**ESSAYS ON RETURN PREDICTABILITY AND ACTIVE
MANAGEMENT**

A Dissertation
presented to
the Faculty of the Graduate School
University of Missouri

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
PRATIK KOTHARI
Michael S. O'Doherty, Dissertation Supervisor

May 2019

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled

ESSAYS ON RETURN PREDICTABILITY AND ACTIVE MANAGEMENT

presented by PRATIK KOTHARI, a candidate for the degree of Doctor of Philosophy, and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Michael S. O'Doherty

Professor Don M. Chance

Professor Stephen P. Ferris

Professor Raynolde Pereira

Professor Xuemin (Sterling) Yan

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. Michael O'Doherty for his time, advice, encouragement, and invaluable feedback in my research process. I would like to thank Dr. Stephen Ferris for his extensive personal and professional guidance throughout the last five years. I would also like to thank the other members of my dissertation committee: Dr. Don M. Chance, Dr. Raynolde Pereira, and Dr. Xuemin (Sterling) Yan for valuable feedback and comments.

I am also grateful to the faculty, graduate students, and staff in the Department of Finance at University of Missouri for their support during my graduate studies.

Finally, this dissertation would not have been possible without the love, encouragement, and most importantly, patience of my wife Megan.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
LIST OF TABLES	vi
LIST OF FIGURES	vii
ABSTRACT	viii
1 Labor Market Frictions and Aggregate Stock Returns	1
1.1 Introduction	1
1.2 Theoretical framework	7
1.2.1 Production economy	8
1.2.2 Adjustment costs	9
1.2.3 Evolution of the labor force	9
1.2.4 Firm value	10
1.2.5 Equilibrium return	11
1.3 Data and summary statistics	13
1.3.1 Aggregate stock market returns	13
1.3.2 Job openings rate	13
1.3.3 Benchmark predictor variables	15

1.3.4	Summary statistics	16
1.3.5	<i>LJOR</i> and other labor market variables	17
1.4	Return predictability results	18
1.4.1	In-sample analysis	19
1.4.2	Comparison with other predictors	23
1.4.3	Out-of-sample analysis	25
1.4.4	Campbell and Thompson (2008) restrictions	29
1.4.5	Out-of-sample encompassing tests	32
1.5	Implications for asset allocation	34
1.6	Conclusion	39
2	Dissecting the Link Between Activeness and Fund Performance	56
2.1	Introduction	56
2.2	Data and Sample Construction	63
2.3	Managerial Activeness Proxies	65
2.4	Results	70
2.4.1	In-Sample Performance	70
2.4.2	Concentrated In-Sample Performance	74
2.4.3	Out-of-Sample Tests	82
2.4.4	Performance relative to a Five-Factor Model	86
2.4.5	Concentrated performance and omitted factors	90
2.4.6	Alternative Benchmark Models	91
2.5	Conclusion	94

APPENDIX

A Model and Data Appendix for Chapter 1	111
A.1 Derivation of expected returns	111
A.2 Construction of the benchmark predictors	113
B Data Appendix for Chapter 2	116
B.1 Construction of the activeness proxies	116
BIBLIOGRAPHY	122
VITA	129

LIST OF TABLES

1.1	Summary statistics	41
1.2	Job openings rate and other labor variables	42
1.3	In-sample predictive regression	43
1.4	In-sample comparison with other predictors	44
1.5	Out-of-sample forecasting results	46
1.6	Out-of-sample forecasting results with restrictions	47
1.7	Out-of-sample encompassing tests	49
1.8	Asset Allocation	50
1.9	Asset allocation with restricted forecasts	51
2.1	List of Performance Predictors	97
2.2	Summary Statistics and Correlations	98
2.3	In Sample Performance	99
2.4	Concentrated Performance	101
2.5	Out of Sample Performance	102
2.6	Performance Relative to Fama-French Five-Factor Model	103
2.7	Five-Factor Model - Factor Exposures	104
2.8	Alternative Benchmark Models	106

LIST OF FIGURES

1.1	Job openings rate	52
1.2	Out-of-sample forecasts	53
1.3	Equity weight	54
1.4	Log cumulative wealth	55
2.1	Net Excess Returns for Long-Short Portfolios	108
2.2	Factor Returns	109
2.3	Factor Returns	110

ABSTRACT

In this dissertation, I consider two topics in return predictability and performance evaluation in active management. In the first essay, I propose a production based asset pricing model with labor inputs. In this model, to either maintain the current size of the workforce (by replacing departing workers) or change the size of the workforce, firms advertise job openings and hire new workers each period. Changes in labor inputs are subject to stochastic adjustment costs coming from search-matching frictions in labor markets. The adjustment costs represent resources spent on hiring activities and increase convexly with the job openings rate (i.e., the ratio of the number of vacancies to the size of the workforce). The expected return for the firm is a ratio of marginal benefits to marginal costs. Accordingly, the model predicts a negative relation between the job openings rate and expected return in the subsequent period. Empirical tests confirm the model's prediction and show that the job openings rate is a strong negative predictor of returns. The job openings rate also outperforms 16 other popular forecasting variables. Forecasts obtained from the job openings rate also leads to large gains in returns for investors who consider it in making asset allocation decisions.

In the second essay, I reexamine the relation between managerial activeness and mutual fund performance. Several recent studies argue that investors can use ex ante observable proxies of managerial activeness (e.g., active share, active weight, return gap, and R2) to predict net of fee risk adjusted performance of mutual fund managers. Using a sample 11 such managerial activeness proxies, first, I show that the predictive power of the activeness proxies is largely concentrated in a three year

period from 1999 to 2001 and excluding this period, only two activeness proxies significantly predict net alpha. Second, none of the activeness proxies are significant predictors in out-of-sample tests. Third, the characteristic-sorted portfolios formed using the activeness proxies have positive exposures to profitability and investment factors. Using the Fama and French (2015) five-factor model as the benchmark, net alpha for these portfolios reduces by about 80% on average.

Chapter 1

Labor Market Frictions and Aggregate Stock Returns

1.1 Introduction

Fama and French (1989) emphasize that expected stock market returns are countercyclical and that investors require a higher risk premium to hold stocks at the end of a recession. Similarly, Cochrane (2009) notes that any potential return predictor should be related to business cycle variations. Because labor market conditions and human capital are integral components of business cycle variations, variables that capture labor market conditions are likely to be associated with variation in expected stock market returns. A large number of papers in asset pricing explore the predictive ability of valuation ratios (e.g., Fama and French (1988) and Campbell and Shiller (1988)) and variables related to macroeconomic conditions (e.g., Fama and Schwert (1977), Fama and Gibbons (1982), Fama (1981), and Lettau and Ludvigson (2001)).¹ Only a few recent studies, however, examine the role of labor market conditions in return predictability (e.g., Danthine and Donaldson (2002), Chen and Zhang (2011), and Belo et al. (2018)).

In this paper, I show that search-matching frictions in the labor market contain

¹See Rapach and Zhou (2013) for a survey of the literature on return predictability.

important information about expected aggregate stock market returns. Kuehn et al. (2017) note that more than 10% of the U.S. workers separate from their employers each quarter. Replacing these workers in the presence of search-matching frictions in the labor market comes with significant costs to firms. Such costs can prevent appropriate adjustments of employment levels in response to productivity shocks. Labor market frictions, thus, have an important role in determining the business cycle fluctuations and are likely to capture variations in expected returns.

I use a standard production-based asset pricing model that includes identical workers and a set of homogeneous firms that incur stochastic adjustment costs associated with search-matching frictions in the labor market. To either maintain the current size of the workforce (by replacing departing workers) or change the size of the workforce, firms advertise job openings and hire new workers each period. The adjustment costs represent the resources spent on hiring activities and increase convexly with the steepness of the turnover in the current workforce. Steeper turnover results in a higher number of job openings per unit of current workforce (i.e., job openings rate) and consequently higher adjustment costs. The expected returns on labor inputs are a ratio of marginal benefits and marginal costs. Adjustment costs, which increase with the job openings rate, are a part of the marginal costs of labor inputs. Accordingly, the model predicts a negative relation between the job openings rate and expected return in the subsequent period.

I follow Barnichon (2010) and Cajner and Ratner (2016) and measure the job openings rate as the total number of job openings in the economy per unit of the labor force. Figure 1.1 plots the monthly time series of this measure from January

1951 to December 2016. The shaded vertical bars represent National Bureau of Economic Research (NBER) dated recessions. As the economy grows during business cycle expansions, the job openings rate steadily increases. The job openings rate reaches a peak just prior to the beginning of recessionary period and then sharply declines through the recessions. From February 1951 to December 2016, there are a total of 11 recessions. In all the 11 cases, peaks in the job openings rate appear just prior to beginning of the recession. Therefore, the job openings rate appears to be a leading indicator of variations in business cycle conditions.

To empirically evaluate the relation between the job openings rate and aggregate stock market returns, I employ standard return predictability regressions (e.g., Fama and French (1989)) at each forecast horizon ranging from one month to two years. In these regressions, the dependent variable is the excess return on the S&P 500 index and the independent variable is the log job openings rate (henceforth, “*LJOR*”) lagged by one month. I find that for each forecast horizon, *LJOR* is a negative and statistically significant predictor of returns at the 1% level. A one unit standard deviation increase in *LJOR* is associated with a 0.50% decrease in expected returns in the subsequent month. I also compare the predictive power of *LJOR* to that of 16 other popular return predictors (henceforth, “benchmark predictor variables”). At each forecast horizon, *LJOR* substantively outperforms the benchmark predictor variables. For example, *LJOR* generates in-sample R^2 statistics of 1.44% and 8.91% for the monthly and annual forecasting horizons, respectively. In comparison, the benchmark predictor variables generate R^2 statistics that range from 0.01% to 0.83% for the monthly horizon and from 0.01% to 4.83% for the annual horizon.

Goyal and Welch (2008) argue that, despite the significant evidence of in-sample predictive ability, popular predictor variables exhibit poor performance in out-of-sample tests. Using their testing methods, I show that the predictive content of *LJOR* is also impressive in the out-of-sample setting. *LJOR* generates a monthly (annual) out-of-sample R^2 statistic of 0.86% (7.30%). In comparison, the benchmark predictor variables generate out-of-sample R^2 statistics that range from -1.26% to 0.35% (-13.67% to 1.83%) for the monthly (annual) forecasting horizon. I also conduct the out-of-sample encompassing tests developed by Harvey et al. (1998). In each tests, the ENC statistic is significant at the 1% level indicating that the forecasts obtained from *LJOR* contain incremental information to the forecasts obtained from each of the benchmark predictors.

I further examine the economic significance of the predictive ability of *LJOR* using an asset allocation analysis. I consider a mean-variance investor with a risk aversion coefficient of three who employs a strategy that invests in the S&P 500 index and a risk-free asset and rebalances on a monthly basis. The asset weights in the portfolio are determined by forecasts obtained from *LJOR* and are constrained so that the weight on the S&P 500 index is between 0% and 150%. That is, I restrict the portfolio to have no short-selling and a maximum of 50% leverage. I compare this strategy to a similar investment strategy that uses the prevailing mean forecast to determine the weight on the S&P 500 index. A mean-variance investor who uses the forecasts obtained from *LJOR* generates an annualized utility gain (i.e., additional certainty equivalent return over that of the prevailing-mean strategy) of 3.36%, which is statistically significant at the 1% level. This utility gain can be interpreted as the

annualized percentage fees an investor would pay to have an access to return forecasts from *LJOR*. In comparison, annualized utility gains associated with the forecasts from the 16 benchmark predictor variables range from -1.00% to 3.17% . The portfolio re-balancing strategy that uses forecasts from *LJOR* generates an annualized Sharpe ratio gain (i.e., additional Sharpe ratio over that of the prevailing-mean strategy) of 0.20 which is statistically significant at the 1% level. In comparison, the Sharpe ratio gains associated with forecasts from the 16 benchmark predictor variables range from -0.07 to 0.20 .

These results contribute to a growing literature on production-based asset pricing. Papers in this literature study relation between production inputs and asset returns. A large portion of this literature examines the role of capital-based or labor-based investment decisions in explaining the cross-section of expected returns. Only a small number of studies examine the ability of labor related variables to predict aggregate stock returns. These papers consider the association between stock market returns and aggregate labor share (i.e., Danthine and Donaldson (2002)), wage rigidity (i.e., Uhlig (2007)), hiring activity (i.e., Chen and Zhang (2011) and Belo et al. (2018)), and sectoral labor reallocation (i.e., Eiling et al. (2016)). Although the labor market characteristics in these studies are inherently related to business cycle variations, only a few of these studies find the evidence of return predictability (e.g., Belo et al. (2018)).

The two studies closest to this paper are Chen and Zhang (2011) and Belo et al. (2018). Both of these papers examine whether stock market returns are predictable using aggregate hiring rates. Chen and Zhang (2011) explore the bi-directional link

between labor and financial markets and find weak evidence of in-sample return predictability using hiring rates of all firms in the economy. They also find that hiring rates are predictable using aggregate stock market returns. Belo et al. (2018) show that the distribution of hiring rates in the economy provides important information regarding aggregate stock market returns. Using a sample period from 1963 to 2015, they show that the aggregate hiring rates of publicly listed firms predict excess aggregate stock market returns for horizons ranging from three to five years. My paper differs from these studies in several ways. The main predictor in this paper is the job openings rate, defined as the number of job openings per unit of the labor force. This variable differs from the hiring rate, which measures the changes in the employed population. In addition, the predictive power of the job openings rate documented in this paper is much larger than that of the hiring rates. Finally, unlike hiring rates, the predictive ability of the job openings rate is also evident at horizons shorter than three years.

This paper also contributes to the macroeconomic literature that examines theoretical links between labor and financial markets and builds on the standard search and matching framework of Diamond (1982) and Mortensen and Pissarides (1994).² For example, Phelps (1994) studies the impact of interest rates on unemployment. Merz and Yashiv (2007) build a theoretical model to link firm value to aggregate investment and hiring activities, and Yashiv (2016) extends this model to incorporate joint adjustment costs of investment and hiring. Hall (2017) shows high discount rates result in high unemployment. These papers feature a neoclassical production-based economy with labor market frictions similar to this study, but do not consider

²Rogerson and Shimer (2011) survey this literature.

implications for return predictability.

Goyal and Welch (2008) note that much of the literature on return predictability is plagued with spurious predictors that lack theoretical motivation. In addition, Cochrane (2009) argues that any potential return predictor should be linked to should also be related to business cycle variations. The job openings rate surpasses both of these hurdles. It is motivated through the standard investment theory framework and is more economically appealing than predictors that lack a theoretical foundation. The job openings rate also captures conditions in the labor market that are integral components of business cycles.

The rest of the paper is organized as follows. Section 1.2 presents a production-based asset pricing model that links the job openings rate to subsequent period expected returns. Section 1.3 describes the data sources and construction of predictor variables. Section 1.4 presents the results from in-sample and out-of-sample tests of return predictability. Section 1.5 analyzes the economic importance of the job openings rate from an asset allocation perspective. Section 1.6 concludes.

1.2 Theoretical framework

This section establishes a theoretical link between the job openings rate and expected returns. I build on the standard production-based asset pricing model with labor inputs and augment it for stochastic adjustment costs associated with the dynamic size of the labor force and labor market frictions.

1.2.1 Production economy

The economy consists of identical firms and workers. Production requires of only labor inputs and follows constant returns to scale. To keep the model tractable and to focus on the effects of labor market frictions, I abstract from modeling investment decisions related to capital inputs and financial frictions. Allowing for both labor and capital inputs, as in Yashiv (2016) and Millard et al. (2017), does not alter the main prediction of the model. The production function of the representative firm is

$$Y_t = e^{x_t} N_t, \tag{1.1}$$

where Y_t is total production or revenue, x_t is aggregate productivity, and N_t represents labor inputs employed by the representative firm.

The representative firm incurs three types of costs in the production process. First, wages associated with the labor inputs are assumed to be paid at a constant rate. Second, the representative firm incurs fixed costs, which are independent of the labor inputs and quantity of production. Third, the labor inputs are subjected to stochastic adjustment costs associated with the changing size of the workforce. These adjustment costs are similar to that of a production-based asset pricing model that contains capital investment as inputs and are further discussed in Section 1.2.2. The representative firm is all-equity financed. All of the firm's operating profits are disbursed to the shareholders in the form of dividends, which are given by the following equation:

$$D_t = Y_t - W_t N_t - F - C_t^{adj}, \tag{1.2}$$

where W_t is the gross compensation rate for the wages paid to workers, F is the fixed cost of production, and C_{adj} are the labor adjustment costs.

1.2.2 Adjustment costs

Adjustment costs represent the stochastic search costs incurred in adjusting the dynamic size of the firm's labor force. These costs include resources spent on hiring activities such as advertising for job openings, screening applications, interviewing applicants, and training new hires. Following Chen and Zhang (2011), Belo et al. (2014), and Belo et al. (2018), these costs take the following functional form:

$$C_t^{adj} = \frac{a}{2} \left(\frac{H_t}{N_t} \right)^2 N_t. \quad (1.3)$$

In the equation above, $a > 0$ is a constant and H_t equals the total number employees hired from the beginning of time t to the beginning of time $t + 1$. The total adjustment costs are convexly increasing in the ratio of the number of employees hired to the current size of the workforce (H_t/N_t). This property is desirable because the adjustment costs increase with the rate of adjustment in the workforce and more rapid changes in the workforce are associated with higher adjustment costs.

1.2.3 Evolution of the labor force

The representative firm makes hiring decisions at the beginning of each time period t . Newly hired workers enter the firm's workforce at the beginning of the next time period $t + 1$. The law of motion for a firm's labor force thus takes the following form:

$$N_{t+1} = N_t(1 - s) + \lambda_t J_t. \quad (1.4)$$

In Equation (1.4), s is the separation rate at which the current workforce leaves the firm voluntarily. This quantity satisfies $0 < s < 1$ and is assumed to be exogenous and constant following Shapiro (1986). J_t represents the number of job openings

advertised by the firm from the beginning to the end of time period t , and λ_t is the matching rate. The matching rate is defined as the probability that a job opening at the representative firm is filled at the end of time period t . I follow the approach of Chen and Zhang (2011) and Belo et al. (2014) and assume λ_t to be exogenous in the model without specifying a particular matching function.³ The last term in Equation (1.4), $\lambda_t J_t$, represents the number of workers hired during the time period t .

1.2.4 Firm value

The firm maximizes the discounted present value (V_t) of all future dividend payments received by its shareholders. This problem can be represented as

$$V_t = \max_{\{J_{t+\tau}, N_{t+1+\tau}\}} E_t \left[\sum_{\tau=0}^{\infty} (M_{t+\tau} D_{t+\tau}) \right], \quad (1.5)$$

where dividend payments $D_{t+\tau}$ are given by Equation (1.2) and $M_{t+\tau}$ is the stochastic discount factor in period $t + \tau$.

I do not model the endogenous supply side of labor coming from households. Also to keep the focus on the production side of the economy, I do not model the equilibrium wage rate W_t by explicitly solving the consumer's problem. Introducing these components in the model requires solving a general equilibrium model (Petrosky-Nadeau et al. (2018)). Instead, I simply assume that the firms are competitive and take prices as given and, hence, the stochastic discount factor is exogenous. This choice is standard in production-based asset pricing models with labor frictions (e.g., Chen and Zhang (2011), Belo et al. (2014), Kuehn et al. (2017), and Belo et al.

³This choice allows me to keep the focus on the production side of the economy and keeps the model tractable. Examples of specifying explicit search functions can be found in Kuehn et al. (2017) and Den Haan et al. (2000).

(2018)).

1.2.5 Equilibrium return

The optimal labor policy for the firm is dynamic and set at the beginning of each time period. The firm selects the number of job openings (J_t) and the subsequent period size of the workforce (N_{t+1}).

The literature on production-based asset pricing outlines two similar approaches to solve the maximization problem in Equation (1.5). The first approach involves specifying an explicit stochastic discount factor ($M_{t+\tau}$) as a function of shocks to aggregate productivity (x_t) (e.g., Kuehn et al. (2017) and Belo et al. (2014)). The second approach, which is followed here, does not explicitly assume a functional form for the stochastic discount factor and uses Lagrangian multipliers (e.g., Liu et al. (2009), Chen and Zhang (2011), and Petrosky-Nadeau et al. (2018)).

Following this approach, the first order conditions with respect to J_t and N_{t+1} for the firm's maximization problem are given by,

$$a \left(\frac{\lambda_t J_t}{N_t} \right) = E_t \left(M_{t+1} e^{x_{t+1}} - W_{t+1} + \frac{a}{2} \left(\frac{\lambda_{t+1} J_{t+1}}{N_{t+1}} \right) - (1-s) Q_{t+1} \right) = Q_t, \quad (1.6)$$

where Q_t and Q_{t+1} are the Lagrangian multipliers. The term left of the first equality represents the marginal adjustment cost associated with changing the size of the workforce in time period t . This cost is incurred at the end of time t regardless of the actual time of hiring, which may fall between the end of time period $t-1$ and the beginning of time period t . The term left of the second equality represents the discounted value of the marginal benefit associated with a change in the size of the labor force in time period $t+1$. In the partial equilibrium defined by the first-order

conditions, the optimal labor policy dictates that the marginal costs in the current time period equal the discounted value of the marginal benefit in the next time period.

Assuming constant returns to scale and combining the first-order conditions with the basic asset-pricing equation $E_t(M_{t+1}R_{t+1}) = 1$ leads to the following relation for expected return:

$$E_t[R_{t+1}] = \frac{E_t(\Pi_{t+1}) + \frac{a}{2}E_t\left(\frac{N_{t+2}}{N_{t+1}}\right)^2}{a\lambda_t\left(\frac{J_t}{N_t}\right)}, \quad (1.7)$$

where $E_t(\Pi_{t+1}) = E_t(e^{x_{t+1}} - W_{t+1} - F - \frac{a}{2}(1-s)^2)$. This derivation follows Liu et al. (2009) and is outlined in Appendix A.1. The term $E_t(\Pi_{t+1})$ in the numerator captures the expected value of the marginal output from labor at time $t+1$ ($e^{x_{t+1}}$) minus the marginal production costs. These production costs are given by the sum of the wages paid (W_{t+1}), the fixed costs of production (F), and the continuation value of employment net of separation costs ($\frac{a}{2}(1-s)^2$). The other term in the numerator ($\frac{a}{2}E_t(N_{t+2}/N_{t+1})^2$) represents the expected value of savings obtained from not employing an additional unit of labor. The entire numerator, thus, sums to the expected value of the subsequent period marginal benefits coming from the decision to change the size of the labor force.

In the denominator, $\frac{J_t}{N_t}$ is equal to the ratio of the number job openings to the size of the labor force and is deemed as the “job openings rate.” The product in the denominator represents the marginal costs associated with the workforce in time t . As such, the expected value of the subsequent period expected return of the firm is given by the ratio of the expected marginal benefits in the subsequent time period to the marginal costs of the changes in the size of the labor force in the current time period. Thus, Equation (1.7) implies that regressing the expected returns on job

openings rate should yield negative slope as the expected return decreases with the job openings rate.

A positive productivity shock to the economy requires firms to increase the size of labor inputs by posting more job openings. Increasing the labor inputs results in higher adjustment costs and therefore lower expected returns. As such, the relation between job openings rate and expected returns is attributable to the presence of search frictions. In an economy without search matching frictions, firms adjust the labor inputs with perfect elasticity and no the exists relation between expected returns and job openings rate.

1.3 Data and summary statistics

This section describes the data sources, market returns, construction of predictor variables, summary statistics, and relation between *LJOR* to other labor market characteristics.

1.3.1 Aggregate stock market returns

I measure the aggregate stock market excess return as the return for the S&P 500 index in excess of the return for the one-month U.S. Treasury bill. The returns data for the S&P 500 index returns and the one-month treasury bill are obtained from Amit Goyal's website: <http://www.hec.unil.ch/agoyal/>. The sample period extends from 1951 to 2016.

1.3.2 Job openings rate

I form the monthly time series of the job openings rate using a combination of three data sources for the aggregate job openings in the U.S. economy over different

time periods from 1951 to 2016. From January 1951 to December 1994, aggregate job openings are measured with the Conference Boards Help Wanted Index (HWI). This index provides a measure of the number of job advertisements in 51 major newspapers and is published monthly. Starting in late 1990s, internet advertising has become more prevalent than newspaper advertising. In response to this trend, the Conference Board started providing the Help Wanted Online Index (online HWI) in January 1995. Barnichon (2010) outlines a method to construct a composite HWI index that combines information from the HWI index and the HWI-online index. This composite index is used to measure the aggregate number of job openings from January 1995 to December 2000. Cajner and Ratner (2016) note that in the post-2000 time period, the composite HWI series diverges from the data published by the Job Openings and Labor Turnover Survey (JOLTS) conducted by the Bureau of Labor Statistics. The data from JOLTS starts in January 2000 and is more reliable as it directly comes from survey of a representative sample and does not rely on the newspaper advertisements. Therefore, for the time period from January 2000 to December 2016, the aggregate number of job openings is measured using the JOLTS data.⁴

I obtain a monthly time series measure of the job openings rate by dividing the aggregate number of job openings by the size of the labor force. Following Kuehn et al. (2017), I lag the job openings rate time series by one month so that the information on the number of job openings and the size of the workforce is publicly available.⁵

⁴This choice is similar to that of Regis Barnichon, who solely relies on JOLTS data in the post-2000 time period.

⁵The monthly time series of the job openings rate can be obtained from Regis Barnichon's website: <http://sites.google.com/site/regisbarnichon/>.

The main predictor variable in this paper, $LJOR$, is simply the log of job openings rate.

Figure 1.1 plots the monthly time series of the job openings rate from February 1951 to December 2016. The shaded regions correspond to the NBER-dated recessions. The job openings rate is pro-cyclical and appears to lead business cycles. During expansions, the job openings rate increases steadily with the growing economy and reaches a peak just prior to the start of recessions. The job openings rate sharply starts declining just prior to beginning each of the 11 recessions from 1951 to 2016. The job openings rate then tends to sharply decrease and reach a trough towards the end of the recessions.

1.3.3 Benchmark predictor variables

I compare the predictive power of the job openings rate with that of 16 other return predictor variables. This section describes these benchmark predictor variables.

I begin with the 14 monthly predictor variables from Goyal and Welch (2008). These predictors are dividend to price ratio (DP), dividend yield (DY), earnings to price ratio (EP), dividend payout ratio (DE), stock return volatility ($RVOL$)⁶, book to market ratio (BM), net equity expansion ($NTIS$), Treasury bill rates (TBL), long-term government bond yield (LTY), long-term government bond returns (LTR), term spread (TMS), default-yield spread (DFY), default-return spread (DFR), and inflation ($INFL$). This group also constitutes the benchmark predictor variables used in Rapach et al. (2016). Monthly data on the predictors from Goyal and Welch

⁶Instead of using $RVOL$, Goyal and Welch (2008) use the sum-squared daily excess stock returns (VOL). However, Rapach et al. (2016) note that this predictor produces a severe outlier in October of 1987.

