Summary: What explains the dynamics of 100 anomalies?

The behavioral finance view on the existence of asset pricing anomalies in the cross-section of expected equity returns is based on two building blocks (e.g. Barberis and Thaler (2003)): investor psychology, which allows mispricings to arise, and limits to arbitrage, which prevent sophisticated market participants from quickly exploiting these inefficiencies. A testable prediction of this theoretically deduced mechanism is that abnormal returns should ceteris paribus be stronger in settings where investor sentiment is high or where arbitrageurs are less capable of aggressively betting against irrationality-induced mispricings. Empirical tests of this fundamental relationship might help academics to enrich or challenge our understanding of the price discovery process and offer practitioners insights into ways to optimize their investment process.

What separates this paper from previous work is the breadth of anomalies taken into account, the focus on time-series (as opposed to cross-sectional) variation in anomaly returns, and the joint analysis of proxies for limits to arbitrage and investor sentiment.

We start by synthesizing information from a very broad range of potential inefficiencies. We identify, categorize, and replicate 100 well-known or recently discovered anomalies related to violations of the law of one price, momentum, technical analysis, short-term reversal, long-term reversal, calendar effects, lead-lag effects among economically linked firms, pairs trading, beta, financial distress, skewness, differences of opinion, industry effects, fundamental analysis, net stock and financing decisions, capital investment and firm growth, innovation, accruals, dividend payments, or earnings surprises. We believe that the resulting data set of more than 65,500 anomaly months covers a reasonably representative universe of anomalies discussed in the literature.

Considering all these anomalies simultaneously in a unified framework offers a number of advantages. Most asset pricing studies concentrate on only one or few anomalies, and methodological or other differences can have a massive impact on inferences (e.g. Fama and French (2008)), making comparisons difficult. In his literature review of predictors of cross-sectional stock returns, Subrahmanyam (2010) thus concludes that the “picture remains murky and suggests a need for clarifying studies” (p. 28). Similarly, Richardson et al. (2010) criticize the “haphazard nature” of this line of research and argue that “to date very few papers have made a serious attempt to bring some structure to the anomaly literature” (p. 422). Our approach aims at progressing on this front.

Our main insights can be summarized as follows. First, from an unconditional perspective, most anomalies produce economically large abnormal returns relative to a Fama and French (1993) model. As a rough estimate, and averaged across time and anomalies, abnormal monthly returns are about 70 to 80 basis points (bp). This is noteworthy as,
compared to many original studies, our data screens are often stricter. Moreover, our sample period is often longer or more recent, and thus partly out-of-sample. This suggests that most anomalous returns uncovered in the literature are unlikely to be primarily driven by statistical biases (see also McLean and Pontiff (2013) and Green et al. (2013)).

Second, investor sentiment as mainly proxied for by the Baker and Wurgler (2006) market-level index is a strong and robust predictor of anomaly returns. These findings complement the insights of Stambaugh et al. (2012) who uncover that the eleven anomalies they consider tend to be more pronounced following high levels of sentiment. In a follow-up study and based on the same set of anomalies, Stambaugh et al. (2014) run simulations to mitigate concerns regarding a spurious-regression bias. Again, they find strong support for the predictive power of sentiment, and argue that "the key is consistency across anomalies" (p. 1). Our approach of substantially increasing the set of anomalies (as well as sentiment proxies) represents a natural extension of their study.

Indeed, for the average anomaly, we find that the long-short spread is roughly 50% larger following months with above median Baker and Wurgler (2006) sentiment than it is following months with below median sentiment. Even though we focus on relatively large and liquid firms for which sentiment is expected to be less relevant (see e.g. Lemmon and Portniaguina (2006)), close to 85% of the anomalies load positively on lagged sentiment. In line with the economic intuition, these results are strongest among return phenomena often attributed to investor overreaction.

