ESSAYS IN EMPIRICAL ASSET PRICING

by

Ziye Nie

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Business Administration

Charlotte

2019

Approved by:

____
Dr. Christopher M. Kirby

____
Dr. I-Husan Ethan Chiang

____
Dr. Yufeng Han

____
Dr. Hwan C. Lin

ABSTRACT

ZIYE NIE. Essays in Empirical Asset Pricing. (Under the direction of DR. CHRISTOPHER M. KIRBY)

The dissertation is composed of three essays that address the cross-sectional relation between firm characteristics and expected stock returns. Chapter 1, "Regime-switching and the Cross-Section of Expected Stock Returns", incorporates regime switching techniques. Under a two-regime-switching model of stock market returns, the good (bad) regime is characterized by a high (low) market mean return and low (high) market volatility. A simple method is proposed to estimate good- and bad-regime means, volatility, and cross-correlations for a large number of individual stocks. We find that the cross-sectional relation between the bad-regime mean return and the expected stock return is significantly negative, and the relation between the average bad-regime cross-correlation and the expected stock return is significantly positive. The observed relations are consistent with hedging hypothesis that investors want to hedge against market downturns and volatile markets. Furthermore, stocks with high (low) one-step-ahead predicted returns estimated using only bad-regime variables earn substantially high (low) subsequent returns and abnormal returns. Chapter 2, "Short-Term Reversals and Trading Activity", takes the interaction between prior returns and prior trading activities into consideration. Using a sample that excludes micro-cap stocks, we find that short-term reversals in monthly stock returns are strongly linked to prior monthly trading activity. Stocks with low turnover display a pronounced reversal effect, whereas those with high turnover display a continuation effect (momentum). The results are similar if we restrict the sample to large-cap stocks. Our analysis suggests that turnover is linked to short-term autocorrelation patterns in returns because it proxies for the flow of news that spurs speculative trading, and that the likelihood of short-term reversals falls as the proportion of turnover that is driven by news increases. Chapter 3, "Portfolio Sorts via Nonparametric Regression: From B-Splines to Basis Portfolios", nests portfolio sorts within the B-spline regression framework.

DEDICATION

The dissertation is dedicated to my Mom, Junren Shi, who encourages me to pursue my research career with her endless love and unconditional support.

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to my academic advisor and dissertation committee chair, Professor Christopher M. Kirby, who guides and inspires me how to do quality research with his patience, profession, and precision. I have learned more than I could have ever imagined through our academic co-authorship.

I appreciate my other dissertation committee members: Professor I-Huan Ethan Chiang, Professor Yufeng Han, and Professor Hwan C. Lin for their valuable suggestions regarding my dissertation and for their availability to help and constant encouragement.

I would like to thank other finance faculty members at the Belk College of Business, especially the director of the PhD program, Professor David C. Mauer, for their support since my first day in the doctoral program.

Last but not least, I would also like to thank my fellow graduate students and my friends, who have helped and supported me during my writing process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Numerous prior studies suggest that a number of firm characteristics generate the cross-sectional variation in expected stock returns. My dissertation takes a closer look at this question from three perspectives.

First, we relate regime switching to the question. The stock market experiences upturns and downturns as well as calm and volatile states interchangeably. To capture the properties, we employ a regime-switching model that allows means and volatility of monthly market returns to depend on an unobserved regime generated by a two-state Markov chain. For the sample period from January 1940 to December 2017, the good regime with an unconditional probability of 78.2% is characterized when the market displays high mean returns and low volatility, and the bad regime with an unconditional probability of 21.2% is characterized when the market displays low mean returns and high volatility. Even though the bad regime has occurred infrequently, investors may want to hedge against it because the bad regime can be viewed as an unfavorable state for investment. For instance, investors may require low overall returns for stocks that have high average returns in the bad regime when the overall market exhibits low returns. As a result, one may expect to find the negative relation between expected stock returns and bad-regime mean returns, all else being equal. On the other hand, investors may require low overall returns for stocks that have low bad-regime correlations with other stocks in consideration of risk reduction that is mostly needed in the bad regime when the overall market is volatile. Thus, one may expect to find a positive relation between the expected return of a stock and its average bad-regime correlations with other stocks, all else being equal.

Then questions become: How does one estimate mean returns, volatility, and cross-correlations in the good and bad regimes for a large cross-section of individual stocks in the condition where the underlying regime is unobserved? Do cross-sectional variations in regime-switching mean returns, volatility, and cross-correlations have important influence on the cross-sectional variation in expected stock returns, as suggested above? If they do, is there any economic significance of the observed relations?We attempt to address the questions in this paper.

We propose a simple method to estimate means, volatility, and cross-correlations in the good and bad regimes for a large number of individual stocks. For a given sample period, means and covariances of individual stock returns in the good (bad) regime can be constructed as probability-weighted average returns and probability-weighted average products of demeaned returns, respectively, where the probability is the normalized good (bad)-regime probability that measures the relative likelihood of an observation drawn from the good (bad) regime across all observations. Good- and bad-regime probabilities for each month can be estimated from a two-regime-switching model that allows regime-dependent means and volatility of market returns using monthly market data. The method allows us to estimate means, volatility, and cross-correlations in the good and bad regimes for a large cross-section of stocks as simply as to calculate their sample counterparts in a setting without regime switching as long as regime probabilities are estimated from the market model upfront.

Then we fit cross-sectional regressions of stock returns on means, volatility, and average cross-correlations in the good and bad regimes month by month. We find evidence that the regime-switching variables have important influence on the cross-section of expected stock returns, and most of the variation in expected returns captured by the regressions comes from the variation in bad-regime variables. Specifically, bad-regime means have significantly negative marginal effect while good-regime means have insignificantly positive marginal effect. Average bad-regime cross-correlations have significantly positive marginal effect while average good-regime cross-correlations have insignificantly positive marginal effect. The first two findings are consistent with the hedging hypothesis that a stock with high mean returns in the bad regime can serve as a good hedge against market downturns, and a stock with low average cross-correlations in the bad regime can serve as a good hedge against volatile markets when risk reduction is mostly needed. The findings also reveal the important role of information contained in the bad regime. In addition, bad-regime volatility have insignificantly positive marginal effect while good-regime volatility have significantly negative effect. The negative relation between good-regime volatility and expected returns is not surprising, given that the good regime is far more common than the bad regime.

Finally, we examine the economic significance of the observed relations between regime-switching variables and expected stock returns using portfolio sorts. In fact, cross-sectional regressions suggest that one can form one-step-ahead predicted returns based only on means, volatility, and cross-correlations in the bad regime since most of the variation in expected returns is captured by bad-regime variables. We predict one-step-ahead returns, as suggested by the cross-sectional regressions, and then form quintile portfolios on the predicted returns. It turns out that stocks with high (low) predicted returns earn high (low) average returns as well as high (low) abnormal returns. The difference in average annualized returns between the two extreme quintiles is as large as 10.9%, and the differences in abnormal returns are even larger. We also examine the robustness of the results using dependent double sorts that control for the market-based variables including short-term reversals, momentum, book-to-market ratio, and size. The double-sort results confirm the cross-sectional predictability of one-step-ahead predicted returns. Furthermore, we check the robustness using cross-sectional regressions that control for both market-based and accounting-based variables that have been known to have important influence on expected stock returns. Specifically, we run cross-sectional regressions of stock returns on one-step-ahead predicted returns and various control variables month by month and find the average coefficient on predicted returns is highly significant across different specifications. For instance, it has a t-statistics of 4.13 in the specification that includes all controls.

Second, we take interaction between prior trading activity and prior returns into consideration. The short-term reversal effect has remained an intriguing puzzle for well over two decades. We show that the likelihood of short-term reversals in monthly stock returns is strongly influenced by prior levels of monthly trading activity. Specifically, the cross-sectional relation between monthly returns and the first lag of monthly returns is highly dependent on prior monthly turnover. Although stocks that have low prior turnover display a pronounced reversal effect, those that have high prior turnover display a continuation effect. In other words, high prior turnover is associated with short-term momentum rather than short-term reversals in monthly stock returns.

We begin with evidence from portfolio sorts using data for July 1963 to December 2018. In particular, we construct a set of 25 value-weighted portfolios by sorting stocks into quintiles using turnover and then sorting the stocks in each turnover quintile into return quintiles. Our interest centers on the performance of long-short portfolios that are formed from the high- and low-prior-return quintiles. To reduce the influence of firms whose economic importance is debatable, the sample used to construct the portfolios excludes stocks whose market equity is below the 20th percentile of the NYSE market equity distribution on a month-by-month basis (the "all-but-micro-cap" sample). For stocks in the bottom quintile of prior turnover, the winners-minus-losers (WML) portfolio has an average return of -0.83% per month, which has a t-statistic of -4.87. In contrast, the WML portfolio for stocks in the top quintile of prior turnover has an average return of 0.58% per month, which has a t-statistic of 2.35. Thus the reversal effect, which is quite strong for low-turnover stocks, is nonexistent for high-turnover stocks.

What explains the link between the autocorrelation properties of returns and prior levels of trading activity? Our analysis points to an explanation for the cross-sectional relation between monthly return reversals and prior trading activity. We start from the premise that a considerable fraction of monthly trading activity is probably motivated by news that changes expectations of future payoffs. This should not be controversial in view of existing models of speculative trading and the price-discovery process. Under the well-known Tauchen and Pitts (1983) model, for example, squared daily stock returns and daily trading volume share a common factor: the rate at which information that alters stock valuations arrives to market. It implies, in other words, that the flow of news drives the dynamics of both volume and volatility.

We therefore conduct a simple test to see whether the role of news in generating turnover might explain our findings. If the interaction between return reversals and prior turnover is linked to the flow of unobserved news, then the Tauchen and Pitts (1983) model predicts that there should be a similar interaction between return reversals and prior volatility. We find that this is indeed the case. The interaction between return reversals and prior volatility is both negative and highly statistically significant. This finding lends indirect support to the hypothesis that turnover acts as proxy for the flow of news that drives speculative trading.

Under this hypothesis, the interaction between short-term reversals and prior turnover has a straightforward interpretation. Consider a stock that falls in the lower tail of the cross-sectional distribution of returns for the month. If the turnover for the stock is low, then it has experienced a below-average return over a period in which the flow of news has been relatively low. The data indicate that this below-average return is likely to be followed by an above-average return over the next month (a short-term reversal effect). Conversely, if the turnover for the stock is high, then it has experienced a below-average return over a period in which the flow of news has been relatively high. The data indicate that this below-average return is likely to be followed by a below-average return over the next month (a short-term momentum effect).

Linking short-term return reversals to prices changes that occur in the absence of much news captures the basic spirit of Campbell et al. (1993) model. More broadly, it is consistent with the implications of the speculative trading model developed by Llorente et al. (2002). The model assumes that there are two basic types of trades: hedging and speculative. Hedging trades convey no signal about future payoffs, so the returns generated by these trades display reversals. Speculative trades, on the other hand, are driven by new information that is only partially incorporated into the stock price in a given trading session. Thus the returns generated by speculative trades display continuations. Because the model implies that the impact of speculative trades is fundamentally different than that of hedging trades, it predicts that cross-sectional differences in the relative importance of speculative trading should lead to differences in the autocorrelation properties of returns across firms.

We use an easily-constructed proxy for the fraction of turnover that is driven by speculative trading to investigate whether our findings are consistent with this prediction. The monthly estimated correlation between the squared demeaned daily returns and daily turnover should be a useful proxy for the relative contribution of news-driven trades to monthly turnover figures for individual stocks.

Conditioning on this correlation lends further credence to the information-flow hypothesis. For instance, we replicate the portfolio sorts after partitioning the stocks in the all-but-micro-cap sample into two categories on a month-by-month basis: those which have a low fraction of news-driven turnover (estimated correlations below the median value) and those that have a high fraction of news-driven turnover (estimated correlations above the median value). For stocks in the bottom quintile of prior turnover, the WML portfolio has an average return of −1.03% per month for the former category (t-statistic of −6.25) and −0.75% per month for the latter category (t-statistic of −3.84). For stocks in the top quintile of prior turnover, however, the WML portfolio has an average return of −0.35% per month for the former category (t-statistic of −1.32) and 1.06% per month for the latter category (t-statistic of 3.69). Thus the reversal effect is much stronger for stocks with a low fraction of news-driven turnover, which is consistent with the predictions of the Llorente et al. (2002) model. Again, the results are similar for WML portfolios formed from large-cap stocks.

To supplement the evidence produced by the portfolio sorts, we fit a series of cross-sectional regressions that control for other price-related anomalies. In particular, we sort stocks into deciles using monthly turnover, and then regress the returns for the stocks in selected deciles on lagged monthly values of returns, log turnover, log realized volatility, log market equity, and a standard measure of price momentum. For stocks in the bottom decile of prior turnover, the average estimated slope on prior returns is −0.70 with a t-statistic of −10.14. For stocks in the top decile of prior turnover, the average estimated slope on prior returns is 0.13 with a t-statistic of 2.75. Hence, we again find that high prior turnover is associated with short-term momentum rather than short-term reversals in monthly stock returns.

Third, we construct basis portfolios that capture the cross-sectional variation associated with some firm characteristic using B-spline regression, which is a class of spline regression that is able to capture the potential nonlinearity in relation between firm characteristics and the expected stock returns. The recent two studies of Freyberger et al. (2019) and Kirby (2019) provide convincing evidence of nonlinearity in the relation between firm characteristics and expected stock returns. The nonlinearity motivates us to construct basis portfolios using B-spline regression, which is a class of spline regression that is able to capture the potential nonlinearity. Notably, B-spline regression is a more general non-parametric technique than the conventional portfolio sorts. In fact, the conventional portfolios sorting procedure is a special case of B-spline regressions of degree zero. For example, the returns for equally-weighted quintile portfolios formed on size are the OLS estimates of B-spline regressions of stock returns on basic functions of degree zero of the size variable using corresponding knot sequence. As a consequence, B-spline regression has a number of advantages over the conventional portfolio sorts.

CHAPTER 1: REGIME SWITCHING AND THE CROSS-SECTION OF EXPECTED STOCK RETURNS

1.1 Introduction

The stock market experiences upturns and downturns as well as calm and volatile states interchangeably. To capture the properties, we employ a regime-switching model that allows means and volatility of monthly market returns to depend on an unobserved regime generated by a two-state Markov chain.[1,2] For the sample period from January 1940 to December 2017, the good regime with an unconditional probability of 78.2% is characterized when the market displays high mean returns and low volatility, and the bad regime with an unconditional probability of 21.2% is characterized when the market displays low mean returns and high volatility. Even though the bad regime has occurred infrequently, investors may want to hedge against it because the bad regime can be viewed as an unfavorable state for investment. For instance, investors may require low overall returns for stocks that have high average returns in the bad regime when the overall market exhibits low returns. As a result, one may expect to find the negative relation between expected stock returns and bad-regime mean returns, all else being equal. On the other hand, investors may require low overall returns for stocks that have low bad-regime correlations with other stocks in consideration of risk reduction that is mostly needed in the bad regime when the overall market is volatile. Thus, one may expect to find a positive relation between the expected return of a stock and its average bad-regime correlations with other stocks, all else being equal.

Then questions become: How does one estimate mean returns, volatility, and cross-correlations in the good and bad regimes for a large cross-section of individual stocks in the condition where the underlying regime is unobserved? Do cross-sectional variations in

[1] The approach follows existing literature that includes Turner et al. (1989) and Ang and Timmermann (2012) to name a few.

[2] The empirical results presented in this paper barely change if we identify regimes using returns on market, size, and value portfolios jointly.

regime-switching mean returns, volatility, and cross-correlations have important influence on the cross-sectional variation in expected stock returns, as suggested above? If they do, is there any economic significance of the observed relations?We attempt to address the questions in this paper.

First, we propose a simple method to estimate means, volatility, and cross-correlations in the good and bad regimes for a large number of individual stocks. For a given sample period, means and covariances of individual stock returns in the good (bad) regime can be constructed as probability-weighted average returns and probability-weighted average products of demeaned returns, respectively, where the probability is the normalized good (bad)-regime probability that measures the relative likelihood of an observation drawn from the good (bad) regime across all observations. Good- and bad-regime probabilities for each month can be estimated from a two-regime-switching model that allows regime-dependent means and volatility of market returns using monthly market data. The method allows us to estimate means, volatility, and cross-correlations in the good and bad regimes for a large cross-section of stocks as simply as to calculate their sample counterparts in a setting without regime switching as long as regime probabilities are estimated from the market model upfront.

Second, we fit cross-sectional regressions of stock returns on means, volatility, and average cross-correlations in the good and bad regimes month by month.[3,4] We find evidence that the regime-switching variables have important influence on the cross-section of expected stock returns, and most of the variation in expected returns captured by the regressions comes from the variation in bad-regime variables. Specifically, bad-regime means have significantly negative marginal effect while good-regime means have insignificantly positive marginal effect. Average bad-regime cross-correlations have significantly positive marginal effect while average good-regime cross-correlations have insignificantly positive marginal effect. The first two findings are consistent with the hedging hypothesis that a stock with high mean

[3]Each month, we estimate the regime-switching variables using data for the month and earlier to avoid look-ahead bias.

[4]The average cross-correlation of a stock is defined as the average of all pair-wise cross-correlations between the stock and other stocks in the market. The average cross-correlation can be viewed as an approximate measure of risk reduction that a stock can provide — the lower it is, the more likely the stock can provide diversification benefits and thus reduce risk.

returns in the bad regime can serve as a good hedge against market downturns, and a stock with low average cross-correlations in the bad regime can serve as a good hedge against volatile markets when risk reduction is mostly needed. The findings also reveal the important role of information contained in the bad regime. In addition, bad-regime volatility have insignificantly positive marginal effect while good-regime volatility have significantly negative effect. The negative relation between good-regime volatility and expected returns is not surprising, given that the good regime is far more common than the bad regime. In fact, the finding is consistent with the negative cross-sectional relation between overall volatility and expected returns, first documented by Ang et al. (2006).

Third, we examine the economic significance of the observed relations between regime-switching variables and expected stock returns using portfolio sorts. In fact, cross-sectional regressions suggest that one can form one-step-ahead predicted returns based only on means, volatility, and cross-correlations in the bad regime since most of the variation in expected returns is captured by bad-regime variables. We predict one-step-ahead returns, as suggested by the cross-sectional regressions, and then form quintile portfolios on the predicted returns. It turns out that stocks with high (low) predicted returns earn high (low) average returns as well as high (low) abnormal returns relative to the Carhart (1997) four-factor model and the Fama and French (2016) five-factor model. The difference in average annualized returns between the two extreme quintiles is as large as 10.9%, and the differences in Carhart four-factor alphas and Fama-French five-factor alphas are even larger — 13.2% and 12.5%, respectively. We also examine the robustness of the results using dependent double sorts that control for the market-based variables including short-term reversals, momentum, book-to-market ratio, and size. The double-sort results confirm the cross-sectional predictability of one-step-ahead predicted returns. Furthermore, we check the robustness using cross-sectional regressions that control for both market-based and accounting-based variables that have been known to have important influence on expected stock returns. Specifically, we run cross-sectional regressions of stock returns on one-step-ahead predicted returns and various control variables month by month and find the average coefficient on predicted returns is highly significant across different specifications. For instance, it has a t-statistics of 4.13 in the specification that includes all controls.

The study is distinct from the existing regime-switching literature in two dimensions. First, several studies focus on time-series predictability of stock returns at market level using regime-switching models. For example, Henkel et al. (2011) documents that the short-horizon performance of market return predictors such as the dividend yield and the short rate appears insignificant during expansions but sizeable during contractions under a regime-switching vector autoregression framework. Zhu and Zhu (2013) find that the regime-switching combination forecasts of stock returns deliver higher gains relative to the historical average. However, we investigate the cross-sectional predictability of stock returns based on regime-switching variables. Most importantly, the findings reveal the important information contained in the bad regime, even though it occurs less frequently. Second, another strand of literature focuses either on the international asset allocation or on the optimal portfolio choice problem among several different asset classes or several well-diversified portfolios. For example, Ang and Bekaert (2015) and Guidolin and Timmermann (2008) solve the international asset allocation using regime-switching models that allow investment opportunity sets are regime-dependent. Tu (2010) takes regime-switching into account when making portfolio decisions among a small section of well-diversified portfolios and finds that ignoring regime-switching may incur large losses. However, little has been done in terms of linking regime-switching techniques to a large cross-section of individual stocks. This study attempts to fill the gap in the regime-switching literature.

The remainder of the paper proceeds as follows. Section 2 describes the data. In Section 3, we describe the two-regime-switching model for the market and discuss the construction of regime-switching variables for individual stocks. Section 4 first investigates the cross-sectional relation between the estimated regime-switching means, volatility, and average cross-correlations and expected stock returns, and then examines the economic significance of the observed relations. Section 5 concludes.

1.2 Data

The data include all common stocks (CRSP share code 10 or 11) listed on NYSE, AMEX, and NASDAQ obtained from the Center for Research in Securities Prices (CRSP). The five Fama and French (2016) factors are obtained from Ken French's website. Other firm characteristics are drawn from COMPUSTAT. We use monthly data for the sample period from January 1940 to December 2017.

1.3 Model and Variables

In this section, we first discuss the regime-switching model of market returns. Second, we describe how to estimate regime-switching means, volatility, and average cross-correlations for individual stocks.

1.3.1 The two-regime-switching model of market returns

Consider the following model that allows means and volatility of market returns are regime-dependent:

$$R_{m,t} = \mu_m^{s_t} + \sigma_m^{s_t}\epsilon_t, \quad \epsilon_t \sim i.i.d.N(0,1), \tag{1.1}$$

where $R_{m,t}$ is the market return for month t, s_t is the unobserved regime for month t, and μ_{m,s_t} and σ_{m,s_t} are the mean and volatility of market returns in regime s_t.

Assume that the random regime $s_t \in \{G, B\}$, and it is generated by an irreducible first-order Markov chain with constant transition probabilities. Let p_{lk} denote the constant transition probability from $s_t = l$ to $s_{t+1} = k$ for $l, k \in \{G, B\}$, i.e. $p_{lk} = \Pr(s_{t+1} = k | s_t = l)$. Let P denote the transition matrix that collects all transition probabilities:

$$P = \begin{bmatrix} p_{GG} & p_{BG} \\ p_{GB} & p_{BB} \end{bmatrix}.$$

Note that the sum of each column is one, and $p_{kk} < 1$ for any $k \in \{G, B\}$ as a result of irreducibility of the Markov chain. The unconditional probability of being in regime k for a month is given by[5]

$$\pi_k = \frac{1 - p_{ll}}{2 - p_{ll} - p_{kk}}. \tag{1.2}$$

Let $\boldsymbol{\theta} = \{\mu_m^G, \mu_m^B, \sigma_m^G, \sigma_m^B, p_{GG}, p_{BB}\}$ denote a set of unknown parameters. θ can be estimated by maximum likelihood estimation.[6] In fact, one never knows for sure about which regime was for month t based on observed data even knowing the true value of $\boldsymbol{\theta}$ because

[5] See Chapter 22 of Hamilton and Press (1994).
[6] See Appendix.

regime is unobserved. The best one can do is to infer the probability about which regime was most likely to be responsible for producing the market return observation for month t. Let $I_t = \{R_{m,t}, R_{mt-1}, \ldots\}$ be an information set containing all monthly market return observations up to month t. The regime probabilities, $\omega^k_{\tau|t} = \Pr(s_\tau = k|I_t; \boldsymbol{\theta})$, denote the probability of being in regime i for month τ conditional on the information set I_t and the knowledge of parameters $\boldsymbol{\theta}$. The regime probabilities can be obtained as by-products of the maximum likelihood estimation.[7]

On one hand, a number of studies that model market returns in a regime-switching context suggest that regimes are characterized by market volatility. For example, Turner et al. (1989) and Ang and Timmermann (2012) model monthly excess returns on S&P 500 index under a two-regime-switching setting. They find that stock market volatility are significantly different across regimes. This is indeed the case. Table 1.1 reports parameter estimates for model (1.1) using monthly value-weighted market returns over the sample period from January 1940 to December 2017. Market volatility in regime B is 7.0% per month that doubles market volatility in regime G (3.4% per month). Following Turner et al. (1989), we test the hypothesis that $\sigma^G_m = \sigma^B_m$ using likelihood ratio test with a modified statistic proposed by Wolfe (1971). The null hypothesis can be rejected at 1% significance level. On the other hand, regimes are also characterized by market mean returns. As shown in Table 1, the market mean in regime G is 1.4% per month that is much higher than the market mean in regime B (-0.7% per month). The hypothesis that $\mu^G_m = \mu^B_m$ can be rejected at 5% significance level. It is clear that market returns exhibit higher means and lower volatility in regime G, and lower means and higher volatility in regime B. Therefore, we term regime G as the "good" regime and regime B as the "bad" regime. And it is relatively less likely for a month to be in the bad regime since the unconditional probability for the bad regime is $(1 - 0.877)/(2 - 0.967 - 0.877) = 21.2\%$, which is relatively low compared to 78.8% — the unconditional probability for the good regime.

[7] See Appendix.

1.3.2 Regime-switching means, volatility, and average cross-correlations of individual stock returns

In this section, we describe how to estimate means, volatility, and cross-correlations in the good and bad regimes for individual stocks. Let $R_{i,t}$ denote the stock return on stock i for month t. Let $\mu^k_{i|t}$ denote the estimated mean return on stock i in regime $k \in \{G, B\}$ conditional on information up to month t. Let $\sigma^k_{ij|t}$ denote the estimated covariance between the return on stock i and the return on stock j in regime k conditional on information up to month t. $\mu^k_{i|t}$ and $\sigma^k_{ij|t}$ are given by the following equations:

$$\mu^k_{i|t} = \sum_{\tau \leq t} R_{i,\tau} q^k_{\tau|t}, \tag{1.3}$$

$$\sigma^k_{ij|t} = \sum_{\tau \leq t} (R_{i,\tau} - \mu^k_{i|t})(R_{j,\tau} - \mu^k_{j|t}) q^k_{\tau|t}, \tag{1.4}$$

where

$$q^k_{\tau|t} = \frac{\omega^k_{\tau|t}}{\sum\limits_{\tau \leq t} \omega^k_{\tau|t}} \in (0, 1), \tag{1.5}$$

and $\omega^k_{\tau|t}$ is the regime probability defined in section 3.1. Equation (1.3) and (1.4) are intuitive as they describe $\mu^k_{i|t}$ and $\sigma^k_{ij|t}$ as a probability-weighted estimator that satisfies a condition where each observation for month t is weighted by its relative likelihood that it was drawn from regime k across all observations up to month t. Derived from equation (1.4), the estimated correlation between the return on stock i and stock j in regime k conditional on information up to month t is

$$\rho^k_{ij|t} = \frac{\sum\limits_{\tau \leq t} (R_{i,\tau} - \mu^k_{i|t})(R_{j,\tau} - \mu^k_{j|t}) q^k_t}{\sqrt{\sum\limits_{\tau \leq t} (R_{i,\tau} - \mu^k_{i|t})^2 q^k_t} \sqrt{\sum\limits_{\tau \leq t} (R_{j,\tau} - \mu^k_{j|t})^2 q^k_t}}. \tag{1.6}$$

Let $\sigma^k_{i|t}$ denote the estimated volatility of stock i in regime k conditional on information up to month t, i.e.

$$\sigma^k_{i|t} = \sqrt{\sum_{\tau \leq t} (R_{i,\tau} - \mu^k_{i|t})^2 q^k_\tau}. \tag{1.7}$$

Assume there are N individual stocks. Let $\rho^k_{i|t}$ denote the average of all estimated pair-wise cross-correlations between the return on stock i and the return on all other $(N-1)$ stocks in regime k conditional on information up to month t, i.e.

$$\rho^k_{i|t} = \frac{1}{N-1}\sum_{j\neq i}\rho^k_{ij|t}. \tag{1.8}$$

Therefore, we can estimate the regime-switching variables for each stock conditional on information up to each month t following a two-step procedure: first, we estimate the market model (1.1) via maximum likelihood estimation using monthly market returns up to month t and obtain the regime probabilities $\omega^k_{\tau|t}$ for all $\tau \leq t$ and $k \in \{G, B\}$; second, we estimate means, volatility, and average cross-correlations based on equation (1.3), (1.7), and (1.8) using monthly individual stock returns up to month t combined with the regime probabilities obtained from the first step. Therefore, we obtain a set of regime-switching variables, $\{\mu^G_{i|t}, \mu^B_{i|t}, \sigma^G_{i|t}, \sigma^B_{i|t}, \rho^G_{i|t}, \rho^B_{i|t}\}$, for each individual stock i for each month t. Note that these variables are estimated using information up to month t without including any forward-looking information.

I turn to examine whether means, volatility, and average cross-correlations of individual stock returns vary across regimes. Table 1.2 reports a set of descriptive statistics of estimated means, volatility, log volatility, and average cross-correlations in the good and bad regimes for individual stocks for the pooled sample from July 1963 to December 2017.[8] First, it is not surprising to find that the mean and the median of bad-regime means for individual stocks are negative and much lower than the counterparts of good-regime means because the overall market exhibits low average returns in the bad regime. Second, bad-regime volatility is, on average, slightly higher than good-regime volatility. Volatility in each regime has large positive skewness, while the log volatility has nearly zero skewness. Third, average bad-regime cross-correlations are, on average, twice as high as average good-regime cross-correlations. There are few studies that have attempted to compare cross-correlations between individual stocks across different market states, however, some have documented

[8]The initial window for estimating the smoothed regime probabilities and the regime-switching variables is set from January 1940 to June 1963, and thus the sample period starts from July 1963.

that correlations between international stock markets are much greater during volatile bear market. The studies include Longin and Solnik (2001), Ang and Bekaert (2015), and Ang and Chen (2002) among others. In this regard, the finding is closely linked to the existing literature.

