What is the best composite liquidity proxy for explaining stock returns? Evidence from the Chinese stock market

Bo Yu
Faculty of Finance
City University of Macau
E-mail: boyu@cityu.edu.mo

Liang Dong[1]
School of Finance
Hunan University of Technology and Business
E-mail: liang.dong@hutb.edu.cn

Zhenjiang Qin
Department of Finance and Business Economics
Faculty of Business Administration
University of Macau
E-mail: zhenjiangqin@um.edu.mo

Keith S. K. Lam
Department of Finance and Business Economics
Faculty of Business Administration
University of Macau
E-mail: keithlam@rs.um.edu.mo

[1] Corresponding author: School of Finance, Hunan University of Technology and Business, Changsha, China. E-mail: liang.dong@hutb.edu.cn. This paper is revised from a part of Dr. Yu Bo's PhD thesis. Liang Dong acknowledges the support from Humanities and Social Science Foundation of the Ministry of Education of China (No. 24YJC790033) and Hunan Provincial Natural Science Foundation under grant (No. 2023JJ40242). Zhenjiang Qin acknowledges the support from University of Macau under grants MYRG2018-00210-FBA, MYRG2022-00007-FBA, and MYRG2022-00146-FBA.

Abstract

Accurate measurement of multidimensional liquidity is crucial for effective asset pricing and risk management. We construct 126 multidimensional composite liquidity proxies by using different combinations of individual single-dimensional liquidity proxies and different proxy combining methods. We propose an approach to select the optimal composite liquidity proxy, with both characteristic-level horseraces and systematic-factor-level comparisons among the competing composite proxies. Our results suggest that the Asymptotic Principal Component (APC) method is the suitable combining method, and the *Amihud-HL-FHT* proxy is the optimal multidimensional liquidity proxy for explaining stock returns in the Chinese stock market. These results remain robust when compared with nested composite proxies, adjusting the significance thresholds, extending the sample period, and using alternative comparison measures.

JEL classification: G12; G15

Keywords: composite liquidity proxy; combining method; horserace tests; stock return; Chinese stock market

1. Introduction

Liquidity is crucial in the stock market as it significantly impacts investors' ability to trade securities efficiently and influences the stability and pricing of financial assets. Since Amihud and Mendelson (1986) documented the significant liquidity-return relation, numerous studies have explored this relationship using various liquidity proxies and different financial market data (e.g., Roll, 1984; Datar et al., 1998; Lesmond et al., 1999; Amihud, 2002; Liu, 2006; Antoniou et al., 2007; Chen et al., 2010; Corwin and Schultz, 2012; Brennan et al., 2013). However, these studies often yield mixed results. Some studies find a significant return-liquidity relation (e.g., Avramov and Chordia, 2006; Liu, 2006; Antoniou et al., 2007; and Chen et al., 2010), while others do not (e.g., Fama and French, 1992; Chen et al., 2010; Hou et al., 2015 and Hou et al., 2018). This inconsistency may be due to the use of single-dimensional liquidity proxies, which capture only one aspect of liquidity, such as transaction costs, price impacts, trading speeds, or trading quantities[2] (e.g., Narayan and Zheng, 2011; Lam and Tam, 2011 and Ho and Chang, 2015).

To address this issue, recent studies (e.g., Korajczyk and Sadka, 2008; Kim and Lee, 2014 and Lam et al., 2019) started to use the multidimensional composite liquidity proxy to improve the reliability of the exploration of the return-liquidity relation. These studies construct the composite liquidity proxies by combining multiple single-dimensional proxies using either the Asymptotic Principal Component (APC) approach (e.g., Korajczyk and Sadka, 2008 and Lam et al., 2019) or the Principal Component Analysis (PCA) method (e.g., Kim and Lee, 2014). However, these studies have not attempted to identify the optimal composite proxy that provides the best return-liquidity explanatory power. They use either all the existing single-dimensional liquidity proxies (e.g., Kim and Lee, 2014) or a few randomly selected commonly single-dimensional liquidity proxies (e.g., Dong et al., 2024) to construct the composite proxy. Few studies attempt to search for the optimal composite proxy from available liquidity proxies in the literature, and there is a lack of methodology to evaluate the best combining method and input individual liquidity proxy to form the composite proxy. Identifying the optimal composite liquidity proxy is

2 Several other liquidity classifications have been used in the literature. For example, Kyle (1985) suggests that market liquidity is a slippery and elusive concept encompassing three transactional properties of stock markets, tightness, depth, and resiliency. Tightness is the cost of turning a position around in a short period, which is close to the trading cost component. Depth is the size of an order flow innovation that is required to change prices by a given amount, which is similar to the price impact component. Resiliency is the speed with which prices recover from a random, uninformative

important because it enhances the accuracy of liquidity measurement, which is crucial for improving market efficiency and informing investment strategies. To fill this gap, this study aims to propose the method and search for the optimal composite liquidity proxy for stock return explanation.

This study investigates the optimal composite liquidity proxy for explaining stock returns in the Chinese stock market using stock-level data from 2007 to 2023. We start with 17 individual liquidity proxies, and select 6 of them to construct 126 competing composite liquidity proxies. Two types of horserace tests are performed to identify the optimal composite liquidity proxy. First, we run the characteristic-level horserace (Liu et al., 2019). Second, we perform the factor-level regressions and pairwise comparisons among the winning composite proxies selected from the characteristic-level horserace. The results of the horserace tests suggest that the *Amihud-HL-FHT* proxy constructed by the APC method stands out as the optimal composite proxy for explaining stock returns. We show that three constituent individual liquidity proxies－*Amihud*, *Liu*, and *FHT*－perform well at capturing the illiquidity variations in the Chinese stock market and provide a comprehensive coverage of the multiple liquidity dimensions.

This study contributes to the literature in three key ways. First, it introduces a well-developed method for constructing and identifying the optimal composite liquidity proxy in the stock market, a critical aspect for precise liquidity measurement and the analysis of market efficiency, which has been scarcely explored in existing research. Second, it empirically demonstrates the superiority of the APC method over the PCA method in integrating multiple individual liquidity proxies from a large number of stocks into a comprehensive market aggregate measure, providing essential guidance for selecting appropriate methods in stock market liquidity studies. Third, it enriches the asset pricing literature by confirming the significant stock-level liquidity-return relation in the Chinese stock market, thereby underscoring the importance of accurate measurement of multidimensional liquidity.

The structure of this study is as follows. Section 2 reviews the literature on liquidity measurement and the pricing effect of liquidity in stock markets. Sections 3 and 4 introduce the sample selection, individual liquidity proxies, and research methodology. Sections 5 and 6 present the results of the empirical tests and robustness tests. Section 7 concludes.

2. Literature Review

Amihud and Mendelson (1986) first suggest a significant relationship between stock returns and liquidity using the bid-ask spread to measure liquidity. Since then, numerous studies have documented significant liquidity-return relations with various individual liquidity proxies in the US and around the world (e.g., Roll, 1984; Datar et al., 1998; Lesmond et al., 1999; Amihud, 2002; Liu, 2006; Antoniou et al., 2007; Chen et al., 2010; Corwin and Schultz, 2012; Brennan et al., 2013). These studies develop new proxies to measure different liquidity dimensions, including transaction costs, price impacts, trading speeds, and trading quantity.

Roll (1984) introduces the *Roll* ratio, an effective bid-ask spread, to capture transaction costs and find a significantly negative relation between the ratio and firm sizes in the US stock market. Corwin and Schultz (2012) suggest a new bid-ask spread proxy, the *HL* ratio, estimated with the daily high and low prices, finding a significant relation between the *HL* ratio and the "variance of stocks' true values". From the trading quantity perspective, Brennan et al. (1988) propose the trading volumes measure, while Datar et al. (1998) advocate a turnover ratio to proxy for the trading quantity dimension of liquidity, finding a significant return-liquidity relation in the US stock market. Following this strand, Chordia et al. (2001) also propose a set of measures related to turnover ratios and trading volumes.

Lesmond et al. (1999) develop two *Zero* ratios to measure transaction speed: *Zero1* (proportion of the trading days with zero returns) and *Zero2* (proportion of days with positive trading volume and zero returns). Both Zero measures document significant return-liquidity relations in the US stock market. Bekaert et al. (2007) use *Zero* ratios and provide significant return-liquidity evidence in 19 emerging markets. More recently, Fong et al. (2017) modify the Zero ratios to develop a new liquidity proxy, the *FHT* ratio, to estimate the effective transaction cost. Liu (2006), using his newly developed *Liu* ratio to proxy for trading speeds and price impacts, finds a significant return-liquidity relation in the US stock market.

Amihud (2002) and Pastor and Stambaugh (2003) propose the *Amihud* and *Pastor* (price reversals) ratios to measure the price impact dimension of illiquidity, respectively, and find significant return-illiquidity evidence. Goyenko et al. (2009), Brennan et al. (2013), and Kang and Zhang (2014) modify the *Amihud* ratio and advocate three new liquidity (illiquidity) ratios, the *Amivest* (the non-zero return version of *Amihud* ratio), *BHS* (the turnover version of Amihud ratio), and Kang ratio (the zero-trading volume version of *Amihud* ratio) respectively, documenting significant liquidity-return relation in the US stock market.

However, numerous other studies also find insignificant or mixed liquidity-return relations using these liquidity proxies (e.g., Fama and French, 1992; Wang and Iorio, 2007; Hou et al., 2015; Hou et al., 2018; Narayan and Zheng, 2011; Lam and Tam, 2011 and Ho and Chang, 2015). For example, Fama and French (1992) and Wang and Iorio (2007) document insignificant liquidity-return relations in US and Chinese stock markets, respectively, using the turnover ratio. Hou et al. (2015) find an insignificant liquidity-return relation using the *Amihud* ratio, and Hou et al. (2018) suggest that most of the 102 liquidity proxies they examined have insignificant relations with stock returns in the US market. Narayan and Zheng (2011) find inconsistent liquidity-return relations in the Chinese stock market using three different single-dimensional liquidity proxies: trading volume, turnover ratios, and trading probability. Lam and Tam (2011) and Ho and Chang (2015) obtain mixed liquidity-return relations in the Hong Kong and Chinese stock markets, respectively, using different individual liquidity proxies.

One possible cause of the mixed results is the use of single-dimensional liquidity proxies, which only capture one aspect of liquidity and may not reflect the complete liquidity-return relation. Recent studies have addressed this issue by constructing composite proxies to explore the pricing effect of liquidity on stock returns (e.g., Korajczyk and Sadka, 2008; Kim and Lee, 2014, and Lam et al., 2019). These studies either employ all available single-dimensional liquidity proxies (e.g., Kim and Lee, 2014) or randomly selected commonly used proxies (e.g., Lam et al., 2019) to construct composite proxies using dimension reduction methods such as the APC and PCA. However, they do not compare the performance of these combing methods or evaluate the composite proxies' explanatory power on the liquidity-return relation. To fill these gaps, we construct composite proxies using different single-dimensional proxies and combining methods, comparing their ability to explain stock returns.

3. Data and individual liquidity proxies

3.1 Data sample

We obtain the daily and monthly trading data and listed companies' accounting data from the China Stock Market and Accounting Research (CSMAR) database. The sample period spans from January 2007 to December 2023, covering the period after the Split-Share Structure Reform in China. The two Chinese stock exchanges were established in 1990 and 1991 respectively, and the market efficiency level and liquidity were quite low in the initial years. However, both market liquidity and efficiency levels have substantially improved following major capital market reforms, exemplified by the Split-Share Structure

Reform, which was implemented in 2005 and mostly accomplished by the end of 2006. Hung et al. (2015) argue that the reform increased the number of tradable shares in the Chinese stock market making the market more efficient, and suggest that liquidity has better explanatory power over stock returns in the period. Hou et al. (2012) also document that the Chinese stock market is more efficient after the Split-Share Structure Reform, as the reform increased the transparency of companies listed in the Chinese stock market.

This study concentrates on non-financial stocks in the Chinese A-share market and Growth Enterprises Market (GEM). We exclude B-share stocks due to their limited number in the market. Financial firms are also excluded, as their financial ratios have different interpretations from those of non-financial firms. The data-cleaning process is as follows: First, previous studies (e.g., Song et al., 2014) suggest that IPO stocks are underpriced in the primary market but overpriced in the secondary market in the first several months after the issuance in China. Thus, we follow Liu et al. (2019) and exclude data in the first six months after stock IPOs to mitigate the influence of the IPO underpricing puzzle in China. Second, Hou et al. (2018) find that a large number of anomalies are primarily driven by microcap stocks. To avoid the influence of microcap stocks, we follow Hou et al. (2018) and exclude stocks with a market capitalization in the bottom 20% in the previous month. Third, trading suspensions in the Chinese stock market tend to be brief, often lasting around an hour for abnormal floating stock prices. However, long trading suspensions indicate that the firm is facing major corporate events, such as mergers and acquisitions, which can cause significant stock price fluctuations. Therefore, we follow Liu et al. (2019) to exclude stocks that have fewer than 15 daily trading records in the past month or fewer than 120 daily trading records in the past year to mitigate the influence caused by long trading suspensions.

3.2 Individual liquidity proxies

By reviewing previous studies on stock market liquidity, we identify 17 individual single-dimensional liquidity proxies. These proxies include the *Pastor*, *Amihud*, *Amivest*, *BHS*, *Kang*, *HL*, *Roll*, *FHT*, *AVG_TO*, *AVG_TV*, *SD_TO*, *SD_TV*, *AVG_CVTO*, *AVG_CVTV*, *Zero1*, *Zero2*, and *Liu* ratios. Table 1 presents the definitions of these proxies (See Appendix A for detailed calculation methods of these proxies). The 17 proxies cover four dimensions of liquidity: the first five proxies relate to price impact, the next three proxies cover trading costs, the following six proxies reflect trading quantity, and the last three proxies capture trading speed. We do not include high-frequency liquidity measures as they are less useful for explaining monthly stock return variations. Moreover, studies such as Dong et al. (2024) and Fong et al. (2017) show high-frequency liquidity measures do not outperform low-frequency ones.

<Insert Table 1 here>

Table 2 presents summary statistics of the 17 single-dimensional liquidity proxies. The number of stock-month observations of the individual proxies varies from 296,335 for the *Liu* ratio to 328,081 for the *AVG_TO*, *AVG_TV*, *Zero1*, and *Zero2* ratios. The differences in the number of observations are due to either the zero-denominator occurrences or missing values. First, stock-level liquidity measured by various proxies exhibits significant cross-sectional and time-series variations. Second, different liquidity proxies vary greatly in terms of level and scale due to their different construction methods. Therefore, we use the standardized versions of the individual proxies to create the composite liquidity proxy. Third, by comparing the statistics before (not tabulated to save space) and after the Split-Share Structure Reform, we observe that the individual liquidity proxies reflect significant liquidity improvement after 2007. For instance, from unreported pre-reform statistics, the mean value of *Amihud* decreased from 158.18 over the period 1992-2006 to 40.3 over the period 2007-2023, and the mean turnover ratio, *AVG_TO*, increased from 0.26 before the reform to 0.34 after the reform. This growth in liquidity is primarily facilitated by the launch of the Split-Share Structure Reform, which urged listed companies to convert their non-tradable shares to tradable shares. This policy increases the trading activities and transparency of the firms, which in turn increases the liquidity of the market (Beltratti et al., 2012; Hung et al., 2015; Chuang et al., 2024).

