Measuring tail risk

Article

Published Version

Dierkes, M., Hollstein, F., Prokopczuk, M. and Würsig, C. M. (2024) Measuring tail risk. Journal of Econometrics, 241 (2). 105769. ISSN 1872-6895 doi: 10.1016/j.jeconom.2024.105769 Available at https://centaur.reading.ac.uk/117476/

Journal of Econometrics 241 (2024) 105769

Contents lists available at ScienceDirect

Journal of Econometrics

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

Measuring tail risk ☆

Maik Dierkes [a,*], Fabian Hollstein [b], Marcel Prokopczuk [a,c], Christoph Matthias Würsig [a]

[a] Leibniz University Hannover, School of Economics and Management, Koenigsworther Platz 1, 30167 Hannover, Germany
[b] Saarland University, School of Human and Business Sciences, Campus C3 1, 66123 Saarbruecken, Germany
[c] University of Reading, Henley Business School, ICMA Centre, Reading RG6 6BA, UK

ARTICLE INFO

JEL classification:
G12
C58
G17
G10

Keywords:
Tail risk
Return forecasting
Tail event forecasting

ABSTRACT

We comprehensively investigate the usefulness of tail risk measures proposed in the literature. We evaluate their statistical as well as their economic validity. The option-implied measure of Bollerslev and Todorov (2011b) ($BT11Q$) performs best overall. While some other tail risk measures excel at specialized tasks, $BT11Q$ performs well in all tests: First, $BT11Q$ can predict both future tail events and future tail volatility. Second, it has predictive power for returns in both the time series and the cross-section, as well as for real economic activity. Finally, a simulation analysis shows that the main driver of performance is measurement error.

1. Introduction

Tail risk can be defined as the risk of ending up in an exceptionally bad state of the world. That is, one in which a low-probability, high-impact, i.e., high-marginal-utility, event occurs. In asset pricing, such a (left)-tail event is typically associated with extremely negative market returns. Several anecdotal and empirical observations suggest that investors are concerned about tail risk. First, previous studies find that the prices of out-of-the-money put options, instruments that provide a positive payoff in the case of a left-tail event, are substantially higher than suggested by theory (Jackwerth, 2000; Bondarenko, 2014). Thus, investors appear to be willing to pay more than standard models suggest to obtain crash insurance. Second, The Economist describes "low-probability, high-impact events" as "a fact of life".[1] Investment practitioners and politicians worry about "fail[ing] to capture [...] the extreme negative tail" (Alan Greenspan) and see as one of their main objectives to "remove [...] tail risks, and the perception of tail risks" (Olivier Blanchard).[2,3]

The apparent interest of investors in tail events has sparked a large literature on various tail risk measures. Such measures come in a variety of fashions, ranging from highly parameterized models to nonparametric approaches. The underlying data vary from

☆ We thank Torben Andersen (the editor), an anonymous associate editor, two anonymous referees, as well as Martin Becker, Victor Todorov, and seminar participants at the 2020 SOFIE Summer School, the 2021 Annual Meeting of the Southern Finance Association, and Leibniz University Hannover for helpful comments and suggestions.

* Corresponding author.
E-mail addresses: maik.dierkes@finance.uni-hannover.de (M. Dierkes), fabian.hollstein@uni-saarland.de (F. Hollstein), prokopczuk@fcm.uni-hannover.de (M. Prokopczuk), wuersig@fcm.uni-hannover.de (C.M. Würsig).

[1] Lead article "The next catastrophe" in the June 25, 2020 issue of The Economist.

[2] The first quote is from a speech by Alan Greenspan in 1999: https://www.federalreserve.gov/boarddocs/speeches/1999/19991014.htm. The second is from an interview with Olivier Blanchard, then Chief Economist at the IMF, for The Economist, January 31, 2009.

[3] In addition, the Chicago Board Options Exchange (CBOE) introduced the VIX Tail Hedge Index (VXTH), designed to cope with extreme downward movements in the stock index.

https://doi.org/10.1016/j.jeconom.2024.105769
Received 9 August 2022; Received in revised form 23 April 2024; Accepted 5 May 2024
Available online 22 May 2024

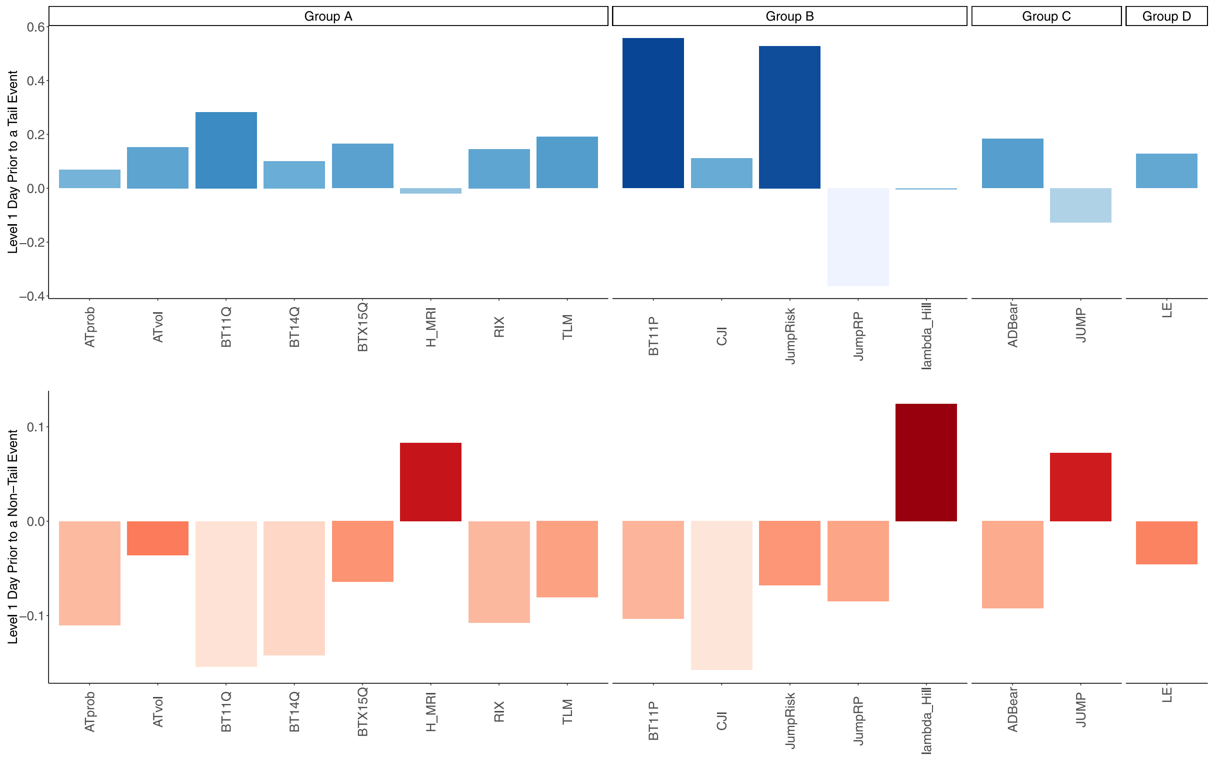

Fig. 1. The top panel of this figure shows the average levels of various tail risk measures one day before (two-sigma or more) left-tail events. In the bottom panel, we show a simple placebo test, reporting the average level of the tail risk measures before (absolute return of 0.02 sigma or less) nontail events. All tail risk measures are standardized to have a mean of zero and a volatility of one. We divide the tail risk measures into four groups: option-implied (Group A), stock-return-based (Group B), option-return-based (Group C), and macroeconomic measures (Group D). The colors indicate the intensity of the tail risk measures prior to the events. The definitions of the tail risk measure acronyms are given in Table 1.

option prices, to historical index and stock returns, to macroeconomic time series. The measures themselves vary in the main aspect of tail risk they attempt to capture, with some capturing tail probabilities, others tail variation, and still others tail risk premia. Finally, some measures capture tail risk under the physical, while others rely on the risk-neutral probability distribution. In short, both investors and policymakers face a difficult choice between different measures with potentially conflicting predictions.

In this paper, we seek to provide some guidance on how best to measure tail risk. Our main contribution is to provide a systematic, coherent, and comprehensive assessment of the tail risk measures proposed in the literature. Knowing how to measure tail risk is very important for academics, investment practitioners, and policymakers. Decisions based on an inaccurate measure could lead to huge investment and welfare losses. Furthermore, assuming that tail risk is a relevant risk factor, it is essential for academics and investors to accurately attribute portfolio performance to tail risk exposures. Thus, there is a great need to identify good tail risk measures.

We analyze a large set of 16 tail risk measures introduced in the literature. Because they are partially based on very different concepts, theories, assumptions, and underlying data, the different tail risk measures are likely to capture different things. Indeed, we find that the first two principal components (PCs) of the tail risk measures explain only 58% of their variation. The correlations between the different measures are moderate at best. In some cases, we even observe negative correlations. Thus, the decision to use a specific measure is nontrivial, with potentially important consequences. Tail risk measures should not be treated as interchangeable.

As a preview, Fig. 1 illustrates the large heterogeneity in the measures. It shows the average levels of the tail risk measures (each standardized to have a mean of zero and a standard deviation of one) one day before the tail events, as well as one day before the placebo (nontail) events. Some of them are high (as they should be), while others are close to or even below their average before a tail event. Similarly, some measures on average indicate that a tail event is likely to occur, when in fact no such event subsequently materializes.

Having documented significant heterogeneity across measures, we go on to define the desirable criteria for a tail risk measure: it should matter both statistically and economically. That is, on the one hand, the tail risk measure should be able to capture both the risk of jumps and provide an indication of the expected magnitude and quadratic variation caused by tail events. Thus, the first two tests we devise are statistical in nature, with (i) a predictive probit regression that predicts two-sigma events and (ii) a prediction of future left-tail variation. On the other hand, several studies show that tail risk is also important for investors (e.g., Rietz, 1988; Barro, 2006; Gourio, 2012; Muir, 2017; Dew-Becker et al., 2021). Therefore, a tail risk measure should be priced in the market. The last of our main tests (iii) is thus economic in nature: we examine whether the measures can predict future market excess returns. All in all, we require a tail risk measure to ideally predict both risk and risk premia.

The overall winner of our analysis is the Bollerslev and Todorov (2011b) option-implied left-tail measure ($BT11Q$). It performs well in predicting the occurrence and in particular the variation associated with future tail events up to one week in advance. In addition, it is able to forecast future market excess returns for several horizons up to one year in advance. $BT11Q$ is among the best measures for each of these tasks, performing consistently well in all tests. In addition, it is relatively simple to implement compared to other tail risk measures, requiring only observed deep out-of-the-money index put option prices.

We document that $BT11Q$ can also predict the magnitude, not just the occurrence, of future tail events. Furthermore, it performs well in predicting cross-sectional stock returns. Moreover, it also predicts real economic activity: $BT11Q$ is a strong negative predictor of the growth of industrial production in the next month and year. We also confirm in the simulated environments of the Pan (2002) and Santa-Clara and Yan (2010) jump–diffusion models that the empirical return predictability results we observe for $BT11Q$ and the other measures are plausible.

Other measures work well for some specialized tasks. For example, the $JumpRisk$ measure of Maheu et al. (2013) works best for predicting future tail events. However, it cannot predict future returns. The $BT11Q$ measure works best for predicting future left-tail variation. The $ADBear$ measure of Lu and Murray (2019) also performs well overall. In particular, it excels at predicting future market excess returns, where it performs best overall, and also performs well in the statistical tests. Overall, we find that the $ADBear$ measure is slightly inferior to the $BT11Q$ measure. Finally, more specialized, we find that the λ_{Hill} measure of Kelly and Jiang (2014) works very well for long-term return forecasting.

We also analyze the sensitivity and robustness of $BT11Q$ and other tail risk measures to their definitions. We find that moderate changes in the target moneyness level do not materially affect the performance of $BT11Q$. Thus, it is not primarily the precise part of the tail that is captured that drives the performance. However, with a time-varying target moneyness defined based on the current at-the-money option-implied volatility, the measure performs significantly less well. This suggests that investors are more interested in the absolute magnitude of a loss than its volatility-adjusted size. For other measures that rely on moneyness cutoffs rather than targets, moderate variations in the cutoff generally do not have a material impact on the results.

Several other tests underscore the robustness of our results. Among other things, we show that the results are qualitatively similar across subsample periods, when predicting the number of jumps, when varying the tail event thresholds, for alternative definitions of left-tail variation, and for different approaches to determining statistical significance. For all tests, the $BT11Q$ measure is among the best.

Why does the $BT11Q$ measure work so well? It seems to combine several desirable features. For one, it uses forward-looking information from options markets. In addition to being forward-looking, options markets have been shown to contain information about future returns that is not readily found in physical risk measures (Andersen et al., 2015).[4] Most stock-return-based and macroeconomic tail risk measures fail, especially in predicting returns. Moreover, the $BT11Q$ measure does not require the estimation of structural parameters. Our simulations show that the $BT11Q$ measure captures its target (the left tail variation) much better than do most others. In particular, similar measures that require parametric or nonparametric optimization perform worse both empirically and in the simulations. Thus, measurement error is an important driver of the relative performance of tail risk measures.

