Time-Varying Skewness and Momentum Crashes*

Daniele Bianchi† Andrea De Polis‡ Ivan Petrella§

Abstract

The return on conventional momentum portfolios exhibits a predominantly negative, time-varying skewness, which deepens during momentum "crashes". This has important implications for the portfolio risk-return trade-off: the relationship between the expected return and volatility is time-varying and depends on conditional skewness. We explore the economic underpinnings of time-varying skewness by timing the capital exposure to a momentum portfolio in response to fluctuations in risk. The results show that a dynamic skewness-adjusted maximum Sharpe ratio strategy outperforms popular volatility targeting approaches. Finally, we show that momentum skewness cannot be fully reconciled with an asymmetric exposure to upside and downside market risk.

Keywords: Time-varying skewness, Momentum investing, Risk-return trade-off, Asset pricing.

JEL codes: G10, G11, G12, C22.

*We are thankful to Juan Arismendi-Zambrano, Nick Baltas (discussant), Victoria Dobrynskaya, Deniz Erdemlioglu, Marcelo Fernandes, Harald Hau, Martin Iseringhausen, Irina Kaminska (discussant), Daniele Massacci, Andrew Patton, Manuela Pedio, Francesco Roccazzella, Olivier Scaillet, Fabio Trojani, Irina Zviadadze and seminar participants at the Geneva Finance Research Institute, University College Dublin, Warwick Business School, Fulcrum Asset Management, King's College Business School, University of St.Andrews School of Economics and Finance, IESEG School of Management, University of Bristol Business School, the FES-ICEF international seminar series at HSE University, the 9th Asset Pricing workshop at University of York, the 74th European meeting of the Econometric Society, the 15th annual SoFiE conference, and the 9th Annual Conference of the International Association for Applied Econometrics at BI Norwegian Business School for their helpful comments and suggestions.

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

‡University of Strathclyde, ESCoE and Fulcrum Asset Management. E-mail: andrea.de-polis@strath.ac.uk

§Warwick Business School, University of Warwick & CEPR. E-mail: ivan.petrella@wbs.ac.uk Web: ivanpetrella.com

1 Introduction

One of the most studied capital markets phenomena is the relationship between the future return on a given asset and its past relative performance, termed *momentum* effect. A simple portfolio that buys the past "winners" and sells the past "losers" has historically delivered a competitive risk-adjusted return in US equity markets and has become central to the market efficiency debate, at least since Jegadeesh (1990).[1] Despite a strong historical performance, a conventional momentum portfolio is subject to rare yet predictable large drawdowns relative to the market, referred to as momentum "crashes" (e.g., Daniel and Moskowitz, 2016).

A popular approach to mitigate the economic impact of these crashes builds upon the intuition that the capital exposure to the momentum portfolio can be dynamically adjusted by timing the risk associated with the strategy performance (e.g., Barroso and Santa-Clara, 2015). We build upon this line of research and offer a novel perspective on the risk associated with momentum investing. Our approach is based on the assumption that if the portfolio return displays time-varying skewness, time-varying volatility alone may not provide a complete representation of the strategy risk. As a result, a capital adjustment which explicitly ignores the skewness dynamics may be sub-optimal as investors with asymmetric risk preferences may even accept a lower return if they can trade it off against lesser downside risk.

The reason why skewness may play an important role in managing momentum risk is intuitive, although often unappreciated: high volatility is not always associated with large negative returns, and significant negative returns can occur in periods when volatility is subdued, and the risk is negatively skewed. In fact, risk is not necessarily symmetric over time (e.g., Bollerslev et al., 2022). Perhaps surprisingly, the explicit role of skewness has mostly been overlooked in managing momentum risk.

Before discussing our main findings, two comments are in order. First, it is essential to highlight

[1] Jegadeesh (1990) first document that stocks that performed well in the past tend to outperform the market. In contrast, stocks that performed poorly tend to underperform. 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 (e.g., Moskowitz and Grinblatt, 1999; Moskowitz et al., 2012; Asness et al., 2013).

that our focus is not to modify the construction of a momentum strategy (e.g., Byun and Jeon, 2023) but rather to design a dynamic capital adjustment to equity momentum portfolios following the blueprint of Daniel and Moskowitz (2016). Second, although our results support the view that the skewness in momentum return can be partly rationalised based on asymmetric risk preferences, our objective is not to provide a structural interpretation of the momentum premium but rather to highlight the importance of conditional skewness to understand better the risk associated with momentum investing.

1.1 Main findings

We estimate a time-varying parameter model that recovers the return distribution's location, scale, and asymmetry over time to tease out the dynamics of conditional skewness in momentum returns. This allows us to explain skewness's role in the momentum risk-return trade-off and derive a skewness-hedging component within a conventional maximum Sharpe ratio strategy. Overall, our paper's contribution is threefold.

First, we uncover a significant, pro-cyclical time variation in the conditional skewness associated with the return on a conventional momentum portfolio á-la Jegadeesh and Titman (1993). The return asymmetry tends to be negligible during economic expansions, while it becomes predominantly negative towards the tail of recession periods. This pattern is exacerbated during momentum crashes, whereby spikes in return volatility coincide with deepening downside risk. This suggests that conditional skewness may be important for timing the return on a momentum factor.

Second, we explore the role of conditional skewness in the risk-return trade-off of a momentum strategy. We highlight that any evidence of negative risk-return trade-off is entirely driven by crash periods, and its dynamic is generally shaped by the strategy's return skewness. This echoes the intuition in Theodossiou and Savva (2016), which argues that in skewed returns, the risk premium features a "pure risk" component and a "skewness risk" premium component. Hence, the former can be interpreted as the prevailing risk-return trade-off absent any conditional skewness. Our results suggest that since the direction of the skewness premium hinges on the sign of the skewness, the

overall risk-return trade-off can range from positive to negative over time, depending on the strength of the returns' asymmetry. This evidence can help to rationalise the rather flat unconditional risk-return trade-off previously reported in the literature (e.g., Barroso and Maio, 2023).

Our third result relates to the economic value of capturing time-varying skewness for managing momentum risk. Consistent with the intuition of distinguishing between "pure risk" and "skewness risk" in the presence of conditional skewness, we derive a dynamic skewness-hedging maximum Sharpe ratio strategy expanding on Daniel and Moskowitz (2016). Specifically, during periods of highly negative (positive) conditional skewness, our approach leads to a decrease (increase) of the capital exposure to a momentum portfolio larger than what would be implied by volatility scaling alone. We show empirically that our approach fares better than leading volatility-managed momentum portfolios, especially regarding the exposure to downside risk. This suggests that, by accounting for time-varying asymmetry, one can reduce the impact of low-probability large drawdowns on momentum profitability without giving up any significant risk-adjusted return.

The main empirical results hold for short-term and intermediate momentum portfolios when considering transaction costs for different leverage constraints in a post-1950 out-of-sample period and in the context of monthly capital adjustments. Notice that our framework is general and can be applied to any other factor or anomaly-based portfolios. Our focus on momentum portfolios is primarily led by the existing evidence that shows how volatility scaling is mainly beneficial for momentum portfolios, while the evidence for other factors/anomalies is far less clear (e.g., Cederburg et al., 2020; Barroso and Detzel, 2021).

Finally, it is worth mentioning that although providing a structural interpretation of the momentum premium is beyond the scope of the paper, we nevertheless attempt to highlight some asset pricing implications that can be drawn from our results. Specifically, we build upon Grundy and Martin (2001) and show that the estimated skewness of momentum return can only be partly reconciled by a CAPM with asymmetric exposure to upside and downside market risk. The latter can be framed as a reduced-form representation of an equilibrium asset pricing model in which a representative agent is endowed with a disappointment-aversion utility function (e.g., Ang et al.,

2006). This poses a challenge for asset pricing models that overlook higher-order moments' role in shaping momentum risk premiums.

1.2 Closely related 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 properties of momentum 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); Ehsani and Linnainmaa (2022). Jacobs et al. (2015) uncover a robust relationship between expected skewness and cross-sectional momentum. Yet Theodossiou and Savva (2016) highlight how the evidence on the shape of the risk-return trade-off in momentum strategies has often been inconclusive. They argue that such ambiguity stems from the fact that volatility and skewness have an offsetting impact on the strategy's expected return.

A second strand of literature we contribute to relates to the role of skewness as an input for investment decisions (e.g., Patton, 2004; Guidolin and Timmermann, 2008; Bollerslev et al., 2022) and asset pricing models (e.g., Harvey and Siddique, 2000; Dittmar, 2002). Building on Barroso and Santa-Clara (2015); Cederburg et al. (2020), recent evidence from Wang and Yan (2021); Hanauer and Windmüller (2023) suggests that by scaling factor portfolio return by downside volatility alone, one can improve upon simple volatility scaling. Our results show that the time-varying interplay between conditional expected return, volatility and skewness can offer novel insights into the dynamic of momentum risk and, thus, significant economic gains compared to popular volatility-managed momentum portfolios.

2 Skewness in US equity momentum

We follow Daniel and Moskowitz (2016) and form portfolios based on all-firm breakpoints; that is, an equal number of firms is present in each decile portfolio, rather than an equal number of NYSE firms as in Fama and French (1996). Stocks are sorted into deciles, ranked based on their performance

over the past J months. A conventional momentum strategy involves 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 as the formation period to avoid the short-term reversal (e.g., Jegadeesh, 1990).

Figure 1 compares the cumulative performance of investing 1\$ in the winners-minus-losers (`WML`) portfolio with a look-back period of 12 months, i.e., the 12_2 momentum of Jegadeesh and Titman (1993), against a buy-and-hold investment in the market portfolio and the risk-free rate. The market portfolio is the value-weighted index of all the CRPS firms, and the risk-free rate is the 1-month T-bill rate.[2] The performance is calculated from the second half of the 1920s holding the investment until the end of 2020. Momentum decile portfolios are rebalanced monthly, but returns are calculated daily as in Daniel and Moskowitz (2016). Despite a strong performance, the 12_2 portfolio has experienced a few severe drawdowns – e.g., -65% assuming a 1\$ investment in the portfolio at the beginning of 1932 and 2009 – followed by prolonged periods underperforming the market.[3]

Despite the risk-adjusted return remaining large and significant, these sporadic but large and persistent losses, dubbed momentum "crashes", induce significant asymmetry in the momentum return's distribution. Table 1 shows that the Sharpe ratio for the momentum strategy is 0.78 annualised, almost double that of the market portfolio.[4] A higher risk-adjusted return is not due to higher exposure to market risk, with the CAPM beta being slightly negative ($\beta = -0.15$, $pval = 0.000$). Yet, the return unconditional skewness, defined as the standardised third moment of the sample distribution, is highly negative and significant as shown by the p-value obtained from the Bai and Ng (2005) test.[5]

[2]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`

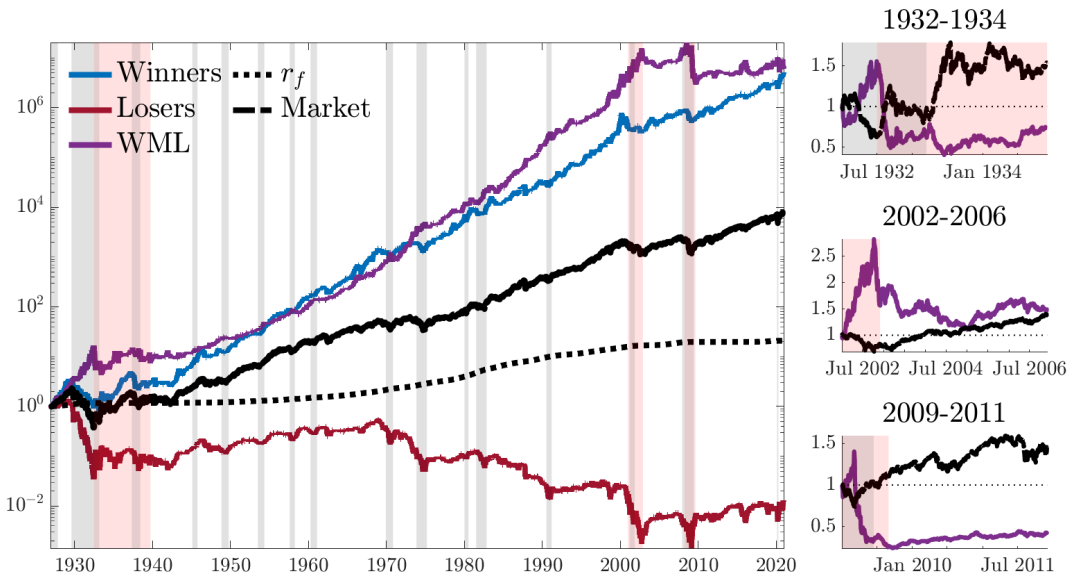
[3]At a monthly frequency, momentum crashes entail losses ranging from -90% to -75% over the same period.

[4]For comparison with Daniel and Moskowitz (2016), we do not consider transaction costs in calculating the performance of the `12_2` portfolio. When adding reasonable transaction costs, the performance of the standard momentum strategy deteriorates (e.g., Novy-Marx and Velikov, 2016; Patton and Weller, 2020; Barroso and Detzel, 2021).

[5]Under the null hypothesis of no return asymmetry, the Bai and Ng (2005) test statistic is $\widehat{\pi}_3 = \frac{\sqrt{T}\widehat{\mu}_3}{s(\widehat{\mu}_3)} \xrightarrow{d} N(0,1)$ with $\widehat{\mu}_3$ a sample estimate of the third central moment of the return distribution and $s\left(\widehat{\mu}_3\right) = \left(\widehat{\boldsymbol{\alpha}}_2 \widehat{\Gamma}_{22} \widehat{\boldsymbol{\alpha}}_2'\right)^{\frac{1}{2}}$. 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 (lower-right) 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 centred first three

Figure 1: **US equity momentum over the last century**

The plot reports the cumulative performance of a 12_2 momentum strategy, the market and a 1-month T-bill rate. The cumulative performance is reported on a logarithmic scale. The right panels zoom in on 1932-1934, 2001-2006 and 2009-2011 momentum crashes by re-scaling the initial investment to 1$ at the beginning of the period. Grey-shaded bands highlight NBER recessions. Red shaded bands indicate momentum crash periods, as indicated in Daniel and Moskowitz (2016).



The presence of asymmetry in the return's distribution is confirmed when discounting the effect of outliers by using the Bowley (1926) measure, defined as $QS_\alpha = \frac{q(\alpha)+q(1-\alpha)-2q(50)}{q(\alpha)-q(1-\alpha)}$, where $q(\alpha)$ is the αth quantile and $q(50)$ the median. The QS_{99} points towards a marked negative asymmetry of -0.108, twice as large as the market portfolio. This indicates that the left tail of the distribution accounts for 55% of the total dispersion of the return, whereas the right tail accounts for 45%; that is, the downside risk is approximately 20% larger than the upside risk over the entire sample.

Table 1 also reports the descriptive statistics for the return on two alternative cross-sectional momentum strategies proposed in the literature, meaning the short-term momentum (6_2), formed based on a six-month look-back period (e.g., Jegadeesh and Titman, 1993), and the intermediate momentum (12_7), formed based on past returns from months $t-12$ through $t-7$ as proposed by Novy-Marx (2012). When examining the return asymmetry from these alternative momentum

moments from their theoretical (Gaussian) counterparts.

Table 1: **A snapshot of the skewness in US equity momentum**

This table reports different descriptive statistics (Panel A) and measures of skewness (Panel B) for past `winners`, `losers`, and `WML` portfolios for three alternative 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 full sample period is from January 1st 1927, to December 31st, 2020, daily.

Panel A: Sample descriptive statistics

	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	

Panel B: Skewness measures

Skewness										
Full sample	0.147	-0.680	-1.230	0.234	-0.714	-1.548	-0.051	-0.744	-0.760	-0.476
	(0.264)	(0.022)	(0.001)	(0.184)	(0.018)	(0.001)	(0.102)	(0.028)	(0.021)	(0.059)
1932-1937	0.644	0.050	-0.215	1.254	-0.395	-1.747	0.233	-0.181	-0.161	0.356
2000-2005	-0.064	-0.185	-0.403	-0.017	-0.284	-1.371	-0.349	-0.371	-0.533	-0.276
2008-2012	0.001	-0.321	-1.742	0.103	-0.285	-0.796	0.289	0.075	-0.378	0.064
QS_{99}										
Full sample	0.021	-0.108	-0.108	0.002	-0.091	-0.096	-0.024	-0.079	-0.089	-0.045
1932-1937	0.117	-0.090	-0.227	0.109	-0.095	-0.239	0.016	-0.009	-0.110	0.057
2000-2005	-0.011	0.031	-0.094	0.016	-0.017	-0.117	-0.030	-0.010	-0.091	-0.100
2008-2012	0.006	-0.114	-0.204	-0.027	-0.090	-0.176	-0.081	-0.031	-0.141	0.084

portfolios, we find evidence that all three long-short strategies present roughly the same skewness profile. The return on the 12_7 strategy have a sample skewness of −0.768 ($pval = 0.021$), which is lower than both the 12_2 strategy ($skew = -1.236$, $pval = 0.001$) and 6_2 ($skew = -1.554$, $pval = 0.001$). The QS_{99} measure is approximately the same across portfolios, i.e., -0.11 for 12_2, -0.096 for 6_2, and -0.089 for 12_7.

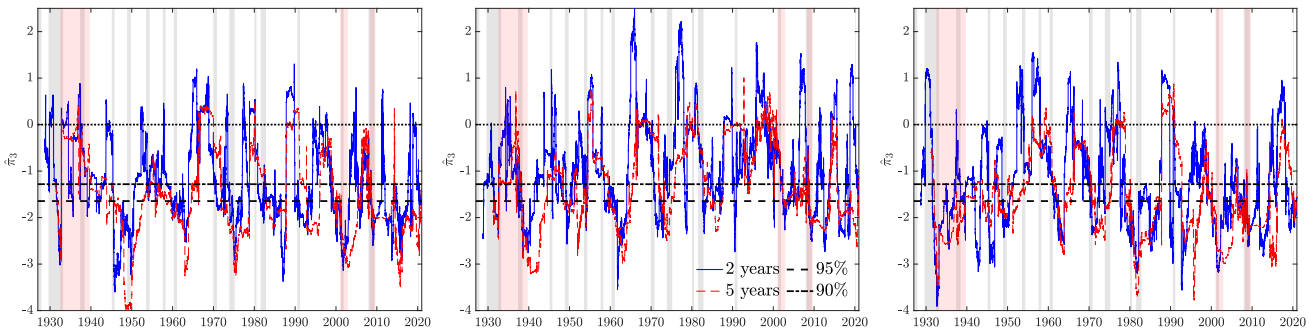
Table 1 also highlights a negative correlation between CAPM alphas and the return skewness. For instance, past winners and the `WML` strategy all show significant, negative (positive) skewness (alphas), with the long-short portfolios displaying, on average, twice the asymmetry and CAPM

alphas compared to past winners. Instead, past losers show a highly negative alpha and a very small, positive skewness. This holds across different momentum portfolios and confirms that the strategy's profitability comes at the cost of substantial downside risk.

A preliminary gauge of the time variation in the asymmetry of the momentum return distribution can be obtained by investigating the realised skewness pattern over different return windows. Panel B of Table 1 reports various skewness measures calculated over five years centred around the three crash periods as indicated by Daniel and Moskowitz (2016). The estimates suggest substantial differences in return asymmetry over different periods. For instance, the 1932 and 2009 crashes exhibit a quantile skewness QS_{99} that is twice as large as the overall sample's. This holds across different momentum portfolios.

Figure 2: **Recursive skewness test**

The three panels report the time series of the Bai and Ng (2005) test statistics for asymmetry, $\widehat{\pi}_3 = \frac{\sqrt{T}\widehat{\mu}_3}{s(\widehat{\mu}_3)}$, over different rolling window of return. We report the testing results by using two and five years of daily return on the `WML` strategy for the `12_2`, `12_7` and the `6_2` momentum. The dashed horizontal lines represent the one-sided test's 90% and 95% critical values. Grey-shaded areas identify NBER recessions, while red-shaded areas highlight momentum crash periods, as indicated in Daniel and Moskowitz (2016). The sample period is daily from January 1st 1927 to December 31st 2020.



(a) `12_2` momentum (b) `12_7` momentum (c) `6_2` momentum

Figure 2 expands on the sub-sample estimates from Table 1 and reports the results of the Bai and Ng (2005) test statistics for asymmetry calculated over different rolling windows of two and five years of daily return. The test statistics consistently show negative values of the standardised third moment of the distribution across alternative momentum portfolios. Yet, they also exhibit a substantial time variation, ranging from near zero to highly negative values. The dashed horizontal

lines represent the 90% and 95% critical values associated with the null hypothesis of no asymmetry against the alternative of negative asymmetry.

