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

Some studies find a significant return- liquidity relation, while others do not. This is due to the use of single dimensional liquidity prices (transaction costs, price impacts, trading seeds, trading quantities). Hence, the need for muti dimensional Liquidity composite liquidity proxy. The study combines multiple single dimensional proxies using either APC (Asymptotic Principal Component) approach or PCA (Principal Component Analysis).

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. Starts with **17** individual liquidity proxies, and select **6** of them to construct **126** competing composite liquidity proxies.

2 tests are performed:

1. Characteristic level horserace
2. Factor level regressions

Then, a pairwise comparison among the winning composite proxies selected from characteristic level horserace.

RESULT: Anihud HL-FHT proxy constructed by APC method stands out as the optimal proxy for explaining stick 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.

CONTRIBUTION:

1. Proposes a robust method to construct the optimal composite liquidity proxy, improving precision in liquidity measurement and market efficiency analysis.
2. Demonstrates that the APC method outperforms PCA in aggregating multiple liquidity proxies into a comprehensive market measure.
3. Confirms a significant liquidity–return relation in the Chinese stock market, emphasizing the need for accurate multidimensional liquidity measurement.

LITERATURE REVIEW

1. Bid -Ask Spread Proxies
   Roll (1984): Roll ratio (effective bid-ask spread) → negative relation with firm size in US stocks. (Captured transaction costs)
   Corwin & Schultz (2012): HL ratio (based on daily high–low prices) → related to variance of true stock values.
2. Trading Quantity Proxies

   Brennan et al. (1988): Trading volume measure.

   Datar et al. (1998): Turnover ratio → significant return-liquidity relation in US stocks.

   Chordia et al. (2001): Additional measures using turnover ratios & trading volumes.

3. Transaction Speed Proxies

   Lesmond et al. (1999): Zero1 (proportion of zero-return days) & Zero2 (zero-return days with positive volume) → significant return-liquidity relation in US stocks.

Bekaert et al. (2007): Used Zero ratios → evidence across 19 emerging markets.

Fong et al. (2017): FHT ratio (modified Zero ratios) → estimates effective transaction cost.

Liu (2006): Liu ratio (trading speed & price impact) → significant return-liquidity relation in US stocks.

4. Price Impact Proxies

   Amihud (2002): Amihud ratio (illiquidity measure) → significant return-illiquidity relation.

   Pastor & Stambaugh (2003): Pastor ratio (price reversals) → return-illiquidity evidence. (Measure the price impact decision)

   Goyenko et al. (2009), Brennan et al. (2013), Kang & Zhang (2014): Modifications of Amihud ratio:

   - Amivest ratio → non-zero return version.
   - BHS ratio → turnover-based version.
   - Kang ratio → zero-trading volume version.
   - All document significant liquidity-return relations in US stocks.

DATA SAMPLE

Data taken from the Chinese stock market(2007-2023).

Proxies Used:

PRICE IMPACT (5 proxies) - Measures how much stock prices move when trades occur

- Pastor - Price reversal effects from signed trade flows
- Amihud - Mean ratio of |daily return| ÷ daily trading amount
- Amivest - "Non-zero returns" version of Amihud ratio
- BHS - "Turnover" version of Amihud ratio
- Kang - "Non-zero volume" version of Amihud ratio

TRADING COST (3 proxies) - Measures transaction costs like bid-ask spreads

- HL - Trading spread from daily high/low prices
- Roll - Effective spread from consecutive price changes
- FHT - Trading cost from zero return days + return volatility

TRADING QUANTITY (6 proxies) - Measures volume and turnover activity levels

- AVG_TO - 12-month average turnover ratios
- AVG_TV - 12-month average trading volumes
- SD_TO - Standard deviation of turnover ratios
- SD_TV - Standard deviation of trading volumes

- AVG_CVTO - Coefficient of variation of turnover ratios
- AVG_CVTV - Coefficient of variation of trading volumes

TRADING SPEED (3 proxies) - Measures how quickly trades can be executed

- Zero1 - Proportion of zero-return trading days
- Zero2 - Proportion of positive volume + zero-return days
- Liu - Turnover ratio adjusted by zero-trading days

CONSTRUCTION OF COMPOSITE LIQUIDITY PROXIES

Correlation matrices were constructed. These were the 6 selected:

Price Impact Dimension

1. Amihud >> BHS, Kang, Amivest as one of the most widely used and later ones are significantly correlated.
2. Pastor also selected as its price reversal measure has a low correlation with Amihud.

