Taming Momentum Crashes

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

Abstract

The returns on US equity momentum exhibit a time-varying conditional skewness which deepens during the so-called "momentum crashes". This has important economic implications for managing the risk associated to a standard momentum factor: a dynamic skewness adjustment within a maximum Sharpe ratio strategy improves upon existing volatility-managed momentum portfolios. The importance of conditional skewness to mitigate momentum risk lies in the fact that the portfolio risk-return trade-off may reflect a non-linear interaction between *both* conditional volatility and skewness. Notably, the dynamics of the conditional skewness in momentum returns can not be fully reconciled by an asymmetric exposure to market risk.

Keywords: Momentum risk, time-varying skewness, volatility-managed portfolios, empirical asset pricing.

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

[‡]Fulcrum Asset Management. 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 the future return on a given asset and its past relative performance, termed the *momentum* factor. By betting on past returns, a momentum portfolio that buys the past "winners" and sells the past "losers" have historically delivered competitive risk-adjusted returns, and have therefore become central to the market efficiency debate.[1] Despite a strong historical performance, standard momentum portfolios can be subject to "crashes", meaning infrequent periods of large and persistent losses (see Daniel and Moskowitz, 2016). These crashes are often associated with time-varying, asymmetric, market betas; momentum portfolios formed during bear markets are likely to bet against high beta stocks, thus potentially generating large losses when the market rebounds (see Grundy and Martin, 2001).

A conventional approach to mitigate downside risk in momentum strategies builds upon the idea that the notional exposure to the winner-minus-loser (`WML`) strategy is levered up (scaled down) based on the conditional volatility of the original portfolio returns (see, e.g., Barroso and Santa-Clara, 2015; Moreira and Muir, 2017; Cederburg et al., 2020; Barroso and Detzel, 2021). This grounds on the widespread empirical evidence that volatility tends to cluster over time and negatively correlates with realised returns, whereas expected returns are often more difficult to measure being buried in the noise of the realised portfolio performance. However, while focusing on conditional volatility may certainly simplify an empirical analysis, it is based on the implicit assumption that momentum returns are conditionally Normal at each time t. That is, skewness risk is ignored.[2]

In this paper we offer some new insight on how to manage the risk associated with standard momentum investing. Our approach is based on the intuition that conditional skewness is potentially time-varying, and its statistical and economic significance is not subsumed by conditional volatility. That is, if returns asymmetry varies over time, volatility may not necessarily represent a full representation of the strategy's risk (see, e.g., Glosten et al., 1993; Wang and Yan, 2021). As a result, adjusting the notional exposure to a momentum portfolio based uniquely on the inverse of

[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 et al. (1995) find that momentum strategies are common among investment funds, while several papers document the pervasiveness of this anomaly across countries – including Rouwenhorst (1998); Fama and French (2012) – and asset classes (see, e.g., Moskowitz and Grinblatt, 1999; Moskowitz et al., 2012; Asness et al., 2013).

[2] By skewness risk we mean the risk that results when observations are not spread symmetrically around an average value, but instead have a skewed distribution. As a result, the conditional mean returns do not represent a robust measure of expected returns, i.e., mean and median differ.

the conditional volatility estimates may be sub-optimal, both in terms of risk-adjusted returns and mean-variance efficiency. Yet, the explicit effect of skewness on momentum risk has been mostly overlooked in the empirical asset pricing literature so far. Our goal is to fill this gap.

Intuitively, the reason why accounting for conditional skewness may be important to reduce large losses in a standard momentum strategy is straightforward, if not often appreciated: simple volatility targeting is based on the implicit assumption that the conditional distribution of the returns is Normal. Put it differently, the conditional probability of a large positive or negative return is symmetric at each time t. However, this assumption explicitly ignores the fact that (1) negative shifts in expected returns can occur even in periods when the volatility does not change significantly, and (2) higher volatility is not always a reflection of large losses. Therefore, standard volatility adjustments could represent, at least *a priori*, an imperfect tool to mitigate momentum risk especially during crash periods.[3]

Following this logic, some extension to simple volatility targeting has been proposed in the literature (see, e.g., Wang and Yan, 2021; Hanauer and Windmüller, 2023). These are often based on the idea that the conditional variance estimates can be replaced with semi-variance, meaning the variance calculated only based on negative returns. However, semi-volatility scaling does not necessarily represent an optimal allocation from the perspective of a mean-variance investor. Rather, it is an agnostic risk-management approach that aims to mitigate the effect of "bad volatility" in face of time-varying momentum risk. Instead, our approach represents an intuitive extension to a simple conditional maximum Sharpe ratio strategy as in Daniel and Moskowitz (2016). Our empirical results suggest that explicitly considering the dynamics of the conditional skewness provides additional economic gains compared to both volatility and semi-volatility targeting.

Before discussing the main empirical results, two comments are in order. First, it is important to note that our goal is not to modify the construction of the momentum factor per se (see, e.g., Novy-Marx and Velikov, 2016; Barroso and Detzel, 2021), but rather to build upon existing volatility adjustments as in Barroso and Santa-Clara (2015); Daniel and Moskowitz (2016); Moreira and Muir (2017); Cederburg et al. (2020), which in turn consider standard momentum factors following the Jegadeesh and Titman (1993) blueprint. Second, although our results support the view that the

[3] In other words, by assuming that the portfolio returns are symmetrically distributed around the conditional mean at each time t, when they may not be, could understate the overall risk of the strategy. This could be economically costly especially during periods of large drawdowns.

momentum premium can be rationalised based on asymmetric risk preferences (see Section 7), the objective of this paper is not to provide a structural explanation of the momentum premium, but rather to isolate the importance of higher-order moments to understand the trade-off between risks and rewards in momentum, and as a result managing the risk associated with it.

1.1 Findings

In order to tease out the dynamics of the conditional skewness in a momentum factor, we propose a time-varying parameter model which allows to recover the location, scale and asymmetry of the portfolio returns' distribution over time. In addition, our modelling framework allows to characterise the momentum expected returns as a direct function of the returns asymmetry. As a result, we can single out a skewness-hedging component within an otherwise standard maximum Sharpe ratio strategy. Notice that our framework is general and can be applied to any factor portfolio returns, above and beyond momentum. The main contribution of our paper is threefold.

First, we find a significant time variation in the conditional skewness associated with momentum returns. The returns' asymmetry is predominantly negative and tend to deepen during momentum crashes. More specifically, its time variation reflects an heterogeneous pattern on the long and short leg of the strategy. For instance, while the skewness of the past winners significantly decreased in the aftermath of the great depression, the returns' asymmetry on the past losers turned highly positive. That is, towards the end of the great depression the momentum portfolio was implicitly buying a lower downside risk, but at the same time selling a substantially higher upside risk, with the latter more than offsetting the former. As a result, the skewness of the `WML` portfolio was increasingly negative in the aftermath of the great depression.

The second main result of the paper relates to managing momentum risk. We improve upon the maximum Sharpe ratio strategy proposed by Daniel and Moskowitz (2016) by leveraging on the flexibility of our modeling framework. Specifically, we derive a skewness-hedging component which adds to a conventional volatility adjustment of the original momentum factor (see, e.g., Moreira and Muir, 2017; Cederburg et al., 2020; Barroso and Santa-Clara, 2015). For a given level of volatility, our extended maximum Sharpe ratio strategy further deleverage (leverage) the notional invested in the original momentum portfolio as the returns' skewness becomes more negative (positive). Economically, our skewness-managed strategy fare better than both benchmark dynamic or constant volatility adjustments as well as semi-volatility targeting. We show in simulation that the economic gain from our time-varying skewness approach is linked to the implicit cost of the mispecification embedded in a simple volatility targeting when in fact the conditional skewness of the returns is non-zero. The more the original portfolio returns depart from a Normal distribution at each time t, the lower the risk-adjusted returns from conventional volatility targeting compared to our skewness-managed momentum.

Our third contribution pertains the role of the conditional skewness for the risk-return trade-off in a standard momentum strategy. Specifically, we delve further into the interplay between expected returns, conditional volatility, and conditional skewness. The results suggest that the reason why the skewness adjustment is not subsumed by the conditional volatility estimates, lies in the role of returns' asymmetry for the dynamics of the strategy expected returns. For instance, a second-order Taylor expansion of the model-implied expected returns points towards a non-exclusive role of the conditional volatility as a measure of momentum risk, and instead highlights a *joint* effect of volatility and skewness on expected returns over time.

We formalise this result and leverage on our modeling framework to explicitly characterise the risk-return trade-off as a non-linear function of the original portfolio returns' asymmetry. The results show that a negative risk-return trade-off in momentum is consistent with the presence of a time-varying, predominantly negative, conditional skewness in the portfolio returns. While the premium pertaining volatility is only mildly negative, the nonlinear interaction with conditional skewness further exacerbates the slope of the risk-return profile, which becomes even more negative. This represents a challenge for common risk-based explanations and structural models of time-varying expected returns (see, e.g., Kelly et al., 2021).

The presence of a negative and time-varying risk-return trade-off may be consistent with a general equilibrium model with asymmetric preferences. We build upon the intuition of Kothari and Shanken (1992) and Grundy and Martin (2001) and show both analytically and in simulation that the time-variation in the conditional skewness of the returns can only be partly reconciled by an asymmetric exposure to aggregate market risk (CAPM). The latter, can be thought of as the reduced-form representation of an equilibrium endowment economy in which a representative agent has a disappointment-aversion utility function (see, e.g., Ang et al., 2006). Furthermore, the estimated conditional skewness only mildly correlates with the skewness on the market portfolio.

1.2 Literature

In addition to Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016), our work contributes to a long-standing literature that seeks to understand the origins and the dynamic properties of momentum risk and returns, such as Jegadeesh (1990); Rouwenhorst (1998); Moskowitz and Grinblatt (1999); Griffin et al. (2003); Moskowitz et al. (2012); Novy-Marx (2012); Asness et al. (2013); Kelly et al. (2021), among others. Jacobs et al. (2015) uncover a robust relation between expected skewness and cross-sectional momentum, specifically in relation to past loser portfolios. Nevertheless, Theodossiou and Savva (2016) highlight how the evidence on the shape of the trade-off between risks and rewards for momentum strategies has often been inconclusive. They argue that such ambiguity is primarily due to the fact that when the skewness of the returns' distribution is negative, the effect of volatility and skewness on expected returns tends to offset, generating a highly uncertain dynamics for the risk-return trade-off.

A second strand of literature we contribute to, relates to the role of higher-order moments in portfolio allocation (see, e.g, Guidolin and Timmermann, 2008) and asset pricing models (see, e.g., Dittmar, 2002; Harvey and Siddique, 2000; Kraus and Litzenberger, 1976). Within the context of momentum strategies, a variety of approaches have been proposed; for instance, Barroso and Santa-Clara (2015); Daniel and Moskowitz (2016) proposed a volatility adjustment that allows to regulate the exposure of investors' capital to momentum risk during crashes and periods of high volatility at large. Wang and Yan (2021); Hanauer and Windmüller (2023) expands on Moreira and Muir (2017); Cederburg et al. (2020) and show that by scaling by factor portfolio returns by downside volatility one can obtain significantly better performance than strategies scaled by total volatility. More recently, Barroso and Detzel (2021) investigate whether transaction costs, arbitrage risk, and short-sale impediments explain the abnormal returns of volatility-managed equity portfolios.

2 Skewness in US equity momentum

We follow Daniel and Moskowitz (2016) and form portfolios based on 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 and French (1996).[4] Momentum decile portfolios are constructed daily but

[4]The interested reader can find additional information on the original momentum strategy at `http://www.kentdaniel.net/data/momentum/mom_data.pdf`.

are rebalanced monthly. The risk free-rate is the daily 1-month T-bill rate, and the market return is the value-weighted index of all the CRPS firms.[5] 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 one-month holding period. We skip the most recent month, which is our formation period, to avoid the short-term reversal documented by Jegadeesh (1990) and Lehmann (1990).

The left panel of Figure 1 compares the cumulative performance of investing 1$ in the `WML` portfolio based on an 11-month ($J = 11$) look-back period – from month $t - 12$ through $t - 1$ – 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. The left panel of Table 1 shows the corresponding descriptive statistics. Clearly, momentum has delivered substantially higher profits 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% annualised, more than twice the 7.8% offered by the market (`MKT`). The Sharpe ratio for the momentum strategy is 0.78 on an annualised basis, which is almost double compared to the 0.42 of the market portfolio.[6]

Figure 1 also suggests that the unconditional market beta on the `WML` is also reasonably low. Table 1 shows that a simple static CAPM beta is indeed slightly negative and significant ($\beta = -0.15$, $pval = 0.000$). This couples with a highly positive and significant annualised Jensen's alpha ($\alpha = 22.2$, $pval = 0.000$). The right panels of Figure 1 show that despite its strong performance, momentum has experienced a few severe downturns. 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%. Daniel and Moskowitz (2016) define these events as "momentum crashes".

The occurrence of sporadic but persistent large negative returns induces significant asymmetry in the distribution of momentum portfolio returns. Table 1 reports both the sample skewness, defined as the standardised third moment of the sample distribution of returns, as well as the corresponding

[5]Both the daily T-bill rate and the daily return on the market portfolio are obtained from Kenneth French data library: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

[6]Notice that, for comparison with Daniel and Moskowitz (2016), we do not consider transaction costs in calculating the performance of the original momentum strategy. When adding reasonable transaction costs the performance of the standard momentum strategy deteriorates (see, e.g., Novy-Marx and Velikov, 2016; Barroso and Detzel, 2021).

p-values for the Bai and Ng (2005) significance test in parenthesis.[7] The statistics suggest that both past `winners` and the `WML` strategy all show significant, negative skewness, with the long-short strategy displaying twice the asymmetry compared to past `winners`. In addition to the realised skewness, we also report the Bowley asymmetry 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 and $q(50)$ the median. This measure is robust to the presence of outliers (see Bowley, 1926). The last row in Table 1 reports the results of $QS_\alpha,\ \alpha = 99$. Even after accounting for outliers, `WML` returns, as well as past `winners`', still display marked negative asymmetry, as twice as large the `MKT` portfolio.

Table 1 also reports descriptive statistics and skewness measures for two alternative momentum strategies which have been investigated in previous literature (see, e.g., Jegadeesh and Titman, 1993; Novy-Marx, 2012). First, we look at the "short-term" momentum (6_2), whereby decile portfolios are formed on the basis of six-month ($J = 6$) look-back periods, from months $t-6$ through $t-1$, and are rebalanced each month. Similarly to the 12_2 case, we skip the most recent month, which is our formation period. Second, we look at an "intermediate" momentum strategy (12_7), whereby decile portfolios are formed monthly based on past returns from months $t-12$ through $t-7$.

Two facts emerge. First, the profitability of the momentum strategy is more tilted towards the intermediate horizon. This is consistent with the intuition in Novy-Marx (2012). However, when the exposure to the market portfolio is netted out, the annualised alpha is largely in favour of the typical 12_2 momentum: $\alpha_{12\_2} = 22.2$ ($pval = 0.000$) versus $\alpha_{12\_7} = 16.9$ ($pval = 0.000$). The short-term momentum has both a lower Sharpe ratio and a lower alpha. Second, the unconditional skewness of the 12_7 strategy is -0.768 ($pval = 0.021$), which is substantially lower than both the 12_2 strategy ($skew = -1.236$, $pval = 0.001$) and 6_2 ($skew = -1.554$, $pval = 0.001$). The robust quantile measure QS_α suggests the discrepancy across strategies is due to the presence of different outlying returns; that is, by accounting for outliers the QS_{99} of all momentum strategies are approximately the same, i.e., -0.11 for 12_2, -0.096 for 6_2 and -0.089 for 12_7.

[7] Under the null hypothesis of no asymmetry in the returns, the Bai and Ng (2005) test statistic is

$$\widehat{\pi}_3 = \frac{\sqrt{T}\widehat{\mu}_3}{s\left(\widehat{\mu}_3\right)} \xrightarrow{d} N(0,1),$$

with $\widehat{\mu}_3$ a sample estimate of the third central moment of the returns distribution and $s\left(\widehat{\mu}_3\right) = \left(\widehat{\boldsymbol{\alpha}}_2\widehat{\Gamma}_{22}\widehat{\boldsymbol{\alpha}}_2'\right)$. Here, $\widehat{\boldsymbol{\alpha}}_2 = \left[1,\ -3\widehat{\sigma}^2\right]$ is a function of the sample variance estimate $\widehat{\sigma}^2$ and $\widehat{\Gamma}_{22}$ a consistent estimate of the 2×2 sub-matrix of $\Gamma = \lim_{T\to\infty} TE\left[\overline{Z}\overline{Z}'\right]$ with $\overline{Z}$ the sample mean of the deviation of the empirical centered first three moments from their theoretical (Gaussian) counterparts.

These sample descriptive statistics show, to a large extent, that the 12_2 strategy provides the largest risk-adjusted performance while at the same time reporting an equally large negative skewness, in fact even larger, compared to the short-term and intermediate momentum strategies. As a result, we follow Barroso and Santa-Clara (2015); Daniel and Moskowitz (2016) and focus on the 12_2 as a baseline momentum strategy to investigate the role of time-varying skewness for the dynamics of risk and return within and outside crash periods.

2.1 A simple test for conditional skewness

The presence of unconditional skewness in the returns' distribution does not mechanically imply that after accounting for time-varying volatility the conditional distribution of the returns is still asymmetric (see, e.g., Glosten et al., 1993; Carriero et al., 2020). A simple preliminary gauge of the time variation in the conditional skewness of the momentum strategy returns can be obtained by testing for the significance of realised skewness estimated recursively over smaller windows of data.

Figure 2 reports the time series of the Bai and Ng (2005) test statistics for asymmetry over different rolling windows. We report the testing results by using one, two and five years of daily returns on the past `losers` (left panel), past `winners` (middle panel) and the `WML` strategy (right panel). The dashed horizontal lines represent the 90% and 95% confidence intervals. The null hypothesis of no asymmetry is often rejected with periods of highly significant skewness. For instance, the returns on past `losers` show a significant and positive skewness towards the end of the great depression of the 1930's, whereas past returns on the `winners` portfolio tend to show a mildly significant negative skewness over the same period. As a result, the skewness of the `WML` strategy returns is significant and negative throughout the momentum crash of the early 1930s, while becomes non-significant again over the following decade.[8] In general, both legs of the momentum strategy and the resulting `WML` portfolio show a highly time-varying conditional skewness, with the `WML` strategy reporting significant spikes of negative skewness over time which tend to coincide with the momentum crashes.

In addition to the simple recursive Bai and Ng (2005) test reported in Figure 2, in Appendix A we report the results of a more robust alternative likelihood-based testing procedure. The basic idea is to assume that a given moment is constant in the data generating process and then look at

[8]Focusing on the great financial crisis of 2008/2009, the skewness of the `WML` returns tends to be negative and significant both in the aftermath of the dot-com crash and the 2008/2009 financial crisis. This is primarily due to the dynamics of the skewness on the past `winners`' portfolio, which is significant and negative around the 2008/2009 period in particular.

the information contained over time in the gradient of the log-likelihood function with respect to that moment. Table A1 reports the results. The null hypothesis of a constant skewness is strongly rejected against the alternative of time variation, with the values of test statistics that are well above 100 for both the long and the short legs of the momentum strategy as well as the `WML` portfolio.

3 Modelling time-varying skewness

We model the conditional distribution of the portfolio return at time t, r_t, as a Skew-t distribution with $\nu > 3$ degrees of freedom and time-varying location, μ_t, scale, σ_t, and shape, ρ_t, parameters (see, e.g., Arellano-Valle et al., 2005; Gómez et al., 2007),

$$r_t = \mu_t + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim Skt_\nu(0, 1, \rho_t), \qquad t = 1, \ldots, T \tag{1}$$

The shape parameter $\rho_t \in (-1, 1)$ fully characterises the asymmetry of the conditional distribution of the strategy returns. Positive (negative) values of ρ_t imply positively (negatively) skewed returns and the ratio $\frac{1+\rho_t}{1-\rho_t}$ defines the probability mass on the right versus on the left of the mode μ_t.