Even though means, volatility, and cross-correlations of individual stock returns vary with regimes, as suggested by Table 1.2, it does not rule out the possibility that cross-sectional variations in these variables are highly correlated across regimes. For example, if stocks that fall in the upper tail of the cross-sectional distribution of returns in the good regime are always those stocks that are winners in the bad regime, then it might be meaningless to differentiate the bad regime from the good regime in the context of this study. Therefore, we examine the time series average of cross-sectional correlations between regime-switching variables and report the results in Table 1.3. For instance, the average correlation between good- and bad-regime means is small (0.16). The average correlation between log good- and bad-regime volatility is 0.82, which indicates the variation in good-regime volatility explains about 67% of the variation in bad-regime volatility. The average correlation between average good- and bad-regime cross-correlations is 0.59, which indicates the variation in good-regime cross-correlations only explains about 35% of the variation in bad-regime cross-correlations. These correlations suggest that cross-sectional variations in means, volatility, and average cross-correlations are not highly correlated across regimes.

1.4 Regime-switching and the Cross-Section of Expected Stock Returns

In this section, we first investigate the cross-sectional relation between expected stock returns and regime variables including means, volatility, and average cross-correlations of stock returns in the good and bad regimes using cross-sectional regressions. Then we construct one-step-ahead predicted returns based on the regime-switching variables, and examine the cross-sectional predictability of the obtained predicted returns.

1.4.1 Regime-switching variables and the cross-section of expected stock returns

In order to investigate the cross-sectional relation between expected stock returns and regime variables, we consider the following cross-sectional regression for month t:

$$R_{i,t} = a_{0t} + a_{1t}\mu^G_{i|t-1} + a_{2t}\mu^B_{i|t-1} + a_{3t}\log(\sigma^G_{i|t-1}) + a_{4t}\log(\sigma^B_{i|t-1}) + a_{5t}\rho^G_{i|t-1} + a_{6t}\rho^B_{i|t-1} + u_{i,t}, \quad (1.9)$$

where $R_{i,t}$ is the month-t return on stock i, independent variables are the regime variables that are obtained based on equation (1.3), (1.7), and (1.8) conditional on information up to month $t-1$. We winsorize all independent variables at 0.5% and 99.5% to reduce the effect of outliers.

I fit cross-sectional regression (1.9) month by month from July 1963 to December 2017. Average estimated coefficients of these regressions, and their heteroskedasticity-consistent t-statistics are reported in panel A of Table 1.4. First, the average slope on good-regime means is statistically positive at 1% significance level while the average slope on bad-regime means is significantly negative at 10% significance level. Second, the average slope on average good-regime cross-correlations is positive but insignificant and the average slope on average bad-regime cross-correlations is positive with a t-statistic of 3.10. Third, the average slope on good-regime volatility is significantly negative at 1% significance level while the average slope on bad-regime volatility is insignificantly positive.

However, numerous empirical studies have shown that there are several market-based variables that have been known to have important variation in expected stock returns, including the short-term reversal effect, momentum effect, value effect, and size effect.[9] Therefore, it is necessary to rule out the effects of these market-based variables in order to isolate the marginal effect of regime-switching variables. Thus, we consider the following cross-sectional regression for month t,

$$\hat{e}_{i,t} = c_{0t} + c_{1t}\mu^G_{i|t-1} + c_{2t}\mu^B_{i|t-1} + c_{3t}\log(\sigma^G_{i|t-1}) + c_{4t}\log(\sigma^B_{i|t-1}) + c_{5t}\rho^G_{i|t-1} + c_{6t}\rho^B_{i|t-1} + v_{i,t}, \quad (1.10)$$

[9] For example, Jegadeesh (1990), Jegadeesh and Titman (1993), Chan et al. (1991), Banz (1981), and Fama and French (1992).

where the dependent variable $\hat{e}_{i,t}$ is the residual of the following cross-sectional regression:

$$R_{i,t} = b_{0t} + b_{1t}R_{i,t-1} + b_{2t}MOM_{i,t-1} + b_{3t}\log(B/M_{i,t-1}) + b_{4t}\log(ME_{i,t-1}) + e_{i,t}, \quad (1.11)$$

where $MOM_{i,t-1}$ is the buy-and-hold return on over the period from month $t-12$ to month $t-2$, $ME_{i,t-1}$ is market equity for month $t-1$, and $B/M_{i,t-1}$ is book-to-market ratio for month $t-1$. We winsorize all independent variables at 0.5% and 99.5% to reduce the effect of outliers.

I fit cross-sectional regression (1.10) month by month from July 1963 to December 2017. Panel B of Table 1.4 reports the average estimated coefficients of these regressions, and their heteroskedasticity-consistent t-statistics. The results disclose a different picture. First, the average slope on good-regime means becomes insignificant. The average estimated coefficient on bad-regime means remains negative with a more significant t-statistic of -2.64. Second, the average slope on average good-regime cross-correlations is insignificant, and the average slope on average bad-regime cross-correlations is positive with a t-statistic of 2.51. Third, the average slope on log good-regime volatility is negative with a t-statistic of -2.77, while the average slope on log bad-regime volatility is insignificantly positive.

The significant negative coefficient on bad-regime implies that investors require low returns for stocks that have high returns in a regime where the overall market is bad. The significant positive coefficient on average bad-regime cross-correlations suggest that investors require low returns for stocks that reduce risk when the overall market is more volatile. A potential explanation to the results is that investors would like to hedge against the bad regime by paying a high price and thus requiring a low return for holding those stocks that display high bad-regime returns and low bad-regime correlations with other stocks. The findings highlight the important role of information contained in the bad regime. Note that the relation between good-regime volatility and expected returns is negative, which is consistent with the negative cross-sectional relation between overall volatility and expected returns, first documented by Ang et al. (2006), given the fact that the good regime is far more common than the bad regime. Overall, the results reported in Table 1.4 suggest that regime-switching means, volatility, and average cross-correlations can explain the variation in the cross-section of expected stock returns, and that most of the variation in expected returns captured by the cross-sectional regressions is related to the variation in bad-regime variables.

1.4.2 Predicted stock returns and the cross-section of expected stock returns

The above findings suggest that one can form one-step-ahead predicted stock returns for month $t+1$ based only on bad-regime variables that are estimated conditionally on the information set at the end of month t since most of the variation in expected returns is captured by bad-regime variables. Therefore, we construct one-step-ahead predicted stock returns as follows:

$$\hat{R}_{i,t+1} = \hat{c}_{2t}\mu^{B}_{i|t} + \hat{c}_{4t}\rho^{B}_{i|t} + \hat{c}_{6t}\log(\sigma^{B}_{i|t}), \tag{1.12}$$

where $\hat{c}_{jt}$ for $j \in \{2, 4, 6\}$ are the estimated coefficients of regression (1.10) for month t. The way that we estimate predicted returns follows Jegadeesh (1990). Then the natural question is, do stocks with high (low) predicted returns earn high (low) subsequent returns? We answer the question using evidence obtained from portfolio sorts and cross-sectional regressions.

1.4.2.1 Evidence from portfolio sorts

I form quintile portfolios for month $t+1$ on one-step-ahead predicted stock returns estimated using information up to month t in the following fashion: first, we form one-step-ahead predict stock returns for month $t+1$ using equation (1.12); second, stocks are ranked in ascending order based on their predicted returns for month $t+1$; third, stocks in the bottom quintile are assigned to quintile low, stocks in the next quintile are assigned to quintile 2, and so on so forth. All portfolios are value-weighted and rebalanced every month.

Table 1.5 summarizes the descriptive statistics for the quintile portfolios. The cross-sectional predictability of the estimated predicted returns shows up clearly in Table 1.5. Average monthly portfolio returns increase monotonically as one moves from portfolio low to portfolio high, and the difference in average returns between portfolio high and low is 0.91% with a t-statistic of 3.02, indicating a remarkable annualized spread of 10.9%. Carhart four-factor alphas and Fama-French five-factor alphas produce a similar pattern as average portfolio returns do and the dispersion in alphas is even more pronounced. The difference in Carhart alphas between the two extreme quintiles is 1.10% with a t-statistic of 3.08, and the difference in FF-5 alphas is 1.10% with a t-statistic of 3.20. Therefore, the Carhart four-factor model and Fama-French five-factor model do not help explain the variation in average returns on the quintile portfolios. In addition, the differences in average market equity, market equity shares, and average book-to-market ratios of the portfolios are very small across quintiles. It is unlikely that size or book-to-market ratio drives the results.

One potential concern in the results presented in Table 1.5 is that predicted stock returns may be correlated with variables that have been identified as cross-sectional predictors of expected stock returns, such as prior-month returns, momentum, book-to-market ratio and size.[10] Then sorting on predicted stock returns may simply capture the effects of these variables. We address the concern using conditional double sort as follows: first, we form one-step-ahead stock returns for month $t+1$ using equation (1.12); second, we sort stocks into control quintiles on the basis of the ranking of corresponding control variables. The control variable is the market equity measured at the end of month t (size), return for month t (prior return), buy-and-hold return over month $t-11$ to $t-1$ (momentum), or book-to-market ratio (B/M ratio) for month t. Third, we sort stocks into predicted return quintiles based on the ranking of the predicted returns within each control quintile, and obtain a set of 5×5 portfolios for each control variable. All portfolios are value-weighted and rebalanced every month.

Table 1.6 presents Carhart four-factor alphas and their heteroskedasticity-consistent t-statistics for value-weighted portfolios formed on predicted returns and various control variables.[11] We first focus on the 25 portfolios formed on predicted returns and size. Alphas increase monotonically from the low-predicted-return quintile to the high-predicted-return quintile within each size quintile. The differences in average returns between high- and low-predicted-return quintiles are large and statistically significant within each size quin-

[10] Although we control for the return, momentum, book-to-market ratio, and size for month $t-1$ when we estimate the marginal effect of regime-switching variables, it is still possible that the obtained one-step-ahead predicted returns for month $t+1$ are correlated with the return, momentum, book-to-market ratio, and size for month t.

[11] Average portfolio returns and FF-5 alphas produce similar patterns as Carhart four-factor alphas do. Results are available upon request.

tile. The smallest one is 0.59% per month, with a t-statistics of 2.49, and the largest one is 0.98% per month, with a t-statistics of 3.67. Taking average over size quintile delivers a spread of 0.75% between the low-predicted-return quintile and the high-predicted-return quintile, indicating an annualized spread of 0.90%. For other control variables, we only report the alphas of quintile portfolios obtained by averaging over control quintiles for 25 portfolios formed on predicted returns and prior-month returns/momentum/book-to-market ratio due to space limitations. It turns out that alphas monotonically increase with predicted returns for the quintile portfolios that controls for prior-month returns, momentum, or book-to-market ratio. The spread between the low-predicted-return quintile and the high-predicted-return quintile is large in magnitude and statistically significant for each control variable. These results suggest that short-term reversal, momentum, book-to-market ratio, and size are unlikely to play an important role in the cross-sectional predictability of the estimated predicted returns.

1.4.2.2 Evidence from cross-sectional regressions

Portfolio sorts, although straightforward, cannot perfectly isolate the effect of predicted returns and fail to control for multiple variables simultaneously. Therefore, we turn to cross-sectional regressions to examine the predictability of one-step-ahead predicted returns that are estimated by bad-regime variables. Specifically, we run the following cross-sectional regression each month:

$$R_{i,t} = \lambda_t + \gamma_t \hat{R}_{i,t} + \beta_t' X_{i,t-1} + \epsilon_{i,t}, \tag{1.13}$$

where $\hat{R}_{it}$ is the one-step-ahead predicted return on stock i for month t estimated using equation (1.12), X_{it-1} is a column vector of control variables that can be obtained using information through month $t-1$. As a robustness check to the results presented in section 1.4.2.1, we include two categories of firm characteristics as control variables that are known as predictors of cross-sectional stock returns: first, market-based firm characteristics that we already take into consideration in the above analysis, including returns for the prior month, momentum, book-to-market ratios, and market capitalization; second, accounting-based firm characteristics, including asset growth rate, return on assets, gross profitability, net operating assets, investment-to-assets, O-score, and total accruals.[12] Again, we winsorize all independent variables at 0.5% and 99.5% to reduce the effect of outliers.

Table 1.7 reports the average estimated coefficients of all cross-sectional regressions, and their Fama-Macbeth t-statistics. Column (1) reports the results for regressions that only include the predicted stock return and a constant as regressors, the average estimated coefficient on the predicted stock return is positive with a t-statistic of 3.54. Next, adding market-based controls barely changes the results. Finally, we add various accounting-based controls on top of market-based controls. The average estimated coefficient on the predicted stock return is still significantly positive with a t-statistic of 4.13, as shown in column (3). Overall, it is credible to draw the conclusion that means, volatility, and average cross-correlations in the bad regime are able to predict stock returns cross-sectionally, and stocks with high one-step-ahead predicted returns estimated by the bad-regime variables earn substantially high average returns.

1.5 Conclusions

I employ a two-regime-switching model that allows means and volatility of monthly market returns to depend on the underlying regime that is generated by an irreducible first-order Markov chain with constant transition probabilities. Though the regime is unobserved, good and bad regime probabilities can be estimated from the market model. Combining monthly return observations with regime probabilities, one can easily estimate regime-switching means, volatility, and cross-correlations for a large number of individual stocks.

By fitting cross-sectional regressions that includes means, volatility, and cross-correlations in the good and bad regimes as regressors, we find that they have important influence on the cross-section of expected stock returns. The relation between bad-regime means and expected stock returns is significantly negative. The relation between average bad-regime cross-correlations and expected stock returns is significantly positive. The observed relations are consistent with hedging hypothesis that investors want to hedge against market down-

[12]See Cooper et al. (2008), Chen et al. (2011), Novy-Marx (2013), Hirshleifer et al. (2004), and Sloan (1996). Note that O-score is a measure of financial distress similar to the measure proposed by Campbell et al. (2008).

turns and volatile markets, and thus require low returns for stocks with high returns during market downturns and for stocks that reduce risk when the market is volatile. Furthermore, we construct one-step-ahead predicted returns based on the regime-switching variables and form quintile portfolios on predicted returns. We find evidence that stocks with high (low) predicted returns earn substantially high (low) subsequent returns.

The findings reveal the important role of information contained in the bad regime that may be ignored in a setting without regime switching and highlight the cross-sectional predictability of regime switching means, volatility, and cross-correlations on expected returns.

References

Ang, A. and Bekaert, G. (2015). International Asset Allocation With Regime Shifts. *The Review of Financial Studies*, 15(4):1137–1187.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.

Ang, A. and Timmermann, A. (2012). Regime changes and financial markets. *Annu. Rev. Financ. Econ.*, 4(1):313–337.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6):2899–2939.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.

Chan, L. K., Hamao, Y., and Lakonishok, J. (1991). Fundamentals and stock returns in japan. *the Journal of Finance*, 46(5):1739–1764.

Chen, L., Novy-Marx, R., et al. (2011). An alternative three-factor model. *Available at SSRN 1418117*.

Cooper, M. J., Gulen, H., and Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance*, 63(4):1609–1651.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2016). Dissecting anomalies with a five-factor model. *The Review of Financial Studies*, 29(1):69–103.

Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *The Review of Financial Studies*, 21(2):889–935.

Hamilton, J. and Press, P. U. (1994). *Time Series Analysis*. Princeton University Press.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99(3):560–580.

Hirshleifer, D., Hou, K., Teoh, S. H., and Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, 38:297–331.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance*, 45(3):881–898.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The journal of Finance*, 56(2):649–676.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1):1–28.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting review*, pages 289–315.

Tu, J. (2010). Is regime switching in stock returns important in portfolio decisions? *Management Science*, 56(7):1198–1215.

Turner, C. M., Startz, R., and Nelson, C. R. (1989). A markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, 25(1):3–22.

Wolfe, J. H. (1971). A monte carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions. Technical report, NAVAL PERSONNEL AND TRAINING RESEARCH LAB SAN DIEGO CA.

Zhu, X. and Zhu, J. (2013). Predicting stock returns: A regime-switching combination approach and economic links. *Journal of Banking and Finance*, 37(11):4120–4133.

Appendix

The regime probabilities, $\omega_{\tau|t}$, for $\tau \in \{1, 2, \ldots, t\}$ are given by the following algorithm:

$$\omega^k_{\tau|\tau} = \frac{\omega^k_{\tau|\tau-1}\eta^k_\tau}{\sum\limits_{l=L,H} \omega^l_{\tau|\tau-1}\eta^l_\tau} \tag{1.14}$$

$$\omega^k_{\tau+1|\tau} = \sum_{l=L,H} \omega^l_{\tau|\tau} p_{lk} \tag{1.15}$$

$$\omega^k_{\tau|t} = \omega^k_{\tau|\tau} \sum_{l=L,H} p_{kl} \frac{\omega^l_{\tau+1|t}}{\omega^l_{\tau+1|\tau}} \tag{1.16}$$

where the transition probability p_{kl} are defined in Section 3.1, $\omega^k_{1|0}$ is the starting value, and η^k_τ is the probability density function of the conditional distribution of monthly market returns for regime k:

$$\eta^k_\tau = f(R_{m,\tau}|s_\tau = k, I_{\tau-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^k_m}} exp\{-\frac{(R_{m,\tau} - \mu^k_m)^2}{2(\sigma^k_m)^2}\}.$$

The regime probability $\omega^k_{\tau|t}$ for $\tau \in \{1, 2, \ldots, t\}$ can be found by iterating on equation (1.14) and (1.15) and substituting the obtained regime probability into equation (1.16), given a starting value $\omega^k_{1|0}$. We set the starting value to be the unconditional probability of being in regime k, i.e.

$$\omega^k_{1|0} = \pi_k = \frac{1 - p_{ll}}{2 - p_{ll} - p_{kk}} \tag{1.17}$$

Note that the regime probability $\omega^i_{\tau|t}$ for $\tau \in \{1, 2, \ldots, t\}$ depends on unknown parameters $\boldsymbol{\theta}$. In fact, the above algorithm allows $\boldsymbol{\theta}$ to be estimated by solving the following maximum likelihood estimation problem:

$$\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}) = \sum_{\tau=1}^{t} \log f(R_{m,\tau}|I_{\tau-1}; \boldsymbol{\theta}) \tag{1.18}$$

where

$$f(R_{m,\tau}|I_{\tau-1}; \boldsymbol{\theta}) = \sum_{l=L,H} \omega^l_{\tau|\tau-1}\eta^l_\tau.$$

Table 1.1

Parameter estimates for the two-regime-switching model of the market return

$$R_{m,t} = \mu_m^{s_t} + \sigma_m^{s_t}\epsilon_t,\ \epsilon \sim i.i.d.N(0,1)$$
$$p_{kk} = \Pr(s_{t+1} = k|s_t = k) \text{ for } k \in \{G, B\}$$

μ_m^G	μ_m^B	σ_m^G	σ_m^B	p_{GG}	p_{BB}
0.014	−0.007	0.034	0.070	0.967	0.877
(0.001)	(0.006)	(0.001)	(0.005)	(0.012)	(0.045)

The table reports parameter estimates for the two-regime-switching model of value-weighted market returns for the sample period from January 1940 to December 2017. The regime s_t is generated by an irreducible first-order Markov chain with constant transition probabilities. The parameters are estimated by maximum likelihood estimation. Standard errors are reported in parentheses.

Table 1.2

Descriptive Statistics of Regime-switching Variables for the Pooled Sample: July1963 to December 2017

Variables	Mean	Median	Std.Dev	Quartile 1	Quartile 3	Skewness
$\mu^G(\%)$	2.03	1.94	2.54	1.20	2.89	0.76
$\mu^B(\%)$	-1.72	-1.24	3.83	-3.35	0.22	0.04
$\sigma^G(\%)$	14.57	12.74	8.53	8.78	18.01	3.09
$\sigma^B(\%)$	16.78	15.11	8.88	10.84	20.65	3.24
$\log(\sigma^G)$	-2.06	-2.06	0.52	-2.43	-1.71	0.10
$\log(\sigma^B)$	-1.90	-1.89	0.48	-2.22	-1.58	-0.07
ρ^G	0.09	0.07	0.06	0.05	0.11	1.42
ρ^B	0.19	0.17	0.12	0.11	0.25	0.59

The table reports descriptive statistics of a set of regime variables for individual stocks for the pooled sample from July 1963 to December 2017. The sample includes all NYSE, AMEX, and NASDAQ stocks. The regime variables include the good-regime mean (μ^G), bad-regime mean (μ^B), good-regime volatility (σ^G) and its logarithm, bad-regime volatility (σ^B) and its logarithm, average good-regime cross-correlation (ρ^G), and the average bad-regime cross-correlation (ρ^B),. The regime variables are estimated month by month according to equation (1.3), (1.7), and (1.8). Note that the regime variables for month t are estimated conditional on information up to month t.

Table 1.3
Average Cross-Sectional Correlations between Regime-switching Variables

	μ^G	μ^B	$\log(\sigma^G)$	$\log(\sigma^B)$	ρ^G	ρ^B
μ^G	1.00	0.16	0.40	0.37	0.04	0.10
μ^B		1.00	−0.19	−0.24	−0.18	−0.33
$\log(\sigma^G)$			1.00	0.82	0.13	0.09
$\log(\sigma^B)$				1.00	0.18	0.25
ρ^G					1.00	0.59
ρ^B						1.00

The table reports the time series average of cross-sectional correlations between regime variables for the sample period from July 1963 to December 2017. The sample includes all NYSE, AMEX, and NASDAQ stocks. The regime variables are estimated month by month according to equation (1.3), (1.7), and (1.8). Note that the regime variables for month t are estimated conditional on information up to month t.

Table 1.4

Cross-Sectional Regressions on Regime-switching Variables

Intercept	μ^G	μ^B	$\log(\sigma^G)$	$\log(\sigma^B)$	ρ^G	ρ^B	R^2_{adj}
Panel A: Stock returns							
1.48	0.05	−0.04	−0.34	0.02	0.39	1.36	0.039
(4.15)	(3.68)	(−1.89)	(−2.59)	(0.17)	(0.66)	(3.10)	
Panel B: Residual stock returns							
0.16	0.01	−0.03	−0.23	0.07	0.80	0.96	0.017
(0.46)	(0.73)	(−2.64)	(−2.77)	(0.57)	(1.41)	(2.51)	

The table reports the average estimated coefficients, the associated **?** t-statistics, and the average adjusted R-squared of the cross-sectional regressions for individual stocks. The sample includes all NYSE, AMEX, and NASDAQ stocks. The regressions are estimated month by month from July 1963 to December 2017. The dependent variable is either the stock return in percent for month t (panel A) or the residual stock return in percent for month t (panel B). The residual stock return is the residual obtained from the cross-sectional regression of the stock return for month t on the return for month $t-1$, return over month $t-12$ to $t-2$, log book-to-market ratio for month $t-1$, and log market equity for month $t-1$. The independent variables are regime means, volatility, and average cross-correlations that are estimated conditional on information up to month $t-1$.

Table 1.5
Descriptive Statistics for Portfolios Formed on Predicted Returns

Quintile	Mean	Std. Dev	Avg ME	%ME Share	B/M	Carhart Alpha	FF-5 Alpha
L (Low)	0.58	7.72	2.47	20.7	0.74	−0.41	−0.39
2	0.84	5.83	2.40	20.1	0.79	−0.09	−0.16
3	1.04	5.30	2.30	19.3	0.80	0.06	0.01
4	1.19	5.38	2.40	20.1	0.78	0.27	0.23
H (High)	1.50	6.90	2.35	19.7	0.74	0.69	0.65
H−L	0.91					1.10	1.04
	(3.02)					(3.08)	(3.20)

The table summarizes the descriptive statistics for five value-weighted portfolios. The sample includes all NYSE, AMEX, and NASDAQ stocks. The sample period starts in July 1963 and ends in December 2017. Portfolios for month $t+1$ are formed by sorting stocks into quintile portfolios on the basis of the ranking of predicted returns for month $t+1$. Predicted returns are estimated from the following steps. First, regression residuals, $\{\hat{e}_{i,t}\}$, are obtained from the following cross-sectional regression for month t:

$$R_{i,t} = b_{0t} + b_{1t}R_{i,t-1} + b_{2t}MOM_{i,t-1} + b_{3t}\log(B/M_{i,t-1}) + b_{4t}\log(ME_{i,t-1}) + e_{i,t}.$$

Second, the estimated coefficients, $\{\hat{a}_{jt}\}_{j=1}^{6}$, are obtained from the following cross-sectional regression for month t:

$$\hat{e}_{i,t} = c_{0t} + c_{1t}\mu^{G}_{i|t-1} + c_{2t}\mu^{B}_{i|t-1} + c_{3t}\log(\sigma^{G}_{i|t-1}) + c_{4t}\log(\sigma^{B}_{i|t-1}) + c_{5t}\rho^{G}_{i|t-1} + c_{6t}\rho^{B}_{i|t-1} + v_{i,t}.$$

Third, predicted returns, $\{\hat{R}_{i,t+1}\}$, are constructed as follows,

$$\hat{R}_{i,t+1} = \hat{c}_{2t}\mu^{B}_{i|t} + \hat{c}_{4t}\log(\sigma^{B}_{i|t}) + \hat{c}_{6t}\rho^{B}_{i|t}.$$

Portfolio high (low) consists of stocks with highest (lowest) predicted return. All five portfolios are value-weighted and rebalanced every month. Mean and Std.Dev are the average and the standard deviation of the monthly percentage portfolio return, ME is the average market equity of the stocks in the portfolios in billions of dollars, (%)ME Share is the ratio of the portfolio's market equity to total market equity of all five portfolios in percent, and B/M is is the average book-to-market ratio of the stocks in the portfolios.Carhart four-factor (Carhart) alphas and Fama-French five-factor (FF-5) alphas are in percent. "H−L" refers to the difference in average monthly portfolio returns, Carhart alphas, or FF-5 alphas between portfolio high and portfolio low. Heteroskedasticity-consistent t-statistics are reported in parentheses.

Table 1.6
Carhart four-factor Alphas for Portfolios Formed on Predicted Returns and Control Variables

		Predicted Return Quintile					
		L (Low)	2	3	4	H (High)	H−L
	S (Small)	−0.53	−0.18	0.03	0.2	0.45	0.98
		(−2.60)	(−1.22)	(0.19)	(1.29)	(2.79)	(3.67)
	2	−0.43	−0.33	−0.02	0.1	0.35	0.78
		(−1.89)	(−2.55)	(−0.20)	(0.79)	(2.49)	(2.66)
Size Quintile	3	−0.32	−0.20	−0.10	0.1	0.36	0.68
		(−2.00)	(−1.89)	(−1.43)	(1.61)	(3.03)	(2.77)
	4	−0.24	−0.18	−0.05	0.1	0.35	0.59
		(−1.64)	(−2.11)	(−0.94)	(1.68)	(3.03)	(2.49)
	B (Big)	−0.25	−0.12	−0.10	0.1	0.47	0.73
		(−1.91)	(−1.80)	(−1.83)	(2.08)	(3.97)	(3.29)
Average over size		−0.36	−0.20	−0.05	0.1	0.40	0.75
		(−2.32)	(−2.43)	(−0.95)	(2.06)	(3.66)	(3.23)
Average over prior return		−0.44	−0.20	0.07	0.3	0.59	1.03
		(−2.95)	(−2.03)	(1.02)	(3.19)	(4.28)	(4.20)
Average over momentum		−0.40	−0.25	−0.09	0.1	0.42	0.82
		(−2.59)	(−2.65)	(−1.09)	(1.60)	(3.32)	(3.74)
Average over B/M ratio		−0.39	−0.21	0.03	0.2	0.51	0.90
		(−2.41)	(−1.85)	(0.44)	(2.17)	(3.65)	(3.60)

The table reports FF-3 alphas (in percent) and heteroskedasticity-consistent t-statistics (in parentheses) for portfolios formed on predicted returns and various control variables. The sample includes all NYSE, AMEX, and NASDAQ stocks. The sample period starts in July 1963 and ends in December 2017. Portfolios for month $t+1$ are formed as follows. First, I obtain predict stock returns for month $t+1$ using the following equation:

$$\hat{R}_{i,t+1} = \hat{c}_{2t}\mu^B_{i|t} + \hat{c}_{4t}\log(\sigma^B_{i|t}) + \hat{c}_{6t}\rho^B_{i|t}.$$

Second, I sort stocks into control quintiles on the basis of the ranking of corresponding control variables. The control variable is the market equity measured at the end of month t (size), return for month t (prior return), buy-and-hold return over month $t-11$ to $t-1$ (momentum), or book-to-market ratio (B/M ratio) for month t. Third, I sort stocks into predicted return quintiles based on the ranking of the predicted returns within each control quintile, and obtain a set of 5×5 portfolios. All portfolios are value-weighted and rebalanced every month. “H−L" refers to the difference in Carhart four-factor alphas between portfolio high and portfolio low.

