Taming Momentum Crashes

Daniele Bianchi[†] Andrea De Polis[‡] Ivan Petrella[‡]

First draft: October 2021. This draft: June 24, 2022

Abstract

We provide evidence that the returns on the US equity momentum strategy exhibit a time-varying skewness which tend to deteriorates during dramatic losses (crashes). This has first order implications to understand the dynamics of risks and rewards associated with momentum portfolios over time: the momentum premium is compensation for *both* volatility and skewness risk. A dynamically adjusted momentum portfolio which hedge in real time for skewness risk, outperforms benchmark constant and dynamic volatility-managed momentum strategies. This result holds both in absolute and risk-adjusted terms, for different levels of transaction costs and risk aversion, and cannot be explained by the exposure to standard equity risk factors.

Keywords: Momentum, Time-varying skewness, Managed portfolios, Empirical asset pricing, Score driven models.

JEL codes: G11, G12, G17, C23

[†]School of Economics and Finance, Queen Mary University of London. E-mail: d.bianchi@qmul.ac.uk Web: whitesphd.com

[‡]Warwick Business School, University of Warwick. E-mail: andrea.depolis.17@mail.wbs.ac.uk

[‡]Warwick Business School, University of Warwick. E-mail: ivan.petrella@wbs.ac.uk Web: ivanpetrella.com

1 Introduction

One of the most studied capital market phenomenon is the relation between asset's return and its recent relative performance history, termed *momentum* effect. By betting on past returns, a momentum strategy has historically delivered competitive Sharpe ratios and high excess returns, and therefore became central to the market efficiency debate and one of the focal point in empirical asset pricing studies.[1] However, despite strong positive average risk-adjusted returns, momentum strategies are subject to "crashes", meaning infrequent periods of large and persistent negative returns (see Daniel & Moskowitz 2016).

These large losses are often associated with a time-varying exposure to systematic risk factors. For instance, Kothari & Shanken (1992) argue that sorted portfolios inherit exposure to market risk from the formation period. As a result, momentum portfolios formed during bear markets are likely to bet against high beta stocks, thus generating adverse performance when these periods are over (see, e.g., Grundy & Martin 2001). Following the major market decline of 1932, during the tail of the great depression, the momentum strategy up-market beta is almost three times larger than the down-market beta; similarly, towards the end of the great financial crisis of 2008/2009 the up-market beta of the momentum strategy is more than five times larger than the down-market beta.

A conventional approach to mitigate the effect of the asymmetric exposure of momentum portfolios to market risk is to leverage up or down the winner-minus-loser (`WML`) portfolio based on the conditional volatility of the strategy returns (see, e.g., Barroso & Santa-Clara 2015, Daniel & Moskowitz 2016). This grounds on the fact that momentum returns volatility is highly predictable, whereas the conditional expected returns, i.e., momentum risk premium, is often hard to measure due to the relatively low signal-to-noise ratio of the realised returns. However, while focusing on volatility to dynamically manage the risk of a momentum strategy certainly simplifies the empirical analysis, assuming risks can be unequivocally mapped by volatility is based on the unspoken assumption that momentum crashes affect only the scale of the returns distribution but not its asymmetry over time. That is, skewness risk is ignored.[2]

In this paper we offer a different method to measure and manage risk associated to the momentum strategy. Our approach is that momentum risk premium does not only represent

[1] Jegadeesh (1990) first document that stocks that performed well in the past have the tendency to outperform the market, while stocks that performed poorly tend to underperform it. Grinblatt, Titman & Wermers (1995) find that momentum strategies are common among investment funds, while several papers document the pervasiveness of this anomaly across countries Rouwenhorst (1998), Fama & French (2012) and asset classes (see, e.g., Moskowitz & Grinblatt 1999, Moskowitz, Ooi & Pedersen 2012, Asness, Moskowitz & Pedersen 2013).

[2] Skewness risk in financial modeling is the risk that results when observations are not spread symmetrically around an average value, but instead have a skewed distribution. As a result, the mean and the median can be different.

a compensation for risk expressed by volatility, but also for the exposure to skenwess risk and its interplay with conditional volatility. We show both analytically and in simulation that considerable time variation in the strategy returns skewness simply arises from the asymmetric exposure of the `WML` portfolio to normally distributed market returns. Yet, if returns asymmetry varies over time, volatility is not an accurate representation of the strategy's risk (see, e.g., Glosten, Jagannathan & Runkle 1993). However, the explicit dynamics and effect of skewness risk on momentum premium has been mostly overlooked so far in the literature.

The importance of the third moment of the distribution of returns has recently been investigated more generally by Harvey & Siddique (2000), that integrate conditional skewness within an otherwise asset pricing model. Jacobs, Regele & Weber (2015) uncover a robust relation between expected skewness and cross-sectional momentum, specifically in relation to past `losers`. Intuitively, the reason why skewness might be as important as volatility to model the momentum strategy risk is straightforward, if not often appreciated: volatility scaling is based on the assumption that the conditional distribution of the momentum returns is normal, that is the conditional mean and median are the same. Assuming returns are spread symmetrically around an average value over time implicitly ignores the fact that negative shifts in the risk premium of the strategy might occur even in periods where the volatility does not change significantly. As such, volatility targeting might represent, at least in theory, an insufficient tool to mitigate the risk of momentum portfolios especially during period of dramatic losses (crashes). In other words, by assuming that the momentum strategy returns are symmetrically distributed around the conditional mean at each time t, when they are not, could cause volatility-managing adjustments to understate the overall risk of the strategy during momentum crashes.

In order to tease out the economic significance of time-varying skewness risk in US equity momentum, we propose a parametric model which allows to recover the location, scale and asymmetry of the strategy returns distribution over time. This allows to provide a deeper understanding of the role of volatility, skewness, and their interplay in shaping the dynamics of momentum risk premium. More specifically, our modeling framework allows to characterize the closed-form dynamics of the strategy risk premium, i.e., conditional expected returns, and conditional volatility as a direct function of the estimated returns asymmetry. As a result, we can single out a skewness-hedging component from an otherwise standard dynamic strategy that maximises the conditional Sharpe ratio (see, e.g., Daniel & Moskowitz 2016).

1.1 Findings

Building upon the intuition of Kothari & Shanken (1992), Grundy & Martin (2001), we show both analytically and in simulation that the asymmetric exposure of the momentum strategy to market risk generates skewness in the conditional distribution of the `WML` portfolio returns. This holds even assuming both the returns on the market portfolio and the error terms are normally distributed. Simple recursive estimates of the up- and down-market betas show that the asymmetry in the exposure to market risk tend to widen during momentum crashes. This suggests that the asymmetry of the conditional distribution of momentum returns might not necessarily be constant over time. We build upon this intuition and investigate the shape of the conditional distribution of the daily returns on a standard momentum strategy in US common stocks over the 1927-2020 time period. The main contribution of the paper is fourfold.

First, we find a significant time variation in the skewness risk associated with momentum investing, which accounts for the large crashes that the strategy has experienced in 1932, 2001 and 2009. This is due to a different pattern of upside and downside risk of the loser and winner portfolios. For instance, although the skewness of the past `winners` portfolios significantly decreased in the aftermath of the great depression, the conditional skewness of the returns on the past `losers` portfolios turned highly positive. That is, the `WML` portfolio implicitly was "buying" a moderately lower downside risk, but at the same time was "selling" a substantially higher upside risk, with the latter more than offsetting the former. As a result, the conditional skewness of the `WML` tend to become more negative during the tail of recessions and throughout periods of large drawdowns.

The second main result pertains the role of time-varying skewness for the dynamics of the momentum premium. Notice that our framework allows to define the conditional expected returns on the momentum strategy as a direct function the location, i.e., mode, the scale and the asymmetry of the returns distribution. A second-order Taylor expansion shows that the risk premium associated to the momentum strategy is not only related to time-varying volatility, but it is also affected by the time variation of returns asymmetry. This result points to a nonlinear interaction between volatility and skewness risk in shaping the momentum premium. For instance, increasing skewness risk tend to dominate volatility risk at times during the momentum crashes of 1930s, 2001 and 2009. On the other hand, volatility plays a dominant role for the dynamics of momentum expected returns primarily within recessions. More generally, we show there is a significant amplification mechanism between skewness and volatility that helps to explain the dynamics of risk premiums for the momentum strategy, especially during the 2001 dot-com bubble and the great financial crisis of 2008/2009. The third main result relates to the dynamics of the risk-return trade-off of the momentum strategy. The evidence on the shape of the trade-off between risks and rewards for momentum strategies has been often inconclusive. Theodossiou & Savva (2016) argue that such ambiguity is primarily due to the fact that returns embed a *linear* combination of skewness and volatility premiums. When the skewness of the predictive distribution is negative, these two premiums offset each other, generating a highly uncertain dynamics for the risk-return trade-off. Our approach allows for a more general, *non-linear*, *time-varying*, interaction between skewness and volatility premiums. We document a time-varying risk-return trade-off which is mostly negative particularly around momentum crashes: while the premium pertaining volatility is only mildly negative, the non-linear interaction with conditional skewness further exacerbates the slope of the risk-return profile, which becomes even more negative. This is due to the effect of conditional skewness on the risk premiums rather than on returns' conditional variance. Larger negative skewness triggers a downward correction of the expected returns on the momentum strategy, whereas we find a less pervasive impact of skewness on the variance of the returns. The limited effect of the skewness on volatility is likely mitigated by the presence of heavier tails in the returns distribution, which are explicitly accounted for by the model.

The last main result of the paper relates to the economic value of explicitly modeling time-varying skewness risk in momentum portfolios. We expand from Daniel & Moskowitz (2016), and derive an economically significant skewness-hedging component from an otherwise standard optimal dynamic strategy that maximises the conditional Sharpe ratio. More specifically, we leverage the direct dependence of the risk premium on volatility and skewness within our model and disentangle two components in the optimal managed portfolios; a first component that shrinks the allocation in the risky momentum strategy as volatility increases, and a second component which, for a given level of volatility, further shrinks the optimal allocation to zero as the returns skewness becomes more negative. We show that managed portfolios that account for time-varying skewness risk fare better than the dynamic or constant volatility adjustments proposed by Barroso & Santa-Clara (2015), Daniel & Moskowitz (2016). Perhaps more importantly, our skewness-managed momentum strategy generates higher Sharpe and Sortino ratios as well as a performance which is positively skewed, that is, unconditionally, the probability of a large gain based on our enhanced momentum strategy is higher than the probability of a large loss, unlike most competing approaches. In addition, the returns on our skew-managed momentum strategy are not spanned by the returns on the market, standard Fama & French (1993) factors, or by the returns on a constant or dynamic volatility managed momentum portfolio.

1.2 Literature

In addition to Barroso & Santa-Clara (2015) and Daniel & Moskowitz (2016), our work relates to a long standing literature that seeks to understand the origins and the dynamic properties of momentum returns, such as Jegadeesh (1990), Rouwenhorst (1998), Moskowitz & Grinblatt (1999), Griffin, Ji & Martin (2003), Moskowitz et al. (2012), Novy-Marx (2012), Asness et al. (2013), among others. Building upon the intuition that market risk exposure in momentum strategies might not be symmetric, in this paper we first show that this generates asymmetry in the returns conditional distribution and then focus on the dynamics and effect of skewness risk in the strategy risk premium.

A second strand of literature we relate to is about the role of higher order moments in portfolio allocation (see, e.g, Guidolin & Timmermann 2008) and asset pricing models (see, e.g., Dittmar 2002, Harvey & Siddique 2000, Kraus & Litzenberger 1976). Within the context of momentum strategies, a variety of approaches have been proposed; for instance, Grundy & Martin (2001), which suggest to dynamically hedge systematic risk factors to reduce the risk of the strategy. However, Daniel & Moskowitz (2016) show that this strategy is not implementable in a realistic setting as it suffers from forward-looking bias. Instead, they proposed an adjusted strategy improving upon the volatility managed portfolio approach of Barroso & Santa-Clara (2015) and Moreira & Muir (2017).

Finally, this paper connects to a third strand of literature which propose the use of dynamic econometric modeling for non-Gaussian distribution. More specifically, we rely on the score-driven framework put forward by Creal, Koopman & Lucas (2013) and Harvey (2013). This setting has proven to be particularly suitable for accommodating parameters' time variation under different distributional assumptions (Koopman, Lucas & Scharth 2016).[3] To the best of our knowledge, our paper is the first to propose the use of such framework within the context of risk managed portfolios. Among the advantages of the score-driven dynamics, the robustness to outlying observations is the perhaps the most prominent. This is particularly relevant when working with daily portfolio returns, as in our case. The filtering framework algorithm which is used to estimate the model parameters over time is, in fact, robust to the information supplied by extreme observations, and turns them into feasible updates for the parameters of interest. As a result, parameter estimates tend to be less volatile and better identified, compared to Hansen (1994) and Harvey & Siddique (1999). In addition, unlike existing methodologies, our score-driven model allows to derive in closed form both conditional expected returns and volatility, as well as the risk-return trade-off which explicitly depend on the conditional asymmetry in the returns distribution. This gives a clear advantage when dissecting the marginal contribution of skewness in the dynamics of risk premiums, conditional variance, as well as on the shape of the trade-off between risks and rewards.

[3]Parameters' updating based on the score of the predictive likelihood always reduces the local Kullback-Leibler divergence between the true conditional density and the model-implied one, even under severe mis-specification (Blasques, Koopman & Lucas 2015).

2 The returns in US equity momentum

We construct momentum decile portfolios based on daily returns of firms listed on the NYSE, Amex and Nasdaq, with CRSP code 10 or 11. Decile portfolios are constructed daily but are rebalanced monthly. We follow Daniel & Moskowitz (2016) and form portfolios based on an all-firms breakpoints, that is an equal number of firms is present in each decile portfolio, rather than an equal number of just NYSE firms as in Fama & French (1996).[4] The risk free rate is the monthly 1-month T-bill rate, and the market return is the value-weighted index of all the CRPS firms; both series are obtained from Kenneth French data library.[5] Decile portfolios are formed on the basis of eleven-month ($J = 11$) look-back period, from months $t - 12$ through $t - 1$, and are rebalanced each month ($K = 1$). We skip the most recent month, which is our formation period, to avoid the short-term reversal documented by Jegadeesh (1990) and Lehmann (1990). Stocks are sorted into deciles, ranked on the basis of their performance over the past J months, then the momentum strategy consists of investing 1\$ in the portfolio of past `winners` (the $10th$ decile) and selling 1\$ of past `losers` (the $1st$ decile), with a $K-$months holding period. This 12_2 long-short portfolio is often referred to as the `winners`-minus-`losers` (`WML`) strategy.

The left panel of Figure (1) compares the cumulative performance of investing 1\$ in the `WML` portfolio against a buy-and-hold investment in the market and the risk-free rate. The performance is calculated from the second half of the 1920s holding the investment until the end of 2020. Clearly, momentum has delivered a substantially higher performance with respect to both the aggregate stock market and the risk-free rate over the last century. The average excess return of the `WML` portfolio is close to 19%, almost three times the 7.5% offered by the market. On the other hand, the volatility is somewhat comparable, with an unconditional annualised standard deviation of 24% for momentum vs 17% of the market. This translates in a Sharpe ratio for the momentum strategy of 0.78 on an annualised basis, almost double with respect to the aggregate stock market (Sharpe ratio: 0.44). Figure 1 also suggests that the market beta on the `WML` is also reasonably low, at least unconditionally. A simple static CAPM regression shows that the market beta is indeed slightly negative and significant (β=-0.15, t-stat=-17.4), which result in a highly positive and significant annualised Jensen's alpha (α=22.2, t-stat=8.3). In other words, a typical momentum strategy á la Jegadeesh & Titman (1993) not only produce a quite high risk-adjusted performance, but also possibly provides a significant diversification with respect to aggregate market fluctuations.

Nevertheless, the right panels of Figure 1 show that despite its strong performance, momentum has experienced few severe downturns. Daniel & Moskowitz (2016) define these

[4]Refer to http://www.kentdaniel.net/data/momentum/mom_data.pdf for additional information.
[5]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

events as momentum *"crashes"*. They are associated with extremely negative monthly returns, ranging from -90% to -75%, in 1932 and 2009, respectively; at the daily frequency, cumulative returns over the same months produced losses of -65% to -70%. Sample statistics suggest that the higher excess returns are associated with higher variance, with momentum bearing about 50% more risk than the market, and significantly lower skewness. Thus, the profitability of this strategy comes at the cost of significant drawdowns.

Intuitively, the occurrence of sporadic but persistent large negative returns induce significant asymmetry in the distribution of momentum portfolio returns. In Table 1 we report different measures of skewness for both the past `winners`, past `losers` and the `WML` portfolio. The first row reports the sample skewness, defined as the standardized third moment of the sample distribution of returns. The p-values for the Bai & Ng (2005) significance test are reported in parentheses. The sample skewness measure suggests that the portfolio of past `losers`, the short-leg of the momentum strategy, does not show any significant asymmetry, as suggested by the p-values. On the other hand, the market, past `winners` and the `WML` strategy all show significant, negative skewness, with the latter displaying twice the asymmetry with respect to past `winners`.[6] These values hint that the strategy is often subject to extreme losses, not being matched by gains of similar magnitude. In addition to the sample skewness, we also report the Bowley skewness measure calculated as $QS_\alpha = \frac{q(\alpha)+q(1-\alpha)-2q(50)}{q(\alpha)-q(1-\alpha)}$, with $q(\alpha)$ being the αth quantile of the data. Such measure is robust to the presence of outliers, which tend to weaken the validity and interpretability of returns asymmetry measures (see Kenney & Keeping 1939, Kim & White 2004). The second and third rows in Table (1) reports the results of two robust specifications of returns asymmetry based on different quantiles, $QS_\alpha,\ \alpha = 95, 90$. Even after accounting for the presence of outliers, `WML` returns, as well as past `winners`', still display marked negative asymmetry. The estimates for the past `winners` and `losers` suggest that, as we remove more outliers, past `winners` tend to be more negatively skewed compared to the strategy. This suggests that the strategy skewness is mostly induced by large losses realized by past `winners`.

2.1 Asymmetric market betas and skewness risk

Grundy & Martin (2001) argue that long-short portfolios formed during bear markets are likely to sell high-beta, and buy low-beta stocks as firms dropping with the market will still be high-beta firms. Daniel & Moskowitz (2016) document the asymmetric nature of the market risk exposure of momentum strategies is at the core of momentum crashes and is primarily due to the past `losers` portfolio: during bear markets and high volatility, the short leg of

[6] Daniel & Moskowitz (2016) report the same measure for monthly momentum returns over the period 1927-2013. They find that `WML` returns display a skewness of -4.70, against a skewness of just -0.82 for past `winners`.

the strategy commands a higher premium, resulting in higher gains as the market rebounds. This pattern in dynamic betas of the loser portfolio implies that momentum strategies in bear markets behave as written call options on the market; that is, when the market falls, they gain a little, but when the market substantially rises – at the tail of, or outside, recession periods –, they lose much. The same intuition is underlying the volatility-managed momentum strategy proposed by Barroso & Santa-Clara (2015). Similarly, Dobrynskaya (2015) finds that unconditionally decile momentum portfolios show a remarkable monotonic increase in the level of downside risk, suggesting that past `winners` tend to be associated with higher downside risk, whereas past `losers` show higher upside risk.

We now build upon this existing evidence and show that the asymmetric exposure to market risk generate negative and possibly highly time varying returns skewness. Let consider a standard conditional CAPM specification which separates up-market and down-market betas (see, e.g., Ang, Chen & Xing 2006, Lettau, Maggiori & Weber 2014),

$$r_t = \alpha + \underbrace{\overline{\beta} m_t I(m_t \geq \mu_m) + \underline{\beta} m_t I(m_t < \mu_m)}_{\beta m_t} + e_t \tag{1}$$

with $m_t \sim \mathcal{N}\left(\mu_m, \sigma_m^2\right)$ the normally distributed market portfolio and $I(m_t \geq \mu_m)$ ($I(m_t < \mu_m)$) an indicator function that takes value one if the market returns are above (below) the mode μ_m and zero otherwise. In Appendix A we show that the distribution of βm_t conditional on the indicator $I(\cdot)$ can be characterised as a split-Normal (or two-piece Normal) distribution (see Johnson, Kotz & Balakrishnan 1995) such that the difference between the expected value $E\left[\beta m_t\right]$ and the mode $\beta \mu_m$ takes the form,,

$$E\left[\beta m_t\right] - \beta \mu_m = \sqrt{\frac{2}{\pi}} \left(\overline{\sigma}_m - \underline{\sigma}_m\right) \propto \sigma_m \left(\overline{\beta} - \underline{\beta}\right) \tag{2}$$

with $\underline{\sigma}_m^2 = \underline{\beta}^2 \sigma_m^2$ and $\overline{\sigma}_m^2 = \overline{\beta}^2 \sigma_m^2$. It is easy to see that under the assumption of equal betas across market states, i.e., $\underline{\beta} = \overline{\beta} = \beta$, there is no asymmetry in the marginal returns of the distribution such that $E\left[\beta m_t\right] = \beta \mu_m$ (see Arnold & Groeneveld 1995). Also, in Appendix A we show that for equal up-market and down-market betas the variance of the systematic component is defined as $V\left[\beta m_t\right] = \beta^2 \sigma_m^2$, such that the marginal distribution of the momentum strategy returns is equal to the standard CAPM formulation $r_t \sim \mathcal{N}\left(\alpha + \beta \mu_m, \beta^2 \sigma_m^2 + \sigma_e^2\right)$.

On the other hand, with asymmetric betas $\underline{\beta} \neq \overline{\beta}$, Eq.(2) shows that for $\overline{\beta} < \underline{\beta}$ ($\overline{\beta} > \underline{\beta}$) the expected value of the systematic CAPM component is lower (higher) than the mode, that is the marginal distribution of the returns is negatively (positively) skewed even assuming that the market returns and the residual e_t are both normally distributed. Figure 2 reports the unconditional estimates of the upside and downside market betas for both the past `losers` and `winners` portfolios as well as the `WML` strategy. The left (right) panel reports the estimates based on daily (monthly) returns.[7] The daily estimates show that the `losers` portfolio are more exposed to upside market risk ($\overline{\beta} = 1.36$) vs downside market risk ($\underline{\beta} = 1.27$), in relative terms compared to the unconditional market beta ($\beta = 1.31$). The opposite holds for the `winners` portfolio ($\overline{\beta} = 1.09$, $\underline{\beta} = 1.22$, $\beta = 1.16$), consistent with the intuition in Grundy & Martin (2001). As a result, the `WML` strategy has a quite sizable and negative up-market beta ($\overline{\beta} = -0.27$) while the down-market beta is close to zero ($\underline{\beta} = -0.04$). The magnitude of the spreads in the upside and downside market betas is even higher at the monthly frequency (right panel of Figure 2).