This modelling framework is particularly flexible since it nests standard distributional choices – with time-varying mean and variance –, as limiting cases. For instance, with both $1/\nu \to 0$ and $\rho_t = 0$, the conditional distribution collapses to a standard Normal. By restricting $\rho_t = 0$ we have the symmetric Student-t distribution, whereas with $1/\nu \to 0$ we retrieve the Skew-Normal distribution of Mudholkar and Hutson (2000). As a result, our model does not impose, but rather allows for, the presence of higher moments in the conditional distribution of the `WML` returns.

We follow Creal et al. (2013) and Harvey (2013) and model the time variation in the conditional location μ_t, scale σ_t and asymmetry ρ_t parameters in a data-driven fashion based on the recursive prediction errors. In order to ensure that the scale σ_t is positive and the asymmetry $\rho_t \in (-1, 1)$, we model $\gamma_t = log\,(\sigma_t)$ and $\delta_t = \text{arctanh}\,(\rho_t)$. The vector of time-varying parameters $f_t = (\mu_t, \gamma_t, \delta_t)'$ is updated at each time t as

$$f_{t+1} = f_t + A s_t, \qquad t = 1, \ldots, T \tag{2}$$

where A contains the parameters regulating the law of motion of the distribution parameters, and s_t

contains the likelihood information from the prediction error $\widehat{\varepsilon}_t$. Specifically, $s_t = \mathcal{S}_t \nabla_t$ is the *scaled score*, with $\nabla_t = J_t' \left[\frac{\partial \ell_t}{\partial \mu_t}, \frac{\partial \ell_t}{\partial \sigma_t^2}, \frac{\partial \ell_t}{\partial \rho_t}\right]'$ being the gradient of the log-likelihood function with respect to the non-linear transformation of the parameters, and J_t the Jacobian matrix for σ_t and ρ_t.

The scaling matrix $\mathcal{S}_t$ is proportional to the diagonal of the information matrix $\mathcal{I}_t = E\left[\nabla_t \nabla_t'\right]$, such that $\mathcal{S}_t = (J_t' \text{diag}(\mathcal{I}_t) J_t)^{-1}$. This makes the dynamics of the model parameters entirely observation-driven in the sense of Cox (1981), i.e., the dynamics of the parameters is a function of past prediction errors only. We assume that A is diagonal; hence, given the specification of the scaled score, the update of each of the time-varying parameters is proportional to the information conveyed by the likelihood with respect to that specific parameter. Blasques et al. (2015, 2022) show that score-driven updates – such as the one outlined in Equation 2 – reduce the local Kullback-Leibler divergence between the true, unobserved, conditional density of the returns and the model-implied estimate, even when the underlying model is potentially mis-specified.[9]

The updating mechanism of the time-varying parameters is based on the score vector. The latter translates new information – which is summarized by the prediction error – into an update of the parameters of the model. Given the specification in Eq.(1) and the conditional log-likelihood in Eq.(5), the elements of the score vector are defined as:

$$s_{\mu,t} = \chi(1+\rho_t^2) w_t \varepsilon_t, \quad s_{\gamma,t} = \chi(\nu+1)(w_t \varepsilon_t^2 - \sigma_t^2), \quad s_{\delta,t} = \chi \text{s}(\varepsilon_t)(1-\text{s}(\varepsilon_t)\rho_t) w_t \frac{\varepsilon_t^2}{3\sigma_t^2}, \tag{3}$$

where $\text{s}(\cdot)$ is the sign function, $\chi = \frac{(\nu+3)}{(\nu+1)}$ and

$$w_t = \frac{\nu+1}{\nu\left(1+\text{s}\left(\varepsilon_t\right)\rho_t\right)^2 + \zeta_t^2}. \tag{4}$$

represents the weights to the standardised prediction errors $\zeta_t = \frac{\varepsilon_t}{\sigma_t}$. For the interested reader, a full derivation of the information matrix $\mathcal{I}_t$, the Jacobian J_t and the elements of the scaled score vector s_t is provided in Appendix C. The updates in the parameters depend only on the standardised prediction error and its square (see, e.g., Hansen, 1994). Notably, the updates on the transformed scale and shape parameters are independent conditional on the prediction error. This means that

[9]Relatedly, Koopman et al. (2016) show that score-driven time-varying parameter models produce similar forecasting precision to parameter-driven state–space models, even if the latter constitute the true data generating process. In this respect, score-driven updates of the time-varying parameters are optimal from an information theoretic perspective.

there is no overlapping information between the conditional volatility and the conditional skewness.

As highlighted by Delle Monache and Petrella (2017), the scalar w_t plays a key role as it serves as an implicit weight of the information contained in the prediction error. For the interested reader we report in Appendix C the key properties of the parameters updating scheme. To summarise, the joint role of the conditional estimates allows for a timely detection of shifts in the shape of the conditional distribution of the returns, while at the same time discounting the effect of outlying observations. In addition, while the scores for the location and shape parameters are negatively correlated, updates of σ_t are unconditionally uncorrelated with revisions of the other parameters. Nevertheless, during crashes, when prediction errors are large and negative, updates on the scale and the shape parameters positively co-move, so that the conditional distribution of the momentum returns features negative shifts in the location, increasing dispersion and deepening negative skewness, consistent with a standard leverage effect (see, e.g., Glosten et al., 1993).

3.1 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). Arellano-Valle et al. (2005) show 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 as a combination of strictly positive densities. For the model specified in Eq.(1), we characterise the conditional log-likelihood as a two-piece distribution (see Fernández and Steel, 1998);

$$\ell_t(r_t|\theta, f_t) = \text{const} - \frac{1}{2}\log\sigma_t^2 - \frac{1+\nu}{2}\begin{cases}\log\left[1+\frac{\varepsilon_t^2}{\nu(1+sgn(\varepsilon_t)\rho_t)^2\sigma_t^2}\right], & r_t \geq \mu_t \\ \log\left[1+\frac{\varepsilon_t^2}{\nu(1-sgn(\varepsilon_t)\rho_t)^2\sigma_t^2}\right], & r_t < \mu_t\end{cases} \tag{5}$$

where $\theta = (\nu, A)$ collects the time-invariant degrees of freedom and the score loadings. Appendix C.1 provides more details. In principle, maximum likelihood estimation of the latent states f_t and static parameters θ can be achieved via a prediction error decomposition (see Blasques et al., 2022). However, given the random-walk nature of the time-varying parameters, the maximum likelihood estimator tend to put a large point mass at zero, an issue known as the "*pile-up problem*" (see, e.g., Sargan and Bhargava, 1983; Anderson and Takemura, 1986; Tanaka and Satchell, 1989; Stock and Watson, 1998). To address this issue, we discipline the parameter space by introducing a minimum set of priors on the score loadings and the degrees of freedom which, while excluding zero, are also quite uninformative, in that any evidence on time variation reflects a strong evidence in the data.

Let a_j the jth element on the diagonal of A, ν is 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. Our prior specifications for these parameters is as follows: an inverse Gamma prior $a_j \sim \mathcal{IG}(a_\kappa, b_\kappa)$ for each element in the diagonal matrix A; a truncated Gamma prior $\nu \sim \mathcal{G}(d_\nu, D_\nu) \cdot I_{(\nu \geq 3)}$ for the degrees of freedom; and a standard multivariate Gaussian $\bar{f}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{M}_0)$ for the initial values of the time-varying parameters. We choose an inverse Gamma prior for the score loadings in line with the properties of the score-driven filters (for further discussion, see Juárez and Steel, 2010; Blasques et al., 2015). We set $a_\kappa = 2$, and $b_\kappa = 1$, so that *a priori* the loadings in A are positive, with a mode of 0.3.

The hyper-parameters for the Gamma prior on the degrees of freedom ν reflect a rather uninformative prior view on the parameters, with $d_\nu = 2$ and $D_\nu = 5$. These values allow the distribution to explore a wide range of feasible values for ν, with a mean of 10 and a mode of 8 (see, e.g., Juárez and Steel, 2010).[10] The initial values of the time-varying parameters are drawn from a multivariate Gaussian distribution, with mean vector $\mathbf{m}_0$, and $\mathbf{M}_0$, both estimated over a pre-sample of data. The small time variation embedded into the prior of the latent states is a prerequisite for the optimality of the score-driven updating (see, Blasques et al., 2014). The posterior distribution is not available in closed form and is numerically evaluated based on draws from the priors and the conditional likelihood in Eq.(5). For each draw $\theta^i = \left(A^i, \nu^i, f_0^i\right) \sim \pi\left(\theta\right)$, we estimate the time-varying parameters $\left\{f_t^i | \theta^i, f_0^i\right\}_{t=0}^T$, and evaluate the log-likelihood $\ell(r|\theta^i) = \sum_{t=1}^T \ell_t(r_t|\theta^i, f_t^i)$. We approximate the corresponding posterior distribution $\pi\left(\theta^i|r\right) \propto \ell(r|\theta^i)\pi\left(\theta^i\right)$ for each draw θ^i. Then, the parameters of the model are estimated as $\theta^* = \arg\max_\theta \pi\left(\theta^i|r\right)$.

3.2 Conditional skewness and expected returns

Existing literature suggests that downside risk predicts future returns better than volatility (see, e.g., Wang and Yan, 2021). Our modelling framework directly embeds this observation and explicitly consider the effect of returns' asymmetry ρ_t, in addition to the location μ_t and scale σ_t parameters, on the dynamics of the strategy expected returns. The latter are of particular relevance for our purpose. Within the context of reduced-form empirical asset pricing models, the ex-ante conditional

[10] In order to ensure the existence of at least the first three moments, we assume $\nu > 3$.

expected returns on zero-cost long-short portfolios are often considered a proxy for the strategy risk premium (see, e.g., Gu et al., 2020).

The expected future returns conditional on the information available at time t can be derived by re-parametrising the Skew-t density in Eq.(1) as a two-piece distribution (see Gómez et al., 2007). This allows us to model the conditional moments as a weighted average of the moments of a Half-t distribution (see Arellano-Valle et al., 2005). Thus, the expected one-step ahead return at time t on the momentum strategy is defined as:

$$\mathbb{E}_t(r_{t+1}) = \mu_t + g(\nu)\rho_t\sigma_t, \qquad \nu > 3 \qquad \text{with} \qquad g(\nu) = \frac{4\nu\mathcal{C}(\nu)}{\nu - 1} \tag{6}$$

where $\mathcal{C}(\cdot)$ is a combination of Gamma functions and constants, and ν is the degrees of freedom parameter. For the interested reader, a full derivation of the moment-generating function and the corresponding expected returns from the two-piece Half-t distribution is provided in Appendix D. Based on Equation 6, the expected future returns on the momentum strategy depend on both the scale and the asymmetry ρ_t of the returns' distribution at each time t. To better understand the role of these parameters on expected returns, 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 in conjunction.

Figure 3 shows the comparative statics as a heat map. For simplicity, we assume 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 heat map, we also report the partial derivative of Equation 6 with respect to ρ_t and ν for varying values of σ_t. Two facts emerge: first, the effect of ρ_t on the expected returns is amplified by the scale σ_t. For instance, for negatively skewed returns, i.e., $\rho_t < 0$, the higher the volatility the lower is the conditional expected returns. This is consistent with the idea that high volatility with negative skewness is associated with large losses from the strategy.

The second property that emerges from Figure 3 is that the interplay between asymmetry and volatility on expected returns is multiplicative. This means that 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 (see the partial derivatives). This means that the more extreme the return observations are, the higher the sensitivity of the conditional expectations to changes in the returns' scale and asymmetry.

The right panel of Figure 3 shows the effect of ρ_t and 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 effect of a change in the shape parameter depends on ν is more visible in the plot of the partial derivative of Equation 6 with respect to ρ_t: the correlation between skewness and expected returns is mitigated by the thinner tails of the returns' distribution. Nevertheless, the conditional skewness still dictates the sign of the expected returns.

4 Time-varying skewness of momentum returns

In order to initially gauge the magnitude and the dynamic of the returns' asymmetry we look at the estimates of μ_t against the conditional mean $\mathbb{E}_t(r_{t+1})$ as per Equation 6. The conditional location parameter μ_t captures the centre of the distribution and is equivalent to the conditional mean only for models with symmetric distributional assumptions – when the returns' asymmetry $\rho_t = 0$. The left panel of Figure 4 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 time-varying expected returns in momentum strategies (see, e.g., Grundy and Martin, 2001; Kelly et al., 2021). Second, there is a major disconnect between μ_t and the conditional mean $\mathbb{E}_t(r_{t+1})$. This is particularly pronounced during the momentum crashes of 1932-1939 and 2008-2009. For instance, while the expected returns from the `WML` portfolio become largely negative in the aftermath of the great depression and the great financial crisis, the location μ_t remains largely in positive territory for both periods.[11]

The top-right panel of Figure 4 shows that the disconnection between μ_t and $\mathbb{E}_t(r_{t+1}) < \mu_t$ is particularly strong for past `winners`. The estimate of the location parameter is primarily positive, while the conditional expected returns are often negative, especially during momentum crashes. Instead, the bottom-right panel of Figure 4 shows that the returns' asymmetry is less pronounced for the past `losers` portfolio. The location and the conditional expected returns tend to align

[11] Recall that for a given σ_t, the disconnect $\mathbb{E}_t(r_{t+1}) < \mu_t$ implies that the returns' asymmetry is negative, i.e., $\rho_t < 0$ (see Eq.6).

fairly closely, with the notable exception of the 2001 and 2008 crashes where expected returns are substantially lower than μ_t. Figure 4 points towards a compounding effect between the large negative skewness in past `winners` and the relatively smaller negative skewness of past `losers`. Overall, the fact that the conditional expected returns are lower than the location of the distribution, and that such a discrepancy tends to exacerbate during momentum crashes, suggests that there is indeed significant, time-varying, pro-cyclical, asymmetry for both legs of the momentum strategy.

The left panel of Figure 5 shows the estimate of the conditional skewness for the `WML` portfolio returns. To increase readability, the figure reports the daily estimates in black, and a five-year smoothed trend in green. The dashed red horizontal line represents the sample mean of the conditional skewness. There is substantial time variation of the momentum returns' asymmetry, especially around recessions and momentum crashes. In particular, momentum returns show a mildly negative skewness ahead of crashes, which then drops quite substantially during crashes. For instance, in August 1932 momentum skewness remains below its mean value 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) portfolios. There is an interesting discrepancy between the asymmetry of the returns on past `winners` versus `losers` portfolios. For instance, the time-varying skewness for past `winners` is in deep negative territory, essentially for the whole sample. On the other hand, past `losers` tend to have a significant upside exposure towards the tail of recessions; that is, the skewness becomes positive at the latest stages of the great depression, the dot-com bubble and the great financial crisis. As a result, the conditional skewness of the momentum strategy tends to become more negative during the tail of recessions and throughout momentum crashes: the `WML` portfolio implicitly is "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.

Figure 6 delves into the two major momentum crash periods of 1932–1939 (top panels) and 2001–2009 (bottom panels), as indicated by Daniel and Moskowitz (2016). The left panels shows

[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 (see Harvey and Luati, 2014; Delle Monache et al., 2021).

the conditional skewness for the `WML` portfolio. To a large extent, consistent with the estimates in Figure 4, 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`' portfolio, which represents the short leg of the momentum strategy. Interestingly, 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 `WML` portfolio. Both legs of the momentum strategy turn out to have negative skewness during the 2001 momentum crash, which results in a relatively mild risk on the downside for the `WML` strategy.

5 Skewness-managed momentum

A list of papers have attempted to improve the profitability of a standard momentum strategy as originally proposed by Jegadeesh and Titman (1993). For instance, Grundy and Martin (2001) propose to hedge the exposure to market risk to attenuate the effect of momentum crashes. Other approaches attempt to time the volatility associated with momentum returns (see Barroso and Santa-Clara, 2015, Daniel and Moskowitz, 2016 and Moreira and Muir, 2017).

Within this setting, during periods of higher (lower) volatility – relative to the unconditional mean volatility – the notional capital exposure to the to the `WML` portfolio is reduced (increased) by an amount proportional to the inverse of the previous month's realised variance. In practice, this should improve the risk-adjusted returns on the momentum strategy by leveraging on a "stop-loss" rule based on conditional volatility estimates. Based on a similar logic, Daniel and Moskowitz (2016) put forward a dynamic strategy aimed at maximising the conditional Sharpe ratio with weights:

$$w_t = \frac{1}{2\gamma}\frac{\mathbb{E}_t(r_{t+1})}{\mathbb{V}_t(r_{t+1})} \tag{7}$$

where γ is a constant calibrated to match the unconditional volatility of the strategy. In the original framework of Daniel and Moskowitz (2016), the conditional expectation $\mathbb{E}_t(r_{t+1})$ is assumed positive and constant during bull markets, while it varies during bear market periods. In addition, the conditional variance $\mathbb{V}_t(r_{t+1})$ is estimated as a weighted average of asymmetric GARCH variance (Glosten et al., 1993), and a six-month realised variance. We expand on Daniel and Moskowitz

(2016) and exploit the flexibility of our modeling framework to decompose the optimal Sharpe ratio weights in two parts:

$$w_t = \frac{1}{2\gamma}\frac{\mathbb{E}_t(r_{t+1})}{\mathbb{V}_t(r_{t+1})} = \frac{1}{2\gamma}\frac{\mu_t + g(\nu)\rho_t\sigma_t}{\mathbb{V}_t(r_{t+1})} = \underbrace{\frac{1}{2\gamma}\frac{\mu_t}{\mathbb{V}_t(r_{t+1})}}_{w_{1,t}} + \underbrace{\frac{1}{2\gamma}\frac{g(\nu)\sigma_t\rho_t}{\mathbb{V}_t(r_{t+1})}}_{w_{2,t}}. \tag{8}$$

where the conditional variance of the momentum returns can be derived analytically based on the moment generating function (see Appendix D) as,

$$\mathbb{V}_t(r_{t+1}) = \sigma_t^2\left(\frac{\nu}{\nu-2} + h(\nu)\rho_t^2\right), \qquad \nu > 2, \tag{9}$$

with $h(\nu) = \frac{3}{\nu-2} - g(\nu)^2 \gg 0$ gauging the interaction between the fat-tailedness of the distribution via ν, and the asymmetry parameter ρ_t.[13] The first component $w_{1,t}$ in Eq.(8) is akin to the maximum conditional Sharpe ratio adjustment proposed by Daniel and Moskowitz (2016) under the assumption of symmetric returns, i.e., $\rho_t = 0$, $\mu_t = \mathbb{E}_t(r_{t+1})$ and $w_{2,t} = 0$. The second component $w_{2,t}$ captures the capital adjustment due to time-varying momentum returns' asymmetry: for a given level of conditional variance, the additional leverage component $w_{2,t}$ becomes more negative (positive) as the asymmetry ρ_t becomes more negative (positive).[14] As a result, during periods of high negative (positive) skewness, our skewness-adjusted momentum portfolio w_t unwinds (leverage up) the investment in the original `WML` factor more than a dynamic volatility targeting does. Put it differently, the $w_{2,t}$ component can be interpreted as skewness hedging component within the context of an otherwise standard maximum Sharpe ratio strategy.

Before discussing the empirical results 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 and Sentana, 2009). This is typically implemented by including higher-order moments in the agents' utility maximisation and asset allocation problem. Differently, our modelling framework focuses on eliciting the effect of conditional skewness in adjusting the capital exposure to an otherwise standard risk-managed momentum strategy. The main advantage is that we can still gauge the benefit of explicitly modelling time-varying returns' asymmetry within the context of the benchmark maximum Sharpe ratio approach, as in Barroso and Santa-Clara (2015)

[13]Notice that for $\rho_t = 0$, $\mathbb{V}_t(r_{t+1})$ reduces to the Student-t variance, $\sigma_t^2\frac{\nu}{\nu-2}$.

