February 17, 2019 Comments Welcome

A Hidden Markov Model of Momentum

Kent Daniel, Ravi Jagannathan, and Soohun Kim†

Abstract

We develop a two-state hidden Markov model where the process driving market returns transitions between turbulent and calm states. A cross-sectional momentum strategy embeds a call option on the market, inducing a state-contingent convex relation between market and momentum returns. In turbulent states, the short side of the momentum strategy has high beta and convexity with respect to the market, as a result of higher effective leverage of the past-loser securities, making momentum crashes more likely. A momentum timing strategy based on this model avoids momentum crashes and achieves superior out-of-sample risk-adjusted performance.

We thank Torben Andersen, Raul Chhabbra, Randolph B. Cohen, Zhi Da, Gangadhar Darbha, Ian Dew-Becker, Francis Diebold, Robert Engel, Bryan T. Kelly, Robert Korajczyk, Tyler Muir, Jonathan Parker, Marco Sammon, Prasanna Tantri, Viktor Todorov, Lu Zhang, and the participants of seminars at the Becker-Friedman Institute Conference honoring Lars Hansen at the University of Chicago, the Fifth Annual Triple Crown Conference in Finance, Fordham University, the Indian School of Business, London Business School, London School of Economics, Nomura Securities, Northwestern University, the Oxford-Man Institute of Quantitative Finance, the Securities and Exchange Board of India, Shanghai Advance Institute for Finance, Shanghai Jiao Tong University, the SoFiE Conference, the University of Virginia, the WFA meetings, the 2016 Value Investing Conference at Western University, and the AFA meetings, for helpful comments on the earlier versions of the paper. Special thanks to Lu Zhang for providing us with time series data on q-factor model. We alone are responsible for any errors and omissions.

†Kent Daniel, Finance and Economics Division, Graduate School of Business, Columbia University and NBER; E-mail: kd2371@columbia.edu,
Ravi Jagannathan, Finance Department, Kellogg School of Management, Northwestern University and NBER, ISB, SAIF; E-mail: rjaganna@kellogg.northwestern.edu,
Soohun Kim, Finance Area, Georgia Institute of Technology; E-mail: soohun.kim@scheller.gatech.edu.

Price momentum can be described as the tendency of securities with relatively high (low) past returns to subsequently outperform (underperform) the broader market. Long-short momentum strategies exploit this pattern by taking a long position in past winners and an offsetting short position in past losers. Momentum strategies have been, and continue to be popular among traders. The majority of quantitative fund managers employ momentum as a component of their overall strategy, and even fundamental managers appear to incorporate momentum when formulating their trading decisions.[1]

Notwithstanding their inherent simplicity, momentum strategies have been profitable across many asset classes and in multiple geographic regions (Asness et al., 2013; Chabot et al., 2014). Over our sample period from 1927:01 to 2017:12 (1,092 months), our baseline momentum strategy produced monthly returns with a mean of 1.17% and a standard deviation of 7.80%, generating an annualized Sharpe ratio of 0.52.[2] Over this same period, the market excess return (Mkt-Rf) had an annualized Sharpe ratio of 0.43. Momentum's CAPM alpha is 1.53%/month ($t = 7.43$).[3]

While the momentum strategy's average risk adjusted return has been high, the strategy has experienced infrequent but large losses. As Chabot et al. (2014) document, this

[1]Swaminathan (2010) shows that most quantitative managers make use of momentum. He further estimates that about one-sixth of the assets under management by active portfolio managers in the U.S. large cap space is managed using quantitative strategies. In addition, Jegadeesh and Titman (1993) motivate their study of price momentum by noting that: "...a majority of the mutual funds examined by Grinblatt and Titman (1989; 1993) show a tendency to buy stocks that have increased in price over the previous quarter." Badrinath and Wahal (2002) show that institutions behave as momentum traders when entering new positions, but as contrarians otherwise.

[2]In our baseline 12-2 momentum strategy, described in more detail later, we rank firms in U.S. based on their cumulative returns from months $t-12$ through $t-2$, and take a long position in the value-weighted portfolio of the stocks in the top decile, and a short position in the value-weighted portfolio of the bottom decile stocks.

[3]Over the period from 1927:01 to 2017:12, Fama and French's (1993) SMB and HML factors had annualized Sharpe Ratios of 0.23 and 0.38, respectively, and the Fama and French three-factor alpha for momentum is 1.75%/month ($t = 8.63$). From 1967:01 to 2016:12, Hou et al.'s (2015) I/A and ROE factors achieved annualized Sharpe ratios of 0.79 and 0.73, respectively, the Hou et al. (2015) four-factor alpha for momentum is 0.43%/month ($t = 1.24$). Lastly, from 1963:07 to 2017:12, the annualized Sharpe ratios of Fama and French's (2015) RMW and CMA factors are 0.39 and 0.50, respectively, and the associated five-factor alpha for momentum is 1.38%/month ($t = 4.37$). The t-statistics are computed using White's (1980) heteroskedasticity-consistent covariance.

phenomenon is robust to the inclusion of pre-CRSP data.[4] The historical distribution of momentum strategy returns is highly left skewed. Consistent with the large estimated negative skewness, over our sample period, there are eight months in which the momentum strategy lost more than 30%, and no month in which it has earned more than 30% (the highest monthly return is 26.16%). Moreover, the strategy's largest losses have been extreme. The worst monthly return was -77.02%, and six monthly losses exceed 40%. Normality can easily be rejected. As Daniel and Moskowitz (2016) document, these large losses cluster, and tend to occur when the market rebounds sharply following a prolonged depressed condition.

To capture these aberrant features of momentum returns, we model momentum return as a nonlinear stochastic function of the market excess return. We show that this helps assess the tail risk in momentum returns. We argue that the way momentum strategy portfolios are constructed necessarily embeds a written call option on the market portfolio, with time varying moneyness. The intuition comes from Merton (1974): following large negative market returns, the effective leverage of the firms on the short side of the momentum strategy (the past-loser firms) becomes extreme. As the firm value falls, the common shares move from being deep in-the-money call options on the firm's underlying assets, to at- or out-of-the-money options, and thus start to exhibit the convex payoff structure associated with call options: the equity value changes little in response to even large down moves in the underlying firm value, but moves up dramatically in response to large up moves. Thus, when the values of the firms in the loser portfolio increase—proxied by positive returns on the market portfolio—the convexity in the option payoff results in outsized gains in the past loser portfolio. Since the momentum portfolio is short these loser firms, this results in dramatic losses for the overall long-short momentum portfolio. We show that in a model economy where firms slowly adjust their leverages towards their target leverages, rare periodic

[4] Using self-collected U.K. historical data from 1867 to 1907 and CRSP data from 1926 to 2012, Chabot et al. (2014) find that, while momentum has earned abnormally high risk-adjusted returns, the strategy also exposed investors to large losses (crashes) during both periods.

momentum crashes occur even when the market prices are determined by rational investors and there are no other financial market frictions.

Interestingly, this same apparent optionality is observed not only in cross-sectional equity momentum strategies, but also in commodity and currency momentum strategies (Daniel and Moskowitz, 2016). Related arguments suggest that effective leverage is likely to be the driver of this same optionality: in the case of commodity momentum, this option-like feature likely arises from the lower bound on variable costs associated with production, the option to shut down, and the lead times involved in adjusting production. In the case of currencies, central bankers tend to lean against the wind and when that effort fails, currency prices tend to crash. Further, those who engage in currency carry trades borrow in the low interest rate currency and lend in the high interest rate currency. When interest rate differentials change, the sudden unwinding of large currency trade positions due to margin calls can lead to large FX momentum strategy losses. For example, in 1997-98, the interest rates on the U.S. Dollar were higher than the interest rates on the Japanese Yen, and the U.S. Dollar was steadily appreciating against the Yen before crashing in October 1998. One explanation for the sharp rise of the Yen against the U.S. Dollar is the drop in U.S. interest rates and the sudden unwinding of Dollar-Yen carry trade positions by hedge funds with weaker capital positions from exposure to 1998 Russian crisis.[5]

For U.S. common stocks, we show that the dynamics of the reported financial leverage are consistent with this hypothesis: going into the five worst momentum crash months, the financial leverage of the loser portfolio averaged 15.39, more than an order of magnitude higher than unconditional average of 1.07.[6] However, a firm's financial leverage is not a good proxy for that firm's *effective* leverage: firms have many fixed costs distinct from the repayment of their debt, including the employees wages, the fixed costs associated with maintenance of property, plant and equipment, etc. If these fixed costs are large, even a firm

[5]See the 69th Annual Report of the Bank for International Settlements, page 107.

[6]These are the averages over the 1964-2017 period for which we have data on the book value of debt.

with zero debt may see its equity start to behave like an out-of-the-money option following large losses. A recent episode that is consistent with the view that non-financial leverage can increase option like features was the collapse of many "dot-com" firms in the 2000-2002 period, where large drops in the values of these firms did not lead to large increases in financial leverage, yet clearly affected the operating leverage of these firms.

Because it is difficult to directly measure the effective leverage—operating plus financial—of the firms that make up the short-side of the momentum portfolio, we instead estimate the leverage dynamics of the momentum portfolio using a hidden Markov model that incorporates this optionality. In the model, we assume that the economy can be viewed as being in one of just two unobserved states: *calm* and *turbulent*. We develop a two-state hidden Markov model (HMM) where the momentum return generating process is different across the two states, and estimate the probability that the economy is in the unobserved turbulent state using maximum likelihood. Our HMM specification can be viewed as a parsimonious dynamic extension of the return generating process in Henriksson and Merton (1981) and Lettau et al. (2014). One striking finding is that, while the momentum returns themselves are highly left-skewed and leptokurtic, the residuals of the momentum return generating process coming out of our estimated HMM specification are Normally distributed.[7] A key component of the HMM specification is the embedded option on the market; by looking for periods in time where the optionality is stronger, we can better estimate whether a momentum "tail event" is more likely. We find that the HMM-based estimate of the turbulent state probability forecasts large momentum strategy losses far better than alternative explanatory variables such as past market and past momentum returns and their realized volatilities or volatility forecasts from a GARCH model.

Interestingly, we find that it is the incorporation of the optionality in the HMM that is key

[7]In contrast, the market-returns residuals have a Student's t-distribution with 5 degrees of freedom. We account for this non-Normality in one our HMM specifications and find that accounting for this non-Normality substantially improves the performance of the model in forecasting tail-events.

to the model's ability to forecast these tail events. A version of the HMM that incorporates all other model components (i.e., the volatilities and mean returns of the both the market and the momentum portfolios), but which does not include the optionality, is not as successful: the model without the optionality produces about 30% more false positives than the baseline HMM specification. This suggests that the historical convexity in the relation between the market and momentum portfolio allows better estimation of the turbulent state probability. Intuitively, increasing leverage in the past loser portfolio, identified by the HMM as an increase in the convexity of the momentum strategy returns, presages future momentum crashes.

There is a vast literature on price momentum. While the focus in this literature has been on documenting and explaining the strategy's high average returns and unconditional risk exposures (Daniel et al., 1998; Barberis et al., 1998; Hong and Stein, 1999; Liu and Zhang, 2008; Han et al., 2009), a more recent literature has focused on characterizing the time variation in the moments. Barroso and Santa-Clara (2015) study the time-varying volatility in momentum strategy returns. Daniel and Moskowitz (2016) find that infrequent large losses to momentum strategy returns are pervasive phenomena. They are present in several international equity markets and commodity markets, and they tend to occur when markets recover sharply from prolonged depressed conditions. Grundy and Martin (2001) examine the time-varying nature of the momentum strategy's exposure to standard systematic risk factors. In contrast to most of this literature, our focus is on the strategy's tail risk. In particular, we show how this tail risk arises, model it with our HMM, estimate this model, and show that it captures these important tail risks well.

Our findings also contribute to the literature characterizing hidden risks in dynamic portfolio strategies and the literature on systemic risk. For example, Mitchell and Pulvino (2001) find that merger arbitrage strategy returns have little volatility and are market neutral during most times. However the strategies effectively embed a written put option on the market, and consequently tend to incur large losses when the market sharply depreciates. When a number of investors follow dynamic strategies that have embedded options on the market of the same type, crashes can be exacerbated and have the potential to trigger systemic responses.

It is well recognized in the literature that payoffs on self-financing zero cost portfolios that have positions in options will in general exhibit spurious positive value (alpha) when alpha is computed using the standard linear market model or linear beta models (Jagannathan and Korajczyk, 1986; Boguth et al., 2011). Using our model for the momentum return generating process that embeds an option on the market, we value the momentum strategy payoffs using the market prices of index options. We find that the alpha of the momentum strategy remains positive and significant after properly accounting for the optionality in the payoffs. However, when the *ex-ante* turbulent state probability is sufficiently high—and there are several historical episodes where it is—the estimated alpha is negative and statistically significant.

The rest of this paper is organized as follows. In Section 1, we examine the various drivers of momentum crashes, and show that these arise as a result of the strong written call option-like feature embedded in momentum strategy returns in certain market conditions. In Section 2, we describe a hidden Markov model for momentum's return generating process that captures this feature of tail risk in momentum strategy returns. In Section 3, we show the ability of our hidden Markov model to predict momentum crashes. In Section 4, we evaluate the conditional alpha of momentum strategy returns based on the estimated parameters of our hidden Markov model and option market prices. We conclude in Section 5.

1 Momentum Crashes

In this section, we describe the return on a particular momentum strategy that we examine in detail in this paper. We show that the distribution of momentum strategy returns is heavily skewed to the left and significantly leptokurtic. We also find that the return on the momentum strategy is non-linearly related to the excess return of the market index portfolio. We show, using a model equilibrium economy, that when firms partially adjust their leverage towards their target leverage, as in Welch (2004), Leary and Roberts (2005), and Flannery and Rangan (2006), and the way momentum portfolios are formed, embeds a written call-like feature in momentum returns. Further, the embedded option like features are accentuated under prolonged depressed market conditions. We use these results to motivate the two-state model that we develop in Section 2.

1.1 Characteristics of Momentum Strategy Returns

Price momentum strategies have been constructed using a variety of metrics. For this study we examine a cross-sectional equity strategy in US common stocks. Our universe consists of all U.S. common stocks in CRSP with sharecodes of 10 and 11 which are traded on the NYSE, AMEX or NASDAQ. We divide this universe into decile portfolios at the beginning of each month t based on each stock's "(12-2)" return: the cumulative return over the 11-month period from months $t-12$ through $t-2$.[8] Our decile portfolio returns are the market-capitalization weighted returns of the stocks in that past return decile. A stock is classified as a "winner" if its (12-2) return would place it in the top 10% of all NYSE stocks, and as a "loser" if its (12-2) return is in the bottom 10%. Most of our analysis will concentrate on the zero-investment portfolio "MOM", which is long the past-winner decile, and short the past-loser decile.

[8] The one month gap between the return measurement period and the portfolio formation date is done both to be consistent with the momentum literature, and to minimize market microstructure effects and to avoid the short-horizon reversal effects documented in Jegadeesh (1990) and Lehmann (1990).

Panel A of Table 1 provides various statistics describing the empirical distribution of the momentum strategy return (MOM) and the three Fama and French (1993) factors.[9] Without risk adjustment, the momentum strategy earns an average return of 1.17%/month and an impressive annualized Sharpe ratio of 0.52. Panels B and C show that after risk adjustment, the average momentum strategy return increases: its CAPM alpha is 1.53%/month (t=7.43) and its Fama and French (1993) three factor model alpha is 1.75%/month (t=8.63).[10] This is not surprising given the negative unconditional exposure of MOM to the three factors.

The focus of our study is the large, asymmetric losses of the momentum strategy: Panel A of Table 1 shows that the MOM returns are highly left-skewed and leptokurtic. The skewness and kurtosis of the MOM returns are -2.34 and 20.43, respectively. In comparison, those of the market excess returns are 0.19 and 10.84. Panel A of Figure 1 illustrates this graphically: we plot the smoothed empirical density for MOM returns (the red line) and a normal density with the same mean and standard deviation (the blue line). The comparison between the two densities reveals that the empirical density of MOM is much more leptokurtic. To highly skewness in tails, we overlay red dots that represent the 24 MOM returns that exceed 20% in absolute value (12 in the left tail and 12 in the right tail). We examine the return characteristics of MOM with respect to market excess returns in Panel B of Figure 1. For a fair comparison, we scale the market excess returns to match the volatility of MOM and denote it as Mkt-Rf*. We plot the empirical density of the scaled market excess returns. The 20 Mkt-Rf* returns that exceed 20% in absolute value (11 in the left tail and 9 in the right tail) are represented by blue dots. We can see the strong left skewness of momentum: the distribution of left tails in MOM is much more left skewed than that of the scaled market returns.

Because one of the objectives of this paper is to show that this skewness is mainly a result of the time-varying non-linear relationship between market and momentum returns

[9]The Mkt-Rf, SMB, and HML return data come from Kenneth French's database.
[10]The t-statistics are computed using White's (1980) heteroskedasticity-consistent covariance.

caused by the time-varying leverage of firms in the loser portfolio. As a way of motivating our model, we next examine the influence of prevailing state variables on market conditions on momentum strategy returns.

To begin, Table 2 describes the prevailed market conditions during the 12 months when the MOM loss exceeded 20%. The first set of columns (R_{MOM}, R^e_{WIN}, R^e_{LOS}, Mkt-Rf) show that the large momentum strategy losses are generally associated with large gains on the past-loser portfolio rather than losses in the past-winner portfolio. During the 12 largest loss months, the loser portfolio earned an average excess return of 49.33% whereas the winner portfolio earned only 8.79%. Interestingly, these loser portfolio gains are associated with large contemporaneous gains in the market portfolio, which earns an average excess return of 18.12% in these months.

The columns of Past Mkt Ret, Past Mkt RV in the table also shows that market return has been strongly negative and volatile in the period leading up to the momentum crashes: the market is down, on average, by more than 34% in the three years leading up to these crashes, and the market volatility is almost three times its normal level in the year leading up to the crash.[11] Given the past losses and high market volatility, it is not surprising that the past-loser portfolio incurred severe losses: the threshold (breakpoint) for a stock to be in the loser portfolio averaged -64.88% in these 12 months, about 2.7 times the average breakpoint. Thus, at the start of the crash months, stocks in the past-loser portfolio are likely very highly levered. The column of BD/DE (book value of debt/market value of equity) of loser portfolio in the table also shows that the average financial leverage, during the five largest loss months after 1964 (when our leverage data starts) is 15.39, more than an order of magnitude higher than the average leverage of the loser portfolio, 1.07.

To summarize, large momentum strategy losses generally occurred in volatile bear markets, when the past-losers have lost a substantial fraction of their market value, and conse-

[11] Realized volatility is computed as the square root of the sum of squared daily returns and expressed as an annualized percentage.

quently have high financial leverage, and probably high operating leverage as well. Thus, following Merton (1974), the equity of these firms behaves like an out-of-the-money call option on the underlying firm value which, in aggregate, is correlated with the market. Consequently when the market recovers sharply, the loser portfolio experiences outsized gains, resulting in the extreme momentum strategy losses we observe.