The left panel of Figure 3 shows the marginal distribution and joint distributions of the returns on the `WML` strategy and the market portfolio. Returns are simulated assuming a two-piece Normal distribution as in Eq.(A2) in Appendix A by using the unconditional $\overline{\beta}$ and $\underline{\beta}$ estimates above. In order to isolate the effect of the upside and downside beta estimates returns on the market portfolio are assumed to be normally distributed with mean zero and volatility equal to the historical standard deviation. The error terms e_t is also normal with zero mean and variance equal to one. Despite both the market returns and the residuals are assumed normally distributed, the negative spread in market betas generate a slightly negatively skewed (`skew` = -0.1) marginal distribution.

As suggested by Daniel & Moskowitz (2016), the spread between upside and downside betas tend to exacerbate during momentum crashes. Thus, we expand the unconditional results in Figure 2 follow Ang et al. (2006), and calculate the downside market beta over time for the `losers`, `winners` portfolios and the `WML` strategy as,

$$\underline{\beta}_t^i = \frac{\text{cov}_t(\tilde{r}_{t+1}^i, \min\{\tilde{m}_{t+1}, 0\})}{\text{var}_t(\min\{\tilde{m}_{t+1}, 0\})} \quad i = \texttt{losers}, \texttt{winners}, \texttt{WML}, \tag{3}$$

where $\tilde{r}_t^i$ and $\tilde{m}_t$ are the demeaned returns for the momentum strategy and the demeaned excess market returns, respectively (see, e.g., Hogan & Warren 1974). The denominator of (3) captures the variance of the downside market excess returns, and is generally referred to as the relative semi-variance. Therefore, high downside betas imply that return is significantly exposed to market's downswings. Upside betas $\overline{\beta}_t^i$ hold a similar interpretation and are computed by substituting the *min* function in (3) with the *max* operator.

Figure 4 reports the estimates for the spread $\mathcal{B}_t = \overline{\beta}_t^{\texttt{WML}} - \underline{\beta}_t^{\texttt{WML}}$ for the periods indicated as

[7]The estimates of the upside, $\overline{\beta}$, and downside, $\underline{\beta}$ betas are based on the following regression:

$$r_t^i = \alpha + \underline{\beta}^i \min(r_t^m, 0) + \overline{\beta}^i \max(r_t^m, 0) + \varepsilon_t, \; i = \texttt{losers}, \texttt{winners}, \texttt{WML}.$$

momentum crashes by Daniel & Moskowitz (2016).[8] To estimate the time-varying downside and upside betas for the momentum strategy returns, we follow Bali & Engle (2010), Tsai, Chen & Yang (2014) and implement a dynamic conditional correlation (DCC) model as originally proposed by Engle (2002). For the easy of exposition we report both the daily DCC estimates of $\mathcal{B}_t$ as well as a smoothed version of the estimates based on a quarterly moving average of the daily estimates. Recessions are highlighted in gray where momentum crashes are color-coded in red shading. Except few nuances, the spread $\mathcal{B}_t$ is primarily negative during the momentum crash of the 30's (left panel). The difference between upside and downside betas tend to spike in 1935 and 1938, although remains persistently large and negative for the entire decade. The momentum crash of the 2001/2002 (right panel) shows a slightly different dynamics, with $\mathcal{B}_t > 0$ during the dot-com bubble collapse, which then switch negative towards the tail of the recession. The momentum crash during the great financial crisis of 2008/2009 is characterised by a large negative spread between upside and downside betas for the `WML` portfolio returns. The $\mathcal{B}_t$ difference is persistently negative and is as large as -2.5 on a daily basis.

As per Eq.2, the differences in the spread $\mathcal{B}_t$ should lead to a higher skewness in the marginal distribution of the momentum returns. The middle and the right panels in Figure 3 shows the simulation results based on the conditional estimates for two specific time stamp of the momentum crashes reported in Figure 4. Consistent with the intuition from an asymmetric CAPM formulation the implied returns skewness of momentum returns markedly differ from the full sample simulation. For instance, in March 1935 – in the middle of the largest momentum crash – the average quarterly difference $\overline{\beta} - \underline{\beta}$ is as large as -1.5. As a result, the marginal distribution of the momentum returns (middle panel) is much more negatively skewed (`skew` = -0.805). Similarly, in the middle of the great financial crisis of 2008/2009, the average quarterly spread $\mathcal{B}_t$ expands to -1.4, with a corresponding `skew` = -0.529 of the simulated momentum returns.

3 Modeling the returns distribution

The asymmetry in the market betas for both the `WML` strategy (see Figure 4) and their dynamic effect on the marginal distribution of the momentum returns (see Figure 3), suggests that returns asymmetry might not be constant over time and possibly more pronounced during periods of large negative (positive) upside (downside) betas. Yet, it is important to notice that downside and upside risk exposure of past `winners` and past `losers` do not linearly map into the long-short strategy exposure to market risk.

[8]For the ease of exposition, the estimates for both the `losers` and the `winners` portfolios are not reported in the main text. They are available upon request to the authors.

We now introduce a flexible modeling framework which allows to capture the time variation in the first three moments of the conditional distribution of the momentum strategy returns, that is the location, scale and asymmetry parameters. Let r_t be the return on the momentum portfolio at time t:

$$r_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim Skt_\nu(0, \sigma_t^2, \rho_t) \tag{4}$$

where the innovations ε_t follow a Skew-t distribution with ν degrees of freedom (see Gómez, Torres & Bolfarine 2007).[9] The time-varying parameters $f_t = (\mu_t, \sigma_t, \rho_t)'$ represent the location μ_t, the scale σ_t, and the asymmetry ρ_t of the conditional distribution of the momentum returns. More specifically, the third moment ρ_t is key for the shape of the distribution of the returns; when $\rho_t > 0$ ($\rho_t < 0$) the conditional distribution of the momentum strategy returns features negative (positive) skewness at time t. The vector of moments $f_t = (\mu_t, \sigma_t, \rho_t)'$ is updated at each time t in a data-driven fashion as in Creal et al. (2013) and Harvey (2013):

$$f_{t+1} = f_t + \kappa s_t \tag{5}$$

where s_t is the scaled score, and κ contains the structural parameters regulating the law of motion of the moments. The scaled score is defined as $s_t = \mathcal{S}_{t|t-1}\nabla_t$, with $\nabla_t = \frac{\partial \ell_t}{\partial f_t'}$ being the gradient of the conditional log-likelihood function with respect to the dynamic location, scale and asymmetry parameters. The scaling matrix $\mathcal{S}_{t|t-1}$ is proportional to the square root of the Moore-Penrose pseudo-inverse of the diagonal of $diag(\mathcal{I}_t) = diag\left(\mathbb{E}\left[-\frac{\partial^2 \ell_t}{\partial f_t \partial f_t'}\right]\right)$, i.e., the information matrix.[10] As a result, the scaled score is a martingale difference sequence with conditional variance equal to one.[11] The full derivation of the scaled scores in s_tm the gradients of the conditional log-likelihood and the scaling matrix $\mathcal{S}_{t|t-1}$ is provided in Appendix B.1.

Notably, our model belongs to the observation-driven class (see Cox 1981, for details), as the dynamics of the moments of the conditional distribution f_t is entirely data driven, i.e., is a function of past obervations only. For instance, the GARCH models of Engle (1982), Bollerslev et al. (1987) are observation-driven model which can be nested within our model specification as special cases.

[9] This distribution is particularly suitable to the case into consideration as it nests the Student-t, skew-Gaussian and Gaussian distributions, as liming cases.

[10] Creal et al. (2013) suggest alternative scaling matrix as the identity matrix or $\mathcal{I}_{t|t-1}^{-1}$. The former choice leaves the score unscaled, with $\mathbf{Var}[\mathbf{s}_t] = \mathcal{I}_{t|t-1}$, whereas the latter scaling matrix sets the variance of the scaled score to the inverse of Information matrix.

[11] Note, however, that the covariance matrix is not diagonal as there is a covariance term between location and scale that is non-zero. We choose to ignore the off-diagonal elements of $\mathcal{I}_t$ for computational stability purposes and to restrict the feedback effect of the parameters to their own score, as in Lucas & Zhang (2016).

Estimation procedure. A feature of observation-driven models is the straightforward computation of the likelihood function (see, e.g., Creal et al. 2013, Harvey 2013). However, the optimization and computation of confidence intervals remain challenging, in particular when models are rich in parameters and non-linearities. For this reason, we opt for a Bayesian estimation procedure for the structural parameters of the model. Perhaps more importantly, Bayesian estimators are not affected by local discontinuities and potentially multiple local minima and/or flat areas of the likelihood. This is because one can more easily explore the entire surface of the likelihood function, particularly in the high-dimensional setting, via simulation from the prior distributions from different initial random samples (see, e.g., Tian, Liu & Wei 2007, Belloni & Chernozhukov 2009).

Let κ be the score loadings, η is the inverse of the Skew-t degrees of freedom and $\bar{f}_0 = [\bar{\mu}_0, \bar{\delta}_0, \bar{\gamma}_0]$ collects the initial values of the time-varying parameters. Prior specifications for these parameters read:

$$\kappa \sim \mathcal{G}^{-1}(a_\kappa, b_\kappa), \qquad \eta \sim \mathcal{G}^{-1}(a_\eta, b_\eta) \cdot I_{(\eta \in H)}, \qquad \bar{f}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{s}_0)$$

We set Inverse Gamma priors for the score loadings such that they reflect the a-priori expectation of small but positive coefficients, in line with the properties of the score-driven filters (for further discussion, see Blasques, Koopman & Lucas 2014). We set $a_\kappa = 4$, and $b_\kappa = 1$, such that the prior distribution has a mean of 0.3 and variance of 0.05. We use an inverse gamma prior for η, the inverse of the degrees of freedom parameter, with $a_\eta = 2$ and $b_\eta = 10$. In line with Juárez & Steel (2010), these values allow the distribution to explore a wide range of feasible values for ν, with a mean of 20 and a median of 10. In order to ensure the existence of, at least, the first three moments, we restrict the support to the $H = [0, 0.33]$ set. The initial values of the time-varying parameters are drawn from a multivariate Gaussian distribution, with mean vector $\mathbf{m}_0$, and $\mathbf{s}_0 = 0.5\mathbf{I}_3$. The mean vector $\mathbf{m}_0$ is obtained by matching the unconditional properties of the data.

Once the prior distributions are specified, the parameters of the time-varying distribution are estimated via simple simulation-based optimization, that is by solving $\arg\max_\theta \ell_t(r_t, f_t; \mathcal{F}_{t-1}, \theta) \times \pi(\theta)$, where θ^* is the vector of parameter that maximizes the posterior distribution of the model. $\pi(\theta)$ contains prior information about the model's parameters, and ℓ is the likelihood function of the data (see Chernozhukov & Hong 2003 for more details).

3.1 Momentum premium dynamics

Our modeling framework allows to explicitly consider the effect of both the location μ_t, scale σ_t and asymmetry ρ_t parameters on the dynamics of the expected returns on the momentum strategy. The conditional expected returns are of particular relevance for our purpose. As a matter of what, within the context of reduced-form empirical asset pricing models, the ex-ante expected returns on zero-cost long-short portfolios is often considered a proxy for the strategy risk premium (see, e.g., Gu, Kelly & Xiu 2020).

Within the context of our modeling framework, the expected value of the returns conditional on information at time t can be derived by reparametrizing the Skew-t density of Gómez et al. (2007) as a two-piece distribution which allows to model the conditional moments as weighted averages of the moments of a Half-t distribution (see Arellano-Valle, Gómez & Quintana 2005, Gómez et al. 2007). Specifically, the moment generating function for the returns distribution is computed as:

$$\mathbb{E}[r^j] = \sum_{k=0}^{j} \binom{j}{k} \sigma^k \mu^{j-k} \hat{\mu}_k. \tag{6}$$

Therefore, the expected return at time t on the strategy returns follows:

$$\mathbb{E}_t(r_{t+1}) = \mu_t - g(\nu)\rho_t\sigma_t, \qquad \nu > 1 \qquad \text{with} \qquad g(\nu) = \frac{4\nu\mathcal{C}(\nu)}{\nu - 1} \tag{7}$$

where $\mathcal{C}(\cdot)$ is a combination of Gamma functions and constants, and ν are the degrees of freedom. A full derivation of the moment generating function, and the corresponding expected returns, from the two-piece Half-t distribution is provided in Appendix C. Based on Eq.(7), the risk premium on the momentum strategy depends on both the scale and the asymmetry of the returns distribution at each time t. The parameter ρ_t is of particular importance since drives the returns asymmetry (or the shape) and maps directly into the value of the conditional skewness.

To better understand the role of the scale and asymmetry parameters on the model-implied risk premium, we implement a comparative static analysis by looking at the changes in expected returns from ρ_t, σ_t and the degrees of freedom ν. Figure (5) shows the results assuming for simplicity that $\mu_t = 0$. The left panel reports the expected value as a function of ρ_t and σ_t. To increase the readability of the heatmap, we also report the partial derivative of (7) with respect to ρ_t and ν for varying values of σ_t. Two facts emerge: first, the effect of the returns asymmetry on expected returns is amplified by the scale σ_t. When the distribution of the returns is negatively skewed, i.e., $\rho_t > 0$, the higher the volatility the lower is the conditional expected returns, i.e., negative risk-return trade-off. On the other hand, when the returns distribution is positively skewed, i.e., $\rho_t < 0$, the conditional expected return increases with σ_t. This is consistent with the idea that high volatility with positive skewness is associated with large positive gains from the strategy. Second, the interplay between asymmetry and volatility is multiplicative, that is the curvature of the expected returns as a function of the scale σ_t increases more than linearly as a function of the returns asymmetry ρ_t. The steepness of the curvature is regulated by the degrees of freedom ν, as suggested by the behaviour of the partial derivatives. This suggests that the more extreme the return observations are, the higher the sensitivity of the conditional expectation to changes in returns scale and asymmetry.

The right panel of Figure (5) shows the effect of the returns asymmetry ρ_t and the thickness of the tails, as proxied by the degrees of freedom ν on the conditional expected returns. Thicker tails, meaning lower values of ν push the strategy expected returns to more extreme values, depending on the sign of the returns asymmetry. For instance, for a small value $\nu = 2$, the expected returns range between -3 and +2 for $\rho_t = 1$ and $\rho_t = -1$, respectively. For higher values of ν, the sensitivity of the expected returns to changes in the asymmetry parameter decreases. The fact that the effect of a change in the shape parameter depends on ν, is perhaps more visible in the plot of the partial derivative of (7) with respect to ρ_t: the effect of the returns asymmetry on the risk premium is mitigated by thinner tails of the returns distribution, while still dictating the sign of the expected returns.

3.2 Time-varying skewness as a sufficient statistics

Given the tight relationship between the distance between up-market and down-market betas and returns asymmetry shown in Section 2.1, one can consider using dynamics betas within a conditional Gaussian framework to capture momentum crashes instead of using our approach. While, in principle, this is a suitable modeling choice, we believe that our framework provides several practical advantages. First, our approach allows to characterise the dynamics of momentum strategy risk and returns on a daily basis, instead of aggregating returns at the conventional monthly frequency. On the one hand, this provides a much more granular representation of risks and returns in momentum portfolios. On the other hand, the returns conditional volatility and skewness can be modeled daily, without the need to use often hard to obtain intra-day returns. The aggregation aspect is particularly relevant when it comes to the properties of the realized variance and skewness of portfolio returns. For instance, in Appendix D we show that standard realised volatility estimates based on the monthly aggregation of the daily squared portfolio returns – as in Barroso & Santa-Clara (2015), Daniel & Moskowitz (2016) – tend to produce distorted monthly realised volatility estimates during returns crashes. Perhaps more importantly, the second main advantage of our modeling framework is that, by modeling the distribution of the residuals as a Skew-t, we can capture sources of asymmetries beyond the spread in the upside vs downside market betas. As a matter of fact, Section 2.1 suggests that while a great deal of asymmetry in the returns distribution can be captured by asymmetric betas, there is still a considerable amount of skewness in the marginal distribution of the momentum returns which could come from either the market portfolio and/or the residuals. This would require to specify a much more complex model with time-varying betas and volatility, as well as non-Normality in the residual terms. Related to this, Appendix B shows that the updating scheme of the location, scale and asymmetry parameters is robust to outlying observations. In other words, unlike standard realised variance and skewness measures which are notoriously sensitive to outliers (see Kenney & Keeping 1939, Kim & White 2004), our modeling framework can track the time-variation of skewness only if it is actually present in the conditional distribution of the returns once outliers are filtered out (see, Harvey 2013, Blasques et al. 2015). More specifically, the score-driven mechanism updates the time-varying moments f_t of the conditional distribution depending on the slope and curvature of the log-likelihood function, in that it maximises the local fit of the model at each point in time. To summarise, we view our approach as a parsimonious, robust, and more comprehensive method to capture returns asymmetries in the conditional distribution of the momentum returns compared to modeling the dynamics of the market betas.

4 Expected returns, volatility and skewness

The location parameter capture the center of the distribution and it is equivalent to the conditional mean for models with symmetric distributional assumptions, i.e., $\rho_t = 0$. To initially gauge the magnitude and the dynamic of returns asymmetry we look at the estimates of the conditional scale parameter μ_t against the conditional mean of the returns as per Eq.(7). The left panel of Figure 6 reports the estimates for the `WML` portfolio. Two things emerge: first, there is a considerable time variation in the location parameter (red line), which supports the idea of a time-varying risk premium (see Eq.(7)). Second, there is a major disconnection between the location parameter and the conditional mean $E_t[r_{t+1}]$. Such disconnection is particularly pronounced during the momentum crashes of 1932-1939 and 2008-2009. For instance, while the expected returns from the `WML` strategy become largely negative in the aftermath of the great depression and the great financial crisis, the location parameter μ_t remains largely in positive territory. For a given level of volatility, Eq.(7) shows that $E_t[r_{t+1}] < \mu_t$ implies that the returns asymmetry is negative, i.e., $\rho_t > 0$.

The top-left panel of Figure 6 shows that the disconnection between μ_t and $E_t[r_{t+1}]$ is particularly evident for past `winners`, whereby the estimate of the location parameter is primarily positive, while the conditional expected returns are often largely negative, especially during momentum crashes as indicated by Daniel & Moskowitz (2016). On the other hand, the asymmetry seems more limited for the past `losers` sub-portfolio. The location and the conditional expected returns tend to align fairly close with the partial exception of the 2001 and 2008 crashes, where expected returns are lower than the conditional mode estimates. Put together, the right panels of Figure 6 suggest that the compounding effect of negative skewness in both past `losers` and past `winners` tend to exacerbate the asymmetry of the `WML` returns. The location parameter (red line) is largely positive, whereas the conditional expected returns (black line) is often in negative territory with large negative spikes around momentum crashes.

The fact that the conditional expected returns are lower than the location of the distribution, and that such discrepancy tend to exacerbate during momentum crashes suggest that there is indeed significant, time-varying, pro-cyclical negative asymmetry for both legs of the momentum strategy. The left panel of Figure (7) shows the estimates of the conditional skewness for the `WML` portfolio returns. The Figure reports the daily skewness estimates in black, and a five year smoothed trend in green. The dashed red horizontal line represents the sample mean of the skewness estimate. The conditional asymmetry of momentum returns shows substantial time variation, especially around recessions and momentum crashes. In particular, the conditional density of momentum returns shows mildly negative skewness ahead of crashes, in line with what implied by the relative upside and downside betas (see Figure 4), and it drops quite substantially during crashes. For instance, in the aftermath of August 1932, momentum skewness remains below-mean for about two decades before it starts increasing steadily. This is in line with the pattern observed for the cumulative performance of the strategy, where a full recovery from the loss incurred in 1932 only happened during the 1950s.[12]

The right panels shows the conditional skewness estimates of the past `winners` (top panel) and the past `losers` (bottom panel) sub-portfolios. The asymmetry of the returns is much more pronounced for past `winners` than past `losers`. For the former, the time-varying skewness is in deep negative territory essentially for the whole sample. Interestingly, past `losers` tend to have a significant upside exposure towards the tail of recessions, that is the skewness becomes positive at the latest stage of both the great depression, the dot-com bubble and the great financial crisis. This explains why the conditional skewness of the `WML` tend to become more negative during the tail of recessions and throughout momentum crashes: although the skewness of the past `winners` portfolios significantly decreased in the aftermath of the great depression, the conditional skewness of the returns on the past `losers` portfolios turned highly positive. In other words the `WML` portfolio implicitly is "buying" a moderately

[12] The average value of the dynamic skewness estimate is roughly -0.3 (red dotted line), which is smaller than the sample skewness of -1.2 (see Table 1). The discrepancy between the sample skewness estimate and the average skewness from our model is due to the robustness of the score-driven model to outliers. The updating mechanism based on Skt innovations allows the mitigate the effect of outliers for the parameters estimates, ensuring the estimated time-varying μ_t, σ_t, ρ_t to be robust to extreme realizations (Harvey & Luati 2014, Delle Monache, De Polis & Petrella 2021).

lower downside risk, but at the same time was "selling" a substantially higher upside risk, with the latter more than offsetting the former.

Figure 8 delve into the two major momentum crash periods of 1932-1939 (top panels) and 2001-2009 (bottom panels) as indicated by Daniel & Moskowitz (2016). The left panels shows the conditional skewness for the `WML` portfolio. To a large extent, consistent with Eq.(7) and the estimates in Figure (6), the asymmetry of the returns is most negative during the momentum crash of 1932-1939, and significantly drops from -0.1 to -0.4 towards the tail of the great financial crisis. The mid and right panels show that this is mostly due to the increasing risk on the upside for the past `losers` sub-portfolio, which represents the short leg of the momentum strategy. Yet, the dynamics of the conditional skewness during the 2009 momentum crash is slightly different. It is really the incremental risk on the downside of past `winners` that seems to weigh more for the negative skewness of the `WML` portfolio. Interestingly, both legs of the momentum strategy turns out to have negative skewness during the 2001 momentum crash, which result in a relatively mild risk on the downside for the `WML` strategy in the aftermath of the dot.com bubble burst. To a large extent, Figure 8 shows that dynamics of skewness during momentum crashes differ across periods. This is possibly due to differences in the marginal effect of volatility vs skewness in shaping the momentum risk premium.