[14]Remember that the asymmetry parameter has a sizable effect on the mean of the returns, whereas it has little impact on their variance since enter squared.

and Daniel and Moskowitz (2016). In fact, compared to a model where $\rho = 0$, our specification captures returns' underlying asymmetry via the first two moments. For instance, Equation 6 shows that a more negative ρ_t implies a lower conditional mean for a given value of the scale σ_t.

5.1 Out-of-sample managed momentum returns

We now compare our skewness-managed strategy – the maximum skewed-Sharpe ratio (`mSSR`) – against a variety of alternative volatility-targeting approaches. First we consider two benchmark approaches based on a conditional maximum Sharpe ratio as in Daniel and Moskowitz (2016) (`mSR`), and constant volatility targeting as in Barroso and Santa-Clara (2015) (`cVol`). These are based on the estimates of $\mathbb{E}_t(r_{t+1}), \mathbb{V}_t(r_{t+1})$ under the restriction $\rho_t = 0$.[15] In addition, for the sake of comparability with our model-implied managed portfolios, we also compare the same exact implementation of Daniel and Moskowitz (2016) (`DM2016`) and Barroso and Santa-Clara (2015) (`BS2015`) which are based on recursive estimates of the realised variance.[16] Finally, we also compare out skewness-adjusted momentum strategy against a constant volatility targeting *with* skewness (`cSVol`) and realised semi-volatility as in Wang and Yan (2021); Hanauer and Windmüller (2023) (`cdVol`).

For each strategy implementation the structural parameters of the model, with or without time-varying skewness, are re-estimated each month; that is, the time-varying parameters μ_t, σ_t and ρ_t are extracted on a daily basis assuming that the structural parameters of the score-driven transition remain constant within a given month. The reason is primarily to reduce the computational cost of the real-time estimation procedure. The random walk dynamics of the parameters in f_t implies that are in general very stable and therefore re-estimating the model daily likely has virtually no impact on the empirical estimates. The initial forecast and portfolio choice is generated in January 1st 1930 – that is, we use three years of daily returns as an initial burn-in sample for the recursive forecasts. For the monthly implementation, we re-scale the model-implied conditional mean and variance from daily to monthly.

[15] To tease out the effect of conditional asymmetry vs volatility alone, the daily returns on the strategies that overlook skewness risk are computed by re-estimating the model constraining $\rho_t = 0$. More specifically, we estimate the model restricting the parameter to be $\rho_t = 0, t = 1, \ldots$, while the conditional volatility evolves over time based on the score-driven dynamics.

[16] Barroso and Santa-Clara (2015) use daily momentum strategy returns to compute six-months realised volatility, rv_t^{126}, measures to implement a volatility-managed momentum strategy. Instead, Daniel and Moskowitz (2016) target a maximum Sharpe ratio strategy, where the mean signal, μ_t, is obtained as the fitted values of a regression of the `WML` portfolio returns on market risk – proxied by rv_t^{126}) in bear market states. The conditional volatility estimate, σ_t^2, is the fitted value of a regression of 22-days `WML` realised volatility, rv_t^{22}, on rv_t^{126} and daily asymmetric GARCH volatility as in (Glosten et al., 1993).

We evaluate the portfolios by means of several distinct sets of 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 test for the significance of the improvements in the SRs compared to the `WML` original factor based on the bootstrap procedure proposed by Ledoit and Wolf (2008). We also compare the portfolios by means of a second measure, namely the Sortino ratio (see, e.g., Sortino and Van Der Meer, 1991). The Sortino ratio entails a penalisation only for returns that fall below a certain threshold. The threshold, commonly referred to as the *minimum accepted return* (MAR), is generally set at the risk-free rate. The denominator of this ratio is `dVol`$= Vol(\tilde{R}_i|\tilde{R}_i < 0)$, with $\tilde{R}_i$ being the excess return of strategy i. To account for downside risk, we also report a series of additional performance measures, such as the maximum drawdown (henceforth MaxDD), the sample skewness, the Value-at-Risk (henceforth VaR), and the Expected Shortfall (henceforth ES).

Panel A of Table (2) reports the results for the daily returns. The Sharpe ratio of the skew-managed strategy, `mSSR`, is 1.57 ($pval = 0.000$) in annualised term. This is about twice as large as the SR delivered by the original momentum factor. Consistent with the existing literature, the dynamic adjustment of Daniel and Moskowitz (2016) and the constant volatility targeting of Barroso and Santa-Clara (2015) do improve in risk-adjusted terms versus the `WML` portfolio, with a SR of 1.37 and 1.26, respectively ($pval = 0.001$). Scaling by semi-volatility and accounting for skewness in the constant volatility targeting strategy also marginally improves upon both `DM2016` and `BS2015`. Nevertheless, our skewness-adjusted maximum Sharpe ratio generates the highest SR of 1.57 annualised ($pval = 0.001$). The same pattern holds for the Sortino ratio. For instance, our `mSSR` strategy produces a 2.5 Sortino ratio against 2 from the `cSVol`, `cVol`, `cdVol`, and `DM2016`. By comparison, the original `WML` factor has a Sortino ratio of 1.

Higher Sharpe and Sortino ratios do not translate in higher MaxDD, ES, or more negative skewness. In fact, the opposite holds. Interestingly, both `DM2016` and `BS2015` do compare favourably against the semi-volatility scaling when it comes to `MaxDD`. In addition, with the exception of our `mSSR` adjustment, the maximum Sharpe ratio adjustment of Daniel and Moskowitz (2016) produces a lower VaR compared to other competing volatility targeting strategies. To a large extent, the Sortino ratios, the MaxDD and the VaR suggest even bigger gains associated with our skew-managed strategy: risk-adjusted returns are not only higher, but may be less exposed to downside risk.

We formally test this assumption by reporting in Table (2) two additional downside risk performance measures across portfolios. Specifically, we compute the Stable Tail Adjusted Return Ratio (henceforth STARR), which replaces volatility with ES as denominator in the Sharpe ratio, and the Rachev ratio (henceforth RR), which is calculated as the ratio between the Expected Longrise over the ES (see, e.g. Fabozzi et al., 2005).[17] Our skewness-adjusted maximum Sharpe ratio portfolios outperform all competing methods, with a STARR (RR) of 14.2 (1.3) against 11.4 and 10.1 (1.1 and 0.97) obtained from Daniel and Moskowitz (2016) and Barroso and Santa-Clara (2015), respectively.

Panel B of Table (2) reports the results for the monthly portfolio implementation. The performance measures on a monthly basis are smaller across the board. This applies to any risk-managed momentum strategy as well as the original `WML` portfolio. Both the Daniel and Moskowitz (2016) and Barroso and Santa-Clara (2015) volatility-managed portfolios improve upon the original momentum strategy. The higher Sharpe ratio and the smaller maximum drawdown is in line with the existing evidence. Notably, while it performs on par in risk-adjusted terms to other competing strategies, our baseline skewness-managed momentum `mSSR` outperforms all competing volatility targeting methods in terms of realised downside risk. For instance, our skewness-adjusted strategy produces the lowest semi-volatility, the highest skewness, the lowest VaR, the lowest ES and the highest STARR measure.

5.1.1 The role of skewness hedging. Panel C in Table (2) disentangle the effect of skewness by separating the performance of the two components $w_{1,t}$ and $w_{2,t}$ separately. We report the same set of statistics as in the main daily portfolio results. The results show some interesting insight into the origins of the performance of skewness adjusting. For instance $w_{1,t}$ in isolation delivers substantially higher SR and Sortino ratios compared to $w_{2,t}$ on itself. This reflects in a marginally lower risk-adjusted performance of w_t compared to $w_{1,t}$.

On the other hand, the returns from the skewness hedging component show a lower VaR, MaxDD, ES and semi-volatility. This result supports the idea that within the context of our model, skewness hedging represents an insurance component in the portfolio construction; that is, a representative investor is willing to give up part of the risk premium to mitigate the effect of low-probability large-impact losses. This is consistent with the fact that investors dislike negative returns and weight losses more than gains (see Kraus and Litzenberger, 1976; Kahneman and Tversky, 2013).

[17]The Expected Longrise is simply the opposite of the ES. It is the expected valued of the area above the 95th percentile of the distribution. The Rachev ratio captures the asymmetry of the returns distribution based on the imbalance between extreme losses vs gains.

5.1.2 Transaction costs and risk aversion.

The last column in Panel A and B in Table 2 shows the average scaling w_t implied by our `mSSR` strategy as well as all competing methods. The results show that our skewness-based scaling implies, on average, a more conservative leverage of the notional invested in the original `WML` factor. Indeed, the average weight w_t is 0.57 for the `mSSR` against 1.6 and 0.97 from the `cdVol` and `DM2016`, respectively. Interestingly, by scaling for a long-term estimates of the realised variance the amount of leverage implied by volatility targeting is substantially reduced. For instance, scaling by the six-month realised variance as in `BS2015` produces the lowest average leverage. This is consistent with some of the existing evidence in the literature, such as Barroso and Detzel (2021) and Bernardi et al. (2022).

We now investigate the implications for turnover and leveraging by evaluating the portfolio performances net of transaction costs and performance fees. Specifically, 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`, `cSvol`, `DM2016`, `BS2015`. Rebalancing the portfolios each day requires adjusting the position in momentum returns by the amount $|w_{i,t+1} - w_{i,t}|$ and cost c. Following DeMiguel et al. (2009), we define the evolution of wealth for strategy i as $W_{i,t+1} = W_{i,t}\mathcal{R}_{i,t}\,(1 - c|w_{i,t+1} - w_{i,t}|)$, 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 et al. (2003) and evaluate the maximum performance fee an investor with constant relative risk aversion (CRRA) utility function would be willing to pay to access the signal from a model. For any pair (i, j) of strategies, the fee $\mathcal{F}$ arises as the solution of,

$$\sum_{t=0}^{T-1} \frac{(\mathcal{R}_{i,t} - \mathcal{F})^{(1-\delta)}}{1-\delta} = \sum_{t=0}^{T-1} \frac{\mathcal{R}_{j,t}^{(1-\delta)}}{1-\delta}, \tag{10}$$

where δ is the degree of relative risk aversion. Finally, as a third measure to further understand the outperformance of our strategy, we also consider a measure of abnormal return measure, $d\mathcal{A}$, as in Modigliani and Modigliani (1997). For any pair of strategies (i, j), we leverage up or down strategy i so as to match the downside-risk profile of strategy j, and we evaluate the annualized, risk-adjusted abnormal returns as follows $d\mathcal{A}_{i,j} = dVol_i(Sortino_i - Sortino_j)$.[18]

[18] Notice that our goal is not to propose an actual trading strategy that can be implemented "off the shelf", but rather to show the economic value of expanding the notion of risk to the third moment within the context of managed momentum portfolios. In other words, although simplistic, we believe that considering different types of cost measures could help to shed further light on the incremental value of controlling for time-varying skewness when constructing managed momentum portfolios.

Table 3 reports the results. We compute both the performance fees and abnormal returns with respect to the unmanaged `WML` strategy. We consider several level of transaction costs, ranging from 0 to 14 bps (see Moreira and Muir, 2017); for the performance fees we set $\delta = 5$ (see, e.g., Rapach et al., 2010; Rapach and Zhou, 2013; Pettenuzzo et al., 2014; Gu et al., 2020; Bianchi et al., 2021 and the references therein). Overall, relative to the simple `WML` strategy, a skewness-managed portfolio realises higher annualised average returns net of notional transaction costs. In addition, such a strategy commands higher performance fees to be access by a CRRA investor. The performance fee of our model remains relatively large compared to the original `WML` factor for several values of transaction costs. For instance, for 14 basis points of notional costs the abnormal performance – relatively to `WML` – of the `mSSR` is 13%, against 11.9% and 9.8% annualised from the `DM2016` and `BS2015`, respectively.

In addition to different transaction costs, we also consider different levels of risk aversion. We repeat the economic evaluation of Table 3 controlling for different levels of risk aversion of $\delta = 2, 7, 15$. For the ease of readability, we report the results in B and discuss the results in the main text. Table B1 shows that explicitly considering time-varying skewness for modelling momentum risk become even more valuable for an investor that is more risk averse.

5.2 Simulation-based evidence of the role of skewness

The results in Table 2 highlight the economic value of skewness hedging compared to more traditional volatility targeting in momentum portfolios. In this section, we provide some practical justification on the origins of such economic gain based on a simple simulation exercise whereby the returns are generated based on different distributional assumptions on the residual term $\varepsilon_t \sim \mathcal{D}$ in a data generating process of the form $y_t = \mu + \sigma_t \varepsilon_t$. Our goal is to show that when the conditional distribution of the returns departs from a Normal, a simple time-varying volatility scaling to mitigate risk is sub-optimal, both from a statistical and economic perspective. Our intuition is that the more misplaced is the assumption of Normality compared to the true data generating process, the larger the cost that will be paid by a mean-variance investor in risk adjusted terms.

We consider five different assumptions for the distribution $\mathcal{D}$: Gaussian ($\mathcal{N}$), Skew-Gaussian ($Skew{-}\mathcal{N}$), Student-t (t), and asymmetric Student-t ($Skew{-}t$) with constant skewness and an asymmetric Student-t with time-varying asymmetry ($Skew_{tv}{-}t$). For simplicity we consider all

Student-t distributions with 5 degrees of freedom and calibrate the unconditional mean μ with the sample mean of the historical `WML` returns. Time-varying volatility σ_t is simulated from a stationary GARCH(1,1) model calibrated based on historical `WML` returns. For each distribution assumption, we simulate 100 paths of 10,000 observations. The $Skew{-}\mathcal{N}$ and $Skew{-}t$ distributions feature -0.12 and -0.3 unconditional skewness, respectively.

Figure 7 shows the simulation results. The panels show the distribution of the Sharpe ratio (a), Sortino ratio (b), and expected shortfall (c), for the simulated data (blue) and three portfolios: constant volatility (`cVol`, red), maximum Sharpe ratio (`mSR`, yellow) and maximum Sharpe ratio with time-varying skewness (`mSSR`, purple). Constant volatility portfolio returns are obtained by standardizing the simulated returns by GARCH(1,1) volatility estimates, while maximum Sharpe ratio strategies are obtained as per Section 5; we target an annualized volatility of 18% across all strategies consistent with the market portfolio returns (see Table 1).

Two interesting facts emerge. First, constant volatility targeting does indeed improve upon the original returns when the unconditional distribution of the returns is non-normal. This is reflected in a higher Sharpe and Sortino ratios when the returns distribution is fat-tailed or negatively skewed, unconditionally. Furthermore, when targeting is based not only on the conditional variance but also on the conditional mean, i.e., maximum Sharpe ratio, the loss-mitigation effect is larger. That is, the `mSR` strategy performs almost on par in risk-adjusted terms to our time-varying skewness adjustment `mSSR`. Second, and perhaps more interestingly, when the skewness of the returns distribution is time varying the gap between our `mSSR` and a more conventional maximum Sharpe ratio portfolio `mSR` becomes large and positive. This suggests that the economic gain from our skewness managed momentum strategy lies in the ability of our modeling framework to capture the dynamics of the conditional skewness.

More generally, the results reported in Figure 7 highlight the costs of overlooking higher order moments in the conditional distribution of the returns, when they are present in the data. Specifically, even when returns do not feature skewness, considering robust volatility estimates tend to improve the upside risk of the returns. When data feature time-varying skewness, a maximum Sharpe ratio strategy that explicitly accounts for it delivers returns with meaningful positive skewness. When skewness and heavy-tails are explicitly modelled and they are featured in the data, the economic gains are even more pronounced.

6 Skewness and momentum risk

A well documented result in the momentum literature is that momentum returns have time-varying volatility (see Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016). This is an intuitive finding because, as shown by Grundy and Martin (2001), momentum investing has a time-varying exposure to market risk, whereby outside of bear markets winners are low-beta stocks and losers have high betas. Following this intuition, existing research, such as Barroso and Santa-Clara (2015); Daniel and Moskowitz (2016); Moreira and Muir (2017); Cederburg et al. (2020), argue that adjusting the capital exposure to momentum based on its conditional volatility would suffice to hedge momentum risk in real time. This contrasts with our finding that the risk of momentum is also significantly linked to conditional skewness (see Section 4), and taking that into account offers substantial economic gains (see Section 5).

However, why does skewness matter beyond conditional volatility? We conjecture that this is because the dynamics of the conditional skewness has first-order implications for the evolution of the strategy risk-return trade-off over time. Equation 6 shows that, within our model, we can distinguish two components for the variation in the conditional expected returns $\mathbb{E}_t(r_{t+1})$: the location parameter, μ_t, and a nonlinear function of higher-order moments, $hom_t = g(\nu)\sigma_t\rho_t$. A negative (positive) ρ_t amplifies the negative (positive) effect of a large value of the scale parameter σ_t (see Section 3.2 for more details). In order to isolate the effect of returns' asymmetry ρ_t on expected returns, we perform a second-order Taylor expansion of Eq. (6) around the mean values $\overline{f} = (\overline{\mu}, \overline{\sigma}, \overline{\rho})'$. Specifically, let $\tilde{f}_t = f_t - \overline{f}$, the expected returns on the momentum strategy can be decomposed as:

$$\mathbb{E}_t(r_{t+1}) = \text{constant} + \tilde{\mu}_t + g(\nu)\overline{\rho}\tilde{\sigma}_t + g(\nu)\overline{\sigma}\tilde{\rho}_t + g(\nu)\tilde{\sigma}_t\tilde{\rho}_t$$

Figure 8 reports the sensitivity of the expected returns to a change in f_t, times the estimated change in $f_t - \overline{f}$, i.e., $\partial\,\mathbb{E}_t(r_{t+1})/\partial f_t \times \left(f_t - \overline{f}\right)$. The different coloured areas represent the five components of the decomposition; in addition to conditional expected returns (black line), we report the effect of the changes in the location μ_t (green), the scale σ_t (light pink), and the asymmetry ρ_t (violet) of the conditional distribution, and the interplay between scale and asymmetry $\sigma_t\rho_t$ (purple). Notice that $\partial\,\mathbb{E}_t(r_{t+1})/\partial\mu_t > 0, \partial\,\mathbb{E}_t(r_{t+1})/\partial\rho_t > 0$, and $\partial\,\mathbb{E}_t(r_{t+1})/\partial\sigma_t < 0$.

When we look at the full sample (top panel), one can see that a higher-than-average volatility has primarily positive effect on expected returns. However, the effect of volatility on expected returns seem to be inverted during recessions and the major momentum crash periods highlighted in pink shaded areas. The bottom panels of Figure 8 zoom in the two main momentum crashes of 1932–1939 and 2001–2009, as indicated by Daniel and Moskowitz (2016). The effects of volatility and asymmetry on the dynamics of expected returns are much cleaner. Contrary to the conventional wisdom, during the momentum crash between 1932 and 1938 higher expected returns do not uniquely depends on higher volatility (see Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016).

Towards the end of the momentum crash period of the 30s the effect of ρ_t becomes fairly relevant; in fact it is dominant from mid-1935 to the end of the 1938 recession. Put it differently, what emerges from Figure 8 is that when volatility is above average, the drag exerted by the conditional skewness increases and becomes potentially predominant. In the aftermath of the 1932 recession this relationship steadily changes: the effect of returns' skewness on expected returns increases, while the volatility term and the second-order term, $\partial\,\mathbb{E}_t(r_{t+1})/\partial\sigma_t\partial\rho_t$, becomes more muted due to a significant decrease in volatility. This suggests that while volatility risk is particularly prominent *within* the recession period, the returns' skewness plays a significant role for the dynamics of the strategy expected returns during the crash following the great depression.