(2008) are from Amimt Goyal’s website: <http://www.hec.unil.ch/agoyal/>. I also extend this list of predictors with two more predictors: output gap (*OG*) and nearness to historical Dow Jones Industrial Average index high (*PMAX*).⁷ The details of construction of the 16 benchmark predictors is available in Appendix A.2.

1.3.4 Summary statistics

Table 1.1 reports summary statistics for the predictor variables. The sample period is from February 1951 to December 2016. Panel A presents the summary statistics for the job openings rate (*JOR*) and log job openings rate (*LJOR*). The job openings rate has a mean (median) of 3.11% (3.07%), which means approximately three job openings per 100 workers in the economy. Also note that the standard deviation of *JOR* is 0.26 and the 1st and 99th percentiles are 0.41 (1.5) and 1.56 (4.78). Thus, *JOR* appears to show a substantial variation from the mean.

Panel B shows summary statistics for the 16 benchmark predictor variables. Consistent with previous literature (e.g., Goyal and Welch (2008)), I find that most of the benchmark predictors exhibit very high autocorrelation. The only exceptions are *LTR* and *DFR*, which have autocorrelation coefficients of 0.05 and -0.08 , respectively. The autocorrelation coefficients for the other benchmark predictor variables range from 0.55 for *INFL* to 0.99 for *DP*, *EP*, *DE*, *BM*, *TBL*, and *LTY*. The autocorrelation coefficients for *LJOR* and *JOR* are 0.98 and 0.99, respectively. As such, similar to the benchmark predictor variables, *JOR* and *LJOR* also appear to be quite persistent.⁸

⁷I review papers on return predictability published in *Journal of Finance*, *Journal of Financial Economics*, and *Review of Financial Studies* from 2006 to 2018 and select the predictor variables that are at monthly frequency and available from 1951 to 2016.

⁸Stambaugh (1999) points that persistent predictor variables can lead to spurious evidence of

Column 8 of Table 1.1 reports the time series correlations between *LJOR* and the benchmark predictor variables. The three highest correlation coefficients in absolute value are those for *TBL* (0.56), *TMS* (0.54), and *EP* (0.38). The absolute values of the correlation coefficients between *LJOR* and the benchmark predictor variables range from 0.04 to 0.56. Therefore, even though *LJOR* is correlated with some of the benchmark predictor variables, it also shows substantial independent variation.

1.3.5 *LJOR* and other labor market variables

In this section I investigate the relation between *LJOR* and other variables related to labor market conditions. The analysis in this section is useful in determining whether *LJOR* exhibits substantial variation that is not captured by other labor market related variables. The other labor market variables I consider are

1. Pay growth: The month-over-month percent growth in non-farm payrolls. The data on non-farm payrolls are obtained from the Bureau of Labor Statistics (BLS).
2. Unemployment rate: The monthly time series of the unemployment rate is obtained from the BLS.
3. Hiring rate: Following Belo et al. (2014), the hiring rate at time t is calculated as $(\frac{N_t - N_{t-1}}{0.5(N_t + N_{t-1})})$, where N_t represents the number of employed workers in the economy at time t . The data on the number of employed workers are obtained from Current Employment Statistics (CES) provided by the BLS.

return predictability in small samples. I address this issue using bootstrap p -value and temporal instability tests in the next section.

I regress the percentage changes in *LJOR* on the percentage changes in the three other labor market variables. The results are reported in Table 1.2. The first three columns correspond to univariate regressions, and the last column corresponds to a multivariate regression. All of the variables are standardized to have a zero mean and one unit standard deviation. Changes in unemployment rate are significantly negatively related to changes in *LJOR*, but changes in pay growth and hiring rate do not show a significant relation with changes in *LJOR*. The last row of Table 1.2 reports the adjusted R^2 from the regressions. These adjusted R^2 are substantially low and range from -10.80% for univariate regression with hiring rates to 6.31% for the multivariate regression. As such, the results from Table 1.2 show that *LJOR* exhibits substantial variation that is not captured by other labor market variables.

1.4 Return predictability results

In this section, I conduct in-sample and out-of-sample tests to assess the predictive power of *LJOR*. I also compare *LJOR* to the 16 benchmark predictor variables and test if the predictive information contained in *LJOR* is distinct from the predictive information contained in the benchmark predictor variables. In addition, only for the in-sample tests in Section 1.4.1 and Section 1.4.2, I standardize all the predictor variables to have zero mean and one unit standard deviation. This transformation facilitates straightforward comparisons of predictive ability as one can directly compare estimated values of the slope coefficients on the predictor variables from the standard return predictability regressions.

1.4.1 In-sample analysis

I employ the standard predictive regression framework from in the literature on return predictability (e.g., Fama and French (1989)) to conduct the in-sample tests,

$$R_{t:t+h} = \beta_0 + \beta_1 x_t + \epsilon_{t:t+h}. \quad (1.8)$$

In Equation (1.8), x_t is the value of the predictor variable at the end of month t and the dependent variable $R_{t:t+h}$ is calculated following Rapach et al. (2016) as

$$r_{t:t+h} = \frac{r_{t+1} + r_{t+2} + \dots r_{t+h}}{h} \quad (1.9)$$

$$R_{t:t+h} = \exp(r_{t:t+h}) - 1,$$

where r_t is the monthly log return on the S&P 500 index in excess of the risk-free rate from the end of month $t - 1$ to the end of month t , and h is the forecast horizon in months. The parameter of interest is the estimated slope coefficient on the predictor variable, β_1 .

The theoretical model in Section 1.2 suggests that the job openings rate is negatively associated with expected returns. Therefore, when Equation (1.8) is estimated using *LJOR* as the predictor variable, the expected sign of the slope coefficient is negative. Inoue and Kilian (2005) argue that when the expected sign on the slope coefficient is guided by theory, a one-sided test for statistical significance is more powerful than a two-sided test. I follow their recommendation and test the null hypothesis that the slope coefficient on *LJOR* is equal to zero against the alternate hypothesis that the slope coefficient on *LJOR* is less than zero.

Several issues that complicate the statistical inference on slope coefficient in Equation (1.8). First, to achieve greater efficiency in estimation Equation (1.8) is estimated using overlapping observations of monthly data. When $h > 1$, that the standard er-

rors from the estimation of the above predictive regression suffer from a downward bias (Nelson and Kim (1993)). To address this issue, I use Newey and West (1987) standard errors with a lag of h correct for heteroscedasticity and autocorrelation.

Second, Stambaugh (1999) points that, when x_t is persistent, a researcher is more likely to find spurious evidence of return predictability, especially in small samples. I follow Rapach et al. (2016) and Huang et al. (2015) and use a bootstrap procedure that accounts for predictor persistence. The bootstrap procedure uses standard errors simulated under the null hypothesis of no return predictability and high predictor persistence. The wild p -values obtained from this procedure are corrected for persistence in the predictor variable, correlations between the dependent variable and innovations in the predictor variables, and non-normal return distributions.⁹

To further alleviate concerns regarding persistence, I conduct the IVX Wald-test developed by Kostakis et al. (2014). This test presents a useful way of inferring statistical significance when the predictor variable is highly persistent and is robust to the presence of unit root, local-to-unit root, near stationarity, or stationarity. In the IVX procedure, I construct an artificial instrumental variable whose degree of persistence can be explicitly controlled. Kostakis et al. (2014) argue that using such an artificial instrument in a standard instrumental variables estimation alleviates concerns about the quality of inference. The Wald-statistic associated with the estimator derived from this procedure asymptotically follows a chi-squared distribution under the null hypothesis of no return predictability.

Third, Goyal and Welch (2008) and Kostakis et al. (2014) argue that return predictability is often spurious and varies widely with the choice of sample period. Both

⁹The details of this procedure can be found in Appendix A.2 of Huang et al. (2015).

of these papers shed doubt on the stability of forecasting models documented in the return predictability literature. Similarly, Paye and Timmermann (2006) point out that when predictor variables are highly persistent, return prediction models can suffer from temporal instability. They suggest that the $\widehat{qLL}$ statistic developed by Elliott and Müller (2006) provides a useful way to test for temporal instability. Following their directive, I estimate $\widehat{qLL}$ statistic for test of the null hypothesis that the slope coefficient on the predictor variable does not suffer from structural breaks throughout the sample period against the alternate hypothesis that the value of the slope coefficient changes throughout the sample period.

The in-sample estimation results of are reported in Table 1.3. The predictor variable is *LJOR* and the sample period is from February 1951 to December 2016. The table shows estimations for the forecast horizons (h) of one, three, six, 12, and 24 months. Column 2 shows the estimated values of the intercept, and column 3 shows the estimated values of the slope coefficients on *LJOR*. The values in parentheses are Newey and West (1987) t -statistics. For each forecast horizon, the estimated values of slope coefficients are economically large and range from -0.50 (for monthly horizon) to -0.28 (for bi-annual horizon). The slope coefficient can be interpreted as the percent decrease in expected returns in the subsequent period for a one unit standard deviation increase in *LJOR*. The estimated slope coefficients are also highly significant as shown by the t -statistics, which range from -3.23 (for monthly horizon) to -3.72 (for bi-annual horizon).

The values for the in-sample R^2 statistic from the estimation of Equation (1.8) for different horizons are reported in column 4 of Table 1.3. For the monthly horizon, the

in-sample R^2 is 1.44%. Consistent with the literature on return predictability, the in-sample R^2 monotonically increases with the forecasting horizon (e.g., Cochrane (2009)). For the annual and bi-annual horizons *LJOR* generates in-sample R^2 of 8.91% and 9.58%, respectively. The economic significance of the estimated R^2 can be assessed using the guideline of Campbell and Thompson (2008) who note that a monthly R^2 of 0.50 % represents a substantial economic value from an investor's perspective. *LJOR* generates a monthly R^2 of almost three times this value.

Column 5 reports the p -values obtained from the wild bootstrap procedure. All of the p -values in this column are less than 0.01 indicating that the slope coefficients on *LJOR* are significant at the 1% level. Column 6 reports the p -values from the IVX Wald-test developed by Kostakis et al. (2014) and further confirms the above inference regarding the statistical significance of the slope coefficient. These p -values are below 0.05 for all the horizons. As such, these results also suggest strong significance of the slope coefficient.

Columns 7 and 8 columns show results of the temporal instability tests. The $\widehat{qLL}$ statistics are reported in column 7 and the corresponding critical values at the 10% significance level are in column 8.¹⁰ If the $\widehat{qLL}$ statistic is less than the critical value, it is deemed significantly different from zero indicating temporal instability in the forecasting model.¹¹ For each horizon, the estimated value of the $\widehat{qLL}$ statistic is greater than the critical value of -7.04 , and the null hypothesis of no structural breaks in slope coefficient cannot be rejected. In other words, I find no substantial evidence

¹⁰The critical values at 1% and 5% significance levels are -11.05 and -8.36 . All the critical values are obtained from Table 1 of Elliott and Müller (2006).

¹¹Note that using a 10% significance level here is more conservative than the typical 5% or 1% significance levels.

that the predictive ability of *LJOR* changes over the February 1951 to December 2016 sample period.

1.4.2 Comparison with other predictors

In this section, I compare the in-sample predictive power of *LJOR* with that of the 16 benchmark predictor variables. Estimates of Equation (1.8) using the benchmark predictor variables are reported in Table 1.4. The results in Panel A (Panel B) correspond to a monthly (annual) forecasting horizon. For brevity, I only report results for estimations for these two horizons. However, results are similar for all the horizons listed in the previous section. The first two rows in each panel present results of estimations from Table 1.3 in which *LJOR* is used as the predictor variable. Baker and Wurgler (2000), Campbell (1987), Fama and French (1989), Fama (1981), Cooper and Priestley (2008), and Li and Yu (2012) document that *NTIS*, *TBL*, *LTY*, *INFL*, *OG*, and *PMAX* are negative predictors of returns. Hence, the expected sign on the slope coefficients of these predictor variables, as noted in column 2, is negative. Column 3 reports estimated slope coefficients on the predictor variable and their Newey and West (1987) *t*-statistics. The significance of the slope coefficients is evaluated using a one-sided test corresponding to the expected sign.

For the monthly horizon (Panel A), the estimates of the slope coefficients on the benchmark predictor variables range from -0.34 (for *TBL*) to 0.38 (for *LTR*). In comparison, the estimated slope coefficient on *LJOR* of -0.50 , which in absolute value, exceeds the estimated slope coefficients for the benchmark predictor variables. Of the 16 benchmark predictor variables, the slope coefficient estimates on *DP*, *RVOL*, *TBL*, *LTR*, and *TMS* are significant at the 5% level. The *t*-statistic for the slope

coefficient on *LJOR* is also greater than those for the benchmark predictor variables. Thus, in both statistical and economic terms, *LJOR* outperforms all of the 16 benchmark predictor variables. Column 4 in Panel A reports estimates of the in-sample R^2 statistics. For the benchmark predictor variables, these estimates range from 0.01% to 0.83%. In comparison, the in-sample R^2 for *LJOR* is 1.44% which is more than one and a half times the largest value obtained from the benchmark predictor variables.

The last four columns report estimation results for the bivariate regression:

$$R_{t:t+h} = \beta_0 + \beta^{BP} BP_{i,t} + \beta^{LJOR} LJOR_t + \epsilon_{t:t+h}, \quad (1.10)$$

where $BP_{i,t}$ is the benchmark predictor variable i at the end of month t . This regression is useful in determining whether the benchmark predictor variables subsume the predictive ability of *LJOR* and vice versa.

The estimated values of the slope coefficients on the benchmark predictor variable and *LJOR* are reported in column 5 and column 6. Only five out of the 16 benchmark predictor variables (*DP*, *EP*, *RVOL*, *BM*, and *LTR*) have estimated slope coefficients that are significantly greater than zero at the 10% significance level. In contrast, for all 16 cases, the estimated slope coefficient on *LJOR* is significant at the 1% level. The estimated values of the slope coefficient on *LJOR* range from -0.45 to -0.66 , which are close to the 0.50 estimate from the univariate regression with *LJOR*.

The last two columns report the R^2 and the partial- R^2 for *LJOR* from the estimation of the bivariate regression. The partial- R^2 can be interpreted as the R^2 contributed by *LJOR* in the bivariate predictive regression. The estimated values of the R^2 from the bivariate regression range from 1.44% to 2.27%. In comparison, the R^2 estimates from the univariate regressions that use only the benchmark predictor

variables range from 0.01% to 0.83%. Therefore, adding *LJOR* drastically improves the performance of the predictive regressions that present the benchmark predictor variables. Furthermore, the estimated values of the partial- R^2 attributable to *LJOR* range from 0.82% to 2.13% and are considerably close to those obtained from a univariate regression with *LJOR*.

In Panel B, I reestimate Equation (1.8) for the annual horizon ($h=12$). For the 16 benchmark predictor variables, the estimates of the slope coefficients and in-sample R^2 range from -0.24 to 0.29 and 0.01% to 4.83% , respectively. In comparison, the slope coefficient and in-sample R^2 estimates for *LJOR* are -0.39 and 8.91% , respectively. Panel B also reports the bivariate regression results for the annual horizon. In all cases, the slope coefficient on *LJOR* is significantly greater than zero. These estimates range from -0.33 to -0.55 and are close to the estimate of -0.39 from the univariate specification. These results suggest that including benchmark variables in the predictive regressions does very little to reduce the predictive power of *LJOR*.

1.4.3 Out-of-sample analysis

Goyal and Welch (2008) point out that most of the predictor variables documented in the return predictability literature suffer from systematically poor out-of-sample performance. In this section, I implement their testing methods to examine if the in-sample predictive power of *LJOR* extends to the out-of-sample setting.

The out-of-sample exercise uses a training period from the start of the sample period (that is, February 1951) to the end of time period $t - 1$ and recursively obtains a return forecast for the subsequent time period as

$$\hat{R}_{t:t+h} = \hat{\beta}_0 + \hat{\beta}_1 x_t, \quad (1.11)$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the coefficients estimated using the predictive regression in Equation (1.8) over the training period and $\hat{R}_{t:t+h}$ is the subsequent period return forecast.

To evaluate the performance of the return forecasts, I compute the out-of-sample R^2 (R_{OS}^2) as

$$R_{OS}^2 = 1 - \left(\frac{\sum_{t=1}^T (R_{t:t+h} - \hat{R}_{t:t+h})^2}{\sum_{t=1}^T (R_{t:t+h} - \bar{R}_{t:t+h})^2} \right), \quad (1.12)$$

where $\bar{R}_{t:t+h}$ represents the historical average of $R_{t:t+h}$ over the training period (i.e., the prevailing mean forecast). The numerator in Equation (1.12) represents the sum of square errors from the forecast obtained using the predictor variable under consideration, and the denominator represents the sum of square errors from the prevailing mean forecast. As Clark and West (2006) and Campbell and Thompson (2008) point out, the expected value of the out-of-sample R^2 under the null hypothesis of no predictability is negative and a zero out-of-sample R^2 is evidence of weak predictability. In contrast, a positive out-of-sample R^2 implies that the predictor variable has a lower mean square forecast error than does the prevailing mean.¹²

Statistical significance of the out-of-sample R^2 estimate is assessed using the Clark and West (2007) statistic (CW statistic), which follows a standard normal distribution. This statistic corresponds to the test of the null hypothesis that the mean square forecast error (MSFE) for the prevailing mean forecast is equal to the MSFE for the forecasts obtained from the predictor variable under consideration.

¹²In calculating $\bar{R}_{t:t+h}$, Campbell and Thompson (2008) use all available returns data starting from 1871 even though the data on given predictor variable may be available in the later part of the sample period. I follow the general convention in the predictability literature (e.g., Goyal and Welch (2008)) and use data from the start of the sample period used in the in-sample tests (that is, February 1951). This choice does not affect the results.

I use data from February 1951 to December 1965 to obtain the initial coefficient estimates and forecast using Equation (1.11) and restrict the out-of-sample testing period from January 1966 to December 2016. Hansen and Timmermann (2015) note that the choice of out-of-sample testing period impacts the power of out-of-sample tests and that researchers must balance between having a reasonably long time series to obtain a reliable first forecast and having a long enough time series to conduct a sufficiently powerful out-of-sample test. They conduct a simulation experiment to show that the power of out-of-sample tests increases with the length of the forecast evaluation window. Following their directive, I use about 15 years of data as the initial training period so that the ratio of the out-of-sample period to the full-sample period is approximately 20%.

Figure 1.2 shows time series plots of out-of-sample forecasts obtained using *LJOR* (solid line), and the prevailing mean forecast (dashed line). The actual returns (dotted line) over the out-of-sample period are also shown. The forecasts are plotted for the annual horizon ($h=12$), and all values are in percentage points. The prevailing mean forecast is almost constant throughout the entire out-of-sample period. The forecast obtained using *LJOR*, however, shows significant variation that is in line with the actual returns in the subsequent time period. In most cases, the forecasts obtained using *LJOR* are quick to adjust to sharp changes in market conditions and, therefore, appear to show a substantial advantage over the prevailing mean forecast.

Table 1.5 reports results from the out-of-sample tests. Column 1 presents the predictor variable that is used to obtain the return forecast using Equation (1.12). Column 2 reports the out-of-sample R^2 estimates for the monthly horizon, and the

values in parentheses are CW statistics. The out-of-sample R^2 for *LJOR* at this horizon is 0.86% and is statistically significant at the 1% level with a CW statistic of 2.51. In comparison, the estimates of the out-of-sample R^2 s for the benchmark predictor variables range from -1.26% to 0.35% , and only of 16 estimates are positive. These results that show poor out-of-sample performance of the benchmark predictor variables is consistent with the findings of Goyal and Welch (2008). The out-of-sample R^2 estimate for *LJOR* is more than twice the highest estimate for the benchmark predictor variables.

For a direct comparison between the in-sample and out-of-sample R^2 estimates, I reestimate Equation (1.8) over the out-of-sample period from January 1966 to December 2016. These estimates for the monthly horizon are reported in column 3 of Table 1.5. For *LJOR*, the in-sample R^2 for the out-of-sample period is 0.94%, which is close to the out-of-sample R^2 of 0.86%. Most of the benchmark predictor variables, however, show large differences in the two R^2 estimates. For example, the only two positive estimates for the out-of-sample R^2 are 0.12% for *RVOL* and 0.35% for *LTR*. In comparison, the in-sample R^2 estimates over the out-of-sample period for these two variables are 0.77% and 1.00%.

The next two columns in Table 1.5 report results of out-of-sample tests for the annual horizon. The out-of-sample R^2 for *LJOR* is 7.30% with a CW statistic of 2.59 and is statistically significant at the 1% level. In comparison, the out-of-sample R^2 estimates for the benchmark predictor variables range from -13.67% for *BM* to 1.83% for *INFL*. Only five of the 16 benchmark predictor variables generate an out-of-sample estimate that is greater than zero. The corresponding CW statistics show that

only three variables have out-of-sample R^2 estimates that are statistically significant at the 10% level. Finally, the in-sample R^2 over the out-of-sample period for *LJOR* is 6.39%, which is close to the out-of-sample R^2 of 7.3%.

1.4.4 Campbell and Thompson (2008) restrictions

Campbell and Thompson (2008) show that imposing theoretically motivated restrictions on the out-of-sample forecasts can substantially improve the forecasting performance. In this section, I re-conduct the out-of-sample analysis imposing their proposed restrictions to analyze: (1) Do these restrictions have any effect on the predictive power of *LJOR*? (2) is the improvement in performance of the benchmark predictors due to these restrictions substantial enough for them to outperform *LJOR*?

Campbell and Thompson (2008) emphasize two restrictions. First, they restrict the signs of the slope coefficients used in the calculation of the out-of-sample forecasts that are consistent with theory. Following this, in Equation (1.11), I restrict $\hat{\beta}_1$ to be consistent with the expected sign on the slope coefficient reported in Table 1.4 (slope restriction). Any $\hat{\beta}_1$ estimates that are inconsistent with the expected sign are set to zero. Second, Campbell and Thompson (2008) also advocate imposing a positive constraint on the forecasted aggregate stock market return ($\hat{R}_{t:t+h} > 0$) (forecast restriction). Any forecast that predicts $\hat{R}_{t:t+h} < 0$ is inconsistent with the theoretical expectation of a positive equity risk premium and is, therefore, set to zero.¹³

Table 1.6 reports results from the out-of-sample tests with the slope and forecast restrictions. I begin with results for the monthly horizon, which are reported in

¹³In the tests conducted in this section, I do not consider Campbell and Thompson (2008)'s restriction on the intercept, $\hat{\beta}_0 > 0$. Imposing a positive restriction on the intercept estimates has little impact on the out-of-sample performance of *LJOR* and the benchmark predictor variables and does not change the conclusions from the results.

Panel A. The first two columns present results of the out-of-sample tests in which only the slope restriction is applied. Imposing this restriction has no effect on the predictive ability of *LJOR*, as the out-of-sample R^2 remains 0.86% with a CW statistic of 2.51. Because, the estimated slope coefficient for *LJOR* in Equation (1.11) is always positive, the slope restriction is never enforced and does not affect the out-of-sample performance of *LJOR*. In comparison, imposing the slope restriction has a positive effect on the performance of five of the 16 benchmark predictor variables. For eight of the 16 benchmark predictor variables, performance remains unchanged. The largest performance improvement is observed in *DFR*, whose out-of-sample R^2 increases by 0.31% (from -0.55% to -0.24%). Nevertheless, the out-of-sample performance of the benchmark predictor variables remains poor as only two out of the 16 benchmark variables have positive out-of-sample R^2 estimates.

The next two columns report results of the out-of-sample analyses in which the forecasted return $\hat{R}_{t:t+h}$ is constrained to be greater than zero. The restriction is enforced for 23% of the out-of-sample testing period and, thus, plays an important role in revising the forecasts obtained from *LJOR*. Imposing this restriction has a slightly positive effect on the out-of-sample R^2 of *LJOR*, with the 0.86% estimate improving to 0.87%. In comparison, the forecast restriction significantly improves the performance of 11 of the 16 benchmark predictor variables. The largest performance improvement is observed for *LTY* whose out-of-sample R^2 increases from -0.83% to 0.63%. For this predictor variable, the forecast is restricted to be greater than zero for 36% of the out-of-sample testing period. However, even with the performance improvements from the forecast restriction, out-of-sample R^2 estimates for the benchmark variables

range from -1.26% to 0.63% with only four of 16 estimates being positive. As such, none of the benchmark predictor variables outperforms *LJOR*.

The last three columns in Panel A show results from the out-of-sample tests with both the slope and forecast restrictions imposed jointly. The estimates of out-of-sample R^2 are slightly better than those when only the forecast restriction is applied. The out-of-sample R^2 estimate for *LJOR* is 0.87% with a CW statistic of 2.37 and is significant at the 1% level. For the benchmark predictor variables, 12 out of the 16 out-of-sample R^2 estimates show improvement when compared to the estimates from unconstrained out-of-sample tests from Table 1.5. The estimated out-of-sample R^2 s for the benchmark predictor variables range from -1.06% to 0.63% . Despite the improved performance from the slope and forecast restrictions, none of the benchmark predictor variables outperforms *LJOR*.

Panel B of Table 1.6 reports results of the out-of-sample tests at the annual forecasting horizon. Throughout the testing period, the slope coefficient $\hat{\beta}_1$ for *LJOR* is always greater than zero. Therefore, the slope restriction has no effect on the out-of-sample performance of *LJOR*. For the benchmark predictor variables, the slope restriction has a positive effect on the out-of-sample R^2 for nine of the 16 benchmark predictor variables. However, only four out of the 16 benchmark predictor variables have positive out-of-sample R^2 estimate and none of them are higher than that of *LJOR*.

Next, I consider the forecast restriction. For *LJOR*, this restriction is enforced for 11% of the testing period, and imposing this restriction, improves the out-of-sample R^2 from 7.30% to 8.88% . The main reason behind this increase can be seen

in the predicted forecasts from September 1977 to May 1980 in Figure 1.2. The predicted forecast from *LJOR* is negative during this time period. During the same time period, the realized market returns, however, are consistently positive. The forecast restriction improves the performance of the out-of-sample forecasts obtained using *LJOR* by restricting the perverse prediction of a negative equity risk premium during this time period to be equal to zero. The forecast restriction also improves the performance of 11 out of the 16 benchmark predictor variables which range from -13.73% to 6.36% . However, even with the improved performance, the benchmark predictor variables fail to outperform the out-of-sample performance of *LJOR*.

The last three columns in Panel B show results in which the slope and the forecast restrictions are simultaneously imposed. Of the 16 benchmark predictor variables, 13 show increases in the out-of-sample R^2 in comparison to the unconstrained case. The estimated out-of-sample R^2 for the benchmark predictor variables range from -13.83% to 6.36% with eight estimates being greater than zero. None of the benchmark predictor variables, however, outperform *LJOR*.

