Anomalies and the Expected Market Return

XI DONG, YAN LI, DAVID E. RAPACH, and GUOFU ZHOU*

Journal of Finance, forthcoming

ABSTRACT

We provide the first systematic evidence on the link between long-short anomaly portfolio returns—a cornerstone of the cross-sectional literature—and the time-series predictability of the aggregate market excess return. Using 100 representative anomalies from the literature, we employ a variety of shrinkage techniques (including machine learning, forecast combination, and dimension reduction) to efficiently extract predictive signals in a high-dimensional setting. We find that long-short anomaly portfolio returns evince statistically and economically significant out-of-sample predictive ability for the market excess return. Economically, the predictive ability of anomaly portfolio returns appears to stem from asymmetric limits of arbitrage and overpricing correction persistence.

Stock return predictability is a fundamental topic in finance. There are two major—and voluminous—lines of research on this issue. The first examines whether firm characteristics can predict the *cross-sectional* dispersion in stock returns, and this line identifies

*XI DONG is at the City University of New York Baruch College. YAN LI is at the Southwestern University of Finance and Economics. DAVID E. RAPACH is at Saint Louis University. GUOFU ZHOU (corresponding author, zhou@wustl.edu) is at Washington University in St. Louis. Rapach was a visiting professor at Washington University in St. Louis when this project was undertaken. We thank Daniel Andrei, Svetlana Bryzgalova, John Cochrane, Andrew Detzel, Victor DeMiguel, Michael Hasler, Dashan Huang, Lawrence Jin, Sicong (Allen) Li, Lin Peng, Joel Peress, Kelly Shue, Jack Strauss, Yajun Wang, Xuemin (Sterling) Yan, and Lu Zhang, as well as conference and seminar participants at the 2020 FinTech Conference in China, Chinese University of Hong Kong-Shenzhen, City University of New York Baruch College, London Business School, Syracuse University, University of Bath, University of North Carolina at Charlotte, University of Liverpool, and Washington University in St. Louis, for helpful comments. We are especially grateful to the Editor, Associate Editor, and two referees for insightful and detailed comments that substantially improved the paper. We have read the *Journal of Finance* disclosure policy and have no conflicts of interest to disclose.

a plethora of equity market anomalies (e.g., Fama and French (2015), Harvey, Liu, and Zhu (2016), McLean and Pontiff (2016), Hou, Xue, and Zhang (2020)). The second line of research investigates the *time-series* predictability of the aggregate market excess return based on a host of economic and financial variables, such as valuation ratios, interest rates, and inflation (e.g., Nelson (1976), Campbell (1987), Fama and French (1988, 1989), Pástor and Stambaugh (2009)). Studies in this vein attempt to understand the variables that affect the equity risk premium.[1]

In this paper, we investigate whether these two leading lines of the finance literature are linked. Specifically, we analyze the ability of long-short anomaly portfolio returns from the cross-sectional literature to predict the market excess return. Our investigation has a number of key features. First, we focus on a cornerstone of the cross-sectional literature, long-short anomaly portfolio returns, as the literature often uses such returns as evidence of cross-sectional mispricing. Second, we employ *out-of-sample* tests, since such tests provide the most rigorous and relevant evidence on stock return predictability (e.g., Goyal and Welch (2008), Martin and Nagel (2019)). Third, we examine the predictive ability of a large number (100) of long-short anomaly portfolio returns representative of those from the cross-sectional literature and simultaneously aggregate the information in the set of anomaly portfolio returns. Because accommodating a large number of anomaly portfolio returns poses problems for the conventional multiple predictive regression approach, we apply a variety of *shrinkage* techniques—including machine learning, forecast combination, and dimension reduction—to guard against *overfitting* the data in a high-dimensional setting. Fourth, we explore economic rationales for the ability of long-short anomaly portfolio returns to predict the market return, thereby complementing our statistical findings with economic analysis.

Empirically, we find that the information in the group of 100 long-short anomaly portfolio returns is indeed useful for forecasting the monthly market excess return on an out-of-sample basis, provided that we rely on strategies that guard against overfitting the data. The out-of-sample R^2 (R^2_{OS}) statistics are economically sizable, ranging from 0.89%

[1] Rapach and Zhou (2013) provide a survey of the literature on market excess return predictability.

to 2.81%, all of which exceed the 0.5% threshold for economic significance suggested by Campbell and Thompson (2008). According to the Clark and West (2007) test, forecasts based on the long-short anomaly returns provide statistically significant reductions in mean squared forecast error (MSFE) vis-à-vis the prevailing mean (i.e., random walk with drift) benchmark forecast. Furthermore, a mean-variance investor with a relative risk aversion coefficient of three would be willing to pay 259 to 638 basis points in annualized portfolio management fees to have access to the market excess return forecasts based on the information in the group of 100 long-short anomaly portfolio returns. Overall, the evidence of out-of-sample market excess return predictability is statistically and economically strong and robust to various shrinkage methods.

Economically, the predictive power of long-short anomaly portfolio returns for the market return can be explained via asymmetric *mispricing correction persistence* (MCP), which arises from asymmetric limits of arbitrage (Shleifer and Vishny (1997)). To analyze the implications of MCP for market return predictability, we specify a data-generating process with stationary underpricing and overpricing components in the prices of the long and short legs, respectively, of the anomaly portfolio.[2] Consider, for example, the short leg. We show that when the return momentum effect from the correction of old overpricing shocks dominates the return reversal effect from the immediate correction of a new overpricing shock, the short-leg return in period t will be positively related to the market return in period $t+1$. Intuitively, the period-t short-leg return measures the correction of the overpricing identified by the anomaly characteristic at the end of period $t-1$ (Akbas et al. (2015), Engelberg, McLean, and Pontiff (2018)). If the overpricing correction process is sufficiently persistent, then there will be positive serial dependence in the short-leg return; because the short-leg return is part of the market return, this in turn means that the short-leg return is positively related to the future market return. The short-leg return can also signal overpricing in market segments beyond the short

[2]We explain our findings in a framework in which anomalies reflect mispricing. Whether our findings can also be explained in a risk-based framework is beyond the scope of the present paper but is an interesting topic for future research.

leg, thereby magnifying the effect on the market return. Indeed, we present empirical evidence that short-leg returns contain relevant overpricing information for a broad swath of the market. Analogous reasoning explains how the long-leg return can be positively related to the future market return.

Our empirical results indicate that long- and short-leg returns are positively related to the future market return. However, the out-of-sample predictive ability of long-leg returns is weak, indicating that MCP is considerably stronger for overpricing vis-à-vis underpricing. Consequently, long-short anomaly returns strongly negatively predict the market return. Via recursive regressions, we find evidence that such negative relationships remain important over time. Stronger MCP for overpricing relative to underpricing is consistent with arbitrage with respect to overpriced shares being less aggressive, due to factors such as short-sale constraints (Miller (1977)), feedback effects (Edmans, Goldstein, and Jiang (2015)), and price drops coupled with declines in liquidity (Dong, Krystyniak, and Peng (2019)). In support of relatively strong overpricing correction persistence, evidence suggests that overvaluation, as reflected in the short legs of anomaly portfolios, drives the profitability of many anomaly returns (e.g., Hong, Lim, and Stein (2000), Stambaugh, Yu, and Yuan (2012, 2015), Avramov et al. (2013)).

If MCP is stronger in the short leg compared to the long leg, it may seem that using short-leg anomaly portfolio returns instead of long-short anomaly returns would work better for forecasting the market return. However, we show econometrically that long-short anomaly returns can provide a stronger predictive signal for the market return. Intuitively, the long- and short-leg returns contain a common component unrelated to the future market return. By taking the difference between the long- and short-leg returns, we filter the noise in the predictor variable, thereby providing a sharper signal for anticipating the market return; such filtering is akin to alleviating the errors-in-variables problem. We also show that aggregating information across long-short anomaly returns can further sharpen the predictive signal by filtering the idiosyncratic noise in the individual predictors.

Methodologically, the most straightforward approach for incorporating the information in a large set of potential predictors is to specify a multiple predictive regression that includes all of the lagged predictors as explanatory variables. However, conventional ordinary least squares (OLS) estimation of a high-dimensional predictive regression is highly susceptible to overfitting. By construction, OLS maximizes the fit of the model over the in-sample estimation period, which can lead to poor out-of-sample performance; intuitively, OLS is prone to misinterpreting noise in the data for a predictive signal. Because we are interested in forecasting the monthly market excess return—which inherently contains a small predictable component—we are operating in a quite noisy environment, thereby exacerbating the danger of overfitting. As anticipated, we find that the conventional forecast based on OLS estimation of the multiple predictive regression that includes all of the long-short anomaly returns exhibits symptoms of extreme overfitting: the forecast is highly volatile and substantially less accurate than the prevailing mean benchmark.

We use a variety of forecasting strategies to guard against overfitting high-dimensional predictive regressions, all of which essentially rely on shrinkage. The first employs the *elastic net* (ENet, Zou and Hastie (2005)), a refinement of the well-known *least absolute shrinkage and selection operator* (LASSO, Tibshirani (1996)), to estimate the forecasting model. The LASSO and ENet are machine-learning techniques that use penalized regression to directly shrink the parameter estimates and thereby avoid overfitting the data.[3] We also consider *forecast combination* (Bates and Granger (1969)). We use a simple combination approach, which takes the arithmetic mean of univariate predictive regression forecasts based on the individual predictors (Rapach, Strauss, and Zhou (2010)), as well as a refinement due to Rapach and Zhou (2020), which incorporates insights from Diebold and Shin (2019) by using machine-learning techniques to select the individual forecasts to

[3]Our study complements recent studies that employ machine-learning techniques to predict stock returns using alternative predictor variables in high-dimensional settings, including Rapach, Strauss, and Zhou (2013), Chinco, Clark-Joseph, and Ye (2019), Rapach et al. (2019), Freyberger, Neuhierl, and Weber (2020), Gu, Kelly, and Xiu (2020), and Avramov, Cheng, and Metzker (2021).

include in the combination forecast. Finally, we employ *dimension-reduction* techniques to combine the predictors into a single variable, which subsequently serves as the explanatory variable in a univariate predictive regression. We consider three dimension-reduction techniques: the first takes the cross-sectional average of the individual predictors; the second extracts the first principal component from the set of predictors (Ludvigson and Ng (2007, 2009)); the third, an implementation of partial least squares (PLS, Wold (1966)), extracts the first target-relevant factor from the set of predictors (Kelly and Pruitt (2013, 2015), Huang et al. (2015)). As indicated previously, the forecasts based on strategies designed to circumvent overfitting all produce statistically and economically significant R^2_{OS} statistics and generate large economic gains for an investor.

To further assess the relevance of asymmetric limits of arbitrage for generating our findings, we form subgroups of anomalies using three proxies for limits of arbitrage: bid-ask spread, idiosyncratic volatility, and market capitalization. Specifically, we use the proxies to construct subgroups of anomalies characterized by more stringent limits of arbitrage in their short legs vis-à-vis their long legs. To the extent that the proxies capture limits of arbitrage, we expect the subgroups of long-short anomaly portfolio returns comprised of anomalies with relatively strong (weak) limits of arbitrage in their short legs to evince more (less) out-of-sample predictive ability for the market return. This is what we find.

Limits of arbitrage emerge from a variety of frictions (Gromb and Vayanos (2010)). In the models of Gârleanu and Pedersen (2013, 2016) and Dong, Kang, and Peress (2020), frictions such as limited risk-bearing capacity and transactions costs induce arbitrageurs to only slowly correct mispricing, resulting in MCP. To examine the relevance of limits of arbitrage, we analyze how various frictions are related to the out-of-sample predictive ability of long-short anomaly portfolio returns. Intuitively, we expect arbitrage activity to be more constrained during periods when frictions are more acute, so that we should find stronger out-of-sample predictive power for long-short anomaly returns during such periods. We consider a number of variables as proxies for frictions that can affect MCP in anomaly portfolios (especially in the short legs), including aggregate liquidity (Pástor and

Stambaugh (2003)), idiosyncratic risk (Ang et al. (2006), Pontiff (2006)), trading noise (Hu, Pan, and Wang (2013)), the VIX, economic uncertainty and risk aversion indices (Jurado, Ludvigson, and Ng (2015), Bekaert, Engstrom, and Xu (forthcoming)), and short costs (Asness et al. (2018)). Consistent with stronger MCP for overpricing vis-à-vis underpricing, we detect significant increases in out-of-sample market return predictability during periods of heightened frictions.

Finally, we provide more direct evidence that the predictive power of long-short anomaly portfolio returns for the market return stems from slow arbitrage, especially with respect to overpricing. Following Chen, Da, and Huang (2019), we construct net arbitrage positions in the overall market by aggregating the value-weighted trades of hedge funds and short sellers. We find that an increase in long-short anomaly portfolio returns predicts a statistically and economically significant decrease in net arbitrage positions in the market, primarily due to increases in short positions. We also examine the tone of public news about US financial markets extracted from news articles from Thomson Reuters by Calomiris and Mamaysky (2019). We find that an increase in long-short anomaly returns leads to more negative market news tone.

Our study is related to Engelberg et al. (2020), which is presently the only other study to systematically investigate connections between cross-sectional and time-series stock return predictability. Engelberg et al. (2020) follow the existing approach in the time-series literature by using the (equal- or value-weighted) average of firm-level values for a given characteristic as a predictor of the market return. Considering a large number of anomaly characteristics from the cross-sectional literature, they find little evidence that characteristics on average have out-of-sample time-series predictive ability for the market return. In contrast, we introduce a new approach that relies on long-short anomaly portfolio returns—a bedrock of the cross-sectional literature—to predict the market return, and we find strong evidence that the aggregate information in long-short anomaly returns is valuable for forecasting the market return. The two papers complement each other by shedding light on the relationship between cross-sectional and time-series stock return predictability.[4]

The remainder of the paper is organized as follows. Section I provides intuition on the predictive power of long-short anomaly portfolio returns for the market return. Section II describes the construction and evaluation of the out-of-sample forecasts. Section III presents the data, while Section IV reports forecast performance results. Section V provides additional results pertaining to asymmetric limits of arbitrage. Section VI concludes.

I. Intuition on Predictability

In this section, we use a stylized data-generating process to provide intuition on the predictive ability of anomaly portfolio returns for the market return.[5]

A. Data-Generating Process

Assume that the prices for the long and short legs of an anomaly portfolio contain a common martingale component with period-t increment f_t, while the long-leg (short-leg) price contains a stationary component $u_{L,t} \leq 0$ ($u_{S,t} \geq 0$) reflecting the level of underpricing (overpricing), which is uncorrelated with the common component. Intuitively, mispricing implies the presence of a stationary component in the price: under the assumption that the mispricing eventually corrects, any mispricing shock only temporarily affects the price, although the correction process can last for multiple periods. The log return (in terms of price changes) in each leg is then given by

$$r_{l,t} = f_t + \Delta u_{l,t} \quad \text{for } l = L, S, \tag{1}$$

[4] In a related but different context, Fama and French (2020) compare cross-sectional and time-series factor models.