A slightly different pattern emerges for the dot-com bubble and the great financial crisis of 2008/2009, as shown in the bottom-right panel of Figure 8. The negative impact of higher volatility on expected returns is rather clear. The term $\partial\,\mathbb{E}_t(r_{t+1})/\partial\sigma_t\times(\sigma_t-\overline{\sigma})$ is largely negative from 2001 to late 2002. However, the joint effect of volatility and skewness on expected returns is highly positive. This produces a positive expected returns, and suggests that that despite the negative effect of a higher volatility on expected returns in the aftermath of the dot-com bubble, this negative effect is offset by a large and positive effect due to a lower-than-average returns' asymmetry (see also left panels in Figure 6). In other words, the relatively lower skewness of the returns mitigates the effect of a large spike in volatility.

To summarise, momentum expected returns during the great financial crisis of 2008/2009 were primarily driven by an outlying spike in volatility throughout the recession period, rather than a reflection of skewness itself. Despite the returns' asymmetry being relatively lower than average during the post-2008 momentum crash, the effect of volatility is so large that the conditional expected returns are essentially dominated by the second moment. Interestingly, outside the two momentum crashes of 2001 and 2009, only the location play any significant role for the dynamics of the momentum expected returns.

For the interested reader, in Appendix B we report the estimates of $\sqrt{\mathbb{V}_t(r_{t+1})}$ for the past `winners`, past `losers` and the `WML` strategy. Returns on the past `losers` tend to be more volatile; in fact, they are almost twice as volatile as the returns for the past `winners`, especially around momentum crashes. This is reflected in a highly time-varying volatility for the `WML` strategy, with volatility spiking around both during the 1929–1932 recession, the dot-com bubble, and the great financial crisis of 2008/2009. This is consistent with the findings in Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016), and suggests that not only might the distribution be 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.

6.1 Skewness and the momentum risk-return trade-off

Figure 8 shows that the role of volatility on the dynamics of momentum expected returns should be red in conjunction with time-varying skewness. For instance, when the strategy asymmetry is close to zero, as in the tail of the dot-com bubble, a larger volatility does not translate in a negative expected return since its effect is offset by a smaller asymmetry in the returns. In addition, the results show that the effect of skewness risk on expected returns changes during momentum crashes; while a lower skewness drags down returns during the post-1932 momentum crash, the opposite holds for the momentum crashes of 2002 and 2009.

Overall, Figure 8 points towards a negative risk-return trade-off, especially during recessions and momentum crashes. Clearly, this trade-off does not uniquely depend on the returns' volatility, but also depends on the dynamics of the returns asymmetry. By rearranging Eq. (6) and using the definition of the conditional variance in Eq. (9), we can re-define the expected returns as:

$$\mathbb{E}_t(r_{t+1}) = \mu_t + \lambda_t \sqrt{\mathbb{V}_t(r_{t+1})}, \tag{11}$$

where λ_t captures the trade-off between risk and return as a nonlinear function of the time-varying asymmetry parameter ρ_t:

$$\lambda_t = \frac{g(\nu)}{\sqrt{\frac{\nu}{\nu-2} + h(\nu)\rho_t^2}}\rho_t \tag{12}$$

Equation 11 implies that the shape of the risk-return trade-off is a function of the returns' asymmetry. This function is slightly concave in ρ_t, and therefore allows for a nonlinear interaction between conditional skewness, volatility and the conditional expected returns.

Figure 9 compares the theoretical and the actual shape of the risk-return trade-off λ_t. The more negative the skewness, the more negative is the trade-off between expected returns and volatility. These results expands and complement the evidence in Theodossiou and Savva (2016), which argues that the sign of the trade-off between risks and rewards for momentum strategies is often unclear. According to them, this inconclusiveness is due to the fact that when the skewness of the returns distribution is negative, the effects of volatility and skewness tend to offset each other, generating a highly uncertain dynamics for the risk-return trade-off. Our results suggest that by allowing for a more general *non-linear*, *time-varying*, interaction between skewness and volatility risk, the trade-off between risk and rewards in momentum investing is persistently negative.

Figure 10 delves deeper into the role of conditional skewness for the risk-return trade-off in the `WML` portfolio. In particular, the figure reports the scatter plots of the expected return and the conditional volatility, the slope, and the interaction $\lambda_t\sqrt{\mathbb{V}_t(r_{t+1})}$. The red and blue crosses highlight observations from the 1932 and 2009 crash periods, respectively. The left panel shows that 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 volatility associated with the crashes.

When we look at the trade-off between $\mathbb{E}_t(r_{t+1})$ and $\sqrt{\mathbb{V}_t(r_{t+1})}$ during momentum crashes, we see that the negative correlation is much more pronounced, with all outlying values of the scale clustered over these two periods. These results indicate a strong performance of the strategy for moderate levels of volatility risk, in line with the intuition in Barroso and Maio (2019). The correlation between expected return and the slope parameter λ_t displays instead a mildly positive correlation: lower values of ρ_t – which imply larger negative skewness and lower λ_t – depress expected returns.

Notably, the relationship between expected returns and the slope on conditional volatility is rather non-linear during momentum crashes.

When the slope and the conditional volatility are interacted, there is a remarkable fit to the conditional expected returns. The right panel of Figure 10 highlights how the interaction of the two parameters results in a strongly negative effect on $\mathbb{E}_t(r_{t+1})$: low values of conditional volatility and ρ_t are associated with very small values of returns' asymmetry. When both the scale increases and the shape of the distribution becomes more tilted to the left, the amplification mechanism leads to a larger negative effect on expected returns than that exerted by the sole asymmetry. This result confirms the intuition that the high expected returns associated with the momentum strategy is not only related to time-varying volatility, but it is also significantly affected by the time variation of returns' conditional skewness.

7 Implications for asset pricing

Grundy and Martin (2001) argue that long-short portfolios formed during bear markets are likely to sell high-beta/buy low-beta stocks as firms dropping with the market will still be high-beta firms. Building upon this intuition, Daniel and Moskowitz (2016) document how 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.[19] 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 evidence and the empirical results in Section 4, and attempt to rationalise, from an asset pricing perspective, the uncovered time-varying skewness risk in momentum returns. Let us consider a standard conditional CAPM specification which separates up-market and down-market betas (see, e.g., Lettau et al., 2014),

$$r_t = \alpha + \underbrace{\overline{\beta} m_t I(m_t \geq \mu_m) + \underline{\beta} m_t I(m_t < \mu_m)}_{\beta m_t} + e_t \tag{13}$$

[19]During bear markets and high volatility, the short leg of the strategy commands a higher premium, resulting in higher gains as the market rebounds.

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. As highlighted by Ang et al. (2006), a state-dependent CAPM formulation as in Equation 13 can be thought of as the reduced form representation of a general equilibrium model in which a representative investor is endowed with a disappointment-aversion utility function that embeds a higher sensitivity to losses versus gains (see, e.g., Gul, 1991).

The distribution of the systematic pricing component, βm_t, conditional on the indicator $I(\cdot)$, can be modelled as a two-piece Normal distribution (see Johnson et al., 1995), such that the difference between the expected value $E[\beta m_t]$ and the mode $\beta\mu_m$ takes the form (see Appendix E),

$$E[\beta m_t] - \beta\mu_m = \sqrt{\frac{2}{\pi}}\left(\overline{\sigma}_m - \underline{\sigma}_m\right) \propto \sigma_m\left(\overline{\beta} - \underline{\beta}\right) \tag{14}$$

with $\underline{\sigma}_m^2 = \underline{\beta}^2\sigma_m^2$ and $\overline{\sigma}_m^2 = \overline{\beta}^2\sigma_m^2$. Under the assumption of equal betas across market states, i.e., $\underline{\beta} = \overline{\beta} = \beta$, we obtain that $E[\beta m_t] = \beta\mu_m$, $V[\beta m_t] = \beta^2\sigma_m^2$; that is, the marginal distribution of the momentum strategy returns is equivalent to a standard CAPM formulation $r_t \sim \mathcal{N}\left(\alpha + \beta\mu_m, \beta^2\sigma_m^2 + \sigma_e^2\right)$ (see Arnold and Groeneveld, 1995). On the other hand, with asymmetric betas $\underline{\beta} \neq \overline{\beta}$ and sign $\left(\underline{\beta}\right)$ = sign $\left(\overline{\beta}\right)$, Equation 14 shows that for $\overline{\beta} < \underline{\beta}$ $\left(\overline{\beta} > \underline{\beta}\right)$ 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. This holds even assuming that the market returns and the residual e_t are both normally distributed.[20]

Appendix B reports the unconditional estimates of the upside and downside market betas for both the past `losers` and `winners` as well as the `WML` strategy. The daily estimates show that the `losers`' portfolio is more exposed to upside market risk ($\overline{\beta} = 1.36$) as compared to 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 findings in Grundy and 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 ($\underline{\beta} = -0.04$).

Panel A in Figure 11 shows the simulated marginal distribution and joint distributions of the

[20]Notice this holds with the sign of the betas being the same, i.e., $sgn\left(\underline{\beta}\right) = sgn\left(\overline{\beta}\right)$. As a matter of fact, under $sgn\left(\underline{\beta}\right) \neq sgn\left(\overline{\beta}\right)$ the distribution of βm_t conditional on the indicator $I(\cdot)$ is no longer a split-Normal but a mixture of Gaussians with different means.

returns on the `WML` strategy and the market portfolio. Returns are drawn from a two-piece Normal distribution by using the unconditional $\overline{\beta}$ and $\underline{\beta}$ estimates outlined above, and assuming the market portfolio and the error term e_t are normally distributed with mean zero and volatility either equal to the historical standard deviation or one, respectively. Interestingly, the negative spread in market betas alone can generate a slightly negatively skewed (`skew` = -0.1) marginal distribution.

We expand the full sample results by calculating the downside market beta for the past `losers` and past `winners` as well as the `WML` at different points in time. Estimates are based on time-varying CAPM with asymmetric betas (see Ang et al., 2006 and Appendix B for more details). The middle and the right parts Panel A show the simulation results based on the conditional estimates for two specific timestamps of the momentum crashes as indicated by Daniel et al. (2021). Consistent with the intuition outlined in Equation 14, the conditional skewness of momentum returns markedly differs from the full sample estimates. 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 more negatively skewed (-0.805). Similarly, sampling a date from the great financial crisis of 2008/2009, the average betas spread is -1.4, with a corresponding returns skewness of -0.529.

7.1 Time-varying skewness and the conditional CAPM

Panel A in Figure 11 suggests a bridge between the asymmetric market exposure highlighted by Grundy and Martin (2001); Daniel and Moskowitz (2016) and our empirical evidence of a time-varying conditional skewness in momentum returns. Nevertheless, our conjecture is that simple model with dynamic betas cannot capture the full extent of the dynamics of the returns' skewness and its implication for momentum risk. Panel B of Figure 11 makes this case in point. The figure shows the sample correlation between the conditional skewness implied by the spread $\overline{\beta}_t - \underline{\beta}_t$ (see Figure B3 in Appendix B) and the empirical estimates reported in Section 4. There is a quite substantial and positive correlation (0.39, $pval = 0.000$) between the skewness implied by the state-dependent CAPM and our estimated skewness over the sample. However, such correlation is flatten out during the two major momentum crashes of 1932 (red dots) and 2009 (blue dots).

This suggests that a conditional CAPM with Gaussian residuals is likely not flexible enough to capture the extent of the time variation in the returns' conditional skewness. Intuitively, the reason could be twofold. First, the CAPM residuals are not normally distributed. A set of unreported results shows that the `WML` portfolio returns, net of market risk exposure, are still quite negatively skewed. Second, our simulation is based on the assumption of normally distributed market returns. Table 1 shows that market returns are indeed quite significantly negatively skewed, at least unconditionally (skewness= -0.476, $pval = 0.059$). Nevertheless, the correlation between the market and momentum conditional skewness is far from obvious. Panel B in Figure 11 shows this case in point. The daily estimates of the conditional skewness of the market and the `WML` are only mildly negatively correlated, with the correlation that is essentially zero once smoother, five-year average, estimates are considered.

These results are instrumental to highlight one key advantage of our modelling framework: by explicitly modelling the conditional skewness of the momentum returns, we can capture sources of asymmetries beyond the spread in the upside and downside market betas. While a fair deal of asymmetry in the returns' conditional distribution can be captured by asymmetric betas, there is still a considerable amount of skewness in momentum returns which can not be reconciled by a state-dependent CAPM or by the correlation between the conditional skewness of the momentum strategy and the market portfolio.

8 Conclusions

We investigate the dynamics of the conditional skewness in daily US equity momentum through the lens of a flexible dynamic parametric model which features time-varying location, scale and asymmetry in the conditional distribution of returns. Empirically, we show that the conditional skewness of the strategy returns is time-varying and deepens during the so-called momentum crashes. This has first-order implications for managing risk in momentum portfolios: an adjusted momentum portfolio that hedges in real time for both volatility and skewness risk outperforms benchmark constant and dynamic volatility-managed momentum strategies. This is due to the role of the conditional skewness for the dynamics of the strategy risk-return trade-off.

References

Anderson, T. W. and Takemura, A. (1986). Why do noninvertible estimated moving averages occur? *Journal of Time Series Analysis*, 7(4):235–254.

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

Arellano-Valle, R. B., Gómez, H. W., and 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. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *The American Statistician*, 49(1):34–38.

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985.

Bai, J. and 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. and Engle, R. F. (2010). The intertemporal capital asset pricing model with dynamic conditional correlations. *Journal of Monetary Economics*, 57(4):377–390.

Barroso, P. and Detzel, A. (2021). Do limits to arbitrage explain the benefits of volatility-managed portfolios? *Journal of Financial Economics*, 140(3):744–767.

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

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

Bernardi, M., Bianchi, D., and Bianco, N. (2022). Smoothing volatility targeting. *arXiv preprint arXiv:2212.07288*.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2):1046–1089.

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

Blasques, F., van Brummelen, J., Koopman, S. J., and Lucas, A. (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics*, 227(2):325–346.

Bowley, A. L. (1926). *Elements of statistics*. Number 8. King.

Calvori, F., Creal, D., Koopman, S. J., and Lucas, A. (2017). Testing for parameter instability across different modeling frameworks. *Journal of Financial Econometrics*, 15(2):223–246.

Carriero, A., Clark, T. E., and Marcellino, M. (2020). Capturing macroeconomic tail risks with bayesian vector autoregressions. Working Paper WP 20-02, Federal Reserve Bank of Cleveland.

Cederburg, S., O'Doherty, M. S., Wang, F., and Yan, X. S. (2020). On the performance of volatility-managed portfolios. *Journal of Financial Economics*, 138(1):95–117.

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

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

Daniel, K., Jagannathan, R., and Kim, S. (2021). Tail risk in momentum strategy returns. Technical report, National Bureau of Economic Research.

Daniel, K. and Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2):221–247.

del Castillo, J. and Daoudi, J. (2009). The mixture of left–right truncated normal distributions. *Journal of statistical planning and inference*, 139(10):3543–3551.

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

Delle Monache, D. and 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., and 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. and Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *Journal of Finance*, 48(5):1749–1778.

Escanciano, J. C. and Lobato, I. N. (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics*, 151(2):140–149.

Fabozzi, F. J., Rachev, S. T., and Menn, C. (2005). *Fat-tailed and skewed asset return distributions: implications for risk management, portfolio selection, and option pricing*. John Wiley & Sons.

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

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

Fernández, C. and 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., and 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., and 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., and 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., and 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., and Wermers, R. (1995). Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. *The American economic review*, pages 1088–1105.

Grundy, B. D. and 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., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

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

Gul, F. (1991). A theory of disappointment aversion. *Econometrica: Journal of the Econometric Society*, pages 667–686.

Hanauer, M. X. and Windmüller, S. (2023). Enhanced momentum strategies. *Journal of Banking & Finance*, 148:106712.

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

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

Harvey, A. and Streibel, M. (1998). Testing for a slowly changing level with special reference to stochastic volatility. *Journal of Econometrics*, 87(1):167–189.

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

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

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

Jacobs, H., Regele, T., and 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. and 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., and Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2*, volume 289. John wiley & sons.

Juárez, M. A. and 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. and Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific.

Kelly, B. T., Moskowitz, T. J., and Pruitt, S. (2021). Understanding momentum and reversal. *Journal of financial economics*, 140(3):726–743.

Koopman, S. J., Lucas, A., and 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. and 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. and Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. *The Journal of finance*, 31(4):1085–1100.

Ledoit, O. and 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., and Weber, M. (2014). Conditional risk premia in currency markets and other asset classes. *Journal of Financial Economics*, 114(2):197–225.

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

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

Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance*, 72(4):1611–1644.

Moskowitz, T. J. and Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance*, 54(4):1249–1290.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2):228–250.

Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon–skew–normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, 83(2):291–309.

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

Novy-Marx, R. and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230.

Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics*, 114(3):517–553.

Psarakis, S. and Panaretoes, J. (1990). The folded t distribution. *Communications in Statistics-Theory and Methods*, 19(7):2717–2734.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. In *Handbook of economic forecasting*, volume 2, pages 328–383. Elsevier.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2):821–862.

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

Sargan, J. D. and Bhargava, A. (1983). Testing residuals from least squares regression for being generated by the gaussian random walk. *Econometrica: Journal of the Econometric Society*, pages 153–174.

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

Stock, J. H. and Watson, M. W. (1998). Median unbiased estimation of coefficient variance in a time-varying parameter model. *Journal of the American Statistical Association*, 93(441):349–358.

Tanaka, K. and Satchell, S. E. (1989). Asymptotic properties of the maximum-likelihood and nonlinear least-squares estimators for noninvertible moving average models. *Econometric Theory*, 5(3):333–353.

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

Tsai, H.-J., Chen, M.-C., and 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*, pages 106–112.

Wang, F. and Yan, X. S. (2021). Downside risk and the performance of volatility-managed portfolios. *Journal of Banking & Finance*, 131:106198.

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

This table reports different descriptive statistics and measures of skewness for both past `winner` and `losers` portfolios, as well as the `WML` momentum strategy for three alternative momentum specifications as in Jegadeesh and Titman (1993) and Novy-Marx (2012). In addition, we report the sample skewness, with p-values for the Bai and Ng (2005) test in parentheses, and the quantile skewness (QS_α), computed as $\frac{q(\alpha)+q(1-\alpha)-2q(50)}{q(\alpha)-q(1-\alpha)}$, with $\alpha = 99$. The sample period is from July 1st 1926 to September 30th 2020, daily.