1.4.5 Out-of-sample encompassing tests

This section presents forecast encompassing tests that analyze if the predictive information contained in *LJOR* forecasts is subsumed by that in the forecasts from the benchmark predictor variables.

The encompassing tests are conducted using the method developed by Harvey et al. (1998). In these tests, an optimal forecast is obtained from a convex combination of the two individual forecasts, one from *LJOR* and one from a benchmark predictor variable. The optimal combination is given as a weighted average of the two individual

forecasts:

$$\hat{R}_{t:t+h}^C = (\lambda_{BP,i})\hat{R}_{t:t+h}^{BP,i} + (\lambda_{LJOR})\hat{R}_{t:t+h}^{LJOR}, \quad (1.13)$$

where $\hat{r}_{t:t+h}^{BP,i}$ is the forecast obtained using the benchmark predictor i , $\hat{r}_{t:t+h}^{LJOR}$ is the forecast obtained using $LJOR$, $\hat{r}_{t:t+h}^C$ is the optimal combination forecast, λ_{LJOR} is the weight for $LJOR$ the forecast, and $\lambda_{BP,i}$ is the weight for benchmark predictor's forecast. Both λ_{LJOR} and $\lambda_{BP,i}$ are restricted such that, $0 \leq \lambda_{LJOR} \leq 1$, $0 \leq \lambda_{BP,i} \leq 1$, and $\lambda_{LJOR} + \lambda_{BP,i} = 1$.

If $\lambda_{LJOR} = 0$, then the optimal combination forecast is the forecast from the benchmark predictor variable. In this case, the predictive forecast obtained from the benchmark predictor variable completely encompasses the forecast obtained using $LJOR$, and $LJOR$ does not contain any additional predictive information. On the contrary, if $0 < \lambda_{LJOR} < 1$, then the optimal combination forecast is formed using components of both forecasts. In this case, both the predictor variables contain distinct predictive information. If $\lambda_{LJOR} = 1$, then the optimal combination forecast is equal to the forecast obtained from $LJOR$, and the forecast obtained using $LJOR$ completely encompass the forecast obtained from the benchmark predictor variable.

The encompassing tests are summarized in Table 1.7. The weights on $LJOR$ and the benchmark predictor variables are in the first two columns for the monthly horizon ($h=1$) and last two columns for the annual ($h=12$) horizon. The values in parentheses are the t -statistics developed by Harvey et al. (1998) (ENC statistic) for the test of the null hypothesis that the corresponding weight is equal to zero against the one-sided (upper-tailed) alternative hypothesis that the estimated weight is positive.

The weight estimate on the forecast obtained from $LJOR$ ranges from 0.55 to 0.85

for the monthly horizon and from 0.67 to 0.92 for the annual horizon. Because the weights on *LJOR* and on the benchmark predictor variables sum to one, the weights on the benchmark predictor variables are less than 0.50 in all 16 cases. Thus, the optimal combination forecasts rely more on the information contained in *LJOR* than on the information contained in the benchmark predictor variables.

The ENC statistics indicate that the estimated weight on *LJOR* is significantly greater than zero in all cases. These results provide further evidence that the predictive content of *LJOR* is not encompassed by that of the benchmark predictor variables. More importantly, the ENC statistics for the weight on benchmark predictor variables show that, for the monthly horizon, only five of the 16 estimates are significant at the 10% level and three at the 5% level. For the annual horizon, five of 16 are significant at the 5% level and eight at the 10% level.

1.5 Implications for asset allocation

The results from the previous section demonstrate that *LJOR* predicts aggregate stock market returns on an out-of-sample basis. In this section, I examine the economic value of the out-of-sample predictability from an asset allocation perspective.

Consider a mean-variance investor who uses the excess return forecasts ($\hat{R}_{t:t+1}$) obtained from Equation (1.11) and allocates between equities and a risk-free asset. The weight on equities in this portfolio is

$$w_t = \left(\frac{1}{\gamma}\right) \left(\frac{\hat{R}_{t:t+1}}{\hat{\sigma}_{t:t+1}^2}\right), \quad (1.14)$$

where γ is the risk aversion coefficient and $\hat{\sigma}_{t:t+1}^2$ is the forecasted variance of excess returns. The investor allocates the rest of his portfolio weight ($1 - w_t$) to the risk-free

asset. The portfolio return on such a strategy is given by $R_{p,t:t+1} = w_t R_{t:t+1} + R_{f,t:t+1}$. In this equation, $R_{t:t+1}$ is the excess return on the market from time the end of time t to the end of time $t+1$ and $R_{f,t:t+1}$ is the risk-free rate. Following Rapach et al. (2016) and others, the volatility forecasts are obtained using a 10-year moving window of monthly excess returns.¹⁴ In addition, following Campbell and Thompson (2008), w_t is constrained to be between 0 and 1.5. These restrictions impose no short-selling and limit portfolio leverage to a maximum of 50%. In addition, the investor is assumed to reestimate the weight on equities at the end of every month so that the portfolio is rebalanced on a monthly basis.

The certainty equivalent return (CER) of this portfolio is

$$CER = \hat{\mu}_p - 0.5\gamma\hat{\sigma}_p^2, \quad (1.15)$$

where $\hat{\mu}_p$ is the realized sample mean for the portfolio return and $\hat{\sigma}_p^2$ is the realized variance. This certainty equivalent return represents the risk-free rate of return that an investor would be willing to accept in lieu of holding the risky portfolio strategy described above. I follow Campbell and Thompson (2008) and Rapach et al. (2016) and use a risk aversion coefficient of three.

The portfolio rebalancing strategy formed using the forecast from a given predictor variable is benchmarked by a similar strategy that uses the prevailing mean forecast. The difference between the certainty equivalent returns of the two strategies is the CER gain. I annualize this difference by multiplying it by 12. This difference can be interpreted as the percentage points that an investor is willing to pay in management fees to have access to the return forecasts from the given predictor over the prevailing

¹⁴Campbell and Thompson (2008) use a five-year moving window of monthly excess returns to obtain the volatility forecasts. The results reported in this section are robust to this choice.

mean forecasts. The statistical significance of the CER gain is tested using the method of DeMiguel et al. (2007). They show that the asymptotic distribution of the CER gain ($f(v)$) is

$$\sqrt{T}(f(\hat{v}) - f(v)) \rightarrow \mathcal{N}\left(0, \frac{\partial f^T}{\partial v} \Theta \frac{\partial f}{\partial v}\right), \quad (1.16)$$

where T is the length of the sample period in months, the vector $v = (\mu_i, \mu_{PM}, \sigma_i^2, \sigma_{PM}^2)$, μ_i and σ_i^2 (μ_{PM} and σ_{PM}^2) are mean and variance of the monthly excess returns for a strategy that uses forecasts from a given predictor i (prevailing mean forecast), and

$$\Theta = \begin{pmatrix} \sigma_i^2 & \sigma_{i,PM} & 0 & 0 \\ \sigma_{i,PM} & \sigma_{PM}^2 & 0 & 0 \\ 0 & 0 & 2\sigma_i^4 & 2\sigma_{i,PM}^2 \\ 0 & 0 & 2\sigma_{i,PM}^2 & 2\sigma_{PM}^4 \end{pmatrix}. \quad (1.17)$$

Here, covariances $\sigma_{i,PM}$ is the covariance between the excess returns of a strategy that uses forecasts from a given predictor i and the prevailing mean strategy.

Column 2 in Table 1.8 reports the CER gains for each predictor variable. As in the out-of-sample tests in the previous sections, the sample period for the asset allocation analysis is from January 1966 to December 2016. The values in parentheses are the z -statistics for testing the null hypothesis that the CER gain is zero against the one-sided (upper-tailed) alternative hypothesis that the CER gain is greater than zero. The strategy that uses forecasts obtained from *LJOR* earns a CER gain of 3.36%. The portfolio rebalancing strategies that use the forecasts obtained from the benchmark predictor variables show CER gains that range from -1.00% to 3.17% . The three highest estimates of CER gains are obtained for strategies formed using *TMS* (3.17%), *TBL* (1.95%), and *LTR* (1.66%). Eight of these 16 strategies show a positive CER gain. However, this difference is only significant for six strategies at the 10% level. None of the CER gains obtained from the strategies formed using

the benchmark predictor variables is higher than the CER gain obtained from the strategy that uses forecasts based on *LJOR*.

The next column in Table 1.8 reports the difference in Sharpe ratios for each portfolio rebalancing strategy and for the prevailing mean strategy. These results allow for a direct measure of the economic value of each predictor that does not rely on an assumed value for the risk aversion coefficient. The values in parentheses are the t -statistics calculated using the method of Jobson and Korkie (1981) and are calculated as

$$\hat{JK} = \frac{\mu_i \sigma_{PM} - \mu_{PM} \sigma_i}{\sqrt{\nu}} \quad (1.18)$$

where the term in the denominator

$$\nu = \frac{1}{T} \left(2\sigma_i^2 \sigma_{PM}^2 - 2\sigma_i \sigma_{PM} \sigma_{i,PM} + \frac{1}{2}\mu_i^2 \sigma_{PM}^2 + \frac{1}{2}\mu_{PM}^2 \sigma_i^2 - \frac{\mu_i \mu_{PM}}{\sigma_i \sigma_{PM}} \sigma_{i,PM}^2 \right) \quad (1.19)$$

The strategy that uses forecasts obtained from *LJOR* shows a gain of 0.20 over the Sharpe ratio of the prevailing mean strategy. This gain has a t -statistic of 1.92 and is significant at the 5% level. In comparison, the Sharpe ratio gains of the 16 strategies that use forecasts obtained from the benchmark predictor variables range from -0.07 to 0.20 , the highest one being for the *TMS* strategy. Nine of the 16 strategies show positive Sharpe ratio gains. However, only the strategy formed using forecasts obtained from *TMS* (*LTR*) has a Sharpe ratio gain that is significant at the 5% (10%) level. Overall, these results show that none of the 16 strategies based on forecasts from benchmark predictors outperform the strategy formed using *LJOR*.

Figure 1.3 provides additional information regarding the performance of the portfolio strategies. Here, the solid line plots the weight on equities for the rebalancing strategies formed using forecasts obtained from *LJOR*, and the dashed line plots

the weight on equities in the benchmark strategy. The shaded vertical bars represent NBER recessions. Due to the smoothness of the prevailing mean forecasts, the weights on equities in the benchmark strategy are stable throughout the sample period. The *LJOR* strategy, on the other hand, shows significant variation in the portfolio weight allocated to equities. More importantly, the weight on equities for this strategy drops around the beginning of recessionary periods and increases towards the end of recessions. For example, the weight on equities in this strategy is 0.00 at the beginning of the recession in November 1973 and sharply increases to 1.50 towards the end of the recession around March 1975. During this time period, the weight for the prevailing mean strategy exhibits an opposite trend. It starts around 1.5 at the beginning of the recession and drops to around 1 towards the end of the recession.

In Figure 1.4, the solid line represents the log cumulative wealth plot for the strategy that uses forecasts obtained from *LJOR* and the dashed line is the log wealth plot for the benchmark strategy that uses the prevailing mean forecasts. Over the sample period, an investor following the strategy that uses *LJOR* forecasts would have done significantly better than an investor following the strategy that uses prevailing mean forecasts. This difference is even more apparent in the post-financial crisis time period. In this time period, the cumulative wealth for the *LJOR* strategy increases at a considerably faster rate than the cumulative wealth for the prevailing mean strategy.

I repeat the tests for economic significance of return predictability by subjecting the out-of-sample return forecasts to Campbell and Thompson (2008) restrictions. The portfolio rebalancing strategies are formed using the exact same assumptions described above with one exception. In the calculation of the return forecast for

the subsequent time period ($\hat{R}_{t:t+1}$) the estimated slope coefficient on the predictor variable and the resulting return forecast are restricted to be greater than zero. That is, I apply both the slope and forecast restrictions described in Section 1.4.4.

Table 1.9 presents the results. The slope and forecast restrictions have little impact on the performance of the portfolio rebalancing strategy that uses forecasts from *LJOR*. Similar to the results in Table 1.8, this strategy generates a CER gain of 3.36% (z -statistic of 4.82) and a Sharpe ratio gain of 0.20 (t -statistic of 1.92). Both the CER gain and the Sharpe ratio gain are significant at the 5% level. In comparison, the CER gains for the strategies that use forecasts obtained from the benchmark predictor variables range from -0.84% to 3.17% . The Sharpe ratio gains on the strategies that use forecasts obtained from the benchmark predictor variables range from -0.07 to 0.20 . Only the strategy formed using *TMS* has a Sharpe ratio gain that is equal to the Sharpe ratio gain of the strategy formed using forecasts obtained from *LJOR*. In all 16 cases, none of the strategies formed using the benchmark predictor variables forecasts outperform the strategy formed using *LJOR* forecasts.

1.6 Conclusion

In this paper, I examine the impact of search frictions in labor market on expected returns in financial markets. Using data from February 1951 to December 2016, I show that the job openings rate, defined as the ratio of number of job vacancies to the size of the workforce, is a strong negative predictor of aggregate stock market excess returns and generates an in-sample R^2 of 1.44% for the monthly horizon and 8.91% for the annual horizon. The predictive power of the job openings rate remains impressive in out-of-sample tests. I also find that the job openings rate outperforms

16 other benchmark predictor variables. In addition, using encompassing tests, I show that the predictive information contained in the job openings rate is distinct from that contained in the benchmark predictor variables.

Using a production-based asset pricing model that is augmented for search costs, I show that labor market frictions play an important role in predicting returns in financial markets. In the model, the predictive power of the job openings rate comes from the marginal costs of hiring activity of firms. These marginal costs proportionally increase with the steepness of turnover in the current labor force, which is captured by the ratio of the number of job openings to the size of the labor force (that is, the job openings rate). Consequently, the job openings rate is negatively associated with subsequent period returns.

These results have significant implications for the nascent but rapidly growing literature on production based asset pricing. Papers in this literature usually only consider capital inputs and financial investment frictions. My results build on this literature and show that labor is an important input for a firm's production and considering adjustment costs associated with labor market frictions can reveal important information about expected returns. The evidence presented in this paper calls for a careful consideration of labor market dynamics in the asset pricing studies.

Table 1.1: Summary statistics

This table provides summary statistics for the predictor variables used in the paper. Panel A presents summary statistics for the job openings rate (JOR), which is measured as the ratio of number of job openings to the size of the labor force in percentage points and log of job openings rate (*LJOR*). Panel B reports the same summary statistics for the benchmark predictor variables. The table reports mean, median, 1st percentile, 99th percentile, standard deviation, first-order autocorrelation coefficient or AR(1), and the correlation coefficient with *LJOR*. *LJOR*, *JOR*, *INFL*, and *OG* are lagged by one month for information on these variables to become publicly available. The sample period is from February 1951 to December 2016.

Panel A: Job openings rate							
Predictor (1)	Mean (2)	Median (3)	1 st Percentile (4)	99 th Percentile (5)	Standard deviation (6)	AR(1) (7)	Correlation with <i>LJOR</i> (8)
<i>LJOR</i>	1.10	1.12	0.41	1.56	0.26	0.98	1.00
<i>JOR</i> (%)	3.11	3.07	1.50	4.78	0.78	0.99	0.99
Panel B: Benchmark predictor variables							
<i>DP</i>	−3.52	−3.48	−4.45	−2.73	0.41	0.99	0.22
<i>DY</i>	0.00	0.00	−0.02	0.02	0.01	0.88	0.19
<i>EP</i>	−2.79	−2.83	−4.29	−1.99	0.42	0.99	0.38
<i>DE</i>	−0.73	−0.69	−1.23	0.74	0.30	0.99	−0.24
<i>RVOL</i>	0.15	0.14	0.06	0.32	0.05	0.96	−0.18
<i>BM</i>	0.52	0.50	0.13	1.12	0.25	0.99	0.30
<i>NTIS</i>	0.01	0.02	−0.05	0.04	0.02	0.98	0.04
<i>TBL</i> (%)	4.34	4.22	0.02	14.72	3.09	0.99	0.56
<i>LTY</i> (%)	6.05	5.56	2.18	13.78	2.75	0.99	0.35
<i>LTR</i> (%)	0.53	0.29	−5.94	8.13	2.77	0.05	−0.04
<i>TMS</i> (%)	1.71	1.65	−1.70	4.35	1.40	0.96	−0.54
<i>DFY</i> (%)	0.97	0.85	0.37	2.62	0.44	0.97	−0.17
<i>DFR</i> (%)	0.02	0.07	−4.03	3.92	1.40	−0.08	−0.08
<i>INFL</i> (%)	0.29	0.29	−0.52	1.26	0.36	0.55	0.35
<i>OG</i>	−0.01	0.00	−0.17	0.14	0.26	0.24	0.04
<i>PMAX</i>	0.90	0.93	0.60	1.00	0.09	0.94	0.05

Table 1.2: Job openings rate and other labor variables

This table reports the results from regressing monthly changes in *LJOR* on monthly changes in pay growth, unemployment rate, and hiring rate. Both the dependent and independent variables are standardized such that the full time series mean and standard deviation are set to zero and one, respectively. Adj. R^2 represents the adjusted R^2 from the regression. The values in parentheses are Newey and West (1987) corrected t -statistics with a lag length of 12. The sample period is from February 1951 to December 2016. * and ** represent significance at the 10% and 5% levels, respectively.

Independent variables	Dependent variable: <i>LJOR</i>			
(1)	(2)	(3)	(4)	(5)
Pay growth	0.03 (0.57)			−0.03 (−0.66)
Unemployment rate		−0.25** (−6.29)		−0.26** (−6.26)
Hiring rate			−0.01 (−0.32)	0.02 (−0.53)
Adj. R ² (%)	−0.02	6.38	−10.80	6.31

Table 1.3: In-sample predictive regression

This table reports results from estimation of the following predictive regression: $R_{t:t+h} = \beta_0 + \beta_1 LJOR_t + \epsilon_{t:t+h}$, where, $R_{t:t+h}$ is the average cumulative excess return on the S& P500 index from the end of month t to the end of month $t+h$ and $LJOR_t$ is the log job openings rate at the end of month t . The results are reported for horizons (h) of one, three, six, 12, and 24 months. Column 2 presents the estimated intercept. Column 3 reports the estimated coefficient on $LJOR$. The values in parentheses are Newey and West (1987) corrected t -statistics with a lag length equal to the horizon (h). Column 4 reports the in-sample R^2 in percentage points. Column 5 presents p -values for the one-sided (lower-tailed) test of the null hypothesis that the slope coefficient on $LJOR$ is equal to zero. These p -values are calculated using the bootstrap approach of Huang et al. (2015). Column 6 presents the p -value for the IVX-Wald statistic developed by Kostakis et al. (2014). Column 7 reports $\widehat{qLL}$ statistics for temporal instability tests developed by Elliott and Müller (2006). The 10% critical values for this statistic are reported in Column 8. A $\widehat{qLL}$ statistic less than the critical value is significantly different from zero and shows evidence of model instability. The sample period is from February 1951 to December 2016. * and ** in columns 2 and 3 represent significance (inferred from the t -statistics) at the 10% and 5% levels, respectively.

Horizon (1)	Regression estimates			Tests for significance of $LJOR$		Tests of temporal instability	
	Intercept (2)	Slope (3)	R^2 (%) (4)	Bootstrapped p -value (5)	IVX p -value (6)	$\widehat{qLL}$ statistic (7)	Critical value (10% level) (8)
1	0.59** (3.97)	-0.50** (-3.23)	1.44	0.000	0.001	-5.49	-7.04
3	0.53** (4.00)	-0.49** (-3.75)	4.04	0.002	0.001	-6.77	-7.04
6	0.52** (3.91)	-0.46** (-3.70)	6.51	0.001	0.001	-6.23	-7.04
12	0.50** (3.84)	-0.39** (-3.54)	8.91	0.003	0.005	-4.95	-7.04
24	0.49** (3.99)	-0.28** (-3.72)	9.58	0.007	0.034	-4.40	-7.04

Table 1.4: In-sample comparison with other predictors

The table reports results from regressions of stock market excess returns on lagged predictor variables. Panel A (Panel B) report the results for the monthly (annual) forecasting horizon. Results in Columns 3 and 4 report the univariate regressions given by Equation (1.8). Columns 5 to 8 present results from the estimation of the bivariate predictive regression given by Equation (1.10). The slope estimates, Newey and West (1987) corrected t -statistics, the regression R^2 , and the partial R^2 attributable to $LJOR$ are reported. * and ** represent significance in a one-sided test at the 10% and 5% levels, respectively.

Panel A: Monthly horizon ($h=1$)							
Predictor (1)	Univariate regression			Bivariate regression			
	Expected Sign (2)	Slope (3)	R^2 (%) (4)	Benchmark predictor (5)	$LJOR$ (6)	R^2 (%) (7)	Partial R^2 for $LJOR$ (%) (8)
<i>LJOR</i>	–	–0.50** (–3.23)	1.44	n.a.	n.a.	n.a.	n.a.
<i>DP</i>	+	0.25** (1.65)	0.37	0.38** (2.47)	–0.58** (–3.74)	2.24	1.88
<i>DY</i>	+	–0.17 (–1.18)	0.17	–0.08 (–0.58)	–0.48** (–3.15)	1.47	1.30
<i>EP</i>	+	0.16 (0.80)	0.15	0.41** (2.04)	–0.66** (–4.14)	2.27	2.13
<i>DE</i>	+	0.12 (0.63)	0.09	0.01 (0.03)	–0.50** (–3.27)	1.44	1.35
<i>RVOL</i>	+	0.35** (2.36)	0.70	–0.26 (1.79)	–0.45** (–2.88)	1.82	1.13
<i>BM</i>	+	0.09 (0.54)	0.05	0.26* (1.51)	–0.57** (–3.61)	1.80	1.75
<i>NTIS</i>	–	–0.08 (–0.41)	0.03	–0.06 (–0.29)	–0.49** (–3.25)	1.45	1.42
<i>TBL</i>	–	–0.34** (–2.15)	0.65	–0.08 (–0.48)	–0.45** (–2.64)	1.46	0.82
<i>LTY</i>	–	–0.22* (–1.44)	0.29	–0.05 (–0.35)	–0.48** (–3.07)	1.45	1.17
<i>LTR</i>	+	0.38** (2.46)	0.83	0.36** (2.33)	–0.48** (–3.16)	2.19	1.36
<i>TMS</i>	+	0.30** (1.96)	0.54	0.05 (0.27)	–0.47** (–2.55)	1.45	0.91
<i>DFY</i>	+	0.10 (0.50)	0.06	0.02 (0.08)	–0.49** (–3.32)	1.44	1.38
<i>DFR</i>	+	0.17 (0.75)	0.17	0.13 (0.56)	–0.49** (–3.10)	1.53	1.37
<i>INFL</i>	–	–0.08 (–0.43)	0.04	0.11 (0.55)	–0.53** (–3.41)	1.49	1.46
<i>OG</i>	–	0.04 (–0.73)	0.01	0.07 (1.61)	–0.50** (–3.25)	1.46	1.45
<i>PMAX</i>	–	–0.26* (–1.33)	0.39	–0.24 (–1.24)	–0.49** (–3.17)	1.77	1.38

(continued)

Table 1.4—*Continued*

Panel B: Annual horizon ($h=12$)							
Predictor (1)	Univariate regression			Bivariate regression			
	Expected Sign (2)	Slope (3)	R^2 (%) (4)	Benchmark predictor (5)	$LJOR$ (6)	R^2 (%) (7)	Partial R^2 for $LJOR$ (%) (8)
$LJOR$	—	−0.39** (−3.54)	8.91	n.a.	n.a.	n.a.	n.a.
DP	+	0.29** (2.05)	4.75	0.40** (2.92)	−0.49** (−4.52)	17.66	13.56
DY	+	−0.06 (−0.62)	0.24	0.01 (0.11)	−0.40** (−3.41)	8.91	8.70
EP	+	0.17 (1.09)	1.71	0.39** (2.59)	−0.55** (−4.67)	16.27	14.82
DE	+	0.15* (1.59)	1.37	0.06 (0.74)	−0.38** (−3.34)	9.13	7.87
$RVOL$	+	0.18** (1.76)	1.90	0.11 (1.12)	−0.37** (−3.25)	9.62	7.87
BM	+	0.14 (0.93)	1.05	0.28** (1.92)	−0.48** (−4.17)	13.05	12.13
$NTIS$	—	−0.01 (−0.05)	0.01	0.02 (0.08)	−0.39** (−3.58)	8.92	8.92
TBL	—	−0.24** (−1.82)	3.22	−0.01 (−0.09)	−0.39** (−3.04)	8.91	5.88
LTY	—	−0.12 (−0.88)	0.78	0.03 (0.26)	−0.41** (−3.60)	8.96	8.25
LTR	+	0.14** (3.50)	1.15	0.13** (3.31)	−0.39** (−3.51)	9.82	8.77
TMS	+	0.29** (2.52)	4.83	0.11 (0.80)	−0.33** (−2.59)	9.38	4.78
DFY	+	0.10 (0.86)	0.64	0.04 (0.33)	−0.39** (−3.34)	8.98	8.40
DFR	+	0.03 (0.59)	0.05	0.00 (−0.08)	−0.39** (−3.51)	8.91	8.86
$INFL$	—	−0.22** (−2.36)	2.84	−0.10 (−1.08)	−0.36** (−3.29)	9.37	6.72
OG	—	−0.06 (−1.08)	0.18	−0.04 (−1.09)	−0.39** (−3.53)	8.99	8.83
$PMAX$	—	−0.22** (−2.40)	2.79	−0.21** (−2.28)	−0.39** (−3.32)	11.32	8.78

Table 1.5: Out-of-sample forecasting results

This table reports results of the out-of-sample forecasting tests. The out-of-sample testing period is from 1966 to 2016. Column 1 presents the predictor variable used for the out-of-sample tests. Column 2 and column 3 report the out-of-sample R^2 and the in-sample R^2 over the out-of-sample testing period for the monthly horizon ($h=1$). Column 4 and column 5 report the out-of-sample R^2 and the in-sample R^2 over the out-of-sample testing period for the annual horizon ($h=12$). The values in parentheses are Clark and West (2007) MSEF-adjusted statistics for one-sided (upper-tail) tests of the null hypothesis that the MSFE for the prevailing mean forecast is less than or equal to the MSFE for the forecast obtained from the predictor variable. * and ** represent significance at the 10% and 5% levels, respectively.