The null hypothesis of no asymmetry is often rejected over the sample; there are multiple periods in which the momentum return shows negative and significant skewness. For instance, the asymmetry of the `12_2` portfolio return is negative and significant throughout the momentum crash of the 1930s, while it becomes non-significant again over the following decade. The rolling window estimates also highlight a substantial co-movement of the Bai and Ng (2005) test statistics across different momentum portfolios. Specifically, the correlation of the test statistics between the 12_2 and the 12_7 (6_2) momentum portfolios is 0.72 (0.65) over the entire sample.

To complement the simple recursive Bai and Ng (2005) test in Figure 2, in Appendix A we report the results of a more formal likelihood-based test whereby we examine whether the conditional skewness of the `12_2` portfolio return remains constant against the alternative of time-varying skewness.[6] The null hypothesis of constant skewness is firmly rejected, with p-values well below the canonical 1% threshold. In the next section, we take stock of this preliminary evidence and introduce a novel modelling framework which allows us to explicitly track the time-varying nature of the return asymmetry of momentum portfolio return.

3 Modelling time-varying skewness

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

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

The shape parameter $\rho_t \in (-1, 1)$ captures the extent of asymmetry of the portfolio return. Positive (negative) values of ρ_t imply a positively (negatively) skewed return at time t, and the ratio $\frac{1+\rho_t}{1-\rho_t}$

[6] In practice, we test the information contained in the score of the log-likelihood function estimated under the null hypothesis of no time-variation. See Appendix A for more details.

defines the probability mass on the right versus on the left of the location $\mathtt{m}_t$.

Equation (1) nests standard distributional assumptions as limiting cases. For instance, by restricting $\rho_t = 0$ we obtain the symmetric Student-t distribution. With $\nu \to \infty$ and $\rho_t = 0$, the conditional distribution coincides with a Normal with time-varying mean and variance. Finally, with $\nu \to \infty$ and $\rho_t \neq 0$ we retrieve the Skew-Normal distribution of Mudholkar and Hutson (2000). As all these parameters are estimated from the data, our model does not impose but allows for time-varying skewness in the return's conditional distribution.

We follow Creal et al. (2013) and Harvey (2013) and assume the dynamics of $\mathtt{m}_t$, σ_t and ρ_t is entirely observation-driven in the sense of Cox (1981), meaning that the time variation of the parameters is a direct function of past prediction errors. In order to ensure that the scale σ_t is positive and the shape $\rho_t \in (-1, 1)$, we adopt the transformations $\gamma_t = log\,(\sigma_t)$ and $\delta_t = \text{arctanh}\,(\rho_t)$. The vector of time-varying parameters $f_t = (\mathtt{m}_t, \gamma_t, \delta_t)'$ is updated at each time t based on the law of motion,

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

where A contains the structural parameters regulating the sensitivity of f_t to the information contained in the *scaled score* $s_t = \mathcal{S}_t \nabla_t$. Here $\mathcal{S}_t$ is a scaling matrix proportional to the diagonal of the information matrix $\mathcal{I}_t = E\left[\nabla_t \nabla_t'\right]$, such that $\mathcal{S}_t = (J_t' \text{diag}(\mathcal{I}_t) J_t)^{-1}$, and $\nabla_t = J_t' \left[\frac{\partial \ell_t}{\partial \mathtt{m}_t}, \frac{\partial \ell_t}{\partial \sigma_t^2}, \frac{\partial \ell_t}{\partial \rho_t}\right]'$ the gradient of the log-likelihood function for the time-varying parameters. The Jacobian matrix J_t maps the transformations γ_t and δ_t into the original time-varying scale σ_t and shape ρ_t parameters.

The scaled score translates the information summarized by the prediction errors at time t into an update of f_t. Specifically, given Eq. (1) and the conditional log-likelihood (see Eq. (B7) in Appendix B.4), the elements of s_t are defined as:

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

where sign$(\cdot)$ is the sign function, $\chi = \frac{(\nu+3)}{(\nu+1)}$ and $w_t = (\nu+1)^{-1}\left(\nu\left(1+\mathrm{s}\left(\varepsilon_t\right)\rho_t\right)^2+\zeta_t^2\right)$ represents the weights to the standardised prediction errors $\zeta_t{=}\frac{\varepsilon_t}{\sigma_t}$. 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 B.1 and Appendix B.2. We assume that the matrix A in Eq. (2) is diagonal so that the update of each time-varying parameter is proportional to the information conveyed by the likelihood of that specific parameter.

The scalar w_t in Eq. (3) plays a key role as it represents the implicit weight of the information contained in the prediction error (e.g., Harvey and Luati, 2014). More generally, Blasques et al. (2015) show that score-driven updates as in Eq. (3) can reduce the local Kullback-Leibler divergence between the actual, unobserved, conditional density and the corresponding estimate, even when the underlying model is potentially mis-specified.[7]

Maximum likelihood estimates (MLE) of f_t and $\theta = (\nu, A)$ can be obtained via a prediction error decomposition (see Blasques et al., 2022). However, given the random-walk specification of f_t, maximum likelihood tends to put a large point mass at the initial condition, an issue known as the "*pile-up problem*" (e.g., Sargan and Bhargava, 1983; Stock and Watson, 1998). To address this issue, we discipline the parameter space by introducing a minimal set of prior conditions on A and ν, such that any evidence of time variation in f_t must reflect strong evidence in the data. As a result, the resulting estimator produces a maximum a posteriori estimate (see, e.g., Kamen and Su, 2012). Appendix B.4 provides a detailed description of the estimation procedure.

3.1 Expected return and parameters updating

The conditional moments of the return distribution implied by Eq. (1) can be derived as a weighted average of the conditional moments of a Half-t distribution (Arellano-Valle et al., 2005; Gómez et al.,

[7] Related to that, 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 actual data generating process. In this respect, score-driven updates of the time-varying parameters are optimal from an information theoretic perspective.

2007). As a result, the expected return $\mathbb{E}_t(r_{t+1})$ can defined as:

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

where ν denotes the degrees of freedom parameter, $\mathcal{C} = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}$ and $\Gamma(.)$ is the Gamma function. Eq. (4) implies that momentum expected return depend on the scale σ_t and the asymmetry ρ_t at each time t. The location parameter $\mathtt{m}_t$ captures the mode of the conditional distribution and is equivalent to $\mathbb{E}_t(r_{t+1})$ under symmetric distributional assumptions – when $\rho_t = 0$. A full derivation of the expected return in Eq. (4) is provided in Appendix C.

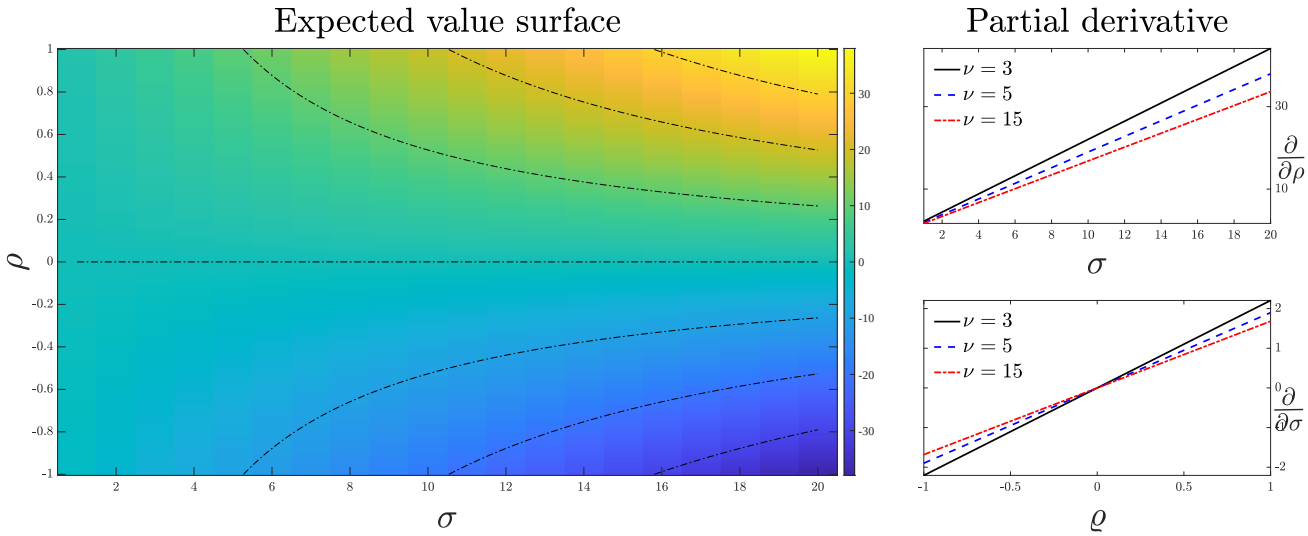
To better understand the role of conditional skewness in our model, Figure 3 shows some comparative statics on the impact of σ_t and ρ_t on $\mathbb{E}_t(r_{t+1})$. Two properties emerge: first, the effect of σ_t on the expected return is amplified by ρ_t. For instance, for a negatively skewed return, i.e., $\rho_t < 0$, the higher the volatility, the lower the expected return (dark blue area). This observation provides an intuitive narrative for the risk associated with momentum investing. The combination of volatility spikes and negative skewness, characterising momentum crashes, can swiftly reverse the strategy's expected return.

The second property that emerges from Figure 3 is that the effect of asymmetry and volatility on the expected return is multiplicative. This means that the curvature of $\mathbb{E}_t(r_{t+1})$ as a function of σ_t increases more than linearly as $|\rho_t|$ increases. The steepness of the curvature is regulated by the degrees of freedom ν (see partial derivative plots). Thicker tails push the portfolio's expected return to more extreme values depending on the conditional asymmetry. It is ρ_t that dictates the sign of the sensitivity of expected return to a change in volatility.

As the time variation of f_t depends on past prediction errors, it is worth discussing how the scale σ_t and asymmetry ρ_t update over time and the effect of these updates on the dynamic of the expected return. Figure 4 displays how new information – measured by the standardised prediction error – translates into changes in the scale and shape parameters. The extent of parameters updating at a given time t depends on the underlying return asymmetry. For instance, when the conditional

Figure 3: **Model-implied expected value surface**

The left panel illustrates the expected value surface for values of ρ_t and σ_t. The smaller panels on the right illustrate the partial derivative of Eq. (4) with respect to ρ_t and σ. Without loss of generality, we report the surface by restricting the location parameter $\mathtt{m}_t$ to zero.



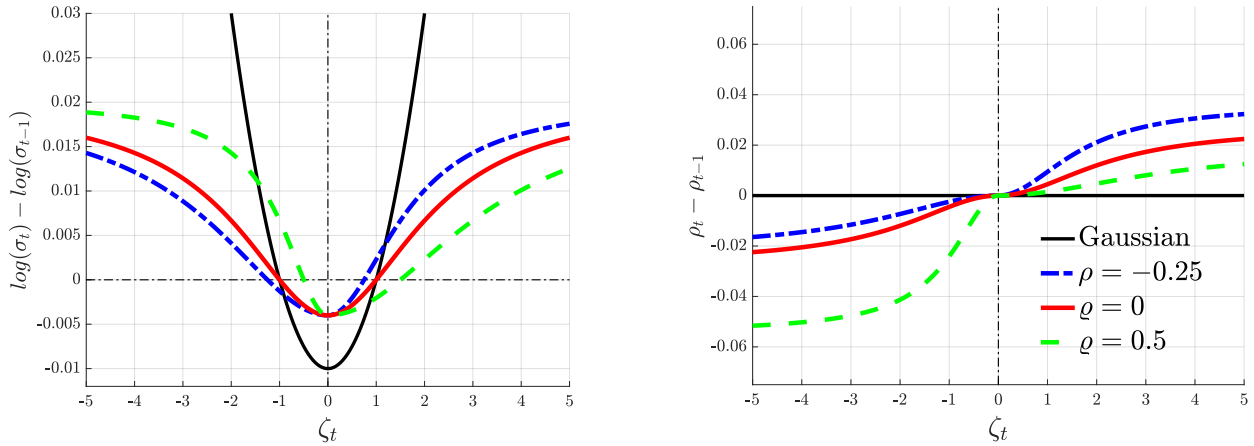
distribution exhibits positive skewness (e.g., $\rho_t = 0.5$, dashed green line), observing a negative prediction error leads to significant adjustments in scale and asymmetry. These adjustments are less pronounced when the return skewness is mildly negative (e.g., $\rho_t = -0.25$, dashed blue line). In the latter case, a negative prediction error is more likely than when the conditional distribution is positively skewed. This determines how σ_t and ρ_t update upon observing the return at time t.

A significant rebound in the strategy's return prompts the model to promptly revise the expected value of future returns due to reassessing the underlying risk profile. This effect becomes more prominent as the strategy return becomes more extreme.[8] As a result, the expected return is responsive to changes in the overall risk profile of the strategy. A reassessment of the risk balance, captured by the asymmetry parameter, especially during periods of high volatility, leads to a quick adaptation of $\mathbb{E}_t(r_{t+1})$ to shifts in the portfolio risk. In the next section, we leverage these model

[8] Except when a return is categorized as a tail event by the model, in which case their informational value is heavily discounted (see Figure B1(a) in Appendix B.3).

Figure 4: **Updating of the scale and asymmetry parameters**

The panels report the news impact curves (NICs) for the scale and shape parameters as functions of the standardised prediction error, ζ_t. We consider values of $\varrho = -0.5, 1, -0.25$. ν is fixed at 5.



(a) Scale updating

(b) Asymmetry updating

features to investigate the role of conditional volatility and skewness in the strategy's risk-return trade-off over time.

4 Time-varying skewness and momentum risk

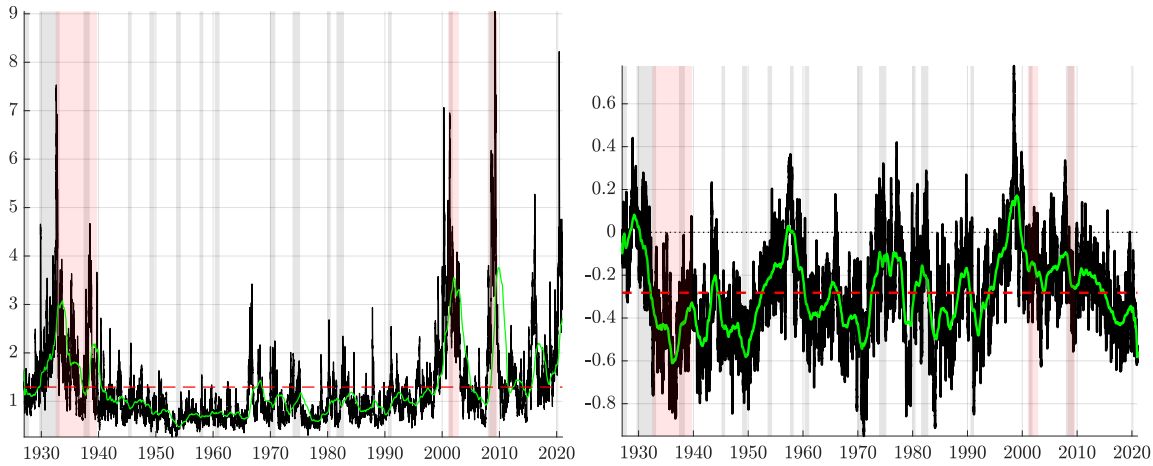
The descriptive statistics in Table 1 suggest that the `12_2` portfolio provides the largest risk-adjusted performance while at the same time reports an equally large negative skewness – even larger – compared to the `12_7` and `6_2` momentum portfolios. In addition, there is a substantial co-movement in the dynamics of skewness across momentum portfolios (see Figure 2). For this reason, we follow Barroso and Santa-Clara (2015); Daniel and Moskowitz (2016); Hanauer and Windmüller (2023) and focus on the `12_2` momentum return. The analysis for the `12_7` and `6_2` portfolios, and the `12_2` long-short strategy based on NYSE breakpoints are discussed in Section 5.1 and Appendix E.

Figure 5 presents the estimates for the `12_2` return conditional volatility $\sqrt{\mathbb{V}_t(r_{t+1})}$ and skewness $Sk_t(r_{t+1})$, respectively. The conditional variance of the return can be derived analytically as $\mathbb{V}_t(r_{t+1}) = \sigma_t^2(\frac{\nu}{\nu-2} + h(\nu)\rho_t^2)$, for $\nu > 3$, where $h(\nu) = \frac{3}{\nu-2} - g(\nu)^2 \gg 0$ gauges the interaction

between the fat-tailedness of the distribution and the asymmetry parameter.[9] For $\rho_t = 0$, the conditional variance $\mathbb{V}_t(r_{t+1})$ reduces to the Student-t variance. The full derivation of $\mathbb{V}_t(r_{t+1})$ and $Sk_t(r_{t+1})$ is provided in Appendix C. We report the daily estimates in black and their two-year average in green to increase readability. The red line represents the sample mean estimate.

Figure 5: **Conditional volatility and skewness in momentum return**

The plot reports the time-varying volatility (left) and skewness (right) estimates for the `12_2 WML` portfolio return. The red dashed lines represent the sample mean, whereas the green lines highlight the 2-year moving averages of the daily estimates. Grey-shaded areas identify NBER recessions, while red-shaded areas highlight momentum crash periods, as indicated by Daniel and Moskowitz (2016). The sample period is from January 1st 1927 to December 31st, 2020.



(a) Conditional volatility

(b) Conditional skewness

The conditional volatility of the momentum return is substantially higher than its sample average during the decade following the Great Depression, the burst of the dot-com bubble, and the period following the great financial crisis of 2008/2009. Perhaps surprisingly, the return conditional volatility remained subdued and relatively stable for almost sixty years, from 1940 to the late '90s, except for a few episodes of short-lived spikes during the '70s and '80s, two decades characterised by a series of economic shocks and subsequent recessions.

Figure 5(b) points towards a pro-cyclical time variation in the asymmetry of the strategy return. Specifically, the conditional skewness tends to zero during economic expansions, while it becomes

[9] The formulation for $Sk_t(r_{t+1})$ is slightly more tedious, and maps directly the asymmetry parameter $\rho_t \in (-1, 1)$ into the unbounded value of the conditional skewness.

more negative during recessions. This pattern is exacerbated during momentum crashes, whereby the return conditional distribution features both an increasing dispersion and a deepening negative skewness akin to a time-varying leverage effect. For instance, the return skewness was largely negative during the crash of 1932- 1939 and significantly dropped from -0.1 to -0.4 towards the end of the great financial crisis. These periods coincide with major peaks in return volatility.

Although with some peculiarities, a similar combination of higher volatility and more negative skewness occurred during the 2001-2002 crash – with skewness collapsing from 0.2 to -0.3 and volatility peaking at 7% daily in annualised terms – and the great financial crisis. Interestingly, the conditional skewness (volatility) remains persistently lower (higher) than its sample mean towards the end of 2020. This coincides with the onset of the COVID-19 pandemic, which may represent the latest episode of a long-lasting flattening in momentum profitability since the early 2000s, as highlighted in Figure 1. Appendix E.1 shows that the estimates of conditional volatility and skewness for the `12_7` and `6_2` momentum portfolios largely align with Figure 5.

4.1 Time-varying skewness and the risk-return trade-off

Figure 5(a) confirms that the strategy's return displays time-varying volatility, a result well-documented in the literature (Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016). In addition, our findings suggest that the momentum return is *also* persistently negatively skewed, with skewness deepening at times as shown in Figure 5(b). In particular, momentum crashes are characterized by increases in volatility with deepening negative skewness of the conditional returns. This is the first indication that conditional volatility alone may not suffice to capture the full extent of momentum risk. But why does skewness matter? We argue that the strength of return asymmetry has important implications for the time variation of the strategy risk-return trade-off. The relationship between volatility and expected returns should be red *in conjunction* with conditional skewness.

We first highlight the role of conditional skewness on the momentum risk-return trade-off by plotting the correlation between the estimated expected return $\mathbb{E}_t(r_{t+1})$ and volatility $\sqrt{\mathbb{V}_t(r_{t+1})}$ over the entire sample. Figure 6(a) supports the evidence shown by Charoenrook and Conrad

(2005); Barroso and Maio (2023) of a mildly negative, if not insignificant, relationship between the momentum expected return and volatility. Yet, such a negative correlation deepens and becomes highly significant during momentum crashes, as highlighted by the red markers in the scatter plot.[10]

Figure 6(b) suggests that one can better understand the sign of the risk-return trade-off by distinguishing between periods with positively and negatively skewed returns. With a positively skewed return distribution, one notes a positive risk-return relationship on average, whereas the opposite arises when the strategy displays negatively skewed returns. Therefore, the often inconclusive evidence on the sign of the unconditional risk-return trade-off may be due to an offsetting impact of conditional volatility and skewness on expected returns.

