Journal of Financial Markets 63 (2023) 100769

Contents lists available at ScienceDirect

Journal of Financial Markets

journal homepage: www.elsevier.com/locate/finmar

Predicting the equity risk premium using the smooth cross-sectional tail risk: The importance of correlation☆

José Afonso Faias *

ARTICLE INFO

JEL classification:
G11
G14
G17

Keywords:
Equity premium
Prediction
Cross-sectional

ABSTRACT

I provide a new monthly cross-sectional measure of stock market tail risk, *SCSTR*, defined as the average of the daily cross-sectional tail risk, rather than the tail risk of the pooled daily returns within a month. Through simulations, I find that *SCSTR* better captures monthly tail risk rather than merely the tail risk on specific days within a month. In an extended period from 1964 until 2018, this difference is important in generating strong in- and out-of-sample predictability and performs better than the historical risk premium and other commonly-used predictors for short- and long-term horizons.

1. Introduction

Studying the properties of extreme events and market crashes has become more important in light of recent economic crises and market bubbles. Attempts have been made to establish a link between such events and firm-specific characteristics (Wang et al., 2009; Bali et al., 2011; Annaert et al., 2013), as well as market-wide factors (Longin and Solnik, 2001; Poon et al., 2004; Fousseni et al., 2018). Regarding the latter, the theory suggests that during periods of increasing tail risk, investors demand higher equity premiums. In addition, the strong persistence of the tail structure (Gardes and Stupfler, 2013; Kelly and Jiang, 2014) indicates that the implied return predictability is substantial. Concurrently, there has been an ongoing debate about equity premium predictability. Bossaerts and Hillion (1999) and Goyal and Welch (2008) show no evidence of predictability, whereas Campbell and Thompson (2008), Lettau and Nieuwerburgh (2008), Ferreira and Santa-Clara (2011), Drechsler and Yaron (2011), and Rapach et al. (2016) support the view of

☆ I thank the editor and the referee for valuable comments that significantly improved the paper. I also thank Rui Albuquerque, Marco Bonomo, Miguel Ferreira, Stefano Giglio, Maria Teresa Gonzalez, Marco Lyrio, Thierry Post, Pedro Santa-Clara, and all participants at the Lubrafin 2016 Meeting, 2018 Foro de Finanzas, 2018 Portuguese Finance Network, 2018 Multinational Finance Society Annual Meeting and the seminar at Nazarbayev University for helpful comments. I am particularly grateful to Pavel Onyshchenko, João Monteiro, Dan Tran, and Pasquale Marotta for outstanding research assistance. I acknowledge the support from FCT - Fundação para a Ciência e a Tecnologia for the projects UID/GES/00407/2013, UID/GES/00407/2020, TDC/IIM-FIN/2977/2014, and PTDC/EGE-OGE/30314/2017.

* Universidade Católica Portuguesa, Católica Lisbon School of Business and Economics, Palma de Cima, 1649-023, Lisbon, Portugal.
E-mail address: jfaias@ucp.pt.

https://doi.org/10.1016/j.finmar.2022.100769
Received 23 July 2019; Received in revised form 8 July 2022; Accepted 12 July 2022
Available online 19 July 2022

predictability. In this paper, I participate in this debate.

As my main contribution, I propose a measure of cross-sectional tail risk that predicts the U.S. stock market equity premium in-sample (IS) and out-of-sample (OOS) tests. My measure, smooth cross-sectional tail risk (*SCSTR*), is estimated by applying Hill's (1975) estimator to the daily cross-section of stock returns and by averaging the obtained tail exponents within a month. As an alternative, Kelly and Jiang (2014) estimate the tail exponent based on pooled daily observations (*CSTR*) over the entire month.[1] Initially, the two estimators may seem to be almost identical, but the two measures have intrinsic and meaningful differences. To clarify these differences, take the fictitious example of a month of 20 days with the same 20 firms on each day. Let us assume an extreme case such that the first day returns for the firms are a ranked sequence from −50% to −31% in steps of 1%, and the remaining days are −5%. Assume that the 5th percentile of returns is the cutoff that defines the distribution tail. For simplicity, assume that the measures are computed as simple averages of returns below the cutoff. The intuition is unchanged for the correct measures, but it facilitates quick understanding in this exercise. In this example, the *CSTR* measure corresponds to 20 observations (5% of the pooled sample of the 400 observations), which, in this case, are given by the first day returns for the 20 firms. The tail risk is due to a specific day, day 1, and is equal to −40.5%. To compute my measure, *SCSTR*, I run the same logic over each day. The 5% threshold corresponds to one firm for each day. Then, on the first day, it is equal to −50%, and on the remaining days, it is −5%. When I average over the 20 days, I get a tail risk value of −7.25%, which is a much smoother value and is related more to the real monthly tail risk than to a specific limited number of days. The two measures yield the same results if, instead, the −50% to −31% range of returns were each observed on a different day of the month. Subsequently, using real data, I show that the difference between the two measures is important through time and that it has surged in the last decade. I compute a sort of Herfindahl index to understand this phenomenon. I compute this index for the proportion of stocks used each day within a month in which the tail risk measure is assessed. Consider the previous hypothetical example. The Herfindahl index takes 5% for each day with the *SCSTR* and takes 100% for one of the days and 0% for the remaining days with the *CSTR*. The difference between the two measures shows how the *CSTR* is concentrated on some days. The numbers are quite different between the two measures, *CSTR* and *SCSTR*. The evolution of the Herfindahl index for *SCSTR* is flat around 5% by construction. The Herfindahl index of *CSTR* is very volatile and consistently increasing since 2005, and it ranges between 5% and 16%. In addition, I run a simulation exercise to understand the fit of *SCSTR* and *CSTR* to the empirical tail distribution. I show that *CSTR* is a better fit for the daily empirical tail distribution for most values of the parameters, but *SCSTR* is a better fit for the monthly empirical tail distribution. Since I am predicting the average monthly market excess return, the latter is more adequate to this end. Does this translate to stronger equity risk premium predictability by *SCSTR*?

Using data for the last 50 years to evaluate the predictive power of various valuation ratios and macroeconomic variables used in previous studies, I show that *SCSTR* has higher OOS performance than the historical average of the equity risk premium and all predictors, even the one proposed by Kelly and Jiang (2014), *CSTR*. I also run the two simultaneously and again *SCSTR* is the one getting the correct and expected sign. Then, I confirm using the orthogonal component of *SCSTR* on *CSTR* that the "pure" *SCSTR* holds stronger predictive power for the equity premium. All in all, it points out the superior and consistent predictive power of *SCSTR* over the *CSTR*. Ever since Bates and Granger (1969), the combination of forecasts has been proven to be a better predictor than individual forecasts (e.g., Stock and Watson, 2004; Timmermann, 2006; Rapach et al., 2010; Elliott and Drive, 2011; Elliott et al., 2013). However, I show that the predictions using *SCSTR* (one variable) beat predictions using a combination of forecasts and that excluding *SCSTR* from the combinations yields worse results in most cases. Next, I study the impact of applying regression restrictions based on financial theory rather than simply applying a pure predictive regression, as in Campbell and Thompson (2008). My measure remains the most robust equity risk premium predictor and remains significant for all horizons.

The next step is to understand the disaggregation of the stock market into industries. Perhaps my results are just a result of a specific sector. I construct the same type of measures for each of the 10 Kenneth French industries available from his website. For this case, I use these measures to predict not only equity premiums for the market but also for each industry. *SCSTR* seems to be innocuous to industry focus. The results are strong both IS and OOS, i.e., for the 1-month horizon, 6 industries' *SCSTR* predicts market and industry excess returns with an OOS R^2 greater than 1%. For a 3-year horizon, again 6 industries predict market and industry excess returns with an OOS R^2 greater than 15%. Regarding *CSTRs*, the number of industries that are OOS significant is scarce, and the magnitude of predictability is quite low.

I address the relation between the good performance of my measure and the business cycle. *SCSTR* is the only one that yields positive and significant R-squares for all horizons in expansion periods (which represent the great majority of periods in the sample). This finding is remarkable since most previous predictors are better during recession periods. Nonetheless, *SCSTR* also yields positive R-squares in recession periods, although only significant in the short horizons. I also study the alternative use of a rolling window for the forecasts, changing the length of the estimation window and addressing concerns related to dynamic betas and illiquidity. The conclusions still hold. All these results provide evidence that *SCSTR* is a robust equity risk premium predictor. However, does this strong statistical predictability translate into economic value to investors?

[1] To account for tail risk in the distribution of stock returns, researchers apply one of the two following approaches. The first one, which is widely applied in risk management, replaces the variance as a measure of risk by using the expected shortfall (ES) (Rockafellar and Uryasev, 2000). Due to the small number of observations in the tails, the accuracy of ES estimates suffers from high variance (Danielsson et al., 1998). Consequently, using ES yields poor out-of-sample performance. Therefore, I turn to extreme value theory (EVT), which offers an alternative methodology for measuring tail risk. Embrechts et al. (1997) propose nonparametric and parametric approaches that more accurately reflect the likelihood of extreme events. I assume that the extreme negative stock returns follow the power law and estimate the tail exponent by applying Hill's (1975) estimator to the daily cross-section of returns.

I use a portfolio choice between the stock market and risk-free asset rebalancing every month. I compare the case of the historical first moment of the distribution versus the predictive power of either *SCSTR* or *CSTR*. I confirm that using predictability information adds value in a power utility setting. This addresses any concerns of not incorporating higher moments on the return distribution. I use both an expansion up to the fourth order of the constant relative risk aversion (CRRA) expected utility function and conditional portfolios to show that *SCSTR* predictability significantly increases the certainty equivalent by 5%–10% when compared to the baseline case of no predictability in the period from 1964 to 2018, whereas *CSTR* only increases by approximately 3%–7% for an investor with a risk aversion parameter of 2. This result shows the practical importance of using *SCSTR* as a predictor of stock market returns.

The final step is to understand whether the time-varying predictability by either *SCSTR* or *CSTR* price stock returns on the cross-section. I first compute the exposure of stock returns to each tail risk measure and sort this sensitivity in each month into 10 deciles. Next, I compute equally- and value-weighted portfolios of each bin on next-month return performance. Finally, I compute long-short portfolios by taking a long position on the highest decile with deep positive sensitivity to each tail risk measure and taking a short position on the lowest decile with deep negative sensitivity to each tail risk measure. The results again show the strong performance of *SCSTR* versus *CSTR*. I confirm that the results hold true even when I compute the sensitivities controlling for a Carhart (1997) factor model. Finally, I compute sensitivities for *SCSTR* and *CSTR* simultaneously and the results lend the correct sign for the case of *SCSTR*. All in all, this shows again the dominance of *SCSTR* now in the case of the cross-section.

The paper is organized as follows. In Section 2, I define the new predictor and contrast it with the previous published measure through a simulation study. In Section 3, I run an equity risk premium predictability exercise and include several robustness tests on the baseline results. In Section 4, I discuss the economic value of the predictability exercise through a portfolio choice application. In Section 5, I discuss cross-sectional results. I conclude in Section 6.

2. Cross-sectional tail risk measures

In this section, I define and estimate my new measure of cross-sectional tail risk. Then, I explain the differences from those defined in the previous studies. Finally, I run a simulation study to assess the conditions under which *SCSTR* is a better estimator of the true tail risk value than *CSTR*.

2.1. Definition

Several studies have developed measures for stock market tail risk in univariate and bivariate frameworks (Kearns and Pagan, 1997; Poon et al., 2004; Frahm et al., 2005; Bollerslev et al., 2013; Faias and Zambrano, 2022). However, these approaches are not suitable in my case because of the necessity of a large estimation period and the heteroskedasticity of high-frequency data (De Vries, 1991; Ghose and Kroner, 1995).

Instead, I use panel data on daily stock returns within a month. The idea is to capture the common component of individual stocks' tail distribution in a single aggregate measure. Under this methodology, the lower tail of the returns distribution is assumed to follow the power law:

$$P\left(R_{i,t+1} < r \,\middle|\, R_{i,t+1} < u_t \text{ and } F_t\right) = \left(\frac{r}{u_t}\right)^{-\frac{a_i}{\lambda_t}},\ r < u_t < 0, \tag{1}$$

where a_i/λ_t is a tail exponent that determines the shape of a tail; u_t is an extreme negative threshold that separates the tail from the body of distribution; and F_t is the set of information available at a point in time. The tail exponent defines the shape and structure of the tail. a_i is a constant that determines the level of the tail risk of a particular asset, while the common dynamics of different assets' tail distributions are captured by the time-varying factor λ_t. The specification of the model implies that $r/u_t > 1$; therefore, the tail exponent a_i/λ_t is always greater than 0 to ensure that the probability that a return falls below the threshold is between 0 and 1. The higher the level of λ_t, the "fatter" the tail of the distribution and the greater the probability of extreme returns.