	`12-2`			`6-2`			`12-7`			`MKT`
	`losers`	`winners`	`WML`	`losers`	`winners`	`WML`	`losers`	`winners`	`WML`	
$\overline{r - r_f}$ (%)	−3.500	15.415	18.915	−0.130	12.928	13.059	−0.075	15.126	15.201	7.786
σ (%)	28.570	23.626	24.104	27.975	23.135	22.942	25.650	23.539	19.913	18.643
SR	−0.123	0.652	0.785	−0.005	0.559	0.569	−0.003	0.643	0.763	0.418
α (%)	−12.640	6.803	22.242	−9.657	4.208	15.341	−9.083	6.326	16.941	
β	1.317	1.162	−0.155	1.313	1.153	−0.159	1.242	1.183	−0.060	
Skewness	0.140	−0.682	−1.236	0.228	−0.717	−1.554	−0.059	−0.747	−0.768	-0.476
	(0.264)	(0.022)	(0.001)	(0.184)	(0.018)	(0.001)	(0.102)	(0.028)	(0.021)	(0.059)
QS_{99}	0.021	−0.108	−0.110	0.001	−0.093	−0.096	−0.025	−0.079	−0.089	-0.045

Table 2: **Managed momentum strategies**

The table reports the ranking of different risk-managed portfolios. Panel A reports the daily returns on our skewness-managed strategy (`mSSR`) against a variety of alternative volatility-targeting approaches. First we consider two benchmark approaches based on a conditional maximum Sharpe ratio (`mSR`), and constant volatility targeting (`cVol`). These are based on the estimates of $\mathbb{E}_t(r_{t+1}), \mathbb{V}_t(r_{t+1})$ under the restriction $\rho_t = 0$. In addition, we also compare the same exact implementation of Daniel and Moskowitz (2016) (`DM2016`) and Barroso and Santa-Clara (2015) (`BS2015`) which are based on recursive estimates of the realised variance. Finally, we also compare out skewness-adjusted momentum strategy against a constant volatility targeting *with* skewness (`cSVol`) and realised semi-volatility adjustment (`cdVol`). Panel B reports the same strategies with returns estimated on a monthly basis. We report in parentheses the bootstrapped p-values for the differences in consecutive Sharpe ratios as in Ledoit and Wolf (2008). The sample period is from July 1nd 1926 to September 30th 2020. Portfolio weights are generated in real-time by recursive forecasts of the conditional mean and variance of the returns based on the model parameters. The first three years are used as burn-in period. Panel C reports a decomposition of the performance of our skewness-managed momentum strategy between the location component w_1 and the skewness hedging component w_2 as in Eq. (8)

Panel A: Daily returns

Strategies	Sharpe	Sortino	dVol	MaxDD	Skew	VaR	ES	STARR	RR	leverage
`mSSR`	1.573 (0.001)	2.513	10.738	0.349	0.153	−3.231	−4.770	14.256	1.309	0.573
`mSR`	1.352 (0.001)	2.118	10.950	0.432	0.133	−3.284	−4.855	12.038	1.266	0.515
`cSVol`	1.414 (0.001)	2.067	11.737	0.563	−0.043	−4.205	−5.338	11.454	0.998	1.119
`cVol`	1.407 (0.001)	2.055	11.745	0.567	−0.044	−4.203	−5.343	11.386	0.996	1.125
`cdVol`	1.394 (0.001)	2.035	11.751	0.558	−0.047	−4.188	−5.349	11.262	0.992	1.652
`DM2016`	1.375 (0.001)	2.011	11.731	0.427	0.021	−3.870	−5.179	11.479	1.077	0.970
`BS2015`	1.262 (0.001)	1.812	11.950	0.462	−0.043	−4.156	−5.384	10.135	0.974	0.243
`WML`	0.737	1.017	12.435	1.137	−0.056	−3.791	−5.387	5.914	0.936	

Panel B: Monthly returns

Strategies	Sharpe	Sortino	dVol	MaxDD	Skew	VaR	ES	STARR	RR	leverage
`mSSR`	1.093 (0.012)	2.561	7.936	0.349	0.405	−0.525	−0.869	2.807	2.101	0.684
`mSR`	1.069 (0.012)	2.420	8.208	0.450	0.376	−0.537	−0.874	2.726	2.153	0.665
`cSVol`	1.141 (0.002)	2.073	10.233	0.364	0.010	−0.842	−1.116	2.280	1.351	0.562
`cVol`	1.141 (0.002)	2.065	10.268	0.366	0.006	−0.854	−1.121	2.269	1.331	0.564
`cdVol`	1.173 (0.001)	2.055	10.611	0.487	−0.067	−0.870	−1.158	2.260	1.176	1.223
`DM2016`	1.159 (0.001)	2.311	9.323	0.547	0.221	−0.684	−0.982	2.632	1.703	0.910
`BS2015`	1.082 (0.001)	1.836	10.957	0.506	−0.010	−0.841	−1.176	2.053	1.209	0.351
`WML`	0.641	0.877	13.599	1.028	−0.092	−0.881	−1.326	1.079	0.925	

Panel C: Skewness-managed portfolio decomposition (daily returns)

Component	Sharpe	Sortino	dVol	MaxDD	Skew	VaR	ES	STARR	RR
w	1.573	2.513	10.738	0.349	0.153	-3.231	-4.770	14.256	1.309
w_1	1.652	2.571	13.623	0.496	0.100	-4.408	-6.206	14.223	1.226
w_2	0.645	0.926	8.833	1.416	-0.002	-3.035	-4.176	12.933	0.974

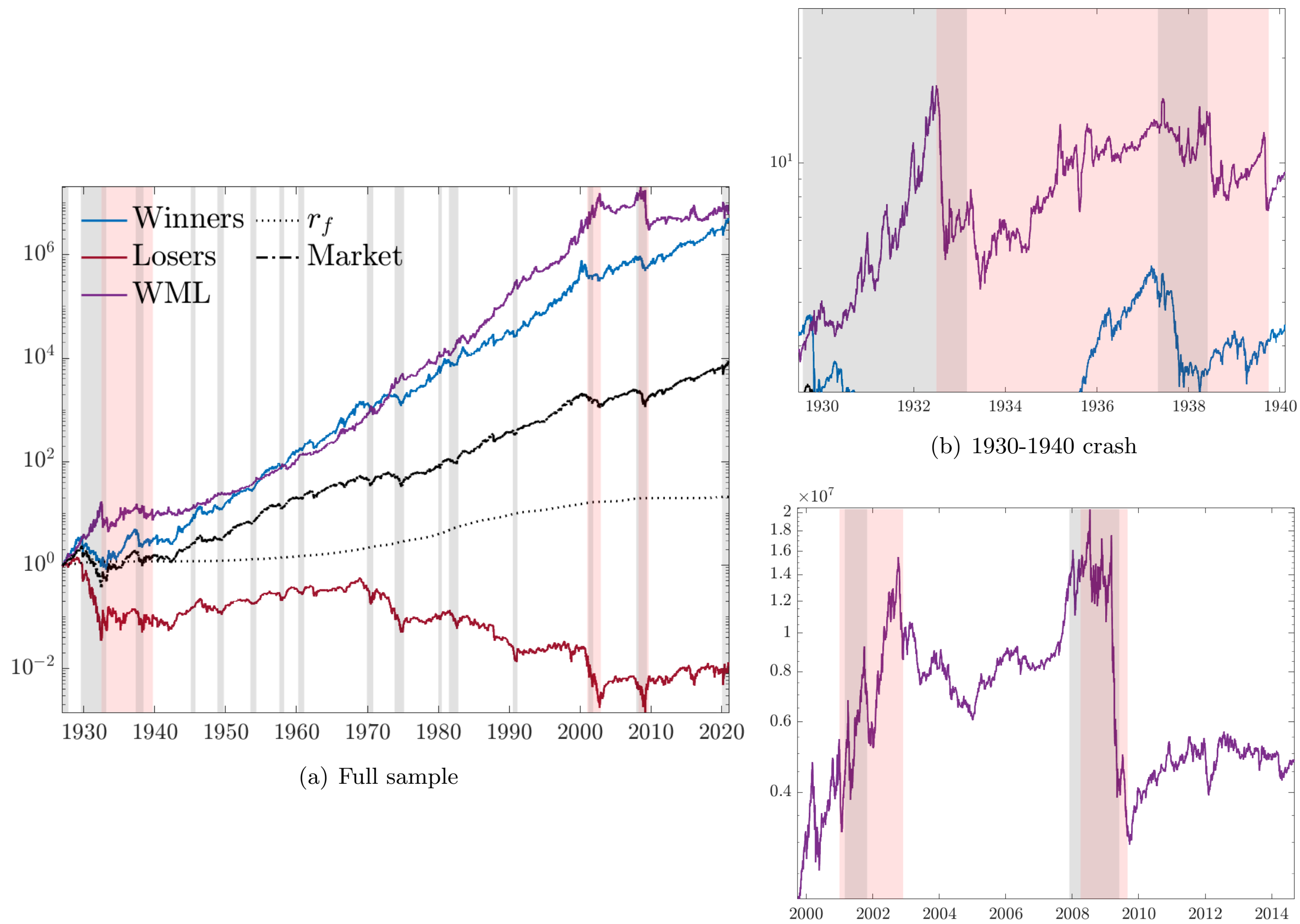
Table 3: **Transaction costs and performance fees**

The table reports the out-of-sample terminal returns net of transaction costs ($\overline{r} - c$, DeMiguel et al., 2009), the performance fee ($\mathcal{F}$) of Fleming et al., 2003), and the downside-abnormal return ($d\mathcal{A}$, Modigliani and Modigliani, 1997). We report the results for our skewness-managed strategy (`mSSR`) against a constant volatility targeting *with* skewness (`cSVol`) and the same exact implementation of Daniel and Moskowitz (2016) (`DM2016`) and Barroso and Santa-Clara (2015) (`BS2015`) which are based on recursive estimates of the realised variance. The performance fees are computed for a risk aversion coefficient of 5. All the measures are reported in annual basis points. The sample period is from July 1nd 1926 to September 30th 2020. Portfolio weights are generated in real-time by recursive forecasts of the conditional mean and variance of the returns based on the model parameters. The first three years are used as burn-in period.

Costs (bps)	mSSR			cSVol			DM2016			BS2015		
	$\overline{r}-c$	$d\mathcal{A}$	$\mathcal{F}$	$\overline{r}-c$	$d\mathcal{A}$	$\mathcal{F}$	$\overline{r}-c$	$d\mathcal{A}$	$\mathcal{F}$	$\overline{r}-c$	$d\mathcal{A}$	$\mathcal{F}$
0	14.328	18.594	10.595	11.597	13.025	7.946	10.902	12.291	7.189	8.957	9.809	5.297
1	13.897	18.057	10.216	11.450	12.860	7.820	10.866	12.251	7.189	8.953	9.805	5.297
5	12.176	15.921	8.450	10.863	12.198	7.189	10.724	12.090	7.063	8.937	9.788	5.297
14	10.024	13.278	6.306	10.130	11.374	6.432	10.547	11.889	6.811	8.918	9.767	5.171

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 and Moskowitz (2016).



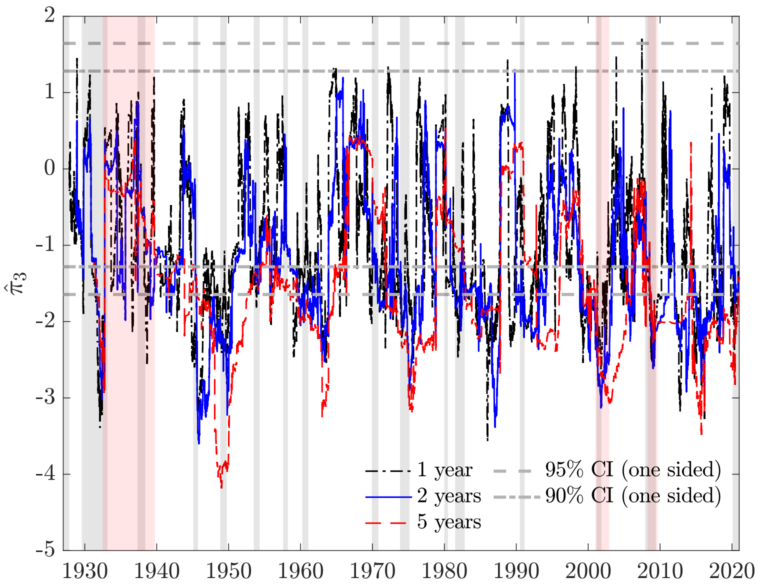
(a) Full sample

(b) 1930-1940 crash

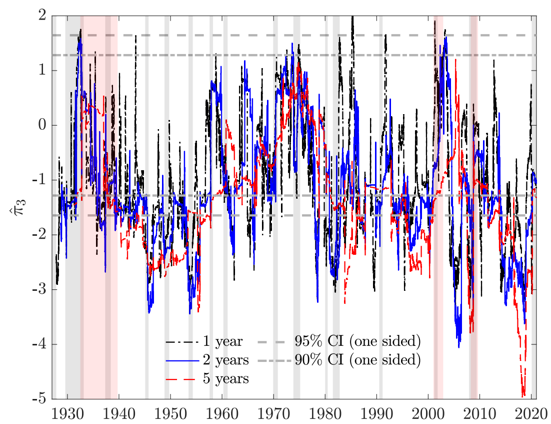
(c) 2000-2009 crash

Figure 2: **Recursive Bai and Ng (2005) skewness test**

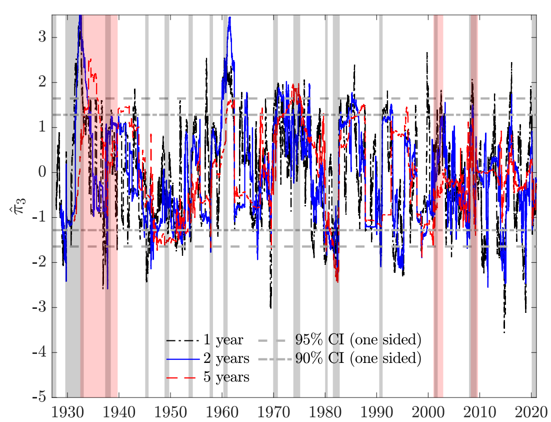
The three panels report the report the time series of the Bai and Ng (2005) test statistics for asymmetry over different rolling window of returns. We report the testing results by using one, two and five years of daily returns on the past `losers` (left panel), past `winners` (middle panel) and the `WML` strategy (right panel). The dashed horizontal lines represent the 90% and 95% confidence intervals. NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods, as indicated in Daniel and Moskowitz (2016). The sample period is from July 1st 1926 to September 30th 2020, daily.



(a) `WML` portfolio



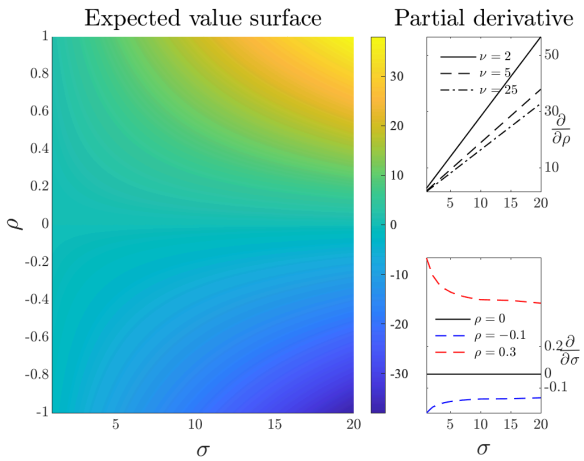
(b) Past `winners`



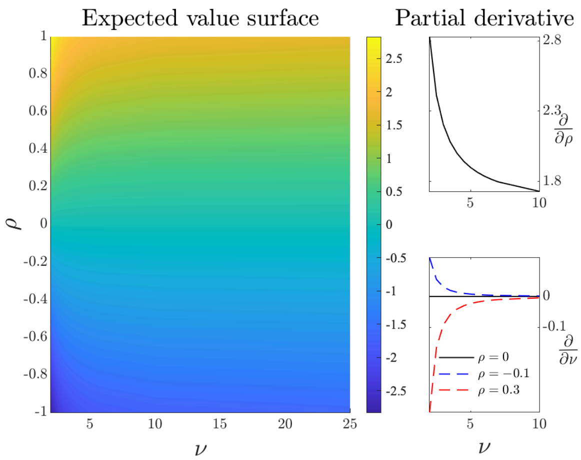
(c) Past `losers`

Figure 3: **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.(6) 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.



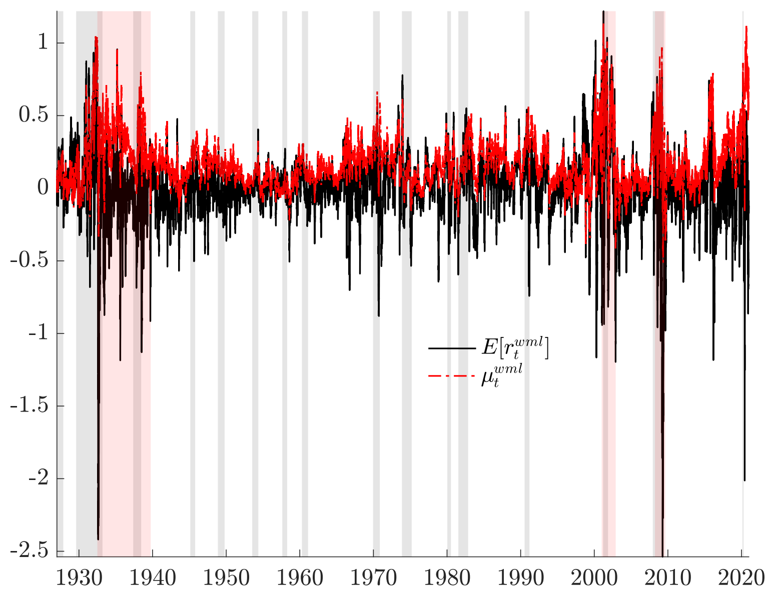
(a) ρ_t and σ_t



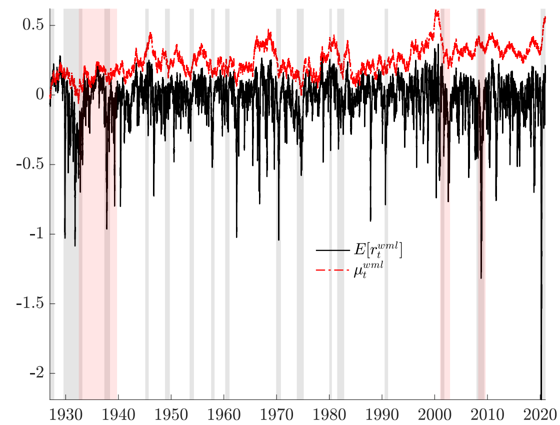
(b) ρ_t and ν

Figure 4: **Conditional expected returns and the location parameter**

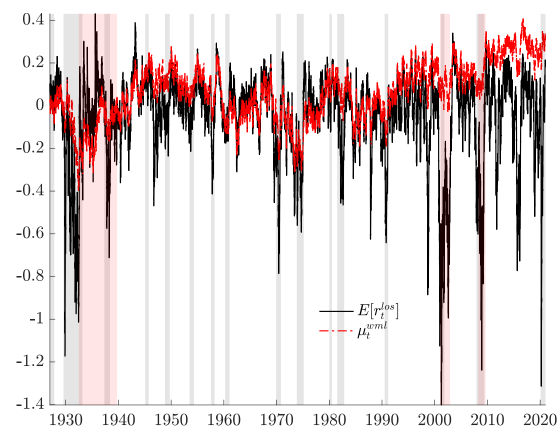
The plot reports the time-varying location parameter μ_t (red line) and the conditional expected returns as in Eq.(6) (black line). We report the values for the `WML` portfolio (left panel), the past **`losers`** (bottom-right panel) and the past **`winners`** (top-right panel) sub-portfolios. NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods, as indicated in Daniel and Moskowitz (2016). The sample period is from July 1st 1926 to September 30th 2020, daily.