Predictor (1)	Monthly horizon ($h=1$)		Annual horizon ($h=12$)	
	Out-of-sample R^2 (%) (2)	In-sample R^2 for out-of-sample period (%) (3)	Out-of-sample R^2 (%) (4)	In-sample R^2 for out-of-sample period (%) (5)
<i>LJOR</i>	0.86** (2.51)	0.94	7.30** (2.59)	6.39
<i>DP</i>	-0.23 (1.01)	0.12	-5.01 (1.45)	1.82
<i>DY</i>	-0.01 (0.68)	0.16	-0.53 (0.27)	0.12
<i>EP</i>	-0.61 (-0.11)	0.04	-6.47 (0.73)	0.41
<i>DE</i>	-0.90 (0.63)	0.02	-6.51 (1.10)	0.70
<i>RVOL</i>	0.12* (1.55)	0.77	0.21* (1.56)	2.52
<i>BM</i>	-1.26 (-1.12)	0.00	-13.67 (-0.02)	0.12
<i>NTIS</i>	-0.97 (0.48)	0.10	-7.27 (0.45)	0.37
<i>TBL</i>	-0.95 (2.04)	0.35	-11.12 (1.69)	1.36
<i>LTY</i>	-0.83 (1.44)	0.07	-12.90 (1.43)	0.01
<i>LTR</i>	0.35** (2.21)	1.00	1.02** (3.01)	1.31
<i>TMS</i>	-0.90 (2.18)	0.67	-4.26 (1.61)	7.03
<i>DFY</i>	-0.64 (-0.10)	0.30	-5.19 (-1.46)	2.25
<i>DFR</i>	-0.55 (-0.08)	0.29	0.52 (0.83)	0.06
<i>INFL</i>	-0.43 (0.33)	0.00	1.83** (1.68)	2.58
<i>OG</i>	-0.05 (-1.14)	0.16	0.98 (1.09)	0.95
<i>PMAX</i>	-0.60 (-0.39)	0.63	-1.29 (1.15)	4.41

Table 1.6: Out-of-sample forecasting results with restrictions

This table reports results of the out-of-sample tests conducted with the Campbell and Thompson (2008) restrictions. The results in Panel A (Panel B) correspond to the monthly (annual) forecasting horizon. In each panel, estimations in columns 2 and 3 impose the restriction that the slope coefficient on the predictor is in line with the expected sign, estimations in columns 4 and 5 impose the restrict the forecasted return to be non-negative, and estimations in the last three columns impose both the above restrictions simultaneously. The out-of-sample R^2 and the Clark and West (2007) MSEF-adjusted statistics are reported. Columns 3, 5, 6, and 7 report the portion of the testing period for which the restrictions are imposed. The forecast evaluation period is from January 1966 to December 2016. * and ** represent significance at 10% and 5% levels, respectively.

Panel A: Monthly horizon ($h=1$)							
Predictor (1)	Expected slope sign		Positive forecast		Both restrictions		
	Out-of-sample R^2 (%) (2)	Slope restriction (3)	Out-of-sample R^2 (%) (4)	Forecast restriction (5)	Out-of-sample R^2 (%) (6)	Slope restriction (7)	Forecast restriction (8)
<i>LJOR</i>	0.86** (2.51)	0.00	0.87** (2.37)	0.23	0.87** (2.37)	0.00	0.23
<i>DP</i>	-0.23 (1.01)	0.00	-0.06 (1.15)	0.15	-0.06 (1.15)	0.00	0.15
<i>DY</i>	-0.16 (0.05)	1.00	-0.01 (0.69)	0.00	-0.16 (0.05)	1.00	0.00
<i>EP</i>	-0.61 (-0.11)	0.00	-0.14 (0.41)	0.02	-0.14 (0.41)	0.00	0.02
<i>DE</i>	-0.76 (0.72)	0.00	-0.22 (0.82)	0.18	-0.20 (0.85)	0.00	0.18
<i>RVOL</i>	0.12* (1.55)	0.00	0.16* (1.57)	0.04	0.16* (1.57)	0.00	0.04
<i>BM</i>	-1.06 (-0.85)	0.06	-1.26 (-1.12)	0.00	-1.06 (-0.85)	0.06	0.00
<i>NTIS</i>	-0.94 (0.51)	0.01	-0.83 (0.57)	0.04	-0.81 (0.60)	0.01	0.04
<i>TBL</i>	-0.95 (2.04)	0.00	0.38** (1.77)	0.32	0.38** (1.77)	0.00	0.32
<i>LTY</i>	-0.83 (1.44)	0.00	0.63** (2.04)	0.36	0.63** (2.04)	0.00	0.36
<i>LTR</i>	0.35** (2.21)	0.00	0.32** (1.98)	0.11	0.32** (1.98)	0.00	0.11
<i>TMS</i>	-0.90 (2.18)	0.00	-0.80 (1.76)	0.13	-0.80 (1.76)	0.00	0.13
<i>DFY</i>	-0.53 (0.08)	0.01	-0.64 (-0.10)	0.00	-0.53 (0.08)	0.01	0.00
<i>DFR</i>	-0.24 (-0.26)	0.40	-0.50 (-0.16)	0.03	-0.26 (-0.42)	0.40	0.01
<i>INFL</i>	-0.43 (0.33)	0.00	-0.11 (0.50)	0.09	-0.11 (0.50)	0.00	0.09
<i>OG</i>	-0.01 (-0.68)	1.00	-0.05 (-1.14)	0.00	-0.01 (-0.68)	1.00	0.00
<i>PMAX</i>	-1.34 (-0.81)	0.01	-0.34 (0.01)	0.00	-0.32 (0.05)	0.01	0.01

(continued)

Table 1.6—*Continued*

Panel B: Annual horizon ($h=12$)							
Predictor (1)	Expected slope sign		Positive forecast		Both restrictions		
	Out-of-sample R^2 (%) (2)	Slope restriction (3)	Out-of-sample R^2 (%) (4)	Forecast restriction (5)	Out-of-sample R^2 (%) (6)	Slope restriction (7)	Forecast restriction (8)
<i>LJOR</i>	7.30** (2.59)	0.00	8.88** (2.66)	0.11	8.88** (2.66)	0.00	0.11
<i>DP</i>	−5.01 (1.45)	0.00	1.81** (1.90)	0.28	1.81** (1.90)	0.00	0.28
<i>DY</i>	1.12** (1.82)	1.00	−0.53 (0.27)	0.00	1.12** (1.82)	1.00	0.00
<i>EP</i>	−6.47 (0.73)	0.00	−2.18 (1.17)	0.09	−2.18 (1.17)	0.00	0.09
<i>DE</i>	−6.51 (1.10)	0.00	0.97* (1.33)	0.16	0.97* (1.33)	0.00	0.16
<i>RVOL</i>	0.21* (1.56)	0.00	0.22* (1.56)	0.00	0.22* (1.56)	0.00	0.00
<i>BM</i>	−13.78 (−0.03)	0.04	−13.73 (−0.04)	0.00	−13.83 (−0.05)	0.04	0.00
<i>NTIS</i>	−5.44 (0.70)	0.14	−7.26 (0.45)	0.00	−5.44 (0.70)	0.14	0.00
<i>TBL</i>	−11.12 (1.69)	0.00	6.36** (2.07)	0.31	6.36** (2.07)	0.00	0.31
<i>LTY</i>	−12.90 (1.43)	0.00	5.96** (2.00)	0.35	5.96** (2.00)	0.00	0.35
<i>LTR</i>	1.02** (3.01)	0.00	1.31** (3.13)	0.01	1.31** (3.13)	0.00	0.01
<i>TMS</i>	−4.26 (1.61)	0.00	−4.23 (1.50)	0.10	−4.23 (1.50)	0.00	0.10
<i>DFY</i>	−3.73 (−0.97)	0.03	−5.12 (−1.45)	0.01	−3.73 (−0.97)	0.03	0.00
<i>DFR</i>	0.82 (0.98)	0.85	0.52 (0.83)	0.00	0.82 (0.98)	0.85	0.00
<i>INFL</i>	1.83** (1.68)	0.00	3.64** (2.08)	0.07	3.64** (2.08)	0.00	0.07
<i>OG</i>	0.99 (1.09)	0.03	0.98 (1.09)	0.00	0.99 (1.09)	0.03	0.00
<i>PMAX</i>	−1.29 (1.15)	0.00	−1.29 (1.15)	0.00	−1.29 (1.15)	0.00	0.00

Table 1.7: Out-of-sample encompassing tests

This table reports the results of the out-of-sample encompassing tests. Column 1 presents the benchmark predictor variable that is competing with *LJOR* in the encompassing tests. Column 2 and column 3 report the weights on *LJOR* and on the competing benchmark predictor variable from the formation of a combination forecast for the monthly horizon ($h=1$). Column 4 and column 5 report the same for the annual horizon ($h=12$). The values in parentheses report the Harvey et al. (1998) ENC-statistic for the one-sided (upper-tail) test of the null hypothesis that the estimated weight is equal to zero. Data from February 1951 to December 1965 is used to obtain the first forecast and therefore, the sample period for the out-of-sample analysis is from January 1966 to December 2016. * and ** represent significance at 10% and 5% levels, respectively.

Predictor (1)	Monthly horizon ($h=1$)		Annual horizon ($h=12$)	
	<i>LJOR</i> (2)	Benchmark predictor (3)	<i>LJOR</i> (4)	Benchmark predictor (5)
<i>DP</i>	0.63** (3.21)	0.37** (1.91)	0.67** (4.40)	0.33** (2.54)
<i>DY</i>	0.69** (2.57)	0.31 (1.20)	0.82** (3.23)	0.18 (0.82)
<i>EP</i>	0.69** (2.79)	0.31* (1.35)	0.71** (3.71)	0.29** (2.24)
<i>DE</i>	0.85** (2.48)	0.15 (0.49)	0.91** (3.20)	0.09 (0.60)
<i>RVOL</i>	0.60** (2.64)	0.40** (1.77)	0.70** (4.10)	0.30** (1.83)
<i>BM</i>	0.79** (3.42)	0.21 (0.99)	0.80** (3.94)	0.20* (1.42)
<i>NTIS</i>	0.70** (3.63)	0.30* (1.58)	0.74** (5.33)	0.26** (1.73)
<i>TBL</i>	0.80** (2.79)	0.20 (0.72)	0.92** (3.35)	0.08 (0.47)
<i>LTY</i>	0.77** (2.65)	0.23 (0.87)	0.90** (3.21)	0.10 (0.80)
<i>LTR</i>	0.55** (3.08)	0.45** (2.45)	0.69** (3.50)	0.31** (1.76)
<i>TMS</i>	0.74** (3.24)	0.26 (1.24)	0.81** (3.20)	0.19* (1.35)
<i>DFY</i>	0.73** (3.19)	0.27 (1.19)	0.85** (3.75)	0.15 (0.84)
<i>DFR</i>	0.73** (2.71)	0.27 (1.04)	0.75** (3.25)	0.25 (1.24)
<i>INFL</i>	0.80** (2.63)	0.20 (0.70)	0.73** (3.13)	0.27* (1.37)
<i>OG</i>	0.67** (2.56)	0.33 (1.27)	0.74** (3.16)	0.26 (1.27)
<i>PMAX</i>	0.73** (3.25)	0.27 (1.25)	0.69** (3.65)	0.31** (1.99)

Table 1.8: Asset Allocation

This table reports the results of the asset allocation tests. Monthly rebalancing strategies that invest in a risk-free asset and the S&P 500 index are formed using the forecasts based on the predictor variable ($\hat{R}_{t:t+1}$) noted in column 1. The weight on the S&P 500 index in these strategies is restricted between 0 and 1.5. Column 2 presents the gain in annualized gain certainty equivalent returns over a similar benchmark strategy that employs the prevailing mean forecasts. The values in parentheses are DeMiguel et al. (2007) z -statistics for the one-sided (upper-tail) test of the null hypothesis that the difference in the gain in certainty equivalent returns is zero. Column 3 shows the gain in Sharpe ratios over the benchmark strategy uses the prevailing mean forecasts. The values in parentheses are Jobson and Korkie (1981) t -statistics for a one-sided (upper-tailed) test of the null hypothesis that the gain in the Sharpe ratios is equal to zero. The investor is assumed to have a risk-aversion coefficient of three. Data from February 1951 to December 1965 is used to obtain the first forecast and therefore, the sample period for the out-of-sample analysis is from January 1966 to December 2016. * and ** represent significance at 10% and 5% levels, respectively.

Predictor (1)	Δ Certainty equivalent return (2)	Δ Sharpe ratio (3)
<i>LJOR</i>	3.36** (4.82)	0.20** (1.92)
<i>DP</i>	-0.61 (-0.98)	-0.07 (-0.77)
<i>DY</i>	-0.05 (-0.17)	0.00 (-0.10)
<i>EP</i>	0.41 (1.01)	0.01 (0.23)
<i>DE</i>	-0.34 (-0.64)	-0.05 (-0.57)
<i>RVOL</i>	-0.22 (-0.57)	0.02 (0.32)
<i>BM</i>	-0.78 (-2.66)	-0.05 (-1.17)
<i>NTIS</i>	0.64* (1.48)	0.06 (1.03)
<i>TBL</i>	1.95** (2.93)	0.11 (1.09)
<i>LTY</i>	1.58** (2.44)	0.08 (0.77)
<i>LTR</i>	1.66** (3.17)	0.11* (1.41)
<i>TMS</i>	3.17** (5.25)	0.20** (2.37)
<i>DFY</i>	-1.00 (-3.70)	-0.03 (-0.81)
<i>DFR</i>	0.35 (1.08)	0.02 (0.33)
<i>INFL</i>	0.75** (1.74)	0.04 (0.57)
<i>OG</i>	-0.06 (-2.42)	-0.00 (-0.88)
<i>PMAX</i>	-0.47 (-1.38)	-0.01 (-0.29)

Table 1.9: Asset allocation with restricted forecasts

This table reports the results of the asset allocation tests in which the out-of-sample return forecasts are subjected to the Campbell and Thompson (2008) slope and forecast restrictions. Portfolio rebalancing strategies similar to those from Table 1.7 are formed using the restricted return forecasts. Column 1 presents the predictive variable that is used to obtain the return forecast. Column 2 presents the gain in annualized certainty equivalent returns over a benchmark strategy that uses the prevailing mean forecasts. The values in parentheses are DeMiguel et al. (2007) z -statistics for the one-sided (upper-tail) test of the null hypothesis that the difference in the gain in certainty equivalent returns is zero. Column 3 shows the gain in Sharpe ratios over the benchmark strategy that employs the prevailing mean forecasts. The values in parentheses are Jobson and Korkie (1981) t -statistics for a one-sided (upper-tailed) test of the null hypothesis that the gain in the Sharpe ratios is equal to zero. The investor is assumed to have a risk-aversion coefficient of three. Data from February 1951 to December 1965 is used to obtain the first forecast and therefore, the sample period for the out-of-sample analysis is from January 1966 to December 2016. * and ** represent significance at 10% and 5% levels, respectively.

Predictor (1)	Δ Certainty equivalent return (2)	Δ Sharpe ratio (3)
<i>LJOR</i>	3.36** (4.82)	0.20** (1.92)
<i>DP</i>	-0.61 (-0.98)	-0.07 (-0.77)
<i>DY</i>	0.03 (0.12)	0.04** (2.22)
<i>EP</i>	0.41 (1.01)	0.01 (0.23)
<i>DE</i>	-0.32 (-0.61)	-0.04 (-0.55)
<i>RVOL</i>	-0.22 (-0.57)	0.02 (0.32)
<i>BM</i>	-0.53 (-1.89)	-0.03 (-0.74)
<i>NTIS</i>	0.67* (1.57)	0.06 (1.08)
<i>TBL</i>	1.95** (2.93)	0.11 (1.09)
<i>LTY</i>	1.58** (2.44)	0.08 (0.77)
<i>LTR</i>	1.66** (3.17)	0.11* (1.41)
<i>TMS</i>	3.17** (5.25)	0.20** (2.37)
<i>DFY</i>	-0.84 (-3.13)	-0.02 (-0.54)
<i>DFR</i>	0.27 (1.19)	0.01 (0.39)
<i>INFL</i>	0.75** (1.74)	0.04 (0.57)
<i>OG</i>	0.00 (0.17)	0.00* (1.50)
<i>PMAX</i>	-0.41 (-1.15)	-0.01 (-0.22)

Figure 1.1: Job openings rate

The figure shows the time series of monthly values of the job openings-to-employment ratio. The series is lagged by one month for the data to become publicly available. The shaded areas indicate recessions as defined by the National Bureau of Economic Research. The sample period is February 1951 to January 2017.

Figure 1.2: Out-of-sample forecasts

This figure plots the monthly time series of the actual and forecasted out-of-sample returns for an annual horizon ($h=12$). The solid line indicates the forecast obtained using the job openings rate, the dotted line indicates realized return, and the dashed line indicates prevailing mean return. The sample period is from January 1966 to December 2016. The shaded vertical bars represent NBER-dated recessions.

Figure 1.3: Equity weight

This figure plots the monthly time series of the equity weight for a mean-variance investor with a relative risk aversion coefficient of three who allocates between equities and risk-free bills using excess return forecast. The weights are restricted to be between 0 and 1.5. The solid line shows the equity weight when the investor uses a forecast obtained from *LJOR* and the dotted line shows the equity weight when the investor uses the prevailing-mean forecast. The sample period is from January 1966 to December 2016. The shaded vertical bars represent NBER-dated recessions.

Figure 1.4: Log cumulative wealth

This figure plots the monthly time series of the log cumulative wealth for a mean-variance investor with a relative risk aversion coefficient of three who allocates between equities and risk-free bills using excess return forecast. The solid line shows log cumulative wealth when the investor uses the forecast obtained from *LJOR* and the dotted line shows the cumulative log wealth when the investor uses prevailing-mean forecast. The investor is assumed to start with \$1 reinvesting all proceeds and rebalancing the portfolio on a monthly basis. The sample period is from January 1966 to December 2016. The shaded vertical bars represent NBER-dated recessions.

Chapter 2

Dissecting the Link Between Activeness and Fund Performance

2.1 Introduction

The literature on performance evaluation in mutual funds presents two prominent views regarding the existence of managerial skill. Using arithmetic of active management proposed by Sharpe (1991), Fama and French (2010) argue that on average, fund managers lack the skill to produce abnormal returns before fees. They also find that fund managers that provide investors with net of fee risk-adjusted returns (net alpha) are rare. On the contrary Berk and Green (2004) argue that in a partial equilibrium, fund managers are skilled. However, the rents of their such skill are consumed in management fees resulting in zero net alpha for every fund. Despite the differences in theoretical motivation and implications, both of these views agree that predicting cross-sectional differences in the net alpha of mutual funds is rather an attribute of a draw of luck. The collective investment experience and empirical evidence from performance attribution studies over the last several decades confirms this argument and suggests that finding skillful managers that deliver positive net

alpha is extremely difficult.¹

Despite the above, several recent studies relate managerial activeness to skill, and therefore, the ability to generate net alpha. These studies show that investors can use ex-ante observable proxies of managerial activeness, to form fund selection strategies (“activeness strategies”) that generate large net alpha spreads for characteristic-sorted fund portfolios.² Although all such activeness proxies measure deviation from the benchmark, these studies also document that their proposed activeness measure is unique and inherently different from all the others published prior. Some studies prove the distinctiveness and superiority of their proposed measure in a horserace of multiple activeness measures. Consequently, readers are left with a overwhelmingly large number distinct activeness proxies that they can use to guide their decisions regarding mutual fund choice.

The goal of this paper is to help investors choose between the different proposed performance predictors in forming their investment strategy. To this purpose, I re-examine the robustness of the predictive ability of 11 ex-ante observable activeness proxies. Specifically, I argue that any investment strategy would also inspire confidence in a potential investor if it had robust performance to the following four propositions.

Proposition 2.1.1. *The proposed investment strategy should generate an economi-*

¹See for example, Jensen (1968), Lehmann and Modest (1987), Samuelson (1989), Malkiel (1995), Carhart (1997), Fama and French (2010), Berk and Van Binsbergen (2015), Ferson and Chen (2015), Cederburg et al. (2018), Jordan and Riley (2016).

²Recent studies claim that fund characteristics such as *Industry Concentration* (Kacperczyk et al. (2005)), *Return Gap* (Kacperczyk et al. (2006)), *Active Share* and *Duration* (Cremers and Petajisto (2009) and Cremers and Pareek (2016)), R^2 (Amihud and Goyenko (2013)), *Contrarian* and *Contemporaneous Herd* (Wei et al. (2014) and Koch (2016)), *Active Weight* (Doshi et al. (2015)), *Volatility* (Jordan and Riley (2015)), and *Leading Herd* (Koch (2016)) predict cross-sectional differences in net alpha.

cally meaningful and statistically significant in-sample performance.

This is the lowest possible threshold and simply requires replicability. The original studies documenting the activeness strategies vary widely in terms of portfolio formation methods, sample periods, and sample selection screens used to identify actively managed domestic mutual funds. Therefore, as a first step, I argue that the evidence of strong performance documented in the original studies should be reconcilable under a consistent set of replication procedures for all the strategies.

Proposition 2.1.2. *The performance of the proposed investment strategy should not be concentrated only in a short part of in-sample time period.*

Any strategy that delivers strong performance for only a short part of the historical data and performs flatly otherwise, is rather misleading in terms of guiding future investment decisions (Goyal and Welch (2008)). Therefore, the performance of a robust investment strategy should be comparable across different in-sample sub-periods.

Proposition 2.1.3. *The strong in-sample performance should continue and be comparable to the out-of-sample performance (i.e., outside the sample period of the original study).*

There can be several reasons for why this might not be the case. For example, in examining the out-of-sample performance cross-sectional asset pricing anomalies, Schwert (2003), Conrad et al. (2003), and McLean and Pontiff (2016) document that the out-of-sample performance can be considerably lower than the in-sample performance due to data mining biases in the original studies. If such is the case for the activeness studies, the documented in-sample predictive power would be largely meaningless in guiding future fund selection decisions.

Proposition 2.1.4. *The performance of the investment strategy should not be explained away by well-known benchmark models other than the traditionally used Carhart (1997) four-factor model.*

In performance evaluation, abnormal return or alpha is defined as the part of returns left unexplained by mechanical trading strategies that span alternative investment opportunity set. Over the last two decades, the cross-sectional asset pricing literature has documented an extensive list of mechanical trading strategies generate significant alpha when benchmarked with the Carhart (1997) four-factor model.³ Therefore, an activeness strategy that can be simply replicated using one or a combination of such mechanical trading strategies will also be deemed to generate significant alpha when benchmarked with the four-factor model. In recent studies, Fama and French (2015) and Fama and French (2016) show that a five-factor asset pricing model that consists of market, size, value, profitability and investment factors and explains a much wider array of such mechanical trading strategies than the four-factor model. I argue, therefore, that the five-factor model serves as a more comprehensive proxy of alternative investment opportunity set. In the same spirit as this argument, Jordan and Riley (2016) and Sheng et al. (2017) also show that, in evaluating mutual fund performance, adding profitability and investment factors to the four-factor model considerably improves the tracking efficiency and explanatory power. Hence, I propose testing the performance of activeness strategies using five-factor as a benchmark.

Using these four criteria, I reexamine the performance of the 11 strategies formed using the managerial activeness proxies listed in Table 2.1. I begin by estimating in-sample performance and note that the net alpha for these activeness strategies

³For an extensive list, see Harvey et al. (2016) and Hou et al. (2017).

are economically large and range from 0.69% to 5.10% per year. Nine of 11 (about 80%) activeness strategies pass the first threshold and exhibit strong in-sample performance.

Figure 2.1 plots the excess returns for the nine activeness strategies.⁴ The shaded area represents a three year time period from 1999 to 2001 (“anomalous time period”), also known for surrounding the crash of the “dot-com bubble” of the 1990s. Note that for all the nine strategies, the returns during these three years are extremely volatile and are marked by large positive spikes. Figure 2.2 shows the cumulative net excess returns for the same strategies. For six out of nine strategies, the steepest period investment growth occurs during with the aforementioned three year period. A simple glance at these figures suggests that the activeness strategies exhibit an extremely concentrated performance in a short three year period and are flat otherwise. I confirm this by re-estimating the net alpha for these strategies excluding data from the anomalous time period and find that only two out nine strategies continue to have significant net alpha.

Next, I conduct an out-of-sample test by estimating net alpha for the activeness strategies in the sample period that is outside the sample period of the original study. My estimates of the out-of-sample net alpha show an average reduction of average 60% to 100% from the in-sample net alpha. Furthermore, none of the strategies have significant out-of-sample net alpha. This weak out-of-sample performance, combined with concentrated in-sample performance, shows that regardless of the time period used, the predictive power of the activeness proxies is largely restricted to a short

⁴In Figure 2.1 and Figure 2.2, I extend the sample period of the original study forward to December 2014, the last month in my sample.

three year time period from January 1999 to December 2001.

Finally, I benchmark the nine activeness strategies with the Fama and French (2015) five-factor model and show that all the nine activeness strategies have positive and significant exposure to either profitability or investment factors. Accounting for these exposures, the net alpha on these strategies declines by, on average, about 80% and only two out of the nine strategies continue to generate significant net alpha.⁵

These results contribute to an extensive literature on performance evaluation of mutual funds. A long-standing question in this literature is whether investors can use ex-ante observable characteristics to successfully identify group skilled managers. The studies argue that easily constructible and ex-ante observable characteristics that measure managerial activeness can feasibly help potential investors find managers that deliver positive alpha after fees. In doing so, these have attracted considerable attention from the investment community. For example, some funds have recently started reporting their high deviation from benchmark (*Active Share*) in promotional materials to differentiate themselves as skilled and attract more capital.⁶ In addition, many data providers (e.g., Factset, Morningstar, and Thomson Reuters Eikon) have started making activeness proxies such as R^2 , *Active Share*, and *Duration* readily available for investors. Therefore, activeness proxies are influencing capital allocation decisions amongst retail and institutional investors.

The results above, however, show that such investors would be largely disappointed by the performance of strategies based on activeness proxies. I find that none of the

⁵The results are similar using other benchmarking models that also include information from profitability and investment factors such as, a six-factor model that adds momentum to the five-factor model, the q-factor model of Hou et al. (2015), and the mispricing-factors model of Stambaugh and Yuan (2017).

⁶For example, see the fiscal year 2016 annual report for ‘Dodge and Cox Funds - Stock Fund (p. 4).

11 strategies examined in this paper, survive the battery of four robustness tests. The tests shows that the predictive power of activeness proxies not only comes from three years of data but also is largely absent in out-of-sample tests. Furthermore, accounting for the omitted profitability and investment factors from the benchmark models, I cast a considerable doubt on the relation between managerial activeness and mutual fund performance.

This study also contributes to a recent and growing literature on the importance of replication and out-of-sample testing. A handful of studies in the cross-sectional asset pricing literature bring attention to this issue. Conrad et al. (2003) argue that data mining itself can be responsible for about 50% of the in-sample predictability. Examining the out-of-sample performance of size, value, momentum, and dividend yield anomalies, Schwert (2003) documents that three out of four anomalies have substantially weak or even reverse of the originally documented relations. McLean and Pontiff (2016) conduct an out-of-sample test for 97 anomalies and conclude that alpha on these anomalies declines by 26% to 32% post-publication. Using a multiple testing framework, Harvey et al. (2016) argue that 27% to 53% of the documented anomalies are false discoveries. Hou et al. (2017) replicate 447 published anomalies and deem 64% of these insignificant after controlling for the microcaps with NYSE breakpoints and value-weighted returns.

