#Display the first five rows of ‘UlogILLIQ’.

#calculate the p-value and t-value of the estimated coefficient:
cwp <- function (object){
  #where ‘<-' means assign a value to a name.
  #Establish a function that needs to have an object, and name it cwp. So when we use
  #cwp function, we will type: cwp(xxx).
  #cwp means ‘coefficients with p-values’; coefficients will be listed with t-values and
  #p-values.
  coef <- coef(object)
  #Assign coef to extract and store the coefficients of its model object.
  #coef is a generic function which extracts model coefficients from objects returned by
  #modeling functions.
  if (length(coef) > 0) {
    #If coef store more than zero value then run the following procedure:
    mask <- object$mask
Let mask store the mask values of the object, ordinarily the values of mask are either true or false, if the coefficients of the object modeling function are estimated by fitting, then the value of mask is true, otherwise, false.

sdev <- sqrt(diag(vcov(object)))
# Assign sdev be the standard deviations of each value in the diagonal of the variance-covariance matrix.
# diag () means extract or replace the diagonal of a matrix, or construct a diagonal matrix.
# vcov means return the variance-covariance matrix of the main parameters of a fitted model object.

t.rat <- rep(NA, length(mask))
# Assign t.rat be a vector that has the length of the mask and each grid in this vector is empty; this is in order to store the results of the following estimation values:

t.rat[mask] <- coef[mask]/sdev
# Assign t.rat[mask] store the coefficients that have ‘true’ mask values and divide them by sdev.

pt <- 2 * pnorm(abs(t.rat))
# Assign pt store the values of p-value.

setmp <- rep(NA, length(mask))
# Assign setmp be a vector that has the length of the mask and each grid in this vector is empty; this is in order to store the results of the following estimation values:

setmp[mask] <- sdev
# Assign setmp[mask] store the sdev of the coefficients that have ‘true’ mask values.

sum <- rbind(coef, setmp, t.rat, pt)
# Assign sum be the matrix of these data: coef, setmp, t.rat, and pt.

dimnames(sum) <- list(c("coef", "s.e.", "t ratio", "p-value"), names(coef))
# Give the new matrix their according names to the data.

return(sum)
# After the performance of the above estimation, we return back to ‘sum’ as the matrix of the results.

} else return(NA)
}
# If ‘coef’ store less than zero value, then the program would return a zero, or NA as the result of ‘there is no results’.
m1$coef
# Display the original coefficients before Kendall’s (1954) bias-correction.
cwp(m1)
# Tell the program to list all the estimated coefficients of AR (1) as well as their
# corresponding t-values and p-values.

# Kendall’s (1954) bias-correction:
coef=m1$coef+(3*m1$coef+1)/length(logILLIQ)
# Here we use the Kendall’s bias-correction procedure to correct the bias.

m2=arima(logILLIQ,order=c(1,0,0),fixed =coef)
# Then take the corrected coefficients into the estimated AR (1) again.

UlogILLIQ=m2$resid
# Let the residual equals to the unexpected market illiquidity in year y.

UlogILLIQ=UlogILLIQ[-1]
# Delete the first value of UlogILLIQ.
y=RM-rf
# Calculate the excess return of year y.
y=y[-1]
# Delete the first value of y.
logILLIQ1=logILLIQ[-length(logILLIQ)]
# Delete the first value of logILLIQ.

m4=lm(y~logILLIQ1+UlogILLIQ+mon1)
# Now take all the data of market excess return, market illiquidity and
# unexpected market illiquidity into the final cross-sectional regression model
# and fitting the model.
m4
summary(m4)
# Then we can check the result.

Likewise, the above programming procedure applies for the illiquidity measure
estimated by Huang-ratio.