1.2 Embedded Option-like Features of Momentum Strategy

In this section, we examine the time-variation in the call-option-like feature of momentum strategy returns. This serves as motivation for the two-state HMM model that we develop in Section 2.

1.2.1 Empirical Evidence

We start with the empirical evidence. In particular, we consider the following augmented return generating process, similar to that considered by Henriksson and Merton (1981) and others:[12]

$$R^e_{p,t} = \alpha_p + \beta^0_p R^e_{\mathrm{MKT},t} + \beta^+_p \max\left(R^e_{\mathrm{MKT},t}, 0\right) + \varepsilon_{p,t}, \tag{1}$$

where $R^e_{\mathrm{MKT},t}$ is the market portfolio returns in excess of the risk free return for month t. We note that α_p, the intercept of the regression, is no longer a measure of the strategy's abnormal return, because the option payoff—$\max(R^e_{\mathrm{MKT},t}, 0)$—is not an excess return. We

[12] To our knowledge, Chan (1988) and DeBondt and Thaler (1987) are the first to document that the market beta of a long-short winner-minus-loser portfolio is non-linearly related to the market return, although they do their analysis on the returns of longer-term winners and losers as opposed to the shorter-term winners and losers we examine here. Rouwenhorst (1998) demonstrates the same non-linearity is present for long-short momentum portfolio returns in non-U.S. markets. Daniel and Moskowitz (2016) show that the optionality is time varying, and is particularly pronounced in high volatility down markets, and is driven by the behavior of the short-side (loser) as opposed to the long (winner) side of their momentum portfolio. Moreover, Boguth et al. (2011), building on the results of Jagannathan and Korajczyk (1986) and Glosten and Jagannathan (1994), note that the interpretation of the measures of abnormal performance in Chan (1988), Grundy and Martin (2001), and Rouwenhorst (1998) are biased. Lettau et al. (2014) propose a downside risk capital asset pricing model (DR-CAPM), which they find explains the cross section of returns in many asset classes better.

return to this issue and estimate the abnormal return of the strategy in Section 4. Here, we concentrate on the time-variation in β^+, which is a measure of the exposure of portfolio p to the payoff on a one-month call option on the stock market or, equivalently, a measure of the convexity in the relation between the market return and the momentum strategy return.

To examine this time-variation, we partition the months in our sample into three groups on the basis of three state variables: the cumulative market return during the 36-month period preceding the portfolio formation month; the realized volatility of daily market returns over the previous 12-month period; and the breakpoints of the loser portfolio (i.e., the return over the (12,2) measurement period of the stock at the 10th percentile). Based on each of these state variables, we partition our sample of 1,092 months into three groups: 'High', 'Medium' and 'Low'. The High (Low) group is the set of months when the state variable is in the top (bottom) 20th percentile at the start of that month. The 'Medium' group contains the remaining months (i.e., the middle 60%). We present the results from sorting on the basis of the past 36-month market return in Table 3; the results from sorting on the other two state variables are presented in Table A1 of the Online Appendix.[13]

Panel A of Table 3 presents the estimates of equation (1) for the momentum strategy returns (R_{MOM}) as well as for the returns of the winner and loser portfolios in excess of the risk-free rate (R^e_{WIN} and R^e_{LOS}). Note that the estimated β^+, the exposure to the market call payoff, is significant only when the past 36-month market returns are in 'Low' group: consistent with the leverage hypothesis, the past-loser portfolio has a positive exposure to the market option payoff of 0.66 ($t = 3.34$). That is, it behaves like a call option on the market. The MOM portfolio, which is short the past-losers, thus has a significantly negative β^+. In the 'Medium' and 'High' groups, the β^+ of the MOM returns and of the long- and

[13]Results are similar when we group based on other variables that capture market conditions: the cumulative market return during the 12 month preceding the portfolio formation month; the realized volatility of daily market returns over the previous six months, and the ratio of the book value of long-term debt to the market value of equity (BD/MV) of the loser stock portfolio.

short-sides are smaller in absolute value and are not statistically significantly negative.[14] In the 'Low' group, the $Adj.R^2$ is 49% for MOM returns, as compared to 6% and 5% in the 'Medium' and 'High' groups, respectively, a result of both the higher β^0 and β^+ in the 'Low' group.

Panel C of Table 3 shows that large MOM losses (crashes) are clustered in months when the option-like feature of β^+ is accentuated; 10 out of 12 momentum losses occur during months in the 'Low' group. Table A1 shows that the results are consistent when we group samples by the other state variables: i.e., realized volatility of market over the past 12 months or return breakpoints for stocks to enter the loser portfolio.

The evidence in Panel D of Table 3 suggests that the large negative skewness of the momentum strategy return distribution is mostly due to the embedded written call option on the market. In the 'Low' group of Panel D, the skewness of the momentum strategy returns is -2.26, but after we control for the non-linear exposure to the market through equation (1), the skewness of the residual drops to -0.91. In the 'Medium' and 'High' groups, the negative skewness of the momentum strategy returns is not that strong and it is not significantly reduced after controlling for the embedded written call option on the market. This is consistent with the results in Panel A, where β^+ is not significantly different from zero in the other two groups. The results reported in Table A1 of the Online Appendix are consistent with the results presented here. The large negative skew in momentum returns is due to the embedded written call option that gets accentuated by market conditions.

The key to our model is the optionality. The optionality arises through the embedded option in the past losers. What drives the optionality is the leverage of the firms in the loser portfolio. However, it is important to recognize, as we highlighted in the introduction, that financial leverage alone does not determine the optionality. It is a combination of financial and operating leverage. Our model highlights that a good proxy for the total effective

[14] We note that β^+ of the winner portfolio exhibits interesting patterns. It is negative and significant for winner stocks in the 'Low' group. Understanding why we see these patterns is left for future research.

leverage is the past returns. But, controlling for past return, it is still likely that financial leverage should proxy for variations in total effective leverage. We examine this conjecture here by empirically investigating whether, after controlling for the formation period return, financial leverage still forecasts optionality.

In Table 4 we examine portfolios of *just* the past-losers, that is the 10% of stocks that have the lowest (12-2) return. We do for the shorter time period over which we have financial leverage data. We then sort these stocks on the basis of their financial leverage as of the start of the formation month. Specifically, in Panel A we sort individual stocks in the loser portfolio by financial leverage (BD/ME)[15] We form three groups: top 30%, middle 40%, bottom 30%. We construct three corresponding portfolios by value-weighting stocks in each group. What we find is that the portfolio of high leverage firms has much greater optionality. The point estimate of β^+ for the high leverage (top 30%) portfolio is 0.40, while it is 0.03 for the low leverage (bottom 30%) portfolio. The difference is almost significant at a 5% level ($t = 1.95$).

In Panel B, instead of sorting on leverage, we sort on the past 36-month individual firm returns. For comparability, we use the same shorter time period as in the analysis presented in Panel A. We again form three value-weighted portfolios. We find that this sort is not as successful as the sort on financial leverage in picking up optionality. However, in unreported results, we show that over the 1930:01-2017:12 time period, the past 36-month return has considerable power to forecast optionality.

The results in Tables 3 and 4 are consistent with the view that the leverage dynamics is an important driver of the embedded written call option on the market-like feature of momentum returns, which in turn leads to momentum crashes when the market rises sharply. Note that the leverage consists of both financial and operating leverage. While many tech-sector firms had low financial leverage during the dot-com bubble crash, they had high operating leverage.

[15]The financial leverage is measured by the ratio of the book value of long-term debt (BD) to the market value of equity (ME) of the loser stock portfolio.

For example, financial leverage of losers was low during two episodes of large momentum losses in 2001:01 and 2002:12.[16] However, as can be seen from Table 5, the optionality is large when we estimate the augmented market model return-generating process for momentum returns given by equation (1) for the 36 monthly returns from 2000:01-2002:12—although it is not statistically significant due to the small sample size.

1.2.2 An Equilibrium Model

Next, we construct a model economy where in equilibrium the way momentum portfolios are formed embeds a written call option on the market-like feature in momentum returns. There are N firms in the economy where N is very large. For convenience, we assume that the interest rate is zero and investors are risk-neutral.[17]

Let A_i denote firm i's value at time t. A_i evolves over time, according to the following geometric Brownian motion:

$$\frac{dA_i}{A_i} = \sigma_c dW_c + \sigma_{\varepsilon,i} dW_i, \tag{2}$$

where $\sigma_c dW_c$ denotes the common shock to the returns on the assets of all firms, and $\sigma_{\varepsilon,i} dW_i$ is firm-specific asset return, which is independent across firms. To highlight the importance of cross-sectional heterogeneity of leverage dynamics, we set $\sigma_{\varepsilon,i}$ is equal to σ_ε for all i. That is, firms are identical except for the leverage, which affects the equity returns but not asset returns. Under these assumptions, the firm-level heterogeneity can be summarized by the differences in leverage, $\frac{A_{i,t}-E_{i,t}}{E_{i,t}}$, where E_i is the value of equity.

We value a firm's equity, E_i, using the option pricing formula of Black and Scholes (1973), following Merton (1974). Note that the valuation formula has six parameters: value of underlying asset, risk free rate (which is set to zero), volatility of underlying asset, strike

[16] Refer to Table 2. In 2001:01 (2002:12), the momentum strategy loses -42.03% (-20.12%) and the financial leverage (BD/MV) of loser portfolio was 0.70 (0.97). The average of financial leverage over all available data from 1964 is 1.07.

[17] We can allow the positive risk premium and the non-zero constant interest rate. The qualitative results do not change.

price (face value of debt) and time to maturity (of debt). We assumed that the value of the underlying asset evolves according to (2). Given (2), and the assumption that $\sigma_\varepsilon = \sigma_{\varepsilon,i}$ for all i, the variance of the underlying asset is $\sigma_c^2 + \sigma_\varepsilon^2$. The strike price is initially set to match the long-run leverage of 751 representative firms which were in the S&P500 index at any time during 2010-2018. The time to maturity is set to be five years initially, and debt and equity holders readjust the time to maturity to be five years at the end of each month. We calibrate the remaining parameter, the variances of underlying assets by calibrating σ_c and σ_ε as follows.

To calibrate σ_c and σ_ε, we measure the annual asset return of firm $i = 1, \cdots, 751$ in year y as $(1 + R_{A,i,y}) = \frac{L_{i,y}+E_{i,y}}{L_{i,y-1}+E_{i,y-1}}$, where L is long-term debt and E is market value of equity. Then, the common shock to the asset returns is computed as $R_{c,y} = \frac{1}{N_y}\sum_i R_{A,i,y}$, where N_y is the number of firms in year y. Note that $R_{c,y}$ and $(R_{A,i,y} - R_{c,y})$ are the discrete time analogs of $\sigma_c dW_c$ and $\sigma_{\varepsilon,i} dW_i$ in (2). Hence, we use the continuously computed sample variance of $R_{c,y}$ and $(R_{A,i,y} - R_{c,y})$ for the annualized values of σ_c^2 and σ_ε^2, respectively. We choose the beginning leverage as the average of $\frac{L_{i,y}}{E_{i,y}}$ over i and y, which is denoted by $\left(\frac{L}{E}\right)_0$.

We let t denote a month, $t = 0, 1, \cdots, T$. Our model economy has T months. At month $t = 0$, $A_{i,0}$ is set to unity for all i. The face value of debt is set to a value such that the ratio of the market value of debt to market value of equity equals $\left(\frac{L}{E}\right)_{i,0} = \left(\frac{L}{E}\right)_0$ for all $i = 1, \cdots, 1000$. As time t changes, asset values evolve according to (2) and equity values, $E_{i,t}$, are determined using the Black and Scholes formula. Equity returns are given by $(1 + R_{i,t}) = \frac{E_{i,t}}{E_{i,t-1}}$. At the end of each month t, the leverage ratio is partially adjusted towards $\left(\frac{L}{E}\right)_0$, by modifying the nominal value of debt and extending the time to maturity by one month, as follows:

$$\left(\frac{L}{E}\right)_{i,t,\text{After}} = \left(\frac{L}{E}\right)_0 + \rho\left(\left(\frac{L}{E}\right)_{i,t,\text{Before}} - \left(\frac{L}{E}\right)_0\right),$$

where ρ is a positive constant smaller than and close to 1.[18] This adjustment is self-financing. To calibrate the annualized ρ, we estimate AR(1) parameters from the annual panel data of the 751 firms over 1964-2018 and take the cross-sectional average of the estimated AR(1) parameters. The asset values of firms do not change due to this leverage adjustment process since the asset values of firms are determined by (2). As a result of the readjustment, some equity holders become debt holders and some debt holders become equity holders. This adjustment takes place without impairing their wealth. To avoid excessive leverage, if $\left(\frac{L}{E}\right)_{t,\text{After}}$ is above 500, close to the maximum value of the ratio of long-term debt to equity of the 751 firms over 1964-2018, the firm readjust a capital structure so that the debt to equity ratio equals $\left(\frac{L}{E}\right)_0$.[19] We repeat this procedure for $t = 1, \cdots, 10120$.

This gives $R_{i,t}$ for $i = 1, \cdots, 1000$ and $t = 1, \cdots, 10120$. From this synthetic data constructed using Monte Carlo simulation, we omit the first 120 (burn in) periods. We construct the market index return as the cross-sectional average of $R_{i,t}$. We classify a stock as a winner (loser) if its past twelve-month return places it in the top (bottom) decile. The winner (loser) portfolio is the equal-weighted portfolio of winners (losers). The momentum strategy takes a long position in the winner portfolio and a short position in the loser portfolio.

In what follows, we first show the non-linear relation of momentum strategy returns on the market returns is more pronounced in prolonged depressed market conditions, as in Panel A of Table 3. We define month t to be in the 'Depressed Market' condition when the previous 36 months market returns are in the bottom 30% percentile. We classify each of the 10,000 months as being in 'Depressed Market' or 'Non Depressed Market' condition and estimate the returns of winner/loser/momentum returns using the specification of Henriksson and Merton given in (1). Table 6 reports the estimation results.

Notice that the results in Table 6 are similar to Panel A of Table 3: The non-linearity of momentum strategy return is accentuated in the depressed market condition. These patterns

[18]As long as ρ is more than $0.95^{1/12}$, results do not change qualitatively.
[19]The exact level of maximum value does not affect results qualitatively.

are obtained due to the leverage dynamics. In Table 6, the convexity parameter of MOM returns, β^{+}, is *-0.34* in the depressed market condition but -0.17 in other months. This difference is mostly due to the embedded call option features of loser portfolio in 'Depressed Market' when compared to 'Non Depressed Market', 0.33 vs 0.19.

The results in Table 6 motivate our hidden Markov model (HMM) specification. The use of past 36-month returns on the market to define depressed market conditions is rather arbitrary. To avoid a taking a stance on how many months we should look back, we assume that there exists an unobservable underlying state S_t which governs the non-linearity of momentum returns as well as the volatility of residual shocks:

$$R_{\mathrm{MOM},t} = \alpha\left(S_t\right) + \beta^0\left(S_t\right) R_{\mathrm{MKT},t} + \beta^+\left(S_t\right) \max(R_{\mathrm{MKT},t}, 0) + \sigma_{\mathrm{MOM}}\left(S_t\right) \varepsilon_{\mathrm{MOM},t}$$
$$R_{\mathrm{MKT},t} = \mu\left(S_t\right) + \sigma_{\mathrm{MKT}}\left(S_t\right) \varepsilon_{\mathrm{MKT},t}.$$

We further assume that S_t can take two different values *turbulent* or *calm*. The *turbulent* state is more likely to identify the depressed market condition but the correspondence is not one to one. We estimate hidden Markov model (HMM) model given above. The estimation method is described in details in Section 2.

From the results of Table 7, we see that the coefficient of β^{+} is much larger in *turblent* state (-0.51) than *calm* state (-0.15). Comparison of results reported in Tables 6 and 7 shows the advantages of HMM. While the difference in β^{+} between 'Depressed Market' and 'Non Depressed Market' is 0.15 (=-0.19+0.34) (See Table 6), the difference between 'turbulent' and 'calm' state is stronger, 0.36 (=-0.15+0.51). Hence, HMM specification is better at identifying the time periods when the embedded option-like features of momentum strategy returns are more pronounced.

Furthermore, we find that our HMM estimation is effective in forecasting the large momentum strategy losses in the simulated economy.[20] For this simulation exercise, we de-

[20]In Section 3, we demonstrate similar results from empirical data.

fine the worst hundred returns (left 1% tail) of 10,000 simulated momentum returns over $t = 121, \cdots, 10120$ as momentum crashes. As we will explain in the next section, the HMM estimation yields $\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1})$, which is the probability that the latent state is turbulent given the returns through $t-1$. We sort the 10,000 sample months by the predicted probability at the start of that month and form three groups: top 30%, middle 40%, bottom 30%. Table 8 shows that the 86% of momentum crashes are clustered in top 30% group. To highlight that HMM is capturing effective leverage, which triggers momentum crashes, we do independent sort on loser portfolios financial leverage and the probability of turbulent period (from our HMM specification) and form nine groups: (top 30%, middle 40%, bottom 30%) $\times$ (top 30%, middle 40%, bottom 30%). Note that financial and *effective* leverages are one and the same in the model economy. As expected, we find that momentum crashes are highly concentrated in the top 30% leverage group, and furthermore, most crashes occur when both the predicted probability of being turbulent and the loser portfolio leverage are high. Hence, this simulation evidence is consistent with our view that our HMM succinctly captures the unobserved effective leverage through the non-linear response of momentum returns to the market. In the following section, we describe our HMM specification in more details.

2 The Two-State Hidden Markov Market Model

Motivated by the findings in the previous section, we model the option-like relationship between the market and the momentum portfolio, with the goal of employing this model to forecast momentum crashes. The evidence above suggests that a model based on Merton (1974), using debt and equity values would not capture these periods. Alternatively, we could form a model with a functional form relating the state-variables explored above (past-market returns, market volatility, etc.) to the convexity of momentum returns. This, however, requires choosing the length of time window over which these state-variables are measured, and that necessarily has to be arbitrary. Given these difficulties, we instead posit a two-state model, with "calm" and "turbulent" states. When the economy is in the turbulent state, the option like feature of momentum returns is accentuated, and momentum crashes are more likely. This naturally leads us to the two-state hidden Markov model (HMM) for identifying time periods when momentum crashes are more likely.

In particular, we develop a two-state hidden Markov model (HMM) in which a single state variable summarizes the market conditions. The "turbulent" state is characterized by higher return volatilities and by more convexity in the market-momentum return relationship.