As a matter of act, Eq.(7) shows that while the nature of the discrepancy between the conditional mode and expected returns is intuitively due to the asymmetric nature of the returns distribution, it is nevertheless also impacted by the dynamics of the scale parameters σ_t, as also shown analytically in Figure (5). Similar to expected returns in Eq.(7), we can derive the conditional variance of the momentum returns based on the moment generating function in Eq.(6) as,

$$Var_t\left[r_{t+1}\right] = \sigma_t^2\left(\frac{\nu}{\nu-2} + h(\nu)\rho_t^2\right), \qquad \nu > 2 \qquad \text{with} \qquad h(\nu) = \frac{3}{\nu-2} - g(\eta)^2, \tag{8}$$

with $h(\nu) \gg 0$ gauging the interaction between the fat-tailedness of the distribution, via ν and the asymmetry parameter ρ_t. Appendix C provides a complete derivation of how to obtain Eq.(8). Notice that for $\rho_t = 0$, Var reduces to the Student-t variance: $\sigma_t^2\frac{\nu}{\nu-2}$. As such, for $\rho_t^2 \in (0,1)$, the Student-t variance is increased by a factor proportional to the degree of asymmetry. Figure 9 shows the estimates of $Vol_t = \sqrt{Var_t\left[r_{t+1}\right]}$ for both the past `winners` (top-right) and `losers` (top-right) sub-portfolios. Returns on the past `losers` portfolio tend to be more volatile, in fact almost twice more volatile than the returns for the past `winners`, especially around momentum crashes. This is reflected in a highly time-varying volatility for the `WML` strategy (right panel), with volatility spikes around both during the 1929-1932 recessions, the dot-com bubble and the great financial crisis of 2008/2009 and subsequent momentum crashes. Such time variation in the strategy risk is consistent with the findings in Barroso & Santa-Clara (2015), Daniel & Moskowitz (2016) and suggest that not only the distribution might become more asymmetric around momentum crashes, but also that both tails of the distribution tend to be thicker in particular towards the end of recession periods.

4.1 Dissecting the momentum premium

Equation (7) states that the conditional expected return of a momentum strategy, $E_t[r_{t+1}]$ is a nonlinear function of the mode μ_t, the scale σ_t and the asymmetry ρ_t of the returns distribution. More specifically, to a first approximation, we can distinguish two sources of expected returns: the location parameter, μ_t, and a nonlinear function of higher-order moments: $hom_t = -g(\nu)\sigma_t\rho_t$. In this setting, positive values of ρ_t (that is, negative skewness) get amplified by a higher value of the scale parameter σ_t (and a constant term, function of the degrees of freedom, $g(\nu)$). We highlighted the interplay between scale and shape of the distribution on the model-implied expected returns in Figure 5.

To decompose the effect of time-varying volatility and skewness on the momentum risk premium, we perform a second-order expansion of Eq.(7) around the unconditional mean values $\overline{f} = (\overline{\mu}, \overline{\sigma}, \overline{\rho})'$. Let $\tilde{f}_t = f_t - \overline{f}$, the conditional expected returns on the momentum strategy can be decomposed as:

$$E_t[r_{t+1}] = \overline{\mu} - \overline{hom} + \tilde{\mu}_t + g(\nu)\overline{\rho}\tilde{\sigma}_t + g(\nu)\overline{\sigma}\tilde{\rho}_t + g(\nu)\tilde{\sigma}_t\tilde{\rho}_t$$

Figure (10) reports the results of this decomposition for the two main momentum crashes of 1932-1939 and 2001-2009 as indicated by Daniel & Moskowitz (2016). Different colored areas represent the five components of the decomposition. More specifically, in addition to the dynamics of momentum risk premium (black line), we report the effect of the changes in the location μ_t (green) , the scale σ_t (light pink), the asymmetry ρ_t (violet) of the conditional distribution, and the interplay between scale and asymmetry $\sigma_t \cdot \rho_t$ (purple). As we primarily focus on the dynamics of skewness risk during momentum crashes, for the ease of exposition we report the risk premium decomposition only for the two major momentum crash periods as indicated by Daniel & Moskowitz (2016).[13]

The left panel reports the momentum crash between 1932 and 1938. The results suggest that during the great depression the risk premium is not uniquely a compensation for high volatility, as suggested by Barroso & Santa-Clara (2015), Daniel & Moskowitz (2016). Outside recessions the effect of ρ_t becomes fairly relevant, in fact dominant from mid 1935 to essentially the mid of the 1938 recession. In other words, what emerges is that when volatility is above

[13] The results for the full the sample are available upon request.

average, the drag exerted by the *hom* component increases. In the Figure this is captured by the pink shaded area, representing the term, $\frac{\partial E_t[r_{t+1}]}{\partial \tilde{\sigma}_t}$, being negative. In the aftermath of the 1932 recession this relation steadily changes: the effect of returns skewness on expected returns increases, while the volatility term and the second-order term, $\frac{\partial E_t[r_{t+1}]}{\partial \tilde{\rho}_t}$, becomes more muted due to a significant decrease in volatility. This suggests that while volatility risk is particularly prominent *within recessions*, the returns skewness play a significant role for the dynamics of the strategy risk premium during the momentum crash after the great depression.

A slightly different pattern emerges for the dot-com bubble, as shown in the right panel of Figure (10). The negative effect of the returns variation on expected returns is rather clear. The partial derivative $\frac{\partial E_t[r_{t+1}]}{\partial \tilde{\sigma}_t}$ is largely negative from 2001 to late 2002. However, the interplay with the asymmetry of the returns, captured by the pink shaded area, representing the term, $\frac{\partial E_t[r_{t+1}]}{\partial \tilde{\sigma}_t \partial \tilde{\rho}_t}$, is highly positive. This suggests that despite the negative effect of a higher-than-average volatility on expected returns in the aftermath of the dot-com bubble, such negative effect is offset by a large and positive effect due to a lower than average returns asymmetry (see also left panels in Figure 8). In other words the relatively low skewness in the returns offset the effect of a large spike in volatility. Interestingly, outside the two momentum crashes of 2001 and 2009, neither volatility nor skewness play any significant role for the dynamics of the momentum risk premium. As far as the great financial crisis of 2008/2009 is concerned, the risk premium decomposition suggests that expected returns primarily represents a compensation for an outlying spike in volatility throughout the recession period. More specifically, despite the returns asymmetry is relatively lower than average during the post-2008 momentum crash, the effect of volatility is so large than the conditional expected returns are essentially dominated by the second moment. This is in line with the conditional skewness and volatility estimates as shown in Figure 7-9.

Revisiting the momentum risk-return trade-off. Figures 10 suggests that the role of volatility for the dynamics of momentum risk premium should be red in conjunction with the returns asymmetry. For instance, when the strategy asymmetry is close to zero, as in the tail of the dot-com bubble, a large volatility does not translate in a negative risk premium but its effect is offset by a smaller-than-average returns skewness. In addition, the results shows that the compensation for skewness risk changes during momentum crashes. That is, while a lower than average skewness drag down returns during the post-1932 momentum crash, the opposite holds for the momentum crashes of 2002 and 2009. This suggests that returns asymmetry plays a key role, both on itself and when interacted with the returns dispersion around the location, i.e., volatility. This is consistent with some of the early literature on the risk-return trade-off; if returns asymmetry varies over time, volatility is not an accurate representation of the strategy's risk (see, e.g., Glosten et al. 1993).

The results from Figure 10 suggest that there is possibly a negative risk-return trade-off, especially during momentum crashes. However, such trade-off does not uniquely depend on the returns volatility, but could be linked to the dynamics of the returns asymmetry as well. By rearranging the terms in Eq.(7) using the conditional variance in Eq.(8), we can define the conditional expected returns:

$$E_t[r_{t+1}] = \mu_t + \lambda(\rho_t)\sqrt{V_t(r_{t+1})}, \tag{9}$$

where $\lambda(\rho_t)$ captures the time-varying risk-return trade-off as a nonlinear function of the time-varying asymmetry of the returns, and is defined as:

$$\lambda(\rho_t) = -\frac{g(\nu)}{\sqrt{\frac{\nu}{\nu-2} + h(\nu)\rho_t^2}}\rho_t \tag{10}$$

As a result, the shape of the risk-return tradeoff directly depends uniquely on the asymmetry of the conditional distribution of the returns as captured by the shape parameter ρ_t. More specifically, the higher the level of skewness, that is the more positive is the value of ρ_t, the more negative is the slope parameter $\lambda\left(\rho_t\right)$ conditional on the degrees of freedom ν.

The left panel in Figure (11) compares the theoretical vs the actual shape of the $\lambda(\rho_t)$. Clearly, the more negative the returns asymmetry, meaning the skewness, the more negative becomes the risk-return trade-off. In addition, the results highlight that such negative relation is non-linear, and is slightly concave in the ρ_t parameter. Specifically, this points to a nonlinear interaction between volatility and skewness risk, in contrast to what argued by Theodossiou & Savva (2016). During periods of increased downside risk our model reduces the probability to realised positive returns, compared to lower returns, thus delivering a negative risk-return trade-off. The right panel of Figure (11) reports the value of $\lambda(\rho_t)$ over time. The blue line shows the actual value of the slope parameter, whereas the dashed horizontal line reports the risk-return trade-off as a function of the unconditional mean $\overline{\rho}$. The evidence shows that the risk-return trade-off is negative, on average. This contributes to an ongoing debate on the sign of the trade-off between risks and rewards for momentum strategies is often unclear. Theodossiou & Savva (2016) argue that such inconclusiveness is due to the fact that returns embed a *linear* combination of skewness and volatility premiums. When the skewness of the predictive distribution is negative, these two premiums offset each other, generating a highly uncertain dynamics for the risk-return trade-off. Our approach allows for a more general, *non-linear*, *time-varying*, interaction between skewness and volatility premiums. Yet, we clearly document a negative risk-return trade-off on average across the sample period.

Yet, the effect of the conditional variance on the expected returns is not constant over time, but is highly volatile and tend to become more negative particularly around momentum crashes. Larger negative skewness triggers a downward correction of the expected returns on the momentum strategy, whereas we find a less pervasive impact of skewness on the variance of the returns. The limited effect of the skewness on volatility is likely mitigated by the presence of heavier tails in the returns distribution, which are explicitly accounted for by the model. Nevertheless, the conditional skewness still affect conditional volatility of the returns, and therefore the risk-return trade-off in relative terms.

In order to disentangle the role of the slope parameter $\lambda(\rho_t)$ vis-á-vis the conditional variance $Var_t[r_{t+1}]$ for the dynamics of expected returns, we look at the correlation between $E_t[r_{t+1}]$ and each of the two components taken singularly. Figure (12) shows the scatter plots of the expected return and the conditional volatility, the slope and the interaction $\lambda(\rho_t)\sqrt{Var_t[r_{t+1}]}$. The red and blue crosses highlight observations form the 1932 and 2009 crash periods, respectively. For low levels of conditional volatility, expected returns cluster around zero, with higher dispersion towards positive figures, and the least square fit (gray line) points to a mildly negative correlation. The mild negative correlation between the conditional volatility and expected returns is mainly due to higher scale values associated to the crashes. By looking at the trade-off between risk premium and $\sqrt{Var_t[r_{t+1}]}$ during momentum crashes, we see that the negative correlation is much more pronounced, with all outlying values of the scale clustered during these two periods. These results indicate a strong performance of the strategy for moderate levels of volatility risk, suggestive of a negative risk-return trade-off, in line with what documented by Barroso & Santa-Clara (2015) and Barroso & Maio (2019). The correlation between expected return and the slope parameter $\lambda(\rho_t)$ displays instead a mildly positive correlation: higher values of ρ_t – which imply higher negative skewness and lower $\lambda(\rho_t)$ – depress expected returns. Notably, the relationship between expected returns and the slope on conditional volatility is rather non-linear during momentum crashes.

Interestingly, when the slope and the conditional volatility are interacted, there is a close to perfect fit to the conditional expected returns. The right panel of Figure (12) highlights how this translate into a distinctively negative correlation with expected returns. The interaction of the two parameters results in a strongly negative effect on $E_t[r_{t+1}]$: low values of conditional volatility and ρ_t are associate to close-to-zero asymmetry. When both the scale increase and the shape of the distribution becomes more negative, the amplification mechanism leads to a larger negative effect on expected returns than that exerted by the sole asymmetry. This result suggests that the high risk premium associated to the momentum strategy is not only related to time-varying volatility, but it is also significantly affected by the time variation of the third moment.

5 A skewness-managed momentum strategy

Few papers have attempted to improve the profitability of momentum investing. Grundy & Martin (2001) propose to hedge the exposure to time-varying volatility to attenuate momentum crashes. However, as noted by Daniel & Moskowitz (2016), this approach is not implementable in real-time, and suffers from a potentially relevant forward-looking bias. Other strategies attempt to time the volatility associated with momentum returns (see Barroso & Santa-Clara 2015 and Moreira & Muir 2017). Specifically, Barroso & Santa-Clara (2015) propose to target the annualized unconditional volatility of the market, by estimating the realised volatility of the momentum strategy on a rolling window basis. Within this setting, during periods of high volatility the strategy unwind the investment in the `WML` portfolio, while during periods of low (compared to the market's) volatility the strategy increases its exposure to the momentum strategy. This should increase the risk-adjusted returns on the momentum strategy. The main drawback is that the trade-off between risk and return is assumed constant over time, that is, expected returns move in lock-step with volatility.

Figures 11-12, however, show that the assumption of a constant trade-off between risk and rewards of a momentum strategy does not hold in practice within the context of a non-Normal modeling framework. Based on a similar consideration, Daniel & Moskowitz (2016) put forward a dynamic strategy aimed at maximising the conditional Sharpe ratio of the strategy, with weights given by:

$$w_t = \frac{1}{2\gamma}\frac{E_t[r_{t+1}]}{Var_t[r_{t+1}]} \tag{11}$$

where γ is a constant controlling the unconditional volatility of the strategy. The conditional expectation $E_t[r_{t+1}]$ is assumed positive and constant during bull markets, while it varies during bear market periods. The conditional volatility $Var_t[r_{t+1}]$ is estimated as a weighted average of asymmetric GARCH volatility (Glosten et al. 1993), and 6-month realized volatility. [14]

We build upon Daniel & Moskowitz (2016) and leverage on the dynamics of the conditional expected returns from our modeling framework as per Equation (7). More specifically, we replace the conditional mean $E_t[r_{t+1}]$ of the momentum strategy returns and decompose the optimal Sharpe ratio weights in two components,

$$w_t = \frac{1}{2\gamma}\frac{E_t[r_{t+1}]}{Var_t[r_{t+1}]} = \frac{1}{2\gamma}\frac{\mu_t - g(\nu)\rho_t\sigma_t}{Var_t[r_{t+1}]} = \underbrace{\frac{1}{2\gamma}\frac{\mu_t}{Var_t[r_{t+1}]}}_{w_{1,t}} - \underbrace{\frac{1}{2\gamma}\frac{g(\nu)\sigma_t\rho_t}{Var_t[r_{t+1}]}}_{w_{2,t}}. \tag{12}$$

[14] Refer to Appendix C in Daniel & Moskowitz (2016) for additional details.

where $w_{1,t}$ would be the equivalent of Daniel & Moskowitz (2016) under the assumption of symmetric returns, i.e., $\rho_t = 0$ (see Eq.7-8), and $w_{2,t}$ capturing the variations in the time-varying asymmetry.[15] For a given level of conditional variance, $w_{2,t}$ becomes more negative as the shape parameter ρ_t becomes more positive (more negative skewness). Therefore, during periods of high negative skewness our strategy unwind the investment in the `WML` portfolio more than a dynamic volatility targeting adjustment, while during periods of low or zero returns asymmetry the strategy tend to coincide with a standard volatility-managed portfolio (see, e.g., Moreira & Muir 2017). For this reason $w_{2,t}$ represents a skewness hedging component within the context of the maximum Sharpe ratio strategy. We argue that, under time-varying returns asymmetry, this should out-perform the adjustments proposed by Barroso & Santa-Clara (2015), Daniel & Moskowitz (2016). Albeit successful in improving upon a standard `WML` portfolio, their approaches ignore the effect of returns asymmetry to gauge the optimal investment in a momentum strategy vis-á-vis a riskless asset.

One comment is in order. A growing body of literature is concerned with the extension of the well-known mean-variance framework towards considering the skewness dynamics (see, e.g., Mencía & Sentana 2009). This is typically implemented by including higher order moments in the agents' utility maximisation and asset allocation problem. However, our modeling framework allows to elicit both the conditional mean and the conditional variance of the momentum returns as a direct function of both the location, the scale and the asymmetry parameters (see Eq.7 and Eq.8). The advantage is twofold: first, we can still gauge the benefit of explicitly modeling time-varying returns asymmetry within the context of the benchmark maximum Sharpe ratio approach as in Barroso & Santa-Clara (2015) and Daniel & Moskowitz (2016). In fact, compared to a model where $\rho = 0$, our specification captures returns' underlying asymmetry via the first two moments. for instance, Eq.(7) shows that a higher ρ_t estimate, meaning a more negative skewness, implies a lower conditional mean for a given value of the scale σ_t. Similarly, a more negative skewness implies a higher volatility of the returns as per Eq.(8). Second, we can construct a portfolio allocation based on real-time predictions of the conditional location, scale and asymmetry of the returns distribution, which defines both $E_t[r_{t+1}]$ and $Var_t[r_{t+1}]$ in Eq.(11).

Figure 13 reports the decomposition of the maximum Sharpe Ratio targeting weights for the two major momentum crashes periods as indicated by Daniel & Moskowitz (2016). The left panel covers the momentum crashes of 1932 and 1940. The right panel reports the momentum crashes of 2001 and 2009. The red dashed line represent the unconditional mean. NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods. The component associated with the central tendency, $w_{1,t}$ is

[15] Recall that the asymmetry parameter has a sizable effect on the mean of the returns, whereas it has little impact on their variance.

reported in blue, while the skewness hedging component $-w_{2,t}$, is in red. The total weight, w_t, is reported in black. Few interesting facts emerge; first, except few nuances, the tendency component $w_{1,t}$ tend to dominate the skewness hedging component, that is $w_t > 0$ for most of the momentum crashes. Second, however the aggregate exposure to momentum is lower than would be without considering the returns asymmetry, meaning $-w_{2,t} < 0$ across both momentum crashes. In other words, the skewness hedging component is non-zero and tends to be larger during the 1932-1939 crash period, in fact twice as large as during the dot-com and 2009 crashes. Third, the investment in the momentum strategy is significantly lower during the crashes of 2001 and 2009 than in the period in between crashes. This suggests that the combination of lower volatility and lower skewness increases the leverage on the `WML` portfolio.

In a set of unreported results, we calculate the Sharpe ratio, the Sortino ratio, the maximum drawdown and the returns skewness of both the total portfolio adjustment w_t, as well as the two components $w_{1,t}$ and $w_{2,t}$ separately. The Sharpe (Sortino) from the total portfolio w_t is 3.57 (7.50) annualised compared to the 2.93 (5.24) annualised from the tendency component $w_{1,t}$ alone. The maximum drawdown of the combined $w_{1,t}$ and $w_{2,t}$ components is almost a third than the tendency component alone. Finally, the aggregate risk managed `WML` shows a positive skewness which is more than double of the strategy without skewness hedging adjustment (0.286 vs 0.135).

5.1 Out-of-sample managed portfolio returns

We compare our skewness managed strategy, the maximum skewed-Sharpe ratio with skewness (`mSSR`), against the approach of Daniel & Moskowitz (2016, `mSR`), and two versions of the strategy in Barroso & Santa-Clara (2015): the constant volatility (`cVol`), and the equivalent which accounts for skewness in variance (`cSVol`). The strategies that overlook skewness are computed by re-estimating the models constraining $\rho = 0$. For each strategy implementation the model's structural parameters are re-estimated each month, that is the time-varying parameters μ_t, σ_t and ρ_t are extracted on a daily basis within a given month, assuming that the structural parameters of the score-driven transition (see Section 3) remain constant. This is to limit the computational cost of the real-time estimation procedure. The first forecast and portfolio choice is generated in January 1st 1930, that is we use three years of daily returns as initial burn-in sample for the recursive forecasts.

We evaluate the portfolios by means of three distinct performance measures. The first is the Sharpe ratio, $SR_i = \frac{\mathbb{E}[\tilde{R}_i]}{Vol(\tilde{R}_i)}$, which measures the reward of the investment once volatility has been accounted for. We evaluate improvements in risk-adjusted returns with respect to the standard `WML` strategy based on the appraisal ratios and we test for the significance of these improvements following Ledoit & Wolf (2008). Despite its widespread use, two common pitfalls of the Sharpe ratio are that it fares badly with dynamic strategies (Marquering & Verbeek 2004), and it overlooks non-Gaussian features of the investment. Therefore, we compare the portfolios by means of a second measure: the Sortino ratio (see, e.g., Sortino & Van Der Meer 1991). The Sortino ratio entails a penalization only for those returns falling below a certain threshold. The threshold, commonly referred to as *minimum accepted return* (MAR), is generally set at the risk free rate. The denominator of this ratio is $Vol(\tilde{R}_i|\tilde{R}_i < 0)$, with $\tilde{R}_i$ being the excess return of strategy i. This accounts for the fact that investors typically dislike negative returns (Kraus & Litzenberger 1976), and weight losses more than gains (Kahneman & Tversky 2013). A third measure we calculate to compare the different strategies performance is the Bowley skewness calculated as $QS_\alpha = \frac{q(\alpha)+q(1-\alpha)-2q(50)}{q(\alpha)-q(1-\alpha)}$, with $q(\alpha)$ being the αth quantile of the data. We choose a value of $\alpha = 5$, which corresponds to a 5% cut-off for the returns distribution.

Table (2) reports the results. The results support the idea that by explicitly considering the time-varying asymmetry of the momentum returns distribution provide some significant economic gain. The Sharpe ratio of the skew-managed strategy, `mSSR`, is 3.5% in annualized term, about four times higher than ratio delivered by the plain `WML` strategy; compared to `mSR` approach, by explicitly considering the dynamics of skewness for the expected returns and volatility delivers a gain of about 20% in terms of risk-adjusted annualized returns, and an appraisal ratio of 0.8. Benchmark volatility-managed strategies, `cSVol` and `cVol`, improve upon the plain `WML` investing, delivering appraisal ratios around 1, but remain less profitable compared to the skew-managed strategies. Notice these strategies are not able to pick up the effect of time-varying skewness, as this has a smaller impact on the second moment compared to its effect on expected returns.