To interpret these findings, we rearrange Eq. (4) and write the expected return as

$$\mathbb{E}_t(r_{t+1}) = \mathtt{m}_t + \lambda_t \sqrt{\mathbb{V}_t(r_{t+1})} \qquad \text{with} \qquad \lambda_t = \frac{g(\nu)}{\sqrt{\frac{\nu}{\nu-2} + h(\nu)\rho_t^2}} \rho_t \tag{5}$$

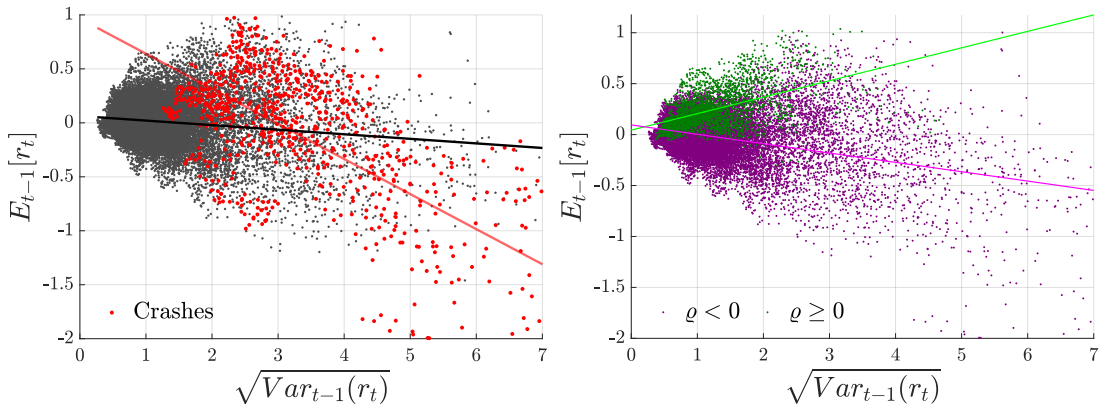
where the slope λ_t is a non-linear function of the time-varying asymmetry ρ_t. Equation (5) echoes the intuition provided by Theodossiou and Savva (2016) whereby in the presence of skewed returns, the mean excess return features a "pure risk" component, which captures the prevailing premium absent any conditional return skewness, and a "skewness risk" component, which aligns with the sign of skewness.

The first component, $\mathtt{m}_t$, in Eq. (5) represents the modal return which should be, theoretically, positively related to risk. The second component takes the value of zero in the absence of skewness and directly maps time-varying asymmetry to the expected returns, and thus can be interpreted as "skewness risk" premium in the jargon of Theodossiou and Savva (2016). Therefore, the direction of the "skewness risk" premium hinges on the sign skewness. As a result, the overall risk-return trade-off can range from positive to zero or negative, depending on the relative strength of the "pure risk" vs"skewness risk", and the sign of the latter.

[10]The regression slopes for the points in Figure 6(b) are -0.05, -0.01, -0.32 for the full sample, the non-crash sample (black) and the crash sample (red), respectively. All coefficients are significant at the 1% level.

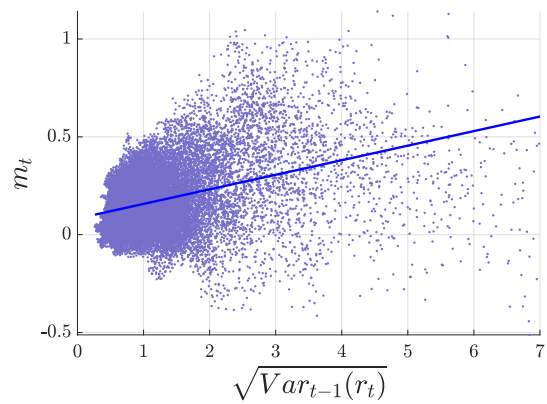
Figure 6: **The dynamic of the risk-return trade-off**

Panel (a) reports the correlation between the estimated expected return $\mathbb{E}_t(r_{t+1})$ and volatility $\sqrt{\mathbb{V}_t(r_{t+1})}$ over the entire sample. We highlight different slopes during momentum crashes (red) and non-crash periods (black). Panel (b) illustrates the risk-return trade-off as a function of the time-varying asymmetry, whereby we colour-code periods with positive vs negative skewness. Panel (c) reports the correlation between the "pure risk" premium component m_t and conditional volatility. Panel (d) shows the change in the price of skewness risk as a function of positive and negative return skewness. The sample is from January 1st 1927 to December 31st 2020.

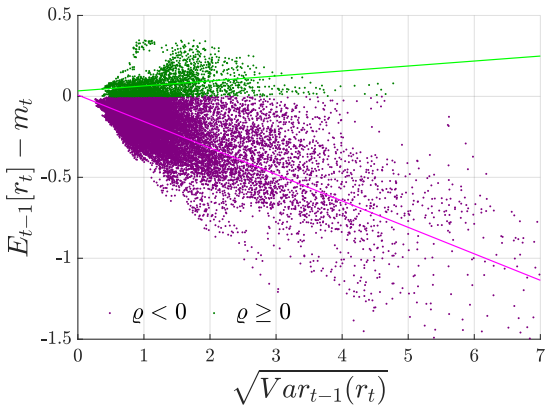


(a) Risk-return trade-off

(b) Pos. vs. Neg. skewness



(c) "Pure risk" premium



(d) "Skewness risk" premium

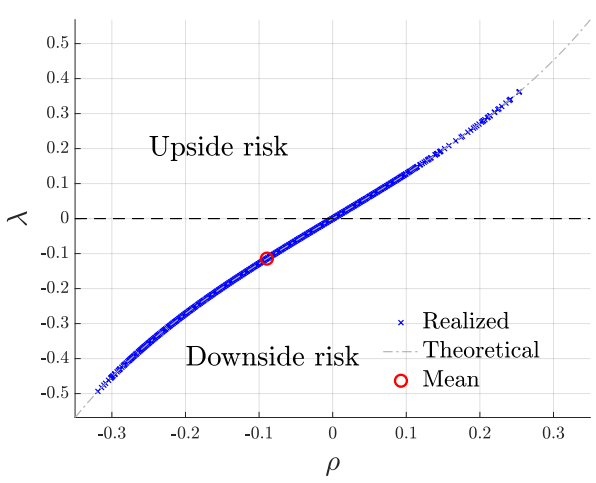
The lower panels of Figure 6 depict a breakdown of the momentum risk-return trade-off into these two components. Consistent with economic theory, Figure 6(c) shows a positive slope in the relationship between the return mode $\mathtt{m}_t$ and conditional volatility $\sqrt{\mathbb{V}_t(r_{t+1})}$. Differently, Figure 6(d) shows that the strategy's "skewness risk" component varies in sign due to changes in skewness over time. On average, the "skewness risk" premium is less pronounced when skewness is positive but stronger when skewness is negative. The two components have a partially offsetting effect on

the expected return when the realised return is negatively skewed, while they reinforce each other for positively skewed returns. As a result, shifts in risk have a more pronounced, negative effect on the expected return when the return asymmetry becomes more negative.

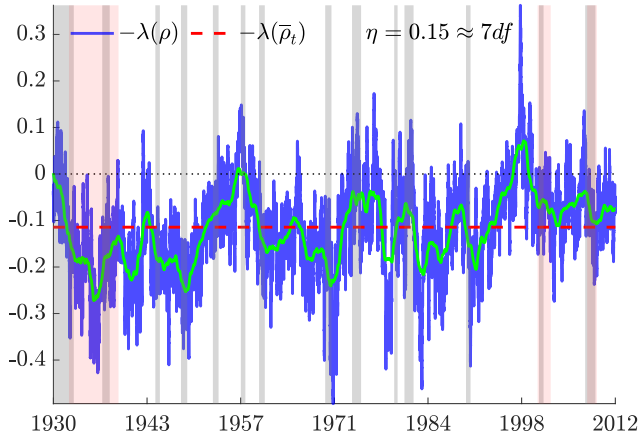
Figure 6 suggests that, in the presence of return asymmetry, the dynamic of the risk-return trade-off does not depend only on conditional volatility. Figure 7(a) formalises this argument by showing the theoretical shape of λ_t as a function of ρ_t (dashed curve), its estimated value over time (blue marks), and its unconditional mean (red circle). The sensitivity of expected returns to "skewness risk" exhibits a more pronounced negative relationship when conditional skewness is negative.

Figure 7: **The role of time-varying skewness**

Panel (a) reports the theoretical shape of λ_t, as a function of ρ_t. The blue marks represent the realized values, and the red circle highlights the mean. In panel (b), we report the time series of λ_t. The blue line represents the daily estimates, the light-blue line a two-year moving average, and the horizontal dashed line the value of λ_t evaluated at the sample mean $\overline{\rho}$. The sample is from January 1st 1927 to December 31st 2020.



(a) λ_t as a function of ρ_t



(b) λ_t dynamics

Figure 7(b) shows that the sign of the risk-return trade-off in momentum investing is generally negative, although it varies considerably over time. Consequently, most of the variation in $\mathbb{E}_t(r_{t+1})$ can be attributed to time-varying risk, provided that the role of risk asymmetry is appropriately accounted for through λ_t. Overall, accounting for time-varying skewness helps to rationalize (a) the unconditional flat relationship between risk and return in momentum, and (b) the pronounced

negative relationship between expected return and volatility during crash episodes. Indeed, the latter are characterized not only by large volatility but also by a more negatively skewed conditional distribution of returns (see Figure 5).

The results in Figures 6-7 help to draw some useful comparisons with Daniel and Moskowitz (2016). They regress the WML return onto the interaction between a bear market indicator and the market variance as a proxy for $\mathbb{E}_t(r_{t+1})$, which implies lower expected return during bear markets. Our model rationalises their findings as a reflection of the role of varying downside risk in the momentum strategy: when volatility is moderate and conditional skewness is negligible, the variation in $\mathbb{E}_t(r_{t+1})$ is small and mostly captured by the "pure risk" component $\mathtt{m}_t$. Instead, when conditional volatility spikes in the lead-up to crash periods, the expected return is depressed due to the interaction with an increasingly negative conditional skewness.

4.2 Sharpe ratio and time-varying skewness

We leverage the definition of expected return in Eq. (5) and investigate further the value of modelling time-varying skewness to understand the dynamic of the momentum risk-adjusted return. Specifically, we decompose the conditional Sharpe ratio, $SR_t = \frac{\mathbb{E}_t(r_{t+1})}{\sqrt{\mathbb{V}_t(r_{t+1})}}$ as

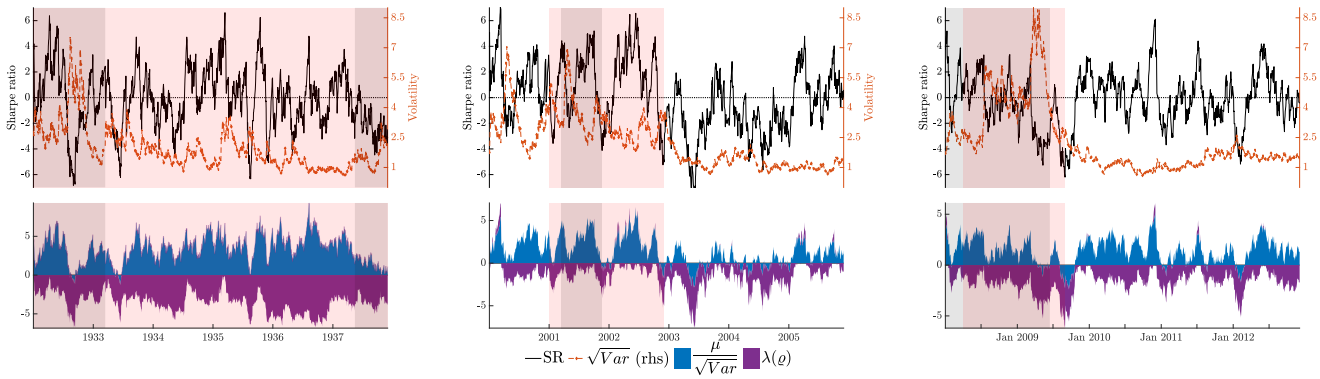
$$SR_t = \frac{\mathtt{m}_t}{\sqrt{\mathbb{V}_t(r_{t+1})}} + \lambda_t \tag{6}$$

such that the time variation of the risk-adjusted return is a function of *both* conditional volatility and skewness. The first component $\frac{\mathtt{m}_t}{\sqrt{\mathbb{V}_t(r_{t+1})}}$ measures the contribution of conditional volatility to the overall strategy risk-adjusted return and is related to the "pure risk" premium component postulated by Theodossiou and Savva (2016) and outlined in Section 4.1. The second component measures the direct role of skewness in the momentum risk-adjusted return based on λ_t (see Eq. (5)).

Figure 8(a) illustrates the dynamics of $\sqrt{\mathbb{V}_t(r_{t+1})}$, SR_t and its two components over the 1932 momentum crash period. As volatility spikes during the great depression, the economic relevance of $\frac{\mathtt{m}_t}{\sqrt{\mathbb{V}_t(r_{t+1})}}$ vanishes, so that SR_t is primarily due to the sensitivity of the "skewness risk" premium

Figure 8: **Sharpe ratio decomposition during crashes**

The plots report in the top panels the conditional Sharpe ratio (black, annualised) against the conditional volatility of the `12_2` momentum return. The bottom panels highlight the decomposition of the conditional Sharpe ratio into a location component ($\mathtt{m}_t/\sqrt{Var}$, blue) and λ_t in purple. The three plots correspond to the 1932, 2001 and 2009 crashes. Grey-shaded bands highlight the NBER recession. Red shaded bands indicate momentum crash periods, as indicated in Daniel and Moskowitz (2016).



(a) 1932 Crash (b) 2001 Crash (c) 2009 Crash

component to a shift in volatility, as captured by λ_t. This leads to an overall negative conditional risk-adjusted return as the negative skewness deepens towards the tail of the great depression. Figure 8(c) shows a similar dynamic during the great financial crisis; that is, as volatility increases, the SR_t becomes negative and primarily driven by conditional skewness. Instead, Figure 8(b) shows that return skewness does not play any role in SR_t during the burst of the dot-com bubble. As volatility trends downward in 2001, the risk-adjusted return becomes positive, with return asymmetry playing a negligible role until 2003, when conditional skewness becomes negative (see Figure 5).

Overall, Figure 8 confirms the intuition that the risk-adjusted return associated with a momentum strategy is not only related to time-varying volatility but is also tightly connected to the dynamic of conditional skewness. When volatility increases and the return distribution shape becomes more tilted towards negative values, the leverage effect leads to an even larger negative conditional Sharpe ratio than that exerted by conditional volatility alone.

5 A skewness-adjusted maximum Sharpe ratio strategy

In this section, we assess the economic significance of modelling time-varying skewness. To this end, we build upon an established literature aiming to improve the profitability of a momentum portfolio by timing the risk associated with the strategy performance (e.g., Barroso and Santa-Clara, 2015 and Hanauer and Windmüller, 2023). Within this setting, during periods of higher (lower) volatility – relative to its sample average – the capital exposure to the `WML` portfolio is reduced (increased) by an amount proportional to the inverse of the previous month's variance. Expanding on this idea, Daniel and Moskowitz (2016) propose a simple dynamic leverage adjustment that maximizes the conditional Sharpe ratio as follows,

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

where $\mathbb{E}_t(r_{t+1})$ and $\mathbb{V}_t(r_{t+1})$ represent some appropriately chosen estimates of the conditional mean and variance of momentum return, and γ is a constant calibrated to match the unconditional volatility of the original momentum return. Eq. (7) illustrates that the capital adjustment ω_t is well described by its dependence on the first two conditional moments of the strategy return (see Appendix C in Daniel and Moskowitz, 2016). Our model directly takes into account the role of the return asymmetry on ω_t in both moments. As a result, the investment rule in Eq. (7) can be separated into two components,

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

The second component, $\omega_{2,t}$, captures the effect of time-varying asymmetry conditional on the scale parameter σ_t. When the conditional distribution of the return is approximately symmetric, i.e., $\rho_t \approx 0$, $\omega_{2,t}$ is economically negligible.[11] Thus, we interpret $\omega_{2,t}$ as a skewness hedging adjustment

[11] With $\rho_t = 0$, we have $\mathtt{m}_t = \mathbb{E}_t(r_{t+1})$ so that $\omega_{1,t}$ would be akin to the adjustment of Daniel and Moskowitz (2016). Note that while ρ_t may have a sizable effect on $\mathbb{E}_t(r_{t+1})$ it has much lower impact on $\mathbb{V}_t(r_{t+1})$ since $\rho_t \in (-1, 1)$ and enters squared in the formula for $\mathbb{V}_t(r_{t+1})$ (see Section 4).

to the original maximum conditional Sharpe ratio strategy of Daniel and Moskowitz (2016). During periods of highly negative (positive) skewness, our dynamic leverage adjustment decreases (increases) the exposure to the `WML` portfolio more than what would be implied by ignoring return asymmetry. For this reason, we label our approach as a maximum "skewness-adjusted" SR (`mSSR`) strategy.[12]

We pit our approach against existing time-series adjustments explored in the literature. First, we consider the constant volatility approach of Barroso and Santa-Clara (2015) (`BS2015` henceforth) whereby the exposure to a momentum strategy is re-scaled based on the six-months realised volatility rv_t^{126} calculated on a rolling-window of daily momentum return. Next, we consider the maximum SR strategy of Daniel and Moskowitz (2016) (`DM2016` henceforth) where $\mathbb{E}_t(r_{t+1})$ is the fitted value of a regression of the `WML` return on the interaction between a bear market indicator and the six-months market realised variance, and $\mathbb{V}_t(r_{t+1})$ is the fitted value of a regression of the 22-days `WML` realised volatility onto rv_t^{126} and a daily asymmetric GARCH estimate. Finally, we consider a semi-volatility targeting as in Wang and Yan (2021); Hanauer and Windmüller (2023) (`cdVol` henceforth).

Each method is tested using a broad range of performance measures. The traditional Sharpe ratio (SR) gives equal weight to the variability of returns above and below the mean and, as such, gives a misleading picture of the underlying risk when a strategy is poised by a marked asymmetry. A simple alternative, the Sortino ratio (Sortino and Van Der Meer, 1991), scales returns by the downside volatility (dVol, i.e. volatility of excess return conditional on being negative). Aside from that, we also calculate a series of performance metrics specifically designed to capture the extent of downside risk across strategies, such as the Stable Tail Adjusted Return Ratio (STARR) – which replaces volatility with the Expected Shortfall (ES) as the denominator in the Sharpe ratio – and the Rachev Ratio (RR) – which represents the ratio between the Expected Longrise (EL) and the ES (e.g., Fabozzi et al., 2005).[13] Lastly, we include the maximum return drawdown (MaxDD), the Value-at-Risk (VaR), and the sample skewness as crude proxies for downside risk.

[12] Kandel and Stambaugh (1996) show that if we assume that only the first two conditional moments matter for portfolio choice, the optimal investment rule under a more general power utility investor would be $\omega_t = \frac{1}{\gamma}\frac{\mathbb{E}_t(r_{t+1})}{\mathbb{V}_t(r_{t+1})} + \frac{1}{2\gamma}$ which is proportional to Eq. (7).

[13] The Rachev ratio captures the asymmetry of the return distribution by assessing the imbalance between extreme losses and gains.

We develop a novel bootstrap procedure to test the performance differential across methods. Specifically, we extend the framework developed by Ledoit and Wolf (2008) to the broad set of downside risk-specific measures outlined above in the presence of both time-series dependence and fat tails. Appendix D contains the complete description of our block-bootstrap method and the simulation results for the optimal choice of the block size.

Before discussing the results, one comment is in order. An alternative approach to gauge the economic performance of time-varying skewness would include higher-order moments in the investors' utility function (e.g., Mencía and Sentana, 2009). While this certainly represents an interesting approach, it prevents a direct comparison with existing volatility targeting methods. Instead, our implementation provides a cleaner setting to gauge the economic value of modelling time-varying return asymmetry while benefiting from the simplicity of the maximum conditional Sharpe ratio adjustment proposed by Daniel and Moskowitz (2016).

Baseline results. We estimate the parameters of the score-driven model, i.e. the degrees of freedom ν and the matrix A in Eq. (2), at the end of every month.[14] Holding those estimates fixed, we update the time-varying parameters $\mathtt{m}_t$, σ_t and ρ_t – used to calculate $\mathbb{E}_t(r_{t+1})$ and $\mathbb{V}_t(r_{t+1})$ in Eq. (7) – on a daily basis for the following month. Therefore, our portfolio strategy avoids look-ahead bias and is fully implementable in real-time. We use three years of daily return as an initial burn-in sample for the recursive estimates and generate the initial portfolio choice on January 1st, 1930. In Section 5.1, we also consider different subsamples with the initial portfolio choice on January 1st, 1950, or January 1st, 1970.