Trading Cost Dimension

3. HL
4. Roll
5. FHT

Trading Speed Dimension

6. Liu >> 2 zero ratios as it has low corelation with other 16 proxies

PCA method:

- Suitable for datasets where the number of observations is much larger than the number of variables. Does not work in high dimensional settings.
- First extracts the stock-level across-measure principal component and then aggregates across stocks to get the market-level proxy.

APC method:

- Works well in high dimensional settings.
- Treats the stock-level proxies as variables and the time series as observations, thus directly estimating the market-level across-measure composite liquidity proxy

Step 1: Generate combinations of liquidity proxies

- Start with six individual liquidity proxies.
- Randomly select different numbers of them (1 to 6) to form combinations.
- This yields 63 combinations

Step 2: Construct aggregate proxies using two methods

- Apply APC method and PCA method to each of the 63 combinations.
- This results in 126 aggregate liquidity proxies (63 from APC, 63 from PCA).
- Using sign analysis, define each combined proxy as an illiquidity measure (since most individual proxies represent illiquidity).

Step 3: Adjust for persistence in illiquidity

- Perform AR(2) regressions (following Pastor & Stambaugh, 2003) on each aggregate measure.
- Use the residuals from these regressions to capture market illiquidity fluctuations.

Step 4: Estimate stock-level sensitivity to market illiquidity

- Run 12-month rolling regressions of stock excess returns on market illiquidity fluctuations.
- The absolute regression coefficient = stock's conditional return sensitivity to liquidity risk (Acharya & Pedersen, 2005).
- Higher value → lower systematic liquidity risk.
- Define these rolling coefficients as Liqbeta (composite liquidity proxy).
    - From APC method → Liqbeta_APC.
    - From PCA method → Liqbeta_PCA.

Step 5: Compare the performance of proxies

1) Use 126 stock-level composite liquidity proxies.
2) Conduct two types of comparison:
    a) Characteristic-level horserace.
    b) Pricing factor-level pairwise comparison.
3) Purpose: identify the optimal composite liquidity proxy in the Chinese stock market.

CHARACTERISTIC LEVEL HORSERACE TEST

Fama Macbeth regressions- univariate and multivariate.

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.

8 firm characteristics introduced as **control variables:**

1. Asset-to-Market Ratio (AM):

$$AM = \ln\left(\frac{Total\ Assets}{Market\ Capitalization}\right)$$

where $Market\ Capitalization = Closing\ Price \times Shares\ Outstanding$

2. Book-to-Market Ratio (BM):

$$BM = \ln\left(\frac{Book\ Equity}{Market\ Capitalization}\right)$$

3. Cash-to-Price Ratio (CP):

$$CP = \frac{\Delta Cash\ \&\ Equivalents}{Market\ Capitalization}$$

where $\Delta Cash\ \&\ Equivalents = Cash_t - Cash_{t-1}$

4. Earning-to-Price Ratio (EP):

$$EP = \frac{Net\ Profit}{Market\ Capitalization}$$

5. Market Size (Size):

$$Size = \ln(Market\ Capitalization)$$

6. Return-on-Equity Ratio (ROE):

$$ROE = \frac{Net\ Earnings}{Book\ Value\ of\ Equity}$$

7. Asset-Growth Ratio (AG):

$$AG = \frac{Total\ Assets_t - Total\ Assets_{t-1}}{Total\ Assets_{t-1}}$$

8. Market Beta (MktBeta):

$$MktBeta = \hat{\beta} \quad \text{from regression } R_{i,t} - R_{f,t} = \alpha + \beta(R_{m,t} - R_{f,t}) + \epsilon_t$$

where regression is estimated using **36-month rolling window.**

Dummy Variables:

Earnings-to-Price Dummy (D_{EP})

$$EP = \begin{cases} \frac{Net\ Profit}{Market\ Capitalization}, & \text{if } Net\ Profit > 0 \\ 0, & \text{if } Net\ Profit \leq 0 \end{cases}$$

$$D_{EP} = \begin{cases} 0, & \text{if } Net\ Profit > 0 \\ 1, & \text{if } Net\ Profit \leq 0 \end{cases}$$

Cash-to-Price Dummy (D_{CP})

$$CP = \begin{cases} \frac{\Delta Cash}{Market\ Capitalization}, & \text{if } \Delta Cash > 0 \\ 0, & \text{if } \Delta Cash \leq 0 \end{cases}$$

$$D_{CP} = \begin{cases} 0, & \text{if } \Delta Cash > 0 \\ 1, & \text{if } \Delta Cash \leq 0 \end{cases}$$

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, βt. The mean of these regression coefficients βt is calculated, which represents the pricing effect of stock-level liquidity risk on stock returns.