[5] We thank the Editor for suggesting the framework for the data-generating process in this section.

where $r_{L,t}$ ($r_{S,t}$) is the long-leg (short-leg) return, $\Delta u_{l,t}$ is the change in mispricing, and $\text{cov}(f_t, \Delta u_{L,t}) = \text{cov}(f_t, \Delta u_{S,t}) = 0$. Using equation (1), the long-short anomaly portfolio return $r_{LS,t}$ can be expressed as

$$r_{LS,t} = \Delta u_{L,t} - \Delta u_{S,t}. \tag{2}$$

For expositional ease, assume that the long and short legs together comprise the market; then,

$$r_{M,t} = f_t + 0.5(\Delta u_{L,t} + \Delta u_{S,t}), \tag{3}$$

where $r_{M,t}$ is the (equal-weighted) market return.[6]

According to the Wold representation theorem, the stationary component in each leg related to mispricing (i.e., the pricing error) can be expressed as

$$u_{l,t} = \sum_{j=0}^{\infty} \psi_{l,j} v_{l,t-j} \quad \text{for } l = L, S, \tag{4}$$

where $\psi_{l,0} = 1$, $v_{L,t} \leq 0$ ($v_{S,t} \geq 0$) is a serially uncorrelated underpricing (overpricing) shock, $\text{var}(v_{l,t}) \geq 0$, $\sum_{j=1}^{\infty} \psi_{l,j}^2 < \infty$ (square summability), and $\psi_{l,j} \geq 0$ for $j \geq 1$ (to ensure that $u_{L,t} \leq 0$ and $u_{S,t} \geq 0$). For simplicity, we assume that $v_{L,t}$ and $v_{S,t}$ are uncorrelated. When $\text{var}(v_{L,t}) = \text{var}(v_{S,t}) = 0$, there is no mispricing, and then the market return in equation (3) reduces to $r_{M,t} = f_t$.

Equation (4) provides a general representation for the mispricing component in each leg, as any stationary autoregressive moving-average (ARMA) process can be expressed via the infinite-order MA process. The equation expresses the current-period level of mispricing as a function of current and past mispricing shocks. The MA process can be interpreted as an impulse-response function: $\psi_{l,j}$ is the response (ceteris paribus)

[6]Following much of the cross-sectional literature, the anomaly portfolios in Section IV are based on stocks sorted into decile portfolios, and each long-short anomaly portfolio goes long (short) the tenth (first) decile portfolio. We discuss the extension to additional market segments in Section I.D.

of $u_{l,t+j}$ for $j \geq 0$ to a period-t unit mispricing shock. For example, consider a unit overpricing shock and suppose that $\psi_{S,1} = 0.9$, $\psi_{S,2} = 0.45$, and $\psi_{S,j} = 0$ for $j \geq 3$. The pricing error is $\psi_{S,1} = 0.9$ at the end of the period immediately after the shock, so that $\psi_{S,1} - \psi_{S,0} = 0.9 - 1 = -0.10$; that is, 10% of the overpricing shock is corrected during the first period after the shock. The pricing error is $\psi_{S,2} = 0.45$ at the end of the second period after the shock, meaning that an additional 45% ($\psi_{S,2} - \psi_{S,1} = -0.45$) of the overpricing shock is corrected. Finally, the remaining 45% of the overpricing shock is corrected during the third period after the shock ($\psi_{S,3} - \psi_{S,2} = -0.45$).

Taking the first difference of equation (4), we obtain the expression for the change in mispricing:

$$\Delta u_{l,t} = \sum_{j=0}^{\infty} \tilde{\psi}_{l,j} v_{l,t-j} \quad \text{for } l = L, S, \tag{5}$$

where $\tilde{\psi}_{l,0} = \psi_{l,0} = 1$ and $\tilde{\psi}_{l,j} = \psi_{l,j} - \psi_{l,j-1}$ for $j \geq 1$. To simplify the exposition, we consider the following *mispricing correction assumption*:

$$\tilde{\psi}_{l,j} = \psi_{l,j} - \psi_{l,j-1} \leq 0 \quad \text{for } j \geq 1, \tag{6}$$

which assumes that arbitrage is sufficiently active to ensure that the mispricing associated with a period-t mispricing shock is not exacerbated in any subsequent period.[7]

B. *Mispricing Correction Persistence*

Consider a predictive regression relating the long- or short-leg return of the anomaly portfolio to next period's market return:

$$r_{M,t+1} = \alpha_l + \beta_l r_{l,t} + \varepsilon_{l,t+1} \quad \text{for } l = S, L, \tag{7}$$

[7] Relaxing this assumption does not change our main intuition, as discussed in footnote 11.

where $\varepsilon_{l,t+1}$ is a zero-mean, serially uncorrelated disturbance term. Using equations (1) and (3), the standardized slope coefficient in equation (7) is given by[8]

$$\tilde{\beta}_l = \frac{0.5\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t})}{[\text{var}(f_t) + \text{var}(\Delta u_{l,t})]^{0.5}} \quad \text{for } l = L, S. \tag{8}$$

Equation (8) indicates that the predictive ability of the long- or short-leg return depends on $\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t})$. Based on equation (5), the latter is given by

$$\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t}) = \left[(\psi_{l,1} - 1) + \sum_{j=1}^{\infty}(\psi_{l,j} - \psi_{l,j-1})(\psi_{l,j+1} - \psi_{l,j})\right]\text{var}(v_{l,t}) \text{ for } l = L, S. \tag{9}$$

Our empirical results generally indicate that $\tilde{\beta}_l > 0$, so that $\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t}) > 0$, especially for the short leg.

To understand the conditions that produce $\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t}) > 0$, we can use equation (5) to write the changes in the level of mispricing for the current and next period as

$$\Delta u_{l,t} = v_{l,t} + \sum_{j=1}^{\infty}(\psi_{l,j} - \psi_{l,j-1})v_{l,t-j} \quad \text{for } l = L, S, \tag{10}$$

$$\Delta u_{l,t+1} = v_{l,t+1} + (\psi_{l,1} - 1)v_{l,t} + \sum_{j=2}^{\infty}(\psi_{l,j} - \psi_{l,j-1})v_{l,t+1-j} \quad \text{for } l = L, S, \tag{11}$$

respectively. Equations (10) and (11) reveal that a new (i.e., period-t) pricing shock affects the current and future changes in mispricing in opposite directions. Consider an overpricing shock ($v_{S,t} > 0$). According to equation (10), $v_{S,t}$ exacerbates the current level of overpricing, corresponding to a positive change in overpricing in period t; according to equation (11), the overpricing induced by the period-t shock is corrected in proportion to $\psi_{S,1} - 1 \leq 0$ in period $t + 1$, corresponding to a non-positive change in overpricing. Thus, the consecutive changes in mispricing due to a new overpricing shock can generate

[8]Section IA.A of the Internet Appendix provides details for the derivations of the standardized slope coefficients for the predictive regressions in Sections I.B to I.D.

negative serial dependence in the short-leg return. In contrast, old pricing shocks ($v_{S,t-j}$ for $j \geq 1$) can produce positive serial dependence in the short leg return. According to equations (10) and (11), the changes in overpricing corresponding to these shocks are non-positive in consecutive periods; for example, the overpricing associated with $v_{S,t-1}$ is corrected in proportion to $\psi_{S,1} - 1 \leq 0$ in t and $\psi_{S,2} - \psi_{S,1} \leq 0$ in $t+1$.

Hence, two opposing effects determine $\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})$: (i) the extent to which the overpricing associated with a new shock is corrected in the next period and (ii) the extent to which the overpricing associated with old overpricing shocks is corrected in the current and next periods. These two effects are evident in the expression in brackets in equation (9), which accounts for all of the consecutive pairs of return responses to new and old overpricing shocks. The first term, $\psi_{S,1} - 1$, is the product of the period-t and period-$(t+1)$ return responses to a new unit overpricing shock, which captures the potential reversal due to the immediate correction of the overpricing induced by the new shock. The second term, $\sum_{j=1}^{\infty}(\psi_{S,j} - \psi_{S,j-1})(\psi_{S,j+1} - \psi_{S,j}) \geq 0$, reflects the potential momentum due to the persistent correction of the overpricing induced by old shocks. For example, $(\psi_{S,1} - 1)(\psi_{S,2} - \psi_{S,1})$ is the product of the period-t and period-$(t+1)$ overpricing corrections corresponding to a period-$(t-1)$ unit overpricing shock; $(\psi_{S,2} - \psi_{S,1})(\psi_{S,3} - \psi_{S,2})$ is the product of the period-t and period-$(t+1)$ overpricing corrections corresponding to a period-$(t-2)$ unit overpricing shock; and so forth.[9]

In order for $\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t}) > 0$, the return momentum generated by the correction of the overpricing induced by old shocks needs to outweigh the magnitude of the return reversal generated by the immediate correction of the overpricing induced by a new shock. When the momentum effect dominates, MCP is sufficiently strong to make $\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t}) > 0$ in equation (9) and $\tilde{\beta}_S > 0$ in equation (8). We can express this

[9]As $\psi_{S,1}$ decreases, the degree of overpricing correction in the period immediately after the shock increases, so that the magnitude of the reversal effect increases. In the extreme, $\psi_{S,1} = 0$, which implies that $\psi_{S,j} = 0$ for $j \geq 2$, so that the overpricing shock fully corrects in one period, and $\psi_{S,1} - 1 = -1$. In this case, the second term in brackets is zero, and $\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t}) = -\text{var}(v_{S,t})$ in equation (9).

condition as

$$\sum_{j=1}^{\infty}(\psi_{S,j}-\psi_{S,j-1})(\psi_{S,j+1}-\psi_{S,j}) > -(\psi_{S,1}-1). \tag{12}$$

An analogous condition to equation (12) holds for $\text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t})$ and $\tilde{\beta}_L$. Equation (12) is consistent with various theoretical models that can explain strong MCP.[10]

The previous example ($\psi_{S,1} = 0.9$, $\psi_{S,2} = 0.45$, and $\psi_{S,j} = 0$ for $j \geq 3$) numerically illustrates the intuition. A unit overpricing shock increases both the price and return by one unit at the time of the shock. Since $\psi_{S,1} = 0.9 < 1$, the price falls in the period after the shock, corresponding to a return of $\psi_{S,1} - 1 = -0.10$, so that the magnitude of the reversal effect is 0.10. The size of the momentum effect associated with old unit overpricing shocks is given by the second term is brackets, which equals $(-0.10)(-0.45)+(-0.45)(-0.45) = 0.2475$. In this case, the magnitude of the momentum effect outweighs that of the reversal effect, so that $\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t}) = 0.1475\text{var}(v_{S,t})$ in equation (9).[11]

Next, consider a predictive regression based on the long-short anomaly portfolio return:

$$r_{M,t+1} = \alpha_{LS} + \beta_{LS} r_{LS,t} + \varepsilon_{LS,t+1}. \tag{13}$$

[10]Andrei and Cujean (2017) show that when information about mispricing spreads among traders at an accelerated rate, immediate correction of a current pricing error is dominated by the corrections of previous pricing errors, resulting in return momentum. From a behavioral perspective, Chan, Jegadeesh, and Lakonishok (1996), Daniel, Hirshleifer, and Subrahmanyam (1998), Hong and Stein (1999), and Da, Gurun, and Warachka (2014) explain return momentum as under-reaction to news. Gârleanu and Pedersen (2013, 2016), and Dong, Kang, and Peress (2020) show that risk and/or cost considerations lead arbitrageurs to slowly allocate capital to correct mispricing.

[11]When the mispricing correction assumption in equation (6) does not hold, mispricing in response to a shock can be exacerbated after anomaly portfolio formation. Equation (9) still applies, so that $\text{cov}(\Delta u_{l,t+1}, \Delta u_{l,t}) > 0$ again requires the expression in brackets—which depends on the successive pairs of return responses to a mispricing shock—to be positive. In this case, we need the momentum effect of either mispricing correction or exacerbation to be stronger than the reversal effect in successive return pairs.

The standardized slope coefficient in equation (13) is given by

$$\tilde{\beta}_{LS} = \frac{0.5[\text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}) - \text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})]}{[\text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t})]^{0.5}}. \tag{14}$$

Empirically, we find that $\tilde{\beta}_{LS} < 0$, which holds when

$$\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t}) > \text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}), \tag{15}$$

that is, when there is stronger MCP with respect to overpricing vis-à-vis underpricing. Our finding that long-short anomaly portfolio returns negatively predict the market return is consistent with existing theoretical and empirical evidence. Miller (1977) argues that, due to short-sale impediments, overpricing should be more prevalent and persistent than underpricing. Edmans, Goldstein, and Jiang (2015) and Dong, Krystyniak, and Peng (2019) further show that feedback effects and price drops in conjunction with declines in liquidity, respectively, can generate overpricing persistence. In line with the relative importance of overpricing, studies such as Hong, Lim, and Stein (2000), Stambaugh, Yu, and Yuan (2012, 2015), and Avramov et al. (2013) find that the short legs of anomaly portfolios are primarily responsible for the profitability of long-short anomaly portfolio returns.

C. *Noise Reduction*

In our setup, the long-short anomaly portfolio return can produce a better predictive signal for predicting the market return than the short-leg return, which we also find empirically. To see this, for simplicity, we assume in this and the next subsection that $\tilde{\beta}_L = 0$, which is in line with the weak predictive ability of long-leg returns in the data (as discussed in Section IV.A). In this case, $\text{cov}(\Delta u_{L,t+1}, \Delta u_{L,t}) = 0$, so that equation (14) becomes

$$\tilde{\beta}_{LS}\big|_{\tilde{\beta}_L=0} = \frac{-0.5\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})}{[\text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t})]^{0.5}}. \tag{16}$$

The magnitude of $\tilde{\beta}_{LS}$ in equation (16) is greater than that of $\tilde{\beta}_S$ in equation (8) when

$$\text{var}(r_{LS,t}) < \text{var}(r_{S,t}), \text{ or, equivalently, } \text{var}(f_t) > \text{var}(\Delta u_{L,t}). \tag{17}$$

Intuitively, using $r_{L,t} - r_{S,t}$ in lieu of $r_{S,t}$ as the predictor in the predictive regression removes the common unpredictable component f_t, thereby filtering noise from the predictor to provide a sharper predictive signal for the market return. In other words, the problem of using the short-leg return alone to predict the market return is akin to the errors-in-variables problem. Consistent with this intuition, we empirically find that $\text{var}(r_{LS,t})$ is considerably smaller than $\text{var}(r_{S,t})$.

A similar intuition explains why the long-short anomaly portfolio return is a better predictor of next month's market return than the market return itself. The standardized slope coefficient for a regression of the market return on its own lag can be expressed as

$$\tilde{\beta}_M\Big|_{\tilde{\beta}_L=0} = \frac{0.5\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})}{[4\text{var}(f_t) + \text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t})]^{0.5}}. \tag{18}$$

Comparing equations (16) and (18), the numerators have the same magnitude. However, the presence of $4\text{var}(f_t)$ in the denominator of equation (18) creates noise in the predictive signal, so that the magnitude of the slope coefficient in equation (16) is larger than that in equation (18). Consistent with this intuition, in Section IV.C, we empirically find that the lagged market return cannot predict the market return on an out-of-sample basis.