Table 1.7
Cross-Sectional Regressions on Predicted Returns and Control Variables

Regressors (month t regression)	(1)	(2)	(3)
Intercept	1.23	0.43	0.33
	(4.39)	(1.66)	(1.35)
$\hat{R}_t$	0.33	0.34	0.28
	(3.54)	(4.50)	(4.13)
R_{t-1}		−0.04	−0.04
		(−10.7)	(−12.1)
MOM_{t-1}		0.72	0.65
		(5.50)	(5.23)
$\log(B/M_{t-1})$		0.19	0.18
		(5.24)	(5.43)
$\log(ME_{t-1})$		−0.04	−0.03
		(−1.30)	(−1.30)
AG_{t-1}			−0.73
			(−8.91)
ROA_{t-1}			0.57
			(3.30)
GP_{t-1}			0.49
			(5.38)
NOA_{t-1}			−0.12
			(−2.53)
ITA_{t-1}			−0.03
			(−0.34)
OSC_{t-1}			−0.00
			(−0.33)
ACC_{t-1}			−0.29
			(−1.83)
R^2_{adj}	0.018	0.047	0.059

The table reports the average estimated coefficients, the associated **?** t-statistics, and the average adjusted R-squared of the cross-sectional regressions. The sample includes all NYSE, AMEX, and NASDAQ stocks. The sample period is from July 1963 to December 2017. The dependent variable is the stock return for month t (R_t). $\hat{R}_t$ is the predicted return for month t estimated based on equation (1.12). MOM_{t-1} is the buy-and-hold return over month $t-12$ to month $t-2$. log $ME_{i,t-1}$ is the log market equity for month $t-1$. log $B/M_{i,t-1}$ is the log book-to-market ratio for month $t-1$. AG_{t-1} is the asset growth rate for month $t-1$. ROA_{t-1} is the return on assets for month $t-1$. GP_{t-1} is the gross profitability for month $t-1$. NOA_{t-1} is net operating assets for month $t-1$. ITA_{t-1} is the investment-to-assets ratio for month $t-1$. OSC_{t-1} is O-score for month $t-1$. ACC_{t-1} is total accruals for month $t-1$.

CHAPTER 2: Short-Term Reversals and Trading Activity[1]

2.1 Introduction

The short-term reversal effect uncovered by Jegadeesh (1990) and Lehmann (1990) has remained an intriguing puzzle for well over two decades. We show that the likelihood of short-term reversals in monthly stock returns is strongly influenced by prior levels of monthly trading activity. Specifically, the cross-sectional relation between monthly returns and the first lag of monthly returns is highly dependent on prior monthly turnover. Although stocks that have low prior turnover display a pronounced reversal effect, those that have high prior turnover display a continuation effect. In other words, high prior turnover is associated with short-term momentum rather than short-term reversals in monthly stock returns.

We begin with evidence from portfolio sorts using data for July 1963 to December 2018. In particular, we construct a set of 25 value-weighted portfolios by sorting stocks into quintiles using turnover and then sorting the stocks in each turnover quintile into return quintiles. Our interest centers on the performance of long-short portfolios that are formed from the high- and low-prior-return quintiles. To reduce the influence of firms whose economic importance is debatable, the sample used to construct the portfolios excludes stocks whose market equity is below the 20th percentile of the NYSE market equity distribution on a month-by-month basis (the "all-but-micro-cap" sample). For stocks in the bottom quintile of prior turnover, the winners-minus-losers (WML) portfolio has an average return of -0.83% per month, which has a t-statistic of -4.87. In contrast, the WML portfolio for stocks in the top quintile of prior turnover has an average return of 0.58% per month, which has a t-statistic of 2.35. Thus the reversal effect, which is quite strong for low-turnover stocks, is nonexistent for high-turnover stocks.

Evidence that stock returns are related to prior trading activity is not new (see, e.g. Brennan et al., 1998; Datar et al., 1998; Chordia et al., 2001). However, we are aware of only

[1]Coauthored with Ethan Chiang and Chris Kirby

one study that empirically links trading activity to changes in the sign of the short-term autocorrelations of stock returns. Connolly and Stivers (2003) use time-series regressions to investigate the impact of *contemporaneous* turnover on the autocorrelation properties of weekly returns for large-cap stock portfolios. They find that there is substantial positive autocorrelation for weeks in which contemporaneous turnover is abnormally high and substantial negative autocorrelation for weeks in which contemporaneous turnover is abnormally low. Although Connolly and Stivers (2003) also look at the relation between prior turnover and return autocorrelations, they find much smaller effects in this case. This leads them to conclude that "this lag relation seems economically small." In contrast, we find that conditioning on prior turnover produces strong and robust evidence of sign changes in the first-order autocorrelations of monthly returns.

What explains the link between the autocorrelation properties of returns and prior levels of trading activity? Existing research suggests one possibility: the impact of short-term liquidity demands. For instance, Campbell et al. (1993) develop a model in which uniformed trading generates short-term price pressure, thereby producing temporary price concessions that are reversed in the following trading session. Because the model implies that trading volume is positively correlated with the amount of uninformed trading activity, it predicts that the reversal effect should be stronger for stocks with high trading volume than for those with low trading volume. But this prediction, which runs counter to the evidence of a negative interaction between reversals and turnover, finds mixed support in the previous literature.

Using a sample of NASDAQ stocks, for example, Conrad et al. (1994) find that the performance of reversal-based trading strategies improves as prior trading activity increases. However, a subsequent study by Cooper (1999) reports the opposite result for a sample of large-cap NYSE and AMEX stocks. Avramov et al. (2006) build on these findings by studying the relation between short-term reversals and trading activity while simultaneously controlling for the effect of liquidity. They report that cross-sectional differences in liquidity have similar implications for the first-order autocorrelations of both weekly and monthly stock returns: the reversal effect is stronger for less liquid stocks. In addition, Avramov et al. (2006) find that low-turnover stocks display a weaker reversal effect than high-turnover stocks at the weekly horizon, but a stronger reversal effect than high-turnover stocks at the monthly horizon. They conjecture that the latter finding, which is broadly consistent with our results, may indicate that turnover is a poor proxy for uninformed trading activity at the monthly horizon.

Our analysis points to a different explanation for the cross-sectional relation between monthly return reversals and prior trading activity. We start from the premise that a considerable fraction of monthly trading activity is probably motivated by news that changes expectations of future payoffs. This should not be controversial in view of existing models of speculative trading and the price-discovery process. Under the well-known Tauchen and Pitts (1983) model, for example, squared daily stock returns and daily trading volume share a common factor: the rate at which information that alters stock valuations arrives to market. It implies, in other words, that the flow of news drives the dynamics of both volume and volatility.

We therefore conduct a simple test to see whether the role of news in generating turnover might explain our findings. If the interaction between return reversals and prior turnover is linked to the flow of unobserved news, then the Tauchen and Pitts (1983) model predicts that there should be a similar interaction between return reversals and prior volatility. We find that this is indeed the case. The interaction between return reversals and prior volatility is both negative and highly statistically significant. This finding lends indirect support to the hypothesis that turnover acts as proxy for the flow of news that drives speculative trading.

Under this hypothesis, the interaction between short-term reversals and prior turnover has a straightforward interpretation. Consider a stock that falls in the lower tail of the cross-sectional distribution of returns for the month. If the turnover for the stock is low, then it has experienced a below-average return over a period in which the flow of news has been relatively low. The data indicate that this below-average return is likely to be followed by an above-average return over the next month (a short-term reversal effect). Conversely, if the turnover for the stock is high, then it has experienced a below-average return over a period in which the flow of news has been relatively high. The data indicate that this below-average return is likely to be followed by a below-average return over the next month (a short-term momentum effect).

Linking short-term return reversals to prices changes that occur in the absence of much news captures the basic spirit of Campbell et al. (1993) model. More broadly, it is consistent with the implications of the speculative trading model developed by Llorente et al. (2002). The model assumes that there are two basic types of trades: hedging and speculative. Hedging trades convey no signal about future payoffs, so the returns generated by these trades display reversals. Speculative trades, on the other hand, are driven by new information that is only partially incorporated into the stock price in a given trading session. Thus the returns generated by speculative trades display continuations. Because the model implies that the impact of speculative trades is fundamentally different than that of hedging trades, it predicts that cross-sectional differences in the relative importance of speculative trading should lead to differences in the autocorrelation properties of returns across firms.

We use an easily-constructed proxy for the fraction of turnover that is driven by speculative trading to investigate whether our findings are consistent with this prediction. Our approach is motivated by the extensions of the Tauchen and Pitts (1983) model developed by Andersen (1996) and Li and Wu (2006). Both extensions introduce liquidity traders within the general Tauchen and Pitts (1983) framework. Andersen (1996) assumes at the outset that there is no covariance between squared daily returns and the component of daily trading volume that is generated by liquidity trades. Li and Wu (2006) relax this assumption and show that the estimated covariance is *negative* for a range of individual stocks. Both models therefore imply that the monthly estimated correlation between the squared demeaned daily returns and daily turnover should be a useful proxy for the relative contribution of news-driven trades to monthly turnover figures for individual stocks.

Conditioning on this correlation lends further credence to the information-flow hypothesis. For instance, we replicate the portfolio sorts after partitioning the stocks in the all-but-micro-cap sample into two categories on a month-by-month basis: those which have a low fraction of news-driven turnover (estimated correlations below the median value) and those that have a high fraction of news-driven turnover (estimated correlations above the median value). For stocks in the bottom quintile of prior turnover, the WML portfolio has an average return of -1.03% per month for the former category (t-statistic of -6.25) and -0.75% per month for the latter category (t-statistic of -3.84). For stocks in the top quintile of prior turnover, however, the WML portfolio has an average return of -0.35% per month for the former category (t-statistic of -1.32) and 1.06% per month for the latter category (t-statistic of 3.69). Thus the reversal effect is much stronger for stocks with a low fraction of news-driven turnover, which is consistent with the predictions of the Llorente et al. (2002) model. Again, the results are similar for WML portfolios formed from large-cap stocks.

To supplement the evidence produced by the portfolio sorts, we fit a series of cross-sectional regressions that control for other price-related anomalies. In particular, we sort stocks into deciles using monthly turnover, and then regress the returns for the stocks in selected deciles on lagged monthly values of returns, log turnover, log realized volatility, log market equity, and a standard measure of price momentum. For stocks in the bottom decile of prior turnover, the average estimated slope on prior returns is -0.70 with a t-statistic of -10.14. For stocks in the top decile of prior turnover, the average estimated slope on prior returns is 0.13 with a t-statistic of 2.75. Hence, we again find that high prior turnover is associated with short-term momentum rather than short-term reversals in monthly stock returns.

We perform several other tests to assess the robustness of our results. One potential concern with respect to the regression evidence is that our specifications employ a fairly small set of controls. We address this concern by expanding the set of controls to include the book-to-market ratio along with a host of anomaly variables that have been used as mispricing indicators in the recent empirical literature (Stambaugh et al., 2015; Stambaugh and Yuan, 2016). Specifically, we employ variables that capture financial distress, share growth, total accruals, net operating assets, gross profitability, asset growth, return on assets, and investment to assets. Including all of these variables as additional controls in the regressions has only a minor impact on the average estimated slopes on prior monthly stock returns. There are still marked differences in the average estimated slopes across turnover deciles.

The role of microstructure effects in generating short-term reversals is another potential robustness issue. For instance, Ball et al. (1995) and Conrad et al. (1997) report that bid-ask bounce makes a significant spurious contribution to the measured profitability of short-term contrarian strategies using samples that overlap with the early part of our sample period. To assess whether this is a concern in our setting, we replicate our portfolio sorts using the most recent 20 years of data in our sample (January 1999 to December 2018). The average returns for the WML portfolios generally have smaller absolute t-statistics in this case, but this is primarily due to the increase in standard errors associated with the large reduction in the number of monthly return observations. Using the subset of all-but-micro-cap stocks that have a high fraction of news-driven turnover, for example, the WML portfolio for stocks in the top quintile of prior turnover has an average return of 1.28% per month with a t-statistic of 2.00. Given that this is larger than the value of 1.06% per month obtained using data for the whole sample period, it seems unlikely that accounting for bid-ask bounce would undermine our key conclusions.

To gain additional insights, we use the spread-based measure of liquidity proposed by Chung and Zhang (2014) to assess robustness to liquidity effects. We find that linear regressions produce only weak evidence of a relation between liquidity and return reversals. But the portfolio sorts reveal the presence of a marked three-way interaction between liquidity, turnover, and reversals. First, we find that the negative interaction between prior turnover and return reversals is stronger for more liquid stocks. Second, we find that conditioning on the fraction of news driven turnover generates more pronounced changes in the relation between prior turnover and return reversals for more liquid stocks. Third, we find that stocks with high liquidity, high turnover, and a high fraction of news driven turnover display the strongest short-term momentum effect. We conclude, therefore, that our results are not driven by the influence of illiquid stocks.

It is notable that increases in liquidity appear to amplify the impact of conditioning on turnover. The evidence of both short-term reversal and short-term momentum effects for large-cap stocks that have low bid-ask spreads suggests that short-term autocorrelation properties of monthly returns hold substantial economic interest for investors. Furthermore, the strong interaction between the magnitude of reversal effect and the correlation between squared daily returns and daily turnover represents an important new piece of the larger puzzle. Because the nature of the evidence makes it difficult to envision a plausible explanation for this interaction that does not involve some type of information-based story, our findings clearly raise the bar for research that seeks to explain the short-term reversal anomaly.

2.2 Data Sources, Sample Selection, and Variable Descriptions

We obtain daily and monthly stock returns along with a number of related items, such as stock prices, trading volume, shares outstanding, exchange codes, and share codes, from the Center for Research in Security Prices (CRSP). The sample begins in July 1963, ends in December 2018, and is restricted to ordinary common equity (share code 10 or 11) for NYSE, AMEX, and NASDAQ firms. We also use monthly returns for the four Carhart (1997) factors and the monthly risk-free rate, which are from the Ken French data library, along with annual values of various items from the Compustat annual industrial file. These items are used to construct the book-to-market ratio and a range of anomaly variables.

2.2.1 Variable descriptions and notation

Let $R_{i,t}$ and $TURN_{i,t}$ denote the return and turnover of stock i for month t. We use portfolio sorts and cross-sectional regressions to assess whether conditioning on $TURN_{i,t-1}$ conveys useful information about the relation between $R_{i,t-1}$ and $R_{i,t}$. To account for the patterns in average returns associated with other price-based anomalies, we employ the market value of equity, monthly realized volatility, and a standard measure of price momentum as controls. These variables are denoted by $ME_{i,t}$, $VOL_{i,t}$, and $MOM_{i,t}$ for stock i in month t. A couple of other variables also feature prominently in our tests: the estimated monthly correlation of daily turnover with daily squared returns and the average monthly value of the daily percentage bid-ask spread. We use $CORR_{i,t}$ and $SPREAD_{i,t}$ to denote these variables for stock i in month t. Finally, we employ a range of standard anomaly variables as part of our robustness checks. The definitions of these variables follow Stambaugh et al. (2015) with minor exceptions. All of the variables are described in more detail in the appendix.

For tests that involve portfolio sorts, we use both average returns and estimates of risk-adjusted expected returns (alphas) to assess portfolio performance. The risk-adjusted expected returns are estimated using the four-factor model of Carhart (1997). Specifically, we use ordinary least squares (OLS) to fit time-series regressions of the form

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_{p,1}(MKT_t - R_{f,t}) + \beta_{p,2}SMB_t + \beta_{p,3}HML_t + \beta_{p,4}UMD_t + \epsilon_{p,t}, \quad (2.1)$$

where $R_{p,t}$ denotes the return on portfolio p for month t, $R_{f,t}$ denotes the risk-free for month t, and MKT_t, SMB_t, HML_t, and UMD_t denote the returns generated by the market, size, value, and momentum factors for month t.

2.2.2 Sample selection

Previous research suggests that market capitalization is cross-sectionally correlated with a number of firm characteristics that might play a part in determining the short-term autocorrelations of individual stock returns. To forestall any concerns that our findings are driven by firms whose economic importance is debatable, we exclude stocks whose market capitalization falls below the 20th percentile of the NYSE market equity distribution from the sample used to develop our baseline results. Following Fama and French (2008), we call this the all-but-micro-cap sample. We also report results using large-cap stocks, which are defined as having a market capitalization at or above the 50th percentile of the NYSE market equity distribution. Our discussion of the short-term autocorrelation properties of micro-cap stocks is confined to a section that deals with robustness issues.

2.3 Portfolio Sorts and Cross-Sectional Regressions

We begin our investigation of the relation between monthly return reversals and trading activity using univariate portfolio sorts. Specifically, we form two sets of quintile portfolios by sorting stocks on prior monthly returns and prior monthly turnover. All of the portfolios are rebalanced monthly and each stock is weighted in proportion to its market equity at the time the portfolio is formed. Table 2.1 presents a range of descriptive statistics for the portfolios. The statistics in panel A are for the full sample of available NYSE, AMEX, and NASDAQ stocks. Those in panels B and C are for the all-but-micro-cap and large-cap samples that will be used for most of the subsequent analysis.

The results are similar for all three samples. First, the portfolios formed on prior returns show evidence of the monthly reversal anomaly first documented by Jegadeesh (1990). That is, low-prior-return portfolios outperform high-prior-return portfolios. But the differences in average returns between the high and low quintiles are fairly small. Using the sample that includes all stocks, for example, the WML portfolio has an average return of −0.25% per month and an estimated four-factor alpha of −0.32% per month, which is statistically significant at the 10% level (a t-statistic of −1.72").

Second, the portfolios formed on prior turnover show no reliable evidence of differences in performance. The average returns of the high-prior-turnover portfolios exceed those of the low-prior-turnover portfolios by a small margin. However, the estimated alphas of WML portfolios are slightly negative, and all of them are statistically insignificant at the 10% level. As might be anticipated, however, the results indicate that the volatility of the portfolio returns is an increasing function of prior turnover. The volatilities for the high turnover portfolios are almost twice as large as those for the low turnover portfolios for the both the all-but-micro-cap and large-cap samples.

Although we find only weak evidence of monthly return reversals in Table 2.1, this is because the univariate portfolio sorts fail to capture the strong interaction between prior returns and prior turnover. To illustrate this point, we use the all-but-micro-cap sample to form a set of 25 portfolios that reveal how monthly reversal effect varies across turnover quintiles. This is accomplished by sorting stocks into turnover quintiles and then sorting the stocks in each turnover quintile into return quintiles. All of the portfolios are rebalanced monthly, and each stock is weighted in proportion to its market equity at the time the portfolio is formed. Table 2.2 presents the results. Panel A reports the average monthly excess return and estimated four-factor alpha for each portfolio, with heteroskedasticity-robust t-statistics in parentheses. Panel B reports the average monthly turnover and average monthly realized volatility of the constituent stocks across all stock-month observations.

Table 2.2 paints a very different picture than Table 2.1. Stocks contained in the bottom quintile of prior turnover display a strong reversal effect. The average return on the WML portfolio is −0.83% per month with a t-statistic of −4.87, and the estimated four-factor alpha of the portfolio is almost identical: −0.82% per month with a t-statistic of −4.72. Thus the four-factor model explains little, if any, of the monthly reversal effect for low-turnover stocks. The reversal effect diminishes as prior turnover increases. Nonetheless, it remains statistically significant at the 10% level for stocks in the bottom four turnover quintiles. For stocks in the top turnover quintile, however, the average return on the WML portfolio is both *positive* and statistically significant at the 5% level: 0.58% per month with a t-statistic of 2.35. The estimated four-factor of the portfolio is somewhat smaller, but it is still statistically significant at the 10% level.

Table 2.3 shows that we obtain similar results if we restrict the analysis to large-cap stocks. We again find that the stocks contained in the bottom quintile of prior turnover display a strong reversal effect. The WML portfolio has an average return of -0.90% per month with a t-statistic of -5.82 and an estimated four-factor alpha of -0.90% per month with a t-statistic of -5.66. For stocks in the top quintile of prior turnover, however, the WML portfolio has an average return of 0.56% per month with a t-statistic of 2.24 and an estimated four-factor alpha of 0.42% per month with a t-statistic of 1.71. Thus the results of the portfolio sorts are insensitive to the presence of small-cap stocks in the sample.

Overall the evidence is indicative of a pronounced negative interaction between prior turnover and the likelihood of monthly return reversals. But the results in Tables 2.2 and 2.3 do not address the question of whether the nature of the interaction effect has evolved over time. To provide insights in this regard, Figure 2.1 plots the average returns of the WML portfolios for the bottom and top quintiles of prior turnover using a rolling 10-year window of monthly observations.[2] The top panel is for the WML portfolios formed from all-but-micro-cap stocks and the bottom panel is for those formed from large-cap stocks.

The plots in Figure 2.1 have several noteworthy features. First, they highlight the extent to which prior turnover conveys information about the short-term autocorrelations of monthly stock returns. Consider the evidence for WML portfolios formed from all-but-

[2]We compute the rolling average returns for month t by averaging the returns for months $t-59$ to $t+60$. The length of the window is reduced as necessary to account for the lack of observations near the beginning and end of the sample (i.e., for months 1 to 59 and 607 to 666). For example, the average returns for July 1963 are computed using a forward-looking window of 61 observations and those for December 2018 are computed using a backward-looking window of 60 observations. This approach is equivalent to using kernel regression with a uniform kernel to estimate the conditional expected portfolio returns at each point in time. To see why, consider the case in which we specify a fractional time index of the form t/T as the regressor. With a uniform kernel and bandwidth h, the kernel estimator at the point $t = s$ is an equally-weighted average of the returns in the window $s - Th$ to $s + Th$. Because the variance of the regressor converges to $1/12$ as T gets large, the Silverman (1986) rule-of-thumb approach for choosing the optimal bandwidth implies that $h = (1.48/\sqrt{12})T^{-1/5}$. This choice of bandwidth for $T = 666$ corresponds to rolling estimation using 13 years of data. Using this bandwidth yields slightly smoother plots than those in Figure 2.1.

micro-cap stocks. The rolling average return for the high-turnover portfolio exceeds that for the low-turnover portfolio for the entire sample period. The gap between the average returns ranges from a low of 0.30% per month to a high of 3.3% per month. The WML portfolios formed from large-cap stocks also display this property. The gap between the rolling average returns of the WML portfolios ranges from a low of 0.13% per month to a high of 3.3% per month.

Second, the plots do not display any clear time trends. The gap between the average returns for the low- and high-turnover portfolios fluctuates, but not in a systematic way. It narrows in the late 1970s, widens for most of the next two and a half decades, and then narrows again near the end of the sample. The lack of readily identifiable time trends is interesting in view of how the market landscape evolved over the course of our sample period. Some of the notable changes include a large decline in trading costs, price decimalization, a reduction in average trade size, the rise of electronic order execution systems, and the growth in high-frequency trading.

Third, the high-turnover portfolios have positive average returns for most 10-year holding periods. The exceptions occur during the interval between the early 1970s and the early 1980s. The average returns are particularly high from around 1990 to around 2010, and then decline to a much lower level around 2015. Nonetheless, they have largely remained positive in recent years. The plots therefore suggest that the short-term momentum effect for high-turnover stocks has persisted across a wide range of market conditions.

These findings bolster the view that prior turnover has a strong influence on the sign of the first-order autocorrelations of monthly returns. As noted earlier, Connolly and Stivers (2003) also investigate the relation between turnover and autocorrelations, but find that conditioning on prior turnover has a relatively small effect for weekly large-cap portfolio returns. Although the contrast in findings could be due solely to the choice of data frequency, we suspect that it has more to do with methodological differences. Because we sort individual stocks on prior turnover, we avoid having to construct measures of "abnormal turnover" for portfolios formed on market capitalization. The evidence suggests that our approach amplifies the "signal-to-noise" ratio, thereby producing a clearer picture of the impact of conditioning on prior turnover. Consequently, we regard our results as more economically interesting and more reliable with respect to the strength of the interaction effect. Our analysis also leads to a richer set of insights concerning the likely origins of this effect. We turn now to a discussion of this issue.

2.3.1 A working hypothesis

Although several hypotheses have been put forward to explain the short-term reversal anomaly, none of them fit the evidence developed thus far. Overreaction, for example, is the favored behavioral story for the negative average returns produced by WML portfolios (see, e.g., De Bondt and Thaler, 1985). But why would we see overreaction for low-turnover stocks and underreaction for high-turnover stocks? Short-term price pressure that stems from uninformed trading is also a widely-discussed mechanism for generating reversals. But these discussions are usually framed in terms of the Campbell et al. (1993) model. Because the model implies that degree of uninformed trading is positively correlated with the trading volume, it predicts that the reversal effect should be stronger for stocks with high prior trading activity, which is at odds with the evidence from the portfolio sorts.

We hypothesize that the negative interaction between monthly return reversals and prior turnover arises from the role of news in driving speculative trading. To see the origins of this hypothesis, consider the results in panel B of Table 2.3. The observed pattern of volatility for the portfolios mirrors the pattern for turnover, which is consistent with the implications of trading models in which news is a key driver of both volume and volatility. For example, the Tauchen and Pitts (1983) model implies that squared daily returns and daily trading volume share a common factor: the rate at which news that alters stock valuations arrives to market.[3] In view of the implications of such models, it seems reasonable to posit that cross-sectional differences in turnover might proxy for cross-sectional differences in speculative trading activity.

[3] More generally, this common-factor structure follows from the mixture-of-distributions hypothesis (MDH). The MDH posits that the daily return R and daily trading volume V are generated by a bivariate mixture model of the form

$$\begin{aligned} R &= \sigma_R \sqrt{I} Z_R, \\ V &= \mu_V I + \sqrt{I} Z_V, \end{aligned}$$

where I is the daily information flow, Z_R and Z_V are standardized shocks, and I, Z_R and Z_V are mutually independent. Hence, it implies that $R^2 = \sigma_R^2 I + U_R$ and $V = \mu_V I + U_V$, where U_R and U_V are mean-zero innovations with $\mathrm{Cov}(U_R, U_V) = 0$. Andersen (1996) discusses the MDH and its extensions in detail.

2.3.2 Evidence from cross-sectional regressions

If the negative interaction between reversals and prior turnover is due to the impact of speculative trading, then there should be a similar interaction between reversals and prior volatility. We use a cross-sectional regression approach to investigate whether this is the case. Specifically, we fit monthly cross-sectional regressions of the form

$$R_{i,t} = \delta_0 + \delta_1 R_{i,t-1} + \delta_2 \log TURN_{i,t-1} + \delta_3 \log VOL_{i,t-1} + \delta_4 \log ME_{i,t-1} + \delta_5 MOM_{i,t-1} + u_{i,t} \tag{2.2}$$

for selected decile groupings of stocks that are formed by conducting month-by-month sorts on either prior turnover or prior realized volatility. This strategy has several noteworthy features. First, we are essentially fitting varying-coefficient models because the coefficient estimates are local to the turnover or volatility neighborhood defined by the decile groupings. Second, we employ log transformations of turnover, realized volatility, and market equity because these variables have highly skewed distributions with long right tails. Third, we use price-based anomaly variables as our principal controls because we view these variables are prime candidates for capturing short-term autocorrelations in individual stock returns. Models that include a much more extensive set of anomaly-based covariates are considered as part of the robustness checks.

Fitting a separate linear regression to each decile of stocks is designed to capture nonlinearity without having to specify a nonlinear model. The basic idea is to estimate the marginal effect of each regressor on expected stock returns while holding either turnover (or volatility) *approximately* constant. An alternative strategy would be to fit global cross-sectional regressions that include an appropriate set of pairwise interactions between the main regressors. This approach leads to similar conclusions about the relation between prior turnover and short-term reversals. However, the interpretation of the coefficient estimates is less straightforward.

Table 2.4 summarizes the regression results. To aid in interpreting the estimates, we demean each of the regressors on a month-by-month basis using its cross-sectional mean for the specific decile under consideration, and then divide each demeaned variable by its monthly cross-sectional standard deviation across all deciles. With this method of standardization, the average estimated intercept for a given decile is the average monthly return on an equally-weighted portfolio of the constituent stocks, and a one unit change in a regressor for a given decile is directly comparable to a one unit change in the regressor for any other another decile (i.e., the estimates are expressed in the same units for every decile).

Panel A reports the average coefficient estimates for turnover deciles 1, 4, 7, and 10. As anticipated, the average estimated slopes on prior returns are indicative of a strong negative interaction between monthly return reversals and prior turnover. Using all-but-micro-cap stocks, for example, the regressions for decile 1 produce an average estimated slope of -0.70 with a t-statistic of -10.14. In contrast, the average estimated slope for decile 10 is 0.13 with a t-statistic of 2.75. Thus the pattern uncovered via portfolio sorts is still evident after controlling for the explanatory power of volatility, market equity, and momentum. Specifically, stocks with low prior turnover display strong short-term reversals and those with high prior turnover display short-term momentum.

The results for large-cap stocks are similar. The average estimated slope for stocks in the bottom decile of prior turnover is -0.55 with a t-statistic of -7.36, whereas that for stocks in the top decile of prior turnover is 0.06 with a t-statistic of 1.10. Although the latter estimate is not statistically significant, it seems likely that it understates the strength of the short-term momentum effect. We say this because interactions are inherently nonlinear. The regressions are designed to capture nonlinearity via local OLS fits. But our local estimation strategy is not optimized to deliver the best tradeoff between bias and variance.