Sortino ratios suggest even bigger gains associated with our skew-managed strategy: risk-adjusted returns are not only higher, but they also are less exposed to crashes and therefore to downside risk. That is, Sortino measures are able to provide a measure of risk-adjusted returns net of the "bad" volatility, while treating "good' volatility as beneficial for maximising returns. A higher Sortino ratio for our skew-managed momentum strategy implies that the effect of bad volatility is mitigated by the explicit effect of time-varying skewness. Quantile skewness values confirm that $mSSR$ portfolio return are characterized by positive unconditional skewness of 0.28, suggesting that the upside potential of the portfolio is about 1.75 times its downside risk.

5.2 Transaction costs and risk aversion

Figure 13 shows that the dynamics of the portfolio weights could take relatively large values, that is the leverage on the `WML` portfolio could be substantial. This is in line with the Daniel & Moskowitz (2016) estimates. Notice that, our goal is not to propose an actual trading strategy that can be implemented off the shelf, but rather to show the dynamic of the returns asymmetry and its economic value within the context of a benchmarj, pseudo-real, portfolio allocation framework. Neverthelss, given the substantial volatility in the risk-managed weights, we now assess the robustness of the results by evaluating the portfolio performances net of transaction costs and performance fees. We implement three difference exercises: first, let $\mathcal{R}_{i,t} = 1 + R_{i,t}$ denote the gross returns at time t for strategy i, $i =$ `mSSR`, `mSR`, `cSvol`, `cVol`, `WML`. Rebalancing the portfolios each day requires to adjust the position in momentum returns by a costly trading of magnitude $|w_{i,t+1} - w_{i,t}|$ and cost c. Following DeMiguel, Garlappi & Uppal (2009), we define the evolution of wealth for strategy i as:

$$W_{i,t+1} = W_{i,t}\mathcal{R}_{i,t}\left(1 - c|w_{i,t+1} - w_{i,t}|\right), \tag{13}$$

such that returns net of transaction fees can be computed as $\overline{r} - c = \frac{W_{i,t+1}}{W_{i,t}} - 1$.

As a second measure of net economic value we follow Fleming, Kirby & Ostdiek (2003) and evaluate the maximum performance fee an investor with mean-variance utility function would be willing to pay to access the investment signal from each model. For any pair (i, j) of strategies, the fee $\mathcal{F}$ arises as the solution to:

$$\sum_{t=0}^{T-1}\left((\mathcal{R}_{i,t} - \mathcal{F}) - \frac{\delta}{2(1+\delta)}(\mathcal{R}_{i,t} - \mathcal{F})^2\right) = \sum_{t=0}^{T-1}\left(\mathcal{R}_{j,t} - \frac{\delta}{2(1+\delta)}\mathcal{R}_{j,t}^2\right), \tag{14}$$

where δ is the degree of relative-risk aversion. Finally, as a third measure to further understand the out-performance of our strategy, we also consider Modigliani & Modigliani (1997) abnormal return measure, $\mathcal{A}$. For any pair of strategies (i, j), we leverage up or down strategy i so as to match the risk profile of strategy j, and we evaluate the annualized, risk-adjusted abnormal returns:

$$\mathcal{A}_{i,j} = Vol_i(SR_i - SR_j). \tag{15}$$

Table 3 reports the results. We compute performance fees and abnormal returns with respect to the unmanaged `WML` strategy. We consider several level of transaction costs, ranging from 0 to 50 bps; for the performance fees we set $\delta = 5$. Overall, relative to the simple `WML` strategy, a skew-managed portfolio realizes higher annualized average net of any level of transaction fees. In addition, such strategy commands higher performance fees and realizes larger abnormal returns. The relative ranking of each strategy remains unaltered regardless the level of transaction costs considered.

We repeat the economic evaluation reported above controlling for different levels of risk aversion. Table 4 report the performance fees $\mathcal{F}$ for risk aversion levels of 2, 7 and 15. These levels compare agents with a strong risk aversion to investors prone to take on more risks. Overall, the main results largely hold: considering time-varying skewness when maximising the Sharpe ratio delivers the highest performance fees across different levels of risk aversion. These results suggest hedging for predictable variations in the returns skewness is economically meaningful, regardless the level of risk aversion.

5.3 Sub-sample performance

As a check on the robustness of our results, we perform the same economic performance analysis over a set of approximately quarter-century sub-samples: 1927 to 1949, 1950 to 1974, 1975 to 1999, and 2000 to 2013. We use the same conditional mean and variance equation in each of the four subsamples. Panels (a)–(d) of Figure 14 plot the cumulative log returns by subsample. For ease of comparison, returns for each of the strategies are scaled to an annualized volatility of 19% in each subsample. In each of the four subsamples, the ordering of performance remains the same. The skew-managed strategy outperforms both the constant and dynamic volatility managed-portfolios as proposed by Barroso & Santa-Clara (2015) and Daniel & Moskowitz (2016), which outperform the static `WML` strategy. As the sub-sample plots show, part of the improved performance of the skew-managed strategy over the static and dynamic volatility-adjusted strategies is the amelioration of big crashes. But, even over sub-periods devoid of those crashes, there is still improvement.

5.4 Spanning tests

A more formal test of the skew-managed portfolio's success is to conduct a series of spanning tests with respect to the other competing momentum strategies and risk factors. Using daily returns, we regress the risk-managed portfolio returns on a host of factors that include the market, the Fama & French (1993) three factors, the static `WML`, the dynamic volatility `mSR` as in Daniel & Moskowitz (2016), and the constant volatility (`cVol`) momentum strategies as in Barroso & Santa-Clara (2015). The annualized alphas and corresponding t-stats (in parenthesis) from these regressions are reported in Table 5.

The first column reports the regression results for all strategies on the market and the static `WML` portfolio. The alphas are highly significant across all different strategies, with the highest alpha at 76% (t-statistic = 33.2) annualised, indicating that the performance of the mSSR managed portfolio is not captured by the market and the static momentum returns. The second column adds the size and value Fama & French (1993) factors in addition to the market. Again, all of the alphas are highly significant, with the mSSR managed portfolio showing the largest risk-adjusted annualised performance (alpha = 75%, t-statistic = 33.15). Hence, our dynamic skew-managed momentum strategy's abnormal performance is not being driven by dynamic exposure to these other factors or to the static momentum portfolio. Columns 3 and 4 replicate the same analysis by replacing the static WML strategy with the constant volatility adjustment as in Barroso & Santa-Clara (2015) (cVol). The alphas significantly drop in magnitude, but for the mSSR remain as high as 50.6% annualised (t-statistic = 28.8), suggesting that the dynamic skew-managed momentum portfolio is not spanned by the constant volatility portfolio. The last two columns replicate the same analysis by replacing the static WML strategy with the dynamic volatility adjustment as in Daniel & Moskowitz (2016) (mSR). Again, the results show that the returns on the skew-managed portfolio strategy cannot be explained by a dynamic volatility targeting strategy. This holds independently of using the market or the Fama & French (1993) factors to calculate the risk-adjusted returns. Interestingly, the alphas on the constant volatility targeting is negative and significant, which suggests dynamic volatility targeting generates a higher performance. This results is consistent with Daniel & Moskowitz (2016).

6 Conclusions

We propose a flexible parametric model to capture the dynamics of the conditional distribution of the momentum returns over time. Our approach is to use both the location, scale and asymmetry parameters of the returns distribution to quantify the interplay between volatility and skewness in shaping the dynamics of the risk premium of the momentum strategy. As a result, our model provides a dynamic characterization of the effect of volatility and skewness on the risk-return trade-off of an otherwise standard momentum strategy. We see both the volatility and the skewness as key ingredients to understand the risk premium dynamics in US equity momentum strategies.

Empirically, we show that the conditional skewness of the momentum returns is time varying and tend to become more negative during momentum crashes. This is particularly relevant for the momentum crash period in the aftermath of the great depression. In addition, by comparing a skew-managed maximum Sharpe ratio portfolio strategy with both a constant and dynamic volatility targeting strategy, our results highlight the economic relevance of explicitly considering the time variation in the returns skewness to capture the risk premium dynamics and make more informed risk managing within the context of momentum strategies.

References

Ang, A., Chen, J. & Xing, Y. (2006), 'Downside risk', *The review of financial studies* **19**(4), 1191–1239.

Arellano-Valle, R. B., Gómez, H. W. & Quintana, F. A. (2005), 'Statistical inference for a general class of asymmetric distributions', *Journal of Statistical Planning and Inference* **128**(2), 427–443.

Arnold, B. C. & Groeneveld, R. A. (1995), 'Measuring skewness with respect to the mode', *The American Statistician* **49**(1), 34–38.

Asness, C. S., Moskowitz, T. J. & Pedersen, L. H. (2013), 'Value and momentum everywhere', *The Journal of Finance* **68**(3), 929–985.

Bai, J. & Ng, S. (2005), 'Tests for skewness, kurtosis, and normality for time series data', *Journal of Business & Economic Statistics* **23**(1), 49–60.

Bali, T. G. & Engle, R. F. (2010), 'The intertemporal capital asset pricing model with dynamic conditional correlations', *Journal of Monetary Economics* **57**(4), 377–390.

Barroso, P. & Maio, P. F. (2019), 'The risk-return tradeoff among equity factors', *Available at SSRN 2909085* .

Barroso, P. & Santa-Clara, P. (2015), 'Momentum has its moments', *Journal of Financial Economics* **116**(1), 111–120.

Belloni, A. & Chernozhukov, V. (2009), 'On the computational complexity of mcmc-based estimators in large samples', *The Annals of Statistics* **37**(4), 2011–2055.

Blasques, F., Koopman, S. J. & Lucas, A. (2014), 'Stationarity and ergodicity of univariate generalized autoregressive score processes', *Electronic Journal of Statistics* **8**(1), 1088–1112.

Blasques, F., Koopman, S. J. & Lucas, A. (2015), 'Information-theoretic optimality of observation-driven time series models for continuous responses', *Biometrika* **102**(2), 325–343.

Bollerslev, T. et al. (1987), 'A conditionally heteroskedastic time series model for speculative prices and rates of return', *Review of Economics and Statistics* **69**(3), 542–547.

Chernozhukov, V. & Hong, H. (2003), 'An mcmc approach to classical estimation', *Journal of Econometrics* **115**(2), 293–346.

Cox, D. R. (1981), 'Statistical analysis of time series: Some recent developments', *Scandinavian Journal of Statistics* pp. 93–115.

Creal, D., Koopman, S. J. & Lucas, A. (2013), 'Generalized autoregressive score models with applications', *Journal of Applied Econometrics* **28**(5), 777–795.

Daniel, K. & Moskowitz, T. J. (2016), 'Momentum crashes', *Journal of Financial Economics* **122**(2), 221–247.

Delle Monache, D., De Polis, A. & Petrella, I. (2021), 'Modeling and forecasting macroeconomic downside risk', *Bank of Italy Temi di Discussione (Working Paper) No* **1324**.

Delle Monache, D. & Petrella, I. (2017), 'Adaptive models and heavy tails with an application to inflation forecasting', *International Journal of Forecasting* **33**(2), 482–501.

DeMiguel, V., Garlappi, L. & Uppal, R. (2009), 'Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?', *The review of Financial studies* **22**(5), 1915–1953.

Dittmar, R. F. (2002), 'Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns', *The Journal of Finance* **57**(1), 369–403.

Dobrynskaya, V. (2015), 'Upside and downside risks in momentum returns', *Higher School of Economics Research Paper No. WP BRP* **50**.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models', *Journal of Business & Economic Statistics* **20**(3), 339–350.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation', *Econometrica: Journal of the Econometric Society* pp. 987–1007.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of financial economics* **33**(1), 3–56.

Fama, E. F. & French, K. R. (1996), 'Multifactor explanations of asset pricing anomalies', *The Journal of Finance* **51**(1), 55–84.

Fama, E. F. & French, K. R. (2012), 'Size, value, and momentum in international stock returns', *Journal of financial economics* **105**(3), 457–472.

Fernández, C. & Steel, M. F. (1998), 'On bayesian modeling of fat tails and skewness', *Journal of the American Statistical Association* **93**(441), 359–371.

Fleming, J., Kirby, C. & Ostdiek, B. (2003), 'The economic value of volatility timing using "realized" volatility', *Journal of Financial Economics* **67**(3), 473–509.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The journal of finance* **48**(5), 1779–1801.

Gómez, H. W., Torres, F. J. & Bolfarine, H. (2007), 'Large-sample inference for the epsilon-skew-t distribution', *Communications in Statistics—Theory and Methods* **36**(1), 73–81.

Griffin, J. M., Ji, X. & Martin, J. S. (2003), 'Momentum investing and business cycle risk: Evidence from pole to pole', *The Journal of finance* **58**(6), 2515–2547.

Grinblatt, M., Titman, S. & Wermers, R. (1995), 'Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior', *The American economic review* pp. 1088–1105.

Grundy, B. D. & Martin, J. S. M. (2001), 'Understanding the nature of the risks and the source of the rewards to momentum investing', *The Review of Financial Studies* **14**(1), 29–78.

Gu, S., Kelly, B. & Xiu, D. (2020), 'Empirical asset pricing via machine learning', *The Review of Financial Studies* **33**(5), 2223–2273.

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

Hansen, B. E. (1994), 'Autoregressive conditional density estimation', *International Economic Review* pp. 705–730.

Harvey, A. C. (2013), *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, Vol. 52, Cambridge University Press.

Harvey, A. & Luati, A. (2014), 'Filtering with heavy tails', *Journal of the American Statistical Association* **109**(507), 1112–1122.

Harvey, C. R. & Siddique, A. (1999), 'Autoregressive conditional skewness', *Journal of Financial and Quantitative Analysis* **34**(4), 465–487.

Harvey, C. R. & Siddique, A. (2000), 'Conditional skewness in asset pricing tests', *The Journal of Finance* **55**(3), 1263–1295.

Hogan, W. W. & Warren, J. M. (1974), 'Toward the development of an equilibrium capital-market model based on semivariance', *Journal of Financial and Quantitative Analysis* pp. 1–11.

Jacobs, H., Regele, T. & Weber, M. (2015), 'Expected skewness and momentum'.

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

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

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995), *Continuous univariate distributions, volume 2*, Vol. 289, John wiley & sons.

Juárez, M. A. & Steel, M. F. (2010), 'Model-based clustering of non-gaussian panel data based on skew-t distributions', *Journal of Business & Economic Statistics* **28**(1), 52–66.

Kahneman, D. & Tversky, A. (2013), Prospect theory: An analysis of decision under risk, *in* 'Handbook of the fundamentals of financial decision making: Part I', World Scientific, pp. 99–127.

Kenney, J. F. & Keeping, E. S. (1939), *Mayhematics of Statistics*, D. van Nostrand.

Kim, T.-H. & White, H. (2004), 'On more robust estimation of skewness and kurtosis', *Finance Research Letters* **1**(1), 56–73.

Koopman, S. J., Lucas, A. & Scharth, M. (2016), 'Predicting time-varying parameters with parameter-driven and observation-driven models', *Review of Economics and Statistics* **98**(1), 97–110.

Kothari, S. P. & Shanken, J. (1992), 'Stock return variation and expected dividends: A time-series and cross-sectional analysis', *Journal of Financial Economics* **31**(2), 177–210.

Kraus, A. & Litzenberger, R. H. (1976), 'Skewness preference and the valuation of risk assets', *The Journal of finance* **31**(4), 1085–1100.

Ledoit, O. & Wolf, M. (2008), ‘Robust performance hypothesis testing with the sharpe ratio’, *Journal of Empirical Finance* **15**(5), 850–859.

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

Lettau, M., Maggiori, M. & Weber, M. (2014), ‘Conditional risk premia in currency markets and other asset classes’, *Journal of Financial Economics* **114**(2), 197–225.

Lucas, A. & Zhang, X. (2016), ‘Score-driven exponentially weighted moving averages and value-at-risk forecasting’, *International Journal of Forecasting* **32**(2), 293–302.

Marquering, W. & Verbeek, M. (2004), ‘The economic value of predicting stock index returns and volatility’, *Journal of Financial and Quantitative Analysis* pp. 407–429.

Mencía, J. & Sentana, E. (2009), ‘Multivariate location–scale mixtures of normals and mean–variance–skewness portfolio allocation’, *Journal of Econometrics* **153**(2), 105–121.

Modigliani, F. & Modigliani, L. (1997), ‘Risk-adjusted performance’, *Journal of portfolio management* **23**(2), 45–54.

Moreira, A. & Muir, T. (2017), ‘Volatility-managed portfolios’, *The Journal of Finance* **72**(4), 1611–1644.

Moskowitz, T. J. & Grinblatt, M. (1999), ‘Do industries explain momentum?’, *The Journal of Finance* **54**(4), 1249–1290.

Moskowitz, T. J., Ooi, Y. H. & Pedersen, L. H. (2012), ‘Time series momentum’, *Journal of Financial Economics* **104**(2), 228–250.

Novy-Marx, R. (2012), ‘Is momentum really momentum?’, *Journal of Financial Economics* **103**(3), 429–453.

Psarakis, S. & Panaretoes, J. (1990), ‘The folded t distribution’, *Communications in Statistics-Theory and Methods* **19**(7), 2717–2734.

Rouwenhorst, K. G. (1998), ‘International momentum strategies’, *The journal of finance* **53**(1), 267–284.

Sortino, F. A. & Van Der Meer, R. (1991), ‘Downside risk’, *Journal of portfolio Management* **17**(4), 27.

Theodossiou, P. & Savva, C. S. (2016), ‘Skewness and the relation between risk and return’, *Management Science* **62**(6), 1598–1609.

Tian, L., Liu, J. S. & Wei, L. J. (2007), ‘Implementation of estimating function-based inference procedures with markov chain monte carlo samplers’, *Journal of the American Statistical Association* **102**(479), 881–900.

Tsai, H.-J., Chen, M.-C. & Yang, C.-Y. (2014), ‘A time-varying perspective on the capm and downside betas’, *International Review of Economics & Finance* **29**, 440–454.

Wallis, K. F. (2014), ‘The two-piece normal, binormal, or double gaussian distribution: its origin and rediscoveries’, *Statistical Science* pp. 106–112.

Table 1: **Skewness in US equity momentum**

This table reports different measures of skewness for both past `winner` and `losers` portfolios, as well as the `WML` momentum strategy and the aggregate market portfolio. The first row reports the sample skewness, with p-values for the Bai & Ng (2005) test in parentheses; the remaining rows report the quantile skewness (QS_α), computed as $\frac{q(\alpha)+q(1-\alpha)-2q(50)}{q(\alpha)-q(1-\alpha)}$, with $\alpha = 95, 90$. The sample period is from July 1st 1926 to September 30th 2020, daily.

	`losers`	`winners`	`WML`	Market
Skewness	0.140	-0.682	-1.230	-0.476
	(0.264)	(0.022)	(0.001)	(0.059)
QS_{95}	-0.035	-0.075	-0.038	-0.063
QS_{90}	-0.038	-0.085	-0.062	-0.079

Table 2: **Risk-adjusted momentum performance**

The table reports the ranking of six different managed portfolios: the mean-variance-skewness (`mSSR`), the mean-variance (`mSR`), constant volatility with skewness (`cSVol`), constant volatility without skewness (`cVol`) and the `WML` strategy. The measure we report are the Sharpe ratio, that is the ratio of annualized excess mean return over the annualized excess return variance, the Appraisal ratio, computes as the square root of the difference between two consecutive Sharpe ratios, the Sortino ratio, which uses the annualized excess return downside variance as denominator, and the quantile skewness at the 5% cut-off. We report in parentheses the bootstrapped p-values for the differences in consecutive Sharpe ratios as in Ledoit & Wolf (2008). The sample period is from July 2nd 1929 to September 30th 2020, daily. Portfolio weights are generated in real-time by recursive forecasts of the conditional mean and variance of the returns based on the model parameters.

	Sharpe	Appraisal	Sortino	MaxDD	Skew
`mSSR`	3.574	1.643	7.501	0.183	0.286
	(0.000)				
`mSR`	2.942	1.438	5.272	0.291	0.168
	(0.000)				
`cSVol`	1.979	1.050	3.000	0.444	−0.046
	(0.000)				
`cVol`	1.958	1.041	2.971	0.442	−0.045
	(0.000)				
`WML`	0.875		1.213	1.133	−0.056

Table 3: **Transaction costs and performance fees**

The table reports the out-of-sample terminal returns net of transaction costs ($\overline{r}-c$, DeMiguel et al. 2009), the performance fee ($\mathcal{F}$, Fleming et al. 2003), and the abnormal return ($\mathcal{A}$) measures of Modigliani & Modigliani (1997). The performance fees are computed for a risk aversion coefficient of 5. All the measures are reported in annual basis points. The first column reports the level of transaction costs, expressed in basis points (bps). The sample period is from July 2nd 1929 to September 30th 2020, daily. Portfolio weights are generated in real-time by recursive forecasts of the conditional mean and variance of the returns based on the model parameters.

	`mSSR`			`mSR`			`cSVol`			`cVol`		
c (bps)	$\overline{r}-c$	$\mathcal{F}$	$\mathcal{A}$	$\overline{r}-c$	$\mathcal{F}$	$\mathcal{A}$	$\overline{r}-c$	$\mathcal{F}$	$\mathcal{A}$	$\overline{r}-c$	$\mathcal{F}$	$\mathcal{A}$
0	0.447	0.426	0.441	0.342	0.322	0.338	0.147	0.133	0.180	0.145	0.130	0.177
1	0.442	0.422	0.436	0.339	0.318	0.335	0.146	0.131	0.179	0.143	0.129	0.175
5	0.424	0.403	0.418	0.325	0.305	0.321	0.140	0.125	0.172	0.137	0.123	0.169
10	0.401	0.380	0.396	0.309	0.289	0.305	0.133	0.118	0.164	0.130	0.115	0.161
50	0.218	0.197	0.212	0.177	0.156	0.173	0.075	0.060	0.100	0.072	0.057	0.097

Table 4: **Transaction costs and risk aversion**

The table reports the performance fees, $\mathcal{F}$, relative to the managed portfolios for different values of risk aversion. We consider $\delta = 1, 7, 15$. The fees are computed with respect to the plain WML strategy. All the measures are reported in annual basis points. The first column reports the level of transaction costs, expressed in basis points (bps). The sample period is from July 2nd 1929 to September 30th 2020, daily. Portfolio weights are generated in real-time by recursive forecasts of the conditional mean and variance of the returns based on the model parameters.