D. Extensions

For expositional ease, we assume that the long and short legs cover the market in equation (3). When the long and short legs are the last and first decile portfolios, respectively—as in our empirical application—the same intuition applies. We address this issue empirically in Section V.B, where we show that the predictive ability of long-short anomaly returns extends across multiple market segments. Thus, our intuition behind market return predictability based on long-short anomaly returns extends beyond serial dependence in the long and short legs (especially the latter) and includes the relevance of long-short anomaly returns for signaling mispricing correction more generally across the market.[12]

Again for expositional ease, we focus on a single long-short anomaly return in the predictive regression in equation (13). In our empirical application, we amalgamate the information in a large number of long-short anomaly portfolio returns to predict the market return. In the spirit of Section I.C, aggregating predictors can help to reduce the noise in individual predictors to better uncover the predictive signal. To see this, we can extend our framework to multiple observable anomalies, each of which contains an idiosyncratic component:

$$r^i_{LS,t} = \Delta u_{L,t} - \Delta u_{S,t} + \omega^i_t \quad \text{for } i = 1, \ldots, n, \tag{19}$$

where ω^i_t is a serially uncorrelated idiosyncratic shock, $\text{cov}(\Delta u_{L,t}, \omega^i_t) = \text{cov}(\Delta u_{S,t}, \omega^i_t) = \text{cov}(f_t, \omega^i_t) = 0$, and $\text{cov}\left(\omega^i_t, \omega^j_t\right) = 0$ for $i \neq j$. The standardized slope coefficient for a predictive regression relating $r^i_{LS,t}$ to $r_{M,t+1}$ is given by

$$\tilde{\beta}^i_{LS}\big|_{\tilde{\beta}_L=0} = \frac{-0.5\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})}{\left[\text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t}) + \text{var}(\omega^i_t)\right]^{0.5}}. \tag{20}$$

Next, consider combining the information in the long-short anomaly returns by taking their period-t cross-sectional average. Based on equation (19), we have

$$\bar{r}_{LS,t} = \frac{1}{n}\sum_{i=1}^{n} r^i_{LS,t} = \Delta u_{L,t} - \Delta u_{S,t} + \frac{1}{n}\sum_{i=1}^{n}\omega^i_t. \tag{21}$$

[12]In a different context, Ehsani and Linnainmaa (forthcoming) find evidence of positive autocorrelation in factor returns and use it to explain cross-sectional momentum. In contrast, we focus on asymmetric MCP in the long and short legs of anomaly portfolios and its implications for market return predictability.

The standardized slope coefficient for a predictive regression relating $\bar{r}_{LS,t}$ to $r_{M,t+1}$ is given by

$$\tilde{\beta}_{LS}^{\text{Avg}}\big|_{\tilde{\beta}_L=0} = \frac{-0.5\text{cov}(\Delta u_{S,t+1}, \Delta u_{S,t})}{\left[\text{var}(\Delta u_{L,t}) + \text{var}(\Delta u_{S,t}) + \frac{1}{n}\overline{\text{var}}(\omega_t^i)\right]^{0.5}}, \tag{22}$$

where $\overline{\text{var}}(\omega_t^i) = \frac{1}{n}\sum_{i=1}^{n} \text{var}(\omega_t^i)$. If $\frac{1}{n}\overline{\text{var}}(\omega_t^i) < \text{var}(\omega_t^i)$, then the slope coefficient in equation (22) will be larger in magnitude than that in equation (20). Intuitively, like portfolio diversification, averaging across predictors can help to filter the idiosyncratic noise, thereby providing a stronger signal for predicting the market return.[13]

II. Methodology

This section describes the construction of the out-of-sample forecasts and their evaluation using both statistical and economic criteria.

A. Forecast Construction

Suppose that we want to forecast the market excess return ($r_{M,t}$) and that we have multiple potential predictors; in our context, the predictors are 100 long-short anomaly portfolio returns. We are interested in generating $\hat{r}_{M,t+1|t}$, a forecast of the month-($t+1$) market excess return based on information available through month t. All of our market excess return forecast are out of sample, as we only use data available through month t to forecast $r_{M,t+1}$.

The prevailing mean forecast, which implicitly assumes that the market excess return is unpredictable (apart from its mean value), is the most popular benchmark in the literature. The prevailing mean forecast is simply the average of the market excess

[13]For simplicity, we assume that the loadings on $\Delta u_{L,t} - \Delta u_{S,t}$ in equation (19) are equal to one for all of the long-short anomaly returns. When we allow for different loadings across anomalies, the strength of the signal in an anomaly will also depend on the magnitude of its loading. A number of the forecasting strategies in Section II.A can be interpreted as effectively placing greater emphasis on long-short anomalies with stronger signals (in a manner that guards against overfitting).

return observations available at the time of forecast formation. Because the predictable component in the monthly market excess return is inherently small (i.e., return data are quite noisy), the prevailing mean forecast is difficult to beat in practice (e.g., Goyal and Welch (2008)). With this in mind, successful out-of-sample strategies effectively shrink the market excess return forecast toward the prevailing mean benchmark to reduce the likelihood of overreacting to the noise in return data.

We compare the prevailing mean benchmark to the seven forecasts summarized below, each of which incorporates the information in the group of predictors. Section IA.B of the Internet Appendix provides details for the construction of each forecast.

Conventional OLS The conventional OLS forecast is based on a fitted multiple predictive regression that includes all of the lagged predictors as explanatory variables. Although it is straightforward to compute, the conventional OLS forecast is likely to perform poorly in practice, especially when the number of predictors is large. By construction, conventional OLS estimation maximizes the fit of the model (i.e., the in-sample R^2 statistic) over the estimation sample, which can lead to in-sample overfitting and poor out-of-sample performance, especially for high-dimensional models. The inherently large unpredictable component in the market excess return exacerbates the overfitting problem.

ENet The ENet forecast is based on the multiple predictive regression fitted via the ENet instead of OLS. The ENet (Zou and Hastie (2005)) relies on penalized regression to guard against overfitting. The ENet penalty term includes both ℓ_1 (LASSO) and ℓ_2 (ridge, Hoerl and Kennard (1970)) components. The ℓ_1 component permits shrinkage to zero, so that the ENet performs variable selection. Based on Flynn, Hurvich, and Simonoff (2013), we use the Hurvich and Tsai (1989) corrected version of the Akaike (1973) information criterion to select the value of the regularization parameter governing the degree of shrinkage. The ENet directly addresses overfitting by shrinking the slope coefficients of the fitted model, which has the effect of shrinking the forecast toward the prevailing mean benchmark.

Simple combination Instead of OLS estimation of the multiple predictive regression, forecast combination begins by computing a set of forecasts based on OLS estimation of univariate predictive regressions that include each lagged predictor (in turn). The simple combination forecast is the arithmetic mean of the individual univariate forecasts. As shown by Rapach, Strauss, and Zhou (2010), the simple combination forecast exerts a strong shrinkage effect.

Combination ENet When the number of predictors is large, the simple combination forecast can be too conservative in the sense that it "overshrinks" the forecast toward the prevailing mean, thereby neglecting too much of the relevant information in the predictor variables. Using insights from Diebold and Shin (2019), Rapach and Zhou (2020) employ the elastic net to refine the simple combination forecast. Instead of averaging across all of the individual univariate predictive regression forecasts, the combination ENet (C-ENet) forecast takes the average of the individual forecasts selected by the ENet in a Granger and Ramanathan (1984) multiple regression relating the actual market excess return to the individual univariate forecasts.

Predictor average An alternative strategy for guarding against overfitting is to first combine the predictors themselves into a small number of variables and then use the reduced set of variables as predictors in a low-dimensional predictive regression. Intuitively, as discussed in Section I.D, we consolidate the predictors to filter the noise in the individual predictors. The predictor average forecast is based on OLS estimation of a univariate predictive regression in which the lagged cross-sectional average of the predictors serves as the explanatory variable.

Principal component We can also combine the predictors by extracting the first principal component from the set of predictors. The principal component forecast uses the lagged principal component as the explanatory variable in a univariate predictive regression estimated via OLS.

PLS The first principal component explains as much variation as possible in the predictors themselves; however, from a forecasting standpoint, we are interested in explaining the target variable. Instead of extracting a factor that explains as much of the variation in the predictors as possible, Kelly and Pruitt (2013, 2015) develop a three-pass-regression filter to construct a target-relevant factor from a set of predictors that is maximally correlated with the target variable; the lagged target-relevant factor then serves as the explanatory variable in a univariate predictive regression estimated via OLS. The three-pass regression filter is essentially a version of PLS.

B. Forecast Evaluation—Statistical Accuracy

We first assess market excess return forecasts in terms of statistical accuracy via MSFE. Denote the errors for the prevailing mean benchmark and a competing forecast by

$$\hat{e}_{0,t|t-1} = r_{M,t} - \hat{r}^{\mathrm{PM}}_{M,t|t-1}, \tag{23}$$

$$\hat{e}_{1,t|t-1} = r_{M,t} - \hat{r}_{M,t|t-1}, \tag{24}$$

respectively, where $\hat{r}^{\mathrm{PM}}_{M,t|t-1}$ is the prevailing mean benchmark forecast and $\hat{r}_{M,t|t-1}$ generically denotes a competing forecast. The sample MSFE is given by

$$\widehat{\mathrm{MSFE}}_j = \frac{1}{T} \sum_{t=1}^{T} \hat{e}^2_{j,t|t-1} \text{ for } j = 0, 1, \tag{25}$$

where T is the number of out-of-sample observations. We test for a difference in the population MSFEs using the Clark and West (2007) procedure, which can be conveniently implemented in a simple regression framework:

$$\underbrace{d_{t|t-1} + \left(\hat{r}^{\mathrm{PM}}_{M,t|t-1} - \hat{r}_{M,t|t-1}\right)^2}_{f_{t|t-1}} = \mu + \varepsilon_t, \tag{26}$$

where $\hat{d}_{t|t-1} = \hat{e}^2_{0,t|t-1} - \hat{e}^2_{1,t|t-1}$ is the period-t loss differential. The t-statistic corresponding to the OLS estimate of μ in equation (26) is used to test

$$H_0\text{: } \text{MSFE}_0 \leq \text{MSFE}_1 \ (\mu \leq 0) \text{ versus } H_A\text{: } \text{MSFE}_0 > \text{MSFE}_1 \ (\mu > 0), \tag{27}$$

where MSFE_j is the population MSFE for $j = 0, 1$.[14] The t-statistic is computed using a heteroskedasticity- and autocorrelation-consistent (HAC) standard error (Newey and West (1987)).

When comparing MSFEs for the prevailing mean benchmark and a competing market excess return forecast, it is common to report the Campbell and Thompson (2008) R^2_{OS} statistic:

$$R^2_{\text{OS}} = 1 - \frac{\widehat{\text{MSFE}}_1}{\widehat{\text{MSFE}}_0}. \tag{28}$$

Equation (28) gives the proportional reduction in the sample MSFE for the competing forecast vis-à-vis the prevailing mean benchmark. Using the Clark and West (2007) statistic to test equatiion (27) is tantamount to testing H_0: $R^2_{\text{OS}} \leq 0$ against H_A: $R^2_{\text{OS}} > 0$ (in population). Because the predictable component in the monthly market excess return is necessarily limited, the R^2_{OS} statistic will be small. Nevertheless, based on the market Sharpe ratio, Campbell and Thompson (2008) suggest that a monthly R^2_{OS} statistic as small as 0.5% can signal economic significance. As described in Section II.C, we also assess the economic significance of market return forecasts more directly by measuring their economic value to an investor.

In addition, we examine whether out-of-sample return predictability (as measured by the R^2_{OS} statistic) is related to market frictions. To the extent that greater frictions

[14] The well-known Diebold and Mariano (1995) and West (1996) (DMW) procedure uses $d_{t|t-1}$ instead of $f_{t|t-1}$ as the dependent variable in equation (26). Clark and McCracken (2001) and McCracken (2007) show that the DMW test tends to be severely undersized when comparing forecasts from *nested* models (as in our application), so that it has little power to detect improvements in forecast accuracy. Clark and West (2007) adjust the DMW test statistic so that its asymptotic distribution is well approximated by the standard normal.

exacerbate limits of arbitrage, we expect anomaly portfolio returns to generate stronger out-of-sample gains during high-friction periods. To test whether out-of-sample return predictability changes with the state of market frictions, we augment the Clark and West (2007) framework in equation (26) as follows:

$$f_{t|t-1} = \mu + \xi I_t + \varepsilon_t, \tag{29}$$

where I_t is an indicator variable that equals one (zero) if market frictions are high (low). We use the t-statistic corresponding to the OLS estimate of ξ in equation (29) to test

$$H_0\text{: } R^2_{\text{OS,high}} \leq R^2_{\text{OS,low}}\ (\xi \leq 0) \text{ versus } H_A\text{: } R^2_{\text{OS,high}} > R^2_{\text{OS,low}}\ (\xi > 0), \tag{30}$$

where $R^2_{\text{OS,high}}$ ($R^2_{\text{OS,low}}$) is the value of the R^2_{OS} statistic during periods of high (low) market frictions. We again compute the t-statistic using a HAC standard error.

C. *Forecast Evaluation—Economic Value*

In addition to statistical accuracy, we compare the benchmark and competing forecasts in terms of their economic value to an investor. Specifically, we consider a mean-variance investor who allocates across equities and risk-free Treasury bills each month. At the end of month t, the investor faces the following objective function:

$$\underset{w_{t+1|t}}{\arg\max}\ w_{t+1|t}\hat{r}_{M,t+1|t} - 0.5\gamma w^2_{t+1|t}\hat{\sigma}^2_{t+1|t}, \tag{31}$$

where γ represents the coefficient of relative risk aversion; $w_{t+1|t}$ and $1 - w_{t+1|t}$ are the allocations to the market portfolio and risk-free bills, respectively, in period $t+1$; $\hat{r}_{M,t+1|t}$ is the investor's market excess return forecast; and $\hat{\sigma}^2_{t+1|t}$ is the investor's forecast of the variance of the market excess return. The well-known solution to equation (31) takes the following form:[15]

$$w^*_{t+1|t} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{M,t+1|t}}{\hat{\sigma}^2_{t+1|t}}\right). \tag{32}$$

We assume that the investor uses the sample variance computed over a 60-month rolling estimation window to forecast the variance in equation (32). The average utility realized by the investor is given by

$$\bar{U}_j = \bar{r}_j - 0.5\gamma\hat{\sigma}^2_j, \quad \text{for } j = 0, 1, \tag{33}$$

where $\bar{r}_0$ ($\bar{r}_1$) and $\hat{\sigma}^2_0$ ($\hat{\sigma}^2_1$) are the mean and variance, respectively, for the portfolio return over the out-of-sample period when the investor uses the prevailing mean (competing) forecast for $\hat{r}_{M,t+1|t}$ in equation (32). Mean-variance utility has the same units as returns, and equation (33) can be interpreted as the certainly equivalent return. Finally, we compute the average utility gain (or increase in certainty equivalent return) when the investor uses the competing forecast in lieu of the prevailing mean benchmark in equation (32):

$$\Delta = \bar{U}_1 - \bar{U}_0. \tag{34}$$

After multiplying the average utility gain in equation (34) by twelve, it can be interpreted as the annualized portfolio management fee (as a proportion of wealth) that the investor would be willing to pay to have access to the information in the competing forecast relative to that in the prevailing mean benchmark. For our empirical analysis in Section IV, we assume that $\gamma = 3$; the results are similar for reasonable alternative values for γ.

[15] To keep the equity allocation in a reasonable range, we impose the restriction that $-1 \leq w_{t+1|t} \leq 2$.