We also see evidence of the low-volatility anomaly in the results. Using all-but-micro-cap stocks, for example, the average estimated slope on log volatility is negative and statistically significant at the 5% level in every case. In contrast, most of the average estimated slopes for log turnover are statistically insignificant at the 5% level. The only exception is for decile 10, which is also the decile for which log volatility has the strongest effect. For the stocks in decile 10, increases in log volatility and log turnover are associated with statistically significant decreases in average returns. Thus, for stocks that display high levels of turnover, changes in turnover and volatility convey similar signals about expected stock returns.

But does this result hold more broadly? Panel B reports the average coefficient estimates for volatility deciles 1, 4, 7, and 10. As predicted by our information-flow hypothesis, the general pattern of the average estimated slopes on prior returns is similar to that in panel A. Using all-but-micro-cap stocks, for example, the regressions for decile 1 produce an average estimated slope of −0.55 with a t-statistic of −6.62, whereas those for decile 10 produce an average estimated slope of 0.01 with a t-statistic of 0.27. Stocks with low prior volatility display strong short-term reversals. But there is no evidence of a reversal effect for those with high prior volatility.

Interestingly, all but two of the average estimated slopes on log turnover in panel B are positive and statistically significant at the 10% level. At first glance this finding may appear to be at odds with the results reported in panel A. But this is not the case. Note in particular that the average estimated intercept in panel A increases across deciles 1, 4, and 7, which indicates that large increases in turnover are associated with increases in average stock returns. It is only at the highest level of turnover than we see a drop in average returns.

This pattern of estimates points to a nonlinear and potentially non-monotonic relation between prior turnover and expected stock returns. Similarly, the average estimated intercept in panel B increases across deciles 1, 4, and 7, but falls for decile 10. The average estimated slope on log volatility also changes sign as we move across deciles. It is positive and statistically significant for decile 1, but negative and statistically significant for decile 10. These findings point to a nonlinear and non-monotonic relation between volatility and expected stock returns.

Overall the estimates support the hypothesis that volatility and turnover share a common factor that helps to explain differences in the first-order autocorrelations of monthly stock returns across firms. If the factor is news that drives speculative trading, then this casts the relation between monthly return reversals and prior turnover in a new light. Consider a stock falls in the lower tail of the cross-sectional distribution of returns. If it has low turnover, then it has experienced a below-average return over a period in which the flow of news has been relatively low. The data indicate that this below-average return is likely to be followed by an above-average return over the next month (a short-term reversal effect). Conversely, if the stock has high turnover, then it has experienced a below-average return over a period in which the flow of news has been relatively high. The data indicate that this below-average return is likely to be followed by a below-average return over the next month (a short-term momentum effect).

The foregoing pattern of effects is broadly consistent with the model of Llorente et al. (2002), which provides a theoretical basis for linking short-term reversals to prices changes that occur in the absence of news-driven speculative trading. Under the model, price changes are generated in three different ways. The first is through the arrival of public news about future payoffs, the second is through hedging trades that are conducted for non-informational reasons, and the third is through speculative trades that are motivated by private information about future payoffs. The price changes due to public news are completely unpredictable. But those due to hedging and speculative trades generate serial correlation in returns.

Hedging trades convey no information about future payoffs, and thus returns generated by hedging trades display reversal effects. Speculative trades, on the other hand, are driven by information that is only partially incorporated into the stock price in a given period (i.e., the equilibrium is less than fully revealing), and thus returns generated by speculative trades display continuation effects. Because the effects of hedging and speculative trades work in opposite directions, the model implies that the relative importance of speculative trading activity should be a key determinant of the short-term autocorrelation properties of returns.

Importantly, the model also implies that the likelihood of return reversals is determined by the interaction between prior returns and prior trading activity. This follows from the relation

$$\mathrm{E}(R_t|\tilde{V}_{t-1}, R_{t-1}) \approx -(\theta_1 + \theta_2 \tilde{V}_{t-1}^2)R_{t-1}, \tag{2.3}$$

where R_t is the dollar return on the risky asset for period t and $\tilde{V}_t$ is the trading volume for period t divided by its unconditional mean (see equation (9) of Llorente et al., 2002). The assumptions employed by Llorente et al. (2002) imply that $\theta_1 \geq 0$ and $\theta_2 \geq 0$, with θ_2 declining as informational asymmetry increases. As they emphasize, however, θ_2 can become negative if there is long-lived private information and the degree of informational asymmetry is sufficiently high (see, e.g., the related model of Wang, 1994). To quote the authors, "when speculative trades are more important, current returns together with high volume predict weaker reversals (or even continuation) in future returns." Hence the model can generate short-term momentum if speculative trades account for a sufficiently large fraction overall trades.

But it should be noted that Llorente et al. (2002) adopt a particularly simple information structure. Specifically, they consider a single stock and assume that its future dividend is the sum of a component that is known to all traders and a component that is known only to a subset of traders. A key implication of this assumption is that all traders agree on how public news affects expected future payoffs. If traders have differences of opinion in this regard, as in the Tauchen and Pitts (1983) model, then public news spurs speculative trading as well. Although Llorente et al. (2002) abstract from this type of setting, it seems reasonable to posit that the interplay between the competing effects of hedging and speculative trades extends to settings with more complex information environments. We therefore focus on the key prediction of the model, which is that cross-sectional differences in the short-term autocorrelations of returns are related to differences in the relative importance of speculative trading.

Testing this prediction requires a measure that captures cross-sectional differences in speculative trading activity. We base our tests on a proxy for the fraction of news-driven turnover that is motivated by the extensions of the Tauchen and Pitts (1983) model developed by Andersen (1996) and Li and Wu (2006). Both extensions introduce liquidity traders within the general Tauchen and Pitts (1983) framework. Andersen (1996) assumes at the outset that there is no covariance between squared daily returns and the component of daily trading volume that is generated by liquidity trades. Li and Wu (2006) relax this assumption and show that the estimated covariance is *negative* for a range of individual stocks. Thus both models imply that the monthly estimated correlation between squared daily returns and daily turnover should be a useful proxy for the relative contribution of news-driven trades to monthly turnover for individual stocks.

2.3.3 A Closer Look at the Relation Between Reversals and Turnover

We begin by considering a simple extension of cross-sectional regressions considered in panel A of Table 2.4. First, we partition the set of available stocks for each month into two groups:

those which have a low fraction of news-driven turnover (estimated correlations below the median value for the month) and those that have a high fraction of news-driven turnover (estimated correlations above the median value for the month). Second, we sort the stocks in each group into turnover deciles and examine the regression evidence. The motivation for this approach is straightforward. Under our information-flow hypothesis, stocks with a low fraction news-driven turnover should display a more pronounced reversal effect than those with a high fraction news-driven turnover. Table 2.5 summarizes the regression results.

Panel A reports the results obtained using all-but-micro-cap stocks. First consider the evidence for stocks that have a low fraction of news-driven turnover. The average estimated slopes on prior returns for deciles 1, 4, and 7 are −0.58, −0.68, and −0.59 with t-statistics of −8.49, −9.38, and −8.56. Thus we see little indication that the reversal effect weakens with increasing turnover from these results. The regressions for decile 10, however, produce an average estimated slope of −0.06 with a t-statistic of −0.98, which suggests that the estimated reversal effect is substantially weaker for stocks in the highest turnover category. Still, we see no evidence of a short-term momentum effect.

Now consider the evidence for stocks that have a high fraction of news-driven turnover. The average estimated slopes on prior returns for deciles 1, 4, and 7 are −0.66, −0.48, and −0.22 with t-statistics of −6.94, −5.52, and −2.54. Hence, they point to a weakening of the reversal effect as turnover increases. More notably, the regressions for decile 10 produce an average estimated slope of 0.22 with a t-statistic of 3.60, indicating that high-turnover display a statistically-significant momentum effect. Thus the results for all-but-micro-cap stocks line up quite well with the predictions of the Llorente et al. (2002) model.

Panel B reports the results obtained using large-cap stocks. In general, they display the same basic patterns as those in panel A. The reversal effect weakens with increasing turnover. However, the changes in the average estimated slope on prior returns are more pronounced for stocks that have a high fraction of news-driven turnover. For example, the regressions for decile 10 produce an average estimated slope of −0.09 with a t-statistic of −1.45 for stocks that have a low fraction of news-driven turnover, and 0.11 with a t-statistic of 1.64 for stocks that have a high fraction of news-driven turnover. So we again conclude that the evidence is in line with the predictions of the Llorente et al. (2002) model.

Table 2.6 shows how conditioning on the fraction of news-driven turnover affects the results of the portfolio sorts. Panel A is for portfolios formed from all-but-micro-cap stocks. For stocks in the bottom quintile of prior turnover, the WML portfolio has an average return of -1.03% per month (a t-statistic of -6.25) for stocks with a low fraction of news-driven turnover and -0.75% per month (a t-statistic of -3.84) for stocks with a high fraction of news-driven turnover. Thus the reversal effect appears to be somewhat weaker in the latter case. For stocks in the top quintile of prior turnover, the WML portfolio has an average return of -0.35% per month (a t-statistic of -1.32) for stocks with a low fraction of news-driven turnover and 1.06% per month (a t-statistic of 3.69) for stocks with a high fraction of news-driven turnover.

The key takeaway from these results is that conditioning on our proxy for the fraction of news-driven turnover clearly matters. Indeed, the results point to a stronger conditioning effect than those of the regressions. This may be because portfolio sorts are fully non-parametric. All of the evidence thus far is indicative of a strong interaction between prior turnover and monthly return reversals. Although our regressions are designed to highlight this interaction, they may not capture it to the same extent as the portfolio sorts.

Panel B shows that repeating the portfolio sorts for large-cap stocks produces almost identical findings. For stocks in the top quintile of prior turnover, for instance, the WML portfolio has an average return of -0.35% per month (a t-statistic of -1.34) for stocks with a low fraction of news-driven turnover and 1.06% per month (a t-statistic of 3.72) for stocks with a high fraction of news-driven turnover. Of course it is important to bear in mind that the portfolio sorts do not control for potential confounding anomalies. Subject to this caveat, however, the results are fully consistent with our information flow hypothesis.

Figure 2.2 shows how conditioning on the fraction of new-driven turnover affects the rolling 10-year average returns on the WML portfolios formed from all-but-micro-cap stocks. The top and bottom panels are for the low and high categories, respectively. In general, the portfolios formed from stocks with a low fraction of news driven turnover have negative average returns, regardless of the level of prior turnover. We see little evidence of short-term momentum for this category of stocks. The average return on the high-prior-turnover portfolio turns positive around 2010 and stays positive through the end of 2018, but it never rises very far above zero.

In contrast, the high-prior-turnover portfolio formed from stocks with a high fraction of news driven turnover has a positive average return for most of the sample period. The value is as high as 3.3% per month during the mid-1990s. So short-term momentum is readily evident for this category of stocks. In addition, the gap between average returns for the low- and high-prior-turnover portfolios is quite wide for much of the sample period, reaching a maximum value of 4.4% per month. The sharp contrast between the plots in the top and bottom panels is consistent with the presence of a strong interaction between the level of turnover, the fraction of turnover driven by news, and monthly return reversals.

As noted earlier, the average return on the high-prior-turnover portfolio increase towards the end of the sample period for stocks with a low fraction of news-driven turnover. Although this is suggestive of some weakening in the reversal effect in the last decade, the evidence is far from definitive. In the bottom panel, for example, the gap between average returns for the low- and high-prior-turnover portfolios widens over the last few years of the sample period. On the whole we can discern little in the way of readily-identifiable time trends.

Figure 2.3 replicates the plots in Figure 2.2 using large-cap stocks. Excluding small-cap stocks from the WML portfolios produces only minor changes in the results. We see the same general patterns as in Figure 2.2. Specifically, the portfolios formed from stocks that have a low fraction of news driven turnover generally have negative average returns, regardless of the level of prior turnover, and the high-prior-turnover portfolio formed from stocks that have a high fraction of news driven turnover has a positive average return for most of the sample period. In short, the rolling average returns for the portfolios formed from large-cap stocks are similar to those of the portfolios that include both small- and large-cap stocks.

2.3.4 Alternative explanations for short-term momentum

The short-term momentum effect for stocks with high prior turnover is one of the most intriguing aspects of our findings. Recall that the Llorente et al. (2002) model implies that short-term momentum arises from the reaction of rational traders to the arrival of new information that alters conditional expectations of future payoffs. The evidence that stocks with a high fraction of news-driven turnover display the strongest momentum is consistent with this feature of the model. But it is important to point out that there may be other mechanisms for generating momentum that could give rise to the same result.

Underreaction to news is one possibility that comes to mind. Not only does an underreaction story have the potential to explain why short-term momentum is concentrated in stocks that have a high fraction of news-driven turnover, it also meshes fairly well with our other findings. Suppose, for example, that speculative traders underreact to news and uninformed traders generate short-term price pressure that leads to price concessions. Under these circumstances, the reversal effect should dominate for stocks with low information flow. But this effect should weaken as we move to stocks with higher information flow, especially those that have a high fraction of news-driven turnover. Thus the net result might be a strong negative interaction between turnover and short-term reversals in conjunction with a short-term momentum effect for stocks that display a sufficiently high level of turnover.

Regardless of how short-term momentum arises, however, it is apparent that our findings raise the bar for research that seeks to explain the short-term autocorrelation properties of monthly stock returns. Any proposed explanation for the short-term reversal anomaly must contend with both the negative interaction between prior turnover and monthly return reversals and the evidence that the prior monthly correlation between squared daily returns and daily turnover conveys substantial information about the likelihood of monthly return reversals. Because it is difficult to envision a plausible explanation for the latter finding that does not entail some type of information-based story, it may be worthwhile to revisit at least some of the empirical results reported by previous studies in the short-term reversal literature.

2.4 Robustness Checks and Further Analysis

The short-term reversal anomaly has attracted its fair share of attention since it was first uncovered by the Jegadeesh (1990) and Lehmann (1990). In a number of cases, prior research identifies methodological or robustness issues that could potentially play a role in our findings. We therefore investigate whether any of the issues or concerns that appear to be most relevant to our analysis have a meaningful impact on our main results.

2.4.1 Regressions using an expanded set of controls

Our baseline regressions control for several price-based anomalies that have been widely studied in the literature. We focus on priced-based anomalies because they seem most likely to be associated with short-term reversals. But there is always a possibility that using a broader set of controls would materially alter our findings. To address this issue, we expand the set of covariates to include the book-to-market ratio along with a wide range of anomaly variables that are used as mispricing indicators in the recent literature (Stambaugh et al., 2015; Stambaugh and Yuan, 2016). The anomaly variable include measures of financial distress, share growth, accruals, net operating assets, gross profitability, asset growth, return on assets, and investment to assets.[4] Table 2.7 summarizes the results of this robustness check.

Expanding the set of controls has only a minor impact on our findings. Using all-but-micro-cap stocks, for example, the average estimated slopes for prior monthly returns are -0.71, -0.64, -0.42, and 0.05 with t-statistics of -7.49,-8.03,-5.78, and 0.88. Thus there are still marked differences in the estimated slopes across turnover deciles. The most notable change from the results in Table 2.4 is that the average estimated slope for the top turnover decile is no longer statistically significant. This is due to both a decrease in the magnitude of the estimate and an increase in its standard error, which suggests that the controls may explain part of the short-term momentum effect for high turnover stocks. But the evidence is not conclusive.

For instance, the regression for the top turnover decile of large-cap stocks produces an average estimated slope on prior returns of 0.05, which is very close to the value of 0.06 reported in Table 2.4. Thus it is unclear whether the decrease in the average estimated slope for all-but-micro-cap stocks is due to a weaker short-term momentum, or whether it

[4]Specifically, we use the estimated probability of bankruptcy from Ohlson (1980), the annual growth rate of the split-adjusted shares outstanding, the annual change in non-cash working capital minus depreciation expense (as a fraction of average total assets for the year), operating assets minus operating liabilities (as a fraction of beginning-of-year total assets), revenues minus cost of goods sold (as a fraction of end-of-year total assets), the annual growth rate of total assets, the ratio of annual earnings to beginning-of-year total assets, and the annual change in gross property, plant, and equipment plus the annual change in inventories (as a fraction of beginning-of-year total assets). Our definitions of these variables match those of Stambaugh et al. (2015) with one minor exception: the return on assets is computed using annual rather than quarterly data. See the appendix for details.

is largely due to the change in sample composition that results from requiring firms to have non-missing values of all the regressors to be included in the regressions. Even if the former effect predominates, however, the basic message of Table 2.7 is consistent with that of Table 2.4. That is, high levels of prior turnover are associated with a dramatic weakening of the short-term reversal effect.

2.4.2 Liquidity effects

Liquidity is another factor that could potentially play a confounding role in our findings. For example, Avramov et al. (2006) report that return reversals are more pronounced for less liquid stocks at both the weekly and monthly horizon. They conduct their analysis using the price-impact criterion of Amihud (2002), which is one of the most widely used proxies for liquidity in empirical research. However, the recent study of Lou and Shu (2017) reports that the cross-sectional explanatory power of the Amihud (2002) criterion for individual stock returns is almost entirely attributable to its dependence on dollar trading volume. This makes the criterion ill suited to our purposes because dollar trading volume is very highly correlated with turnover for most stocks.

Following the recommendations of Fong et al. (2017), we use the monthly average of the daily proportional bid-ask spread to investigate liquidity effects. Its value for stock i in month t is denoted by $SPREAD_{i,t}$. Fong et al. (2017) report that, in general, $SPREAD_{i,t}$ is the most accurate low-frequency proxy for liquidity. This is consistent with the results of Chung and Zhang (2014), who use data for 1993 to 2007 to investigate the performance of a wide range of different liquidity proxies (including the Amihud (2002) criterion) and find that $SPREAD_{i,t}$ has the highest correlation with the daily time-weighted average quoted spread from the Trade and Quote files of the NYSE.

Table 2.8 shows how conditioning on liquidity affects the results of the portfolio sorts. We form the portfolios in two steps. First, we partition the set of available stocks into two categories for each month t: those with low liquidity ($SRREAD_{i,t}$ above the median value for month t) and those with high liquidity ($SPREAD_{i,t}$ below the median value for month t). Second, we construct low- and high-liquidity versions of the portfolios considered in Table 2.2 for each category of stocks. Due to the limited availability of bid and ask prices in the CRSP daily stock file, the sample period is January 1993 to December 2018.

Panel A reports the results for all-but-micro-cap stocks. To assess the overall effect of liquidity, we look at the mean values of the average portfolio returns across turnover quintiles. The values for the low- and high prior-return portfolios are 0.82 and 0.28 for low-liquidity stocks and 0.62 and 0.39 for high-liquidity stocks. So these statistics suggest that the reversal effect becomes weaker as liquidity increases. But we also find evidence of three-way interaction between liquidity, turnover, and reversals. The WML portfolios formed from low- and high-prior-turnover stocks with high liquidity have average returns of −1.25 and 1.18 with t-statistics of −4.55 and 2.64, whereas those formed for stocks with low liquidity have average returns of −0.86 and 0.06 with t-statistics of −3.23 and 0.13. Thus prior turnover becomes more informative about the likelihood of short-term reversals as liquidity increases.

Using large-cap stocks for the portfolio sorts yields similar results. The WML portfolios formed from low- and high-prior-turnover stocks with high liquidity have average returns of −0.93 and 1.09 with t-statistics of −3.61 and 2.24, whereas the values for stocks with low liquidity are −0.78 and 0.41 with t-statistics of −2.85 and 0.90. Hence, the tendency for high-turnover stocks to display short-term momentum increases with liquidity, and the effect of conditioning on prior turnover becomes more pronounced as liquidity increases.

Table 2.9 uses cross-sectional regressions to provide additional evidence on liquidity effects. The results in panel A, which are for selected turnover deciles that are formed from either low-liquidity or high-liquidity stocks, are consistent with the evidence from the portfolio sorts. The average estimated slope on prior monthly returns displays marked differences across turnover deciles for both low- and high-liquidity stocks. Specifically, it ranges from −0.65 for decile one (t-statistic of −4.15) to 0.01 for decile ten (t-statistics of −0.15) in the case of low-liquidity stocks, and from −0.69 for decile one (t-statistic of −4.84) to 0.16 for decile 10 (t-statistics of 1.94) in the case of high-liquidity stocks. The short-term reversal effect weakens as turnover increases, but the interaction with turnover is stronger for high-liquidity stocks. For these stocks we see a statistically-significant short-term momentum effect for decile ten.

The results in panel B are for selected liquidity deciles that are formed from either low-turnover or high-turnover stocks. The estimates suggest that the overall relation between liquidity and short-term reversals is relatively weak, which is again consistent with the evidence from the portfolio sorts. The average estimated slope on prior monthly returns ranges from -0.45 to -0.25 for low-turnover stocks and from -0.15 to 0.02 for high-turnover stocks. At first glance, therefore, cross-sectional differences in liquidity do not appear to be very informative about the likelihood of subsequent return reversals.

To complete our analysis of liquidity effects, we investigate whether conditioning on liquidity alters our conclusions regarding the relation between return reversals and the prior monthly correlation between squared daily returns and daily turnover. Specifically, we replicate the portfolio sorts described in Table 2.6 for the low- and high-liquidity categories of all-but-micro-cap stocks. The results are summarized in Table 2.10.

Panel A shows that low-liquidity stocks with a low fraction of news-driven turnover display a reversal effect for every quintile of prior turnover, although it is not statistically significant for the top quintile. The average returns for the WML portfolios range from -1.55% to -0.52% per month. In comparison, the range is -0.50% to 0.27% per month for low-liquidity stocks with a high fraction of news-driven turnover. This points to a substantial weakening of the reversal effect as the fraction of news-driven turnover increases.

Panel B shows that high-liquidity stocks also display ample evidence of return reversals. But the reversal effect is more pronounced among the subset of stocks that have a low fraction of news-driven turnover. For stocks with a high fraction of news-driven turnover, the WML portfolio for the top prior turnover quintile has an average return of 1.63% per month with a t-statistic of 3.11. This is considerably higher than the average return on the corresponding portfolio formed from low-liquidity stocks. Hence, the evidence indicates that stocks with high liquidity, high turnover, and a high fraction of news driven turnover display a stronger short-term momentum effect than stocks with low liquidity, high turnover, and a high fraction of news driven turnover.

These findings highlight the importance of accounting for nonlinear phenomena. Although the regression estimates in panel B of Table 2.9 provide only weak evidence of a linear relation between liquidity and return reversals, the average portfolio returns in Table 2.10 point to strong nonlinear effects in the data. In particular, they are indicative of a marked three-way interaction between liquidity, turnover, and reversals. Because the effect of conditioning on the fraction of new-driven turnover becomes more pronounced as liquidity increases, it is apparent that our conclusions are not driven by low liquidity stocks.

2.4.3 Portfolio sorts for a low-trading-cost sample period

Several studies raise concerns about the role of bid-ask spreads in generating short-term reversals. For instance, Ball et al. (1995) and Conrad et al. (1997) report that bid-ask bounce makes a significant spurious contribution to the measured profitability of short-term contrarian trading strategies. If bid-ask bounce influences our results, then research on trading costs suggests that its impact is likely to be of most concern for the early part of our sample period. Jones (2002), for example, estimates that quoted proportional spreads for Dow Jones stocks declined from around 0.60% in the 1980s to around 0.20% at the end of the 20th century. To see if the relation between reversals and prior turnover during the early part of the sample period drives our findings, we replicate the portfolio sorts described in Table 2.6 using the most recent 20 years of data. The results, which are shown in Table 2.11, are consistent with those for the full sample period.

Using all-but-micro-cap stocks, for example, the WML portfolio for stocks in the bottom quintile of prior turnover has an average return of −0.91% per month (a t-statistic of −2.76) for stocks with a low fraction of news-driven turnover and −0.74% per month (a t-statistic of −1.89) for stocks with a high fraction of news-driven turnover. In comparison, the WML portfolio for stocks in the top quintile of prior turnover has an average return of 0.11% per month (a t-statistic of 0.20) for stocks with a low fraction of news-driven turnover and 1.28% per month (a t-statistic of 2.00) for stocks with a high fraction of news-driven turnover.

The t-statistics in Table 2.11 are generally smaller in magnitude than those in Table 2.6. However, this is mainly due to the reduction in the number of observations. The shorter sample period is roughly one-third as long as the full sample period. Despite the relatively large standard errors of the average returns in Table 2.11, we still find that the portfolios formed from stocks with a high-fraction of news driven turnover produce clear evidence of a strong interaction between prior turnover and return reversals. It seems unlikely, therefore, that the sensitivity of measured WML returns to bid-ask bounce is more than a minor concern.

2.4.4 Short-term reversals among micro-cap stocks

The evidence indicates that sensitivity to liquidity effects is not a concern in our setting. But they might be a key robustness issue for studies that include micro-cap stocks in the analysis. These stocks, which have been excluded from consideration thus far, account for almost 60% of the available firm-year observations for our 1963-to-2018 sample period. In general, we would expect micro-cap stocks to be much less liquid than small- and large-cap stocks, making them prime candidates for strong return reversals. This may be one of the reasons that the short-term momentum effect for small- and large-cap stocks has gone undetected in prior research. If micro-cap stocks display strong return reversals, then including them in the analysis could mask the evidence of short-term momentum for high-turnover stocks.

Consider the results in Table 2.12, which uses portfolio sorts to illustrate the relation between prior turnover and short-term reversals for micro-cap stocks. As anticipated, the reversal effect is particularly pronounced for these stocks, both in terms of magnitude and statistical significance. The WML portfolios for the bottom four quintiles of prior turnover have average returns that range from -1.48% to -1.84% per month, and the smallest unsigned t-statistic is 8.01. The reversal effect is weaker for micro-cap stocks in the top quintile of prior turnover, but there is no evidence whatsoever of short-term momentum. The WML portfolio has an average return of -0.68 with a t-statistic of -2.50.

Because micro-cap stocks are of questionable importance from a capital markets perspective, the presence of such a strong short-term reversal effect among these stocks would raise robustness concerns in many applications. However, this is clearly not the case with respect to the documented interaction between prior turnover and return reversals. Not only does the interaction appear to be more pronounced for small- and large-cap stocks than for micro-cap stocks, it is also more pronounced for high-liquidity stocks than for low-liquidity stocks. The evidence in this regard should allay any concerns about the economic significance of our findings.

2.5 Conclusions

The likelihood of short-term return reversals is strongly linked to prior trading activity for both small- and large-cap stocks. Stocks with low prior turnover have the strongest tendency to display monthly return reversals. Those with high prior turnover display short-term momentum. We posit that these findings arise from the interplay between short-term price pressure generated by uninformed traders and short-term continuations generated by the actions of speculative traders. By conditioning on a proxy for the fraction of turnover that is driven by news, we show that the predictions of our hypothesis are consistent with the observed negative interaction between prior turnover and reversals in monthly stock returns.

The pronounced reversal effect in monthly returns for liquid, low-turnover, large-cap stocks is an intriguing phenomenon that belies the view that short-term reversals are of little economic significance. In addition, the link between the strength of reversal effect and the correlation between squared daily returns and daily turnover is a telling finding. Because the nature of the evidence makes it difficult to envision a plausible explanation for this interaction that does not involve some type of information-based story, our findings clearly raise the bar for research that seeks to explain the short-term reversal anomaly.

More broadly, the evidence suggests that market capitalization probably plays an important confounding role in research on short-term reversals. Our analysis reveals that the reversal effect is particularly strong for micro-cap stocks, which account for almost 60% of the available firm-year observations for our sample period. Hence, the presence of these stocks in the dataset tends to mask evidence of short-term momentum. This may be one of the reasons why the short-term momentum effect for high-turnover stocks has gone undetected in prior research. Without adequate controls for the influence of market capitalization, studies of the short-term reversal anomaly are likely to produce incomplete and potentially misleading findings with respect to the autocorrelation properties of individual stock returns.

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, 51(1):169–204.

Avramov, D., Chordia, T., and Goyal, A. (2006). Liquidity and autocorrelations in individual stock returns. *The Journal of Finance*, 61(5):2365–2394.

Ball, R., Kothari, S., and Wasley, C. E. (1995). Can we implement research on stock trading rules? *Journal of Portfolio Management*, 21(2):54–63.

Brennan, M. J., Chordia, T., and Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics*, 49(3):345–373.

Campbell, J. Y., Grossman, S. J., and Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, 108(4):905–939.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.

Chordia, T., Subrahmanyam, A., and Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, 59(1):3–32.

Chung, K. H. and Zhang, H. (2014). A simple approximation of intraday spreads using daily data. *Journal of Financial Markets*, 17:94–120.

Connolly, R. and Stivers, C. (2003). Momentum and reversals in equity-index returns during periods of abnormal turnover and return dispersion. *The Journal of Finance*, 58(4):1521–1556.

Conrad, J., Gultekin, M. N., and Kaul, G. (1997). Profitability of short-term contrarian strategies: Implications for market efficiency. *Journal of Business & Economic Statistics*, 15(3):379–386.

Conrad, J. S., Hameed, A., and Niden, C. (1994). Volume and autocovariances in short-horizon individual security returns. *The Journal of Finance*, 49(4):1305–1329.

Cooper, M. (1999). Filter rules based on price and volume in individual security overreaction. *The Review of Financial Studies*, 12(4):901–935.

Datar, V. T., Naik, N. Y., and Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, 1(2):203–219.

De Bondt, W. F. and Thaler, R. (1985). Does the stock market overreact? *The Journal of finance*, 40(3):793–805.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2008). Average returns, b/m, and share issues. *The Journal of Finance*, 63(6):2971–2995.