2.1 A Hidden Markov Model of Market and Momentum Returns

Let S_t denote the unobserved underlying state of the economy at time t, which is either "*calm*" (C) or "*turbulent*" (T) in our setting. Our specification for the return generating process of the momentum strategy is as follows:

$$R_{\text{MOM},t} = \alpha(S_t) + \beta^0(S_t) R^e_{\text{MKT},t} + \beta^+(S_t) \max\left(R^e_{\text{MKT},t}, 0\right) + \sigma_{\text{MOM}}(S_t)\, \varepsilon_{\text{MOM},t}, \qquad (3)$$

where $\varepsilon_{\text{MOM},t}$ is an $i.i.d$ random process with zero mean and unit variance. Equation (3) is similar to equation (1). However, the option-like feature, $\beta^+(S_t)$, the sensitivity of momentum strategy return to the market return, $\beta^0(S_t)$, and the volatility of momentum specific shock, $\sigma_{\text{MOM}}(S_t)$, all differ across the unobserved turbulent and calm states of the economy. We also let the intercept, $\alpha(S_t)$, vary across the two hidden states of the economy. We assume that the return generating process of the market return in excess of the risk-free rate is given by:

$$R^e_{\text{MKT},t} = \mu\left(S_t\right) + \sigma_{\text{MKT}}\left(S_t\right) \varepsilon_{\text{MKT},t}, \qquad (4)$$

where $\varepsilon_{\mathrm{MKT},t}$ is an *i.i.d* random process with zero mean and unit variance. That is, $\mu(S_t)$ and $\sigma_{\mathrm{MKT}}(S_t)$ represent the state dependent mean and volatility of the market excess return.

Finally, we assume that the transition of the economy from one hidden state to another is Markovian, with the transition probability matrix given as:

$$\Pi = \begin{bmatrix} \Pr(S_t = C|S_{t-1} = C) & \Pr(S_t = T|S_{t-1} = C) \\ \Pr(S_t = C|S_{t-1} = T) & \Pr(S_t = T|S_{t-1} = T) \end{bmatrix}, \tag{5}$$

where S_t, the unobservable random state at time t which, in our setting, is either *Calm*(C) or *Turbulent*(T) and $\Pr(S_t = s_t|S_{t-1} = s_{t-1})$ denotes the probability of transitioning from state s_{t-1} at time $t-1$ to state s_t at time t.[21]

2.2 Maximum Likelihood Estimation

We now estimate the parameters of the hidden Markov model in equations (3), (4), and (5), which we summarize with the 14-element parameter vector θ^0:

$$\theta^0 = \left\{ \begin{array}{c} \alpha(C), \beta^0(C), \beta^+(C), \sigma_{\mathrm{MOM}}(C), \\ \alpha(T), \beta^0(T), \beta^+(T), \sigma_{\mathrm{MOM}}(T), \\ \mu(C), \sigma_{\mathrm{MKT}}(C), \mu(T), \sigma_{\mathrm{MKT}}(T), \\ \Pr(S_t = C|S_t = C), \Pr(S_t = \mathrm{T}|S_t = \mathrm{T}) \end{array} \right\}. \tag{6}$$

The observable variables are the time series of excess returns on the momentum portfolio and on the market, which we summarize in the vector $\mathbf{R}_t$:

$$\mathbf{R}_t = \left(R_{\mathrm{MOM},t}, R^e_{\mathrm{MKT},t}\right)'.$$

[21] Here, we use $\Pr(x)$ to denote the probability mass of the event x when x is discrete, and the probability density of x when x is continuous.

We let $\mathbf{r}_t$ denote the realized value of $\mathbf{R}_t$.

We follow Hamilton (1989) and estimate the HMM parameters by maximizing the log likelihood of the sample, given distributional assumptions for $\varepsilon_{\mathrm{MOM},t}$ and $\varepsilon_{\mathrm{MKT},t}$, in (3) and (4), respectively. As shown in Appendix A, when the likelihood is misspecified, the ML estimator of θ^0 can be inconsistent. Hence, we choose the distribution of $\varepsilon_{\mathrm{MOM},t}$ and $\varepsilon_{\mathrm{MKT},t}$ so that the unconditional variance, skewness, and kurtosis of momentum and market excess returns implied by our HMM specification are closer to their sample analogues. As we discuss later in more detail, while the momentum returns $R_{\mathrm{MOM},t}$ are highly skewed and leptokurtic, the momentum return residuals ($\varepsilon_{\mathrm{MOM},t}$) appear normally distributed. Interestingly, the market return residual ($\varepsilon_{\mathrm{MKT},t}$) is non-normal—it is better characterized as Student's t-distributed with d.f.=5.

Let $\widehat{\theta}_{\mathrm{ML}}$ denote the vector of HMM parameters that maximizes the log likelihood function of the sample given by:

$$\mathcal{L} = \sum_{t=1}^{T} \log\left(\Pr\left(\mathbf{r}_t | \mathcal{F}_{t-1}\right)\right), \tag{7}$$

where $\mathcal{F}_{t-1}$ denotes the agent's time $t-1$ information set (i.e., all market and momentum excess returns up through time $t-1$).

Given the hidden-state process that governs returns, the time-t element of this equation, which is the likelihood of observing $\mathbf{r}_t$, is:

$$\Pr\left(\mathbf{r}_t | \mathcal{F}_{t-1}\right) = \sum_{s_t \in \{C,T\}} \Pr\left(\mathbf{r}_t, S_t = s_t | \mathcal{F}_{t-1}\right), \tag{8}$$

where the summation is over the two possible values of the unobservable state variable S_t. The joint likelihood inside the summation can be written as:

$$\begin{aligned}\Pr\left(\mathbf{r}_t, S_t = s_t | \mathcal{F}_{t-1}\right) &= \Pr\left(\mathbf{r}_t | S_t = s_t, \mathcal{F}_{t-1}\right) \Pr\left(S_t = s_t | \mathcal{F}_{t-1}\right) \\ &= \Pr\left(\mathbf{r}_t | S_t = s_t\right) \Pr\left(S_t = s_t | \mathcal{F}_{t-1}\right). \end{aligned} \tag{9}$$

The first term in equation (9) is the state dependent likelihood of $\mathbf{r}_t$, which can be computed, given distributional assumptions for $\varepsilon_{\text{MOM},t}$ and $\varepsilon_{\text{MKT},t}$ in (3) and (4), as:

$$\Pr\left(\mathbf{r}_t | S_t = s_t\right) = \Pr\left(\varepsilon_{\text{MOM},t} | S_t = s_t\right) \cdot \Pr\left(\varepsilon_{\text{MKT},t} | S_t = s_t\right),$$

where

$$\begin{aligned}
\varepsilon_{\text{MOM},t} &= \frac{1}{\sigma_{\text{MOM}}\left(s_t\right)}\left(r_{\text{MOM},t} - \alpha\left(s_t\right) - \beta^0\left(s_t\right) r^e_{\text{MKT},t} - \beta^+\left(s_t\right) \max\left(r^e_{\text{MKT},t}, 0\right)\right) \\
\varepsilon_{\text{MKT},t} &= \frac{1}{\sigma_{\text{MKT}}\left(s_t\right)}\left(r^e_{\text{MKT},t} - \mu\left(s_t\right)\right).
\end{aligned}$$

The second term in equation (9) can be written as a function of the time $t-1$ state probabilities as:

$$\begin{aligned}
\Pr\left(S_t = s_t | \mathcal{F}_{t-1}\right) &= \sum_{s_{t-1} \in \{C,T\}} \Pr\left(S_t = s_t, S_{t-1} = s_{t-1} | \mathcal{F}_{t-1}\right) \\
&= \sum_{s_{t-1} \in \{C,T\}} \Pr\left(S_t = s_t | S_{t-1} = s_{t-1}, \mathcal{F}_{t-1}\right) \Pr\left(S_{t-1} = s_{t-1} | \mathcal{F}_{t-1}\right) \\
&= \sum_{s_{t-1} \in \{C,T\}} \Pr\left(S_t = s_t | S_{t-1} = s_{t-1}\right) \Pr\left(S_{t-1} = s_{t-1} | \mathcal{F}_{t-1}\right), \qquad (10)
\end{aligned}$$

where the third equality holds since the transition probabilities depend only on the hidden state. We can compute the expression on the left hand side of equation (10) using the elements of the transition matrix, $\Pr\left(S_t = s_t | S_{t-1} = s_{t-1}\right)$. The right hand side of equation (10)—the conditional state probability $\Pr\left(S_{t-1} = s_{t-1} | \mathcal{F}_{t-1}\right)$—comes from Bayes' rule:

$$\begin{aligned}
\Pr\left(S_t = s_t | \mathcal{F}_t\right) &= \Pr\left(S_t = s_t | \mathbf{r}_t, \mathcal{F}_{t-1}\right) \\
&= \frac{\Pr\left(\mathbf{r}_t, S_t = s_t | \mathcal{F}_{t-1}\right)}{\Pr\left(\mathbf{r}_t | \mathcal{F}_{t-1}\right)}. \qquad (11)
\end{aligned}$$

where the numerator and denominator of equation (11) come from equations (9) and (8), respectively.

Thus, given time 0 state probabilities, we can calculate the conditional state probabilities for all $t \in \{1, 2, \cdots, T\}$. In our estimation, we set $\Pr(S_0 = s_0 | \mathcal{F}_0)$ to their corresponding steady state values implied by the transition matrix.[22] Table 9 gives the Maximum Likelihood parameter estimates and standard errors of the hidden Markov model parameter vector in equation (6) with our assumption that the momentum return residual ($\varepsilon_{\text{MOM},t}$) is drawn from standard normal distribution and the market returns residual is drawn from a Student's t-distribution with d.f.=5, the choice of which will be discussed later. We show that Quasi-Maximum Likelihood estimator of HMM parameters does not need to be consistent when the wrong likelihood is maximized in Appendix A.

The results in Table 9 suggest that our HMM does a good job of picking out two distinct states: Notice that β^+, while still negative in the calm state, is more than twice as large in the turbulent state. Similarly, the estimated momentum and market return volatilities, $\sigma_{\text{MOM}}(S_t)$ and $\sigma_{\text{MKT}}(S_t)$, are more than twice as large in the turbulent state. We see also that the calm state is more persistent than the turbulent state, at least based on point estimates.

An implication of the large $\beta^+(T)$ is that MOM's response to up moves in the market is considerably more negative than the response to down-moves in the market. In the turbulent state, MOM's up market beta is -1.47 (=-0.40-1.07), but its down market beta is only -0.40. The combination of this feature and the higher volatilities means that the left tail risk is high when the hidden state is turbulent.

One rather striking feature of the results in Table 9 is the large differences in the market parameter estimates across the two states. For the calm state, the point estimates of the annualized expected excess return and volatility of the market are, respectively, 13.6%/year and 13.9%/year, giving an annualized Sharpe ratio of 0.98. In contrast, the corresponding estimated parameters for the turbulent state are -6.0%/year and 28.9%/year. We caution

[22]The vector of steady state probabilities is given by the eigenvector of the transition matrix given in equation (5).

the reader that the hidden state is not observed, so these returns are not directly achievable. We also note that these results are consistent with prior evidence on the inverse relationship between market volatility and market Sharpe ratios (Glosten et al., 1993; Breen et al., 1989; Moreira and Muir, 2017).

A natural question that arises is whether our HMM specification is consistent with the highly non-normal momentum return distribution in our sample. We therefore compare the unconditional sample moments of the momentum strategy returns and market excess returns implied by the HMM return generating process with their sample counterparts. For this purpose, we consider the following distributions for ($\varepsilon_{\mathrm{MOM},t}$, $\varepsilon_{\mathrm{MKT},t}$): (Normal, Student's t), (Normal, Normal), (Student's t, Normal), and (Student's t, Student's t). For each of these pairs of distributions, we estimate our HMM model, generate a 1,092 month-long time series of momentum strategy and market excess returns using Monte Carlo simulation, and obtain their first four moments. We repeat this exercise 10,000 times to obtain the distribution of the first four momentums implied by the HMM specification.

Panel A of Table 10 gives the results for our baseline specification of normal ($\varepsilon_{\mathrm{MOM},t}$) and Student's t ($\varepsilon_{\mathrm{MKT},t}$), which is the only case where all of the first four realized moments of momentum strategy returns, over our sample period 1,092 months (1927:01-2017:12), fall inside the 95% confidence interval for the HMM-implied moments.[23] These findings are consistent with the hypothesis that the significant left skewed and leptokurtic sample momentum strategy returns are due to the non-linear exposure to market returns. In contrast, when we use normal distributions for $\varepsilon_{\mathrm{MKT},t}$, the sample skewness of momentum strategy returns lies outside of the 99% (95%) confidence interval of our HMM-implied moments, as can be seen in Panel B (C). Furthermore, the sample kurtosis of market excess returns exceeds the 99.5th percentile value of its HMM-implied distribution. If we use Student's

[23] If we perform the non-parametric test by Kolmogorov-Smirnov on the similarity between the empirical CDFs of realized momentum returns and simulated momentum returns, any of the four distributional assumptions is not rejected with 5% significance level due to the low power of the test. Hence, we examine the distribution for each of the first four moments.

t-distributions for both $\varepsilon_{\text{MOM},t}$ and $\varepsilon_{\text{MKT},t}$, the realized skewness of momentum returns lies outside of the 95% confidence intervals of our HMM-implied moments and the confidence intervals for kurtosis becomes much wider. These findings suggest that $\varepsilon_{\text{MOM},t}$ is drawn from a normal distribution and $\varepsilon_{\text{MKT},t}$ is drawn from a Student's t-distribution with 5 degrees of freedom.

We now proceed to examine the extent to which the estimated state probabilities can forecast the momentum "crashes" observed in our sample.

3 Using the HMM to Predict Momentum Crashes

In this section, we examine the predictability of momentum crashes based on the estimated probability of the economy being in the hidden turbulent state in a given month, $\Pr(S_t = \text{T}|\mathcal{F}_{t-1})$. It is evident from the results in Table 9 that when the hidden state is turbulent, the written call option-like features of the momentum strategy returns become accentuated, and both the momentum strategy and market excess returns become more volatile. Hence, we should expect that the frequency with which extreme momentum strategy losses occur should increase with $\Pr(S_t = \text{T}|\mathcal{F}_{t-1})$.

Figure 2 presents scatter plots of realized momentum strategy returns on the vertical axis against $\Pr(S_t = \text{T}|\mathcal{F}_{t-1})$, the estimated probability that the hidden state is turbulent, on the horizontal axis. Momentum strategy losses exceeding 20% are in red and momentum strategy gains exceeding 20% are in green. Panel A is based on in-sample estimates using the full sample of 1,092 months from 1927:01 to 2017:12. Consistent with results in the preceding section, the large losses, highlighted in red, occur only when the estimated turbulent state probability is high. The large gains (the green dots) are fairly evenly distributed across the different state probabilities.

The analysis reflected in Panel A is in-sample, meaning that the full-sample parameters (i.e., those presented in Table 9) are used to estimate the state probability at each point in time. In Panel B, the turbulent state probability is estimated fully out-of-sample; the parameters are estimated using the same ML procedure, but only up through the month prior to portfolio formation.[24] Also, the first out of sample month is 1984:09, giving us a sufficiently large period over which to estimate the parameters. To further challenge the HMM estimation, we estimate the HMM parameters using only data for the period from 1937:01; this means that the market crash of 1929-1932 and the large momentum crash in 1932 are not included in the parameter estimation period. In Panel B, just as in Panel A, there is strong association between momentum crashes that are worse than -20% (red dots) and high values of the (out-of-sample) estimated turbulent state probability. In contrast, large momentum gains more than 20% (green dots) are dispersed more evenly across high and low values of the estimated state probability.

Table 11 highlights that momentum crashes are not well forecast by the financial leverage of loser portfolio alone. Over the 1964:07-2017:01 period (for which we have financial leverage data) there are only 5 months with momentum losses greater than 20%. We do an independent sort on probability of turbulent state (Top 20% and Bottom 80%) and financial leverage (Top 20% and Bottom 80%). We then give the number of crash months in the 2 by 2 table. All the 5 crashes occur when the underlying state being turbulent is high. However, two out of five occur when the financial leverage is not high. In fact, these two months were in the Dot-com burst. The momentum returns tend to react to market in a quite non-linear manner as shown in Table 5. These findings justify our HMM approach of capturing the *effective* leverage through the dynamic written call option-like feature in momentum strategy returns.

Table 12 presents the number of large negative and large positive momentum strategy

[24] We chose to model the market return distribution as a Student's t-distribution based on the full sample, so in this sense we are not completely out of sample. However, we conjecture that estimating this distribution out of sample would not appreciably change our findings.

returns during months when $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ is above a certain threshold. All twelve momentum crashes, defined as losses exceeding 20%, happen when the $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ is more than 80%. However, only eight out of twelve momentum gains exceeding 20% are found when the $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ is more than 80%, and two out of those large gains happen when the $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ is less than 30%.

Most quantitative fund managers operate with mandates that impose limits on their portfolios' return-volatilities. Barroso and Santa-Clara (2015) demonstrate the benefit of such mandates: when exposure to the momentum strategy is varied over time to keep its volatility constant, the Sharpe ratio significantly improves. A natural question that arises is whether managing the volatility of the portfolio to be within a targeted range is the best way to manage the portfolio's exposure to left tail risk. We add to this literature by focusing on tail risk (i.e., the probability of very large losses). As discussed above, left tail risk is related to the left skewness of returns, and there are no a priori reasons to believe that changes in left skewness move in lock step with changes in the volatility of momentum strategy returns. We therefore let the data speak, by comparing the performance of the two tail risk measures: the volatility of momentum strategy returns (measured either by realized volatility or by GARCH) and the probability of the economy being in a turbulent state computed based on the estimated HMM parameters in predicting momentum crashes.

In Table 13, we compare the number of false positives in predicting momentum crashes across different tail risk measures. The number of false positives of a given tail risk measure is computed as follows. Suppose we classify months in which momentum strategy returns lost more than a threshold X. Let Y denote the lowest value attained by a given tail risk measure during those momentum crash months. For example, consider all months during which the momentum strategy lost more than 20% (X=20%). Among those months, the lowest value, attained by the tail risk measure of $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$, is 84.4% ($Y$=84%). During months when the tail risk measure is above the threshold level of Y, we count the number of months when momentum crashes did not occur and we denote it as the number of false positives. Clearly, the tail risk measure that has the least number of false positives is preferable. We consider different values of threshold X=10%, 20%, 30%, 40%.

In Panel A of Table 13, we use $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ as a tail risk measure. The results in Panel A-1 are from our original HMM model specified in (3), (4) and (5). To emphasize the importance of the option-like feature $\beta^+(S_t)$ in (3), we impose the restriction $\beta^+(S_t) = 0$ and report the associated results in Panel A-2. We find that the inclusion of $\beta^+(S_t)$ substantially improves the performance. For example, when X=20%, the number of false positives with the option-like feature is 124, much smaller than 163 from the model without the option-like feature.

