May 12, 2016 Comments Welcome

Tail Risk in Momentum Strategy Returns

Kent Daniel, Ravi Jagannathan and Soohun Kim†

Abstract

Momentum strategy returns are highly left skewed and leptokurtic. These features are explained by the leverage dynamics of the portfolio of past losers: under certain market conditions, the stocks in the past loser portfolio become highly levered, embedding a call option on the market. We estimate this leverage using a hidden Markov model that uses the convexity in the relation between the market and the momentum portfolio to, effectively, estimate the leverage and the likelihood of tail-events. The estimated HMM predicts these tail events better than alternative models both in- and out-of-sample. Moreover, the momentum residuals from this model are normally distributed; all of the non-normality of the momentum portfolio results from the time-varying interaction with the market return. The dramatic momentum crashes appear to be fully a result of the dynamics of the relation between the market and momentum returns, rather than black-swan like shocks.

†Columbia Business School and NBER (Daniel), Northwestern University, Kellogg School of Management and NBER (Jagannathan), and Georgia Institute of Technology, Scheller College of Business (Kim). This paper originally appeared under the title "Risky Cycles in Momentum Returns." We thank Raul Chhabbra, Randolph B. Cohen, Zhi Da, Gangadhar Darbha, Francis Diebold, Robert Engel, Bryan T. Kelly, Robert Korajczyk, Jonathan Parker, Prasanna Tantri, the seminar participants at the Fourth Annual Conference of the Society for Financial Econometrics, Northwestern University, the Indian School of Business, the Securities and Exchange Board of India, Shanghai Advance Institute for Finance, Shanghai Jiao Tong University, the Fifth Annual Triple Crown Conference in Finance, Fordham University, University of Chicago, University of Virginia, WFA 2012 meeting, and Nomura Securities for helpful comments on the earlier version of the paper. In particular, we thank the editor and two anonymous referees for insightful comments and helpful suggestions. We alone are responsible for any errors and omissions.

1 Introduction

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 a component of their overall strategy, and even fundamental managers appear to incorporate momentum in formulating their trading decisions.[1]

Notwithstanding their inherent simplicity, momentum strategies have been profitable across many asset classes and in multiple geographic regions.[2] Over our sample period of 1044 months from 1927:01 to 2013:12, our baseline momentum strategy produced monthly returns with a mean of 1.18% and a standard deviation of 7.94%, generating an annualized Sharpe ratio of 0.52.[3] In contrast, over this same period the three Fama and French (1993) factor portfolios – Mkt-Rf, SMB, and HML – had annualized Sharpe Ratios of 0.41, 0.26, and 0.39, respectively. The profitability of this momentum strategy after adjusting for exposure to economy wide systematic risks is still higher: the CAPM alpha is 1.52%/month ($t = 7.10$), and the Fama and French (1993) three-factor alpha is 1.76%/month ($t = 8.20$).[4]

While the momentum strategy's average risk adjusted return has been high, the strategy has experienced infrequent but large losses. The historical distribution of momentum strategy returns is highly left skewed. Consistent with the large estimated negative skewness, over our sample there are eight months in which the momentum strategy has lost more than 30%, and none in which it has earned more than 30% (the highest monthly return is 26.18%). Moreover, the magnitude of momentum's largest losses has been extreme. The

[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."

[2]Asness et al. (2013) provide extensive cross-sectional evidence on momentum effects. Chabot et al. (2014) find the momentum effect in the Victorian era UK equity market.

[3]Our baseline 12-2 momentum strategy, described in more detail later, ranks firms based on their cumulative returns from months $t-12$ through $t-2$, and takes 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.

[4]The t-statistics are computed using the heteroskedasticity-consistent covariance estimator by White (1980).

worst monthly return was -79.57%, and six monthly losses exceed 40%. normality can be easily rejected. Also, as Daniel and Moskowitz (2015) document, these large losses cluster, and tend to occur when the market rebounds sharply following a prolonged depressed condition.

The focus of this paper is modeling time variation in the tail risk of momentum strategies. 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 here follows 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 of these firms become at- or out-of-the-money call options on the firm's underlying assets, and 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 the dramatic losses for the overall long-short momentum portfolio.

We show that the dynamics of reported financial leverage are consistent with this hypothesis: going into the five worst momentum crash months, financial leverage of the loser portfolio averaged 47.2, more than an order of magnitude higher than unconditional average of 3.97.[5] Of course, 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 wages of crucial employees, the fixed costs associated with maintenance of property, plant and equipment, etc. If these fixed costs are large, even a firm with zero debt may see its equity start to behave like an out-of-the-money option following large losses. One recent episode consistent with this 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

[5]These are the averages over the 1964-2013 period over which we have data on the book value of debt.

the leverage dynamics of the momentum portfolio using 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. 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 approximately normally distributed.[6] 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. Consistent with this, 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 GARCH models.

Interestingly, we find that it is the incorporation of the optionality in the HMM that is key to the model's ability to forecast these tail events. A version of the HMM which 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 20% more false positives than the baseline HMM specification, suggesting 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.

The literature examining price momentum is vast. While the focus in this literature has been on documenting and explaining the strategy's high average returns[7] and unconditional

[6]Interesting the market-returns residuals have a Student-t distribution with 5 degrees of freedom. We account for this non-normality in one our HMM specifications and show that accounting for this non-normality substantially improves the performance of the model in forecasting tail-events.

[7] See Daniel et al. (1998), Barberis et al. (1998), Hong and Stein (1999) and Liu and Zhang (2008) for examples.

risk exposures, 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 (2015) 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 momentum strategy's exposure to standard systematic risk factors. In contrast to most of this literature, our focus here 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 better than other forecasting techniques suggested by the literature.

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 depreciates sharply. When a number of investors follow dynamic strategies that have embedded options on the market of the same type, crashes can be exacerbated with the potential to trigger systemic responses.

While our focus is in modeling systematic stochastic variations in the tail risk of momentum returns—which we find is due to its embedded option on the market like features—our findings also have implications for estimating the abnormal returns to the momentum strategy. It is well recognized in the literature that payoffs on self financing zero cost portfolios that have positions in options can exhibit spurious positive value (alpha) when alpha is computed using the market model or linear beta models in general.[8] We therefore calculate an option-adjusted abnormal performance for the momentum strategy. As might be anticipated, we find that alpha of the momentum strategy is generally strongly positive and statistically significant. 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.

[8] See Glosten and Jagannathan (1994) for an example.

The rest of this paper is organized as follows. In Section 2, 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 3, we describe a hidden Markov model for momentum return generating process that captures this feature of tail risk in momentum strategy returns. In Section 4, we show the ability of our hidden Markov model to predict momentum crashes. In Section 5, we evaluate the conditional alpha of momentum strategy returns based on the estimated parameters of our hidden Markov model and option market prices. Section 6 concludes.

2 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 market index portfolio. The nature of non-linear relationship depends on market conditions. This examination motivates the two-state model that we develop in Section 3.

2.1 Characteristics of Momentum Strategy Returns

Price momentum strategies have been constructed using variety of metrics. For this study we examine a cross-sectional equity strategy in US common stocks. Our universe consists of all US 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 $t-2$.[9] 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

[9]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).

the zero-investment portfolio "MOM" which is long the past-winner decile, and short the past-loser decile.

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.[10] Without risk adjustment the momentum strategy earns an average return of 1.18%/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 1.52%/month ($t = 7.10$) and its Fama and French (1993) three factor model alpha is 1.76%/month ($t = 8.20$).[11] 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. Figure 1.A illustrates this graphically: we plot the smoothed empirical density for MOM returns (the dashed red line) and a normal density with the same mean and standard deviation. Overlayed on the density function plot are red dots that represent the 25 MOM returns that exceed 20% in absolute value (13 in the left tail and 12 in the right tail). Figure 1.B overlays the empirical density market excess returns which are scaled to match the volatility of MOM returns over this sample period. 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.

Consistent with the results in Table 1, Figure 1 reveals that both the market and momentum strategy are leptokurtic. However, Panel B in particular shows the strong left skewness of momentum. Again, one of the objectives of this paper is to show that this skewness is completely a result of the time-varying non-linear relationship between market and momentum returns that is a result of the time-varying leverage of the firms in the loser portfolio. As a way of motivating our model, we next examine the influence of prevailing various state variables on market conditions on momentum strategy returns.

To begin, Table 2 lays out the MOM returns in the 13 months when the MOM loss exceeded 20%, and measures of various market conditions that prevailed during the months. The first set of columns show that the large momentum strategy losses are generally asso-

[10] We obtain the data of the three factors in Fama and French (1993) model of Mkt-Rf, SMB and HML from Kenneth French's database.

[11] The t-statistics are computed using the heteroskedasticity-consistent covariance estimator by White (1980).

ciated with to large gains on the past-loser portfolio rather than losses in the past-winner portfolio. During the 13 large loss months, the loser portfolio earned an excess return 45.69% whereas the winner portfolio earned only by 6.32%. Interestingly, these loser portfolio gains are associated with large contemporaneous gains in the market portfolio, which earns an average excess return of 16.14% in these months. However, the table also shows that market return is strongly negative and volatile in the period leading up to the momentum crashes: the market is down, on average, by more than 37% 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.[12] Given the past losses high volatility of the market, it is not surprising that the past-loser portfolio has suffered severe losses: the threshold (breakpoint) for a stock to be in the loser portfolio averaged -63.77% in these 13 months, about 2.7 times the average breakpoint. Thus, at the start of the crash months are likely very highly levered. Table 2 also shows that the average financial leverage (book value of debt/market value of equity), during the 5 loss months after 1964 (when our leverage data starts) is 47.2, more than an order of magnitude higher than the average leverage of the loser portfolio of 3.97.

To summarize, large momentum strategy losses generally occur have occurred in volatile bear markets, when the past-losers have lost a substantial fraction of their market value, and consequently have high financial leverage, and probably high operating leverage as well. Thus, following Merton (1974), the equity of these firms are likely to behave like out-of-the-money call options on the underlying firm values which, in aggregate, are 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.

2.2 Time Varying Option-like Features of Momentum Strategy

Motivated by the evidence in the preceding Section, we here 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 will develop in Section 3.

In particular, we consider the following augmented market model return generating pro-

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

cess, similar to that considered by Henriksson and Merton (1981) and others.[13]

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

where $R^e_{\mathrm{MKT},t}$ is the market portfolio returns in excess of risk free return for month t. We note that α, 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 (Glosten and Jagannathan, 1994). We return to this issue and estimate the abnormal return of the strategy in Section 5. For the moment, we concentrate on the time-variation in β^+, which is a measure of the exposure of the portfolio p to the payoff on a one-month call option on the stock market or, equivalently, a measure of the convexity in the relationship 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 preceding the portfolio formation month; the realized volatility of daily market returns over the previous 12 months; 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 1044 months into 'High', 'Medium' and 'Low' groups. 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 C.5 in the Online Appendix.[14]

[13] To our knowledge, Chan (1988) and DeBondt and Thaler (1987) first document that the market beta of a long-short winner-minus-loser portfolio is non-linearly related to the market return, though 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-US markets. Daniel and Moskowitz (2015) 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) needs for caution.

[14] Results are similar when we group sample with other market conditions – cumulative market return during the 12 month preceding the portfolio formation month, the realized volatility of daily market returns over the previous 6 months and the ratio of the book value of debt to the market value of equity (BD/MV)

Panel A presents the estimates of equation (1) for the momentum strategy returns (R_{MOM}), and for the returns of the winner and loser portfolio in excess of the risk free rate (R^e_{WIN} and R^e_{LOS}). First, note that the estimated β^+, the exposure to the market call payoff is significant only when the the past 36-month market returns are in 'Low' group: consistent with the leverage hypothesis, the the past-loser has a positive exposure to the market option payoff of 0.72 (t=3.60). That is, it behaves like a call option on the market. The MOM portfolio, which is short the past-losers, thus has a a significantly negative β^+..[15] In contrast, in the 'Medium' and 'High' group, β^+ of the MOM returns and of the long- and short-sides are smaller in absolute value and are not statistically significantly negative.[16] Interestingly the Low State, the $Adj.R^2$ is 48% for MOM returns, as compared to 6% in both the 'Medium' and 'High' states, a result of both the higher β^0 and β^+ in the Low state.

Panel C shows that large MOM losses (crashes) are clustered in months when the option-like feature of β^+ is accentuated; 11 out of 13 momentum losses occur during months in the 'Low' group. Table C.5 shows that the results when the grouping is on 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 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 A, the skewness of the momentum strategy returns is -2.33, but after we control for the non-linear exposure to the market through equation (1), the skewness of residual drops to -0.48. In 'Medium' and 'High' group, the negative skewness of 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; β^+ is not significantly different from zero in the other two groups. The results reported in Table C.5 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.

---

of the loser stock portfolio.

[15] The t-statistics are computed using the heteroskedasticity-consistent covariance estimator by White (1980).

[16] We note that β^+ of winner and loser portfolios exhibit interesting patterns: It is negative and significant for winner stocks in 'Low' group. It is negative and statistically significant for loser stocks in the 'High' group. Understanding why we see these patterns is left for future research.

The above results suggest that the embedded written call option on the market is the key driver of momentum crashes, and that this optionality is a result of the high leverage of the past-loser firms. However this leverage will not always be apparent in the *financial* leverage of the past-loser portfolio. For example, it is likely that the *operating* leverage of many of the firms that earned low returns in the post-March 2000 collapse of the tech sector was quite high, even though these firms' financial leverage was insignificant. The evidence is consistent with this: the financial leverage of the loser portfolio was low during two episodes of large momentum losses in 2001:01 and 2002:12.[17] However, as can be seen from Table 4, 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.