Fong, K. Y., Holden, C. W., and Trzcinka, C. A. (2017). What are the best liquidity proxies for global research? *Review of Finance*, 21(4):1355–1401.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance*, 45(3):881–898.

Jones, C. M. (2002). A century of stock market liquidity and trading costs. *Working Paper*.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, 105(1):1–28.

Li, J. and Wu, C. (2006). Daily return volatility, bid-ask spreads, and information flow: Analyzing the information content of volume. *The Journal of Business*, 79(5):2697–2739.

Llorente, G., Michaely, R., Saar, G., and Wang, J. (2002). Dynamic volume-return relation of individual stocks. *The Review of Financial Studies*, 15(4):1005–1047.

Lou, X. and Shu, T. (2017). Price impact or trading volume: Why is the amihud (2002) measure priced? *The Review of Financial Studies*, 30(12):4481–4520.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, pages 109–131.

Silverman, B. W. (1986). Density estimation for statistics and data analysis.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948.

Stambaugh, R. F. and Yuan, Y. (2016). Mispricing factors. *The Review of Financial Studies*, 30(4):1270–1315.

Tauchen, G. E. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica: Journal of the Econometric Society*, pages 485–505.

Wang, J. (1994). A model of competitive stock trading volume. *Journal of political Economy*, 102(1):127–168.

Appendix. Variable Definitions

The variables used for the analysis are described in detail below. They are constructed using data from CRSP (daily and monthly stock files) and Compustat (annual industrial file).

Variables used for the baseline analysis

All of the data items used to construct these variables are from CRSP. The CRSP items names are shown in roman capital letters.

1. $R_{i,t}$: Denotes the return for stock i in month t. It is RET from the monthly stock file.

2. $TURN_{i,t}$: Denotes turnover for stock i in month t. $TURN = \text{VOL}/(10 \times \text{SHROUT})$, where VOL and SHROUT are trading volume and shares outstanding from the monthly stock file.

3. $ME_{i,t}$: Denotes the market equity of firm i in month t. $ME = |\text{PRC}| \times (\text{SHROUT}/1000)$, where PRC is the stock price from the monthly stock file.

4. $VOL_{i,t}$: Denotes the realized volatility for stock i in month t. We compute this variable as

$$VOL_{i,t} = \left(\sum_{n=1}^{N_{i,t}} (R_{i,t_n} - \hat{m}_t(R_{i,t_n}))^2 \right)^{1/2},$$

where R_{i,t_n} is the stock return for day n of month t (RET from the daily stock file), $N_{i,t}$ is the number of days with non-missing daily returns for stock i in month t, and $\hat{m}_t(R_{i,t_n}) = (1/N_{i,t}) \sum_{n=1}^{N_{i,t}} R_{i,t_n}$. We treat $VOL_{i,t}$ as missing if $N_{i,t} < 11$.

5. $MOM_{i,t}$: Denotes the momentum for stock i in month t, which is measured using the stock return over the first 11 months of the prior year. That is, $MOM_{i,t} = (\prod_{n=1}^{11}(1 + R_{i,t-n})) - 1$.

6. $CORR_{i,t}$: Denotes the correlation between squared demeaned daily stock returns and daily turnover for stock i in month t. We compute this variable as

$$CORR_{i,t} = \frac{\sum_{n=1}^{N_{i,t}}(TURN_{i,t_n} - \hat{m}_t(TURN_{i,t_n}))(S_{i,t_n} - \hat{m}_t(S_{i,t_n}))}{(\sum_{n=1}^{N_{i,t}}(TURN_{i,t_n} - \hat{m}_t(TURN_{i,t_n}))^2)^{1/2}(\sum_{n=1}^{N_{i,t}}(S_{i,t_n} - \hat{m}_t(S_{i,t_n}))^2)^{1/2}},$$

where $\hat{m}_t(TURN_{i,t_n}) = (1/N_{i,t})\sum_{n=1}^{N_{i,t}} TURN_{i,t_n}$, $\hat{m}_t(S_{i,t_n}) = (1/N_{i,t})\sum_{n=1}^{N_{i,t}} S_{i,t_n}$, and $S_{i,t_n} = (R_{i,t_n} - \hat{m}_t(R_{i,t_n}))^2$. We treat $CORR_{i,t}$ as missing if $N_{i,t} < 11$.

Additional variables used for the robustness checks

Most of these variables are constructed using one or more annual Compustat data items. To account for the delay between the end of a firm's fiscal year and the date its annual report is disseminated, we match Compustat information for firms whose fiscal year ends in month t with stock returns for months $t+5$ to $t+16$ (i.e. all financial statement items are lagged by at least four months with respect to the interval covered by the returns).[5] Hence, the t subscripts in the variable definitions denote information observed at the end of month $t-5$ or earlier. Compustat mnemonics are shown in roman capital letters, and lag($\cdot$) is used to denote the first annual lag of the argument.

1. $BTM_{i,t}$: Denotes the book-to-market equity ratio for stock i in month t. Market equity is $ME_{i,t}$. Book equity is derived from Compustat. Following Fama and French (1992), it is defined as shareholders equity (SEQ), plus balance-sheet deferred taxes and investment tax credit (TXDITC), if available, minus the book value of preferred stock, which is either its redemption value (PSTKRV), liquidation value (PSTKL), or par value (PSTK), in this order of preference. If the value of SEQ is missing, we substitute common equity plus preferred stock (CEQ plus PSTK), if available, or assets minus liabilities (AT minus LT), if available, in this order of preference. We treat book equity as missing if it is less than zero.

2. $AG_{i,t}$: Denotes asset growth for firm i in month t. The continuously-compounded growth rate of firm assets over the prior fiscal year (log(AT/lag(AT))).

3. $ROA_{i,t}$: Denotes the return on assets for firm i in month t. The ratio of income before extraordinary items to beginning-of-year assets for the prior fiscal year (IB divided by lag(AT)).

4. $GP_{i,t}$: Denotes gross profitability for firm i in month t. The ratio of revenue minus

[5] We exclude firms with less than two years of Compustat data to mitigate the well-known biases that arise from the way in which firms are added to the file.

cost of goods sold to assets for the prior fiscal year (REV minus COGS as a fraction of AT).

5. $NOA_{i,t}$: Denotes net operating assets for firm i in month t. The ratio of operating assets minus operating liabilities for the prior fiscal year to beginning-of-year total assets. Operating assets are assets (AT) minus cash and short-term investments (CHE). Operating liabilities are assets (AT), minus debt in current liabilities (DLC), minus long-term debt (DLTT), minus minority interest (MIB), minus par value of preferred stock (PSTK), minus common equity (CEQ). Missing values of MIB and PSTK are set to zero.

6. $ITA_{i,t}$: Denotes the investment-to-assets ratio for firm i in month t. Investment is the change in gross property, plant, and equipment over the prior fiscal year (PPEGT minus lag(PPEGT)) plus the change in inventories over the prior fiscal year (INVT minus lag(INVT)). Assets are the beginning-of-year assets (lag(AT)).

7. $ACC_{i,t}$: Denotes total accruals for firm i in month t. The ratio of the change in non-cash working capital over the prior fiscal year minus the depreciation expense (DP) for the year to the average value of assets for the year (the average of AT and lag(AT)). Non-cash working capital is current assets (ACT), minus cash and short-term investments (CHE), minus current liabilities (LCT), plus debt in current liabilities (DLC), plus taxes payable (TXP). Missing values of DLC and TXP are set to zero.

8. $NNS_{i,t}$: Denotes share growth for firm i in month t. The continuously-compounded growth rate of split-adjusted shares outstanding (AJEX times CSHO) over the prior fiscal year.

9. $OSC_{i,t}$: Denotes the O-score (estimated bankruptcy probability) for firm i in month

t. The formula is

$$
\begin{aligned}
OSC \quad = \quad & -1.32 - 0.407 \log(\mathrm{AT}) + 6.03 \left(\frac{\mathrm{DLC} + \mathrm{DLTT}}{\mathrm{AT}} \right) - 1.43 \left(\frac{\mathrm{ACT} - \mathrm{LCT}}{\mathrm{AT}} \right) + \\
& 0.076 \left(\frac{\mathrm{LCT}}{\mathrm{ACT}} \right) - 1.72 \times 1_{(\mathrm{LT}>\mathrm{AT})} - 2.37 \left(\frac{\mathrm{NI}}{\mathrm{AT}} \right) - 1.83 \left(\frac{\mathrm{PI}}{\mathrm{LT}} \right) + \\
& 0.285 \times 1_{(\mathrm{NI}<0 \ \& \ \mathrm{lag(NI)}<0)} - 0.521 \left(\frac{\mathrm{NI} - \mathrm{lag(NI)}}{|\mathrm{NI}| + |\mathrm{lag(NI)}|} \right),
\end{aligned}
$$

where NI is net income, PI is pre-tax income, and $1_{(\cdot)}$ denotes the indicator function.

10. $SPREAD_{i,t}$: Denotes the average proportional bid-ask spread for firm i in month t. We compute this variable as

$$SPREAD_{i,t} = \frac{1}{N_{i,t}} \sum_{n=1}^{N_{i,t}} 2 \left(\frac{A_{i,t_n} - B_{i,t_n}}{A_{i,t_n} + B_{i,t_n}} \right),$$

where A_{i,t_n} and B_{i,t_n} denote the closing bid and ask prices for day n of month t (ASK and BID from the daily stock file). Following Chung and Zhang (2014), we treat the summand for day t_n as missing if $A_{i,t_n} = B_{i,t_n} = 0$ or if $A_{i,t_n} - B_{i,t_n} > (A_{i,t_n} + B_{i,t_n})/4$.

Table 2.1
Portfolios Formed on Returns and on Turnover

Panel A: All stocks

	Portfolios formed on prior returns						Portfolios formed on prior turnover					
Quintile	Mean	Vol	Avg ME	ME Share	4-Fac Alpha	t-stat	Mean	Vol	Avg ME	ME Share	4-Fac Alpha	t-stat
Low (L)	0.59	6.70	0.78	8.86	0.07	0.64	0.44	3.64	0.27	3.09	0.01	0.07
2	0.73	5.07	1.89	21.53	0.23	3.30	0.46	4.18	0.96	11.02	−0.07	−1.06
3	0.55	4.33	2.39	27.28	0.05	1.23	0.55	4.24	2.84	32.42	0.03	0.61
4	0.51	4.29	2.42	27.65	0.01	0.12	0.59	4.86	2.84	32.38	−0.02	−0.45
High (H)	0.33	5.17	1.28	14.67	−0.25	−2.54	0.56	6.44	1.85	21.09	−0.17	−2.07
H−L	−0.25				−0.32	−1.72	0.11				−0.17	−1.35

Panel B: All-but-micro-cap stocks

Decile	Mean	Vol	Avg ME	ME Share	4-Fac Alpha	t-stat	Mean	Vol	Avg ME	ME Share	4-Fac Alpha	t-stat
Low (L)	0.62	6.13	3.02	14.68	0.11	1.08	0.42	3.71	3.67	17.83	−0.03	−0.50
2	0.64	4.64	4.63	22.48	0.12	2.08	0.51	3.95	6.27	30.46	0.04	0.74
3	0.56	4.27	4.97	24.12	0.05	1.29	0.58	4.56	4.76	23.13	0.05	1.34
4	0.45	4.27	4.85	23.55	−0.04	−0.77	0.56	5.25	3.52	17.10	−0.04	−0.92
High (H)	0.33	5.05	3.12	15.16	−0.24	−2.50	0.57	6.98	2.36	11.48	−0.13	−1.27
H−L	−0.30				−0.34	−1.98	0.15				−0.10	−0.69

Panel C: Large-cap stocks

Decile	Mean	Vol	Avg ME	ME Share	4-Fac Alpha	t-stat	Mean	Vol	Avg ME	ME Share	4-Fac Alpha	t-stat
Low (L)	0.62	5.79	6.48	16.90	0.11	1.20	0.42	3.62	12.02	31.34	−0.02	−0.38
2	0.65	4.49	8.43	21.98	0.15	2.87	0.57	4.14	10.02	26.13	0.10	2.41
3	0.50	4.22	8.61	22.46	0.02	0.37	0.57	4.57	7.04	18.35	0.10	2.46
4	0.43	4.20	8.42	21.97	−0.05	−0.98	0.53	5.26	5.33	13.91	−0.05	−0.97
High (H)	0.32	4.87	6.40	16.69	−0.22	−2.43	0.52	6.95	3.94	10.27	−0.14	−1.26
H−L	−0.30				−0.33	−1.99	0.10				−0.12	−0.81

Each panel reports descriptive statistics for two sets of value-weighted quintile portfolios. The statistics are the average excess percentage return (Mean), the volatility of the excess percentage return (Vol), the average market equity of the stocks in the portfolio across all stock-month observations (Avg ME) in billions of dollars, the percentage of total market equity across all stock-month observations that is attributable to the stocks in the portfolio (ME share), the intercept produced by a time series regression the excess percentage returns on the four **?** factors (4-Fac Alpha), and the heteroskedasticity-robust t-statistic of the intercept (t-stat). To estimate the alpha of a portfolio, we use ordinary least squares to fit a time-series regression of its monthly excess return on the monthly realizations of the four **?** factors.The returns for month $t+1$ are for portfolios formed in month t. We form the first set of portfolios by sorting stocks into return quintiles, and the second set of portfolios by sorting stocks into turnover quintiles. Each stock is weighted in proportion to its market equity at the time the portfolios are formed. The sample used to form the portfolios in a given month is either all available NYSE, AMEX, and NASDAQ stocks (all stocks), the subset of stocks whose market equity is larger than the 20th percentile of the NYSE market-equity distribution (all-but-micro-cap stocks), or the subset of stocks whose market equity is larger than the 50th percentile of the NYSE market-equity distribution (large-cap stocks). The sample period is July 1963 to December 2018.

Table 2.2

Portfolios Formed on Returns After Sorting on Turnover Using All-But-Micro-Cap Stocks

Panel A: Average excess portfolio returns and estimated four-factor alphas

	Prior return quintile						Prior return quintile					
	L	2	3	4	H	H−L	L	2	3	4	H	H−L
Prior turnover quintile	Average excess return						Estimated four-factor alpha					
Low (L)	0.85	0.60	0.43	0.39	0.02	−0.83	0.37	0.11	−0.04	−0.07	−0.45	−0.82
	(4.17)	(3.66)	(2.65)	(2.47)	(0.12)	(−4.87)	(3.17)	(1.11)	(−0.39)	(−0.73)	(−3.88)	(−4.72)
2	0.93	0.76	0.56	0.29	0.06	−0.87	0.46	0.27	0.09	−0.18	−0.39	−0.85
	(4.55)	(4.45)	(3.51)	(1.81)	(0.35)	(−5.31)	(3.99)	(3.17)	(1.09)	(−2.31)	(−4.15)	(−5.13)
3	0.94	0.81	0.50	0.46	0.32	−0.61	0.40	0.27	−0.05	−0.06	−0.19	−0.59
	(4.01)	(4.08)	(2.77)	(2.68)	(1.69)	(−3.42)	(3.50)	(3.19)	(−0.69)	(−0.81)	(−1.85)	(−3.28)
4	0.70	0.61	0.63	0.59	0.33	−0.37	0.09	−0.00	0.04	−0.04	−0.31	−0.40
	(2.74)	(2.78)	(3.02)	(2.81)	(1.51)	(−1.83)	(0.73)	(−0.05)	(0.45)	(−0.47)	(−2.54)	(−2.02)
High (H)	0.12	0.53	0.73	0.73	0.71	0.58	−0.49	−0.19	−0.03	0.05	−0.08	0.41
	(0.38)	(1.80)	(2.58)	(2.65)	(2.40)	(2.35)	(−2.98)	(−1.39)	(−0.23)	(0.37)	(−0.46)	(1.71)
Average	0.71	0.66	0.57	0.49	0.29		0.16	0.09	0.00	−0.06	−0.28	
	(3.14)	(3.49)	(3.20)	(2.80)	(1.53)		(1.97)	(2.00)	(0.03)	(−1.29)	(−3.56)	

Panel B: Turnover and volatility characteristics

	Average monthly turnover of stocks (%)					Average monthly volatility of stocks (%)				
L (Low)	2.13	2.07	2.08	2.12	2.15	9.38	7.47	7.23	7.58	9.85
2	4.56	4.56	4.55	4.56	4.58	9.66	7.67	7.44	7.80	10.22
3	6.96	6.91	6.90	6.91	6.97	10.56	8.43	8.15	8.53	11.11
4	10.60	10.45	10.41	10.46	10.61	12.13	9.90	9.62	10.01	12.80
H (High)	26.11	21.57	21.22	21.82	26.95	16.91	13.11	12.75	13.42	18.23

Each panel reports descriptive statistics for a set of 25 value-weighted portfolios. Panel A reports the average monthly excess return and estimated alpha for each portfolio, with heteroskedasticity-robust t-statistics shown in parentheses. To estimate the alpha of a portfolio, we use ordinary least squares to fit a time-series regression of its monthly excess return on the monthly realizations of the four **?** factors. Panel B reports the average monthly turnover and average monthly realized volatility of the constituent stocks across all stock-month observations. The portfolio returns for month $t+1$ are for portfolios formed on stock returns and turnover in month t. First, we sort stocks into turnover quintiles. Second, we sort the stocks contained in each turnover quintile into return quintiles. Third, we weight the month $t+1$ excess returns of the stocks in each of the 25 groups in proportion to the stock's month-end market equity for month t. The sorts and portfolios exclude stocks whose market equity is smaller than the 20th percentile of the NYSE market-equity distribution for month t. The sample period is July 1963 to December 2018.

Table 2.3

Portfolios Formed on Returns After Sorting on Turnover Using Large-Cap Stocks

Panel A: Average excess portfolio returns and estimated four-factor alphas

	Prior return quintile						Prior return quintile					
	L	2	3	4	H	H−L	L	2	3	4	H	H˘L
Prior turnover quintile	Average excess return						Estimated four-factor alpha					
Low (L)	0.87	0.58	0.41	0.26	−0.03	−0.90	0.41	0.11	−0.04	−0.17	−0.48	−0.90
	(4.63)	(3.57)	(2.70)	(1.68)	(−0.18)	(−5.82)	(3.66)	(1.15)	(−0.43)	(−1.84)	(−4.93)	(−5.66)
2	0.96	0.70	0.61	0.41	0.18	−0.78	0.45	0.25	0.13	−0.07	−0.26	−0.71
	(4.87)	(3.86)	(3.62)	(2.47)	(1.08)	(−5.07)	(4.22)	(2.97)	(1.69)	(−0.83)	(−2.92)	(−4.55)
3	0.80	0.81	0.56	0.48	0.25	−0.55	0.36	0.29	0.05	−0.01	−0.24	−0.60
	(3.55)	(4.17)	(2.99)	(2.74)	(1.36)	(−3.08)	(3.19)	(3.27)	(0.58)	(−0.08)	(−2.38)	(−3.40)
4	0.68	0.61	0.59	0.54	0.36	−0.32	0.12	0.04	−0.01	−0.07	−0.27	−0.39
	(2.77)	(2.78)	(2.83)	(2.51)	(1.63)	(−1.65)	(1.07)	(0.40)	(−0.12)	(−0.71)	(−2.20)	(−2.04)
High (H)	0.04	0.50	0.64	0.80	0.60	0.56	−0.54	−0.15	−0.02	0.15	−0.13	0.42
	(0.14)	(1.71)	(2.33)	(2.90)	(2.09)	(2.24)	(−3.23)	(−1.09)	(−0.18)	(0.98)	(−0.71)	(1.71)
Average	0.67	0.64	0.56	0.50	0.27		0.16	0.11	0.02	−0.03	−0.28	
	(3.13)	(3.39)	(3.18)	(2.83)	(1.48)		(2.07)	(2.35)	(0.54)	(−0.68)	(−3.48)	

Panel B: Turnover and volatility characteristics

	Average monthly turnover of stocks (%)					Average monthly volatility of stocks (%)				
L (Low)	2.92	2.96	2.97	2.96	2.95	7.61	6.37	6.23	6.46	8.07
2	5.41	5.36	5.36	5.38	5.41	7.94	6.81	6.67	6.89	8.34
3	7.59	7.54	7.52	7.55	7.61	8.84	7.59	7.46	7.68	9.25
4	10.97	10.79	10.79	10.82	10.97	10.29	8.86	8.68	8.96	10.83
H (High)	24.83	21.04	20.77	21.26	26.11	14.47	11.74	11.50	11.91	15.59

Each panel reports descriptive statistics for a set of 25 value-weighted portfolios. Panel A reports the average monthly excess return and estimated alpha for each portfolio, with heteroskedasticity-robust t-statistics shown in parentheses. To estimate the alpha of a portfolio, we use ordinary least squares to fit a time-series regression of its monthly excess return on the monthly realizations of the four factors. Panel B reports the average monthly turnover and average monthly realized volatility of the constituent stocks across all stock-month observations. The portfolio returns for month $t+1$ are for portfolios formed on stock returns and turnover in month t. First, we sort stocks into turnover quintiles. Second, we sort the stocks contained in each turnover quintile into return quintiles. Third, we weight the month $t+1$ excess returns of the stocks in each of the 25 groups in proportion to the stock's month-end market equity for month t. The sorts and portfolios exclude stocks whose market equity is smaller than the 50th percentile of the NYSE market-equity distribution for month t. The sample period is July 1963 to December 2018.

Table 2.4

Average Regression Estimates for Turnover and Volatility Deciles

Panel A: Regressions estimated by turnover decile

Regressors in month t	Prior turnover decile using all-but-micro-cap stocks				Prior turnover decile using large-cap stocks			
	1	4	7	10	1	4	7	10
Intercept	0.92	1.06	1.17	0.92	0.89	1.02	1.04	0.90
	(5.67)	(5.68)	(5.22)	(2.87)	(5.70)	(5.96)	(5.10)	(2.98)
R_{t-1}	−0.70	−0.59	−0.35	0.13	−0.55	−0.60	−0.33	0.06
	(−10.14)	(−8.64)	(−5.49)	(2.75)	(−7.36)	(−8.36)	(−5.20)	(1.10)
$\log TURN_{t-1}$	0.04	0.64	−0.01	−0.50	0.07	−0.86	−0.46	−0.40
	(0.91)	(1.64)	(−0.03)	(−3.28)	(1.69)	(−1.87)	(−0.90)	(−2.36)
$\log VOL_{t-1}$	−0.12	−0.14	−0.20	−0.55	−0.14	0.00	−0.19	−0.41
	(−2.40)	(−2.03)	(−2.60)	(−5.84)	(−2.60)	(0.03)	(−2.62)	(−4.29)
$\log ME_{t-1}$	−0.10	−0.13	−0.21	−0.21	−0.06	−0.08	−0.12	−0.07
	(−2.12)	(−2.85)	(−4.07)	(−2.62)	(−1.45)	(−1.98)	(−2.37)	(−0.85)
MOM_{t-1}	0.28	0.37	0.38	0.26	0.19	0.22	0.37	0.29
	(2.91)	(4.33)	(4.58)	(4.38)	(1.85)	(2.22)	(4.40)	(4.85)
R-squared	0.086	0.099	0.088	0.086	0.132	0.131	0.123	0.132

Panel B: Regressions estimated by volatility decile

	Prior volatility decile using all-but-micro-cap stocks				Prior volatility decile using large-cap stocks			
Intercept	0.95	1.21	1.23	0.45	0.90	1.10	1.13	0.54
	(7.31)	(6.64)	(5.36)	(1.27)	(6.90)	(6.42)	(5.38)	(1.66)
R_{t-1}	−0.55	−0.67	−0.32	0.01	−0.53	−0.64	−0.44	0.00
	(−6.62)	(−9.87)	(−5.22)	(0.27)	(−6.32)	(−9.81)	(−6.85)	(0.02)
$\log TURN_{t-1}$	0.08	0.09	0.13	0.02	0.06	0.15	0.22	−0.06
	(2.75)	(1.94)	(2.19)	(0.27)	(1.68)	(2.97)	(3.70)	(−0.78)
$\log VOL_{t-1}$	0.21	−0.07	−0.57	−0.91	0.14	−0.14	−0.08	−0.64
	(4.31)	(−0.21)	(−1.39)	(−6.27)	(2.48)	(−0.34)	(−0.18)	(−4.00)
$\log ME_{t-1}$	−0.07	−0.20	−0.22	−0.21	−0.04	−0.13	−0.18	−0.04
	(−1.97)	(−4.43)	(−3.99)	(−2.20)	(−1.16)	(−3.20)	(−3.83)	(−0.48)
MOM_{t-1}	0.14	0.37	0.32	0.27	0.03	0.23	0.40	0.36
	(1.63)	(4.06)	(3.95)	(3.98)	(0.29)	(2.44)	(4.50)	(5.18)
R-squared	0.085	0.081	0.078	0.075	0.120	0.108	0.112	0.120

We fit the regressions for each month between Jul 1963 and Dec 2018 using selected decile groupings of stocks (1, 4, 7, and 10). The deciles are formed by sorting stocks on either turnover (panel A) or realized volatility (panel B). The dependent variable is the percentage stock return for month t. We use R_{t-1} to denote the stock return for month $t-1$, $\log TURN_{t-1}$ to denote the log turnover for month $t-1$, $\log VOL_{t-1}$ to denote the log realized volatility for month $t-1$, $\log ME_{t-1}$ to denote the log market equity at the end of month $t-1$, and MOM_{t-1} to denote the stock return for the 11-month interval from $t-12$ to $t-2$. Each regressor is standardized on a month-by-month basis (i.e., each variable has a mean of zero and a variance of one in every cross section). The regression for month t excludes stocks for which ME_{t-1} is smaller than either the 20th percentile (all-but-micro-caps) or the 50th percentile (large-caps) of the NYSE market-equity distribution for month $t-1$. We report the average values of the estimated coefficients, **?** t-statistics (in parentheses), and the average R-squared statistic.

Table 2.5
Average Regression Estimates for Turnover Deciles by Proportion of News-Driven Turnover

Panel A: Using all-but-micro-cap stocks

Regressors	Prior turnover decile for stocks in low-proportion category				Prior turnover decile for stocks in high-proportion category			
in month t	1	4	7	10	1	4	7	10
Intercept	0.97	1.08	1.18	1.13	0.89	1.11	1.08	0.76
	(6.01)	(5.88)	(5.61)	(3.67)	(5.28)	(5.63)	(4.53)	(2.27)
R_{t-1}	−0.58	−0.68	−0.59	−0.06	−0.66	−0.48	−0.22	0.22
	(−8.49)	(−9.38)	(−8.56)	(−0.98)	(−6.94)	(−5.52)	(−2.54)	(3.60)
log $TURN_{t-1}$	0.04	0.62	−0.38	−0.20	0.04	0.30	0.42	−0.52
	(0.82)	(1.18)	(−0.63)	(−1.10)	(0.76)	(0.53)	(0.59)	(−2.26)
log VOL_{t-1}	−0.06	−0.14	−0.13	−0.36	−0.13	−0.17	−0.33	−0.68
	(−1.16)	(−2.16)	(−1.58)	(−3.04)	(−2.05)	(−2.13)	(−3.37)	(−5.72)
log ME_{t-1}	−0.13	−0.13	−0.15	−0.19	−0.08	−0.18	−0.22	−0.26
	(−2.19)	(−2.51)	(−2.73)	(−2.20)	(−1.38)	(−3.65)	(−3.38)	(−2.42)
MOM_{t-1}	0.21	0.29	0.28	0.20	0.39	0.43	0.41	0.31
	(2.05)	(3.00)	(3.32)	(2.84)	(3.06)	(3.71)	(4.14)	(4.39)
R-squared	0.125	0.132	0.130	0.132	0.125	0.129	0.121	0.123

Panel B: Using large-cap stocks

Regressors in month t	1	4	7	10	1	4	7	10
Intercept	0.93	1.07	1.13	1.05	0.85	0.98	1.01	0.78
	(5.82)	(6.25)	(5.76)	(3.60)	(5.41)	(5.43)	(4.71)	(2.47)
R_{t-1}	−0.47	−0.55	−0.37	−0.09	−0.54	−0.56	−0.18	0.11
	(−5.85)	(−6.56)	(−4.77)	(−1.45)	(−5.18)	(−5.91)	(−2.11)	(1.64)
log $TURN_{t-1}$	0.10	0.12	−0.85	−0.08	0.08	0.65	1.23	−0.54
	(1.75)	(0.18)	(−1.10)	(−0.42)	(1.21)	(0.89)	(1.59)	(−2.13)
log VOL_{t-1}	−0.13	−0.06	−0.02	−0.30	−0.08	0.03	−0.32	−0.62
	(−1.95)	(−0.74)	(−0.28)	(−2.49)	(−1.19)	(0.39)	(−3.14)	(−4.87)
log ME_{t-1}	−0.02	−0.15	−0.05	−0.15	−0.11	−0.09	−0.17	−0.09
	(−0.35)	(−3.13)	(−0.79)	(−1.62)	(−1.96)	(−1.68)	(−2.46)	(−0.74)
MOM_{t-1}	0.02	0.24	0.21	0.18	0.28	0.16	0.44	0.39
	(0.18)	(1.93)	(2.00)	(2.43)	(1.83)	(1.20)	(4.06)	(4.97)
R-squared	0.203	0.204	0.195	0.197	0.200	0.191	0.187	0.193

We fit the regressions for each month between Jul 1963 and Dec 2018 using selected decile groupings of stocks (1, 4, 7, and 10). The deciles are formed by sorting stocks on turnover. We form two different sets of deciles for each month t: one using stocks for which the estimated correlation between squared demeaned daily returns and daily turnover for month $t-1$ is less than or equal to the median estimated correlation (stocks with a low fraction of news-driven turnover), and one using stocks for which the estimated correlation is greater than the median estimated correlation (stocks with a high fraction of news-driven turnover). The dependent variable is the percentage stock return for month t. We use R_{t-1} to denote the stock return for month $t-1$, log $TURN_{t-1}$ to denote the log turnover for month $t-1$, log VOL_{t-1} to denote the log realized volatility for month $t-1$, log ME_{t-1} to denote the log market equity at the end of month $t-1$, and MOM_{t-1} to denote the stock return for the 11-month interval from $t-12$ to $t-2$. Each regressor is standardized on a month-by-month basis (i.e., each variable has a mean of zero and a variance of one in every cross section). The regression for month t excludes stocks for which ME_{t-1} is smaller than either the 20th percentile (all-but-micro-caps) or the 50th percentile (large-caps) of the NYSE market-equity distribution for month $t-1$. We report the average values of the estimated coefficients, **?** t-statistics (in parentheses), and the average R-squared statistic.

