Journal of Banking & Finance 57 (2015) 65–85

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

What explains the dynamics of 100 anomalies?

Heiko Jacobs *

Finance Department, University of Mannheim, L5, 2, 68131 Mannheim, Germany

ARTICLE INFO

Article history:
Received 7 October 2014
Accepted 21 March 2015
Available online 11 April 2015

JEL classification:
G12
G14

Keywords:
Anomalies
Limits to arbitrage
Sentiment
Return predictability
Behavioral finance

ABSTRACT

Are anomalies strongest when investor sentiment or limits of arbitrage are considered to be greatest? We empirically explore these theoretically deducted predictions. We first identify, categorize, and replicate 100 long-short anomalies in the cross-section of expected equity returns. We then comprehensively study their interaction with popular proxies for time-varying market-level sentiment and arbitrage conditions. We find a powerful (relatively weak) role of the variation in proxies for sentiment (arbitrage constraints). In this context, the predictive power of sentiment is mostly restricted to the short leg of strategy returns. Our insights collectively suggest that the dynamics of sentiment combined with the base level (and not primarily the variations) of limits to arbitrage provide at least a partial explanation for inefficiencies.

1. Introduction

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 deducted mechanism is that abnormal returns should ceteris paribus be stronger in settings where many investors behave irrationally or where arbitrageurs are less capable of aggressively betting against mispricings (see e.g. the discussions in Baker and Wurgler, 2007; Brav et al., 2010; or Hanson and Sunderan, 2014). 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.

However, the empirical evidence is in fact far from conclusive. We aim to revisit this controversial debate. What separates this paper from previous work is the breadth of anomalies taken into account as well as the focus on time-series (as opposed to cross-sectional) variation in market-level (as opposed to anomaly-level or stock-level) arbitrage constraints. This approach enables us to yield some novel insights into the following questions: When considered jointly and based on the same stock universe and the same methodology, which type of phenomena yields the highest seemingly abnormal returns in which situations? Judging from the "big picture", to what extent can variations in market-wide sentiment on the one hand and variations in market-wide limits to arbitrage on the other hand be deemed to be good explanations for the dynamics of anomaly returns?

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 and 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.

Of course, the 100 anomalies are not fully independent. For instance, our data set contains many "enhanced" momentum strategies proposed in the literature, which are different from, but still closely related to the approach in the seminal study of Jegadeesh and Titman (1993). Nevertheless, the average correlation of the Fama and French (1993) model adjusted equally weighted

* Tel.: +49 621 181 3453; fax: +49 621 181 1534.
E-mail address: jacobs@bank.bwl.uni-mannheim.de

http://dx.doi.org/10.1016/j.jbankfin.2015.03.006
0378-4266/

100 anomaly returns is only .12, suggesting that we are able to capture a diverse set of return phenomena. This insight is consistent with Green et al. (2014) who uncover that the cross-section of expected returns is surprisingly multidimensional. Their findings are derived from 100 return predictive signals as well, which however only partly overlap with our anomaly data base.

Forming a common basis for all anomalies 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 large-scale analysis is motivated by the lack of comparability, consensus, or even existence of previous work regarding the impact of investor sentiment and particularly of limits to arbitrage on individual anomalies, as we outline in the literature review below.

For instance, a critical issue in this context appears to be the treatment of micro caps and small caps. As Fama and French (2008) highlight: "From a general economic perspective, it is important to know whether anomalous patters in returns are marketwide or limited to illiquid stocks that represent a small portion of market wealth" (p. 1655). Importantly, small stocks have been argued to obstruct the view on the economic importance of arbitrage constraints (e.g. Brav et al., 2010). In light of these concerns, our baseline analysis applies the same filter rules on size and liquidity as e.g. Jegadeesh and Titman (2001). This results in excluding about 50% of the firm months of common stocks in the CRSP database, which however account for a maximum of a few percent of the total market capitalization. Our approach thus enables us to rely on a stock universe which is comparable across anomalies and which represents the economically meaningful fraction of the market. In sum, our approach helps to assess to what extent prior results dealing with specific anomalies can be generalized.

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 on firm size are often stricter. Moreover, our sample period is on average about 20 years longer (due to an often earlier start date and typically more recent data), 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 (for further discussions see Green et al., 2013; Green et al., 2014; Harvey et al., 2015; McLean and Pontiff, 2015).

Second, market-level investor sentiment is a strong predictor of anomaly returns. This finding complements 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. Instead of relying on simulations for a limited set of anomalies, we test for generalizability by providing out-of-sample evidence for many anomalies not covered in their papers.

For more than 80% of the anomalies, the role of sentiment goes in the predicted direction, even though findings are only significant for about 40%. Eliminating noise by focusing on the "big picture" nevertheless reveals a powerful role of sentiment: 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. This is particularly noteworthy as we focus on relatively large and liquid firms for which sentiment is expected to be less relevant (see e.g. Lemmon and Portniaguina, 2006). In line with Stambaugh et al. (2012) and prominent theories, these results are strongest among return phenomena often attributed to investor overreaction, and they are mainly driven by the short leg of the portfolio. With respect to an aggregate anomaly, a one standard deviation increase in lagged sentiment leads to an insignificant return increase of less than 3 bp in the long leg, but to an highly significant return decrease of close to 18 bp in the short leg.[1]

Third, and in contrast to our findings for investor sentiment, 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. An eyeball test also suggests that the proxies tend to capture periods that one would intuitively classify as phases of relatively high limits to arbitrage (such as the great depression in the 1930ies or the recent financial crisis). In general, these variables turn out to have a low correlation (.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, 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. A notable exception is the role of idiosyncratic risk in some specifications. This time-series evidence is consistent with the view that idiosyncratic volatility may be the most important deterrent for arbitrage activity in the cross-section (e.g. Pontiff, 2006; Stambaugh et al., 2014).

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, use additional controls, focus on the long or short leg of the anomaly portfolios, consider quarterly instead of monthly returns, rely on alternative proxies for arbitrage constraints, consider composite anomalies, include or even focus on small firms,

[1] Taken together, these findings are consistent with the overpricing argument formally developed in Miller (1977). Many investors are reluctant or simply unable to go short (e.g. Almazan et al., 2004). In conjunction with such permanent short-sale constraints, mispricing in the sense of overpricing (induced by high investor sentiment) should be more prevalent than mispricing in the sense of underpricing (induced by low investor sentiment). This asymmetric effect suggests that anomalies, to the extent they reflect mispricing, should be more pronounced following positive than following negative sentiment, and that the short leg should be more sensitive to sentiment than the long leg (see also e.g. Stambaugh et al., 2015 and the references therein).