PRICING FACTOR-LEVEL COMPARISONS

A systematic Liquidity Factor is constructed (LIQ) . LIQ is incorporated in CAPMs to enhance their return explanator power on the returns of various testing portfolios. Next, factor level Fama-Macbeth regressions are preformed and pairwise comparison to identify the optimal composite liquidty proxy.

FACTOR- LEVEL FAMA-MACBETH CROSS SETIONAL REGRESSIONS

Step 1: Construct systematic liquidity factor (LIQ)

- Use Liqbeta_APC or Liqbeta_PCA (from the characteristic-level horserace).
- Form portfolios sorted by liquidity.
- Calculate the return spread between most illiquid and most liquid portfolios → this is the LIQ factor.

Step 2: Augment asset pricing models

- Add the LIQ factor to well-known models:
  - CAPM → becomes L2F (2-factor model).
  - FF3F → becomes L4F (4-factor model).
  - FF4F → becomes L5F (5-factor model).
  - FF5F → becomes L6F (6-factor model).
  - FF6F → becomes L7F (7-factor model).

Step 3: Ensure robustness with multiple testing portfolios
Form nine sets of testing portfolios:

1. 1 set: Decile portfolios sorted by Liqbeta.
2. 4 sets: 25 portfolios each, based on 5×5 double-sorting:
   - Liqbeta–Size
   - Liqbeta–EP
   - Liqbeta–Beta
   - Liqbeta–ROE
3. 4 sets: 27 portfolios each, based on 3×3×3 triple-sorting:
   - Liqbeta–Size–Beta
   - Liqbeta–Size–EP
   - Liqbeta–Size–ROE
   - Liqbeta–EP–ROE

Step 4: Calculate portfolio returns

- Compute value-weighted average excess returns (portfolio return – risk-free rate).
- These serve as the dependent variables in factor regressions.

2 stages in Factor- level Fama-Macbeth cross sectional regressions

- 36 month rolling regression for each portfolio p:

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

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.

- Cross-sectional regression of portfolio excess returns on the factor loadings estimated from the first stage.

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

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.

PAIRWISE COMPARISON

We want to check which liquidity proxy (used to build the LIQ factor) gives the best performing model.
Instead of time-series metrics like GRS, adj. R^2, and alpha ratios, we design new criteria for cross-sectional regressions.

Appraisal Metrics Used

1. N_LIQ → Number of significant LIQ coefficients
2. N_A → Number of insignificant intercepts (good if more are insignificant → means model captures risk better)
3. MDR → Mean Incremental Adjusted R^2 (how much adj. R^2 improves after adding LIQ compared to base model, e.g., L6F – FF5F)

Procedures to compare performance of liquidity models

1. Factor-level Fama-MacBeth regressions
    - Run regressions using models with LIQ factors built from different composite liquidity proxies.
    - Compute appraisal metrics (N_LIQ, N_A, MDR).
2. Round-robin matches
    - Each composite liquidity proxy is paired against all others.
    - Winner is decided based on *one* of the appraisal metrics.
3. Winning probabilities
    - For each proxy, calculate winning probability under each metric:
        - PROBL (wins on N_LIQ)
        - PROBA (wins on N_A)

        - PROBR (wins on MDR)
    - Overall mean winning probability (MPB) = average of the three.
4. Statistical tests
    - Shapiro-Wilk → check if MPB is normally distributed
    - Sign test → check if MPB is significantly > 0
    - Wilcoxon rank-sum → check if best proxy's MPB is significantly better than others
    - Bootstrapping (per Cooper & Miao, 2019)
        - Resample 30% MPB with replacement, repeat 1,000 times
        - Compute simulated MPB (MPB_SIMU) and subsample means (MPB_SUB)
        - Test if optimal proxy's MPB is significantly higher.
5. Additional test (if >1 winner remains)
    - Use Win-or-Loss (WOL):
        - If a proxy wins in at least 2/3 metrics vs another proxy → WOL = 1, else 0
    - Compute average WOL across all comparisons → gives MPBM
    - Test if optimal proxy has significantly higher MPBM than others.