The literature contains studies that compare different risk measures in several areas. For example, there is a large literature comparing the ability of different approaches to predict future volatility (e.g., Andersen and Bollerslev, 1998; Hansen and Lunde, 2005; Jiang and Tian, 2005; Brownlees and Gallo, 2010). There are also studies looking at how best to forecast covariances (e.g., Symitsi et al., 2018) and beta (e.g., Faff et al., 2000; Hollstein and Prokopczuk, 2016; Hollstein et al., 2019). Surprisingly, however, to the best of our knowledge, to date no such study exists about tail risk. Given the plethora of different measures that have been proposed over the last decade, we believe there is an urgent need for such a study. Our main contributions are therefore to (i) define the criteria that a good tail risk measure should satisfy and (ii) comprehensively analyze the measures proposed in previous studies based on these criteria. Importantly, we use the same methodology to analyze and evaluate all measures.

The rest of the paper is organized as follows: In Section 2, we present the tail risk measures considered. Section 3 describes our methodology and data. In Section 4, we present the results of our main analysis, and in Section 5 we examine the impact of variations in the tail risk measure definitions. Simulation results are shown in Section 6. Section 7 concludes. Detailed descriptions of the tail risk measures and extensive further analyses and robustness checks are provided in the Online Appendix.

2. Tail risk measures

Our goal is to analyze as comprehensive a set of tail risk measures as possible. The selection of measures is based on two main criteria: (i) relevance/importance and (ii) (public) availability of the underlying data for the measure. Based on these criteria, we have compiled the following list.[5]

[4] In fact, David Einhorn refers to the traditional Value-at-Risk (VaR) approach based on historical return data as "an airbag that works all the time except when you have a car accident" (https://www.valuewalk.com/wp-content/uploads/2014/05/Grants-Conference-04-08-2008.pdf).

[5] Other relevant measures include Andersen et al. (2015), Agarwal et al. (2017), Andersen et al. (2017), Seo and Wachter (2018) and Weller (2018). We do not use the measure of Andersen et al. (2015) because the model is highly parameterized, making the estimation computationally very intensive. For Andersen et al. (2017), the weekly options are only available for a limited time period starting in 2011, making a meaningful empirical evaluation infeasible. Finally, we do not have access to the data underlying the measures in Agarwal et al. (2017), Seo and Wachter (2018), and Weller (2018).

Table 1
Description of the tail risk measures.
This table lists the main tail risk measures used in this study. The column "Acronym" defines the symbol used in this paper to refer to the measure. "Source" provides the reference of the original paper, "Description" gives the main cornerstones of the definitions of the tail risk measures. "Interpretation" characterizes the main quantity of which the measure provides an estimate. We assign the measures to the various categories to which they most naturally (but rather broadly defined) belong. Finally, "Freq" denotes the frequency with which the different measures are available. "D" indicates that we observe a measure every trading day. "W", "M", and "Q" denote weekly, monthly and quarterly observation frequencies, respectively.

Acronym	Source	Description	Interpretation	Freq
Group A — Option-Implied Measures				
ATprob	Andersen et al. (2020, 2021)	Probability of a daily loss of 10% or more	Left-tail probability	D
ATvol	Andersen et al. (2020, 2021)	Nonparametric tail variation estimate	Left-tail variation	D
BT11Q	Bollerslev and Todorov (2011b)	Left tail approximation measure under $\mathbb{Q}$	Left-tail variation	D
BT14Q	Bollerslev and Todorov (2014)	Inverse left-tail shape parameter	Tail intensity decay	W
BTX15Q	Bollerslev et al. (2015)	Parametric tail variation estimate (with time-varying tail shape and jump intensity)	Left-tail variation	W
H_MRI	Gormsen and Jensen (2022)	First PC of risk-neutral higher moments	Higher moment risk	D
RIX	Gao et al. (2018, 2019)	Left-tail volatility as the difference of two volatility indexes	Jump intensity	D
TLM	Vilkov and Xiao (2015)	Expected shortfall inferred from parameterized tail distribution	Expected shortfall	D
Group B — Stock-Return-Based Measures				
BT11P	Bollerslev and Todorov (2011b)	Left-tail approximation measure under $\mathbb{P}$	Left-tail variation	D
CJI	Christoffersen et al. (2012)	Parametric model-implied conditional jump intensity	Jump intensity	D
JumpRisk	Maheu et al. (2013)	Parametric model-implied conditional jump intensity	Jump intensity	D
JumpRP	Maheu et al. (2013)	Parametric model-implied conditional jump risk premium	Jump risk premium	D
λ_{Hill}	Kelly and Jiang (2014)	Left-tail shape parameter derived from the cross-section of stock returns	Tail intensity decay	M
Group C — Option-Return-Based Measures				
ADBear	Lu and Murray (2019)	Return of bear spread put option positions	(Change in) Left-tail prob.	D
JUMP	Cremers et al. (2015)	Return of vega-neutral, gamma-positive option portfolio	Jump risk premium	D
Group D — Macroeconomic Measures				
LE	Adrian et al. (2019)	Left entropy of expected GDP growth	Left entropy	Q

In the following, we present the main tail risk measures analyzed in this study. In order to keep the paper focused, only the main mechanisms of the different measures are described in this section. The technical details can be found in Section A1 of the Online Appendix. We categorize the measures into four main groups, mainly based on their underlying data: (i) option-implied measures, (ii) stock-return-based measures, (iii) option-return-based measures, and (iv) tail risk measures based on macroeconomic data.

In Table 1, we summarize the measure acronyms and provide brief descriptions, further information on how to interpret the different measures, and the estimation frequency. Whenever possible, we define the acronyms for the tail risk measures in the same way as in the original studies. In cases where this would result in names that could not be uniquely identified, we rely on bibliographic information about the study to generate generic acronyms based on the author names, years, and the probability measure under which they are estimated.

2.1. Underlying asset price dynamics

Let X denote an asset price. Many of the tail risk measures, especially among the option-implied ones, assume a dynamic continuous-time representation of the type:

$$\frac{dX_t}{X_t} = a_t dt + \sigma_t dW_t + \int_{\mathbb{R}} (e^x - 1)\,\tilde{\mu}^{\mathbb{P}}(dt, dx). \tag{1}$$

There are only a few restrictions on the drift (a_t) and the diffusive process (W_t; a Brownian motion). They are required to follow càdlàg paths but are otherwise left unspecified. $\mu(dt, dx)$ is a simple counting measure of the jumps in X. The jump compensator is $dt \otimes \nu_t^{\mathbb{P}}(dx)$, so that $\tilde{\mu}^{\mathbb{P}}(dt, dx) = \mu(dt, dx) - dt \otimes \nu_t^{\mathbb{P}}(dx)$. The quadratic variation of the log price process over the interval $[t, T]$ is:

$$QV_{[t,T]} = \int_t^T \sigma_s^2 ds + \int_t^T \int_{\mathbb{R}} x^2 \mu(dt, dx). \tag{2}$$

The first part on the right-hand side of Eq. (2) refers to the diffusive volatility due to small price changes. The second part captures the variation due to tail events.

The corresponding dynamics under the risk-neutral probability measure $\mathbb{Q}$ are:

$$\frac{dX_t}{X_t} = \left(r_{f,t} - \delta_t\right) dt + \sigma_t dW_t^{\mathbb{Q}} + \int_{\mathbb{R}} (e^x - 1)\, \tilde{\mu}^{\mathbb{Q}}(dt, dx). \quad (3)$$

$r_{f,t}$ and δ_t are the instantaneous risk-free rate and dividend yield, respectively. $W_t^{\mathbb{Q}}$ is a Brownian motion under the risk-neutral probability measure. In addition, $\tilde{\mu}^{\mathbb{Q}}(dt, dx) = \mu(dt, dx) - dt \otimes v_t^{\mathbb{Q}}(dx)$ with the latter part denoting the jump compensator under $\mathbb{Q}$.

The predictable components of the tail variation under the real-world and risk-neutral probability measures ($\mathbb{P}$ and $\mathbb{Q}$, respectively) are:

$$TV^{\mathbb{P}} = \int_t^T \int_{\mathbb{R}} x^2 v_s^{\mathbb{P}}(dx)ds \text{ and } TV^{\mathbb{Q}} = \int_t^T \int_{\mathbb{R}} x^2 v_s^{\mathbb{Q}}(dx)ds. \quad (4)$$

For this study, our particular interest is in the left tail, hence the left-tail variation, which is defined as:

$$LTV^{\mathbb{P}} = \int_t^T \int_{x<-k_t} x^2 v_s^{\mathbb{P}}(dx)ds \quad (5)$$

and

$$LTV^{\mathbb{Q}} = \int_t^T \int_{x<-k_t} x^2 v_s^{\mathbb{Q}}(dx)ds. \quad (6)$$

k_t is a cutoff that separates continuous returns from tail events. This cutoff can be either fixed or time-varying, depending on a multiple of the current volatility (see Section A1 of the Online Appendix for details on how k_t is specified for different measures).

2.2. Option-implied measures

$\boldsymbol{BT11Q}$ (Bollerslev and Todorov, 2011b) is a measure of the left-tail variation similar to the predictable component of the left-tail variation under $\mathbb{Q}$ (Eq. (6)). The main idea behind this and other, similar tail risk measures is that close-to-maturity deep out-of-the-money put options will only end up in-the-money if there is a tail event during the remaining life span of the option. Thus, the value of the put option is primarily determined by investors' expectations of future left-tail events. $BT11Q$ is based on constant-moneyness put options with a ratio of strike to futures price of 0.9.

$\boldsymbol{BT14Q}$ (Bollerslev and Todorov, 2014) builds on a similar basic setup to Bollerslev and Todorov (2011b), but with a special focus on the parameters governing the tail shape and its time variation. The main measure promoted in the study is the inverse of the smoothed time-varying tail shape parameter α_t^-. The $BT14Q$ tail risk measure is estimated from one week of pooled short-maturity out-of-the-money put prices. The larger the $BT14Q$ measure, the slower is the decay of the left tail intensity, thus, the "fatter" is the tail for a given tail intensity level ϕ_t^-.

$\boldsymbol{BTX15Q}$, $\boldsymbol{ATprob}$, and $\boldsymbol{ATvol}$ (Bollerslev et al., 2015; Andersen et al., 2020, 2021) are further extensions of the main setup used for $BT11Q$ and $BT14Q$. Bollerslev et al. (2015) note that the time-varying tail shape parameter is not the only determinant of the jump compensator and the tail variation. The level of the tail intensity must also be estimated. For $BTX15Q$, Bollerslev et al. (2015) do this using a parametric optimization also based on one week of pooled short-maturity out-of-the-money put prices. Alternatively, $ATvol$ uses similar basic equations but with a nonparametric estimation.[6] Both $BTX15Q$ and $ATvol$ provide parametric estimates of the left-tail variation under $\mathbb{Q}$ in Eq. (6), which follows from the tail intensity process in Equation (A.2) in the Online Appendix. Finally, $ATprob$ is defined as the probability of a daily loss of 10% or more. It thus captures the cumulative left-tail intensity. The measure is also based on the nonparametric estimation of the tail shape and level shift parameters. Both $ATvol$ and $ATprob$ are estimated on a daily basis.

$\boldsymbol{H\_MRI}$ (Gormsen and Jensen, 2022) is a measure of higher-moment risk. It is defined as the first principal component (PC) of the standardized option-implied skewness and kurtosis for a constant time-to-maturity. The moments are computed using out-of-the-money put and call options using the inference techniques of Breeden and Litzenberger (1978) and Bakshi et al. (2003).[7]

$\boldsymbol{RIX}$ (Gao et al., 2018, 2019) is a left-tail variation index. The measure is constructed as the difference between the downside variance of the holding period return as in Bakshi et al. (2003) and the integrated downside variance as in Britten-Jones and Neuberger (2000). Intuitively, the difference between the two can be interpreted as tail risk, since the former gives more weight to deep out-of-the-money put option prices than the latter. Du and Kapadia (2013) show that under asset price dynamics similar to Eq. (1), their measure is proportional to the expected number of jumps.

$\boldsymbol{TLM}$ (Vilkov and Xiao, 2015) is a parameterized expected shortfall measure under $\mathbb{Q}$. To infer the tail parameters, the authors optimize over the difference between the theoretical (using Extreme Value Theory, EVT) and observed prices of deep out-of-the-money put options. The resulting density can be used to calculate the expected shortfall.

[6] Both $ATvol$ and $ATprob$ are based on the website "tailindex.com", previously maintained by Torben Andersen and Victor Todorov, but now closed down. A similar implementation to $ATvol$ is now available from the CBOE: https://www.cboe.com/us/indices/dashboard/LTV/.

[7] The authors show that the first PC is positively related to the kurtosis and negatively related to the skewness. The measure is negatively correlated with volatility. Thus, it tends to be low during volatile periods.

2.3. Stock-return-based measures

$\boldsymbol{BT11P}$ (Bollerslev and Todorov, 2011b) is a left-tail measure under the objective probability measure corresponding to $BT11Q$. It is estimated from high-frequency intraday returns above a certain threshold. The authors use EVT-based approximations and reduced-form modeling to compute a forward-looking tail measure.

$\boldsymbol{CJI}$ (Christoffersen et al., 2012) is the tail intensity from a parametric dynamic volatility with separate dynamic jumps (DVSDJ) model. It is estimated using daily return data. To obtain the unobservable measures, Christoffersen et al. (2012) use a filtering technique along with maximum likelihood estimation.

$\boldsymbol{JumpRisk}$ and $\boldsymbol{JumpRP}$ (Maheu et al., 2013) are the conditional tail intensity and the conditional equity premium due to jumps (tail risk premium), respectively. Both measures are derived from a parametric Generalized Autoregressive Conditional Heteroskedastic (GARCH) jump mixture model. The tail risk premium is calculated as the first derivative of the equity risk premium with respect to the tail intensity. Since they argue that risk premia in the model behave inversely to the current state of volatility and tail risk, we define $JumpRP$ as the inverse of the corresponding Maheu et al. (2013) measure.