In the next section, we model the option-like relation 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.) and the convexity. However, this requires choosing the length of the time window over which these state-variables are measured, and that necessarily has to be rather 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 return is accentuated, and momentum crashes are likely. This naturally leads us to the two-state hidden Markov model (HMM) for estimating this state, which we explore in the next Section.

3 Model

In this section 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. We

[17] Refer to Table 2. In 2001:01 (2002:12), the momentum strategy loses -41.97% (-20.40%) and the financial leverage (BD/MV) of loser portfolio was 0.68 (2.32). The average of financial leverage over all available data from 1964 is 3.97.

then show that an estimated HMM allows *ex-ante* estimation of this this state variable based on the history of returns on the momentum strategy and the market portfolio.

3.1 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 return generating process of the momentum strategy is as follows:

$$R_{\text{MOM},t} = \alpha(S_t) + \beta^0(S_t)R^e_{\text{MKT},t} + \beta^+(S_t)\max\left(R^e_{\text{MKT},t}, 0\right) + \sigma_{\text{MOM}}(S_t)\,\varepsilon_{\text{MOM},t}, \tag{2}$$

where $\varepsilon_{\text{MOM},t} \sim i.i.d\ \mathcal{N}(0,1)$. Equation (2) 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 returns in excess of risk free rate is given by:

$$R^e_{\text{MKT},t} = \mu\left(S_t\right) + \sigma_{\text{MKT}}\left(S_t\right)\varepsilon_{\text{MKT},t}, \tag{3}$$

where $\varepsilon_{\text{MKT},t} \sim i.i.d\ \mathcal{N}(0,1)$. That is, $\mu\left(S_t\right)$ and $\sigma_{\text{MKT}}\left(S_t\right)$ 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 as given below:

$$\Pi = \begin{bmatrix} \Pr(S_t = C|S_{t-1} = C) & \Pr(S_t = T|S_{t-1} = C) \\ \Pr(S_t = C|S_{t-1} = T) & \Pr(S_t = T|S_{t-1} = T) \end{bmatrix}, \tag{4}$$

where S_t, the unobservable random state at time t which, in our setting, is either *Calm*(C) or *Turbulent*(T) and $\Pr\left(S_t = s_t|S_{t-1} = s_{t-1}\right)$ denotes the probability of transitioning from state s_{t-1} at time $t-1$ to state s_t at time t.[18]

[18]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.

3.2 Quasi Maximum Likelihood Estimation

We now estimate the set of parameters of the hidden Markov model in equations (2), (3), and (4), which we summarize with the 14-element parameter vector θ^0:

$$\theta^0 = \left\{ \begin{array}{c} \alpha(C), \beta^0(C), \beta^+(C), \sigma_{\mathrm{MOM}}(C), \\ \alpha(T), \beta^0(T), \beta^+(T), \sigma_{\mathrm{MOM}}(T), \\ \mu(C), \sigma_{\mathrm{MKT}}(C), \mu(T), \sigma_{\mathrm{MKT}}(T), \\ \Pr(S_t = C | S_t = C), \Pr(S_t = \mathrm{T} | S_t = \mathrm{T}) \end{array} \right\}. \tag{5}$$

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

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

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 under the assumption that $\varepsilon_{\mathrm{MOM},t}$ and $\varepsilon_{\mathrm{MKT},t}$ in (2) and (3) are jointly normally distributed. When $\varepsilon_{\mathrm{MOM},t}$ and $\varepsilon_{\mathrm{MKT},t}$ are not normally distributed, the resulting estimator is referred to as QML (Quasi Maximum Likelihood). As we discuss in Appendix A, when this assumption is violated, the QML estimator of θ^0 can be inconsistent.[19] As we discuss later in more detail, while the momentum returns $R_{\mathrm{MOM},t}$ are highly skewed and leptokurtic, the momentum return residuals ($\varepsilon_{\mathrm{MOM},t}$) appears normally distributed. Interestingly, the market return residual is non-normal—it is better characterized as Student-t distributed with (d.f.=5)–but we show in Appendix A that given these distribution for the residuals, the QML estimator provides reasonably well behaved estimates.

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

$$\mathcal{L} = \sum_{t=1}^{T} \log\left(\Pr\left(\mathbf{r}_t | \mathcal{F}_{t-1}\right)\right), \tag{6}$$

where $\mathcal{F}_{t-1}$ denotes the agent's time $t-1$ information set (*i.e.* all market and momentum

[19] In contrast to what it sometimes assumed in the literature, QML in this setting is not consistent. See especially Appendix A.1, where we demonstrate this result.

excess returns up through time $t-1$).

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

$$\Pr\left(\mathbf{r}_t|\mathcal{F}_{t-1}\right) = \sum_{s_t \in \{C,T\}} \Pr\left(\mathbf{r}_t, S_t = s_t|\mathcal{F}_{t-1}\right), \tag{7}$$

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:

$$\begin{aligned}\Pr\left(\mathbf{r}_t, S_t = s_t|\mathcal{F}_{t-1}\right) &= \Pr\left(\mathbf{r}_t|S_t = s_t, \mathcal{F}_{t-1}\right)\Pr\left(S_t = s_t|\mathcal{F}_{t-1}\right)\\ &= \Pr\left(\mathbf{r}_t|S_t = s_t\right)\Pr\left(S_t = s_t|\mathcal{F}_{t-1}\right).\end{aligned} \tag{8}$$

The first term of equation (8) is the state dependent likelihood of $\mathbf{r}_t$ which, under the distributional assumptions from (2) and (3), is given by

$$\Pr\left(\mathbf{r}_t|S_t = s_t\right) = \frac{1}{\sqrt{2\pi}\sigma_{\text{MOM}}\left(s_t\right)}\exp\left\{-\frac{\left(\varepsilon_{\text{MOM},t}\right)^2}{2}\right\} \times \frac{1}{\sqrt{2\pi}\sigma_{\text{MKT}}\left(s_t\right)}\exp\left\{-\frac{\left(\varepsilon_{\text{MKT},t}\right)^2}{2}\right\},$$

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 of equation (8)—the likelihood that the unobserved state S_t is $s_t \in \{C, T\}$ conditional on $\mathcal{F}_{t-1}$—can be written as a function of the time $t-1$ state probabilities as:

$$\begin{aligned}\Pr\left(S_t = s_t|\mathcal{F}_{t-1}\right) &= \sum_{s_{t-1}\in\{C,T\}} \Pr\left(S_t = s_t, S_{t-1} = s_{t-1}|\mathcal{F}_{t-1}\right)\\ &= \sum_{s_{t-1}\in\{C,T\}} \Pr\left(S_t = s_t|S_{t-1} = s_{t-1}, \mathcal{F}_{t-1}\right)\Pr\left(S_{t-1} = s_{t-1}|\mathcal{F}_{t-1}\right)\\ &= \sum_{s_{t-1}\in\{C,T\}} \Pr\left(S_t = s_t|S_{t-1} = s_{t-1}\right)\Pr\left(S_{t-1} = s_{t-1}|\mathcal{F}_{t-1}\right),\end{aligned} \tag{9}$$

where 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 (9) using the elements of the transition matrix, $\Pr(S_t = s_t | S_{t-1} = s_{t-1})$. The right hand side of equation (9)—the conditional state probability $\Pr(S_{t-1} = s_{t-1} | \mathcal{F}_{t-1})$—comes from Bayes' rule:

$$\begin{aligned} \Pr(S_t = s_t | \mathcal{F}_t) &= \Pr(S_t = s_t | \mathbf{r}_t, \mathcal{F}_{t-1}) \\ &= \frac{\Pr(\mathbf{r}_t, S_t = s_t | \mathcal{F}_{t-1})}{\Pr(\mathbf{r}_t | \mathcal{F}_{t-1})}. \end{aligned} \tag{10}$$

where the numerator and denominator of equation (10) come from equations (8) and (7), 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.[20] Table 5 gives the Quasi Maximum Likelihood parameter estimates and standard errors of the hidden Markov model parameter vector in equation (5).[21]

The parameters in Table 5 suggest that 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, at least based on the 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 -1.45 (=-0.20-1.25), but its down market beta is only -0.25. The combination of this with the higher volatilities means that the left tail risk is high when the hidden state is turbulent.

One question the reader might have at this point is whether these parameters estimates (and the associated standard errors) are reliable, particularly given the highly-non-normal

[20]The vector of steady state probabilities is given by the eigenvector of the transition matrix given in equation (4).

[21]We use standard Quasi Maximum Likelihood standard errors for inference. While the consistency of the parameter estimates we obtain depend on the conditional normality assumption, as verified in Appendix A.1, in the following subsection we show that biases due to deviations from normality are small.

momentum returns distribution. We explore this question in Appendix A. As noted earlier, despite the extreme skewness of the $R_{\text{MOM},t}$, the momentum residuals return ($\varepsilon_{\text{MOM},t}$) appears normally distributed. All of the skewness in the MOM return arises as a result of the optionality on the market. While the conditional market return residual remains non-normal–it is better characterized as Student-t distributed with (d.f.=5)–we show in Appendix A that given these distribution for the residuals, the QML estimator provides reasonably well behaved estimates.

We now examine the extent to which the estimated state probability can forecast the momentum tail-events or "crashes" we see in the return data.

4 Predicting Momentum Crashes using the HMM

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 Table 5 that when the hidden state is turbulent, the written call option-like features of momentum strategy returns become accentuated, and in addition 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 all 1044 months of data during 1927:01-2013:12. 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 in in-sample, meaning that the full-sample parameters (i.e., those presented in Table 5) are used to estimated 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 by the same QML procedure, but only up through the month prior to portfolio formation. Here the sample is 1980:09-2013:12, 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 from the slightly less volatile period following 1937:01.[22] In Panel B, just as in Panel A, there is again strong association between momentum crashes 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 7 presents the number of large negative and large positive momentum strategy returns during months when $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ is above a certain threshold. Notice that all thirteen momentum crashes, defined as losses exceeding 20%, happen when the $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ is more than 60%. 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 60%, and three out of those large gains happen when the $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ is less than 10%.

We examine the extent to which our tail risk measure succinctly summarizes the information about the likelihood of large momentum strategy losses using the following probit and logit models:

$$\Pr(R_{\mathrm{MOM},t} < Threshold) = F(a + \mathbf{b}' X_{t-1}), \tag{11}$$

where the functional form of F depends on whether we use the probit or the logit model.[23] For the predictors of X_{t-1}, we use our tail risk measure as well as the market return during the preceding 36 months, the realized volatility of daily market return during the preceding 12 months, the interaction of preceding market return and preceding market return volatility, and financial leverage (BD/ME) of the loser portfolio – variables that characterize market conditions when large losses in momentum strategy returns end to occur, (see Tables 2 and 3). We report results in Table 8 with $Threshold$=-10%.[24] Panel A reports the results for the longer sample, 1929:07-2013:12.[25] The $i'th$ entry of the coefficient vector b represents

[22]We can confirm that the parameter estimates converge more quickly if we start our parameter estimation with the return data beginning in 1927.

[23] For the probit model, we use the CDF of the standard normal distribution as $F(x)$. For the logit model, we use the logistic function: $F(x) = \frac{\exp(x)}{1+\exp(x)}$.

[24]If we lower the threshold below -10%, no single variable appears to be significant due to the rare occurrence of the events of $R_{\mathrm{MOM},t} < Threshold$. For the purpose of comparison across variables, we use -10% as $Threshold$.

[25]Since we use the past market returns over the previous 36 months, the sample period becomes shorter.

the coefficient corresponding to $i'th$ predictor variable. We arrange the predictor variables in the following sequence: our tail risk measure $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ as the first variable; market return during the preceding 36 months when as the second variable; realized volatility of daily market return during the preceding 12 months as the third variable; product of second and third variables as the fourth variable; and financial leverage (BD/ME) of the loser portfolio as the fifth variable. Let $t(b_i)$ denote the t-statistic associated with the coefficient b_i. All the variables are significant when used individually. When we use all variables in the estimation, only our tail risk measure remains to be statistically significant. We find similar results in Panel B using the shorter sample, 1964:01-2013:12 when financial leverage (BD/ME) of the loser portfolio is available.

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. As we saw before, left tail risk is related to 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 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.

Table 9 compares 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 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 67% ($Y$=67%). 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. Table 9 gives the number of false positives for different tail risk measures and different values of threshold X=10%, 20%, 30%, 40%.

In Panel A, 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 (2), (3) and (4). To emphasize the importance of option-like feature $\beta^+(S_t)$ in (2), we impose the restriction $\beta^+(S_t) = 0$ and report the associated results in Panel A-2. Also, motivated by the findings reported in Appendix B, we extend our HMM model to the hybrid case where residuals for momentum strategy returns, $\varepsilon_{\mathrm{MOM},t}$, are drawn from normal distribution and residuals for market excess returns, $\varepsilon_{\mathrm{MKT},t}$, are drawn from Student-t (d.f.=5) distribution. Results of which are reported in Panel A-3.