III. Data

We consider 100 long-short anomaly portfolio returns that are representative of anomalies from the cross-sectional literature and that can be replicated using publicly available data from CRSP, Compustat, and I/B/E/S. The anomalies are from numerous categories, such as value versus growth, profitability, investment, issuance activity, momentum, and trading frictions. Table IA.I in the Internet Appendix lists the 100 anomalies.[16]

For each anomaly, at the beginning of each month, we sort stocks into value-weighted decile portfolios based on the relevant characteristic. We consider all stocks traded on the NYSE, AMEX, and NASDAQ, after excluding stocks with a price below $5. The long-short anomaly portfolio goes long (short) the tenth (first) decile portfolio, where the long (short) leg is expected to generate relatively high (low) returns. We exclude anomalies that are interactions of two separate signals (since such anomalies are essentially based on multiple anomalies) and those that are indicator variables (such as IPOs). We also use anomalies that have a non-missing return starting in 1985.[17] The market excess return is the CRSP value-weighted market return minus the risk-free return (also from CRSP).

The sample period spans 1970:01 to 2017:12. Table I reports summary statistics for the anomaly portfolio returns. Of the 100 representative anomalies, 75, 71, 56, and 49 generate alphas with t-statistics (in magnitude) above 1.645, 1.96, 2.58, and three, respectively, in the context of the Fama and French (1993) three-factor model. The average correlation between decile rankings of any two anomalies is 0.05, which is similar to those reported in McLean and Pontiff (2016) and Green, Hand, and Zhang (2017), so that a stock considered overvalued or undervalued by one anomaly has a nearly equal chance of being considered overvalued or undervalued by another anomaly. This is consistent with the argument of Stambaugh, Yu, and Yuan (2012, 2015) and Akbas et al. (2015) that individual anomaly characteristics can be noisy indicators of whether a stock is mis-

[16]Section IA.C of the Internet Appendix provides detailed definitions and relevant studies for the anomalies.

[17]The exception is the change in forecasted earnings per share, which begins in 1989.

Table I
Summary Statistics

The table reports summary statistics for monthly anomaly portfolio returns for 100 anomalies. The sample period is 1970:01 to 2017:12. For each anomaly, we sort stocks into value-weighted decile portfolios according to the characteristic underlying the anomaly. The long-short anomaly portfolio goes long (short) the tenth (first) decile portfolio.

Number of anomalies	100
Fama and French (1993) three-factor model alpha	
Number of long-short anomaly portfolio returns with $\lvert t\text{-stat.}\rvert \geq 1.645$	75
Number of long-short anomaly portfolio returns with $\lvert t\text{-stat.}\rvert \geq 1.96$	71
Number of long-short anomaly portfolio returns with $\lvert t\text{-stat.}\rvert \geq 2.58$	56
Number of long-short anomaly portfolio returns with $\lvert t\text{-stat.}\rvert \geq 3$	49
Average correlation across anomaly decile rankings	0.05
Average correlation across monthly anomaly excess returns	
Long leg	0.76
Short leg	0.82
Long-short	0.08
Long-leg anomaly portfolio excess returns	
Average of sample means	0.71%
Average of sample standard deviations	5.16%
Short-leg anomaly portfolio excess returns	
Average of sample means	0.33%
Average of sample standard deviations	6.20%
Long-short anomaly portfolio returns	
Average of sample means	0.38%
Average of sample standard deviations	4.37%

priced. It also suggests that aggregating the information in anomalies can help to reduce the noise in individual anomalies.

On average, the long- and short-leg portfolio returns of any two anomalies are strongly correlated (0.76 and 0.82, respectively). In the context of Section I, this likely reflects the common component in the long- and short-leg returns unrelated to mispricing (f_t). The long-short returns are relatively weakly correlated on average (0.08). This is consistent with the presence of idiosyncratic components in the individual long-short anomaly returns, as discussed in Section I.D. In support of the condition in equation (17), the average of the sample standard deviations for the long-short portfolio returns is about

70% of that for the short-leg excess returns, indicating that long-short returns provide a sharper predictive signal than short-leg returns, as discussed in Section I.C. Figure IA.1 in the Internet Appendix provides histograms for the sample autocovariances for the long- and short-leg excess returns. Consistent with stronger MCP for overpricing vis-à-vis underpricing in equation (15), the mean of the autocovariances for the short-leg excess returns is considerably larger than that for the long-leg excess returns (4.39 and 2.14, respectively), and the difference in means is significant at the 1% level.[18]

IV. Out-of-Sample Results

In constructing the out-of-sample forecasts, we use the first ten years (1970:01 to 1979:12) of the full sample period as the initial in-sample estimation period. The subsequent five years (1980:01 to 1984:12) serve as the initial holdout out-of-sample period for computing the C-ENet forecast, so that 1985:01 to 2017:12 (396 observations) constitutes the out-of-sample period for forecast evaluation. Because the methods in Section II.A require non-missing predictor data (with the exception of the predictor average), for an anomaly with a missing return in a month, we fill in the missing return with the cross-sectional average for the available anomaly returns in that month.

A. *Forecast Accuracy*

Figure 1 depicts monthly market excess return forecasts based on the 100 long-short anomaly portfolio returns and strategies described in Section II.A. The conventional OLS forecast is highly volatile. In fact, the conventional OLS forecast is more than 6% (72% annualized) in magnitude for a number of months; such extreme forecasts point to severe

[18]We also generate a forecast of the short-leg excess return via a univariate predictive regression with the lagged short-leg excess return serving as the predictor. As shown in Table IA.II in the Internet Appendix, there is significant out-of-sample evidence of serial dependence: the R^2_{OS} statistics are positive for all anomalies, and 73, 30, and seven are significant at the 10%, 5%, and 1% levels, respectively. Table IA.III in the Internet Appendix reports the full matrix of cross-autocorrelations for the 100 long-short anomaly portfolio returns.

overfitting. The other forecasts are substantively less volatile than the conventional OLS forecast. In other words, as intended, the ENet, simple combination, C-ENet, predictor average, principal component, and PLS strategies all work to shrink the forecasts toward the prevailing mean (while still incorporating information from the 100 long-short anomaly portfolio returns). The simple combination forecast exerts a particularly strong shrinkage effect (as discussed in Section II.A). Excluding the conventional OLS forecast, the PLS forecast evinces the most volatility, with the volatilities for the ENet, C-ENet, predictor average, and principal component forecasts falling between those for the simple combination and PLS forecasts. Observe that the market excess return forecasts in Figure 1 are typically more volatile around business-cycle recessions, indicating that the long-short anomaly portfolio returns themselves are more volatile around recessions.

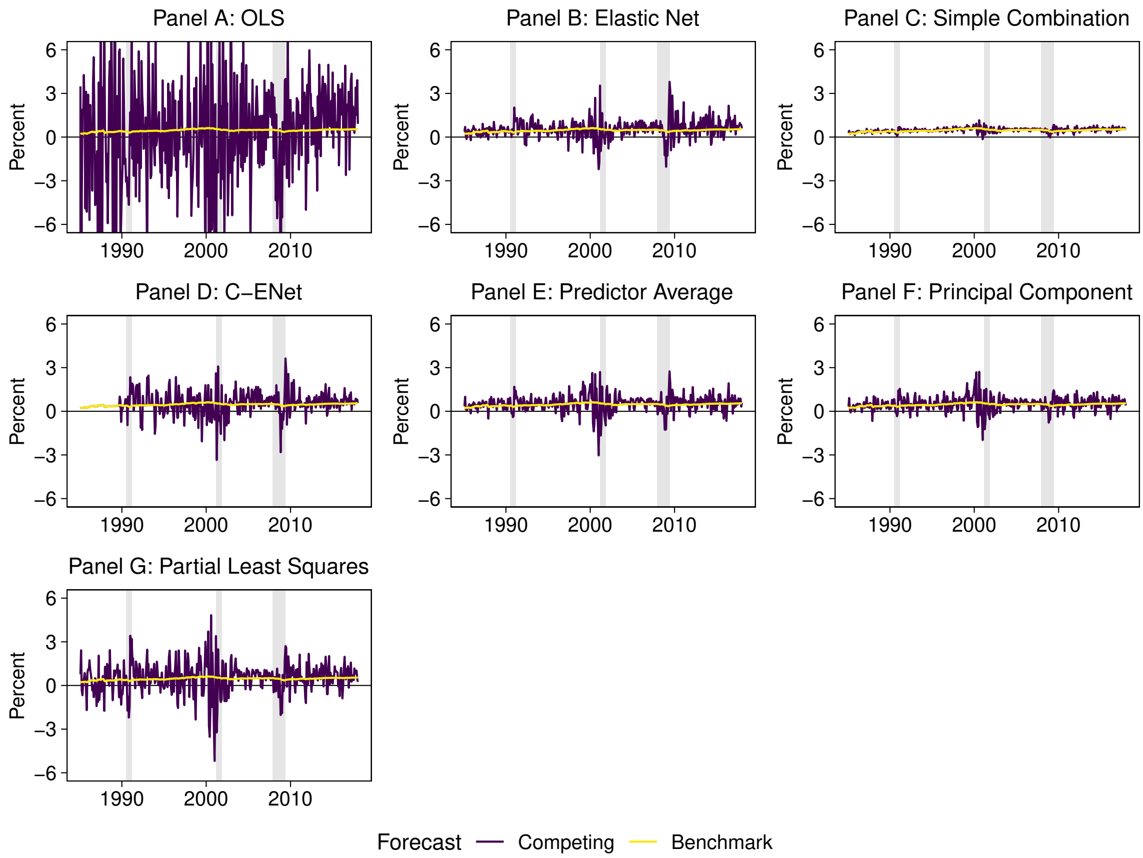

Figure 1. Market excess return forecasts based on 100 long-short anomaly portfolio returns. Each panel depicts the competing market excess return forecast in the panel heading and prevailing mean benchmark forecast. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

Table II reports R^2_{OS} statistics for monthly market excess return forecasts based on the 100 long-short anomaly portfolio returns. The conventional OLS forecast produces a negative R^2_{OS} statistic that is extremely large in magnitude, confirming that the forecast is plagued by overfitting.[19] In sharp contrast, the other forecasts all generate positive R^2_{OS} statistics, so that they are more accurate than the prevailing mean benchmark over the 1985:01 to 2017:12 out-of-sample period. Using the Campbell and Thompson (2008) threshold of 0.5%, the monthly R^2_{OS} statistics for all six of the forecasts designed to guard against overfitting are economically significant; indeed, the R^2_{OS} statistics for the ENet, C-ENet, predictor average, and PLS forecasts are quite sizable (2.03%, 2.81%, 1.89%, and 2.06%, respectively), making them among the highest in the literature to date. Furthermore, according to the Clark and West (2007) statistics, all six of the forecasts designed to guard against overfitting provide statistically significant improvements in MSFE vis-à-vis the prevailing mean benchmark.

Overall, we find evidence that the information in the group of 100 long-short anomaly portfolio returns is quite useful for predicting the monthly market excess return, provided that we employ forecasting strategies that guard against overfitting. The fact that the ENet, simple combination, C-ENet, predictor average, principal component, and PLS forecasts all generate statistically and economically significant improvements in out-of-sample accuracy indicates that our results are robust and not overly reliant on a particular method for circumventing overfitting.

Table II also reports results for monthly market excess return forecasts based on long- and short-leg anomaly portfolio excess returns. A clear pattern emerges: the short-leg excess returns provide some evidence of statistically and economically significant predictive ability, while there is no statistically or economically significant evidence of predictive ability for the long-leg excess returns. As discussed in Section I.B, this pattern is con-

[19]In a univariate predictive regression setting, Campbell and Thompson (2008) find that truncating OLS forecasts from below at zero helps to improve out-of-sample performance. In our application, although this approach improves the accuracy of the conventional OLS forecast, it still performs substantially worse than the prevailing mean benchmark ($R^2_{\text{OS}} = -31.50\%$). Thus, in our high-dimensional setting, truncating the OLS forecast appears insufficient for guarding against overfitting.

Table II
R^2_{OS} Statistics

The table reports Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percent for market excess return forecasts based on 100 anomaly portfolio returns. The out-of-sample period is 1985:01 to 2017:12. The OLS (ENet) forecast is based on ordinary least squares (elastic net) estimation of a multiple predictive regression that includes all 100 of the anomaly portfolio returns. Combine is the arithmetic mean of univariate predictive regression forecasts based on the 100 individual anomaly portfolio returns (in turn). C-ENet is the arithmetic mean of the univariate predictive regression forecasts selected by the elastic net in a Granger and Ramanathan (1984) regression. Avg is a univariate predictive regression forecast based on the cross-sectional average of the 100 anomaly portfolio returns. PC (PLS) is a univariate predictive regression forecast based on the first principal component (target-relevant factor) extracted from the 100 anomaly portfolio returns. Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for the positive R^2_{OS} statistics.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Anomaly Portfolio	OLS	ENet	Combine	C-ENet	Avg	PC	PLS
Long-short return	−2,513.86	2.03**	0.89***	2.81***	1.89**	1.25**	2.06***
Long-leg excess return	−344,960.22	−0.90	0.29	−0.68	0.26	0.24	0.41
Short-leg excess return	−13,284.68	1.81*	0.72*	0.39*	0.75*	0.74*	0.84*

sistent with asymmetric limits of arbitrage and stronger MCP for overpricing vis-à-vis underprcing.

Comparing the results in Table II for the long-short and short-leg excess returns, the former perform better than the latter for forecasting the monthly market excess return. As explained in Section I.C (and supported by the average volatilities reported in Table I), the long-short return can provide a sharper predictive signal when the long and short legs contain a sizable common component unrelated to the future market excess return, as the long-short return filters the common component to produce a less-diluted signal.

B. *Economic Value*

As described in Section II, we also measure the marginal economic benefit of the predictive ability of long-short anomaly portfolio returns for a mean-variance investor with a relative risk aversion coefficient of three who allocates between the market portfolio and risk-free Treasury bills. Figure 2 portrays log cumulative excess returns for portfolios using market excess return forecasts based on the 100 long-short anomaly returns. The figure also depicts the log cumulative excess return for the portfolio based on the prevailing mean benchmark forecast, as well as that for the market portfolio. Figure 2 reveals that portfolios that incorporate the information in the 100 long-short anomaly returns generally exhibit superior performance compared to both the portfolio based on the prevailing mean benchmark forecast and the market portfolio.

Figure 2 also reports annualized average utility gains when the investor uses a competing forecast of the market excess return in lieu of the prevailing mean benchmark. For the conventional OLS forecast based on the 100 long-short anomaly returns (not reported in the figure), there is a large loss of −4.97%, further manifesting the overfitting problem for the conventional approach. In contrast, the forecasts designed to guard against overfitting all generate substantial utility gains. The smallest annualized gain is for the simple combination forecast (2.59%), which is still economically sizable.[20] The annualized gain reaches as high as 6.38% for the PLS forecast, and it is above 600 basis points for the ENet and C-ENet forecasts (6.26% and 6.06%, respectively).