4. Methodologies identifying the optimal composite liquidity proxy

4.1 Construction of the composite liquidity proxies

To reduce data complexity, we start with the 17 individual liquidity and select a subset of these proxies for constructing the composite liquidity proxy. This selection is based on analyzing the correlation matrix of the individual liquidity proxies. As shown in Table 3, the proxies exhibit varying degrees of correlation, indicating that the various dimensions of liquidity have common components. The correlation coefficients also vary across a large range, ranging from -59.04% to 89.85%, demonstrating that the different liquidity dimensions have unique dynamics. Notably, the correlation coefficient between the two zero-trading related ratios, *Zero1* and *Zero2*, is nearly 100%. This suggests the two methods of calculating the proportion of zero-return days yield similar results in the Chinese market. Our goal is to select a parsimonious subset of the individual proxies that effectively capture the various liquidity dimensions.

<Insert Table 3 here>

For the price impact dimension, we select *Amihud* over *Amivest*, *BHS*, and *Kang*. The *Amihud* ratio is one of the most widely used liquidity proxies in the literature, and the latter three proxies are all significantly correlated with *Amihud* as they are essentially variants of the *Amihud* ratio. We also select *Pastor*, as it is a price reversal measure which has a low correlation with *Amihud* (correlation coefficient=-2.34%). For the trading cost dimension, we select three proxies: *HL*, *Roll*, and *FHT*. These transaction cost proxies are estimated using very different methods and are lowly correlated (average correlation coefficient=7.8%). For the trading quantity dimension, the six proxies are all related to trading volumes or turnover rates, resulting in an average absolute correlation coefficient of 33%. Since these proxies from the trading quantity dimension are closely related with ratios such as *Amihud*, *Amivest*, and *Kang*, which are also calculated with trading volumes or turnover rates, we exclude them from the target subset. Although these proxies are not directly selected, the *Amihud* ratio should capture the trading quantity information. Lastly, for the trading speed dimension, we select the *Liu* ratio over the two *Zero* ratios, as it has a low correlation with all the other 16 proxies (average absolute correlation coefficient = 4.6%). In comparison, the *Zero* ratios have an average absolute correlation of 25% with other proxies, and their correlation with *FHT* is as high as 85.1%. Therefore, we end up with a subset of six individual liquidity proxies as the input for constructing the composite liquidity proxy: *Pastor*, *Amihud*, *HL*, *Roll*, *FHT*, and *Liu*.

Next, we combine an arbitrary number (one to six) of the six selected individual proxies to construct the market-aggregate composite liquidity proxy. Two common methods for combining multiple proxies are Principal Component Analysis (PCA) and the Asymptotic Principal Component (APC) method. The PCA method is suitable for datasets where the number of observations is much larger than the number of variables, but it suffers from the “curse of dimensionality” when the number of variables or features becomes significantly large. In contrast, the APC method works well in high-dimensional settings. To achieve the objective of constructing a market-aggregate composite liquidity proxy, the PCA method first extracts the stock-level across-measure principal component and then aggregates across stocks to get the market-level proxy. The APC method, on the other hand, treats the stock-level proxies as variables and the time series as observations, thus directly estimating the market-level across-measure composite liquidity proxy. In this study, we adopt both methods and examine which method is better for constructing the composite liquidity proxy.

First, we randomly select an arbitrary number of the six individual liquidity proxies as the input for the combination procedures, which yields 63 ($= C_6^1 + C_6^2 + C_6^3 + C_6^4 + C_6^5 + C_6^6$) different combinations of individual proxies. Second, we combine the individual liquidity proxies using the APC method (Appendix B provides the detailed process of APC) and the PCA method, resulting in 126 market aggregate liquidity proxies—63 from APC and 63 from PCA. Based on sign analysis, we determine the combined market aggregate proxy as an illiquidity measure, since most individual proxies are illiquidity measures. Third, to deal with the persistence of market-level illiquidity, we follow Pastor and Stambaugh (2003) to perform AR (2) regressions on the market aggregate measures and use the residuals from the AR(2) regressions as proxies for market illiquidity fluctuations.

Fourth, we perform 12-month rolling regressions of stock excess returns on market illiquidity fluctuations. The absolute regression coefficient represents a stock's conditional return sensitivity to market illiquidity fluctuations, as interpreted in Acharya and Pedersen (2005), with a higher value suggesting a lower systematic liquidity risk. These rolling coefficients are used as the stock-level liquidity risk proxy (the *composite liquidity proxy, Liqbeta,* hereafter), which encompasses multiple liquidity dimensions. We denote the regression coefficients based on the APC method as *Liqbeta_APC* and those based on the PCA method as *Liqbeta_PCA*. Using these 126 stock-level composite liquidity proxies, we conduct two levels of comparisons, the characteristic-level horserace, and the pricing factor-level pairwise comparison, to identify the optimal composite liquidity proxy in the Chinese stock market.

4.2 Firm characteristic-level horserace tests

We follow Liu et al. (2019) to conduct the firm characteristics-level horserace tests among the 126 composite liquidity proxies, including 63 *Liqbeta_APCs* and 63 *Liqbeta_PCAs*. The horserace tests consist of univariate and multivariate Fama-MacBeth style cross-sectional regressions on the composite proxies. A *Liqbeta_APC* or *Liqbeta_PCA* is selected as the winner in the characteristic-level horserace tests if it has a significant explanatory effect on stock returns in both the univariate and multivariate regressions.

The models of the univariate and multivariate Fama-MacBeth type regressions are shown in Equations (1) and (2), respectively:

$$R_{i,t} - R_{f,t} = a_t + \beta_t FC_{i,t-1} + \varepsilon_{i,t} \quad (1)$$

$$R_{i,t} - R_{f,t} = a_t + \beta_t FC_{i,t-1} + \sum \gamma_t OCF_{i,t-1} + \varepsilon_{i,t} \quad (2)$$

where $R_{i,t}$ and $R_{f,t}$ are the monthly stock return for stock *i* and the monthly risk-free rate in month *t*, respectively, with the risk-free rate represented by the one-month deposit interest rate; $FC_{i,t-1}$ represents the liquidity beta, *Liqbeta_APC* or *Liqbeta_PCA*, for stock *i* in month *t-1*; $OCF_{i,t-1}$ represents other firm characteristics in month *t-1*; a_t and $\varepsilon_{i,t}$ are the intercept and residual terms, respectively.

Following Liu et al. (2019), we introduce eight other firm characteristics as control variables in the multivariate regressions. These variables are defined as follows: The asset-to-market ratio (*AM*) is the natural logarithm of the total assets divided by the market capitalization, where market capitalization is calculated as the product of the closing price and the number of outstanding shares. The book-to-market ratio (*BM*) is the natural logarithm of the book equity divided by the market capitalization. The cash-to-price ratio (*CP*) is the net change in cash and cash equivalents between the two most recent financial periods divided by the market capitalization. The earning-to-price ratio (*EP*) is the most recently reported net profit divided by the market capitalization. The market size (*Size*) is the natural logarithm of market capitalization. The return-to-equity ratio (*ROE*) is the most recently reported net earnings divided by the book values of the equities. The asset-growth ratio (*AG*) is the percentage change in the total assets of the two most recent financial periods. The market beta (*MktBeta*) is the coefficient of stock returns on the market premium, estimated from 36-month rolling regressions. In addition, we follow Hou et al. (2011) and Liu et al. (2019) to construct dummy variables *D_EP* and *D_CP* to address the different pricing effects of positive and negative earnings and cash flows. Specifically, if the earnings (cash flows) is positive, *EP* (*CP*) equals its original value, and *D_EP* (*D_CP*) equals zero. If the earnings (cashflows) is negative, *EP* (*CP*) equals zero and *D_EP* (*D_CP*) equals one. These additional variables and dummy variables are incorporated into the multivariate regression model to control for the impact of firm-specific characteristics on stock returns.

The regressions specified in Equations (1) and (2) are performed for the cross-section of individual stock returns on a monthly basis, resulting in the time-series of monthly regression coefficients, β_t. The mean of these regression coefficients β_t is calculated, which represents the pricing effect of stock-level liquidity risk on stock returns. Based on these regressions, we identify the composite liquidity proxies that have significant pricing effects as candidates for the subsequent round of selection.

4.3 Pricing factor-level comparisons

Stock-level regressions often suffer from high noise due to idiosyncratic risks affecting individual stock returns, which can obscure the true relationship between liquidity and returns. To address this issue, we construct a systematic liquidity factor, *LIQ*, using the selected composite liquidity proxies. The *LIQ* factor is incorporated into common asset pricing models to enhance their return explanatory power on the returns of various testing portfolios. Subsequently, we perform factor-level Fama-MacBeth regressions and conduct pairwise comparisons to identify the optimal composite liquidity proxy.

4.3.1 Factor-level Fama-MacBeth cross-sectional regressions

We construct the systematic liquidity factors, *LIQ*, in a manner similar to the Fama-French factors by calculating the return difference between the most illiquid and most liquid portfolios using the *Liqbeta_APC* or *Liqbeta_PCA* proxies selected from the characteristic-level horserace. Then, we add the *LIQ* factor to well-known asset pricing models, including CAPM, FF3F, FF4F, FF5F, and FF6F[3]. After incorporating the systematic liquidity factor to these models, we obtain five liquidity-augmented factor pricing models, namely, L2F, L4F, L5F, L6F, and L7F, respectively[4] (Appendix C presents the construction details of the pricing factor).

To ensure the robustness of the analysis, we employ various sorting approaches to form nine different sets of testing portfolios. Specifically, the testing assets include a set of decile portfolios based on *Liqbeta* sorting, four sets of 25 portfolios based on 5×5 double-sorting on *Liqbeta-Size*, *Liqbeta-EP*, *Liqbeta-Beta*, and *Liqbeta-ROE*, and four sets of 27 portfolios based on 3×3×3 triple-sortings on*Liqbeta-Size-Beta*, *Liqbeta-Size-EP*, *Liqbeta-Size-ROE*, and *Liqbeta-EP-ROE*. The value-weighted average portfolio returns in excess of the risk-free rate are calculated, which are the dependent variables in the regressions.

There are two stages in the factor-level Fama-MacBeth cross-sectional regressions. The first stage is the 36-month rolling regression for each portfolio *p* as follows:

---

3 FF3F is the *MP-SMB-HML* three-factor model; FF4F is the *MP-SMB-HML-WML* four-factor model; FF5F is the *MP-SMB-HML-RMW-CMA* five-factor model; and FF6F is the *MP-SMB-HML-WML-RMW-CMA* six-factor model.

4 L2F is the *MP-LIQ* two-factor model; L4F is the *MP-SMB-HML-LIQ* four-factor model; L5F is the *MP-SMB-HML-LIQ-WML* five-factor model; L6F is the *MP-SMB-HML-LIQ-RMW-CMA* six-factor model; and L7F is the *MP-SMB-HML-LIQ-WML-RMW-CMA* seven-factor model.

$$R_{p,t} - R_{f,t} = a_{p,t} + \sum \boldsymbol{\beta_{p,t}} \boldsymbol{F_t} + \varepsilon_{p,t} \quad (3)$$

where $R_{p,t}$ and $R_{f,t}$ are the value-weighted portfolio excess return of portfolio *p* and the risk-free rate in month *t*, respectively; $\boldsymbol{F_t}$ is a vector of the monthly returns of the pricing factors; $\boldsymbol{\beta_{p,t}}$ is a vector of factor loadings; $a_{p,t}$, and $\varepsilon_{p,t}$ are the intercept and residual, respectively. From these rolling regressions, we obtain the time-varying factor loadings of the testing portfolios.

The second stage of the Fama-MacBeth regression is the cross-sectional regression of portfolio excess returns on the factor loadings estimated from the first stage. The regression equation is as follows:

$$R_{p,t} - R_{f,t} = a_t + \sum \boldsymbol{\beta_{p,t-1}} \boldsymbol{\lambda_t} + \varepsilon_{p,t} \quad (4)$$

where $R_{p,t}$ and $R_{f,t}$ are the same as those in Equation (3); $\boldsymbol{\beta_{p,t-1}}$ represents the factor loadings from the first stage for portfolio *p* in month *t-1*; $\boldsymbol{\lambda_t}$ is the regression coefficients in the second stage regressions in month *t*. The second-stage regressions are performed each month, and the premiums of the tested pricing factors are estimated as the time-series average of $\boldsymbol{\lambda_t}$. To adjust for the "error-in-variables" bias, we use the Shanken (1992) *t*-values to determine the significance of the premiums of the pricing factors.

4.3.2 Pairwise comparisons

Based on the factor-level regression results, we compare the performances of the liquidity-augmented factor models with the *LIQ* factor calculated from various competing composite liquidity proxies. The common performance metrics used for horseraces in time-series regressions include the Gibbons, Ross, and Shanken (GRS) test statistic, adjusted R^2s, and alpha-based appraisal ratios (e.g., Fama and French, 2015). These metrics, however, are not suitable for cross-sectional regressions. Therefore, we design the comparison criteria for the liquidity models in the spirit of the time-series regressions performance metrics, awarding models with more insignificant alphas, higher adjusted R^2, and more significant *LIQ* coefficients from the second-stage regressions. Specifically, we use three appraisal metrics: the number of significant coefficients of *LIQ* factor (*N_LIQ*), the number of insignificant intercepts (*N_A*), and the mean incremental adjusted R^2 (*MDR*). The incremental adjusted R^2 is the increase in adjusted R^2 caused by adding the *LIQ* factor to the non-liquidity factor pricing models. For instance, the incremental adjusted R^2 for the L6F model is the difference between the adjusted R^2 of the L6F model and that of the FF5F model. These metrics provide a comprehensive assessment of the model performance in cross-sectional regressions, emphasizing the importance of significant liquidity factors and the improvement in explanatory power provided by the liquidity-augmented models.

With the three appraisal metrics, we compare the performance of the liquidity models using the following procedures:

(1) **Factor-level Fama-MacBeth regressions:** We perform the factor-level Fama-MacBeth regressions with liquidity models where the *LIQ* factor is formed using various competing composite liquidity proxies. The appraisal metrics, *N_LIQ*, *N_A*, and *MDR*, are calculated based on the regression results.

(2) **Round-robin matches**: A random composite liquidity proxy is put to pairwise "round-robin matches" with each of the remaining composite liquidity proxies. The result of a match depends on the comparisons of one of the three appraisal metrics.

(3) **Calculating winning probabilities**: For a composite proxy, we calculate the mean winning probability based on each of the appraisal metrics' winning probabilities: *PROBL*, *PROBA*, and *PROBR*. These metric-specific winning probabilities are calculated as the number of "wins" based on one specific metric divided by the total number of matches. The overall mean winning probability, *MPB*, is the mean of the three metric-specific probabilities.

(4) **Statistical tests:** We conduct the Shapiro-Wilk normality test to examine whether *MPB* is normally distributed, the Sign test to examine if *MPB* is significantly different from zero, and the Wilcoxon rank-sum test to determine whether the *MPB* of the best composite proxy is significantly greater than other models. To substantiate the reliability of the tests, we follow Cooper and Miao (2019) and perform bootstrapping simulations to examine the significance of the differences between the simulated *MPB* (*MPB_SIMU*) of the optimal composite proxy and other composite proxies. For a particular composite proxy, we resample 30% MPB from the sample with replacements. Then, we repeat the selection process 1,000 times and calculate the composite proxy's mean *MPB* in each subsample (*MPB_SUB*).