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 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 Panel B, C and D. For example, in our 1930:01-2013:12 sample,[26] we find 137 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,[27] the number of false positives increases to 187 months. The result of Panel A-2 shows that the necessity of the option features in our HMM specification. If we impose that $\beta^+(S_t) = 0$ while estimating our HMM model, the performance becomes worse. The number false positives increases from 137 to 164. When we relax the restriction that the residuals in momentum strategy returns and market excess returns are drawn from identical distribution and impose that $\varepsilon_{\mathrm{MOM},t}$ is drawn from normal distribution and $\varepsilon_{\mathrm{MKT},t}$ is drawn from Student-t (d.f.=5) distribution, the number of false positives declines sharply

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

[27]Barroso and Santa-Clara (2015) used this measure to imposing the volatility target of the momentum strategy.

from 137 to 114 as reported in Panel A-3.

This establishes the link between the tail risk of 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.

5 The Momentum Strategy's Option Adjusted Alpha

We have shown that the two-state HMM effectively picks out 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.

These results raise the question of how the alpha of the momentum strategy varies over time with changes in market conditions. While not the focus of our paper, in this section we briefly examine this question, based on the estimated HMM model from Section 3. We calculate the alpha from the perspective of an investor who can freely invest in the risk free asset, the market index portfolio, and in 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.

Specifically, we assume that how the investor values payoffs on risky assets has the stochastic discount factor representation.[28] Let M_t denote the stochastic discount factor, and $\mathcal{F}_{t-1}$ 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

[28]In our derivations, we follow the framework in Hansen and Jagannathan (1991) and Glosten and Jagannathan (1994).

followings relations hold:

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

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.

Regress M_t based on a constant, the market excess return, and the payoff on the 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

$$\begin{aligned} \widetilde{M_t} &= \lambda_{0,t-1} + \lambda_{1,t-1}R^e_{\mathrm{MKT},t} + \lambda_{2,t-1}\left(R^e_{\mathrm{MKT},t},0\right) & (13) \\ \mathbb{E}\left[\widetilde{e}_t|\mathcal{F}_{t-1}\right] &= \mathbb{E}\left[R^e_{\mathrm{MKT},t}\widetilde{e}_t|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\max\left(R^e_{\mathrm{MKT},t},0\right)\widetilde{e}_t|\mathcal{F}_{t-1}\right] = 0. & (14) \end{aligned}$$

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.

In a similar manner, 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 (2), 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 a standard normal distribution.

We consider the following conditional regression given the information set $\mathcal{F}_{t-1}$ that 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

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

Specifically, the vector of regression coefficients $\begin{bmatrix}\alpha_{t-1} & \beta^0_{t-1} & \beta^+_{t-1}\end{bmatrix}'$ is determined as

$$\begin{bmatrix}\alpha_{t-1} & \beta^0_{t-1} & \beta^+_{t-1}\end{bmatrix}' = \left(\mathbb{E}\left[\mathbf{x}_t\mathbf{x}_t'|\mathcal{F}_{t-1}\right]\right)^{-1}\mathbb{E}\left[\mathbf{x}_tR_{\text{MOM},t}|\mathcal{F}_{t-1}\right]$$

where $\mathbf{x}_t = \begin{bmatrix}1 & R^e_{\text{MKT},t} & \max\left(R^e_{\text{MKT},t},0\right)\end{bmatrix}'$, and

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

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

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

where the quantity in parenthesis is the excess return on one-period call option on the market.[29]

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

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

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

With Assumption 1, the following proposition holds.

[29] 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 $R^e_{\text{MKT},t}=0$.

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.

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

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, and 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. □

In what follows we 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. We then assess the validity of Assumption 1 by examining whether the residual in the equation (16) is uncorrelated with various risk factors proposed in the literature. Figure 3 plots the time series of α^* calculated based on the estimated HMM model for the sample period 1996:01-2013:12. Notice that the sample average of the α^*_{t-1}'s is 1.15%/month, which is significantly positive. However, α^*_{t-1} is negative during 1998:09-1998:10 (Russian crisis), 2002:08-2002:10 (dot-com bubble bursts), 2008:10-2008:12 and 2009:02-2009:04 (financial crisis), and 2011:10 (sovereign debt crisis) – time periods when 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 alphas as follows. First, we simulate 10,000 sets of parameters from the asymptotic distributions obtained from QML estimator, reported in Table 5. Then, for each set of parameters, we estimate the probability for the hidden state being turbulent based on the realized market excess returns and momentum strategy returns in our sample period 1996:01-2013:12 . With the simulated parameters, the estimated probabilities, and the time series of risk-free returns and call option prices, we construct the time series of α^*'s for the period 1996:01-2013:12 as described earlier. Finally, for each month, 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.

In Figure 3, we plot the time series of estimated α^*_{t-1} along with 95% the corresponding confidence intervals. In 167 of the 216 months in the sample period 1996:01-2013:12, the option-adjusted alpha is significantly positive. While the option adjusted alpha is negative during the other 49 months, only during two months – both occur during the recent financial crisis period 2008:12 and 2009:03 – they are statistically significantly different from zero.

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

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

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 3-Month LIBOR (LIBOR), Term Spread (the yield spread between the 10-year treasury bond and 3-month T-bill, TERM), Credit Spread (the yield spread between Moody's BAA bond and AAA bond, CREDIT), and 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).[30] Specifically, we estimate the regression

[30]We obtain MKT, SMB, HML, CMA and RMW from Ken French's data library: `http://mba.tuck.`

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 11. Except for ROE factor in Hou et al. (2015), we do not find any significant correlation between the residual we computed and systematic risk factors. These findings are mostly consistent with the Assumption 1 that the residual is conditionally uncorrelated with the systematic risk factors.

6 Conclusion

There is a vast literature documenting that the rather simple strategy of buying past winners and selling past losers, commonly referred to as the momentum strategy, generates abnormally high 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. This makes the momentum strategy susceptible to large losses when the market sharply recovers.

We document several patterns in the data. Following prolonged depressed and volatile market conditions, stocks in the loser portfolio become highly levered, behaving like out of the money call option on the stock market. When the market recovers, the stocks in the loser portfolio rise much more in value than the stocks in the winner portfolio. Since the momentum strategy takes a short position in stocks in the loser portfolio, momentum crashes tend to occur when the market recovers sharply from depressed market conditions, causing extremely large momentum strategy losses. That is more likely if the market is turbulent during those periods.

`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_2014.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.

Motivated by these empirical observations, we model the time-varying systematic tail risk of momentum strategy returns using a two-state hidden Markov model (HMM) where the embedded option-like features of momentum strategy returns become accentuated in the hidden *turbulent* state. We show that when the economy is in the latent *turbulent* state, the volatilities of market excess returns and momentum strategy returns are more than doubled In the *calm* state, both the momentum strategy returns and the market returns are less volatile, and the option-like features of momentum strategy returns become attenuated.

We 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 60% in only 179 of the 1044 months in our 1927:01-2013:12 sample. Yet in each of the 13 severe loss months, the *ex-ante* probability that the hidden state is turbulent exceeds 60 percent. Interestingly, the average momentum strategy return during those 179 months is only -0.94% per month.

We derive the conditional alpha of the momentum strategy for a given month based on the information available till the end of the previous month using HMM return generating process for momentum strategy returns and market excess returns and the price of call options on the market and the risk free rate. During 1996:01-2013:12 (216 months), for which we have call option prices, the average conditional alpha is 1.15%/month, which is significantly positive. However, the conditional alpha is negative during 49 out of the 216 months and significantly negative for two months 2008:12 and 2009:03 of the financial crisis period.

We show that QML estimator of HMM parameters need not to be consistent when the wrong normal likelihood is maximized. We find that the normally distributed residuals for momentum strategy returns and Student-t (d.f.=5) distributed residuals for market excess returns best describe the data. Our HMM model has the least number of false positives in predicting momentum crashes when compared to models on historical realized volatility, GARCH or past market returns.

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.

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.

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.

Chabot, Benjamin, Eric Ghysels, and Ravi Jagannathan, 2014, Momentum Trading, Return Chasing and Predictable Crashes, *Manuscript* .

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 D., and Tobias J. Moskowitz, 2015, Momentum crashes, forthcoming *Journal of Financial Economics* .

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.

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.

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.

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.

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

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, *The 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.

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

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

Swaminathan, Bhaskaran, 2010, Qunatitative money management: A practical application of behavioral finance, *Working Paper* .

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

Wooldridge, Jeffrey M., 1999, Chapter 45 Estimation and Inference for Dependent Processes, *Handbook of Econometrics, Vol 4* 2111–2245.

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), small size minus big size (SMB) factor, high book-to-market ratio minus low book-to-market ratio (HML) factor, 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 and the Fama and French (1993) three factor model, respectively, along with the corresponding risk exposures (betas). The sample period is 1927:01-2013:12. The t-statistics are computed using the heteroscedasticity 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	1.18	7.94	0.52	-2.43	21.22	26.18	-79.57
Mkt-Rf	0.64	5.43	0.41	0.16	10.35	38.04	-29.10
Mkt-Rf*	0.94	7.94	0.41	0.16	10.35	55.74	-42.64
SMB	0.24	3.24	0.26	2.05	23.46	37.45	-16.39
HML	0.39	3.52	0.39	1.92	18.69	34.08	-12.68
Panel B: Risk Adjusted MOM Returns							
	α	$\beta_{\text{Mkt-Rf}}$	$Adj.R^2$				
estimate	1.52	-0.52	0.13				
(t-stat)	(7.10)	(-4.82)					
	α	$\beta_{\text{Mkt-Rf}}$	β_{SMB}	β_{HML}	$Adj.R^2$		
estimate	1.76	-0.38	-0.23	-0.70	0.23		
(t-stat)	(8.20)	(-5.33)	(-2.11)	(-4.95)			

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 1927:01-2013:12 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 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 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 Mths	12 Mths	12 Mths	6 Mths
Panel A: Momentum Crash Months												
1931:06	−29.03	8.17	37.20	13.79	−0.64	n.a.	−74.07	n.a.	−36.67	−45.68	22.74	20.90
1932:07	−60.37	13.95	74.32	33.60	−33.12	n.a.	−88.35	n.a.	−81.52	−65.87	41.73	40.23
1932:08	−79.57	14.36	93.93	36.46	−30.61	n.a.	−86.25	n.a.	−76.45	−51.19	41.60	38.91
1932:11	−22.68	−20.83	1.85	−5.61	50.00	n.a.	−50.00	n.a.	−67.25	−27.00	44.65	50.32
1933:04	−41.94	28.77	70.71	38.04	55.17	n.a.	−54.55	n.a.	−72.52	−12.66	45.55	39.33
1933:05	−28.03	19.27	47.30	21.38	114.00	n.a.	−41.94	n.a.	−61.25	46.97	45.32	40.36
1938:06	−33.34	10.45	43.79	23.72	−9.16	n.a.	−68.93	n.a.	8.60	−39.09	32.45	29.26
1939:09	−44.57	7.92	52.49	16.96	51.22	n.a.	−33.33	n.a.	−16.29	−0.96	19.64	19.54
2001:01	−41.97	−6.94	35.03	3.12	94.09	0.08	−55.17	0.68	37.67	−11.58	24.48	22.15
2002:11	−20.40	2.12	22.52	5.96	64.42	0.12	−48.34	2.32	−30.71	−13.63	23.99	30.04
2009:03	−39.31	4.81	44.12	8.95	7.25	0.06	−79.52	70.55	−38.37	−42.63	42.11	56.03
2009:04	−45.89	−0.13	45.76	10.19	−1.76	0.07	−82.44	106.89	−34.05	−37.00	43.60	56.15
2009:08	−24.80	0.21	25.01	3.32	15.98	0.10	−66.10	55.57	−15.22	−18.90	45.14	32.77
Panel B: Averages Across Momentum Crash Months												
	−39.38	6.32	45.69	16.14	28.99	0.09	−63.77	47.20	−37.23	−24.56	36.39	36.62
Panel C: Average Across All Available Sample Months												
	1.18	1.24	0.06	0.64	76.08	0.10	−23.24	3.97	38.37	11.95	14.82	14.58

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 counts 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 the heteroscedasticity-consistent covariance estimator by White (1980). The sample period is 1929:07-2013:12.

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	3.00	1.12	-1.88	2.62	1.03	-1.59	0.59	0.65	0.06
$t(\alpha)$	(3.28)	(2.88)	(-2.93)	(5.20)	(3.44)	(-6.03)	(1.04)	(1.82)	(0.17)
β^0	-0.44	0.97	1.41	-0.08	1.27	1.35	0.20	1.40	1.19
$t(\beta^0)$	(-3.14)	(12.45)	(16.60)	(-0.42)	(11.80)	(13.23)	(1.89)	(17.93)	(13.45)
β^+	-1.01	-0.29	0.72	-0.49	-0.26	0.23	0.26	-0.12	-0.38
$t(\beta^+)$	(-3.17)	(-2.02)	(3.60)	(-1.51)	(-1.37)	(1.39)	(0.94)	(-0.78)	(-1.92)
$Adj.R^2$	0.48	0.77	0.83	0.06	0.71	0.72	0.06	0.81	0.62
B: CAPM Estimates									
α	0.05	0.27	0.22	1.81	0.60	-1.21	1.08	0.41	-0.67
$t(\alpha)$	(0.08)	(1.05)	(0.45)	(7.42)	(4.62)	(-7.42)	(2.74)	(1.90)	(-2.45)
β	-1.02	0.80	1.82	-0.32	1.14	1.47	0.31	1.35	1.04
$t(\beta)$	(-6.60)	(13.38)	(18.01)	(-3.40)	(21.17)	(27.93)	(3.45)	(30.06)	(15.37)
$Adj.R^2$	0.43	0.76	0.82	0.05	0.71	0.71	0.06	0.81	0.62
C: Number of Momentum Losses worse than -20%									
	11			2			0		
D: Conditional Skewness									
R^e_p	-2.33	-0.21	1.74	-0.98	-0.59	0.07	0.17	-0.73	-0.61
ε_p	-0.48	-0.59	0.59	-0.76	-0.21	0.94	-0.12	1.07	0.81

Table 4: Option-like Feature of Momentum Returns during Dotcom 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 the heteroscedasticity-consistent covariance estimator by White (1980).