Although largely ignored by the market return predictability literature, it is important to assess the statistical significance of the average utility gains. To do so, we use the moving-block bootstrap procedure recommended by McCracken and Valente (2018) to compute critical values for testing H_0: $\Delta \leq 0$ against H_A: $\Delta > 0$ in equation (34). The utility gains for the portfolios based on the ENet, predictor average, and principal component (simple combination, C-ENet, and PLS) forecasts are significant at the 5% (1%) level. Overall, the utility gains provide further convincing evidence that long-short

[20]Pástor and Stambaugh (2000) use a 2% annualized utility gain as a threshold for economic significance when comparing factor models that include the market excess return.

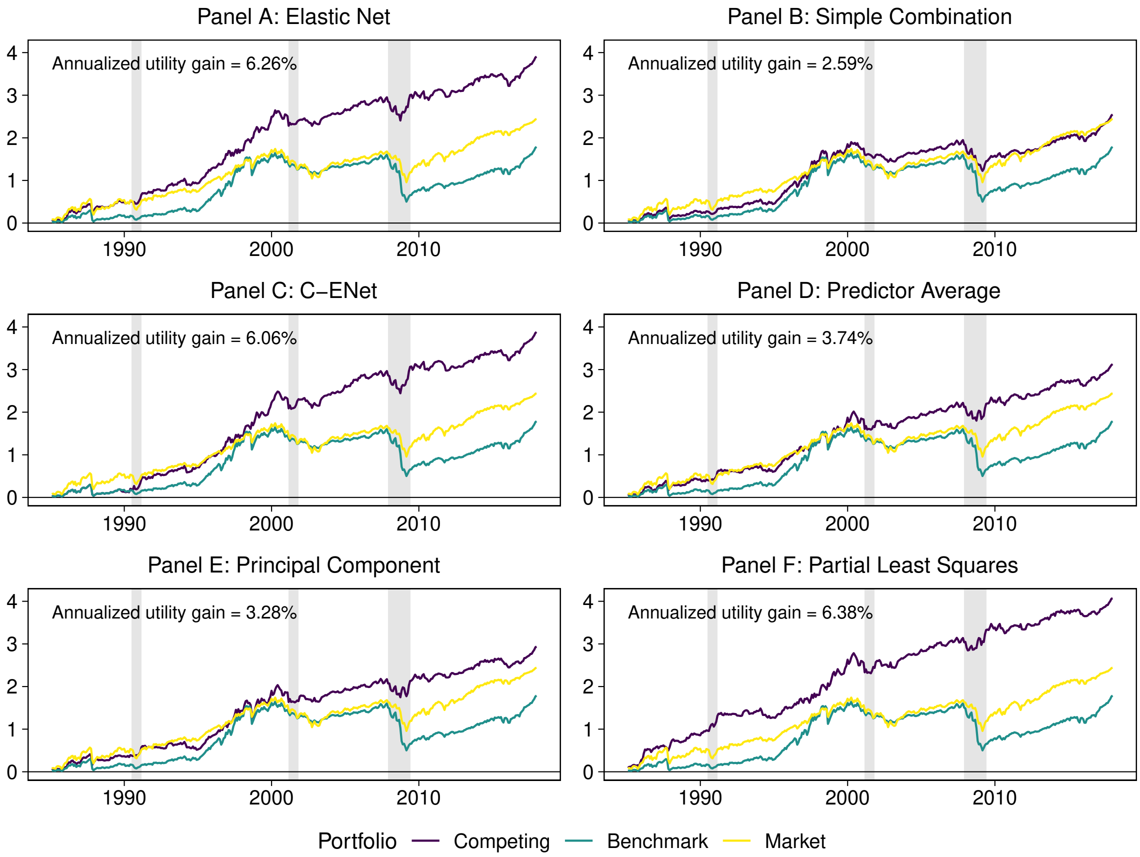

Figure 2. Log cumulative excess returns for portfolios constructed using market excess return forecasts based on 100 long-short anomaly returns. Each panel depicts the log cumulative excess return for a portfolio constructed using the market excess return forecast in the panel heading and the prevailing mean benchmark forecast, as well as the log cumulative excess return for the CRSP value-weighted market portfolio. Each panel also reports the annualized average utility gain when the investor uses the competing forecast in the panel heading in lieu of the prevailing mean benchmark forecast. Vertical segments delineate business-cycle recessions as dated by the National Bureau of Economic Research.

anomaly portfolio returns contain valuable information for predicting the market excess return.[21]

For the portfolios based on the ENet, C-ENet, predictor average, principal component, and PLS forecasts, the information in the long-short anomaly returns appears especially valuable for improving performance during the latter half of the Great Recession. It is worth noting that most of the 100 anomalies were published before that time, suggesting

[21] Table IA.IV in the Internet Appendix reports Sharpe ratios for the portfolios, which follow the same pattern as the utility gains in Figure 2.

that the utility gains accruing to the information in the group of anomalies are not limited to the anomalies' pre-publication periods.[22]

C. Additional Results

Because the long- and short-leg anomaly returns are parts of the market return, the lagged market return itself potentially contains relevant information for forecasting the monthly market return. For the 1970:01 to 2017:12 sample period, the autocorrelation for the market excess return is 0.08 (significant at the 10% level). However, the lagged market excess return does not evince out-of-sample predictive ability: for the 1985:01 to 2017:12 out-of-sample period, a univariate predictive regression forecast based on the lagged market excess return fails to outperform the prevailing mean benchmark forecast ($R^2_{\text{OS}} = -0.58$). As explained in I.C, this is not surprising, since the market return is a noisier predictor than the long-short anomaly portfolio return.

Long-short anomaly portfolio returns also evince stronger predictive power for the monthly market excess return than popular predictors from the literature for the 1985:01 to 2017:12 out-of-sample period. Table IA.VI in the Internet Appendix reports R^2_{OS} statistics for univariate predictive regression forecasts based on 16 popular predictors, including a variety of economic variables (e.g., valuation ratios, interest rates, interest rate spreads, and inflation) from Goyal and Welch (2008) and technical indicators from Neely et al. (2014).[23] Among the 16 predictors, three produce a positive R^2_{OS} statistic:

[22]Table IA.V in the Internet Appendix reports R^2_{OS} statistics and annualized average utility gains for the 1985:01 to 2001:12 and 2002:01 to 2017:12 subsamples. The results indicate substantial out-of-sample gains across both subsamples, again provided that we use strategies designed to guard against overfitting.

[23]Data used to construct the predictors are from Amit Goyal's website and Federal Reserve Economic Data. The predictors are the log dividend-price ratio; log earnings-price ratio; return volatility, computed using the approach in Mele (2007); three-month Treasury bill yield (deviation from a twelve-month moving average); ten-year Treasury bond yield (deviation from a twelve-month moving average); ten-year Treasury bond yield minus the three-month Treasury bill yield; Baa-rated corporate bond yield minus the Aaa-rated corporate bond yield; long-term Treasury bond return; corporate bond return minus the long-term Treasury bond return; net equity expansion; consumer price index inflation; industrial production growth; MA(1,12) (MA(3,12)), an indicator variable that takes a value of one if the S&P 500 index

the ten-year Treasury bond yield (0.54%, significant at the 10% level) and two of the technical indicators, MA(1,12) and MOM(6) (0.45% and 0.19%, respectively). When we aggregate the information in the 16 predictors using the methods in Section II.A, only the simple combination forecast outperforms the prevailing mean benchmark (R^2_{OS} = 0.14%).[24] Hence, these predictors appear less useful than long-short anomaly returns for predicting the market excess return.

To directly compare the information in the popular predictors to that in the long-short anomaly returns, Table IA.VIII in the Internet Appendix reports Harvey, Leybourne, and Newbold (1998) forecast encompassing test results for monthly market excess return forecasts based on the two sets of predictors. Forecasts based on the popular predictors do not encompass those based on the long-short anomaly returns, so that the anomaly returns contain information useful for forecasting the monthly market excess return beyond that contained in the popular predictors. Thus, the information in long-short anomaly returns appears quite different from that in popular predictors from the literature.[25] Moreover, as shown in Table II and Figure 2, the differential information in long-short anomaly returns delivers statistically and economically significant out-of-sample gains.

(three-month moving average of the S&P 500 index) is greater than the twelve-month moving average of the S&P 500 index and zero otherwise; MOM(6) (MOM(12)), an indicator variable that takes a value of one if the S&P 500 index is greater than its value six (twelve) months ago and zero otherwise.

[24]Table IA.VII in the Internet Appendix reports annualized average utility gains for forecasts based on the 16 predictors. Only the simple combination and principal component forecasts generate positive annualized utility gains (0.29% and 1.16%, respectively).

[25]Table IA.IX in the Internet Appendix reports R^2 statistics for univariate regressions of the ENet, simple combination, C-ENet, predictor average, principal component, and PLS forecasts based on the long-short anomaly returns on forecasts based on the 16 individual popular predictors (in turn). Further confirming the weak links between market excess return forecasts based on the long-short anomaly returns and those based on popular predictors, only twelve of the 96 R^2 statistics are above 5%, and none are above 15%. Among the popular predictors, the forecast based on the corporate bond return minus the long-term Treasury bond return is generally the most strongly related to the forecasts based on the long-short anomaly returns.

The out-of-sample predictive ability of long-short anomaly portfolio returns at longer horizons is also of interest. Table IA.X in the Internet Appendix reports R^2_{OS} statistics for market excess return forecasts based on the 100 anomalies for horizons of two, three, six, and twelve months.[26] Predictability tends to fall as the horizon increases, and we find no significant evidence of predictability at the twelve-month horizon. The results are consistent with our economic explanation, as we expect the mispricing related to anomalies to correct reasonably quickly, so that long-short anomaly returns are primarily short-horizon predictors.

As an external validity test, we examine the out-of-sample predictive ability of long-short anomaly portfolio returns for industry excess returns.[27] Specifically, we consider five industry portfolios (spanning the market) from Kenneth French's Data Library: Consumer, Manufacturing, Hi-Tech, Health, and Other.[28] Table IA.XI in the Internet Appendix reports R^2_{OS} statistics for the five industry excess returns. The conventional OLS forecast continues to perform poorly, as the R^2_{OS} statistics are negative and large in magnitude for all of the industries. As with the market excess return, the forecasting strategies designed to guard against overfitting perform substantially better, producing positive R^2_{OS} statistics for three to five of the industries. For these strategies, 21 of the 30 R^2_{OS} statistics are significant at conventional levels. Out-of-sample return predictability is especially strong for Consumer, Manufacturing, and Other, and a number of the R^2_{OS} statistics are above 3% for these industries.[29] The results for the industry excess returns

[26]To compute market excess return forecasts beyond the one-month horizon, we straightforwardly modify the methods in Section II using the multi-period return, $r^{(h)}_{M,t} = (1/h)\sum_{j=1}^{h} r_{M,t+(j-1)}$, as the target variable. The HAC standard error used to compute the t-statistic for the intercept term in equation (26) accounts for the autocorrelation induced in the forecast errors by the overlap in the multi-period return observations.

[27]We thank a referee for suggesting this external validity test.

[28]Available at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[29]Out-of-sample return predictability is the weakest for Health. This is likely due in part to the fact that the Health industry is a relatively small segment of the market. For 1970:01 to 2017:12, the average

provide further evidence that long-short anomaly portfolio returns are genuine predictors of the market excess return.

As a final assessment of out-of-sample return predictability in this section, we explore the effects of the selection of the group of long-short portfolio returns used to forecast the market excess return.[30] We begin with a set of more than 18,000 long-short portfolio returns from Yan and Zheng (2017), which are constructed by sorting on individual firm characteristics from Compustat.[31] From the complete set of long-short portfolio returns, we form two subsets with insignificant and significant alphas, respectively.[32] Based on the evidence in Yan and Zheng (2017), long-short portfolio returns with highly significant alphas are likely to be meaningful anomalies and not the result of data mining. We randomly draw two groups of 100 long-short portfolio returns from the respective subsets; for each group, we generate out-of-sample market excess return forecasts beginning in 1985:01 using the predictor average and compute the R^2_{OS} statistics. Repeating this process 1,000 times, we have empirical distributions for the R^2_{OS} statistics for market excess return forecasts based on randomly selected groups of 100 insignificant and significant long-short portfolio returns.

Figure 3 reports histograms for the R^2_{OS} statistics for the two subsets of long-short portfolio returns. The R^2_{OS} statistics corresponding to the significant long-short portfolio returns are clustered to the right of those corresponding to the insignificant long-short portfolio returns. The difference between the means (1.91) for the two distributions is significant at the 1% level. The histograms in Figure 3 indicate that groups of significant long-short portfolio returns generally improve out-of-sample market excess return forecasts. Overall, groups of long-short portfolio returns that are significantly more likely to

market capitalization share for Health is 9%, while the shares range from 19% to 30% for the other four industries.

[30]We thank a referee for suggesting this insightful analysis.

[31]We thank Xuemin Yan and Lingling Zheng for providing the data from Yan and Zheng (2017).

[32]For the purpose of forming the subsets, insignificant (significant) alphas are those with t-statistics below one (above 2.58) in magnitude in the Fama and French (1993) three-factor model.

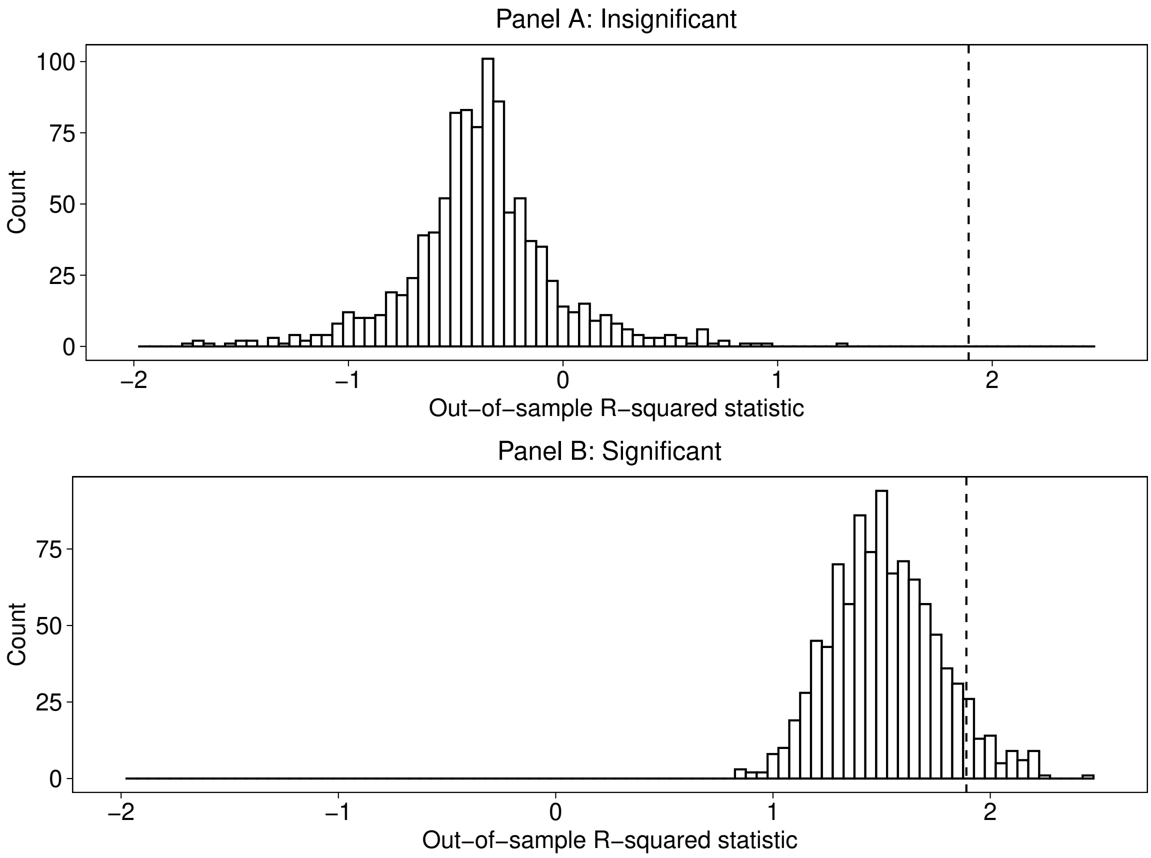

Figure 3. Histograms for R^2_{OS} statistics. The figure depicts histograms for Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percent for market excess return forecasts based on the predictor average. The forecasts are based on 1,000 randomly selected groups of 100 long-short portfolio returns from Yan and Zheng (2017) with insignificant (Panel A) and significant (Panel B) Fama and French (1993) three-factor model alphas. The vertical dashed line delineates the R^2_{OS} statistic for the predictor average forecast based on the long-short anomaly portfolio returns in Table II.

constitute anomalies appear more informative for predicting the market excess return, providing further evidence for the importance of anomalies.