Table 2 reports the results. A p-value below the conventional 5% threshold indicates that the performance differential of a given method compared to `mSSR` is statistically different from zero. The Sharpe ratio of `mSSR` is 1.57 annually. This is about twice as large as the original momentum portfolio (0.737, $pval = 0.000$). Not surprisingly, the `DM2016` and `BS2015` strategies do improve in risk-adjusted terms vis-á-vis the `WML` portfolio, with an SR of 1.37 and 1.26, respectively. Yet,

[14] The random walk dynamics of the parameters f_t imply that these are quite stable and, therefore, re-estimating the model daily would have a limited impact on the results.

Table 2: **Managed momentum portfolios**

Panel A reports a series of performance measures on the daily returns on our skewness-adjusted maximum conditional Sharpe ratio strategy (`mSSR`) against a variety of alternative managed-momentum portfolios. We consider Daniel and Moskowitz (2016) (`DM2016`) and Barroso and Santa-Clara (2015) (`BS2015`) which are based on recursive estimates of the realised variance, as well as a semi-volatility targeting as proposed by Wang and Yan (2021) and Hanauer and Windmüller (2023). Specifically, the performance measures in each of the columns are, in order, the Sharpe ratio, the Sortino ratio, the Stable Tail Adjusted Return ratio (STARR), the Rachev ratio (RR), the Value-at-Risk (VaR), the Expected Shortfall (ES), the volatility of excess return conditional on being negative (dVol), the sample skewness of excess returns, and the maximum drawdown (MaxDD). In parentheses we report the bootstrapped p-values testing the difference in performance of `mSSR` against each of the alternatives. In the last column we report the implied leverage. Panel B reproduces the performance of `mSSR`, denoted with ω_t, and reports performance measures associated with the two components of the portfolio (see Eq. (8)): the location component ω_{1t} and the skewness hedging component ω_{2t}. The out-of-sample period is from January 1st 1930 to December 31st 2020.

Panel A: Performance metrics

Strategies	Sharpe	Sortino	STARR	RR	VaR	ES	dVol	Skew	MaxDD	Leverage
`mSSR`	1.573	2.513	14.256	1.309	−3.231	−4.770	10.738	0.153	0.349	0.573
`cdVol`	1.432 (0.139)	2.095 (0.037)	11.604 (0.003)	1.005 (0.027)	−4.186 (0.000)	−5.333 (0.580)	11.720	−0.035	0.565	1.527
`DM2016`	1.375 (0.107)	2.011 (0.035)	11.479 (0.020)	1.077 (0.051)	−3.870 (0.000)	−5.179 (0.087)	11.731	0.021	0.427	0.970
`BS2015`	1.262 (0.014)	1.812 (0.005)	10.135 (0.002)	0.974 (0.021)	−4.156 (0.000)	−5.384 (0.086)	11.950	−0.043	0.462	0.243
`WML`	0.737 (0.000)	1.017 (0.000)	5.914 (0.000)	0.936 (0.000)	−3.791 (0.000)	−5.387 (0.002)	12.435	−0.056	1.137	

Panel B: `mSSR` performance decomposition

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

while the SR from `DM2016` is statistically equivalent to our `mSSR` adjustment ($pval = 0.107$), the `BS2015` produces significantly lower risk-adjusted return ($pval = 0.014$). Scaling by semi-volatility also improves upon the original `WML` portfolio, with a SR of 1.43 that is statistically equivalent to `mSSR` ($pval = 0.139$). Note the average leverage implied by `DS2016` and `cdVol` is almost two and three times larger than `mSSR`, respectively.

The competitive SR obtained from our skewness-adjusted method does not translate into higher downside risk. The Sortino ratio of 2.5 obtained from `mSSR` is both economically and statistically

larger than all the other strategies. In comparison, the original `WML` factor has a Sortino ratio of 1.017 ($pval = 0.000$). More generally, our `mSSR` strategy outperforms, both economically and statistically, all competing methods in terms of realised downside risk. For example, `mSSR` produces a STARR of 14.2 compared to 11.4 ($pval = 0.020$) and 10.1 ($pval = 0.002$) obtained from `DM2016` and `BS2015`, respectively. Similar conclusions can be drawn also by looking at other downside risk measures. For instance, `mSSR` produces an RR of 1.3, compared to 1.1 ($pval = 0.051$) and 0.97 ($pval = 0.021$) obtained from `DM2016` and `BS2015`, respectively.

When separating the performance of $\omega_{1,t}$ and $\omega_{2,t}$ in Eq. (8) an interesting observation arises. The first component $\omega_{1,t}$ yields a higher SR than $\omega_{2,t}$. This implies that adjusting for skewness may harm the portfolio's overall performance. However, $\omega_{2,t}$ results in lower VaR, ES, and dVol. Consequently, combining the two leads to significant enhancements in all downside risk measures and overall higher performance when measured with the STARR and RR metrics. The combined portfolio exhibits a lower maximum drawdown and a more positively skewed return distribution than either component individually. This supports the notion that $\omega_{2,t}$ serves as an insurance component against downside risk, which could be particularly beneficial for investors who prioritize avoiding losses over maximizing gains (e.g., Kraus and Litzenberger, 1976; Kahneman and Tversky, 2013).

5.1 Additional results

We expand the main economic evaluation and implement a variety of alternative exercises. First, we examine the models' performance based on short-term and intermediate momentum strategies. Next, we investigate the role of leverage constraints on the capital adjustment implied by (7) and transaction costs on the strategy's profitability. Finally, we investigate the robustness of the main results by focusing on the post-1950 subsample or by implementing the capital adjustment monthly instead of daily.

Alternative momentum portfolios. Table 3 reports the performance of our `mSSR` strategy for both the short-term `6_2` and intermediate `12_7` momentum portfolios. The results are consistent with the `12_2` momentum portfolio. A dynamic leverage adjustment that accounts for time-varying

skewness has a more decisive effect on mitigating the exposure to downside risk than achieving a higher risk-adjusted return. For instance, our `mSSR` strategy produces a significantly higher Sortino ratio, STARR, RR, and a lower VaR and ES. In addition, the average leverage from our approach is substantially lower than the one required by `cdVol` and the `DM2016` strategy. This implies less binding liquidity constraints.

Table 3: **Alternative momentum portfolios**

The table reports a series of performance measures on the daily returns on our skewness-adjusted maximum conditional Sharpe ratio strategy (`mSSR`) against a variety of alternative managed-momentum portfolios. We consider Daniel and Moskowitz (2016) (`DM2016`) and Barroso and Santa-Clara (2015) (`BS2015`) which are based on recursive estimates of the realised variance, as well as a semi-volatility targeting as proposed by Wang and Yan (2021) and Hanauer and Windmüller (2023). Specifically, the performance measures in each of the columns are, in order, the Sharpe ratio, the Sortino ratio, the Stable Tail Adjusted Return ratio (STARR), the Rachev ratio (RR), the Value-at-Risk (VaR), the Expected Shortfall (ES), the volatility of excess return conditional on being negative (dVol), the sample skewness of excess returns, and the maximum drawdown (MaxDD). In parentheses we report the bootstrapped p-values testing the difference in performance of `mSSR` against each of the alternatives. In the last column we report the implied leverage. Panel A reports the results for the intermediate momentum `12_7` portfolio, whereas Panel B reports the results for the short-term momentum `6_2` portfolio. The out-of-sample period is from January 1st 1930 to December 31st 2020.

Panel A: Intermediate momentum `12_7`

Strategies	Sharpe	Sortino	STARR	RR	VaR	ES	dVol	Skew	MaxDD	leverage
`mSSR`	1.380	2.155	12.169	1.249	−3.355	−4.901	10.982	0.130	0.547	0.641
`cdVol`	1.274 (0.161)	1.873 (0.071)	10.533 (0.014)	1.043 (0.335)	−4.155 (0.000)	−5.229 (0.359)	11.671	−0.021	0.442	1.785
`DM2016`	1.209 (0.143)	1.779 (0.092)	9.852 (0.338)	1.092 (0.147)	−4.034 (0.000)	−5.306 (0.334)	11.658	0.028	0.563	0.998
`BS2015`	1.134 (0.046)	1.646 (0.018)	9.342 (0.146)	1.027 (0.089)	−4.139 (0.000)	−5.248 (0.983)	11.820	−0.023	0.577	0.280
`WML`	0.734 (0.000)	1.030 (0.000)	5.974 (0.017)	0.984 (0.000)	−3.782 (0.000)	−5.308 (0.015)	12.218	−0.038	1.040	

Panel B: Short-term momentum `6_2`

Strategies	Sharpe	Sortino	STARR	RR	VaR	ES	dVol	Skew	MaxDD	leverage
`mSSR`	1.210	1.843	10.593	1.237	−3.382	−4.928	11.242	0.112	0.472	0.503
`cdVol`	1.123 (0.369)	1.621 (0.222)	9.905 (0.008)	0.990 (0.496)	−4.273 (0.000)	−5.360 (0.541)	11.886	−0.047	1.027	1.589
`DM2016`	1.189 (0.861)	1.746 (0.662)	9.455 (0.048)	1.073 (0.694)	−3.972 (0.000)	−5.422 (0.204)	11.653	0.033	0.544	0.843
`BS2015`	1.004 (0.105)	1.430 (0.056)	7.857 (0.063)	0.966 (0.123)	−4.251 (0.000)	−5.510 (0.454)	12.013	−0.045	0.953	0.253
`WML`	0.517 (0.000)	0.697 (0.000)	4.065 (0.024)	0.930 (0.000)	−3.752 (0.004)	−5.488 (0.006)	12.697	−0.054	1.799	

The similarity of the performance for different momentum portfolios does not come as a surprise.

Figure E1 in Appendix E.1 shows that the estimates of conditional volatility and skewness for the `12_7` and `6_2` momentum portfolios are similar to the `12_2` portfolio. For instance, both alternative momentum portfolios experienced spikes in return volatility during the great depression, which coincided with deepening negative skewness. Return asymmetry tends to deteriorate during economic recessions while becoming negligible, or at times positive, during economic expansions, especially if an upturn in economic activity occurs for prolonged periods. Overall, conditional volatility and skewness dynamics are rather consistent across momentum portfolios.

For completeness, Figure E2 in Appendix E.1 also reports the estimates of conditional skewness based on the NYSE breakpoints as in Fama and French (1996). The trajectory of the conditional skewness is broadly consistent with the `12_2` portfolio constructed from the all-firm breakpoints used in the main empirical analysis. The correlation between the two estimates is as high as 70%, even when looking at noisy daily estimates.

Leverage constraints. The capital adjustment in Eq. (7) is unconstrained, and ω_t can be large for small levels of conditional volatility or can take negative values. For instance, the `DM2016` implementation leaves open the possibility of negative ω_t as far as there is a negative correlation between volatility and return during recessions (see Figure 5 in Daniel and Moskowitz, 2016). Table 2 confirms that some adjustments can require, on average, as much as 1.5 times more capital than the one invested in the original `WML` portfolio. This can exacerbate liquidity needs or lead to a switch from momentum to a reversal strategy and thus can cast doubt on the actual feasibility of implementing Eq. (7) under reasonable liquidity or operational constraints (e.g., Harvey et al., 2018; Patton and Weller, 2020).

For this reason, we investigate the performance across models under a leverage cap and forbidding negative weights. We follow Moreira and Muir (2017); Cederburg et al. (2020); Barroso and Detzel (2021); Wang and Yan (2021) and implement two sets of leverage constraints. Panel A in Table 4 reports the results for the baseline `12_2` momentum portfolio with an x5 leverage constraint, i.e., $\omega_t \in (0, 5)\ \forall t$. In parentheses, we report the bootstrap p-values for all performance measures.

Table 4: **Accounting for portfolio constraints**

The table reports a series of performance measures on the daily returns on our skewness-adjusted maximum conditional Sharpe ratio strategy (`mSSR`) against a variety of alternative managed-momentum portfolios. We consider Daniel and Moskowitz (2016) (`DM2016`) and Barroso and Santa-Clara (2015) (`BS2015`) which are based on recursive estimates of the realised variance, as well as a semi-volatility targeting as proposed by Wang and Yan (2021) and Hanauer and Windmüller (2023). Specifically, the performance measures in each of the columns are, in order, the Sharpe ratio, the Sortino ratio, the Stable Tail Adjusted Return ratio (STARR), the Rachev ratio (RR), the Value-at-Risk (VaR), the Expected Shortfall (ES), the volatility of excess return conditional on being negative (dVol), the sample skewness of excess returns, and the maximum drawdown (MaxDD). In parentheses we report the bootstrapped p-values testing the difference in performance of `mSSR` against each of the alternatives. In the last column we report the implied leverage. All strategies are constrained such that $\omega_t \in (0, 5)\ \forall t$ in the firs panel, and $\omega_t \in (0, 1.5)\ \forall t$ in the second panel. The out-of-sample period is from January 1st 1930 to December 31st 2020.

Strategies	Sharpe	Sortino	STARR	RR	VaR	ES	dVol	Skew	MaxDD	Leverage
Panel A: x5 leverage cap										
`mSSR`	1.671	2.688	15.174	1.348	−3.062	−4.761	10.666	0.193	0.261	0.703
`cdVol`	1.402 (0.013)	2.095 (0.004)	11.605 (0.000)	1.005 (0.935)	−4.186 (0.000)	−5.333 (0.040)	11.720	−0.037	0.565	1.506
`DM2016`	1.373 (0.005)	2.007 (0.002)	11.453 (0.000)	1.078 (0.240)	−3.873 (0.148)	−5.182 (0.012)	11.732	0.023	0.428	0.868
`BS2015`	1.262 (0.000)	1.812 (0.000)	10.135 (0.000)	0.974 (0.279)	−4.156 (0.062)	−5.384 (0.002)	11.950	−0.043	0.462	0.213
`WML`	0.737 (0.000)	1.017 (0.000)	5.914 (0.000)	0.936 (0.006)	−3.791 (0.000)	−5.387 (0.000)	12.435	−0.056	1.137	
Panel B: x1.5 leverage cap										
`mSSR`	1.573	2.513	14.256	1.309	−3.231	−4.770	10.738	0.153	0.349	0.573
`cdVol`	1.394 (0.140)	2.035 (0.041)	11.262 (0.000)	0.992 (0.587)	−4.188 (0.003)	−5.349 (0.177)	11.751	−0.047	0.558	1.652
`DM2016`	1.375 (0.100)	2.011 (0.037)	11.479 (0.000)	1.077 (0.081)	−3.870 (0.113)	−5.179 (0.093)	11.731	0.021	0.427	0.970
`BS2015`	1.262 (0.012)	1.812 (0.004)	10.135 (0.000)	0.974 (0.087)	−4.156 (0.071)	−5.384 (0.021)	11.950	−0.043	0.462	0.243
`WML`	0.737 (0.000)	1.017 (0.000)	5.914 (0.000)	0.936 (0.002)	−3.791 (0.002)	−5.387 (0.000)	12.435	−0.056	1.137	

Unlike the unconstrained case, the SR significantly favours our `mSSR` strategy. The second best approach, `cdVol`, produces an SR of 1.402 against a 1.671 obtained from `mSSR`. The null hypothesis that the SRs are statistically equivalent is strongly rejected ($pval = 0.013$). Our `mSSR` also retains an advantage in pure exposure to downside risk. For instance, the Sortino ratio, the STARR, the VaR, and the ES significantly favour `mSSR`. The realised sample skewness and the maximum drawdown also favour our `mSSR` strategy compared to other methods. Panel B shows that the results are broadly consistent with a more restrictive x1.5 leverage constraint, i.e., $\omega_t \in (0, 1.5)\ \forall t$.

Subsample analysis. In the main empirical analysis, the out-of-sample period is from January 1st, 1930, to December 31st, 2020. Given all the institutional changes in the US stock market, one may wonder if the data from the pre-World War II period may be relevant for investors today. In addition, the results so far indicate that the momentum crash of the 1930s had a substantial role in shaping the risk-return trade-off of the momentum portfolio to this date. This is due to the compounding nature of the return of a buy-and-hold strategy in the `WML` (see Section 2). To mitigate concerns about sample selection, we investigate the performance of different dynamic leverage strategies post-World War II. Specifically, we use three years of daily return as an initial burn-in sample for the recursive estimates to generate the initial portfolio choice on January 1st, 1950 and 1970.

Table 5 reports the results. The performance in the two subsamples is notably stronger across methods, including the original `WML` portfolio, especially regarding SR and Sortino ratios. Nevertheless, all time-series capital adjustments still improve upon the `WML` portfolio. Consistent with the longer out-of-sample results (see Table 2), our `mSSR` produces a comparable, statistically equivalent SR, but at the same time, significantly lower exposure to downside risk. The gap in favour of our `mSSR` regarding downside risk mitigation is confirmed for all performance metrics. Comparing the results for the post-1970 sample to the post-1950, it is noticeable that while all alternatives produce relatively stable performance, `mSSR` shows a marked improvement both in terms of Sortino, STARR and Ratichev ratio, without any marked difference in the downside risk metrics.

Transaction costs. The last column of Table 2 shows that our `mSSR` strategy implies, on average, quite conservative leverage. The average ω_t is 0.57 for the `mSSR`, vs 1.5 and 0.97 for `cdVol` and `DM2016`, respectively. Perhaps not surprisingly, a smoother volatility estimate, such as the six-month realised variance used by `BS2015`, helps to reduce turnover (e.g., Barroso and Detzel, 2021 and Bernardi et al., 2022). The same conclusion can be drawn for alternative momentum portfolios (see Table 3) and for a monthly capital adjustment (see Table 8).

To investigate the economic cost of turnover across strategies, we evaluate the performance for all methods after transaction costs.[15] Specifically, we implement three different notions of trading

[15]Notice that our goal is not to propose an actual trading strategy that can be implemented "off the shelf", but

Table 5: **Subsample analysis**

The table reports a series of performance measures on the daily returns on our skewness-adjusted maximum conditional Sharpe ratio strategy (`mSSR`) against a variety of alternative managed-momentum portfolios. We consider Daniel and Moskowitz (2016) (`DM2016`) and Barroso and Santa-Clara (2015) (`BS2015`) which are based on recursive estimates of the realised variance, as well as a semi-volatility targeting as proposed by Wang and Yan (2021) and Hanauer and Windmüller (2023). Specifically, the performance measures in each of the columns are, in order, the Sharpe ratio, the Sortino ratio, the Stable Tail Adjusted Return ratio (STARR), the Rachev ratio (RR), the Value-at-Risk (VaR), the Expected Shortfall (ES), the volatility of excess return conditional on being negative (dVol), the sample skewness of excess returns, and the maximum drawdown (MaxDD). In parentheses we report the bootstrapped p-values testing the difference in performance of `mSSR` against each of the alternatives. The out-of-sample is from January 1st, 1950, to December 31st, 2020.

Strategies	Sharpe	Sortino	STARR	RR	VaR	ES	dVol	Skew	MaxDD	Leverage
Panel A: Post-1950 sample										
`mSSR`	1.805	2.944	16.523	1.344	−3.212	−4.722	10.515	0.162	0.323	0.628
`cdVol`	1.646 (0.219)	2.447 (0.063)	13.468 (0.295)	1.029 (0.002)	−4.186 (0.000)	−5.283 (0.040)	11.537	−0.032	0.567	1.602
`DM2016`	1.657 (0.290)	2.481 (0.071)	13.867 (0.204)	1.117 (0.118)	−3.915 (0.000)	−5.166 (0.041)	11.458	0.032	0.432	1.169
`BS2015`	1.500 (0.027)	2.195 (0.010)	12.091 (0.037)	1.001 (0.003)	−4.184 (0.000)	−5.363 (0.008)	11.722	−0.037	0.465	0.257
`WML`	0.883 (0.000)	1.224 (0.000)	7.108 (0.000)	0.969 (0.002)	−3.713 (0.002)	−5.372 (0.004)	12.376	−0.036	1.192	
Panel B: Post-1970 sample										
`mSSR`	1.985	3.318	18.512	1.390	−3.160	−4.636	10.263	0.179	0.305	0.542
`cdVol`	1.649 (0.030)	2.457 (0.011)	13.555 (0.034)	1.036 (0.168)	−4.201 (0.000)	−5.257 (0.386)	11.511	−0.034	0.562	1.417
`DM2016`	1.676 (0.057)	2.528 (0.025)	14.052 (0.032)	1.128 (0.823)	−3.906 (0.000)	−5.156 (0.114)	11.373	0.037	0.449	1.017
`BS2015`	1.456 (0.002)	2.125 (0.001)	11.684 (0.001)	0.992 (0.916)	−4.224 (0.000)	−5.386 (0.065)	11.749	−0.042	0.464	0.224
`WML`	0.809 (0.000)	1.115 (0.000)	6.382 (0.000)	0.954 (0.001)	−3.887 (0.000)	−5.481 (0.003)	12.446	−0.052	1.060	

costs: first, we follow DeMiguel et al. (2009) and calculate the evolution of wealth for strategy i as $W_{i,t+1} = W_{i,t}\mathcal{R}_{i,t}\left(1 - c|\omega_{i,t+1} - \omega^{+}_{i,t}|\right)$ with $\mathcal{R}_{i,t} = 1+R_{i,t}$ the gross return at time t, $\omega^{+}_{i,t} = \mathcal{R}_{i,t}\omega_{i,t-1}$ the time t weights after accruing the return (e.g., Detzel et al., 2023) and $|\omega_{i,t+1} - \omega^{+}_{i,t}|$ and c the turnover for a given period and the transaction costs. Thus, a strategy performance net of transaction costs can be computed as $r - c = \frac{W_{i,t+1}}{W_{i,t}} - 1$. Second, we follow Della Corte et al. (2008) and evaluate

rather to show the economic value of expanding the notion of risk to the third moment. Thus, although simplistic, we believe that considering different cost measures could help shed further light on the incremental value of accounting for time-varying skewness in momentum return.

the maximum performance fee an investor with a quadratic utility function would be willing to pay to access a given managed-momentum strategy. Specifically, for any pair (i, j) of strategies, the fee $\mathcal{F}$ arises as the solution of,

$$\sum_{t=0}^{T-1}(\mathcal{R}_{i,t} - \mathcal{F}) - \frac{\delta(\mathcal{R}_{i,t} - \mathcal{F})^2}{2(1-\delta)} = \sum_{t=0}^{T-1}\mathcal{R}_{j,t} - \frac{\delta\mathcal{R}_{j,t}^2}{\delta(1-\delta)}, \tag{9}$$

where δ is the degree of relative risk aversion. Third, we consider a measure of abnormal return as in Modigliani and Modigliani (1997). That is, for any pair of strategies (i, j), we leverage up or down strategy i to match the downside-risk profile of strategy j, and we evaluate the annualised abnormal return as follows $d\mathcal{A}_{i,j} = dVol_i(Sortino_i - Sortino_j)$.