In mutual funds, Cederburg et al. (2018) reevaluate the performance of six strategies formed using managerial activeness proxies. Using conditional models that rely on high-frequency beta instruments, they decompose the performance of these strategies into factor timing, volatility timing and conditional alpha. In most cases, they

find that the performance of these strategies is largely attributable to factor timing. Jones and Mo (2017) follow the framework of McLean and Pontiff (2016) and evaluate out-of-sample performance of 16 predictors of cross-sectional mutual fund performance. They document, on average, a 50% reduction in the net alpha outside the original sample of the study. Their findings differ from the findings in this paper in several ways. The results in this paper give a precise estimate of decline in the predictive power of each of the activeness proxies, whereas they study the average decline performance of all predictors taken as a group. In the economic interpretation of their results, Jones and Mo (2017) find that higher levels of arbitrage activity causes the decline. The results in this paper, however, suggest that even the in-sample predictability is absent outside three year time period and is an artifact of omitted benchmark factors.

The remainder of the paper is organized as follows. Section 2.2 details the sample construction process. Section 2.3 describes the selection of managerial activeness proxies and the construction of the mutual fund portfolios. Section 2.4 presents the main results related to the performance of the mutual fund portfolios, and Section 2.5 concludes.

2.2 Data and Sample Construction

I obtain data on monthly fund returns from the CRSP (Center for Research in Security Prices) Survivor-Bias-Free US Mutual Fund Database for the period between January 1984 and December 2014. The CRSP data start in 1962, but I choose exclude all the data until December 1983 from the analysis in this paper. The choice to restrict the sample period to post-1983 is motivated by the fact that in pre-1984 sample

period, not all funds in the sample report monthly returns. In addition, there is also a selection bias in the data such that funds which report monthly returns during pre-1984 have higher average reported returns than the broader universe (Elton et al. (2001), Fama and French (2010)). Furthermore, construction of some of the activeness measures (e.g., strategy on *Volatility*) requires the use daily fund returns which are also obtained from CRSP and are only available starting in September 1998.

These returns are net of fee, expenses and brokerage commissions. I convert them to excess returns by subtracting the risk-free rate.⁷ Moreover, the CRSP database also contains information on fund characteristics such as total net assets, expense ratio, turnover, and the percentage of portfolio that is in equities. I use the MFLINKS database to combine multiple share classes into fund-level portfolios. For these funds, I also compute returns, expenses, turnover ratio, and other characteristics by taking an average weighted by most recent total net assets of a share class.

Several of the activeness proxies also require data on fund holdings. For this purpose, I use the Thomson Reuters Mutual Fund Holdings Database and link it to the CRSP mutual fund database using MFLINKS. The holdings data also contains information on investment objective codes of individual funds and stock-level identifiers. I use these to link fund holdings to CRSP Monthly Stock File and obtain information on returns for individual fund holdings, industry (SIC codes), as well as market capitalizations.

I limit the analysis to actively managed domestic equity funds identified with

⁷I obtain data on risk-free rate, and factor returns for the Carhart (1997) four-factor model and Fama and French (2015) five-factor model from Ken French’s website (<http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>). In addition, the returns for q-factors are provided by Lu Zhang. The data on mispricing-factors are obtained from Yu Yuan’s website (<http://www.saif.sjtu.edu.cn/facultylist/yyuan/>). I thank Ken French, Lu Zhang, and Yu Yuan for sharing their data.

the Lipper, Strategic Insight and Wiesenberger objective codes obtained from the CRSP database. Whenever none of these objective codes are available, I use the investment objective code from Thomson Reuters.⁸ In addition, I also eliminate index funds, balanced funds, international funds (either by their stated style or by their name), sector funds, and funds that hold less than 70% in common stocks.⁹ Following Elton et al. (1996) and Evans (2010), I include two additional screens to address the potential of incubation bias. First, I eliminate funds with total net assets less than \$10 millions. Second, I delete any observations that precede the offer date in the CRSP database. The final sample contains 3,437 funds and 497,657 fund-month observations.

2.3 Managerial Activeness Proxies

To examine the link between cross-sectional fund performance predictability and managerial activeness, I first need to establish a comprehensive definition of activeness that covers all the broad variety of documented measures in the literature. For this purpose, I follow Cremers and Petajisto (2009) who use a counterexample to define activeness. They state that passive management, which is easier to define than active management, consists of simply following the benchmark. Using this argument, they define active management as a “deviation from the passive management” or the

⁸I include funds with Lipper objective codes that are in EI, EIEI, ELCC, G, GI, LCCE, LCGE, LCVE, LSE, MC, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, MR, S, SCCE, SCGE, SCVE, SESE, and SG, Strategic Insight codes that are in AGG, GMC, GRI, GRO, ING, and SCG, and Wiesenberger codes that are in G, GCI, G-I, G-I-S, GS, G-S, G-S-I, I, IEQ, IFL, I-G, I-G-S, I-S, I-SG, LTG, MCG, S, SCG, S-G-I, S-I, and S-I-G.

⁹I remove all the funds with any of the following in their names 'ind', 'index', 's&p', 'dow', 'wilshire', 'russell', 'etf', 'exchange', 'idx', 'dfa', 'nasdaq', 'spdr', 'balance', 'international', 'sector', 'emerging market', 'emerging-market', 'global', 'commodity', 'commodities', 'euro', 'currency', 'currencies', 'world', 'target', '2005', '2010', '2015', '2020', '2025', '2030', '2035', '2040', '2045', '2050', '2055', '2060', '2065', '2070'.

benchmark.

Consistent with the above definition, there are several managerial activeness proxies documented in mutual funds literature. Although all such proxies measure deviation from passive management, they widely vary from each other in terms of their choice of benchmarks used to define passive management. For example, Cremers and Petajisto (2009) and Cremers and Pareek (2016) use either the self-declared benchmark or the benchmark that produces lowest value of their proposed activeness measure. On the other hand, Amihud and Goyenko (2013) measure activeness as a deviation from the Carhart (1997) four-factor model.

I read through all the studies on mutual funds published in seven journals (Journal of Finance, Journal of Financial Economics, Review of Financial Studies, Journal of Financial and Quantitative Analysis, Management Science, Review of Finance, and Review of Asset Pricing Studies) and make a comprehensive list of all the characteristic variables that are shown to predict net of fee mutual fund performance. For this sample of characteristic variables, I exclude all the characteristics that do not proxy for managerial activeness or deviation from a specified benchmark. In addition, I impose two additional filters on the sample of these activeness proxies. First, I exclude any activeness proxy where the original study does not document a significant (at 10%) net alpha for the characteristic-sorted portfolios when benchmarked with the Carhart (1997) four-factor model.¹⁰ Second, I limit the sample of characteristics to the variables that can be constructed using either the past returns, the holdings

¹⁰For example, Koch (2016) proposes tendency to follow a herd as a negative predictor of net alpha. However, he does not find significant net alpha on characteristic-sorted portfolios formed using this measure.

data or a combination of both.¹¹ The final sample of characteristics consists of the 11 managerial activeness proxies that are listed in Table 2.1. This table also consists that proposed each of the activeness proxies and the sample period used in these study.

I follow the original studies to construct each of these activeness proxies and relegate the exact details of these papers. I construct each of the fund-level activeness proxies on either a monthly or a quarterly basis based on the data availability.¹² For each of the activeness proxies in the sample, I construct hypothetical equal-weighted long, short, and long-short mutual funds portfolios using quintile sorts of lagged values. In addition, if the construction of a particular activeness proxy entails the use only holdings data (*Duration*, *Contemporaneous Herd*, *Industry Concentration*, *Return Gap*, *Contrarian*, *Active Weight*, and *Active Share*), I rebalance the characteristic-sorted portfolios on a quarterly basis. The remaining portfolios, which are formed using activeness proxies whose construction entails the use of either the returns data or a combination of the returns and holdings data (R^2 , *Volatility*, *volatility gap*, and *Return Gap*), are rebalanced monthly.

In last two columns, Table 2.1 lists the quintiles used to form the corresponding long, short and long-short characteristic-sorted portfolios in each strategy. The

¹¹For example, Kacperczyk and Seru (2007) use deviation from public information proxied by consensus analyst recommendations and require use of the IBES data. As such, I do not include their proposed variable in this study.

¹²Appendix B.1 details the definitions and construction of these proxies. For *Active Share* I obtain the quarterly data from Antti Petajisto's website (<http://www.petajisto.net/data.html>) from January 1984 to December 2009. I further supplement this with annual data from Martijn Cremers's website from January 2010 to December 2014 <http://activeshare.nd.edu/data/>. Martijn Cremers also provides annual data from on *Duration* for the period of January 1984 to December 2014. I leverage Mike Simutin's code for the construction of *Active Weight* http://www-2.rotman.utoronto.ca/simutin/dfp_code.asp. I thank Antti Petajisto, Martijn Cremers, and Mike Simutin for sharing their data and code.

construction of these long-short portfolios is such that the long (short) side of the strategy consists of funds that are expected to outperform (underperform) based on the findings of the original study. For example, R^2 , *Duration*, *Volatility* and *Contemporaneous Herd* are negatively related to fund performance. Therefore, the long side of the strategies consists of funds in the lowest quintile, and the short side consists of funds in the highest quintile. The only exception is *volatility gap*. Following Huang et al. (2011), the long side of portfolios formed on *volatility gap* consists of funds in the third quintile and the short side consists of funds in the highest quintile. All the other remaining activeness proxies (i.e. *Industry Concentration*, *Return Gap*, *Contrarian*, *Active Weight*, *Active Share*, and *Leading Herd*) are related to fund performance positively. Therefore, for the strategies based on these activeness proxies, the long side consists of funds in the highest quintile and the short side consists of funds in the lowest quintile.

Table 2.2 presents the summary statistics. The sample period is from January 1984 to December 2014. For *Volatility*, however, I use the sample period from September 1999 to December 2014 due to restricted availability of daily returns data. Panel A presents the time-series averages of the cross-sectional mean, median, 10th percentile, and 90th percentile for each activeness proxy. Comparing the summary statistics to the original studies, despite the differences in the sample periods and empirical methods used to calculate the cross-sectional summary statistics, I find that all the variable mean and medians are close to those in the original studies. The only notable exception is *Duration*. The mean *Duration* in the sample of Cremers and Pareek (2016) is 1.36, which is considerably lower than the mean of 5.77 in Table 2.2. This

is partly due to the differences in the sample selection criteria. Cremers and Pareek (2016) use a sample period from January 1990 to December 2013. However, for the pre-1990 and post-2013 periods, the *Duration* for mutual funds in my sample is much higher as compared to the rest of the sample period. Furthermore, as mentioned above, I take time-series averages of cross-sectional summary statistics to address the unbalanced nature of the panel data. This procedure puts higher weights on the observations from pre-1990 time period resulting in higher values for summary statistics.

Panel B of reports the time-series averages of cross-sectional correlations between the activeness proxies. As noted before, since all the activeness proxies measure deviation from some benchmark, one might expect these variables to be substantially correlated. However, although some of the activeness proxies tend to show a sizable correlation, most of them show substantial independent variation. For example, the largest pairwise correlation in absolute value is 0.51 between R^2 and *Active Share*, which is substantially high. The second and third largest pairwise correlations are between *Industry Concentration* and *Active Share* (0.37) and between R^2 and *Industry Concentration* (-0.35). Evidently, although these three variables overlap in terms of their information content to a certain extent, they also seem to contain independent information about the fund's strategy. In addition, out of the fifty-five pairwise correlations between the 11 activeness proxies, only 11 are above the magnitude of 0.10 in absolute value. As such, these variables appear to capture distinct aspects of managerial activeness.

2.4 Results

In this section, I re-examine the performance the 11 activeness strategies subjected to the four robustness criteria stated previously in Proposition 2.1.1 to Proposition 2.1.4. Throughout the performance evaluation tests, I consider three different sample periods for each of the above 11 activeness strategies. First, I consider the sample periods used in the original studies that introduce a particular activeness proxy (original sample period). Second, I extend the sample periods of the original studies forward to December 2014, the last month of the sample period (forward-extended sample period). Third, whenever data is available, I extend the sample periods of the original studies forward to December 2014 as well as backward to January 1984 (backward and forward-extended sample period).

Section 2.4.1 presents the results of the in-sample performance evaluation, in Section 2.4.2, I conduct tests for concentrated in-sample performance, Section 2.4.3 reports the results of out-of-sample tests, Section 2.4.4 reports the performance when benchmarked with Fama and French (2015) five-factor model, and Section 2.4.6 explores alternative benchmark models.

2.4.1 In-Sample Performance

Proposition 2.1.1 states that any robust investment strategy should be replicable under a general set of procedures and have a strong in-sample performance. Following this, I begin with evaluating the performance of these strategies over the original sample period as well as the forward-extended sample period.¹³

¹³Although the forward-extended sample period contains observations outside the original sample period, I deem it as in-sample. I make this choice to distinguish the performance in forward-extended sample period from the sample period that is truly outside the original sample period and is used

Table 2.3 presents the results. I report the average net excess returns and the net alpha estimates from the regression of net excess returns of the activeness strategies on the Carhart (1997) four-factor benchmark model. The average returns and net alphas are presented for the hypothetical long, short and long-short equal-weighted portfolios. Panel A contains results using the original sample period and Panel B contains results using the forward-extended. Unless specified otherwise, here and throughout the remainder of the study, all return and net alpha estimates are reported in percent per year.

The estimates in both the sample periods provide substantial evidence that the activeness proxies contain predictive information regarding future cross-sectional fund performance. I note that in both the sample periods, the average net returns for long side are greater than the short side for nine out of 11 strategies. The only exceptions being strategies formed using *Duration* and *Contrarian*. After adjusting for the exposures to the four factors in the benchmark model, all the 11 strategies generate positive net alpha in both the sample periods. For instance, a long-short strategy that consists of the funds from the lowest R^2 quintile in the long side and funds in the highest R^2 quintile in the short side earns an average return of 2.59% (1.49%) in the original sample period (forward-extended sample period). The same strategy generates a net alpha of 1.76% (1.18%) in the original (forward-extended) sample period. Similarly, a long-short strategy formed using *Volatility* earns net alpha of 5.00% (5.10%) in the original (forward-extended) sample period. Comparing the average returns and net alpha in both the sample periods, I observe that extending the sample period of the original study forward to December 2014 seems to have little

later in the out-of-sample tests.

to no impact on the performance of the activeness strategies. In fact, for five out of 11 strategies, the net alpha in the forward-extended sample period is higher than the net alpha in the original sample period.

Note that the above mentioned estimates of net alpha for the activeness strategies may not be exactly same as those in the original studies that introduce these strategies. In estimating the in-sample performance of the strategies, the aim is not to achieve a perfect to the decimal-point estimate of the net alphas from the original study. Perhaps, as McLean and Pontiff (2016) (p. 9) point in their replication of asset pricing anomalies, “such task could be impossible since the CRSP data changes over time”. In addition, the original studies often use different criteria in selecting a sample of actively managed domestic equity mutual funds. For example, some studies manually check the fund names to eliminate index funds and other studies use a delete funds whose name includes some specific words such as “index” “S&P”, “DOW”, “Wilshire”, and “Russell”. In addressing the incubation bias, Amihud and Goyenko (2013) eliminate fund-month observations with total net assets below \$15 millions, Cremers and Pareek (2016) use a screen of \$10 millions, Huang et al. (2011) use a screen of \$5 millions, and Kacperczyk et al. (2005) or Koch (2016) do not use a size screen at all. Similarly, to limit their sample to domestic equity funds, the original studies require the sample funds to contain an average of anywhere between 60.00% to 80.00% of the portfolio to be held in domestic equities. As such, the results in the original studies are based on a sample of fund-month observations that is different from the sample in this study (and from other studies) and hence, the results in this paper may not match to the original studies exactly.

Furthermore, and most importantly, the original studies also differ in portfolio formation process and benchmarking methods. In forming portfolios, the original studies use different types of sorting procedures (using quintiles vs. deciles sorts) and rebalancing frequency (i.e. monthly, quarterly, vs. annually). Although all the original studies that I consider here use the Carhart (1997) four-factor model as the primary benchmark, some studies use time-series averages of net alpha estimates of individual funds calculated using lagged beta loadings and others use a full regression approach similar to the one used here.

Nevertheless, despite the above noted differences in the sample selection, portfolio formation process and benchmarking methods, the in-sample results consistently produce positive net alpha estimates for all the 11 activeness strategies. These results capture the essence of the original studies which is “proxies of managerial activeness predict cross-sectional differences in net of fees mutual fund performance”. Furthermore, to compensate for the differences in methods allowing for the maximum benefit of doubt to the original studies and increasing hurdle for the tests throughout the remainder of the paper, I use a one-tailed test with 10% significance level for the long-short net alpha. That is, I deem the predictor to be informative of future fund performance if the t-statistic of the long-short net alpha is above 1.28.¹⁴

Using the above relaxed criteria for significance tests, results in Table 2.3 show that out of 11, eight (nine) long-short strategies have significant net alpha in the original sample period (forward-extended sample period). The strategy based on *volatility gap* has a significant long-short net alpha in the extended sample period but not in the

¹⁴I use one-tailed test because the goal is to determine whether the funds in long side of the strategy outperform the funds in short side of the strategy. This is inherently different than testing whether the long-short strategy itself has an alpha that is significantly different from zero.

original sample period. Strategies based on *Industry Concentration* and *Contrarian*, do not produce a significant long-short net alpha in either of the sample periods.¹⁵ I drop these two predictors from the remaining tests and proceed with the remaining nine that significantly predict cross-sectional net of fees fund performance in either of the two sample periods. Therefore, throughout the rest of the paper, the sample of predictors consists of R^2 , *Duration*, *Volatility*, *Contemporaneous Herd*, *volatility gap*, *Return Gap*, *Active Weight*, *Active Share*, and *Leading Herd*.

2.4.2 Concentrated In-Sample Performance

I continue with the nine strategies that have a significant in-sample net alpha and subject them to the second robustness criteria in Proposition 2.1.2. This proposition entails that performance of a robust investment strategy should not only come from a short part of the in-sample period. Instead, the performance should be comparable across different sub-periods that constitute the in-sample period.

To examine such a possibility of concentrated in-sample performance, I plot the time-series of returns for the long-short portfolios of the nine investment strategies over the forward-extended sample period in Figure 2.1. The shaded area in the each of the plots refers to a three year time period from January 1999 to December 2001 or the anomalous time period. These return plots show an interesting patterns during the anomalous time period that is consistent across all the nine strategies. First, this three year period is marked by extremely large and positive (as well as negative in

¹⁵This means that 82.81% of the strategies that I consider have significant in-sample net alpha. This success rate is similar to that of McLean and Pontiff (2016) who report that 88.00% of the strategies that they consider produce significant net alphas in-sample. It is also similar to Jones and Mo (2017) who replicate 16 strategies based on mutual fund performance predictors and note that 75% of the strategies produce significant in-sample net alpha.

some cases) return spikes and as such, performance of the activeness strategies appears to be extremely volatile during this time period. Second, the highest returning month during the entire forward-extended sample period always appears to be during these three years. In fact, for strategies formed using *volatility gap* and *Leading Herd* all five out of five highest returning months over the entire sample period fall during the above mentioned three years and for strategies formed using R^2 , *Duration*, *Contemporaneous Herd*, *Contemporaneous Herd*, *Return Gap*, and *Active Weight* this proportion is four out of five. For example, consider the excess returns on long-short portfolios formed on *Return Gap*. This strategy generates 11.57% in February 2000, the highest ever return for this strategy. The highest returning month being outside the anomalous time period is December 1993 with a considerably lower return of 2.97%.

Figure 2.2 plots the cumulative net excess returns for the same nine strategies. The cumulative net excess returns represent the growth of \$1 borrowed at risk-free rate and invested in the nine strategies at the start of the respective forward-extended sample periods. Same as in Figure 2.1 the shaded area in the each of the plots represents a three year time period from January 1999 to December 2001. These plots show patterns that are similar to the ones mentioned above. For six out of nine strategies (R^2 , *Volatility*, *Contemporaneous Herd*, *Return Gap*, *Active Weight* and *Active Share*) the steepest period investment growth overlaps with the aforementioned three year period. In addition, for the same six strategies, the investment growth rate appears to be almost flat outside these three years. For every strategy except *volatility gap* and *Duration*, the investment growth rate during this three year period is positive.

Both the return plots and the cumulative investment growth plots of the nine

strategies evince a highly concentrated positive performance for the nine activeness strategies from January 1999 to December 2000. In addition, it also appears that the performance of these strategies remains mostly flat outside this short time period. Although helpful in terms of identifying unusual patterns in returns, these plots do not tell the reader anything about the risk-adjusted performance. Before I conclude on the inconsistent in-sample performance, I must analyze the abnormal returns of the nine activeness strategies during and outside this short anomalous three year time period. Therefore, I re-estimate the in-sample net alphas for each of the nine activeness strategies during and outside the anomalous three year time period using a conditional as well as an unconditional modeling approach. The unconditional model assumes constant factor exposures throughout the entire in-sample period and is estimated using eq. (2.1). On the contrary, the conditional model allows factor exposures to be different for the anomalous three year sub-period and rest of the sample period and is estimated using eq. (2.2).

$$R_t = \alpha_b d_b + \alpha_{nb} d_{nb} + \sum \beta_k F_{k,t} + \varepsilon_t \quad (2.1)$$

$$R_t = \alpha_b d_b + \alpha_{nb} d_{nb} + \sum \beta_{k,b} F_{k,t} d_b + \sum \beta_{k,nb} F_{k,t} d_{nb} + \varepsilon_t \quad (2.2)$$

Here, d_b is a dummy variable which equals to one from January 1999 to December 2001 and is zero otherwise, d_{nb} is a dummy variable that equals zero from January 1999 to December 2001 and is one otherwise, β_k , $\beta_{k,b}$ and $\beta_{k,nb}$ are the unconditional factor loadings, conditional factor loadings during the time period January 1999 to December 2001, and conditional factor loadings during outside the above three years,

and $F_{k,t}$ represent excess returns on market, size, value and momentum factors.

The parameters of interest here are α_b and α_{nb} and they each represent the net alpha for the strategy under consideration during and outside the three year anomalous time period. Note that in estimation of α_b , I mainly use 36 months of data. Therefore, it might be the case that the estimates of α_b are not as efficient as estimates of α_{nb} (which are estimated using a much longer time-series) and are marked by high standard errors. Therefore, I primarily focus on the significance of α_{nb} , i.e. the net alpha of the strategy excluding any observations from January 1999 to December 2001. If α_{nb} is significant, regardless of the significance of α_b , I can conclude that the performance of the strategy under consideration is robust to excluding the anomalous three year time period.

To clarify any confusion regarding the terminology used here, note that I use the concept of conditional modelling in the context of different factor exposures in different sub-sample periods. This differs from the commonly adopted conception of conditionality which involves using predictive instrumental variables to capture the time-variation in the factor exposures, the risk premia associated with the factors, or both (e.g., Ferson and Mo (2016) and Cederburg et al. (2018)) based on the contemporaneous realizations of states of the world. The conditional modeling approach assumes constant factor exposures for the entire sub-sample periods regardless of the states of the world. It simply allows for the possibility that the out-of-sample factor exposures could differ than the in-sample ones. In addition, at this point I would also like to point out that, the conditional modelling approach is exactly equivalent to that of running two separate unconditional regressions in the two sub-sample periods.

Similar to Section 2.4.1, the tests are conducted using the original sample period and the forward-extended sample period. In each of the sample periods, I estimate the net alpha for the nine activeness strategies during and outside the three year anomalous period from January 1999 to December 2001. The results are presented in Table 2.4. Panel A reports the results from the original sample period and Panel B reports results from the forward-extended sample period. All net alpha estimates are in percentage per year. In both the panels, I first present results from an unconditional model followed by results from the unconditional model.

I start with performance of the investment strategy formed using R^2 . Results from Table 2.3 show that in twenty-one years of the original sample period of this strategy, from January 1990 to December 2010, this strategy yielded an annual net alpha of 1.76% with a t-statistic of 2.04. The unconditional (conditional) estimates in Table 2.4 Panel A, however, show that same strategy generated an impressive annualized net alpha of 12.53% (10.37%) during the three years from January 1999 to December 2001. During the remaining nineteen years, the annual net alpha on this strategy estimated using an unconditional (conditional) model was 0.02% (0.34%) with a t-statistic of 0.00 (0.41), almost close to zero. Comparing results from Panel B show that in the forward-extended time period, the R^2 strategy generates a net alpha of -0.40% (-0.13) using the unconditional (conditional) approach with a t-statistic of -0.5 (-0.17). This indicates that the R^2 strategy mostly fails to produce any abnormal return whatsoever outside a short three year anomalous time period.

A similar pattern is observed for strategies as well. On average, the nine strategies generate net alpha of 5.52% (3.81%) during the anomalous time period when

estimated using unconditional (conditional) model. However, the average net alpha outside this three year period is 0.86% (0.78%), only about 16% (20%) of the net alpha during the anomalous time period. Furthermore, the average reduction in net alpha calculated using the entire original sample period (i.e., the alpha from Table 2.3) to net alpha in sub-period of the original sample period that is outside the three year anomalous period (α_{nb}) is about 42%. This reduction is about 47% if I use the forward-extended sample period.

Only two out of the nine strategies, *Duration* and *Volatility*, show significant net alpha outside the three year anomalous time period in either of the estimations. However, for strategy based on *Volatility*, the net alpha during the three years of anomalous time period is roughly double of the net alphas during the rest of the time period. Therefore, even the strategy formed using *Volatility* exhibits a significantly concentrated performance during this time period.

To summarize, in both Panel A and Panel B, I use combination unconditional and conditional estimations in the original sample period as well as the forward-extended sample period and conduct four different tests. The results show that for most of the activeness strategies, the net alpha outside the anomalous three year period from January 1999 to December 2001 is insignificant. Out of the nine different activeness strategies, only the portfolios formed using *Duration* and *Volatility* yield a significant net alpha in all four tests. None of the other predictors is robust to all four model specifications.

Note that the three years from January 1999 to December 2001 are also known for surrounding the peak of the late 1990's dot-com bubble and therefore, represent

a unique time period in the history of financial markets and money management industry in the United States. Using data from 1871 to 2000, Shiller (2015) notes that the S&P 500 price appreciation was the highest from 1992 to 2000, and steepest in the last two years of the decade. This run-up was even steeper than the price appreciation in stock market prior to the 2008 financial crisis. Not surprisingly, this run-up was followed by an even steeper decline after the burst of the bubble in March 2000.