(a) `WML` portfolio



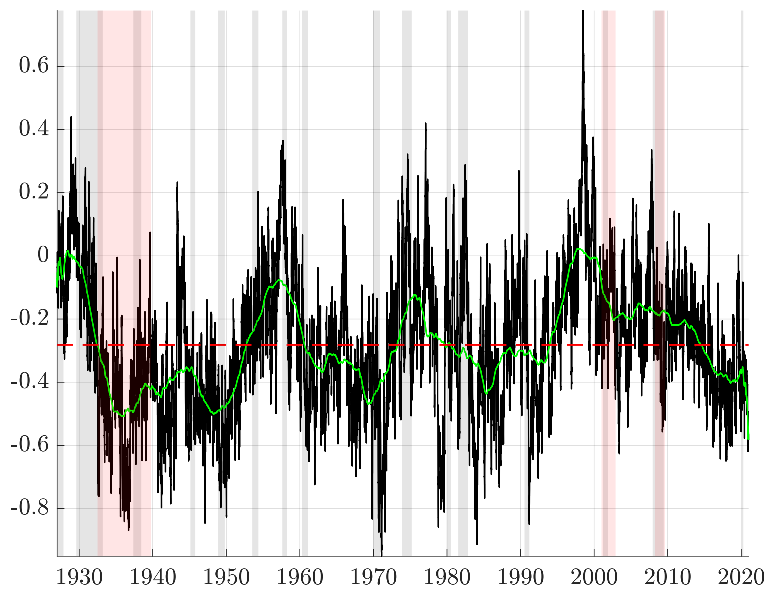
(b) Past **`winners`**



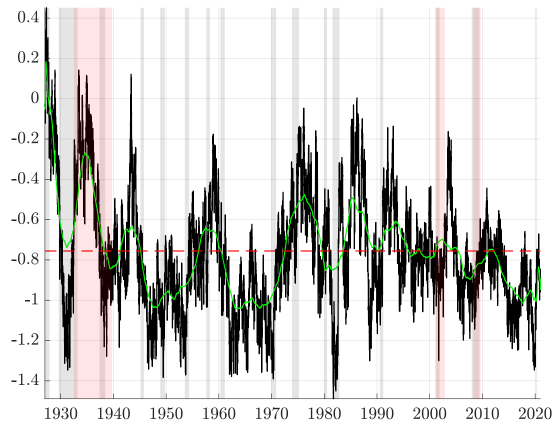
(c) Past **`losers`**

Figure 5: **Conditional skewness**

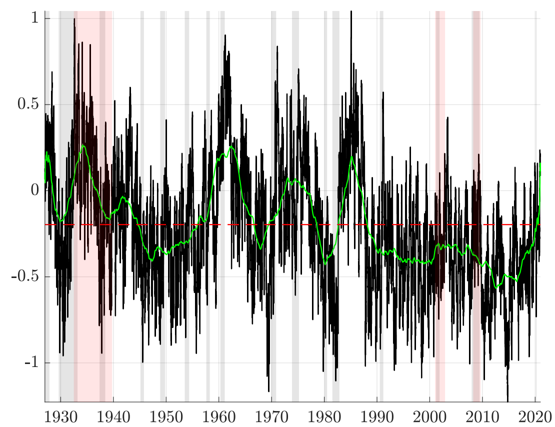
The plot reports the time-varying skewness estimates for the `WML` portfolio (left panel), the past `losers` (bottom-right panel) and the past `winners` (top-right panel) sub-portfolio returns. The horizontal red dashed line represents the sample mean of the conditional skewness estimates. The green 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 and Moskowitz (2016). The sample period is from July 1st 1926 to September 30th 2020, daily.



(a) `WML` portfolio



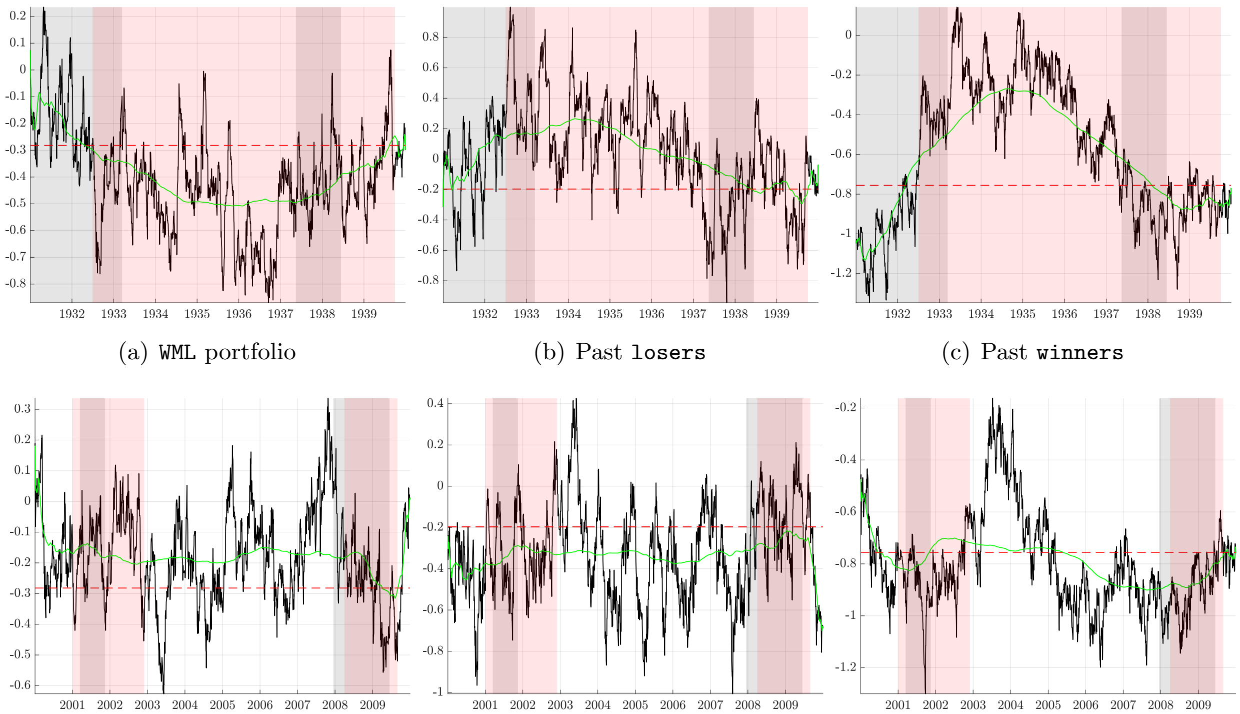
(b) Past `winners`



(c) Past `losers`

Figure 6: **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 and Moskowitz (2016).

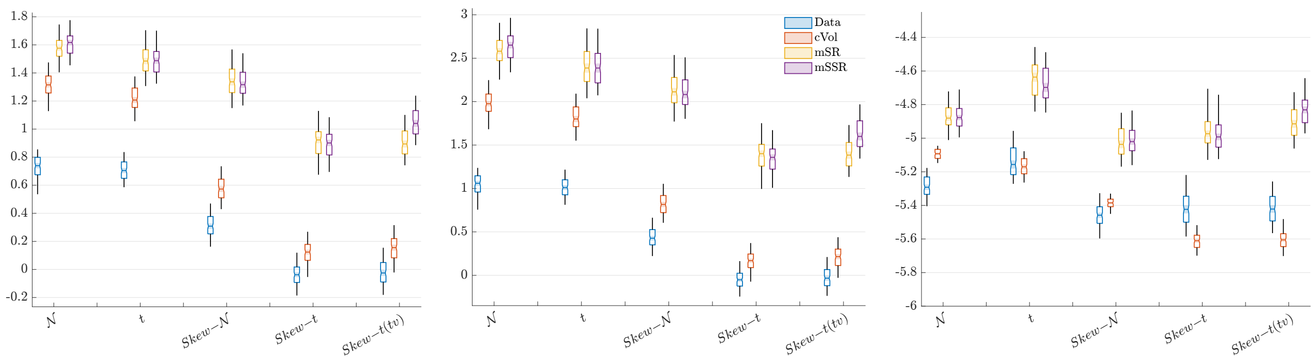


(a) `WML` portfolio (b) Past `losers` (c) Past `winners`

(d) `WML` portfolio (e) Past `losers` (f) Past `winners`

Figure 7: **Simulation study**

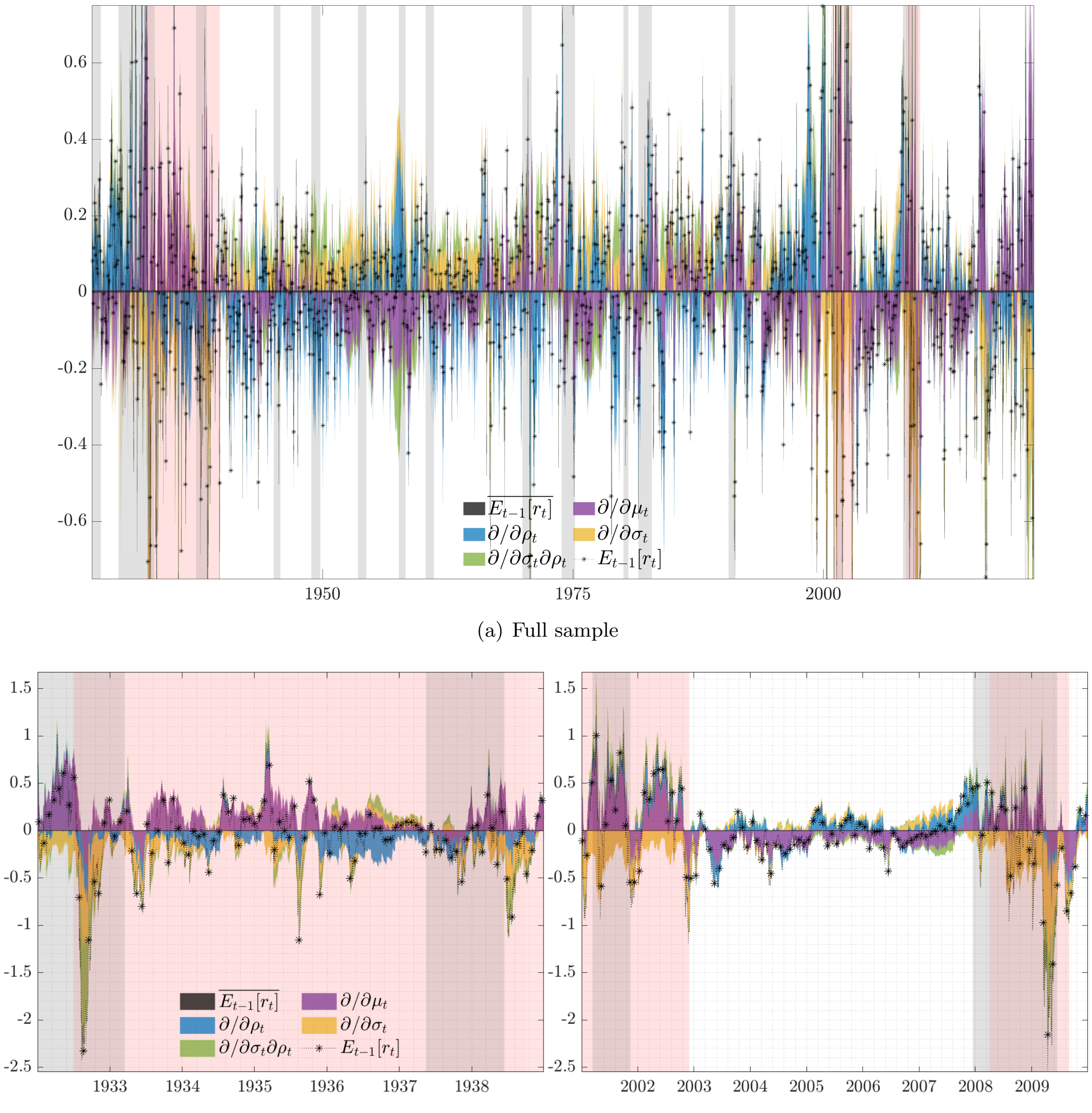
The box plots report the distributions of the sample Sharpe ratios (a), Sortino ratios (b) and maximum drawdown (c) of 100 return paths (Data, blue), simulated assuming the following distributions: Gaussian ($\mathcal{N}$), Skew-Gaussian ($Skew{-}\mathcal{N}$), Student-t (t), and asymmetric Student-t ($Skew{-}t$) with constant skewness and an asymmetric Student-t with time-varying asymmetry ($Skew_{tv}{-}t$). For each distribution assumption, we simulate 100 paths of 10,000 observations. The $Skew{-}\mathcal{N}$ and $Skew{-}t$ distributions feature -0.12 and -0.3 unconditional skewness, respectively. We report the performance of three portfolios: a constant volatility targeting (`cVol`, red), a maximum Sharpe ratio (`mSR`, yellow) and maximum Sharpe ratio with time-varying skewness (`mSSR`, purple). Constant volatility portfolio returns are obtained by standardizing the simulated returns by GARCH(1,1) volatility estimates, while maximum Sharpe ratio strategies are obtained as per Section 5; we target an annualized volatility of 18% across all strategies consistent with the market portfolio returns.



(a) Sharpe ratio (b) Sortino ratio (c) Expected Shortfall

Figure 8: **Expected returns decomposition**

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 and Moskowitz (2016).



(a) Full sample

(b) 1932-1939 period

(c) 2001-2009 period

Figure 9: **Conditional skewness and the risk-return trade-off**

The figure reports the theoretical shape of $\lambda(\rho_t)$ (dashed curve), its realized value (blue marks), and the unconditional sample mean (red circle). The sample period is from July 1st 1926 to September 30th 2020, daily.

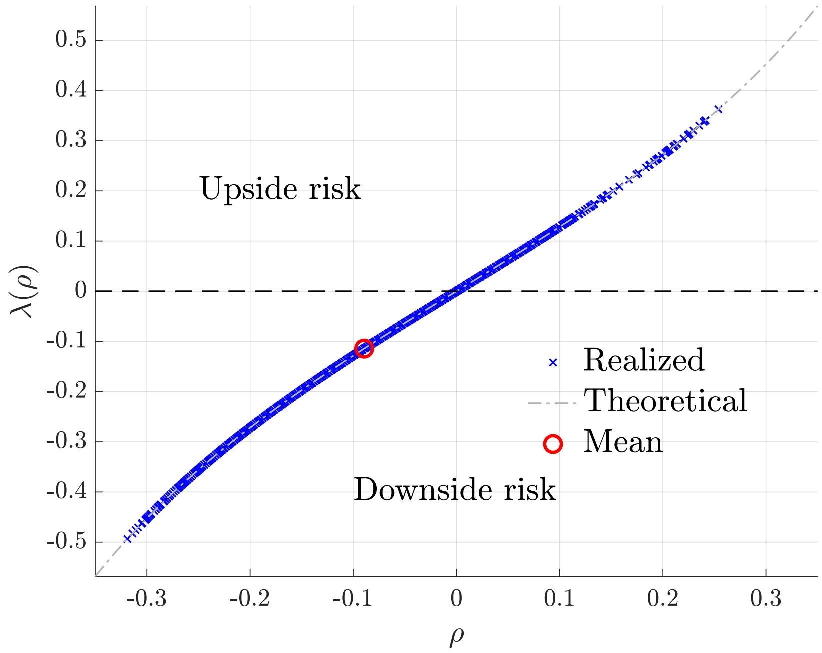
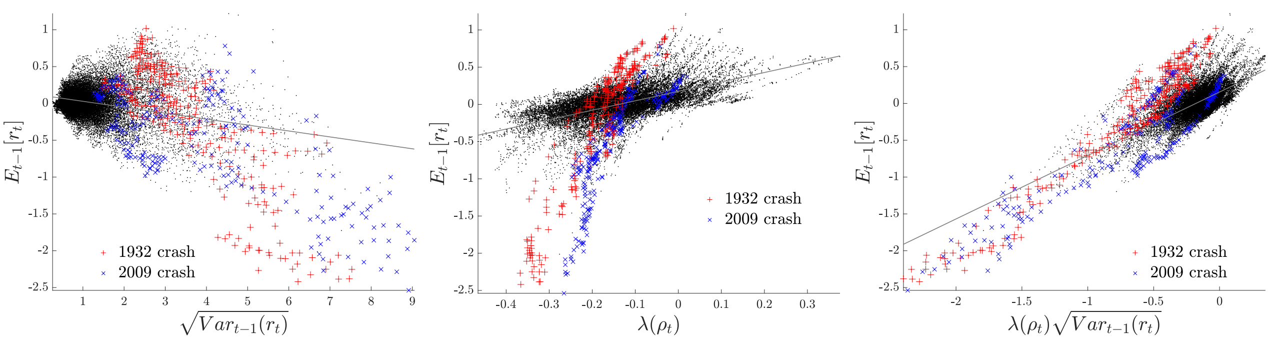


Figure 10: **Decomposing the risk-return trade-off**

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.(11). Red crosses highlight observations relative to the 1932 crash. Blue crosses highlight observations relative to the 2009 crash.

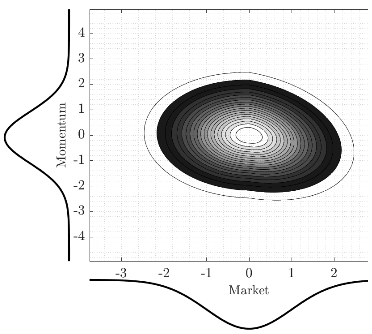


(a) Expected returns vs volatility (b) Expected returns vs slope (c) Realised vs fitted expected returns

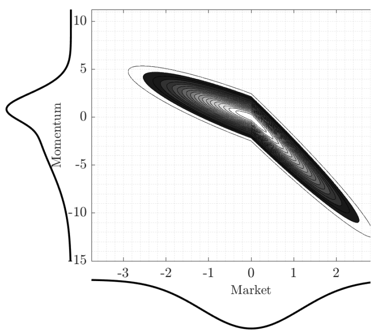
Figure 11: **State-dependent CAPM and simulated momentum returns**

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.(E2) in Appendix E. 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.

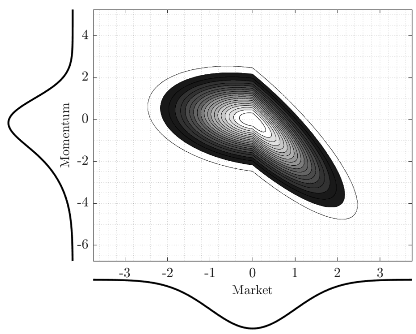
Panel A: Joint and marginal distribution of simulated `WML` and `MKT` returns



(a) **Full sample**

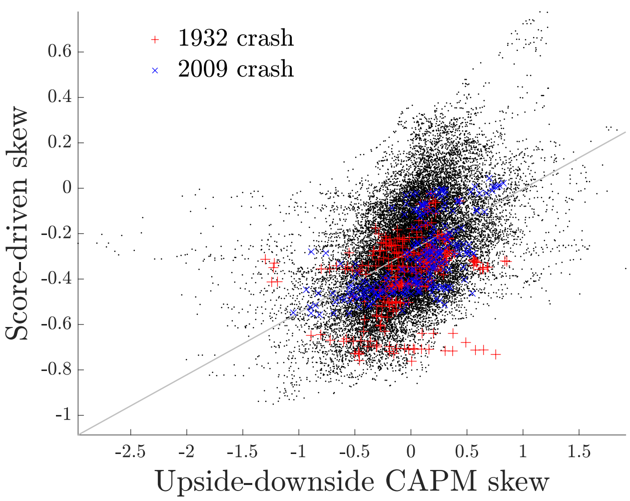


(b) **22/03/1935**

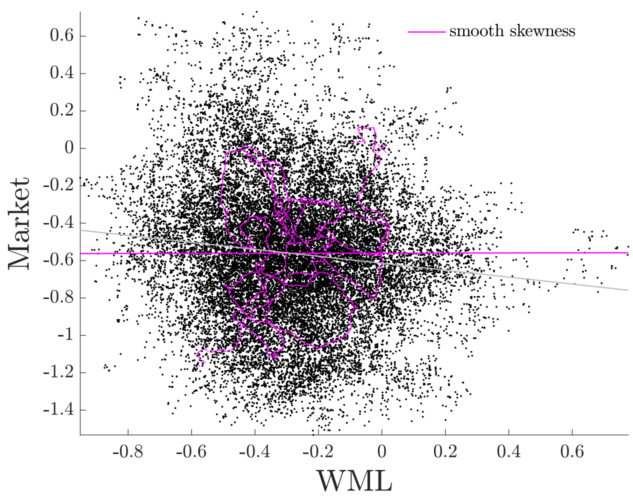


(c) **30/09/2008**

Panel B: Skewness from our model vs state-dependent CAPM



(d) Model-implied skewness of the `MKT` vs `WML`



(e) Model-implied vs CAPM-implied skewness of `WML`

Internet Appendix to

"Taming Momentum Crashes"

This appendix provide a set of additional empirical results as well as outline our econometric model. More specifically, we provide detailed derivations for the scaled-scores vector and the conditional mean and variance of the returns under the distributional assumptions outlined in the main text. The appendix also provides a simple mapping between the asymmetric CAPM beta and conditional returns asymmetry. Finally, we also report some simulation results to highlight the role of conditional skewness for risk-adjusted returns. Both the additional empirical analysis and the simulation results are referred to in the main text where appropriate.