In Panel B, we use various estimates of the volatility of momentum strategy returns as tail risk measures. Specifically, we estimate the volatility of the momentum strategy returns using GARCH (1,1), and realized volatility of daily momentum strategy returns over the previous 3, 6, 12, and 36 months of estimation windows. In Panel C, we use the volatility of the market return estimated using GARCH(1,1) which can be viewed as realized volatility using all past returns, and realized volatility of the daily market return during the preceding 3, 6, 12, and 36 month windows as tail risk measures. In Panel D, we use the market return during the preceding 3, 6, 12 and 36 month windows as tail risk measures. When X=20%, we find that the number of false positives in Panel A-1 is always smaller than other cases in Panels B, C and D. For example, in our 1930:01-2017:12 sample,[25] we find 124 false positives when we use the tail risk measure based on our main specification of HMM. In contrast, if we use the realized volatility of daily momentum strategy returns over the previous six months,the number of false positives increases to 210 months.

We further examine the interaction between our risk measure $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ and other risk measures in Table 13. To this end, Figure 3 provides three-dimensional scatter plots with

[25]Since we utilize momentum returns over the previous 36 months to construct risk measures, the sample period becomes shorter.

momentum return on the vertical axis over the plane of our risk measure $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ and an alternative risk measure. Because we find similar patterns for all alternative risk measures in Table 13, we pick two measures as examples. For the left scatter plot, we use the realized volatility of daily momentum returns over the previous six months as an alternative risk measure. For the right, we use the past 36-month market returns. Momentum strategy losses exceeding 20% are in red and momentum strategy gains exceeding 20% are in green. The scatter plots illustrate that once you control the level of $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$, alternative measures do not help in predicting large momentum losses.

Note that the results in Table 13 and Figure 3 do not necessarily imply that our HMM is better than other approaches in every dimension of momentum risk management. To highlight this, we construct a simple momentum timing strategy using $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ and compare it to other strategies. Let $\overline{\Pr}$ denote the maximum level of $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ such that all momentum strategy losses exceeding 20% are captured. Then, we enter (exit) momentum strategy if $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1}) < (>) \overline{\Pr}$. We call this modified momentum strategy as HMM-MOM. In comparison, we consider two momentum risk management strategies, motivated by Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016). One uses $k/\sigma_{\mathrm{MOM},t}$ and the other $k/\sigma^2_{\mathrm{MOM},t}$ for the exposure to the momentum. We compute $\sigma^2_{\mathrm{MOM},t}$ as the annualized average of squared daily momentum returns over the previous six months. We adjust k so that the average exposure to the momentum strategy becomes one. Table 14 reports the mean, standard deviation, annual Sharpe ratio, and annual CAPM appraisal ratio of four strategies: MOM, HMM-MOM, $k/\sigma_{MOM,t}$, and $k/\sigma^2_{MOM,t}$. The performance of using $k/\sigma_{\mathrm{MOM},t}$ or $k/\sigma^2_{\mathrm{MOM},t}$ for momentum exposures is superior to our simple HMM-MOM strategy in terms of Sharpe ratio or CAPM appraisal ratio.

This establishes the link between the tail risk of the momentum strategy returns and the probability of the economy being in the hidden turbulent state. In the next section we examine how the alpha of the momentum strategy return varies over time as the probability of the economy being in the turbulent state changes.

4 Momentum's Option Adjusted Alpha

We have shown that the two-state HMM effectively identifies changes in the market environment that lead to dramatic shifts in the distribution of market and momentum returns. Moreover, even when estimated out-of-sample, the HMM does a far more effective job of forecasting momentum tail events or "crashes" than alternative methods.

In this section, we examine how the alpha of the momentum strategy varies over time with changes in market conditions. Based on the HMM model from Section 2, we calculate the alpha from the perspective of an investor who can freely invest in the risk free asset, the market index portfolio, and at-the-money call options on the market index portfolio without any frictions, but whose pricing kernel is otherwise uncorrelated with innovations in the momentum strategy. Given this assumption, our valuation requires the prices of traded options on the market portfolio, which we proxy with one month, at-the-money index options on the S&P 500. Our derivation for a pricing equation follows the framework in Hansen and Jagannathan (1991) and Glosten and Jagannathan (1994).

Specifically, we assume that how the investor values payoffs on risky assets has the following stochastic discount factor representation. Let M_t denote the stochastic discount factor, and $\mathcal{F}_{t-1}$ denote the investor's information set at time t–1. Since the investor has frictionless access to the risk free asset, the market portfolio, and call options on the market portfolio, the followings relations hold:

$$\begin{aligned} 1 &= \mathbb{E}\left[M_t(1+R_{\mathrm{f},t})|\mathcal{F}_{t-1}\right] \\ 0 &= \mathbb{E}\left[M_t R^e_{\mathrm{MKT},t}|\mathcal{F}_{t-1}\right] \\ V_{c,t\text{-}1} &= \mathbb{E}\left[M_t \max\left(R^e_{\mathrm{MKT},t},0\right)|\mathcal{F}_{t-1}\right], \end{aligned}$$

where $R_{\mathrm{f},t}$ is the risk free rate from $t-1$ to t and $V_{c,t-1}$ is the market price of the call option, which pays $\max\left(R^e_{\mathrm{MKT},t}, 0\right)$ at the end of time t.

We regress M_t, which is based on a constant, the market excess return, and the payoff of a call option on the market based on the information set $\mathcal{F}_{t-1}$. Let $\widetilde{M}_t$ be the fitted part of M_t and $\widetilde{e}_t$ be the residual in that conditional regression. Then we can write M_t as follows:

$$M_t = \widetilde{M}_t + \widetilde{e}_t \tag{12}$$

where

$$\widetilde{M}_t = \lambda_{0,t-1} + \lambda_{1,t-1} R^e_{\mathrm{MKT},t} + \lambda_{2,t-1}\left(R^e_{\mathrm{MKT},t}, 0\right) \tag{13}$$

$$\mathbb{E}\left[\widetilde{e}_t|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[R^e_{\mathrm{MKT},t}\widetilde{e}_t|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\max\left(R^e_{\mathrm{MKT},t}, 0\right)\widetilde{e}_t|\mathcal{F}_{t-1}\right] = 0. \tag{14}$$

The residual $\widetilde{e}_t$ represents the risk that the investor cares about that is not an affine function of the risk free return, market excess return, and the payoff of the call option on the market excess return. $\widetilde{M}_t$ is the dynamic analogue of the stochastic discount factor implied by the downside risk capital asset pricing model proposed by Lettau et al. (2014).

In a similar manner, we regress the momentum strategy return on a constant, the market excess return, and the call option payoff on the market given the information set $\mathcal{F}_{t-1}$. Recall that when the hidden state S_t is turbulent, which occurs with the probability of $\Pr\left(S_t = T|\mathcal{F}_{t-1}\right)$, the momentum strategy return and market excess return generating processes are given by equation (3), where S_t is either calm or turbulent, and where $\varepsilon_{\mathrm{MOM},t}$ and $\varepsilon_{\mathrm{MKT},t}$ are assumed to be drawn from standard normal and Student's t-distributions, respectively.

We consider the following conditional regression given the information set $\mathcal{F}_{t-1}$, which includes the risk free return and the price of the call option on the market:

$$R_{\mathrm{MOM},t} = \alpha_{t-1} + \beta^0_{t-1} R^e_{\mathrm{MKT},t} + \beta^+_{t-1} \max\left(R^e_{\mathrm{MKT},t}, 0\right) + \epsilon_{\mathrm{MOM},t}, \tag{15}$$

where

$$\mathbb{E}\left[\epsilon_{\mathrm{MOM},t}|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\epsilon_{\mathrm{MOM},t} R^e_{\mathrm{MKT},t}|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\epsilon_{\mathrm{MOM},t} \max\left(R^e_{\mathrm{MKT},t}, 0\right)|\mathcal{F}_{t-1}\right] = 0.$$

Specifically, the vector of regression coefficients $\left[\alpha_{t-1} \;\; \beta^0_{t-1} \;\; \beta^+_{t-1}\right]'$ is determined as

$$\left[\alpha_{t-1} \;\; \beta^0_{t-1} \;\; \beta^+_{t-1}\right]' = \left(\mathbb{E}\left[\mathbf{x}_t \mathbf{x}_t' | \mathcal{F}_{t-1}\right]\right)^{-1} \mathbb{E}\left[\mathbf{x}_t R_{\mathrm{MOM},t} | \mathcal{F}_{t-1}\right],$$

where $\mathbf{x}_t = \left[1 \;\; R^e_{\mathrm{MKT},t} \;\; \max\left(R^e_{\mathrm{MKT},t}, 0\right)\right]'$, and

$$\begin{aligned}
\mathbb{E}\left[\mathbf{x}_t \mathbf{x}_t' | \mathcal{F}_{t-1}\right] &= \Pr\left(S_t = C | \mathcal{F}_{t-1}\right) \mathbb{E}\left[\mathbf{x}_t \mathbf{x}_t' | S_t = C\right] + \Pr\left(S_t = T | \mathcal{F}_{t-1}\right) \mathbb{E}\left[\mathbf{x}_t \mathbf{x}_t' | S_t = T\right] \\
\mathbb{E}\left[\mathbf{x}_t R_{\mathrm{MOM},t} | \mathcal{F}_{t-1}\right] &= \Pr\left(S_t = C | \mathcal{F}_{t-1}\right) \mathbb{E}\left[\mathbf{x}_t R_{\mathrm{MOM},t} | S_t = C\right] \\
&\quad + \Pr\left(S_t = T | \mathcal{F}_{t-1}\right) \mathbb{E}\left[\mathbf{x}_t R_{\mathrm{MOM},t} | S_t = T\right].
\end{aligned}$$

Furthermore, the regression equation of (15) can be expressed in terms of excess returns as follows:

$$R_{\mathrm{MOM},t} = \alpha^*_{t-1} + \beta^0_{t-1} R^e_{\mathrm{MKT},t} + \beta^+_{t-1} V_{c,t-1} \left(\frac{\max\left(R^e_{\mathrm{MKT},t}, 0\right)}{V_{c,t-1}} - (1 + R_{\mathrm{f},t}) \right) + \epsilon_{\mathrm{MOM},t}, \tag{16}$$

where the quantity in parenthesis is the excess return on one-period call option on the market.[26]

$$\alpha^*_{t-1} = \alpha_{t-1} + (1 + R_{\mathrm{f},t}) \beta^+_{t-1} V_{c,t-1}. \tag{17}$$

[26]The strike price of the option is the level of the market index times $(1 + R_{f,t})$, which means that the option will be at-the-money at expiration if $R^e_{\mathrm{MKT},t} = 0$.

We denote α^*_{t-1} as the option adjusted alpha of the momentum strategy return. When Assumption 1 holds, $\frac{\alpha^*_{t-1}}{1+R_{\mathrm{f},t}}$ gives the marginal value of the momentum strategy return from the perspective of the marginal investor.

Assumption 1. $\mathbb{E}\left[\widetilde{e}_t \epsilon_{\mathrm{MOM},t}|\mathcal{F}_{t-1}\right] = 0$, *where* $\widetilde{e}_t$ *and* $\epsilon_{\mathrm{MOM},t}$ *are given in equations (12) and (16), respectively.*

This assumption represents that an investor's residual component in SDF, $\widetilde{e}_t$, is not correlated with the residual returns in the momentum returns, $\epsilon_{\mathrm{MOM},t}$. With Assumption 1, Proposition 1 holds.

Proposition 1. *The value of momentum strategy return to the investor whose stochastic discount factor is* M_t*, is* $\frac{\alpha^*_{t-1}}{1+R_{\mathrm{f},t}}$.

Proof.

$$
\begin{aligned}
& \mathbb{E}\left[M_t R_{\mathrm{MOM},t}|\mathcal{F}_{t-1}\right] \\
= \; & \alpha^*_{t-1}\mathbb{E}\left[M_t|\mathcal{F}_{t-1}\right] + \beta^0_{t-1}\mathbb{E}\left[M_t R^e_{\mathrm{MKT},t}|\mathcal{F}_{t-1}\right] \\
& +\beta^+_{t-1}V_{c,t-1}\left(\frac{\mathbb{E}\left[M_t \max\left(R^e_{\mathrm{MKT},t},0\right)|\mathcal{F}_{t-1}\right]}{V_{c,t-1}} - (1+R_{\mathrm{f},t})\mathbb{E}\left[M_t|\mathcal{F}_{t-1}\right]\right) \\
& +\mathbb{E}\left[M_t \epsilon_{\mathrm{MOM},t}|\mathcal{F}_{t-1}\right] \\
= \; & \frac{\alpha^*_{t-1}}{1+R_{\mathrm{f},t}} + \mathbb{E}\left[\left(\widetilde{M}_t + \widetilde{e}_t\right)\epsilon_{\mathrm{MOM},t}|\mathcal{F}_{t-1}\right] = \frac{\alpha^*_{t-1}}{1+R_{\mathrm{f},t}} + \mathbb{E}\left[\widetilde{e}_t \epsilon_{\mathrm{MOM},t}|\mathcal{F}_{t-1}\right] \\
= \; & \frac{\alpha^*_{t-1}}{1+R_{\mathrm{f},t}},
\end{aligned}
$$

where the first equality follows from equation (16). The second equality follows from the assumption that the investor, whose stochastic discount factor is M_t, agrees with the market prices of the risk free asset, market excess return, the call option payoff and the decomposition in equation (12). The third equality follows from equation (13) and the properties of the conditional regression residual $\epsilon_{\mathrm{MOM},t}$. The last equality follows from Assumption 1. □

We next compute the time series of the estimated option adjusted alpha, α^*, in (17) based on the time series of risk-free returns and the prices of call options. In Figure 4, we plot the time series of α^* calculated based on the estimated HMM model in an out-of-sample manner for the sample period 1996:01-2016:04. Notice that the sample average of the α^*_{t-1}'s is 1.37%/month, which is significantly positive. However, α^*_{t-1} is negative during 1998:10 (Russian crisis), 2008:10-2008:12, and 2009:02-2009:04 (financial crisis), which are months when option prices were high and the market was more likely to be in the hidden turbulent state.

We compute the confidence intervals for the estimated option adjusted alpha for each month t for our sample period 1996:01-2016:04 as follows. First, we estimate the parameters of our HMM model using data up to month $t-1$ using ML method. Next, we simulate 10,000 sets of parameters from the asymptotic distributions obtained from the ML estimator. Then, for each set of parameters, we predict the probability that the hidden state is turbulent based on the realized market excess returns and momentum strategy returns in month t. With the simulated parameters, the estimated probabilities, and the risk-free returns and one-month at-the-money call option prices at the end of the month $t-1$, we construct α^*'s for each month t over the period 1996:01-2016:04. Finally, we find the 95% confidence intervals of α^* by choosing the top and bottom 2.5% quantiles from the simulated 10,000 α^* in each month. The 95% confidence intervals in Figure 4 are computed in this manner. It shows that in 209 of the 244 months in the sample period 1996:01-2016:04, the option-adjusted alpha is significantly positive. While the option adjusted alpha is negative during seven months, only during two months – both occur during the recent financial crisis period 2008:12 and 2009:03, are they statistically significantly different from zero.

We provide evidence that the estimated option-adjusted alpha is strongly associated with large losses in momentum strategy returns. We classify the 244 sample months into two groups - one group of months with significantly positive option-adjusted alphas at the five percent significance level and another group of months without. Then, we examine whether the empirical frequency of large losses in momentum strategy returns significantly differs across the two groups using the chi-square test (Pearson, 1900). Table 15 reports the number of months in the two-way table for each threshold level of -10%, -20%, and -30% per month, as well as the chi-square test statistics and associated p-values. For each level of the three thresholds, the chi-square test result confirms that the option-adjusted alpha gives an active signal of the future occurrence of momentum crashes.

Besides, an investor can enhance the performance of his/her momentum based portfolio by exploiting the option-adjusted alpha. As shown in Table 15, whether the option-adjusted alpha of the momentum strategy returns is significantly positive at the 5% percent significance level or not is strongly associated with the occurrence of large losses in momentum strategy returns. Thus, we consider a strategy of holding the momentum portfolio only during months with significant alpha and switching into treasury bonds in other months. Furthermore, because the momentum crashes tend to coincide with the massive market upswings as documented in Table 2, we also investigate another strategy of holding the momentum portfolio over the months with significant alpha and switching into the market portfolio in other months. Table 16 reports the annualized summary statistics of the returns from the proposed momentum timing strategies and those of the conventional momentum strategy returns over the 244 months in the sample period 1996:01-2016:04. Both momentum timing strategies significantly reduce the standard deviations while the average returns are enhanced substantially, more than doubling the annual Sharpe ratios relative to that from holding the momentum strategy returns throughout.

Lastly, to assess the reasonableness of Assumption 1, we construct the time series of the residuals, $\epsilon_{\mathrm{MOM},t}$ in equation (16), based on the estimated parameter values as follows.

$$\epsilon_{\mathrm{MOM},t} = R_{\mathrm{MOM},t} - \alpha^*_{t-1} - \beta^0_{t-1} R^e_{\mathrm{MKT},t} - \beta^+_{t-1} V_{c,t\text{-}1} \left(\frac{\max\left(R^e_{\mathrm{MKT},t}, 0\right)}{V_{c,t-1}} - (1 + R_{\mathrm{f},t}) \right).$$

We regress the residual on commonly used economy-wide risk factors in the literature: the three factors of market excess returns (MKT), small minus big size (SMB), high minus low book to market (HML) in Fama and French (1993); robust minus weak (RMW) and conservative minus aggressive (CMA) factor in Fama and French (2015); investment to assets (I/A) and return on equity (ROE) factor in Hou et al. (2015); quality minus junk (QMJ) factor in Asness et al. (2014); liquidity risk factor (LIQ) in Pastor and Stambaugh (2003); funding liquidity risk factor (FLS) in Chen and Lu (2015); betting against beta (BAB) risk factor in Frazzini and Pedersen (2014); changes in the 3-Month LIBOR (LIBOR), Term Spread (the yield spread between the 10-year Treasury bond and the 3-month T-bill, TERM), Credit Spread (the yield spread between Moody's BAA bond and AAA bond, CREDIT), and the TED Spread (the yield spread between the 3-month LIBOR and 3-month T-bill, TED); and returns of variance swap (VAR-SWAP) across different horizons (Dew-Becker et al., 2015); and the changes in VIX as well as the changes in left jump variations (LJV) embedded in option prices measured by Bollerslev et al. (2015).[27] Specifically, we estimate the following regression equation

$$\epsilon_{\text{MOM},t} = \text{intercept} + \text{coeff} \times \text{systematic factor}_t + e_t$$

and report coeff (t-stat) and R^2 in Table 17. Except for the HML and ROE factors, we do not find any significant correlation between the computed residuals and the systematic risk factors.[28] These findings suggest the need for using dynamic versions of the Fama

[27] We obtain MKT, SMB, HML, CMA and RMW from Ken French's data library: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`; QMJ and BAB data come from Andrea Frazzini's library: `http://www.econ.yale.edu/~af227/data_library.htm`; LIQ from Lubos Pastor: `http://faculty.chicagobooth.edu/lubos.pastor/research/liq_data_1962_2017.txt`; LIBOR, TERM, CREDIT, and TED from FRED: `https://research.stlouisfed.org/fred2/` and VIX from the CBOE: `http://www.cboe.com/micro/vix/historical.aspx`. Finally, we thank Zhuo Chen, Ian Dew-Becker, and Grant Thomas Clayton for sharing FLS, VAR-SWAP, and LJV, respectively, and Lu Zhang for supplying the I/A and ROE data.