V. Asymmetric Limits of Arbitrage

In this section, we further examine the relevance of asymmetric limits of arbitrage for market excess return predictability based on long-short anomaly portfolio returns.

A. Slope Coefficients

We first examine recursive estimates of the slope coefficients in the predictive regressions underlying the predictor average, principal component, and PLS forecasts. These predictive regressions provide a convenient means for capturing the information in all 100 of the long-short anomaly portfolio returns in a single slope coefficient.

The predictive regression underlying the predictor average forecast is given by

$$r_{M,t+1} = \alpha + \beta \bar{r}_{LS,t} + \varepsilon_{t+1}, \tag{35}$$

where $\bar{r}_{LS,t} = \frac{1}{n} \sum_{i=1}^{n} r^{i}_{LS,t}$ and $r^{i}_{LS,t}$ is the long-short portfolio return for the ith anomaly for $i = 1, \ldots, n$ ($n = 100$). For the principal component forecast, we replace $\bar{r}_{LS,t}$ in equation (35) with Z_t, the first principal component extracted from the 100 long-short anomaly portfolio returns. We set the sign of the principal component so that it is positively correlated with $\bar{r}_{LS,t}$, meaning that an increase in Z_t can be interpreted as a general increase in the long-short anomaly portfolio returns. For the PLS forecast, we replace $\bar{r}_{LS,t}$ in equation (35) with Z^{*}_{t}, the first target-relevant factor extracted from the 100 long-short anomaly portfolio returns, and we again set the sign of Z^{*}_{t} so that it is positively correlated with $\bar{r}_{LS,t}$. When computing the recursive slope coefficients estimates based on Z_t and Z^{*}_{t}, the principal components and target-relevant factors are computed using data available at the time of forecast formation, so that there is no look-ahead bias in the recursive estimates.

Panels A to C of Figure 4 depict recursive estimates of the standardized slope coefficients in equation (35) and their 90% confidence intervals when the explanatory variable is $\bar{r}_{LS,t}$, Z_t, and Z^{*}_{t}, respectively.[33] The recursive estimates are always negative, consistent with stronger MCP for overpricing vis-à-vis underpricing. As expected, the 90% confidence intervals tend to narrow as the estimation sample lengthens. For the predictor average (principal component) regression, the recursive estimates become significant starting in the mid-to-late 1990s (early 2000s) and remain significant thereafter; for the PLS regression, the recursive estimates are always significant. The recursive estimates

[33]The horizontal axes in Figure 4 correspond to the forecast month, so that the month-t estimate is based on data available through month $t - 1$. Confidence intervals for the slope coefficient estimates for the predictive regressions with Z_t and Z^{*}_{t} are based on the formulas for the standard errors in Bai and Ng (2006) and Kelly and Pruitt (2015), respectively.

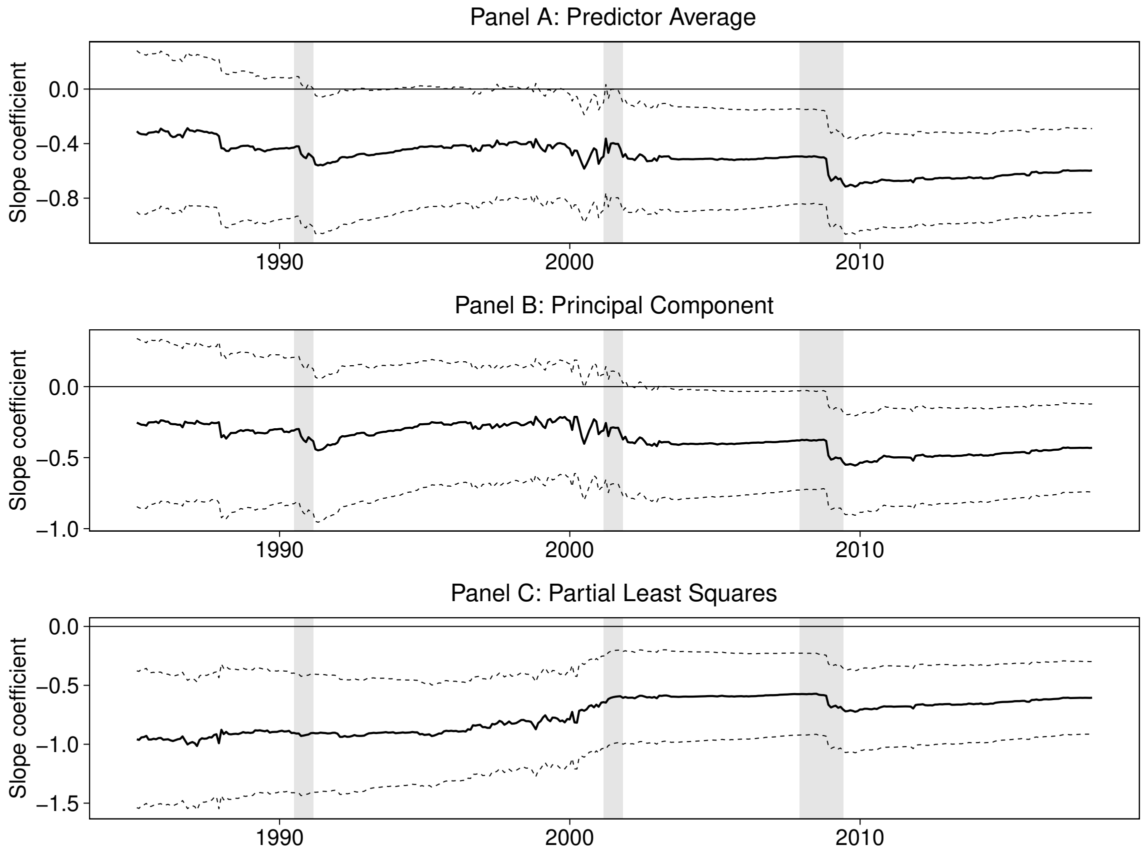

Figure 4. Recursive slope coefficient estimates. Solid lines depict standardized recursive slope coefficient estimates used to compute the predictor average, principal component, and partial least squares forecasts of the market excess return based on 100 long-short anomaly portfolio returns. Dashed lines delineate 90% confidence intervals. Vertical bars delineate business-cycle recessions as dated by the National Bureau of Economic Research.

are quite stable in all three panels of Figure 4 from the early to late 2000s. They then become larger in magnitude around the Global Financial Crisis and concomitant Great Recession and remain relatively stable thereafter, suggesting that asymmetric limits of arbitrage have not substantially diminished in importance since the early 2000s.[34]

[34]Table IA.XII in the Internet Appendix reports standardized OLS slope coefficient estimates for univariate predictive regressions based on the 100 individual long-short anomaly portfolio returns (in turn) for the 1970:01 to 2017:12 sample period. Further supporting stronger MCP for overpricing vis-à-vis underpricing, the vast majority (78) of the slope coefficient estimates are negative, and 41, 32, and 12 are significant at the 10%, 5%, and 1% levels, respectively.

B. Market Segments

According to the intuition in Section I, an important part of the predictive ability of long-short anomaly portfolio returns stems from their lead-lag covariances with segments of the market return. To glean insight into this mechanism, we examine the cross-autocovariances between long-short anomaly portfolio returns and decile excess returns. This allows us to investigate the relevance of the information contained in long-short anomaly returns for different segments of the market related to the anomalies.

To incorporate information from the entire group of 100 anomalies and construct market segments without overlapping stocks, we proceed as follows.[35] For a given month, we sort stocks according to each anomaly characteristic (in turn). For each stock, we then take the average of its ranks across the anomalies; we take the average of the ranks to account for the fact that some stocks have missing observations for some characteristics in a given month. We then sort stocks into value-weighted decile portfolios according to their average ranks, with the first (tenth) decile constituting the most overvalued (undervalued) stocks. The long-short portfolio again goes long (short) the tenth (first) decile portfolio.

Panel A of Figure 5 plots estimates of $\text{cov}(r_{j,t+1}, r_{LS,t})$ for the decile excess returns, indexed by $j = S, D_2, \ldots, D_9, L$. The cross-autocovariances are all negative. The first decile excess return ($r_{S,t+1}$) displays the largest covariance in magnitude with $r_{LS,t}$, so that MCP appears most relevant in the short leg. Nevertheless, the covariances of many of the other decile excess returns with the lagged long-short anomaly return are often significant at conventional levels, so that the mispricing detected in the long-short anomaly return has important implications for the market more broadly. The magnitudes of the cross-autocovariances tend to decrease from the first to the tenth deciles, and the cross-autocovariance for the tenth decile excess return ($r_{L,t+1}$) is the smallest in magnitude (and statistically insignificant). Lagged long-short anomaly returns thus appear more relevant

[35]We thank a referee for suggesting this informative procedure.

for segments of the market that contain stocks that are relatively more overpriced on average according to the anomalies, consistent with asymmetric limits of arbitrage.[36]

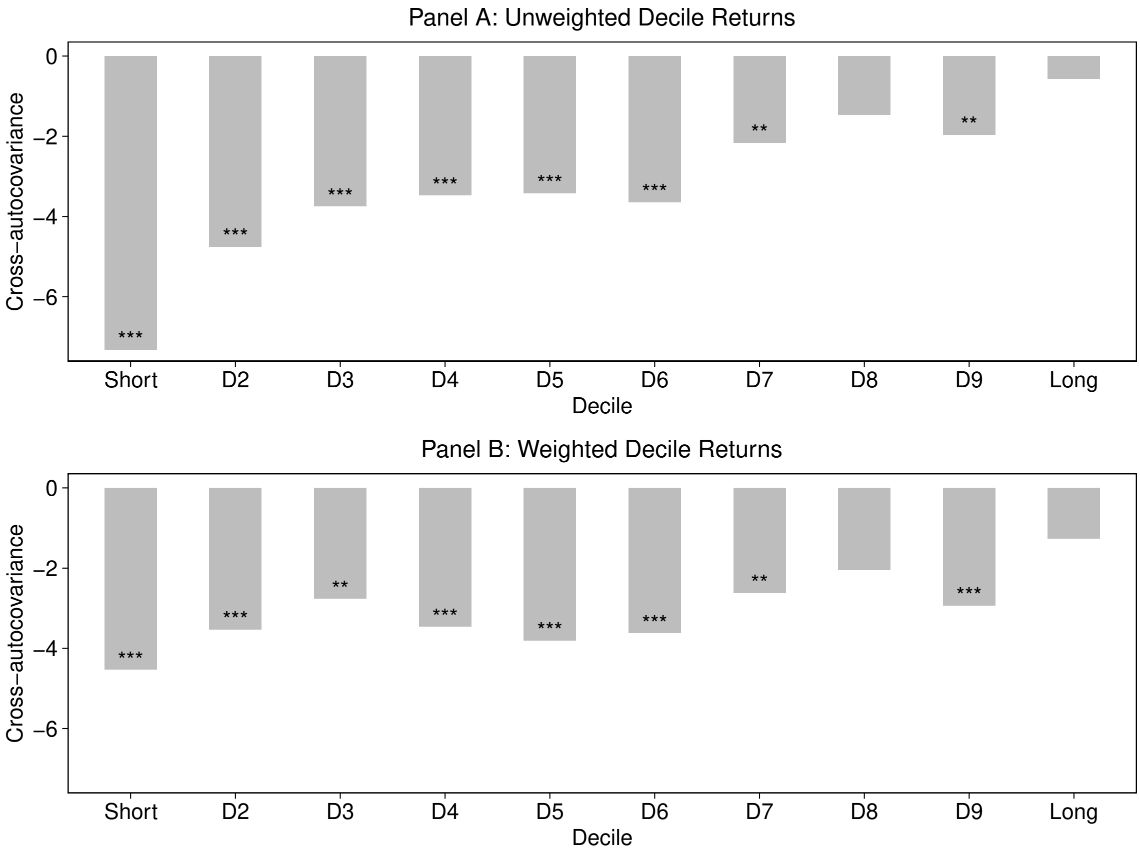

Figure 5. Cross-autocovariances for long-short anomaly and market segment returns. Each panel depicts estimates of cross-autocovariances for the month-$(t+1)$ decile excess return and month-t long-short anomaly return. The deciles are constructed by sorting stocks according to their average rank across 100 anomaly characteristics. The decile excess returns in Panel B are weighted by the deciles' market capitalization shares; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

For the deciles constructed based on the average of anomaly ranks, and in the spirit of Lo and MacKinlay (1990) and Lewellen (2002), we can obtain an exact decomposition of the covariance between the long-short anomaly return and next month's market excess

[36]The long-short portfolio return constructed from average ranks also evinces out-of-sample predictive ability for the market excess return. The R^2_{OS} statistic is 1.29%, which is significant at the 5% level according to the Clark and West (2007) test.

return as follows:

$$\begin{aligned}\operatorname{cov}(r_{M,t+1}, r_{LS,t}) = &\operatorname{cov}(w_{S,t+1} r_{S,t+1}, r_{LS,t}) + \sum_{j=2}^{9} \operatorname{cov}\big(w_{D_j,t+1} r_{D_j,t+1}, r_{LS,t}\big) \\ &+ \operatorname{cov}(w_{L,t+1} r_{L,t+1}, r_{LS,t}),\end{aligned} \tag{36}$$

where $w_{j,t+1}$ is the jth decile's market capitalization share. Panel B of Figurre 5 depicts estimates of the cross-autocovariances on the right-hand-side of equation (36), which weight the decile returns by their market capitalization shares.[37] The results are similar to those in Panel A, in that the cross-autocovariances are all negative, the cross-autocovariance for the first decile is the largest in magnitude, the cross-autcovariances are significant for many market segments, and they tend to decrease in magnitude from the first to the tenth deciles.