Table 6: **Accounting for transaction costs**

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 downside volatility targeting (`cdVol`) 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 out-of-sample period is from January 1st 1930 to December 31st 2020. Portfolio weights are generated by recursively estimating the conditional mean and variance of the returns based on the model parameters. The first three years are used as a burn-in period.

Costs (bps)	`mSSR`			`cdVol`			`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	11.894	13.378	8.198	10.902	12.291	7.189	8.957	9.809	5.297
1	13.914	18.075	10.216	11.706	13.164	7.946	10.862	12.244	7.189	8.962	9.815	5.297
5	12.258	16.010	8.450	10.956	12.313	7.189	10.702	12.059	6.937	8.982	9.839	5.171
10	10.187	13.455	6.432	10.017	11.254	6.180	10.503	11.827	6.685	9.008	9.867	5.171

We calculate the fee $\mathcal{F}$ relative to the original `WML` portfolio and consider a risk aversion of $\delta = 5$ (e.g., Rapach and Zhou, 2013; Pettenuzzo et al., 2014; Bianchi et al., 2021). We explore different levels of transaction costs, ranging from 0 to 10 bps. The latter represents a non-trivial execution cost to trade an "ETF-like" momentum portfolio. The cost of building the actual momentum portfolio is arguably higher, although symmetric for all managed momentum strategies and therefore irrelevant for any relative comparisons between strategies (see e.g., DeMiguel et al., 2020).

Table 6 reports the results. Our `mSSR` strategy produces a higher annualised average return

after transaction costs and commands higher performance fees than all competing strategies. The performance fee remains largely in favour of mSSR for different levels of transaction costs. For instance, with 10 basis points of rebalancing costs the abnormal performance $d\mathcal{A}_{i,j}$ – relatively to WML – of the mSSR is 13.3% (annualised), against 11.9%, 11.3% and 9.8% for the DM2016, BS2015 and cdVol, respectively.

Appendix E reports the results obtained by repeating the economic evaluation of Table 6 for different levels of risk aversion of $\delta = 2, 7, 15$. Interestingly, a time-varying skewness adjustment becomes even more valuable for a more risk-averse investor. This supports the intuition that the insurance component represented by the skewness adjustment becomes more valuable as investors dislike the potential losses of the original WML portfolio even more.

Spanning tests. We estimate a series of factor-spanning regressions whereby the daily return on each adjusted momentum portfolio is regressed onto a host of factors, including the market, size and value factors from Fama and French (1993), the original WML portfolio, the dynamic volatility adjustment of Daniel and Moskowitz (2016) and the constant volatility adjustment of Barroso and Santa-Clara (2015). Table 7 reports the annualised alphas and corresponding t-stats (in parenthesis).

Table 7: **Factor spanning regressions**

The table reports the results of a series of factor spanning regressions where the dependent variables are the returns on different managed momentum portfolios based on our skewness-adjusted maximum conditional Sharpe ratio strategy (mSSR), the original maximum Sharpe ratio adjustment ofDaniel and Moskowitz (2016) (DM2016), the constant volatility targeting of Barroso and Santa-Clara (2015) (BS2015), and semi-volatility targeting as proposed by Wang and Yan (2021) and Hanauer and Windmüller (2023). Each portfolio returns is regressed onto the market and other Fama and French (1993) common risk factors in addition to the original WML portfolio. The sample size is from January 1st 1930 to December 31st 2020, daily.

	Mkt+WML	FF3+WML	Mkt+cdVol	FF3+cdVol	Mkt+DM2016	FF3+DM2016	Mkt+BS2015	FF3+BS2015
mSSR	26.374	26.366	15.249	15.234	27.471	27.453	27.923	27.906
	(13.765)	(13.762)	(9.355)	(9.348)	(13.749)	(13.743)	(13.944)	(13.939)
cdVol	15.410	15.431			23.521	23.517	23.774	23.773
	(13.961)	(13.980)			(12.015)	(12.015)	(12.142)	(12.143)
DM2016	25.254	25.274	22.248	22.265			3.614	3.625
	(12.741)	(12.751)	(11.424)	(11.434)			(5.306)	(5.322)
BS2015	22.527	22.535	19.779	19.784	−0.206	−0.216		
	(11.519)	(11.523)	(10.269)	(10.273)	(−0.309)	(−0.323)		

The first column reports the results controlling for the market (Mkt) and the WML portfolio. The

intercept is highly significant for all strategies, with the highest value for `mSSR` (26.3% annualised, t-stat = 13.7). This indicates that the `Mkt` and `WML` portfolios cannot explain the performance of `mSSR`. The second column adds the size and value factors as further regressors. All the alphas are again highly significant, with our `mSSR` strategy showing the largest intercept (26.3% annualised, t-stat = 13.7). Hence, conventional equity factors cannot explain the performance of `mSSR`. Columns 3 and 4 replace the `WML` portfolio with the return on the `cdVol` strategy. The intercepts drop in magnitude but remain highly significant throughout. This is surprising since `cdVol` should capture the downside risk exposure. Overall, the evidence suggests that the performance of managed momentum strategies remains substantially high even after controlling for asymmetric volatility.

The last four columns in Table 7 replace the `WML` portfolio with the return on either the Daniel and Moskowitz (2016) or the Barroso and Santa-Clara (2015) strategy. The results show that these managed momentum portfolios cannot fully explain the return on our skewness-adjusted strategy. Interestingly, the intercept on `BS2015` is no longer statistically significant when conditioning on the Daniel and Moskowitz (2016) managed portfolio, whereas the intercept of `DM2016` is still significant when conditioning on the Barroso and Santa-Clara (2015). This confirms that a maximum Sharpe ratio strategy generally subsumes constant volatility targeting.

Monthly capital adjustment. The performance outlined in Table 2 concerns daily adjustments based on Eq. (7). Specifically, the momentum decile portfolios are rebalanced monthly, but the capital adjustment implied by ω_t is implemented daily. To enhance the comparability with existing studies, we investigate the performance across methods for a lower-frequency monthly implementation of the capital adjustment ω_t.[16]

Table 8 reports the results for the `12_2` momentum portfolio. Overall, the effectiveness of the leverage adjustment decreases quite substantially when implemented monthly; that is, the performance at a monthly frequency is notably lower.[17] This corroborates the intuition in Novy-Marx

[16] To mitigate the effect of aggregation from high to lower frequency return in a portfolio context (e.g., Boguth et al., 2016), for the monthly adjustment, in practice, we re-scale and aggregate the estimated conditional mean and variance from daily to monthly.

[17] Therefore, once the loss in performance is fully accounted for, it is unclear whether a lower-frequency implementation of Eq. (7) would effectively reduce transaction costs.

Table 8: **Monthly capital adjustment**

The table reports the performance on our skewness-adjusted maximum conditional Sharpe ratio strategy (`mSSR`) against a variety of alternative managed-momentum portfolios. We consider Daniel and Moskowitz (2016) (`DM2016`) and Barroso and Santa-Clara (2015) (`BS2015`) which are based on aggregating daily recursive estimates of the realised variance, as well as a semi-volatility targeting as proposed by Wang and Yan (2021) and Hanauer and Windmüller (2023). The capital adjustment in Eq. (7) is implemented monthly instead of daily. We report in parentheses the bootstrapped p-values for the difference in the performance of `mSSR` against each of the alternatives. The out-of-sample period is from January 1930 to December 2020.

Strategies	Sharpe	Sortino	STARR	RR	VaR	ES	dVol	Skew	MaxDD	Leverage
`mSSR`	1.093	2.561	2.807	2.101	−0.525	−0.869	7.936	0.405	0.349	0.684
`cdVol`	1.292 (0.000)	2.832 (0.326)	3.159 (0.000)	1.490 (0.060)	−0.739 (0.002)	−0.983 (0.655)	9.134	0.027	0.402	1.049
`DM2016`	1.159 (0.392)	2.311 (0.352)	2.632 (0.550)	1.703 (0.579)	−0.684 (0.095)	−0.982 (0.208)	9.323	0.221	0.547	0.910
`BS2015`	1.082 (0.869)	1.836 (0.003)	2.053 (0.858)	1.209 (0.069)	−0.841 (0.001)	−1.176 (0.040)	10.957	−0.010	0.506	0.351
`WML`	0.641 (0.006)	0.877 (0.000)	1.079 (0.001)	0.925 (0.006)	−0.881 (0.005)	−1.326 (0.024)	13.599	−0.092	1.028	

and Velikov (2019); monthly rebalancing reduces the strategy's profitability due to the signal's staleness at a monthly frequency. This is the case for our `mSSR`, which highlights the importance of high-frequency variation in downside risk.[18]

6 Implications for asset pricing models

Grundy and Martin (2001) document that the asymmetric nature of the momentum exposure to market risk is at the core of momentum crashes. We build upon this intuition and attempt to provide a more structured asset pricing interpretation of the uncovered time-varying skewness risk in momentum return. Consider a state-dependent CAPM specification which separates up-market and down-market betas for the momentum return r_t,

$$r_t = \alpha + \underbrace{\overline{\beta} m_t I(m_t \geq \mu_m) + \underline{\beta} m_t I(m_t < \mu_m)}_{\beta m_t} + e_t \tag{10}$$

[18] At the same time, our `mSSR` produces the lowest realised downside risk, as shown by the lowest VaR and ES, the highest RR and the largest positive return skewness. Appendix E.3 reports the results for the `12_7` and `6_2` portfolios. The results are consistent with the `12_2` momentum portfolio.

with $m_t \sim \mathcal{N}\left(\mu_m, \sigma_m^2\right)$ the normally distributed market portfolio and $I(m_t \geq \mu_m)$ (or $I(m_t < \mu_m)$) an indicator function that takes value one if the market return m_t are above (or below) the mode μ_m and zero otherwise. Ang et al. (2006) suggest that a conditional CAPM as Eq. (10) can be interpreted as a reduced form representation of a general equilibrium model with a disappointment-aversion utility function in which a representative investor has a higher sensitivity to losses versus gains (e.g., Kraus and Litzenberger, 1976; Gul, 1991; Kahneman and Tversky, 2013). Conditional on $I(\cdot)$, the systematic component βm_t can be characterised by a two-piece continuous distribution (e.g., 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 G),

$$E[\beta m_t] - \beta\mu_m = \sqrt{\frac{2}{\pi}}\left(\overline{\sigma}_m - \underline{\sigma}_m\right) \propto \sigma_m\left(\overline{\beta} - \underline{\beta}\right) \tag{11}$$

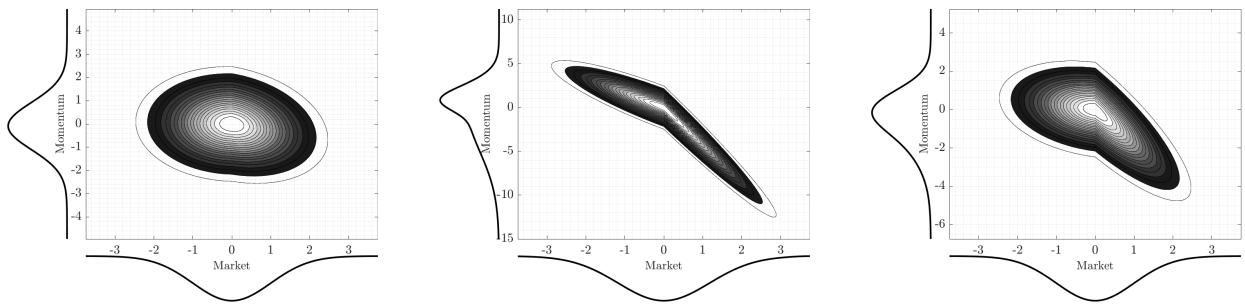
with $\underline{\sigma}_m^2 = \underline{\beta}^2\sigma_m^2$ and $\overline{\sigma}_m^2 = \overline{\beta}^2\sigma_m^2$. Under the assumption of equal betas between market states, $\underline{\beta} = \overline{\beta} = \beta$, we obtain that $E[\beta m_t] = \beta\mu_m$, $V[\beta m_t] = \beta^2\sigma_m^2$, such that the marginal distribution of the momentum strategy return is equivalent to a standard CAPM formulation $r_t \sim \mathcal{N}\left(\alpha + \beta\mu_m, \beta^2\sigma_m^2 + \sigma_e^2\right)$. Instead, with asymmetric betas $\underline{\beta} \neq \overline{\beta}$ and $\text{sign}\left(\underline{\beta}\right) = \text{sign}\left(\overline{\beta}\right)$, Eq. (11) suggests that for $\overline{\beta} < \underline{\beta}$ ($\overline{\beta} > \underline{\beta}$) the expected value of the systematic CAPM component is lower (higher) than the mode (e.g., Arnold and Groeneveld, 1995). As a result, even if the market return m_t and the residual e_t are both normally distributed, the marginal distribution of r_t can still be negatively (positively) skewed, with the level of skewness that depends on how far apart are the state-dependent betas.[19]

Appendix F.1 reports the sample estimates of the upside and downside market betas for daily return on the past losers, past winners, and the `WML` portfolios. Consistent with Grundy and Martin (2001), the estimates show that the past losers portfolio is more exposed to upside market risk ($\overline{\beta} = 1.36$) compared to downside market risk ($\underline{\beta} = 1.27$). The opposite holds for past winners ($\overline{\beta} = 1.09$, $\underline{\beta} = 1.22$). Thus, the `WML` strategy shows a sizable, negative up-market beta ($\overline{\beta} = -0.27$),

[19] Notice this holds with the sign of the betas being the same, i.e., $sgn\left(\underline{\beta}\right) = sgn\left(\overline{\beta}\right)$. Instead, 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 normal distributions with different means.

Figure 9: **State-dependent CAPM and simulated return**

This figure reports the marginal distribution of the return on the momentum strategy (y-axis), the return on the market portfolio (x-axis) and the corresponding joint distribution. return are simulated assuming a two-piece Normal distribution (see Appendix G). The left panel shows the joint distribution for the full sample, whereas the middle and the right panels show the joint return distributions for two different timestamps.



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

while the down-market beta is closer to zero ($\underline{\beta} = -0.04$). We plug these estimates in Eq. (10) and simulate the marginal and joint distribution of the `WML` and the market portfolio returns. To isolate the effect of the state-dependent betas on return skewness, we assume e_t is normally distributed with mean zero and volatility equal to one, whereas m_t is normally distributed with mean zero and volatility equal to the sample standard deviation of the market return.

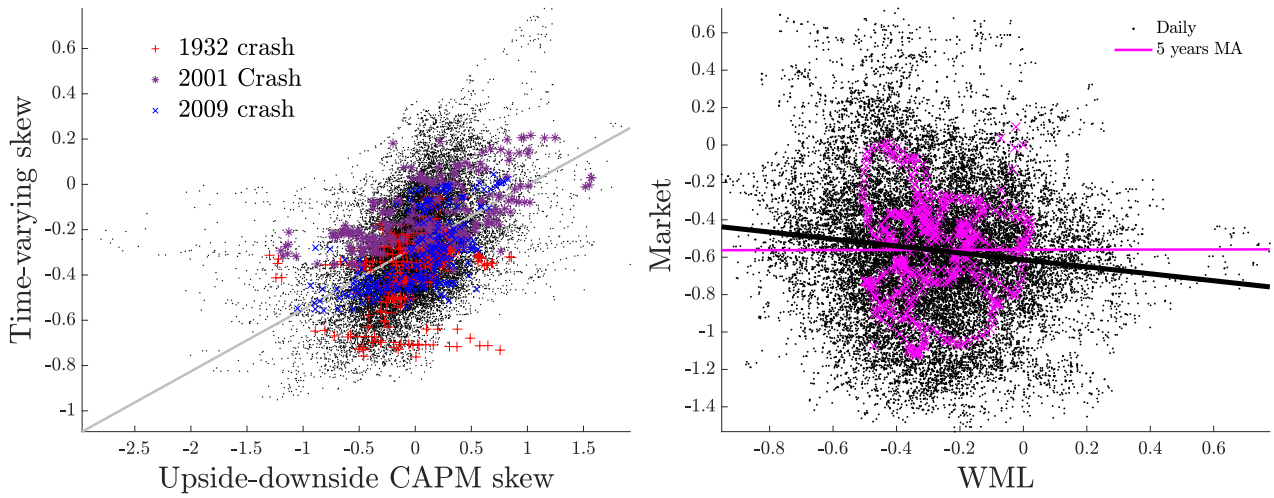
Figure 9 shows the simulated return distribution. The left column shows that the negative spread in betas can generate a slightly negatively skewed marginal return distribution. The middle and right columns expand the simulation by calculating $\overline{\beta}$ and $\underline{\beta}$ for two specific timestamps within crash periods. The skewness of momentum return 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 return (middle panel) is more negatively skewed (-0.805). Similarly, in September 2008, during the great financial crisis, the average betas spread was -1.3, which generated a conditional skewness of -0.529.

6.1 Evidence from a time-varying CAPM

The simulation results suggest that an asymmetric exposure to market risk may have the potential to rationalize the origin of conditional skewness in momentum return. It is, therefore, natural to ask to what extent the pattern of time-varying skewness we document across alternative momentum factors is just a reflection of potentially time-varying, state-dependent betas. To answer this question Figure 10(a) shows the sample correlation between the conditional skewness implied by the spread $\overline{\beta}_t - \underline{\beta}_t$ (see Figure F2) and our estimated time-varying skewness for the `12_2` portfolio (see Figure 5). The conditional beta estimates are based on time-varying CAPM with asymmetric betas (see Ang et al., 2006 and Appendix E).

Figure 10: **Momentum skewness and aggregate market return**

The left panel reports the conditional skewness estimated from our model against the conditional skewness generated from an asymmetric CAPM based on the simulation study. The right panel compares the conditional skewness of the `WML` portfolio and the aggregate market portfolio estimated from our model time-varying parameter model.



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

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

A positive correlation (0.39, $pval = 0.000$) exists between the skewness implied by the state-dependent CAPM and our estimated conditional skewness. However, such correlation flattens during the two major momentum crashes of 1932 (red dots) and 2009 (blue dots). This suggests that an asymmetric CAPM with normally distributed market return and residuals is likely not flexible

enough to capture the extent of the time variation in the conditional skewness of momentum return. The reason is twofold. First, the CAPM residuals are not normally distributed. A set of unreported results shows that the WML portfolio, net of market risk exposure, is still negatively skewed. Second, the excess return on the market is also not normally distributed. Table 1 shows that excess market return is negatively skewed, at least unconditionally (skewness= -0.476, $pval = 0.059$).

The correlation between the market and momentum conditional skewness is far from obvious. Figure 10(b) shows that the estimates of the conditional skewness of the market and the WML portfolio are only mildly negatively correlated (light grey line), with the correlation disappearing once two-year average estimates are compared (magenta line).

The fact that a state-dependent CAPM cannot reconcile the dynamics of return skewness is instrumental in highlighting one key advantage of our framework. By accounting for the conditional skewness of the return, our model can parsimoniously summarize sources of non-normality beyond asymmetric market betas. Indeed, while state-dependent CAPM betas can capture a fair deal of asymmetry in the return conditional distribution, there is still a considerable amount of skewness in momentum return that the correlation between the conditional skewness of the momentum strategy and the market portfolio cannot reconcile.