A similar power law rule has been applied to stock returns (Jondeau and Rockinger, 1999; Poon et al., 2004; Gardes and Stupfler, 2013). The common time-varying factor in the tail distribution of returns requires maximum likelihood estimation. However, a simpler methodology can be applied, as the results are qualitatively the same and quantitatively nearly identical (Kelly and Jiang, 2014). Kelly and Jiang (2014) estimate the monthly tail exponent by applying Hill's (1975) power law estimator to the pooled cross-section of daily stock returns in month m:

$$CSTR_m = \frac{1}{K_m}\sum_{i=1}^{K_m} ln\frac{R_{i,m}}{u_m}, \tag{2}$$

where K_m is the number of daily returns that fall below the threshold u_m for month m. They use the 5th percentile of pooled daily returns as the threshold u_m (Poon et al., 2004; Gabaix et al., 2006). Note that only the observations in the tails of the pooled cross-section are used for tail exponent estimation because the body of distribution may not follow the power law. For this reason, a large sample is needed to ensure a sufficient number of observations for measurement.

I propose a new measure of monthly *CSTR* by applying Hill's (1975) estimator to the cross-section of daily returns during day d and, afterward, by averaging the obtained daily tail exponents (λ_d) over month m:

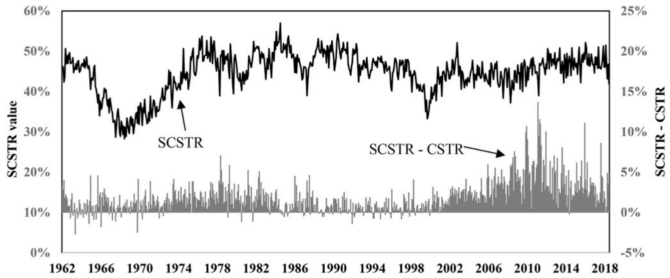

Fig. 1. Evolution of cross-sectional tail risk measures.
The top black line reports the evolution of cross-sectional tail risk, *SCSTR*, for the U.S. market. The relative difference between *SCSTR* and *CSTR* is reported at the bottom of the figure using bars. The variables *SCSTR* and *CSTR* are explained in Section 2.

$$SCSTR_m = \frac{1}{N_m} \sum_{d=1}^{N_m} \lambda_d, \tag{3}$$

where:

$$\lambda_d = \frac{1}{K_d} \sum_{i=1}^{K_d} ln \frac{R_{i,d}}{u_d} \tag{4}$$

and u_d is the bottom 5th percentile of the sorted cross-section of returns for day d, K_d is the number of returns that fall below the threshold u_d, and N_m is the number of days in month m.

There are two advantages to averaging daily tail exponents over the tail exponent of the pooled daily cross-section within a month. My new measure better captures the cross-section and time series effects. Specifically, it avoids market tail risk being inflated by individual stocks with the highest tail exponents. Additionally, I assign the same weight to all the daily tail risk estimates within a month. Consequently, *SCSTR* accounts for the entire cross-section of market tail risk and its spread over the month rather than simply pooling the most negative individual stock returns, which can all be contained within a few days of a month, as in *CSTR*.

2.2. Estimation

To estimate the monthly tail risk of the U.S. market, I extract the daily returns from the Center for Research in Security Prices (CRSP) for all common NYSE/AMEX/NASDAQ stocks (share codes 10 and 11) for the period from July 1962 to December 2018. It was inappropriate to consider the period before July 1962 owing to the insufficient number of observations in the CRSP cross-section.[2] In the beginning of the sample, the addition of AMEX to CRSP data almost doubled the number of stocks to 2000; in December 1972, after the inclusion of the NASDAQ, the number of stocks in the sample reached 5400. I observe the largest cross-section width at the end of 1997 when the number of stocks in CRSP reached approximately 7500.

For estimation purposes, I assume that the size, price, and liquidity of stocks do not bias the tail risk estimate, as in Longin and Solnik (2001), Poon et al. (2004), and Kelly and Jiang (2014). Thus, the only filter applied to the data is the availability of returns at the date of estimation. I compute the *CSTR* and *SCSTR* measures by using raw returns.[3] Their evolution is shown in Fig. 1. The top solid line shows my *SCSTR* measure, and the bar in the bottom of the figure presents the relative difference between *SCSTR* and *CSTR*. *SCSTR* starts at a high level, which is possibly related to the Flash Crash of 1962, when the S&P 500 Index declined 22.5% from December 1961 until June 1962. However, during the most recent financial crisis of 2007–2009, there was no similar increase in the tail risk measure. This finding is consistent with that of Brownlees et al. (2011), who argue that the crisis was followed by inflated volatility that could be predicted by standard volatility forecasting models. Therefore, the effect of soaring volatility could have been captured by changes in the constant percentile threshold u_t in the tail risk estimation. While the distribution became wider, the structure of the tail was not different from that in the previous periods. On average, *SCSTR* is higher than *CSTR* due to the larger number of less extreme observations and the lower (in absolute value) average daily threshold. The difference is greater since the 2007–2009 financial crisis. In the previous period, the relative difference was, on average, 1%, with a maximum of 7%. After 2002, the average was 11%, with a maximum of 47%. This result is found since my measure is able to capture the frequency of market-wide extreme events during each month instead of concentrating on a few days; thus, it can better capture cross-sectional effects. To make this point clearer, I compute a measure of the concentration of days on the stocks chosen for the monthly tail estimate. The measure is embodied by the Herfindahl Index (HHI) as follows:

[2] The number of observations in the tails of the daily cross-section for the period before July 1962 is well below 50, which would clearly endanger the reliability of the tail risk estimates.

[3] In the next section, I show that when I run *SCSTR* on the residuals of a daily Fama-French model, the economic magnitude of the results remain.

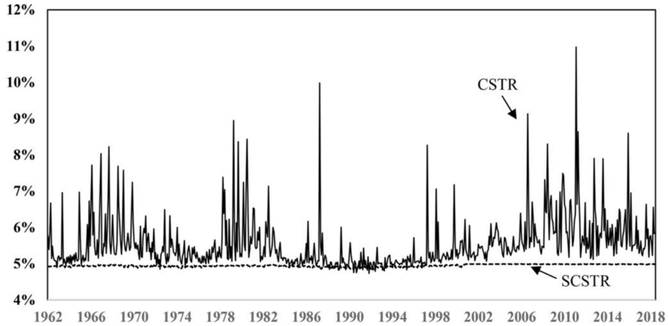

Fig. 2. Concentration on specific days of tail risk measures.
This figure reports the evolution of a sort of Herfindahl index as defined in Section 2 for *CSTR* (solid black line) and *SCSTR* (dashed line).

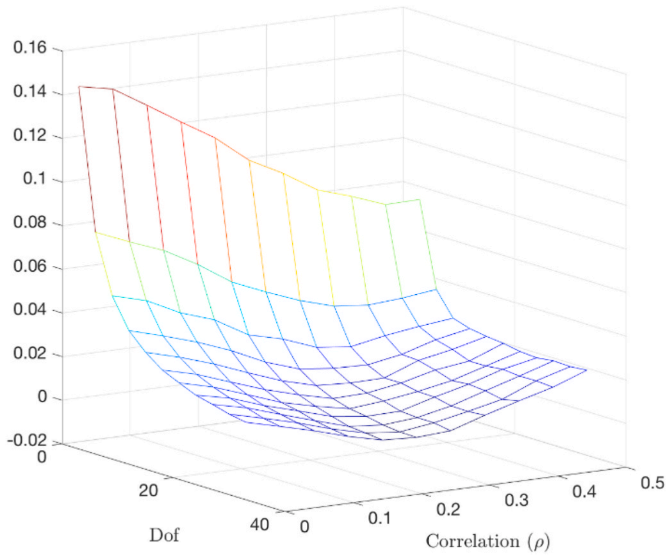

Fig. 3. Difference in probabilistic MSE between simulated daily *SCSTR* and *CSTR*.
This figure presents the difference in pMSE between *SCSTR* and *CSTR* when compared to the realized daily tail risk estimated from equation (1) for different values of degrees of freedom (Dof) and correlation within stocks (ρ). Degrees of freedom of the marginal and copula distributions are set to the same value and equal to Dof.

$$HHI_m = \sqrt{1 \bigg/ N_d \sum_{i=1}^{N_d} w_i^2}, \tag{5}$$

where w_i is the proportion of stocks below the 5% threshold out of the total of stocks available on day i of month m and N_d is the number of days in each month. Fig. 2 presents this index for the two measures of tail risk, *CSTR* and *SCSTR*. As expected, for *SCSTR*, it is flat at approximately 5%. Such a result is not observed for *CSTR*, since it may be driven by a few days in the month. It is clear from the figure that this series is quite volatile. It ranges between 5% and 11%. For most months (97.6%), the HHI index based on the *CSTR* exceeds that of the *SCSTR*. There is a surge in the HHI for *CSTR* after 2002. In fact, the average number of months with an HHI greater than 5.5% (6.0%) after 2002 was 66% (30%), whereas before this period, it was merely 24% (11%). This result is clearly related to the difference between the levels of *CSTR* and *SCSTR*.

2.3. Simulation study

In this subsection, I use numerical methods to explore the properties of my estimator *SCSTR* and compare them with those of the *CSTR* measure. First, I explain the methodology used to simulate the data. Then, I compare both measures in two settings: one derived from simulated daily returns and another from simulated monthly returns. The latter simulates the actual approach I take in this paper. Finally, I empirically analyze which estimator—the *SCSTR* or the *CSTR*—is closer to the empirical data.

For the simulated data, I start by using a t-copula to generate an n-dimensional multivariate distribution $X = (X_1, X_2, \ldots, X_n)$ and

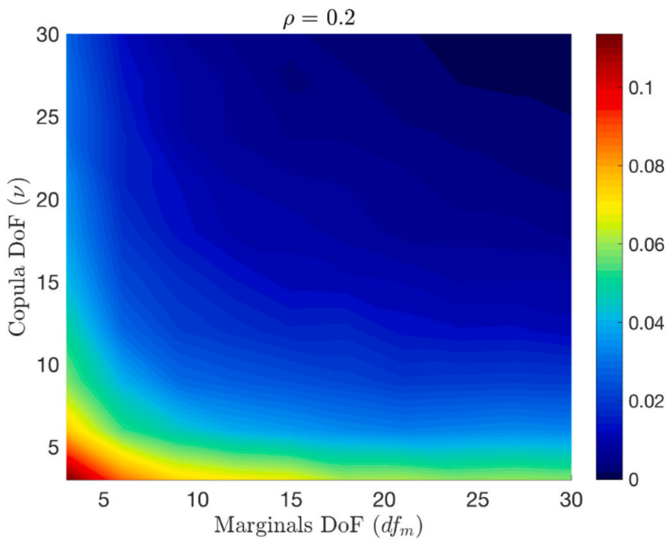

Fig. 4. Difference in probabilistic MSE between simulated daily *SCSTR* and *CSTR* conditional on average realized correlation.
This figure presents the difference in pMSE between *SCSTR* and *CSTR* when compared to the realized daily tail risk estimated from equation (1) but for a specific correlation within stocks value equal to 0.2. In contrast to Fig. 3, now I fix the correlation value to the average empirical correlation observed in the data and allow different degrees of freedom for the copula (υ) and for the marginals (df_m).

then reverse-transform the uniform marginals into *t*-distributions (*df* stands for degrees of freedom). The *t*-copula is defined by a pair with two parameters (R, ν). R is the correlation matrix of the n covariates. Following the equicorrelation method of Engle and Kelly (2012), R is defined by unique off-diagonal values ρ with 1 *s* on the diagonal. Consequently, each simulated dataset represents a single history of returns of n stocks over time horizon T. ν represents the degrees of freedom of the copula itself, which generates fat tails in the joint distribution, independent of the marginal distributions. The fat tails of the joint distribution are determined by the parameter space (df, ν, ρ).

For the simulation procedure, I set $n = 1000$ stocks and $T = 2000$ days (100 months). I obtain the simulated monthly returns from the same simulated daily returns. For the degrees of freedom, df and ν are set in the range of 3–30; note that as these values increase, the distributions and the copula approach Gaussian form. For the correlation ρ, I choose values between 0 and 0.5. I report in the figures the average of the computed values for 100 iterations. Next, I use simulated data for comparison.

In my first exercise, I compare the predictive performance using daily returns. For each month, I estimate the common tail risk using both methods to compute $\widehat{\lambda}_i$ ($i = CSTR, SCSTR$). Then, I feed this estimated tail risk back into equation (1) to compute the predictive probabilities of the 5th-percentile returns. I assume that stocks are driven by the same process, implying that in equation (1), $a_i = a = 1$ for all stocks. Using these predictive probabilities, I can construct the estimated cumulative distribution function (cdf) of the 5th-percentile tail distribution. High predictive performance will minimize the difference between the estimated cdf and the empirical cdf of the realized tail distribution of the simulated data. I capture the differences in probabilities by the probabilistic mean squared error (pMSE).

To study the effect of ρ, I set df and ν to the same degrees of freedom, which I designate as Dofs. Fig. 3 presents the difference in pMSE between the *SCSTR* and the *CSTR* for different parameter values. A positive (negative) value means that compared to the *CSTR*, the *SCSTR* achieves a closer (further) estimated cdf than the empirical cdf of the simulated data. Overall, the *SCSTR* outperforms the *CSTR* by an order of magnitude, on average, of approximately 0.04 across different values of Dofs. Thus, on average, the *SCSTR* estimates are 0.04 closer to the realized probabilities. The difference quickly decreases as the Dofs increase. When ρ is approximately 0, the difference is minimal.