Table 2.6

Portfolios Formed on Returns After Sorting on Turnover for Stocks Grouped by Proportion of News-Driven Turnover

Prior turnover quintile	Prior return quintile for stocks with a low proportion of news-driven turnover						Prior return quintile for stocks with a a high proportion of news-driven turnover					
	L	2	3	4	H	H−L	L	2	3	4	H	H−L
Panel A: Average excess portfolios returns using all-but-micro-cap stocks												
Low (L)	0.98	0.72	0.62	0.41	−0.05	−1.03	0.79	0.45	0.48	0.37	0.03	−0.75
	(4.71)	(4.09)	(3.79)	(2.50)	(−0.28)	(−6.25)	(3.52)	(2.64)	(2.83)	(2.10)	(0.17)	(−3.84)
2	1.06	0.84	0.49	0.28	−0.03	−1.09	0.82	0.83	0.57	0.49	0.12	−0.70
	(4.86)	(4.59)	(2.86)	(1.67)	(−0.16)	(−5.89)	(3.52)	(4.45)	(3.30)	(2.91)	(0.62)	(−3.36)
3	1.14	0.80	0.58	0.28	0.01	−1.13	0.81	0.66	0.75	0.49	0.47	−0.34
	(4.96)	(4.21)	(3.25)	(1.57)	(0.07)	(−6.19)	(3.18)	(3.20)	(3.60)	(2.54)	(2.12)	(−1.51)
4	0.87	0.91	0.63	0.70	0.25	−0.62	0.52	0.75	0.71	0.64	0.31	−0.21
	(3.28)	(4.14)	(3.14)	(3.43)	(1.16)	(−2.81)	(1.85)	(2.96)	(2.96)	(2.64)	(1.31)	(−0.93)
High (H)	0.84	0.97	0.61	0.57	0.49	−0.35	−0.21	0.15	0.55	0.74	0.86	1.06
	(2.56)	(3.54)	(2.27)	(2.23)	(1.79)	(−1.32)	(−0.59)	(0.48)	(1.80)	(2.37)	(2.67)	(3.69)
Average	0.98	0.85	0.59	0.45	0.13		0.55	0.57	0.61	0.55	0.36	
	(4.32)	(4.60)	(3.42)	(2.63)	(0.75)		(2.28)	(2.87)	(3.19)	(2.88)	(1.79)	
Panel B: Average excess portfolios returns using large-cap stocks												
Low (L)	1.16	0.62	0.57	0.27	−0.02	−1.18	0.68	0.49	0.49	0.42	−0.01	−0.70
	(5.79)	(3.41)	(3.36)	(1.69)	(−0.10)	(−6.61)	(3.36)	(2.82)	(2.95)	(2.48)	(−0.08)	(−3.68)
2	1.06	0.69	0.58	0.17	−0.01	−1.06	0.79	0.65	0.60	0.51	0.31	−0.48
	(5.19)	(3.81)	(3.33)	(0.96)	(−0.03)	(−6.53)	(3.61)	(3.35)	(3.25)	(2.83)	(1.61)	(−2.49)
3	1.05	0.77	0.67	0.38	−0.00	−1.05	0.80	0.73	0.53	0.60	0.44	−0.36
	(4.62)	(3.82)	(3.59)	(2.06)	(−0.02)	(−5.66)	(3.31)	(3.36)	(2.52)	(3.10)	(2.04)	(−1.67)
4	0.84	0.80	0.64	0.59	0.27	−0.56	0.36	0.68	0.63	0.57	0.33	−0.03
	(3.26)	(3.67)	(3.15)	(2.90)	(1.26)	(−2.59)	(1.29)	(2.82)	(2.64)	(2.27)	(1.35)	(−0.13)
High (H)	0.84	0.79	0.59	0.57	0.49	−0.35	−0.12	0.25	0.50	0.72	0.94	1.06
	(2.59)	(2.85)	(2.15)	(2.15)	(1.78)	(−1.34)	(−0.35)	(0.79)	(1.65)	(2.36)	(3.00)	(3.72)
Average	0.99	0.73	0.61	0.40	0.15		0.50	0.56	0.55	0.56	0.40	
	(4.56)	(4.01)	(3.55)	(2.33)	(0.82)		(2.23)	(2.82)	(2.88)	(3.01)	(2.05)	

Each panel reports average excess returns for two sets of 25 value-weighted portfolios. The portfolio returns for month $t+1$ are for portfolios formed on stock returns and turnover in month t. First, we use the estimated correlation between squared demeaned daily returns and daily turnover for month t to partition stocks into two categories: those which have an estimated correlation that is less than or equal to the median estimated correlation (stocks with a low fraction of news-driven turnover) and those that have an estimated correlation that is greater than the median estimated correlation (stocks with a high fraction of news-driven turnover). Second, we sort stocks contained in each category into turnover quintiles. Third, we sort the stocks contained in each turnover quintile into return quintiles. Fourth, we weight the month $t + 1$ excess returns of the stocks in each of the two sets of 25 groups in proportion to the stock's month-end market equity for month t. The sorts and portfolios exclude stocks whose market equity is smaller than either the 20th percentile (panel A) or the 50th percentile (panel B) of the NYSE market-equity distribution for month t. The sample period is July 1963 to December 2018.

Table 2.7

Average Regressions Estimates by Turnover Decile Using Anomaly Variables as Controls

Regressors in month t	Prior turnover decile using all-but-micro-cap stocks 1	4	7	10	Prior turnover decile using large-cap stocks 1	4	7	10
Intercept	0.91	1.08	1.18	0.93	0.87	1.05	1.05	0.91
	(4.61)	(5.14)	(4.77)	(2.63)	(4.44)	(5.43)	(4.61)	(2.70)
R_{t-1}	−0.71	−0.64	−0.42	0.05	−0.68	−0.65	−0.39	0.05
	(−7.49)	(−8.03)	(−5.78)	(0.88)	(−6.94)	(−7.11)	(−4.75)	(0.75)
$\log TURN_{t-1}$	0.01	0.01	−0.14	−0.52	0.03	−1.39	0.37	−0.21
	(0.23)	(0.01)	(−0.28)	(−2.90)	(0.41)	(−2.41)	(0.59)	(−1.04)
$\log VOL_{t-1}$	−0.00	−0.08	−0.20	−0.41	−0.12	−0.01	−0.11	−0.38
	(−0.02)	(−1.09)	(−2.31)	(−4.11)	(−1.57)	(−0.17)	(−1.40)	(−3.65)
$\log ME_{t-1}$	−0.00	−0.09	−0.19	−0.18	−0.09	−0.06	−0.10	−0.18
	(−0.06)	(−1.60)	(−3.12)	(−2.11)	(−1.56)	(−1.21)	(−1.60)	(−1.98)
MOM_{t-1}	0.29	0.31	0.35	0.23	0.12	0.14	0.31	0.25
	(2.16)	(3.29)	(3.84)	(3.77)	(0.85)	(1.25)	(2.79)	(3.79)
BTM_{t-1}	0.22	0.28	0.13	0.21	0.29	0.14	0.11	0.30
	(4.09)	(3.50)	(1.53)	(2.03)	(3.29)	(1.64)	(1.41)	(2.41)
AG_{t-1}	−0.03	−0.16	0.08	−0.08	0.03	0.05	0.14	−0.12
	(−0.21)	(−1.60)	(0.86)	(−0.88)	(0.19)	(0.41)	(1.16)	(−1.08)
ROA_{t-1}	0.60	0.25	−0.12	0.17	−0.10	0.10	0.08	0.11
	(3.01)	(1.76)	(−0.89)	(1.80)	(−0.24)	(0.38)	(0.46)	(1.06)
GP_{t-1}	0.11	0.10	0.09	0.01	0.14	0.01	0.13	−0.00
	(2.02)	(1.71)	(1.28)	(0.11)	(2.05)	(0.10)	(1.66)	(−0.02)
NOA_{t-1}	−0.13	−0.18	−0.25	−0.19	0.02	−0.13	−0.19	−0.04
	(−1.25)	(−1.69)	(−2.51)	(−2.09)	(0.10)	(−0.96)	(−1.55)	(−0.36)
ITA_{t-1}	−0.06	0.12	−0.03	−0.09	−0.06	0.01	−0.20	0.02
	(−0.59)	(1.36)	(−0.33)	(−1.14)	(−0.44)	(0.09)	(−1.89)	(0.23)
ACC_{t-1}	−0.13	−0.10	−0.05	−0.17	−0.14	−0.04	−0.01	−0.16
	(−2.09)	(−1.78)	(−0.92)	(−3.22)	(−1.85)	(−0.58)	(−0.22)	(−2.44)
NNS_{t-1}	−0.07	−0.01	−0.06	−0.06	0.25	0.01	−0.04	−0.11
	(−0.66)	(−0.16)	(−0.75)	(−0.82)	(1.64)	(0.08)	(−0.43)	(−1.11)
OSC_{t-1}	0.08	0.03	−0.00	−0.10	0.05	−0.13	0.12	−0.07
	(1.00)	(0.33)	(−0.04)	(−1.00)	(0.55)	(−1.20)	(1.10)	(−0.61)
R-squared	0.229	0.215	0.198	0.193	0.362	0.329	0.302	0.307

We fit the regressions for each month between Jan 1973 and Dec 2018 using selected decile groupings of stocks (1, 4, 7, and 10). The deciles are formed by sorting stocks on turnover. The dependent variable is the percentage stock return for month t. We use R_{t-1} to denote the stock return for month $t-1$, $\log TURN_{t-1}$ to denote the log turnover for month $t-1$, $\log VOL_{t-1}$ to denote the log realized volatility for month $t-1$, $\log ME_{t-1}$to denote the log market equity at the end of month $t-1$, MOM_{t-1} to denote the stock return for the 11-month interval from $t-12$ to $t-2$, $BTM_{t-1}$to denote the book-to-market equity ratio for month $t-1$, AG_{t-1} to denote the annual growth rate of total assets for month $t-1$, ROA_{t-1} to denote the ratio of annual earnings to beginning-of-year total assets for month $t-1$, GP_{t-1} to denote revenues minus cost of goods sold (as a fraction of end-of-year total assets) for month $t-1$, NOA_{t-1} to denote operating assets minus operating liabilities (as a fraction of beginning-of-year total assets) for month $t-1$, ITA_{t-1} to denote the annual change in gross property, plant, and equipment plus the annual change in inventories (as a fraction of beginning-of-year total assets) for month $t-1$, ACC_{t-1} to denote the annual change in non-cash working capital minus depreciation expense (as a fraction of average total assets for the year) for month $t-1$, NNS_{t-1} to denote the annual growth rate of the split-adjusted shares outstanding for month $t-1$, and OSC_{t-1} to denote the estimated probability of bankruptcy from **?**. Each regressor is standardized on a month-by-month basis (i.e., each variable has a mean of zero and a variance of one in every cross section). The regression for month t excludes stocks for which ME_{t-1} is smaller than either the 20th percentile (all-but-micro-caps) or the 50th percentile (large-caps) of the NYSE market-equity distribution for month $t-1$. We report the average values of the estimated coefficients, **?** t-statistics (in parentheses), and the average R-squared statistic.

Table 2.8

Portfolios Formed on Returns After Sorting on Turnover for Stocks Grouped by Liquidity

Prior turnover quintile	Prior return quintile for low-liquidity stocks						Prior return quintile for high-liquidity stocks					
	L	2	3	4	H	H−L	L	2	3	4	H	H−L
Panel A: Average excess portfolios returns using all-but-micro-cap stocks												
Low (L)	0.87	0.71	0.60	0.36	0.01	−0.86	1.16	0.87	0.62	0.45	−0.09	−1.25
	(2.61)	(2.85)	(2.58)	(1.54)	(0.04)	(−3.23)	(4.19)	(3.76)	(2.83)	(2.09)	(−0.36)	(−4.55)
2	1.02	1.06	0.57	0.64	0.05	−0.97	0.82	0.90	0.57	0.48	0.43	−0.39
	(2.77)	(3.84)	(2.08)	(2.38)	(0.16)	(−3.33)	(2.33)	(3.07)	(2.41)	(2.11)	(1.75)	(−1.27)
3	0.88	0.92	0.68	0.52	0.06	−0.82	0.78	0.83	0.60	0.55	0.47	−0.32
	(2.35)	(2.77)	(2.33)	(1.75)	(0.19)	(−2.59)	(1.96)	(2.70)	(2.00)	(2.00)	(1.34)	(−0.84)
4	0.81	0.94	0.88	0.44	0.71	−0.09	0.66	0.85	0.95	0.65	0.33	−0.33
	(1.95)	(2.71)	(2.84)	(1.36)	(2.08)	(−0.28)	(1.53)	(2.30)	(2.70)	(1.81)	(0.90)	(−0.88)
High (H)	0.51	0.58	0.53	0.42	0.58	0.06	−0.35	0.28	0.85	0.92	0.84	1.18
	(0.94)	(1.28)	(1.23)	(1.04)	(1.19)	(0.13)	(−0.65)	(0.55)	(1.68)	(1.81)	(1.71)	(2.64)
Average	0.82	0.84	0.65	0.48	0.28		0.62	0.74	0.72	0.61	0.39	
	(2.21)	(2.86)	(2.41)	(1.77)	(0.93)		(1.74)	(2.52)	(2.63)	(2.23)	(1.35)	
Panel B: Average excess portfolios returns using large-cap stocks												
Low (L)	0.99	0.86	0.50	0.30	0.21	−0.78	0.91	0.89	0.52	0.31	−0.02	−0.93
	(3.31)	(3.42)	(1.92)	(1.19)	(0.73)	(−2.85)	(3.24)	(3.82)	(2.45)	(1.47)	(−0.09)	(−3.61)
2	1.22	0.80	0.77	0.51	0.01	−1.22	0.91	0.85	0.66	0.24	0.62	−0.29
	(3.98)	(2.79)	(2.84)	(1.91)	(0.02)	(−4.77)	(2.66)	(2.85)	(2.54)	(0.91)	(2.39)	(−0.97)
3	0.97	0.76	0.93	0.83	0.24	−0.74	0.78	0.78	0.63	0.50	0.25	−0.53
	(2.79)	(2.35)	(3.25)	(2.82)	(0.80)	(−2.65)	(2.01)	(2.27)	(2.06)	(1.74)	(0.68)	(−1.45)
4	0.47	0.52	0.53	0.58	0.33	−0.14	0.78	1.04	1.08	0.55	0.50	−0.28
	(1.16)	(1.45)	(1.64)	(1.69)	(0.92)	(−0.41)	(1.84)	(2.79)	(2.66)	(1.48)	(1.28)	(−0.73)
High (H)	0.27	0.82	0.48	0.80	0.68	0.41	−0.21	0.52	0.89	0.58	0.88	1.09
	(0.54)	(1.80)	(1.13)	(2.15)	(1.52)	(0.90)	(−0.38)	(0.98)	(1.75)	(1.08)	(1.69)	(2.24)
Average	0.79	0.75	0.64	0.61	0.29		0.63	0.82	0.75	0.44	0.45	
	(2.41)	(2.59)	(2.43)	(2.30)	(1.03)		(1.83)	(2.71)	(2.63)	(1.54)	(1.48)	

Each panel reports average excess returns for two sets of 25 value-weighted portfolios. The portfolio returns for month $t+1$ are for portfolios formed on stock returns and turnover in month t. First, we use the number of daily returns that are zero for month t to partition stocks into two categories: those for which the number of zero daily returns is less than or equal to the median number of zero daily returns (high-liquidity stocks) and those for which the number of zero daily returns is greater than the median number of zero daily returns (low-liquidity stocks). Second, we sort stocks contained in each category into turnover quintiles.

Third, we sort the stocks contained in each turnover quintile into return quintiles. Fourth, we weight the month $t+1$ excess returns of the stocks in each of the two sets of 25 groups in proportion to the stock's month-end market equity for month t. The sorts and portfolios exclude stocks whose market equity is smaller than either the 20th percentile (panel A) or the 50th percentile (panel B) of the NYSE market-equity distribution for month t. The sample period is January 1993 to December 2018.

Table 2.9

Average Regression Estimates for Turnover and Spread-Based Liquidity Deciles

Panel A: Regressions by turnover decile

Regressors in month t	Prior turnover decile using low-liquidity stocks				Prior turnover decile using high-liquidity stocks			
	1	4	7	10	1	4	7	10
Intercept	0.82	0.99	1.10	0.92	1.00	0.96	0.93	0.76
	(3.50)	(3.47)	(3.36)	(1.92)	(4.80)	(3.92)	(2.87)	(1.43)
R_{t-1}	−0.65	−0.34	−0.23	−0.01	−0.69	−0.37	−0.21	0.16
	(−4.15)	(−2.39)	(−1.58)	(−0.15)	(−4.84)	(−3.09)	(−1.96)	(1.94)
$TURN_{t-1}$	−0.14	−1.80	−0.35	−0.59	0.09	0.03	−0.61	−0.55
	(−1.52)	(−2.00)	(−0.32)	(−1.56)	(0.99)	(0.04)	(−0.69)	(−2.00)
$SPREAD_{t-1}$	−0.07	0.02	0.58	−0.20	0.12	0.03	0.06	−0.18
	(−1.34)	(0.16)	(1.95)	(−0.79)	(2.15)	(0.49)	(0.75)	(−1.47)
log VOL_{t-1}	−0.22	−0.23	−0.39	−0.31	−0.17	−0.25	−0.19	−0.35
	(−2.43)	(−1.89)	(−2.99)	(−1.79)	(−2.03)	(−1.80)	(−1.27)	(−1.98)
log ME_{t-1}	−0.16	−0.18	−0.21	−0.06	−0.04	−0.09	−0.13	−0.21
	(−1.61)	(−2.32)	(−2.65)	(−0.53)	(−0.55)	(−1.31)	(−1.40)	(−1.41)
MOM_{t-1}	0.20	0.20	−0.00	0.08	0.06	0.28	0.24	0.15
	(1.07)	(1.31)	(−0.01)	(0.63)	(0.30)	(1.70)	(1.68)	(1.36)
R-squared	0.132	0.113	0.119	0.128	0.129	0.130	0.126	0.127

Panel B: Regressions by spread-based liquidity decile

	Prior liquidity decile using low-turnover stocks				Prior liquidity decile using high-turnover stocks			
Intercept	0.88	1.07	0.98	0.81	0.97	1.01	0.98	1.04
	(4.32)	(4.32)	(3.52)	(2.79)	(2.44)	(2.77)	(2.71)	(2.25)
R_{t-1}	−0.45	−0.32	−0.25	−0.29	0.01	−0.08	−0.15	0.02
	(−4.32)	(−3.59)	(−2.60)	(−3.38)	(0.06)	(−0.60)	(−1.15)	(0.15)
$TURN_{t-1}$	0.14	0.20	0.13	0.00	−0.20	−0.27	0.05	0.04
	(1.59)	(2.47)	(1.61)	(0.03)	(−1.76)	(−2.61)	(0.44)	(0.31)
$SPREAD_{t-1}$	0.93	2.80	0.18	−0.06	−3.58	0.21	−1.13	0.07
	(0.78)	(1.37)	(0.13)	(−0.98)	(−2.50)	(0.10)	(−0.89)	(0.71)
log VOL_{t-1}	−0.04	−0.13	−0.20	−0.24	−0.10	−0.23	−0.32	−0.42
	(−0.45)	(−1.27)	(−1.66)	(−2.10)	(−0.73)	(−1.49)	(−2.14)	(−2.47)
log ME_{t-1}	−0.03	−0.11	−0.26	−0.25	−0.21	−0.31	−0.19	−0.02
	(−0.41)	(−1.29)	(−2.11)	(−1.56)	(−1.83)	(−2.80)	(−1.55)	(−0.14)
MOM_{t-1}	0.16	0.20	0.16	0.02	0.35	0.37	0.52	0.36
	(1.48)	(1.67)	(1.39)	(0.13)	(1.99)	(2.34)	(2.85)	(1.58)
R-squared	0.140	0.123	0.114	0.118	0.154	0.134	0.123	0.123

We fit the regressions for each month between Jan 1993 and Dec 2018 using selected decile groupings of stocks (1, 4, 7, and 10). The deciles are formed by sorting stocks on either turnover (panel A) or a spread-based liquidity proxy (panel B). In each case, we form two different sets of deciles for each month t: one using stocks for which the alternative sorting characteristic (the spread in panel A and turnover in panel B) is less than or equal to its median value, and one using stocks for which this characteristic is above its median value. The dependent variable is the percentage stock return for month t. We use R_{t-1} to denote the stock return for month $t-1$, $SPREAD_{t-1}$ to denote the average value of the daily percentage bid-ask spread for month $t-1$, log $TURN_{t-1}$ to denote the log turnover for month $t-1$, log VOL_{t-1} to denote the log realized volatility for month $t-1$, log ME_{t-1} to denote the log market equity at the end of month $t-1$, and MOM_{t-1} to denote the stock return for the 11-month interval from $t-12$ to $t-2$. Each regressor is standardized on a month-by-month basis (i.e., each variable has a mean of zero and a variance of one in every cross section). The regression for month t excludes stocks for which ME_{t-1} is smaller than the 20th percentile of the NYSE market-equity distribution for month $t-1$. We report the average values of the estimated coefficients, **?** t-statistics (in parentheses), and the average R-squared statistic.

Table 2.10

Portfolios Formed on Returns After Sorting on Turnover For Stocks Grouped by Liquidity and Proportion of News-Driven Turnover

Prior turnover quintile	Prior return quintile for stocks with a low proportion of news-driven turnover						Prior return quintile for stocks with a high proportion of news-driven turnover					
	L	2	3	4	H	H−L	L	2	3	4	H	H−L
Panel A: Average excess portfolios returns using low-liquidity stocks												
Low (L)	0.76	0.84	0.61	0.42	−0.02	−0.78	0.64	0.52	0.39	0.49	0.21	−0.42
	(2.19)	(3.11)	(2.50)	(1.73)	(−0.07)	(−2.59)	(1.64)	(1.89)	(1.50)	(1.84)	(0.64)	(−1.23)
2	1.11	1.12	0.70	0.63	−0.08	−1.19	0.59	0.73	0.77	0.49	0.12	−0.47
	(2.88)	(3.97)	(2.67)	(2.26)	(−0.25)	(−3.67)	(1.49)	(2.21)	(2.49)	(1.58)	(0.33)	(−1.36)
3	1.24	0.76	0.89	0.47	−0.32	−1.55	0.70	1.09	0.61	0.98	0.55	−0.14
	(3.21)	(2.39)	(3.10)	(1.57)	(−0.93)	(−4.60)	(1.69)	(2.87)	(1.76)	(2.91)	(1.58)	(−0.38)
4	1.08	0.60	0.96	0.49	0.31	−0.77	0.89	0.92	0.38	0.12	0.40	−0.50
	(2.62)	(1.75)	(2.94)	(1.55)	(0.87)	(−2.17)	(1.79)	(2.34)	(1.02)	(0.32)	(1.03)	(−1.14)
High (H)	1.18	0.68	0.81	0.66	0.66	−0.52	0.57	−0.10	0.32	0.79	0.83	0.27
	(2.10)	(1.61)	(2.12)	(1.60)	(1.45)	(−1.10)	(0.96)	(−0.21)	(0.65)	(1.70)	(1.53)	(0.48)
Average	1.07	0.80	0.79	0.54	0.11		0.68	0.63	0.49	0.57	0.42	
	(2.91)	(2.84)	(3.11)	(2.04)	(0.37)		(1.71)	(1.94)	(1.65)	(1.98)	(1.29)	
Panel B: Average excess portfolios returns using high-liquidity stocks												
Low (L)	1.30	0.69	0.68	0.64	−0.01	−1.31	0.78	0.57	0.60	0.49	0.12	−0.66
	(4.42)	(2.66)	(2.82)	(2.41)	(−0.03)	(−4.26)	(2.30)	(2.29)	(2.59)	(2.15)	(0.47)	(−1.85)
2	1.20	0.76	0.52	0.18	−0.03	−1.23	0.81	0.82	0.62	0.52	0.71	−0.09
	(3.48)	(2.93)	(2.10)	(0.65)	(−0.11)	(−4.14)	(2.05)	(2.67)	(2.10)	(2.04)	(2.22)	(−0.24)
3	1.39	1.07	0.80	0.55	0.06	−1.33	0.54	0.67	0.70	0.59	0.22	−0.32
	(3.59)	(3.08)	(2.98)	(1.98)	(0.20)	(−3.52)	(1.21)	(1.95)	(2.07)	(1.86)	(0.59)	(−0.76)
4	1.03	0.93	0.89	0.57	0.22	−0.81	0.33	0.95	1.18	0.87	0.49	0.17
	(2.40)	(2.74)	(2.76)	(1.71)	(0.66)	(−2.02)	(0.69)	(2.22)	(2.71)	(2.17)	(1.27)	(0.40)
High (H)	0.54	1.20	0.56	0.42	0.67	0.13	−0.50	0.01	0.72	0.73	1.13	1.63
	(0.95)	(2.54)	(1.17)	(0.96)	(1.39)	(0.29)	(−0.89)	(0.01)	(1.29)	(1.21)	(2.10)	(3.11)
Average	1.09	0.93	0.69	0.47	0.18		0.39	0.60	0.76	0.64	0.54	
	(3.09)	(3.35)	(2.75)	(1.83)	(0.67)		(1.02)	(1.92)	(2.47)	(2.14)	(1.68)	

Each panel reports average excess returns for two sets of 25 value-weighted portfolios. The portfolio returns for month $t+1$ are for portfolios formed on stock returns and turnover in month t. First, we use the estimated correlation between squared demeaned daily returns and daily turnover for month t to partition stocks into two categories: those which have an estimated correlation that is less than or equal to the median estimated correlation (stocks with a low fraction of news-driven turnover) and those that have an estimated correlation that is greater than the median estimated correlation (stocks with a high fraction of news-driven turnover). Second, we sort stocks contained in each category into turnover quintiles. Third, we sort the stocks contained in each turnover quintile into return quintiles. Fourth, we weight the month $t+1$ excess returns of the stocks in each of the two sets of 25 groups in proportion to the stock's month-end market equity for month t. The sorts and portfolios exclude stocks whose market equity is smaller than either the 20th percentile (panel A) or the 50th percentile (panel B) of the NYSE market-equity distribution for month t. The sample period is July 1963 to December 2018.