	`mSSR`			`mSR`			`cSVol`			`cVol`		
c (bps)	$\delta=2$	$\delta=7$	$\delta=15$	$\delta=2$	$\delta=7$	$\delta=15$	$\delta=2$	$\delta=7$	$\delta=15$	$\delta=2$	$\delta=7$	$\delta=15$
0	0.426	0.411	0.400	0.358	0.133	0.126	0.121	0.104	0.130	0.123	0.119	0.101
1	0.422	0.406	0.396	0.354	0.131	0.124	0.120	0.102	0.129	0.122	0.117	0.099
5	0.403	0.388	0.377	0.336	0.125	0.119	0.114	0.096	0.123	0.116	0.111	0.093
10	0.380	0.365	0.355	0.312	0.118	0.111	0.107	0.089	0.115	0.109	0.104	0.086

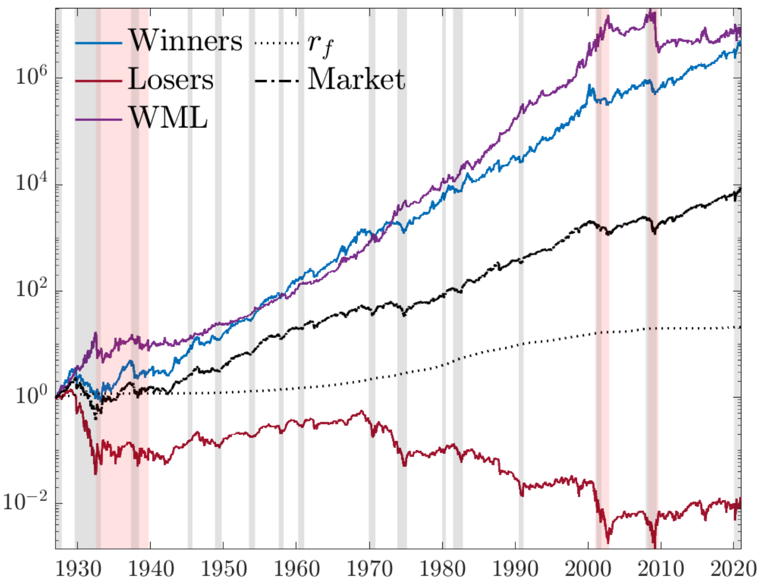
Table 5: **Spanning tests**

The table reports results of a series of spanning regressions where the dependent variables are the returns on different managed momentum portfolios based on constant volatility (`cVol`) as in Barroso & Santa-Clara (2015), dynamic volatility (`mSR`) as in Daniel & Moskowitz (2016), and skew-managed constant volatility (`cSVol`) and dynamic volatility (`mSSR`) strategies. Each portfolio returns is regressed onto the market and other Fama & French (1993) common risk factors in addition to the static `WML` portfolio. The sample size is from 1st August 1929 to September 30th 2020, daily.

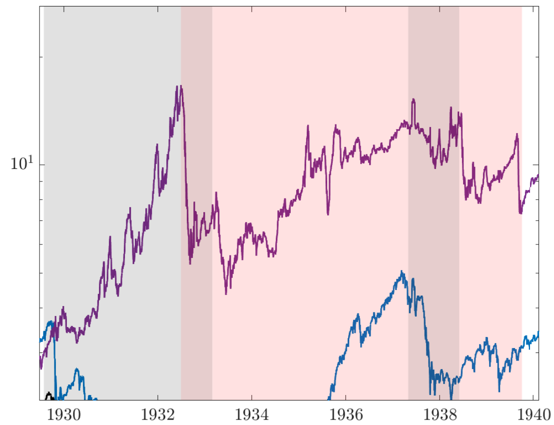
	Mkt+WML	FF3+WML	Mkt+cVol	FF3+cVol	Mkt+mSR	FF3+mSR
mSSR	76.004	75.331	50.639	49.864	17.285	16.751
	(33.283)	(33.150)	(28.856)	(28.620)	(20.016)	(19.643)
mSR	50.698	49.916	25.063	24.744		
	(28.258)	(28.006)	(22.953)	(22.738)		
cSVol	20.006	19.314	0.328	0.331	−7.075	−6.573
	(20.208)	(19.755)	(9.428)	(9.497)	(−8.439)	(−7.883)
cVol	19.650	18.957			−7.293	−6.791
	(19.844)	(19.389)			(−8.649)	(−8.096)

Figure 1: **Cumulative performance of the `WML` strategy**

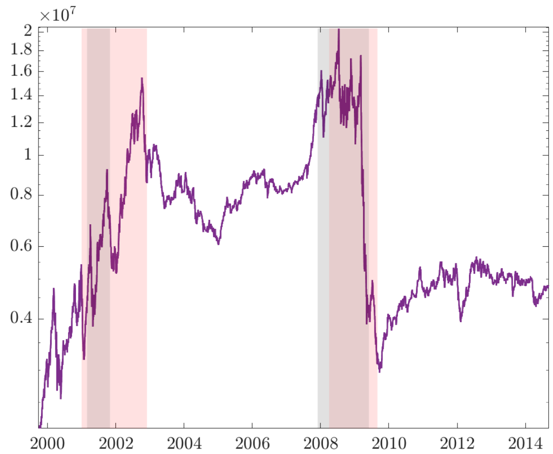
The plot reports the cumulative performance of a 12_2 momentum strategy, the market and treasury bond returns. The cumulative performance is reported in log-scale. Gray shaded bands highlight NBER recession. Red shaded bands indicate momentum crash periods, as indicated in Daniel & Moskowitz (2016).



(a) Full sample



(b) 1930-1940 crash



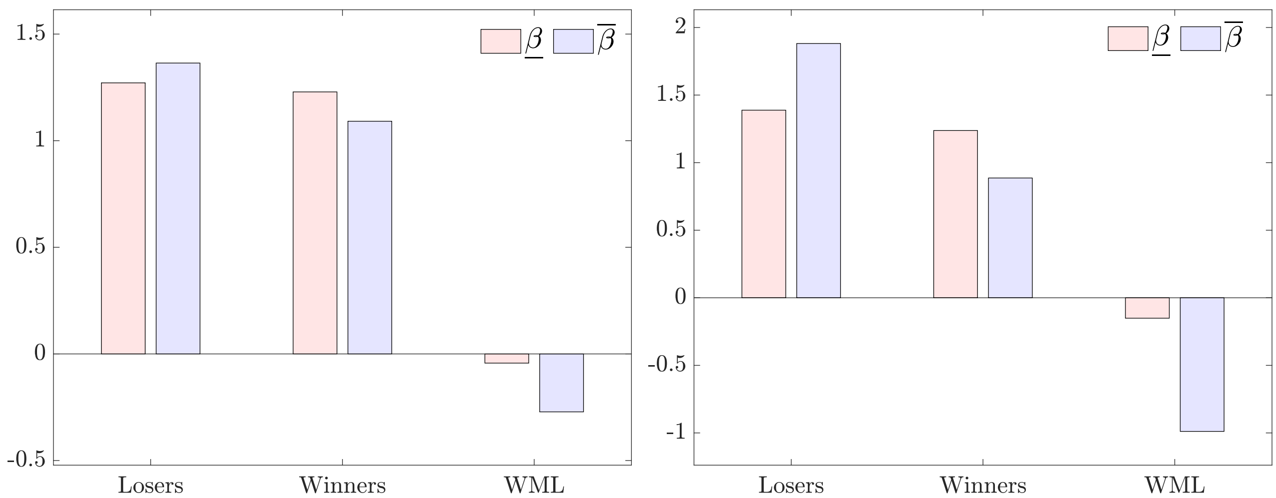
(c) 2000-2009 crash

Figure 2: **Static upside vs downside market betas**

The figures plot the upside, $\overline{\beta}$, and downside, $\underline{\beta}$, for the `losers`, `winners` and `WML` portfolios give by the following regression:

$$r_t^i = \alpha + \underline{\beta}^i \min(r_t^m, 0) + \overline{\beta}^i \max(r_t^m, 0) + \varepsilon_t,\ i = \texttt{losers}, \texttt{winners}, \texttt{WML}.$$

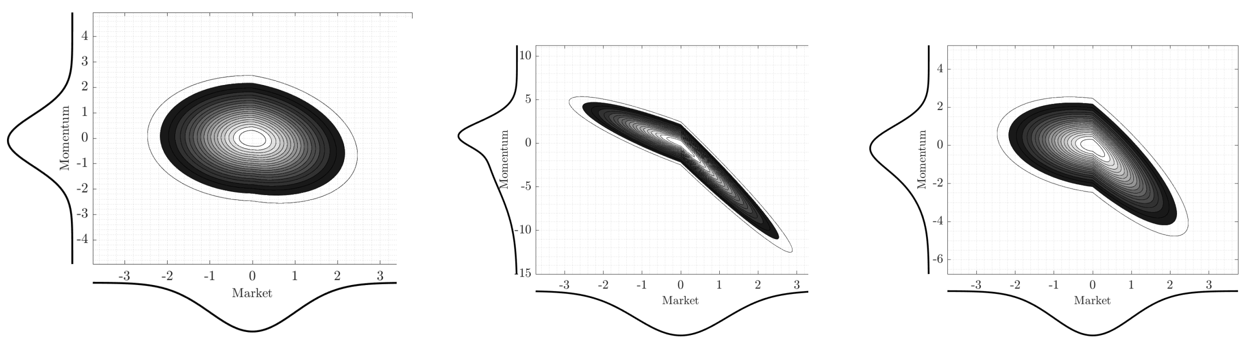
The sample period is from July 1st 1926 to September 30th 2020. The left (right) panel reports the estimates based on daily (monthly) returns.



(a) Daily (b) Monthly

Figure 3: **Simulated momentum strategy returns**

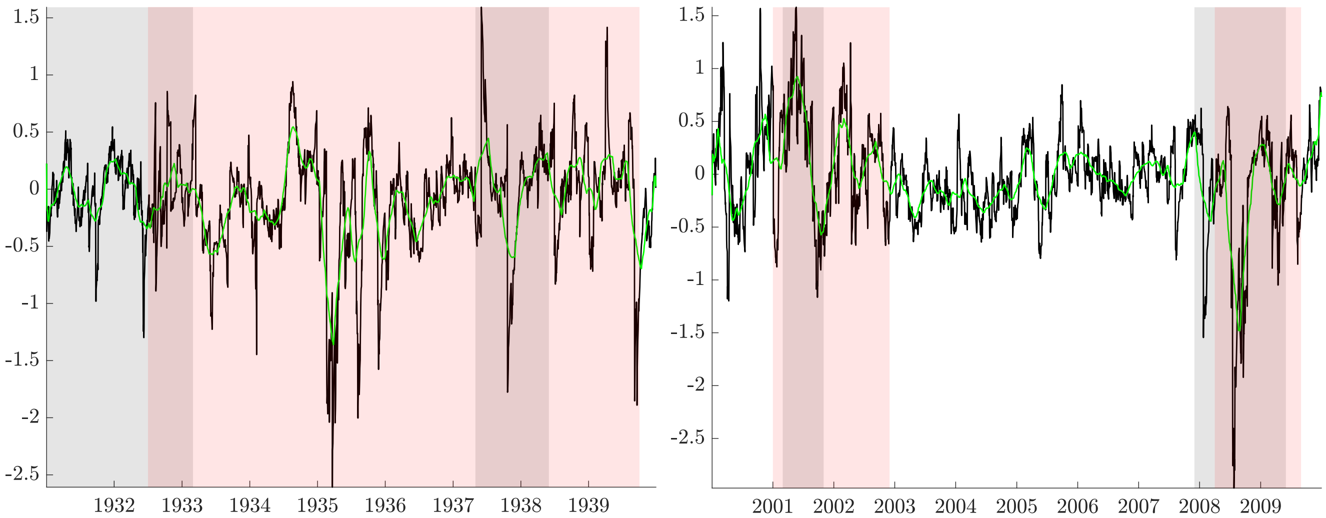
This figure reports the marginal distribution of the returns on the momentum strategy (y-axis) and the returns on the market portfolio (x-axis) and the corresponding joint distribution. Returns are simulated assuming a two-piece Normal distribution as in Eq.(A2) in Appendix A. The left panel shows the joint distribution for the full sample whereby the middle and the right panels show the joint distributions of the market and momentum returns during momentum crashes.



(a) **Sample** (b) **22/03/1935** (c) **30/09/2008**

Figure 4: **Momentum crashes and the exposure to downside and upside risk**

The plots report the relative downside (dot-dashed red) and upside (dashed blue) betas, and the risk exposure asymmetry in solid black. Top panels span the 1927-1940 period, while bottom panels cover from 2000 to 2020. Gray shaded bands highlight NBER recession. Red shaded bands indicate momentum crash periods, as indicated in Daniel & Moskowitz (2016).

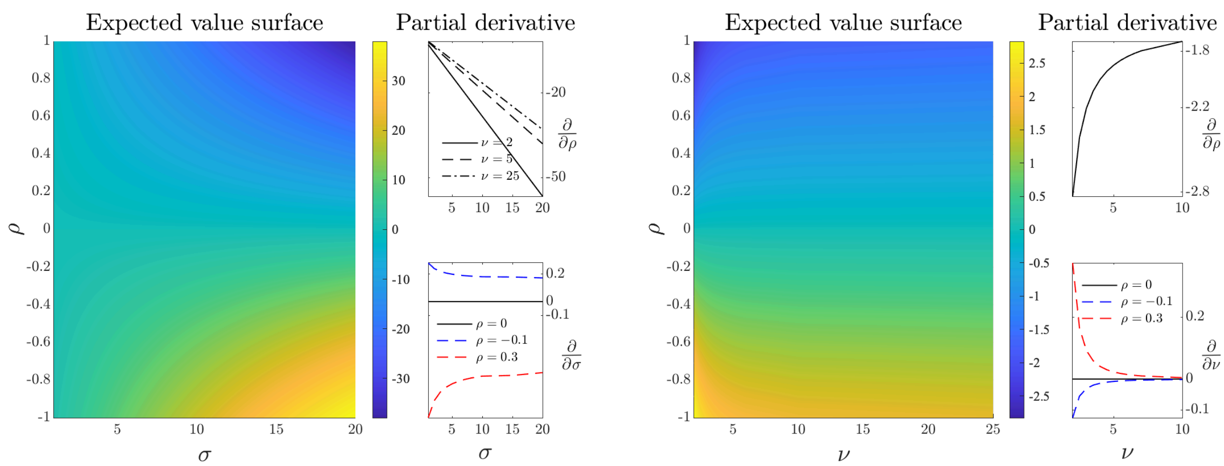


(a) 1927-1940 period

(b) 2000-2020 period

Figure 5: **The role of scale and shape parameters for expected returns**

The left panel reports the expected value surface for values of ρ_t and σ_t; the smaller windows in the plot illustrate the partial derivative of Eq.(7) with respect to ρ_t and ν for varying values of σ_t. Similarly, the right panel reports the expected value surface for values of ρ_t and ν; the smaller windows in the plot illustrate the partial derivative of the expected value with respect to ρ_t and ν for varying values of ν. Both surfaces are reported for zero location.

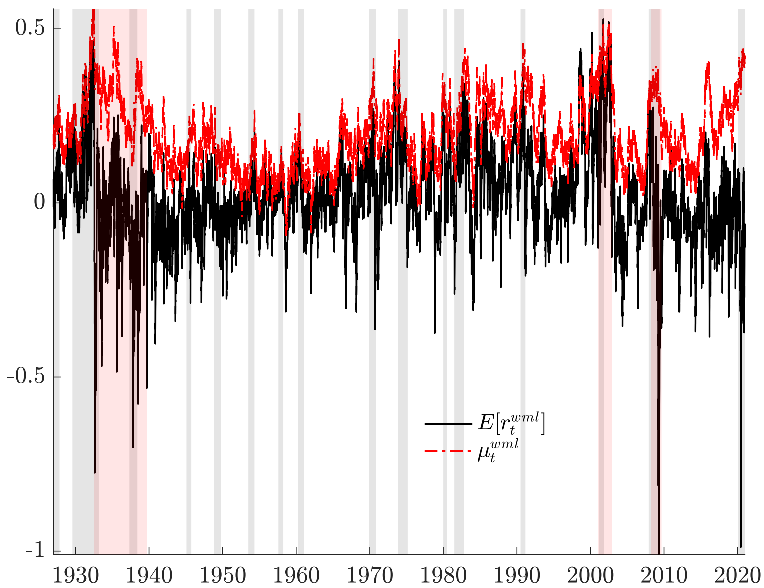


(a) ρ_t and σ_t

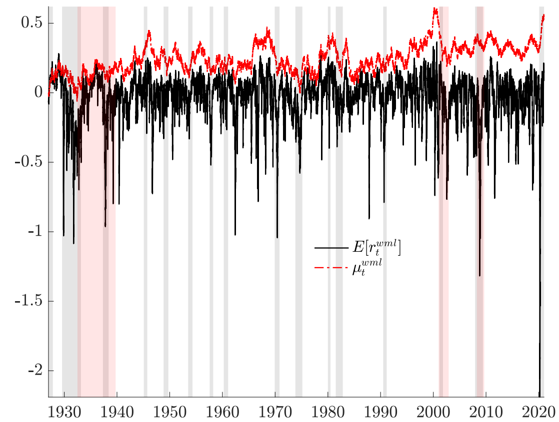
(b) ρ_t and ν

Figure 6: **Conditional expected momentum returns and the location parameter**

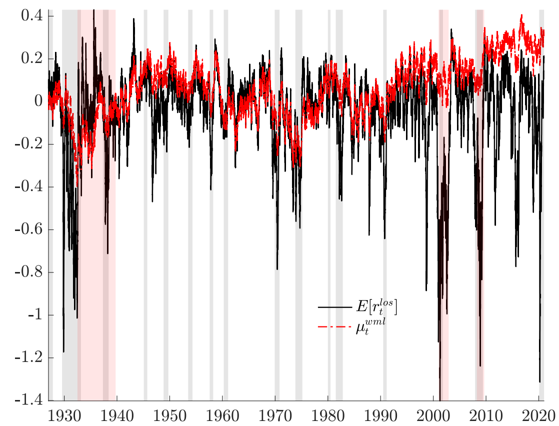
The plot reports the time-varying location parameter μ_t (red line) and the conditional expected returns as in Eq.(7) (black line). We report the values for the `WML` portfolio (left panel), the past `losers` (bottom-right panel) and the past `winners` (top-right panel) sub-portfolios. NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods, as indicated in Daniel & Moskowitz (2016). The sample period is from July 1st 1926 to September 30th 2020, daily.



(a) `WML` portfolio



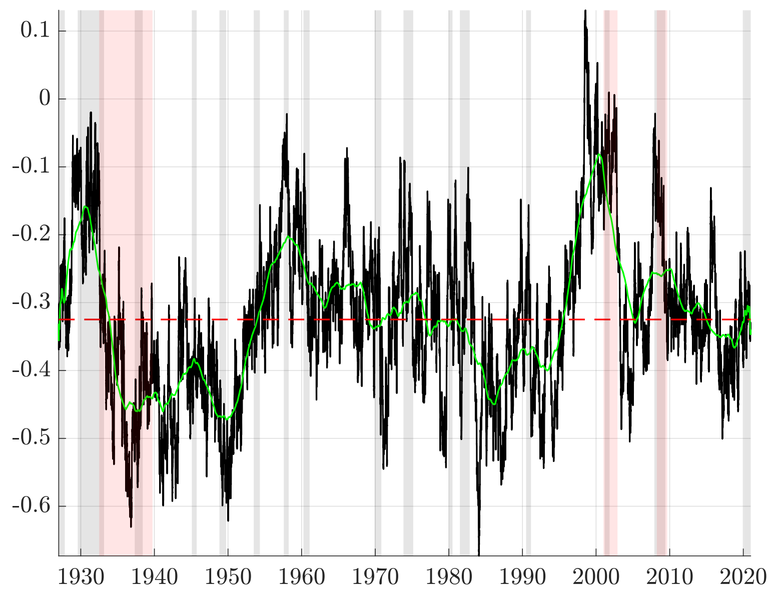
(b) Past `winners`



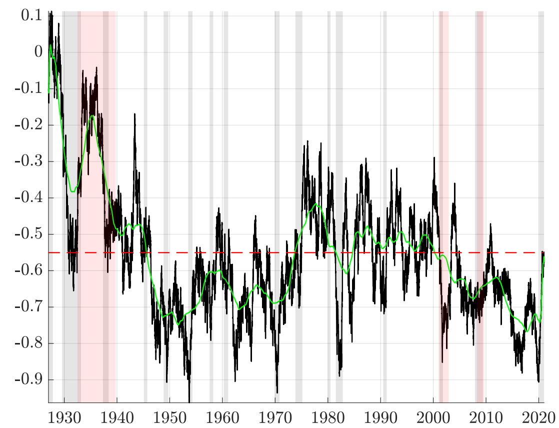
(c) Past `losers`

Figure 7: **Conditional skewness of momentum returns**

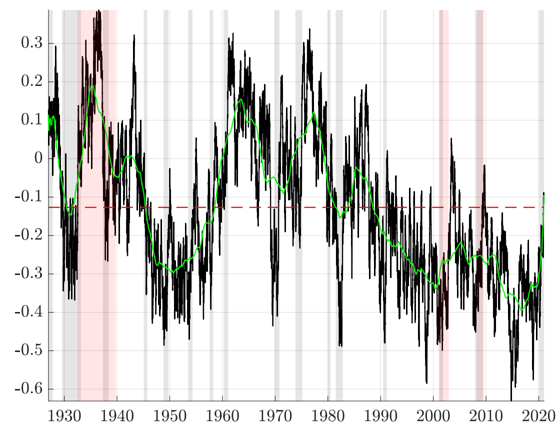
The plot reports the time-varying skewness estimates for the `WML` portfolio (left panel), the past `losers` (bottom-right panel) and the past `winners` (top-right panel) sub-portfolio returns. The horizontal red dashed line represents the sample mean of the conditional skewness estimates. The green line represents a smoothed representation of the daily conditional skewness estimates. NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods, as indicated in Daniel & Moskowitz (2016). The sample period is from July 1st 1926 to September 30th 2020, daily.