I further investigate the effects of Dofs in Fig. 4. I fix the correlation value to the average empirical correlation observed in the data, $\rho = 2$, and I analyze the effect of different values of df and ν with a heatmap. The difference is only persistent at the 10% significance level for an extreme fat tail distribution (both Dof $\leq$5), which quickly vanishes as Dofs increase in either direction. In the Online Appendix, I show that the *SCSTR* outperforms when negative returns are realized across different days, and these returns are strongly negative but mostly still in the range of 2–4 standard deviations.[4]

[4] For the return patterns that drive the difference in predictive performance, I set $\rho = 0.2$ and $df = \nu = 4$ to induce strong fat tails. In the Online Appendix, I show the simulated daily returns in a month, flattened out in 20 sequential blocks of 1000 observations per day for two anecdotal cases. In the first case, I use a simulated month in which there is the largest difference in pMSE between the *CSTR* and the *SCSTR*. In this case, the pMSE of the *CSTR* (*SCSTR*) is 0.056 (0.271). The *SCSTR* figure resembles a "black swan," i.e., concentrated extreme negative returns that are more than 5 standard deviations away from the mean. By contrast, I also use the case in which the difference is the smallest. In this case, the pMSE of the *CSTR* (*SCSTR*) is 0.053 (0.047).

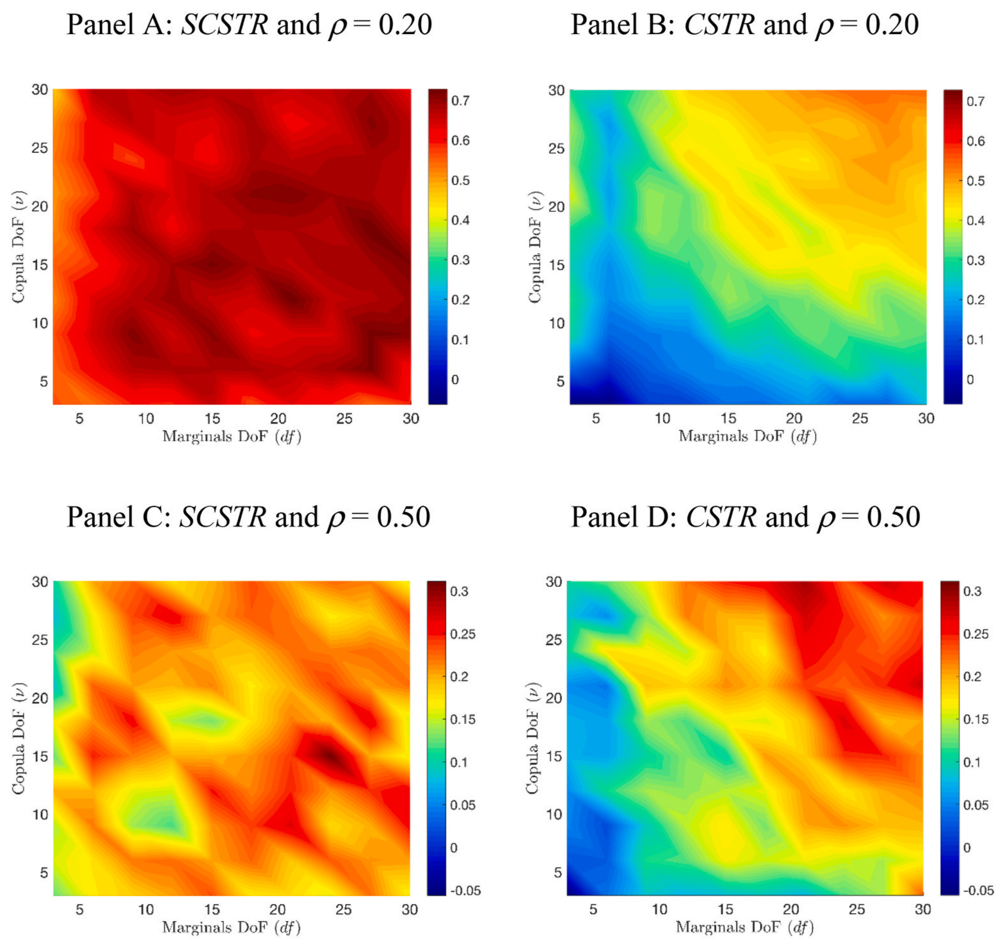

Fig. 5. Correlation of the simulated *SCSTR* and *CSTR* distributions with the monthly empirical tail distribution.
This figure presents the correlation of the simulated *SCSTR* and *CSTR* with the realized monthly tail distribution for different values of copula degrees of freedom (υ) and marginal degrees of freedom (df). Panels A and B (Panels C and D) present the case when the within stock correlation is 0.20 (0.50).

Until now, I have analyzed the performance of the *SCSTR* and the *CSTR* in fitting the realized daily tail risk distribution. In a second exercise, I run a similar analysis on the monthly tail return empirical distribution. Specifically, I compute monthly cumulative returns for each month (from the same simulated daily returns processes) and apply the Hill estimator to estimate the tail risk at the monthly level. This estimate will serve as the benchmark for the monthly tail risk. However, because daily returns behave differently from monthly cumulative returns, it is not possible to directly compare predictive performance as in the previous analysis. Thus, I measure the correlation of the two series $\widehat{\lambda}_i$ against the monthly estimates.

Fig. 5 presents the correlation of the empirical monthly tail distribution for *SCSTR* and the *CSTR*. Panels A and B (Panels C and D) show the case when the correlation within stocks is 0.20 (0.50). I again use heatmaps. A warmer (colder) color means a higher (lower) correlation. The *SCSTR* is more strongly correlated with the empirical monthly tail risk. When ρ is 0.20, the correlation with the *SCSTR* is high, between 0.45 and 0.80. In the case of the *CSTR*, the correlation is significantly lower with a maximum of 0.50 but only when Dofs are quite large. In some cases, this correlation is even negative when both Dofs define an extreme fat tail region. When ρ is 0.50, this difference is less pronounced, and both correlations are also weaker (maximum of 0.30). In the normal distribution parameter space, the correlation is similar between the two measures. However, in the region of the extreme fat tail, the *SCSTR* is still more correlated with the empirical monthly tail distribution.

In summary, the *SCSTR* and the *CSTR* are both based on the same Hill estimator. The main difference is the implementation with respect to the data used, resulting in two different types of behavior. The *CSTR* pools all daily observations within a month and then applies the Hill estimator, with the main assumption that all stocks share common tail-risk dynamics. Thus, the *CSTR* measure is more likely to be driven by a few days that have very large negative returns. In practice, these extreme returns are concentrated in only one day or a few days. In contrast, the *SCSTR* measure applies the Hill estimator repeatedly for all days in a month and then averages them to estimate the month's tail risk. Thus, this measure is more likely to capture clusters of negative days, in which 5% of the stocks are likely to have negative returns larger than two standard deviations. The resulting difference is that the *CSTR* will better capture tail risk

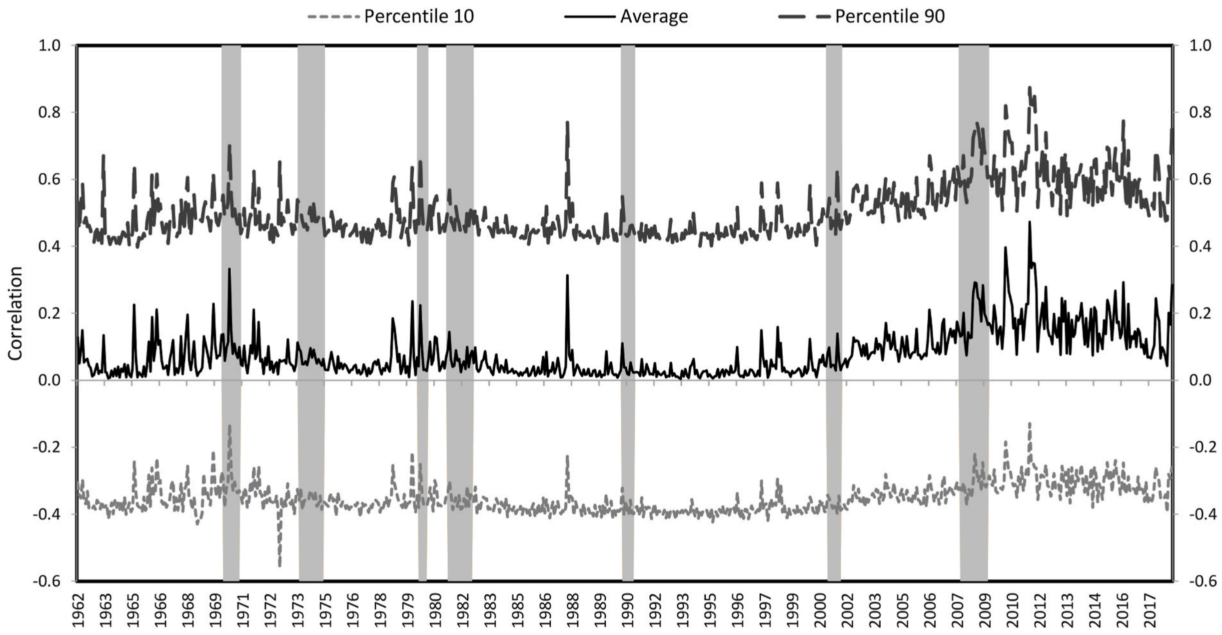

Fig. 6. Evolution of stock correlations.
This figure presents the average stock return correlations (in solid black), as well as the 10th and 90th percentiles of monthly stock correlations (in dashed gray) for the stocks in my sample for the period from 1964 to 2018. Each pairwise correlation is computed with daily returns within the month. The gray bars denote the NBER-defined U.S. recessions.

generated by isolated extreme days, while the *SCSTR* will better capture overall monthly tail risk, in which there are more likely clusters of days with strong but not very deep negative returns.

The next step is to empirically understand the dependence assumption between stocks. Within each month, I compute the Pearson correlation between the daily returns of any given stock *i* and all other stocks in the sample, excluding pairwise repetitions.[5] I then compute the average monthly correlation, as well as the 10th and 90th percentiles of these correlations. The results are presented in Fig. 6. The average stock return correlation is approximately 0.20, with a slight upward increasing trend and mean reverting behavior around this value. Overall, 80% of the monthly correlations are between −0.20 and 0.50. Since the empirical results suggest that the average pairwise correlation coefficient of the U.S. equity market is close to 0.20 and that 90% fall below 0.50, the outcome of the simulations supports my previous findings that the *SCSTR* is a more robust monthly stock market tail risk proxy.

3. Equity premium prediction

In this section, I focus on the predictability of the U.S. equity premium by the *SCSTR*. I test whether *SCSTR* strongly predicts the stock market at different horizons and compare it with traditional predictors and *CSTR*. I run several robustness tests.

3.1. Equity risk premium and traditional predictors

As an independent variable, I use the CRSP Value-Weighted Index return in excess of the short-term risk-free rate extracted from Kenneth French's Data Library.[6] I evaluate the predictive power of tail risk and compare it to the historical average and other commonly used equity premium predictors. I use variables related to stock market characteristics, interest rates, and broad macroeconomic indicators, including the following:

Book-to-Market (*B/M*): The book-to-market ratio of the Dow Jones Industrial Average (Kothari and Shanken, 1997; Pontiff and Schall, 1998).

Dividend-Price Ratio (*D/P*): The difference between the logarithm of the 12-month moving sum of dividends paid on the S&P 500 Index and the logarithm of prices.

Dividend Yield (*D/Y*): The difference between the logarithm of the 12-month moving sum of dividends paid on the S&P 500 Index and the logarithm of lagged prices (Fama and French, 1988, 1989; Lewellen, 2004).

Dividend Payout Ratio (*D/E*): The difference between the logarithm of the 12-month moving sum of dividends paid on the S&P

[5] I exclude firms that, in a given month, have fewer than 15 valid daily returns.
[6] The data library is available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. I also checked the results using the S&P 500 Index as a market benchmark and obtain similar qualitative and quantitative results. These results are available upon request.

Table 1
Descriptive statistics of predictors.

	Mean	Std.Dev	Max	Min	AR (1)
SCSTR	0.45	0.05	0.57	0.29	0.84
CSTR	0.43	0.05	0.57	0.27	0.78
B/M	0.49	0.26	1.21	0.12	0.99
D/P	−3.61	0.40	−2.75	−4.52	0.99
D/Y	−3.60	0.40	−2.75	−4.53	0.99
DFS	1.02	0.44	0.34	0.32	0.97
TMS	1.83	1.47	4.55	−3.65	0.96
D/E	−0.76	0.31	1.37	−1.24	0.99
E/P	−2.84	0.43	−1.9	−4.84	0.99
SVAR	0.21	0.44	7.09	0.01	0.47
NTIS	0.97	2.00	5.11	−5.76	0.98
INFL	0.31	0.36	1.81	−1.91	0.62
LTY	6.49	2.72	14.82	1.75	0.99
TBILL	4.66	3.21	1.63	0.01	0.99
LTR	0.60	2.95	15.23	−11.24	0.03

This table reports the descriptive statistics (mean, standard deviation, maximum, and minimum) of the predictors for the period between July 1962 and December 2018. The definition of the predictors is in Section 3. I also report the autocorrelation coefficient of order 1 for each of the variables.