Table 2.11

Portfolios Formed on Returns After Sorting on Turnover for the 1999 to 2018 Sample Period

Panel A: Average excess portfolios returns using all-but-micro-cap stocks

Prior	Prior return quintile for stocks with a low proportion of news-driven turnover						Prior return quintile for stocks with a high proportion of news-driven turnover					
turnover quintile	L	2	3	4	H	H−L	L	2	3	4	H	H−L
Low (L)	0.86	0.89	0.81	0.63	−0.05	−0.91	0.72	0.67	0.30	0.31	−0.01	−0.74
	(2.24)	(2.88)	(2.72)	(2.23)	(−0.16)	(−2.76)	(1.69)	(2.43)	(1.06)	(1.07)	(−0.04)	(−1.89)
2	1.06	0.89	0.39	0.21	−0.15	−1.21	0.45	0.74	0.43	0.52	0.08	−0.37
	(2.68)	(3.03)	(1.39)	(0.74)	(−0.43)	(−3.12)	(1.08)	(2.36)	(1.58)	(2.00)	(0.24)	(−0.88)
3	0.95	0.77	0.63	0.16	0.17	−0.79	0.67	0.67	0.57	0.63	0.62	−0.05
	(2.47)	(2.48)	(2.31)	(0.53)	(0.50)	(−2.22)	(1.52)	(1.95)	(1.69)	(2.01)	(1.48)	(−0.10)
4	0.52	0.99	0.81	0.41	0.02	−0.51	0.44	0.61	0.83	0.71	0.33	−0.11
	(1.07)	(2.54)	(2.50)	(1.22)	(0.05)	(−1.09)	(0.83)	(1.43)	(2.05)	(1.60)	(0.78)	(−0.22)
High (H)	0.68	0.72	0.18	0.32	0.58	−0.11	−0.26	−0.48	0.30	0.86	1.02	1.28
	(1.08)	(1.44)	(0.36)	(0.68)	(1.13)	(−0.20)	(−0.40)	(−0.84)	(0.55)	(1.43)	(1.61)	(2.00)
Average	0.82	0.85	0.56	0.35	0.11		0.40	0.44	0.49	0.61	0.41	
	(1.99)	(2.80)	(2.04)	(1.23)	(0.35)		(0.92)	(1.35)	(1.57)	(1.90)	(1.13)	

Panel B: Average excess portfolios returns using large-cap stocks

Prior turnover quintile	L	2	3	4	H	H−L	L	2	3	4	H	H−L
Low (L)	1.02	0.47	0.61	0.34	−0.07	−1.09	0.59	0.38	0.44	0.31	−0.07	−0.66
	(3.05)	(1.59)	(2.10)	(1.26)	(−0.24)	(−3.25)	(1.78)	(1.38)	(1.58)	(1.20)	(−0.23)	(−1.99)
2	1.10	0.72	0.88	−0.16	−0.17	−1.27	0.70	0.70	0.33	0.72	0.57	−0.13
	(3.26)	(2.54)	(3.20)	(−0.50)	(−0.56)	(−4.21)	(1.87)	(2.02)	(1.07)	(2.54)	(1.84)	(−0.35)
3	0.85	0.77	0.71	0.56	0.03	−0.82	0.41	0.71	0.28	0.76	0.32	−0.09
	(2.19)	(2.16)	(2.18)	(1.76)	(0.08)	(−2.24)	(0.94)	(1.98)	(0.79)	(2.32)	(0.79)	(−0.20)
4	0.35	0.70	0.84	0.41	−0.09	−0.44	0.09	0.44	0.68	0.75	0.33	0.25
	(0.74)	(1.88)	(2.47)	(1.11)	(−0.22)	(−0.98)	(0.16)	(1.05)	(1.65)	(1.51)	(0.73)	(0.49)
High (H)	0.49	0.80	0.31	0.25	0.56	0.07	−0.41	−0.24	0.34	0.93	1.09	1.50
	(0.76)	(1.48)	(0.59)	(0.50)	(1.05)	(0.12)	(−0.65)	(−0.38)	(0.61)	(1.57)	(1.76)	(2.41)
Average	0.76	0.69	0.67	0.28	0.05		0.27	0.40	0.41	0.69	0.45	
	(1.98)	(2.28)	(2.34)	(0.97)	(0.16)		(0.68)	(1.16)	(1.30)	(2.17)	(1.29)	

Each panel reports average excess returns for two sets of 25 value-weighted portfolios. The portfolio returns for month $t + 1$ are for portfolios formed on stock returns and turnover in month t. First, we use the estimated correlation between squared demeaned daily returns and daily turnover for month t to partition stocks into two categories: those which have an estimated correlation that is less than or equal to the median estimated correlation (stocks with a low fraction of news-driven turnover) and those that have an estimated correlation that is greater than the median estimated correlation (stocks with a high fraction of news-driven turnover). Second, we sort stocks contained in each category into turnover quintiles. Third, we sort the stocks contained in each turnover quintile into return quintiles. Fourth, we weight the month $t + 1$ excess returns of the stocks in each of the two sets of 25 groups in proportion to the stock's month-end market equity for month t. The sorts and portfolios exclude stocks whose market equity is smaller than either the 20th percentile (panel A) or the 50th percentile (panel B) of the NYSE market-equity distribution for month t. The sample period is January 1999 to December 2018.

Table 2.12

Portfolios Formed on Returns After Sorting on Turnover Using Micro-Cap Stocks

Panel A: Average excess portfolio returns and estimated four-factor alphas

	Prior return quintile						Prior return quintile					
	L	2	3	4	H	H−L	L	2	3	4	H	H−L
Prior turnover quintile	Average excess return						Estimated four-factor alpha					
Low (L)	1.21	0.58	0.47	0.47	−0.27	−1.48	0.60	−0.01	−0.05	−0.06	−0.85	−1.45
	(4.77)	(3.06)	(2.81)	(2.92)	(−1.52)	(−8.85)	(4.17)	(−0.14)	(−0.52)	(−0.64)	(−7.95)	(−9.25)
2	1.55	0.98	0.79	0.68	−0.25	−1.80	0.88	0.30	0.11	−0.02	−0.93	−1.81
	(5.37)	(4.17)	(3.74)	(3.31)	(−1.18)	(−10.07)	(5.94)	(2.72)	(1.04)	(−0.19)	(−8.40)	(−10.45)
3	1.54	1.09	0.82	0.65	−0.31	−1.84	0.86	0.35	0.06	−0.10	−1.03	−1.89
	(4.68)	(3.97)	(3.35)	(2.78)	(−1.26)	(−9.28)	(5.21)	(3.08)	(0.62)	(−0.98)	(−9.22)	(−10.01)
4	1.66	1.06	0.91	0.65	−0.05	−1.71	1.02	0.32	0.17	−0.14	−0.75	−1.77
	(4.52)	(3.39)	(3.24)	(2.47)	(−0.19)	(−8.01)	(6.00)	(2.38)	(1.57)	(−1.47)	(−6.07)	(−8.59)
High (H)	1.05	0.87	0.83	0.59	0.37	−0.68	0.46	0.13	0.01	−0.24	−0.40	−0.85
	(2.48)	(2.39)	(2.45)	(1.76)	(1.09)	(−2.50)	(2.16)	(0.81)	(0.09)	(−1.72)	(−2.31)	(−3.27)
Average	1.40	0.91	0.77	0.61	−0.10		0.76	0.22	0.06	−0.11	−0.79	
	(4.45)	(3.51)	(3.25)	(2.71)	(−0.44)		(5.87)	(2.40)	(0.76)	(−1.65)	(−9.24)	

Panel B: Turnover and size characteristics

	Average monthly turnover of stocks (%)						Average market equity of stocks ($ billions)					
Low (L)	0.49	0.45	0.41	0.45	0.48		0.03	0.05	0.04	0.05	0.05	
2	1.34	1.32	1.31	1.33	1.35		0.04	0.06	0.07	0.07	0.06	
3	2.59	2.56	2.56	2.57	2.60		0.05	0.07	0.08	0.08	0.07	
4	4.81	4.75	4.74	4.76	4.86		0.06	0.09	0.10	0.10	0.08	
High (H)	18.72	13.56	13.23	14.03	21.22		0.08	0.10	0.11	0.11	0.09	

The table reports descriptive statistics for a set of 25 value-weighted portfolios. Panel A reports the average monthly excess return and estimated four-factor alpha for each portfolio, with heteroskedasticity-robust t-statistics shown in parentheses. To estimate the alpha of a portfolio, we use ordinary least squares to fit a time-series regression of its monthly excess return on the monthly realizations of the four **?** factors. Panel B reports the average monthly turnover and average market equity of the constituent stocks across all stock-month observations. The portfolio returns for month $t+1$ are for portfolios formed on stock returns and turnover in month t. First, we sort stocks into turnover quintiles. Second, we sort the stocks contained in each turnover quintile into return quintiles. Third, we weight the month $t+1$ excess returns of the stocks in each of the 25 groups in proportion to the stock's month-end market equity for month t. The sorts and portfolios exclude stocks whose market equity is larger than the 20th percentile of the NYSE market-equity distribution for month t. The sample period is July 1963 to December 2018.

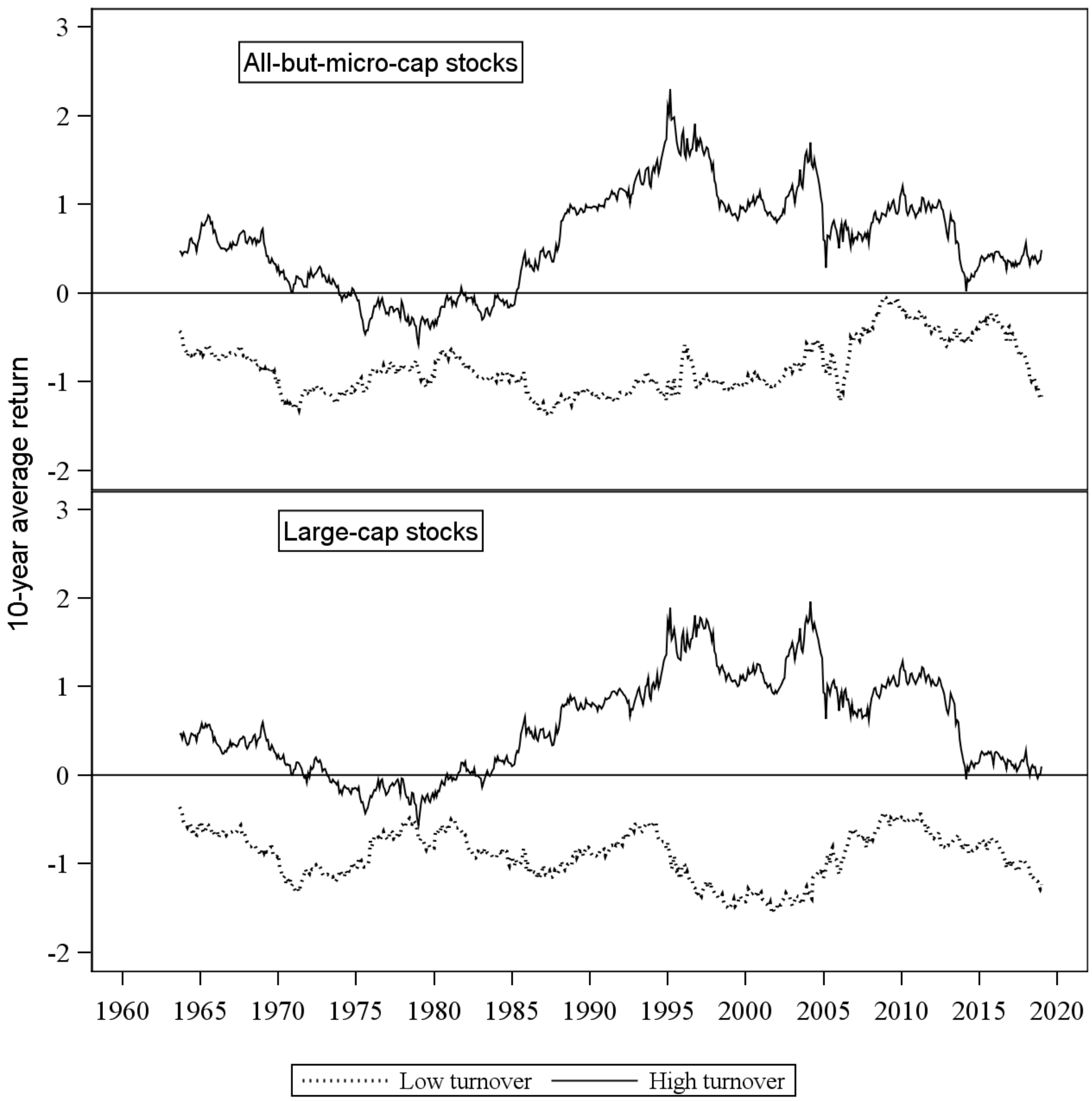

Figure 2.1. Rolling average returns for WML portfolios

The figure plots rolling 10-year average returns for winners-minus-losers (WML) portfolios that are formed using stocks that fall into the bottom and top quintiles of prior turnover. The sample period is July 1963 to December 2018. We compute the rolling average returns for month t by averaging the returns for months $t-59$ to $t+60$. The length of the window is reduced as necessary to account for the lack of observations near the beginning and end of the sample (i.e., for months 1 to 59 and 505 to 564). For example, the average returns for July 1963 are computed using a forward-looking window of 61 observations and those for December 2018 are computed using a backward-looking window of 60 observations. The top panel is for all-but-micro-cap stocks (market equity equal to or greater than the 20th percentile of the NYSE market equity distribution) and the bottom panel is for large-cap stocks (market equity equal to or greater than the 50th percentile of the NYSE market equity distribution).

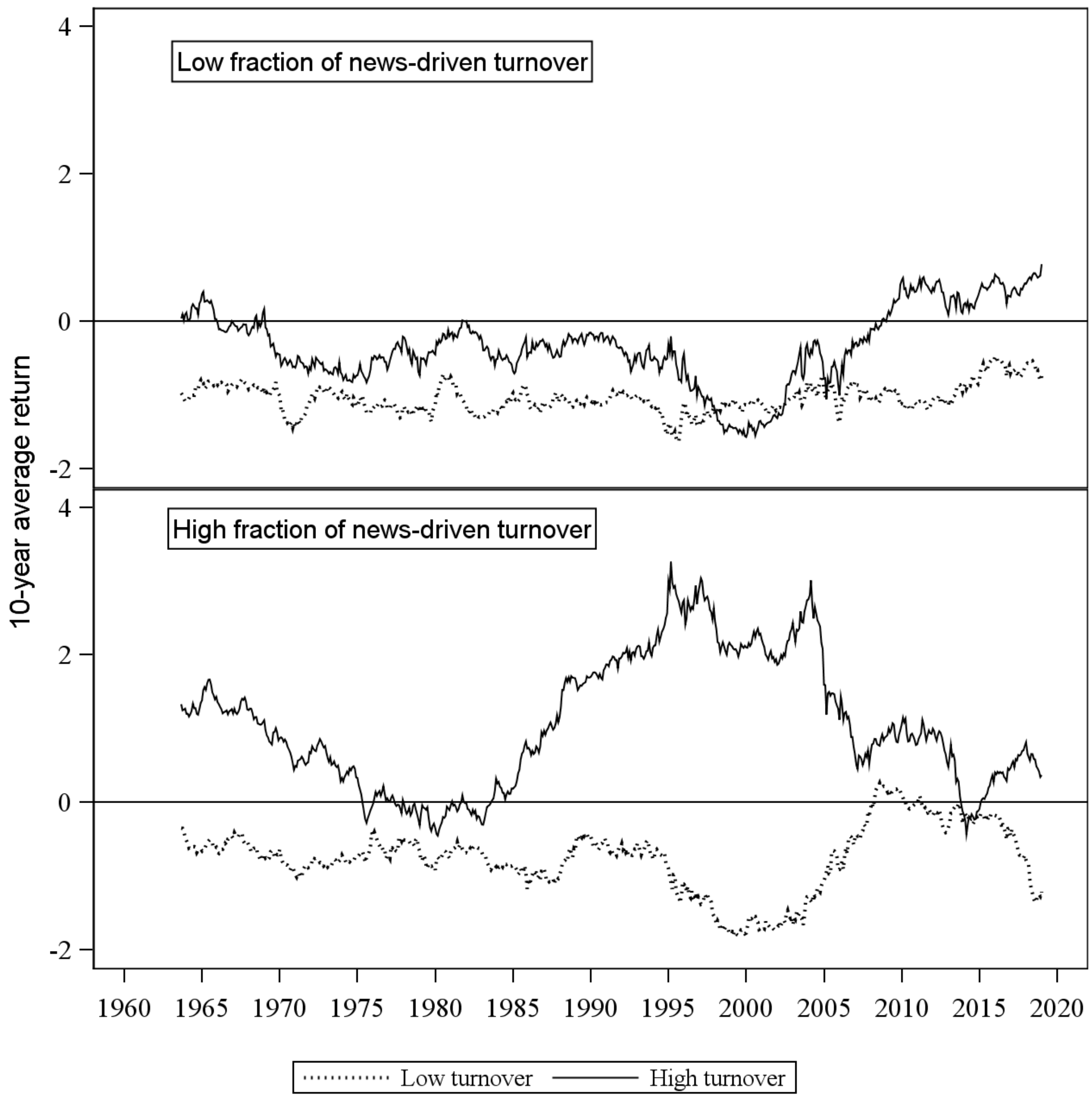

Figure 2.2. Rolling average returns for WML portfolios formed from all-but-micro-cap stocks

The figure plots rolling 10-year average returns for winners-minus-losers (WML) portfolios that are formed using "all-but-micro-cap" stocks (market equity equal to or larger than the 20th percentile of the NYSE market equity distribution) that fall into the bottom and top quintiles of prior turnover. The sample period is July 1963 to December 2018. We compute the rolling average returns for month t by averaging the returns for months $t-59$ to $t+60$. The length of the window is reduced as necessary to account for the lack of observations near the beginning and end of the sample (i.e., for months 1 to 59 and 505 to 564). For example, the average returns for July 1963 are computed using a forward-looking window of 61 observations and those for December 2018 are computed using a backward-looking window of 60 observations. The top panel is for stocks with a low fraction of news-driven turnover (monthly correlation between squared demeaned daily returns and daily turnover less than its median value) and the bottom panel is for stocks with a high fraction of news-driven turnover (monthly correlation between squared demeaned daily returns and daily turnover greater than its median value).

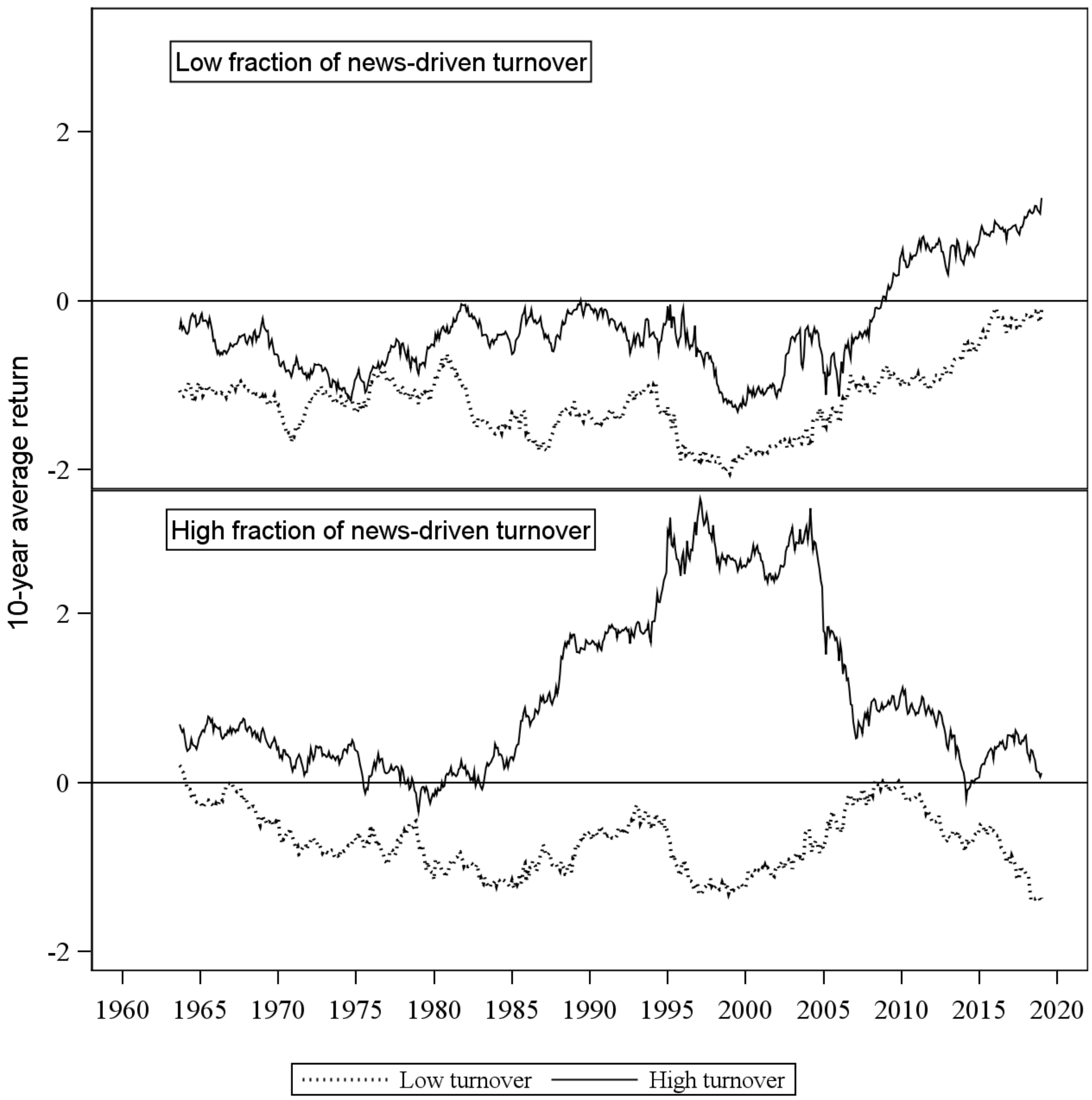

Figure 2.3. Rolling average returns for WML portfolios formed from large-cap stocks

The figure plots rolling 10-year average returns for winners-minus-losers (WML) portfolios that are formed using large stocks (market equity equal to or larger than the 50th percentile of the NYSE market equity distribution) that fall into the bottom and top quintiles of prior turnover. The sample period is July 1963 to December 2018. We compute the rolling average returns for month t by averaging the returns for months $t-59$ to $t+60$. The length of the window is reduced as necessary to account for the lack of observations near the beginning and end of the sample (i.e., for months 1 to 59 and 505 to 564). For example, the average returns for July 1963 are computed using a forward-looking window of 61 observations and those for December 2018 are computed using a backward-looking window of 60 observations. The top panel is for stocks with a low fraction of news-driven turnover (monthly correlation between squared demeaned daily returns and daily turnover less than its median value) and the bottom panel is for stocks with a high fraction of news-driven turnover (monthly correlation between squared demeaned daily returns and daily turnover greater than its median value).

CHAPTER 3: Portfolio Sorts via Nonparametric Regression: From B-Splines to Basis Portfolios

3.1 Introduction

Numerous prior studies suggest that a number of firm characteristics generate the cross-sectional variation in expected stock returns. Our study attempts to form basis portfolios that represent the cross-sectional variations in expected stock returns associated with some firm characteristic. We term these portfolios as basis portfolios.

The recent two studies of Freyberger et al. (2019) and Kirby (2019) provide convincing evidence of nonlinearity in the relation between firm characteristics and expected stock returns. The nonlinearity motivates us to construct basis portfolios using B-spline regression, which is a class of spline regression that is able to capture the potential nonlinearity. Notably, B-spline regression is a more general non-parametric technique than the conventional portfolio sorts. In fact, the conventional portfolios sorting procedure is a special case of B-spline regressions of degree zero. For example, the returns for equally-weighted quintile portfolios formed on size are the OLS estimates of B-spline regressions of stock returns on basic functions of degree zero of the size variable using the corresponding knot sequence.[1] As a consequence, B-spline regression has a number of advantages over the conventional portfolio sorts.

First, basis portfolios obtained using B-spline regression can capture more general relation between one firm characteristic and the expected stock returns than the conventional univariate portfolio sorts. A conventional univariate portfolio sort assumes a piecewise constant relation, whereas B-spline regression assumes a piecewise polynomial relation, such as piecewise constant, piecewise linear, piecewise quadratic, or higher order depending on the degree of the B-spline functions. Second, basis portfolios obtained using B-spline regression capture the incremental cross-sectional variation associated with one firm characteristic

[1] Similarly, the returns for value-weighted portfolios corresponds to the weighted least squares estimates.

while controlling for the influence of one or more additional firm characteristics under an addictive regression assumption. However, a conventional portfolio sort suffers from the curse of dimensionality. Third, we are able to obtain the optimal number of basis portfolios that represent the cross-sectional variation in expected stock returns associated with one firm characteristic using B-spline regression. The “optimal" is from the standpoint of maximizing the explanatory power of B-splines for the cross-sectional variation in expected stock returns generated by the characteristic. But we typically do not have any criteria to quantify the question, such as whether it is better to do quintile sorts or decile sorts, using the conventional portfolio sorts technique.

3.2 Portfolio sorts via B-spline regression

3.2.1 B-splines

B-splines are class of polynomial splines constructed from linear combinations of local basis functions. In general, we can represent a B-spline of degree $d \geq 0$ whose domain is $x \in [a, b)$ as

$$S_d(x, d, \kappa) = \sum_{j=1}^{J} \delta_j B_{j,d}(x, \boldsymbol{\kappa}), \tag{3.1}$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_J)'$ is the $J \times 1$ vector of control points, $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{J+d+1})'$ is the $(J + d + 1) \times 1$ vector of knots, and $B_{j,d}(x, \boldsymbol{\kappa})$ is a local basis function of degree d. The knot sequence $\{a = \kappa_1 \leq \kappa_2 \leq ... \leq \kappa_{J+d+1} = b\}$ determines the set of x values for which each local basis function is nonzero. We assume throughout that $\kappa_1 = \kappa_2 = \cdots = \kappa_{d+1}$ and $\kappa_{J+1} = \kappa_{J+2} = \cdots = \kappa_{J+d+1}$ for all $d > 0$. Hence, $S_d(x, d, \kappa)$ has $J - d - 1$ interior (non-endpoint) knots because the initial $d+1$ and final $d+1$ knots consist of repeated values.

The B-spline basis functions for a given choice of d are constructed using the De Boor (1972) algorithm. We start with the set of $J + d$ basis functions of degree zero,

$$B_{j,0}(x, \boldsymbol{\kappa}) = \begin{cases} 1, & \text{if } \kappa_j \leq x < \kappa_{j+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{3.2}$$

and construct the J basis functions of degree d via the recurrence relation

$$B_{j,d}(x, \boldsymbol{\kappa}) = \omega_{j,d}(x, \boldsymbol{\kappa}) B_{j,d-1}(x, \boldsymbol{\kappa}) + (1 - \omega_{j+1,d}(x, \boldsymbol{\kappa})) B_{j+1,d-1}(x, \boldsymbol{\kappa}), \tag{3.3}$$

where

$$\omega_{j,d}(x, \boldsymbol{\kappa}) = \begin{cases} (x - \kappa_j)/(\kappa_{j+d} - \kappa_j), & \text{if } \kappa_j \neq \kappa_{j+d}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

Because $B_{j,0}(x, \boldsymbol{\kappa}) = 0$ for $\kappa_j = \kappa_{j+1}$, the B-spline basis of degree zero forms a partition of unity, i.e., $\sum_{j=1}^{J} B_{j,0}(x, \boldsymbol{\kappa})$ for any $x \in [a, b)$. Thus it follows from the recurrence relation that the B-spline basis functions of any degree $d > 0$ forms a partition of unity, i.e., $\sum_{j=1}^{J} B_{j,d}(x, \boldsymbol{\kappa})$ for any $x \in [a, b)$. In addition, a minimum of $J - d - 1$ of the basis functions are zero for any given value of x. For example, if $x \in [\kappa_j, \kappa_{j+1})$, then all the basis functions other than $B_{j-d,d}(x, \boldsymbol{\kappa}), \ldots, B_{j,d}(x, \boldsymbol{\kappa})$ are zero by construction.

B-splines are piecewise polynomials, the order and the smoothness of which depend on the choice of d. B-splines of degree zero are piecewise constant and discontinuous at the knots, those of degree one are piecewise linear and continuous but not differentiable at the knots, those of degree two are piecewise quadratic and differentiable at the knots, etc.

3.2.2 B-spline regression and estimation

B-spline regression is a well-developed nonparametric estimation technique. It can be regarded as regressing the dependent variable on a set of basis functions of the independent variable. Since we focus on how stock returns are influenced by firm characteristics, the discussion of B-spline regression is formulated using the following notations. Suppose x_n denotes the value of some firm characteristic, such as market capitalization (size), for firm $n \in \{1, 2, \ldots, N\}$. To investigate the cross-sectional relation between firm size and expected stock returns, we could fit a B-spline regression of the from

$$r_n = \sum_{j=1}^{J} \delta_j B_{j,d}(x_n, \boldsymbol{\kappa}) + e_n, \tag{3.5}$$

where r_n is the stock return for firm n. The model does not include an intercept because it would be redundant given that $\sum_{j=1}^{J} B_{j,d}(x_n, \boldsymbol{\kappa}) = 1$ for all n.

The coefficient vector $\boldsymbol{\delta}$ can be estimated using ordinary least squares, weighted least sqaures, or any estimation methods that can be used in the linear regression framework. For example, the OLS estimator of $\boldsymbol{\delta}$ can be expressed as $\hat{\boldsymbol{\delta}} = \boldsymbol{(B'B)^{-1}B'r}$, where $\boldsymbol{B}$ is an $N \times J$ design matrix with nth row $(B_{1,d}(x_n, \boldsymbol{\kappa}), B_{2,d}(x_n, \boldsymbol{\kappa}), \ldots, B_{J,d}(x_n, \boldsymbol{\kappa}))$ and $\boldsymbol{r} = (r_1, r_2, \ldots, r_N)'$ is the $N \times 1$ vector of returns. It is straightforward that $\hat{\boldsymbol{\delta}}$ is a vector of portfolio returns since each element of $\hat{\boldsymbol{\delta}}$ is linear combination of individual stock returns.

