DATA COLLECTION & PREPROCESSING

DATA SOURCE:

- NSE 500 constituent stocks
- Time Period: 2008-01-02 to 2024-12-31
- Total Years Covered: 17.0 years

DATASET STATISTICS:

- Raw Daily Observations: 1,513,231
- Stocks Successfully Processed: 570
- Final Monthly Dataset: 73,138 stock-month observations
- Average Trading Days per Month: 20.7

QUALITY CONTROLS APPLIED:

- Minimum 15 trading days per month requirement
- Removed zero volume trading days
- Minimum 5 years of data per stock
- Comprehensive missing value treatment

INDIVIDUAL LIQUIDITY PROXIES CALCULATED

1. Amihud Ratio (2002)

   Dimension: Price Impact

   Description: Measures price movement per unit of trading volume

   Completion Rate: 100.0% (73,138 observations)

   Mean Value: 0.052094

2. High-Low Spread (Corwin-Schultz 2012)

   Dimension: Trading Cost

   Description: Bid-ask spread estimate using daily high-low prices

   Completion Rate: 100.0% (73,138 observations)

Mean Value: 0.019449

3. Roll Spread (1984)

Dimension: Trading Cost

Description: Effective bid-ask spread using price change covariance

Completion Rate: 100.0% (73,138 observations)

Mean Value: 10.344931

4. Zero Trading Ratio (Lesmond et al. 1999)

Dimension: Trading Speed

Description: Proportion of days with zero returns

Completion Rate: 100.0% (73,138 observations)

Mean Value: 0.011917

5. FHT Measure (Fong et al. 2017)

Dimension: Trading Cost

Description: Trading cost using zero returns and volatility

Completion Rate: 100.0% (73,138 observations)

Mean Value: 0.000453

6. Average Trading Volume (Standardized)

Dimension: Trading Quantity

Description: Cross-sectionally standardized average volume

Completion Rate: 100.0% (73,138 observations)

Mean Value: 0.000000

7. Volume Volatility (Standardized)

Dimension: Trading Quantity

Description: Cross-sectionally standardized volume variability

Completion Rate: 100.0% (73,138 observations)

Mean Value: -0.000000

8. Volume Coefficient of Variation

Dimension: Trading Quantity

Description: Volume variability relative to average

Completion Rate: 100.0% (73,138 observations)

Mean Value: 0.841735

DESCRIPTIVE STATISTICS

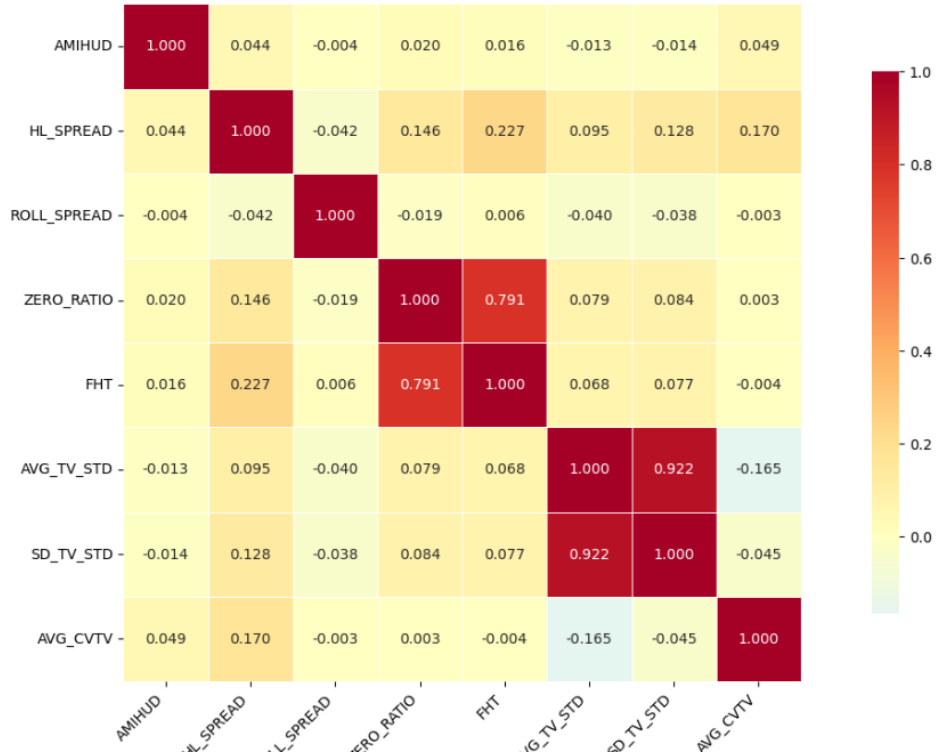
SUMMARY STATISTICS FOR ALL PROXIES:

	AMIHUD	HL_SPREAD	ROLL_SPREAD	ZERO_RATIO	FHT	AVG_TV_STD	SD_TV_STD	AVG_CVTV
count	73138.000000	73138.000000	73138.000000	73138.000000	73138.000000	73138.000000	73138.000000	73138.000000
mean	0.052094	0.019449	10.344931	0.011917	0.000453	0.000000	-0.000000	0.841735
std	1.564168	0.007911	56.826356	0.039545	0.002634	0.998611	0.998611	0.473356
min	0.000000	0.000000	0.000000	0.000000	0.000000	-0.585734	-0.599914	0.067845
25%	0.000025	0.014094	0.000000	0.000000	0.000000	-0.355655	-0.335078	0.520176
50%	0.000117	0.017793	0.933200	0.000000	0.000000	-0.239313	-0.218200	0.716979
75%	0.000728	0.022868	6.717124	0.000000	0.000000	-0.073758	-0.060927	1.023514
max	188.903144	0.134038	3540.603671	0.761905	0.217931	21.055096	21.330891	4.522060

CORRELATION ANALYSIS RESULTS

CORRELATION MATRIX:

(Values range from -1 to +1; closer to ±1 indicates stronger relationship)



HIGH CORRELATION PAIRS (>0.70):

• ZERO_RATIO ↔ FHT: 0.791 (Positive)

• AVG_TV_STD ↔ SD_TV_STD: 0.922 (Positive)

PROXY SELECTION FOR COMPOSITE INDEX

SELECTION CRITERIA:

• Include all four liquidity dimensions (Impact, Cost, Speed, Quantity)

• Avoid highly correlated proxies (correlation >70%)

• Prioritize most cited academic measures

• Ensure sufficient diversity for robust composite

FINAL SELECTED PROXIES (6 total):

1. AMIHUD (Price Impact)

2. HL_SPREAD (Trading Cost)

3. ROLL_SPREAD (Trading Cost)

4. ZERO_RATIO (Trading Speed)

5. FHT (Trading Cost)

6. AVG_TV_STD (Trading Quantity)

COMBINATION POSSIBILITIES:

• Selected proxies will generate 63 possible combinations

• Using both PCA and APC methods: 126 total combinations

• All combinations will be tested in horserace analysis

CURRENT PROJECT STATUS

COMPLETED PHASES:

✅ Phase 1: Data Collection & Preprocessing

✅ Phase 2: Individual Liquidity Proxy Calculation

✅ Phase 3: Data Quality Validation

✅ Phase 4: Correlation Analysis & Proxy Selection

NEXT PHASES:

🔄 Phase 5: Composite Index Construction (PCA & APC methods)

⏳ Phase 6: Horserace Testing & Statistical Validation

⏳ Phase 7: Optimal Index Selection

⏳ Phase 8: Indian Market Validation & Crisis Testing

⏳ Phase 9: Research Paper Documentation