R^e_p :	R_{MOM}	R^e_{WIN}	R^e_{LOS}
α	3.41	1.57	-1.84
$t(\alpha)$	(0.92)	(0.85)	(-0.71)
β^0	-0.42	1.25	1.67
$t(\beta^0)$	(-0.71)	(3.06)	(4.31)
β^+	-1.35	-0.54	0.82
$t(\beta^+)$	(-1.26)	(-0.81)	(1.11)

Table 5: Quasi Maximum Likelihood Estimates of HMM Parameters

We maximize the likelihood of data with the assumption that both of $\varepsilon_{\text{MOM},t}$ in (2) and $\varepsilon_{\text{MKT},t}$ in (3) are drawn from a standard normal distribution. The parameters are estimated using data for the period 1927:01-2013:12. α, σ_{MOM}, 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 (%)	2.12	(6.88)	4.30	(3.54)
β^0	0.37	(2.71)	-0.20	(-1.25)
β^+	-0.54	(-2.49)	-1.25	(-3.87)
σ_{MOM} (%)	4.22	(12.65)	11.59	(11.65)
μ	1.00	(6.71)	-0.49	(-0.82)
σ_{MKT} (%)	3.60	(22.78)	8.94	(9.11)
$\Pr(S_t = s_{t-1} \mid S_{t-1} = s_{t-1})$	0.96	(8.19)	0.88	(6.20)

Table 6: Distribution of HMM-implied moments of Pseudo Residuals

Monte Carlo simulation is performed as follows. First, we take the estimated parameters of our HMM model in Table 5 as given and generate the time series of moment strategy returns and market excess returns of length of 1044 months (the number of months during 1927:01-2013:12 in our sample) with a distributional assumption. Second, using this time series, we re-estimate our HMM parameters, construct the time series of $\Pr(S_t = C|\mathcal{F}_{t-1})$ and $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$, and obtain the simulated time series of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ and $\widehat{\varepsilon}_{\mathrm{MKT},t}$ defined in (A.7) and (A.8). Finally, we compute the first four moments of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ and $\widehat{\varepsilon}_{\mathrm{MKT},t}$. We then repeat this exercise 10,000 times and generate the sampling distribution of four moments of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ and $\widehat{\varepsilon}_{\mathrm{MKT},t}$.

	Pseudo Residuals of MOM: $\widehat{\varepsilon}_{\mathrm{MOM},t}$						Pseudo Residuals of MKT: $\widehat{\varepsilon}_{\mathrm{MKT},t}$					
	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: Using Residuals with Normal Distribution												
mean	0.04	-0.06	-0.04	0.00	0.05	0.06	-0.04	-0.05	-0.04	-0.02	0.00	0.00
std.dev	1.19	1.15	1.17	1.26	1.36	1.39	1.15	1.07	1.08	1.13	1.19	1.22
skewness	-0.16	-1.40	-0.98	-0.14	0.66	0.78	-0.38	-1.02	-0.68	-0.20	0.16	0.30
kurtosis	5.06	4.75	5.15	7.49	12.71	15.47	6.19	3.54	3.79	4.97	7.74	8.39
Panel B: Using Residuals with Student-t distribution d.f. 5												
mean	0.04	-0.09	-0.06	0.00	0.06	0.08	-0.04	-0.08	-0.07	-0.03	0.00	0.01
std.dev	1.19	1.22	1.27	1.43	1.63	1.73	1.15	1.13	1.16	1.26	1.42	1.50
skewness	-0.16	-4.54	-2.54	-0.22	1.32	2.67	-0.38	-4.44	-1.92	-0.24	1.30	2.75
kurtosis	5.06	6.47	6.88	11.26	39.95	77.98	6.19	5.04	5.56	9.11	36.63	78.65

Table 7: 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 1927:01-2013:12.

$\Pr(S_t{=}T\vert\mathcal{F}_{t-1})$ is more than	# Extreme losses during turbulent months /# Extreme losses in the sample					# of Months
	$\leq -20\%$	$\leq -17.5\%$	$\leq -15\%$	$\leq -12.5\%$	$\leq -10\%$	
80%	11/13	16/21	19/32	23/37	26/56	123
70%	12/13	17/21	20/32	24/37	28/56	152
60%	13/13	18/21	22/32	26/37	31/56	179
50%	13/13	19/21	23/32	27/37	32/56	208
40%	13/13	19/21	23/32	27/37	33/56	229
30%	13/13	20/21	24/32	28/37	37/56	270
20%	13/13	20/21	27/32	31/37	40/56	307
10%	13/13	21/21	30/32	34/37	46/56	403
$\Pr(S_t{=}T\vert\mathcal{F}_{t-1})$ is more than	# Extreme gains during turbulent months /# Extreme gains in the sample					# of Months
	$\geq 20\%$	$\geq 17.5\%$	$\geq 15\%$	$\geq 12.5\%$	$\geq 10\%$	
80%	5/12	6/15	11/28	18/45	27/74	123
70%	5/12	7/15	13/28	20/45	31/74	152
60%	8/12	10/15	16/28	23/45	35/74	179
50%	8/12	10/15	18/28	27/45	40/74	208
40%	8/12	10/15	18/28	29/45	44/74	229
30%	8/12	10/15	18/28	29/45	46/74	270
20%	9/12	11/15	19/28	31/45	49/74	307
10%	9/12	11/15	20/28	34/45	57/74	403

Table 8: Probit/Logit Regressions for Predicting Momentum Crashes

We examine the extent to which our tail risk measure succinctly summarizes the information about the likelihood of large momentum strategy losses using the following probit and logit models: $\Pr\left(R_{\mathrm{MOM},t} < Threshold\right) = F(a + \mathbf{b}' X_{t-1})$. For the predictors of X_{t-1}, we use our tail risk measure as well as the past market returns over the previous 36 months, the realized volatility of daily market returns over the previous 12 months, the interaction of past market returns and the volatility, and financial leverage (BD/ME) of the loser portfolio which are related to the large losses in momentum strategy returns, as shown in Tables 2 and 3. We report results below with $Threshold$ = -10%. Panel A reports the results for the longer sample, 1929:07-2013:12. b_i represents the coefficient on the predictors of our tail risk measure when $i = 1$, the past market returns over the previous 36 months when $i = 2$, the realized volatility of daily market returns over the previous 12 months when $i = 3$, the interaction of past market returns and the market volatility when $i = 4$ and financial leverage (BD/ME) of the loser portfolio when $i = 5$. $t(b_i)$ is the t-statistic of the corresponding coefficient. Panel B reports the results for the shorter sample, 1964:01-2013:12 when financial leverage (BD/ME) of the loser portfolio is available.

Panel A: The longer sample, 1929:07-2013:12										
	A-1: Probit					A-2: Logit				
b_1	1.48				1.06	3.03				2.20
$t(b_1)$	(7.47)				(3.96)	(7.41)				(4.01)
b_2		-0.01			0.02		-0.02			0.03
$t(b_2)$		(-4.87)			(1.81)		(-5.18)			(1.63)
b_3			0.05		0.00			0.09		0.00
$t(b_3)$			(6.96)		(-0.17)			(7.20)		(-0.35)
b_4				0.00	0.00				0.00	0.00
$t(b_4)$				(-4.65)	(-0.30)				(-5.31)	(-0.12)
a	-2.12	-1.38	-2.43	-1.52	-2.27	-3.98	-2.40	-4.47	-2.71	-4.19
$t(a)$	(-19.41)	(-18.16)	(-16.61)	(-22.63)	(-10.81)	(-15.42)	(-16.20)	(-14.68)	(-19.17)	(-09.51)

Continued on next page

Panel B: The shorter sample, 1964:01-2013:12												
	B-1: Probit						B-2: Logit					
b_1	1.41					1.00	2.87					2.09
$t(b_1)$	(5.42)					(3.09)	(5.47)					(3.14)
b_2		-0.01				0.01		-0.02				0.01
$t(b_2)$		(-2.70)				(0.61)		(-2.85)				(0.41)
b_3			0.05			0.00			0.10			-0.01
$t(b_3)$			(5.06)			(-0.45)			(5.19)			(-0.67)
b_4				0.00		0.00				0.00		0.00
$t(b_4)$				(-2.16)		(0.35)				(-2.49)		(0.57)
b_5					0.03	0.01					0.05	0.03
$t(b_5)$					(3.85)	(1.38)					(4.02)	(1.42)
a	-2.04	-1.38	-2.46	-1.48	-1.75	-2.12	-3.81	-2.37	-4.46	-2.56	-3.15	-3.77
$t(a)$	(-15.21)	(-12.76)	(-12.00)	(-15.66)	(-18.13)	(-6.05)	(-12.16)	(-11.01)	(-10.92)	(-13.45)	(-15.01)	(-5.56)

Table 9: 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 $\Pr(S_t = \text{Turbulent}|\mathcal{F}_{t-1})$ as a tail risk measure. The results in Panel A-1 are from our original HMM model specified in (2), (3) and (4). To emphasize the importance of option-like feature $\beta^+(S_t)$ in (2), we impose the restriction $\beta^+(S_t) = 0$ and report the associated results in Panel A-2. We examine the hybrid case where $\varepsilon_{\text{MOM},t}$ is drawn from Normal distribution and $\varepsilon_{\text{MKT},t}$ is drawn from Student-t (d.f.=5) distribution. Results of which are reported in Panel A-3. 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 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				
$\Pr(S_t = \text{T}|\mathcal{F}_{t-1})$	144	142	137	931
A-2: Without the option-like feature $\beta^+(S_t) = 0$				
$\Pr(S_t = \text{T}|\mathcal{F}_{t-1})$	171	169	164	930
A-3: Extension to Normal ($\varepsilon_{\text{MOM},t}$) and Student-t ($\varepsilon_{\text{MKT},t}$)				
$\Pr(S_t = \text{T}|\mathcal{F}_{t-1})$	121	119	114	902
Panel B: Momentum Strategy Returns Volatility				
GARCH(1,1)	263	261	256	829
RV(3 Months)	234	232	227	922
RV(6 Months)	194	192	187	892
RV(12 Months)	154	152	147	866
RV(36 Months)	180	178	173	951

Continued on next page

Panel C: Market Returns Volatility				
GARCH(1,1)	188	186	181	809
RV(3 Months)	166	164	159	889
RV(6 Months)	183	181	176	920
RV(12 Months)	191	189	184	796
RV(36 Months)	179	177	172	858
Panel D: Past Market Returns				
3 Months	618	616	980	948
6 Months	131	129	918	944
12 Months	249	247	966	938
36 Months	491	489	484	930

Table 10: Momentum and Market excess returns in Calm and Turbulent Markets

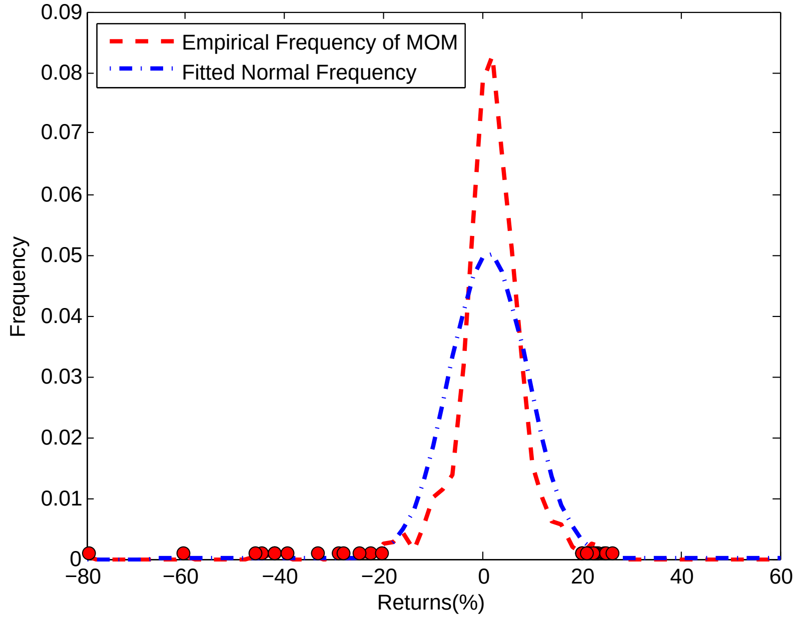
We classify 1044 months in our sample period of 1927:01-2013:12 into two mutually exclusive groups, depending on whether the tail risk measure for each month is *higher* than a given *threshold* value (Turbulent market) or *lower* than a given *threshold* value (Calm market). The properties of momentum strategy returns and market excess returns within each group, for various choices of the threshold tail risk measure, are given in the table below. SD and SR represent the standard deviation and the Sharpe ratio, respectively. N represents the number of months in each group out of 1044 months over 1927:01-2013:12. Mean and SD are presented in percentage.