Note that the columns of $\boldsymbol{B}$ are mutually orthogonal for $d = 0$. Thus it is easy to see that the jth element of $\hat{\boldsymbol{\delta}}$ for a B-spline of degree zero is

$$\hat{\delta}_j = \frac{\sum_{n=1}^{N} I_{[\kappa_j, \kappa_{j+1})}(x_n) r_n}{\sum_{n=1}^{N} I_{[\kappa_j, \kappa_{j+1})}(x_n)} \tag{3.6}$$

where $I_A x$ denotes the indicator function for $x \in A$. In other words, $\hat{\delta}_j$ is simply the return on an equally-weighted portfolio that contains the stocks of all firms for which $x_n \in [\kappa_j, \kappa_{j+1})$. We can therefore view the B-spline regression as a procedure that sorts the set of stocks used for the analysis into equally-weighted size portfolios.[2]

3.2.3 B-splines vs. basis portfolios

let the fitted B-spline denoted by

$$S_d(x, \hat{\boldsymbol{\delta}}, \boldsymbol{\kappa}) = \sum_{j=1}^{J} \hat{\delta}_j B_{j,d}(x, \boldsymbol{\kappa}), \tag{3.7}$$

as a function of x, which describes the cross-sectional relation between a firm characteristic and expected stock returns. It is easy to see that $S_d(x, \hat{\boldsymbol{\delta}}, \boldsymbol{\kappa})$ is a portfolio return for any x because δ_j is a portfolio return.

Our objective is to pick one point for each non-empty interval $[\kappa_j, \kappa_{j+1})$ as $J - d$ basis portfolios for the characteristics. A natural choice is to use the midpoint of $[\kappa_j, \kappa_{j+1})$, i.e. $(\kappa_j + \kappa_{j+1})/2$, as the point to evaluate the B-spline.

[2] Similarly, we can obtain value-weighted portfolios by using weighted least squares is instead of OLS to fit the regression.

3.2.4 Optimal number of basis portfolios

The advantage of nesting standard univariate portfolio sorts within the spline regression framework is that we bring all of the usual regression tools to bear on the analysis.

Suppose, for instance, that we want to decide whether it is better to do quintile sorts or decile sorts. We can use cross-validation to answer this question. Note that $S_0(x, \boldsymbol{\delta}, \kappa)$ has two endpoint knots (κ_1 =min$(x_1, x_2, \ldots, x_N)$) and $J-1$ interior knots.[3] If we assume for simplicity that all of these knots are fixed for the purposes of cross-validation, then the leave-out-one cross-validation criterion for the $d=0$ case can be expressed as

$$CV(J) = \sum_{n=1}^{N} (r_n - S_0(x_n, \hat{\boldsymbol{\delta}}_{-n}, \boldsymbol{\kappa}))^2 \tag{3.8}$$

where $\hat{\boldsymbol{\delta}}_{-n}$ denotes the OLS estimate of $\boldsymbol{\delta}$ obtained by excluding the data for firm n. In our example, $S_0(x_n, \hat{\boldsymbol{\delta}}_{-n}, \boldsymbol{\kappa})$ is simply the return on the size portfolio constructed by finding the value of j for which $x_n \in [\kappa_j, \kappa_{j+1})$, assigning a weight of zero to the return on stock n, and assigning equal weights to the returns of all other stocks in same size category, i.e., the category defined by $[\kappa_j, \kappa_{j+1})$. The choice of J that delivers the lowest value of $CV(J)$ is optimal from the standpoint of maximizing the explanatory power $S_0(x_n, \hat{\boldsymbol{\delta}}_{-n}, \boldsymbol{\kappa})$ for the cross-section of expected stock returns.

Of course we typically want to construct a time series of size portfolio returns. This can be accomplished by generalizing the cross-validation procedure in an obvious fashion. Let N_t denote the number of stocks for time period $t \in 1, 2, \ldots, T$. Instead of focusing on a single time period, we can formulate the leave-out-one cross-validation criterion as

$$CV(J) = \sum_{t=1}^{T} \sum_{n=1}^{N_t} (r_{n,t} - S_{0,t}(x_{n,t}, \hat{\boldsymbol{\delta}}_{-n,t}, \boldsymbol{\kappa}_t))^2 \tag{3.9}$$

where $r_{n,t}$ is the stock return for firm n in period t, $x_{n,t}$ is the size of firm n in period t, $S_{0,t}(x_{n,t}, \hat{\boldsymbol{\delta}}_{-n,t}, \boldsymbol{\kappa}_t))$ is the B-spline function for the period t regression, and $\hat{\boldsymbol{\delta}}_{-n,t}$ is the OLS estimate of $bm\hat{\delta}_t$ obtained by excluding the data for firm n from the period t regres-

[3]Technically, κ_{J+d+1} should be infinitesimally larger than $\max(x_1, x_2, ..., x_N)$ because we want the last interval $[\kappa_{J+d}, \kappa_{J+d+1})$ to contain the maximum value of x.

sion. Minimizing $CV(J)$ is equivalent to maximizing the explanatory power of the sequence $\{S_{0,t}(x_{n,t}, \boldsymbol{\delta}_t, \boldsymbol{\kappa}_t)\}_{t=1}^{T}$ under the constraint that J is the same for all t. Once J is selected via cross-validation, the time-series of returns for the jth size portfolio is given by $\{\hat{\delta}_{j,t}\}_{t=1}^{T}$.

3.2.5 Controlling for the influence of additional covariates

Suppose we want to investigate the cross-sectional relation between firm size and expected stock returns while controlling for the influence one or more additional covariates, such as the book-to- market ratio, gross profitability, etc. This can be accomplished via a simple two-step procedure. To illustrate, consider the case of a single additional covariate denoted by c_n. The first step is designed to extract the component of size that is unrelated to the additional covariate. We do this by fitting a B-spline regression of the form

$$x_n = \sum_{j=1}^{J} \gamma_j B_{j,d}(c_n, \boldsymbol{\lambda}) + u_n, \tag{3.10}$$

which $B_{j,d}(c_n, \boldsymbol{\lambda})$ is the jth local basis function for the covariate under consideration. Because the residual $\hat{u}_n$ is cross-sectionally uncorrelated with all of the basis functions used to fit the regression, it captures the cross-sectional variation in size that is unrelated to the cross-sectional variation is the covariate. The second step mirrors the approach used for the case in which there are no additional covariates, except that we replace x_n with the residual $\hat{u}_n$. Specifically, we fit a B-spline regression of the form

$$r_n = \sum_{j=1}^{J} \alpha_j B_{j,d}(\hat{u}_n, \boldsymbol{\kappa}) + e_n, \tag{3.11}$$

where $B_{j,d}(\hat{u}_n, \boldsymbol{\kappa})$ is the jth local basis function using the residuals from step one. The optimal value of J for each step can be chosen by cross-validation.

For the general case, we can implement the first step of the procedure using a B-spline regression of the form

$$x_n = \sum_{k=1}^{K} \sum_{j=1}^{J} \gamma_{j,k} B_{j,d}(c_{k,n}, \boldsymbol{\lambda}_{\boldsymbol{k}}) + u_n, \tag{3.12}$$

where K is the total number of additional covariates (characteristics) of interest. Note that this is a B-spline variant of the general class of additive regression models. It is similar to a standard linear regression in the sense that the marginal effect of each covariate is assumed to be additive (i.e., there are no interaction effects). However, it allows the marginal effect of $c_{k,n}$ to vary over the cross-sectional distribution of the covariate.

3.3 Data and Variables

We obtain monthly stock returns from the Center for Research in Security Prices (CRSP) and annual accounting data from the Compustat Annual Industrial file. The sample period starts in July 1963 and ends in December 2018. We consider five firm characteristics: market equity (ME), book-to-market ratio (B/M), gross profitability (GP), investment-to-asset ratio (ITA), and the momentum return (MOM).

In addition, we transform the value of each characteristic for each firm into the normalized firm's cross-sectional rank for each month. Specifically, let $x_{n,t}$ denote the value of some firm characteristic for firm n in month t. Then normalized cross-sectional rank of the firm characteristic for firm n in month t, $\tilde{x}_{n,t}$, is given by

$$\tilde{x}_{n,t} = \frac{x_{n,t}}{(N_t + 1)}, \tag{3.13}$$

where N_t denote the total number of firms in month t.

A rank transformation has the following properties. First, the extreme value effect can be mitigated by using firms' rank to fit regressions. Second, rank transformation is consistent with the basic spirit of portfolio sorts. Third, the values of the rank always lie in between zero and one for each month, which makes easier to aggregate the effect across months.

3.4 Empirical Results

We use quintile sorts as an example to illustrate the basic idea of B-splines estimated based on some firm characteristic without controlling for additional covariates.

Figure 3.1 $\sim$ Figure 3.5 plot the time series average of B-splines estimated by regressing stock returns for month t on firm characteristics for month $t-1$ month by month. Note that the knot sequence is chosen such that it maps the NYSE breakpoints.[4]

[Figure 3.1 $\sim$ Figure 3.5 go about here.]

[4] We use CRSP breakpoints for momentum returns.

We first focus on degree of zero (blue lines). The B-spline regression of degree zero is equivalent to the conventional quintile portfolio sorts as discussed in 3.2.2. In this regard, the plots are consistent with prior studies that investigate the cross-sectional relation between firm characteristics and expected stock returns using portfolio sorts. We find that ME has a negative effect on expected stock returns, whereas B/M, GP, ITA, and MOM have a positive effect. It is also noted that the relation estimated by B-splines of degree zero is monotonic for all firm characteristics examined.

However, we find strong nonlinear relation for ME, B/M, and MOM in case of B-spline regression of degree one and two (red lines and orange lines). More interestingly, the relation is even non-monotonic. Note that we do not control for the influence of any covariates, which will be addressed in the future version of the project as well as the empirical results for the basis portfolios.

References

De Boor, C. (1972). On calculating with b-splines. *Journal of Approximation theory*, 6(1):50–62.

Freyberger, J., Neuhierl, A., and Weber, M. (2019). Dissecting characteristics nonparametrically. *Review of Financial Studies*, forthcoming.

Kirby, C. (2019). Firm characteristics, cross-sectional regression estimates, and intertemporal asset pricing tests. Working Paper.

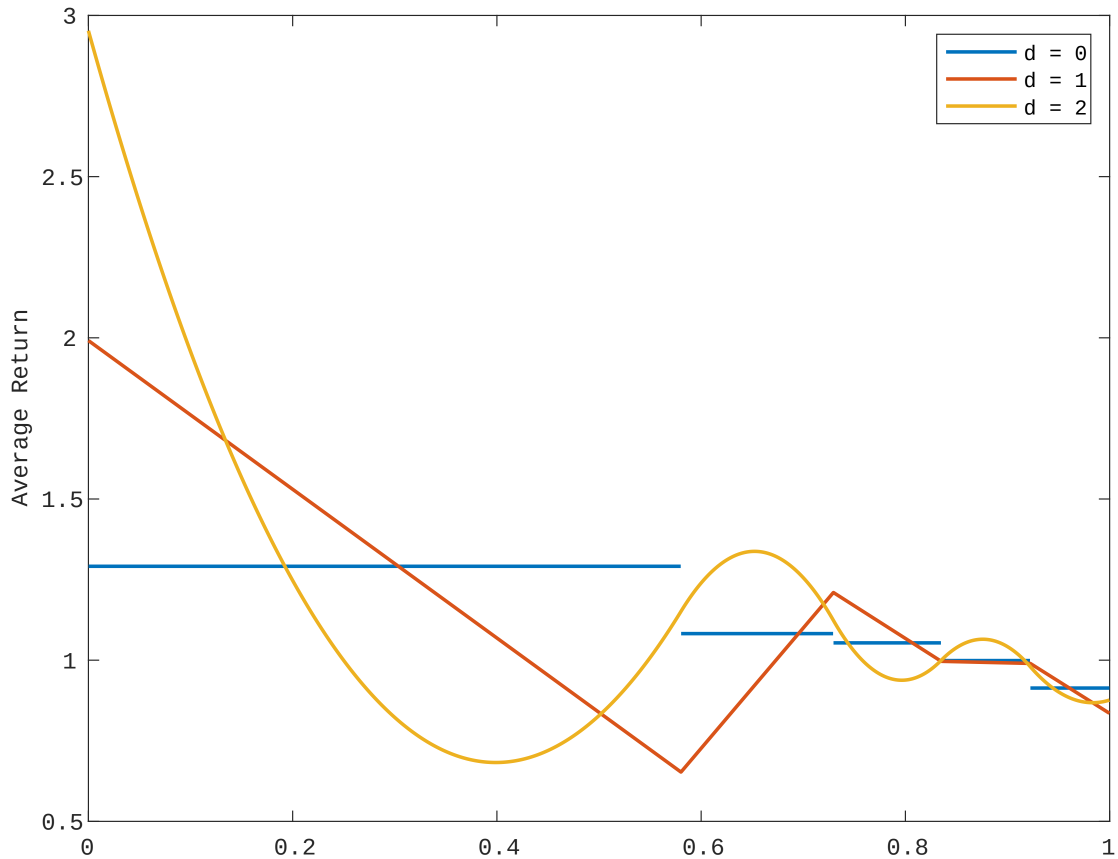

Figure 3.1. Average Return vs Market Equity
The figure plots the time series average of B-splines estimated by regressing stock returns on market equity month by month.

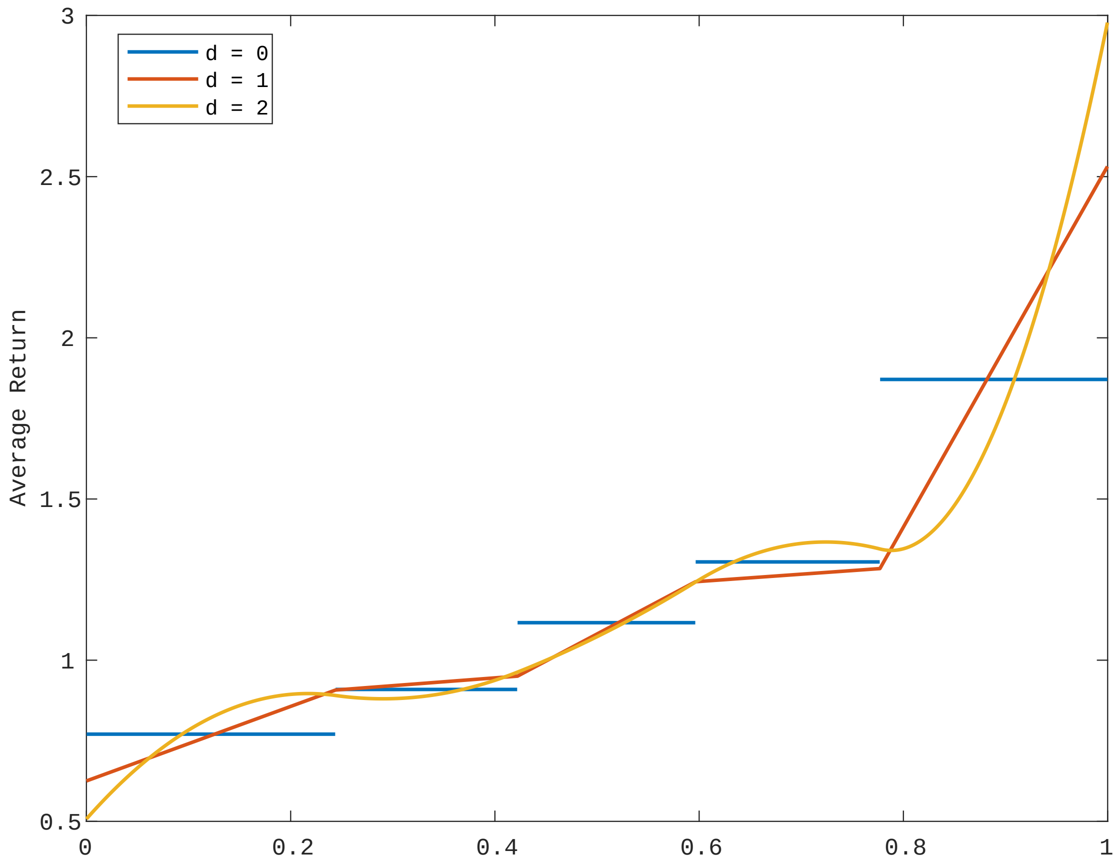

Figure 3.2. Average Return vs B/M Ratio
The figure plots the time series average of B-splines estimated by regressing stock returns on book-to-market ratio month by month.

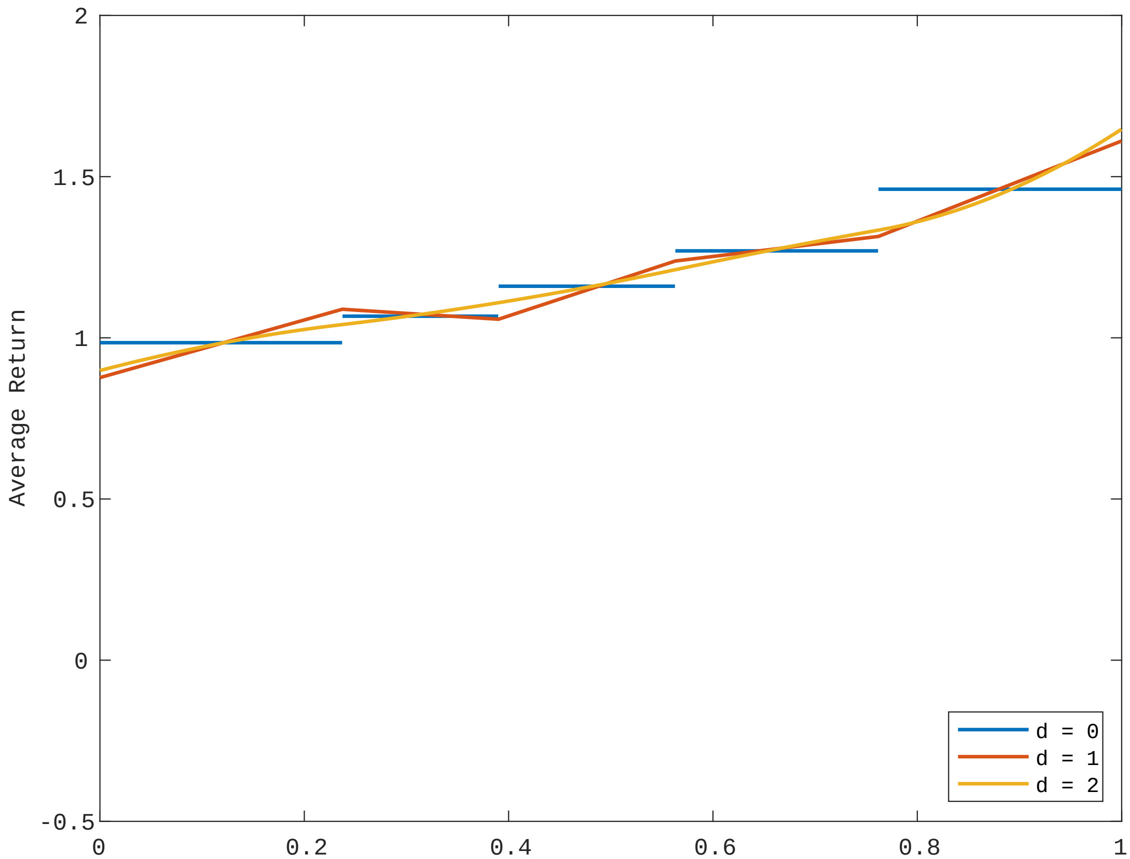

Figure 3.3. Average Return vs Gross Profitability
The figure plots the time series average of B-splines estimated by regressing stock returns on gross profitability month by month.

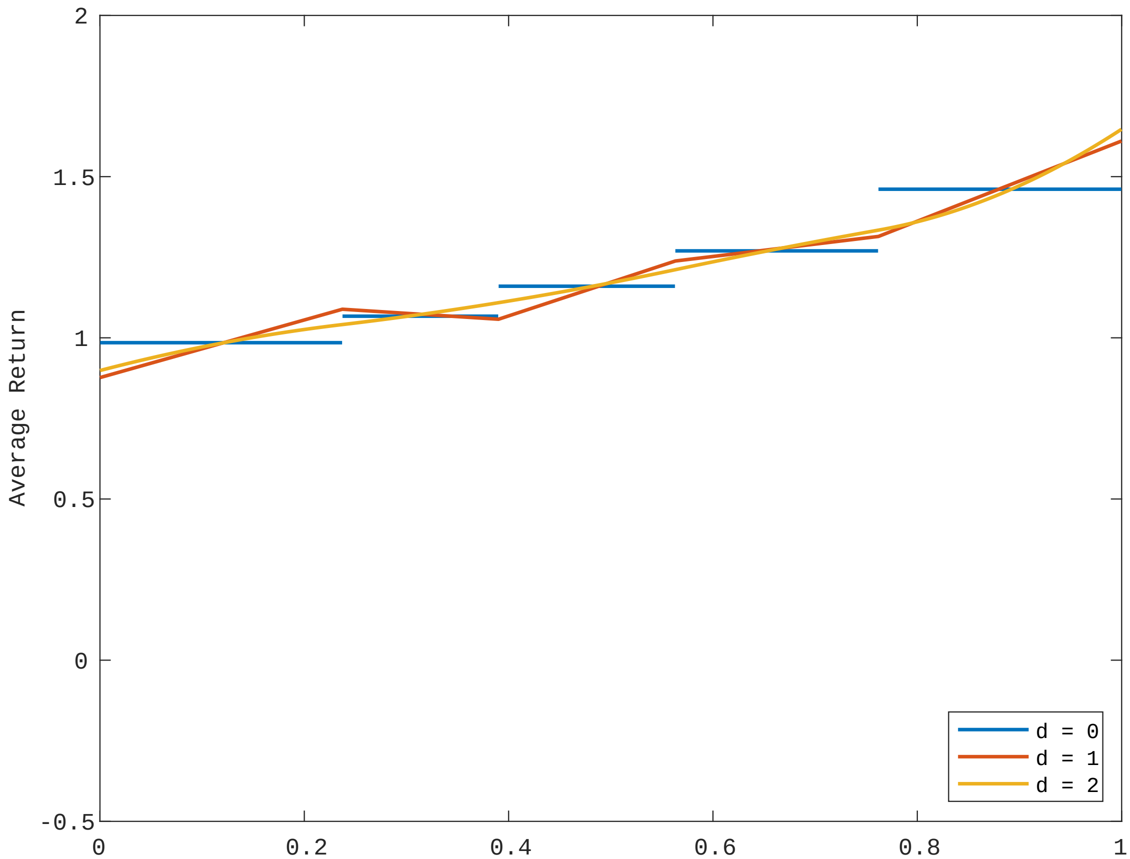

Figure 3.4. Average Return vs Investment to Asset
The figure plots the time series average of B-splines estimated by regressing stock returns on investment to asset ratio month by month.

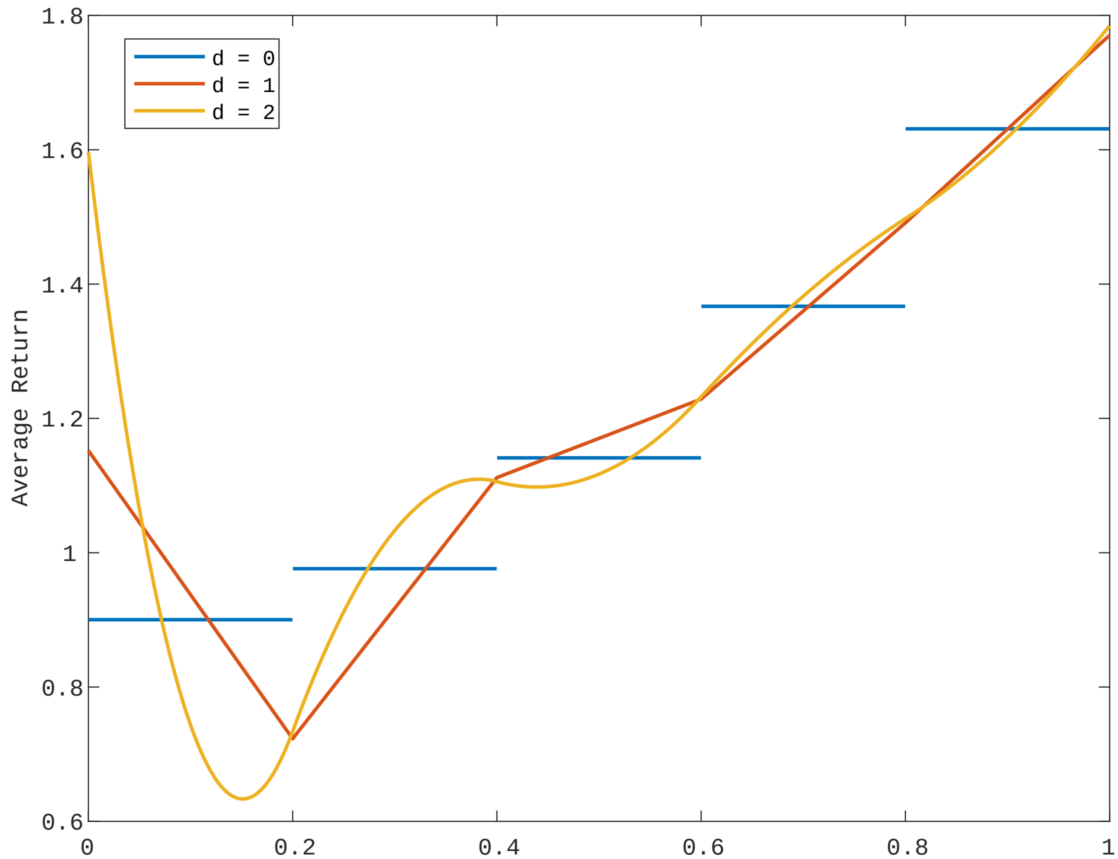

Figure 3.5. Average Return vs Momentum
The figure plots the time series average of B-splines estimated by regressing stock returns on momentum returns month by month.

CONCLUSIONS

Numerous prior studies suggest that a number of firm characteristics generate the cross-sectional variation in expected stock returns. My dissertation takes a closer look at this question from three perspectives.

First, we relate regime switching to the question. Specifically, we employ a two-regime-switching model that allows means and volatility of monthly market returns to depend on the underlying regime that is generated by an irreducible first-order Markov chain with constant transition probabilities. Though the regime is unobserved, good and bad regime probabilities can be estimated from the market model. Combining monthly return observations with regime probabilities, one can easily estimate regime-switching means, volatility, and cross-correlations for a large number of individual stocks. By fitting cross-sectional regressions that includes means, volatility, and cross-correlations in the good and bad regimes as regressors, we find that they have important influence on the cross-section of expected stock returns. The relation between bad-regime means and expected stock returns is significantly negative. The relation between average bad-regime cross-correlations and expected stock returns is significantly positive. The observed relations are consistent with hedging hypothesis that investors want to hedge against market downturns and volatile markets, and thus require low returns for stocks with high returns during market downturns and for stocks that reduce risk when the market is volatile. Furthermore, we construct one-step-ahead predicted returns based on the regime-switching variables and form quintile portfolios on predicted returns.We find evidence that stocks with high (low) predicted returns earn substantially high (low) subsequent returns. The findings reveal the important role of information contained in the bad regime that may be ignored in a setting without regime switching and highlight the cross-sectional predictability of regime switching means, volatility, and cross-correlations on expected returns.

Second, we take interaction between prior trading activity and prior returns into consideration. We find that the likelihood of short-term return reversals is strongly linked to prior trading activity for both small- and large-cap stocks. Stocks with low prior turnover have the strongest tendency to display monthly return reversals. Those with high prior turnover display short-term momentum. We posit that these findings arise from the interplay between short-term price pressure generated by uninformed traders and short-term continuations generated by the actions of speculative traders. By conditioning on a proxy for the fraction of turnover that is driven by news, we show that the predictions of our hypothesis are consistent with the observed negative interaction between prior turnover and reversals in monthly stock returns. The pronounced reversal effect in monthly returns for liquid, low-turnover, large-cap stocks is an intriguing phenomenon that belies the view that short-term reversals are of little economic significance. In addition, the link between the strength of reversal effect and the correlation between squared daily returns and daily turnover is a telling finding. Because the nature of the evidence makes it difficult to envision a plausible explanation for this interaction that does not involve some type of information-based story, our findings clearly raise the bar for research that seeks to explain the short-term reversal anomaly. More broadly, the evidence suggests that market capitalization probably plays an important confounding role in research on short-term reversals. Our analysis reveals that the reversal effect is particularly strong for micro-cap stocks, which account for almost 60% of the available firm-year observations for our sample period. Hence, the presence of these stocks in the dataset tends to mask evidence of short-term momentum. This may be one of the reasons why the short-term momentum effect for high-turnover stocks has gone undetected in prior research. Without adequate controls for the influence of market capitalization, studies of the short-term reversal anomaly are likely to produce incomplete and potentially misleading findings with respect to the autocorrelation properties of individual stock returns.

Third, we construct basis portfolios that capture the cross-sectional variation associated with some firm characteristic using B-spline regression, which is a class of spline regression that is able to capture the potential nonlinearity in relation between firm characteristics and the expected stock returns.

REFERENCES

Campbell, J., Grossman, S., and Wang, J., 1993. Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics* 108, 905–939.

Llorente, G., Michaely, R., Saar, G., Wang, J., 2002. Dynamic volume-return relation of individual stocks. *Review of Financial Studies* 15, 1005–1047.

Tauchen, G., Pitts, M., 1983. The price variability-volume relationship on speculative markets. *Econometrica* 51, 485-505.

Kirby, Chris., 2019. Firm Characteristics, Cross-Sectional Regression Estimates, and Intertemporal Asset Pricing Tests. Working paper.

Freyberger, Joachim., Neuhierl, Andreas., and Weber, Michael., 2019. Dissecting characteristics nonparametrically. *Review of Financial Studies*, forthcoming.