To further explore the relevance of overpricing correction persistence, we examine return predictability for portfolios of stocks that are most overvalued according to the predictor average.[38] The predictor average (i.e., the cross-sectional average of the long-short anomaly portfolio returns) for a given month is effectively a portfolio of all of the individual available stocks for the month (although some stocks may receive zero weights). Each month, we back out the portfolio weights corresponding to the predictor average and sort stocks according to the predictor average weights. We then form a portfolio (NEG10-PA) comprised of the 10% of stocks with the largest negative weights (in magnitude) in the predictor average portfolio; the weights for the stocks in the NEG10-PA portfolio are proportional to the absolute values of their weights in the predictor average portfolio. Based on a univariate predictive regression, we then use the excess return on the NEG10-PA portfolio to predict the excess return on a portfolio of the same stocks with the same weights in the next month. For the 1985:01 to 2017:12 out-of-sample period, the R^2_{OS} statistic for the predictive regression forecast is 1.87%, which is significant at the 5% level. We then repeat the same analysis using the same 10% of stocks each month, but

[37]We scale the cross-autocovariances in Figure 5 to makes them comparable across Panels A and B.

[38]We thank a referee for suggesting this instructive analysis.

instead use portfolio weights proportional to market capitalization. The R^2_{OS} statistic in this case is nearly a third smaller (0.65%) and insignificant at conventional levels. In sum, we find stronger out-of-sample return predictability for a portfolio of overvalued stocks (as detected by the predictor average) when we weight the stocks according to their relative importance in the predictor average portfolio, rather than simply by their market capitalization. This indicates that the information in the long-short anomaly portfolio returns is linked to overpricing correction persistence.

C. Subgroups

Next, we generate out-of-sample forecasts based on subgroups of anomalies formed according to proxies for asymmetric limits of arbitrage.[39] We form subgroups based on three proxies: bid-ask spread (BA), idiosyncratic volatility (IDIO), and market capitalization (SIZE).[40] For a given month and anomaly characteristic, we sort stocks into deciles as described in Section III and compute the average values for a given proxy for the stocks in the long and short legs, respectively. We then compute the long-leg average value for the proxy minus the short-leg average value for the proxy for each month. Finally, we compute the time-series average of the differences for 1970:01 to 1984:12. We denote the time-series averages for the differences for the bid-ask spread, idiosyncratic volatility, and size proxies by DTSA-BA, DTSA-IDIO, and DTSA-SIZE, respectively.

We form the subgroups BA-NEG, IDIO-NEG, and SIZE-NEG (BA-POS, IDIO-POS, and SIZE-POS), which are the subgroups of anomalies for which DTSA-BA, DTSA-IDIO, and DTSA-SIZE, respectively, are negative (positive). In line with our out-of-sample focus, we exclude data from the forecast evaluation period when determining the subgroups.[41] Stocks with larger bid-ask spreads, greater idiosyncratic volatility, and smaller market capitalization are typically viewed as having stronger limits of arbitrage. Hence,

[39] We thank a referee for recommending the formation of informative subgroups along these lines.

[40] Idiosyncratic volatility is the standard deviation of the residuals for a regression of the daily excess return on the three Fama and French (1993) factors for the previous month.

[41] Table IA.XIII in the Internet Appendix lists the anomalies included in the subgroups.

we expect greater market return predictability based on BA-NEG, IDIO-NEG, and SIZE-POS compared to BA-POS, IDIO-POS, and SIZE-NEG, respectively, as the former three subgroups represent anomalies with asymmetrically strong limits of arbitrage in their short legs vis-à-vis their long legs. To combine the information in the different proxies, we also form a subgroup that is the union of the anomalies in BA-NEG, IDIO-NEG, and SIZE-POS, as well as a subgroup comprised of the complement of the union. We expect the union subgroup to evince stronger predictive ability relative to its complement.

Table III reports R^2_{OS} statistics for the different subgroups using the strategies designed to guard against overfitting.[42] Although the bid-ask spread, idiosyncratic volatility, and size are quite noisy proxies for limits of arbitrage, the results generally support the relevance of asymmetric limits of arbitrage and stronger MCP for overpricing vis-à-vis underpricing. For example, Panel A reports results for the two subgroups formed using the bid-ask spread. The R^2_{OS} statistics for the BA-NEG subgroup are all positive and sizable in magnitude, well above the Campbell and Thompson (2008) threshold for economic significance. The R^2_{OS} statistics for the ENet, C-ENet, and PLS forecasts are particularly large (3.57%, 3.65%, and 2.94%, respectively). All six of the forecasts based on the BA-NEG subgroup are significant at conventional levels according to the Clark and West (2007) statistics. For each forecasting strategy, the R^2_{OS} statistic for the BA-POS subgroup is always lower than the corresponding statistic for the BA-NEG subgroup. Two of the R^2_{OS} statistics for the BA-POS subgroup are negative, and only two are above the 0.5% threshold. The results in Panel A align with the intuition in Section I, with relatively stronger limits of arbitrage in the short leg producing greater market return predictability. The results for the subgroups based on idiosyncratic volatility and size in Panels B and C, respectively, deliver similar messages to those for the subgroups based on the bid-ask spread in Panel A.

Panel D of Table III reports R^2_{OS} statistics for the union of the BA-NEG, IDIO-NEG, and SIZE-POS subgroups. All of the R^2_{OS} statistics for the union subgroup are statistically

[42]The conventional OLS forecasts based on the subgroups are always negative and large in magnitude, so that overfitting remains a serious problem.

Table III
R^2_{OS} Statistics for Subgroups

The table reports Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics in percent for market excess return forecasts based on long-short anomaly portfolio returns for the subgroup in the first column. The out-of-sample period is 1985:01 to 2017:12. The ENet forecast is based on elastic net estimation of a multiple predictive regression that includes all of the long-short anomaly portfolio returns in the subgroup. Combine is the arithmetic mean of univariate predictive regression forecasts based on the individual long-sort anomaly portfolio returns in the subgroup (in turn). C-ENet is the arithmetic mean of the univariate predictive regression forecasts selected by the elastic net in a Granger and Ramanathan (1984) regression. Avg is a univariate predictive regression forecast based on the cross-sectional average of the long-short anomaly portfolio returns in the subgroup. PC (PLS) is a univariate predictive regression forecast based on the first principal component (target-relevant factor) extracted from the long-short anomaly portfolio returns in the subgroup. Based on the Clark and West (2007) test, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for the positive R^2_{OS} statistics.

(1) Subgroup	(2) ENet	(3) Combine	(4) C-ENet	(5) Avg	(6) PC	(7) PLS
Panel A: Bid-Ask Spread						
BA-NEG	3.57***	1.13***	3.65***	1.97**	1.39**	2.94***
BA-POS	−0.28	0.48**	−1.41	0.97**	0.85*	0.29**
Panel B: Idiosyncratic Volatility						
IDIO-NEG	1.63***	0.97***	2.43***	1.53**	1.24**	1.85***
IDIO-POS	1.17*	0.42**	−0.43	−0.33	0.12	−0.14
Panel C: Market Capitalization						
SIZE-NEG	0.39	0.40**	−1.01	0.43	0.19	−1.20
SIZE-POS	2.39***	1.04**	2.91***	1.60**	1.15**	1.59***
Panel D: Bid-Ask Spread, Idiosyncratic Volatility, Market Capitalization						
Union of BA-NEG, IDIO-NEG, SIZE-NEG	2.67***	0.98***	3.10***	1.93**	1.23**	2.12***
Complement of union	0.37	−0.05	−0.09	−0.26	−0.31	−0.83

and economically significant. For the complement subgroup, five of the R^2_{OS} statistics are negative, and the positive statistic is neither statistically nor economically significant. Continuing the pattern in Panels A to C, the R^2_{OS} statistics for the subgroup that we expect to have stronger predictive ability (union) are always larger than the corresponding statistics for the subgroup that we expect to have weaker predictive ability (complement).

We also form subgroups via a double sort to identify subgroups of anomalies with both high limits of arbitrage overall and high asymmetric limits of arbitrage with respect to their short legs. To do so, for BA, we first identify anomalies with relatively low and high average values for the proxy for the stocks in their long and short legs (BA-LOW and BA-HIGH, respectively). For each of the BA-LOW and BA-HIGH subgroups, we then form NEG and POS subgroups as described above, resulting in four subgroups based on BA (BA-LOW-NEG, BA-LOW-POS, BA-HIGH-NEG, and BA-HIGH-POS). For compactness, we take the union of the BA-LOW-NEG, BA-LOW-POS, and BA-HIGH-POS subgroups to form the BA-REST group. We then compare the results between the BA-HIGH-NEG and BA-REST subgroups. In a similar manner, we form IDIO-HIGH-NEG and IDIO-REST (SIZE-LOW-POS and SIZE-REST) subgroups based on IDIO (SIZE). We expect the BA-HIGH-NEG, IDIO-HIGH-NEG, and SIZE-LOW-POS subgroups to exhibit stronger predictive ability compared to the BA-REST, IDIO-REST, and SIZE-REST subgroups, respectively, as the former three subgroups are the anomalies with both high limits of arbitrage overall and high asymmetric limits of arbitrage with respect to their short legs. As shown in Table IA.XIV in the Internet Appendix, this is indeed what we find.

D. Market Frictions

As argued by Gârleanu and Pedersen (2013, 2016) and Dong, Kang, and Peress (2020), frictions such as limited risk-bearing capacity and transaction costs induce arbitrageurs to slowly respond to mispricing, resulting in MCP. If our predictability finding is driven by arbitrageurs slowly correcting mispricing in the presence of asymmetric limits of arbitrage and stronger MCP for overpricing vis-à-vis underpricing, then long-short anomaly portfolio returns should contain more relevant information for predicting the market excess return during times of high frictions. We investigate this issue by testing for an increase in the R^2_{OS} statistic during periods of high frictions using equation (29).

We consider a variety of proxies for market frictions from the literature. We begin with the level of and innovations to aggregate liquidity from Pástor and Stambaugh (2003). We also consider idiosyncratic volatility, which is widely believed to be a major implementation cost of short arbitrage (e.g., Pontiff (2006)). We measure aggregate idiosyncratic risk for a given month by first calculating the idiosyncratic volatility of individual stocks following Ang et al. (2006) and then computing the value-weighted average of the idiosyncratic volatilities for the individual stocks. Furthermore, we use trading noise (Hu, Pan, and Wang (2013)), which tracks the shortage in arbitrage capital via noise in Treasury security prices, as well as short fees (Asness et al. (2018)), which measure the cost of shorting stocks. For the latter, we again aggregate to the market level by computing the value-weighted average of short fees for individual stocks.[43]

In addition to the above variables, we consider risk, uncertainty, and risk aversion indices as proxies for the frictions affecting the risk-bearing capacity of arbitrageurs. Along with the VIX, we use macroeconomic, financial, and real uncertainty indices from Jurado, Ludvigson, and Ng (2015), who construct the indices by estimating stochastic volatility series for the residuals from fitted dynamic diffusion-index models based on macroeconomic and financial variables. Finally, we consider the jointly estimated economic uncertainty and risk aversion indices from Bekaert, Engstrom, and Xu (forthcoming). For all of the proxies, we delineate high- and low-friction regimes using the sample median.

Table IV reports differences in R^2_{OS} statistics (in percentage points) between high- and low-friction regimes for market excess return forecasts based on the 100 long-short anomaly portfolio returns. In support of the relevance of asymmetric limits of arbitrage and stronger MCP for overpricing relative to underpricing, the R^2_{OS} statistics are almost always higher during high-friction periods, and the vast majority of the increases are significant according to the augmented Clark and West (2007) test. Furthermore, the magnitudes of the increases in the R^2_{OS} statistics are economically sizable, exceeding ten

[43] The short fee is based on the cost-to-borrow score from Markit. Short fee data are available beginning in 2002:01.

Table IV

R^2_{OS} Statistic Differences Between High- and Low-Friction Regimes

The table reports percentage-point increases in Campbell and Thompson (2008) out-of-sample R^2 (R^2_{OS}) statistics for market excess return forecasts based on 100 long-short anomaly portfolio returns. The out-of-sample period is 1985:01 to 2017:12. The increase is computed between high- and low-friction periods, which are defined using the sample median of the variable in the first column. The ENet forecast is based on elastic net estimation of a multiple predictive regression that includes all 100 of the long-short anomaly portfolio returns. Combine is the arithmetic mean of univariate predictive regression forecasts based on the 100 individual long-short anomaly portfolio returns (in turn). C-ENet is the arithmetic mean of the univariate predictive regression forecasts selected by the elastic net in a Granger and Ramanathan (1984) regression. Avg is a univariate predictive regression forecast based on the cross-sectional average of the 100 long-short anomaly portfolio returns. PC (PLS) is a univariate predictive regression forecast based on the first principal component (target-relevant factor) extracted from the 100 long-short anomaly portfolio returns. We use an augmented version of the Clark and West (2007) statistic to test for an increase in the R^2_{OS} statistic; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Friction	ENet	Combine	C-ENet	Avg	PC	PLS
Aggregate liquidity	4.48**	0.86**	5.20**	3.92**	3.64**	9.23**
Liquidity innovations	6.53**	1.42**	6.56**	7.82***	6.11***	13.21***
Idiosyncratic volatility	2.73*	1.06**	4.21**	4.29**	3.57**	2.05**
Trading noise	3.67**	1.49***	5.11**	8.12***	5.83***	12.42***
Short fees	10.34*	2.50**	12.56*	9.63**	8.50**	14.70**
VIX	−0.25	1.26**	4.87**	6.06**	6.10**	8.93**
Financial uncertainty	−1.28	0.02*	1.18*	3.24**	2.80**	3.94**
Macro uncertainty	1.20	0.95**	3.28*	3.97**	4.09**	7.21**
Real uncertainty	2.02*	1.10**	6.19**	4.47**	4.31**	8.02***
Economic uncertainty	1.29	0.88**	3.93*	3.67**	3.48**	3.59**
Risk aversion uncertainty	3.52**	0.60**	6.36**	4.68**	4.19**	7.38**

percentage points for some of the forecasts for frictions proxied by liquidity innovations, trading noise, and short fees.

E. Arbitrage Trading

Next, we examine whether long-short anomaly portfolio returns predict arbitrageurs' trading in the broad market and the tone of market-wide news. If the overpricing correction signaled by long-short anomaly returns during month t is only partial and continues to spread throughout the market in the subsequent month, then we expect an increase in long-short anomaly returns in month t to lead to a reduction in arbitrageurs' net positions (i.e., long minus short positions) in the broad market in month $t + 1$, which is likely to be primarily driven by an increase in their short positions; we also expect more negative market-wide news tone in month $t + 1$.

We first examine changes in arbitrageurs' aggregate net position in the broad market, where we follow Chen, Da, and Huang (2019) in constructing the net arbitrage trading measure. Specifically, we compute aggregate value-weighted long positions for hedge funds across all stocks in the market, where hedge funds' long positions in individual stocks are based on quarterly reported holdings in the 13F database. Agarwal et al. (2013) identify hedge funds by combining the information in the 13F database and five hedge fund databases. The 13F holdings data cover the largest number of institutional investors, as all institutional investment managers that have discretion over $100 million or more in Section 13(f) securities need to report their quarter-end holdings in these securities. A 13F-filing institution is classified as a hedge fund if its major business is sponsoring and/or managing hedge funds according to information from various sources (such as institution websites, SEC filings, industry publications, and news articles). Our final sample consists of 1,710 unique hedge funds.[44] We construct the value-weighted short-selling positions across all stocks in the market by aggregating short interest in individual stocks. A stock's short interest in a month is the total number of uncovered shares sold short for transactions settled on or before the 15th of the month (from Compustat) divided by the total number of shares outstanding (from CRSP). We use short interest from the last month of each quarter to match the quarterly data frequency for hedge-fund holdings. The sample period covers 1980:1 to 2017:4.