(5) **Additional tests:** If more than one composite proxy is selected, we conduct an additional test, which uses an alternative measure of winning, the "win or loss" (*WOL*) measure. *WOL* counts the instance when a composite liquidity proxy wins a best-of-three match based on the comparisons of the three appraisal metrics with another competing proxy. That is, *WOL* takes the value of one if a composite proxy has two or three appraisal metrics, outperforming another composite proxy and zero otherwise. We take the average of a composite proxy's *WOL* against all the other competing proxies to obtain the probability of winning a best-of-three match, *MPBM*. Subsequently, we examine whether the optimal composite proxy has a significantly greater *MPBM* than other composite proxies as described in step (4).

With the above procedure, we are able to identify the optimal composite liquidity proxy for explaining stock returns in the Chinese market as the one that has a greater winning probability against all the other competing proxies. Figure 1 is a flow chart which demonstrates the aforementioned procedure.

<Insert Figure 1 here>

5. Empirical Results

5.1 Characteristic-level horserace tests

The characteristic-level horserace tests include both univariate and multivariate cross-sectional regressions of excess stock returns on the firm characteristics, including the composite liquidity proxy. The cross-sectional regressions are performed for each of the 126 competing composite liquidity proxies, comprising the 63 *Liqbeta_APC* and 63 *Liqbeta_PCA* as specified in Section 4.1. Table 4 reports the *Liqbeta_APC* and *Liqbeta_PCA*, which are significant at the 10% level in the univariate and multivariate regressions[5].

<Insert Table 4 here>

Panels A and B of Table 4 summarize the significant *Liqbeta_APC* (Panel A) and *Liqbeta_PCA* (Panel B) from the univariate and multivariate regressions, respectively. Panel A shows that 18 out of 63 (28.6%) *Liqbeta_APCs* are significantly related to stock returns in univariate regressions. This count is reduced to 15 out of 63 (23.8%) *Liqbeta_APCs* that are significant in both the univariate and multivariate regressions. Panel B demonstrates that only three *Liqbeta_PCAs are* significantly associated with stock returns in the multivariate regressions, and no *Liqbeta_PCAs* are significant in both the univariate and multivariate regressions. These results indicate that the APC method is a better combining method than the PCA method, as the APC method provides more composite proxies with significant explanation power on the return-liquidity relation in the multivariate regression. A possible reason might be that APC directly combines stock-level individual liquidity proxies into the market aggregate liquidity, taking into consideration the inter-correlations of liquidity across stocks, while the PCA method achieves the market

5 We choose 10% as the threshold significant level in the characteristic-level horserace rather than 5% so that there will be more winner composite proxies for the factor-level pairwise comparisons, thus resulting in larger sample size in the non-parametric tests and increasing the reliability in our results.

aggregate liquidity by averaging stock-level first components, which might lose information regarding the intricate interactive dynamics of stock-level liquidity proxies.

From the characteristic-level regressions, we obtain 15 *Liqbeta_APCs* that can explain the cross-sectional variations in stock returns in both the univariate and multivariate Fama-MacBeth regressions. As mentioned in Section 4, these 15 *Liqbeta_APCs* are the winners in the characteristic-level horserace. We will employ these *Liqbeta_APCs* to construct the systematic liquidity factor, *LIQ*, and use them to conduct the factors-level regression and pairwise comparisons to further search for the optimal composite liquidity proxy.

5.2 Factor-level Fama-MacBeth regressions

We construct the systematic liquidity factor, *LIQ*, using each of the 15 winner *Liqbeta_APCs* from the characteristics-level horserace. The *LIQ* factor is added to the CAPM and FF-series pricing models to enhance their explanatory power for stock returns. Then, we conduct Fama-MacBeth cross-sectional regressions with the liquidity-augmented factor pricing models~~,~~ by regressing value-weighted portfolio excess returns on the *LIQ* and other pricing factors. Finally, we use three appraisal metrics from the Fama-MacBeth regressions—*N_LIQ*, *N_A*, and *MDR*—to perform the factor-level regression and pairwise comparisons among the 15 competing composite liquidity proxies. We present the comparison results in Table 5.

<Insert Table 5 here>

The 15 composite proxies are listed based on the ranking of their *MPB*, the mean probability that a composite proxy can win in a round of an appraisal metric comparison. The composite proxy of *Amihud-HL-FHT* ranks first with the highest *MPB* of 0.9048, which suggests that the *Amihud-HL-FHT* proxy can outperform better than 13.6 (0.9048×15) of the other 14 composite proxies on average. For the *N_LIQ* ranking, it has 10 significant *LIQ* factor premiums, which takes the joint first place in the *PROBL* ranking. Its *PROBL* is 0.9286, indicating that the composite proxy can win over the remaining 14 proxies by comparing their *N_LIQ*. From the perspective of the regression intercept, more insignificant intercepts suggest that the *LIQ*-augmented factor models offer a thorough explanation of portfolio returns. The *N_A* ranking of the *Amihud-HL-FHT* proxy has four insignificant intercepts, which takes the joint third place in the ranking. Its *PROBA* of 0.7857, which indicates that the proxy has more insignificant intercepts than 12 of the other 14 composite proxies. Lastly, the incremental adjusted R^2 caused by adding the *LIQ* factor to non-liquidity factor models reveals the additional explanatory power of liquidity on portfolio returns. The *MDR* of *Amihud-HL-FHT* is 0.0459, ranking first among the 15 proxies. Its *PROBR* is 1.0000, indicating that the composite proxy can always offer higher incremental explanatory power than the other 14 proxies. Overall, with individual metrics rankings of *N_LIQ*, *N_A*, and *MDR*, the *Amihud-HL-FHT* proxy takes first place in the overall average ranking of *MPB*.

The second-best composite proxy is *Amihud-HL-Pastor-FHT*, which has the same rankings of *PROBL* and *PROBA* as *Amihud-HL-HL* but a lower *PROBR* of 0.9286. Even though some of the composite proxies rank quite high in individual appraisal metrics, their rankings on other metrics are relatively low, pulling down their overall *MPB* rankings. For example, the *Amihud-HL-FHT-Pastor-Liu* proxy ranks first in the number of insignificant intercepts, but its rankings of *N_LIQ* and *MDR* are only second and fifth, respectively, lowering its overall average *MPB* ranking to the third place.

Next, we perform statistical tests to examine whether the *MPB* of a composite proxy is significantly higher than the other proxies. Five non-parametric tests are performed to examine whether the *Amihud-FHT-Liu* proxy has an *MPB* significantly higher than the other 14 composite proxies. First, we use the Shapiro-Wilk normality tests to inspect whether *MPB* is normally distributed. The results in Table 5 suggest that the *MPB* of all composite proxies are not normally distributed, as the test *p*-values (P_SWs) are all less than 5%. Second, we conduct the Sign test to test the null hypothesis that the MPB equal zero. Results Table 5 shows that all proxies' *p*-values (P_ST) are generally less than 5%, suggesting that these proxies' *MPB* are significantly different from zero. Third and fourth, we conduct another Sign test and a Wilcoxon rank-sum test, to examine whether there is a significant difference between *MPB* of *Amihud-HL-FHT* and other proxies. The columns of XPL_WIL, XPR_WIL, XP2_WIL, and P_ST2 represent the *p*-values of the one-way (left), one-way (right), and two-way Wilcoxon rank-sum tests, and the second Sign test, respectively. The test *p*-values for the top six composite proxies are greater than 5%, suggesting that the differences in MPB between *Amihud-HL-FHT* and the other five following-up proxies are statistically insignificant.

In addition, we conduct bootstrapping simulations on the *MPB* as described in Section 4.3.2. The simulated *MPB* (*MPB_SIMU*) and their *p*-values (*P_VALUE_SIMU*) are presented in the last two columns of Table 5. The values of the *MPB_SIMU* are generally similar to the *MPB*, which shows the robustness of our previous computations. The simulated *p*-values for the top six composite proxies are also greater than 5%, which is consistent with the results in the non-parametric tests. Overall, the *MPB* and simulated results in Table 5 suggest that the top six composite proxies have greater *MPB* than the other nine proxies, but their *MPB* are not significantly different from each other. Therefore, these six proxies are selected as the candidates for the best composite liquidity proxies up to this step.

5.3 Pairwise comparisons with an alternative winning probability

In order to select the optimal composite liquidity proxy among the top six proxies in Table 5, we repeat the pairwise comparison procedures using an alternative winning probability measure, the mean probability that a composite proxy can win a best-of-three match (*MPBM*). The results of the *MPBM* comparisons are presented in Table 6.

<Insert Table 6 here>

The results show that the composite proxy of *Amihud-HL-FHT* proxy has the greatest *MPBM* of 1.0000. This indicates that this proxy can always outperform all the other 14 competing composite proxies. The rest of the composite proxies are ranked from second to ninth place (with some tied proxies), with *MPBM* decreasing from 0.9286 to 0.0000. Therefore, the *Amihud-HL-FHT* proxy stands out to be the optimal composite proxy according to the *MPBM* ranking.

We repeat the non-parametric tests in Table 5 on the *MPBM* to examine the statistical significance of the out-performance of the *Amihud-HL-FHT* proxy over all the other competing proxies. The *P_SW* and *P_ST* in Table 6 are all less than 5%, suggesting that the *MPBM* of all proxies are significantly greater than zero and not normally distributed. Thus, we apply the Wilcoxon rank-sum tests and the second Sign test to examine whether there are significant differences between the *MPBM* of *Amihud-HL-FHT* and the other proxies. Table 6 shows that the *p*-values of the tests, *PL_WIL*, *PR_WIL, P2_WIL*[6], and *P_ST2*, are all less than 5%, which suggests that the *MPBM* of *Amihud-HL-FHT* is statistically greater than those of the other proxies.

In addition, we also repeat the bootstrapping simulation results on the *MPBM*. Similar to Table 5, the results in Table 6 indicate that the *MPBM* and simulated *MPBM* (*MPBM_SIMU*) are close to each other, which shows the robustness of our previous results. The simulated *p*-values (*P_VALUE_SIMU*) in Table 6 are all less than 5%, which further suggests that there are significant differences between

6 In Table 6, we report the simple *p*-values (*P2_WIL*, *PL_WIL*, and *PR_WIL*) rather than the exact *p*-values (*XP2_WIL*, *XPL_WIL*, and *XPR_WI*L) for the Wilcoxon rank-sum tests as the exact *p*-value is suitable for small samples ($N < 30$). Since our sample size is 1,000, substantially greater than 30, the simple *p*-values are reported.

MPBM_SIMU of *Amihud-HL-FHT* and other composite proxies. *Amihud-HL-FHT* outperforms all the other composite proxies with an *MPBM_SIMU* of 1.000, which is significantly greater than those of the other composite proxies. The bootstrapping simulation test results arrive at a similar conclusion and further supported the claim that *Amihud-HL-FHT* is the best composite liquidity proxy for explaining stock returns in the Chinese market.

5.4 Further exploration of the optimal composite liquidity proxy

One might wonder why the combination of *Amihud-HL-FHT* stands out as the optimal composite proxy. In this section, we provide some intuitive explanations. Figure 1 plots the time series of the cross-sectional averages of the six individual liquidity proxies. Based on the principle of "good ingredients in, good results out", we argue that the average *Amihud*, *HL*, and *FHT* proxies are themselves good proxies capturing the liquidity information in the Chinese market. For example, the three proxies spiked up around the global financial crisis in 2008, the market crash in 2015, and the 2019-2020 COVID pandemic, indicating heightened illiquidity in during the market crashes. Conversely, during periods of market rebound and liquidity recovery, the proxies showed a decline. The plots also reveal that the *Amihud* ratio captures larger-scale variations in price impact, while the *FHT* and *HL* measures reflect more oscillations in trading costs. Thus, together, these proxies provide comprehensive coverage of multidimensional and multi-frequency liquidity information in the market.

<Insert Figure 1 here>

Comparing the two price impact proxies, *Amihud* and *Pastor*, we find that *Pastor*'s plot follows a similar general trend to *Amihud*'s, though with different fluctuation scales. Thus, the *Aimhud* ratio should capture a significant amount of information presented in *Pastor*. Similarly, the trading cost proxy, *Roll*, mirrors the major trends of the *HL* measure. Lastly, the *Liu* measure, which is the turnover-rate adjusted ratio of zero-trading days to total trading days, captures the illiquidity information from dimensions of the trading quantity and trading speed. The time-series of market average *Liu* demonstrates a general improvement in trading speed in the Chinese market, except for an interruption during 2014-2016 when the market experienced its largest crash and subsequent rebound. However, the time-series variation of *Liu* is much smaller than that of the other proxies, potentially limiting its ability to explain stock return variations.

Next, we verify whether the optimal composite liquidity proxy captures the information of the six individual liquidity proxies, by regressing the optimal *Liqbeta_APC* on the individual liquidity proxies. In particular, we conduct both univariate (regressing the optimal *Liqbeta_APC* on each one of the individual liquidity proxies) and multivariate (regressing the optimal *Liqbeta_APC* on the six individual liquidity proxies) regressions. The regression results are presented in Table 7.

<Insert Table 7 here>

The results in Table 7 suggest that the "good ingredients" indeed result in the "good results". Specifically, the *Amihud-HL-FHT* proxy is significantly related to the *Amihud*, *FHT*, *HL*, *Liu*, and *Roll* at the 5% significance level in both the univariate and multivariate regressions. Overall, we provide some more details to justify why the *Amihud-HL-FHT* can be the optimal proxy.

6. Robustness tests

This study conducts four robustness tests to examine the reliability of our findings in the previous section. First, we repeat the characteristic-level horserace and the factor-level comparisons between the optimal composite liquidity proxy and its nested composite proxies to re-examine whether *Amihud-HL-FHT* is the optimal composite proxy. A given composite liquidity proxy's nested composite proxies are formed by adding or removing an individual liquidity proxy from it. And if the composite proxy is the optimal one, it should outperform all its nested proxies. However, if the composite proxy is not the optimal proxy, it could underperform the nested composite proxies if the less efficient individual liquidity proxies are removed or the more efficient individual proxies are added. Therefore, we examine whether *Amihud-HL-FHT* outperforms its six nested proxies. The comparison results are presented in Table 8.

<Insert Table 8 here>

Table 8 shows that the nested proxies of *Amihud-HL* and *Amihud-HL-FHT-Roll* lose to *Amihud-HL-FHT* as they don't have significant relations with stock returns in univariate and multivariate Fama-MacBeth cross-sectional regressions. In addition, *Amuhud-FHT*, *HL-FHT*, and *Amihud-HL-FHT-Liu* lose as they have smaller *N_LIQ* and *N_A* than *Amihud-HL-FHT*. Moreover, *Amihud-HL-FHT-Pastor* also loses because it has a lower incremental R^2. In general, the results in Table 8 further support that

Amihud-HL-FHT is the optimal composite liquidity proxy in the Chinese stock market, which is consistent with the results reported in Tables 5 and 6.