$\boldsymbol{\lambda_{Hill}}$ (Kelly and Jiang, 2014) is the common time-varying tail shape parameter, derived from the cross-sectional distribution of individual stock returns. The tail threshold is defined as the fifth percentile of all daily unsystematic returns in the cross-section over the past month. The measure is computed using the Hill (1975) power law estimator.

2.4. Option-return-based measures

$\boldsymbol{ADBear}$ (Lu and Murray, 2019) is the excess return of a bear spread portfolio of S&P 500 options. The bear spread portfolio is designed to pay \$1 if the excess return of the S&P 500 is below a threshold K_2. To generate this payoff, the portfolio is long a put option with strike price K_1 and short a put option with strike K_2, where $K_1 > K_2$. The payoff is scaled by $K_1 - K_2$. The resulting portfolio pays \$0 above K_1 and \$1 below K_2. The authors set K_2 and K_1 to be at a return of -1.5 and -1 standard deviations off from the current S&P 500 forward price, respectively, and hold the portfolio for five days.

$\boldsymbol{JUMP}$ (Cremers et al., 2015) is the return of a vega-neutral and gamma-positive portfolio created from market-neutral straddles written on the S&P 500. We use the daily returns resulting from a strategy with daily rebalancing.

2.5. Macroeconomic measures

$\boldsymbol{LE}$ (Adrian et al., 2019) is a measure of the left entropy of the expected distribution of gross domestic product (GDP) growth. The authors model the conditional GDP growth distribution using interpolated quantile regressions with the National Financial Conditions Index (NFCI) as the explanatory variable.

3. Data and methodology

3.1. Data

The tail risk measures introduced in the previous section require several types of data. We obtain both options and stock return data from several sources. First, we obtain data on S&P 500 option prices as well as the corresponding Greeks and the risk-free interest rate and dividend yield from OptionMetrics. To clean the option data, we follow the steps outlined in Carr and Wu (2003, 2009). First, we remove strike prices that are duplicated per day, keeping the one with the higher open interest. Second, the bid prices must be strictly positive and ask prices cannot be lower than bid prices. Some measures impose a cutoff level for short-maturity options. To be consistent, we follow Carr and Wu (2003, 2009) and choose 8 days.

Second, we use the 1-minute prices of the S&P 500 from Thomson Reuters Tick History (TRTH). We follow the steps recommended by Barndorff-Nielsen et al. (2009) to clean the data. First, we use only data with a timestamp that falls during the exchange trading hours, i.e., between 9:30 AM and 4:00 PM EST. Second, we remove recording errors in prices. Specifically, we filter out prices that differ by more than 10 mean absolute deviations from a rolling centered median of 50 observations. We then use the nearest previous entry to assign prices to each 1-minute interval.

Third, we obtain prices from the Center for Research in Security Prices (CRSP) for all stocks traded on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), and the National Association of Securities Dealers Automated Quotations (NASDAQ) that are classified as ordinary common stocks (CRSP share codes 10 or 11). In addition, we obtain data on the S&P 500 index from the same source. We use the total return of the S&P 500 as the market return and subtract the 1-month Treasury bill rate from Kenneth French's website to obtain excess returns.[8]

Finally, we obtain data on the NFCI from the Chicago Federal Reserve and on the GDP from the Bureau of Economic Analysis (BEA). We collect additional data from Amit Goyal's webpage (10-year, 3-month, and 1-month Government Bond yields), the St. Louis FRED (AAA and BAA rated corporate bond yields, industrial production), and Martin Lettau's webpage (CAY).[9]

[8] The website is https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[9] Amit Goyal's webpage can be accessed as http://www.hec.unil.ch/agoyal/. Martin Lettau's webpage is https://sites.google.com/view/martinlettau/datawebpage.

Our sample period runs from 1996 to 2017.[10] Since the goal of this study is to compare different tail risk measures, we limit our attention to this period, even for those measures for which data would be available for longer time series.

3.2. Empirical test design

What characterizes a good tail risk measure? Obviously, it should be good at predicting future tail events. To test this property, we devise two statistical tests to assess the ability of the measures to predict future tail events. In addition, a good tail risk measure should also be important in economic terms. That is, it should command a risk premium, i.e., it should be priced by investors (Rietz, 1988; Barro, 2006). To analyze the economic content, we test the ability of the measures to predict future aggregate market returns. The following sections describe these tests in more detail.

3.2.1. Statistical tests

The first test we use is a simple prediction of realized tail events. We use a binary probit model (Vilkov and Xiao, 2015). We define the threshold based on the VIX. The binary dummy variable is defined as follows:

$$D_{t+\Delta t} = \begin{cases} 1 & \text{if } R_{t+\Delta t} \leq -2\sigma_t, \\ 0 & \text{if otherwise,} \end{cases} \tag{7}$$

where $R_{t+\Delta t}$ is the market excess return over the period from t to $t+\Delta t$, where Δt is measured in trading days. $\sigma_t = VIX_t/100\sqrt{\Delta t/252}$ is the conditional volatility. VIX_t is the level of the VIX at the end of day t.

To test whether the tail risk measure can capture the realization of a 2-sigma tail event, we run the following regression:

$$D_{t+\Delta t} = a + b \cdot TRM_t + c \cdot VIX_t + \epsilon_{t+\Delta t}, \tag{8}$$

where TRM_t is the tail risk measure observation at time t. In this regression, we control for the current level of the VIX.

While the probit model captures the occurrence of tail events, it does not account for how much the observed returns exceed the specified threshold and how much quadratic variation they account for. Predicting the quadratic variation due to left-tail events may therefore be even more important for investors. Thus, in a second test, we examine the ability of the measures to predict the future realized left-tail variation. This measure yields particularly high values when the size of the (ex-post) tail realizations is very large or when there are many tail events in the period under study. Based on Mancini (2001), Bollerslev and Todorov (2011b) propose the following left-tail variation measure, which is a special case of the truncated variance:

$$\begin{aligned} LTV_t^{\mathbb{P}} &= \sum_{i=1}^{n-1} r_i^2 \cdot 1_{r_i<(-\upsilon_{t,i}\Psi^{0.49})} \\ LTV_{t+\Delta t}^{\mathbb{P}} &= \sum_{i=t}^{t+\Delta t} LTV_i^{\mathbb{P}}, \end{aligned} \tag{9}$$

where r_i denotes an intraday log return. Following Mancini (2001) and Bollerslev and Todorov (2011a) we include only intraday returns. Ψ is the length of each intraday sampling interval as a fraction of a day. Following Bollerslev and Todorov (2011a), we use market excess returns during $n = 390$ 1-minute intervals each day to estimate Eq. (9).[11] $1_{r_i<(-\upsilon_{t,i}\Psi_n^{0.49})}$ denotes a dummy variable that equals 1 if the realized intraday return r_i is less than $-\upsilon_{t,i}\Psi_n^{0.49}$. $\upsilon_{t,i}$ is a time-varying threshold adjusted by a time-of-day (TOD) factor that accounts for the predictable variation in intraday returns[12]:

$$\upsilon_{t,i} = 4\sqrt{BV_t \wedge RV_t} \cdot TOD_i \cdot \Psi^{0.49}. \tag{10}$$

BV_t and RV_t are the bi-power and realized variation, respectively. To test whether the tail risk measure can capture the future left-tail variation, we run the following regression:

$$LTV_{t+\Delta t}^{\mathbb{P}} = a + b \cdot TRM_t + c \cdot LTV_t^{\mathbb{P}} + d \cdot VIX_t + \epsilon_{t+\Delta t}. \tag{11}$$

We control for both the lagged left-tail variation $LTV_t^{\mathbb{P}}$ and the current conditional volatility, measured by VIX_t. We do this to see if the tail risk measures contribute to predicting the left-tail variation beyond its own lag and the VIX.

3.2.2. Economic tests

Our main economic test examines the ability of the various tail risk measures to predict future market excess returns. If tail risk is a relevant risk factor in the market, then the equity risk premium should include compensation for tail risk. Thus, when tail risk is high, the equity risk premium should be higher than in calm times when tail risk is low. Therefore, a measure of tail risk that is priced in the market should be able to positively forecast future market excess returns.

[10] The starting date, 1996, is dictated by the fact that both the OptionMetrics and TRTH databases do not begin before that date. The end date of our sample period is limited by the availability of data when we began this project.

[11] In Section A3.8 of the Online Appendix, we show that the results are qualitatively similar when also including overnight returns and with 5-minute returns.

[12] Additional implementation details, beyond those provided in the following paragraphs, can be found in Section A1.2 of the Online Appendix.

We use the following regression model to test whether the tail risk measures can predict returns:

$$R_{t+\Delta t} = a + b \cdot TRM_t + c \cdot Controls_t + \epsilon_{t+\Delta t}. \tag{12}$$

Since there are several variables that have been previously documented to predict future stock returns, we follow Bollerslev et al. (2009) and use several control variables in the vector $Controls_t$: the consumption, wealth, income ratio (CAY), the default spread (DFSP), the log dividend price ratio (log(D/P)), the stochastically detrended risk-free rate (RREL), the term spread (TMSP), and the variance risk premium (VRP).

3.2.3. Further methodological details

Throughout this paper, we report partial rather than "full" R^2s. We do this to emphasize the marginal contribution of each tail risk measure to the explanatory power of a model that may include multiple variables.[13] For the probit regressions, we obtain the average contribution from the dominance analysis as described in Azen and Budescu (2003). A predictor is dominant if it contributes more to the prediction than another. We report the partial R^2 measure based on general dominance, which is derived from the mean average incremental contribution at each level. For all other tests, we use the partial R^2 of Lindeman et al. (1980). This measure uses a simple unweighted average of the average contributions of different models of different sizes. In both cases, the partial R^2s add up to the total R^2.

For statistical inference, we rely on the wild bootstrap procedure of Rapach et al. (2013), which we describe in detail in Section A2.1 of the Online Appendix. The bootstrap preserves the contemporaneous correlation structure in the data, controls for the Stambaugh (1999) bias, and allows for conditional heteroskedasticity in stock returns. To account for autocorrelation, we base all t-statistics in the original and the bootstrap samples on robust Newey and West (1987) standard errors with 25 lags (252 lags for annual horizons), as recommended by Lazarus et al. (2018). As a robustness test, in Section A3.9 of the Online Appendix, we also present the results when using alternative inference methods, such as a block bootstrap. These are qualitatively similar.

Finally, to reduce the dimensionality in multiple regressions, we follow Bekaert et al. (2011) and use the general-to-specific PcGets search algorithm. This algorithm eliminates insignificant predictor variables in several steps. We outline the details of the procedure in Section A2.2 of the Online Appendix. As a robustness test, we also present the results of a jackknife procedure (Bekaert et al., 2011) in Section A3.5 of the Online Appendix.

4. Main analysis

4.1. Summary statistics

In Table 2, we present the summary statistics of the 16 different tail risk measures. We find that the main characteristics of the measures in our sample are consistent with those documented in the literature. The measures are very heterogeneous in their means and standard deviations. To account for this, and to make the results comparable across measures, we standardize all of them to have a mean of zero and a standard deviation of one for the following tests. Importantly, all but one of the measures have positive skewness and have significant excess kurtosis. This observation is consistent with interpreting the measures as capturing the risk of low-probability, high-impact events. As these events become increasingly likely, a tail risk measure should show a clear peak.

An important feature to distinguish between the different tail risk measures is their persistence. The (daily) first-order autocorrelation exceeds 0.99 for CJI, and $JumpRP$. It is also above 0.90 for $ATvol$, $BT11Q$, H_MRI, RIX, TLM, and $JumpRisk$.[14] The high autocorrelations imply that the tail risk measured by these variables is very persistent and changes little from day to day. On the other hand, $JUMP$ has an autocorrelation close to zero. This low autocorrelation would imply that the tail risk is highly variable even over short windows. Part of this is certainly due to the large noise in the estimation and the construction of the measure as a daily return. It seems more akin to the first difference in tail risk. The first-order autocorrelations of the remaining measures are all above 0.60, suggesting that tail risk is quite persistent by most measures.[15] However, whether low, medium, or high persistence is a desirable property of a good tail risk measure is an empirical question that we seek to answer in this section.

The Figures A2, A3, and A4 of the Online Appendix show the time series of the standardized tail risk measures. For better visualization, we average all daily observations of the tail risk measures over a month. For most measures, we observe significant peaks in October 2008, the height of the financial crisis immediately following the bankruptcy of Lehman Brothers. In particular, all of the Andersen–Bollerslev–Todorov measures show this peak. For some of the other measures, however, we do not observe it. For example, for H_MRI and λ_{Hill} there is a trough rather than a peak in the time series at that time. In addition, even for the Andersen–Bollerslev–Todorov measures, we observe substantially different behavior in the time series, with strong peaks in some measures that seem to be largely absent in others. This visual inspection of the tail risk measures thus suggests that they may not be very strongly correlated with each other and thus may contain quite different information.

[13] This is particularly important because our analyses also include control variables. In addition, for the analyses with multiple tail risk measures, we can assess the contribution of each individual variable.

[14] The autocorrelation of the λ_{Hill} measure in our sample is somewhat lower than that reported by Kelly and Jiang (2014) (0.75 vs. 0.93). However, this seems to depend on the sample period. For their full sample period (1963–2010), we also get an autocorrelation of 0.93.