Since hedge funds' aggregate long positions and short interest display trends, we compute deviations in the long positions of hedge funds and short interest from trailing

[44]We thank Vikas Agarwal, Wei Jiang, Yuehua Tang, and Baozhong Yang for providing information for identifying hedge funds.

four-quarter moving averages. Chen, Da, and Huang (2019) show that the difference between the long and short arbitrage trading measures (i.e., net arbitrage trading) effectively aggregates the actions of arbitrageurs.

Table V reports standardized slope coefficient estimates for predictive regressions that relate quarterly arbitrage trading measures to lagged long-short anomaly portfolio returns, where we use long-short anomaly returns from the last month in the quarter. As in Section V.A, in order to capture the aggregate information across the 100 anomalies in a single slope coefficient, we estimate predictive regressions with $\bar{r}_{LS,t}$, Z_t, and Z_t^* serving as the predictor (in turn). The estimates for the net market position regressions in the second column are all significant (at the 5% or 1% level). The estimates are all negative, so that an increase in long-short anomaly returns leads arbitrageurs to decreases their net positions in the market. The coefficient estimates are also economically sizable: a one-standard-deviation increase in the predictor corresponds to a 4.81% to 5.81% reduction in the net arbitrage market position (in terms of value-weighted shares outstanding).

The third and fourth columns of Table V report results for arbitrageurs' long and short positions, respectively. The coefficient estimates are all negative (positive) for the long (short) trading measure, indicating that arbitrageurs reduce (increase) their long (short) positions in the market in response to an increase in long-short anomaly returns. The estimates are all significant at the 5% level for the regressions for the short market position, while they are all insignificant for the regressions for the long market position. Comparing the results in the third and fourth columns, the coefficient estimates are larger in magnitude for the short side than for the long side, consistent with stronger MCP for overpricing. Decomposing the net change in the second column, approximately one-third comes from declines in long positions and two-thirds from increases in short positions.

Finally, the last column of Table V reports predictive regression results for monthly US financial market news tone. The news-tone measure, from Calomiris and Mamaysky (2019), is based on English-language news articles from Thomson Reuters. Monthly news tone is an aggregation of differences in word tone among news articles on the topic of US financial markets. We detrend news tone by computing deviations from a trailing twelve-

Table V
Arbitrage Trading and News Tone

The table reports ordinary least squares estimates of standardized slope coefficient estimates for univariate predictive regressions that use the strategy in the first column to combine the information in 100 long-short anomaly portfolio returns to predict the variable in the column heading. Predictor average uses the the cross-sectional average of the 100 long-short anomaly portfolio returns. Principal component (partial least squares) uses the first principal component (target-relevant factor) extracted from the 100 long-short anomaly portfolio returns. The sample period for the second through fourth columns is 1980:1 to 2017:4; the sample period for the fifth column is 1996:01 to 2017:12. The dependent variable is standardized in the fifth column; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

(1) Strategy	(2) Arbitrage Net Market Position	(3) Arbitrage Long Market Position	(4) Arbitrage Short Market Position	(5) Financial Market News Tone
Predictor Average	-5.81^{***}	-1.91	3.89^{**}	-0.36^{***}
Principal component	-4.81^{**}	-1.18	3.63^{**}	-0.32^{***}
Partial least squares	-5.32^{***}	-1.62	3.70^{**}	-0.35^{***}

month moving average. In order to incorporate information from the 100 anomalies in a single slope coefficient, we again estimate predictive regressions with $\bar{r}_{LS,t}$, Z_t, and Z_t^* serving as the predictor (in turn). We standardize both the dependent variable and predictor before estimating the predictive regressions. The sample period is 1996:01 to 2017:12.[45]

To the extent that asymmetric limits of arbitrage generate relatively strong overpricing correction persistence, we expect an increase in long-short anomaly returns to lead to more negative market-wide news tone in the next month; in other words, public news is more likely to confirm that shares were broadly overvalued. The results in the last column of Table V support this proposition. The coefficient estimates are all significantly negative (at the 1% level), and the economic magnitudes of the estimates are sizable: a one-standard-deviation increase in the predictor results in a 0.32- to 0.36-standard-deviation

[45]We thank Charles Calomiris and Harry Mamaysky for providing updated data through 2017.

drop in market news tone. Overall, the results in Table V are consistent with the notion that long-short anomaly portfolio returns predict market returns because arbitrageurs slowly correct market-wide overpricing.

VI. Conclusion

Is cross-sectional stock return predictability related to the time-series predictability of the aggregate market excess return? Our paper provides the first positive systematic answer to this question. Specifically, we find that a representative group of 100 long-short anomaly portfolio returns from the cross-sectional literature contains valuable information for predicting the market excess return on an out-of-sample basis, provided that we use forecasting strategies that guard against overfitting the data. We explore economic channels underlying the predictive ability of long-short anomaly portfolio returns, explaining how asymmetric limits of arbitrage—by giving rise to overpricing correction persistence—can account for the patterns of return predictability in the data. We also provide evidence supporting the relevance of asymmetric limits of arbitrage and overpricing correction persistence for explaining our results. In light of our positive findings for the equity market, it would be interesting in future research to examine links between cross-sectional and time-series return predictability in other markets, such as bonds and currencies, by applying our methods for extracting predictive signals from large sets of cross-sectional anomalies.

REFERENCES

Agarwal, Vikas, Wei Jiang, Yuehua Tang, and Baozhong Yang, 2013, Uncovering hedge fund skill from the portfolio holdings they hide, *Journal of Finance* 68, 739–783.

Akaike, Hirotogu, 1973, Information theory and an extension of the maximum likelihood principle, in B. N. Petrov, and F. Csaki, eds., *Proceedings of the 2nd International Symposium on Information Theory*, 267–281 (Akadémiai Kiadó, Budapest).

Akbas, Ferhat, Will J. Armstrong, Sorin Sorescu, and Avanidhar Subrahmanyam, 2015, Smart money, dumb money, and capital market anomalies, *Journal of Financial Economics* 118, 355–382.

Andrei, Daniel, and Julien Cujean, 2017, Information percolation, momentum and reversal, *Journal of Financial Economics* 123, 617–645.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Asness, Clifford, Andrea Frazzini, Ronen Israel, Tobias J. Moskowitz, and Lasse H. Pedersen, 2018, Size matters, if you control your junk, *Journal of Financial Economics* 129, 479–509.

Avramov, Doron, Si Cheng, and Lior Metzker, 2021, Machine learning versus economic restrictions: Evidence from stock return predictability, Working Paper.

Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov, 2013, Anomalies and financial distress, *Journal of Financial Economics* 108, 139–159.

Bai, Jushan, and Serena Ng, 2006, Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions, *Econometrica* 74, 1133–1150.

Bates, John M., and Clive W. J. Granger, 1969, The combination of forecasts, *Journal of the Operational Research Society* 20, 451–468.

Bekaert, Geert, Eric Engstrom, and Nancy R. Xu, forthcoming, The time variation in risk uncertainty and appetite, *Management Science* .

Calomiris, Chalres W., and Harry Mamaysky, 2019, How news and its context drive risk and returns around the world, *Journal of Financial Economics* 133, 299–336.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies, *Journal of Finance* 51, 1681–1713.

Chen, Yong, Zhi Da, and Dayong Huang, 2019, Arbitrage trading: The long and the short of it, *Review of Financial Studies* 32, 1608–1646.

Chinco, Alex, Adam D. Clark-Joseph, and Mao Ye, 2019, Sparse signals in the cross-section of returns, *Journal of Finance* 74, 449–492.

Clark, Todd E., and Michael W. McCracken, 2001, Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* 105, 85–110.

Clark, Todd E., and Kenneth D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.

Da, Zhi, Umit G. Gurun, and Mitch Warachka, 2014, Frog in the pan: Continuous information and momentum, *Review of Financial Studies* 27, 2171–2218.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839–1885.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–263.

Diebold, Francis X., and Minchul Shin, 2019, Machine learning for regularized survey forecast combination: Partially-egalitarian Lasso and its derivatives, *International Journal of Forecasting* 35, 1679–1691.

Dong, Xi, Namho Kang, and Joel Peress, 2020, Slow arbitrage: Fund flows and mispricing in the frequency domain, Working Paper.

Dong, Xi, Karolina Krystyniak, and Lin Peng, 2019, Liquidity shocks and institutional trading, Working Paper.

Edmans, Alex, Itay Goldstein, and Wei Jiang, 2015, Feedback effects, asymmetric trading, and the limits to arbitrage, *American Economic Review* 105, 3766–3797.

Ehsani, Sina, and Juhani T. Linnainmaa, forthcoming, Factor momentum and the momentum factor, *Journal of Finance* .

Engelberg, Joseph, R. David McLean, and Jeffrey Pontiff, 2018, Anomalies and news, *Journal of Finance* 73, 1971–2001.

Engelberg, Joseph, R. David McLean, Jeffrey Pontiff, and Matthew C. Ringgenberg, 2020, Do cross-sectional predictors contain systematic information?, Working Paper.

Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 21, 3–25.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and the expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F., and Kenneth R. French, 2020, Comparing cross-section and time-series factor models, *Review of Financial Studies* 33, 1891–1926.

Flynn, Cheryl J., Clifford M. Hurvich, and Jeffrey S. Simonoff, 2013, Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models, *Journal of the American Statistical Association* 108, 1031–1043.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *Review of Financial Studies* 33, 2326–2377.

Gârleanu, Nicolai, and Lasse Heje Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 66, 2309–2340.

Gârleanu, Nicolai, and Lasse Heje Pedersen, 2016, Dynamic portfolio choice with frictions, *Journal of Economic Theory* 165, 487–516.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Granger, Clive W. J., and Ramu Ramanathan, 1984, Improved methods of combining forecasts, *Journal of Forecasting* 3, 197–204.

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang, 2017, The characteristics that provide independent information about average U.S. monthly stock returns, *Review of Financial Studies* 30, 4389–4436.

Gromb, Denis, and Dimitri Vayanos, 2010, Limits of arbitrage, *Annual Review of Financial Economics* 2, 251–275.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *Review of Financial Studies* 33, 2223–2273.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ...And the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Harvey, David I., Stephen J. Leybourne, and Paul Newbold, 1998, Tests for forecast encompassing, *Journal of Business and Economic Statistics* 16, 254–259.

Hoerl, Arthur E., and Robert W. Kennard, 1970, Ridge regression: Applications to nonorthogonal problems, *Technometrics* 12, 69–82.

Hong, Harrison, Terence Lim, and Jeremy C. Stein, 2000, Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies, *Journal of Finance* 55, 265–295.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

Hu, Grace Xing, Jun Pan, and Jiang Wang, 2013, Noise as information for illiquidity, *Journal of Finance* 68, 2341–2382.

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2015, Investor sentiment aligned: A powerful predictor of stock returns, *Review of Financial Studies* 28, 791–837.

Hurvich, Clifford M., and Chih-Ling Tsai, 1989, Regression and time series model selection in small samples, *Biometrika* 76, 297–307.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng, 2015, Measuring uncertainty, *American Economic Review* 105, 1177–1216.

Kelly, Brian, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *Journal of Finance* 68, 1721–1756.

Kelly, Brian, and Seth Pruitt, 2015, The three-pass regression filter: A new approach to forecasting using many predictors, *Journal of Econometrics* 186, 294–316.

Lewellen, Jonathan, 2002, Momentum and autocorrelation in stock returns, *Review of Financial Studies* 15, 533–563.

Lo, Andrew W., and A. Craig MacKinlay, 1990, When are contrarian profits due to stock market overreaction?, *Review of Financial Studies* 3, 175–205.

Ludvigson, Sydney C., and Serena Ng, 2007, The empirical risk-return relation: A factor analysis approach, *Journal of Financial Economics* 83, 171–222.

Ludvigson, Sydney C., and Serena Ng, 2009, Macro factors in bond risk premia, *Review of Financial Studies* 22, 5027–5067.

Martin, Ian, and Stefan Nagel, 2019, Market efficiency in the age of big data, NBER Working Paper No. 26586.

McCracken, Michael W., 2007, Asymptotics for out of sample tests of Granger causality, *Journal of Econometrics* 140, 719–752.

McCracken, Michael W., and Giorgio Valente, 2018, Asymptotic inference for performance fees and the predictability of asset returns, *Journal of Business and Economic Statistics* 36, 426–437.

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy return predictability?, *Journal of Finance* 71, 5–32.

Mele, Antonio, 2007, Asymmetric stock market volatility and the cyclical behavior of expected returns, *Journal of Financial Economics* 86, 446–478.

Miller, Edward M., 1977, Risk, uncertainty, and divergence of opinion, *Journal of Finance* 32, 1151–1168.

Neely, Christopher J., David E. Rapach, Jun Tu, and Guofu Zhou, 2014, Forecasting the equity risk premium: The role of technical indicators, *Management Science* 60, 1772–1791.

Nelson, Charles R., 1976, Inflation and rates of return on common stocks, *Journal of Finance* 31, 471–483.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Pástor, Ľuboš, and Robet F. Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics* 56, 335–381.

Pástor, Ľuboš, and Robet F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Pástor, Ľuboš, and Robet F. Stambaugh, 2009, Predictive systems: Living with imperfect predictors, *Journal of Finance* 64, 1583–1628.

Pontiff, Jeffrey, 2006, Costly arbitrage and the myth of idiosyncratic risk, *Journal of Accounting and Economics* 42, 35–52.

Rapach, David E., Jack K. Strauss, Jun Tu, and Guofu Zhou, 2019, Industry return predictability: A machine learning approach, *Journal of Financial Data Science* 1, 9–28.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821–862.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2013, International stock return predictability: What is the role of the United States?, *Journal of Finance* 68, 1633–1662.

Rapach, David E., and Guofu Zhou, 2013, Forecasting stock returns, in Graham Elliott, and Allan Timmermann, eds., *Handbook of Economic Forecasting*, volume 2A, 328–383 (Elsevier, Amsterdam).

Rapach, David E., and Guofu Zhou, 2020, Time-series and cross-sectional stock return forecasting: New machine learning methods, in Emmanuel Jurczenko, ed., *Machine Learning for Asset Management: New Developments and Financial Applications*, 1–34 (Wiley, Hoboken, NJ).

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.

Stambaugh, Robert F., Jiafeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 102, 288–302.

Stambaugh, Robert F., Jiafeng Yu, and Yu Yuan, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *Journal of Finance* 70, 1903–1948.

Tibshirani, Robert, 1996, Regression shrinkage and selection via the LASSO, *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267–288.

West, Kenneth D., 1996, Asymptotic inference about predictive ability, *Econometrica* 64, 1067–1084.

Wold, Herman, 1966, Estimation of principal components and related models by iterative least squares, in P. R. Krishnajah, ed., *Multivariate Analysis*, 391–420 (Academic Press, New York).

Yan, Xuemin, and Lingling Zheng, 2017, Fundamental analysis and the cross-section of stock returns: A data-mining approach, *Review of Financial Studies* 30, 1382–1423.

Zou, Hui, and Trevor Hastie, 2005, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67, 301–320.