Second, we repeat the tests in Table 5 using appraisal metrics of the *N_LIQ*, *N_A*, and *MDR* calculated with an increased significance threshold from 5% to 10%, to investigate whether *Amihud-HL-FHT* remains the optimal composite proxy. Specifically, we use 10% as the significant level to count the number of significant *LIQ* factor premiums (*N_LIQ*) and the insignificant intercepts (*N_A)*. The comparison results are shown in Table 9. It shows that *Amihud-HL-FHT* has the highest *MPB* at 0.8810 and the highest *MPB_SIMU* at 0.8816, which suggests it remains the optimal composite proxy with the 10% significance threshold. Thus, the results suggest that *Amihud-HL-FHT* is the optimal composite proxy as the pairwise comparison procedure using appraisal metrics calculated with the 10% significance level, confirming the results in Tables 5, 6, and 8.

<Insert Table 9 here>

Third, we replicate the pairwise comparisons using a longer sample period from January 2005 to December 2023, which covers the process of the Split-Share Structure Reform. Since the APC method extracts the market aggregate liquidity component by taking the entire sample period as a whole, lengthening the sample period with a volatile time with liquidity changes might affect the identification result of optimal composite liquidity proxy. We present the results in Table 10. The results show that *Amihud-HL-FHT* has the highest *MPB* at 0.8333. The second-best composite proxy, *Amihud-FHT-Pastor-Liu*, adds the *Pastor* measure to the *Amihud-HL-FHT* proxy and obtains the same *MPB* of 0.8333. However, its simulated winning probability is slightly lower (*MPB_SIMU*=0.8331) than that of *Amihud-HL-FHT* (*MPB_SIMU*=0.8346). Therefore, the results suggest that the composite proxy of *Amihud-HL-FHT* outperforms the other composite proxies, which is consistent with and further supports the results from Tables 5 to 9.

<Insert Table 10 here>

Fourth, we conduct a comparison among the 21 winning composite proxies in Table 3 using their total rank of the three appraisal metrics to check whether *Amihud-HL-FHT* is still the optimal composite proxy with a different comparison method, where we select the optimal composite proxy based on the sum of the ranks of the *N_LIQ*, *N_A*, and *MDR*. In particular, we sort these composite proxies by the *N_LIQ* in descending order. The composite proxy with the greatest number of significant liquidity factor premiums (*N_LIQ*) is ranked as the first. If two proxies have the same value of *N_LIQ*, both of them are given the same rank. We repeat this process by ranking the composite proxies with *N_A* and *MDR* in descending order, respectively. We then obtain the sum of the three ranks and obtain the total rank (*TOTAL_RANK*). The composite proxy with the lowest total ranking is the optimal composite proxy. We present the comparison results in Table 11. The *Amihud-HL-FHT* proxy has the lowest TOTAL_RANK of five and stands out as the optimal composite proxy. This finding further confirms our main results.

<Insert Table 11 here>

7. Conclusion

This study investigates the optimal composite liquidity proxy for explaining stock returns in the Chinese stock market using stock-level data from 2007 to 2023. Our results suggest that the multidimensional composite proxy of *Amihud-HL-FHT* is the most optimal.

With preliminary screening, we begin with 126 competing composite liquidity proxies. We first conduct the stock characteristic-level horserace to identify the top-performing composite proxies. These winner proxies were then used in factor-level regressions and pairwise comparisons. The results indicate that the APC method is superior in combining individual proxies, ultimately identifying *Amihud-HL-FHT* as the best composite proxy, with five other follow-up proxies. Additional tests using an alternative winning probability measure (*MPBM*) confirmed that *Amihud-HL-FHT* is the optimal composite proxy. The robustness of our findings was substantiated through four tests: comparing the performance of *Amihud-HL-FHT* with its nested proxies, adjusting the significance threshold of appraisal metrics calculation from 5% to 10%, extending the sample period to cover the Split-Share Structure Reform, and using the total ranks across three appraisal metrics as pairwise comparisons. All tests reaffirmed the superiority of the *Amihud-HL-FHT* composite proxy.

Our study provides valuable insights into identifying the optimal combining method and composite proxy for the Chinese stock market. Future research should explore the implications of these findings on asset pricing, performance appraisal, and trading strategies related to stock-level liquidity. Additionally, investigating the application of the optimal composite proxy in different market conditions and its impact on investment decision-making would be worthwhile avenues for further study.

Reference

Acharya, V. V., and Pedersen, L. H., 2005, "Asset pricing with liquidity risk", *Journal of Financial Economics*, 77(2), 375-410.

Amihud, Y., 2002, "Illiquidity and stock returns: Cross-section and time-series effects", *Journal of Financial Markets,* 5(1), 31-56.

Amihud, Y., and Mendelson, H., 1986, "Asset pricing and the bid-ask spread", *Journal of Financial Economics*, 17(2), 223-249.

Antoniou, A., Lam, H. Y., and Paudyal, K., 2007, "Profitability of momentum strategies in international markets: The role of business cycle variables and behavioural biases", *Journal of Banking & Finance*, 31 (3), 955–972.

Avramov, D., and Chordia, T., 2006, "Asset pricing models and financial market anomalies", *The Review of Financial Studies*, 19 (3), 1001–1040.

Bekaert, G., Harvey, C. R., and Lundblad C., 2007, "Liquidity and expected returns: Lessons from emerging markets", *Review of Financial Studies*, 20(6), 1783-1831.

Beltratti, A., Bortolotti, B., and Caccavaio, M., 2012, "The stock market reaction to the 2005 split share structure reform in China", *Pacific-Basin Finance Journal*, 20(4), 543–560.

Brennan, M. J., Chordia, T., and Subrahmanyam, A., 1988, "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns", *Journal of Financial Economics*, 49(3), 345-373.

Brennan, M., Huh, S.W., and Subrahmanyam, A., 2013, "An analysis of the Amihud illiquidity premium", *Review of Asset Pricing Studies*, 3(1), 133–176.

Chen, X., Kim, K. A., Yao, T., and Yu, T., 2010, "On the predictability of Chinese stock returns", *Pacific-Basin Finance Journal,* 18(4), 403-425.

Chordia, T., Roll, R., and Subrahmanyam, A., 2001, "Market liquidity and trading activity", *Journal of Finance*, 56(2), 501-530.

Chuang, O. C., Chuang, H. C., Wang, Z., and Xu, J., 2024, "Profitability of technical trading rules in the Chinese stock market", *Pacific-Basin Finance Journal*, 84, 102278.

Connor, G., and Korajczyk, R. A., 1986, "Performance-measurement with the arbitrage pricing theory – a new framework for analysis", *Journal of Financial Economics*, 15(3), 373-394.

Cooper, I., and Maio, P., 2019, "New evidence on conditional factor models", *Journal of Financial and Quantitative Analysis*, 54(5), 1975-2016.

Corwin S.A., and Schultz, P., 2012, "A simple way to estimate bid-ask spreads from daily high and low prices", *Journal of Finance*, 67(2), 719-759.

Datar, V.T., Naik, N.Y., and Radcliffe, R., 1998, "Liquidity and stock returns: An alternative test", *Journal of Financial Markets*, 1(2), 203-219.

Dong, L., Yu, B., Qin, Z. and Lam, K.S., 2024, "Liquidity risk and expected returns in China's stock market: A multidimensional liquidity approach", *Research in International Business and Finance*, 69, 102247.

Fama, E. F., and French K. R., 1992, "The cross-section of expected stock returns", *Journal of Finance*, 47(2), 427-465.

Fama, E.F., and MacBeth, J.D., 1973, "Risk, return, and equilibrium: Empirical tests", *Journal of Political Economy*, 81(3), 607-636.

Fama, E. F., and French, K. R., 2015, "A five-factor asset pricing model", *Journal of Financial*

Economics, 116(1), 1-22.

Fong, K., Holden, C., and Trzcinka, C.A., 2017, "What are the best liquidity proxies for global research?", *Review of Finance,* 21(4), 1355-14507.

Guo, B., Zhang, W., Zhang, Y., and Zhang, H., 2017, "The five-factor asset pricing model tests for the Chinese stock market". *Pacific-Basin Finance Journal*, 43, 84-106.

Goyenko, R.Y., Holden, C.W., and Trzcinka, C.A., 2009, "Do liquidity measures measure liquidity?", *Journal of Financial Economics*, 92(2), 153–181.

Han, C., Wang, Y., and Xu Y., 2019, "Efficiency and multifractality analysis of the Chinese stock market: Evidence from stock indices before and after the 2015 stock market crash", *Sustainability*, 11(6): 1699.

Ho, T.W., and Chang, S.H., 2015, "The pricing of liquidity risk on the Shanghai stock market". *International Review of Economics and Finance*, 38, 112-130.

Hou, K., Xue, C., and Zhang, L., 2015, "Digesting anomalies: an investment approach", *Review of Financial Studies*, 28(3), 650-705.

Hou, K., Xue, C., and Zhang, L., 2018, "Replicating anomalies", *Review of Financial Studies,* 33(5), 2019-2133.

Hou, K., Karolyi, G. A., and Kho, B. C., 2011, "What factors drive global stock returns?", *Review of Financial Studies,* 24(8), 2527-2574.

Hou, W., Kuo, J. M., and Lee, E., 2012, "The impact of state ownership on share price informativeness: The case of the Split Share Structure Reform in China", *The British Accounting Review*, 44(4): 248–261.

Hung, C. H. D., Chen, Q., and Fang, V., 2015, "Non-tradable share reform, liquidity, and stock returns in China", *International Review of Finance*, 15(1): 27–54.

Kang, W., and Zhang, H., 2014, "Measuring liquidity in emerging markets", *Pacific-Basin Finance Journal*, 27, 49-71.

Kim, S. H., and Lee, K. H., 2014, "Pricing of liquidity risks: evidence from multiple liquidity measures", *Journal of Empirical Finance*, 25(2), 112-133.

Korajczyk, R.A., and Sadka, R., 2008, "Pricing the commonality across alternative measures of liquidity". *Journal of Financial Economics*, 87(1), 45-72.

Lam, K.S.K., and Tam, H.K., 2011, "Liquidity and asset pricing: Evidence from the Hong Kong stock market", *Journal of Banking & Finance*, 35, 2217-2230.

Lam, K.S.K., Tam, H.K., and Dong, L., 2019, "Liquidity and expected returns: Evidence from the Chinese stock market", *China Accounting and Finance Review*, 21(4), 1-42.

Lesmond, D., Ogden J., and Trzcinka C., 1999, "A new estimate of transaction costs", *Review of Financial Studies*, 12(5), 1113-1141.

L'Her, J. F., Masmoudi, T., and Suret, J. M., 2004, "Evidence to support the four-factor pricing model from the Canadian stock market", *Journal of International Financial Markets, Institutions and Money*, 14 (4), 313–328.

Li, C., 2017, "Log-periodic view on critical dates of the Chinese stock market bubbles", *Physica A: Statistical Mechanics and its Applications*, 465, 305-311.

Liang, P., and Willett, T. D., 2015, "Chinese stocks during 2000–2013: Bubbles and busts or fundamentals?", *The Chinese Economy*, 48(3), 199-214.

Liu, W.M., 2006, "A liquidity-augmented capital asset pricing model", *Journal of Financial Economics*, 82(3), 631-671.

Liu, J., Stambaugh, R. F., and Yuan, Y., 2019. “Size and value in China”. *Journal of Financial Economics,* 134(1), 48-69.

Martinez, M. A., Nieto, B., Rubio, G., and Tapia, M., 2005, “Asset pricing and systematic liquidity risk: An empirical investigation of the Spanish stock market”, *International Review of Economics and Finance*, 14 (1), 81–103.

Narayan, P.K., and Zheng, X., 2011, “The relationship between liquidity and returns on the Chinese stock market”, *Journal of Asian Economics*, *22*(3), 259-266.

Pastor, L., and Stambaugh, R. F., 2003, “Liquidity risk and expected stock returns”, *Journal of Political Economy*, 111(3), 624-685.

Shanken, J., 1992, “On the estimation of beta-pricing models”, *Review of Financial Studies*, 5(1), 1-33.

Song, S., Tan, J. S., and Yi, Y., 2014, “IPO initial returns in China: underpricing or overvaluation?”, *China Journal of Accounting Research*, 7(1), 31-49.

Sornette, D., Demos, G., Zhang, Q., Cauwels, P., and Zhang, Q., 2015, “Real-time prediction and post-mortem analysis of the Shanghai 2015 stock market bubble and crash”, *Swiss Finance Institute Research Paper*, 15-31.

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

Wang, Y., and Di Iorio, A., 2007, “The cross section of expected stock returns in the Chinese A-share market”, *Global Finance Journal*, 17(3), 335–349.

Yan, W., Powell, J. G., Shi, J., and Xu, W., 2007, “Chinese stock market cyclical regimes: 1991–2006”, *Economics Letters*, 97(3), 235-239.

Zhang, F., Tian, Y., and Wirjanto, T. S., 2009, “Empirical tests of the float-adjusted return model”, *Finance research letters*, 6(4), 219-229.

Table 1

List of Individual Liquidity Proxies

ID	Proxy	Dimension	Definition	Source
1	*Pastor*	Price Impact	The price reversal effects caused by signed trade flows.	Pastor and Stambaugh, 2003
2	*Amihud*	Price Impact	The mean ratio of the absolute daily stock return to the stock's daily trading amount.	Amihud, 2002
3	*Amivest*	Price Impact	The “non-zero-returns” version of the *Amihud* ratio, which equals the average ratio of the trading volume measured with dollars to the absolute value of daily “non-zero-returns”.	Goyenko et al., 2009
4	*BHS*	Price Impact	The "turnover" version of the *Amihud* ratio, which is the mean ratio of absolute daily stock return to the stock's daily turnover.	Brennan et al., 2013
5	*Kang*	Price Impact	The "non-zero-volume" version of the *Amihud* ratio, which is the mean ratio of absolute daily stock return to the stock's daily trading volume, adjusted by the ratio of zero-volume trading days over total trading days.	Kang and Zhang, 2014
6	*HL*	Trading Cost	The trading spread estimated with the daily high and low prices.	Corwin and Schultz, 2012
7	*Roll*	Trading Cost	The effective spread with two multiplied square roots of the covariation between observable changed trading prices in two consecutive days.	Roll, 1984
8	*FHT*	Trading Cost	The trading cost estimated with the number of zero return days and trading days and the standard deviation of non-zero returns.	Fong et al., 2017
9	*AVG_TO*	Trading Quantity	The 12-month average turnover ratios.	Datar et al., 1998
10	*AVG_TV*	Trading Quantity	The 12-month average trading volumes.	Brennan et al., 1988
11	*SD_TO*	Trading Quantity	The 12-month standard deviation of the turnover ratios.	Chordia et al., 2001
12	*SD_TV*	Trading Quantity	The 12-month standard deviation of the trading volumes.	Chordia et al., 2001
13	*AVG_CVTO*	Trading Quantity	The ratio of the 12-month standard deviation of the turnover ratios to the 12-month average turnover ratios.	Chordia et al., 2001
14	*AVG_CVTV*	Trading Quantity	The ratio of the 12-month standard deviation of the trading volumes to the 12-month average trading volumes.	Chordia et al., 2001
15	*Zero1*	Trading Speed	The proportion of the number of zero-return trading days in a month to the number of trading days in a month.	Lesmond et al., 1999
16	*Zero2*	Trading Speed	The proportion of the number of positive trading volume and zero-return days in the month to the number of positive trading volume days in the month.	Lesmond et al., 1999
17	*Liu*	Trading Speed	The turnover adjusted by zero-trading days.	Liu, 2006

Table 2

Summary Statistics of Individual Liquidity Proxies

This table presents the statistics of the 17 single-dimensional liquidity proxies for the period from Jan 2007 to Dec 2023. *Pastor* is the price reversal in Pastor and Stambaugh (2003); *Amihud* is the monthly mean of the absolute value of stock returns above trading volume in Amihud (2002); *Amivest*, *BHS* and *Kang* are the non-zero return, turnover ratio and zero-volume versions of Amihud ratios in Goyenko et al. (2009), Brennan et al. (2013) and Kang and Zhang (2014) respectively; *HL* is the high-low price spread in Corwin and Schultz (2012); *Roll* is the effective bid-ask spread established in Roll (1984); *FHT* is the senior version of zero ratios in Fong et al. (2017); *AVG_TO* is the turnover ratio in Datar et al. (1998); *AVG_TV* is trading volume in Brennan et al. (1988); *SD_TO*, *SD_TV*, *AVG_CVTO* and *AVG_CVTV* are the standard deviation of turnover ratios, the standard deviation of trading volume, the coefficient of variations of turnover ratios and the coefficient of variations of trading volume, in Chordia et al. (2001); *Zero1* and *Zero2* are ratios of zero-return trading days and zero-return days with positive trading volume respectively used in Lesmond et al. (1999); *Liu* is the turnover ratio adjusted zero ratio in Liu (2006).

