Replicating Anomalies

Kewei Hou, The Ohio State University
Chen Xue, University of Cincinnati
Lu Zhang, The Ohio State University and NBER

Executive Summary

Kewei Hou, Chen Xue, and Lu Zhang replicate the published anomalies literature in finance and accounting with a data library of 447 anomaly variables. To control for microcaps — stocks smaller than the 20th percentile of market equity of NYSE stocks) — the researchers form testing deciles with NYSE breakpoints and value-weighted returns. An anomaly is treated as a replication success if the average return of its high-minus-low decile is significant at the 5 percent level ($t \geq 1.96$).

Their key finding is that most anomalies fail to replicate. Out of 447 anomalies, 286 (64 percent) fail to replicate at the 5 percent level. Imposing the cutoff t-value of 3 per Harvey, Liu, and Zhu (2016) raises the number of replication failures further to 380 (8 percent).

The biggest casualty is the liquidity literature. In the trading frictions category containing mostly liquidity variables, 95 out of 102 variables (93 percent) fail to replicate. Prominent variables include short-term reversal, share turnover, the coefficient of variation for dollar trading volume, absolute return-to-volume, liquidity betas, idiosyncratic, total, and systematic volatilities, zero trading days, and bid-ask spread.

The distress anomaly is virtually nonexistent. The failure probability, O-score, Z-score, and credit rating produce mostly insignificant average return spreads.

Other influential anomaly variables that fail to replicate include debt-to-market, five-year sales growth, the dispersion in analysts' forecast, the G-index of corporate governance, earnings attributes (persistence, smoothness, value relevance, and conservatism), accrual quality, total accruals, and operating profitability.

Even for replicated anomalies, magnitudes are much lower than originally reported. Famous examples include price momentum, earnings momentum formed on standardized unexpected earnings, abnormal returns around earnings announcements, and revisions in analysts' earnings forecasts, cash flow-to-price, operating accruals, customer momentum, and asset growth.

Why does the replication differ so much from original studies? The key word is microcaps. At the end of 2014, microcaps represented only 1.4 percent of the total market cap of the U.S. market, but accounted for 60 percent of the number of stocks. Microcaps have not only the highest equal-weighted returns, but also the largest cross-sectional dispersions in returns and anomaly variables. Many studies overweight microcaps with equal-weighted returns in portfolio sorts, and often together with NYSE-Amex-NASDAQ breakpoints. Hundreds of studies use cross-sectional regressions, assigning even higher

weights to microcaps than equal-weights in sorts. With a linear functional form, regressions are particularly susceptible to outliers, which most likely are microcaps. Due to high costs in trading these stocks, anomalies in microcaps are more apparent than real. With tiny market equity, the economic importance of microcaps is small.

The low replication rate of 36 percent is not due to the extended sample. Repeating their replication in the shorter samples in original studies, the researchers find that 293 anomalies (66 percent) fail to replicate at the 5 percent level. Imposing the t-cutoff of 3 raises the number to 387 (86.6 percent).

Overall, the results indicate that capital markets are more efficient than previously reported.