500 Index and the logarithm of the 12-month moving sum of earnings on the S&P 500 Index (Lamont, 1998).

Earnings-Price Ratio (*E/P*): The difference between the logarithm of the 12-month moving sum of earnings on the S&P 500 Index and the logarithm of prices (Campbell and Shiller, 1988, 1998).

Long-term Yield and Rate (*LTY, LTR*): Long-term U.S. government bond yield and rate (Fama and Schwert, 1977; Campbell, 1987; Ang and Bekaert, 2007).

Default Yield and Rate Spread (*DFS*): The difference between BAA- and AAA-rated corporate bond yields (Keim and Stambaugh, 1986; Campbell, 1987; Fama and French, 1989).

Term Spread (*TMS*): The difference between long-term U.S. government bond yield and the Treasury bill rate (Fama and French, 1989).

Realized Stock Variance (*SVAR*): The sum of squared daily returns on the S&P 500 Index during a month (Guo, 2006).

Net Equity Expansion (*NTIS*): The ratio of 12-month moving sums of net equity issues by NYSE-listed stocks to the total end-of-year market capitalization of NYSE stocks (Baker and Wurgler, 2000; Boudoukh et al., 2007).

Inflation (*INFL*): The lagged Consumer Price Index provided by the Bureau of Labor Statistics (Lintner, 1975; Nelson, 1976).

T-Bill Rate (*TBILL*): The 3-month Treasury Bill rate of the secondary market from the economic research database at the Federal Reserve Bank at St. Louis (FRED) (Campbell, 1987; Hodrick, 1992).[7]

I consider the same time frame for all predictions for ease of comparison and consistency. The cross-section width requirement for the tail risk estimation implies that the predictive regressions only start in July 1962 and end in December 2018. I present the descriptive statistics for all predictors in Table 1. I also evaluate how the predictors relate to each other by computing the correlation matrix of the predictors and present the results in Table 2. I find that *SCSTR* has a relatively low correlation with most of the other variables, yet it is strongly positively correlated with *CSTR*. The correlation is moderate with *LTY*, *NTIS, BM, TBILL*, and *TMS*, and there is a different pattern of correlation with other variables between *SCSTR* and *CSTR*. For example, *SCSTR* moderately correlates with *TBILL*, but *CSTR* has a much lower correlation. The diagonal part of the matrix shows the serial correlation coefficients of order one. Most variables are strongly persistent; however, many of them use overlapping data, which implies that these serial correlations are artificially inflated. The exceptions are *SCSTR*, *CSTR*, *DFS*, *TMS*, *SVAR*, *INFL, LTR*, and *LTY*.

3.2. Standard predictability results

I apply the widely used methodology of comparing the sum of squared errors (SSE) of the prediction with the SSE of the sample equity risk premium average (e.g.,Campbell and Thompson 2008; Goyal and Welch, 2008; Rapach et al., 2010; Ferreira and Santa-Clara, 2011; Li et al., 2015; Faias and Castel-Branco, 2018).

First, I obtain in-sample (IS) results. I run a predictive regression for the entire sample of as follows:

$$ERP_t = \alpha + \beta x_{t-1} + \varepsilon_t, \tag{6}$$

where x_{t-1} is the predictor at time *t* - 1 of the equity risk premium ERP_t at time *t*. I estimate the parameters in equation (6) using ordinary least squares (OLS). Then, I compute the R^2 of this regression:

[7] The predictors are obtained from Amit Goyal's website: http://www.hec.unil.ch/agoyal. Apart from the variables used in this paper, there are others that I could have used: cross-sectional beta premium (Polk et al., 2006), investment-to-capital ratio (Cochrane, 1991), and consumption-wealth ratio (Lettau and Ludvigson, 2005). However, there is evidence that all of them fail OOS (Goyal and Welch, 2008). Due to data unavailability, I exclude these variables from the equity premium predictors.

Table 2
Correlation matrix of predictors.

	SCSTR	CSTR	B/M	D/P	D/Y	DFS	TMS	D/E	E/P	SVAR	NTIS	INFL	LTY	TBILL	LTR
SCSTR	0.86*	0.93*	0.04	0.13*	0.16*	0.16*	0.33*	−0.08*	0.18*	−0.10*	−0.33*	−0.09*	−0.19*	0.01	0.11*
CSTR		0.81*	0.07	0.17*	0.21*	0.06	0.25*	−0.04	0.19*	−0.21*	−0.21*	−0.05	0.30*	0.14*	0.08*
B/M			0.99*	0.90*	0.90*	0.40*	−0.28*	0.04	0.81*	−0.07	0.25*	0.47*	0.65*	0.68*	0.02
D/P				0.99*	0.99*	0.39*	−0.19*	0.28*	0.73*	−0.04	0.16*	0.37*	0.68*	0.66*	0.04
D/Y					0.99*	0.39*	−0.19*	0.28*	0.72*	−0.07	0.16*	0.36*	0.68*	0.66*	0.04
DFS						0.97*	0.22*	0.32*	0.13*	0.32*	−0.28*	0.04	0.42*	0.25*	0.12*
TMS							0.96*	0.21*	−0.33*	0.12*	−0.07	−0.29*	−0.09*	−0.53*	−0.01
D/E								0.99*	−0.46*	0.19*	0.00	−0.11*	0.03	−0.07	0.01
E/P									0.99*	−0.17*	0.15*	0.42	0.60*	0.66*	0.03
SVAR										0.47*	−0.20*	−0.18*	−0.03	−0.08*	0.14*
NTIS											0.98*	0.18	0.19*	0.20*	−0.05
INFL												0.58*	0.39*	0.46*	−0.16*
LTY													0.99*	0.89*	0.02
TBILL														0.99*	0.02
LTR															0.04

This table presents the Pearson correlation coefficients of predictors (off-diagonal elements) and the one-month serial correlation (diagonal underlined elements) from July 1962 until December 2018. The definition of the predictors is in Section 3. * Significant at the 5% level.

Table 3
Predictability results.

	1M			3M			1Y			3Y		
	R2	β	Wald	R2	β	Wald	R2	β	Wald	R2	β	Wald
Panel A: In-Sample Results												
SCSTR	1.11***	0.09	6.76	1.68**	0.20	3.19	7.84**	0.89	5.36	22.36**	2.29	4.64
CSTR	0.76**	0.07	4.67	0.73	0.13	1.73	4.62*	0.65	3.45	14.11*	1.72	3.26
B/M	0.01	0.00	0.08	0.07	0.01	0.19	0.33	0.04	0.29	0.01	−0.01	0.00
D/P	0.17	0.01	1.11	0.55	0.01	1.34	2.04	0.06	1.46	3.39	0.11	0.67
D/Y	0.23	0.00	1.58	0.59	0.01	1.46	2.1	0.06	1.53	3.22	0.11	0.67
DFS	0.19	0.43	0.74	0.58	1.37	0.87	2.23	5.36	0.92	2.73	9.03	0.21
TMS	0.55*	0.22	3.36	1.29*	0.62	2.91	6.45**	2.8	4.43	18.49**	7.18	5.32
D/E	0.06	0.00	0.54	0.37	0.02	1.07	1.28	0.06	1.33	2.6	0.13	1.73
E/P	0.04	0.00	0.21	0.06	0.00	0.11	0.25	0.02	0.14	0.28	0.03	0.03
SVAR	1.14***	−1.08	8.31	0.12	−0.64	0.66	0.54	2.72	0.87	0.73	4.82	0.64
NTIS	0.04	−0.04	0.11	0.03	−0.07	0.01	0.57	−0.61	0.14	0.64	−1.03	0.00
INFL	0.52*	−0.89	3.76	0.65	−1.8	2.56	2.97**	−7.82	5.33	1.4	−8.21	1.02
LTY	0.14	−0.06	1.34	0.25	−0.15	0.93	0.11	−0.2	0.26	0.15	0.36	0.03
TBILL	0.43*	−0.09	3.22	0.90	−0.23	2.42	2.08	−0.73	1.63	2.93	−1.36	0.84
LTR	1.03**	0.15	6.44	0.59*	0.21	3.59	1.14**	0.59	6.09	0.65	0.67	1.19
	1M			3M			1Y			3Y		
	R2	β	Wald	R2	β	Wald	R2	β	Wald	R2	β	Wald
Panel B: Out-of-Sample Results												
SCSTR	1.26	0.09	2.45	1.95*	0.20	3.75	9.13***	0.79	6.67	21.51**	2.06	5.57
CSTR	0.68	0.08	0.26	0.59	0.16	0.64	4.40*	0.71	2.84	11.19	1.91	2.25
B/M	−0.40	0.00	0.43	−0.41	0.02	0.31	−2.80	0.09	0.00	−3.99	0.09	0.11
D/P	−0.46	0.01	1.51	−0.33	0.03	1.21	−4.21	0.14	0.63	−4.26	0.26	0.32
D/Y	−0.44	0.01	1.25	−0.28	0.03	1.30	−4.16	0.14	0.55	−3.83	0.25	0.25
DFS	−0.23	0.82	2.15	0.17*	2.13	3.22	2.23*	5.18	3.31	3.22***	6.28	8.34
TMS	−0.29	0.32	1.16	−0.02	0.84	1.19	6.78*	2.81	3.65	24.37***	4.47	7.90
D/E	−1.21	0.01	0.25	−1.26	0.03	0.48	0.40	0.07	0.62	1.98	0.00	1.41
E/P	−0.50	0.00	0.65	−0.53	0.01	0.26	−1.87	0.06	0.04	−5.46	0.14	0.00
SVAR	−1.65	−0.46	0.58	−13.98	2.07	0.18	−14.12	5.42	1.45	−27.03	8.64	0.74
NTIS	−0.65	−0.21	0.01	−1.21	−0.57	0.03	−12.88	−2.45	0.00	−0.79	−2.07	0.12
INFL	−0.27	−1.64	0.15	−0.34	−3.28	0.93	3.76**	−7.21	5.36	1.45	−4.07	2.47
LTY	−0.73	−0.06	1.31	−0.91	−0.09	0.94	−0.16	0.34	0.28	−0.87	2.20	4.80
TBILL	−0.78	−0.13	0.01	−0.88	−0.31	0.01	2.15	−0.49	1.01	2.99	0.61	1.85
LTR	−0.01	0.19	2.80	0.38	0.31	1.63	0.81***	0.82	16.07	0.53	0.67	2.02

This table presents the in- and out-of-sample R^2 values (in %) of the equity premium predictions for 1-month, 3-month, 1-year and 3-year horizons. I also report the slope of each standardized variable in the predictive regression (average slopes in the case of the out-of-sample results). The definition of the predictors can be found in Section 3. I use the Wald test proposed by Kostakis et al. (2015), which considers the persistence of the regressors to test the null hypothesis of no predictability for the in-sample results. The test of equal forecast ability of Clark and West (2007) is used for the out-of-sample results. I also apply Hodrick (1992) standard error correction for overlapping data using 36 lags for the 3-year forecast horizn and 12 lags for other horizons. The sample period is between July 1962 and December 2018. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

$$R_{IS}^2 = 1 - \frac{\sum_{t=2}^{T}(ERP_t - \widehat{ERP}_t)^2}{\sum_{t=2}^{T}(ERP_t - \overline{ERP}_t)^2}, \tag{7}$$

where T is the size of the sample, $\widehat{ERP}_t$ is the prediction value from equation (6), and $\overline{ERP}_t$ is the sample average of the risk premium. If the R^2 is positive, then the predictor forecasts the value of the equity risk premium better than the historical risk premium average. The higher the R^2, the better the quality of the forecast.

Panel A of Table 3 presents the IS results. I conclude that the *SCSTR* better predicts the equity risk premium for all time horizons (but 3 years for *TMS*) than the other commonly-used predictors. In addition, only *SCSTR* and *TMS* are significant in all time horizons. I also compute the slopes of these regressions for the standardized variables for ease of comparison. The slopes are quite close in value between these predictors. Note that *CSTR* loses against *TMS* for all horizons greater than one month, and that *SCSTR* consistently outperforms *CSTR* and achieves a higher relative incremental R^2, ranging from 38% (for 1 month) to 130% (for 3 months). For all predictors, the IS R^2 is always positive, and with several exceptions, it increases with the forecast horizon.

Next, I evaluate the OOS predictive power, which is closer to real-time forecasting. To predict the value of the risk premium OOS at time $t+1$, I use only the data available until time t instead of the entire sample using the OLS estimates from equation (6). Hence, the regression is re-estimated before every prediction. The OOS R^2 is given by:

$$R_{OOS}^2 = 1 - \frac{\sum_{t=m+1}^{T}(ERP_t - \widehat{ERP}_t)^2}{\sum_{t=m+1}^{T}(ERP_t - \overline{ERP}_t)^2}. \tag{8}$$

Table 4
Out-of-sample results after slope and forecast restrictions.