ID	Proxy	N	Mean	Standard Deviation	*t*-value	Minimum	Maximum
1	*Pastor*	385,373	-0.0015	0.0141	-66.57	-0.5331	2.0134
2	*Amihud*	394,373	40.03	56.73	443.11	0.16	1,570.96
3	*Amivest*	394,459	227.05	632.51	225.45	0.77	79,273.41
4	*BHS*	394,460	0.0244	0.0671	228.35	0.0011	14.5492
5	*Kang*	394,455	-0.8409	0.1031	-5,123.54	-1.3326	-0.2530
6	*HL*	400,263	0.0094	0.0043	1,388.05	0.0000	0.1220
7	*Roll*	400,056	0.2285	0.6838	211.32	0.0000	68.6798
8	*FHT*	400,052	0.0010	0.0017	380.30	0.0000	0.0567
9	*AVG_TO*	400,265	0.3439	0.2817	772.32	0.0017	7.8312
10	*AVG_TV*	400,265	363.07	617.27	0.37	2.99	16,268.53
11	*AVG_CVTO*	395,237	0.5283	0.2008	1654.17	0.0004	2.5925
12	*AVG_CVTV*	395,237	0.6174	0.2369	1638.37	0.0000	3.0181
13	*SD_TO*	395,237	0.1847	0.1716	676.83	0.0001	3.0435
14	*SD_TV*	395,237	221.55	377.33	0.37	0.00	12,033.84
15	*Zero1*	400,265	0.0276	0.0455	383.98	0.0000	0.6842
16	*Zero2*	400,265	0.0276	0.0455	383.98	0.0000	0.6842
17	*Liu*	369,227	3.2273	6.3268	309.96	0.0000	80.8501

Table 3

Correlations Matrix of Individual Liquidity Proxies

This table presents the correlation matrix between single-dimensional liquidity proxies for the period from January 2007 to December 2023. *Pastor* is the price reversal in Pastor and Stambaugh (2003); *Amihud* is the monthly mean of the absolute value of stock returns above trading volume in Amihud (2002); *Amivest*, *BHS* and *Kang* are the non-zero return, turnover ratio and zero-volume versions of Amihud ratios in Goyenko et al. (2009), Brennan et al. (2013) and Kang and Zhang (2014) respectively; *HL* is the high-low price spread in Corwin and Schultz (2012); *Roll* is the effective bid-ask spread established in Roll (1984); *FHT* is the senior version of zero ratios in Fong et al. (2017); *AVG_TO* is the turnover ratio in Datar et al. (1998); *AVG_TV* is trading volume in Brennan et al. (1988); *SD_TO*, *SD_TV*, *AVG_CVTO* and *AVG_CVTV* are the standard deviation of turnover ratios, the standard deviation of trading volume, the coefficient of variations of turnover ratios and the coefficient of variations of trading volume, in Chordia et al. (2001); *Zero1* and *Zero2* are ratios of zero-return trading days and zero-return days with positive trading volume respectively used in Lesmond et al. (1999); *Liu* is the turnover ratio adjusted zero ratio in Liu (2006).

Proxies	Pastor	Amihud	Amivest	BHS	Kang	HL	Roll	FHT	AVG_TO	AVG_TV	AVG_CVTO	AVG_CVTV	SD_TO	SD_TV	Zero1	Zero2	Liu
Pastor	1.0000	-0.0234	0.0255	0.0178	-0.0736	0.0159	0.0122	-0.0110	0.0189	0.0373	-0.0357	-0.0207	0.0087	0.0353	-0.0155	-0.0155	0.0124
Amihud	*-0.0234*	*1.0000*	*-0.1705*	*0.2469*	*0.7039*	0.0521	-0.0557	0.0733	-0.1926	-0.2472	0.0692	0.0577	-0.1634	-0.2393	0.0340	0.0340	0.1029
Amivest	0.0255	-0.1705	1.0000	-0.0235	-0.4730	0.0467	0.2976	-0.0746	0.0771	0.7189	-0.0894	-0.0509	0.0431	0.5464	-0.0830	-0.0830	-0.0252
BHS	0.0178	0.2469	-0.0235	1.0000	0.1574	0.0038	0.0033	0.0111	-0.1472	-0.0378	-0.0028	-0.0206	-0.1394	-0.0467	0.0146	0.0146	0.0534
Kang	-0.0736	0.7039	-0.4730	0.1574	1.0000	0.0013	-0.1437	0.1051	-0.2558	-0.6316	0.1200	0.0661	-0.2024	-0.5793	0.0801	0.0801	0.0817
HL	0.0159	0.0521	0.0467	0.0038	0.0013	1.0000	0.2081	-0.0933	0.2774	0.0791	0.0412	0.1427	0.2285	0.1184	-0.2489	-0.2489	0.1252
Roll	0.0122	-0.0557	0.2976	0.0033	-0.1437	0.2081	1.0000	-0.0914	0.0335	0.2276	-0.1028	-0.0710	-0.0125	0.1404	-0.1143	-0.1143	0.0267
FHT	-0.0110	0.0733	-0.0746	0.0111	0.1051	-0.0933	-0.0914	1.0000	-0.0752	-0.0821	0.0859	0.0638	-0.0341	-0.0618	0.8508	0.8508	-0.0323
AVG_TO	0.0189	-0.1926	0.0771	-0.1472	-0.2558	0.2774	0.0335	-0.0752	1.0000	0.2637	0.0753	0.1570	0.8601	0.3327	-0.1577	-0.1577	0.0408
AVG_TV	0.0373	-0.2472	0.7189	-0.0378	-0.6316	0.0791	0.2276	-0.0821	0.2637	1.0000	-0.0857	-0.0250	0.1945	0.8739	-0.1010	-0.1010	-0.0276
AVG_CVTO	-0.0357	0.0692	-0.0894	-0.0028	0.1200	0.0412	-0.1028	0.0859	0.0753	-0.0857	1.0000	0.9040	0.4622	0.1324	0.0616	0.0616	-0.0274
AVG_CVTV	-0.0207	0.0577	-0.0509	-0.0206	0.0661	0.1427	-0.0710	0.0638	0.1570	-0.0250	0.9040	1.0000	0.4975	0.2280	0.0167	0.0167	-0.0064
SD_TO	0.0087	-0.1634	0.0431	-0.1394	-0.2024	0.2285	-0.0125	-0.0341	0.8601	0.1945	0.4622	0.4975	1.0000	0.3648	-0.1089	-0.1089	0.0032
SD_TV	0.0353	-0.2393	0.5464	-0.0467	-0.5793	0.1184	0.1404	-0.0618	0.3327	0.8739	0.1324	0.2280	0.3648	1.0000	-0.0889	-0.0889	-0.0194
Zero1	-0.0155	0.0340	-0.0830	0.0146	0.0801	-0.2489	-0.1143	0.8508	-0.1577	-0.1010	0.0616	0.0167	-0.1089	-0.0889	1.0000	1.0000	-0.0733
Zero2	-0.0155	0.0340	-0.0830	0.0146	0.0801	-0.2489	-0.1143	0.8508	-0.1577	-0.1010	0.0616	0.0167	-0.1089	-0.0889	1.0000	1.0000	-0.0733
Liu	0.0124	0.1029	-0.0252	0.0534	0.0817	0.1252	0.0267	-0.0323	0.0408	-0.0276	-0.0274	-0.0064	0.0032	-0.0194	-0.0733	-0.0733	1.0000

Table 4

Characteristic-Level Horserace Tests

This table reports the results of the firm characteristic-level horserace tests. The horserace includes two Fama-MacBeth type cross-sectional regressions, univariate regression and multivariate regressions. The univariate regression has only one explanatory variable, *Liqbeta_APC* or *Liqbeta_PCA*. The multivariate regression also contains other firm characteristics, including the asset-to-market ratio (*AM*), book-to-market ratio (*BM*), cash-to-price ratio (*CP*), earning-to-price ratio (*EP*), the natural logarithm of market caps (*Size*), return-to-equity ratio (*ROE*), asset-growth ratio (*AG*), market beta (*Beta*), and the dummy variables for negative EP and CP (D_EP and D_CP). Significant *Liqbeta_APCs* in the regressions are presented in Panel A, and Significant *Liqbeta_PCAs* are presented in Panel B. The sample period is from January 2007 to December 2023. * represents significance at 10% level; ** represents significance at 5% level; *** represents significance at 1% level.

Panel A: Significant *Liqbeta_APC* in univariate and multivariable cross-sectional regressions

ID	x1	x2	x3	x4	x5	x6	Univariate		Multivariate	
							Coefficient	t-value	Coefficient	t-value
1	*Amihud*						-0.0001	-3.05***	0.0000	-2.73***
2	*Pastor*						-0.0001	-1.76*	-0.0001	-1.78*
3	*FHT*						-0.0019	-2.52**	-0.0014	-2.49**
4	*Amihud*	*FHT*					-0.0003	-1.66*	-0.0002	-1.67*
5	*Liu*	*FHT*					-0.0004	-1.78*	-0.0003	-1.75*
6	*HL*	*FHT*					-0.0012	-2.75***	-0.0010	-3.06***
7	*FHT*	*Pastor*					-0.0018	-2.52**	-0.0013	-2.49**
8	*Amihud*	*HL*	*FHT*				-0.0005	-2.49**	-0.0004	-2.68***
9	*Liu*	*Pastor*	*FHT*				-0.0004	-1.80*	-0.0003	-1.77*
10	*Liu*	*FHT*	*HL*				-0.0006	-2.27**	-0.0005	-2.68***
11	*Pastor*	*FHT*	*HL*				-0.0012	-2.74***	-0.0010	-3.06***
12	*Amihud*	*Liu*	*HL*	*FHT*			-0.0005	-2.27**	-0.0004	-2.55**
13	*Amihud*	*HL*	*Pastor*	*FHT*			-0.0005	-2.48**	-0.0004	-2.67***
14	*Liu*	*Pastor*	*FHT*	*HL*			-0.0006	-2.29**	-0.0005	-2.70***
15	*Amihud*	*Liu*	*HL*	*FHT*	*Pastor*		-0.0004	-2.28**	-0.0004	-2.56**
16	*Amihud*	*Pastor*					<-0.0001	-2.05**	Insignificant	
17	*Amihud*	*HL*					-0.0002	-1.79*	Insignificant	
18	*Amihud*	*HL*	*Pastor*				-0.0002	-1.82*	Insignificant	

Panel B Significant *Liqbeta_PCA* in univariate and multivariable cross-sectional regressions

ID	x1	x2	x3	x4	x5	x6	Univariate		Multivariate	
							Coefficient	t-value	Coefficient	t-value
1	*Amihud*	*Roll*					Insignificant		0.0207	1.83*
2	*Liu*	*Roll*					Insignificant		-0.0182	-1.71*
3	*Amihud*	*Liu*	*Roll*				Insignificant		-0.0209	-1.90*

Table 5

Factor-Level Regression and Pairwise Comparisons on *MPB*

This table presents the results of the factor-level regression based pairwise comparisons. Factor-level fama-macbeth cross-sectional regressions of portfolio excess returns on the factors in the liquidity-augmented factor pricing models. The *LIQ* factor is constructed by each of the competing composite liquidity proxies. Three appraisal metrics are calculated: *N_LIQ*, *N_A*, and *MDR*. The probabilities that a proxy can win in terms of a particular metric are the *PROBL* (*N_LIQ*), *PROBA* (*N_A*), and *PROBR* (*MDR*). The mean winning probability of a composite proxy is *MPB*. A normality test, four non-parametric tests, and one bootstrapping simulation are also performed. *P_SW* is the *p*-value of the Shapiro-Wilk normality test for the null hypothesis that the *MPB* is normally distributed. *P_ST* is the p-value of the Sign test for the null hypothesis that the *MPB* equals zero. *XPL_WIL, XPR_WIL, XP2_WIL*, and *P_ST2* are one-way(left), one-way (right), and two-way exact *p*-values of the Wilcoxon rank-sum test and the *p*-value of the second Sign test, respectively. The simulated *MPB* (*MPB_SIMU*) and their *p*-values (*P_VALUE_SIMU*) are presented. The simulated *p*-values are for the null hypothesis that the *MPB* of the optimal composite proxy equals the *MPB* of other composite proxies. The sample period is from January 2007 to December 2023.