[15] In statistical tests, we use bootstrap procedures (described in Section A2.1 of the Online Appendix) to ensure that the inference is robust to this persistence in the explanatory variables.

Table 2
Summary statistics.
This table shows the summary statistics of the tail risk measures considered. The definitions of the tail risk measure acronyms are given in Table 1. We assign the tail risk measures to different categories based on their underlying data. We present several time-series statistics. "*Mean*" denotes the time-series average, "*SD*" is the standard deviation. For the remainder of the paper, we standardize the tail risk measures to have a mean of zero and a standard deviation of one. "*Median*", "*Min*", and "*Max*" denote the median, the lowest, and the highest values of the measures, respectively. "*Skewness*" and "*Kurtosis*" denote the skewness and kurtosis of the distributions of the measures. Finally, "$AR(1)$" shows the first-order autocorrelation of the measures. All measures except for RIX, $BT14Q$, λ_{Hill}, and LE are available at the daily frequency. $BT14Q$ is weekly, λ_{Hill} and RIX are monthly, and LE is quarterly. The values of $BT11P$ are multiplied by 10^9.

	Mean	*SD*	*Median*	*Min*	*Max*	*Skewness*	*Kurtosis*	$AR(1)$
Group A — Option-Implied Measures								
$ATprob$	0.8237	0.7408	0.5976	0.0754	10.0623	3.4242	23.768	0.7341
$ATvol$	0.0790	0.0359	0.0706	0.0021	0.3985	2.4111	13.425	0.9317
$BT11Q$	0.0308	0.0506	0.0146	0.0001	0.7945	5.1971	42.812	0.9314
$BT14Q$	0.0679	0.0244	0.0635	0.0351	0.3155	3.8270	29.511	0.8319
$BTX15Q$	0.0090	0.0044	0.0079	0.0023	0.0430	2.5132	14.374	0.6192
H_MRI	−0.0000	1.3923	−0.3127	−2.4190	9.4042	2.2517	10.564	0.9650
RIX	0.0204	0.0069	0.0015	0.0001	0.1234	8.5753	102.36	0.9529
TLM	0.0015	0.0059	0.0192	0.0106	0.0588	1.5628	7.309	0.9761
Group B — Stock-Return-Based Measures								
$BT11P$	0.0626	0.0035	0.0008	−0.0058	0.1218	15.295	364.68	0.8521
CJI	0.1562	0.0621	0.0458	0.0028	0.4575	3.1576	15.675	0.9979
$JumpRisk$	1.0029	0.0885	0.1292	0.0431	0.6471	1.7382	6.8921	0.9720
$JumpRP$	0.4426	0.3018	0.9184	0.4580	2.1240	1.2681	4.3748	0.9924
λ_{Hill}	−0.0855	0.0275	0.4450	0.3447	0.5054	−0.5789	3.8619	0.7538
Group C — Option-Return-Based Measures								
$ADBear$	−0.0014	0.7180	−0.2841	−0.9962	9.9967	2.8659	20.274	0.6953
$JUMP$	0.0885	0.0455	−0.0073	−0.7931	1.1417	6.4229	159.62	−0.0480
Group D — Macroeconomic Measures								
LE	0.8237	0.1690	0.0331	−0.0266	1.0478	3.5552	17.734	0.7904

Table 3
Correlations.
This table shows the time-series correlations between the tail risk measures considered. The definitions of the tail risk measure acronyms are given in Table 1. To ensure comparability of the correlations, we use a daily sample with constant extrapolation. The last row shows the correlations of the tail risk measures with the VIX.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Group A — Option-Implied Measures																
(1) $ATprob$		0.41	0.61	0.47	0.42	−0.36	0.48	0.67	0.37	0.53	0.47	0.13	−0.19	0.26	0.06	0.25
(2) $ATvol$			0.70	0.64	0.63	−0.03	0.70	0.78	0.47	0.63	0.42	0.13	−0.12	0.14	0.06	0.45
(3) $BT11Q$				0.80	0.67	−0.31	0.91	0.83	0.73	0.78	0.56	0.08	−0.26	0.26	0.09	0.47
(4) $BT14Q$					0.85	−0.38	0.79	0.76	0.62	0.79	0.46	0.24	−0.38	0.08	0.03	0.52
(5) $BTX15Q$						−0.17	0.65	0.69	0.49	0.60	0.40	0.13	−0.24	0.12	0.04	0.38
(6) H_MRI							−0.23	−0.37	−0.18	−0.38	−0.22	−0.35	0.44	−0.17	−0.05	−0.26
(7) RIX								0.76	0.74	0.73	0.49	0.02	−0.21	0.17	0.07	0.45
(8) TLM									0.53	0.81	0.60	0.24	−0.27	0.24	0.07	0.51
Group B — Stock-Return-Based Measures																
(9) $BT11P$										0.52	0.47	−0.05	−0.16	0.16	0.07	0.32
(10) CJI											0.49	0.35	−0.36	0.06	0.00	0.76
(11) $JumpRisk$												−0.33	−0.27	0.31	0.15	0.14
(12) $JumpRP$													−0.18	0.00	−0.04	0.48
(13) λ_{Hill}														−0.04	−0.01	−0.22
Group C — Option-Return-Based Measures																
(14) $ADBear$															0.20	0.01
(15) $JUMP$																−0.00
Group D — Macroeconomic Measures																
(16) LE																
ρ_{VIX}	0.69	0.67	0.87	0.81	0.65	−0.48	0.79	0.93	0.59	0.87	0.65	0.25	−0.36	0.26	0.08	0.53

Table 3 shows the correlations of the tail risk measures. Consistent with the time-series plots, we find that the correlations are indeed much lower than one would expect from different measures that are broadly designed to capture essentially the same underlying risk. In particular, the correlation between measures across different groups is typically low.

Table 4
Principal components.
This table reports the results of a principal component analysis (PCA) of the standardized tail risk measures. The definitions of the tail risk measure acronyms are given in Table 1. We use a daily sample of the tail risk measures, with constant extrapolation. We show the first two PCs among all measures and within the different subgroups. The column "$CumVar$" displays the cumulative variance explained by the PCs. The last column shows the correlation of each PC with the VIX.

	$AT prob$	$AT vol$	$BT11Q$	$BT14Q$	$BTX15Q$	H_MRI	RIX	TLM	$BT11P$	CJI	$JumpRisk$	$JumpRP$	λ_{Hill}	$ADBear$	$JUMP$	LE	$CumVar$	ρ_{VIX}
Full sample																		
PC1	0.24	0.28	0.34	0.33	0.29	−0.16	0.32	0.34	0.26	0.33	0.24	0.07	−0.13	0.09	0.03	0.22	0.46	0.95
PC2	0.08	0.06	0.11	−0.07	0.02	0.24	0.12	0.01	0.18	−0.22	0.35	−0.62	0.20	0.27	0.21	−0.39	0.58	−0.04
Group A — Option-Implied Measures																		
PC1	0.29	0.35	0.41	0.39	0.36	−0.17	0.39	0.40									0.65	0.93
PC2	0.35	−0.35	−0.02	−0.02	−0.17	−0.84	−0.14	0.04									0.79	0.22
Group B — Stock-Return-Based Measures																		
PC1									0.51	0.57	0.52	0.04	−0.37				0.44	0.87
PC2									0.15	−0.28	0.40	−0.82	0.25				0.70	−0.15
Group C — Option-Return-Based Measures																		
PC1														0.71	0.71		0.60	0.21
PC2														0.71	−0.71		1.00	0.14

Among the option-implied measures, we generally observe the highest correlations. For example, $BT11Q$ has correlations of 0.91 and 0.83 with RIX and TLM, respectively. On the other hand, H_MRI is negatively correlated with all other option-implied measures.[16] For the stock-return-based measures, the correlations are generally lower. Interestingly, the correlations of CJI with most option-implied measures are also relatively high.[17] The correlations of the option-return-based measures with all the others are rather low. It is interesting to note that the only macroeconomic measure in our dataset, even though it is measured at a low frequency and is not directly based on stock or option data, is quite highly correlated with several of the other measures. For example, the correlations of LE with $BT14Q$, TLM, and CJI are all above 0.5.

Table 3 also shows the correlations of the tail risk measures with the VIX, a simple measure of the current conditional volatility. It would be natural to find that there is some correlation of tail risk with volatility. However, the tail risk measures should capture the risk of ending up in particularly bad states of the world on top of the "normal" daily variation. We find that many tail risk measures have high correlations with the VIX, e.g., $BT11Q$ (0.87), $BTX14Q$ (0.81), RIX (0.79), TLM (0.93), and CJI (0.87). These high correlations imply that the tail risk measures may provide only little additional insights into tail risk beyond what is captured by the VIX. To account for this, we control for volatility in our empirical tests.

In Table 4 we present a principal component (PC) analysis of the tail risk measures. We compute the first two PCs among all measures, as well as the respective first two PCs within each group of measures. Consistent with our previous results in this section, the commonality among the different measures is rather low. The first PC of all measures can only explain 46% of the variation. Together with the second PC, the proportion increases to only 58%. Thus, it is difficult to capture the information contained in the different tail risk measures with just a few PCs.

The largest loadings of the first PC are on $BT11Q$ (0.34), $BT14Q$ (0.33), RIX (0.32), TLM (0.34), and CJI (0.33). Thus, these measures appear to be the most representative of the common variation in the tail risk measures. Within the subgroups, the degree of commonality is somewhat greater. The first two PCs in each subgroup capture at least 70% of the variation in the tail risk measures. The highest loadings of the first PC among the option-implied measures are again on $BT11Q$ (0.41), $BT14Q$ (0.39), RIX (0.39), and TLM (0.40). Among the stock-return-based measures, the highest PC loadings are on $BT11P$ (0.51), CJI (0.57), and $JumpRisk$ (0.52). However, the ability to capture common variation in the tail risk measures may be a misguided objective for selecting a particular measure. Rather, we should judge the measures based on their ability to predict future tail events and capture risk premia.

4.2. Statistical tests

We start with the statistical tests. We use three different forecast horizons: (i) one day ($Daily$), (ii) one week ($Weekly$), and (iii) one month ($Monthly$).[18] We do not consider longer horizons for this analysis because it seems unrealistic to be able to predict realized tail events or variation in the distant future. Starting with the probit model, we examine how well the tail risk measures

[16] This is consistent with Gormsen and Jensen (2022), who show that H_MRI tends to be low when volatility is high.

[17] λ_{Hill} has negative correlations with almost all other measures except H_MRI. The latter observation is consistent with Kelly and Jiang (2014), who show that λ_{Hill} loads negatively on skewness and positively on kurtosis, as does (by construction) H_MRI.

[18] Four of the measures are not available at a daily frequency. For these measures, we constantly extrapolate the most recent weekly, monthly, or quarterly observation until new information becomes available.

perform in predicting future tail events. For each measure and forecast horizon, we run separate regressions of the (horizon-specific) dummy variables on the lagged standardized tail risk measures and the VIX.

In Figure A1 of the Online Appendix, we illustrate the timing of realized left-tail events. We plot these separately for the daily, weekly, and monthly horizons. There is some clustering of realized left-tail events during specific crisis periods, such as the bursting of the dot-com bubble and the 2007–2008 financial crisis. Interestingly, we find that not all daily left-tail realizations lead to weekly or monthly left-tail observations. Similarly, some weekly and monthly tail events occur without being driven by a single or multiple daily tail observations.

The probit regression results are shown in Table 5. At the daily level, we find that many tail risk measures have some predictive power for future tail events. The three measures that yield the highest partial R^2s with statistically significant positive slope coefficients are, in order, $JumpRisk$, $BT11P$, and $BT11Q$. At the weekly horizon, the performance of the measures becomes somewhat weaker. Only $JumpRisk$, $BT11Q$, and $ADBear$ (ordered by partial R^2) are also significant positive predictors of future tail events at this horizon. At the monthly horizon, $JumpRisk$, λ_{Hill}, and $JUMP$ are significant positive predictors of future tail events.

It is important to note that we require the tail risk measures to be positively related to future tail events. That is, a high tail risk measure should be associated with a higher probability of a future tail event. For example, at the monthly horizons, H_MRI and RIX actually yield slope coefficients that are significantly negative. Such results are likely inconsistent with being a good tail risk measure.[19]

In addition to the individual tail risk measures, we also repeat the probit regressions with the first PC of all measures and among the different subgroups. We find that the first PC of all measures and that only using stock-return-based measures significantly predict tail events at the daily frequency. At the weekly horizon none of the PCs significantly predicts future tail events and at the monthly horizon only the PC derived from the options-return-based measures does so.

We also report the results of multiple probit regressions in Table 6. For each horizon, we select the measures with PcGets. For the daily forecast horizon, the algorithm selects only $JumpRisk$, and for the weekly horizon, it selects only $BT11Q$. Both yield significant positive slope coefficients. At the monthly forecast horizon, PcGets selects six measures, of which $BTX15Q$, $BT11P$, $JumpRisk$, and λ_{Hill} yield significant positive slope coefficients.

Next, we move from a left-hand-side variable that only indicates whether or not there is a tail event to one that also contains information about the variation it causes. That is, we predict the realized left-tail variation (also standardized to have a mean of zero and a standard deviation of one). We present the results in Table 7. We examine the same time horizons as before and control for the lagged left-tail variation measure and the VIX.