	1M	3M	1Y	3Y
SCSTR	1,25**	1,95**	9.13***	21,51***
CSTR	0,68	0,59	4,40*	11,19*
B/M	−0,43	−0,44	−2,86	−3,98
D/P	−0,23	−0,33	−4,21	−4,32
D/Y	−0,29	−0,28	−4,16	−3,94
DFS	−0,23	−0,17	2,23	3,22
TMS	−0,29	−0,02	6,78**	24,44***
D/E	−0,34	−0,25	0,41	1,98
E/P	−0,42	−0,53	−1,87	−5,46
SVAR	−2,28	−13,95	−14,12	−27,03
NTIS	−0,34	−0,97	−10,57	−0,79
INFL	−0,44	−0,52	3,65***	1,45
LTY	−0,05	−0,01	−0,16	−0,60
TBILL	−0,29	−0,31	2,40	2,99
LTR	0,58	0,67	2,50**	0,53***

This table presents the out-of-sample R^2 values (in %) of the equity premium predictions for 1-month, 3-month, 1-year and 3-year horizons. The definition of the predictors can be found in Section 3. All predictions use an expanding window with a starting length of 240 months. The test of equal forecast ability of Clark and West (2007) is used. I also apply Hodrick (1992) standard error correction for overlapping data using 36 lags for the three-year forecast horizon and 12 lags for other horizons. The sample period is between July 1962 and December 2018. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

For the OOS forecast, I require *m* periods for the initial estimation period for the initial prediction and then either roll over the estimation period (rolling window) or expand it for the subsequent forecasts (recursive or expanding window) so that I obtain $q = T - m$ OOS observations. I use an expanding window, as in Kelly and Jiang (2014), with an initial estimation period of 20 years.[8] The large estimation period is essential to obtain reliable regression coefficients at the beginning of the evaluation period and follows Goyal and Welch (2008). In addition, the number of OOS predictions should be sufficiently long to be representative. To test the statistical significance of the IS and OOS predictions, I could use the Clark and West (2007) test of equal forecast ability. This test helps identify whether the mean squared percentage error (MSPE) of prediction is significantly lower than the MSPE of the historical equity risk premium average. In practice, this test is identical to testing the null hypothesis of $R^2 \leq 0$ against the alternative hypothesis of $R^2 > 0$. I apply Hodrick's (1992) standard error correction for overlapping data by using 36 lags for the three-year forecast horizon and 12 lags for other horizons.[9] However, the Clark and West (2007) test may be invalid since the regressors are highly persistent. Thus, I adopt Kostakis et al.'s (2015) extended instrumental variable (IVX) instrumentation procedure. The key idea behind the IVX method is to create an instrument that relies solely on the regressor (hence the name IVX) but falls within the class of near-stationary processes and uses it to remove endogeneity. In addition, it presents good finite-sample properties. I present the later test results in Panel A of Table 3 for IS analysis and Clark and West (2007) for the OOS analysis.[10]

Panel B of Table 3 exhibits the OOS R^2s. The slopes of the IS and OOS predictive regressions using *SCSTR* are all positive and significant. The IS slopes for the 1-month, 3-month, 1-year, and 3-year horizons are 0.09, 0.20, 0.89, and 2.29, respectively. The OOS slopes are quite stable and range between 0.08 and 0.14 for a 1-month horizon, 0.16 to 0.24 for a 3-month horizon, 0.66 to 1.01 for a 1-year horizon, and 1.64 to 2.53 for a 3-year horizon. In Table 1 report the average slopes. These results show that *SCSTR* is positively and strongly related to equity return premiums at different horizons. As expected, most of the predictors exhibit a significant reduction in R^2 and a loss of significance in comparison with the IS results. The only variable with positive and significant results for all horizons (but 1-month horizon) is *SCSTR*. I observe that the only positive predictors for a one1-month horizon are *SCSTR* and *CSTR*, but the former with a value of R^2 that is approximately twice the one from the latter. In this timespan, *CSTR* has no predictability for any of the selected horizons at a 5% significance level. My results also support the view that valuation ratios lose their predictive power over time. From analyzing the cumulative SSE difference, I conclude that *NTIS*, *INFL*, and *DFS* perform well until the recent financial crisis but then suffer from a large shock in the second half of 2008. This result indicates that during periods of unexpected extreme downturns, these predictors are unable to outperform the historical equity risk premium mean. I reach the same conclusion as Ang and Bekaert (2007) that term spread is a robust predictor for long-term horizons.

[8] To ensure the robustness of the results, I also evaluate OOS predictions using 120- and 360-month estimation periods with both rolling and expanding windows. All results are qualitatively similar, with minor exceptions, and some are presented later in this section.

[9] Richardson and Smith (1991) argue that overlapping return observations produce a moving average structure in the errors of the forecast and hence jeopardize the reliability of the tests based on OLS and even Newey and West (1987) standard errors. According to Ang and Bekaert (2007), Hodrick's (1992) standard error correction yields the most conservative test results.

[10] I report the IS results in Panel A of Table 3 with Kostakis et al. (2015) procedure and OOS results in Panel B with Clark and West (2007). Nevertheless, it does not affect my conclusions and I present the results of a traditional Clark and West (2007) test of equal forecast ability in the remaining tables for both IS and OOS for ease of comparison with previous studies.

Table 5
Bivariate predictive regression: just tail measures.

	1M			3M			1Y			3Y		
	R2	β1	β2	R2	β1	β2	R2	β1	β2	R2	β1	β2
		SCSTR	*CSTR*		*SCSTR*	*CSTR*		*SCSTR*	*CSTR*		*SCSTR*	*CSTR*
In-Sample	1.20***	0.16*	−0.07	2.61***	0.58***	−0.38**	9.42***	1.91***	−1.04***	25.84***	4.68***	−2.42***
Out-of-Sample	0.78**	0.18	−0.09	2.84***	0.54**	−0.35	11.04***	1.38***	−0.60	25.72***	2.74***	−0.70

This table presents the in- and out-of-sample R^2 values (in %) and corresponding slope coefficients of a bivariate regression using simultaneously *SCSTR* and *CSTR* on market premium predictions for 1-month, 3-month, 1-year and 3-year horizons. The sample period is between July 1962 and December 2018. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

3.3. Imposing theoretical restrictions on predictive regressions

Despite the low predictive power of commonly-used predictors in the baseline OOS evaluation, it is premature to conclude that *CSTR* outperforms other variables. Running predictive regressions with unexpected shocks during the initial estimation period might generate unexpected estimates. For instance, such regressions can yield slope coefficients that are inconsistent with financial theory or yield a negative risk premium forecast. To minimize the effect of such estimation errors, I further set simple regression restrictions as in Campbell and Thompson (2008).

I suggest that investors would rather apply their knowledge of financial theory than simply use a mechanical rule such as applying the pure predictive regression framework. First, I assume that they would not use perverse coefficients of a regression if they contradict common theory. Every time the slope sign is different from the theoretically expected regression coefficient over the entire sample, I set the slope equal to zero. For example, I expect an increase in *D/P* to be followed by an increase in prices and, therefore, positive returns. The same intuition should be applied to tail risk, stock variance or interest rate spreads given the risk-averse behavior of investors. Second, I assume that equity risk premium forecasts should be positive, as market returns are subject to systematic risk, which should be compensated with a positive premium. Campbell and Thompson (2008) propose setting ERP to zero when the prediction is negative. As historical equity premiums also contain useful information, I suggest using the historical risk premium average instead of zero.

Table 4 exhibits the OOS results after I apply the regression restrictions. After imposing slope and forecast restrictions on the predictive regressions, I obtain mixed results. The forecast accuracy of the valuation ratios (*D/P*, *D/Y*, *D/E*, and *E/P*) and long-term yield increased for short-term forecast horizons. For most predictors, the OOS R^2 decreased for the 1- and 3-year forecasts. These results imply that the statistical methodology is more powerful than imposing theoretical restrictions in such cases. Several variables are almost unaffected by restrictions or exhibit small, not significant changes in predictive power (*SCSTR*, *CSTR*, *B/M*, *TMS*, *DFS*, *NTIS*, and *SVAR*). The only variable that is significantly negatively affected by restrictions is *LTY* for 1- and 3-year horizons and *TBILL* for a 3-year horizon.

SCSTR remains the most robust equity risk premium predictor and is significant for all forecast horizons. None of the predictors yields a comparable R^2 value for horizons up to 1 year. As before, only *TMS* yields better predictive ability for the 3-year horizon, but it lacks predictive power for short-term horizons.

3.4. Bivariate regressions

One fundamental question is the power of each of the tail measures – *SCSTR* and *CSTR* – against each other. Do these two variables convey the same information for investors? Does *CSTR* subsume the information of *SCRST* and the later just contains marginal non-relevant information? I run *SCSTR* and *CSTR* simultaneously and present the results in Table 5. I find that the slope coefficient for *SCSTR* is positive (always significant at least for a 10% significance level), whereas that the slope coefficient for *CSTR* is negative. This shows that *SCSTR* is the variable driving predictability with the correct sign, whereas after controlling for *SCSTR*, *CSTR* obtains a negative sign. This result reinforces the importance of considering *SCSTR* in predicting equity risk premiums.

3.5. Combination of individual predictors

The next step is to study the improvement in predictive power by combining the forecasts of different variables and to determine whether any of the combinations outperforms *SCSTR* on its own.

The intuition behind the benefit of combining different predictions is straightforward and was implemented in Rapach et al. (2010). Every predictor carries some set of useful information for forecasting future excess returns. The combination of forecasts aims to capture all the partitioned information into a single aggregated forecast that potentially has higher predictive power. Timmermann (2006), Rapach et al. (2010), and Elliott and Drive (2011) argue that simple combined methods commonly perform better than more sophisticated approaches. I follow this approach, as errors introduced by estimation of the combination weights might overpower any gains from setting the weights closer to their optimal values. The general combination of forecasts takes the following form:

$$ERP_{comb,t+1} = \sum_{i=1}^{N} \omega_{i,t} ERP_{i,t+1}, \tag{9}$$

where N is the number of individual predictions to combine, and $\omega_{i,t}$ is the ex ante combining weight of the ith individual prediction defined at time t and is ranked by prediction size merely for the ease of nomenclature.

I suggest five simple combination methods. The first, combination (1), is the average of all individual forecasts. Combination (2) is a trimmed mean that assigns zero weights to the largest and the smallest predictions. Combination (3) is a trimmed mean that assigns zero weights to the three largest and three smallest predictions. Combination (4) is the median of all individual predictions. Combination (5) is the discounted mean squared prediction error (DMSPE). It is a function of the historical forecasting performance of the individual models over the holdout OOS period and is given by:

$$\omega_{i,t} = \varphi_{i,t}^{-1} \Big/ \sum_{j=1}^{N} \varphi_{j,t}^{-1}, \tag{10}$$

where:

Table 6
Out-of-sample results: combining predictors.

	1M	3M	1Y	3Y
Panel A: Excluding *SCSTR*				
(1) Average	0.14	0.98	3.47*	5.11
(2) Trimmed Average 1	0.20	0.84	2.61	3.12
(3) Trimmed Average 3	0.30	0.88	1.91	2.11
(4) Median	0.42	0.86	1.70	1.38
(5) DMSPE θ = 0.90	0.16	1.36*	6.38***	12.26***
	1M	3M	1Y	3Y
Panel B: Including *SCSTR*				
(1) Average	0.19	1.08	4.10**	6.79*
(2) Trimmed Average 1	0.27	0.99	3.31**	4.98
(3) Trimmed Average 3	0.36	0.99	2.58*	3.67
(4) Median	0.45	0.95	2.16*	2.45
(5) DMSPE θ = 0.90	0.21	1.44*	6.96***	14.15***

This table presents the out-of-sample R^2 values (in %) of the equity premium predictions for 1-month, 3-month, 1-year and 3-year horizons. Panel A reports the results of combining all 13 predictors excluding the tail risk measures, and Panel B reports the results of combining all 13 predictors including *SCSTR* (see Section 3). Combination (1) is a simple average of predictions. Combination (2) is an average of all predictions excluding the highest and lowest values. Combination (3) excludes the three largest and three smallest predictions from the combination and averages the rest. Combination (4) is the median of all predictions. Combination (5) is the DMSPE method, which assigns greater weights to individual predictive regression model forecasts that have lower mean squared predictive error values (better forecasting performance) over the holdout out-of-sample period and is further explained in Section 4.2. All predictions are made using an expanding window of 240 months. The test of equal forecast ability of Clark and West (2007) is used. I also apply Hodrick (1992) standard error correction for overlapping data using 36 lags for the 3-year forecast horizon and 12 lags for other horizons. The sample period is between July 1962 and December 2018. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

$$\varphi_{i,t} = \sum_{s=m}^{t-1} \theta^{t-1-s} \left(r_{s+1} - \widehat{r}_{i,s+1} \right)^2 \tag{11}$$

and θ is a discount factor that I set to 0.90.[11] The DMSPE method assigns greater weights to individual predictive regression model forecasts that have lower mean squared predictive error values (better forecasting performance) over the holdout OOS period. For further details on the DMSPE method, see Rapach et al. (2010).

Table 6 presents the results of the forecast combinations. Panel A shows the OOS R^2 when I combine across all variables excluding the tail measures, and in Panel B I consider the same variables as in Panel A and add *SCSTR* to the regression. First, adding *SCSTR* yields better or similar results relative to the case of excluding it. Second, combining *SCSTR* with other predictors yields a lower R^2 than individual forecasts by *SCSTR* and a few variables lose significance. For example, for a 1-year horizon, the best combination yields a 35% lower R^2 than the one from the individual forecast. The results indicate that while combining tail risk with other predictors, the estimation noise distorts the predictive power more than it benefits from a reduction in the volatility of the estimates.