	Market Condition														
	Calm							Turbulent							
	Momentum			Market				Momentum			Market			
Threshold	Mean	SD	SR	Mean	SD	SR	N	Mean	SD	SR	Mean	SD	SR	N
10%	1.65	4.67	1.22	0.79	4.12	0.66	641	0.44	11.30	0.13	0.42	7.02	0.13	403
20%	1.64	4.97	1.14	0.70	4.18	0.58	737	0.10	12.39	0.03	0.51	7.63	0.14	307
30%	1.60	5.12	1.08	0.65	4.27	0.53	774	−0.02	12.92	0.00	0.63	7.86	0.17	270
40%	1.54	5.20	1.03	0.71	4.34	0.57	815	−0.10	13.77	−0.02	0.42	8.22	0.11	229
50%	1.58	5.23	1.04	0.66	4.38	0.52	836	−0.40	14.27	−0.10	0.59	8.43	0.14	208
60%	1.62	5.36	1.05	0.63	4.41	0.49	865	−0.94	14.97	−0.22	0.73	8.85	0.17	179
70%	1.56	5.91	0.91	0.65	4.55	0.49	892	−1.01	14.94	−0.23	0.64	9.02	0.15	152
80%	1.53	6.00	0.88	0.60	4.68	0.44	921	−1.39	16.09	−0.30	1.02	9.31	0.22	123

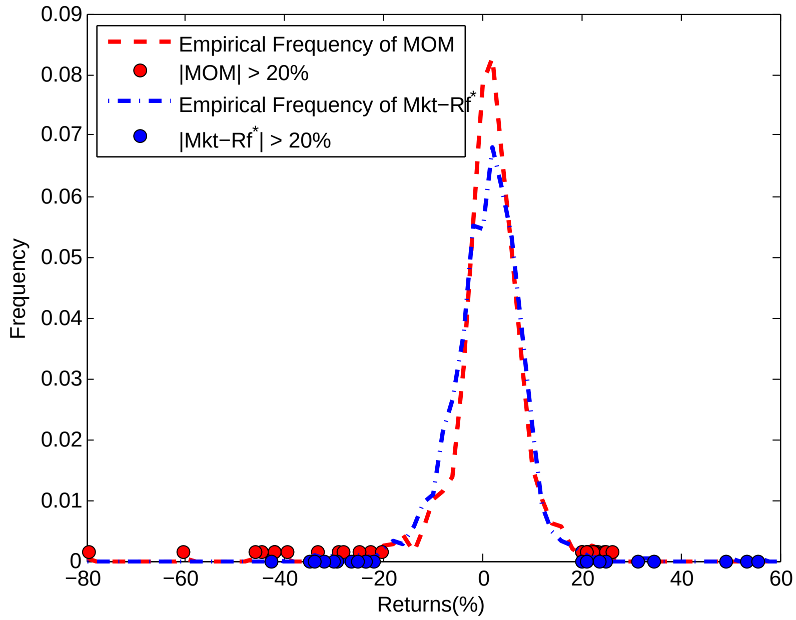
Table 11: Systematic Risk in Momentum Strategy Returns

When the specification of (16) describes the systematic risk of momentum strategy returns, $\epsilon_{\text{MOM},t}$ should be unrelated to various economy wide factors examined in the literature. To examine whether this is the case, we regress $\epsilon_{\text{MOM},t}$ on various systematic risk factors, $\epsilon_{\text{MOM},t}$ = intercept + coeff $\times$ systematic factor$_t$ + e_t. Results are tabulated below using the data from 1996:01 to 2013:12 (216 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.

Panel A: General Factors						
systematic factor	coeff	t(coeff)	R^2(%)	First Month	Last Month	N
MKT	-0.14	-0.92	0.57	1996:01	2013:12	216
SMB	0.25	0.96	1.09	1996:01	2013:12	216
HML	-0.50	-1.89	3.96	1996:01	2013:12	216
RMW	0.03	0.08	0.01	1996:01	2013:12	216
CMA	0.03	0.07	0.01	1996:01	2013:12	216
I/A	-0.20	-0.45	0.28	1996:01	2013:12	216
ROE	1.06	3.25	14.75	1996:01	2013:12	216
QMJ	0.56	1.86	4.39	1996:01	2013:12	216
Panel B: Liquidity Related Factors						
systematic factor	coeff	t(coeff)	R^2(%)	First Month	Last Month	N
LIQ	0.26	1.31	1.57	1996:01	2013:12	216
FLS	0.07	0.56	0.71	1996:01	2012:10	202
BAB	0.24	0.94	1.46	1996:01	2012:03	195
Δ LIBOR	0.05	1.19	1.66	1996:01	2013:12	216
Δ TERM	-0.04	-1.33	1.09	1996:01	2013:12	216
Δ CREDIT	0.03	0.70	0.24	1996:01	2013:12	216
Δ TED	0.00	-0.12	0.00	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	-0.34	0.03	1996:02	2013:09	212
VAR-SWAP 3M	0.01	0.91	0.15	1996:03	2013:08	210
VAR-SWAP 6M	0.02	1.60	0.55	1996:08	2013:08	203
VAR-SWAP 12M	0.01	0.59	0.09	1997:03	2013:08	193
Δ VIX	0.00	1.12	0.49	1996:01	2013:12	216
Δ LJV	-0.54	-0.28	0.04	1996:01	2013:12	216



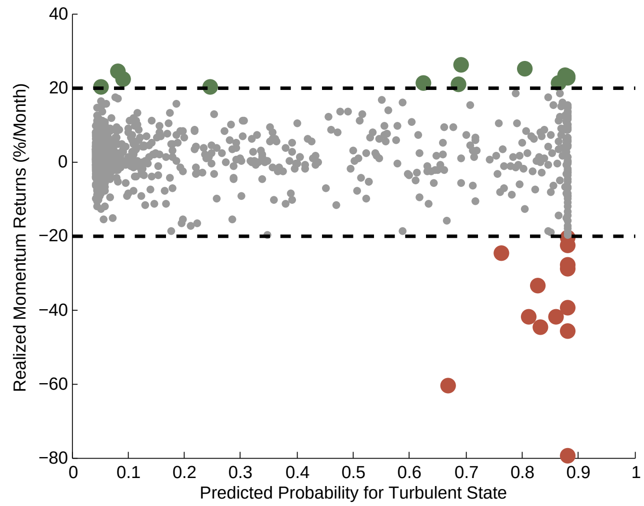
A Momentum Strategy Returns – Smoothed Density Function



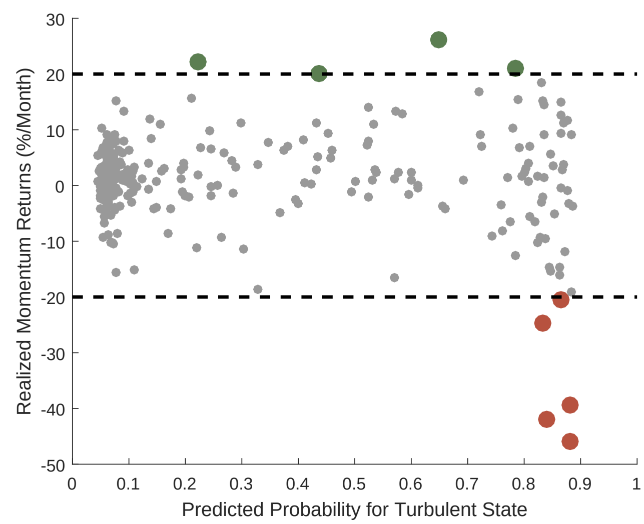
B Smoothed Density Functions–MOM and scaled excess market returns

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 25 MOM returns (13 in left tails and 12 in right tails) that exceed 20% in absolute value. Panel B plots the 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 1927:01-2013:12.



A In Sample Prediction



B Out of Sample Prediction

Figure 2: Momentum Returns and Probability of the Hidden State being 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 1044 months (1927:01-2013:12). For each month t of the last 400 months in 1980:09-2013:12, 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.

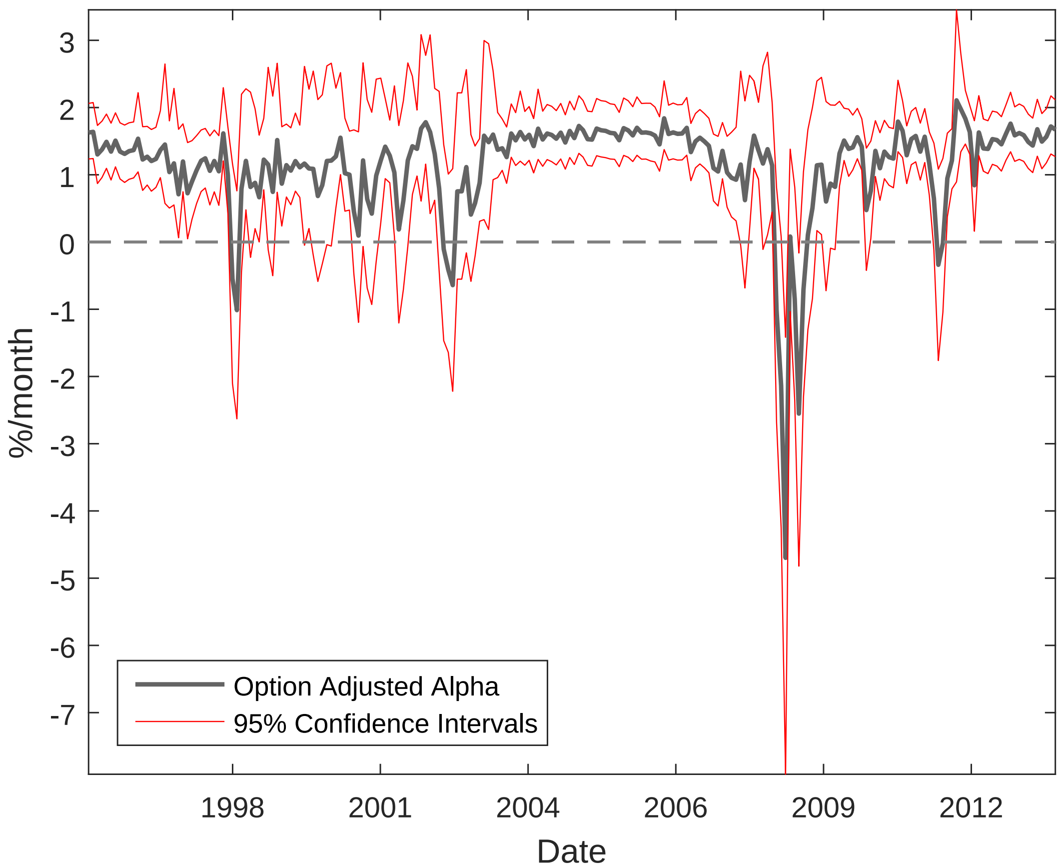


Figure 3: Time Series of Option Adjusted Alpha

Option adjusted alpha, α^*, is computed by (17). The sample period is 1996:01 to 2013:12 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. 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^*.

Appendix A QML Estimation Robustness

In Subsection 3.2, we estimate the parameters of our HMM by maximizing the likelihood function in equation (6) under the assumption that the residuals in (2) and (3) are jointly normally distributed. Here, we show that our HMM model behaves properly along several dimensions even when residuals are not drawn from normal distribution and examine the nature of deviations from normality.

To summarize, we find that point estimates for HMM parameters are fairly robust against the deviation from the normality assumption. We show in Appendix A.1 that the QML estimator that maximizes the (wrong) normal likelihood will not in general provide consistent estimates of the true HMM parameters. However, we find that the QML estimates are reasonably well behaved in finite samples when true residuals are drawn from Student-t distribution. We further show that the momentum residuals (in equation (2)) appear normally distributed, but the market residuals (in equation (3)) are better characterized as Student-t (d.f.=5).

In Table 5, we examine the effect of deviations from normality on QML point estimates using Monte Carlo simulations when the residuals in (2) and (3) are drawn from Student-t distribution with d.f. (degrees of freedom) 10 and 5. We simulate momentum strategy returns and market excess returns over 1044 months using HMM parameters in Table 5 and the residuals drawn from Student-t distribution. With the simulated data, we re-estimate the parameters of our HMM by maximizing the misspecified normal likelihood as described in the previous subsection. By repeating this exercise 1,000 times, we construct the sampling distribution of QML point estimates, the results of which are reported in Table B.1. When the residuals in (2) and (3) are drawn from Student-t with d.f. 10, the mean of QML estimator is quite close to the true value. Also, when the residuals are drawn from Student-t with d.f. 5, although the variances in calm state, $\sigma_{\text{MOM}}(C)$ and $\sigma_{\text{MKT}}(C)$, tend to be underestimated, the magnitude of the biases are small.

Next, we examine the sensitivity of the ranking based on the conditional probability that the (unobserved) state is Turbulent ($\Pr(S_t = \text{T}|\mathcal{F}_{t-1})$) when the residuals in the return generating processes are drawn from Student-t distributions. We generate a time series of length 50,000 months of momentum strategy and market excess returns by Monte Carlo simulation when these returns are generated by the hidden Markov model with parameters in Table 5 where residuals in (2) and (3) are drawn from a bivariate Student-t distribution with d.f. (degree of freedom) 10 and 5. Then, we compute the tail risk measure $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ for each of simulated 50,000 months using i) the true likelihood (Student-t) function and ii) the misspecified likelihood (normal) function. This gives a set of two $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ values for each month in the simulated time series. Given $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$, we classify month t as belonging to the 'Low' ('High') group if $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ is below the 30th percentile (above the 70th percentile). Months with $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ falling in between the 30th and 70th percentile are classified as belonging to the 'Med' group. Table B.2 present a 3×3 matrix, summarizing the joint distribution of $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ inferred through the true likelihood (Student-t) function and the misspecified likelihood (normal) function. When the residuals are drawn from Student-t distribution with d.f. 10, 96.9% (73.8+15.3+7.8) of simulated sample belong to the same groups whether we use the correct Student-t or the wrong normal distribution to compute $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$. When the residuals are drawn from Student-t distribution with d.f. 5, 93.4% (72.7+13.9+6.9) of samples are consistently classified by either true likelihood (Student-t) function or misspecified likelihood (normal) function.