[28] The momentum strategy returns have a Fama and French (1993) three-factor alpha of 1.28%/month ($t = 2.22$) which is significantly positive and different from zero at 95% significance level during the period 1996:01-2016:04. Interestingly, the Hou et al. (2015) four-factor alpha is only 0.15%/month – which is not

and French (1993) three-factor model and the Hou et al. (2015) four-factor model to value momentum strategy returns in any given month based on all available information. We leave such extensions for future research.

5 Conclusion

In this paper, we propose the return generating process of momentum strategy returns as a non-linear market model in which parameters vary across two hidden states, *calm* and *turbulent*. Our model can be viewed as a parsimonious dynamic extension of Henriksson and Merton (1981) and Lettau et al. (2014), addressing the findings in the literature that momentum returns are stronger during some periods and weaker in others, with rare but severe crashes that are more likely under prolonged depressed and volatile market conditions.

There is a vast literature documenting that the simple strategy of buying past winners and selling past losers, commonly referred to as a momentum strategy, generates high average risk adjusted returns. However, such a strategy also experiences infrequent but large losses. We provide an explanation for the phenomenon, *i.e.*, why we see such large losses occurring at periodic but infrequent intervals. We show that the way momentum portfolios are formed embeds features that resemble a written call option on the market portfolio into the momentum strategy returns. These features become accentuated in prolonged bear markets, when the market is volatile due to increased financial and operating leverage. These dynamics lead to large momentum strategy losses when the market recovers.

The intuition for the optionality follows Merton (1974): after large negative market returns, the effective leverage of the firms on the short side of the momentum strategy (the past-loser firms) becomes extreme. As the firm values fall, driven both by the market and other factors, the common shares move from being deep in-the-money call options on the

statistically significantly different from zero during this sample period. The four-factor model by Hou et al. (2015) seems to capture momentum strategy return risk profile well during this sample period.

firm's underlying assets to at- or even out-of-the-money options. Thus, the firm value starts to exhibit the convex payoff structure associated with call options: the equity value changes little in response to even large down moves in the underlying firm value, but moves up dramatically in response to up moves. Therefore, when the values of the firms in the loser portfolio increase, as proxied by positive returns on the market portfolio, the convexity in the option payoff results in outsized gains in the past loser portfolio. Since the momentum portfolio is short these loser firms, this results in dramatic losses for the overall long-short momentum portfolio.

High leverage of the past-loser portfolio is the driver of the tail-risk of the momentum strategies in our model. We show that, just preceding the five worst momentum crash months in the 1964-2017 period, the average financial leverage of the past loser portfolio was 15.39, compared to an unconditional average over this period of 1.07. However *effective leverage*—both financial and operating leverage effect—drives the optionality of the past loser portfolio, and effective leverage is difficult to measure. Consequently, we estimate the time-varying tail risk of momentum strategies using the two-state hidden Markov model (HMM) where the embedded option-like features of momentum strategy returns become accentuated in the hidden *turbulent* state. Empirically, we find that the behavior of both the market and the momentum portfolio are consistent with this model. When the economy is in the latent *turbulent* state, the levels of market and momentum strategy volatility are more than double their values in the calm state. In the *turbulent* state, the option-like features of the momentum strategy are pronounced; while in the *calm* state they are dramatically attenuated.

We also find that momentum crashes tend to occur more frequently during months in which the hidden state is more likely to be turbulent. The turbulent state occurs infrequently in the sample: the probability that the hidden state is turbulent exceeds 80% in only 156 of the 1,092 months in our 1927:01-2017:12 sample. Yet in each of the 12 severe loss months, the *ex-ante* probability that the hidden state is turbulent exceeds 80 percent. Interestingly, the average momentum strategy return during those 156 months is only -0.86% per month.

We derive the conditional option-adjusted alpha of the momentum strategy for a given month based on the information available until the end of the previous month using the HMM return generating process for the momentum strategy returns and market excess returns, the price of call options on the market, and the risk free rate. Over the 224 month period for which we have call option prices (1996:01-2017:12), the average conditional option-adjusted alpha is 1.36%/month, which is significantly positive. In fact, the conditional option-adjusted alpha is significantly positive at the 5% percent significance level during 209 out of the 244 months and significantly negative for two months 2008:12 and 2009:03 of the financial crisis period.

Because our HMM model identifies periods when momentum crashes are more likely, an investor can manage the tail risk embedded in his/her momentum based portfolio with our model. Momentum timing strategy using the conditional option-adjusted alpha of the momentum strategy achieves superior out-of-sample risk-adjusted performance.

References

Asness, Clifford S., Andrea Frazzini, and Lasse H. Pedersen, 2014, Quality minus junk, Working paper, SSRN.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, 2013, Value and momentum everywhere, *Journal of Finance* 68, 929–985.

Badrinath, Swaminathan G, and Sunil Wahal, 2002, Momentum trading by institutions, *Journal of Finance* 57, 2449–2478.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Barroso, Pedro, and Pedro Santa-Clara, 2015, Momentum has its moments, *Journal of Financial Economics* 116, 111–120.

Black, Fischer, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–659.

Boguth, Oliver, Murray Carlson, Adlai Fisher, and Mikhail Simutin, 2011, Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas, *Journal of Financial Economics* 102, 363–389.

Bollerslev, Tim, Viktor Todorov, and Lai Xu, 2015, Tail risk premia and return predictability, *Journal of Financial Economics* 118, 113–134.

Breen, William, Lawrence R Glosten, and Ravi Jagannathan, 1989, Economic significance of predictable variations in stock index returns, *Journal of Finance* 44, 1177–89.

Chabot, Benjamin, Eric Ghysels, and Ravi Jagannathan, 2014, Momentum trading, return chasing and predictable crashes, Working paper.

Chan, K.C., 1988, On the contrarian investment strategy, *Journal of Business* 61, 147–163.

Chen, Zhuo, and Andrea Lu, 2015, A market-based funding liquidity measure, Working paper, Tsinghua University and University of Melbourne.

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

Daniel, Kent, and Tobias J. Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221–247.

DeBondt, Werner F. M., and Richard H. Thaler, 1987, Further evidence on investor overreaction and stock market seasonality, *Journal of Finance* 42, 557–581.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, 2015, The price of variance risk, Working paper, Northwestern University, University of Chicago, Penn State University, and the Federal Reserve Board.

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

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

Flannery, Mark J, and Kasturi P Rangan, 2006, Partial adjustment toward target capital structures, *Journal of financial economics* 79, 469–506.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Glosten, Lawrence R., and Ravi Jagannathan, 1994, A contingent claim approach to performance evaluation, *Journal of Empirical Finance* 1, 133–160.

Glosten, Lawrence R., Ravi Jagannathan, and David Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

Grinblatt, Mark, and Sheridan Titman, 1989, Mutual fund performance: an analysis of quarterly portfolio holdings, *Journal of Business* 62, 393–416.

Grinblatt, Mark, and Sheridan Titman, 1993, Performance measurement without benchmarks: An examination of mutual fund returns, *Journal of Business* 66, 47–68.

Grundy, Bruce, and J. Spencer Martin, 2001, Understanding the nature of the risks and the source of the rewards to momentum investing, *Review of Financial Studies* 14, 29–78.

Hamilton, James D, 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357–84.

Han, Bing, Dong Hong, and Mitch Warachka, 2009, Forecast accuracy uncertainty and momentum, *Management Science* 55, 1035–1046.

Hansen, Lars P., and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Henriksson, Roy D., and Robert C. Merton, 1981, On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills, *Journal of Business* 54, 513–533.

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

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: an investment approach, *Review of Financial Studies* 28, 650–705.

Jagannathan, Ravi, and Robert A Korajczyk, 1986, Assessing the market timing performance of managed portfolios, *Journal of Business* 59, 217–35.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *Journal of Finance* 45, 881–898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Leary, Mark T, and Michael R Roberts, 2005, Do firms rebalance their capital structures?, *The journal of finance* 60, 2575–2619.

Lehmann, Bruce N., 1990, Fads, martingales, and market efficiency, *Quarterly Journal of Economics* 105, 1–28.

Lettau, Martin, Matteo Maggiori, and Michael Weber, 2014, Conditional risk premia in currency markets and other asset classes, *Journal of Financial Economics* 114, 197 – 225.

Liu, Laura Xiaolei, and Lu Zhang, 2008, Momentum profits, factor pricing, and macroeconomic risk, *Review of Financial Studies* 21, 2417–2448.

Merton, Robert C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449–470.

Mitchell, Mark, and Todd Pulvino, 2001, Characteristics of risk and return in risk arbitrage, *Journal of Finance* 56, 2135–2175.

Moreira, Alan, and Tyler Muir, 2017, Volatility managed portfolios, *Journal of Finance* 72, 1611–1644.

Pastor, Lubos, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Pearson, Karl, 1900, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50, 157–175.

Rouwenhorst, K. Geert, 1998, International momentum strategies, *Journal of Finance* 53, 267–284.

Swaminathan, Bhaskaran, 2010, Quantitative money management: A practical application of behavioral finance, Working paper.

Welch, Ivo, 2004, Capital structure and stock returns, *Journal of political economy* 112, 106–131.

White, Halbert, 1980, A heteroskedasticity-consistent covariance estimator and direct test for heteroskedasticity, *Econometrica* 48, 817–838.

Wooldridge, Jeffrey, 1986, Estimation and inference for dependent processes, in R. F. Engle, and D. McFadden, eds., *Handbook of Econometrics*, volume 4, first edition, chapter 45, 2639–2738 (Elsevier).

Table 1: Summary Statistics of Momentum Strategy Returns

Panel A reports the mean, standard deviation (SD), annualized Sharpe ratio (SR), skewness (skew), kurtosis (kurt), maximum (max), and minimum (min) of momentum strategy returns (MOM) along with those of market excess returns (Mkt-Rf), and scaled market excess returns (Mkt-Rf*) with the standard deviation equal to that of momentum strategy returns. Panel B reports the average risk adjusted monthly return (alpha), calculated as the intercept from time series regressions of the MOM return on the Market along with the corresponding risk exposures (betas). The sample period is 1927:01-2017:12. The t-statistics are computed using the heteroskedasticity consistent covariance estimator (White, 1980). The mean, SD, max and min in Panel A and α in Panel B are reported in percentage per month.

Panel A: Summary Statistics							
	mean	SD	SR	skew	kurt	max	min
MOM	1.17	7.80	0.52	-2.34	20.43	26.16	-77.02
Mkt-Rf	0.66	5.35	0.43	0.19	10.84	38.85	-29.13
Mkt-Rf*	0.96	7.82	0.43	0.19	10.84	56.74	-42.54
Panel B: Risk Adjusted MOM Returns							
	α	$\beta_{\text{Mkt-Rf}}$	$Adj.R^2$				
ESTIMATE	1.53	-0.54	0.13				
(t-stat)	(7.43)	(-5.15)					

Table 2: Market Conditions during Momentum Crashes

Panel A presents the momentum strategy returns (R_{MOM}), and the excess returns of winner portfolio, loser portfolio, and market portfolio, denoted by R^e_{WIN}, R^e_{LOS} and Mkt-Rf, respectively, during months with the momentum crashes worse than -20% during 1927:01-2017:12, along with the breakpoints for the winner and loser portfolios, i.e., the threshold values for the cumulative returns over the measurement period from month $t-12$ to $t-2$, i.e., (*12-2 Ret*) for entering the winner and loser portfolios , and the ratio of the book value of long-term debt to the market value of equity (BD/ME) of the winner and loser portfolios, the cumulative market returns in percentage during the 36 and 12 months preceding the month in which the momentum portfolios are formed, and the realized volatility of daily market returns during the 12 and 6 months preceding the month in which the momentum portfolios are formed. Sample averages of the variables across thirteen months in which the momentum crashes were realized are reported in Panel B and the averages of those variables across all available data are reported in Panel C. The book value of long term debt (BD) is available from 1964 onwards. Realized volatility is computed as the square root of the sum of squared daily returns and reported as annualized percentage. All of the variables except BD/ME are reported in percentage.

					Winner Portfolio		Loser Portfolio		Past Mkt Ret		Past Mkt RV	
Date	R_{MOM}	R^e_{WIN}	R^e_{LOS}	Mkt-Rf	Break-points	BD /ME	Break-points	BD /ME	36 Mos.	12 Mos.	12 Mos.	6 Mos.
Panel A: Momentum Crash Months												
1931:06	−29.26	8.04	37.30	13.90	−14.38	n.a.	−74.07	n.a.	−36.04	−45.51	22.86	21.05
1932:07	−60.17	14.10	74.27	33.84	−41.04	n.a.	−88.24	n.a.	−81.46	−65.84	42.04	40.56
1932:08	−77.02	16.93	93.95	37.06	−38.57	n.a.	−86.25	n.a.	−76.32	−51.05	41.95	39.22
1933:04	−41.92	28.78	70.70	38.85	28.57	n.a.	−54.35	n.a.	−72.31	−12.47	45.96	39.68
1933:05	−26.87	19.30	46.17	21.43	69.23	n.a.	−41.94	n.a.	−60.80	48.04	45.77	40.79
1938:06	−33.20	10.47	43.67	23.87	−18.88	n.a.	−68.93	n.a.	8.65	−39.18	32.52	29.21
1939:09	−45.16	7.91	53.07	16.88	35.71	n.a.	−33.33	n.a.	−16.39	−1.03	19.65	19.58
2001:01	−42.03	−7.00	35.03	3.13	66.62	0.14	−55.17	0.70	37.48	−11.71	24.65	22.30
2002:11	−20.12	2.12	22.24	5.96	38.83	0.31	−48.16	0.97	−30.81	−13.62	24.63	30.94
2009:03	−39.76	4.81	44.57	8.95	−3.06	0.25	−79.47	29.61	−38.38	−42.63	41.78	55.33
2009:04	−45.58	−0.13	45.45	10.19	−10.70	0.28	−82.44	34.86	−34.06	−37.00	43.23	55.39
2009:08	−25.38	0.20	25.58	3.33	3.81	0.68	−66.21	10.84	−15.23	−18.91	44.50	32.06
Panel B: Averages Across Momentum Crash Months												
	−40.54	8.79	49.33	18.12	9.68	0.33	−64.88	15.39	−34.64	−24.24	35.79	35.51
Panel C: Average Across All Available Sample Months												
	1.17	1.23	0.06	0.66	54.23	0.32	−23.45	1.07	38.82	12.03	14.80	14.55

Table 3: Option-like Feature of Momentum Returns and Market Conditions

We partition the months in our sample into three groups on the basis of the cumulative market return during the 36 months immediately preceding the momentum portfolio formation date. The 'High' ('Low') group consists of all months in which this variable is in the top (bottom) 20th percentile. The rest of the months are classified as 'Medium'. We estimate equation (1): using ordinary least squares for the months within each group, and report the results in Panel A. The dependent variable is either: the momentum strategy returns (R_{MOM}), or the returns of the winner or loser portfolio in excess of risk-free return (R^e_{WIN} and R^e_{LOS}). For comparison, in Panel B we report the estimates for the CAPM, without the exposure to the call option on the market in (1). Panel C reports the number of momentum losses worse than 20% within each group. Panel D reports the skewness of $R^e_{p,t}$ with that of estimated ε of (1). α is reported in percentage per month. The t-statistics are computed using White's (1980) heteroskedasticity-consistent covariance and reported in parentheses. The sample period is 1929:07-2017:12.

State Variable: Past 36 Months Market Returns									
	Low			Medium			High		
R^e_p :	R_{MOM}	R^e_{WIN}	R^e_{LOS}	R_{MOM}	R^e_{WIN}	R^e_{LOS}	R_{MOM}	R^e_{WIN}	R^e_{LOS}
A: Henriksson-Merton Estimates									
α	2.87	1.14	-1.73	2.45	0.91	-1.53	0.76	0.67	-0.09
$t(\alpha)$	(3.32)	(3.13)	(-2.77)	(4.95)	(3.14)	(-5.77)	(1.33)	(1.94)	(-0.21)
β^0	-0.48	0.97	1.45	-0.12	1.25	1.38	0.19	1.39	1.20
$t(\beta^0)$	(-3.44)	(12.82)	(16.78)	(-0.67)	(11.81)	(13.77)	(1.54)	(18.85)	(10.41)
β^+	-0.95	-0.29	0.66	-0.41	-0.23	0.19	0.22	-0.13	-0.35
$t(\beta^+)$	(-3.15)	(-2.16)	(3.34)	(-1.30)	(-1.21)	(1.16)	(0.80)	(-0.83)	(-1.62)
$Adj.R^2$	0.49	0.78	0.83	0.06	0.71	0.70	0.05	0.81	0.63
B: CAPM Estimates									
α	0.13	0.30	0.17	1.78	0.55	-1.23	1.18	0.43	-0.76
$t(\alpha)$	(0.21)	(1.22)	(0.37)	(7.37)	(4.34)	(-7.48)	(3.06)	(2.04)	(-2.78)
β	-1.02	0.81	1.83	-0.33	1.14	1.47	0.28	1.34	1.06
$t(\beta)$	(-7.00)	(14.20)	(18.84)	(-3.55)	(21.27)	(28.67)	(3.09)	(31.03)	(14.26)
$Adj.R^2$	0.44	0.77	0.82	0.05	0.71	0.70	0.05	0.81	0.62
C: Number of Momentum Losses worse than -20%									
	10			2			0		
D: Conditional Skewness									
R^e_p	-2.26	-0.21	1.71	-0.89	-0.58	0.02	0.05	-0.72	-0.70
ε_p	-0.91	-0.99	0.90	-0.89	-0.45	0.80	-0.15	1.03	0.80

Table 4: Option-like Feature due to High Effective Leverage within Loser Stocks

We sort stocks in the loser portfolios based on two proxies for effective leverage (i) financial leverage (Panel A) and (ii) past 36-month loss of individual stocks (Panel B) and form three groups: top 30%, middle 40%, bottom 30%. We construct three corresponding portfolios by value-weighting stocks in each group. We estimate equation (1) using the excess returns on these portfolios α is reported in percentage per month. The sample period is 1964:07-2017:12.