The unreported correlation matrix of predictions and the individual OOS results suggest that it might be more beneficial not to include predictors with poor performance in the combinations. However, doing so creates an IS bias of choosing ex post good predictors for a combination. Instead, I suggest testing whether a model can provide useful information beyond that already contained in tail risk.[12] Consider an optimal forecast combination between two models i and j:

$$\widehat{ERP}^{*}_{t+1} = (1-\theta) ERP_{i,t+1} + \theta ERP_{j,t+1}, \tag{12}$$

where $0 \leq \theta \leq 1$ and $ERP_{k,t+1}$ is the forecast at time $t+1$ of the equity risk premium of predictor k at time t. If $\theta = 0$, then model i encompasses model j, which does not contain any additional useful information. Harvey et al. (1998) develop a test to check whether the forecast given by model i encompasses the forecast given by alternative model j or, in other words, if $\theta = 0$. In unreported results, I compute the p-values of these tests. The results indicate that none of the other individual models can provide additional useful information over the one provided by tail risk forecasts. Therefore, I should not expect any noticeable improvement even by using more complex forecast combination methodologies to reduce the weights of poor performing models. Moreover, the results of other variables provide evidence that combinations can be beneficial for other predictors.

[11] I also compute for θ equal to 0.99 and 1, and the results are similar.

[12] In unreported results, I check the results when considering jointly *SCSTR* and each one of the predictors, allocating a weight of 50% to each one. With no exception, the bivariate regression with *SCSTR* improves IS and OOS all the R^2 values of the univariate predictive regressions. For example, for a 1month forecast, the OOS R^2 improves from −0.13% for the predictive regression using *LTR* to 1.05% using the combination between *LTR* and *SCSTR* in the predictive regression.

Table 7
Out-of-sample results: business cycle.

	Contractions				Expansions			
	1M	*3M*	*1Y*	*3Y*	*1M*	*3M*	*1Y*	*3Y*
SCSTR	3.36**	3.20*	6.04	6.42	0.63**	1.41*	10.60***	22.80***
CSTR	3.53**	3.03*	3.9	−25.14	−0.17	−0.48	4.63	14.30**
B/M	0.74	1.62	5.62	−0.37	−0.74	−1.31	−6.85	−4.30
D/P	2.31**	5.54***	18.03***	17.77***	−1.28	−2.91	−14.69	−6.15
D/Y	3.40**	6.30***	18.64***	15.11***	−1.58	−3.17	−14.91	−5.45
DFS	−0.14	1.88	10.09	33.01**	−0.26	−0.58	−1.48	0.67
TMS	1.16	1.77	2.06	22.68	−0.72	−0.82	9.01***	24.51***
D/E	−2.13	1.17	12.18	45.16	−0.93	−2.33	−5.16	−1.72
E/P	−0.55	−1.36	−1.91	−41.71	−0.49	−0.16	−1.85	2.35
SVAR	3.41	−0.22	3.65	12.77***	−3.14	−20.04	−22.5	−30.44
NTIS	−2.45	−1.41	1.18	13.07	−0.11	−1.13	−19.51	−1.97
INFL	−6.46	−7.37	1.68	2.15	1.56***	2.76***	4.75***	1.39
LTY	−3.05	−3.23	1.92	−12.7	−0.5	0.11	−1.14	0.14
TBILL	−3.98	−5.13	−4.03	2.48	0.17	0.98	5.06**	3.03
LTR	4.62	−0.93	1.93	0.64	−1.39	0.95**	0.29*	0.52**

This table presents the out-of-sample R^2 values (in %) of the equity premium predictions for 1-month, 3-month, 1-year and 3-year horizons. The forecast period starts in July 1962 and ends in December 2018. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

3.6. Predictive power and the real economy

Thus far, the results provide evidence that tail risk is the most powerful individual predictor that outperforms a set of popular predictors commonly used in the academic literature. Moreover, different forecast combinations are also unable to outperform predictions made by tail risk. In this section, I link the predictive power with the state of the real economy. My intent is to test whether the high performance of the tail risk measure is driven by the business cycle.

To define the business cycle, I use the periods of expansions and contractions provided by the National Bureau of Economic Research (NBER).[13] According to NBER, during the OOS forecast period, troughs (peaks) of business cycles occur in November 1982 (July 1990), March 1991 (March 2001), November 2001 (December 2007), and June 2009. Consequently, the period comprises 45 months of recession and 348 months of expansion. Note that the long-term forecasting (1-year and 3-year horizons) for recession periods is affected by the reduced number of observations, and the results should be treated with caution.

Table 7 presents the OOS prediction results in contraction versus expansion periods. Most predictors yield higher predictive power during contractions versus expansions considering short horizon forecasts (except *D/E*, *E/P*, *NTIS*, *INFL*, *LTY*, *TBILL,* and *LTR*) and long horizon forecasts (except *SCSTR*, *CSTR*, *TMS, E/P*, *LTY*, TBILL, and *LTR*). The higher performance of predictors during recessions is consistent with the findings of Fama and French (1989) and Cochrane (2005), who claim that increasing risk aversion during such periods inflates the risk premium demanded by investors, consequently triggering the high predictability of equity premiums. *SCSTR* outperforms *CSTR* in all horizons and both states of the economy except for the 1-month horizon in recessions. *SCSTR* is the only predictor that has a significant and positive R^2 for all horizons in expansions. Most of the incremental value of *SCSTR* is derived from expansions for longer horizons and from recessions for shorter horizons.

3.7. Industry analysis

In this subsection, I analyze the importance of industry decomposition. Is there any industry-specific factor that drives the previous results? Does it still work at the industry level? I estimate the same type of variables – *SCSTR* and *CSTR* – but at the industry level. Then, I examine their predictability not only for the industry excess returns but also for the market excess returns. Since I require a somewhat large enough sample for this analysis, I use the 10-industry Kenneth French classification. There are between 37 and 624 stocks for the day with minimum stocks in any industry and between 58 and 1554 for the day with maximum stocks in any industry. Panels A and B of Table 8 give the R^2 results for IS and OOS, respectively. I analyze this for four horizons as before and discriminate between columns (A) and (B), with the former being the industry predictability and the latter being market risk premium predictability. There is no doubt that the magnitude and significance of the results are stronger for *SCSTR*, both IS and OOS. For example, in the OOS results, almost no R^2 is significant when using *CSTR*, but the results are quite strong in the case of *SCSTR*.

Next, I combine industry tail risk measures to predict the equity risk premium. I construct two types of combinations. I compute the average of industry tail risk measures and denote *AISCSTR* and *AICSTR* for *SCSTR* and *CSTR,* respectively. I also compute the maximum of industry tail risk measures to understand the maximum impact of industry tail risk at each point in time. I denote these measures as *MISCSTR* and *MICSTR* for *SCSTR* and *CSTR,* respectively. I present the results in Table 9, which point to the same direction as before. The R^2 of combined industry measures based on *SCSTR* hold more predictability for the equity premium than the same type of measures based on *CSTR*. As an example, the OOS of R^2 for *AISCSTR* and 1-month horizon is 1.05% and significant at the 1% level,

[13] The data are available online at http://www.nber.org/cycles.html.

Table 8
Predictability results of industry tail measures.

	SCSTR								*CSTR*							
	1M		3M		1Y		3Y		1M		3M		1Y		3Y	
	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)
Panel A: In-Sample Results																
NoDur	0.99***	0.85***	1.45**	0.68	4.04*	3.50**	12.10*	12.99***	0.83**	0.67**	0.95*	0.35	2.29	1.98	6.43	7.77**
Durbl	0.50*	0.71**	0.52	0.95**	0.46	2.15*	1.29	6.63**	0.60*	0.54*	0.43	0.58	0.44	1.34	0.12	2.97
Manuf	0.41*	0.53**	0.96*	1.122**	4.51**	5.36**	19.85***	18.56***	0.28	0.40*	0.17	0.23	1.16	1.65	6.61	6.40*
Enrgy	0.42*	0.75**	0.68	1.60**	3.50**	6.00***	7.41***	17.03***	0.48*	0.75**	0.25	1.15*	1.82	3.20*	4.70***	8.93**
Hitec	0.73**	0.95***	0.65	0.70*	5.51**	6.01***	15.85**	18.60***	0.60**	0.87***	0.28	0.38	3.47*	3.59**	12.02**	13.49***
Telcm	0.99**	0.32	1.96**	0.26	5.78***	1.16	10.08**	3.17	1.29***	0.26	2.07**	0.31	6.62***	1.77	10.69***	3.78*
Shops	0.98***	0.89**	1.01*	1.29**	6.27**	8.50***	13.55**	22.24***	0.77**	0.76**	0.46	0.52	3.62*	5.37**	5.9	12.63***
Hlth	0.35*	0.50**	0.00	0.05	0.31	1.65*	0.60	5.77***	0.00	0.04	0.02	0.00	0.60	1.46	1.01	5.42**
Utils	0.45	0.47	0.87	1.06*	5.39**	3.73**	17.44***	11.78**	0.24	0.34	0.38	0.57	3.42**	2.79**	15.14***	11.46***
Other	1.27***	0.94**	1.47*	1.23*	6.68**	6.75**	21.45***	20.23***	1.06*	0.61**	1.18*	0.66	5.51**	4.66**	17.95***	15.03***

	SCSTR								*CSTR*							
	1M		3M		1Y		3Y		1M		3M		1Y		3Y	
	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)
Panel B: Out-of-Sample Results																
NoDur	1.02**	0.72**	1.89*	0.51	4.59	3.37**	9.08	11.30**	0.71	0.49**	1.04	0.03	1.72	1.2	0.85	4.67*
Durbl	0.44**	0.85**	0.44	1.22**	0.3	2.64*	0.57	4.78**	0.46	0.55	0.26	0.6	−0.22	0.35	−1.25	−1.02
Manuf	0.00	0.16	0.81*	1.00	6.04***	6.65***	22.61***	17.86***	−0.36	−0.17	−0.33	−0.24	0.74	1.07	4.58	2.75
Enrgy	−0.36	0.24	−1.52	0.52*	−0.29	5.78***	8.70***	17.87***	0.11	0.44	−0.6	0.85	−1.00	1.74	5.38***	9.26*
Hitec	0.54**	0.93***	0.67	0.84	6.55**	7.70***	17.55***	20.14***	0.46**	1.01***	0.2	0.42	4.26**	4.96**	12.66**	14.50**
Telcm	0.99***	0.17	2.30***	0.12	6.41***	1.31	10.54**	3.77	1.24**	0.09	2.41**	0.11	7.24**	2.52	10.44*	4.38***
Shops	1.25***	1.05**	1.41***	1.59**	10.58***	9.68***	19.90***	20.90***	0.93***	0.80*	0.45	0.41	4.12*	4.56**	5.04	9.03**
Hlth	0.49*	0.49	−0.97	−0.12	−1.17	2.49**	−0.23	6.69***	−0.31	−0.15	−1.05	−0.31	−0.37	2.00	0.49	5.80**
Utils	−0.07	−0.1	0.69*	1.00*	6.77***	4.73**	21.10***	14.70**	−0.79	−0.32	−0.09	0.31	3.72**	3.50**	14.49***	13.11***
Other	1.27***	0.97**	1.43*	1.22*	5.79***	7.61***	15.59***	19.48***	1.05**	0.55	1.18	0.57	4.59*	4.94*	11.23***	13.08**

This table presents the R^2 values (in %) of the industry tail measures on the industry and market premium predictions for 1-month, 3-month, 1-year and 3-year horizons. Column (A) shows the R^2 value for the industry portfolio, and column (B) shows the R^2 value for the market portfolio. The sample period is between July 1962 and December 2018. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

Table 9
Predictability by combining industry tail measures.

	1M	3M	1Y	3Y
Panel A: In-Sample Results				
AISCSTR	1.11***	1.40**	6.81***	20.97***
AICSTR	0.77**	0.70***	4.20***	13.03***
MISCSTR	1.00**	1.89	4.86**	15.92***
MICSTR	0.55*	0.66	3.95**	11.91***
	1M	3M	1Y	3Y
Panel B: Out-of-Sample Results				
AISCSTR	1.05***	1.77**	8.85***	22.11***
AICSTR	0.67*	0.71	4.54**	11.43**
MISCSTR	0.19	2.08***	5.98***	16.94***
MICSTR	0.29	0.66	4.25**	10.09***

This table presents the in- and out-of-sample R^2 values (in %) of industry tail measures on the market risk premium prediction for 1-month, 3-month, 1-year and 3-year horizons. The variables *AISCSTR* and *AICSTR* are the average of the *SCSTR* and *CSTR*, respectively, across industries in each month. The variables M*ISCSTR* and *MICSTR* are the maximum values of the *SCSTR* and *CSTR*, respectively, across industries in each month. The sample period is between July 1962 and December 2018. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

Table 10
Portfolio choice.