RESULTS

Characteristic-level horserace tests

- Setup
    - Regressions: univariate + multivariate cross-sectional regressions of excess stock returns.
    - Independent variables: firm characteristics + composite liquidity proxy.
    - 126 composite proxies tested:
        - 63 Liqbeta_APC
        - 63 Liqbeta_PCA
- Threshold
    - Significance set at 10% (instead of 5%).
    - Rationale: allows more winner proxies → increases sample size for later pairwise tests.
- Results
    - Panel A (APC method)
        - 18/63 (28.6%) Liqbeta_APCs significant in univariate regressions.
        - 15/63 (23.8%) remain significant in both univariate & multivariate regressions.
    - Panel B (PCA method)
        - Only 3 Liqbeta_PCAs significant in multivariate regressions.
        - None significant in both univariate & multivariate regressions.
- Interpretation
    - APC method > PCA method.
        - APC combines proxies while considering inter-correlations across stocks.
        - PCA aggregates via first components (averaging) → may lose interaction information.
- Outcome
    - 15 Liqbeta_APCs emerge as winners.
    - These will be used to construct systematic liquidity factor (LIQ) for next steps.

Factor-level Fama-MacBeth regressions

- Construction
    - LIQ factor built from each of the 15 winning Liqbeta_APCs.
    - Added to CAPM + FF-series models.
    - Value-weighted portfolio excess returns regressed on LIQ + other pricing factors.
- Evaluation Metrics (from Fama-MacBeth regressions)
    1. N_LIQ → number of significant LIQ coefficients.
    2. N_A → number of insignificant intercepts (better = fewer unexplained returns).
    3. MDR → mean incremental adjusted R^2 (gain from adding LIQ factor).
- Ranking Method
    - Composite proxies ranked by MPB = mean winning probability across metrics.
    - MPB = average of PROBL (LIQ wins), PROBA (intercept wins), PROBR (R^2 wins).
- Key Results
    - Amihud-HL-FHT proxy ranked 1st:
        - MPB = 0.9048 → beats ~13.6 of 14 proxies on avg.
        - N_LIQ: 10 sig. premiums → joint 1st (PROBL = 0.9286).
        - N_A: 4 insignificant intercepts → joint 3rd (PROBA = 0.7857).
        - MDR = 0.0459 → 1st (PROBR = 1.0000).
    - 2nd best: Amihud-HL-Pastor-FHT.
    - Other findings:
        - Some proxies strong on one metric but weak on others → lowers overall MPB.
        - Ex: Amihud-HL-FHT-Pastor-Liu → top on N_A but weaker on others.
- Statistical Tests
    - Shapiro-Wilk: MPB not normally distributed (all $p < 5\%$).
    - Sign Test: MPBs ≠ 0 (all significant).
    - Wilcoxon rank-sum + second Sign test:
        - Differences between Amihud-HL-FHT and other top 5 proxies not significant.
    - Bootstrapping (1000 resamples, 30% MPB each)
        - Simulated MPB (MPB_SIMU) ≈ actual MPB → robustness confirmed.
        - For top 6 proxies → no significant difference in MPB.
- Conclusion
    - Top 6 proxies selected as candidates for optimal composite liquidity proxy.

Pairwise comparisons with alternative winning probability

- Alternative measure: MPBM = probability of winning a best-of-three match across 3 metrics.
- Results:
    - Amihud-HL-FHT: MPBM = 1.0000 → always wins vs all others.
    - Others ranked lower, MPBM decreasing from 0.9286 → 0.0000.
- Statistical Tests (MPBM)
    - Shapiro-Wilk + Sign test: MPBM ≠ 0 and not normal.
    - Wilcoxon + 2nd Sign test: Amihud-HL-FHT significantly > all others ($p < 5\%$).
    - Bootstrapping confirms robustness: MPBM_SIMU = 1.000 for Amihud-HL-FHT.
- Conclusion

- Amihud-HL-FHT stands out as the optimal composite liquidity proxy.

Why Amihud-HL-FHT stands out

- Time-series evidence
    - Amihud, HL, FHT proxies are strong on their own.
    - Capture market stress (2008 GFC, 2015 crash, COVID-19 crisis).
    - Amihud → price impact (broad variations).
    - HL & FHT → trading costs (oscillations).
    - Together → multi-dimensional & multi-frequency coverage of liquidity.
- Comparison with other proxies
    - Pastor proxy → trend similar to Amihud → overlapping info.
    - Roll proxy → mirrors HL.
    - Liu proxy → captures trading quantity & speed.
        - Shows general liquidity improvements (except 2014–16 crash).
        - But smaller variation → weaker explanatory power.
- Regression checks
    - Optimal Amihud-HL-FHT regressed on 6 individual proxies.
    - Significant relations (5% level) with Amihud, FHT, HL, Liu, Roll (both univariate & multivariate).
    - Confirms "good ingredients → good results" logic.