Starting with the daily frequency, we find that $BT11Q$ turns out to be the best predictor. It has the largest slope coefficient and the highest partial R^2. The slope coefficient of 0.41 indicates that, all else being equal, an one standard deviation increase in $BT11Q$ increases the left-tail variation by 0.41 standard deviations. The measures H_MRI, RIX, $BT11P$, $JumpRisk$, $ADBear$, and $JUMP$ are also significant positive predictors of future left-tail variation at the daily frequency. At the weekly horizon, $BT11Q$ also performs best, while $BT14Q$, H_MRI, RIX, $BT11P$, and $JumpRisk$ also show some predictability for future left-tail variation. Finally, at the monthly horizon, only $JumpRisk$, $ADBear$, and LE yield significant positive predictive slope coefficients.

Turning to the PCs, we find that only the first PC of the stock-return-based measures has predictive power for future left-tail variation at all horizons. The other first PCs have predictive power for two of the three horizons.

In Table 8 we present the results of the multiple regressions to predict the future left-tail variation. Here, $BT11Q$ turns out to be the best predictor of realized left-tail variation for the daily and weekly horizons, for which it is the only measure selected. At the monthly horizon, however, the PcGets algorithm eliminates all tail risk measures. The lagged left-tail variation measure is the only one selected (see Table A19 of the Online Appendix).

Thus, overall, the statistical analysis places $BT11Q$ and $JumpRisk$ together in pole position in the horse race of tail risk measures. Both perform well in predicting both future tail events and future left-tail variation. $JumpRisk$ is best at predicting the former, while $BT11Q$ is particularly good at the latter.

4.3. Economic tests

For the last of our main tests, we turn to whether tail risk is priced in the market. While some of the tail risk measures are designed for slightly different purposes, most of the studies seem to argue that their tail risk measure is priced. Therefore, this analysis is equitable. We examine whether the tail risk measures have predictive power for future market excess returns over various time horizons. For this analysis, we include an annual forecast horizon in addition to the daily, weekly, and monthly horizons. We do this for two reasons. First, it is common in the return predictability literature to also consider longer horizons. Second, long-horizon returns may also be affected by tail-risk expectations, while for the statistical tests we would require observing actual tail event realizations, which are rare at long horizons. In the analysis we are interested in the marginal effect of the tail risk measures, controlling for several other predictor variables (see the details in Section 3.2.2). We present the results in Table 9. As Kelly and Jiang (2014), we use annualized returns in percentage points.

Looking first at the top performers from the statistical analysis, we find that $BT11Q$ again performs very well, while $JumpRisk$ clearly does not. $BT11Q$ significantly predicts future market excess returns at the daily, weekly, and annual horizons. Furthermore,

[19] The negative predictive coefficient may be due to investors' subjective beliefs about tail risk, which may show overreaction or underreaction (Baron and Xiong, 2017).

Table 5
Prediction of tail events.
This table reports the coefficients of the predictive probit regressions. We run single probit regressions of a dummy variable for each lagged tail risk measure:

$$D_{t+\Delta t} = a + b \cdot TRM_t + c \cdot VIX_t + \epsilon_{t+\Delta t}.$$

$D_{t+\Delta t}$ equals 1 if the realized market excess return falls below the threshold defined by minus two times the current conditional volatility, and 0 otherwise. The conditional volatility is defined as the level of the VIX at the end of the previous day (VIX_t). TRM_t is the current observation of a tail risk measure. We use three different forecast horizons Δt: (i) one day ($Daily$), (ii) one week ($Weekly$), and (iii) one month ($Monthly$). In parentheses, we present robust Newey and West (1987) standard errors with 25 lags. The R^2 columns show the partial McFadden R^2s, obtained by dominance analysis (in percentage points) "$PCOneAll$", "$PCOneOption$", "$PCOneStReturn$", and "$PCOneOpReturn$" denote the first PCs of all measures, option-implied, stock-return-based, and option-return-based tail risk measures, respectively. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2
Group A — Option-Implied Measures						
$ATprob$	−0.04	0.05	−0.06	0.07	−0.08	0.10
	(0.06)		(0.09)		(0.10)	
$ATvol$	0.01	0.09	−0.00	0.07	−0.05	0.07
	(0.07)		(0.07)		(0.12)	
$BT11Q$	0.10*	0.37	0.13*	0.52	−0.07	0.14
	(0.06)		(0.09)		(0.12)	
$BT14Q$	−0.06	0.09	−0.05	0.08	0.07	0.34
	(0.06)		(0.08)		(0.09)	
$BTX15Q$	0.02	0.12	0.04	0.21	0.07	0.38
	(0.06)		(0.08)		(0.11)	
H_MRI	0.04	0.04	−0.05	0.19	−0.29*	1.91
	(0.07)		(0.07)		(0.18)	
RIX	−0.02	0.08	−0.01	0.08	−0.87***	1.32
	(0.05)		(0.08)		(0.36)	
TLM	0.03	0.15	−0.02	0.12	−0.19	0.23
	(0.19)		(0.29)		(0.38)	
Group B — Stock-Return-Based Measures						
$BT11P$	0.07***	0.93	0.02	0.14	0.04	0.26
	(0.02)		(0.03)		(0.03)	
CJI	−0.09	0.11	−0.20	0.28	−0.03	0.13
	(0.10)		(0.17)		(0.10)	
$JumpRisk$	0.20***	1.93	0.18**	1.46	0.28**	3.68
	(0.06)		(0.09)		(0.14)	
$JumpRP$	−0.18***	1.48	−0.18**	1.55	−0.07	0.27
	(0.07)		(0.10)		(0.15)	
λ_{Hill}	0.02	0.02	−0.02	0.07	0.11**	0.39
	(0.06)		(0.08)		(0.06)	
Group C — Option-Return-Based Measures						
$ADBear$	0.05*	0.21	0.06*	0.37	0.02	0.06
	(0.03)		(0.05)		(0.05)	
$JUMP$	−0.06	0.19	−0.05*	0.13	0.05***	0.29
	(0.06)		(0.04)		(0.02)	
Group D — Macroeconomic Measures						
LE	0.01	0.07	−0.07	0.10	0.05	0.31
	(0.05)		(0.07)		(0.06)	
$PCOneAll$	0.17*	0.37	0.06	0.17	0.11	0.25
	(0.10)		(0.16)		(0.17)	
$PCOneOption$	−0.01	0.13	0.04	0.16	−0.08	0.15
	(0.12)		(0.17)		(0.32)	
$PCOneStReturn$	0.20***	1.10	0.11	0.39	0.18	0.84
	(0.06)		(0.10)		(0.14)	
$PCOneOpReturn$	−0.00	0.01	0.02	0.04	0.05**	0.26
	(0.05)		(0.04)		(0.03)	
$Controls$	Yes		Yes		Yes	

its partial R^2s are in the top 4 of the individual tail risk measures for each of these horizons. For example, at the daily frequency, a one-standard-deviation increase in $BT11Q$ implies an increase in the annualized market excess return by 28.67 percentage points, all else equal. The partial R^2 is 0.45%. On the other hand, $JumpRisk$ cannot predict future market excess returns for any of the horizons examined.

Apart from $BT11Q$, $ATprob$, RIX, TLM, and $ADBear$ also have good and consistent predictive power for future market excess returns. In fact, $ADBear$ and TLM have significant positive predictive power at all the forecast horizons considered. It is also clear that option-based tail risk measures have significantly better predictive abilities than stock-return-based or macroeconomic

Table 6
Multiple prediction of tail events.
This table reports the coefficients of the predictive probit regressions. We run multiple probit regressions of a dummy variable on lagged tail risk measures:

$$D_{t+\Delta t} = a + b \cdot TRM_t + c \cdot VIX_t + \epsilon_{t+\Delta t}.$$

$D_{t+\Delta t}$ equals 1 if the realized market excess return falls below the threshold defined by minus two times the current conditional volatility, and 0 otherwise. The conditional volatility is defined as the level of the VIX at the end of the previous day (VIX_t). TRM_t is a vector of the current observations of the tail risk measures. We use four different forecast horizons Δt: (i) one day ($Daily$), (ii) one week ($Weekly$), and (iii) one month ($Monthly$). For each forecast horizon, we first perform a variable selection based on the PcGets algorithm. A blank space indicates that a variable was not selected. In parentheses, we present robust Newey and West (1987) standard errors with 25 lags. The R^2 columns present the partial McFadden R^2s, obtained by dominance analysis (in percentage points). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2
Group A — Option-Implied Measures						
$ATprob$						
$ATvol$						
$BT11Q$			0.07**	0.59		
			(0.04)			
$BT14Q$						
$BTX15Q$					0.38***	1.47
					(0.15)	
H_MRI					−0.37**	2.33
					(0.16)	
RIX					−1.90***	3.56
					(0.53)	
TLM						
Group B — Stock-Return-Based Measures						
$BT11P$					0.39***	1.05
					(0.14)	
CJI						
$JumpRisk$	0.15***	2.02			0.36**	4.22
	(0.04)				(0.16)	
$JumpRP$						
λ_{Hill}					0.26***	1.39
					(0.08)	
Group C — Option-Return-Based Measures						
$ADBear$						
$JUMP$						
Group D — Macroeconomic Measures						
LE						
$Controls$	Yes		Yes		Yes	

measures. The only notable stock-return-based measure is λ_{Hill}, whose predictive power seems to start only at the annual forecast horizon. However, with a partial R^2 of 7.03%, the long-term predictive ability of the measure is very strong.[20]

The PCs also perform well in predicting market excess returns. The first PC of all measures and that of the option-implied measures predict returns at the daily, weekly, and annual horizons. The first PC from the option-return-based measures predicts returns for all horizons.

The results for the multiple return forecasts are shown in Table 10 and more or less confirm our previous results. The selected measures with the highest impact at the daily and weekly forecast horizons are RIX and $ADBear$. At the monthly forecast horizon, it is $BT11Q$, and at the annual horizon $ATvol$ yields the highest partial R^2 associated with a significantly positive selected measure.

We also conduct a further evaluation of the economic value of the tail risk measures. First, we perform an analysis where we regress the future return not only on the tail risk measure, but also on a dummy interaction of the tail risk measure with our tail dummy variable from Eq. (7). This analysis allows us to test whether the size of the tail event is also predictable. We present the results in Tables A1 and A2 of the Online Appendix. We find that $BT11Q$, among others, again performs well for all horizons in this more granular analysis.

[20] Kelly and Jiang (2014) also report a good performance of λ_{Hill} for the 3- and 5-year forecast horizons in their 1963–2010 sample period.

Table 7
Predictability of left-tail variation.
This table reports the coefficients of a predictive regression for future left-tail variation. We run single regressions of the standardized realized left-tail variation for each lagged tail risk measure:

$$LTV^{\mathbb{P}}_{t+\Delta t} = a + b \cdot TRM_t + c \cdot LTV^{\mathbb{P}}_t + d \cdot VIX_t + \epsilon_{t+\Delta t}.$$

TRM_t is the current observation of a tail risk measure. We control for the lagged left-tail variation $LTV^{\mathbb{P}}_t$ and the current level of the VIX (VIX_t). We use three different forecast horizons Δt: (i) one day ($Daily$), (ii) one week ($Weekly$), and (iii) one month ($Monthly$). In parentheses, we present robust Newey and West (1987) standard errors with 25 lags. Statistical inference is based on the wild bootstrap of Rapach et al. (2013). The R^2 columns show the Lindeman et al. (1980) partial R^2 of each tail risk measure (in percentage points). "$PCOneAll$", "$PCOneOption$", "$PCOneStReturn$", and "$PCOneOpReturn$" denote the first PCs of all measures, option-implied, stock-return-based, and option-return-based tail risk measures, respectively. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2
Group A — Option-Implied Measures						
$ATprob$	0.01	0.88	−0.04	1.92	0.03	1.94
	(0.06)		(0.07)		(0.05)	
$ATvol$	0.05	1.23	0.04	2.84	−0.06	2.53
	(0.04)		(0.05)		(0.08)	
$BT11Q$	0.41**	4.83	0.46**	10.27	0.09	6.82
	(0.20)		(0.23)		(0.12)	
$BT14Q$	−0.00	1.16	0.08*	4.16	−0.11	4.11
	(0.04)		(0.05)		(0.11)	
$BTX15Q$	−0.01	0.72	0.03	2.60	−0.03	2.71
	(0.03)		(0.04)		(0.08)	
H_MRI	0.05**	0.31	0.04*	0.84	−0.03	1.02
	(0.02)		(0.03)		(0.04)	
RIX	0.20***	3.17	0.21**	6.94	−0.22	5.36
	(0.05)		(0.09)		(0.24)	
TLM	−0.08	1.44	−0.13	3.84	−0.13	3.79
	(0.09)		(0.15)		(0.20)	
Group B — Stock-Return-Based Measures						
$BT11P$	0.16**	2.38	0.23**	8.07	0.05	6.90
	(0.04)		(0.06)		(0.09)	
CJI	−0.08	1.15	−0.02	3.44	−0.02	3.62
	(0.06)		(0.07)		(0.06)	
$JumpRisk$	0.12**	2.08	0.17**	6.20	0.19*	7.17
	(0.06)		(0.10)		(0.15)	
$JumpRP$	−0.09**	0.45	−0.12***	1.05	−0.07*	0.82
	(0.05)		(0.06)		(0.05)	
λ_{Hill}	0.02	0.13	0.02	0.39	0.00	0.43
	(0.02)		(0.02)		(0.02)	
Group C — Option-Return-Based Measures						
$ADBear$	0.04*	0.42	0.04	0.91	0.07**	1.00
	(0.03)		(0.04)		(0.04)	
$JUMP$	0.05**	0.29	0.01	0.09	0.02	0.08
	(0.03)		(0.01)		(0.02)	
Group D — Macroeconomic Measures						
LE	−0.00	0.49	0.03	1.64	0.08**	2.48
	(0.02)		(0.03)		(0.05)	
$PCOneAll$	0.31***	2.54	0.43***	7.20	0.05	6.39
	(0.11)		(0.14)		(0.24)	
$PCOneOption$	0.21***	2.33	0.25**	6.05	−0.17	5.27
	(0.08)		(0.13)		(0.31)	
$PCOneStReturn$	0.17***	2.26	0.29***	7.72	0.22*	7.90
	(0.08)		(0.10)		(0.15)	
$PCOneOpReturn$	0.06**	0.58	0.03	0.66	0.06***	0.68
	(0.04)		(0.03)		(0.03)	
$Controls$	Yes		Yes		Yes	

Second, we analyze the impact of tail risk measures on the cross-section of stock returns. We first estimate the sensitivities of the stocks to tail risk using a rolling historical window and then sort the stocks into portfolios based on these sensitivities. We present the main results for value-weighted portfolios in Table A3 of the Online Appendix. We find that for $BT11Q$, the difference between the high and low portfolios is −11.87% per year on average. These results are consistent with a mechanism in which stocks that perform well following a high tail risk observation are highly desirable to investors, and trade at a premium. The vast majority of the other tail risk measures do not yield significant negative high–low portfolio excess returns.