Analyzing the role of institutional investors in this unprecedented run-up, Brunnermeier and Nagel (2004) and Griffin et al. (2011) point that various institutional investors such as banks, insurance companies, hedge funds, mutual funds, and other independent investment advisors aggressively overweighted hi-tech stocks in their portfolios during the last two years of the run-up. They also note that institutional investors dramatically reduced their portfolio exposures to hi-tech stocks towards the end of the run-up (and perhaps in some cases, after the burst). Similarly, studying the effect of managerial incentives on herding behavior and bubbles, Dass et al. (2008) argue that portfolios of high incentive managers were overweighted in hi-tech stocks during the run-up and underweighted prior to the burst. On the contrary, the portfolios low incentive showed an opposite pattern. Although each of the studies reach different conclusions regarding the cause of the dot-com bubble, they point that there was a substantial cross-sectional deviation in percent hi-tech stocks held by mutual fund managers.

The findings from the above studies directly relate to some of the managerial activeness proxies. Consider the following illustration. A sub-group of active funds

with portfolios overweighted in hi-tech stocks during the run-up of the bubble would also appear to deviate far from the market portfolio. Unsurprisingly, such a group of managers will tend to rank very highly on some the activeness measures such as *Active Weight* (defined as the absolute differences between market portfolio weight and the fund portfolio weight, summed across all the holdings). Furthermore, it is not surprising that this group of managers would have done extremely well provided they rapidly reduced exposures to hi-tech stocks right around the burst of the bubble. On the other hand, a sub-group of fund managers that ranked low on activeness measures would be following an opposite strategy. Overall, a hypothetical long-short portfolio of mutual funds formed on active weight would do extremely well during this period.

In the above illustration, correctly anticipating the steep run-up or decline of the hi-tech stocks can be classified as sector timing and therefore, skill. Consequently, one can argue that the managerial activeness proxies such as *Active Weight* are extremely useful to investors in terms of selecting such a skilled group of fund managers, at least during the dot-com bubble.

The same illustration, however, begs the question. Why do these proxies fail to predict cross-sectional differences in fund performance in either before or after the three year period surrounding the peak of the dot-com bubble? Perhaps, the predictive power of the activeness proxies is limited to extremely volatile economic conditions marked by sharp market run-ups and declines. However, if this is the case, why didn't the same activeness proxies predict fund performance equally well during (or before) the financial crisis of 2008? I try to answer these questions in the Section 2.4.5.

2.4.3 Out-of-Sample Tests

Proposition 2.1.3 calls for out-of-sample performance tests. A standard approach to assessing out-of-sample performance is to compare the in-sample performance to either performance in a sample period that starts after the original sample period (post-sample period) or a combination of the sample period that is prior and after the original sample period (pre- and post-sample period). If the strategy's out-of-sample performance is comparable to in-sample performance, it is deemed to be robust to both sample periods.

Following this, I use two different sample periods that partly overlap with the original sample period and estimate the in- and out-of-sample net alpha of the long-short portfolios described in the previous section using the Carhart (1997) four-factor model as the benchmark model. First, to conduct a post-sample performance evaluation, I use the forward extended sample period which is comprised of the original sample period as well as the post-sample period and estimate the net alpha in both of these sub-periods. Second, to conduct a pre- and post-sample performance evaluation, I use the backward and forward-extended sample period which is comprised of the pre- and post-sample periods as well as the original sample period and estimate the net alpha in both of these sub-periods. In both these sample periods, the net alpha for the sub-period that consists of the original sample period represents the in-sample performance of the strategy. Similarly, the net net alpha estimates for the sub-period that is outside the original sample period is considered as the out-of-sample performance of the strategy.

Similar to previous section, to estimate net alphas for each of the sub-sample

periods, I use a conditional as well as an unconditional modeling approach. The net alphas estimated using the unconditional model assumes that the factor exposures remain constant for the in-sample and the out-of-sample sub-periods. In this approach I estimate the in-sample alpha (α_{in}) and the out-of-sample alpha (α_{out}) for each of the long-short portfolios using the following regression:

$$R_t = \alpha_{in}d_{in} + \alpha_{out}d_{out} + \sum \beta_k F_{k,t} + \varepsilon_t \quad (2.3)$$

On the contrary to the unconditional model, the conditional model allows for different factor exposures for the in-sample and the out-of-sample sub-periods. In this approach, I estimate the in-sample net alpha (α_{in}) and out-of-sample net alpha (α_{out}) using the following regression:

$$R_t = \alpha_{in}d_{in} + \alpha_{out}d_{out} + \sum \beta_{k,in}F_{k,t}d_{in} + \sum \beta_{k,out}F_{k,t}d_{out} + \varepsilon_t \quad (2.4)$$

Here, in both the equations, R_t is the excess long-short portfolio returns net of fees, $F_{k,t}$ are the factor returns, d_{in} is a dummy variable which equals to one in-sample and is zero otherwise, d_{out} is a dummy variable that equals to one out-of-sample and is zero otherwise¹⁶, β_k , $\beta_{k,in}$ and $\beta_{k,out}$ represent the unconditional, conditional in-sample, and conditional out-of-sample factor loadings respectively.

Table 2.5 presents the results from the out-of-sample tests. The individual sample periods for each of the strategies are also reported. In each of the panels, the first four rows consist of estimates and t-statistics from an unconditional model and the four rows consist of estimates and t-statistics from the unconditional model.

¹⁶That is, when I estimate both these models using the forward-extended sample period, d_{in} (d_{out}) is one (zero) for observations that are in the original sample period (post-sample period) and is zero otherwise. Similarly, when I use the backward and forward-extended sample period, d_{in} (d_{out}) is one (zero) for observations that are in the original sample period (pre- and post-sample period) and is zero otherwise.

Panel A presents the results of the post-sample analysis in which I use the forward-extended sample period. I observe remarkable reduction from the in-sample to the out-of-sample net alpha for the nine strategies. For example, consider the strategy based on R^2 . Using the unconditional (conditional) approach, the in-sample net alpha is estimated to be 1.93% (1.76%) with a t-statistic of 2.28 (2.12) suggesting that during the period from January 1990 to December 2010, R^2 is a statistically as well as economically significant predictor of the cross-sectional net of fees performance. However, this same strategy generates an unconditional (conditional) out-of-sample net alpha of -2.81% (-3.40%). Therefore, the reduction from in-sample to out-of-sample is about 250% for the unconditional model and 290% for the conditional model. This suggests that out-of-sample, R^2 is negatively related to future fund performance.

I observe similar differences between in-sample and out-of-sample performance of the other strategies under consideration. On average, the in-sample net alpha of the strategies estimated using the unconditional model is 1.52% and ranges from 0.79% to 5.02%. On the contrary, their out-of-sample counterparts average to 0.64% and range from -2.81% to 6.18%. This accounts for an average reduction of about 60% from in-sample to out-of-sample. In addition, two out of the nine strategies under investigation produce negative out-of-sample net alphas. On the contrary to these results, only five strategies produce out-of-sample net alphas that are lower than the in-sample net alphas. However, none of the out-of-sample net alphas are significant.

These differences are even more apparent using the conditional modeling approach. The average of the in-sample net alphas estimated using the conditional model is

1.47% and ranges from 0.74% to 5.00%. In contrast, the out-of-sample net alphas produced by these strategies average to -0.13% and fall in the range from -3.40 to 1.44% . Following this, the percent reduction from in-sample to out-of-sample averages to about 100%. In addition, three out of the nine strategies produce negative out-of-sample net alphas and seven out of the nine strategies yield lower out-of-sample net alphas compared to their in-sample counterparts. Similar to the unconditional model, none of the out-of-sample net alphas are significant in a one-tailed test at 10% significance level.

Panel B reports the results for pre- and post-sample analysis. Here, I use the backward- and forward-extended sample periods, i.e., January 1984 to December 2014. The only exception is strategy based on *Volatility*. For this strategy, the sample period in this panel is from September 1999 to December 2014. Comparing results of this panel to those in Panel A, I see an even worse out-of-sample performance. For net alphas estimated using both the unconditional and the conditional models, the average reduction from in-sample to out-of-sample net alpha is approximately 90%. Four (two) out of nine strategies yield negative out-of-sample net alphas using the unconditional (conditional) modelling approach. In addition, for the net alphas estimated using the unconditional model, only two out of nine strategies have higher out-of-sample net alphas than the in-sample net alphas. For the net alphas estimated using the conditional model, none of the out-of-sample net alphas are higher than the in-sample net alphas. Same as before, none of the out-of-sample net alphas estimated using either the unconditional or the conditional model are significant.

In short, the results above show that the strong in-sample net alpha for the nine

strategies is, on average, reduced by approximately between 60% to 100% depending on the estimation model and sample period used. In addition, none of the nine activeness strategies generate significant out-of-sample net alpha. Recall that the results in Section 2.4.2 show that, even in-sample, the nine activeness proxies largely lack failed to predict fund performance outside a short three year time period from January 1999 to December 2000. Both the lack of consistent in-sample performance accompanied with poor out-of-sample performance suggest that the predictive power of activeness proxies was specific to the economic conditions that occurred from 1999 to 2001. The out-of-sample tests also show that this unusual and short spur during this time period does not repeat itself in either the pre- or the post-sample periods.

2.4.4 Performance relative to a Five-Factor Model

Proposition 2.1.4 states that the performance of a robust investment strategy should not be explained away by benchmark models other than the traditionally used Carhart (1997) four-factor model. As I mention previously, the Fama and French (2015) five-factor model serves as a more comprehensive benchmark for an alternative investment opportunity set of a mutual fund investor than the four-factor model and therefore, in this section, I benchmark the nine activeness strategies using the five-factor model.

Table 2.6 presents the results. I also report the net alphas estimated using the four-factor model from Table 2.3 for comparison. Panel A presents results using the original sample period and Panel B presents results using forward-extended sample period. In both the panels, for eight out of nine strategies the net alphas estimated using the five-factor model are lower than those estimated using the five-factor model.

For example, the strategy formed using *Volatility* has a net alpha of 5.00% (5.10%) in the original (extended) sample period with a t-statistic of 2.70 (2.92) when estimated using a four-factor model. The net alpha for the same strategy, however, is -0.03% (0.33%) in the original (extended) sample period when benchmarked with a five-factor model. This represents close to a 100.00% reduction regardless of the sample period under consideration.¹⁷ In both the sample periods, three out of nine strategies produce a negative net alpha. In addition, the average reduction in the four-factor net alpha to the five-factor is approximately 84.00% (79.2%) in the original (extended) sample period.

In both the sample periods only the strategies formed using *Return Gap* and *Active Weight* generate a significant five-factor net alpha. Notably, the five-factor net alpha for strategy based on *Return Gap* is almost double of the four-factor net alpha in both the sample periods. The four-factor and five-factor net alphas of strategies formed on *Active Weight* are approximately the same in both the sample periods. The rest of the seven strategies perform considerably worse in terms of both, economical and statistical significance of the net alpha when benchmarked with the five-factor model. In addition, the five-factor net alphas for the remaining seven strategies are insignificant regardless of the sample period under consideration.

One could argue that the insignificance of the net alphas estimated from the five-factor model is attributable to a bad model fit, and therefore, inefficiently estimated high standard errors. Both the panels in Table 2.6 report the adjusted- R^2 from the four-factor as well as the five-factor model regressions. For strategies formed

¹⁷This finding is in-line with that of Jordan and Riley (2015) who show that the net alpha of investment strategy formed using *Volatility* are largely attributable omitted exposures to profitability and investment factors.

using *Duration* and *Return Gap* the model fit does not improve with the five-factor model. For the remaining six strategies, in both the sample periods, the adjusted- R^2 from the five-factor model is at least marginally higher than the one from the four-factor model. This affirms that the worse performance of the activeness strategies when benchmarked with the five-factor models is not because of poor model fit. Furthermore, this is also supported by the fact that the difference between magnitude of the four-factor and five-factor net alpha is economically meaningful and cannot be simply a result of inefficient standard errors.¹⁸

The results above suggest that the strong in-sample net alpha for the activeness strategies is mostly attributable to omitted benchmark exposures to profitability and investment factors from the five-factor model. To demonstrate that the importance of accounting for these two omitted factors in benchmarking these strategies, I report the factor exposures estimated using the five-factor model in Table 2.7. Panel A presents the results using the original sample period of the study and Panel B presents the results using the forward-extended sample period.

In both the sample periods, all the strategies have a significant and positive exposure to either profitability or the investment factor. For example, consider the five-factor exposures of the strategy formed using R^2 . This portfolio has a profitability factor exposure (β_{RMW}) of 0.12 (0.11), and an investment factor exposure (β_{CMA}) of 0.15 (0.12) in the original (forward-extended) sample period. Both the exposures betas are highly significant with t-statistics of 3.91 and 3.35 (3.83 and 2.81). In addi-

¹⁸In the next section, I show that for the three strategies based on *Duration* and *Return Gap* the lower five-factor model adjusted- R^2 is a result of not including the momentum factor. A six-factor model that includes momentum produces better model fit. Nevertheless, the six-factor model does not produce a significant reduction in the net alpha of the activeness strategies over the five-factor model. Therefore, I use the five-factor model as the primary benchmark.

tion, these exposure betas are higher than the exposures to the market factor, and the value factor in magnitude. Since both these exposures are positive and significant, unsurprisingly, when either of the profitability or the investment factors perform well, so does the R^2 strategy. A four-factor benchmark model ignores these exposures and falsely deems the R^2 strategy to have a significant net alpha. Similarly, the remaining strategies also have positive exposure to profitability and investment factors with the only exception being the strategy formed using *Return Gap*. The portfolio formed using sorts on *Return Gap* has a negative exposure to the profitability factor.¹⁹ A similar pattern is observed across all the other activeness strategies as well.

To summarize, the traditionally used four-factor model grossly mismeasures performance of the nine strategies and falsely identifies the managerial activeness proxies as predictors of future cross-sectional net of fee fund performance. The five-factor model cures the systematic problems of the four-factor model by accounting for omitted exposures to profitability and investment factors. In unreported results, I also find that an alternative five-factor model that uses the three factors based on one way sorts from the Fama and French (1993) and the profitability and investment factors from the Fama and French (2015) five-factor model does equally well at explaining the performance of long-short portfolios formed using the predictor variables.²⁰

¹⁹I use a two-tailed test for the factor exposures since under the alternate hypothesis, any strategy has non-zero exposure to the benchmark factors. Therefore, I deem the betas to be significant if the t-statistic is higher than 1.67 (significance level of 10%).

²⁰The construction of size and value factors in the original three-factor model is based on one-way sorts which is slightly different from that in the updated five-factor model in which these factors are constructed using conditional four-way sorts.

2.4.5 Concentrated performance and omitted factors

In the previous section, I show that the activeness strategies do not generate significant net alpha when benchmarked with the five-factor model. In addition, these strategies have positive and significant exposures to the profitability and investment factors. In this section, I show that the concentrated performance of the activeness strategies from January 1999 to December 2001 is partly attributable to not including profitability and investment factors in the four-factor model.

Figure 2.4.5 plots the excess returns on each of the five factors from the Fama and French (2015) model. The sample period is January 1984 to December 2014. The grey area in the figure refers to the anomalous three year period from January 1999 to December 2001. The return plots show that this three year period is also marked by extreme returns on the profitability and investment factors. In fact, during the entire time period from January 1984 to December 2014, five out of five highest returning months for the profitability factor and four out of five highest returning months for the investment factor fall during the anomalous three year period. Coincidentally, any funds that maintain high exposures to either of these factors during this time period would also have appeared to do extremely well. The common four-factor benchmark that does not include profitability and investment factors would misclassify such fund managers as skilled.

The results in Section 2.4.2 show that the predictive power of the activeness proxies is largely concentrated in a short unusual time period from January 1999 to December 2000. The above findings suggest that the main source behind the extremely positive and concentrated performance of the activeness strategies during this anomalous time

period is omitted benchmark factors. The traditionally used four-factor model does not account for positive exposures of the activeness strategies to profitability and investment factors. Both the profitability and investment factors perform extremely well generating highest ever returns during the anomalous time period. Consequently, not accounting for exposures to these factors, the nine activeness strategies appear to generate positive abnormal returns using the four-factor model as the benchmark. Therefore, either excluding this anomalous time period from sample or adjusting for exposures to profitability and investment factors, most of the activeness strategies do not seem to generate a significant net alpha.

2.4.6 Alternative Benchmark Models

In this section, I consider alternative benchmark models that also use the information from the profitability and investment factors. The results are presented in Table 2.8. Panel A presents the results using the original sample period and Panel B presents results using the forward-extended sample period. Similar to Table 2.6, for comparison, I also present the net alpha and adjusted- R^2 from the four-factor model. I consider three other alternative benchmark models described below.

The first benchmark is a six-factor model (similar to Jordan and Riley (2016) and Sheng et al. (2017)) that adds the momentum factor of Carhart (1997) to the five-factor model and simultaneously accounts for the exposures to profitability, investment and momentum factors. This model performs consistently better than the four-factor model and the five-factor model in terms of the model fit across all nine activeness strategies. The average adjusted- R^2 from the six-factor model is approximately 57% (55%) in the original (forward-extended) sample period. The four-factor

model and the five-factor models produce an average adjusted- R^2 of approximately 52% (50%) and 54% (52%).

Similar to the net alpha from the five-factor model, only strategies formed on *Return Gap* and *Active Weight* generate significant net alphas in both the sample periods mentioned above. The average reduction in from the four-factor net alphas to the six-factor net alphas for the nine strategies is 72.00% (67.00%) in the original (extended) sample period. This average reduction is lower than the average reduction from four-factor model to five-factor model which equals of 84.00% (79.2%). This suggests that although the six-factor model generates a marginally better model fit than the five-factor model, it does not produce any meaningful reduction in the net alpha the nine activeness strategies over the five-factor model.

Next, I consider the q-factor model of Hou et al. (2015). This model consists of a market-factor, a size factor, a profitability factor and an investment factor. As opposed to the five-factor model, the q-factor model does not include a value factor. The construction of the factors in the q-factor model is slightly different than the five-factor model. For example, in the factors are rebalanced monthly in the q-factor model, whereas, in the five-factor model, the factors are rebalanced annually. In addition, Hou et al. (2015) use return on equity to build the profitability factor whereas, Fama and French (2015) use operating profitability.

The results from benchmarking with the q-factor model are similar to that of benchmarking with the five-factor model, only slightly weaker. For example, two (three) out of nine strategies continue to have significant long-short net alpha in the original (extended) sample period when benchmarked with the q-factor model.

However, the average reduction in net alpha from the four-factor model to the q-factor model is considerably lower than the one to the five-factor model. The q-factor model, on average, reduces the four-factor net alpha by about 55% in both the original and the extended sample periods, about 30% less reduction compared to the five-factor model. The model fit measured by the adjusted- R^2 is also lower than the five-factor model. In untabulated results, I also find that similar to the six-factor model, adding a momentum factor to the q-factor model does very little to improve benchmarking of these portfolios.

Finally, I benchmark the nine activeness strategies using the mispricing-factors model of Stambaugh and Yuan (2017). This model consists of a market factor, a size factor, a management factor and a performance factor. The management and performance factors aggregate information across 11 prominent asset pricing anomalies from Stambaugh et al. (2012) by averaging rankings of stocks within clusters exhibiting the highest co-movement returns. Broadly, they classify the 11 anomalies into two different clusters. The management factor is constructed by aggregating information in a cluster which contains anomalies related to net stock issues, composite equity issues, accruals, net operating assets, asset growth, and investment-to-assets. Whereas, the performance factor is constructed by aggregating information in a cluster which contains anomalies related to distress, o-score, momentum, gross profitability, and return on assets. Therefore, although the mispricing factors model does not explicitly include profitability and investment factors, it still includes information that is similar to these factors.

I find that benchmarking the activeness strategies with the mispricing-factors

model, overall reduction in net alpha from the four-factor model to the mispricing factors model is, on average, about 45% (51%) in the original (extended) sample. Similar to the q-factor model, this model also generally produces a lower adjusted- R^2 than the five-factor model.

All of the benchmarking models that I consider here account for the profitability and investment factors in some form. These alternative benchmark models produce results similar to that of the five-factor model. Overall, when benchmarking with the three alternative models, two to three out of nine activeness strategies generate significant net alpha. However, the alternative benchmark models produce generally lower model fit as compared to the five-factor model. The only exception being the six-factor model that adds momentum to the original five-factor model. The six-factor model serves as a marginally better benchmark in terms of model fit, however, it does not produce economically different results than the five-factor model. These results confirm that the four-factor model grossly mismeasures the performance of the nine strategies due to the omitted benchmark factors.

2.5 Conclusion

The studies on managerial activeness offer investors useful tools to guide their capital allocation decisions while making investments in mutual funds. These studies measure activeness as deviation from benchmark and claim that their proposed measure has the ability to predict cross-sectional differences in net of fee mutual fund performance. Therefore, forming fund selection strategies using ex-ante observable characteristics, these studies essentially provide investors with ways of easily identifying positive net present value opportunities in mutual fund industry.

I systematically reevaluate performance of 11 such strategies formed using activeness proxies under four robustness criteria. Starting with testing replicability under consistent set of methods, I show that nine out of the 11 strategies generate significant in-sample net alpha. Taking a closer look at these nine strategies, I find that their performance is highly concentrated in a short unusual three year time period. Excluding this period from in-sample estimation, two out of nine strategies generate significant net alpha. Next, I also find that none of the nine strategies have significant out-of-sample net alpha. Finally, I show that the large in-sample net alpha is mismeasured and is attributed to omitting profitability and investment factors from the benchmark model. None of the 11 proxies that I consider are robust to all four of the above tests. I further also show that the concentrated performance during the anomalous three year time period is also due to not including profitability and investment factors in the benchmark model. Together, the results from these four tests evince that managerial activeness and skill are largely unrelated.

The main finding of this paper is that the predictive power of activeness proxies is largely a result of outliers in the historical data and/or misspecified benchmark model. Both these effects together can cause severe biases in inferences from performance attribution studies and therefore, call for re-examination of other performance predictors documented in the mutual fund literature. For example, the literature has identified plenty of other ex-ante observable fund characteristics other than activeness (e.g. holding characteristics similarities (Cohen et al. (2005)), sentiment exposure (Massa and Yadav (2015)), location of the fund (Christoffersen and Sarkissian (2009)), incentive structures (Massa and Patgiri (2008)) that predict future fund per-

formance. These characteristics supposedly can help investors select a skilled group of fund managers and earn positive net of fees risk adjusted return. If the findings in these studies hold a four-stage battery of tests proposed in this paper, they would make an extremely valuable comment regarding the existence of managerial skill.

In addition to predicting fund performance, there are several empirical facts in the mutual fund literature which also could be susceptible to a mismeasured performance. For example, convex relationship between flow and performance (Sirri and Tufano (1998)), and the non existence of diminishing economies of scale in performance at fund level (e.g. Reuter and Zitzewitz (2013) and Pástor et al. (2015)) can also be a result of mismeasured performance. Perhaps, the battery of tests proposed in this paper could prove useful in reevaluating similar evidence documented in past research as well as for the future research in mutual funds.

Table 2.1: List of Performance Predictors

This table lists the mutual fund performance predictors used in this study. The sample of performance predictors is limited to the ones that proxy for managerial activeness. For each predictor variable, its abbreviation, the source in the academic literature, the sample period from the original study and the relationship to future fund performance are reported. All variables are defined in Appendix B.1.

Predictor	Source	Original Sample Period	Long Quintile	Short Quintile
R^2 (%)	Amihud and Goyenko (2013)	1990-2010	low	high
Duration, DUR (years)	Cremers and Pareek (2016)	1990-2013	low	high
Volatility, VOL (% per month)	Jordan and Riley (2015)	2000-2013	low	high
Contemporaneous Herd, CH (%)	Koch (2016)	1990-2009	low	high
Volatility Gap, VG (% per month)	Huang et al. (2011)	1983-2009	mid	high
Industry Concentration, IC (%)	Kacperczyk et al. (2005)	1984-1999	high	low
Return Gap, RG (% per month)	Kacperczyk et al. (2006)	1984-2003	high	low
Contrarian, Con	Wei et al. (2014)	1995-2012	high	low
Active Weight, AW (%)	Doshi et al. (2015)	1984-2013	high	low
Active Share, AS (%)	Cremers and Pareek (2016)	1990-2012	high	low
Leading Herd, LH (%)	Koch (2016)	1990-2009	high	low

Table 2.2: Summary Statistics and Correlations

This table reports summary statistics and correlations for the performance predictors listed in Table 2.1. The sample period for Volatility is January 2000-December 2014, and the sample period for the remaining characteristics is January 1984-December 2014. Panel A presents the mean, median, standard deviation, 10th percentile, and 90th percentile, and Panel B presents rank correlations. The statistics are computed for each month, and the values reported in the table are time-series averages.

Panel A: Distributional Statistics				
Predictor	P10	Mean	Median	P90
R^2 (%)	80.44	89.77	92.52	96.88
Duration, DUR (years)	1.80	5.77	4.93	11.18
Volatility, VOL (% per month)	0.99	1.28	1.24	1.64
Contemporaneous Herd, CH (%)	-3.72	11.26	11.36	24.81
Volatility Gap, VG (% per month)	-2.89	-1.81	-1.88	0.64
Industry Concentration, IC (%)	0.32	4.65	2.11	8.60
Return Gap, RG (% per month)	-1.67	0.02	0.02	0.42
Contrarian, CON	-1.44	-0.37	-0.37	0.69
Active Weight, AW (%)	27.75	40.88	40.93	54.90
Active Share, AS (%)	64.45	82.45	84.92	97.15
Leading Herd, LH (%)	-10.32	-2.35	-2.67	6.02

Panel B: Correlation Matrix											
Predictor	R^2	DUR	VOL	CH	VG	IC	RG	CON	AW	AS	LH
R^2 (%)	1.00										
Duration, DUR (years)	0.05	1.00									
Volatility, VOL (% per month)	-0.01	-0.19	1.00								
Contemporaneous Herd, CH (%)	-0.05	-0.07	0.07	1.00							
Volatility Gap, VG (% per month)	0.03	0.08	-0.02	-0.01	1.00						
Industry Concentration, IC (%)	-0.35	-0.03	0.28	-0.01	-0.05	1.00					
Return Gap, RG (% per month)	0.01	-0.05	0.04	-0.01	-0.10	0.06	1.00				
Contrarian, CON	-0.01	0.20	-0.17	0.01	0.00	-0.06	-0.06	1.00			
Active Weight, AW (%)	-0.12	-0.01	-0.10	-0.04	0.00	0.10	0.02	0.05	1.00		
Active Share, AS (%)	-0.51	-0.11	0.23	0.11	-0.08	0.37	0.05	0.04	0.09	1.00	
Leading Herd, LH (%)	0.03	0.01	-0.05	0.03	0.00	-0.03	-0.01	0.13	0.00	0.05	1.00

Table 2.3: In Sample Performance

This table presents the performance of long, short and long-short equal-weighted portfolios formed using quintile sorts of the lagged predictor variables. For R^2 , Volatility, Return Gap, and Volatility Gap portfolios are rebalanced monthly. For all other variables, portfolios are rebalanced on quarterly. The long and short quintiles for each predictor are presented in Table 2.1. R_t represents excess portfolio returns net of fees and $\bar{R}_t$ represents a time-series average of R_t . α is estimated using a regression of R_t on the Fama and French (1993) three factors and a Carhart (1997) momentum factor (the four-factor model). Panel A presents results using the sample period from the original study, and Panel B extends the sample period in Panel A through December 2014. $\bar{R}_t$ and α are reported in percentage per year and the corresponding t -statistics is in the parentheses. ** and * indicate significance at 5% and 10% level using a one tailed test.