ID	x1	x2	x3	x4	x5	x6	*N_LIQ*	*N_A*	*MDR*	*PROBL*	*PROBA*	*PROBR*	*MPB*	*P_SW*	*P_ST*	*XPL_WIL*	*XPR_WIL*	*XP2_WIL*	*P_ST2*	*MPB_SIMU*	*P_VALUE_SIMU*
1	***Amihud***	***HL***	***FHT***				**10**	**4**	**0.0459**	**0.9286**	**0.7857**	**1.0000**	**0.9048**	**0.0000**	**0.0001**	**N/A**	**N/A**	**N/A**	**N/A**	**0.9109**	**1.0000**
2	***Amihud***	***HL***	***FHT***	***Pastor***			**10**	**4**	**0.0453**	**0.9286**	**0.7857**	**0.9286**	**0.8810**	**0.0000**	**0.0002**	**N/A**	**0.5000**	**1.0000**	**N/A**	**0.8792**	**0.5180**
3	***Amihud***	***HL***	***FHT***	***Pastor***	***Liu***		**8**	**6**	**0.0334**	**0.8571**	**1.0000**	**0.7143**	**0.8571**	**0.0001**	**0.0001**	**N/A**	**0.3857**	**0.7714**	**1.0000**	**0.8541**	**0.3820**
4	*Amihud*	*HL*	*FHT*	*Liu*			7	5	0.0325	0.6429	0.9286	0.6429	0.7381	0.0014	0.0002	N/A	0.1043	0.2087	0.1250	0.7342	0.1580
5	*HL*	*FHT*	*Pastor*				7	3	0.0420	0.6429	0.7143	0.7857	0.7143	0.0010	0.0005	N/A	0.0930	0.1861	0.0625	0.7139	0.1580
6	*HL*	*FHT*	*Pastor*	*Liu*			4	2	0.0450	0.5714	0.5714	0.8571	0.6667	0.0007	0.0005	N/A	0.0610	0.1220	0.0625	0.6683	0.1240
7	*HL*	*FHT*					7	2	0.0000	0.6429	0.5714	0.4286	0.5476	0.0015	0.0039	0.0141	N/A	0.0282	0.0156	0.5449	0.0420
8	*HL*	*FHT*	*Liu*				3	1	0.0302	0.5000	0.4286	0.5000	0.4762	0.0006	0.0078	N/A	0.0078	0.0155	0.0156	0.4647	0.0280
9	*FHT*	*Pastor*	*Liu*				2	1	0.0324	0.2143	0.4286	0.5714	0.4048	0.0068	0.0078	N/A	0.0005	0.0011	0.0020	0.4149	0.0070
10	*Pastor*						2	0	N/A	0.2143	0.0000	0.0000	0.0714	0.0000	0.2500	0.0000	N/A	0.0000	0.0002	0.0696	0.0000
11	*Amihud*	*FHT*					2	0	N/A	0.2143	0.0000	0.0000	0.0714	0.0000	0.2500	0.0000	N/A	0.0000	0.0002	0.0722	0.0000
12	*FHT*	*Pastor*					2	0	N/A	0.2143	0.0000	0.0000	0.0714	0.0000	0.2500	0.0000	N/A	0.0000	0.0002	0.0713	0.0000
13	*FHT*						1	0	N/A	0.0714	0.0000	0.0000	0.0238	0.0000	1.0000	0.0000	N/A	0.0000	0.0002	0.0243	0.0000
14	*FHT*	*Liu*					1	0	N/A	0.0714	0.0000	0.0000	0.0238	0.0000	1.0000	0.0000	N/A	0.0000	0.0002	0.0243	0.0000
15	*Amihud*						0	0	N/A	0.0000	0.0000	0.0000	0.0000	N/A	N/A	0.0000	N/A	0.0000	0.0002	0.0000	0.0000

Table 6
Factor-Level Regression and Pairwise Comparisons on *MPBM*

This table presents the results of the factor-level regression based pairwise comparisons with the alternative winning probability, *MPBM*. Factor-level Fama-MacBeth cross-sectional regressions of portfolio excess returns on the factors in the five liquidity-augmented factor pricing models. The *LIQ* factor is constructed by each of the competing composite liquidity proxies. Three appraisal metrics are calculated: *N_LIQ*, *N_A*, and *MDR*. The probabilities that a proxy can win in terms of a particular metric are the *PROBL* (*N_LIQ*), *PROBA* (*N_A*), and *PROBR* (*MDR*). The variable "win or loss" (*WOL*) calculates a composite proxy's probability of winning a best-of-three match in terms of the three appraisal metrics. If a composite proxy wins a best-of-three match, *WOL* equals one and zero; otherwise, it equals zero. We take the average of a composite proxy's *WOLs* to obtain the composite proxy's mean probability of winning a match (*MPBM*). A normality test, four non-parametric tests, and one bootstrapping simulation are also performed. *P_SW* is the *p*-value of the Shapiro-Wilk normality test for the null hypothesis that the *MPB* is normally distributed. *P_ST* is the p-value of the Sign test for the null hypothesis that the *MPB* equals zero. *PL_WIL, PR_WIL, P2_WIL*, and *P_ST2* are one-way(left), one-way (right), and two-way simple *p*-values of the Wilcoxon rank-sum test and the *p*-value of the second Sign test, respectively. The simulated *MPB* (*MPB_SIMU*) and their *p*-values (*P_VALUE_SIMU*) are presented. The simulated *p*-values are for the null hypothesis that the *MPB* of the optimal composite proxy equals the *MPB* of other composite proxies. The sample period is from January 2007 to December 2023.

ID	x1	x2	x3	x4	x5	x6	*N_LIQ*	*N_A*	*MDR*	*PROBL*	*PROBA*	*PROBR*	*MPBM*	*P_SW*	*P_ST*	*PL_WIL*	*PR_WIL*	*P2_WIL*	*P_ST2*	*MPBM_SIMU*	*P_VALUE_SIMU*
1	***Amihud***	***HL***	***FHT***				**10**	**4**	**0.0459**	**0.9286**	**0.7857**	**1.0000**	**1.0000**	**N/A**	**0.0000**	**N/A**	**N/A**	**N/A**	**N/A**	**1.0000**	**1.0000**
2	***Amihud***	***HL***	***FHT***	***Pastor***			**10**	**4**	**0.0453**	**0.9286**	**0.7857**	**0.9286**	**0.9286**	**0.0000**	**0.0000**	**N/A**	**0.0000**	**0.0000**	**0.0000**	**0.9281**	**0.0000**
3	***Amihud***	***HL***	***FHT***	***Pastor***	***Liu***		**8**	**6**	**0.0334**	**0.8571**	**1.0000**	**0.7143**	**0.8571**	**0.0000**	**0.0000**	**N/A**	**0.0000**	**0.0000**	**0.0000**	**0.8501**	**0.0000**
4	*Amihud*	*HL*	*FHT*	*Liu*			7	3	0.0420	0.6429	0.7143	0.7857	0.7143	0.0000	0.0000	N/A	0.0000	0.0000	0.0000	0.7144	0.0000
5	*HL*	*FHT*	*Pastor*				7	5	0.0325	0.6429	0.9286	0.6429	0.7143	0.0000	0.0000	N/A	0.0000	0.0000	0.0000	0.7084	0.0000
6	*HL*	*FHT*	*Pastor*	*Liu*			7	2	0.0000	0.6429	0.5714	0.4286	0.5714	0.0000	0.0000	0.0000	N/A	0.0000	0.0000	0.5673	0.0000
7	*HL*	*FHT*					4	2	0.0450	0.5714	0.5714	0.8571	0.5714	0.0000	0.0000	N/A	0.0000	0.0000	0.0000	0.5761	0.0000
8	*HL*	*FHT*	*Liu*				2	1	0.0324	0.2143	0.4286	0.5714	0.4286	0.0000	0.0000	N/A	0.0000	0.0000	0.0000	0.4443	0.0000
9	*FHT*	*Pastor*	*Liu*				3	1	0.0302	0.5000	0.4286	0.5000	0.4286	0.0000	0.0000	N/A	0.0000	0.0000	0.0000	0.4188	0.0000
10	*Pastor*						2	0	N/A	0.2143	0.0000	0.0000	0.2143	0.0000	0.0000	0.0000	N/A	0.0000	0.0000	0.2088	0.0000
11	*Amihud*	*FHT*					2	0	N/A	0.2143	0.0000	0.0000	0.2143	0.0000	0.0000	0.0000	N/A	0.0000	0.0000	0.2168	0.0000
12	*FHT*	*Pastor*					2	0	N/A	0.2143	0.0000	0.0000	0.2143	0.0000	0.0000	0.0000	N/A	0.0000	0.0000	0.2137	0.0000
13	*FHT*						1	0	N/A	0.0714	0.0000	0.0000	0.0714	0.0000	0.0000	0.0000	N/A	0.0000	0.0000	0.0729	0.0000
14	*FHT*	*Liu*					1	0	N/A	0.0714	0.0000	0.0000	0.0714	0.0000	0.0000	0.0000	N/A	0.0000	0.0000	0.0728	0.0000
15	*Amihud*						0	0	N/A	0.0000	0.0000	0.0000	0.0000	N/A	N/A	0.0000	N/A	0.0000	0.0000	0.0000	0.0000

Table 7

Regression of *Amihud-HL-FHT* on the Six Individual Liquidity Proxies

This table presents the regression result of the optimal composite liquidity proxy, *Amihud-HL-FHT*, on the six standardized individual liquidity proxies: *Amihud*, *FHT*, *HL*, *Liu*, *Pastor*, and *Roll*. The tests include six univariate regressions and one multivariate regression. The sample period is from January 2007 to December 2023. * represents significance at 10% level; ** represents significance at 5% level; *** represents significance at 1% level.

Regressions	Types	*Amihud*	*FHT*	*HL*	*Liu*	*Pastor*	*Roll*
Univariate	Coefficient	-0.4144	-0.1218	0.5114	0.2134	0.0481	0.2382
	t-value	-7.18	-7.28***	10.66***	5.22***	1.31	7.35***
Multivariate	Coefficient	-0.4367	-0.0742	0.4172	0.2289	0.0375	0.0506
	t-value	-7.29***	-4.91***	9.72***	5.92***	1.05	2.09**

Table 8

Comparisons between the Optimal Composite Proxy and Its Nested Proxies

The characteristic-level horserace and the factor-level regression and pairwise comparisons are performed on the optimal composite proxy and its nested composite proxies for robustness. The nested composite proxies of the optimal composite proxy are constructed by adding or removing a single-dimensional liquidity proxy to or from the optimal composite proxy. *UNI* and *T_UNI* represent the coefficient and its *t*-value of the univariate Fama-MacBeth cross-sectional regression. *MUL* and *T_MUL* represent the coefficient and its t-value of the multivariate Fama-MacBeth cross-sectional regression. The definitions of the *N_LIQ, N_A*, and *MDR* are the same as those in Tables 4 and 5. *SCORE_LIQ, SCORE_A*, and *SCORE_R* are the composite proxy's scores of *N_LIQ, N_A*, and *MDR* comparison in the match, respectively. The *TOTAL_SCORE* is the sum of the *SCORE_LIQ, SCORE_A*, and *SCORE_R*. *PROB* is the mean probability that the composite proxy can win a metric comparison in the match and equals the *TOTAL_SCORE* divided by 3. The sample period is from January 2007 to December 2023.

Match	x1	x2	x3	x4	x5	x6	*UNI*	*T_UNI*	*MULTI*	*T_MUL*	*N_LIQ*	*N_A*	*MDR*	*SCORE_LIQ*	*SCORE_A*	*SCORE_R*	*TOTAL_SCORE*	*PROB*
1	Amihud	HL					-0.0002	-1.79	-0.0001	-1.60	N/A	N/A	N/A	0	0	0	0	0.0000
	Amihud	**HL**	**FHT**				**-0.0005**	**-2.49**	**-0.0004**	**-2.68**	**10**	**4**	**0.0459**	**1**	**1**	**1**	**3**	**1.0000**
2	Amihud	FHT					-0.0003	-1.66	-0.0002	-1.67	2	N/A	N/A	0	0	0	0	0.0000
	Amihud	**HL**	**FHT**				**-0.0005**	**-2.49**	**-0.0004**	**-2.68**	**10**	**4**	**0.0459**	**1**	**1**	**1**	**3**	**1.0000**
3	HL	FHT					-0.0012	-2.75	-0.0010	-3.06	7	2	0.0000	0	0	0	0	0.0000
	Amihud	**HL**	**FHT**				**-0.0005**	**-2.49**	**-0.0004**	**-2.68**	**10**	**4**	**0.0459**	**1**	**1**	**1**	**3**	**1.0000**
4	Amihud	HL	FHT	Pastor			-0.0005	-2.48	-0.0004	-2.67	10	4	0.0453	0	0	0	0	0.0000
	Amihud	**HL**	**FHT**				**-0.0005**	**-2.49**	**-0.0004**	**-2.68**	**10**	**4**	**0.0459**	**0**	**0**	**1**	**1**	**0.3333**
5	Amihud	HL	FHT	Liu			-0.0005	-2.27	-0.0004	-2.55	7	5	0.0325	0	1	0	1	0.3333
	Amihud	**HL**	**FHT**				**-0.0005**	**-2.49**	**-0.0004**	**-2.68**	**10**	**4**	**0.0459**	**1**	**0**	**1**	**2**	**0.6667**
6	Amihud	HL	FHT	Roll			-0.0004	-0.57	0.0002	0.55	N/A	N/A	N/A	0	0	0	0	0.0000
	Amihud	**HL**	**FHT**				**-0.0005**	**-2.49**	**-0.0004**	**-2.68**	**10**	**4**	**0.0459**	**1**	**1**	**1**	**3**	**1.0000**

Table 9

Factor-Level Regression and Pairwise Comparisons with Appraisal Metrics Calculated with 10% Significance Level

The characteristic-level horserace and the factor-level regression and pairwise comparisons are performed by using the appraisal metrics *N_LIQ* and *N_A* using the 10% significance level. The definitions of the variables are the same as those reported in Table 4. The sample period is from January 2007 to December 2023.

ID	x1	x2	x3	x4	x5	x6	*N_LIQ*	*N_A*	*MDR*	*PROBL*	*PROBA*	*PROBR*	*MPB*	*P_SW*	*P_ST*	*XPL_WIL*	*XPR_WIL*	*XP2_WIL*	*P_ST2*	*MPB_SIMU*	*P_VALUE_SIMU*
1	***Amihud***	***HL***	***FHT***				**22**	**16**	**0.0414**	**1.0000**	**1.0000**	**0.6429**	**0.8810**	**0.0001**	**0.0001**	**N/A**	**N/A**	**N/A**	**N/A**	**0.8816**	**1.0000**
2	***Amihud***	***HL***	***FHT***	***Pastor***			**20**	**14**	**0.0398**	**0.9286**	**0.9286**	**0.5714**	**0.8095**	**0.0002**	**0.0002**	**N/A**	**0.3727**	**0.7455**	**N/A**	**0.8077**	**0.3680**
3	***HL***	***FHT***	***Pastor***				**16**	**7**	**0.0441**	**0.7857**	**0.6429**	**0.8571**	**0.7619**	**0.0020**	**0.0001**	**N/A**	**0.1601**	**0.3201**	**0.5000**	**0.7624**	**0.2120**
4	*HL*	*FHT*	*Pastor*	*Liu*			8	5	0.0491	0.5714	0.5000	1.0000	0.6905	0.0003	0.0001	N/A	0.0558	0.1116	0.0625	0.6921	0.1140
5	*Amihud*	*HL*	*FHT*	*Liu*			10	7	0.0431	0.6429	0.6429	0.7857	0.6905	0.0039	0.0002	N/A	0.1002	0.2004	0.1250	0.6824	0.1260
6	*Amihud*	*HL*	*FHT*	*Pastor*	*Liu*		10	7	0.0427	0.6429	0.6429	0.7143	0.6667	0.0036	0.0005	N/A	0.0876	0.1752	0.1250	0.6647	0.0980
7	*HL*	*FHT*					16	7	0.0363	0.7857	0.6429	0.5000	0.6429	0.0022	0.0010	0.1044	N/A	0.2087	0.1250	0.6420	0.1160
8	*HL*	*FHT*	*Liu*				7	5	0.0488	0.5000	0.5000	0.9286	0.6429	0.0009	0.0002	N/A	0.0424	0.0848	0.0313	0.6322	0.0770
9	*FHT*	*Pastor*	*Liu*				3	2	0.0338	0.2143	0.3571	0.4286	0.3333	0.0007	0.0313	N/A	0.0008	0.0017	0.0020	0.3503	0.0050
10	*FHT*	*Liu*					3	2	0.0309	0.2143	0.3571	0.3571	0.3095	0.0002	0.0625	0.0008	N/A	0.0017	0.0020	0.3038	0.0030
11	*FHT*	*Pastor*					4	1	0.0250	0.3571	0.2857	0.2857	0.3095	0.0005	0.0313	0.0005	N/A	0.0011	0.0020	0.3100	0.0010
12	*Pastor*						4	0	N/A	0.3571	0.0000	0.0000	0.1190	0.0001	0.0625	0.0000	N/A	0.0000	0.0002	0.1179	0.0000
13	*FHT*						2	0	N/A	0.0714	0.0000	0.0000	0.0238	0.0000	1.0000	0.0000	N/A	0.0000	0.0002	0.0243	0.0000
14	*Amihud*	*FHT*					2	0	N/A	0.0714	0.0000	0.0000	0.0238	0.0000	1.0000	0.0000	N/A	0.0000	0.0002	0.0240	0.0000
15	*Amihud*						0	0	N/A	0.0000	0.0000	0.0000	0.0000	N/A	N/A	0.0000	N/A	0.0000	0.0002	0.0000	0.0000

Table 10

Factor-Level Pairwise Comparison using the different estimation window

The characteristic-level horserace and the factor-level regression and pairwise comparisons are performed by extending the sample period to cover the Split-Share-Structure Reform during 2005-2006. The definitions of the variables are the same as those reported in Table 4. The sample period is from January 2005 to December 2023.