	A: Using Financial Leverage as a Proxy of Effective					B: Using Past 36-Month Loss as a Proxy of Effective				
	All Loser Stocks	Top 30%	Middle 40%	Bottom 30%	Top− Bottom	All Loser Stocks	Top 30%	Middle 40%	Bottom 30%	Top− Bottom
α	-1.42	-1.29	-1.25	-0.07	-1.22	-1.42	-1.11	-1.35	-0.73	-0.38
$t(\alpha)$	(-4.46)	(-2.69)	(-3.68)	(-0.16)	(-2.90)	(-4.46)	(-3.33)	(-3.79)	(-2.01)	(-1.15)
β^0	1.31	1.32	1.27	1.41	-0.09	1.31	1.27	1.45	1.32	-0.05
$t(\beta^0)$	(13.86)	(9.68)	(14.01)	(14.53)	(-0.83)	(13.86)	(14.16)	(14.91)	(13.11)	(-0.64)
β^+	0.29	0.40	0.17	0.03	0.37	0.29	0.23	0.05	0.03	0.20
$t(\beta^+)$	(1.47)	(1.39)	(0.87)	(0.14)	(1.95)	(1.47)	(1.23)	(0.23)	(0.13)	(1.36)

Table 5: Option-like Feature of Momentum Returns during Dot-com Crash

We estimate equation (1) with the momentum strategy return (R_{MOM}) and the winner and loser portfolio excess returns (R^e_{WIN} and R^e_{LOS}) as a candidate dependent variable. We use 36 monthly data on returns during 2000:01-2002:12. α is reported in percentage per month. The t-statistics are computed using White's (1980) heteroskedasticity-consistent covariance.

R^e_p :	R_{MOM}	R^e_{WIN}	R^e_{LOS}
α	3.37	1.56	-1.81
$t(\alpha)$	(0.91)	(0.84)	(-0.70)
β^0	-0.44	1.25	1.68
$t(\beta^0)$	(-0.73)	(3.04)	(4.32)
β^+	-1.30	-0.53	0.77
$t(\beta^+)$	(-1.23)	(-0.81)	(1.07)

Table 6: Option-like Feature of Momentum Returns in Simulated Data

We estimate equation (1) with the momentum strategy return (R_{MOM}) and the winner and loser portfolio returns (R_{WIN} and R_{LOS}) as a candidate dependent variable using 10,000 simulated data $t = 121, \cdots, 10120$. We partition the 10,000 simulated data into two groups on the basis of the cumulative market return during the 36 months immediately preceding the momentum portfolio formation date. The 'Depressed Market' group consists of all months in which this variable is in the bottom 30th percentile. The rest of the months are classified as 'Non Depressed Market'.

	Depressed Market			Non Depressed Market		
R^e_p :	R_{MOM}	R_{WIN}	R_{LOS}	R_{MOM}	R_{WIN}	R_{LOS}
α	1.89	0.14	-1.75	0.79	-0.06	-0.85
β^0	-0.82	0.91	1.73	-0.40	1.06	1.46
β^+	-0.34	-0.02	0.33	-0.17	0.01	0.19

Table 7: HMM estimates from Simulated Data

We maximize the likelihood of simulated momentum and market return data over $t = 121, \cdots, 12000$. The details of estimation assumptions are described in Section 2. The estimates of α, σ_{MOM}, μ, and σ_{MKT} are reported in percentage per month.

	Hidden State	
	$S_t = \text{Calm}(C)$	$S_t = \text{Turbulent}(T)$
Parameter	estimates	estimates
α (%)	0.44	2.43
β^0	−0.15	−0.82
β^+	−0.15	−0.51
σ_{MOM} (%)	2.02	3.07
μ (%)	0.32	−0.80
σ_{MKT} (%)	6.58	9.02

Table 8: Distribution of Momentum Crashes in Simulated Data

We define momentum crashes as the 1% left-tail of simulated momentum returns over $t = 121, \cdots, 10120$. We report the conditional distribution of momentum crashes over nine groups: (Top 30%, Middle 40%, Bottom 30%) of $\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1})$ × (Top 30%, Middle 40%, Bottom 30%) of loser portfolio leverage.

	Loser Portfolio Leverage			
$\Pr(S_t = \text{Turbulent}\|\mathcal{F}_{t-1})$	Top 30%	Middle 40%	Bottom 30%	total
Top 30%	0.84	0.02	0.00	0.86
Middle 40%	0.10	0.02	0.00	0.12
Bottom 30%	0.02	0.00	0.00	0.02
total	0.96	0.04	0.00	

Table 9: Maximum Likelihood Estimates of HMM Parameters

We maximize the likelihood of data with the assumption that $\varepsilon_{\text{MOM},t}$ in (3) is drawn from a standard normal distribution and $\varepsilon_{\text{MKT},t}$ in (4) is drawn from a Student's t-distribution with d.f.=5. The parameters are estimated using data for the period 1927:01-2017:12. The estimates of α, σ_{MOM}, μ, and σ_{MKT} are reported in percentage per month.

	Hidden State			
	S_t = Calm(C)		S_t = Turbulent(T)	
Parameter	estimates	(t-stat)	estimates	(t-stat)
α (%)	1.86	(7.17)	3.62	(3.52)
β^0	0.32	(3.54)	−0.40	(−2.53)
β^+	−0.41	(−2.95)	−1.07	(−4.48)
σ_{MOM} (%)	4.42	(26.99)	10.68	(20.41)
μ (%)	1.13	(8.65)	−0.50	(−1.01)
σ_{MKT} (%)	4.01	(29.30)	8.33	(16.63)
$\Pr(S_t = s_{t-1} \mid S_{t-1} = s_{t-1})$	0.99	(11.12)	0.95	(11.79)

Table 10: Momentum and Market Excess Returns: Sample Moments vs HMM-implied Moments

We compare the HMM-implied moments of momentum strategy returns and market excess returns with the corresponding moments in our sample. After we estimate HMM parameters, we generate $\varepsilon_{\mathrm{MOM},t}$ and $\varepsilon_{\mathrm{MKT},t}$ in our HMM specification in (3) and (4) using Monte Carlo simulation from various combinations of normal and Student's t-distributions. Then, we construct a 1,092-month time series of momentum strategy and market excess returns using HMM specification and compute their first four moments. We then repeat this exercise 10,000 times to obtain the distribution of the first four momentums.

	Momentum Strategy Returns: $R_{\mathrm{MOM},t}$						Market Excess Returns: $R^e_{\mathrm{MKT},t}$					
	Realized Moments	Quantiles (%) of Simulated Moments					Realized Moments	Quantiles (%) of Simulated Moments				
		0.5	2.5	50	97.5	99.5		0.5	2.5	50	97.5	99.5
Panel A: Normal ($\varepsilon_{\mathrm{MOM},t}$) and Student's t ($\varepsilon_{\mathrm{MKT},t}$)												
mean	1.17	0.56	0.76	1.22	1.69	1.83	0.66	0.12	0.29	0.74	1.06	1.20
std.dev	7.80	5.89	6.16	7.72	9.35	9.74	5.35	4.45	4.62	5.44	6.34	6.75
skewness	-2.34	-4.51	-2.55	-0.62	0.07	0.36	0.19	-2.59	-1.71	-0.28	1.51	3.58
kurtosis	20.43	5.37	5.84	8.37	32.57	67.62	10.84	5.18	5.55	8.61	34.25	67.01
Panel B: Normal ($\varepsilon_{\mathrm{MOM},t}$) and Normal ($\varepsilon_{\mathrm{MKT},t}$)												
mean	1.17	0.49	0.61	1.10	1.60	1.72	0.66	0.13	0.26	0.66	1.01	1.07
std.dev	7.80	6.07	6.43	7.65	8.88	9.29	5.35	4.49	4.66	5.33	6.08	6.26
skewness	-2.34	-1.47	-1.28	-0.61	-0.08	0.13	0.19	-0.85	-0.75	-0.36	0.07	0.22
kurtosis	20.43	5.60	6.00	7.81	11.38	13.84	10.84	4.52	4.75	5.72	7.43	8.29
Panel C: Student's t ($\varepsilon_{\mathrm{MOM},t}$) and Normal ($\varepsilon_{\mathrm{MKT},t}$)												
mean	1.17	0.69	0.77	1.27	1.74	1.95	0.66	0.15	0.26	0.66	0.99	1.09
std.dev	7.80	5.97	6.48	7.83	9.26	9.56	5.35	4.39	4.59	5.36	6.15	6.39
skewness	-2.34	-2.26	-1.58	-0.46	0.73	2.28	0.19	-0.91	-0.80	-0.39	0.02	0.17
kurtosis	20.43	6.20	6.80	9.71	26.04	52.72	10.84	4.49	4.70	5.72	7.36	8.12
Panel D: Student's t ($\varepsilon_{\mathrm{MOM},t}$) and Student's t ($\varepsilon_{\mathrm{MKT},t}$)												
mean	1.17	0.72	0.92	1.40	1.86	1.99	0.66	0.15	0.31	0.74	1.12	1.18
std.dev	7.80	5.77	6.09	7.72	9.30	10.16	5.35	4.32	4.55	5.36	6.27	6.62
skewness	-2.34	-3.68	-2.06	-0.47	0.80	2.04	0.19	-2.99	-1.49	-0.30	0.94	2.07
kurtosis	20.43	6.26	6.79	10.37	30.80	71.93	10.84	5.05	5.54	8.35	28.78	54.69

Table 11: Distribution of Momentum Crashes over 1964:07-2017:12

We define momentum crashes as momentum returns worse than -20%. There were five momentum crashes in the 1964:07-2017:12 period. We report the conditional distribution of momentum crashes over four groups: (Top 20%, Bottom 80%) of $\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1}) \times$ (Top 20%, Bottom 80%) of loser portfolio financial leverage.

	Loser Portfolio Financial Leverage		
$\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1})$	Top 20%	Bottom 80%	total
Top 20%	3	2	5
Bottom 80%	0	0	0
total	3	2	

Table 12: Extreme Losses/Gains Conditional on $\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1})$

This table presents the fraction of the total number of extreme losses/gains greater than a given value that occur when $\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1})$ is larger than a given threshold. The sample period is 1927:01-2017:12.

$\Pr(S_t = T \mid \mathcal{F}_{t-1})$	# Extreme losses during turbulent months /# Extreme losses in the sample					# of
is more than	$\leq -20\%$	$\leq -17.5\%$	$\leq -15\%$	$\leq -12.5\%$	$\leq -10\%$	Months
90%	10/12	15/24	18/31	22/37	25/57	114
80%	12/12	19/24	23/31	27/37	30/57	156
70%	12/12	19/24	23/31	27/37	31/57	179
60%	12/12	19/24	23/31	27/37	33/57	210
50%	12/12	20/24	25/31	29/37	37/57	231
40%	12/12	21/24	27/31	31/37	40/57	254
30%	12/12	21/24	28/31	32/37	42/57	278
20%	12/12	21/24	29/31	33/37	43/57	306
10%	12/12	21/24	31/31	35/37	47/57	368
$\Pr(S_t = T \mid \mathcal{F}_{t-1})$	# Extreme gains during turbulent months /# Extreme gains in the sample					# of
is more than	$\geq 20\%$	$\geq 17.5\%$	$\geq 15\%$	$\geq 12.5\%$	$\geq 10\%$	Months
90%	5/12	7/16	11/27	19/44	27/71	114
80%	8/12	11/16	15/27	24/44	33/71	156
70%	8/12	11/16	15/27	27/44	36/71	179
60%	8/12	11/16	16/27	28/44	38/71	210
50%	9/12	12/16	19/27	32/44	43/71	231
40%	10/12	13/16	20/27	33/44	46/71	254
30%	10/12	13/16	20/27	34/44	47/71	278
20%	12/12	15/16	22/27	36/44	51/71	306
10%	12/12	16/16	23/27	37/44	54/71	368

Table 13: False Positives In Predicting Momentum Crashes

We compare the number of false positives in predicting momentum crashes across different tail risk measures. The number of false positives of a given tail risk measure is computed as follows. Suppose we classify months in which momentum strategy returns lost more than a threshold X. Let Y denote the lowest value attained by a given tail risk measure during those momentum crash months. During months when the tail risk measure is above the threshold level of Y, we count the number of months when momentum crashes did not occur and we denote it as the number of false positives. We consider X=10%, 20%, 30%, 40%. In Panel A, we use $\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1})$, computed by using in-sample parameter estimates, as a tail risk measure. The results in Panel A-1 are from our original HMM model in (3), (4) and (5). To emphasize the importance of option-like feature $\beta^+(S_t)$ in (3), we impose the restriction $\beta^+(S_t) = 0$ and report the associated results in Panel A-2. In Panel B, we use various estimates of the volatility of momentum strategy returns as tail risk measures. Specifically, we estimate the volatility of the momentum strategy returns using GARCH (1,1) and realized volatility of daily momentum strategy returns over the previous 3, 6, 12, and 36 months. In Panel C, we use the volatility of the market return estimated using GARCH(1,1) and realized volatility of the daily market return during the preceding 3, 6, 12, and 36 months as the tail risk measures. In Panel D, we use the market return during the preceding 3, 6, 12 and 36 month windows as tail risk measures.

	Momentum Crash Threshold (-X)			
Tail Risk Measure	$\leq -40\%$	$\leq -30\%$	$\leq -20\%$	$\leq -10\%$
Panel A: HMM				
A-1: Main Specification				
$\Pr(S_t = \text{T}|\mathcal{F}_{t-1})$	130	128	124	929
A-2: Without the option-like feature $\beta^+(S_t) = 0$				
$\Pr(S_t = \text{T}|\mathcal{F}_{t-1})$	169	167	163	919
Panel B: Momentum Strategy Returns Volatility				
GARCH(1,1)	262	260	256	876
RV(3 Months)	250	248	244	936
RV(6 Months)	216	214	210	861
RV(12 Months)	236	234	230	892
RV(36 Months)	166	164	176	995

Continued on next page

Table 13 – continued from previous page

Panel C: Market Returns Volatility				
GARCH(1,1)	189	187	183	831
RV(3 Months)	170	168	164	936
RV(6 Months)	185	183	179	967
RV(12 Months)	195	193	189	843
RV(36 Months)	180	178	174	888
Panel D: Past Market Returns				
3 Months	667	665	1030	995
6 Months	130	128	968	991
12 Months	252	250	1019	985
36 Months	511	509	505	979

Table 14: Risk-Managed Momentum Portfolio Returns

This table reports the performance of various momentum strategies. The 'MOM' represents the conventional momentum strategy. The 'HMM-MOM' is constructed as follows. Let $\overline{\Pr}$ denote the maximum level of $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ such that all momentum strategy losses exceeding 20% are captured. Then, we enter (exit) momentum strategy if $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1}) < (>) \overline{\Pr}$. For the last two strategies of 'k/σ_{MOM}' and '$k/\sigma^2_{\mathrm{MOM}}$', we use $k/\sigma_{\mathrm{MOM},t}$ and $k/\sigma^2_{\mathrm{MOM},t}$ for the exposure to the momentum, respectively. We compute $\sigma^2_{\mathrm{MOM},t}$ as the annualized average of squared daily momentum returns over the previous six months. We adjust k so that the average exposure to the momentum strategy becomes one. We report the mean (%/Month), standard deviation (SD %/Month), annual Sharpe ratio (SR), and annual CAPM appraisal ratio (AR) of the four strategies. The sample period is 1927:01-2017:12.

	Mean	SD	SR	AR
MOM	1.17	7.80	0.52	0.73
HMM-MOM	1.31	5.09	0.89	0.92
k/σ_{MOM}	1.50	5.41	0.96	1.05
$k/\sigma^2_{\mathrm{MOM}}$	1.67	5.40	1.07	1.10

Table 15: Option Adjusted Alphas and Momentum Crashes

We classify the 244 months in the sample period 1996:01-2016:04 into two groups - one group of months with significantly positive option-adjusted alphas and another group of months without. We examine whether the empirical frequency of large losses in momentum strategy returns significantly differs across the two groups using chi-square test by Pearson (1900). This table reports the number of months in the two-way table for each threshold level of -10%, -20%, and -30% per month, and the chi-square test statistics as well as the associated p-values.

	Significantly positive Option-adjusted Alpha		χ^2-stat (p-value)
	YES	NO	
$R_{\text{MOM}} > -10\%$	195	26	16.658
$R_{\text{MOM}} < -10\%$	13	10	(0.000)
	YES	NO	
$R_{\text{MOM}} > -20\%$	208	31	29.493
$R_{\text{MOM}} < -20\%$	0	5	(0.000)
	YES	NO	
$R_{\text{MOM}} > -30\%$	208	33	17.549
$R_{\text{MOM}} < -30\%$	0	3	(0.000)

Table 16: Momentum Timing

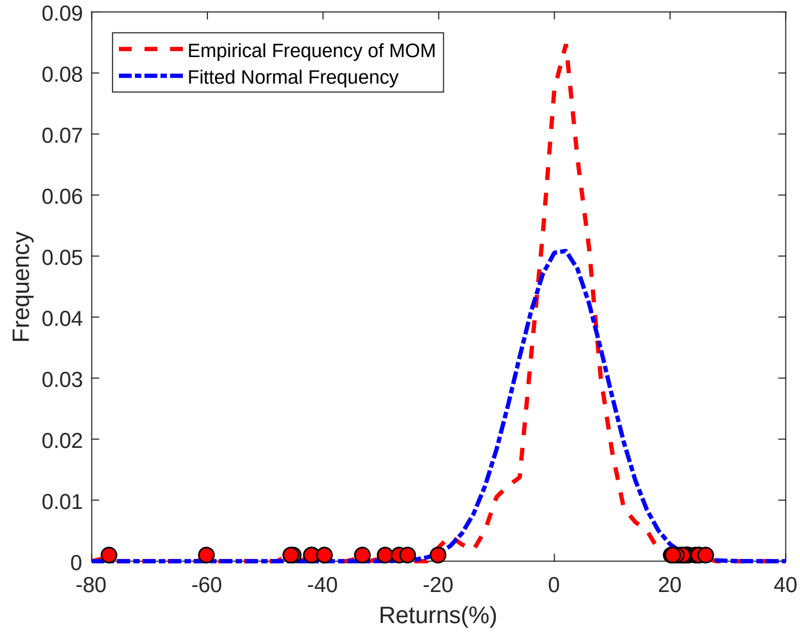
We consider two momentum timing strategies (i) a strategy of holding momentum portfolio only during months with significant alpha and switching into treasury bonds in other months and (ii) another strategy of holding momentum portfolio over the months with significant alpha and switching into the market portfolio in other months. This table reports the annualized summary statistics of the returns from the proposed momentum timing strategies and those of conventional momentum strategy returns over the 244 months in the sample period 1996:01-2016:04.