Panel A: Sample Period from Original Study												
	R^2	DUR	VOL	CH	VG	IC	RG	CON	AW	AS	LH	
	1990-2010	1990-2013	2000-2013	1990-2009	1983-2013	1984-1999	1984-2003	1995-2012	1984-2013	1990-2012	1990-2009	
$\bar{R}_t$												
L	7.07** (2.15)	6.80** (2.24)	5.70** (1.68)	6.19** (1.85)	6.02** (1.99)	10.69** (2.53)	9.40** (2.25)	5.46* (1.43)	7.39** (2.51)	8.71** (2.45)	6.00** (1.75)	
S	4.48* (1.31)	7.25** (2.08)	2.36 (0.37)	5.47* (1.51)	5.69** (1.72)	9.79** (2.75)	6.70** (1.79)	6.86** (1.65)	6.72** (2.24)	5.93** (1.94)	5.28* (1.53)	
L-S	2.59** (2.47)	-0.44 (-0.55)	3.34 (0.81)	0.72 (0.96)	0.34 (0.39)	0.90 (0.61)	2.70** (2.34)	-1.40 (-0.73)	0.67** (1.99)	2.78** (1.89)	0.72 (1.12)	
α												
L	0.34 (0.40)	-0.49 (-1.21)	1.78** (2.05)	-0.28 (-0.44)	-0.56 (-1.14)	-0.10 (-0.11)	-0.06 (-0.08)	-0.61 (-0.85)	-0.42 (-0.92)	0.03 (0.04)	-0.30 (-0.46)	
S	-1.42 (-3.46)	-1.32 (-2.30)	-3.23 (-2.45)	-1.24 (-1.78)	-1.19 (-1.89)	-0.85 (-2.27)	-1.50 (-2.05)	-1.37 (-1.59)	-1.11 (-2.17)	-1.17 (-4.24)	-1.04 (-1.58)	
L-S	1.76** (2.04)	0.83** (2.10)	5.00** (2.70)	0.95* (1.61)	0.64 (1.07)	0.75 (0.88)	1.44** (2.06)	0.76 (0.67)	0.69** (2.03)	1.20* (1.59)	0.74* (1.33)	

(continued)

Table 2.3—Continued

Panel B: Sample period extended through December 2014												
	R^2	DUR	VOL	CH	VG	IC	RG	CON	AW	AS	LH	
	1990-2014	1990-2014	2000-2014	1990-2014	1983-2014	1984-2014	1984-2014	1995-2014	1984-2014	1990-2014	1990-2014	
$\overline{R}_t$												
L	7.43** (2.53)	6.85** (2.33)	5.88** (1.83)	7.36** (2.47)	7.37** (2.69)	7.81** (2.45)	8.82** (2.76)	6.79** (1.93)	7.40** (2.59)	8.57** (2.49)	7.18** (2.35)	
S	5.94** (1.94)	7.24** (2.15)	2.57 (0.43)	6.70** (2.07)	7.15** (2.39)	6.55** (2.46)	6.84** (2.32)	8.24** (2.16)	6.72** (2.30)	6.12** (2.08)	6.61** (2.14)	
L-S	1.49* (1.61)	-0.38 (-0.49)	3.31 (0.86)	0.66 (0.98)	0.22 (0.30)	1.26 (1.10)	1.98** (2.62)	-1.45 (-0.84)	0.68** (2.06)	2.45** (1.71)	0.56 (1.00)	
α												
L	-0.30 (-0.38)	-0.54 (-1.38)	1.67** (2.04)	-0.42 (-0.75)	-0.64 (-1.48)	-0.46 (-0.70)	-0.20 (-0.35)	-0.64 (-0.95)	-0.46 (-1.04)	-0.08 (-0.11)	-0.34 (-0.60)	
S	-1.49 (-3.92)	-1.38 (-2.48)	-3.43 (-2.71)	-1.25 (-2.02)	-1.32 (-2.34)	-1.14 (-4.14)	-1.36 (-2.65)	-1.31 (-1.64)	-1.17 (-2.35)	-1.18 (-4.40)	-1.08 (-1.83)	
L-S	1.18* (1.50)	0.84** (2.18)	5.10** (2.92)	0.82* (1.57)	0.68* (1.30)	0.68 (1.12)	1.15** (2.24)	0.67 (0.64)	0.71** (2.13)	1.10* (1.51)	0.75* (1.53)	

Table 2.4: Concentrated Performance

This table highlights the concentrated in-sample performance of the predictors during the three year period surrounding the end of Dot-Com bubble from January 1999 to December 2014. α_b is the net alpha on long-short portfolios during this time-period, and α_{nb} is the net alpha otherwise. α_b refers to performance during this time-period, and α_{nb} refers to performance otherwise. α_b and α_{nb} are estimated using two different regression models: $R_t = \alpha_b(d_b) + \alpha_{nb}(d_{nb}) + \sum \beta_k F_{k,t} + \varepsilon_t$ and $R_t = \alpha_b d_b + \alpha_{nb} d_{nb} + \sum \beta_{k,b} F_{k,t} d_b + \sum \beta_{k,nb} F_{k,t} d_{nb} + \varepsilon_t$. Here, R_t represents net excess returns of long-short portfolios, $F_{k,t}$ are factors from the the four-factor model, d_b is a dummy variable which equals to one from January 1999 to December 2001 and is zero otherwise, d_{nb} is a dummy variable that equals 1- d_b . β_k , $\beta_{k,b}$ and $\beta_{k,nb}$ are the unconditional factor loadings, conditional factor loadings from January 1999 to December 2001, and outside this time-period. Panel A presents the results for the sample period from the original study. Panel B extends this period forward to December 2014. α_b and α_{nb} are reported in percentage per year and the corresponding t -statistics is in the parentheses. ** and * indicate significance at 5% and 10% level using a one tailed test.

Panel A: Sample Period From Original Study									
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH
	1990-2010	2000-2013	1984-2003	1984-2009	1990-2013	1990-2013	1984-2013	1990-2009	1990-2009
$R_t = \alpha_b d_b + \alpha_{nb} d_{nb} + \sum \beta_k F_{k,t} + \varepsilon_t$									
α_b	12.53** (5.85)	-0.23 (0.22)	9.42** (1.86)	5.31** (3.53)	2.18 (1.27)	6.46** (3.91)	7.04** (7.07)	6.13** (2.93)	0.90 (0.62)
α_{nb}	0.02 (0.00)	0.99** (2.33)	4.36** (2.20)	0.24 (0.41)	0.44 (0.68)	0.49 (0.67)	-0.01 (0.00)	0.52 (0.64)	0.71 (1.20)
$R_t = \alpha_b d_b + \alpha_{nb} d_{nb} + \sum \beta_{k,b} F_{k,t} d_b + \sum \beta_{k,nb} F_{k,t} d_{nb} + \varepsilon_t$									
α_b	10.37** (5.12)	0.00 (0.00)	-0.11 (0.00)	6.70** (4.55)	1.25 (0.93)	6.36** (3.77)	5.84** (5.75)	4.27** (2.56)	-0.38 (0.26)
α_{nb}	0.34 (0.41)	0.84** (2.04)	3.57** (2.19)	-0.08 (0.14)	0.31 (0.65)	0.71 (0.97)	0.09 (0.26)	0.70 (1.14)	0.62 (1.09)
Panel B: Sample Period Extended through December 2014									
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH
	1990-2014	2000-2014	1984-2014	1984-2014	1990-2014	1990-2014	1984-2014	1990-2014	1990-2014
$R_t = \alpha_b d_b + \alpha_{nb} d_{nb} + \sum \beta_k F_{k,t} + \varepsilon_t$									
α_b	12.85** (6.07)	-0.28 (0.24)	9.76** (1.97)	5.15** (3.50)	2.21* (1.34)	7.03** (4.50)	7.08** (7.17)	6.17** (3.00)	1.01 (0.72)
α_{nb}	-0.40 (0.50)	0.98** (2.42)	4.46** (2.40)	0.24 (0.45)	0.52 (0.94)	0.50 (0.95)	0.03 (0.10)	0.41 (0.55)	0.71* (1.37)
$R_t = \alpha_b d_b + \alpha_{nb} d_{nb} + \sum \beta_{k,b} F_{k,t} d_b + \sum \beta_{k,nb} F_{k,t} d_{nb} + \varepsilon_t$									
α_{bb}	10.37** (5.15)	0.00 (0.00)	-0.11 (0.00)	6.70** (4.63)	1.25 (0.97)	6.36** (4.21)	5.84** (5.80)	4.27** (2.55)	-0.38 (0.28)
α_{nb}	-0.13 (0.17)	0.84** (2.13)	3.64** (2.37)	-0.04 (0.10)	0.34 (0.82)	0.66* (1.34)	0.11 (0.35)	0.63 (1.04)	0.58 (1.16)

Table 2.5: Out of Sample Performance

This table presents the in- and out-of-sample performance of the predictors. α_{in} is the in-sample performance and α_{out} is the out-of-sample performance. α_{in} and α_{out} are estimated using two different regression models: $R_t = \alpha_{in}d_{in} + \alpha_{out}d_{out} + \sum \beta_k F_{k,t} + \varepsilon_t$ and $R_t = \alpha_{in}d_{in} + \alpha_{out}d_{out} + \sum \beta_{k,in}F_{k,t}d_{in} + \sum \beta_{k,out}F_{k,t}d_{out} + \varepsilon_t$. Here, R_t represents excess long-short portfolio returns net of fees, $F_{k,t}$ are factors from the four-factor model, d_{in} is a dummy variable which equals to one in-sample and is zero otherwise, d_{out} is a dummy variable that equals $1-d_{in}$. β_k , $\beta_{k,in}$ and $\beta_{k,out}$ represent the unconditional, conditional in-sample, and conditional out-of-sample factor loadings. Panel A (post-sample performance) presents the results for the sample period from the original study that is extended to December 2014. Panel B (pre and post-sample performance) uses the sample period from January 1984 to December 2014, except for long-short portfolio formed on volatility which use a sample period of September 1999 to December 2014. α_{in} and α_{out} are reported in percentage per year and the corresponding t -statistics is in the parentheses. ** and * indicate significance at 5% and 10% level using a one tailed test.

Panel A: Sample Period Extended through December 2014 (Post-Sample)									
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH
	1990-2014	1990-2014	2000-2014	1990-2014	1983-2014	1984-2014	1984-2014	1990-2014	1990-2014
$R_t = \alpha_{in}d_{in} + \alpha_{out}d_{out} + \sum \beta_k F_{k,t} + \varepsilon_t$									
α_{in}	1.93** (2.28)	0.81** (2.09)	5.02** (2.76)	0.90* (1.58)	0.68 (1.18)	1.70** (2.63)	0.70** (2.07)	1.22* (1.63)	0.79* (1.49)
α_{out}	-2.81 (-1.47)	1.25 (0.67)	6.18 (0.93)	0.42 (0.33)	0.72 (0.57)	0.25 (0.32)	0.94 (0.52)	-1.65 (-0.46)	0.50 (0.41)
$R_t = \alpha_{in}d_{in} + \alpha_{out}d_{out} + \sum \beta_{k,in}F_{k,t}d_{in} + \sum \beta_{k,out}F_{k,t}d_{out} + \varepsilon_t$									
α_{in}	1.76** (2.12)	0.83** (2.13)	5.00** (2.75)	0.95** (1.66)	0.64 (1.14)	1.44** (2.39)	0.69** (2.05)	1.20* (1.61)	0.74* (1.41)
α_{out}	-3.40 (-1.62)	1.38 (0.66)	1.44 (0.20)	1.33 (0.93)	0.02 (0.00)	0.45 (0.59)	0.11 (0.00)	-2.51 (-0.62)	-0.05 (0.00)
Panel B: Sample Period from January 1984 to December 2014 (Pre and Post-Sample)									
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH
	1984-2014	1984-2014	1999-2014	1984-2014	1984-2014	1984-2014	1984-2014	1984-2014	1984-2014
$R_t = \alpha_{in}d_{in} + \alpha_{out}d_{out} + \sum \beta_k F_{k,t} + \varepsilon_t$									
α_{in}	1.83** (2.33)	0.62* (1.48)	4.90** (2.61)	0.79* (1.42)	0.68 (1.18)	1.71** (2.65)	0.70** (2.07)	1.22* (1.59)	0.83* (1.62)
α_{out}	-1.13 (-0.99)	0.03 (0.00)	-1.38 (-0.22)	-0.24 (-0.28)	0.72 (0.57)	0.31 (0.40)	0.94 (0.52)	-1.13 (-0.81)	0.33 (0.45)
$R_t = \alpha_{in}d_{in} + \alpha_{out}d_{out} + \sum \beta_{k,in}F_{k,t}d_{in} + \sum \beta_{k,out}F_{k,t}d_{out} + \varepsilon_t$									
α_{in}	1.76** (2.25)	0.83** (2.23)	5.00** (2.73)	0.95** (1.73)	0.64 (1.14)	1.44** (2.39)	0.69** (2.05)	1.20** (1.68)	0.74* (1.46)
α_{out}	-1.19 (-1.00)	0.24 (0.33)	3.36 (0.46)	-0.23 (-0.28)	0.02 (0.00)	0.46 (0.63)	0.11 (0.00)	0.07 (0.00)	0.16 (0.20)

Table 2.6: Performance Relative to Fama-French Five-Factor Model

This table presents the performance of the predictors with respect to the Fama and French (2015) five-factor model. The α estimates from Table 2.3 (i.e. the Carhart (1997) four-factor model) are also reported for comparison. Panel A presents the results for the sample period from the original study. Panel B extends this period forward to December 2014. α is reported in percentage per year and the corresponding t -statistics is in the parentheses. R^2 represents adjusted- R^2 from the regression. ** and * indicate significance at 5% and 10% level using a one tailed test.

Panel A: Sample Period From Original Study									
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH
	1990-2010	2000-2013	1984-2003	1984-2009	1990-2013	1990-2013	1984-2013	1990-2009	1990-2009
Carhart (1997) four-factor model									
α	1.76**	0.83**	5.00**	0.95*	0.64	1.44**	0.69**	1.20*	0.74*
$t(\alpha)$	(2.04)	(2.10)	(2.70)	(1.61)	(1.07)	(2.06)	(2.03)	(1.59)	(1.33)
R^2	37.56	77.38	80.63	44.23	54.63	68.18	5.19	75.55	29.47
Fama and French (2015) five-factor models									
α	1.03	-0.62	-0.03	0.51	-1.34	3.26**	0.57*	0.10	0.02
$t(\alpha)$	(1.21)	(-1.33)	(0.00)	(0.85)	(-2.34)	(3.66)	(1.62)	(0.14)	(0.00)
R^2	43.15	70.87	86.47	44.74	60.21	50.24	5.24	82.02	19.76
Panel B: Sample Period Extended through December 2014									
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH
	1990-2014	2000-2014	1984-2014	1984-2014	1990-2014	1990-2014	1984-2014	1990-2014	1990-2014
Carhart (1997) four-factor model									
α	1.18*	0.84**	5.10**	0.82*	0.68*	1.15**	0.71**	1.10*	0.75*
$t(\alpha)$	(1.50)	(2.18)	(2.92)	(1.57)	(1.30)	(2.24)	(2.13)	(1.51)	(1.53)
R^2	32.84	77.24	80.40	41.90	52.03	57.39	5.96	75.75	27.55
Fama and French (2015) five-factor models									
α	0.53	-0.56	0.33	0.44	-1.02	2.41**	0.58**	0.05	0.02
$t(\alpha)$	(0.68)	(-1.24)	(0.22)	(0.81)	(-2.01)	(3.98)	(1.69)	(0.10)	(0.00)
R^2	37.72	70.87	86.56	42.57	57.25	44.14	6.14	82.04	18.65

Table 2.7: Five-Factor Model - Factor Exposures

This table presents factor exposures of the long-short portfolios to the Fama and French (2015) five factors. Panel A presents the results for the sample period from the original study. Panel B extends this period forward to December 2014. β_{MKT} is the exposure to the market factor, β_{SMB} is the exposure to the size factor, β_{HML} is the exposure to the value factor, β_{RMW} is the exposure to the profitability factor, and β_{CMA} is the exposure to the investment factor. Corresponding to these β_k , the $t(\beta_k)$ i.e. the t-statistics are also reported.

Panel A: Sample Period From Original Study										
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH	
	1990-2010	2000-2013	1984-2003	1984-2009	1990-2013	1990-2013	1984-2013	1990-2009	1990-2009	
β_{MKT}	-0.05	-0.04	-0.31	-0.02	0.01	0.00	-0.01	0.04	0.00	
β_{SMB}	0.26	-0.16	-0.33	-0.12	-0.11	0.18	-0.02	0.64	0.06	
β_{HML}	-0.01	0.11	0.26	0.03	0.12	-0.20	0.00	0.02	0.08	
β_{RMW}	0.12	0.10	0.51	0.08	0.21	-0.05	0.02	0.12	0.00	
β_{CMA}	0.15	0.03	0.33	0.05	0.02	0.06	0.02	0.01	0.05	
$t(\beta_{MKT})$	-2.94	-3.74	-8.95	-1.22	0.97	0.10	-0.94	2.48	0.10	
$t(\beta_{SMB})$	11.43	-12.31	-6.99	-7.62	-6.89	7.77	-1.87	33.49	3.65	
$t(\beta_{HML})$	-0.33	6.54	4.62	1.55	5.48	-5.44	0.14	0.95	3.79	
$t(\beta_{RMW})$	3.91	6.06	8.53	3.74	10.10	-1.57	1.60	4.84	0.14	
$t(\beta_{CMA})$	3.35	1.06	4.28	1.70	0.53	1.36	0.85	0.40	1.71	

(continued)

Table 2.7—Continued

Panel B: Sample Period Extended through December 2014										
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH	
	1990-2014	2000-2014	1984-2014	1984-2014	1990-2014	1990-2014	1984-2014	1990-2014	1990-2014	
β_{MKT}	-0.07	-0.04	-0.31	-0.02	0.00	0.00	-0.01	0.03	0.00	
β_{SMB}	0.26	-0.16	-0.33	-0.12	-0.12	0.15	-0.02	0.64	0.05	
β_{HML}	-0.01	0.11	0.25	0.04	0.11	-0.16	0.00	0.03	0.08	
β_{RMW}	0.11	0.10	0.51	0.08	0.20	-0.07	0.02	0.12	0.01	
β_{CMA}	0.12	0.02	0.33	0.05	0.01	0.03	0.02	0.01	0.06	
$t(\beta_{MKT})$	-3.82	-3.92	-9.37	-1.55	0.26	0.10	-0.92	2.45	-0.33	
$t(\beta_{SMB})$	11.46	-12.51	-7.42	-7.49	-7.65	8.65	-2.23	34.18	3.24	
$t(\beta_{HML})$	-0.37	6.76	4.74	1.97	5.56	-6.98	0.10	1.05	3.96	
$t(\beta_{RMW})$	3.83	6.18	9.04	4.03	10.38	-2.91	1.69	4.89	0.30	
$t(\beta_{CMA})$	2.81	1.01	4.46	1.73	0.20	1.00	0.97	0.36	2.01	

Table 2.8: Alternative Benchmark Models

This table presents the performance of the predictors when benchmarked with a six-factor model that adds the momentum factor to the Fama and French (2015) model (6-Factors) from Table 2.6, the q-factor model of Hou et al. (2015) (Q-Factors), and the Stambaugh and Yuan (2017) mispricing factors model (M-Factors). α estimates from Table 2.3 (i.e. the Carhart (1997) four-factor model) are also reported for comparison. Panel A presents the results for the sample period from the original study. Panel B extends this period forward to December 2014. α is reported in percentage per year. R^2 represents adjusted- R^2 from the regression. ** and * indicate significance at 5% and 10% level using a one tailed test.

Panel A: Sample Period From Original Study											
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH		
	1990-2010	2000-2013	1984-2003	1984-2009	1990-2013	1990-2013	1984-2013	1990-2009	1990-2009		
α	4-Factors	1.76**	0.83**	5.00**	0.95*	0.64	1.44**	0.69**	1.20*	0.74*	
	6-Factors	0.81	-0.08	-0.24	0.40	-0.79	1.94**	0.55*	0.33	0.43	
	Q-Factors	1.32*	-0.06	2.68	0.20	-1.14	2.79**	0.73**	0.46	0.60	
	M-Factors	0.26	0.62	3.12	0.46	0.00	1.54**	0.38*	0.60	0.61	
R^2	4-Factors	37.56	77.38	80.63	44.23	54.63	68.18	5.19	75.55	29.47	
	6-Factors	44.12	82.34	86.98	45.19	67.00	70.27	5.06	82.60	30.84	
	Q-Factors	38.26	61.28	75.67	42.60	44.70	44.82	4.56	77.51	14.56	
	M-Factors	41.25	66.22	76.27	45.74	39.75	47.31	6.75	75.13	22.93	

(continued)

Table 2.8—Continued

Panel B: Sample Period Extended through December 2014										
	R^2	DUR	VOL	CH	VG	RG	AW	AS	LH	
	1990-2014	2000-2014	1984-2014	1984-2014	1990-2014	1990-2014	1984-2014	1990-2014		1990-2014
α	1.18*	0.84**	5.10**	0.82*	0.68*	1.15**	0.71**	1.10*		0.75*
4-Factors										
6-Factors	0.34	-0.02	0.19	0.32	-0.57	1.75**	0.56*	0.28		0.40
Q-Factors	0.90	-0.06	2.67*	0.17	-0.90	2.05**	0.74**	0.39		0.56
M-Factors	-0.10	0.62	3.12	0.37	0.05	0.97**	0.42*	0.48		0.54
R^2										
4-Factors	32.84	77.24	80.40	41.90	52.03	57.39	5.96	75.75		27.55
6-Factors	38.44	82.13	87.04	43.12	63.77	60.24	5.94	82.62		29.41
Q-Factors	32.26	61.33	75.88	40.11	43.13	36.09	5.33	77.57		13.70
M-Factors	35.23	66.24	76.40	43.21	38.73	39.21	7.42	74.88		19.89

Figure 2.1: Net Excess Returns for Long-Short Portfolios

This figure shows net excess returns for the long-short equal-weighted portfolios formed using quintile sorts of the lagged predictor variables. I extend the sample period from the original study to December 2014. The gray shaded area refers to the three year period surrounding the end of Dot-Com bubble, i.e., January 1999 to December 2001. Returns are reported in percentage per month.

Figure 2.2: Factor Returns

This figure shows excess returns for the Fama and French (2015) five factors and the momentum factor of Carhart (1997) from January 1984 to December 2014. MKT refers to the return on market factor, SMB refers to the return on size factor, HML to the return on the value factor, RMW to the return on the profitability factor, and CMA to the return on the investment factor. The gray shaded area refers to the three year period surrounding the end of Dot-Com bubble, i.e., January 1999 to December 2001. Returns are reported in percentage per month.

Figure 2.3: Factor Returns

This figure shows excess returns for the Fama and French (2015) five factors and the momentum factor of Carhart (1997) from January 1984 to December 2014. MKT refers to the return on market factor, SMB refers to the return on size factor, HML to the return on the value factor, RMW to the return on the profitability factor, and CMA to the return on the investment factor. The gray shaded area refers to the three year period surrounding the end of Dot-Com bubble, i.e., January 1999 to December 2001. Returns are reported in percentage per month.

Appendix A

Model and Data Appendix for Chapter 1

A.1 Derivation of expected returns

This appendix outlines the derivation Equation (1.7) using Lagrangian multipliers approach. This derivation follows the standard production based q-theory models similar to Liu et al. (2009).

Substituting Equation (1.1), Equation (1.2), Equation (1.3), and $H_t = \lambda_t J_t$, in the firm's value maximization problem defined by the Equation (1.5) yields

$$V_t = \max_{J_{t+\tau}, N_{t+1+\tau}} E_t \left[\sum_{\tau=0}^{\infty} M_{t+\tau} (e^{x_{t+\tau}} N_{t+\tau} - W_{t+\tau} N_{t+\tau} - F - \frac{a}{2} \left(\frac{H_{t+\tau}}{N_{t+\tau}} \right)^2 N_{t+\tau}) \right], \quad (\text{A.1})$$

where V_t is the cum-dividend value of the firm at time t . Rewriting this equation as a Lagrangian maximization constrained by the law of motion of the labor force in Equation (1.4) leads to

$$V_t = E_t \left[\sum_{\tau=0}^{\infty} M_{t+\tau} \left(e^{x_{t+\tau}} N_{t+\tau} - W_{t+\tau} N_{t+\tau} - F - \frac{a}{2} \left(\frac{H_{t+\tau}}{N_{t+\tau}} \right)^2 N_{t+\tau} - Q_{t+\tau} (N_{t+\tau+1} - (1-s)N_t - \lambda_{t+\tau} J_{t+\tau}) \right) \right], \quad (\text{A.2})$$

where $Q_{t+\tau}$ represents the Lagrangian multiplier associated with the constraint on

the size of the labor force. Expanding this infinite summation yields

$$\begin{aligned}
V_t = & e^{x_t} N_t - W_t N_t - F - \frac{a}{2} \left(\frac{\lambda_t J_t}{N_t} \right)^2 N_t - Q_t \left(N_{t+1} - (1-s)N_t - \lambda_t J_t \right) \\
& + E_t \left[M_{t+1} \left(e^{x_{t+1}} N_{t+1} - W_{t+1} N_{t+1} - F - \frac{a}{2} \left(\frac{\lambda_t J_t}{N_{t+1}} \right)^2 N_{t+1} \right. \right. \\
& \left. \left. - Q_{t+1} \left(N_{t+2} - (1-s)N_{t+1} - \lambda_{t+1} J_{t+1} \right) \right) \right] + \dots
\end{aligned} \tag{A.3}$$

The first-order conditions for the value maximization problem defined in Equation (A.3) are given by Equation (1.6). The ex-dividend value of the firm is $P_t = V_t - D_t$. Substituting the current period value of dividends in Equation (1.2) in this definition gives

$$V_t = P_t + e^{x_t} N_t - W_t N_t - F - \frac{a}{2} \left(\frac{\lambda_t J_t}{N_t} \right)^2 N_t. \tag{A.4}$$

Combining Equation (A.3) and Equation (A.4) and recursively substituting the first-order conditions given by Equation (1.6) yields the following equation for the ex-dividend value P_t :

$$P_t = Q_t N_{t+1}. \tag{A.5}$$

To explicitly link the maximization problem to expected returns for the representative firm, note that by the definition of the ex-dividend value, the return for the representative firm is

$$R_{t+1} = \frac{V_{t+1}}{P_t} = \frac{P_{t+1} + D_{t+1}}{P_t}. \tag{A.6}$$

Substituting Equation (1.2), Equation (A.5), and Equation (1.4) and taking expectations on both sides yields (with some additional algebra)

$$E_t(R_{t+1}) = \frac{(e^{x_t} - W_{t+1} - F - \frac{a}{2}(1-s)^2) + \frac{a}{2} E_t \left(\frac{N_{t+2}}{N_{t+1}} \right)^2}{a \lambda_t \left(\frac{J_t}{N_t} \right)}, \tag{A.7}$$

which can be rewritten as Equation (1.7).