ID	x1	x2	x3	x4	x5	x6	*N_LIQ*	*N_A*	*MDR*	*PROBL*	*PROBA*	*PROBR*	*MPB*	*P_SW*	*P_ST*	*XPL_WIL*	*XPR_WIL*	*XP2_WIL*	*P_ST2*	*MPB_SIMU*	*P_VALUE_SIMU*
1	***Amihud***	***HL***	***FHT***				**17**	**13**	**0.0391**	**1.0000**	**1.0000**	**0.5000**	**0.8333**	**0.0001**	**0.0001**	**0.5000**	**N/A**	**1.0000**	**1.0000**	**0.8346**	**1.0000**
2	***Amihud***	***HL***	***FHT***	***Pastor***			**15**	**10**	**0.0407**	**0.9286**	**0.9286**	**0.6429**	**0.8333**	**0.0008**	**0.0001**	**N/A**	**N/A**	**N/A**	**N/A**	**0.8331**	**0.5530**
3	***HL***	***FHT***	***Pastor***				**8**	**2**	**0.0690**	**0.7857**	**0.6429**	**0.8571**	**0.7619**	**0.0020**	**0.0001**	**0.3000**	**N/A**	**0.5999**	**0.5000**	**0.7633**	**0.3310**
4	*Amihud*	*HL*	*FHT*	*Liu*			6	4	0.0557	0.6429	0.7857	0.7857	0.7381	0.0002	0.0001	0.2507	N/A	0.5015	0.6250	0.7337	0.2970
5	*HL*	*FHT*					8	1	0.0804	0.7857	0.4286	1.0000	0.7381	0.0047	0.0001	0.2101	N/A	0.4202	0.2500	0.7363	0.2600
6	*Amihud*	*HL*	*FHT*	*Pastor*	*Liu*		6	4	0.0396	0.6429	0.7857	0.5714	0.6667	0.0036	0.0005	N/A	0.1788	0.3575	0.2500	0.6620	0.1980
7	*HL*	*FHT*	*Pastor*	*Liu*			4	2	0.0456	0.5714	0.6429	0.7143	0.6429	0.0028	0.0005	N/A	0.1322	0.2643	0.1250	0.6476	0.1760
8	*Pastor*						2	1	0.0801	0.3571	0.4286	0.9286	0.5714	0.0032	0.0002	0.0238	N/A	0.0476	0.0156	0.5736	0.0520
9	*HL*	*FHT*	*Liu*				3	1	0.0319	0.5000	0.4286	0.4286	0.4524	0.0002	0.0156	0.0309	N/A	0.0618	0.0156	0.4363	0.0410
10	*Amihud*	*FHT*					2	0	N/A	0.3571	0.0000	0.0000	0.1190	0.0001	0.0625	0.0000	N/A	0.0000	0.0002	0.1208	0.0000
11	*FHT*						1	0	N/A	0.0714	0.0000	0.0000	0.0238	0.0000	1.0000	0.0000	N/A	0.0000	0.0002	0.0243	0.0000
12	*Liu*	*FHT*					1	0	N/A	0.0714	0.0000	0.0000	0.0238	0.0000	1.0000	0.0000	N/A	0.0000	0.0002	0.0243	0.0000
13	*FHT*	*Pastor*					1	0	N/A	0.0714	0.0000	0.0000	0.0238	0.0000	1.0000	0.0000	N/A	0.0000	0.0002	0.0222	0.0000
14	*FHT*	*Pastor*	*Liu*				1	0	N/A	0.0714	0.0000	0.0000	0.0238	0.0000	1.0000	0.0000	N/A	0.0000	0.0002	0.0256	0.0000
15	*Amihud*						0	0	N/A	0.0000	0.0000	0.0000	0.0000	N/A	N/A	0.0000	N/A	0.0000	0.0002	0.0000	0.0000

Table 11

Factor-Level Regression and Pairwise Comparisons using the Total Rank

The characteristic-level horserace and the factor-level regression and pairwise comparisons are performed using the total rank of the three metric-specific ranks: *RANK_LIQ*, *RANK_A*, and *RANK_R*. These metric-specific ranks are obtained by ranking the *N_LIQ, N_A*, and *MDR* in descending order. The sum of the three metric-specific ranks is the total rank (*TOTAL_RANK*). The sample period is from January 2007 to December 2023.

ID	x1	x2	x3	x4	x5	x6	*N_LIQ*	*N_A*	*MDR*	*RANK_LIQ*	*RANK_A*	*RANK_R*	*TOTAL_RANK*
1	***Amihud***	***HL***	***FHT***				**10**	**4**	**0.0459**	**1**	**3**	**1**	**5**
2	***Amihud***	***HL***	***FHT***	***Pastor***			**10**	**4**	**0.0453**	**1**	**3**	**2**	**6**
3	***Amihud***	***HL***	***FHT***	***Pastor***	***Liu***		**8**	**6**	**0.0334**	**3**	**1**	**5**	**9**
4	*Amihud*	*HL*	*FHT*	*Liu*			7	5	0.0325	4	2	6	12
5	*Pastor*	*HL*	*FHT*				7	3	0.0420	4	5	4	13
6	*HL*	*FHT*	*Pastor*	*Liu*			4	2	0.0450	7	6	3	16
7	*HL*	*FHT*					7	2	0.0000	4	6	9	19
8	*FHT*	*Pastor*	*Liu*				2	1	0.0324	9	8	7	24
9	*HL*	*FHT*	*Liu*				3	1	0.0302	8	8	8	24
10	*Pastor*						2	0	N/A	9	10	10	29
11	*Amihud*	*FHT*					2	0	N/A	9	10	10	29
12	*FHT*	*Pastor*					2	0	N/A	9	10	10	29
13	*FHT*						1	0	N/A	13	10	10	33
14	*FHT*	*Liu*					1	0	N/A	13	10	10	33
15	*Amihud*						0	0	N/A	15	10	10	35

Figure 1

Procedure of the construction and selection of the composite liquidity proxy

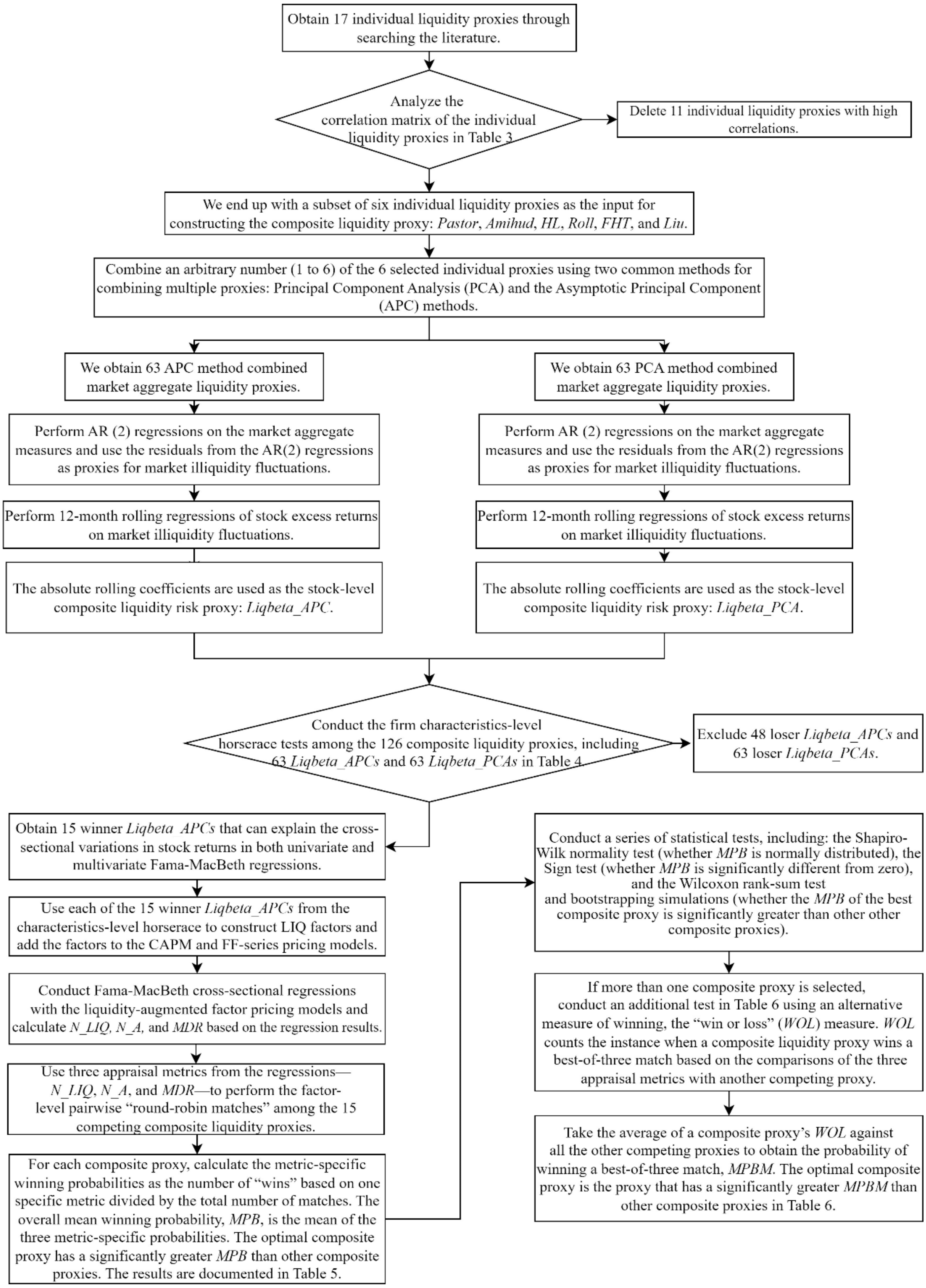
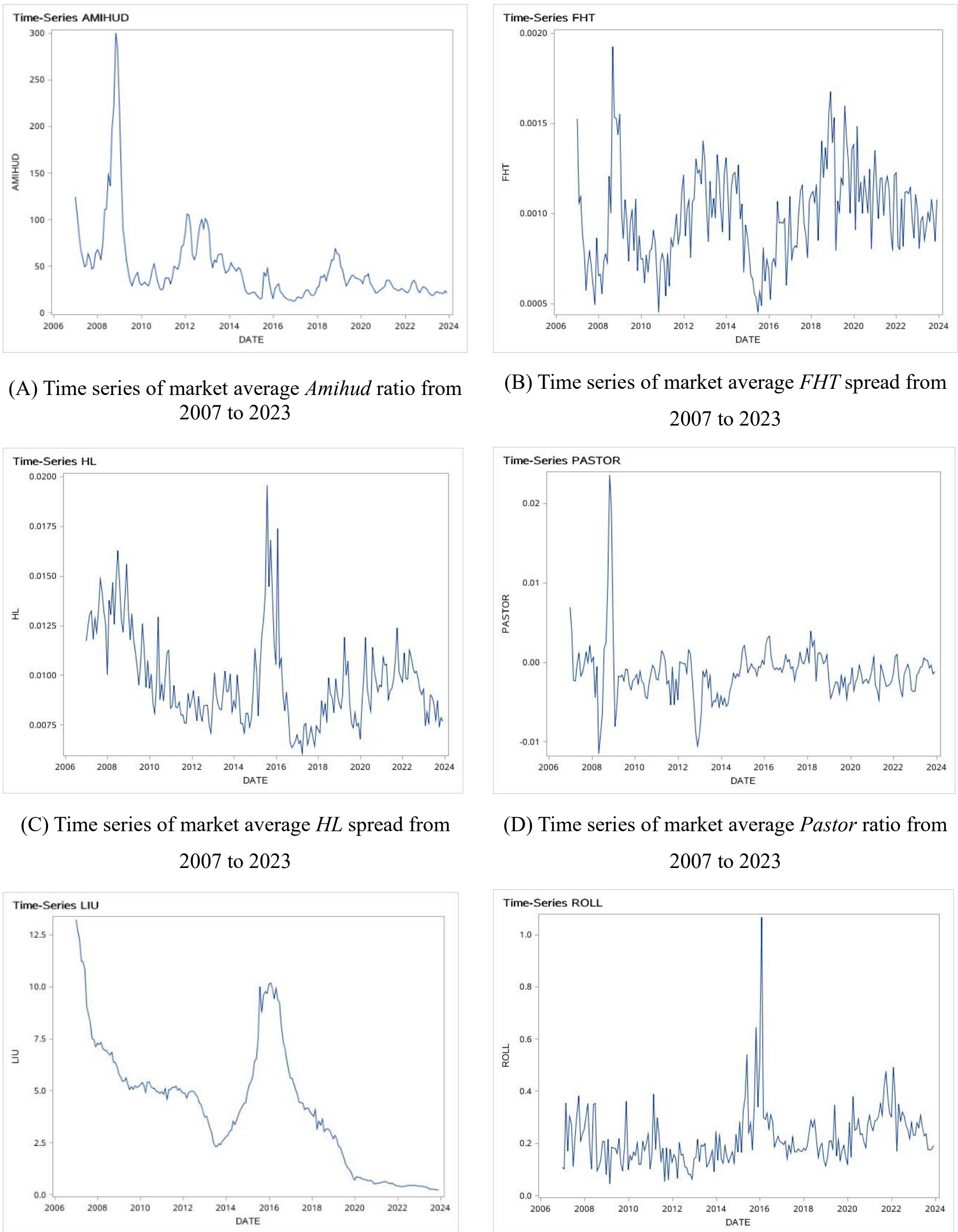


Figure 2

Time-Series of the Six Individual Liquidity Proxies



(A) Time series of market average *Amihud* ratio from 2007 to 2023

(B) Time series of market average *FHT* spread from 2007 to 2023

(C) Time series of market average *HL* spread from 2007 to 2023

(D) Time series of market average *Pastor* ratio from 2007 to 2023

(E) Time series of market average *Liu* spread from 2007 to 2023

(F) Time series of market average *Roll* spread from 2007 to 2023

Appendix A

Construction of the individual liquidity proxies

We construct 17 individual liquidity proxies: *Pastor* (price reversal/impact), *Amihud* (price impact), *Amivest* (price impact), *BHS* (price impact), *Kang* (price impact), *HL* (trading cost), *Roll* (trading cost), *FHT* (trading cost), *TO* (trading quantity), *TV* (trading quantity), *SDTO* (trading quantity), *SDTV* (trading quantity), *CVTO* (trading quantity), *CVTV* (trading quantity), *Zero1&2* (trading speed), and *Liu* (trading speed). These individual liquidity proxies are defined and calculated as follows:

A. Price impact

(1) *Pastor* (Pastor and Stambaugh, 2003): The Pastor ratio captures the price reversal effects due to signed trade flows. The formula of the Pastor ratio is as follows:

$$r'_{i,d+1,t} = \theta_{i,t} + \phi_{i,t} sign\left(r'_{i,d,t}\right) v_{i,d,t} + \varepsilon_{i,d+1,t}, \quad d = 1, \ldots, D(D \geq 8) \ , \qquad (A.1)$$

where $r_{i,d,t}$ is stock *i*'s return on day *d*; $r'_{i,d,t} = r_{i,d,t} - r_{m,d,t}$ where $r_{m,d,t}$ is the daily market return for day *d*; and $v_{i,d,t}$ is stock *i*'s trading amount on day *d*.