	Momentum Timing (i)	Momentum Timing (ii)	Conventional Momentum
annual mean (%/year)	16.49	18.17	9.23
annual std.dev (%/year)	22.60	24.24	31.72
skewness	0.17	-0.04	-1.31
kurtosis	2.22	1.32	5.44
annual Sharpe ratio	0.73	0.75	0.29

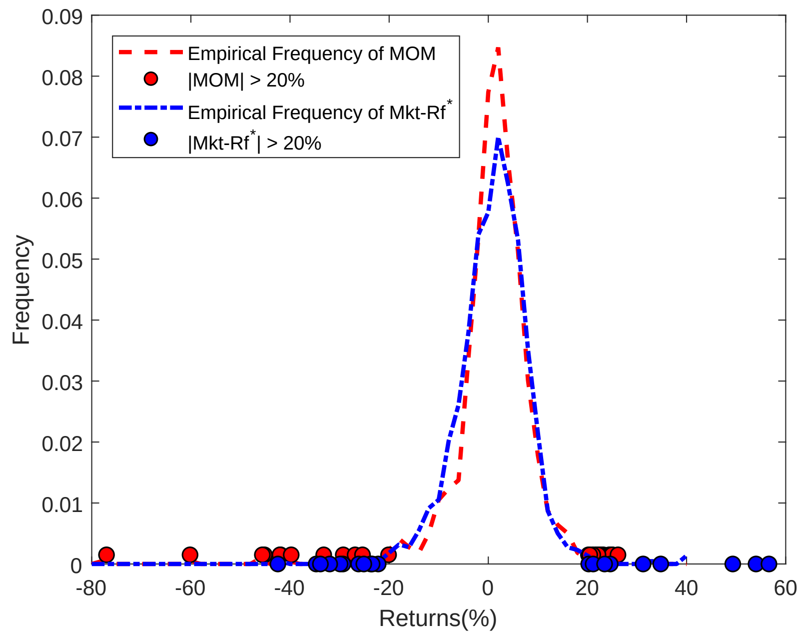
Table 17: SYSTEMATIC RISK IN MOMENTUM STRATEGY RETURNS

This table presents the results of regressing $\epsilon_{\text{MOM},t}$ in (16) on various systematic risk factors, $\epsilon_{\text{MOM},t}$ = intercept + coeff $\times$ systematic factor$_t$ + e_t. Results are obtained by using the data from 1996:01 to 2016:04 (244 months) where we can reconstruct $\epsilon_{\text{MOM},t}$ from the market prices of call option on S&P 500 from OptionMetrics. Details on systematic factors are described in the main text.

systematic factor	coeff	t(coeff)	R^2(%)	First Month	Last Month	N
Panel A: General Factors						
MKT	-0.09	-0.30	0.02	1996:01	2016:04	244
SMB	-0.35	-0.63	0.13	1996:01	2016:04	244
HML	-2.73	-4.40	5.21	1996:01	2016:04	244
RMW	-0.04	-0.12	0.02	1996:01	2016:04	244
CMA	-0.15	-0.39	0.15	1996:01	2016:04	244
I/A	-0.34	-0.78	0.75	1996:01	2016:04	244
ROE	1.03	3.34	13.76	1996:01	2016:04	244
Panel B: Liquidity Related Factors						
systematic factor	coeff	t(coeff)	R^2(%)	First Month	Last Month	N
QMJ	0.55	1.95	4.17	1996:01	2016:04	244
LIQ	0.32	1.83	2.06	1996:01	2016:04	244
FLS	0.07	0.55	0.68	1996:01	2012:10	202
BAB	0.20	0.79	1.05	1996:01	2012:03	195
Δ LIBOR	0.05	1.23	1.73	1996:01	2013:12	216
Δ TERM	-0.04	-1.41	1.23	1996:01	2013:12	216
Δ CREDIT	0.03	0.60	0.18	1996:01	2013:12	216
Δ TED	-0.04	-0.39	0.09	1996:01	2013:12	216
Panel C: Tail Risk Related Factors						
systematic factor	coeff	t(coeff)	R^2(%)	First Month	Last Month	N
VAR-SWAP 1M	0.00	-0.45	0.05	1996:02	2013:09	212
VAR-SWAP 3M	0.01	0.95	0.17	1996:02	2013:08	210
VAR-SWAP 6M	0.02	1.45	0.49	1996:08	2013:08	203
VAR-SWAP 12M	0.01	0.62	0.11	1997:03	2013:08	193
Δ VIX	0.00	0.60	0.13	1996:01	2016:04	244
Δ LJV	-0.80	-0.44	0.10	1996:01	2013:12	216



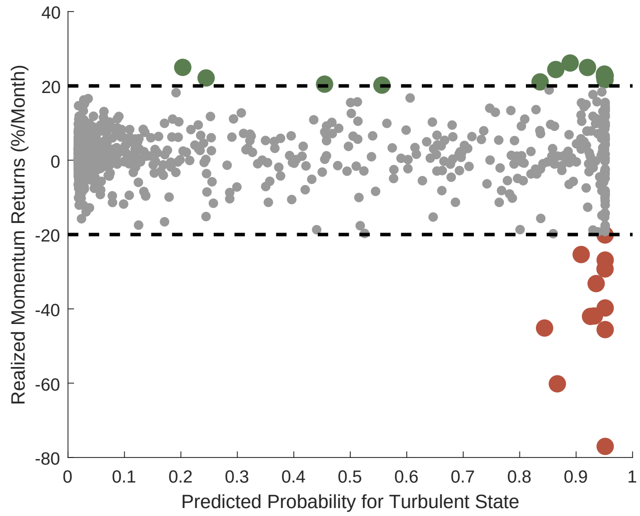
A Momentum Strategy Returns – Smoothed Density Function



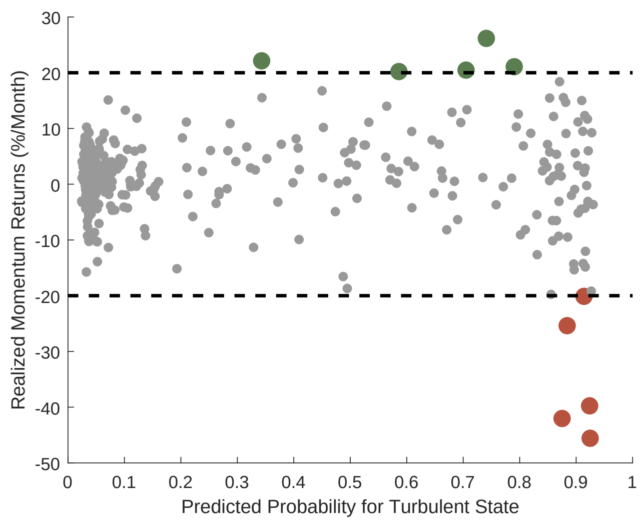
B Smoothed Density Functions–MOM and scaled excess market returns

Figure 1: Empirical Frequency of Momentum Strategy Returns (MOM)

Panel A plots the smoothed empirical density of the MOM and the normal density with the same mean and standard deviation. To highlight the left skew of momentum strategy returns, we represent 24 MOM returns (12 in left tails and 12 in right tails) that exceed 20% in absolute value. Panel B plots the empirical density of MOM along with the empirical density of scaled market excess returns, Mkt-Rf*, with standard deviation equal to that of momentum strategy returns. The sample period is 1927:01-2017:12.



A In Sample Prediction



B Out of Sample Prediction

Figure 2: Momentum Returns and Probability that the Hidden State is Turbulent

The figure presents a scatter plot of momentum strategy return on the vertical axis and $\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1})$, the probability that the hidden state is turbulent, on the horizontal axis. Momentum strategy returns below -20% are highlighted in red, and returns of exceeding 20% are in green. Figure (a) is based on in-sample estimates using all 1,092 months (1927:01-2017:12). For each month t of the last 400 months in 1984:09-2017:12, we skip first 10 years over 1927:01-1936:12 and estimate our HMM using data from 1937:01 till month $t-1$ to compute $\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1})$. Figure (b) reports out-of-sample results.

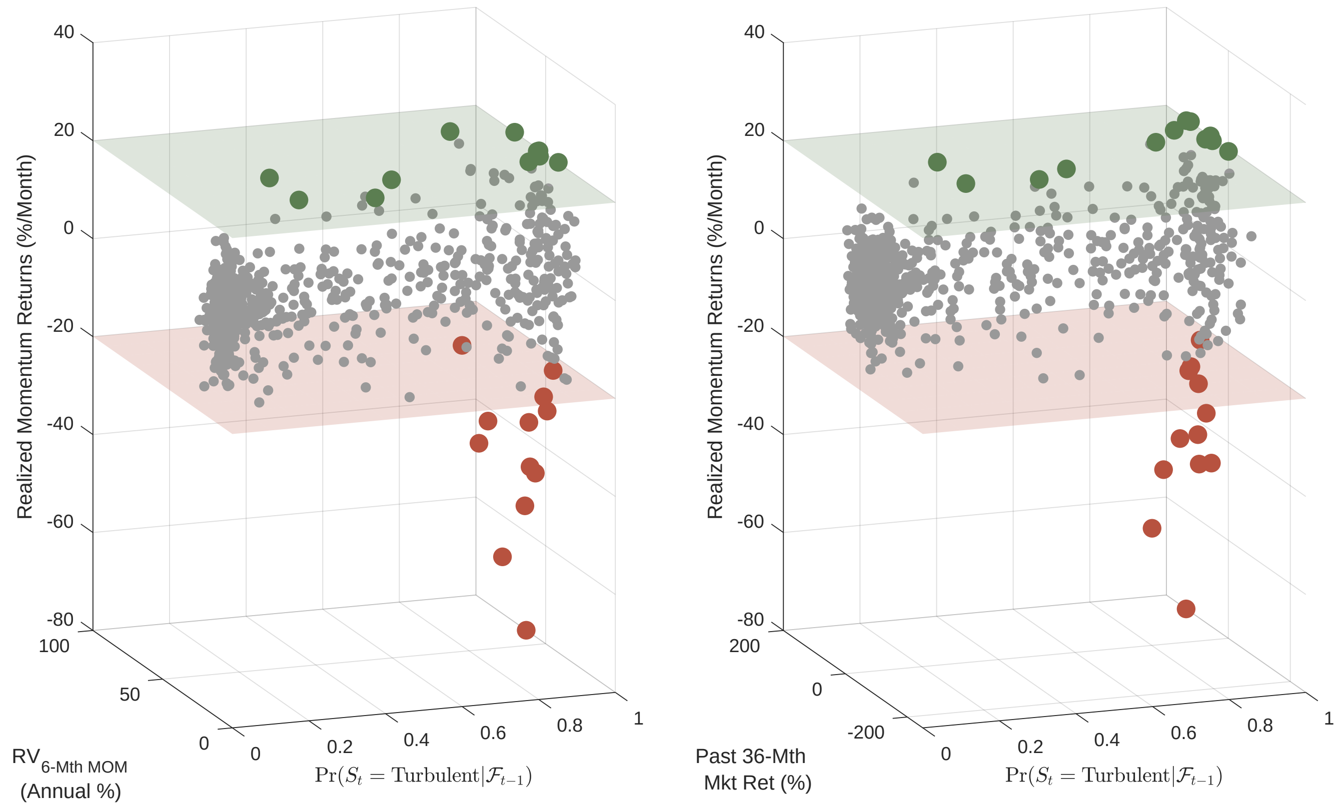


Figure 3: Probability that the Hidden State is Turbulent VS Alternative Risk Measures

The left (right) figure presents a three-dimensional scatter plot of momentum strategy return over $\Pr\left(S_t = \text{Turbulent}|\mathcal{F}_{t-1}\right)$ and realized volatility of past six-month daily momentum returns in annual percentage (past 36-month market returns in percentage). Momentum strategy returns below -20% are highlighted in red, and returns of exceeding 20% are in green. The sample period of left (right) figure is 1927:07-2017:12 (1930:01-2017:12).

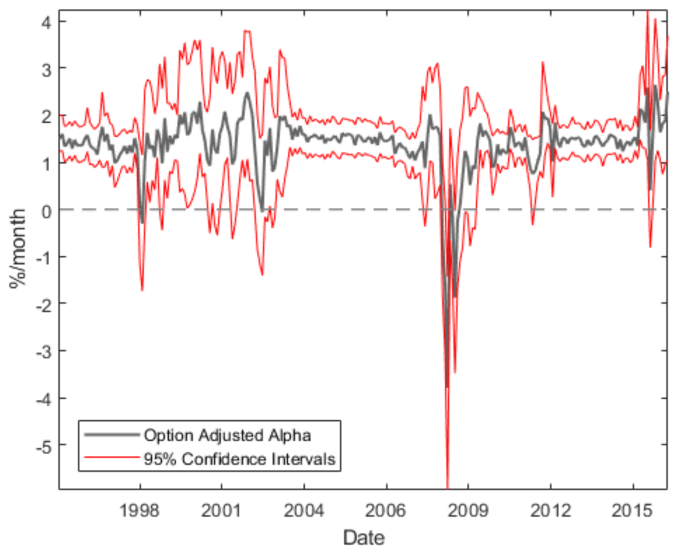


Figure 4: Time Series of Option Adjusted Alpha

Option adjusted alpha, α^*, is computed by (17). The sample period is 1996:01 to 2016:04 where we can find the market price of call option on S&P 500 from OptionMetrics. 95% confidence intervals are computed as follows. First, we simulate 10,000 sets of parameters from the asymptotic distributions of parameters obtained from ML estimator using the previous historical market excess returns and momentum returns. Then, for each set of parameters, we compute the monthly time series of α^*. Lastly, in each month, we find 95% confidence intervals of α^* by choosing top and bottom 2.5% quantiles from the simulated 10,000 observations of α^*.

A Hidden Markov Model of Momentum

Online Appendices

Appendix A Inconsistency of QML

In many settings, it useful to assume that residuals are drawn from normal distributions in estimating a statistical model. When the true distribution of the residual is not normal, these estimates are Quasi-Maximum Likelihood (QML). Wooldridge (1986) provides sufficient conditions for the consistency and asymptotic normality of QML estimators. These conditions are not satisfied in our case. Below, we provide an example where the HMM return generating process innovations are drawn from a non-normal distribution and the resulting QML estimator—obtained by maximizing the misspecified normal likelihood—gives an asymptotically biased (inconsistent) estimate of the true parameter value.

Suppose R_t follows the process given below:

$$R_t = \sigma\left(S_t\right)\varepsilon_t, \tag{A.1}$$

where $\sigma\left(S_t\right)$ is either σ_{H} or σ_{L}, depending on the realization of hidden state of S_t which is either H or L. The transition probability matrix that determines the evolution of the hidden state S_t is given by

$$\Pi = \begin{bmatrix} \Pr(S_t = \mathrm{H}|S_{t-1} = \mathrm{H}) & \Pr(S_t = \mathrm{L}|S_{t-1} = \mathrm{H}) \\ \Pr(S_t = \mathrm{H}|S_{t-1} = \mathrm{L}) & \Pr(S_t = \mathrm{L}|S_{t-1} = \mathrm{L}) \end{bmatrix} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}. \tag{A.2}$$

An econometrician observes the time series of $\{R_t\}_{t=1}^{T}$ but not the underlying state. The parameters p and σ_{L} are known. The econometrician estimates the unknown parameter σ_{H} by QML, that is by assuming that ε_t is drawn from the standard normal distribution, whereas ε_t is either 1 or -1 with equal probability. In what follows, we show that when

$$\sigma_{\mathrm{H}} = 1.5, \sigma_{\mathrm{L}} = 1, \text{ and } p = 0.52, \tag{A.3}$$

the QML estimator of σ_{H} is inconsistent.

The misspecified normal log likelihood of $\{R_t\}_{t=1}^{T}$ is given by

$$\frac{1}{T}\sum_{t=1}^{T}\log\left(\mathcal{L}\left(R_t\right)\right), \tag{A.4}$$

where

$$\mathcal{L}\left(R_t\right) = \lambda_{t-1}\phi\left(R_t|\sigma_{\mathrm{H}}\right) + \left(1-\lambda_{t-1}\right)\phi(R_t|\sigma_{\mathrm{L}}), \tag{A.5}$$

$\phi\left(x|\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)$ is the density function of $\mathcal{N}\left(0,\sigma^2\right)$, and λ_{t-1} is the probability for $S_t = H$ given the information set $\mathcal{F}_{t-1} = \{R_1, R_2, \cdots, R_{t-1}\}$ when the econometrician uses the (incorrect) normal density for inference. When the true likelihood is used, let λ_{t-1}^* denote the probability of $S_t = H$ given $\mathcal{F}_{t-1}$. Since S_t is hidden, both λ_{t-1} and λ_{t-1}^* are weighted averages of p and $1-p$ and the following should be satisfied:

$$1-p \leq \lambda_{t-1}, \lambda_{t-1}^* \leq p \tag{A.6}$$

for every $\mathcal{F}_{t-1}$.

The QML estimate $\widehat{\sigma}_{\mathrm{H}}$ is obtained by maximizing (A.4), giving rise to the first order condition:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\log\left(\mathcal{L}\left(R_t\right)\right)}{\partial\sigma_{\mathrm{H}}}|_{\sigma_{\mathrm{H}}=\widehat{\sigma}_{\mathrm{H}}}=0. \tag{A.7}$$

If $\widehat{\sigma}_{\mathrm{H}}$ converges to σ_{H}^0, the LHS of (A.7) converges to the true expectation as

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\log\left(\mathcal{L}\left(R_t\right)\right)}{\partial\sigma_{\mathrm{H}}}|_{\sigma_{\mathrm{H}}=\widehat{\sigma}_{\mathrm{H}}}\xrightarrow{p}\mathbb{E}\left[\frac{\partial\log\left(\mathcal{L}\left(R_t\right)\right)}{\partial\sigma_{\mathrm{H}}}\right]_{\sigma_{\mathrm{H}}=\sigma_{\mathrm{H}}^0} \tag{A.8}$$

under mild regularity conditions. Noting that the RHS of (A.7) is always zero, it follows that

$$\mathbb{E}\left[\frac{\partial\log\left(\mathcal{L}\left(R_t\right)\right)}{\partial\sigma_{\mathrm{H}}}\right]_{\sigma_{\mathrm{H}}=\sigma_{\mathrm{H}}^0}=\mathbb{E}\left[\mathbb{E}\left[\frac{\partial\log\left(\mathcal{L}\left(R_t\right)\right)}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right]\right]_{\sigma_{\mathrm{H}}=\sigma_{\mathrm{H}}^0}=0. \tag{A.9}$$

We show the inconsistency of $\widehat{\sigma}_{\mathrm{H}}$ by verifying that (A.9) cannot hold. When $\sigma_{\mathrm{H}}=\sigma_{\mathrm{H}}^0$, there exists $\delta>0$ such that $\mathbb{E}\left[\frac{\partial\log(\mathcal{L}(R_t))}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right]<-\delta$ for every $\mathcal{F}_{t-1}$, implying that $\mathbb{E}\left[\frac{\partial\log(\mathcal{L}(R_t|\sigma_{\mathrm{H}}))}{\partial\sigma_{\mathrm{H}}}\right]<-\delta$.