(a) `WML` portfolio



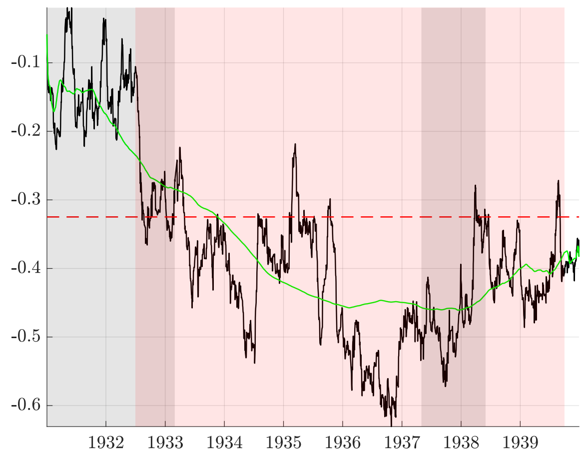
(b) Past `winners`



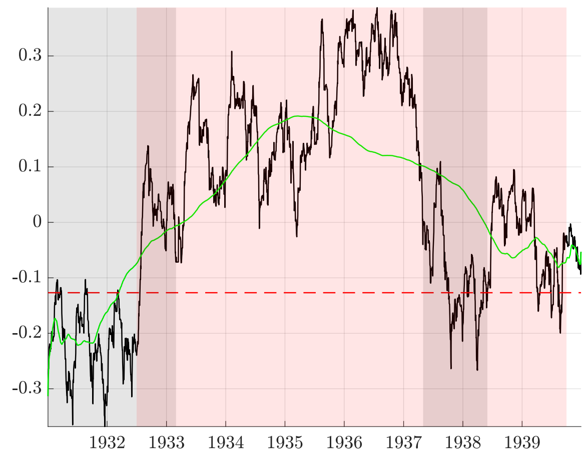
(c) Past `losers`

Figure 8: **Conditional skewness during momentum crashes**

The plot reports the time-varying skewness estimates for the `WML` portfolio (left panels), the past `losers` (mid panels) and the past `winners` (right panels) sub-portfolio returns. We report the results for both the 1930-1940 (top panels) and the 2001-2009 (bottom panels) periods. The horizontal red dashed line represents the sample mean of the conditional skewness estimates. The green line represents a smoothed representation of the daily conditional skewness estimates. NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods, as indicated in Daniel & Moskowitz (2016).

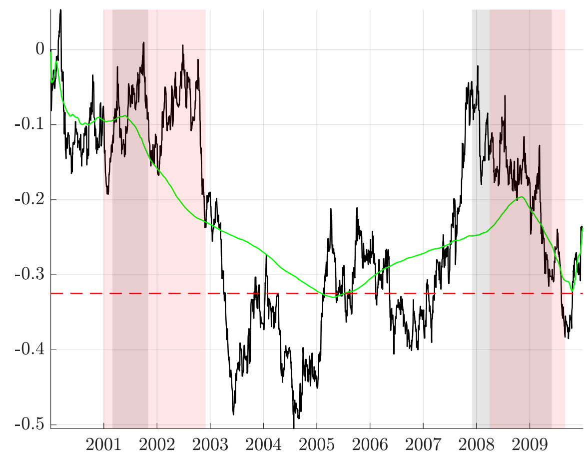


(a) `WML` portfolio

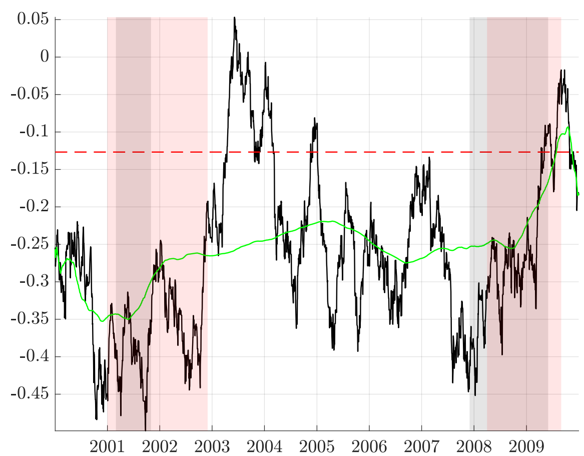


(b) Past `losers`

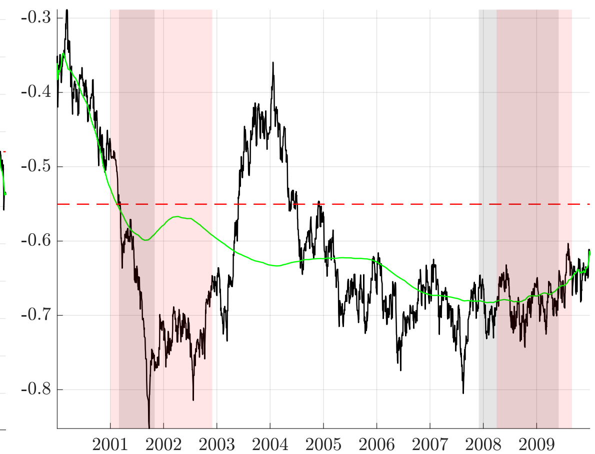
(c) Past `winners`



(d) `WML` portfolio



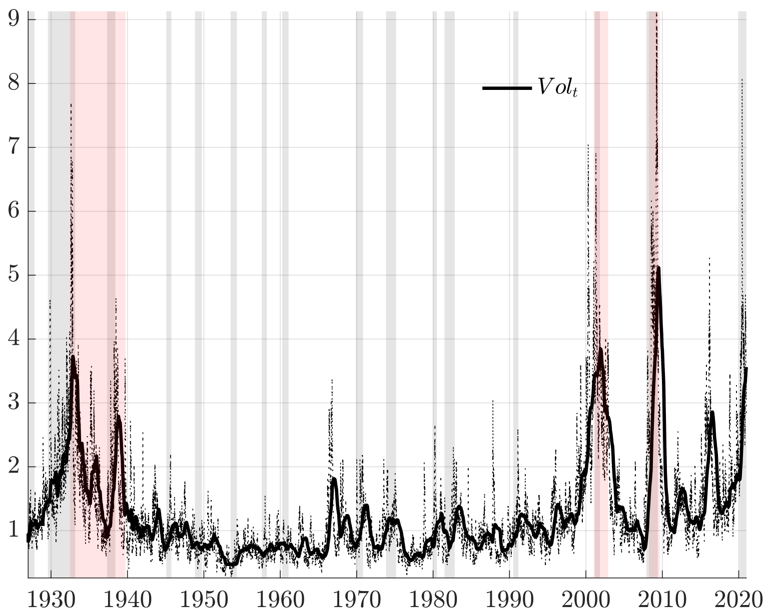
(e) Past `losers`



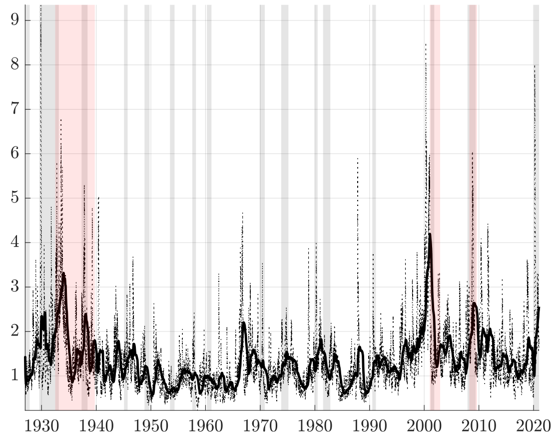
(f) Past `winners`

Figure 9: **Conditional volatility of momentum returns**

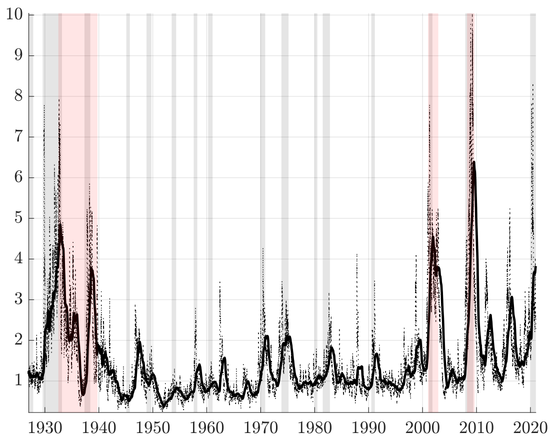
The plot reports the time-varying volatility estimates based on Eq.(8) for the `WML` portfolio (left panel), the past `losers` (bottom-right panel) and the past `winners` (top-right panel) sub-portfolios. NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods, as indicated in Daniel & Moskowitz (2016). The sample period is from July 1st 1926 to September 30th 2020, daily.



(a) `WML` portfolio



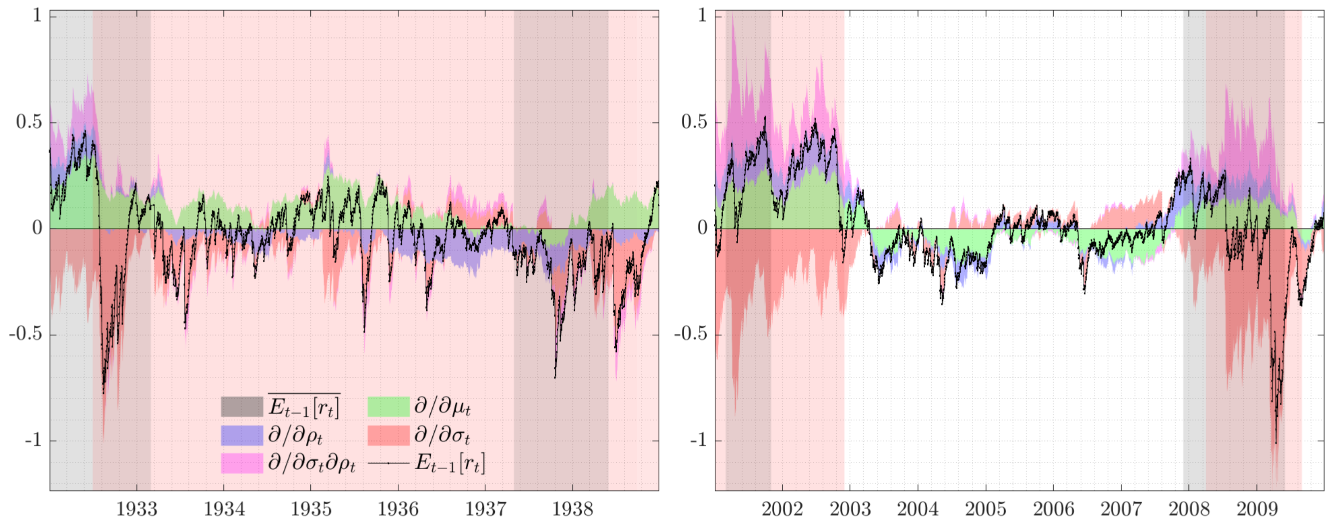
(b) Past `winners`



(c) Past `losers`

Figure 10: **Risk premium decomposition over momentum crashes**

The figure reports the decomposition of the expected returns, in black, into a location component (green shaded area), a scale component (pink shaded area), an asymmetry component (purple shaded area) and the interplay between volatility and asymmetry (magenta shaded area) and the *hom* component (purple shaded area). NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods, as indicated in Daniel & Moskowitz (2016).

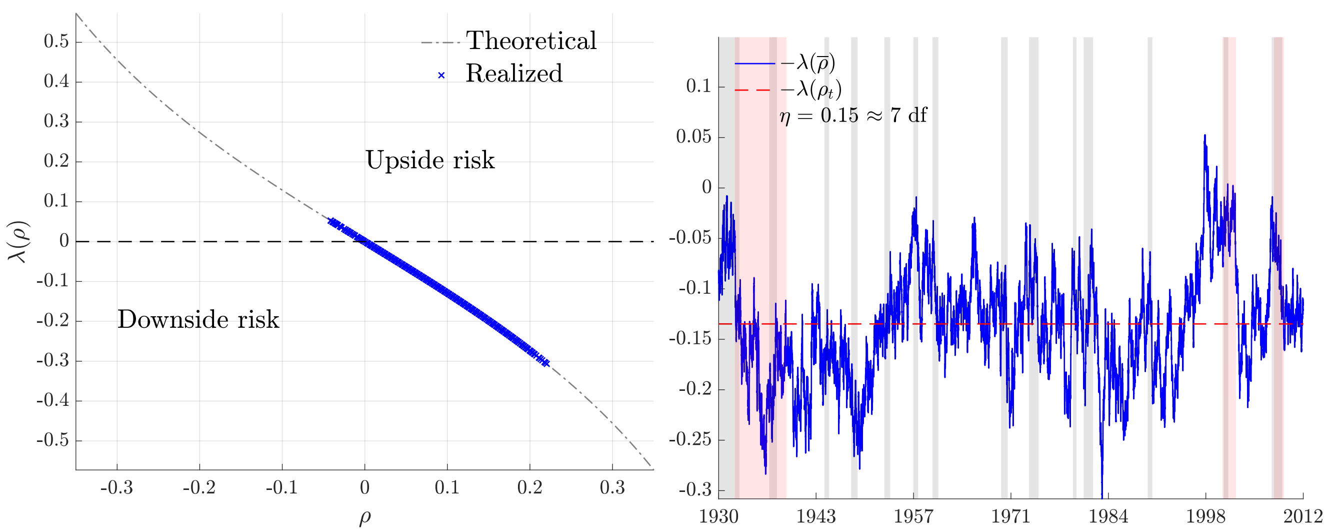


(a) 1932-1939 period

(b) 2001-2009 period

Figure 11: **Risk-return trade-off of the momentum strategy**

The left panel of the figure reports the theoretical shape of $\lambda(\rho_t)$ (dashed curve) and its realized value (blue marks). The right panel of the figure shows the time series of the slope parameter $\lambda(\rho_t)$ for $t = 1, \ldots, T$. The sample period is from July 1st 1926 to September 30th 2020, daily.

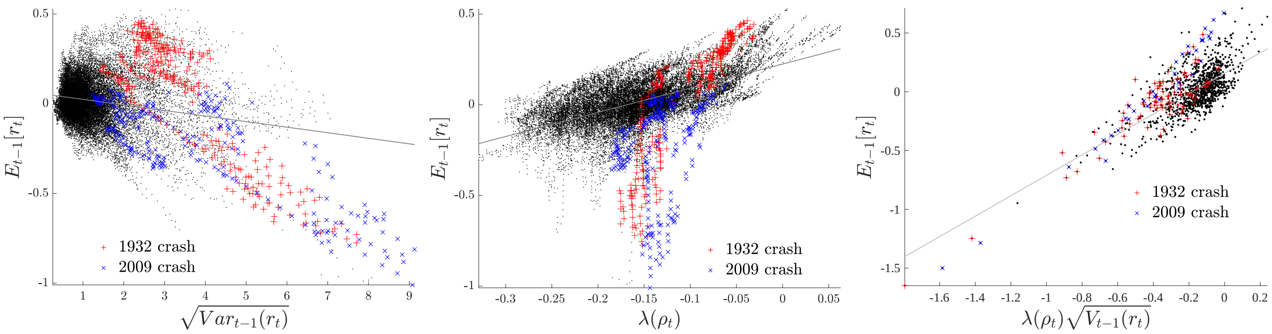


(a) The $\lambda\,(\rho)$ parameter

(b) Time series dynamics

Figure 12: **Decomposing the risk-return trade-off**

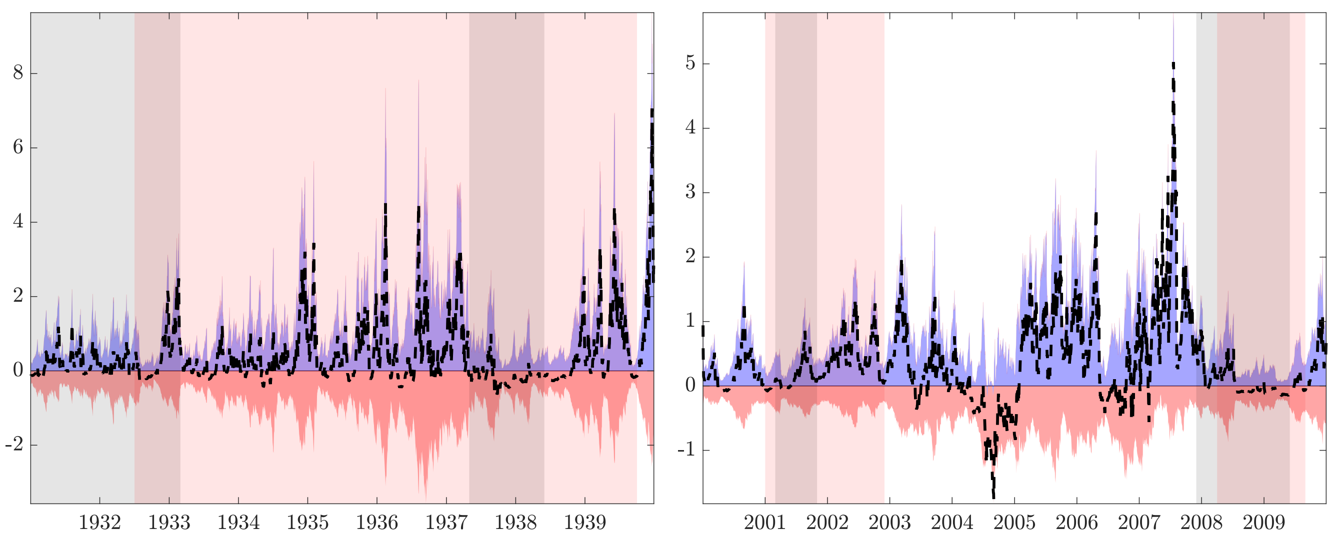
The three panels report the the correlation plots between the expected return and the volatility, the slope parameter $\lambda(\rho_t)$ and the fitted value of the risk-return trade-off in Eq.(**??**). Red crosses highlight observations relative to the 1932 crash. Blue crosses highlight observations relative to the 2009 crash.



(a) Expected returns vs volatility (b) Expected returns vs slope (c) Realised vs fitted expected returns

Figure 13: **Managed portfolio weights**

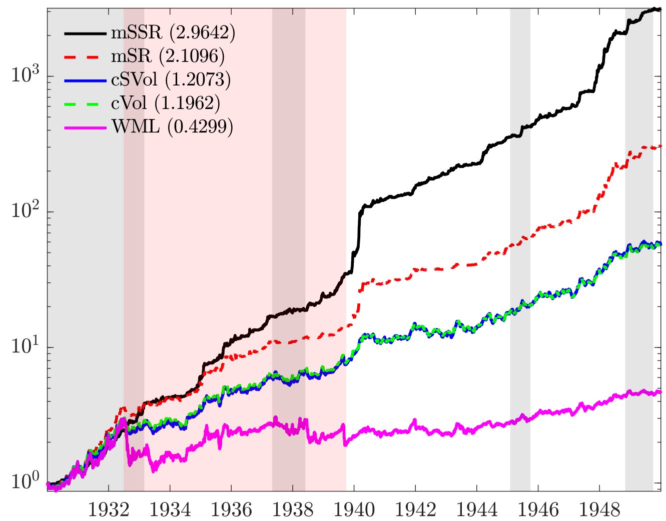
The plot reports the decomposition of the maximum Sharpe Ratio targeting weights. The component associated with the central tendency, $w_{1,t}$, is reported in blue, while that hedging the overall uncertainty, $w_{2,t}$, is in red. The total weight, w_t, is reported in black. The left panel covers the momentum crashes of 1932 and 1940. The right panel reports the momentum crashes of 2001 and 2009. The red dashed line represent the unconditional mean. NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods as indicated by Daniel & Moskowitz (2016).



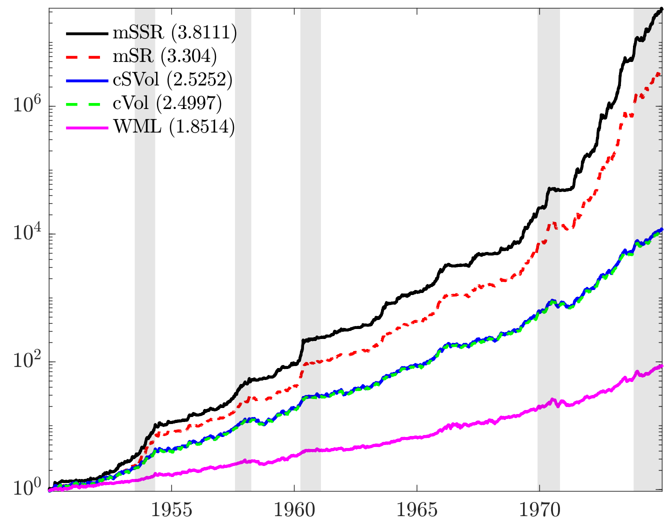
(a) 1932-1939 crash period (b) 2001-2009 crash period

Figure 14: **Cumulative returns across sub-samples**

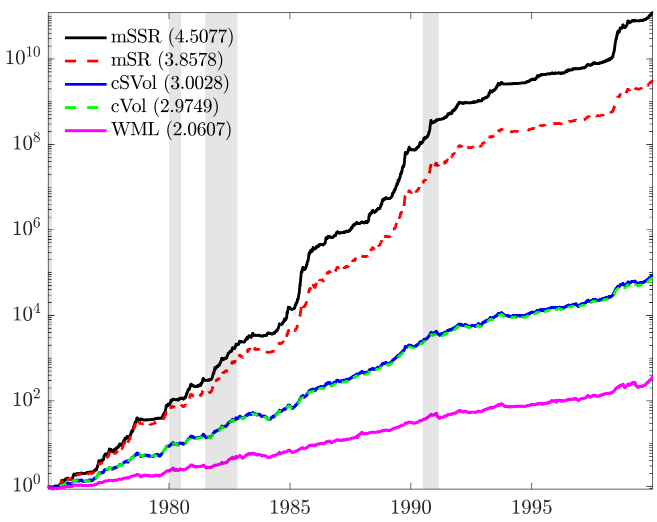
This figures shows the compounded returns assuming 1$ initial investment of the standard `WML` portfolio and all different competing risk-managed strategies based on constant volatility targeting (see Barroso & Santa-Clara 2015), dynamic volatility targeting (see Daniel & Moskowitz 2016), and our skew-managed momentum strategy (see Eq.(12)). The figure reports four different sub-samples; the three major momentum crashes and two samples without large losses in momentum.