Third, we analyze the impact of tail risk on industrial production growth. The results are presented in Table A4 of the Online Appendix. We find that many measures, including $BT11Q$, significantly negatively predict future industrial production growth at

Table 8
Multiple predictability of left-tail variation.
This table reports the coefficients of a predictive regression for future left-tail variation. We run multiple regressions of the realized left-tail variation on the lagged tail risk measures:

$$LTV^{\mathbb{P}}_{t+\Delta t} = a + b \cdot TRM_t + c \cdot LTV^{\mathbb{P}}_t + d \cdot VIX_t + \epsilon_{t+\Delta t}.$$

TRM_t is a vector of the current observations of the tail risk measures. We control for the lagged left-tail variation $LTV^{\mathbb{P}}_t$ and the current level of the VIX (VIX_t). We use three different forecast horizons Δt: (i) one day ($Daily$), (ii) one week ($Weekly$), and (iii) one month ($Monthly$). For each forecast horizon, we first perform a variable selection based on the PcGets algorithm. A blank space indicates that a variable was not selected. In parentheses, we present robust Newey and West (1987) standard errors with 25 lags. Statistical inference is based on the wild bootstrap of Rapach et al. (2013). The R^2 columns show the Lindeman et al. (1980) partial R^2 of each tail risk measure (in percentage points). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2
Group A — Option-Implied Measures						
$ATprob$						
$ATvol$						
$BT11Q$	0.25***	6.09	0.43***	18.55		
	(0.08)		(0.15)			
$BT14Q$						
$BTX15Q$						
H_MRI						
RIX						
TLM						
Group B — Stock-Return-Based Measures						
$BT11P$						
CJI						
$JumpRisk$						
$JumpRP$						
λ_{Hill}						
Group C — Option-Return-Based Measures						
$ADBear$						
$JUMP$						
Group D — Macroeconomic Measures						
LE						
$Controls$	Yes		Yes		Yes	

both the monthly and annual horizons. A more detailed discussion and the full results of these further analyses can be found in Sections A3.1, A3.2, and A3.3 of the Online Appendix.

5. Different tail risk measure definitions

An important difference between $BT11Q$ and other measures with similar basic setups ($ATprob$, $ATvol$, $BT14Q$, $BTX15Q$, and TLM) is that it uses a fixed target moneyness while the others rely on a time-varying cutoff. Thus, an important driver of the results may be which part of the tail the measures capture.

It is therefore worth examining the robustness of our main results to some variations in the option-implied tail risk measures. The main $BT11Q$ measure uses out-of-the-money put options with a fixed moneyness level (defined as the strike price over the current futures price) of $K/F_{t,\tau} = 0.9$ (see Section A1 of the Online Appendix for more details). Therefore, as a first check, we vary this fixed threshold and also consider $K/F_{t,\tau} = 0.8875$ and $K/F_{t,\tau} = 0.9125$. Second, we also consider a version of the $BT11Q$ measure based on time-varying moneyness with $K/F_{t,\tau} = e^{-2.5\sigma_{ATM,\tau}\sqrt{\tau}}$, where $\sigma_{ATM,\tau}$ is the at-the-money option-implied volatility and τ is the option's time-to-maturity. We denote this tail risk measure as $BT11Q_{var}$, while the others have their target moneyness levels in the subscript. Finally, we also examine the measures $ATprob$, $ATvol$, $BT14Q$, $BTX15Q$, and TLM with the fixed moneyness cutoff of $K/F_{t,\tau} = 0.9$. We mark these measures with "fixed" in the subscript.

We present the results in Table 11. For $BT11Q_{0.8875}$ and $BT11Q_{0.9125}$, the correlations with $BT11Q$ are almost perfect, with 1.00 when rounded to two decimal places. Correspondingly, the results for the two alternative fixed moneyness levels are qualitatively

Table 9
Return predictability.
This table reports the coefficients of a return predictability regression. We run single regressions of the market excess return for each lagged tail risk measure:

$$R_{t+\Delta t} = a + b \cdot TRM_t + c \cdot Controls_t + \epsilon_{t+\Delta t}.$$

$R_{t+\Delta t}$ is the excess return over the period Δt. TRM_t is the current observation of a tail risk measure. We use the following control variables ($Controls_t$): the consumption, wealth, income ratio, the default spread, the log dividend price ratio, the stochastically detrended risk-free rate, the term spread, and the variance risk premium. We use four different forecast horizons Δt: (i) one day ($Daily$), (ii) one week ($Weekly$), (iii) one month ($Monthly$), and (iv) one year ($Annual$). In parentheses, we present robust Newey and West (1987) standard errors with lag length chosen to be the maximum of 25 and the number of overlapping observations. Statistical inference is based on the wild bootstrap of Rapach et al. (2013). The R^2 columns show the Lindeman et al. (1980) partial R^2 of each tail risk measure (in percentage points). "$PCOneAll$", "$PCOneOption$", "$PCOneStReturn$", and "$PCOneOpReturn$" denote the first PCs of all measures, option-implied, stock-return-based, and option-return-based tail risk measures, respectively. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2	$Annual$	R^2
Group A — Option-Implied Measures								
$ATprob$	17.50**	0.25	9.50**	0.45	4.63**	0.47	0.94	0.29
	(8.80)		(4.69)		(2.03)		(0.92)	
$ATvol$	4.75	0.01	1.23	0.01	1.02	0.05	3.80**	2.79
	(5.38)		(4.68)		(3.20)		(1.72)	
$BT11Q$	28.67***	0.45	12.14*	0.38	4.39	0.21	3.33***	2.24
	(7.48)		(6.96)		(4.73)		(0.96)	
$BT14Q$	10.75	0.03	8.07	0.08	1.82	0.18	2.33*	0.66
	(8.20)		(5.69)		(4.82)		(1.23)	
$BTX15Q$	6.48	0.02	3.12	0.02	−0.66	0.16	2.50*	1.21
	(6.01)		(4.54)		(3.22)		(1.36)	
H_MRI	−9.28***	0.03	−4.55*	0.04	−1.98	0.10	2.55	2.02
	(3.99)		(3.37)		(3.00)		(2.11)	
RIX	30.52***	0.51	14.42*	0.53	2.30	0.08	3.39*	2.60
	(6.55)		(5.95)		(4.56)		(1.16)	
TLM	18.25***	0.16	11.10**	0.32	6.47*	0.43	3.72***	1.80
	(5.63)		(4.99)		(3.92)		(1.20)	
Group B — Stock-Return-Based Measures								
$BT11P$	18.39	0.23	3.68	0.04	−2.04	0.22	0.65	0.39
	(8.76)		(7.11)		(3.89)		(0.39)	
CJI	13.79*	0.03	10.51	0.11	8.00	0.34	3.82*	1.56
	(8.97)		(7.87)		(6.11)		(2.34)	
$JumpRisk$	6.95	0.04	3.03	0.04	0.74	0.01	−0.78	0.64
	(5.40)		(5.86)		(4.26)		(1.58)	
$JumpRP$	1.03	0.01	3.80	0.05	4.00	0.20	1.10	0.60
	(5.87)		(6.09)		(4.66)		(2.21)	
λ_{Hill}	−4.63	0.01	−3.99	0.07	−4.79*	0.28	2.65**	7.03
	(3.79)		(3.65)		(2.96)		(1.25)	
Group C — Option-Return-Based Measures								
$ADBear$	20.97***	0.46	13.49***	1.05	3.81**	0.36	0.60*	0.05
	(4.73)		(3.34)		(1.61)		(0.43)	
$JUMP$	7.21	0.06	5.89***	0.24	0.92	0.04	0.13	0.02
	(7.10)		(1.64)		(0.73)		(0.12)	
Group D — Macroeconomic Measures								
LE	−7.97	0.02	−10.26	0.16	−3.96	0.33	−5.70**	2.48
	(7.90)		(7.81)		(5.79)		(2.25)	
$PCOneAll$	23.60***	0.21	11.14*	0.23	3.13	0.12	2.61**	1.34
	(7.66)		(6.87)		(5.43)		(1.15)	
$PCOneOption$	22.09***	0.21	11.11*	0.25	3.73	0.12	3.58***	1.87
	(7.04)		(5.92)		(4.68)		(1.21)	
$PCOneStReturn$	16.04**	0.10	6.23	0.06	2.09	0.10	−0.28	0.51
	(6.99)		(7.52)		(5.76)		(1.21)	
$PCOneOpReturn$	18.27***	0.36	12.47***	0.95	3.02**	0.26	0.47*	0.04
	(5.55)		(2.75)		(1.38)		(0.32)	
$Controls$	Yes		Yes		Yes		Yes	

similar to those for the main $BT11Q$ measure. Thus, which exact fixed part of the tail the measure captures does not seem to have a primary impact on the results.

The choice of fixed versus time-varying target or cutoff moneyness has additional implications. On the one hand, time-varying target or cutoff moneyness helps to isolate tail risk from diffusive risk by pushing the threshold further out during periods of high volatility. However, given the negligible impact of diffusive volatility on short-dated out-of-the-money options, this is unlikely to be a significant effect. On the other hand, it is possible that investors only care about the magnitude of the return and do not adjust it for the prevailing volatility. Thus, a return of −10% would be equally worrisome in both high- and low-volatility conditions. For these investors, a measure with a fixed target is obviously preferable.

Table 10
Multiple return predictability.
This table reports the coefficients of a return predictability regression. We run multiple regressions of the market excess return on lagged tail risk measures:

$$R_{t+\Delta t} = a + b \cdot TRM_t + c \cdot Controls_t + \epsilon_{t+\Delta t}.$$

$R_{t+\Delta t}$ is the excess return over the period Δt. TRM_t is a vector of the current observations of the tail risk measures. We use the following control variables ($Controls_t$): the consumption, wealth, income ratio, the default spread, the log dividend price ratio, the stochastically detrended risk-free rate, the term spread, and the variance risk premium. We use four different forecast horizons Δt: (i) one day ($Daily$), (ii) one week ($Weekly$), (iii) one month ($Monthly$), and (iv) one year ($Annual$). For each forecast horizon, we first perform a variable selection based on the PcGets selection algorithm. A blank space indicates that a variable was not selected. In parentheses, we present robust Newey and West (1987) standard errors with lag length chosen to be the maximum of 25 and the number of overlapping observations. Statistical inference is based on the wild bootstrap of Rapach et al. (2013). The R^2 columns show the Lindeman et al. (1980) partial R^2 of each tail risk measure (in percentage points). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Daily	R^2	*Weekly*	R^2	*Monthly*	R^2	*Annual*	R^2
Group A — Option-Implied Measures								
$ATprob$								
$ATvol$	−25.98***	0.16	−16.78***	0.34			3.92**	4.28
	(10.11)		(6.39)				(1.78)	
$BT11Q$					12.75***	0.82		
					(4.38)			
$BT14Q$								
$BTX15Q$								
H_MRI								
RIX	56.10***	0.68	31.38***	0.86				
	(11.85)		(6.91)					
TLM								
Group B — Stock-Return-Based Measures								
$BT11P$			−12.29*	0.20	−11.05***	0.89		
			(5.76)		(2.98)			
CJI	−30.19**	0.18					9.37**	3.21
	(10.78)						(3.37)	
$JumpRisk$							−6.14**	3.35
							(2.70)	
$JumpRP$								
λ_{Hill}								
Group C — Option-Return-Based Measures								
$ADBear$	17.06***	0.39	13.04***	1.02			1.54***	0.24
	(4.69)		(3.12)				(0.43)	
$JUMP$								
Group D — Macroeconomic Measures								
LE							−9.49***	4.31
							(2.54)	
Controls	*Yes*		*Yes*		*Yes*		*Yes*	

Moreover, the fixed and time-varying cutoff measures theoretically capture different things: while both attempt to approximate the integral in Eq. (6), they do so with different cutoff parameters k_t, either fixed or time-varying. As a result, the time-series dynamics of the two measures can be very different. Any difference in the dynamics can have a big impact on the prediction results. Finally, at-the-money option-implied volatility is a biased measure of true volatility, and the bias may vary over time. This introduces additional noise into measures based on time-varying cutoffs. Thus, the potential drawbacks of time-varying target moneyness are likely to outweigh the benefits. For time-varying cutoff moneyness, the effects are likely to be smaller because the estimates still rely on multiple options beyond the cutoff.