A A likelihood-based test on conditional skewness

The basic idea is to assume a given moment is constant in the data generating process and then look at the information contained over time in the gradient of the log-likelihood function (or score) with respect to that moment (see Harvey (2013)). Assume the conditional distribution of the portfolio returns being a Skew-t of Gómez et al. (2007) with time-varying location μ_t and scale σ_t, but fixed shape ρ parameters which pins down the degree of asymmetry in the conditional distribution of the returns, that is $r_t|\mathcal{F}^{t-1} \sim Skt_\nu(\mu_t, \sigma_t^2, \rho)$. The gradient associated with the transformed shape (asymmetry) parameter $\delta = \operatorname{arctanh} \rho$ is defined as

$$\nabla_{\delta,t} = \frac{\mathrm{s}\left(\varepsilon_t\right)\left(1-\rho^2\right)}{\left(1+\mathrm{s}\left(\varepsilon_t\right)\rho\right)} w_t \zeta_t^2, \tag{A1}$$

with $\zeta_t = \epsilon_t/\sigma_t$ the standardised residuals $\epsilon_t = r_t - \mu_t$, and $w_t = (1+\nu)/\left(\nu\left(1+\mathrm{s}\left(\varepsilon_t\right)\right)\rho + \zeta_t^2\right)$ the weight given to the squared of standardised residuals at each time t (see Section 3 and Appendix C.1 for more details). By looking at the autocorrelation properties of the score in Eq.(A1), a Lagrange multiplier principle (LM) can be employed to formally test for the time variation of ρ (see, e.g., Calvori et al., 2017). More specifically, tests for the time variation of ρ can be carried out using the score autocorrelation function and implementing otherwise standard Portmanteau (P) and Ljung-Box (Q) tests for the null hypothesis of absence of autocorrelation in the score $\nabla_{\delta,t}$, that is no time variation in ρ. The optimal lag-length for the P and Q tests is selected following the methodology by Escanciano and Lobato (2009). In addition to the Portmanteau and Ljung-Box tests, we also report the results from a general test for the null of constant parameters against a random-walk alternative based on the LM principle as proposed by Nyblom (1989). In our case, the test statistics

Table A1: **Score-based test for time variation in conditional skewness**

P is the portmanteau test, Q is the Ljung-Box extension and N corresponds to the Nyblom test. The lag length for the Portmanteau and Ljung-Box tests are selected following Escanciano and Lobato (2009). P and Q are distributed as a χ^2_1, while N is distributed as a Cramer von-Mises distribution with 1 degree of freedom. $*p < 10\%, **p < 5\%, ***p < 1\%$.

Portfolios	Autocorrelation tests		
	P	Q	N
`losers`	$> 100^{***}$	$> 100^{***}$	3.374^{***}
`winners`	$> 100^{***}$	$> 100^{***}$	7.991^{***}
`WML`	$> 100^{***}$	$> 100^{***}$	6.751^{***}

reads as follows:

$$N = \sigma_{\nabla}^{-2} T^{-2} \sum_{j=1}^{T} \left(\sum_{k=j}^{T} \nabla_{\delta,k} \right)^2, \tag{A2}$$

where $\nabla_{\delta,k}$ denotes the score of the distribution with respect to the transformed shape parameter $\delta = \text{arctanh}\ \rho$ at time k and σ^2_{∇} represents the sample variance of the score. Harvey and Streibel (1998) showed that although the Nyblom (1989) test is regarded as a test against a random walk alternative it can also be interpreted as a general test against the alternative hypothesis of time variation of a given model parameters (see, e.g., Delle Monache et al., 2021). Table A1 reports the results. The null hypothesis of a constant skewness is strongly rejected against the alternative of time variation, with the values of test statistics which are well above 100 for both the long and the short legs of the momentum strategy as well as the `WML` portfolio. The Nyblom test statistic follows a Cramer-von Mises distribution with a 5% critical value of 0.462. The last column in Table A1 shows that the Nyblom test suggests that the asymmetry, meaning the shape parameter, of the conditional distribution of each portfolios and the `WML` strategy is possibly not constant over time.

B Additional Results

In this section, we report a set of additional results related to the dynamics of the conditional volatility estimates and the relationship between asymmetric betas and conditional skewness.

B.1 Risk aversion

In this Section, we repeat the economic evaluation of Table 3 controlling for different levels of risk aversion. Table B1 reports 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

Table B1: **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			cSVol			DM2016			BS2015		
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	15.514	15.009	16.775	12.739	12.108	13.748	12.108	11.225	11.730	10.090	9.333	10.342
1	15.135	14.631	16.396	12.613	11.982	13.495	11.982	11.225	11.730	10.090	9.333	10.342
5	13.369	12.865	14.631	11.982	11.351	12.991	11.856	11.099	11.604	10.090	9.333	10.342
10	11.225	10.721	12.486	11.225	10.595	12.234	11.730	10.973	11.351	10.090	9.333	10.216

hedging for predictable variations in the returns skewness is economically meaningful, regardless the level of risk aversion.

B.2 Conditional volatility estimates

Figure B1 shows the estimates of $Vol_t = \sqrt{Var_t\left[r_{t+1}\right]}$ based on Eq.(9) 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 and Santa-Clara (2015); Daniel and Moskowitz (2016) and suggest that not only the distribution is more asymmetric around momentum crashes, but also that both tails of the distribution tend to be thicker in particular towards the end of recessions.

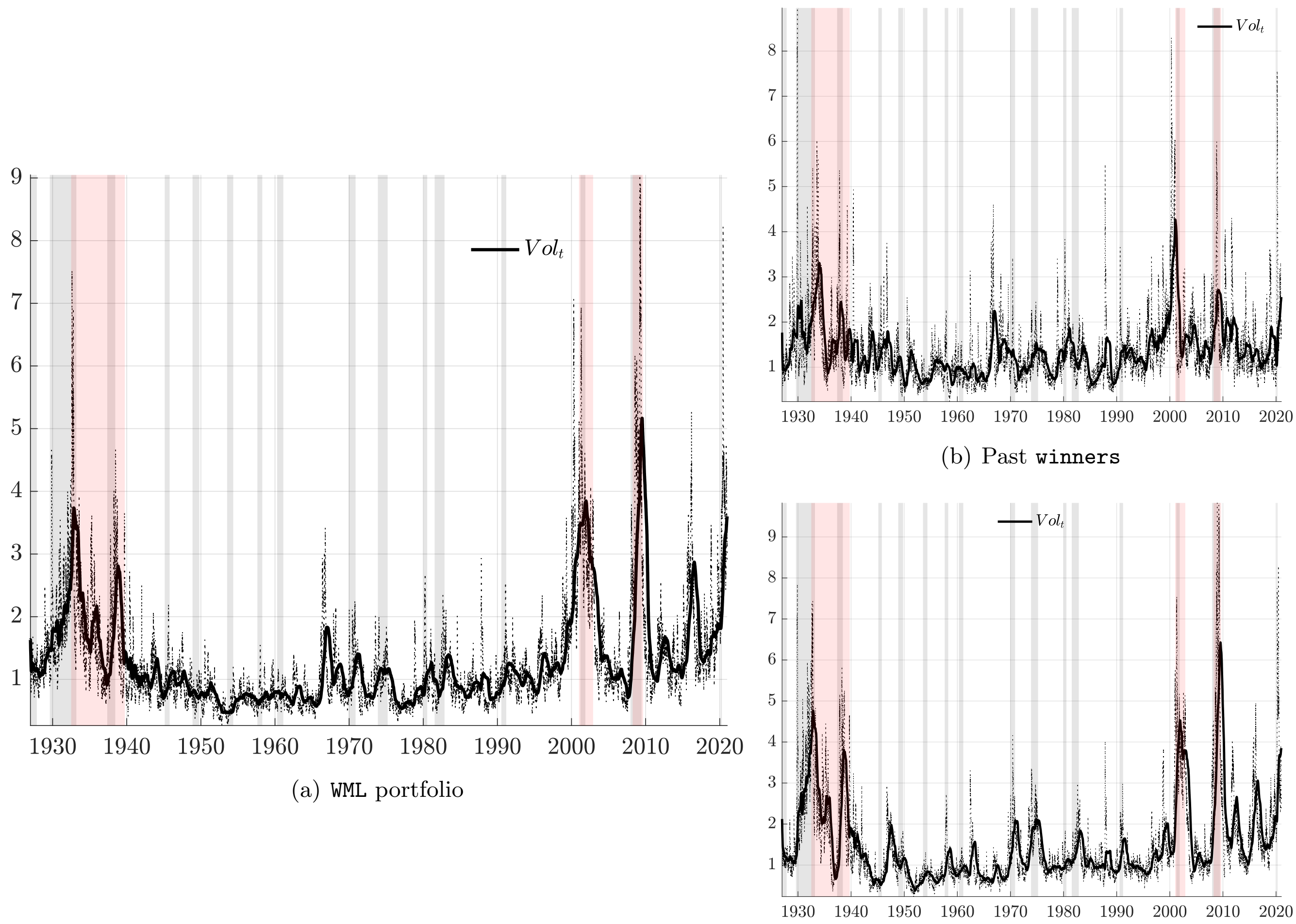
One comment is in order. Section 3.2 shows that larger negative skewness triggers a downward correction of the expected returns on the momentum strategy, whereas the impact of skewness on the variance of the returns is less pervasive. This is primarily due to the fact that $\rho_t \in (-1, 1)$ and enters squared in the conditional variance – see Eq. (9) –, while enters in level in the dynamics of expected returns – see Eq. (6). 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.

B.3 State-dependent CAPM estimates

Figure B2 reports the unconditional estimates of the upside and downside market betas for both the past `losers` and `winners` as well as the `WML` strategy. The left (right) panel reports the estimates based on daily (monthly) returns. The estimates of the upside, $\overline{\beta}$, and downside, $\underline{\beta}$ betas are based

Figure B1: **Conditional volatility of momentum returns**

The plot reports the time-varying volatility estimates based on Eq.(9) 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 and Moskowitz (2016). The sample period is from July 1st 1926 to September 30th 2020, daily.



(a) `WML` portfolio

(b) Past `winners`

(c) Past `losers`

on the following regression:

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

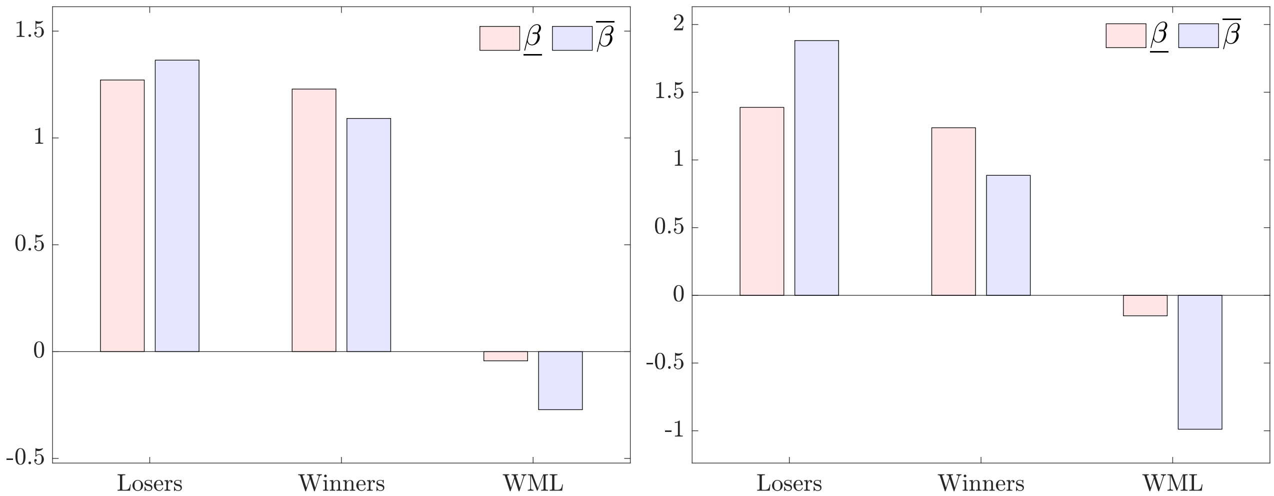
The daily estimates show that the `losers`' portfolio is more exposed to upside market risk ($\overline{\beta} = 1.36$) as compared to downside market risk ($\underline{\beta} = 1.27$), in relative terms compared to the unconditional market beta ($\beta = 1.31$). The opposite holds for the `winners`' portfolio ($\overline{\beta} = 1.09$, $\underline{\beta} = 1.22$, $\beta = 1.16$), consistent with the findings in Grundy and 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.

Figure B2: **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

B.3.1 Time-varying market betas. We follow Ang et al. (2006), and calculate the downside market beta over time for the `losers`, `winners` portfolios and the `WML` at different points in time based on a time-varying CAPM with asymmetric betas as follows,

$$\underline{\beta}_t^i = \frac{\text{cov}_t(\tilde{r}_{t+1}^i, \min\{\tilde{m}_{t+1}, 0\})}{\text{var}_t(\min\{\tilde{m}_{t+1}, 0\})} \quad i = \texttt{losers}, \texttt{winners}, \texttt{WML}, \tag{B1}$$

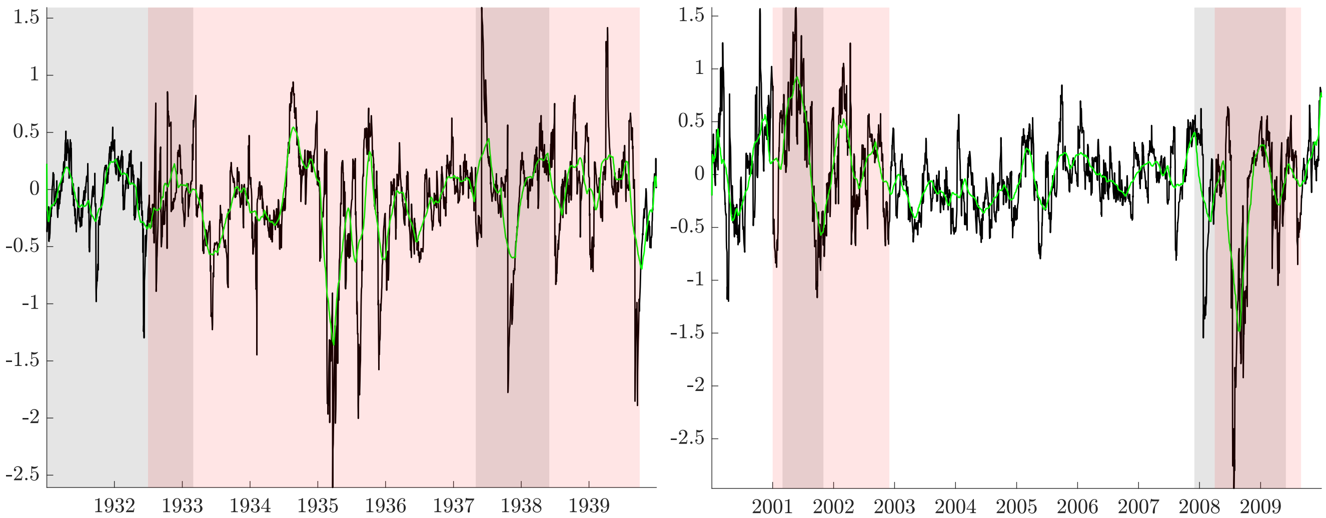
where $\tilde{r}_t^i$ and $\tilde{m}_t$ are the demeaned returns for the momentum strategy and the demeaned excess market returns, respectively (see, e.g., Hogan and Warren, 1974). The denominator of Eq. (B1) captures the variance of the downside market excess returns, and is generally referred to as 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 Equation (B1) with the *max* operator.

Fig. B3 reports the estimates for the spread $\mathcal{B}_t = \overline{\beta}_t^{\texttt{WML}} - \underline{\beta}_t^{\texttt{WML}}$ for the periods indicated as momentum crashes by Daniel and Moskowitz (2016).[21] To estimate the time-varying downside and upside betas for the momentum strategy returns, we follow Bali and Engle (2010); Tsai et al. (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.

[21]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.

Figure B3: **Momentum crashes and the exposure to downside and upside risk**

The plots report the spread between the upside and downside betas, $\mathbf{B}_t$. The left panel span the 1927-1940 period, while the right panel cover from 2000 to 2020. Gray shaded bands highlight NBER recession. Red shaded bands indicate momentum crash periods, as indicated in Daniel and Moskowitz (2016).



(a) 1927-1940 period (b) 2000-2020 period

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.

C Modeling framework

Assume that the return y_t is generated by the observation density $\mathcal{D}(\theta, f_t)$, with θ collecting the static parameters of the distribution and f_t a series of time-varying parameters which characterize the first three moments of the conditional distribution:

$$f_{t+1} = f_t + As_t, \qquad t = 1, \ldots, T \tag{C1}$$

where A contains the structural parameters regulating the law of motion of the distribution parameters, and s_t containing the likelihood information from the prediction error $\widehat{\varepsilon}_t$. Specifically, $s_t = \mathcal{S}_t \nabla_t$ is the *scaled score*, with $\nabla_t = J_t' \left[\frac{\partial \ell_t}{\partial \mu_t}, \; \frac{\partial \ell_t}{\partial \sigma_t^2}, \; \frac{\partial \ell_t}{\partial \rho_t} \right]'$ being the gradient of the log-likelihood function with respect to the (nonlinear transformation of the) location, squared scale and asymmetry parameters, J_t the Jacobian matrix associated to the non-linear transformation of the parameters for σ_t and ρ_t and

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

the scaling matrix proportional to the square-root generalized inverse of the Information matrix $\mathcal{I}_{t-1}$.[22] Within this framework, the parameters are updated in the direction of the steepest ascent, in order to maximize the local fit of the model. In the following, we are going to derive both gradient of the log-likelihood function and the Jacobian matrix in order to define the scaled-scores vector.