	SR	Δ SR	CE	Δ CE (p.p)
Panel A: Truncated Power Utility Function and Predicting Expected Return				
No constraints				
Baseline	0.35		6.30%	
SCSTR	0.60	0.25	10.96%	4.66%
CSTR	0.49	0.14	8.98%	2.68%
Constraints on stock market weight between −100% and 250%				
Baseline	0.35		6.30%	
SCSTR	0.57	0.22	10.89%	4.59%
CSTR	0.51	0.16	9.22%	2.92%
	SR	Δ SR	CE	Δ CE (p.p)
Panel B: Power Utility Function and State Variable in Weight				
No constraints				
Baseline	0.39		7.54%	
SCSTR	0.64	0.25	15.22%	7.68%
CSTR	0.52	0.13	9.95%	2.41%
Constraints on stock market weight between −100% and 250%				
Baseline	0.39		7.54%	
SCSTR	0.64	0.25	14.03%	6.49%
CSTR	0.51	0.12	10.42%	2.88%

This table presents the out-of-sample portfolio choice results of a trading strategy timing the market rebalanced every month. Panel A reports the case when a fourth-order expansion of a power utility function is used for the portfolio choice and the mean is predicted from either *SCSTR* or *CSTR* and compared to the baseline case of using the historical mean return. Panel B reports the case when a power utility function is used and either a state variable *SCSTR* or *CSTR* is used on the weight. The risk aversion parameter is set to 2. For all strategies, the level and change in the Sharpe ratio and power utility certainty equivalent are reported. All numbers are annualized. The sample period is from July 1962 to December 2018. The forecasts begin 20 years after the sample start.

whereas for *AICSTR* is 0.67% and significant only at the 10% level. For a 3-year horizon, the R^2 is 22.11% for *AISCSTR* and only 11.43% for *AICSTR*. The results also hold true for the maximum of industry tail measures although not as strong as for the average.

Overall, even at the industry level, my tail risk measure exhibits stronger predictability both in economic and statistical terms and demonstrates that my previous results for the market are not driven by any specific industry.

4. Added economic value

In this section, I analyze two asset allocation exercises between the stock market and the risk-free asset to understand the implications for the economic value of the tail risk measures' strong statistical predictability of the expected stock market return. I focus on the monthly frequency and use OOS forecasts to determine the weights for allocating wealth between the risky asset (the market) and riskless asset (the T-bill rate). Then, I compute two metrics: the Sharpe ratio of the strategy that considers only the first two moments of the distribution and the certainty equivalent return that a risk-averse investor is willing to pay for switching from the benchmark historical mean to the proposed model but based on a power utility function. This power utility function considers higher moments of the return distribution, which is crucial in this setting of tail risk measures. I set the risk aversion parameter equal to 2 as in Ferreira and

Table 11
Tail risk sensitivity cross-sectional results.

Deciles	Obs	Beta	*SCSTR*				*CSTR*			
			VW		EW		VW		EW	
			Average	*t*-stat	Average	*t*-stat	Average	*t*-stat	Average	*t*-stat
Panel A: Deciles Based Strategy										
D1	290	−1.06	0.2184	6.70	0.1344	3.43	0.2376	6.82	0.1524	3.77
D2	344	−0.20	0.1860	8.23	0.1248	4.57	0.1908	8.30	0.1380	4.98
D3	379	−0.02	0.1668	8.34	0.1284	5.42	0.1740	8.81	0.1392	5.94
D4	395	0.10	0.1692	7.97	0.1404	6.22	0.1692	8.07	0.1464	6.48
D5	389	0.21	0.1980	8.58	0.1440	6.13	0.1812	7.97	0.1416	6.00
D6	382	0.32	0.1944	7.76	0.1416	5.62	0.1884	7.41	0.1440	5.71
D7	371	0.45	0.2220	7.94	0.1584	5.73	0.2004	7.17	0.1476	5.45
D8	349	0.64	0.2412	7.61	0.1656	5.39	0.2472	8.01	0.1536	5.00
D9	319	0.94	0.2580	7.12	0.1704	4.85	0.2544	7.29	0.1536	4.46
D10	261	2.04	0.3588	7.94	0.2436	4.97	0.3492	8.13	0.2208	4.66
D10 - D1			Value	*t*-stat	Value	*t*-stat	Value	*t*-stat	Value	*t*-stat
AveCarhart (1997) factrage return			0.1404	3.48	0.1092	2.35	0.1116	2.83	0.0684	1.50
or alpha			0.0813	1.23	0.1205	2.27	−0.0218	−0.34	0.0532	0.98
Sharpe ratio			0.2042	1.49	0.3267	2.38	0.3334	1.62	0.2053	1.50

D10 - D1	*SCSTR*				*CSTR*			
	VW		EW		VW		EW	
	Value	*t*-stat	Value	*t*-stat	Value	*t*-stat	Value	*t*-stat
Panel B: Other Specifications for D10 – D1								
Simple Regression + Carhart (1997) factor model								
Average return	0.1116	2.93	0.0912	2.18	0.0684	1.81	0.0576	1.36
Carhart (1997) factor alpha	0.1142	1.86	0.1180	2.41	0.0429	0.74	0.0731	1.51
Sharpe ratio	0.2496	1.82	0.2993	2.19	0.0997	0.73	0.1778	1.30
Simultaneous Regression								
Average return	0.1572	4.14	0.1464	3.23	−0.1128	−2.84	−0.1416	−2.87
Carhart (1997) factor alpha	0.2056	2.88	0.2105	3.86	−0.1776	−2.60	−0.2167	−3.27
Sharpe ratio	0.4082	2.98	0.4604	3.35	−0.3929	−2.86	−0.4149	−3.02
Simultaneous Regression + Carhart (1997) factor model								
Average return	0.1236	3.43	0.1356	3.44	−0.0744	−2.10	−0.1092	−2.56
Carhart (1997) factor alpha	0.1887	2.94	0.1678	3.61	−0.1702	−2.53	−0.1347	−2.61
Sharpe ratio	0.3989	2.91	0.5001	3.64	−0.3369	−2.46	−0.3681	−2.68

Q5 - Q1	*SCSTR*				*CSTR*			
	VW		EW		VW		EW	
	Value	t-stat	Value	t-stat	Value	t-stat	Value	t-stat
Panel C: Other Specifications for Q5 – Q1								
Simple Regression								
Average return	0.0948	3.25	0.0720	2.24	0.0756	2.66	0.0372	1.17
Carhart (1997) factor alpha	0.0761	1.78	0.0762	2.09	−0.0187	−0.41	0.0226	0.61
Sharpe ratio	0.2495	1.82	0.3075	2.24	0.0192	0.14	0.1580	1.15
Simple Regression + Carhart (1997) factor model								
Average return	0.0912	3.62	0.0636	2.20	0.0696	2.78	0.0336	1.16
Carhart (1997) factor alpha	0.1072	2.49	0.0819	2.40	0.0398	0.93	0.0466	1.36
Sharpe ratio	0.2957	2.16	0.2981	2.18	0.0973	0.71	0.1542	1.13

This table reports monthly return statistics for portfolios formed on the tail risk sensitivity of each firm's stock returns. Each month stocks are sorted into decile portfolios based on predictive tail loadings that are estimated from monthly data over the previous ten years. Portfolios are based on NYSE/AMEX/NASDAQ stocks with CRSP share codes 10 and 11. Panel A reports equal- and value-weighted average out-of-sample one-month holding period portfolio returns. The first columns report the average number of firms per month in each decile, the second column the time-series average of the average beta for each decile. The next columns report for each case the average raw return and associated *t*-statistics. The table also reports high minus low zero net investment portfolio that is long decile portfolio ten (D10) and short decile one (D1). For this strategy, it reports annualized average return, annualized Carhart (1997) alpha and annualized Sharpe ratio. Panel B reports only the statistics for returns of the D10 - D1 strategies for three specifications. The first computes stock returns tail risk sensitivities controlling for the Carhart (1997) factor model. The second computes both tail risk sensitivities – of *SCSTR* and *CSTR* – simultaneously. The third repeats the second also controlling for the Carhart (1997) factor model. Panel C reports the long-short strategy based on quintiles that is long on quintile five (Q5) and short decile one (Q1). Stocks with prices below $5 at the portfolio formation date are excluded.

Santa-Clara (2011). As my predictor is a tail risk variable, a method that does not consider higher moments of the return distribution, such as skewness and kurtosis, the optimization process may lead to incorrect conclusions. Faias and Santa-Clara (2017) also use a power utility function in a context with extreme outcomes. In that regard, I run two exercises to account for tail risk measures.

I first follow Gao and Nardari (2018) by considering that the weight can be obtained by solving the following problem:

$$\omega_t^* = argmax\left\{\varphi_0 + \varphi_1 m_{p,t+1}^{(1)} + \varphi_2 m_{p,t+1}^{(2)} + \varphi_3 m_{p,t+1}^{(3)} + \varphi_4 m_{p,t+1}^{(4)}\right\}, \tag{13}$$

where $\varphi_0 = \frac{1}{1-\gamma}$, $\varphi_1 = 1$, $\varphi_2 = -\frac{\gamma}{2}$, $\varphi_3 = -\frac{\gamma(\gamma+1)}{6}$, $\varphi_4 = -\frac{\gamma(\gamma+1)(\gamma+2)}{24}$ and $m_{p,t+1}^{(i)}$ is the ith-order expected noncentral moment of portfolio returns. This formula results from the expansion up to the fourth order of the expected power utility function. I use the predicted value by either the *SCSTR* or the *CSTR* for the centralized moment of the stock market, and for the other three centralized moments of the portfolio return distribution, I use the moment estimated using historical information up to month t. Note that the noncentralized moments of order 2, 3, and 4 also depend on the first centralized moment and hence will impact all the terms. The results are reported in Panel A of Table 10. First, I find that the *SCSTR* improves the baseline case. In addition, the annualized CE using the predictability by the *SCSTR* versus the baseline case of no predictability increases by approximately 5%, whereas with the *CSTR*, it increases by approximately 3%. I also check the influence of extreme weights in my conclusions by constraining the stock market weight to be between −100% and 250%. I observe no relevant change in my conclusions.

As a second exercise and following Brandt and Santa-Clara (2006), I run conditional portfolios. Here, the weights are modeled as a linear function of the state variables. I run the model using either *SCTSR* or *CSTR* as state variables. The results are in Panel B of Table 10. Crucially, the results with *SCSTR* are always better than those using *CSTR*. Adding *SCSTR* increases the CE by approximately 100% from the case when there is no variable (CRRA utility investor only). However, *CSTR* only increases the CE by 50%. This is a clear difference between the variables. I also included the stock market weight constraint to be between −100% and 250% to make it closer to reality. However, the results barely change.

In summary, I use two different economic models to assess the added economic value of the found statistical predictability by *SCSTR*. The models that incorporate *SCSTR* provide tangible economic gains to a risk-averse investor who implements a trading strategy based on OOS forecasting of the equity premium.

5. Cross-sectional analysis

I show above that *SCSTR* is a time-varying measure that induces changes in expectations about future expected returns and expected risk. In this section, I explore the predictive power of *SCSTR* on the cross-section of U.S. stock returns and compare it to that induced by *CSTR*.

I analyze all U.S. common stocks with common share codes 10 and 11, available from the CRSP from 1962 to 2018 with at least 36 months of data, which I require to compute the rolling betas (sensitivities), yielding a total of 2905 stocks. This restriction implies having the first beta in the beginning of February 1965, leading to a total of 636 months of portfolio returns until December 2018. The factors (i.e., market risk premium, size, value, and momentum) used for the Carhart (1997) factor model are taken from Kenneth French's online data library.

First, I compute the exposure of each stock return to each tail risk measure for every month. I regress the monthly return of each stock i on the tail risk variables:

$$r_{i,t} = \beta_{i,t}^0 + \beta_{i,t}^{TR} TR_{t-1} + \varepsilon_{i,t}, \tag{14}$$

where $r_{i,t}$ is the monthly returns of stock i, and TR is either *SCSTR* or *CSTR*. I use $\beta_{i,t}^{TR}$ to rank stocks in each month and form portfolios. These regressions are rebalanced every month using a 10-year rolling window as the formation period, as in Kelly and Jiang (2014). I run rolling regressions of the previous 120 observations (10 years) requiring a minimum of 36 observations in this window. Then, for each predictor and for every month, I sort the stocks into deciles based on exposure to the tail risk variables. I form equally-weighted (EW) and value-weighted (VW) portfolios and assess the next month's performance. Finally, I compute a long-short strategy by taking a long position of stocks on decile 10 (high positive exposure) and a short position of stocks on decile 1 (high negative exposure).

The results are in Table 11. There are approximately 300 stocks in each bin. The beta (sensitivity) from D1 is on average −1.06, whereas that from D10 is on average 2.04. There is a beta of on average 0 between D3 and D4. Most of the stocks have a positive tail beta. The sensitivity is stronger in terms of absolute value for the positive betas than for the negative betas. The strategy achieves the lowest next month's average return in D3. It increases monotonically away from D3, i.e., from D3 to D1, and from D3 to D10. For example, for the VW strategy based on *SCSTR,* this value is approximately 17%. In D1, the average return corresponds to approximately 22%, which represents a difference of 5% from D3. In D10, the average return corresponds to approximately 36%, which represents a difference of 14% from D3. These results show that the long-short strategy in the table (D10-D1) delivers a lower bound than a strategy based on tail risk can deliver. Nevertheless, the self-financed long-short strategy corresponding to taking a long position on high positive betas (portfolio 10) and a short position on high negative betas (portfolio 1) based on *SCSTR* delivers an annualized average return of 14% (and significant) for the VW portfolio and 11% (and significant) for the EW portfolio, which compares well with 11% (and significant) and 7% (not significant), respectively, for *CSTR*. These results indicate that the *SCRST* contains more information and translates to approximately 4pp more in terms of the annualized average return.