A.2 Construction of the benchmark predictors

This section provides details of the construction of the 16 benchmark predictor variables.

1. Dividend to price ratio (DP): The difference between the log dividends and the log prices of the S&P 500 index.
2. Dividend yield (DY): The difference between the log of dividends and the log lagged prices of the S&P 500 index.
3. Earnings to price ratio (EP): The difference between log earnings and log prices of the S&P 500 index.
4. Dividend payout ratio (DE): The difference between log dividends and log earnings of the S&P 500 index.
5. Stock return volatility ($RVOL$): A 12-month moving standard deviation of the S&P 500 excess returns (Mele (2007) and Rapach et al. (2016)).
6. Book to market ratio (BM): The ratio of the book value to market value for the Dow Jones Industrial Average.
7. Net equity expansion ($NTIS$): The ratio of the 12-month cumulative sum of net issues by NYSE listed stocks to their aggregate market capitalization.
8. Treasury bill rates (TBL): The yields on the 3-Month U.S. treasury bill.

9. Long-term government bond yield (*LTY*): The long-term government bond yield obtained from Ibbotsons Stocks, Bonds, Bills and Inflation Yearbook.
10. Long-term government bond returns (*LTR*): The return on long-term government bonds obtained from Ibbotsons Stocks, Bonds, Bills and Inflation Yearbook
11. Term spread (*TMS*): The difference between long-term government bond yield and the T-bill.
12. Default-yield spread (*DFY*): The difference between Moodys BAA- and AAA-rated corporate bond yields.
13. Default-return spread (*DFR*): The difference between long-term corporate bond returns and long-term government bond returns.
14. Inflation (*INFL*): Calculated using the Consumer Price Index for all urban consumers. Following Rapach et al. (2016), for the information on to be publicly available, I lag this variable by one month.
15. Output gap (*OG*): The deviations of the log of industrial production from a trend that incorporates both linear and quadratic components (Cooper and Priestley (2008)). For the information on to be publicly released, I lag OG by one month.¹
16. Nearness to historical Dow Jones Industrial Average index high (*PMAX*): The

¹Data on seasonally adjusted industrial production are obtained from the Federal Reserve website: <https://fred.stlouisfed.org/series/INDPRO>. In addition, the detrending is performed on a recursive basis such that data from January 1951 to month t is used to construct OG at the end of month t .

ratio of the current Dow Jones Industrial Average index and its historical high
(Li and Yu (2012)).

Appendix B

Data Appendix for Chapter 2

B.1 Construction of the activeness proxies

This appendix provides the details regarding the construction the managerial activeness variables.

1. R^2 The R^2 is estimated from a regression of net of fee fund excess fund returns on the four-factor model of Carhart (1997). Following Amihud and Goyenko (2013), the estimation is conducted using the prior 24 months of data. In addition, I also require the mutual funds to have a valid return for each of these 24 months. Similar to Amihud and Goyenko (2013), I trim the estimated R^2 sample to exclude the fund-observations that fall in the 0.5% tails on either side of the distribution.
2. Duration This measure was first introduced by Cremers and Pareek (2014) and Cremers and Pareek (2016). They calculate the fund *duration* as the average of duration of all the stocks in the portfolio, weighted by the market value of the stock holdings. The duration of stock j and fund i at time $T - 1$ is calculated as:

$$DUR_{i,j,T-1} = \sum_{t=T-W}^{T-1} \left(\frac{(T-t-1)a_{i,j,t}}{H_{i,j} + B_{i,j}} \right) \left(\frac{(W-1)H_{i,j}}{H_{i,j} + B_{i,j}} \right) \quad (\text{B.1})$$

Here $B_{i,j}$ is the percentage of shares of stock i bought by fund j between the time period $T - W$ and $T - 1$, $H_{i,j}$ is the percentage of total shares outstanding of stock i held by fund j at time $T - W$, and $a_{i,j,t}$ equals the percentage of total outstanding shares of stock i that are traded by fund j between the time period from $t1$ and t , such that, $a_{i,j,t} > 0$ if the trade is a buy and $a_{i,j,t} < 0$ if the trade is a sale. I obtain annual data on this measure from Martijn Cremers's website (<http://activeshare.nd.edu/data/>). I convert the annual data to quarterly by assuming the fund *duration* remains the same for next twelve months unless reported otherwise.

3. Volatility I calculate *volatility* as the standard deviation of daily fund returns over the prior twelve months. Following Jordan and Riley (2015), I require the funds to have a valid return observation for each trading day in the twelve months.
4. Leading herd and contemporaneous herd Koch (2016) calculates *leading herd* as the cosine of the angle between the vector representing one-quarter ahead changes in fund weights and the vector representing change in herd weights. Similarly, *contemporaneous herd* is calculated as the cosine of the angle between vector representing changes in fund weights and change in herd weights. For fund i , at time t , these measures are calculated as follows:

$$\text{Contemporaneous Herd}_{i,t} = \frac{\Delta w_{i,t} \bullet \Delta h_{i,t}}{\|\Delta w_{i,t}\| \|\Delta h_{i,t}\|} \quad (\text{B.2})$$

$$\text{Leading Herd}_{i,t} = \frac{\Delta w_{i,t+1} \bullet \Delta h_{i,t}}{\|\Delta w_{i,t+1}\| \|\Delta h_{i,t}\|} \quad (\text{B.3})$$

Here, changes in $\Delta w_{i,t}$ and $\Delta w_{i,t+1}$ represent the in fund weights from time $t - 1$

to t and t to $t+1$ respectively and $\Delta h_{i,t}$ represents the changes in herd weights at from time $t-1$ to t . The herd weights are calculated as an equal-weighted average of the weights of all other funds in the cross-section i.e. $h_{i,t} = \frac{1}{F-1} \sum_{-f} w_{i,t}$ where, F is the number of funds in the cross-section. Similar to Koch (2016), I orthonormalize the herd vectors $\Delta h_{i,t+1}$, $\Delta h_{i,t}$, $\Delta h_{i,t-1}$ using the Gram-Schmidt process for a given time t for each fund in the cross-section. Furthermore, I also require the funds to report holdings in March, June, September and December only and drop funds that do not report holdings in consecutive quarters. I also drop any fund-month observations with less than 10 stocks at the beginning or end of the quarter. In addition, I lag the *leadingherd* measure by one quarter since it uses fund holdings weights from $t+1$.

5. Volatility gap *Volatility gap* is measured as the difference between the current holdings volatility ($\sigma_{i,t}^H$) which is defined as the volatility of fund's most recently disclosed positions and the volatility of fund's realized returns ($\sigma_{i,t}^R$). Similar to Huang et al. (2011), both the $\sigma_{i,t}^H$ and $\sigma_{i,t}^R$ are estimated as the standard deviation of monthly returns over past 36 months. I exclude funds that did not disclose their holdings and the funds that do not have valid returns over any of the 36 months in the estimation period. In the estimation of $\sigma_{i,t}^H$, I use the total return of the Barclays Capital Aggregate Bond Index for bonds and preferred stocks, and the Treasury bill rate for cash holdings and other assets.

6. Industry Concentration Kacperczyk et al. (2005) compute the *industry concen-*

tration for fund i at the end of month t as

$$IC_{i,t} = \sum_{j=1}^m (w_{i,t}^j - \bar{w}_t^j)^2 \quad (\text{B.4})$$

where $w_{i,t}^j$ is the weight of industry j in the portfolio of fund i at time t and $\bar{w}_t^j$ is the weight in industry j in the market portfolio. I begin with the Fama and French (1997) 48-industry classification which is sourced from Kenneth French's website and classify these into 10 groups mentioned in the Kacperczyk et al. (2005) (Table AI). In addition, I also require funds to hold at least 10 stocks.

7. Return gap Following Kacperczyk et al. (2006), I define *return gap* as the funds' monthly gross return minus the return of its most recently reported holdings, averaged over the prior 12 months. I require funds to hold at least 10 stocks.
8. Contrarian I begin the construction of *contrarian* by first calculating a herding measure $HM_{j,t}$ for stock j at in month t similar following Lakonishok et al. (1992) as,

$$HM_{j,t} = |p_{j,t} - \bar{p}_{j,t}| - E(|p_{j,t} - \bar{p}_t|) \quad (\text{B.5})$$

where $p_{j,t}$ is the proportion of mutual funds buying stock j out of all funds trading that stock during quarter t , $\bar{p}_t$ is the cross-sectional mean of $p_{j,t}$ over all stocks traded by all funds during quarter t , and $E(|p_{j,t} - \bar{p}_t|)$ is the expected value of $|p_{j,t} - \bar{p}_{j,t}|$ under the null hypothesis of no herding as defined in Lakonishok et al. (1992).¹ In order to calculate a meaningful estimate of the herding

¹This is commonly referred to as the "adjustment factor" and represents the probability of a fund being net buyer of any stock under the null hypothesis of no herding. Suppose, N is the number of funds trading stock j in quarter t and B (S) is the number of funds buying (selling) that particular stock during the same quarter. Assuming that B follows a binomial distribution, $p_{k,t}$ is the probability that k out of N funds are net buyers in this stock. In this case, the adjustment factor is calculated as $\sum_{k=0}^B \binom{N}{k} p_{k,t}^k (1 - p_{k,t})^{(N-k)} \left| \frac{k}{N} - p_{k,t} \right|$.

measure, I require a stock to be traded by at least 10 funds during a given quarter and be listed for least 12 months. Using this stock-level herding measure, I calculate the conditional fund-level buying and selling herding measures separately as

$$BHM_{j,t} = HM_{j,t} \mid p_{j,t} > \bar{p}_t \quad (\text{B.6})$$

$$SHM_{j,t} = HM_{j,t} \mid p_{j,t} < \bar{p}_t \quad (\text{B.7})$$

These measures are then combined to form a single variable $HERD_{j,t}$ which is a herding index ranked from -1 to -5 if $SHM_{j,t} > 0$ and 1 to 5 if $BHM_{j,t} > 0$.

Finally, I use this herding index to measure fund-level *contrarian* $CON_{j,t}$ for fund i in quarter t as,

$$CON_{j,t} = - \sum_{j=1}^m w_{i,j,t} HERD_{j,t} \quad (\text{B.8})$$

where, $w_{i,j,t}$ is the portfolio weight adjusted for flow-driven trading as defined in Wei et al. (2014) (eqn. 5).

9. Active Share Following Cremers and Petajisto (2009) and Cremers and Pareek (2016), *active share* for mutual fund i , holding m stocks in the portfolio, at the end of month t is defined as:

$$AS_{i,t} = \frac{1}{2} \sum_{j=1}^m |w_{i,j,t} - w_{b,j,t}| \quad (\text{B.9})$$

where $w_{i,j,t}$ is the weight of fund i in stock j at time t , $w_{b,j,t}$ is the weight of stock j in the benchmark index portfolio at time t , and summation is defined across all m stocks in the portfolio of the fund. Cremers and Petajisto (2009) consider a total of 19 candidate benchmark indexes and assign benchmark to each

fund such the active share is minimized. I obtain the quarterly *activeshare* data from Antti Petajisto's website(<http://www.petajisto.net/data.html>) from January 1984 to December 2009 and further supplement this with and annual data from Martijn Cremers's website(<http://activeshare.nd.edu/data/>) from January 2010 to December 2014.

10. Active Weight *Activeweight* is defined as the absolute value of difference between the fund portfolio weights and the market portfolio weights summed across all the holdings of a fund and divided by two. Following Doshi et al. (2015), for fund i at time t , *activeweight* is defined as:

$$AW_{i,t} = \frac{1}{2} \sum_{j=1}^m |w_{i,j,t} - w_{m,j,t}| \quad (\text{B.10})$$

where $w_{i,j,t}$ is the weight of fund i in stock j at time t , $w_{m,j,t}$ is the weight of stock j in the market portfolio, and summation is defined across all stocks m in the portfolio of the fund. I require funds to hold at least 10 stocks. I leverage Mike Simutin's code (http://www-2.rotman.utoronto.ca/simutin/dfp_code.asp) to calculate this measure.

Bibliography

- Amihud, Yakov, and Ruslan Goyenko, 2013, Mutual fund's r^2 as predictor of performance, *The Review of Financial Studies* 26, 667–694.
- Baker, Malcolm, and Jeffrey Wurgler, 2000, The equity share in new issues and aggregate stock returns, *The Journal of Finance* 55, 2219–2257.
- Barnichon, Regis, 2010, Building a composite help-wanted index, *Economics Letters* 109, 175–178.
- Belo, Frederico, Andres Donangelo, Xiaoji Lin, and Ding Luo, 2018, Labor hiring, aggregate dividends, and return predictability in the time series, Unpublished Working Paper.
- Belo, Frederico, Xiaoji Lin, and Santiago Bazdresch, 2014, Labor hiring, investment, and stock return predictability in the cross section, *Journal of Political Economy* 122, 129–177.
- Berk, Jonathan B, and Richard C Green, 2004, Mutual fund flows and performance in rational markets, *Journal of political economy* 112, 1269–1295.
- Berk, Jonathan B, and Jules H Van Binsbergen, 2015, Measuring skill in the mutual fund industry, *Journal of Financial Economics* 118, 1–20.
- Brunnermeier, Markus K, and Stefan Nagel, 2004, Hedge funds and the technology bubble, *The Journal of Finance* 59, 2013–2040.
- Cajner, Tomaz, and David Ratner, 2016, A cautionary note on the help wanted online data, *FEDS Notes, Board of Governors of the Federal Reserve System* .
- Campbell, John Y, 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.
- Campbell, John Y, and Robert J Shiller, 1988, Stock prices, earnings, and expected dividends, *The Journal of Finance* 43, 661–676.
- Campbell, John Y, and Samuel B Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *The Review of Financial Studies* 21, 1509–1531.
- Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of finance* 52, 57–82.

- Cederburg, Scott, NE Savin, Ashish Tiwari, et al., 2018, Conditional benchmarks and predictors of mutual fund performance, *Critical Finance Review* 7, 331–372.
- Chen, Long, and Lu Zhang, 2011, Do time-varying risk premiums explain labor market performance?, *Journal of Financial Economics* 99, 385–399.
- Christoffersen, Susan EK, and Sergei Sarkissian, 2009, City size and fund performance, *Journal of Financial Economics* 92, 252–275.
- Clark, Todd E, and Kenneth D West, 2006, Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis, *Journal of Econometrics* 135, 155–186.
- Clark, Todd E, and Kenneth D West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.
- Cochrane, John H, 2009, *Asset Pricing:(Revised Edition)* (Princeton university press).
- Cohen, Randolph B, Joshua D Coval, and L’uboš Pástor, 2005, Judging fund managers by the company they keep, *The Journal of Finance* 60, 1057–1096.
- Conrad, Jennifer, Michael Cooper, and Gautam Kaul, 2003, Value versus glamour, *The Journal of Finance* 58, 1969–1995.
- Cooper, Ilan, and Richard Priestley, 2008, Time-varying risk premiums and the output gap, *The Review of Financial Studies* 22, 2801–2833.
- Cremers, KJ Martijn, and Antti Petajisto, 2009, How active is your fund manager? a new measure that predicts performance, *The Review of Financial Studies* 22, 3329–3365.
- Cremers, Martijn, and Ankur Pareek, 2014, Short-term trading and stock return anomalies: Momentum, reversal, and share issuance, *Review of Finance* 19, 1649–1701.
- Cremers, Martijn, and Ankur Pareek, 2016, Patient capital outperformance: The investment skill of high active share managers who trade infrequently, *Journal of Financial Economics* 122, 288–306.
- Danthine, Jean-Pierre, and John B Donaldson, 2002, Labour relations and asset returns, *The Review of Economic Studies* 69, 41–64.
- Dass, Nishant, Massimo Massa, and Rajdeep Patgiri, 2008, Mutual funds and bubbles: The surprising role of contractual incentives, *The Review of Financial Studies* 21, 51–99.
- DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2007, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *The Review of Financial Studies* 22, 1915–1953.

- Den Haan, Wouter J, Garey Ramey, and Joel Watson, 2000, Job destruction and propagation of shocks, *American Economic Review* 90, 482–498.
- Diamond, Peter A, 1982, Wage determination and efficiency in search equilibrium, *The Review of Economic Studies* 49, 217–227.
- Doshi, Hitesh, Redouane Elkamhi, and Mikhail Simutin, 2015, Managerial activeness and mutual fund performance, *The Review of Asset Pricing Studies* 5, 156–184.
- Eiling, Esther, Raymond Kan, and Ali Sharifkhani, 2016, Sectoral labor reallocation and return predictability, Unpublished Working Paper.
- Elliott, Graham, and Ulrich K Müller, 2006, Efficient tests for general persistent time variation in regression coefficients, *The Review of Economic Studies* 73, 907–940.
- Elton, Edwin J, Martin J Gruber, and Christopher R Blake, 1996, Survivor bias and mutual fund performance, *The Review of Financial Studies* 9, 1097–1120.
- Elton, Edwin J, Martin J Gruber, and Christopher R Blake, 2001, A first look at the accuracy of the crsp mutual fund database and a comparison of the crsp and morningstar mutual fund databases, *The Journal of Finance* 56, 2415–2430.
- Evans, Richard B, 2010, Mutual fund incubation, *The Journal of Finance* 65, 1581–1611.
- Fama, Eugene F, 1981, Stock returns, real activity, inflation, and money, *The American Economic Review* 71, 545–565.
- Fama, Eugene F, and Kenneth R French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.
- Fama, Eugene F, and Kenneth R French, 1989, Business conditions and 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, 1997, Industry costs of equity, *Journal of financial economics* 43, 153–193.
- Fama, Eugene F, and Kenneth R French, 2010, Luck versus skill in the cross-section of mutual fund returns, *The journal of finance* 65, 1915–1947.
- 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, 2016, Dissecting anomalies with a five-factor model, *The Review of Financial Studies* 29, 69.
- Fama, Eugene F, and Michael R Gibbons, 1982, Inflation, real returns and capital investment, *Journal of Monetary Economics* 9, 297–323.

- Fama, Eugene F, and G William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.
- Ferson, Wayne, and Yong Chen, 2015, How many good and bad fund managers are there, really?, Unpublished Working Paper.
- Ferson, Wayne, and Haitao Mo, 2016, Performance measurement with selectivity, market and volatility timing, *Journal of Financial Economics* 121, 93–110.
- Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455–1508.
- Griffin, John M, Jeffrey H Harris, Tao Shu, and Selim Topaloglu, 2011, Who drove and burst the tech bubble?, *The Journal of Finance* 66, 1251–1290.
- Hall, Robert E, 2017, High discounts and high unemployment, *American Economic Review* 107, 305–30.
- Hansen, Peter Reinhard, and Allan Timmermann, 2015, Equivalence between out-of-sample forecast comparisons and wald statistics, *Econometrica* 83, 2485–2505.
- Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2016, ...and the cross-section of expected returns, *The Review of Financial Studies* 29, 5–68.
- Harvey, David S, Stephen J Leybourne, and Paul Newbold, 1998, Tests for forecast encompassing, *Journal of Business & Economic Statistics* 16, 254–259.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2017, Replicating anomalies, Forthcoming in *Review of Financial Studies*.
- Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2015, Investor sentiment aligned: A powerful predictor of stock returns, *The Review of Financial Studies* 28, 791–837.
- Huang, Jennifer, Clemens Sialm, and Hanjiang Zhang, 2011, Risk shifting and mutual fund performance, *The Review of Financial Studies* 24, 2575–2616.
- Inoue, Atsushi, and Lutz Kilian, 2005, In-sample or out-of-sample tests of predictability: Which one should we use?, *Econometric Reviews* 23, 371–402.
- Jensen, Michael C, 1968, The performance of mutual funds in the period 1945–1964, *The Journal of finance* 23, 389–416.
- Jobson, J Dave, and Bob M Korkie, 1981, Performance hypothesis testing with the sharpe and treynor measures, *The Journal of Finance* 36, 889–908.
- Jones, Christopher S, and Haitao Mo, 2017, Out-of-sample performance of mutual fund predictors, Unpublished Working Paper.

- Jordan, Bradford D, and Timothy B Riley, 2015, Volatility and mutual fund manager skill, *Journal of Financial Economics* 118, 289–298.
- Jordan, Bradford D, and Timothy B Riley, 2016, Skill and persistence in mutual fund returns: Evidence from a six-factor model, Unpublished Working Paper.
- Kacperczyk, Marcin, and Amit Seru, 2007, Fund manager use of public information: New evidence on managerial skills, *The Journal of Finance* 62, 485–528.
- Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2005, On the industry concentration of actively managed equity mutual funds, *The Journal of Finance* 60, 1983–2011.
- Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2006, Unobserved actions of mutual funds, *The Review of Financial Studies* 21, 2379–2416.
- Koch, Andrew, 2016, Herd behavior and mutual fund performance, *Management Science* .
- Kostakis, Alexandros, Tassos Magdalinos, and Michalis P Stamatogiannis, 2014, Robust econometric inference for stock return predictability, *The Review of Financial Studies* 28, 1506–1553.
- Kuehn, Lars-Alexander, Mikhail Simutin, and Jessie Jiaxu Wang, 2017, A labor capital asset pricing model, *The Journal of Finance* 72, 2131–2178.
- Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, 1992, The impact of institutional trading on stock prices, *Journal of financial economics* 32, 23–43.
- Lehmann, Bruce N, and David M Modest, 1987, Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons, *The journal of finance* 42, 233–265.
- Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *The Journal of Finance* 56, 815–849.
- Li, Jun, and Jianfeng Yu, 2012, Investor attention, psychological anchors, and stock return predictability, *Journal of Financial Economics* 104, 401–419.
- Liu, Laura Xiaolei, Toni M Whited, and Lu Zhang, 2009, Investment-based expected stock returns, *Journal of Political Economy* 117, 1105–1139.
- Malkiel, Burton G, 1995, Returns from investing in equity mutual funds 1971 to 1991, *The Journal of finance* 50, 549–572.
- Massa, Massimo, and Rajdeep Patgiri, 2008, Incentives and mutual fund performance: higher performance or just higher risk taking?, *The Review of Financial Studies* 22, 1777–1815.
- Massa, Massimo, and Vijay Yadav, 2015, Investor sentiment and mutual fund strategies, *Journal of Financial and Quantitative Analysis* 50, 699–727.

- McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The 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.
- Merz, Monika, and Eran Yashiv, 2007, Labor and the market value of the firm, *American Economic Review* 97, 1419–1431.
- Millard, Stephen P, Alexandra Varadi, and Eran Yashiv, 2017, The interaction of financial frictions and labor market frictions in a dsge model, Unpublished Working Paper.
- Mortensen, Dale T, and Christopher A Pissarides, 1994, Job creation and job destruction in the theory of unemployment, *The Review of Economic Studies* 61, 397–415.
- Nelson, Charles R, and Myung J Kim, 1993, Predictable stock returns: The role of small sample bias, *The Journal of Finance* 48, 641–661.
- Newey, Whitney K, and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
- Pástor, L’uboš, Robert F Stambaugh, and Lucian A Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23–45.
- Paye, Bradley S, and Allan Timmermann, 2006, Instability of return prediction models, *Journal of Empirical Finance* 13, 274–315.
- Petrosky-Nadeau, Nicolas, Lu Zhang, and Lars-Alexander Kuehn, 2018, Endogenous disasters, Forthcoming in *American Economic Review*.
- Phelps, Edmund S, 1994, *Structural slumps: The modern equilibrium theory of unemployment, interest, and assets* (Harvard University Press).
- Rapach, David, and Guofu Zhou, 2013, Forecasting stock returns, in *Handbook of Economic Forecasting*, volume 2, 328–383 (Elsevier).
- Rapach, David E, Matthew C Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.
- Reuter, Jonathan, and Eric Zitzewitz, 2013, How much does size erode mutual fund performance? a regression discontinuity approach, Unpublished Working Paper.
- Rogerson, Richard, and Robert Shimer, 2011, Search in macroeconomic models of the labor market, in *Handbook of Labor Economics*, volume 4, 619–700 (Elsevier).
- Samuelson, Paul A, 1989, The judgment of economic science on rational portfolio management: Indexing, timing, and long-horizon effects, *The Journal of Portfolio Management* 16, 4–12.

- Schwert, G William, 2003, Anomalies and market efficiency, *Handbook of the Economics of Finance* 1, 939–974.
- Shapiro, Matthew D, 1986, The dynamic demand for capital and labor, *The Quarterly Journal of Economics* 101, 513–542.
- Sharpe, William F, 1991, The arithmetic of active management, *Financial Analysts Journal* 47, 7–9.
- Sheng, Jinfei, Mikhail Simutin, and Terry Zhang, 2017, Cheaper is not better: On the superior performance of high-fee mutual funds, Unpublished Working Paper.
- Shiller, Robert J, 2015, *Irrational Exuberance*, student edition, 3 edition (Princeton University Press).
- Sirri, Erik R, and Peter Tufano, 1998, Costly search and mutual fund flows, *The journal of finance* 53, 1589–1622.
- Stambaugh, Robert F, 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.
- Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.
- Stambaugh, Robert F, and Yu Yuan, 2017, Mispricing factors, *The Review of Financial Studies* 30, 1270–1315.
- Uhlig, Harald, 2007, Explaining asset prices with external habits and wage rigidities in a dsge model, *American Economic Review* 97, 239–243.
- Wei, Kelsey D, Russ Wermers, and Tong Yao, 2014, Uncommon value: The characteristics and investment performance of contrarian funds, *Management Science* 61, 2394–2414.
- Yashiv, Eran, 2016, Capital values and job values, *Review of Economic Dynamics* 19, 190–209.

VITA

Pratik Kothari was born January 30th, 1990. He earned a Bachelor of Technology in Biotechnology in 2011 from Indian Institute of Technology Roorkee. Pratik then attended the University of Texas at Austin, where he received a Master of Science in Finance. Pratik worked as a Valuation Associate in Houston, TX before joining the doctoral program in the Department of Finance in 2014. He graduated with his Ph.D. in May 2019 and will join Cornerstone Research, Inc. as an Associate in the fall of 2019.