As expected, the correlation between the fixed and time-varying target versions of $BT11Q$ is only modest at 0.53. For the $BT11Q_{var}$ measure, the results are quite different from those for $BT11Q$, and significantly worse. $BT11Q_{var}$ can predict future tail events only at the daily horizon, the left-tail variation not at all, and future market excess returns only at the annual horizon. Thus, the time-varying target moneyness seems to rather add noise than help to extract more precise economic content for $BT11Q$. We therefore recommend using the original measure with fixed moneyness.

The correlations between the fixed- and time-varying cutoff moneyness versions of the other measures are higher, being 0.79 for $ATprob$, 0.93 for $ATvol$, 0.86 for $BT14Q$, 0.92 for $BTX15Q$, and 0.97 for TLM. Accordingly, the fixed-moneyness results for these measures are generally qualitatively similar to those with the time-varying moneyness cutoff. The only notable difference is that $ATprob_{fixed}$ outperforms the standard $ATprob$ measure, particularly in predicting future tail events. Overall, however, the $ATprob_{fixed}$ measure still does not perform as well as $BT11Q$. For the other measures, the moneyness cutoff appears to have a

Table 11
Different tail risk measure definitions.
This table reports the robustness analyses for different moneyness target and cutoff definitions of the option-implied tail risk measures. For $ATprob$, $ATvol$, $BT14Q$, $BTX15Q$, and TLM, we use the fixed cutoff moneyness level of 0.9 instead of the time-varying one on which the main measures are based. We denote the fixed-cutoff measures with a superscript "fixed". Furthermore, we vary the target moneyness of $BT11Q$, using 0.8875 ($BT11Q_{0.8875}$) or 0.9125 ($BT11Q_{0.9125}$), and we also use a variable moneyness level, defined as $e^{-2.5\sigma_{ATM,t}\sqrt{\tau}}$ ($BT11Q_{var}$). Panel A shows the results for predicting tail events. Panel B shows the predictability of the left-tail variation, and Panel C presents the results for the return predictability. The corresponding methodologies are described in the respective main tables. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Prediction of Tail Events

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2
$ATprob_{fixed}$	−0.10	0.12	0.14**	0.47	0.17***	0.68
	(0.09)		(0.08)		(0.05)	
$ATvol_{fixed}$	0.03	0.17	0.03	0.14	−0.16	0.51
	(0.08)		(0.08)		(0.16)	
$BT11Q_{0.8875}$	0.09*	0.35	0.10	0.38	−0.10	0.15
	(0.06)		(0.09)		(0.11)	
$BT11Q_{0.9125}$	0.10*	0.33	0.16**	0.61	−0.04	0.14
	(0.06)		(0.08)		(0.12)	
$BT11Q_{vary}$	0.09**	0.70	−0.04	0.04	−0.15**	0.50
	(0.05)		(0.08)		(0.07)	
$BT14Q_{fixed}$	−0.03	0.05	0.04	0.20	0.10	0.56
	(0.05)		(0.07)		(0.08)	
$BTX15Q_{fixed}$	0.06	0.31	0.03	0.15	0.05	0.32
	(0.06)		(0.08)		(0.10)	
TLM_{fixed}	−0.06	0.14	−0.36	0.25	−0.41	0.33
	(0.30)		(0.49)		(0.54)	
$Controls$	Yes		Yes		Yes	

Panel B. Predictability of Left-Tail Variation

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2
$ATprob_{fixed}$	−0.03	1.33	0.22	5.80	0.26*	5.81
	(0.05)		(0.21)		(0.17)	
$ATvol_{fixed}$	0.04*	1.16	−0.03	1.98	−0.11	1.92
	(0.03)		(0.07)		(0.11)	
$BT11Q_{0.8875}$	0.39**	4.92	0.42*	10.11	0.06	6.65
	(0.19)		(0.22)		(0.12)	
$BT11Q_{0.9125}$	0.49**	5.35	0.53**	10.80	0.15	7.12
	(0.25)		(0.26)		(0.12)	
$BT11Q_{vary}$	−0.00	0.44	−0.03	1.06	−0.05	0.97
	(0.03)		(0.03)		(0.04)	
$BT14Q_{fixed}$	−0.02	0.68	0.00	2.20	−0.04	2.56
	(0.03)		(0.03)		(0.05)	
$BTX15Q_{fixed}$	−0.01	0.49	0.00	1.48	−0.02	1.67
	(0.03)		(0.04)		(0.06)	
TLM_{fixed}	0.02	1.72	−0.07	4.61	−0.43	4.65
	(0.14)		(0.32)		(0.48)	
$Controls$	Yes		Yes		Yes	

Panel C. Return Predictability

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2	$Annual$	R^2
$ATprob_{fixed}$	30.34***	0.49	14.11**	0.58	6.39*	0.53	1.82*	0.77
	(8.14)		(7.20)		(4.31)		(1.12)	
$ATvol_{fixed}$	−2.18	0.01	0.02	0.01	2.16	0.06	3.99**	3.49
	(4.77)		(3.96)		(3.03)		(1.68)	
$BT11Q_{0.8875}$	28.38***	0.45	12.45*	0.41	4.09	0.18	3.34**	2.32
	(7.58)		(7.05)		(4.71)		(0.99)	
$BT11Q_{0.9125}$	29.49***	0.46	12.20*	0.37	4.74	0.24	3.27***	2.10
	(7.28)		(6.79)		(4.74)		(0.93)	
$BT11Q_{vary}$	5.31	0.04	4.44	0.18	2.06	0.20	3.50**	4.14
	(5.25)		(3.27)		(2.18)		(1.48)	
$BT14Q_{fixed}$	11.84**	0.06	5.77	0.05	0.66	0.24	2.34*	0.55
	(6.14)		(4.49)		(3.38)		(1.50)	
$BTX15Q_{fixed}$	8.20*	0.05	4.33	0.08	−0.64	0.10	2.17*	0.90

(continued on next page)

Table 11 (*continued*).

	(5.62)		(3.96)		(2.70)		(1.17)	
TLM_{fixed}	24.94***	0.25	15.10**	0.50	7.30*	0.49	3.69***	1.79
	(6.83)		(5.93)		(4.67)		(1.10)	
Controls	*Yes*		*Yes*		*Yes*		*Yes*	

negligible effect on the results. Thus, for measures with a cutoff point, moderate changes in the exact part of the tail that they look at also do not seem to have a primary effect on the results.

6. Simulation evidence

To better understand why some option-implied tail risk measures work better than others, we perform a simulation analysis using the simulated environments of two important classes of standard option pricing models.[21] First, we simulate stock index returns and option prices using the model of Pan (2002). For simplicity, we follow Carr and Wu (2003) and assume that the interest rate r_f and the dividend yield δ are constant. The joint data-generating process for the index price X and its volatility σ is as follows:

$$dX_t = \left[r_f - \delta + \eta_x \sigma_t^2 + \lambda \sigma_t^2 \left(\mu - \mu^*\right)\right] X_t dt + \sigma_t X_t dW_t^{(1)} + dZ_t - \mu X_t \lambda \sigma_t^2 dt, \tag{13}$$

$$d\sigma_t^2 = \kappa_\sigma \left(\bar{\sigma}^2 - \sigma_t^2\right) dt + \sigma_\sigma \sigma_t \left(\rho dW_t^{(1)} + \sqrt{1-\rho^2} dW_t^{(2)}\right). \tag{14}$$

$W^{(1)}$ and $W^{(2)}$ are adapted Brownian motions. The Brownian shocks are correlated with the correlation coefficient ρ. Price jumps are captured by the pure-jump process Z, which consists of random jump arrival times and random jump sizes. Jumps follow a Poisson distribution. For small time intervals Δt, the conditional probability of a jump is approximately $\lambda \sigma_t^2 \Delta t$. If there is a jump event, the expected relative jump size is $\mu = e^{\mu_J + \sigma_J^2/2} - 1$.

For the parameters, we use the estimates of Pan (2002), reported in Tables 3 and 6 of her paper. In particular, we set $\eta_x = 3.6$, $\lambda = 12.3$, $\mu = -0.008$, $\sigma_J = 0.0387$, and $\mu^* = -0.192$. Also, $\kappa_\sigma = 6.4$, $\bar{\sigma}^2 = 0.0153$, $\sigma_\sigma = 0.30$, $\rho = -0.53$, $r_f = 0.058$, and $\delta = 0.025$. As starting values, we use the unconditional averages and $X_0 = 1$.

We simulate this system $n = 10{,}000$ times, each with 5-minute data for 510 daily time-series observations. We discard the first 252 days as a burn-in period. The next 6 days are used to compute the option-based tail risk measures used in this study, and the last 252 days provide various return forecast windows for evaluating the tail risk measures. We compute option prices using a Fourier transform of the risk-neutral densities (Carr and Madan, 1999). We compute 25 out-of-the-money put (and call) options with moneyness between 0.7 and 1 (1 and 1.3).[22] In addition, we also compute the boundary put option price with a moneyness of $e^{-2.5\sigma_{ATM,\tau}\sqrt{\tau}}$. We follow Chong and Todorov (2023) and assume that option prices are observed with random measurement errors according to:

$$\hat{O}_t(K_j) = O_t(K_j)\left(1 + \left(0.01 + 0.004\left|\frac{K_j}{F_{t,\tau}} - 1\right|\right) z_{t,j}\right), \tag{15}$$

where $\hat{O}_t(K_j)$ is the observed option price (put or call) with strike price K_j, $O_t(K_j)$ is the model-implied option price with no measurement error, and $z_{t,j}$ is the realization of a standard normally distributed random variable.[23]

Finally, we compute the tail risk measures as described in Section A1 of the Online Appendix. For measures based on short-term options (without a precise target time-to-maturity), we compute option prices with 10 business days to maturity. For those referring to "options expiring in the next month" we use 21 business days to maturity, and for those expiring in the "month after next" we use 42 business days. We track future market excess returns over different windows separately in each simulated system. We thus have $n = 10{,}000$ observations of each tail risk measure along with the simulated future returns.

We present the results in Panels A and B of Table 12. To establish a baseline, we perform an analysis based on "infeasible" jump measures computed directly from the model state variables and the conditional distribution at time t. We consider jump intensity, different definitions of jump variation under $\mathbb{P}$ and $\mathbb{Q}$, the 5% value-at-risk and expected shortfall, and the probability of a 10% loss, each also under $\mathbb{P}$ and $\mathbb{Q}$. We find that almost all of these infeasible measures have significant predictability for future returns in the Pan (2002) model. Their median correlation is 0.92. Only the jump variation with a time-varying threshold based on ten times the at-the-money implied volatility has low correlations with the other measures and appears to have no discernible return predictability. Thus, the $ATvol$ and $BTX15Q$ measures that proxy this infeasible measure are unlikely to perform well. On the other hand, other tail risk measures that proxy one of the remaining infeasible measures should be capable of predicting future returns in the Pan (2002) model.

[21] We thank an anonymous associate editor for suggesting that we pursue this analysis.

[22] We need the call options primarily to compute the H_MRI measure, as well as the model-free option-implied variance needed to compute the variance risk premium.

[23] As a slight refinement of Chong and Todorov (2023), we model measurement errors that are more volatile for options further out-of-the-money. The authors use a flat specification with $0.015z_{t,j}$, which they choose to approximate the bid–ask spread of index options.

Table 12
Return predictability simulation.
This table reports the coefficients of return predictability regressions in simulated environments. We simulate 10,000 5-minute price paths in the Pan (2002) and Santa-Clara and Yan (2010) models. For each price path, we aggregate the information to the daily level and compute option prices of out-of-the-money put options with moneyness between 0.7 and 1. Using these options, we compute the values of all option-based tail risk measures and track future market excess returns in the system. Finally, we run single regressions of the simulated market excess returns for each lagged tail risk measure:

$$R_{t+\Delta t} = a + b \cdot TRM_t + c \cdot Controls_t + \epsilon_{t+\Delta t}.$$

$R_{t+\Delta t}$ is the excess return over the period Δt. TRM_t is the current observation of a tail risk measure. The results for these are in Panels B and D. In Panels A and C, we also consider a number of infeasible measures calculated directly from the model state variables and density. These include the jump intensity ($Jumpint$), the jump variation (JV) under various definitions, the 5% value-at-risk and expected shortfall, and the probability of a loss of 10% or more. We consider measures under $\mathbb{P}$ and $\mathbb{Q}$, as indicated by the last capital letter in each measure. The different jump variation measures include one that uses a constant threshold $k = log(0.9)$, two with time-varying thresholds k_t based on 2.5 and 10 times the current at-the-money implied volatility (subscripts var and $var10$, respectively), and one that follows the definition of Bollerslev and Todorov (2011b) (subscript BT; see Equation (A.1) in the Online Appendix). $Controls_t$ contains the simulated variance risk premium. We use four different forecast horizons Δt: (i) one day ($Daily$), (ii) one week ($Weekly$), (iii) one month ($Monthly$), and (iv) one year ($Annual$). In parentheses, we present robust Newey and West (1987) standard errors with 34 lags. The R^2 columns show the Lindeman et al. (1980) partial R^2 of each tail risk measure (in percentage points). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. In Panel E, we also show the correlations of the simulated tail risk measures with their corresponding infeasible measures (see Table 1; we use the exact matches for the jump variation) and the state variables that drive the risk premia in each model. Blank spaces indicate that there is no clear corresponding infeasible measure.