7 Conclusions

We investigate the dynamic of skewness in equity momentum return through the lens of a flexible model that features time-varying location, scale, and asymmetry parameters. Empirically, we uncover a pro-cyclical time variation of skewness, which tends to deepen during the so-called momentum crashes. This has important implications for the dynamic of the risk-return trade-off in momentum portfolios and, ultimately, for managing the risk associated with momentum investing.

References

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.

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. (2023). 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. (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.

Boguth, O., Carlson, M., Fisher, A., and Simutin, M. (2016). Horizon effects in average returns: The role of slow information diffusion. *The Review of Financial Studies*, 29(8):2241–2281.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2022). Realized semibetas: Disentangling "good" and "bad" downside risks. *Journal of Financial Economics*, 144(1):227–246.

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

Byun, S.-J. and Jeon, B. (2023). Momentum crashes and the 52-week high. *Financial Analysts Journal*, 79(2):120–139.

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.

Charoenrook, A. A. and Conrad, J. S. (2005). Identifying risk-based factors. In *AFA 2006 Boston meetings paper*.

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

De Polis, A. (2023). *Conditional asymmetries and downside risks in macroeconomic and financial time series*. PhD thesis, University of Warwick.

Della Corte, P., Sarno, L., and Thornton, D. L. (2008). The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value. *Journal of Financial Economics*, 89(1):158–174.

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.

DeMiguel, V., Martin-Utrera, A., Nogales, F. J., and Uppal, R. (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies*, 33(5):2180–2222.

Detzel, A., Novy-Marx, R., and Velikov, M. (2023). Model comparison with transaction costs. *The Journal of Finance*.

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.

Ehsani, S. and Linnainmaa, J. T. (2022). Factor momentum and the momentum factor. *The Journal of Finance*, 77(3):1877–1919.

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. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

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.

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.

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.

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.

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

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., Hoyle, E., Korgaonkar, R., Rattray, S., Sargaison, M., and Van Hemert, O. (2018). The impact of volatility targeting. *The Journal of Portfolio Management*, 45(1):14–33.

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

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.

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.

Kamen, E. W. and Su, J. K. (2012). *Introduction to optimal estimation*. Advanced Textbooks in Control and Signal Processing, Springer Science & Business Media.

Kandel, S. and Stambaugh, R. F. (1996). On the predictability of stock returns: an asset-allocation perspective. *The Journal of Finance*, 51(2):385–424.

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.

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.

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.

Novy-Marx, R. and Velikov, M. (2019). Comparing cost-mitigation techniques. *Financial Analysts Journal*, 75(1):85–102.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of financial econometrics*, 2(1):130–168.

Patton, A. J. and Weller, B. M. (2020). What you see is not what you get: The costs of trading market anomalies. *Journal of Financial Economics*, 137(2):515–549.

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

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

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.

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

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

Supplementary Appendix for:

Time-Varying Skewness and Momentum Crashes

This appendix provides an in-depth description of the econometric model and the circular block bootstrap procedure implemented in Section 5 for performance testing. 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. In addition, we provide an extensive discussion on the simulation results for the optimal choice of the block size of the circular block bootstrap procedure. We also report a set of additional results concerning the time-varying skewness for alternative momentum portfolio constructions and the sample estimates of the state-dependent CAPM used to simulate momentum returns based on an asymmetric exposure to market risk (see Section 6). All additional results are referred to in the main text where appropriate.

A Likelihood-based test for conditional skewness

We outline a likelihood-based test for conditional skewness and its corresponding results. The basic idea of the test 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 $\mathtt{m}_t$ and scale σ_t, but fixed shape parameter ρ. The latter pin down the degree of asymmetry in the conditional distribution of the returns that is $r_t|\mathcal{F}^{t-1} \sim Skt_\nu(\mathtt{m}_t, \sigma_t^2, \rho)$. The gradient associated with the transformed shape (asymmetry) parameter $\delta = \text{arctanh}\ \rho$ 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 $\epsilon_t = r_t - \mathtt{m}_t$, 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 B.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 Portmanteau (P) and Ljung-Box (Q) tests for the null

Table A1: **Likelihood-based test for conditional skewness**

The table reports the results for a likelihood-based test for the time variation of the conditional skewness in the returns of the `12_2` momentum portfolio. Here, 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^{***}

hypothesis of absence of autocorrelation in the score $\nabla_{\delta,t}$, i.e., 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 read 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 Nyblom (1989) 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 test statistics which are well above 100 and p-values below the 0.01 threshold 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 likely not constant over time.

B Modelling 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:

$$f_{t+1} = f_t + As_t, \qquad t = 1, \dots, T \tag{B1}$$

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 $\nabla_t = J_t' \left[\frac{\partial \ell_t}{\partial \mathtt{m}_t}, \frac{\partial \ell_t}{\partial \sigma_t^2}, \frac{\partial \ell_t}{\partial \rho_t} \right]'$ 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 $\mathcal{I}_{t-1}$.[20] 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.

B.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), $y_t | Y_{t-1} \sim skt_\nu(\mathtt{m}_t, \sigma_t^2, \rho_t)$, the log-likelihood takes the form

$$\begin{aligned} \ell_t(r_t|\theta, \mathcal{F}_{t-1}) &= \log \mathcal{C}(\nu) - \frac{1}{2}\log \sigma_t^2 - \frac{1+\nu}{2} \log\left[1 + \frac{\varepsilon_t^2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2 \sigma_t^2} \right], \\ \log \mathcal{C}(\nu) &= \log \Gamma\left(\frac{\nu+1}{2} \right) - \log \Gamma\left(\frac{\nu}{2} \right) - \frac{1}{2}\log \nu - \frac{1}{2}\log \pi, \end{aligned} \tag{B2}$$

where $\Gamma(\cdot)$ is the Gamma function and $\nu > 3$ are the degrees of freedom. Differentiating (B2) with respect to location, scale and asymmetry we obtain the gradient vector $\nabla_t = \left[\frac{\partial \ell_t}{\partial \mathtt{m}}, \frac{\partial \ell_t}{\partial \sigma_t^2}, \frac{\partial \ell_t}{\partial \rho_t} \right]'$. Recall

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

that $\varepsilon_t = y_t - \mathtt{m}_t$, $\zeta_t = \frac{\varepsilon_t}{\sigma_t}$ and let

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

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

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

Proof. Define

$$g(\mathtt{m}_t) = a \log f(\mathtt{m}_t, \sigma_t^2, \rho_t),$$

such that $\frac{\partial \ell_t}{\partial \mathtt{m}_t} = \frac{\partial g(\mathtt{m}_t)}{\partial \mathtt{m}_t} = a\frac{f'_{\mathtt{m}_t}}{f(\mathtt{m}_t, \sigma_t^2, \rho_t)}$. For

$$f'_{\mathtt{m}_t} = -\frac{2}{\nu(1+\mathrm{s}(\varepsilon_t)\rho_t)^2\sigma_t^2}\varepsilon_t,$$

it follows:

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

. □

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

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

such that $\frac{\partial \ell_t}{\partial \sigma_t^2} = \frac{\partial g(\sigma_t^2)}{\partial \sigma_t^2} = -\frac{1}{2\sigma_t^2} + a\frac{f'_{\sigma_t^2}}{f(\mathtt{m}_t, \sigma_t^2, \rho_t)}$, 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

$$g(\rho_t) = a \log f(\mathtt{m}_t, \sigma_t^2, \rho_t),$$

such that $\frac{\partial \ell_t}{\partial \rho_t} = \frac{\partial g(\rho_t)}{\partial \sigma_t^2} = a\frac{f'_{\rho_t}}{f(\mathtt{m}_t, \sigma_t^2, \rho_t)}$, 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:

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

. □

B.2 Scaled scores

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

$$\frac{\partial \ell_t}{\partial \gamma_t} = \frac{\partial \ell_t}{\partial \sigma_t^2}\frac{\partial \sigma_t^2}{\partial \gamma_t}, \quad \frac{\partial \ell_t}{\partial \delta_t} = \frac{\partial \ell_t}{\partial \rho_t}\frac{\partial \rho_t}{\partial \delta_t}, \tag{B3}$$

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 $f_t = (\mathtt{m}_t, \gamma_t, \delta_t)'$ with the associated Jacobian matrix

$$J_t = \frac{\partial(\mathtt{m}_t, \sigma_t^2, \rho_t)}{\partial f_t'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2\sigma_t^2 & 0 \\ 0 & 0 & 1-\rho_t^2 \end{bmatrix}. \tag{B4}$$

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:

$$\mathcal{I}_t = \mathbb{E}_{t-1}[\nabla_t \nabla_t'] = \begin{bmatrix} \frac{(1+\nu)}{(\nu+3)(1-\rho_t^2)\sigma_t^2} & 0 & \frac{4(1+\nu)}{\sigma_t(1-\rho_t^2)(3+\nu)} \\ 0 & \frac{1}{2(3+\nu)\sigma_t^4} & 0 \\ \frac{4(1+\nu)}{\sigma_t(1-\rho_t^2)(3+\nu)} & 0 & \frac{3(1+\nu)}{(1-\rho_t^2)(3+\nu)} \end{bmatrix}. \tag{B5}$$

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

$$\mathbf{s}_t = (J_t' \mathrm{diag}(\mathcal{I}_t) J_t)^{-1} J_t' \nabla_t = \begin{bmatrix} s_{\mathtt{m}t} \\ s_{\sigma t} \\ s_{\rho t} \end{bmatrix} = \chi \begin{bmatrix} (1-\rho_t^2) w_t \varepsilon_t \\ (\nu+1)(w_t \varepsilon_t^2 - \sigma_t^2) \\ \mathrm{s}(\varepsilon_t)(1-\mathrm{s}(\varepsilon_t)\rho_t) w_t \frac{\varepsilon_t^2}{3\sigma_t^2} \end{bmatrix}. \tag{B6}$$

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

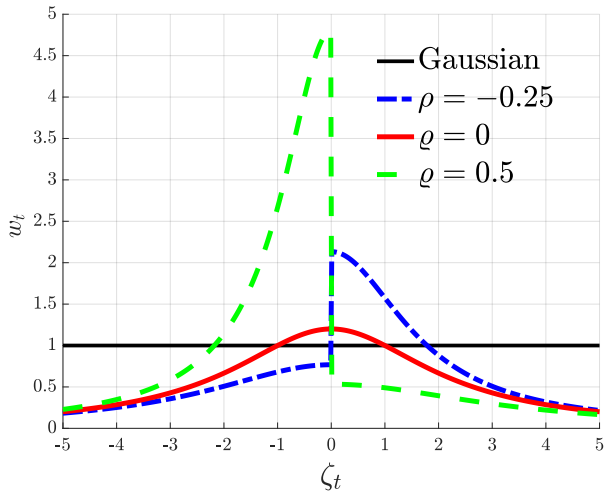
B.3 Model properties

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 of its key properties in turn. Figure B1(a) 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 larger the asymmetry, i.e., $\rho_t \to 1$, the larger the asymmetric effect of prediction errors.

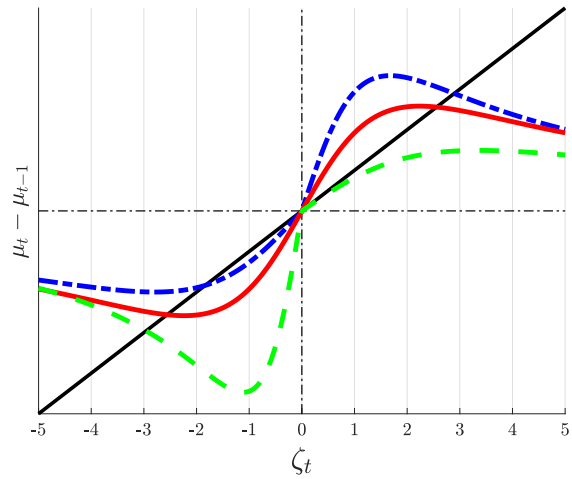
The remaining plots display Engle and Ng (1993)'s *news impact curve*, i.e. how new information – measured by the standardised prediction error – translates into updates of the parameters of the

Figure B1: **Updating weights and news impact curves**

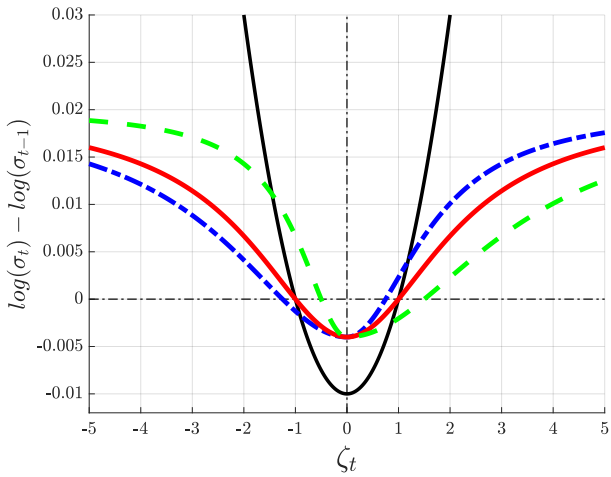
The figure reports the weighting scheme implied by w_t, and the news impact curve (NICs) for different values of the prediction error $\zeta_t = \varepsilon_t/\sigma_t$. We consider the Gaussian (black), the Student-t with $\nu = 5$ (red), and positively (blue) and negatively (green) Skew-t with $\nu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



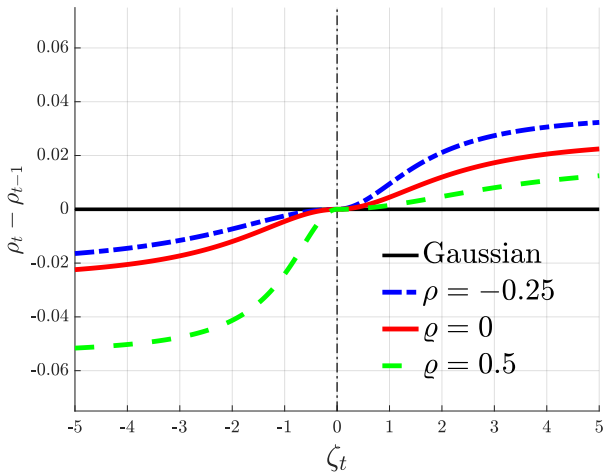
(a) Updating weights



(b) Location



(c) Scale



(d) Shape

model. Figure B1(b) shows that the location parameter updates in the direction of the prediction error. Updates of the scale parameter (Figure B1(c)) 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, Figure B1(d) shows that the shape parameter updates in the opposite direction of the prediction error so that for negative prediction errors, the distribution becomes more left-skewed.

Overall, 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. 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 (Figure B1(b)), 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 the 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. 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.

B.4 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 a Skew-t distribution can be expressed as a combination of strictly positive densities. For our modelling framework, we follow Fernández and Steel (1998) and characterise the conditional log-likelihood as a two-piece distribution;

$$\ell_t(r_t|\theta, f_t) = \text{const} - \frac{1}{2}\log\sigma_t^2 - \frac{1+\nu}{2}\begin{cases}\log\left[1+\frac{\varepsilon_t^2}{\nu(1+sgn(\varepsilon_t)\rho_t)^2\sigma_t^2}\right], & r_t \geq \mathtt{m}_t \\ \log\left[1+\frac{\varepsilon_t^2}{\nu(1-sgn(\varepsilon_t)\rho_t)^2\sigma_t^2}\right], & r_t < \mathtt{m}_t\end{cases} \tag{B7}$$

where $\theta = (\nu, A)$ collects the time-invariant degrees of freedom and the score loadings. 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 tends 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 are quite uninformative in that any evidence of time variation must reflect 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 $\bar{f}_0 = [\bar{\mathfrak{m}}_0, \bar{\delta}_0, \bar{\gamma}_0]$ collects the initial values of the time-varying parameters. Our prior specifications for these parameters are 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 multivariate Gaussian $\bar{f}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{M}_0)$ for the initial values of the time-varying parameters. The inverse Gamma prior for the score loadings is in line with the properties of the score-driven filters (for further discussion, see Juárez and Steel, 2010; Blasques et al., 2015). We set $a_\kappa = 3$, and $b_\kappa = 1$, so that *a priori* the loadings in A are positive, with a mode of 0.25. This corresponds to a quite uninformative prior centred on the possibility of a smooth update of the time-varying moments in Eq. (2).

The hyper-parameters for the Gamma prior on the degrees of freedom ν reflect a rather uninformative view on the parameters, with $d_\nu = 3$ and $D_\nu = 5$. These values allow to explore a wide range of feasible values for ν with a mode at 8 (see, e.g., Juárez and Steel, 2010).[21] 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 calibrated over an initial training period of one year of daily returns. A 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. (B7). Specifically, for each draw $\theta^i = \left(A^i, \nu^i, f_0^i\right) \sim \pi\left(\theta\right)$ for $i = 1, \ldots, M$, we simulate the time-varying parameters $\left\{f_t^i | \theta^i, f_0^i\right\}_{t=0}^T$, and evaluate the log-likelihood $\ell(r|\theta^i) = \sum_{t=1}^T \ell_t(r_t|\theta^i, f_t^i)$, such that the parameters of the model are estimated as $\theta^* = \arg\max_\theta \ell(r|\theta)$; that is, by optimizing the conditional likelihood given the prior hyper-parameters.

C Moments of the skew-t distribution

In this Section, we provide derivations for the moment of the *Skew-t*; for further details see (De Polis, 2023, Chapter 5). In what follow, to simplify the notation, we drop the time subscript from the

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

time-varying parameters. Consider the *Skew-t* distribution proposed by Gómez et al. (2007):

$$p(y|\mathtt{m},\sigma,\rho,\nu) = \frac{\mathcal{C}}{\sigma}\left[1+\frac{1}{\nu}\left(\frac{y-\mathtt{m}}{\sigma(1+sgn(y-\mathtt{m})\rho)}\right)^2\right]^{-\frac{1+\nu}{2}}, \tag{C8}$$

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 Eq. (C8) as a two-piece distribution (Fernández and Steel, 1998):

$$p(y|\mathtt{m},\sigma,\rho,\nu) = \begin{cases} \frac{\mathcal{C}}{\sigma}\left[1+\frac{1}{\nu}\left(\frac{y-\mathtt{m}}{\sigma_+}\right)^2\right]^{-\frac{1+\nu}{2}}, & y \geq \mathtt{m} \\ \frac{\mathcal{C}}{\sigma}\left[1+\frac{1}{\nu}\left(\frac{y-\mathtt{m}}{\sigma_-}\right)^2\right]^{-\frac{1+\nu}{2}}, & y < \mathtt{m} \end{cases} \tag{C9}$$

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

$$P(y \geq \mathtt{m}) = \frac{\sigma_+}{\sigma_+ + \sigma_-} = \frac{1+\rho}{2}, \qquad P(y < \mathtt{m}) = \frac{\sigma_-}{\sigma_+ + \sigma_-} = \frac{1-\rho}{2}. \tag{C10}$$

The two-piece formulation allows to consider separately the two half of the distribution when taking expectations: for $y = \mathtt{m} + \sigma\zeta$, 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).[22] Specifically, the r-th moment of ζ is defined as:

$$\mathbb{E}[\zeta^r] = \hat{\mu}_r = \frac{1}{2}\left[(1+\rho)^{r+1} + (-1)^r(1-\rho)^{r+1}\right] d_r(\nu), \tag{C11}$$

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 Eq. (C11), the moments of y are then computed as:

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

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

Therefore, the expected value y is given by:

$$\begin{aligned}\mathbb{E}[y] &= \mathtt{m} + \hat{\mu}_1\sigma \\ &= \mathtt{m} + \frac{4\nu\mathcal{C}(\nu)}{\nu-1}\rho\sigma, \qquad \nu > 1 \end{aligned} \tag{C12}$$

the variance is calculated as:

$$\begin{aligned}\mathbb{E}[y^2] &= \mathtt{m}^2 + 2\mathtt{m}\sigma\hat{\mu}_1 + \sigma^2\hat{\mu}_2 \\ &= \mathtt{m}^2 + 2\mathtt{m}\sigma\frac{4\nu\mathcal{C}(\nu)}{\nu-1}\rho + \sigma^2\frac{(1+3\rho^2)\nu}{\nu-2}, \qquad \nu > 2 \end{aligned} \tag{C13}$$

$$\begin{aligned}Var(y) &= \mathbb{E}[y^2] - \mathbb{E}[y] \\ &= \mathtt{m}^2 + 2\mathtt{m}\sigma\frac{4\nu\mathcal{C}(\nu)}{\nu-1}\rho + \sigma^2\frac{(1+3\rho^2)\nu}{\nu-2} - \left(\mathtt{m} + \frac{4\nu\mathcal{C}(\nu)}{\nu-1}\rho\sigma\right)^2 \\ &= \sigma^2\left(\frac{(1+3\rho^2)\nu}{\nu-2} - \left(\frac{4\nu\mathcal{C}(\nu)}{\nu-1}\rho\right)^2\right) \\ &= \sigma^2\left[\frac{\nu}{\nu-2} + \left(\frac{3}{\nu-2} - \left(\frac{4\nu\mathcal{C}(\nu)}{\nu-1}\right)^2\right)\rho^2\right], \qquad \nu > 2 \end{aligned} \tag{C14}$$

and the skewness is defined as:

$$\begin{aligned}\mathbb{S}kew(Y) &= Var(y)^{-\frac{3}{2}}\,\mathbb{E}[y^3] \\ &= \frac{g(\nu)\rho\left[\rho^2\left((5-2g(\nu)^2)\nu^2 + (10g(\nu)^2-19)\nu - 12g(\nu)^2\right) - \nu(\nu+1)\right]}{(\nu-3)(\nu-2)\left(\frac{\nu}{\nu-2} + h(\nu)\rho^2\right)^{\frac{3}{2}}}, \qquad \nu > 3, \end{aligned} \tag{C15}$$

with $g(\nu) = \frac{4\mathcal{C}(\nu)\nu}{\nu-1}$ and $h(\nu) = \frac{3}{\nu-2} - g(\nu)^2$.