The average liquidity measure over 3 months is calculated as follows:

$$\hat{\gamma}_i = {}^{1}/_{N} \sum \hat{\gamma}_{i,t} \ (N = 3)$$

(2) *Amihud* (Amihud, 2002): The Amihud ratio is the mean ratio of the absolute daily stock return to its trading amount. We delete the highest 0.5% tail observations and take the average of the ratio over three months.

$$Amihud_{i,t} = \left({}^{1}/_{N_{i,t}}\right) \Sigma_d \left(\left|r_{i,d,t}\right| / vol_{i,d,t}\right) \qquad (A.2)$$

where $r_{i,d,t}$ represents stock *i*'s return on day *d*; $vol_{i,d,t}$ represents stock *i*'s daily trading amount on day *d*, and $N_{i,t}$ represents the number of trading days that stock *i* has in the month.

(3) *Amivest* (Goyenko et al., 2009): It is a non-zero-return days version of the Amihud ratio, which equals the average ratio of the trading volume measured with dollars, TV_{ei}, to the absolute value of daily "non-zero-returns" on day t, $|r_t|$.

$$\text{Amivest}_{i,t} = \frac{1}{N} \Sigma \frac{TV_{ei}}{|r_t|} \ . \qquad (A.3)$$

(4) *BHS* (Brennan et al., 2013): It is the turnover ratio revised version of the Amihud ratio.

$$\text{BHS} = \frac{1}{N} \Sigma \frac{|r_{i,d}|}{TO_{i,d}} \qquad (A.4)$$

where $\left|r_{i,d}\right|$ represents stock *i*'s absolute value of returns for day *d*; $TO_{i,d}$ represents the daily turnover ratio of stock i on day d.

(5) *Kang* (Kang and Zhang, 2014): It is the zero-trading volume days revised version of the Amihud ratio.

$$\text{Kang}_{i,m} = \left[\ln \left(\frac{1}{N_{i,m}} \sum_{t=1}^{N_{t,m}} \frac{|R_{i,t}|}{Vol_{i,t}}\right)\right] \times (1 + ZeroVol_{i,m}) \qquad (A.5)$$

where $N_{i,m}$ is the number of non-zero-volume days that stock *i* has in the month; $|R_{i,t}|$ represents stock *i*'s absolute return for day *t*; $Vol_{i,t}$ represents stock i's trading amount for day *t*; $ZeroVol_{i,m}$ represents the zero-volume trading days over the total trading days in the month.

B. Trading costs

(6) *HL* (Corwin and Schultz, 2012): the trading cost estimated by the daily high and low prices.

$$HL = \frac{2(e^{\alpha}-1)}{1+e^{\alpha}} \quad , \tag{A.6}$$

where $\alpha = \frac{(\sqrt{2}\beta-\sqrt{\beta})}{(3-2\sqrt{2})} - \sqrt{\left[\gamma / (3-2\sqrt{2})\right]}$ and, $\beta = E\left\{\sum_{j=0}^{1}\left[ln\left(H_{t+j}^{0}/L_{t+j}^{0}\right)\right]^2\right\}$ and γ = [ln($H^0_{t,t+1}$/$L^0_{t,t+1}$)]^2} $\gamma = \left[\ln\left(H_{t,t+1}^{0}/L_{t,t+1}^{0}\right]^2\right.$ with, H_{t+j}^{0} and L_{t+j}^{0} are the stock's highest and lowest prices on day $t+j$, and $H_{t,t+1}^{0}$ and $L_{t,t+1}^{0}$ are the highest and lowest stock prices for the period ranging from day *t* to *t+1*.

(7) *Roll* (Roll, 1984): the effective spread with two multiplied square roots of the covariation between observable changed trading prices at time t and t-1.

$$\mathrm{S}_{Roll} = 2 \times \sqrt{-COV\ \ (\Delta P_t, \Delta P_{t-1})} \quad , \tag{A.7}$$

where *S* is the spread; ΔP_t is the changed trading prices at time *T* and equals to $\Delta P_t \equiv P_t - P_{t-1}$.

$$\mathrm{S}_{Roll} = \begin{cases} 0 & \text{if } COV\ \ (\Delta P_t, \Delta P_{t-1}) \geq 0 \\ 2 \times \sqrt{-COV\ \ (\Delta P_t, \Delta P_{t-1})} & \text{if } COV\ \ (\Delta P_t, \Delta P_{t-1}) \leq 0 \end{cases} \quad . \tag{B.8}$$

(8) *FHT* (Fong et al., 2017): the transaction cost with trading days and the standard deviation of non-zero returns.

$$\mathrm{FHT} \equiv \mathrm{S} \equiv 2\sigma N^{-1}\left(\frac{1+z}{2}\right) \quad , \tag{A.9}$$

where z represents the "zero ratio" which equals the number of zero-return days in the month, *ZRD*, over the number of days in the month (the sum of trading days, *TD*, and non-trading days, *NTD*); σ represents the standard deviation of the nonzero return; N^{-1} represents the inverse function of the cumulative normal distribution.

C. Trading Quantity

(9) Turnover ratio: (Datar et al., 1998)

Turnover ratio (*AVG_TO*): the 12-month average turnover ratio, which equals the monthly trading volume above the number of shares outstanding at the end of the month.

(10) Trading Volume: (Brennan et al., 1988)

Trading volume (*AVG_TV*): the 12-month average monthly trading volumes.

(11) Standard deviation of turnover ratio: (Chordia et al., 2001)

Standard deviation of turnover ratio (*SD_TO*): the 12-month standard deviation of the monthly turnover ratios.

(12) Standard deviation of trading volume: (Chordia et al., 2001)

Standard deviation of trading volume (*SD_TV*): the 12-month standard deviation of the monthly trading volumes.

(13) Coefficient of variation of turnover ratio: (Chordia et al., 2001)

The coefficient of variation of turnover (*AVG_CVTO*): the ratio of the 12-month standard deviation above the 12-month average monthly turnover ratio.

(14) Coefficient of variation of Trading Volume (*CVTV*): (Chordia et al., 2001)

The coefficient of variation of trading volume (AVG_CVTV): the ratio of the 12-month standard deviation above the 12-month average monthly trading volume.

D. Trading Speed

(15-16) *Zero* (Lesmond et al., 1999): *Zero1* measures transaction costs with the proportion of the number of zero-return trading days in a month to the number of trading days in a month. *Zero2* measures transaction costs with the proportion of the number of trading days with positive trading volume and zero-return in the month to the number of positive trading volume days in the month.

$$(15)\ Zero1 = \frac{\text{The number of zero−return trading days}}{\text{The total number of trading days in the month}}\ , \qquad (A.10)$$

$$(16)\ Zero2 = \frac{\text{Number of positive trading volume and zero−return days}}{\text{Number of positive trading volume days in the month}}\ . \qquad (A.11)$$

(17) *Liu* (Liu, 2006): The *Liu* ratio is the turnover ratio adjusted by zero-trading ratio and is calculated as follows:

$$Liu = \left[\text{ZTRD} + \left({}^{1}/_{(12-\text{month turnover})}\right)/Deflator\right] \times 21 \times \frac{12}{NoTD}\ , \qquad (A.12)$$

where *ZTRD* represents the number of trading days with zero trading volume in a lag of 12 months; the 12-month turnover represents the sum total of the daily turnover ratio, which equals the daily stock trading volume above the closing balance of the number of shares outstanding, for the past 12 months; *NoTD* represents the number of trading days in lag 12 months; and *Deflator* equals to 5,000 in this study.

Appendix B

Estimation of market-aggregate liquidity proxy with APC method

We employ the Asymptotic Principal Components (APC) method in Korajczyk and Sadka (2008) to combine different single-dimensional proxies.

In an approximate factor-model setting for a balanced panel (complete data), Connor and Korajczyk (1986) show that n-consistent estimates (up to a linear transformation) of the latent factors, F^i, are obtained by calculating the eigenvectors corresponding to the k largest eigenvalues of

$$\Omega^{i,u}_{t,\tau} = (L^{i'} L^i)_{t,\tau}/n. \tag{B.1}$$

They refer to these estimates as Asymptotic Principal Components (APC). Note that Ω is a $T\times T$ matrix so that the computational burden of the eigenvector decomposition is independent of the cross-sectional sample size, n. This implies that factor estimates can be obtained for very large cross-sectional samples. Standard approaches to principal components or factor analysis are often unimplementable on large cross-sections because they require eigenvector decompositions of $n\times n$ matrices.

The APC approach applies an alternative estimator of the factor model that accommodates missing data. From Connor and Korajczyk (1986), we estimate each element of Ω by averaging over the observed data. Let L^i be the data for liquidity measure i with missing data replaced by zeros. Define N^i to be an $n\times T$ matrix, for which $N^i_{j;t}$ is equal to one if $L^i_{j;t}$ is observed and is zero if $L^i_{j;t}$ is missing. Define

$$\Omega^{i,u}_{t,\tau} = \left(L^{i'} L^i\right)_{t,\tau}/(N^{i'} N^i)_t. \tag{B.2}$$

In Eq. (B.2), $\Omega^{i,\mu}{}_{t,\tau}$ is the unbalanced panel equivalent of Ω^i, in which the (t, τ) element is defined over the cross-sectional averages over the observed data only. While Ω^i in a balanced panel is guaranteed to be positive semi-definite, $\Omega^{i,\mu}$ is not. In large cross-sections, however, we do not encounter cases in which $\Omega^{i,\mu}$ is not positive definite. The estimates of the (within measure) latent factors, $\bar{F}^i$, are obtained by calculating the eigenvectors for the k largest eigenvalues of $\Omega^{i,\mu}$.

We then estimate the (across measure) common factor(s) (F_t) across all ten measures of liquidity. This is done by stacking the liquidity measures into $L' = [L^{1\prime};L^{2\prime}; \ . \ . \ . \ ; \ L^{8\prime}]$, forming Ω^μ using L, and extracting the eigenvectors of Ω^μ (the cross-sectional measure). Same as Korajczyka and Sadka (2008), we choose the sign so that the factors represent liquidity, rather than illiquidity. Due to the autocorrelation in F_t, we also follow Korajczyka and Sadka (2008) by fitting an AR(2) model for F_t to create the measure of systematic liquidity fluctuation.

Appendix C

Construction of Fama-French five factors, momentum, and the liquidity systematic factor

We follow the factor construction process in Lam et al. (2019) and revise the process based on our sample period, portfolio construction frequency, and so on.

Following previous studies, we use monthly return data on non-financial companies only with appropriate adjustments for capital changes. We employ value-weighted market returns with cash dividends reinvested as the proxy for the market index. For the risk-free rate, we use the 1-month China Central Bank deposit rate in this study.

To avoid the look-ahead bias, accounting data at the quarter-end in calendar month *t-1* are matched to stock returns for the period between month *t* to *t+2*. Firm size is measured by the market capitalization or market value of equity. It is defined as the product of stock prices and the number of shares outstanding at the end of the previous month. The Earnings-to-price ratio (*EP*) is the most recently reported net profit divided by the product of the closing price and the number of outstanding shares at the end of the previous month.

Following Liu et al. (2019), for each month, stocks are assigned to one of two portfolios of *Size* (Small (*S*) and Big (*B*)) based on their firm sizes at the end of month *t-1*. The same stocks are independently sorted into three portfolios (Low (*L*), Medium (*M*), and High (*H*)) based on their *EP*. Six portfolios (*S/L, S/M, S/H, B/L, B/M*, and *B/H*) are then formed at the intersection of Size and *EP* and in the way of having approximately equal numbers of stocks. The value-weighted returns on the six portfolios are calculated monthly.

SMB (small minus big) is the simple average of the returns on the small-stock portfolios minus the returns on the big-stock portfolios:

$$SMB = [(S/L - B/L) + (S/M - B/M) + (S/H - B/H)]/3. \tag{C.1}$$

Similarly, *HML* (high minus low) is the simple average of the returns on the high-*EP* portfolios minus the returns on the low-*EP* portfolios:

$$HML = [(S/H - S/L) + (B/H - B/L)]/2. \tag{C.2}$$

In addition, *CMA* (conservative minus aggressive) is the simple average of the returns on the low-investment portfolios minus the returns on the high-investment portfolios:

$$CMA = [(S/C - S/A) + (B/C - B/A)]/2. \tag{C.3}$$

Furthermore, *RMW* (robust minus weak profitability) is the simple average of the returns on the high-profitability portfolios minus the returns on the low-profitability portfolios:

$$RMW = [(S/R - S/W) + (B/R - B/W)]/2. \tag{C.4}$$

We follow L'her et al.'s (2004) approach to construct the momentum factor. For each month, stocks are ranked by their size and prior performance. The size is based on the *ME* value at the end of the previous month, whereas the prior performance is based on the compounded stock return from month t-12 to month t-2. Excluding the most recent month's (*t-1*) return can attenuate the continuation effect caused by the bid-ask spread. Winners (*W*) are the top 30% of the total stocks with the highest average prior performance. Losers (*L*) are the bottom 30% of the total stocks with the lowest average prior performance. Neutral is the remaining 40% of the stocks. Six portfolios (*S/L, S/M, S/W, B/L, B/M,* and *B/W*) are formed at the intersection of size and prior performance. The value-weighted returns on the six portfolios are calculated each month. *WML* (winner minus loser) is the simple average of the returns on the winner-stock portfolios minus the returns on the loser-stock portfolios:

$$WML = [(S/W - S/L) + (B/W - B/L)]/2. \tag{C.5}$$

The construction of *LIQ* is similar to *SMB, HML, WML, RMW*, and *CMA*. First, all stocks are divided into two portfolios, small (*S*) and big (*B*), based on their ranking of firm sizes in the previous month. Second, we divide all the stocks into three portfolios, bottom (*LB*), middle (*LM*), and top (*LT*), according to the ranking of *Liqbetas* in the previous month. These three *Liqbeta* portfolios contain the bottom (with low *Liqbeta*) 30%, the middle 40%, and the top (high *Liqbeta*) 30% stocks, respectively. We take the intersection of the firm size and *Liqbeta* portfolios so that we get six portfolios (*SLB*, *SLM*, *SLT*, *BLB*, *BLM*, and *BLT*) and calculate the portfolio returns in each month. The LIQ factor equals the mean difference between the two low *Liqbeta* portfolios (*SLT* + *BLT*) and the two high *Liqbeta* portfolios (*SLB* + *BLB*), and the equation is as follows:

$$LIQ = [(SLT + BLT) - (SLB + BLB)]/2. \tag{C.6}$$