C.1 Score derivations

The scaled score s_t is a non-linear function of past observations and past parameters' values. For $\ell_t = \log \mathcal{D}(\theta, f_t)$ being the Skew-t of Gómez et al. (2007), $y_t|Y_{t-1} \sim skt_\nu(\mu_t, \sigma_t^2, \rho_t)$, the log-likelihood takes the form

$$\ell_t(r_t|\theta, \mathcal{F}_{t-1}) = \log \mathcal{C}(\nu) - \frac{1}{2}\log \sigma_t^2 - \frac{1+\nu}{2}\log\left[1 + \frac{\varepsilon_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2}\right], \tag{C2}$$
$$\log \mathcal{C}(\nu) = \log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\nu - \frac{1}{2}\log\pi,$$

where $\Gamma(\cdot)$ is the Gamma function and $\nu > 3$ are the degrees of freedom. Differentiating (C2) with respect to location, scale and asymmetry we obtain the gradient vector $\nabla_t = \left[\frac{\partial \ell_t}{\partial \mu}, \frac{\partial \ell_t}{\partial \sigma_t^2}, \frac{\partial \ell_t}{\partial \rho_t}\right]'$. Recall that $\varepsilon_t = y_t - \mu_t$, $\zeta_t = \frac{\varepsilon_t}{\sigma_t}$ and let

$$f(\mu_t, \sigma_t^2, \rho_t) = 1 + \frac{\varepsilon_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2} = \frac{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2 + \varepsilon_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2}$$

To avoid overburdening the notation, in what follows $\frac{\partial f(x)}{\partial x} = f_x'$ and $a = -\frac{1+\nu}{2}$. The score with respect to the location parameter reads

$$\frac{\partial \ell_t}{\partial \mu_t} = w_t \frac{\zeta_t}{\sigma_t}, \qquad \text{with} \qquad w_t = \frac{\nu+1}{\nu\left(1+\mathrm{s}\left(\varepsilon_t\right)\rho_t\right)^2 + \zeta_t^2}.$$

Proof. Define

$$g(\mu_t) = a \log f(\mu_t, \sigma_t^2, \rho_t),$$

such that $\frac{\partial \ell_t}{\partial \mu_t} = \frac{\partial g(\mu_t)}{\partial \mu_t} = a\frac{f_{\mu_t}'}{f(\mu_t, \sigma_t^2, \rho_t)}$. For

$$f_{\mu_t}' = -\frac{2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2}\varepsilon_t,$$

[22] Refer to Creal et al. (2013) for additional details on this choice.

it follows:

$$\begin{aligned}
\frac{\partial \ell_t}{\partial \mu_t} &= \frac{1+\nu}{2}\frac{2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2}\cdot\varepsilon_t\cdot\frac{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2+\varepsilon_t^2}\\
&= \frac{(1+\nu)}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2+\varepsilon_t^2}\varepsilon_t\\
&= \omega_t\frac{\zeta_t}{\sigma_t}
\end{aligned}$$

. □

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

$$g(\sigma_t^2) = -\frac{\log\sigma_t^2}{2} + a\log f(\mu_t,\sigma_t^2,\rho_t),$$

such that $\frac{\partial \ell_t}{\partial \sigma_t^2} = \frac{\partial g(\sigma_t^2)}{\partial \sigma_t^2} = -\frac{1}{2\sigma_t^2} + a\frac{f'_{\sigma_t^2}}{f(\mu_t,\sigma_t^2,\rho_t)}$, with $f'_{\sigma_t^2} = -\frac{\varepsilon_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^4}$. It follows that:

$$\begin{aligned}
\frac{\partial \ell_t}{\partial \sigma_t^2} &= -\frac{1}{2\sigma_t^2} - \frac{1+\nu}{2}\cdot\left[-\frac{\varepsilon_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^4}\cdot\frac{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2+\varepsilon_t^2}\right]\\
&= -\frac{1}{2\sigma_t^2} - \frac{1+\nu}{2}\cdot\left[-\frac{\varepsilon_t^2}{\sigma_t^2}\cdot\frac{1}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2+\varepsilon_t^2}\right]\\
&= -\frac{1}{2\sigma_t^2} + \frac{w_t\zeta_t^2}{2\sigma_t^2} = \frac{(w_t\zeta_t^2-1)}{2\sigma_t^2}
\end{aligned}$$

. □

The score with respect to the shape parameter reads as

$$\frac{\partial \ell_t}{\partial \rho_t} = \frac{\mathrm{s}(\varepsilon_t)}{(1+\mathrm{s}(\varepsilon_t)\rho_t)}w_t\zeta_t^2.$$

Proof. Define

$$g(\rho_t) = a\log f(\mu_t,\sigma_t^2,\rho_t),$$

such that $\frac{\partial \ell_t}{\partial \rho_t} = \frac{\partial g(\rho_t)}{\partial \sigma_t^2} = a\frac{f'_{\rho_t}}{f(\mu_t,\sigma_t^2,\rho_t)}$, with $f'_{\rho_t} = -\frac{2(\mathrm{s}(\varepsilon_t)+\rho_t)\varepsilon_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^4\sigma_t^2}$. It follows that:

$$\begin{aligned}
\frac{\partial \ell_t}{\partial \rho_t} &= \frac{1+\nu}{2}\cdot\frac{2(\mathrm{s}(\varepsilon_t)+\rho_t)\varepsilon_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^4\sigma_t^2}\cdot\frac{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2+\varepsilon_t^2}\\
&= \frac{(\mathrm{s}(\varepsilon_t)+\rho_t)\varepsilon_t^2}{(1+\mathrm{s}(\varepsilon_t)\rho_t)^2}\frac{w_t}{\sigma_t^2} = \frac{\mathrm{s}(\varepsilon_t)}{(1+\mathrm{s}(\varepsilon_t)\rho_t)}w_t\zeta_t^2
\end{aligned}$$

. □

C.2 Scaled scores

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

$$\frac{\partial \ell_t}{\partial \gamma_t} = \frac{\partial \ell_t}{\partial \sigma_t^2}\frac{\partial \sigma_t^2}{\partial \gamma_t}, \quad \frac{\partial \ell_t}{\partial \delta_t} = \frac{\partial \ell_t}{\partial \rho_t}\frac{\partial \rho_t}{\partial \delta_t}, \tag{C3}$$

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

$$J_t = \frac{\partial(\mu_t, \sigma_t^2, \rho_t)}{\partial f_t'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2\sigma_t^2 & 0 \\ 0 & 0 & 1-\rho_t^2 \end{bmatrix}. \tag{C4}$$

The Fisher information matrix is computed as the expected value of outer product of the gradient vector. Given the degrees of freedom $\nu > 3$ this is computed as:

$$\mathcal{I}_t = \mathbb{E}_{t-1}[\nabla_t \nabla_t'] = \begin{bmatrix} \frac{(1+\nu)}{(\nu+3)(1-\rho_t^2)\sigma_t^2} & 0 & \frac{4(1+\nu)}{\sigma_t(1-\rho_t^2)(3+\nu)} \\ 0 & \frac{1}{2(3+\nu)\sigma_t^4} & 0 \\ \frac{4(1+\nu)}{\sigma_t(1-\rho_t^2)(3+\nu)} & 0 & \frac{3(1+\nu)}{(1-\rho_t^2)(3+\nu)} \end{bmatrix}. \tag{C5}$$

As a result, the vector of scaled scores reads as:

$$\mathbf{s}_t = (J_t'\text{diag}(\mathcal{I}_t)J_t)^{-1}J_t'\nabla_t = \begin{bmatrix} s_{\mu t} \\ s_{\sigma t} \\ s_{\rho t} \end{bmatrix} = \chi \begin{bmatrix} (1-\rho_t^2)w_t\varepsilon_t \\ (\nu+1)(w_t\varepsilon_t^2 - \sigma_t^2) \\ \text{s}(\varepsilon_t)(1-\text{s}(\varepsilon_t)\rho_t)w_t\frac{\varepsilon_t^2}{3\sigma_t^2} \end{bmatrix}. \tag{C6}$$

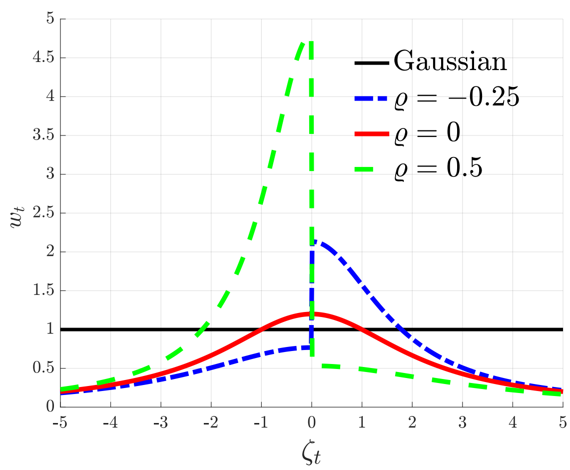
with $\chi = \frac{(\nu+3)}{(\nu+1)}$ and $w_t = \frac{\nu+1}{\nu(1+\text{s}(\varepsilon_t)\rho_t)^2+\zeta_t^2}$.

C.3 Model properties

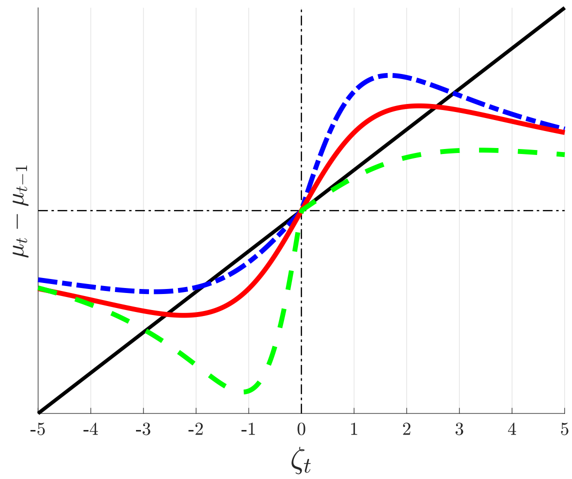
As highlighted by Delle Monache and Petrella (2017), the scalar factor w_t plays a key role as it serves as an implicit weight of the information contained in the prediction error. We summarise some its key properties in turn. The top-left panel of Fig. C1 plots the weights associated with the prediction error for alternative model parametrisations. Under a Normal distribution assumption, prediction errors are assumed to carry the same information regardless of their magnitude, i.e., $w_t = 1, \forall t$. When we consider thick tails but no asymmetry (red line), the weights tend to discount symmetrically extreme prediction errors, as is typical of Student-t distributions (see, e.g., Delle Monache and Petrella, 2017). When the distribution is negatively skewed (dashed blue line), positive prediction errors are less likely and as such command a more significant update of the parameters when they occur. The opposite holds when the distribution is positively skewed (green dashed line); large negative prediction errors are less likely and so command a larger update on the parameters. The asymmetric effect of prediction errors increases as the skewness grows larger, i.e., $\rho_t \to 1$.

Figure C1: **News impact curves**

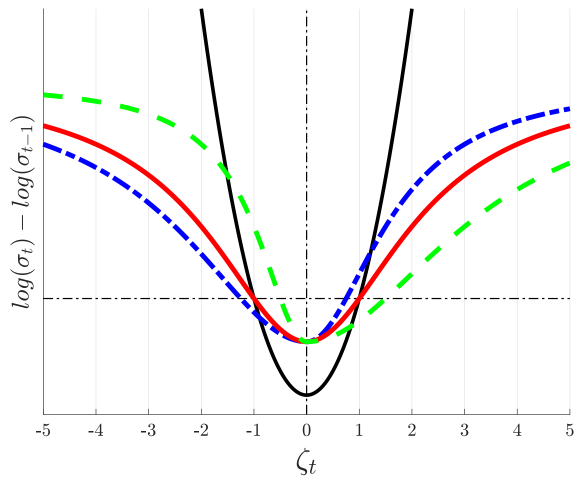
The figure reports the weighting scheme implied by w_t, and the news impact curve (NICs) for different values of the prediction error $\zeta_t = \varepsilon_t/\sigma_t$. We consider the Gaussian (black), the Student-t with $\nu = 5$ (red), and positively (blue) and negatively (green) Skew-t with $\nu = 5$.



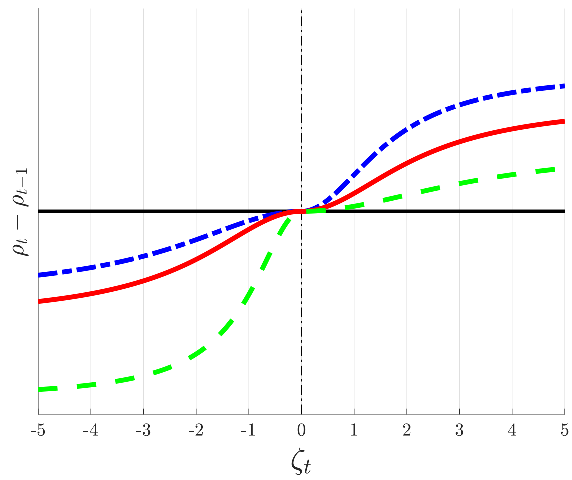
(a) Updating weights



(b) Location



(c) Scale



(d) Shape

The remaining plots display Engle and Ng (1993)'s *news impact curve*, i.e. how new information – measured by the standardised prediction error – translates into updates of the parameters of the model. The location parameter (top-right panel) updates in the direction of the prediction error. Updates of the scale parameter (bottom-left panel) are positive whenever the prediction error is larger than the scale of the distribution, appropriately adapted to account for the difference in positive and negative dispersion. Finally, the shape parameter (bottom-right panel) updates in the opposite direction of the prediction error, so that for negative prediction errors the distribution becomes more left skewed. Yet the "news" contained in a given prediction error depends on how "unlikely" a priori is such news, given the ex-ante conditional distribution of returns, and whether the prediction error is perceived to be a tail observation. In fact, when the underlying distribution is fat tailed, prediction errors that are large – given the scale of the underlying distribution –

are discounted, as they are partially characterised as "outliers" and, as such, are associated with smaller updates of the underlying distribution. For the location parameter (top-right panel), this property translates into the typical S-shaped function of the location in contrast with a classical linear updating in a Gaussian setting (see, e.g., Harvey and Luati, 2014). The asymmetry of the distribution also plays a key role in mapping the prediction errors onto the updating mechanism. When the distribution is left skewed, a positive (negative) prediction error is ex-ante less (more) likely, and therefore when observed it commands stronger (weaker) revisions in the underlying distribution. The opposite holds for right-skewed returns.

The joint role of the conditional estimates in the updating mechanism of the parameters allows for timely detecting of shifts in the shape of the conditional distribution of the returns, while at the same time discounting the effect of outlying observations. In addition, while the scores for the location and shape parameters are negatively correlated, updates of σ_t are (unconditionally) uncorrelated with revisions of the other parameters. Yet, during crashes, when prediction errors are large and negative, updates on the scale and the shape parameters positively co-move, so that the conditional distribution of the momentum returns features negative shifts in the location, increasing dispersion and deepening negative skewness.

D Moments of the Skew-t distribution

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):

$$p(y|\mu, \sigma, \rho, \nu) = \frac{\mathcal{C}}{\sigma}\left[1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma(1+sgn(y-\mu)\rho)}\right)^2\right]^{-\frac{1+\nu}{2}}, \tag{D7}$$

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 D7 as a two-piece distribution (Fernández and Steel, 1998):

$$p(y|\mu, \sigma, \rho, \nu) = \begin{cases} \frac{\mathcal{C}}{\sigma}\left[1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma_+}\right)^2\right]^{-\frac{1+\nu}{2}}, \; y \geq \mu \\ \frac{\mathcal{C}}{\sigma}\left[1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma_-}\right)^2\right]^{-\frac{1+\nu}{2}}, \; y < \mu \end{cases} \tag{D8}$$

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

$$P(y \geq \mu) = \frac{\sigma_+}{\sigma_+ + \sigma_-} = \frac{1+\rho}{2}, \qquad P(y < \mu) = \frac{\sigma_-}{\sigma_+ + \sigma_-} = \frac{1-\rho}{2}. \tag{D9}$$

The two-piece formulation allows to consider separately the two half 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).[23] Specifically:

$$\mathbb{E}[\zeta^r] = \hat{\mu}_r = \frac{1}{2}\left[(1+\rho)^{r+1} + (-1)^r(1-\rho)^{r+1}\right] d_r(\nu), \tag{D10}$$

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 et al., 1995). Starting from D10, 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 y is given by:

$$\begin{aligned}\mathbb{E}[y] &= \mu + \hat{\mu}_1 \sigma \\ &= \mu + \frac{4\nu\mathcal{C}(\nu)}{\nu - 1}\rho\sigma, \qquad \nu > 1\end{aligned} \tag{D11}$$

and the variance is calculated as:

$$\begin{aligned}\mathbb{E}[y^2] &= \mu^2 + 2\mu\sigma\hat{\mu}_1 + \sigma^2\hat{\mu}_2 \\ &= \mu^2 + 2\mu\sigma\frac{4\nu\mathcal{C}(\nu)}{\nu-1}\rho + \sigma^2\frac{(1+3\rho^2)\nu}{\nu-2}, \qquad \nu > 2\end{aligned} \tag{D12}$$

$$\begin{aligned}Var(y) &= \mathbb{E}[y^2] - \mathbb{E}[y] \\ &= \mu^2 + 2\mu\sigma\frac{4\nu\mathcal{C}(\nu)}{\nu-1}\rho + \sigma^2\frac{(1+3\rho^2)\nu}{\nu-2} - \left(\mu + \frac{4\nu\mathcal{C}(\nu)}{\nu-1}\rho\sigma\right)^2 \\ &= \sigma^2\left(\frac{(1+3\rho^2)\nu}{\nu-2} - \left(\frac{4\nu\mathcal{C}(\nu)}{\nu-1}\rho\right)^2\right) \\ &= \sigma^2\left[\frac{\nu}{\nu-2} + \left(\frac{3}{\nu-2} - \left(\frac{4\nu\mathcal{C}(\nu)}{\nu-1}\right)^2\right)\rho^2\right], \qquad \nu > 2\end{aligned} \tag{D13}$$

E Asymmetric betas and returns asymmetry

In this section we provide some simple intuition on how a state-dependent CAPM with asymmetric market betas can generate asymmetry in the marginal distribution of returns. Let consider the conditional regression model in Eq.(13),

$$r_t = \alpha + \underbrace{\overline{\beta} m_t I(m_t \geq \mu_m) + \underline{\beta} m_t I(m_t < \mu_m)}_{\beta m_t} + e_t \tag{E1}$$

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

with $m_t \sim \mathcal{N}\left(\mu_m, \sigma_m^2\right)$ 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. Theoretically, Ang et al. (2006) show that this upside vs downside CAPM formulation can be rationalised based on a disappointment aversion utility function that embeds downside risk following Gul (1991).

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; del Castillo and Daoudi, 2009),

$$f\left(\beta m_t\right) = \begin{cases} C \exp\left\{-\frac{1}{2\underline{\sigma}_m^2}\left(\underline{\beta} m_t - \beta\mu_m\right)^2\right\} & \text{if } m_t \leq \mu_m \\ C \exp\left\{-\frac{1}{2\overline{\sigma}_m^2}\left(\overline{\beta} m_t - \beta\mu_m\right)^2\right\} & \text{if } m_t > \mu_m \end{cases} \tag{E2}$$

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

$$E\left[\beta m_t\right] = \sqrt{\frac{2}{\pi}}\left(\overline{\sigma}_m - \underline{\sigma}_m\right) + \beta\mu_m, \tag{E3}$$

Notice that for $\underline{\beta} = \overline{\beta} = \beta$, then we have $\overline{\sigma}_m^2 = \underline{\sigma}_m^2 = \sigma_m^2$, such that $E\left[\beta m_t\right] = \beta\mu_m$. That is, the mean and the mode of the conditional distribution of the momentum returns coincide, i.e., $E\left[r_t\right] = \alpha + \beta\mu_m$. Similarly, the variance of the split-Normal in Eq.(E2) takes the form ,

$$V\left[\beta m_t\right] = \left(1 - \frac{2}{\pi}\right)\left(\overline{\sigma}_m^2 - \underline{\sigma}_m^2\right)^2 + \overline{\sigma}_m\underline{\sigma}_m \tag{E4}$$

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 $V\left[\beta m_t\right] = \sqrt{\beta^2\sigma_m^2}\sqrt{\beta^2\sigma_m^2} = \beta^2\sigma_m^2$. 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 $r_t \sim \mathcal{N}\left(\alpha + \beta\mu_m, \beta^2\sigma_m^2 + \sigma_e^2\right)$. 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 $E\left[\beta m_t\right]$ and the mode $\beta\mu_m$, which is given by

$$\begin{aligned} E\left[\beta m_t\right] - \beta\mu_m &= \sqrt{\frac{2}{\pi}}\left(\overline{\sigma}_m - \underline{\sigma}_m\right) \propto \sqrt{\overline{\beta}^2\sigma_m^2} - \sqrt{\underline{\beta}^2\sigma_m^2}, \\ &= \sigma_m\left(\sqrt{\overline{\beta}^2} - \sqrt{\underline{\beta}^2}\right) \\ &= \sigma_m\left(\overline{\beta} - \underline{\beta}\right) \end{aligned} \tag{E5}$$

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.