D Bootstrap testing procedure

Consider the return on two strategies a and b, $r_{a,t}$ and $r_{b,t}$ respectively, in excess of some benchmark return. We observe a strictly stationary bivariate return distribution, $r_t = [r_{a,t},\ r_{b,t}]'$, for which T observations are available. Assume now that for each return series we wand to evaluate an *almost everywhere* differentiable function $g(\theta)$, such that $\nabla = \frac{\partial g(\theta)}{\partial\theta'} \neq 0,\ \forall\theta\,(\theta \in \Theta \wedge \theta \notin \Phi)$. We now want to draw inference on $g(\theta)$ in order to compare the performance of the return series. Define $d(\theta) = g_a(\theta_a) - g_b(\theta_b)$, such that $\theta = [\theta_a,\ \theta_b]'$, and let $\hat{\theta}$ be the estimator of θ, such that under (mild) regularity conditions

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \Psi),$$

where Ψ is a symmetric, positive semi-definite covariance matrix. By the *Delta method*,

$$\sqrt{T}(d(\hat{\theta}) - d(\theta)) \xrightarrow{d} \mathcal{N}\left(0, \nabla\Psi\nabla'\right).$$

Given $\hat{\Psi}$, a consistent estimator of Ψ (see, e.g., Andrews and Monahan, 1992; Newey and West, 1994), the standard error for $d(\hat{\theta})$ is given by

$$s(\hat{d}) = \sqrt{\frac{\nabla\hat{\Psi}\nabla'}{T}}.$$

Consider now resampling pairs of return series $r_r^* = [r_{a,t}^*, r_{b,t}^*]$ using a (circular) block bootstrap method, with block size b and $l = \lfloor T/b \rfloor$. As it is generally, θ contains sample moments;[23] hence, we can define y_t^*, a set of moment conditions of the form $y_{j,t}^{*(n)} = r_{j,t}^* - \theta_j^{*(n)}$, where (n) indicates the n^{th} element of θ_j, $j = a, b$. Therefore,

$$\hat{\Psi} = \frac{1}{l}\sum_{k=0}^{l}\zeta_k\zeta_k', \tag{D1}$$

where $\zeta_k = \frac{1}{\sqrt{b}}\sum_{t=1}^{b} y_{(k-1)b+t}^*$.

D.1 Performance measures

We first consider the benchmark performance measure, the Sharpe ratio, as an instructive case. For this measure, our results are those of Ledoit and Wolf (2008). We then consider several measures fit to measure the exposure of a portfolio to downside risks. Specifically, we consider the Sortino ratio (Sortino and Van Der Meer, 1991; Satchell, 2001), the Value-at-risk, the Expected Shortfall, the Stable Tail Adjusted Return Ratio (STARR) and the Rachev ratio (Fabozzi et al., 2005). In what follows we lay down the performance measure, the gradient and the moment conditions necessary to derive the covariance matrix of $d(\theta)$, as in Eq. (D1).

Sharpe Ratio. Give a time series of returns, the Sharpe ratio is defines as the ratio between the sample average and the sample standard deviation. Let $d(\theta) = \frac{\mu_a}{\sqrt{\gamma_a^2 - \mu_a^2}} - \frac{\mu_b}{\sqrt{\gamma_b^2 - \mu_b^2}}$, where $\mu_j = \frac{1}{T}\sum_{t=0}^{T} r_{j,t}$, $\gamma_j^2 = \frac{1}{T}\sum_{t=0}^{T} r_{j,t}^2$, and $\theta = [\mu_a\ \mu_b\ \gamma^2\ \gamma^2]'$, so that

$$\nabla = \left[\frac{\gamma_a^2}{(\gamma_a^2 - \mu_a^2)^{3/2}},\ -\frac{\gamma_b^2}{(\gamma_b^2 - \mu_b^2)^{3/2}},\ -\frac{1}{2}\frac{\mu_a}{(\gamma_a^2 - \mu_a^2)^{3/2}},\ \frac{1}{2}\frac{\mu_b}{(\gamma_b^2 - \mu_b^2)^{3/2}}\right].$$

[23] Consider the Sharpe ratio: $g(\theta) = \frac{\mu}{\sqrt{(\gamma - \mu^2)}}$, then $\theta = [\mu\ \gamma]$, where $\mu = \frac{1}{T}\sum_t r_t$ and $\gamma = \frac{1}{T}\sum_t r_t^2$.

Define $\mu_j^* = \frac{1}{T}\sum_{t=0}^{T} r_{j,t}^*$ and $\gamma_j^{*2} = \frac{1}{T}\sum_{t=0}^{T} r_{j,t}^{*2}$ the bootstrapped first and second moment of $(r_{a,t}^*, r_{b,t}^*)$, and define $y_t^* = [r_{a,t}^* - \mu_a^*, \;\; r_{b,t}^* - \mu_b^*, \;\; r_{a,t}^{*2} - \gamma_a^{*2}, \;\; r_{b,t}^{*2} - \gamma_b^{*2}]'$.

Sortino Ratio. The Sortino ratio of Sortino and Van Der Meer (1991) is the ratio of the average return in excess of some pre-specified threshold m, the so called *minimum accepted return*, and a measure of downside volatility, ς. Define $\delta(r_{j,t}, m) = \min(r_{j,t}, m)^2$ and $\varsigma_j = \frac{1}{T}\sum_{t=0}^{T}\delta(r_{j,t}, m)$, and let $d(\theta) = \frac{\mu_a}{\sqrt{\varsigma_a}} - \frac{\mu_b}{\sqrt{\varsigma_b}}$, where $\mu_j = \frac{1}{T}\sum_{t=0}^{T} h(r_{j,t}, m)$ with $h(r_{j,t}, m) = r_{j,t} - m$, and $\theta = [\mu_a \; \mu_b \; \varsigma_a \; \varsigma_b]'$, so that

$$\nabla = \left[\frac{1}{\varsigma_a}, \;\; -\frac{1}{\varsigma_b}, \;\; -\frac{\mu_a}{\varsigma_a^{3/2}}, \;\; \frac{\mu_b}{\varsigma_b^{3/2}}\right].$$

Define $\mu_j^* = \frac{1}{T}\sum_{t=0}^{T} h(r_{j,t}^*, m)$ and $\varsigma_j^* = \frac{1}{T}\sum_{t=0}^{T}\delta(r_{j,t}^*, m)$ the bootstrapped mean in excess of m and second partial-moment of $(r_{a,t}^*, r_{b,t}^*)$, and define

$$y_t^* = [h(r_{a,t}^*, m) - \mu_a^*, \;\; h(r_{b,t}^*, m) - \mu_b^*, \;\; \sqrt{\delta(r_{a,t}^*, m)} - \varsigma_a^*, \;\; \sqrt{\delta(r_{b,t}^*, m)} - \varsigma_b^*]'.$$

Value-at-Risk. Define the α-level Value at Risk (VaR), v_t^α as

$$v_{j,t}^\alpha \equiv \inf\{x \in \mathbb{R} / P(r_{j,t} \leq x) \geq \alpha\}, \tag{D2}$$

and let $d(\theta) = v_{a,t}^\alpha - v_{b,t}^\alpha$, where $\theta = [v_{a,t}^\alpha, \;\; v_{b,t}^\alpha]'$ and $\nabla = [1, \;\; -1]$. Consider the *Hit* loss function, $g(r_{j,t}, v_{j,t}^\alpha) = I(r_{j,t} < v_{j,t}^\alpha) - \alpha$; when the VaR is correctly specified, that is when $v_{j,t}^\alpha$ is the $\alpha-$quantile of the (un-)conditional distribution of the data, we can express it as the following moment condition

$$\mathbb{E}[g(r_{j,t}, v_{j,t}^\alpha) r_{j,t}] = 0,$$

in that $\mathbb{E}\left[I(r_{j,t} < v_{j,t}^\alpha)\right] = \mathbb{P}\left(r_{j,t} \leq v_{j,t}^\alpha\right) = \alpha$ by Eq. (D2). Hence,

$$y_t^* = [g(r_{a,t}^*, v_{a,t}^\alpha), \;\; g(r_{b,t}^*, v_{b,t}^\alpha)]'.$$

Expected shortfall & Value-at-Risk. The *Basel III* accord (Basel Committee on Banking Supervision, 2010) has shifted the focus from the VaR to the Expected shortfall (e_t^α, ES), defined as the expected return on an asset, conditional on the return being below its VaR,

$$e_{j,t}^\alpha = \frac{1}{\alpha}\int_0^\alpha v_{j,t}^s ds.$$

The pitfall of this measure is, however, its lack of *elicitability*, that is the ES is not the minimizer of the expectation of any loss function, which makes the definition of a suitable moment condition a

difficult task.[24] Fissler and Ziegel (2016) show that the VaR and the ES are *jointly* elicitable, that is

$$(VaR_t^\alpha, ES_t^\alpha) = \arg\min_{(v_t^\alpha, e_t^\alpha)} \mathbb{E}_{t-1}\, L_t^{FZ},$$

where $L_t^{FZ} = -\frac{I(r\leq v_t^\alpha)}{e_t^\alpha}(v_t^\alpha - r_t) + \frac{v_t^\alpha}{e_t^\alpha} + \log(-e_t^\alpha) - 1$, where I is an indicator function; then the loss function differential is

$$d(\theta) = -\frac{1}{\alpha e_{a,t}^\alpha} I(r_{a,t} \leq v_{a,t}^\alpha)(v_{a,t}^\alpha - r_{a,t}) + \frac{1}{\alpha e_{b,t}^\alpha} I(r_{b,t} \leq v_{b,t}^\alpha)(v_{b,t}^\alpha - r_{b,t}) + \frac{v_{a,t}^\alpha}{e_{a,t}^\alpha} - \frac{v_{b,t}^\alpha}{e_{b,t}^\alpha} + \log\left(\frac{e_{a,t}^\alpha}{e_{b,t}^\alpha}\right),$$

with $\theta = \left[v_{a,t}^\alpha,\ \ v_{b,t}^\alpha,\ \ e_{a,t}^\alpha,\ \ e_{b,t}^\alpha\right]'$. Following Patton et al. (2019), for $r_{j,t} \neq v_{j,t}^\alpha$:

$$\nabla = \left[\frac{\lambda_{a,t}^v}{\alpha e_{a,t}^\alpha v_{a,t}^\alpha},\ \ -\frac{\lambda_{b,t}^v}{\alpha e_{b,t}^\alpha v_{b,t}^\alpha},\ \ -\frac{\lambda_{a,t}^v + \alpha\lambda_{a,t}^e}{\alpha (e_{a,t}^\alpha)^2},\ \ \frac{\lambda_{b,t}^v + \alpha\lambda_{b,t}^e}{\alpha (e_{b,t}^\alpha)^2}\right]$$

with $\lambda_{j,t}^v = -v_{j,t}^\alpha\left(I(r_{j,t} \leq v_{j,t}^\alpha) - v_{j,t}^\alpha\right)$ and $\lambda_{j,t}^e = \frac{1}{\alpha} I(r_{j,t} \leq v_{j,t}^\alpha) r_{j,t} - e_{j,t}^\alpha$. Hence,

$$y_t^* = \left[-\frac{\lambda_{a,t}^v}{v_{a,t}^\alpha},\ \ -\frac{\lambda_{b,t}^v}{v_{b,t}^\alpha},\ \ \lambda_{a,t}^e,\ \ \lambda_{b,t}^e\right]'.$$

Stable Tail Adjusted Return Ratio. The STARR replaces the denominator of the Sharpe Ratio with a coherent measure of risk, e.g., the ES. Given all the above, define

$$d(\theta) = \frac{\mu_a}{e_{a,t}^\alpha} - \frac{\mu_b}{e_{b,t}^\alpha}$$

with $\theta = \left[\mu_a,\ \ \mu_b,\ \ e_{a,t}^\alpha,\ \ e_{b,t}^\alpha\right]'$, and

$$\nabla = \left[\frac{1}{e_{a,t}^\alpha},\ \ -\frac{1}{e_{b,t}^\alpha},\ \ -\frac{\mu_a}{(e_{a,t}^\alpha)^2},\ \ \frac{\mu_b}{(e_{b,t}^\alpha)^2}\right].$$

Now, define

$$y_t^* = [r_{a,t}^* - \mu_a^*,\ \ r_{b,t}^* - \mu_b^*,\ \ \lambda_{a,t}^e,\ \ \lambda_{b,t}^e]'.$$

[24] For example, the VaR is elicitable by means of the *tick* loss function; the mean and the median by mean of quadratic and absolute loss functions, respectively.

Rachev Ratio. Define the $(1-\alpha)$-level Expected Longrise ($e^{(1-\alpha_t)}$, ES) as

$$e_{j,t}^{(1-\alpha)} = \frac{1}{1-\alpha}\int_0^{1-\alpha} v_{j,t}^s ds,$$

and consider

$$d(\theta) = \frac{e_{a,t}^{(1-\alpha)}}{e_{a,t}^{\alpha}} - \frac{e_{a,t}^{(1-\alpha)}}{e_{b,t}^{\alpha}}$$

with $\theta = \left[e_{a,t}^{(1-\alpha)},\ e_{b,t}^{(1-\alpha)},\ e_{a,t}^{\alpha},\ e_{b,t}^{\alpha}\right]'$, and

$$\nabla = \left[\frac{1}{e_{a,t}^{\alpha}},\ -\frac{1}{e_{b,t}^{\alpha}},\ -\frac{e_{a,t}^{(1-\alpha)}}{(e_{a,t}^{\alpha})^2},\ \frac{e_{b,t}^{(1-\alpha)}}{(e_{b,t}^{\alpha})^2}\right].$$

Now, define

$$y_t^* = [\tilde{\lambda}_{a,t}^e,\ \tilde{\lambda}_{b,t}^e,\ \lambda_{a,t}^e,\ \lambda_{b,t}^e]';$$

where $\tilde{\lambda}_{a,t}^e = \frac{1}{1-\alpha} I(r_{j,t} \leq v_{j,t}^{(1-\alpha)}) r_{j,t} - e_{j,t}^{(1-\alpha)}$.

D.2 Simulation study

In this Section we report results for the asymptotic properties of the test derived above. We consider two different sizes for the simulated history of returns, $T = 2500$ and $T = 5000$. We report the empirical rejection probabilities (*erp*) for all the loss functions at the $\alpha = 1, 5, 10\%$ for $N = 2000$ replications and $M = 500$ bootstrap replication of the data. We consider three different data generating processes (DGPs): i) Normal iid returns with unit mean and variance, ii) heavy-tailed returns with unit mean and variance, for which a *Student-t* distribution with 5 degrees of freedom is employed, and iii) bivariate t-GARCH(1,1) simulated from a diagonal BEKK model of Engle and Kroner (1995). All returns are simulated with unit mean and variances.

Choice of the block size. Give a set of reasonable block sizes, we select the one that minimizes the difference between the empirical rejection probability and a specified acceptance level, generally set a 5%. The procedure is akin to the one of Ledoit and Wolf (2008), who target coverage levels. We first pre-whiten the data by fitting a parametric linear model to the data, e.g. a VAR(1), and we bootstrap the residuals by means of Politis and Romano (1994) stationary bootstrap in order to remove any non-linear dependence not captured by the linear model. We then generate $K = 2000$ pseudo-samples from the VAR(1) estimates and the bootstrapped residuals to compute the empirical

rejection probability (or empirical p-value) as

$$erp = \frac{1 + \sum_{m \in M} I(\hat{p}_m \geq \bar{p})}{1 + M},$$

where $\hat{p}_m$ is the p-value for the m^{th} sample, $\bar{p}$ is the p-value for the original data and I is an indicator function. The size properties of the tests are reported in Table D1. Overall, all tests seem to present adequate sizes, with the exception of the VaRs and ESs appear to be slightly oversized.

Table D1: Calibration of the bootstrap block size

T	2500			5000		
	10%	5%	1%	10%	5%	1%
	Gaussian-iid					
Sharpe	0.100	0.052	0.010	0.081	0.038	0.008
Sortino	0.098	0.056	0.009	0.092	0.039	0.008
VaR(5%)	0.086	0.045	0.008	0.098	0.042	0.011
ES(5%)	0.084	0.042	0.008	0.101	0.051	0.012
ES(5%)	0.093	0.044	0.012	0.101	0.051	0.012
Rachev	0.102	0.051	0.006	0.085	0.037	0.006
STARR	0.104	0.052	0.011	0.087	0.038	0.006
	t_5-*iid*					
Sharpe	0.086	0.043	0.007	0.100	0.049	0.006
Sortino	0.083	0.041	0.007	0.097	0.052	0.011
VaR(5%)	0.097	0.042	0.009	0.126	0.065	0.016
ES(5%)	0.120	0.054	0.012	0.153	0.094	0.028
ES(5%)	0.118	0.055	0.013	0.148	0.089	0.029
Rachev	0.096	0.047	0.010	0.114	0.063	0.013
STARR	0.086	0.046	0.007	0.098	0.047	0.006
	Bivariate t_5-*GARCH*					
Sharpe	0.092	0.037	0.006	0.107	0.053	0.012
Sortino	0.091	0.039	0.004	0.105	0.047	0.010
VaR(5%)	0.143	0.083	0.016	0.169	0.097	0.029
ES(5%)	0.130	0.056	0.012	0.143	0.071	0.019
ES(5%)	0.141	0.068	0.019	0.135	0.072	0.014
Rachev	0.092	0.042	0.008	0.104	0.045	0.009
STARR	0.097	0.046	0.008	0.096	0.044	0.008

E Additional Results

In this section, we report a set of additional results related to the dynamics of the conditional volatility and skewness estimates for alternative momentum strategies, as well as additional robustness checks for the portfolio implementation outlined in Section 5.

E.1 Conditional estimates for alternative momentum factors

Figure E1 presents the estimates of the conditional volatility $\sqrt{\mathbb{V}_t(r_{t+1})}$ and skewness $Sk_t(r_{t+1})$ for the short-term momentum `6_2` and intermediate momentum `12_7` strategies. The dynamics of the conditional volatility and skewness for the `12_7` and `6_2` momentum portfolios are broadly consistent with the benchmark `12_2` implementation (see Figure 5 in the main text). For instance, both the short and intermediate momentum experience spikes in returns volatility during the great depression which coincide with deepening negative conditional skewness.

Again, similar to the `12_2` portfolio, conditional skewness tend to deteriorates during economic recessions while becomes zero, in fact at times positive, during economic expansions, especially if upturn in economic activity are for prolonged periods. Overall, except for few nuances, the dynamics of both volatility and skewness is rather consistent across different momentum portfolios.