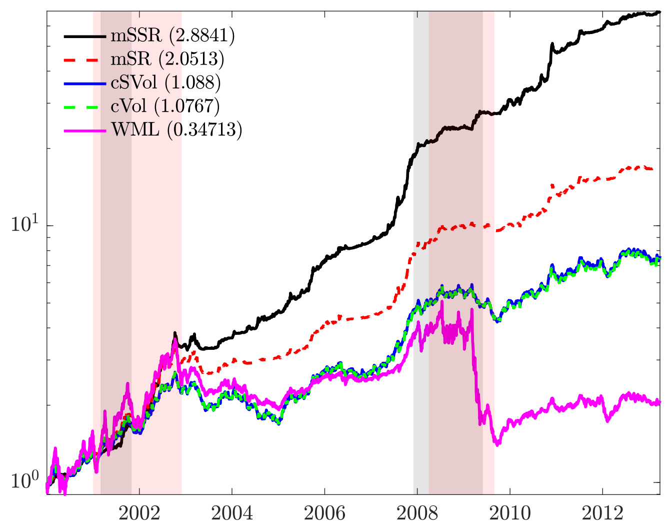
(a) 1932-1939 crash period



(b) 1950-1975 period



(c) 1975-2000 period



(d) 2001-2009 crashes period

A From asymmetric betas to skewness

Let consider the conditional regression model in Eq.(1),

$$r_t = \alpha + \underbrace{\overline{\beta} m_t I(m_t \geq \mu_m) + \underline{\beta} m_t I(m_t < \mu_m)}_{\beta m_t} + e_t \tag{A1}$$

with $m_t \sim \mathcal{N}(\mu_m, \sigma_m^2)$ the normal distributed market portfolio and $I(m_t \geq \mu_m)$ ($I(m_t < \mu_m)$) an indicator function that takes value one if the market returns are above (below) the mode μ_m and zero otherwise. The distribution of βm_t conditional on the indicator $I(\cdot)$ can be defined as a split-Normal (or two-piece Normal) distribution of the form (see Johnson et al. 1995),

$$f(\beta m_t) = \begin{cases} C \exp\left\{-\frac{1}{2\underline{\sigma}_m^2}\left(\underline{\beta} m_t - \beta\mu_m\right)^2\right\} & \text{if } m_t \leq \mu_m \\ C \exp\left\{-\frac{1}{2\overline{\sigma}_m^2}\left(\overline{\beta} m_t - \beta\mu_m\right)^2\right\} & \text{if } m_t > \mu_m \end{cases} \tag{A2}$$

with $C = \sqrt{\frac{2}{\pi}}\left(\underline{\sigma}_m + \overline{\sigma}_m\right)^{-1}$ and $\underline{\sigma}_m^2 = \underline{\beta}^2\sigma_m^2$ and $\overline{\sigma}_m^2 = \overline{\beta}^2\sigma_m^2$. Following Wallis (2014), the expected value of the distribution takes the form

$$E[\beta m_t] = \sqrt{\frac{2}{\pi}}\left(\overline{\sigma}_m - \underline{\sigma}_m\right) + \beta\mu_m, \tag{A3}$$

Notice that for $\underline{\beta} = \overline{\beta} = \beta$, then we have $\overline{\sigma}_m^2 = \underline{\sigma}_m^2 = \sigma_m^2$, such that $E[\beta m_t] = \beta\mu_m$. That is, the mean and the mode of the conditional distribution of the momentum returns coincide, i.e., $E[r_t] = \alpha + \beta\mu_m$. Similarly, the variance of the split-Normal in Eq.(A2) takes the form ,

$$V[\beta m_t] = \left(1 - \frac{2}{\pi}\right)\left(\overline{\sigma}_m^2 - \underline{\sigma}_m^2\right)^2 + \overline{\sigma}_m\underline{\sigma}_m \tag{A4}$$

such that for no asymmetry in the betas estimates the first component $\left(1 - \frac{2}{\pi}\right)\left(\overline{\sigma}_m^2 - \underline{\sigma}_m^2\right)^2 = 0$, and we are left with $V[\beta m_t] = \sqrt{\beta^2\sigma_m^2}\sqrt{\beta^2\sigma_m^2} = \beta^2\sigma_m^2$. As a result, for $\underline{\beta} = \overline{\beta} = \beta$, and given $e_t \sim N(0, \sigma_e^2)$, we obtain that the marginal distribution of the momentum strategy returns is $r_t \sim \mathcal{N}(\alpha + \beta\mu_m, \beta^2\sigma_m^2 + \sigma_e^2)$. Now let us assume that $\underline{\beta} \neq \overline{\beta}$, and indicator of the asymmetry of the returns distribution can be defined as the difference between the expected value $E[\beta m_t]$ and the mode $\beta\mu_m$, which is given by

$$\begin{aligned} E[\beta m_t] - \beta\mu_m &= \sqrt{\frac{2}{\pi}}\left(\overline{\sigma}_m - \underline{\sigma}_m\right) \propto \sqrt{\overline{\beta}^2\sigma_m^2} - \sqrt{\underline{\beta}^2\sigma_m^2}, \\ &= \sigma_m\left(\sqrt{\overline{\beta}^2} - \sqrt{\underline{\beta}^2}\right) \\ &= \sigma_m\left(\overline{\beta} - \underline{\beta}\right) \end{aligned} \tag{A5}$$

that is, for $\overline{\beta} = \underline{\beta}$ there is no returns asymmetry, whereas for $\overline{\beta} < \underline{\beta}$ ($\overline{\beta} > \underline{\beta}$) the expected value is lower (higher) than the mode, that is the marginal distribution of the returns is negatively (positively) skewed.

B Econometric framework

Assume that the variable y_t is generated by the observation density $\mathcal{D}(\theta, f_t)$, with θ collecting the static parameters of the distribution. The score-driven setting postulates the dynamics of the time-varying parameters, f_t, being:

$$f_{t+1} = \varpi + \sum_{i=0}^{p-1} \alpha_i s_{t-i} + \sum_{j=0}^{q-1} \beta_j f_{t-j}, \tag{B1}$$

which we refer to as $GAS(p, q)$ dynamics.[16] The scaled score s_t is a non-linear function of past observations and past parameters' values. For $\ell_t = \log \mathcal{D}(\theta, f_t)$, we define:

$$s_t = \mathcal{S}_t \nabla_t, \quad \nabla_t = \frac{\partial \ell_t}{\partial f_t}, \quad \mathcal{S}_t = \mathcal{I}_t^{-\frac{1}{2}} = -\,\mathbb{E}\left(\frac{\partial^2 \ell_t}{\partial f_t \partial f_t'}\right)^{-\frac{1}{2}},$$

where ∇_t corresponds to the gradient vector of the log-likelihood function, ℓ_t, and the scaling matrix $\mathcal{S}_{t-1}$ is proportional to the square-root generalized inverse of the Information matrix $\mathcal{I}_{t-1}$.[17] Within this framework, the parameters are updated in the direction of the steepest ascent, in order to maximize the local fit of the model.

B.1 Score-driven Skew-t model

In this appendix we provide details on the scaled scores for the Skew-t distribution of Gómez et al. (2007), the use of the link functions and the matrix representation of the model. Refer to Appendix B in Delle Monache et al. (2021) for detailed derivations.

Scaled scores. Assume the conditional distribution of y_t being the Skew-t of Gómez et al. (2007), $y_t|Y_{t-1} \sim skt_\nu(\mu_t, \sigma_t^2, \rho_t)$, with log-likelihood

$$\ell_t(y_t|\theta, Y_{t-1}) = \log \mathcal{C}(\eta) - \frac{1}{2}\log \sigma_t^2 - \frac{1+\eta}{2\eta}\log\left[1 + \frac{\eta \varepsilon_t^2}{(1 - sgn(\varepsilon_t)\rho_t)^2 \sigma_t^2}\right], \tag{B2}$$
$$\log \mathcal{C}(\eta) = \log \Gamma\left(\frac{\eta+1}{2\eta}\right) - \log \Gamma\left(\frac{1}{2\eta}\right) - \frac{1}{2}\log\left(\frac{1}{\eta}\right) - \frac{1}{2}\log \pi,$$

where $\Gamma(\cdot)$ is the Gamma function, $sgn(.)$ is the sign function, and $\eta = 1/\nu$ is the inverse of the degrees of freedom.

Differentiating the log-likelihood function in (C18) with respect to location, scale and

[16] Creal et al. (2013) refer to the Generalized Autoregressive Score (GAS) model, while Harvey (2013) names it Dynamics Conditional Score (DCS) model. We stick to the former notation throughout this appendix.
[17] Refer to Creal et al. (2013) for additional details on this choice.

asymmetry we obtain the gradient vector $\nabla_t = \left[\frac{\partial \ell_t}{\partial \mu_t}, \frac{\partial \ell_t}{\partial \sigma_t^2}, \frac{\partial \ell_t}{\partial \rho_t}\right]'$, with elements:

$$\frac{\partial \ell_t}{\partial \mu_t} = \frac{1}{\sigma_t^2}\left[\frac{(1+\eta)\sigma_t^2 \varepsilon_t}{(1-sgn(\varepsilon_t)\rho_t)^2\sigma_t^2 + \eta\varepsilon_t^2}\right] = \frac{1}{\sigma_t^2} w_t \varepsilon_t, \tag{B3}$$

$$\frac{\partial \ell_t}{\partial \sigma_t^2} = \frac{1}{2\sigma_t^2}\left[\frac{(1+\eta)\varepsilon_t^2}{(1-sgn(\varepsilon_t)\rho_t))^2\sigma_t^2 + \eta\varepsilon_t^2} - 1\right] = \frac{1}{2\sigma_t^4}(w_t\varepsilon_t^2 - \sigma_t^2), \tag{B4}$$

$$\frac{\partial \ell_t}{\partial \rho_t} = -\frac{sgn(\varepsilon_t)}{(1-sgn(\varepsilon_t)\rho_t)}\frac{(1+\eta)}{(1-sgn(\varepsilon_t)\rho_t))^2\sigma_t^2 + \eta\varepsilon_t^2}\varepsilon_t^2 = -\frac{1}{\sigma_t^2}\frac{sgn(\varepsilon_t)}{(1-sgn(\varepsilon_t)\rho_t)} w_t \varepsilon_t^2, \tag{B5}$$

where

$$w_t = \frac{(1+\eta)}{(1-sgn(\varepsilon_t)\rho_t)^2 + \eta\zeta_t^2}, \tag{B6}$$

and $\zeta_t = \frac{\varepsilon_t}{\sigma_t}$ is the standardized innovations. The first panel in Figure fig. B1 illustrates the shape of the weights in eq. (B6) against the standardized innovations: when the distribution is Gaussian (black), w_t is constant and implies no discounting of extreme realizations; once fat tails are factored in (red), large realizations, in absolute value, get down-weighted by means of the classic outlier-discounting typical of the symmetric Student-t distribution (see, e.g., Delle Monache & Petrella 2017). When the distribution is positively (negatively) skewed, i.e., for $\rho_t < 0$ ($\rho_t > 0$), negative (positive) prediction errors are less likely in expectation, and as such command a larger update of the parameters. This asymmetric treatment of the prediction error is more pronounced as skewness grows larger (i.e., $|\rho_t| \to 1$).

The Fisher information matrix is computed as the expected values of outer product of the gradient vector, for a fixed number of degrees of freedom:

$$\mathcal{I}_{t|t-1} = \mathbb{E}_{t-1}[\nabla_t \nabla_t'] = \begin{bmatrix} \frac{(1+\eta)}{(1+3\eta)(1-\rho_t^2)\sigma_t^2} & 0 & -\frac{4c(1+\eta)}{\sigma_t(1-\rho_t^2)(1+3\eta)} \\ 0 & \frac{1}{2(1+3\eta)\sigma_t^4} & 0 \\ -\frac{4c(1+\eta)}{\sigma_t(1-\rho_t^2)(1+3\eta)} & 0 & \frac{3(1+\eta)}{(1-\rho_t^2)(1+3\eta)} \end{bmatrix}. \tag{B7}$$

Given we model $\gamma_t = \log \sigma_t$ and $\delta_t = atanh(\rho_t)$, for the chain rule we have:

$$\frac{\partial \ell_t}{\partial \gamma_t} = \frac{\partial \ell_t}{\partial \sigma_t^2}\frac{\partial \sigma_t^2}{\partial \gamma_t}, \quad \frac{\partial \ell_t}{\partial \delta_t} = \frac{\partial \ell_t}{\partial \rho_t}\frac{\partial \rho_t}{\partial \delta_t}, \tag{B8}$$

where $\frac{\partial \sigma_t^2}{\partial \gamma_t} = 2\sigma_t^2$ and $\frac{\partial \rho_t}{\partial \delta_t} = (1-\rho_t^2)$. We can thus define the vector of interest as $f_t = (\mu_t, \gamma_t, \delta_t)'$ with the associated Jacobian matrix

$$J_t = \frac{\partial(\mu_t, \sigma_t^2, \rho_t)}{\partial f_t'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2\sigma_t^2 & 0 \\ 0 & 0 & 1-\rho_t^2 \end{bmatrix}. \tag{B9}$$

As such, the scaled score reads:

$$\mathbf{s}_t = (J_t' \text{diag}(\mathcal{I}_t) J_t)^{-\frac{1}{2}} J_t' \nabla_t = \begin{bmatrix} s_{\mu t} \\ s_{\sigma t} \\ s_{\rho t} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{(1+3\eta)(1-\rho_t^2)}{(1+\eta)}} w_t \varepsilon_t \\ \sqrt{\frac{(1+3\eta)}{2}} (w_t \zeta_t^2 - 1) \\ -sgn(\varepsilon_t) \sqrt{\frac{(1+sgn(\varepsilon_t)\rho_t)(1+3\eta)}{3(1-sgn(\varepsilon_t)\rho_t)(1+\eta)}} w_t \zeta_t^2 \end{bmatrix}. \tag{B10}$$

The reminder of fig. B1 plots the updating mechanism for the time-varying parameters, as per eq. (B10). The location updates in the direction of the prediction error, while heavy tails introduce the typical S-shaped influence function (see, e.g. Harvey & Luati 2014); in our case this adapts to the asymmetry of the conditional distribution. The shape updates in the opposite direction of the prediction error, such that for negative innovations the distribution becomes more left skewed. On the contrary, updates of the scale do not depend on the sign of the innovations, but on their magnitude. Notice that, while the scores for the location and shape parameters are negatively correlated ($Corr(s_{\mu,t}, s_{\delta,t}) = -\frac{4\mathcal{C}}{\sqrt{3}}$), updates of σ_t are (unconditionally) uncorrelated with revisions of the other parameters, as suggested by the Information matrix.

In practice, to prevent numerical instability, it is often the case to replace the scaling matrix $\dot{\mathcal{S}}_t = (J_t' \text{diag}(\mathcal{I}_t) J_t)^{-\frac{1}{2}}$ with its smoothed estimator, $\ddot{\mathcal{S}}_t = (1-\lambda)\ddot{\mathcal{S}}_{t-1} + \lambda \dot{\mathcal{S}}_t, 0 < \lambda < 1$.

C Moments and score-vector derivation

In this Section, to simplify the notation, we drop the time subscript from the time-varying parameters. Consider the *Skew-t* distribution proposed by Gómez et al. (2007):

$$p(y|\mu, \sigma, \rho, \nu) = \frac{\mathcal{C}}{\sigma} \left[1 + \frac{1}{\nu} \left(\frac{y-\mu}{\sigma(1 - sgn(y-\mu)\rho)} \right)^2 \right]^{-\frac{1+\nu}{2}}, \tag{C11}$$

where $\mathcal{C} = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}$. Arellano-Valle et al. (2005) shows that any symmetric density on $\mathbb{R}$ can be uniquely determined from a density on $\mathbb{R}^+$, and a $Skew-t$ distribution can then be expressed in terms of strictly positive densities. Specifically, we can re-parametrize the density in C11 as a two-piece distribution (Fernández & Steel 1998):

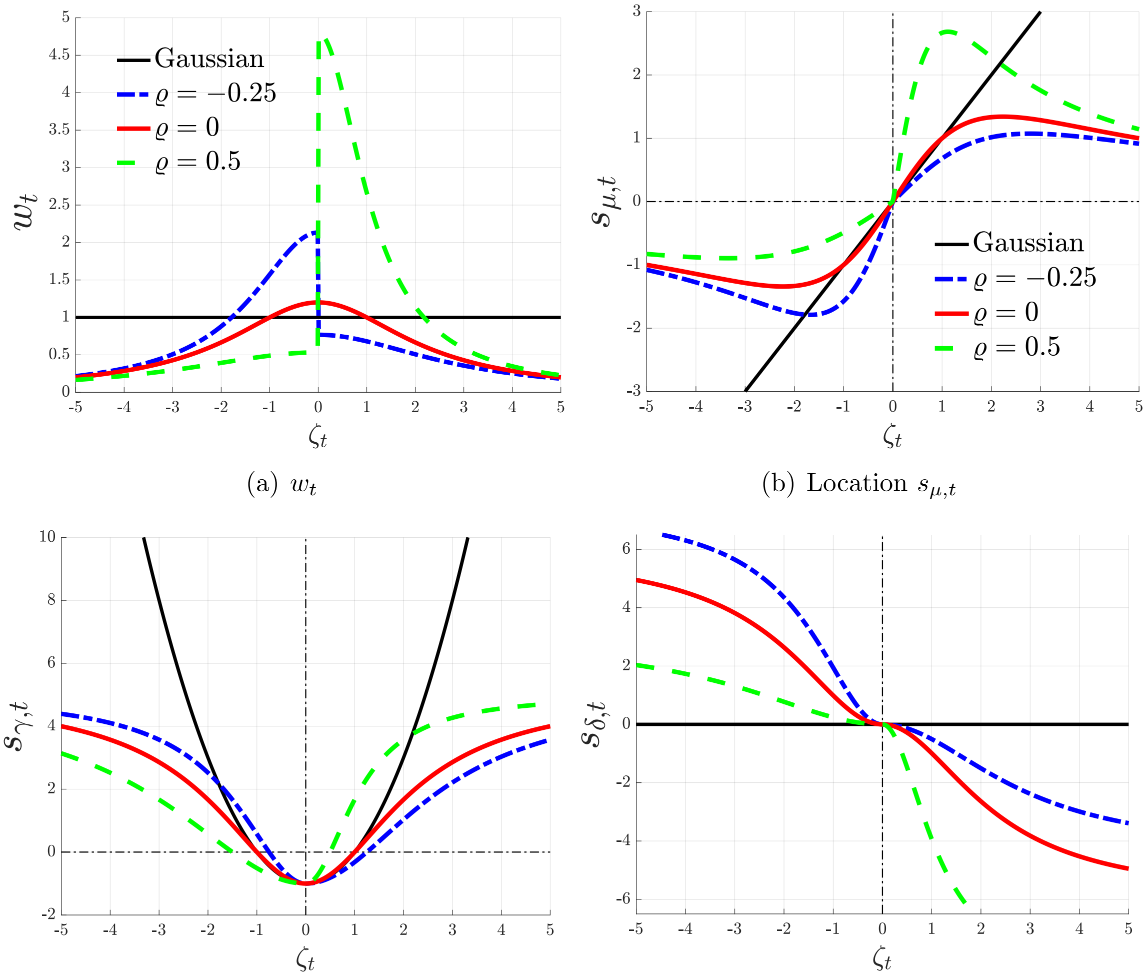
$$p(y|\mu, \sigma, \rho, \nu) = \begin{cases} \frac{\mathcal{C}}{\sigma} \left[1 + \frac{1}{\nu} \left(\frac{y-\mu}{\sigma_+} \right)^2 \right]^{-\frac{1+\nu}{2}}, & y \geq \mu \\ \frac{\mathcal{C}}{\sigma} \left[1 + \frac{1}{\nu} \left(\frac{y-\mu}{\sigma_-} \right)^2 \right]^{-\frac{1+\nu}{2}}, & y < \mu \end{cases} \tag{C12}$$

where $\sigma_+ = (1-\rho)\sigma$ and $\sigma_- = (1+\rho)\sigma$ are the scale parameters of the two *Half-t* densities on each side

$$P(y \geq \mu) = \frac{\sigma_+}{\sigma_+ + \sigma_-} = \frac{1-\rho}{2}, \qquad P(y < \mu) = \frac{\sigma_-}{\sigma_+ + \sigma_-} = \frac{1+\rho}{2}. \tag{C13}$$

Figure B1: **Properties of the score vector updating scheme**

The figure reports the theoretical properties of the score vector for the dynamics of the location, scale and asymmetry parameters as a function of the standardized residuals.



(a) w_t (b) Location $s_{\mu,t}$

(c) Log-volatility $s_{\gamma,t}$ (d) Asymmetry $s_{\delta,t}$

The two-piece formulation allows to consider separately the two halves of the distribution when taking expectations: for $y = \mu + \sigma\zeta$, where $\zeta \sim Skt_\nu(0, 1, \rho)$, the moments of y are weighted averages of the moments of $|\zeta|$, where $|\zeta| \sim Ht_\nu$, is an *Half-t* distribution (see, e.g., Gómez et al. 2007).[18] Specifically:

$$\mathbb{E}[\zeta^r] = \hat{\mu}_r = \frac{1}{2}\left[(1-\rho)^{r+1} + (-1)^r(1+\rho)^{r+1}\right] d_r(\nu), \tag{C14}$$

where $d_r(\nu) = \int_{-\infty}^{\infty} |\zeta|^r p(\zeta) d\zeta < \infty$ is the r^{th} moment of the *Half-t* distribution (Johnson

[18]Notice that the Half-t distribution is a special case of the folded-f distribution (Psarakis & Panaretoes 1990).

et al. 1995). Starting from C14, the moments of y are the computed as:

$$\mathbb{E}[y^j] = \sum_{k=0}^{j} \binom{j}{k} \sigma^k \mu^{j-k} \hat{\mu}_k.$$

Therefore, the expected value and variance of y are given by:

$$\begin{aligned} \mathbb{E}[y] &= \mu + \hat{\mu}_1 \sigma \\ &= \mu - \frac{4\nu \mathcal{C}(\nu)}{\nu - 1} \rho\sigma, \qquad \nu > 1 \end{aligned} \tag{C15}$$

$$\begin{aligned} \mathbb{E}[y^2] &= \mu^2 + 2\mu\sigma\hat{\mu}_1 + \sigma^2 \hat{\mu}_2 \\ &= \mu^2 - 2\mu\sigma^2 \frac{4\nu\mathcal{C}(\nu)}{\nu - 1}\rho + \sigma^2 \frac{(1+3\rho^2)\nu}{\nu - 2}, \qquad \nu > 2 \end{aligned} \tag{C16}$$

$$\begin{aligned} Var(y) &= \mathbb{E}[y^2] - \mathbb{E}[y] \\ &= \mu^2 - 2\mu\sigma^2 \frac{4\nu\mathcal{C}(\nu)}{\nu - 1}\rho + \sigma^2 \frac{(1+3\rho^2)\nu}{\nu - 2} - \left(\mu - \frac{4\nu\mathcal{C}(\nu)}{\nu - 1}\rho\sigma\right)^2 \\ &= \sigma^2 \left(\frac{(1+3\rho^2)\nu}{\nu - 2} - \left(\frac{4\nu\mathcal{C}(\nu)}{\nu - 1}\rho\right)^2 \right) \\ &= \sigma^2 \left[\frac{\nu}{\nu - 2} - \left(\frac{3}{\nu - 2} - \left(\frac{4\nu\mathcal{C}(\nu)}{\nu - 1}\right)^2 \right) \rho^2 \right], \qquad \nu > 2 \end{aligned} \tag{C17}$$

C.1 Score derivations

Consider the log-likelihood function

$$\begin{aligned} \ell_t(y_t|\theta, Y_{t-1}) &= \log \mathcal{C}(\eta) - \frac{1}{2}\log\sigma^2 - \frac{1+\eta}{2\eta} \log\left[1 + \frac{\eta\varepsilon_t^2}{(1 - sgn(\varepsilon_t)\rho)^2\sigma^2}\right], \\ \log\mathcal{C}(\eta) &= \log\Gamma\left(\frac{\eta+1}{2\eta}\right) - \log\Gamma\left(\frac{1}{2\eta}\right) - \frac{1}{2}\log\left(\frac{1}{\eta}\right) - \frac{1}{2}\log\pi, \end{aligned} \tag{C18}$$

where $\Gamma(\cdot)$ is the Gamma function, $sgn(.)$ is the sign function, and $\eta = 1/\nu$ is the inverse of the degrees of freedom.