In what follows, we examine the nature of deviations from normality. Since we do not directly observed the residuals of our HMM (because S_t is not directly observed) we examine the moments of the *pseudo residuals* – the probability weighted residuals of the two hidden states. We can write the residuals of our HMM return generating process as follows:

$$\varepsilon_{\mathrm{MOM},t} = \mathbf{I}(S_t = C)e_{\mathrm{MOM},t}(C) + \mathbf{I}(S_t = T)e_{\mathrm{MOM},t}(T), \tag{A.1}$$

$$\varepsilon_{\mathrm{MKT},t} = \mathbf{I}(S_t = C)e_{\mathrm{MKT},t}(C) + \mathbf{I}(S_t = T)e_{\mathrm{MKT},t}(T), \tag{A.2}$$

where $\mathbf{I}(\cdot)$ is an indicator function and

$$e_{\mathrm{MOM},t}(C) = \frac{1}{\sigma_{\mathrm{MOM}}(C)}\left(R_{\mathrm{MOM},t} - \alpha(C) - \beta^0(C)\,R^e_{\mathrm{MKT},t} - \beta^+(C)\max\left(R^e_{\mathrm{MKT},t}, 0\right)\right) \tag{A.3}$$

$$e_{\mathrm{MOM},t}(T) = \frac{1}{\sigma_{\mathrm{MOM}}(T)}\left(R_{\mathrm{MOM},t} - \alpha(T) - \beta^0(T)\,R^e_{\mathrm{MKT},t} - \beta^+(T)\max\left(R^e_{\mathrm{MKT},t}, 0\right)\right) \tag{A.4}$$

$$e_{\mathrm{MKT},t}(C) = \frac{1}{\sigma_{\mathrm{MKT}}(C)}\left(R^e_{\mathrm{MKT},t} - \mu(C)\right) \tag{A.5}$$

$$e_{\mathrm{MKT},t}(T) = \frac{1}{\sigma_{\mathrm{MKT}}(T)}\left(R^e_{\mathrm{MKT},t} - \mu(T)\right). \tag{A.6}$$

If we could observe the hidden states, $\varepsilon_{\text{MOM},t}$ and $\varepsilon_{\text{MKT},t}$ could be constructed from the observed momentum strategy returns and market excess returns. However, since $\mathbf{I}(S_t = C)$ and $\mathbf{I}(S_t = T)$ are not observed, we cannot construct a consistent estimator of $\varepsilon_{\text{MOM},t}$ and $\varepsilon_{\text{MKT},t}$ in (A.1) and (A.2).

We therefore define pseudo residuals – the probability weighted averages of sample counter parts of $e_{\text{MOM},t}(C)$, $e_{\text{MOM},t}(T)$, $e_{\text{MKT},t}(C)$ and $e_{\text{MKT},t}(T)$, in (A.3), (A.4), (A.5) and (A.6) – as follows:

$$\widehat{\varepsilon}_{\text{MOM},t} = \Pr(S_t = C|\mathcal{F}_{t-1})\widehat{e}_{\text{MOM},t}(C) + \Pr(S_t = T|\mathcal{F}_{t-1})\widehat{e}_{\text{MOM},t}(T) \quad \text{(A.7)}$$

$$\widehat{\varepsilon}_{\text{MKT},t} = \Pr(S_t = C|\mathcal{F}_{t-1})\widehat{e}_{\text{MKT},t}(C) + \Pr(S_t = T|\mathcal{F}_{t-1})\widehat{e}_{\text{MKT},t}(T) \quad \text{(A.8)}$$

by replacing the state indicator functions of $\mathbf{I}(S_t = C)$ and $\mathbf{I}(S_t = T)$ with the inferred probabilities of $\Pr(S_t = C|\mathcal{F}_{t-1})$ and $\Pr(S_t = T|\mathcal{F}_{t-1})$, respectively, and using QML estimates for our HMM of $\widehat{\alpha}(C), \widehat{\alpha}(T), \widehat{\beta}^0(C), \widehat{\beta}^0(T), \widehat{\beta}^+(C), \widehat{\beta}^+(T), \widehat{\sigma}_{\text{MOM}}(C), \widehat{\sigma}_{\text{MOM}}(T)$, $\widehat{\mu}(C), \widehat{\mu}(T), \widehat{\sigma}_{\text{MKT}}(C), \widehat{\sigma}_{\text{MKT}}(T)$:

$$\widehat{e}_{\text{MOM},t}(C) = \frac{1}{\widehat{\sigma}_{\text{MOM}}(C)}\left(R_{\text{MOM},t} - \widehat{\alpha}(C) - \widehat{\beta}^0(C)\, R^e_{\text{MKT},t} - \widehat{\beta}^+(C)\max\left(R^e_{\text{MKT},t}, 0\right)\right) \quad \text{(A.9)}$$

$$\widehat{e}_{\text{MOM},t}(T) = \frac{1}{\widehat{\sigma}_{\text{MOM}}(T)}\left(R_{\text{MOM},t} - \widehat{\alpha}(T) - \widehat{\beta}^0(T)\, R^e_{\text{MKT},t} - \widehat{\beta}^+(T)\max\left(R^e_{\text{MKT},t}, 0\right)\right) \quad \text{(A.10)}$$

$$\widehat{e}_{\text{MKT},t}(C) = \frac{1}{\widehat{\sigma}_{\text{MKT}}(C)}\left(R^e_{\text{MKT},t} - \widehat{\mu}(C)\right) \quad \text{(A.11)}$$

$$\widehat{e}_{\text{MOM},t}(T) = \frac{1}{\widehat{\sigma}_{\text{MOM}}(T)}\left(R^e_{\text{MKT},t} - \widehat{\mu}(T)\right). \quad \text{(A.12)}$$

Note that the pseudo residuals of $\widehat{\varepsilon}_{\text{MOM},t}$ and $\widehat{\varepsilon}_{\text{MKT},t}$ will in general not normally distributed even when the true residuals are normally distributed and we replace the estimated $\widehat{e}_{\text{MOM},t}(C)$, $\widehat{e}_{\text{MOM},t}(T)$, $\widehat{e}_{\text{MKT},t}(C)$, and $\widehat{e}_{\text{MOM},t}(T)$ in (A.9) - (A.12) with the population counterparts $e_{\text{MOM},t}(C)$, $e_{\text{MOM},t}(T)$, $e_{\text{MKT},t}(C)$, and $e_{\text{MKT},t}(T)$ in (A.3) - (A.6) due to the unobservability of the hidden state S_t.

Substituting $\widehat{e}_{\text{MOM},t}(C)$, $\widehat{e}_{\text{MOM},t}(T)$, $\widehat{e}_{\text{MKT},t}(C)$, and $\widehat{e}_{\text{MKT},t}(T)$ in the RHS of (A.7) and

(A.8) with the expressions of (A.9) - (A.12) and rearranging terms, we get

$$
\begin{aligned}
\widehat{\varepsilon}_{\mathrm{MOM},t} &= \left(\tfrac{\lambda_{t-1}}{\widehat{\sigma}_{\mathrm{MOM}}(C)} + \tfrac{1-\lambda_{t-1}}{\widehat{\sigma}_{\mathrm{MOM}}(T)}\right) R_{\mathrm{MOM},t} - \left(\tfrac{\lambda_{t-1}\widehat{\alpha}(C)}{\widehat{\sigma}_{\mathrm{MOM}}(C)} + \tfrac{(1-\lambda_{t-1})\widehat{\alpha}(T)}{\widehat{\sigma}_{\mathrm{MOM}}(T)}\right) \\
&\quad - \left(\tfrac{\lambda_{t-1}\widehat{\beta}^{0}(C)}{\widehat{\sigma}_{\mathrm{MOM}}(C)} + \tfrac{(1-\lambda_{t-1})\widehat{\beta}^{0}(T)}{\widehat{\sigma}_{\mathrm{MOM}}(T)}\right) R^{e}_{\mathrm{MKT},t} \\
&\quad - \left(\tfrac{\lambda_{t-1}\widehat{\beta}^{+}(C)}{\widehat{\sigma}_{\mathrm{MOM}}(C)} + \tfrac{(1-\lambda_{t-1})\widehat{\beta}^{+}(T)}{\widehat{\sigma}_{\mathrm{MOM}}(T)}\right) \max\left(R^{e}_{\mathrm{MKT},t}, 0\right), \\
\widehat{\varepsilon}_{\mathrm{MKT},t} &= \left(\tfrac{\lambda_{t-1}}{\widehat{\sigma}_{\mathrm{MKT}}(C)} + \tfrac{1-\lambda_{t-1}}{\widehat{\sigma}_{\mathrm{MkT}}(T)}\right) R^{e}_{\mathrm{MkT},t} - \left(\tfrac{\lambda_{t-1}\widehat{\mu}(C)}{\widehat{\sigma}_{\mathrm{MKT}}(C)} + \tfrac{(1-\lambda_{t-1})\widehat{\mu}(T)}{\widehat{\sigma}_{\mathrm{MKT}}(T)}\right),
\end{aligned}
$$

where $\lambda_{t-1} = \Pr(S_t = C|\mathcal{F}_{t-1})$. We will examine whether the empirical distribution of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ and $\widehat{\varepsilon}_{\mathrm{MKT},t}$ constructed from estimated HMM parameters and observed momentum strategy returns and market excess returns matches its' analogue constructed from Monte Carlo simulation where $R_{\mathrm{MOM},t}$ and $R^{e}_{\mathrm{MKT},t}$ are generated by our HMM model in (2), (3), and (4) when true residuals of $\varepsilon_{\mathrm{MOM},t}$ and $\varepsilon_{\mathrm{MKT},t}$ are drawn from normal or Student-t (d.f.=5) distribution.

Monte Carlo simulation is performed as follows. First, we take the estimated parameters of our HMM model in Table 5 as given and generate the time series of moment strategy returns and market excess returns of length of 1044 months (the number of months during 1927:01-2013:12 in our sample) with a distributional assumption. Second, using this time series, we re-estimate our HMM parameters, construct the time series of $\Pr(S_t = C|\mathcal{F}_{t-1})$ and $\Pr(S_t = T|\mathcal{F}_{t-1})$, and obtain the simulated time series of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ and $\widehat{\varepsilon}_{\mathrm{MKT},t}$ defined in (A.7) and (A.8). Finally, we compute the first four moments of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ and $\widehat{\varepsilon}_{\mathrm{MKT},t}$. We then repeat this exercise 10,000 times and generate the sampling distribution of four moments of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ and $\widehat{\varepsilon}_{\mathrm{MKT},t}$, summarized in Table 6. Panel A (B) of Table 6 reports the results using normal distribution (Student-t distribution with d.f.=5). First three moments of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ lie within 95% confidence region for the corresponding moments obtained by Monte Carlo simulation using normal distribution while the kurtosis of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ lies just to the left of the 95% interval with p-value 4%. In contrast, if we use Student-t (d.f.=5) distribution, the standard deviation and kurtosis of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ fall outside of 99% confidence interval. Hence, the empirical behavior of $\widehat{\varepsilon}_{\mathrm{MOM},t}$ fits better with normal distribution. Regarding the behavior of $\widehat{\varepsilon}_{\mathrm{MKT},t}$, both distributions fits well to the data, while the standard deviation of $\widehat{\varepsilon}_{\mathrm{MKT},t}$ is slightly off the 95% confidence interval from the simulated distribution of using Student-t distribution. In summary, we find that the diagnosis using pseudo residuals supports that the residuals in our HMM model are close to normal distribution.

Another diagnostic is to compare the empirical moments of momentum strategy returns and market excess returns with our HMM-implied moments for various combination of normal and Student-t distributions for residuals. Details are given in Appendix B. We find that normally distributed residuals for momentum strategy returns and Student-t distributed residuals for market excess returns help match the empirical moments. We treat QML estimates obtained with this distributional assumption as a distinct alternative specification when we compare models later.

A.1 Inconsistency of QML

In this paper, we estimate the parameters for our HMM specification under the assumption that shocks are drawn from —em i.i.d. normal distributions. We showed that the estimates are reasonably well behaved even when shocks are drawn from Student-t distributions, even though the parameters are estimated under the normality assumption, i.e., our estimates are Quasi-Maximum Likelihood (QML). Wooldridge (1999) 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 non-normal distribution and the 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:

$$R_t = \sigma\left(S_t\right)\varepsilon_t, \tag{A.13}$$

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

$$\Pi = \begin{bmatrix} \Pr(S_t = \mathrm{H}|S_{t-1} = \mathrm{H}) & \Pr(S_t = \mathrm{L}|S_{t-1} = \mathrm{H}) \\ \Pr(S_t = \mathrm{H}|S_{t-1} = \mathrm{L}) & \Pr(S_t = \mathrm{L}|S_{t-1} = \mathrm{L}) \end{bmatrix} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}. \tag{A.14}$$

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

$$\sigma_{\mathrm{H}} = 1.5, \sigma_{\mathrm{L}} = 1, \text{ and } p = 0.52, \tag{A.15}$$

the QML estimator of σ_{H} is consistent.