E.2 Time-series estimates of the location parameter $\mathtt{m}_t$

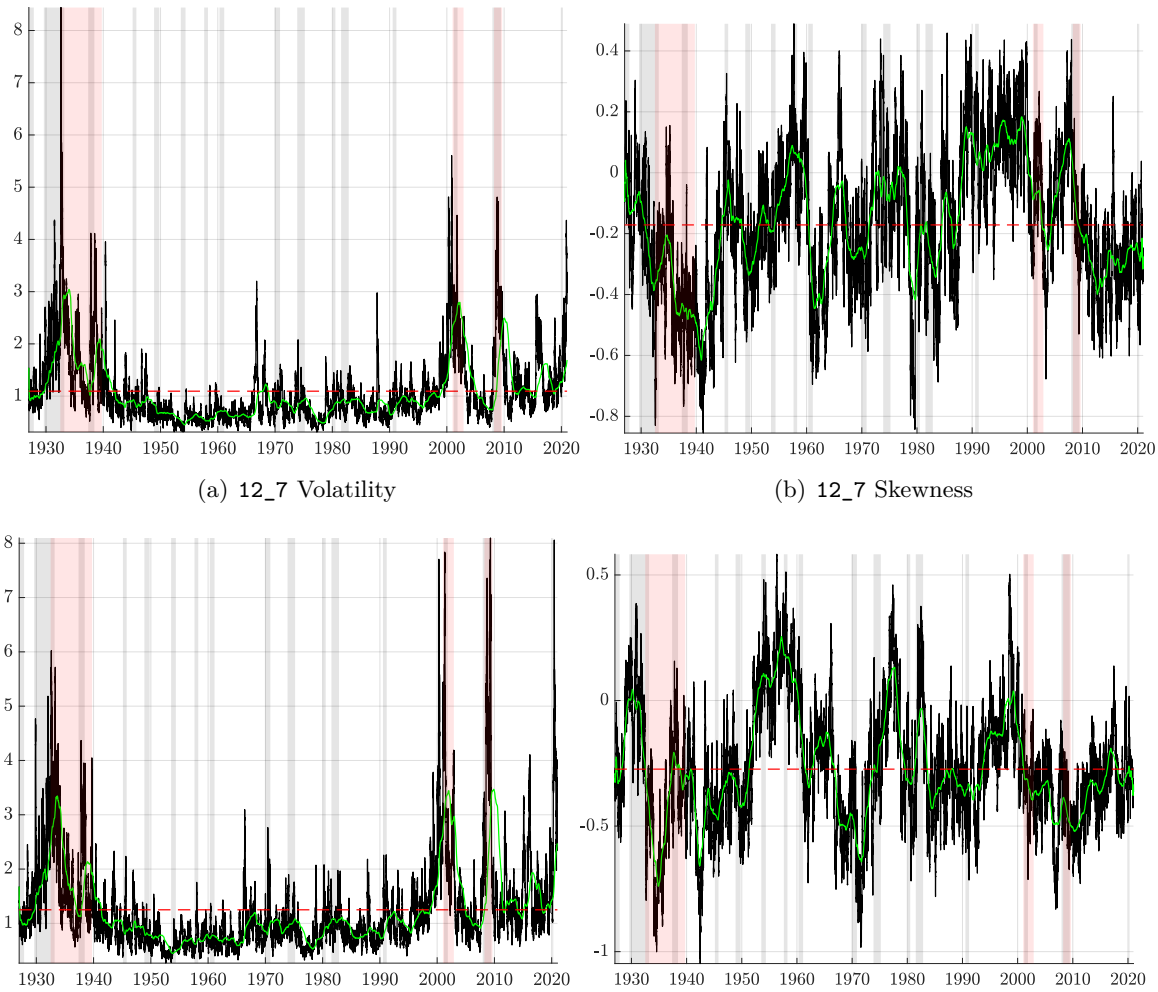
The location parameter $\mathtt{m}_t$ captures the centre of the distribution and is equivalent to the conditional mean only under symmetric distributional assumptions – when the returns' asymmetry $\rho_t = 0$ as from Eq. (4). Figure Fig. E3 reports the estimates of $\mathtt{m}_t$ for `12_2` momentum portfolio. The estimates for the `12_7` and `6_2` momentum portfolios are similar (see the dynamics of conditional skewness in Figure E1) and are available upon request. Two things emerge: first, the dynamics of the location parameter $\mathtt{m}_t$ (red line) is much more stable than the conditional mean $\mathbb{E}_t(r_{t+1})$. This implies that the majority of the variation in the dynamics of expected returns is primarily driven by the interplay between conditional volatility and skewness (see Figure 8(a)).

Second, there is a major disconnect between $\mathtt{m}_t$ and the conditional mean $\mathbb{E}_t(r_{t+1})$ (black line). 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 $\mathtt{m}_t$ remains persistently in positive territory for both periods.[25]

[25] Recall that for a given σ_t, the disconnect $\mathbb{E}_t(r_{t+1}) < \mathtt{m}_t$ implies that $\rho_t < 0$ (see Eq. (4)).

Figure E1: **Conditional estimates for short-term and intermediate momentum**

The plot reports the time-varying volatility (left) and skewness (right) estimates for the `12_7` and `6_2 WML` portfolio returns. The red dashed lines represent the sample mean, whereas green lines highlight 2-year moving averages of the daily 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 January 1st 1927 to December 31th 2020.



(a) `12_7` Volatility

(b) `12_7` Skewness

(c) `6_2` Volatility

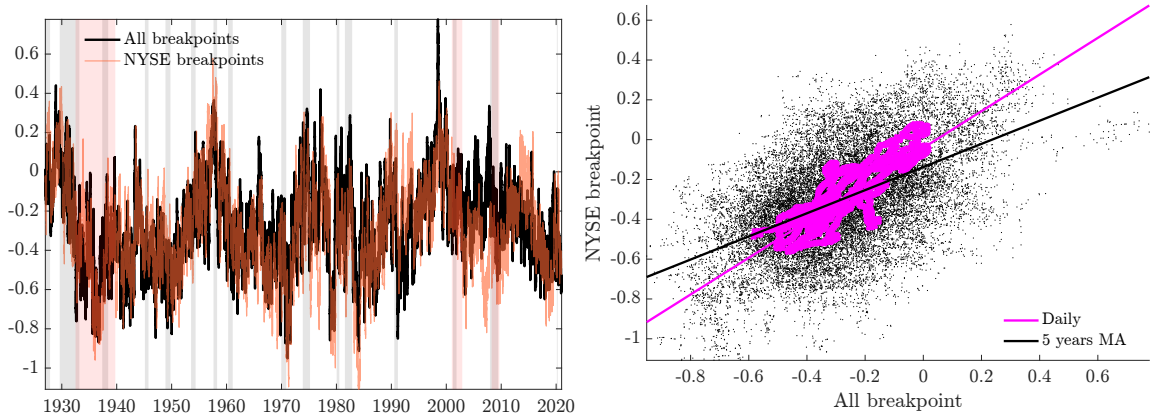
(d) `6_2` Skewness

E.3 Additional portfolio results for monthly rebalancing

Table 8 reports the results for the monthly rebalancing for the `12_2` benchmark momentum implementation. In Table E1, we report additional monthly performances for the short-term `6_2` and intermediate `12_7` momentum portfolios (see Section 2 for a description). Similar to the daily results, the monthly performance across different momentum implementations is fairly comparable to the benchmark `12_2` momentum portfolio.

Figure E2: **Conditional estimates for momentum based on NYSE breakpoints**

The left panel reports the skewness estimates for the `12_2` strategy implemented based on the NYSE breakpoints as in Fama and French (1996). NBER recession are identified by gray shaded areas, while red shaded areas highlight momentum crashes periods, as indicated in Daniel and Moskowitz (2016). The right panel report the correlation between the estimate in the main text and the estimates based on the NYSE breakpoints. The sample period is from January 1st 1927 to December 31st 2020.



(a) Conditional skewness estimates

(b) Correlation over the sample

This is due to a fairly similar dynamics in the conditional skewness and volatility, which, except for few nuances during the 2001 burst of the dot-com bubble and the crash over the great financial crisis, are quite comparable across different momentum implementations.

E.4 Risk aversion

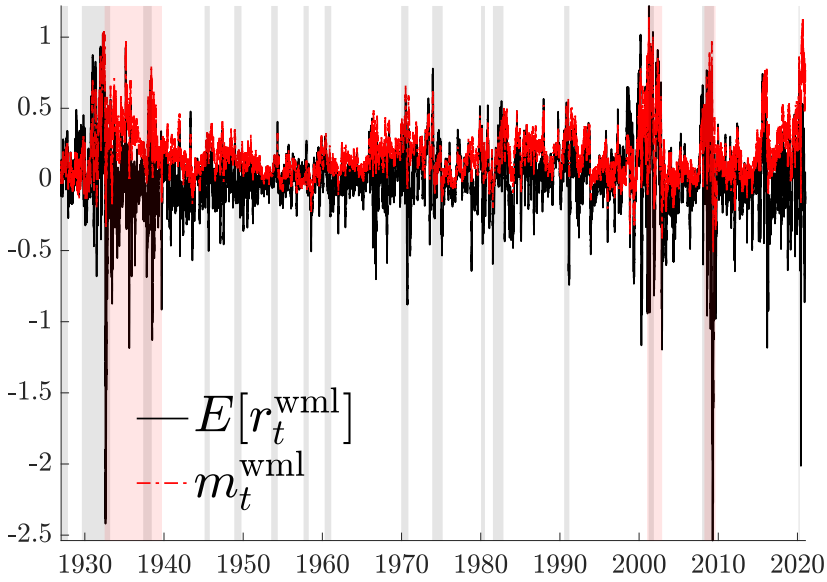
In this Section, we repeat the economic evaluation of Table 6 controlling for different levels of risk aversion. Table E2 reports the performance fees $\mathcal{F}$ for risk aversion levels of 2, 7 and 15. These levels compare agents with a strong risk aversion to investors prone to take on more risks. 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 hedging for predictable variations in the returns skewness is economically meaningful, regardless the level of risk aversion.

F Auxiliary results for the asset pricing implications

In this section we report a series of estimates of the conditional CAPM specification as in Eq. (11), both static and dynamic. These estimates are used to implement the simulation study in Section 6 which compares the conditional skewness estimates from our model against a state-dependent CAPM regression.

Figure E3: **Expected returns and the location parameter**

The plot reports the time-varying location parameter $\mathtt{m}_t$ (red line) and the conditional expected returns as in Equation (4) (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 December 31th 2020, daily.



F.1 State-dependent CAPM estimates

Figure F1 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 on the following regression:

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

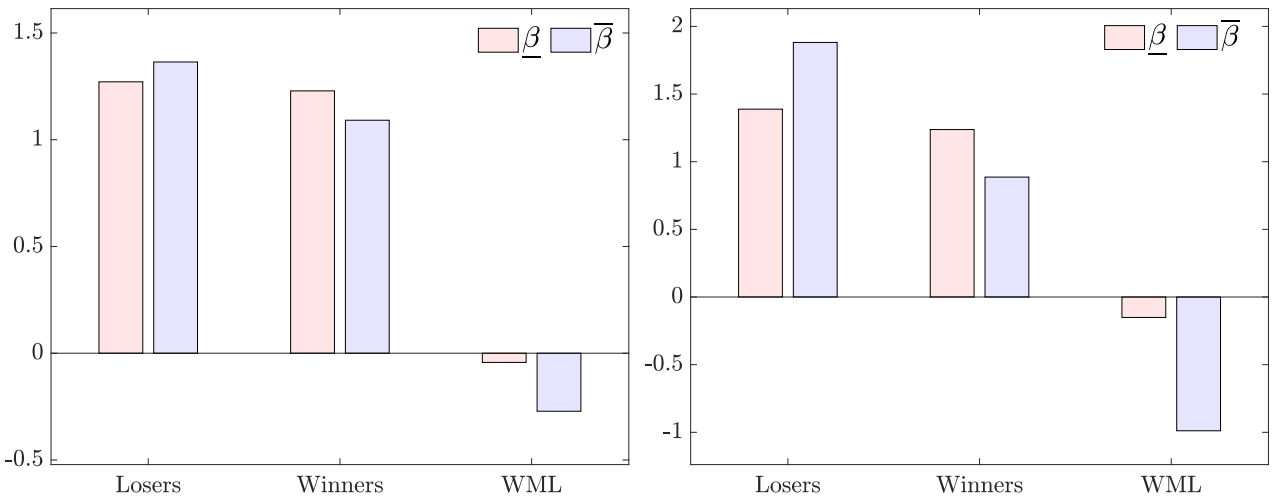
The daily estimates show that the `losers`' portfolio is more exposed to upside market risk ($\overline{\beta} = 1.36$) as compared to downside market risk ($\underline{\beta} = 1.27$), in relative terms compared to the unconditional market beta ($\beta = 1.31$). The opposite holds for the `winners`' 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 F1: **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

Table E1: **Monthly rebalancing for alternative momentum portfolios**

The table reports monthly returns on our skewness-adjusted maximum conditional Sharpe ratio strategy (`mSSR`) against a variety of alternative managed-momentum portfolios. We consider Daniel and Moskowitz (2016) (`DM2016`) and Barroso and Santa-Clara (2015) (`BS2015`) which are based on aggregating daily recursive estimates of the realised variance, as well as a semi-volatility targeting as proposed by Wang and Yan (2021) and Hanauer and Windmüller (2023). We report in parentheses the bootstrapped p-values for all performance measures (see Appendix D). The out-of-sample period is from January 1930 to December 2020. Panel A reports the results for intermediate momentum `12_7`. Panel B reports the results for short-term momentum `6_2`.

Panel A: Intermediate momentum `12_7`

Strategies	Sharpe	Sortino	STARR	RR	VaR	ES	dVol	Skew	MaxDD	Leverage
mSSR	1.004	2.238	2.914	2.340	−0.485	−0.769	8.340	0.411	0.436	0.771
`cdVol`	1.389 (0.000)	2.942 (0.188)	3.498 (0.000)	1.785 (0.912)	−0.727 (0.000)	−0.886 (0.020)	8.777	0.055	0.301	1.260
`DM2016`	1.058 (0.523)	1.959 (0.356)	2.544 (0.551)	1.784 (0.495)	−0.609 (0.081)	−0.928 (0.129)	10.041	0.240	0.541	1.095
`BS2015`	1.061 (0.534)	1.835 (0.218)	2.181 (0.504)	1.386 (0.306)	−0.780 (0.001)	−1.085 (0.106)	10.749	0.029	0.611	0.405
`WML`	0.668 (0.137)	0.904 (0.028)	1.393 (0.057)	1.144 (0.043)	−0.745 (0.016)	−1.071 (0.058)	13.750	−0.029	1.251	

Panel B: Short-term momentum `6_2`

Strategies	Sharpe	Sortino	STARR	RR	VaR	ES	dVol	Skew	MaxDD	Leverage
mSSR	1.015	2.125	2.805	2.256	−0.581	−0.807	8.879	0.347	0.470	0.710
`cdVol`	1.266 (0.013)	2.484 (0.336)	2.831 (0.008)	1.506 (0.522)	−0.784 (0.006)	−0.998 (0.376)	9.475	0.031	0.824	1.104
`DM2016`	1.113 (0.329)	2.147 (0.948)	2.417 (0.575)	1.668 (0.715)	−0.728 (0.050)	−1.027 (0.182)	9.635	0.226	0.318	0.841
`BS2015`	0.964 (0.613)	1.580 (0.103)	1.816 (0.666)	1.154 (0.138)	−0.913 (0.000)	−1.184 (0.069)	11.343	−0.063	1.046	0.365
`WML`	0.494 (0.008)	0.651 (0.008)	0.844 (0.003)	0.859 (0.012)	−0.852 (0.032)	−1.305 (0.044)	14.110	−0.163	2.117	

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,

$$\underset{\sim}{\beta}_t^i = \frac{\text{cov}_t(\tilde{r}_{t+1}^i, \min\{\tilde{m}_{t+1}, 0\})}{\text{var}_t(\min\{\tilde{m}_{t+1}, 0\})} \quad i = \texttt{losers}, \texttt{winners}, \texttt{WML}, \tag{F1}$$

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. (F1) 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

Table E2: **The role of 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 January 1nd 1927 to December 31th 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			cdVol			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	12.234	9.586	5.171	9.838	7.189	2.775	8.829	6.180	1.766	6.937	4.162	0.000
1	11.856	9.081	4.793	9.586	6.937	2.523	8.829	6.054	1.766	6.937	4.162	0.000
5	10.090	7.441	3.027	8.829	6.180	1.766	8.703	5.928	1.640	6.811	4.162	0.000
10	7.946	5.297	0.883	7.820	5.171	0.757	8.450	5.802	1.387	6.811	4.162	0.000

downswings. Upside betas $\overline{\beta}_t^i$ hold a similar interpretation and are computed by substituting the *min* function in Eq. (F1) with the *max* operator.

Figure F2 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).[26] 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. 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.

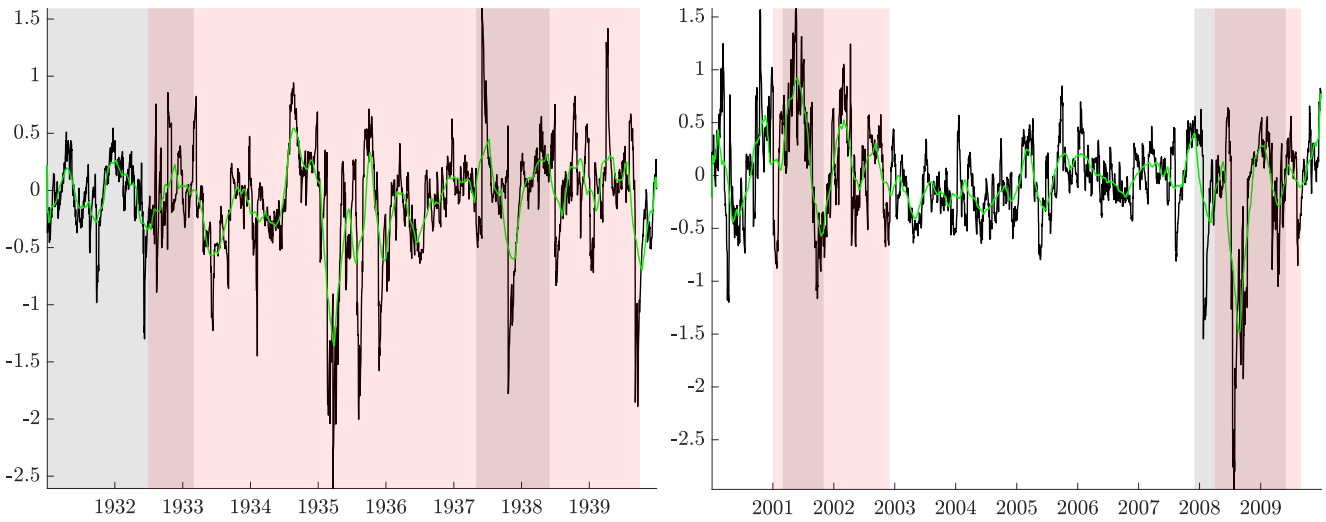
G State-dependent betas and returns asymmetry

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

[26] 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 F2: **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

conditional regression model in Eq.(10),

$$r_t = \alpha + \underbrace{\overline{\beta} m_t I(m_t \geq \mu_m) + \underline{\beta} m_t I(m_t < \mu_m)}_{\beta m_t} + e_t \tag{G1}$$

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 mean 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),

$$f\left(\beta m_t\right) = \begin{cases} C \exp\left\{-\frac{1}{2\underline{\sigma}_m^2}\left(\underline{\beta} m_t - \beta \mu_m\right)^2\right\} & \text{if } m_t \leq \mu_m \\ C \exp\left\{-\frac{1}{2\overline{\sigma}_m^2}\left(\overline{\beta} m_t - \beta \mu_m\right)^2\right\} & \text{if } m_t > \mu_m \end{cases} \tag{G2}$$

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

$$E\left[\beta m_t\right] = \sqrt{\frac{2}{\pi}} \left(\overline{\sigma}_m - \underline{\sigma}_m\right) + \beta \mu_m, \tag{G3}$$

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.(G2) takes the form ,

$$V\left[\beta m_t\right] = \left(1 - \frac{2}{\pi}\right) \left(\overline{\sigma}_m^2 - \underline{\sigma}_m^2\right)^2 + \overline{\sigma}_m \underline{\sigma}_m \tag{G4}$$

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

$$\begin{aligned} E\left[\beta m_t\right] - \beta \mu_m &= \sqrt{\frac{2}{\pi}} \left(\overline{\sigma}_m - \underline{\sigma}_m\right) \propto \sqrt{\overline{\beta}^2 \sigma_m^2} - \sqrt{\underline{\beta}^2 \sigma_m^2}, \\ &= \sigma_m \left(\sqrt{\overline{\beta}^2} - \sqrt{\underline{\beta}^2}\right) \\ &= \sigma_m \left(\overline{\beta} - \underline{\beta}\right) \end{aligned} \tag{G5}$$

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.

References

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

Andrews, D. W. and Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica: Journal of the Econometric Society*, pages 953–966.

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.

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.

Basel Committee on Banking Supervision (2010). Basel iii: A global regulatory framework for more resiliant banks and banking systems. Technical report, Basel Committee on Banking Supervision.

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.

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.

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. 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.

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 Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric theory*, 11(1):122–150.

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.

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.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and osband's principle.

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.

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.

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

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.

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.

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.

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

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4):631–653.

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

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of econometrics*, 211(2):388–413.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313.

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

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.

Satchell, S. E. (2001). Lower partial-moment capital asset pricing models: a re-examination. In *Managing Downside Risk in Financial Markets*, pages 156–168. Elsevier.

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.

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.