Hereafter, we will evaluate the conditional expectation at $\sigma_{\mathrm{H}}=\sigma_{\mathrm{H}}^0$. From (A.5), note that $\mathbb{E}\left[\frac{\partial\log(\mathcal{L}(R_t))}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right]$ is decomposed as follows:

$$\begin{aligned}\mathbb{E}\left[\frac{\partial\log\left(\mathcal{L}\left(R_t\right)\right)}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\frac{\lambda_{t-1}}{\mathcal{L}\left(R_t\right)}\frac{\partial\phi\left(R_t|\sigma_{\mathrm{H}}\right)}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right]\\ &+ \mathbb{E}\left[\frac{1}{\mathcal{L}\left(R_t\right)}\left(\phi\left(R_t|\sigma_{\mathrm{H}}\right)-\phi\left(R_t|\sigma_{\mathrm{L}}\right)\right)|\mathcal{F}_{t-1}\right]\frac{\partial\lambda_{t-1}}{\partial\sigma_{\mathrm{H}}}.\end{aligned} \tag{A.10}$$

To determine the sign of each component in RHS of (A.10), we need the conditional distribution of R_t. Since λ_{t-1}^* is the true probability of $S_t=H$ given $\mathcal{F}_{t-1}$ and ε_t in (A.1) is drawn from a binomial distribution of 1 or -1 with equal probability, the probability mass of R_t over $(-\sigma_{\mathrm{H}},-\sigma_{\mathrm{L}},\sigma_{\mathrm{L}},\sigma_{\mathrm{H}})$ equals $\left(\frac{\lambda_{t-1}^*}{2},\frac{1-\lambda_{t-1}^*}{2},\frac{1-\lambda_{t-1}^*}{2},\frac{\lambda_{t-1}^*}{2}\right)$.

First, we determine the sign of $\mathbb{E}\left[\frac{\lambda_{t-1}}{\mathcal{L}(R_t)}\frac{\partial\phi(R_t|\sigma_{\mathrm{H}})}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right]$. From the properties of the normal density, it follows that $\frac{\partial\phi(x|\sigma)}{\partial\sigma}=\phi(x|\sigma)\left(-\frac{1}{\sigma}+\frac{x^2}{\sigma^3}\right)$ and $\phi(-x|\sigma)=\phi(x|\sigma)$. Hence

$$\begin{aligned}\mathbb{E}\left[\frac{\lambda_{t-1}}{\mathcal{L}}\frac{\partial\phi\left(R_t|\sigma_{\mathrm{H}}\right)}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right] &= \frac{\lambda_{t-1}^*}{2}\sum_{R_t=-\sigma_{\mathrm{H}},\sigma_{\mathrm{H}}}\frac{\lambda_{t-1}}{\mathcal{L}\left(R_t\right)}\phi(R_t|\sigma_{\mathrm{H}})\left(-\frac{1}{\sigma_{\mathrm{H}}}+\frac{R_t^2}{\sigma_{\mathrm{H}}^3}\right)\\ &+ \frac{1-\lambda_{t-1}^*}{2}\sum_{R_t=-\sigma_{\mathrm{L}},\sigma_{\mathrm{L}}}\frac{\lambda_{t-1}}{\mathcal{L}\left(R_t\right)}\phi(R_t|\sigma_{\mathrm{H}})\left(-\frac{1}{\sigma_{\mathrm{H}}}+\frac{R_t^2}{\sigma_{\mathrm{H}}^3}\right)\\ &= \frac{\left(1-\lambda_{t-1}^*\right)\lambda_{t-1}}{\mathcal{L}(\sigma_{\mathrm{L}})}\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}})\left(-\frac{1}{\sigma_{\mathrm{H}}}+\frac{\sigma_{\mathrm{L}}^2}{\sigma_{\mathrm{H}}^3}\right)\\ &< -\left(1-p\right)^2\frac{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}})}{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}})}\left(\frac{\sigma_{\mathrm{H}}^2-\sigma_{\mathrm{L}}^2}{\sigma_{\mathrm{H}}^3}\right),\end{aligned} \tag{A.11}$$

where the last inequality is from (A.6) and $\mathcal{L}(\sigma_{\mathrm{L}})<\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}})$.

Next, from the property, $\phi(-x|\sigma) = \phi(x|\sigma)$, and the fact that $\phi(x|\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$, the sign of $\mathbb{E}\left[\frac{1}{\mathcal{L}}\left(\phi(R_t|\sigma_{\mathrm{H}}) - \phi(R_t|\sigma_{\mathrm{L}})\right)|\mathcal{F}_{t-1}\right]$ is determined as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{\mathcal{L}}\left(\phi(R_t|\sigma_{\mathrm{H}}) - \phi(R_t|\sigma_{\mathrm{L}})\right)|\mathcal{F}_{t-1}\right] \\
=\;& \frac{\lambda^*_{t-1}}{2} \sum_{R_t=-\sigma_{\mathrm{H}},\sigma_{\mathrm{H}}} \left(\frac{\phi(R_t|\sigma_{\mathrm{H}}) - \phi(R_t|\sigma_{\mathrm{L}})}{\mathcal{L}(R_t)}\right) + \frac{1-\lambda^*_{t-1}}{2} \sum_{R_t=-\sigma_{\mathrm{L}},\sigma_{\mathrm{L}}} \left(\frac{\phi(R_t|\sigma_{\mathrm{H}}) - \phi(R_t|\sigma_{\mathrm{L}})}{\mathcal{L}(R_t)}\right) \\
=\;& \lambda^*_{t-1}\left(\frac{\phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{H}}) - \phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{L}})}{\mathcal{L}(\sigma_{\mathrm{H}})}\right) + \left(1-\lambda^*_{t-1}\right)\left(\frac{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}}) - \phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}})}{\mathcal{L}(\sigma_{\mathrm{L}})}\right) \\
=\;& \lambda^*_{t-1}\left(\frac{\phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{H}}) - \phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{L}})}{\mathcal{L}(\sigma_{\mathrm{H}})}\right) + \left(1-\lambda^*_{t-1}\right)\frac{\mathcal{L}(\sigma_{\mathrm{H}})}{\mathcal{L}(\sigma_{\mathrm{L}})}\left(\frac{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}}) - \phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}})}{\mathcal{L}(\sigma_{\mathrm{H}})}\right) \\
>\;& \frac{1}{\mathcal{L}(\sigma_{\mathrm{H}})}\left(\lambda^*_{t-1}\left(\phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{H}}) - \phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{L}})\right) + \left(1-\lambda^*_{t-1}\right)\left(\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}}) - \phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}})\right)\right) \\
>\;& \frac{1}{\mathcal{L}(\sigma_{\mathrm{H}})}\left((1-p)\left(\phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{H}}) - \phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{L}})\right) + p\left(\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}}) - \phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}})\right)\right) > 0,
\end{aligned}
\tag{A.12}
$$

where the last three inequalities can be verified by (A.6) and the given parameter values of (A.3).

Finally, we show that $\frac{\partial\lambda_{t-1}}{\partial\sigma_{\mathrm{H}}} \leq 0$ by induction. We assume that λ_0 is determined as the steady state distribution determined by (A.2). Since λ_0 does not depend on σ_{H}, the following holds:

$$
\frac{\partial\lambda_0}{\partial\sigma_{\mathrm{H}}} = 0. \tag{A.13}
$$

Next, we show that $\frac{\partial\lambda_{t-1}}{\partial\sigma_{\mathrm{H}}} \leq 0$ implies $\frac{\partial\lambda_t}{\partial\sigma_{\mathrm{H}}} \leq 0$. Note that the process of $\{\lambda_t\}_{t=0}^{T}$ is constructed by the following recursion:

$$
\widetilde{\lambda}_t = \frac{\lambda_{t-1}\phi(R_t|\sigma_{\mathrm{H}})}{\lambda_{t-1}\phi(R_t|\sigma_{\mathrm{H}}) + (1-\lambda_{t-1})\phi(R_t|\sigma_{\mathrm{L}})}, \tag{A.14}
$$

and

$$
\lambda_t = p\widetilde{\lambda}_t + (1-p)\left(1-\widetilde{\lambda}_t\right). \tag{A.15}
$$

Equation (A.14) describes how the econometrician updates the probability on the hidden state of S_t using the misspecified normal likelihood after observing R_t. Equation (A.15) shows how the econometrician predicts the hidden state of S_{t+1} with the given information set $\mathcal{F}_t$ through the transition matrix given in (A.2). Combining (A.14) and (A.15), we get

$$
\frac{\lambda_t + p - 1}{2p-1} = \frac{\lambda_{t-1}\phi(R_t|\sigma_{\mathrm{H}})}{\lambda_{t-1}\phi(R_t|\sigma_{\mathrm{H}}) + (1-\lambda_{t-1})\phi(R_t|\sigma_{\mathrm{L}})}. \tag{A.16}
$$

Taking the derivative of (A.16) with respect to σ_{H}, we obtain the following:

$$
\frac{1}{2p-1}\frac{\partial\lambda_t}{\partial\sigma_{\mathrm{H}}} = \frac{\partial\frac{\lambda_{t-1}\phi(R_t|\sigma_{\mathrm{H}})}{\lambda_{t-1}\phi(R_t|\sigma_{\mathrm{H}})+(1-\lambda_{t-1})\phi(R_t|\sigma_{\mathrm{L}})}}{\partial\lambda_{t-1}}\frac{\partial\lambda_{t-1}}{\partial\sigma_{\mathrm{H}}} + \frac{\partial\frac{\lambda\phi(R_t|\sigma_{\mathrm{H}})}{\lambda\phi(R_t|\sigma_{\mathrm{H}})+(1-\lambda)\phi(R_t|\sigma_{\mathrm{L}})}}{\partial\phi(R_t|\sigma_{\mathrm{H}})}\frac{\partial\phi(R_t|\sigma_{\mathrm{H}})}{\partial\sigma_{\mathrm{H}}}. \tag{A.17}
$$

To determine the sign of each component in RHS of (A.17), we use the following properties:

$$\frac{\partial \frac{\lambda m}{\lambda m+(1-\lambda)n}}{\partial \lambda} = \frac{mn}{(\lambda m+(1-\lambda)n)^2} > 0 \tag{A.18}$$

$$\frac{\partial \frac{\lambda m}{\lambda m+(1-\lambda)n}}{\partial m} = \frac{\lambda(1-\lambda)n}{(\lambda m+(1-\lambda)n)^2} > 0 \tag{A.19}$$

for $m, n > 0$ and $\lambda \in (0,1)$. Further, using the properties of $\frac{\partial \phi(x|\sigma)}{\partial \sigma} = \phi(x|\sigma)\left(-\frac{1}{\sigma}+\frac{x^2}{\sigma^3}\right)$ and $\phi(x|\sigma) = \phi(-x|\sigma)$, we have that

$$\begin{aligned}\frac{\partial \phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{H}})}{\partial \sigma_{\mathrm{H}}} &= \phi(\sigma_{\mathrm{H}}|\sigma_{\mathrm{H}})\left(-\frac{1}{\sigma_{\mathrm{H}}}+\frac{\sigma_{\mathrm{H}}^2}{\sigma_{\mathrm{H}}^3}\right) = 0\\ \frac{\partial \phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}})}{\partial \sigma_{\mathrm{H}}} &= \phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}})\left(-\frac{1}{\sigma_{\mathrm{H}}}+\frac{\sigma_{\mathrm{L}}^2}{\sigma_{\mathrm{H}}^3}\right) < 0,\end{aligned}$$

implying

$$\frac{\partial \phi(R_t|\sigma_{\mathrm{H}})}{\partial \sigma_{\mathrm{H}}} \leq 0 \tag{A.20}$$

for every possible realization of R_t from $\{-\sigma_{\mathrm{H}}, -\sigma_{\mathrm{L}}, \sigma_{\mathrm{L}}, \sigma_{\mathrm{H}}\}$. With the assumption that $\frac{\partial \lambda_{t-1}}{\partial \sigma_{\mathrm{H}}} \leq 0$, inequalities of (A.18), (A.19), and (A.20) ensure that RHS of (A.17) is non-positive. Hence, with $p > 1/2$ as assumed in (A.3), it follows that $\frac{\partial \lambda_t}{\partial \sigma_{\mathrm{H}}} \leq 0$. Combining (A.13) with this finding, we conclude that

$$\frac{\partial \lambda_{t-1}}{\partial \sigma_{\mathrm{H}}} \leq 0, \tag{A.21}$$

for every possible information set of $\mathcal{F}_{t-1}$.

Recall that we want to show that (A.10) is strictly negative. Finally, combining (A.11), (A.12), and (A.20), we conclude that

$$\mathbb{E}\left[\frac{\partial \log(\mathcal{L}(R_t|\sigma_{\mathrm{H}}))}{\partial \sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right] < -\delta, \tag{A.22}$$

where

$$\delta = (1-p)^2 \frac{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}})}{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}})}\left(\frac{\sigma_{\mathrm{H}}^2-\sigma_{\mathrm{L}}^2}{\sigma_{\mathrm{H}}^3}\right) > 0, \tag{A.23}$$

completing the proof that QML estimate of $\widehat{\sigma}_{\mathrm{H}}$ in (A.7) will not converge to the true parameter value.

Appendix B Additional Table

Table A1: Option-like Feature of Momentum Returns and Market Conditions

We partition the months in our sample into three groups: ‘High’ group is made up of months when variable describing the market conditions (past market returns, realized volatility of the market, or leverage of loser portfolio stocks) was in the top 20th percentile and the ‘Low’ group corresponds to months when the market condition variable was in the bottom 20th percentile. The rest of the months are classified as ‘Medium’. The sample period is 1927:07-2017:12. In Panel A, we group the months based on the realized volatility of daily market returns over the previous 12 months. In Panel B, we use the breakpoints of the loser portfolio for grouping. We then pool the months within each group and analyze the behavior of momentum strategy returns. Specifically, we estimate equation (1) with ordinary least squares using momentum strategy returns (R_{MOM}) and the returns of winner and loser portfolio in excess of risk free return (R^e_{WIN} and R^e_{LOS}) as LHS variables and report results in Panel A-1-i and B-1-i. For comparison, we report the estimates for the CAPM, without the exposure to the call option on the market in (1), in Panel A-1-ii and B-1-ii. Then, we count the numbers of large momentum losses worse than negative 20% within the groups and report those in Panel A-2 and B-2. Finally, we compare the skewness of $R^e_{p,t}$ with that of estimated ε of (1) in Panel A-3 and B-3. α is reported in percentage per month. The t-statistics are computed using the heteroscedasticity-consistent covariance estimator by White (1980).

Continued on next page

Panel A: Past 12 Months Realized Volatility of Market Returns									
	High			Medium			Low		
LHS	R_{MOM}	R^e_{WIN}	R^e_{LOS}	R_{MOM}	R^e_{WIN}	R^e_{LOS}	R_{MOM}	R^e_{WIN}	R^e_{LOS}
A-1: Option-like features									
A-1-i: Henriksson-Merton Estimates									
α	3.10	1.07	-2.03	1.77	0.67	-1.10	2.41	1.28	-1.13
$t(\alpha)$	(3.33)	(2.90)	(-2.94)	(5.11)	(3.67)	(-4.41)	(5.31)	(4.06)	(-4.30)
β^0	-0.58	0.95	1.53	0.11	1.33	1.22	0.50	1.53	1.03
$t(\beta^0)$	(-4.83)	(14.09)	(17.99)	(1.09)	(25.90)	(16.50)	(2.94)	(14.97)	(8.95)
β^+	-0.89	-0.28	0.61	-0.21	-0.16	0.05	-0.66	-0.42	0.25
$t(\beta^+)$	(-3.35)	(-2.35)	(3.32)	(-1.10)	(-1.65)	(0.37)	(-2.26)	(-2.21)	(1.36)
$Adj.R^2$	0.49	0.74	0.83	0.00	0.78	0.66	0.03	0.72	0.66
A-1-ii: CAPM Estimates									
α	0.40	0.23	-0.18	1.41	0.40	-1.01	1.56	0.75	-0.81
$t(\alpha)$	(0.60)	(0.83)	(-0.35)	(6.44)	(3.45)	(-6.59)	(5.35)	(4.12)	(-4.37)
β	-1.08	0.79	1.87	0.01	1.26	1.25	0.13	1.30	1.17
$t(\beta)$	(-8.63)	(15.54)	(21.93)	(0.20)	(42.26)	(29.56)	(1.36)	(23.79)	(18.96)
$Adj.R^2$	0.45	0.73	0.82	0.00	0.78	0.66	0.00	0.72	0.66
A-2: Number of Momentum Losses worse than -20%									
	12			0			0		
A-3: Conditional Skewness									
R^e_p	-1.86	-0.23	1.44	-0.21	-0.63	-0.27	-0.12	-0.14	0.17
ε_p	-0.77	-1.08	0.85	-0.21	0.25	0.36	-0.16	0.58	0.28

Continued on next page

Table A1 – continued from previous page

Panel B: Breakpoints of Loser Portfolio									
	Low			Medium			High		
LHS	R_{MOM}	R^e_{WIN}	R^e_{LOS}	R_{MOM}	R^e_{WIN}	R^e_{LOS}	R_{MOM}	R^e_{WIN}	R^e_{LOS}
B-1: Option-like features									
B-1-i: Henriksson-Merton Estimates									
α	2.57	0.90	-1.67	2.65	1.10	-1.55	0.87	0.33	-0.55
$t(\alpha)$	(2.71)	(2.38)	(-2.38)	(5.68)	(5.84)	(-4.49)	(1.49)	(0.89)	(-1.60)
β^0	-0.69	0.91	1.60	0.17	1.37	1.20	0.52	1.47	0.95
$t(\beta^0)$	(-5.87)	(14.52)	(18.01)	(1.36)	(25.52)	(13.18)	(2.98)	(11.10)	(13.07)
β^+	-0.84	-0.28	0.57	-0.57	-0.32	0.25	-0.15	-0.08	0.08
$t(\beta^+)$	(-3.21)	(-2.46)	(3.03)	(-1.94)	(-2.95)	(1.16)	(-0.46)	(-0.40)	(0.37)
$Adj.R^2$	0.51	0.71	0.83	0.03	0.79	0.67	0.16	0.81	0.75
B-1-ii: CAPM Estimates									
α	0.12	0.10	-0.02	1.72	0.58	-1.14	0.61	0.20	-0.41
$t(\alpha)$	(0.17)	(0.35)	(-0.04)	(8.55)	(5.52)	(-7.75)	(2.02)	(0.89)	(-2.66)
β	-1.15	0.76	1.91	-0.11	1.21	1.32	0.44	1.43	0.99
$t(\beta)$	(-9.54)	(15.64)	(22.87)	(-1.32)	(34.86)	(21.39)	(4.57)	(25.75)	(15.94)
$Adj.R^2$	0.48	0.70	0.82	0.01	0.79	0.67	0.16	0.81	0.76
B-2: Number of Momentum Losses worse than -20%									
		11			1			0	
B-3: Conditional Skewness									
R^e_p	-1.64	-0.02	1.42	-1.22	-0.73	0.43	0.01	-0.52	0.18
ε_p	-0.46	-0.09	0.83	-1.11	-0.30	0.77	-0.21	0.27	1.05