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

$$\frac{1}{T}\sum_{t=1}^{T} \log\left(\mathcal{L}\left(R_t\right)\right), \tag{A.16}$$

where

$$\mathcal{L}\left(R_t\right) = \lambda_{t-1}\phi\left(R_t|\sigma_{\mathrm{H}}\right) + \left(1-\lambda_{t-1}\right)\phi(R_t|\sigma_{\mathrm{L}}), \tag{A.17}$$

$\phi\left(x|\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)$ 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:

$$1 - p \leq \lambda_{t-1}, \lambda^*_{t-1} \leq p \tag{A.18}$$

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

The QML estimate $\widehat{\sigma}_{\mathrm{H}}$ is obtained by maximizing (A.16), giving rise to the first order condition:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\log\left(\mathcal{L}\left(R_t\right)\right)}{\partial\sigma_{\mathrm{H}}}|_{\sigma_{\mathrm{H}}=\widehat{\sigma}_{\mathrm{H}}} = 0. \tag{A.19}$$

If $\widehat{\sigma}_{\mathrm{H}}$ converges to σ^0_{H}, the true value of σ_{H}, the following should hold:

$$\mathbb{E}\left[\mathbb{E}\left[\frac{\partial\log\left(\mathcal{L}\left(R_t\right)\right)}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right]\right]_{\sigma_{\mathrm{H}}=\sigma^0_{\mathrm{H}}} = 0. \tag{A.20}$$

We show the inconsistency of $\widehat{\sigma}_{\mathrm{H}}$ by verifying that (A.20) cannot hold. When $\sigma_{\mathrm{H}} = \sigma^0_{\mathrm{H}}$, 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

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

Hereafter, we will evaluate the conditional expectation at $\sigma_{\mathrm{H}} = \sigma_{\mathrm{H}}^0$. From (A.17), 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:

$$\begin{aligned}\mathbb{E}\left[\frac{\partial\log\left(\mathcal{L}\left(R_t\right)\right)}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\frac{\lambda_{t-1}}{\mathcal{L}\left(R_t\right)}\frac{\partial\phi\left(R_t|\sigma_{\mathrm{H}}\right)}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right] \\ &+ \mathbb{E}\left[\phi\left(R_t|\sigma_{\mathrm{H}}\right)-\phi\left(R_t|\sigma_{\mathrm{L}}\right)|\mathcal{F}_{t-1}\right]\frac{\partial\lambda_{t-1}}{\partial\sigma_{\mathrm{H}}}. \qquad (A.21)\end{aligned}$$

To determine the sign of each component in RHS of (A.21), 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.13) 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

$$\begin{aligned}\mathbb{E}\left[\frac{\lambda_{t-1}}{\mathcal{L}}\frac{\partial\phi\left(R_t|\sigma_{\mathrm{H}}\right)}{\partial\sigma_{\mathrm{H}}}|\mathcal{F}_{t-1}\right] &= \frac{\lambda_{t-1}^*}{2}\sum_{R_t=-\sigma_{\mathrm{H}},\sigma_{\mathrm{H}}}\frac{\lambda_{t-1}}{\mathcal{L}\left(R_t\right)}\phi(R_t|\sigma_{\mathrm{H}})\left(-\frac{1}{\sigma_{\mathrm{H}}}+\frac{R_t^2}{\sigma_{\mathrm{H}}^3}\right) \\ &+ \frac{1-\lambda_{t-1}^*}{2}\sum_{R_t=-\sigma_{\mathrm{L}},\sigma_{\mathrm{L}}}\frac{\lambda_{t-1}}{\mathcal{L}\left(R_t\right)}\phi(R_t|\sigma_{\mathrm{H}})\left(-\frac{1}{\sigma_{\mathrm{H}}}+\frac{R_t^2}{\sigma_{\mathrm{H}}^3}\right) \\ &= \frac{\left(1-\lambda_{t-1}^*\right)\lambda_{t-1}}{\mathcal{L}(\sigma_{\mathrm{L}})}\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}})\left(-\frac{1}{\sigma_{\mathrm{H}}}+\frac{\sigma_{\mathrm{L}}^2}{\sigma_{\mathrm{H}}^3}\right) \\ &< -\left(1-p\right)^2\frac{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}})}{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}})}\left(\frac{\sigma_{\mathrm{H}}^2-\sigma_{\mathrm{L}}^2}{\sigma_{\mathrm{H}}^3}\right), \qquad (A.22)\end{aligned}$$

where the last inequality is from (A.18) 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 $\phi\left(x|\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)$, the sign of $\mathbb{E}\left[\phi\left(R_t|\sigma_{\mathrm{H}}\right)-\phi\left(R_t|\sigma_{\mathrm{L}}\right)|\mathcal{F}_{t-1}\right]$ is determined as follows:

$$\begin{aligned}&\mathbb{E}\left[\phi\left(R_t|\sigma_{\mathrm{H}}\right)-\phi\left(R_t|\sigma_{\mathrm{L}}\right)|\mathcal{F}_{t-1}\right] \\ =\ & \frac{\lambda_{t-1}^*}{2}\sum_{R_t=-\sigma_{\mathrm{H}},\sigma_{\mathrm{H}}}\left(\phi\left(R_t|\sigma_{\mathrm{H}}\right)-\phi\left(R_t|\sigma_{\mathrm{L}}\right)\right)+\frac{1-\lambda_{t-1}^*}{2}\sum_{R_t=-\sigma_{\mathrm{L}},\sigma_{\mathrm{L}}}\left(\phi\left(R_t|\sigma_{\mathrm{H}}\right)-\phi\left(R_t|\sigma_{\mathrm{L}}\right)\right) \\ =\ & \lambda_{t-1}^*\left(\phi\left(\sigma_{\mathrm{H}}|\sigma_{\mathrm{H}}\right)-\phi\left(\sigma_{\mathrm{H}}|\sigma_{\mathrm{L}}\right)\right)+\left(1-\lambda_{t-1}^*\right)\left(\phi\left(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}}\right)-\phi\left(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}}\right)\right) \\ >\ & \left(1-p\right)\left(\phi\left(\sigma_{\mathrm{H}}|\sigma_{\mathrm{H}}\right)-\phi\left(\sigma_{\mathrm{H}}|\sigma_{\mathrm{L}}\right)\right)+p\left(\phi\left(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}}\right)-\phi\left(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}}\right)\right) > 0, \qquad (A.23)\end{aligned}$$

where the last two inequalities can be verified by (A.18) and the given parameter values of (A.15).

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.14). Since λ_0 does not depend on σ_{H}, the following holds:

$$\frac{\partial \lambda_0}{\partial \sigma_{\mathrm{H}}} = 0. \tag{A.24}$$

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:

$$\widetilde{\lambda}_t = \frac{\lambda_{t-1}\phi\left(R_t|\sigma_{\mathrm{H}}\right)}{\lambda_{t-1}\phi\left(R_t|\sigma_{\mathrm{H}}\right) + \left(1-\lambda_{t-1}\right)\phi\left(R_t|\sigma_{\mathrm{L}}\right)}, \tag{A.25}$$

and

$$\lambda_t = p\widetilde{\lambda}_t + (1-p)\left(1-\widetilde{\lambda}_t\right). \tag{A.26}$$

Equation (A.25) 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.26) 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.14). Combining (A.25) and (A.26), we get

$$\frac{\lambda_t + p - 1}{2p-1} = \frac{\lambda_{t-1}\phi\left(R_t|\sigma_{\mathrm{H}}\right)}{\lambda_{t-1}\phi\left(R_t|\sigma_{\mathrm{H}}\right) + \left(1-\lambda_{t-1}\right)\phi\left(R_t|\sigma_{\mathrm{L}}\right)}. \tag{A.27}$$

Taking the derivative of (A.27) with respect to σ_{H}, we obtain the following:

$$\frac{1}{2p-1}\frac{\partial \lambda_t}{\partial \sigma_{\mathrm{H}}} = \frac{\partial \frac{\lambda_{t-1}\phi(R_t|\sigma_{\mathrm{H}})}{\lambda_{t-1}\phi(R_t|\sigma_{\mathrm{H}})+(1-\lambda_{t-1})\phi(R_t|\sigma_{\mathrm{L}})}}{\partial \lambda_{t-1}}\frac{\partial \lambda_{t-1}}{\partial \sigma_{\mathrm{H}}} + \frac{\partial \frac{\lambda\phi(R_t|\sigma_{\mathrm{H}})}{\lambda\phi(R_t|\sigma_{\mathrm{H}})+(1-\lambda)\phi(R_t|\sigma_{\mathrm{L}})}}{\partial \phi\left(R_t|\sigma_{\mathrm{H}}\right)}\frac{\partial \phi\left(R_t|\sigma_{\mathrm{H}}\right)}{\partial \sigma_{\mathrm{H}}} \tag{A.28}$$

To determine the sign of each component in RHS of (A.28), we use the following properties:

$$\frac{\partial \frac{\lambda m}{\lambda m+(1-\lambda)n}}{\partial \lambda} = \frac{mn}{\left(\lambda m+(1-\lambda)n\right)^2} > 0 \tag{A.29}$$

$$\frac{\partial \frac{\lambda m}{\lambda m+(1-\lambda)n}}{\partial m} = \frac{\lambda(1-\lambda)n}{\left(\lambda m+(1-\lambda)n\right)^2} > 0 \tag{A.30}$$

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

$$\frac{\partial \phi(R_t|\sigma_{\mathrm{H}})}{\partial \sigma_{\mathrm{H}}} \leq 0 \tag{A.31}$$

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.29), (A.30), and (A.31) ensure that RHS of (A.28) is non-positive. Hence, with $p > 1/2$ as assumed in (A.15), it follows that $\frac{\partial \lambda_t}{\partial \sigma_{\mathrm{H}}} \leq 0$. Combining (A.24) with this finding, we conclude that

$$\frac{\partial \lambda_{t-1}}{\partial \sigma_{\mathrm{H}}} \leq 0, \tag{A.32}$$

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

Recall that we want to show that (A.21) is strictly negative. Finally, combining (A.22), (A.23), and (A.31), we conclude that

$$\mathbb{E}\left[\frac{\partial \log(\mathcal{L}(R_t|\sigma_{\mathrm{H}}))}{\partial \sigma_{\mathrm{H}}} | \mathcal{F}_{t-1}\right] < -\delta, \tag{A.33}$$

where

$$\delta = (1-p)^2 \frac{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{H}})}{\phi(\sigma_{\mathrm{L}}|\sigma_{\mathrm{L}})} \left(\frac{\sigma_{\mathrm{H}}^2 - \sigma_{\mathrm{L}}^2}{\sigma_{\mathrm{H}}^3} \right) > 0, \tag{A.34}$$

completing the proof that QML estimate of $\widehat{\sigma}_{\mathrm{H}}$ in (A.19) will not converge to the true parameter value.

Appendix B Explaining the moments of the momentum strategy returns

In this appendix, we examine the extent to which we can match the unconditional sample moments of momentum strategy returns and market excess returns based on the HMM return generating process. For this purpose, we consider the following distributions for the pair of ($\varepsilon_{\text{MOM},t}$, $\varepsilon_{\text{MKT},t}$): (Normal, Normal), (Student-t, Student-t), (Normal, Student-t), and (Student-t, Normal). We assume that the QML estimates of HMM parameters in Table 5 are the true parameters.

We generate a 1044 month-long time series of momentum strategy and market excess returns based using monte carlo simulation and obtain their first four moments. We then repeat this exercise 10,000 times to obtain the distribution of the first four momentums.Table B.3 summarize the distribution of first four moments of momentum strategy returns and market excess return obtained in this way for the four sets of distributions. Panel A of Table B.3 gives the result for normal ($\varepsilon_{\text{MOM},t}$) and normal ($\varepsilon_{\text{MKT},t}$). We find that the skewness (-2.43) and kurtosis (21.22) of momentum strategy returns, over our sample period 1044 months (1927:01-2013:12), fall outside of the 99% confidence interval of our HMM-implied moments obtained by simulation. However, once we use Student-t (d.f.=5) distribution for $\varepsilon_{\text{MKT},t}$, those sample moments lie within the 95% confidence interval of our HMM-implied moments, as shown in Panel C of Table B.3. If we use Student-t for both $\varepsilon_{\text{MOM},t}$ and $\varepsilon_{\text{MKT},t}$, those sample moments lie within 95% confidence intervals of our HMM-implied moments. However, the intervals becomes too wide. When we compare Panel B with Panel C, the 95% confidence intervals of skewness and kurtosis of momentum strategy returns are (-2.85,0.77) and (7.98,43.38) when $\varepsilon_{\text{MOM},t}$ has Student-t distribution which are much wider than the corresponding 95% confidence intervals of (-2.63,0.01) and (6.69,29.77) when $\varepsilon_{\text{MOM},t}$ has a normal distribution.

Motivated by this finding, we estimate the HMM parameters assuming that $\varepsilon_{\text{MOM},t}$ is drawn from a normal distribution and $\varepsilon_{\text{MKT},t}$ is drawn from a Student-t (d.f.=5) distribution. Results are reported in Table B.4. We find that the point estimates in Table B.4 are reasonably close to those in Table 5. This is consistent with our findings, reported earlier, that the QML estimates obtained by maximizing the wrong normal likelihood are reasonably robust.