I compute two additional measures. The first is the Carhart (1997) factor alpha, so I confirm that the results are robust to the inclusion of common exposures to the market risk premium, size, value, and momentum. The results are now weaker for the VW strategy based on *SCSTR* still delivering 8% alpha but nonsignificant. For the EW portfolio, the results are stronger, with an alpha of 12%, and quite significant. For the portfolios based on *CSTR*, the VW portfolio delivers a negative alpha, while the EW portfolio delivers a positive 5% but is also nonsignificant. I also compute annualized Sharpe ratios to consider the volatility of the portfolios. The only significant Sharpe ratio is the one from EW portfolios based on *SCSTR* with a value of 0.33.

Next, I run robustness tests on these results. First, I compute the sensitivities $\beta_{i,t}^{TR}$ for each stock and each month by incorporating the Carhart (1997) factor model into the same regression. Moreover, there may be time-varying covariance between these factors and each tail risk measure, which may lead to changes in the sensitivities $\beta_{i,t}^{TR}$. The results are in Panel B of Table 11. Only the average return of the VW portfolio is significant with a value of approximately 7%. After controlling for common factors, *SCSTR* still delivers significant abnormal returns that are 60% higher than the strategy based on *CSTR*. I next run a bivariate regression using the two tail risk measures simultaneously:

$$r_{i,t} = \beta_{i,t}^{0} + \beta_{i,t}^{SCSTR} SCSTR_{t-1} + \beta_{i,t}^{CSTR} CSTR_{t-1} + \varepsilon_{i,t}. \tag{15}$$

The results are in Panel B of Table 11. The first important result is that while the strategy based on *SCSTR* still delivers positive performance measures, the strategy based on *CSTR* inverts the relationship and delivers negative performance measures. This strategy replicates the same directional results of the time series predictability in Table 5 (i.e., *SCSTR* obtains the correct sign, whereas the *CSTR* does not). Both the VW and EW strategies based on *SCSTR* deliver a strong significant positive average return of approximately 15%, a Carhart (1997) factor alpha of approximately 21%, and a Sharpe ratio of approximately 0.40. Next, I run the same exercise—simultaneously regression—but control for the Carhart (1997) factor model. The results for the average return and Carhart (1997) factor alpha drop by approximately 2% but continue to be strongly significant. Last, I run the same analysis based on a smaller partition of the stocks. Based on the quintiles, the self-financed long-short strategies based on *SCSTR* deliver an average return of 9% and 7% for VW and EW strategies, respectively. The performance is lower than that using deciles, which was expected. Nevertheless, both of them are strongly significant. There is still significance for Carhart's (1997) factor alpha and Sharpe ratio for all portfolios based on *SCSTR*. The average returns of the same type of strategies based on *CSTR* are 2% and 3% lower for the VW and EW strategies, respectively. Note that the portfolios based on *CSTR* are mostly not significant (for all performance measures) except for the average returns of the VW portfolio.

Overall, the results show that *SCSTR* contains information on price stock returns in the cross-section, and the results are robust to controlling for *CSTR* and for the Carhart (1997) factor model.

6. Conclusion

In this paper, I propose a new measure, *SCSTR*, by averaging daily cross-sectional tail exponents within a month. I test this measure and other variables, including valuation ratios, stock market characteristics and interest rates, for OOS predictive power during the period between July 1982 and December 2018. *SCSTR* is the only variable that yields a significant and positive OOS R^2 over short- and long-term forecast horizons. In addition, it also consistently outperforms a set of popular predictors.

I run a simulation study to understand in which scenarios *SCSTR* outperforms *CSTR*. I find that *SCSTR* better captures monthly tail risk rather than merely the tail risk on specific days within a month. Subsequently, I use this statistical predictability to understand the economic value of using *SCSTR* by running an asset allocation exercise between the stock market and the riskless asset. Conditional portfolios based on *SCSTR* deliver an annualized certainty-equivalent of about 15% compared to a baseline result assuming no predictability of 8%. For all scenarios and models considered, the strategy based on *SCSTR* delivers a certainty-equivalent that is at least 30% better than the ones based on *CSTR* or assuming no predictability. Lastly, I explore the implications for the cross-section of stock returns. I show that a self-financed long-short strategy based on the sensitivity of firms' stock returns to *SCSTR* delivers an average return of 10% from 1982 to 2018.

There is clear evidence of the superiority of *SCSTR* over *CSTR* for the U.S. stock market not only from a statistical but also from an economic point of view.

References

Ang, A., Bekaert, G., 2007. Stock return predictability: is it there? Rev. Financ. Stud. 20, 651–707.
Annaert, J., De Ceuster, M., Verstegen, K., 2013. Are extreme returns priced in the stock market? European evidence. J. Bank. Finance 37, 3401–3411.
Baker, M., Wurgler, J., 2000. The equity share in new issues and aggregate stock returns. J. Finance 55, 2219–2257.
Bali, T.G., Cakici, N., Whitelaw, R., 2011. Maxing out: stocks as lotteries and the cross-section of expected returns. J. Financ. Econ. 99, 427–446.
Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. Operat. Res. Q. 20, 451–468.
Bollerslev, T., Todorov, V., Li, S.J., 2013. Jump tails, extreme dependencies, and the distribution of stock returns. J. Econom. 172, 307–324.
Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: what do we learn? Rev. Financ. Stud. 12, 405–428.
Boudoukh, J., Michaely, R., Richardson, M.P., Roberts, M.R., 2007. On the importance of measuring payout yield: implications for empirical asset pricing. J. Finance 62, 877–915.
Brandt, M.W., Santa-Clara, P., 2006. Dynamic portfolio selection by augmenting the asset space. J. Finance 61, 2187–2217.
Brownlees, C., Engle, R., Kelly, B.T., 2011. A practical guide to volatility forecasting through calm and storm. J. Risk 14, 3–22.
Campbell, J.Y., 1987. Stock returns and the term structure. J. Financ. Econ. 18, 373–399.
Campbell, J.Y., Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Rev. Financ. Stud. 1, 195–228.
Campbell, J.Y., Shiller, R.J., 1998. Valuation ratios and the long-run stock market outlook. J. Portfolio Manag. 24, 11–26.
Campbell, J.Y., Thompson, S.B., 2008. Predicting the equity premium out of sample: can anything beat the historical average? Rev. Financ. Stud. 21, 1456–1508.
Carhart, M.M., 1997. On persistence in mutual fund performance. J. Finance 52, 57–82.
Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. J. Econom. 138, 291–311.
Cochrane H., J., 1991. Production-based asset pricing and the link between stock returns and economic fluctuations. J. Finance 46 (1), 209–237.
Cochrane, J.H., 2005. Financial markets and the real economy. Found. Trends® Finance 1, 1–101.
Danielsson, J., Hartmann, P., de Vries, C.G., 1998. The cost of conservatism: extreme returns, value-at-risk, and the Basle' multiplication factor. J. Risk 11, 101–113.

De Vries, C.G., 1991. On the relationship between GARCH and stable processes. J. Econom. 48, 313–324.
Drechsler, I., Yaron, A., 2011. What's vol got to do with it? Rev. Financ. Stud. 24, 1–45.
Elliott, G., Drive, G., 2011. Averaging and the Optimal Combination of Forecasts. Working Paper.
Elliott, G., Gargano, A., Timmermann, A., 2013. Complete subset regressions. J. Econom. 177, 357–373.
Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer.
Engle, R., Kelly, B., 2012. Dynamic equicorrelation. J. Bus. Econ. Stat. 30, 212–228.
Faias, J.A., Castel-Branco, T., 2018. Out-of-sample stock return prediction using higher-order moments. Int. J. Theor. Appl. Finance 21, 1–27.
Faias, J.A., Santa-Clara, P., 2017. Optimal option portfolio strategies: deepening the puzzle of index option mispricing. J. Financ. Quant. Anal. 52, 277–303.
Faias, J.A., Zambrano, J.A., 2022. Equity Premium Predictability from Cross-Sectorial Downturns. Review of Asset Pricing Studies, Forthcoming.
Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. J. Financ. Econ. 22, 3–25.
Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. J. Financ. Econ. 25, 23–49.
Fama, E.F., Schwert, G.W., 1977. Asset returns and inflation. J. Financ. Econom. 5, 115–146.
Ferreira, M.A., Santa-Clara, P., 2011. Forecasting stock market returns: the sum of the parts is more than the whole. J. Financ. Econ. 100, 514–537.
Fousseni, C.-Y., Ruenzi, S., Weigert, F., 2018. Crash sensitivity and the cross-section of expected stock returns. J. Financ. Quant. Anal. 53, 1059–1100.
Frahm, J., Junker, M., Schmidt, T., 2005. Estimating the tail-dependence coefficient: properties and pitfalls. Insur. Math. Econ. 37, 80–100.
Gabaix, X., Gopikrishnan, P., Plerou, V., Stanley, H., 2006. Institutional investors and stock market volatility. Q. J. Econ. 121, 461–504.
Gao, X., Nardari, F., 2018. Do commodities add economic value in asset allocation? New evidence from time-varying moments. J. Financ. Quant. Anal. 53, 365–393.
Gardes, L., Stupfler, G., 2013. Estimation of the conditional tail index using a smoothed local hill estimator. Extremes 17, 45–75.
Ghose, D., Kroner, K.F., 1995. The relationship between GARCH and symmetric stable processes: finding the source of fat tails in financial data. J. Empir. Finance 2, 225–251.
Goyal, A., Welch, I., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21, 1509–1531.
Guo, H., 2006. On the out-of-sample predictability of stock market returns. J. Bus. 79, 645–670.
Harvey, C.R., Leybourne, S.J., Newbold, P., 1998. Tests for forecast encompassing. J. Bus. Econ. Stat. 16, 254–259.
Hill, B., 1975. A simple general approach to inference about the tail of a distribution. Ann. Stat. 3, 1163–1174.
Hodrick, R., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Rev. Financ. Stud. 5, 357–386.
Jondeau, E., Rockinger, M., 1999. The Tail Behavior of Stock Returns: Emerging versus Mature Markets. Banque de France Working Paper.
Kearns, P., Pagan, A., 1997. Estimating the density tail index for financial time series. Rev. Econ. Stat. 79, 171–175.
Keim B., D., Stambaugh F., R., 1986. Predicting returns in the stock and bond markets. J. Financ. Econ. 17 (2), 357–390.
Kelly, T.B., Jiang, H., 2014. Tail risk and asset prices. Rev. Financ. Stud. 27, 2841–2871.
Kostakis, A., Magdalinos, T., Stamatogiannis, M.P., 2015. Robust econometric inference for stock return predictability. Rev. Financ. Stud. 28, 1506–1553.
Kothari, S., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: a time-series analysis. J. Financ. Econ. 44, 169–203.
Lamont, O., 1998. Earnings and expected returns. J. Finance 53, 1563–1587.
Lettau, M., Ludvigson, S., 2005. Expected returns and expected dividend growth. J. Financ. Econ. 76, 583–626.
Lettau, M., Nieuwerburgh, S.V., 2008. Reconciling the return predictability evidence. Rev. Financ. Stud. 21, 1607–1652.
Lewellen, J., 2004. Predicting returns with financial ratios. J. Financ. Econ. 74, 209–235.
Li, J., Tsiakas, I., Want, W., 2015. Predicting exchange rates out of sample: can economic fundamentals beat the random walk? J. Financ. Econom. 13, 293–341.
Lintner, J., 1975. Inflation and security returns. J. Finance 30, 259–280.
Longin, F., Solnik, B., 2001. Extreme correlation in international equity markets. J. Finance 56, 649–676.
Nelson, C.R., 1976. Inflation and the rates of return on common stock. J. Finance 31, 471–483.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Polk, C., Thompson, S., Vuolteenaho, T., 2006. Cross-sectional forecasts of the equity premium. J. Financ. Econ. 81, 101–141.
Pontiff, J., Schall, L.D., 1998. Book-to-market ratios as predictors of market returns. J. Financ. Econ. 49, 141–160.
Poon, S.-H., Rockinger, M., Tawn, J., 2004. Extreme value dependence in financial markets: diagnostics, models, and financial implications. Rev. Financ. Stud. 17, 581–610.
Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev. Financ. Stud. 23, 821–862.
Rapach, D.E., Ringgenberg, M.C., Zhou, G., 2016. Short interest and aggregate stock returns. J. Financ. Econ. 121, 46–65.
Richardson, M., Smith, T., 1991. Tests of financial models in the presence of overlapping observations. Rev. Financ. Stud. 4, 227–254.
Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. J. Risk 2, 21–41.
Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. J. Forecast. 23, 405–430.
Timmermann, A., 2006. Forecast combinations. Handb. Econ. Forecast. 1.
Wang, J., Liu, Z., Meric, G., Meric, I., 2009. Stock market crashes, firm characteristics, and stock returns. J. Bank. Finance 33, 1563–1574.