Third, we find little evidence that the time variation in proxies for market-wide limits to arbitrage has predictive power for the dynamics of anomaly returns. Building on a literature review, our baseline analysis considers the Vix, average idiosyncratic volatility, the Ted spread, the Moody’s credit spread, average bid-ask spreads, and market illiquidity. These variables have a solid theoretical foundation, capture different aspects of limits to arbitrage (e.g. funding liquidity, transaction costs, holdings costs), and are widely employed in the literature. In general, these variables turn out to have a low correlation (0.0-0.2) with the Baker and Wurgler (2006) sentiment measure, and thus make quite distinct predictions.

We run regressions analogously to the ones for investor sentiment. We indeed find that the few relatively unambiguous deviations from the law of one price exhibit a strong positive link to proxies for time-varying limits of arbitrage. To a lesser extent, these insights also hold for short-term reversal, pairs trading, and net stock and financing anomalies. However, and with the notable exception of idiosyncratic risk in some specifications, the proxies turn out to be at best loosely related to the large time-variation of most other anomalies. In fact, anomaly returns only load sporadically on market-wide arbitrage risk factors in a statistically and economically significant matter in the predicted direction.
The overall relatively low predictive power of most proxies for the magnitude of most anomalies is persistent. Among others, we run predictive and contemporaneous regressions, use the raw level of the proxies or a more reduced form, use changes instead of levels, consider quarterly instead of monthly returns, rely on alternative proxies for arbitrage constraints, consider composite anomalies, include or even focus on small firms, and control for outliers. Inferences do not change. Nevertheless, individually, the proxies might be too noisy. Among others, we thus use principal component analysis to capture as much of the joint variation across the different proxies as possible. This approach mirrors the construction of the Baker and Wurgler (2006) index, which is based on the principal component of six market-wide sentiment proxies such as aggregate turnover or the number of IPOs. Again, inferences do not change: all else equal, time-variation in aggregate sentiment (limits to arbitrage) goes along with or predicts a large (only small) fraction of the variation in anomaly returns.

Fourth, in an attempt to understand the low predictive power of the limits to arbitrage proxies better, we uncover that market-level arbitrage conditions lack explanatory power for time-varying anomaly-level arbitrage popularity. In deriving this novel finding, we draw on recent studies which argue that unobservable changes in arbitrage activity are likely to manifest themselves in observable changes in the behavior or characteristics of stocks which a specific quantitative trading strategy would typically speculate on.

These findings do not imply that market-wide limits to arbitrage do not matter. The fact that investor sentiment has such strong predictive power for anomaly returns is highly suggestive of a certain level of ever-present limits to arbitrage. This is also the line of reasoning behind Stambaugh et al. (2012). They focus on variation in investor sentiment only and implicitly assume potentially time stable limits to arbitrage, particularly for the short leg of the portfolio. For instance, such constraints stem from the fact that many institutional investors are not allowed to go short (Almazan et al. (2004)) or face other frictions (e.g. Baker et al. (2011), Gromb and Vayanos (2010)). What our findings suggest is that the dynamics around these persistent constraints appear to often have a rather weak impact, at least for some forms of market-wide limits to arbitrage measured at a monthly frequency. In other words, the ever-present level of aggregate arbitrage constraints might in many settings play a more important role in explaining potential inefficiencies than its fluctuations over time. The apparently weak link between market-level conditions and anomaly-level popularity lends further support for this line of reasoning.

In contrast, the dynamics of investor sentiment appear to be quite important relative to the mean level of deviations from full rationality. Alternatively, given the many facets of limits to arbitrage, measures of investor sentiment themselves might represent some form of limits to arbitrage in that they could proxy for noise trader risk (e.g. Shleifer and Vishny (1997), Barberis and Thaler (2003)) or other risk factors (Sibley et al. (2013)).
References


Sibley, S. E., Y. Xing, and X. Zhang, 2013, “Is sentiment sentimental?,” Unpublished working paper, Purdue University, Rice University.