Table B.1: Sensitivity of HMM QML Estimates to Deviations from Normality

Given the estimated parameters in Table 5 as true parameters, we generate momentum returns and market excess returns by simulation using the return generating equations (2), (3), and (4) when the residuals in the equations drawn from an *i.i.d.* bivariate Student-t distribution with the d.f. (degrees of freedom) 10 and 5. Given the simulated sample over 1044 months, we estimate the parameters using quasi maximum likelihood, maximizing the likelihood when residuals are drawn from an i.i.d. bivariate normal distribution. By repeating this exercise 1,000 times, we construct the sampling distribution of estimated parameters, the properties of which are reported in this table. $\Pr(\cdot|\cdot)$ represents the probability for the same underlying state to be realized, $\Pr(S_t = s_{t-1} | S_{t-1} = s_{t-1})$. α, σ_{MOM}, and σ_{MKT} are reported in percentage per month.

	Hidden State									
	$S_t = \text{Calm}(C)$					$S_t = \text{Turbulent}(T)$				
Para-			quantile					quantile		
meter	true	mean	10%	50%	90%	true	mean	10%	50%	90%
Panel A: When the true distribution is Student-t with d.f.=10										
α	2.12	2.11	1.78	2.11	2.45	4.30	4.25	2.77	4.25	5.77
β^0	0.37	0.36	0.22	0.36	0.49	−0.20	−0.19	−0.38	-0.19	0.00
β^+	−0.54	−0.52	-0.73	−0.52	−0.34	−1.25	−1.22	−1.56	-1.23	−0.88
σ_{MOM}	4.22	4.08	3.91	4.08	4.26	11.59	11.68	10.85	11.67	12.57
μ	1.00	1.00	0.82	1.00	1.17	−0.49	−0.46	−1.20	-0.45	0.30
σ_{MKT}	3.60	3.49	3.34	3.49	3.64	8.94	8.99	8.32	8.98	9.69
$\Pr(\cdot\|\cdot)$	0.96	0.95	0.94	0.95	0.96	0.88	0.85	0.81	0.85	0.89
Panel B: When the true distribution is Student-t with d.f.=5										
α	2.12	2.13	1.82	2.14	2.41	4.30	4.14	2.58	4.15	5.79
β^0	0.37	0.35	0.22	0.35	0.49	−0.20	−0.15	−0.34	-0.16	0.04
β^+	−0.54	−0.53	-0.72	−0.53	−0.33	−1.25	−1.15	−1.51	-1.15	−0.80
σ_{MOM}	4.22	3.79	3.58	3.79	3.98	11.59	12.04	10.82	11.90	13.22
μ	1.00	0.99	0.83	0.99	1.15	−0.49	−0.38	−1.16	-0.35	0.37
σ_{MKT}	3.60	3.24	3.07	3.24	3.42	8.94	9.16	8.23	9.10	10.09
$\Pr(\cdot\|\cdot)$	0.96	0.93	0.91	0.93	0.95	0.88	0.78	0.71	0.79	0.85

Table B.2: PROPERTIES OF $\Pr(S_t = T|\mathcal{F}_{t-1})$ CONSTRUCTED USING MISSPECIFIED (NORMAL) LIKELIHOOD

We examine the sensitivity of the ranking of $\Pr(S_t = T|\mathcal{F}_{t-1})$ constructed using misspecified (normal) likelihood, where the joint process of the momentum strategy returns and the market excess returns follow the hidden Markov model with the parameters in Table 5 and residuals are drawn from t-distribution with d.f. (degree of freedom) 10 and 5. Specifically, we simulate the joint process of momentum strategy returns and market excess returns over 50,000 months and compute the tail risk measure $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ for each of simulated 50,000 months using the true likelihood (t-distribution) function and the misspecified likelihood (normal distribution) function. This table gives 3 (Low,Med,High) by 3 (Low,Med,High) matrix, summarizing the joint distribution of $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ inferred through the true likelihood (t-distribution) function and the misspecified likelihood (normal distribution) function. We classify a month as Low (High) group if the inferred probability of $\Pr(S_t = \mathrm{T}|\mathcal{F}_{t-1})$ is below 30% (above 70%). Months with the inferred probability between 30% and 70% are classified as Med group. Number are reported in percentage.

Misspecified Likelihood of normal	True Likelihood of Student-t with d.f. 10			Misspecified Likelihood of normal	True Likelihood of Student-t with d.f. 5		
	Low	Med	High		Low	Med	High
Low	73.8	0.0	0.8	Low	72.7	0.0	2.2
Med	0.0	15.3	0.8	Med	0.3	13.9	1.2
High	0.9	0.5	7.8	High	1.5	1.4	6.9

Table B.3: 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. We generate $\varepsilon_{\text{MOM},t}$ and $\varepsilon_{\text{MKT},t}$ in our HMM specification of (2)) and (3)) using monte carlo simulation from various combinations of Normal and Student-t distributions. Then, we construct a 1044 month-long 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: $R_{\text{MOM},t}$						Market Excess Returns: $R^e_{\text{MKT},t}$					
	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 ($\varepsilon_{\text{MOM},t}$) and Normal ($\varepsilon_{\text{MKT},t}$)												
mean	1.18	0.47	0.65	1.14	1.61	1.75	0.64	0.18	0.28	0.65	0.99	1.09
std.dev	7.94	6.21	6.59	7.76	8.95	9.36	5.43	4.56	4.75	5.41	6.06	6.26
skewness	-2.43	-1.46	-1.23	-0.59	0.00	0.19	0.16	-0.89	-0.76	-0.34	0.07	0.21
kurtosis	21.22	5.88	6.27	8.08	11.64	13.56	10.35	4.52	4.77	5.78	7.47	8.27
Panel B: Student-t ($\varepsilon_{\text{MOM},t}$) and Student-t ($\varepsilon_{\text{MKT},t}$)												
mean	1.18	0.60	0.76	1.26	1.73	1.88	0.64	0.15	0.28	0.65	0.99	1.09
std.dev	7.94	6.17	6.55	7.79	9.27	9.83	5.43	4.48	4.68	5.38	6.28	6.70
skewness	-2.43	-5.03	-2.85	-0.76	0.77	2.35	0.16	-4.16	-2.12	-0.33	1.40	3.46
kurtosis	21.22	7.23	7.98	12.51	43.48	96.63	10.35	5.81	6.41	10.09	38.77	95.11
Panel C: Normal ($\varepsilon_{\text{MOM},t}$) and Student-t ($\varepsilon_{\text{MKT},t}$)												
mean	1.18	0.60	0.75	1.24	1.72	1.81	0.64	0.15	0.28	0.65	0.99	1.09
std.dev	7.94	6.14	6.57	7.82	9.10	9.64	5.43	4.48	4.68	5.38	6.28	6.70
skewness	-2.43	-4.84	-2.63	-0.74	0.01	0.22	0.16	-4.16	-2.12	-0.33	1.40	3.46
kurtosis	21.22	6.24	6.69	9.31	29.77	69.22	10.35	5.81	6.41	10.09	38.77	95.11
Panel D: Student-t ($\varepsilon_{\text{MOM},t}$) and Normal ($\varepsilon_{\text{MKT},t}$)												
mean	1.18	0.46	0.63	1.13	1.61	1.74	0.64	0.18	0.28	0.65	0.99	1.09
std.dev	7.94	6.14	6.52	7.74	9.10	9.61	5.43	4.56	4.75	5.41	6.06	6.26
skewness	-2.43	-3.23	-1.91	-0.60	0.93	2.35	0.16	-0.89	-0.76	-0.34	0.07	0.21
kurtosis	21.22	6.90	7.50	10.91	30.80	61.93	10.35	4.52	4.77	5.78	7.47	8.27

Table B.4: Quasi Maximum Likelihood Estimates of HMM Parameters when $\varepsilon_{\text{MOM},t}$ has a Normal distribution and $\varepsilon_{\text{MKT},t}$ has a Student-t (d.f.=5) distribution

We maximize the likelihood of data with the assumption that $\varepsilon_{\text{MOM},t}$ in (2) has a Normal distribution and $\varepsilon_{\text{MKT},t}$ in (3) has a Student-t (d.f=5) distribution. The parameters are estimated using data for the period 1927:01-2013:12. α, σ_{MOM}, and σ_{MKT} are reported in percentage per month.

	Hidden State			
	$S_t = Calm(C)$		$S_t = \text{Turbulent}(T)$	
Parameter	estimates	(t-stat)	estimates	(t-stat)
α (%)	1.95	(6.87)	4.05	(3.31)
β^0	0.34	(3.43)	−0.32	(−1.11)
β^+	−0.46	(−2.57)	−1.14	(−3.51)
σ_{MOM} (%)	4.31	(5.11)	11.02	(14.19)
μ	1.11	(8.76)	−0.38	(−0.59)
σ_{MKT} (%)	4.04	(12.92)	8.36	(5.46)
$\Pr(S_t = s_{t-1} \mid S_{t-1} = s_{t-1})$	0.98	(3.21)	0.94	(4.16)

Appendix C Additional Tables

Table C.5: 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'. For Panel A, the sample period is 1929:07-2013:12. For Panel B and C, the sample period is 1927:07-2013:12. In Panel A, we group the sample on the basis of cumulative market return during the 36 months preceding the month in which the momentum portfolios are formed. In Panel B, we group the months based on the realized volatility of daily market returns over the previous 12 months. In Panel C, 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, B-1-i, and C-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, B-1-ii, and C-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, B-2, and C-2. Finally, we compare the skewness of $R^e_{p,t}$ with that of estimated ε of (1) in Panel A-3, B-3, and C-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 B: 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}
B-1: Option-like features									
B-1-i: Henriksson-Merton Estimates									
α	2.90	1.07	-1.83	1.93	0.77	-1.16	2.40	1.38	-1.02
$t(\alpha)$	(2.96)	(2.71)	(-2.58)	(5.73)	(4.05)	(-5.01)	(4.55)	(4.21)	(-2.98)
β^0	-0.59	0.94	1.52	0.16	1.35	1.19	0.54	1.55	1.02
$t(\beta^0)$	(-4.83)	(13.78)	(17.78)	(1.72)	(25.36)	(18.24)	(3.00)	(14.91)	(8.23)
β^+	-0.91	-0.27	0.63	-0.25	-0.19	0.06	-0.63	-0.46	0.17
$t(\beta^+)$	(-3.23)	(-2.14)	(3.39)	(-1.38)	(-1.93)	(0.51)	(-1.92)	(-2.39)	(0.79)
$Adj.R^2(\%)$	0.49	0.74	0.83	0.00	0.78	0.68	0.03	0.73	0.57
B-1-ii: CAPM Estimates									
α	0.12	0.23	0.11	1.48	0.43	-1.04	1.58	0.78	-0.80
$t(\alpha)$	(0.18)	(0.82)	(0.20)	(6.69)	(3.58)	(-6.96)	(4.87)	(4.16)	(-3.56)
β	-1.10	0.78	1.88	0.05	1.27	1.22	0.19	1.30	1.11
$t(\beta)$	(-8.43)	(14.68)	(21.61)	(0.78)	(41.31)	(29.99)	(1.83)	(23.47)	(15.56)
$Adj.R^2$	0.45	0.73	0.82	0.00	0.78	0.68	0.01	0.72	0.57
B-2: Number of Momentum Losses worse than -20%									
	13			0			0		
B-3: Conditional Skewness									
LHS	-1.88	-0.21	1.42	-0.17	-0.65	-0.23	0.00	-0.13	0.16
ε	-0.62	-0.86	0.70	-0.11	0.33	0.41	-0.01	0.59	0.48

Continued on next page

Panel C: 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}
C-1: Option-like features									
C-1-i: Henriksson-Merton Estimates									
α	2.67	0.96	-1.71	2.79	1.21	-1.58	0.81	0.32	-0.50
$t(\alpha)$	(2.67)	(2.35)	(-2.38)	(5.82)	(6.23)	(-4.50)	(1.40)	(0.84)	(-1.51)
β^0	-0.65	0.91	1.56	0.22	1.39	1.17	0.52	1.48	0.96
$t(\beta^0)$	(-5.46)	(14.09)	(17.98)	(1.83)	(25.31)	(13.89)	(2.96)	(10.89)	(13.27)
β^+	-0.92	-0.29	0.63	-0.61	-0.35	0.26	-0.14	-0.09	0.05
$t(\beta^+)$	(-3.31)	(-2.37)	(3.34)	(-2.07)	(-3.18)	(1.23)	(-0.42)	(-0.44)	(0.27)
$Adj.R^2$	0.50	0.70	0.83	0.03	0.80	0.67	0.16	0.81	0.75
C-1-ii: CAPM Estimates									
α	-0.07	0.10	0.16	1.76	0.62	-1.14	0.57	0.17	-0.40
$t(\alpha)$	(-0.09)	(0.33)	(0.31)	(8.50)	(5.77)	(-7.58)	(1.86)	(0.76)	(-2.50)
β	-1.15	0.75	1.91	-0.08	1.22	1.30	0.45	1.43	0.98
$t(\beta)$	(-9.05)	(14.61)	(22.14)	(-0.95)	(34.29)	(20.93)	(4.67)	(25.29)	(16.17)
$Adj.R^2$	0.47	0.69	0.82	0.00	0.79	0.67	0.17	0.81	0.75
C-2: Number of Momentum Losses worse than -20%									
	12			1			0		
C-3: Conditional Skewness									
LHS	-1.70	-0.02	1.44	-1.21	-0.73	0.50	0.04	-0.51	0.07
ε	-0.39	0.06	0.75	-0.72	-0.05	0.69	-0.09	0.31	0.70