Differentiating (C18) with respect to location, scale and asymmetry we obtain the gradient vector $\nabla_t = \left[\frac{\partial \ell_t}{\partial\mu}, \frac{\partial \ell_t}{\partial\sigma^2}, \frac{\partial \ell_t}{\partial\rho}\right]'$. Recall that $\varepsilon_t = y_t - \mu$, $\zeta_t = \frac{\varepsilon_t}{\sigma}$ and $\omega_t = \frac{(1+\eta)}{(1-sgn(\varepsilon_t)\rho)^2 + \eta\zeta_t^2}$, and let

$$\begin{aligned} f(\mu, \sigma^2, \rho) &= 1 + \frac{\eta\varepsilon_t^2}{(1 - sgn(\varepsilon_t)\rho)^2\sigma^2} \\ &= \frac{(1 - sgn(\varepsilon_t)\rho)^2)\sigma^2 + \eta\varepsilon_t^2}{(1 - sgn(\varepsilon_t^2)\rho)^2\sigma^2}. \end{aligned} \tag{C19}$$

To avoid overburdening the notation, in what follows $\frac{\partial f(x)}{\partial x} = f'_x$ and $a = -\frac{1+\eta}{2\eta}$.

The score with respect to the location parameter reads

$$\frac{\partial \ell_t}{\partial \mu_t} = w_t \frac{\zeta_t}{\sigma}.$$

Proof. Define

$$g(\mu) = a \log f(\mu, \sigma^2, \rho),$$

such that $\frac{\partial \ell_t}{\partial \mu} = \frac{\partial g(\mu)}{\partial \mu} = a \frac{f'_\mu}{f(\mu,\sigma^2,\rho)}$. For

$$f'_\mu = \frac{2\eta}{(1 - sgn(\varepsilon_t)\rho)^2 \sigma^2} \varepsilon_t,$$

it follows:

$$\begin{aligned}
\frac{\partial \ell_t}{\partial \mu} &= -\frac{1+\eta}{2\eta} \frac{2\eta}{(1 - sgn(\varepsilon_t)\rho)^2 \sigma^2} \cdot \varepsilon_t \cdot \frac{(1 - sgn(\varepsilon_t)\rho)^2 \sigma^2}{(1 - sgn(\varepsilon_t)\rho)^2 \sigma^2 + \eta \varepsilon_t^2} \\
&= \frac{(1+\eta)}{(1 - sgn(\varepsilon_t)\rho)^2 \sigma^2 + \eta \varepsilon_t^2} \varepsilon_t \\
&= \omega_t \frac{\zeta_t}{\sigma}
\end{aligned}$$

. □

The score with respect to the squared scale parameter reads

$$\frac{\partial \ell_t}{\partial \sigma_t^2} = \frac{(w_t \zeta_t^2 - 1)}{2\sigma_t^2}.$$

Proof. Define

$$g(\sigma^2) = -\frac{\log \sigma^2}{2} + a \log f(\mu, \sigma^2, \rho),$$

such that $\frac{\partial \ell_t}{\partial \sigma^2} = \frac{\partial g(\sigma^2)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + b \frac{f'_{\sigma^2}}{f(\mu,\sigma^2,\rho)}$.
For

$$f'_{\sigma^2} = -\frac{\eta \varepsilon_t^2}{(1 - sgn(\varepsilon_t)\rho)^2 \sigma_t^4},$$

it follows:

$$\begin{aligned}
\frac{\partial \ell_t}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} - \frac{\eta+1}{2\eta} \cdot \left[-\frac{\eta \varepsilon_t^2}{(1 - sgn(\varepsilon_t)\rho)^2 \sigma_t^4} \cdot \frac{(1 - sgn(\varepsilon_t^2)\rho)^2 \sigma^2}{(1 - sgn(\varepsilon_t)\rho)^2)\sigma^2 + \eta \varepsilon_t^2} \right] \\
&= \frac{1}{2\sigma^2} \left(\frac{(\eta+1)\varepsilon_t^2}{(1 - sgn(\varepsilon_t^2)\rho)^2 \sigma^2 + \eta \varepsilon_t^2} - 1 \right) \\
&= \frac{(w_t \zeta_t^2 - 1)}{2\sigma^2}
\end{aligned}$$

. □

The score with respect to the shape parameter reads

$$\frac{\partial \ell_t}{\partial \rho_t} = -\frac{sgn(\varepsilon_t)}{(1 - sgn(\varepsilon_t)\rho_t)} w_t \zeta_t^2.$$

Proof. Define

$$g(\rho) = a \log f(\mu, \sigma^2, \rho),$$

such that $\frac{\partial \ell_t}{\partial \rho} = \frac{\partial g(\rho)}{\partial \sigma^2} = a \frac{f'_\rho}{f(\mu,\sigma^2,\rho)}$.
For

$$f'_\rho = \frac{2(sgn(\varepsilon_t) - \rho)\eta\varepsilon_t^2}{(1 - sgn(\varepsilon_t)\rho)^4\sigma^2},$$

it follows:

$$\begin{aligned}
\frac{\partial \ell_t}{\partial \rho} &= -\frac{\eta + 1}{2\eta} \cdot \frac{2(sgn(\varepsilon_t) - \rho)\eta\varepsilon_t^2}{(1 - sgn(\varepsilon_t)\rho)^4\sigma^2} \cdot \frac{(1 - sgn(\varepsilon_t^2)\rho)^2\sigma^2}{(1 - sgn(\varepsilon_t)\rho)^2)\sigma^2 + \eta\varepsilon_t^2} \\
&= \frac{\eta + 1}{(1 - sgn(\varepsilon_t)\rho)^2) + \eta\zeta_t^2} \cdot \frac{(sgn(\varepsilon_t) - \rho)\zeta_t^2}{(1 - sgn(\varepsilon_t)\rho)^2} \\
&= -\frac{sgn(\varepsilon_t)}{(1 - sgn(\varepsilon_t)\rho)} \omega_t \zeta_t^2
\end{aligned}$$

. □

C.2 Transformed parameters' scores

Here we provide a proof of the equivalence between the the score vector arising from the restriction imposed on the scale and shape parameters. To ensure positive scale and bounded shape, we model $\gamma = \log \sigma$, $\delta = atanh(\rho)$.

Proof. Let consider $\gamma = \log \sigma$ being the log-scale, it follows that $\sigma = \exp \gamma$, and the gradient is

$$\frac{\partial \ell}{\partial \gamma} = \frac{\partial \ell}{\partial \sigma^2}\frac{\partial \sigma^2}{\partial \gamma} = \frac{\partial \ell}{\partial \sigma^2} 2\sigma^2.$$

□

Proof. For the shape parameter $\rho = \tanh \delta$, we model $\delta = atanh\rho$ and the gradient is

$$\frac{\partial \ell}{\partial \delta} = \frac{\partial \ell}{\partial \rho}\frac{\partial \rho}{\partial \delta} = \frac{\partial \ell}{\partial \rho}(1 - \rho^2).$$

□

Therefore, the score vector with respect to μ, γ, and δ will be equal to:

$$\begin{bmatrix} s_{\mu,t} \\ s_{\gamma,t} \\ s_{\delta,t} \end{bmatrix} = \sqrt{\frac{(1+3\eta)}{(1+\eta)}} \begin{bmatrix} \sqrt{(1-\rho^2)} w_t \zeta_t \\ \sqrt{\frac{(1+\eta)}{2}} (w_t \zeta_t^2 - 1) \\ -\sqrt{\frac{(1-\rho^2)}{3}} \frac{sgn(\varepsilon_t)}{(1-sgn(\varepsilon_t)\rho)} w_t \zeta_t^2 \end{bmatrix}, \tag{C20}$$

where

$$w_t = \frac{(1+\eta)}{(1-sgn(\varepsilon_t)\rho)^2 + \eta\zeta_t^2}, \quad \zeta_t = \frac{\varepsilon_t}{\sigma}. \tag{C21}$$

D Returns aggregation and the realised moments

Barroso & Santa-Clara (2015) and Daniel & Moskowitz (2016) study the properties of momentum strategies, and exploit them to implement "henhanced" strategies robust to crashes. The former use daily momentum strategy returns to compute 6-months realized volatility measures to implement a volatility-managed momentum strategy. This approach increases the profitability of the strategy, as measured by the Sharpe Ratio (SR), reduces portfolio's kurtosis and rises the skewness. Daniel & Moskowitz (2016) target a maximum SR strategy, where the weights are given by $\frac{\mu_t}{\lambda\sigma_t^2}$, where λ is a constant aimed a achieving a specific volatility target, μ_t is obtained as the fitted values of a regression of the Winners-Minus-Losers (WML) portfolio returns on market risk (proxied by 6-months realized volatility) in bear market states. The volatility estimate, σ_t^2, is the fitted value of a regression of 22-days WML realized volatility on 6-months realized volatility and daily GJR volatility (Glosten et al. 1993). Therefore, σ_t^2 is a measure of volatility net of a long-term component and leverage effect, that proxies, to some extent, return skewness. The authors show that a momentum strategy based on these signals outperforms volatility-managed momentum portfolios, both in- and out-of-sample.

A common approach taken by both Barroso & Santa-Clara (2015) and Daniel & Moskowitz (2016) is to recover measures of monthly returns risk using realized volatility measures. These measure are then used in monthly-level analyses. Nevertheless, momentum returns sampled at the daily frequency might be troublesome when aggregated to the monthly frequency, or if used to compute realized moments. While the two separate legs of the WML strategy can be aggregated relying on standard techniques, a problem arises when dealing with WML returns.

Consider the simple net excess returns for past winners, $X_{i,t}$, and past losers, $Y_{i,t}$, sampled in day i and month t; assume that in one month we observe D daily returns. The 1-month holding period (HP) returns can be computed as:

$$X_t = \prod_{i=1}^{D} \left(1 + X_{i,t}\right) - 1; \quad Y_t = \prod_{i=1}^{D} \left(1 + Y_{i,t}\right) - 1.$$

Now, consider $Z_{i,t} = X_{i,t} - Y_{i,t}$, the long-short (LS) portfolio that buys X_t and sells Y_t.

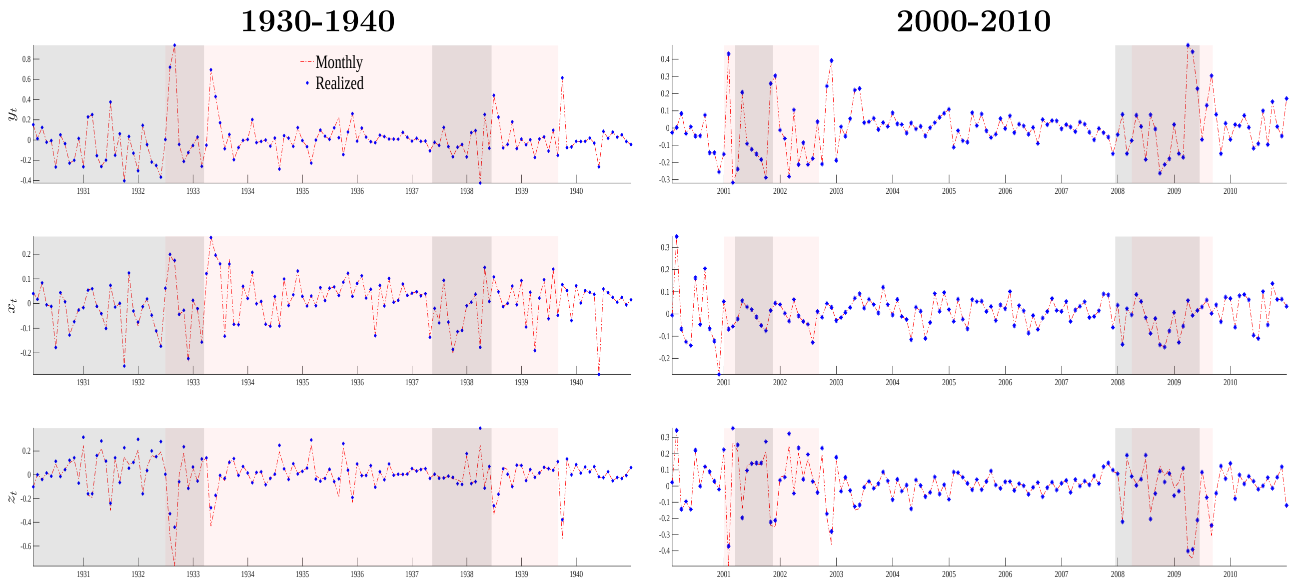


Figure D2: Different aggregation rules

Monthly returns are reported with red dash-dotted lines. Realized simple returns, that is the exponential transform of the sum of log returns within each month, correspond to the blue stars. NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods.

Following Daniel & Moskowitz (2016), the monthly return is given by

$$Z_t = \prod_{i=1}^{D} \left(1 + X_{i,t} - Y_{i,t} + \tfrac{r^f_{t-1}}{D}\right) - 1,$$

that is, at the end of each period the investments in both the long and short side of the portfolio are adjusted to set their value equal to the total value of the portfolio ($V_{t+1} = V_t(1 + X_t - Y_t + r^f_t)$). Notice that by this definition, the LS-strategy's 1-month HP return and the WML monthly return will be different as $Z_t \neq X_t - Y_t$, with differences amplified around crashes periods. Given $z_{i,t} = \log(1 + Z_{i,t})$, $z^r_t = \sum_{i=1}^{D} z_{i,t}$ are the realized monthly return. Dissecting this measure, we can rewrite:

$$z^r_t = \sum_{i=1}^{D} z_{i,t} = \sum_{i=1}^{D} \log(1 + Z_{i,t}) = \sum_{i=1}^{D} \log\left(1 + X_{i,t} - Y_{i,t} + \tfrac{r^f_{t-1}}{D}\right).$$

Differently, the monthly log-returns of the portfolio that buys X_t and sells Y_t is $\tilde{z}_t = \log(1 + \tilde{Z}_t) = \log(1 + X_t - Y_t)$. Expanding this measure we get:

$$\tilde{z}_t = \log\left(1 + \prod_{i=1}^{D}(1 + X_{i,t}) - \prod_{i=1}^{D}(1 + Y_{i,t}) + r^f_{t-1}\right).$$

The difference between the two methods lies in the order in which the cross-sectional and time aggregations are preformed. In the first case, the portfolio aggregation is performed day-by-day, and then log-returns are cumulated over time; in the second case, first we compute the monthly returns for the two legs of the strategy, and then the monthly portfolio is constructed.[19] Figure D2 illustrate the two aggregation approaches for the three portfolios. As both procedures are correct and admissible, they should deliver the same value throughout the sample, as it is the case for the two legs of the strategy (x_t and y_t). Nevertheless, as evident from the bottom panels, this equivalence does not hold for z_t during crash periods, where monthly strategy returns are systematically bigger in absolute value, therefore emphasising their severity for the longer-term investor.

Figure D3 zooms in these periods, illustrating the departure of the two methodologies as crashes occur in 1932 and 2009, for both realized returns and volatilities. Months in red represent the worst returns within the year. Months realizing the worst returns also experience the greatest distances between realized and monthly returns, with differences in 1932 being the most pronounced. Interestingly, as returns recover from the crashes, the equivalence between the two methods restores. When the log-transformation is applied to these returns, the wedge between the two measures further increases. Figure D4 illustrate how negative returns exceeding 50% worryingly decrease, possibly overstating the severity of the returns. Therefore, as illustrated in the bottom panels in Figure D3, volatilities are more sensitive to the aggregation method, which could results in a severe mischaracterization of risk, especially during crashes. Such differences would be further amplified if we consider higher-order realized moments, such as realized skewness.

As a consequence, realized volatility measures computed for the WML strategy returns may be ill-suited to study momentum strategy performances (Barroso & Santa-Clara 2015, Daniel & Moskowitz 2016). Hence, in this work we propose a parametric model which captures the dynamics of the first three moments of the conditional distribution of daily momentum returns. This methodology is robust to the use of log-transformation, and the filtering procedure we implement allows to manage outlier returns associated to momentum crashes.

[19] Eventually, the difference boils down to the difference between the product of a sum and the sum of a product.

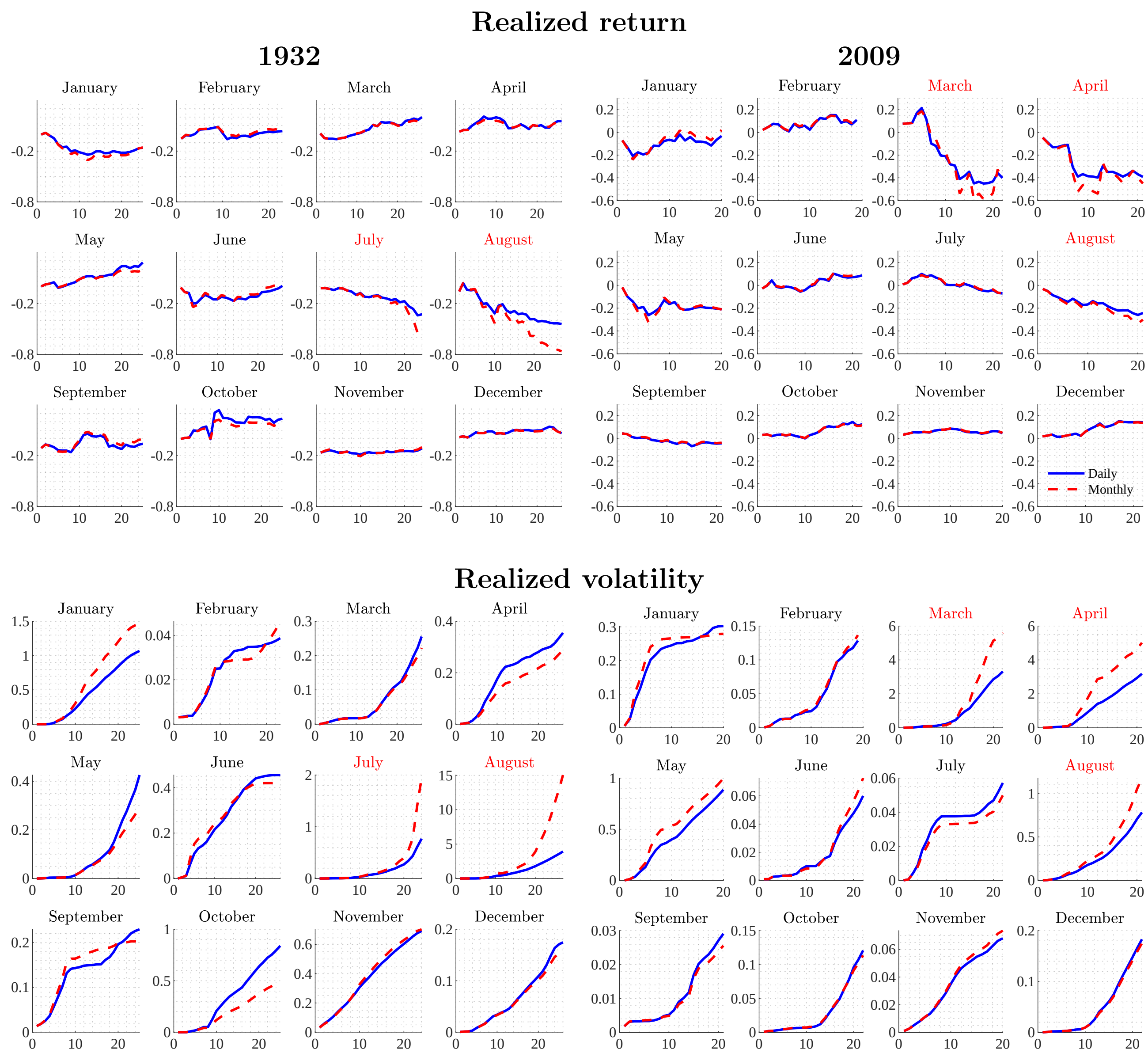


Figure D3: Aggregation during crashes

The top panels report two series of realized returns: the cumulative product of the strategy returns over each month (daily) and the difference between the cumulative products of the two legs of the strategy over time (monthly). The bottom panels report the realized volatilities associated with the realized returns.

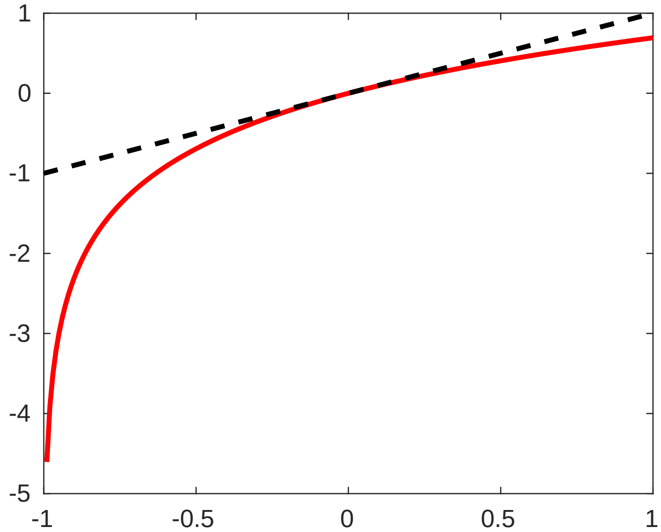


Figure D4: Logarithmic approximation