Panel A. Infeasible Measures in the Pan (2002) Model

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2	$Annual$	R^2
$Jumpint$	8.80***	0.19	7.21***	0.64	6.23***	2.14	1.40***	1.22
	(2.67)		(1.25)		(0.54)		(0.13)	
JVP	9.62***	0.23	7.42***	0.68	6.33***	2.21	1.36***	1.15
	(3.03)		(1.39)		(0.60)		(0.14)	
JVP_{var}	7.86***	0.15	6.75***	0.56	5.85***	1.89	1.35***	1.13
	(2.34)		(1.11)		(0.49)		(0.12)	
JVP_{var10}	−0.97**	0.00	−0.77**	0.01	−0.61***	0.02	0.02	0.00
	(0.46)		(0.36)		(0.20)		(0.12)	
JVP_{BT}	9.86***	0.25	7.46***	0.69	6.27***	2.18	1.32***	1.08
	(3.12)		(1.44)		(0.62)		(0.15)	
$VaRP$	7.88***	0.16	6.78***	0.56	5.91***	1.93	1.36***	1.15
	(2.38)		(1.13)		(0.49)		(0.12)	
ESP	7.85***	0.15	6.77***	0.56	5.88***	1.91	1.36***	1.15
	(2.39)		(1.13)		(0.50)		(0.12)	
$P10lossP$	9.35***	0.22	7.33***	0.66	6.34***	2.22	1.39***	1.20
	(2.95)		(1.35)		(0.59)		(0.14)	
JVQ	9.67***	0.24	7.43***	0.68	6.32***	2.21	1.36***	1.14
	(3.04)		(1.40)		(0.60)		(0.14)	
JVQ_{var}	8.59***	0.18	7.12***	0.62	6.15***	2.09	1.39***	1.20
	(2.56)		(1.20)		(0.52)		(0.13)	
JVQ_{var10}	−0.91**	0.00	−0.72**	0.01	−0.53***	0.02	0.03	0.00
	(0.43)		(0.37)		(0.20)		(0.11)	
JVQ_{BT}	9.89***	0.25	7.47***	0.69	6.26***	2.17	1.31***	1.07
	(3.13)		(1.44)		(0.62)		(0.15)	
$VaRQ$	8.02***	0.16	6.85***	0.58	5.97***	1.96	1.37***	1.16
	(2.43)		(1.15)		(0.50)		(0.12)	
ESQ	7.94***	0.16	6.83***	0.57	5.94***	1.95	1.37***	1.16
	(2.41)		(1.14)		(0.50)		(0.12)	
$P10lossQ$	9.33***	0.22	7.32***	0.66	6.34***	2.22	1.39***	1.20
	(2.94)		(1.35)		(0.59)		(0.14)	
$Controls$	Yes		Yes		Yes		Yes	

Panel B. Simulated Options in the Pan (2002) Model

	$Daily$	R^2	$Weekly$	R^2	$Monthly$	R^2	$Annual$	R^2
$ATprob$	9.29***	0.22	7.31***	0.66	6.22***	2.14	1.23***	0.95
	(3.28)		(1.51)		(0.64)		(0.15)	
$ATvol$	−6.05***	0.09	−5.46***	0.37	−4.88***	1.33	−1.20***	0.89
	(2.06)		(0.98)		(0.46)		(0.12)	
$BT11Q$	9.80***	0.24	7.43***	0.68	6.28***	2.19	1.32***	1.09
	(3.11)		(1.43)		(0.62)		(0.14)	
$BT14Q$	8.68***	0.19	6.13***	0.47	4.70***	1.24	1.05***	0.69
	(2.43)		(1.17)		(0.51)		(0.14)	
$BTX15Q$	−5.74***	0.08	−5.48***	0.37	−4.96***	1.37	−1.21***	0.91
	(2.09)		(1.00)		(0.46)		(0.12)	
H_MRI	−5.37***	0.07	−5.41***	0.36	−4.47***	1.11	−0.98***	0.60
	(1.86)		(0.89)		(0.39)		(0.12)	

(continued on next page)

Table 12 (*continued*).

RIX	9.39***	0.22	7.43***	0.68	6.34***	2.22	1.39***	1.19
	(2.87)		(1.33)		(0.57)		(0.14)	
TLM	7.85***	0.15	6.61***	0.54	5.81***	1.86	1.37***	1.17
	(2.36)		(1.12)		(0.50)		(0.12)	
ADBear	1.45	0.00	2.12**	0.05	1.41***	0.10	0.26**	0.05
	(2.31)		(1.17)		(0.51)		(0.13)	
JUMP	1.74	0.01	0.10	0.00	0.98**	0.05	0.14	0.01
	(2.12)		(0.93)		(0.43)		(0.13)	
Controls	*Yes*		*Yes*		*Yes*		*Yes*	

Panel C. Infeasible Measures in the Santa-Clara and Yan (2010) *Model*

	Daily	R^2	*Weekly*	R^2	*Monthly*	R^2	*Annual*	R^2
Jumpint	11.83***	0.09	5.75***	0.12	4.72***	0.37	0.61***	0.06
	(4.53)		(1.98)		(0.85)		(0.25)	
JVP	12.46***	0.10	6.05***	0.13	4.76***	0.37	0.61***	0.06
	(4.48)		(1.99)		(0.85)		(0.25)	
JVP_{var}	11.75***	0.09	5.79***	0.12	4.70***	0.36	0.61***	0.05
	(4.51)		(1.95)		(0.84)		(0.25)	
JVP_{var10}	−7.71**	0.04	−3.32**	0.04	−1.39**	0.03	−0.07	0.00
	(3.76)		(1.57)		(0.78)		(0.26)	
JVP_{BT}	12.41***	0.10	6.04***	0.13	4.79***	0.38	0.61***	0.06
	(4.47)		(1.98)		(0.85)		(0.25)	
VaRP	15.31***	0.14	7.01***	0.17	4.08***	0.28	0.41*	0.03
	(3.87)		(1.77)		(0.80)		(0.25)	
ESP	14.93***	0.14	7.41***	0.19	5.20***	0.45	0.58***	0.05
	(4.19)		(1.88)		(0.82)		(0.24)	
P10lossP	14.85***	0.14	6.97***	0.17	4.97***	0.41	0.59***	0.05
	(4.39)		(2.03)		(0.86)		(0.25)	
JVQ	5.92*	0.02	3.06**	0.03	2.82***	0.13	0.49**	0.04
	(4.03)		(1.76)		(0.80)		(0.25)	
JVQ_{var}	5.13*	0.02	2.75*	0.03	2.66***	0.12	0.48**	0.03
	(4.00)		(1.72)		(0.79)		(0.25)	
JVQ_{var10}	−9.63***	0.06	−4.26***	0.06	−2.45***	0.10	−0.13	0.00
	(2.93)		(1.31)		(0.67)		(0.25)	
JVQ_{BT}	11.55***	0.09	5.66***	0.11	4.59***	0.35	0.61***	0.05
	(4.50)		(1.98)		(0.85)		(0.25)	
VaRQ	15.00***	0.14	6.43***	0.14	4.49***	0.34	0.53**	0.04
	(4.71)		(2.08)		(0.86)		(0.25)	
ESQ	13.49***	0.12	6.78***	0.16	4.92***	0.40	0.61***	0.06
	(4.32)		(1.88)		(0.83)		(0.25)	
P10lossQ	14.29***	0.13	6.76***	0.16	4.94***	0.40	0.60***	0.05
	(4.42)		(2.02)		(0.86)		(0.25)	
Controls	*Yes*		*Yes*		*Yes*		*Yes*	

Panel D. Simulated Options in the Santa-Clara and Yan (2010) *Model*

	Daily	R^2	*Weekly*	R^2	*Monthly*	R^2	*Annual*	R^2
ATprob	9.64**	0.06	5.37***	0.10	3.90***	0.25	0.65***	0.06
	(4.46)		(2.00)		(1.22)		(0.26)	
ATvol	4.90	0.02	3.72**	0.05	3.19***	0.16	0.47**	0.03
	(3.85)		(1.66)		(0.74)		(0.25)	
BT11Q	12.47***	0.10	6.06***	0.13	4.82***	0.38	0.61***	0.06
	(4.48)		(1.98)		(0.85)		(0.25)	
BT14Q	0.40***	0.00	0.10***	0.00	−0.14***	0.00	0.02***	0.00
	(0.04)		(0.02)		(0.01)		(0.00)	
BTX15Q	0.40***	0.00	0.10***	0.00	−0.14***	0.00	0.02***	0.00
	(0.04)		(0.02)		(0.01)		(0.00)	
H_MRI	−4.30*	0.01	−3.04***	0.03	−1.65***	0.04	0.16	0.00
	(3.02)		(1.25)		(0.62)		(0.24)	
RIX	13.09***	0.11	6.32***	0.14	4.96***	0.41	0.62***	0.06
	(4.50)		(2.00)		(0.86)		(0.25)	
TLM	−14.62***	0.13	−6.27***	0.14	−4.04***	0.27	−0.63***	0.06
	(4.78)		(2.17)		(0.90)		(0.24)	
ADBear	−0.67	0.01	0.46	0.00	0.45	0.01	0.02	0.00
	(10.44)		(2.53)		(0.91)		(0.25)	

(*continued on next page*)

Table 12 (*continued*).

$JUMP$	−7.31 (17.59)	0.04	−1.38 (3.72)	0.01	−0.60 (1.09)	0.00	0.02 (0.26)	0.00
$Controls$	Yes		Yes		Yes		Yes	

Panel E. Correlations with Infeasible Measures and State Variables

	Pan (2002) Model		Santa-Clara and Yan (2010) Model				
	Infeasible	$\lambda\sigma_t^2$	Infeasible	Y_t	Y_t^2	λ_t	Y_tZ_t
$ATprob$	0.94	0.87	0.68	0.10	0.12	0.69	0.55
$ATvol$	−0.13	−0.83	0.27	−0.20	−0.20	0.75	0.42
$BT11Q$	1.00	0.94	0.99	0.20	0.20	1.00	0.80
$BT14Q$	0.71	0.72	0.02	−0.02	−0.02	−0.01	−0.02
$BTX15Q$	−0.13	−0.84	0.03	−0.02	−0.02	−0.01	−0.02
H_MRI		−0.79		−0.62	−0.56	−0.00	−0.34
RIX	0.99	0.99	0.99	0.26	0.26	0.99	0.83
TLM	1.00	0.98	−0.60	−0.66	−0.67	−0.59	−0.74
$ADBear$	0.25	0.27	0.16	0.17	0.19	0.12	0.19
$JUMP$		−0.01		0.11	0.10	0.03	0.07

Since the infeasible measures generally exhibit material predictability in the context of the Pan (2002) model, the performance of the tail risk measures appears to be driven primarily by how well the various measures capture these infeasible quantities. However, there is still some secondary heterogeneity across measures. The best-performing infeasible measures include the jump variation measures, the probability of a 10% loss, and the jump intensity. Based on these results, $ATprob$, $BT11Q$, $BT14Q$, RIX, and TLM should be particularly good predictors of future returns in the Pan (2002) model if they are good proxies for their corresponding infeasible measures. Interestingly, among the jump variation measures, the version underlying the $BT11Q$ measure of Bollerslev and Todorov (2011b) has the highest predictability for short horizons.

Panel B of Table 12 shows that, consistent with our empirical results, the measures $ATprob$, $BT11Q$, and RIX perform best in this simulated environment. They yield the highest positive coefficients and the largest partial R^2s among all measures for all forecast horizons. The coefficients and partial R^2s for these measures are generally similar in magnitude to those for the corresponding infeasible measures, indicating that they capture their targets well within the Pan (2002) model. The $BT14Q$, TLM, and $ADBear$ measures also yield positive coefficients for all horizons. However, $ADBear$, which also performs well empirically, yields only very small partial R^2s in the simulated environment of the Pan (2002) model.

In addition, also consistent with the empirical results, H_MRI is a significant negative predictor of future market excess returns over short time periods. Finally, as also observed empirically, the remaining measures ($ATvol$, $BTX15Q$, and $JUMP$) appear to have little to no ability to positively predict future market excess returns in the simulated environment of the Pan (2002) model. Thus, the fact that our empirical results are largely reproduced in the simulated sample reassuringly confirms the superiority of the simple $BT11Q$ measure.

As a second alternative, we also simulate the model of Santa-Clara and Yan (2010). Note that unlike the Pan (2002) model, Santa-Clara and Yan (2010) introduce a non-affine option pricing model that allows for separate and imperfectly correlated processes for stochastic volatility and stochastic jump intensity. It includes a risk premium for both the jump intensity and the jump size. The model dynamics are:

$$dX_t = \left[r_f + \phi_t - \lambda_t\mu_Q\right] X_t dt + Y_t X_t dW_t^{(1)} + Q_t X_t dN_t, \tag{16}$$

$$dY_t = \left[\mu_Y + \kappa_Y Y_t\right] dt + \sigma_Y dW_t^{(2)}, \tag{17}$$

$$dZ_t = \left[\mu_Z + \kappa_Z Z_t\right] dt + \sigma_Z dW_t^{(3)}, \tag{18}$$

$$log(1+Q_t) \sim \mathbb{N}\left(log\left(1+\mu_Q\right) - \frac{1}{2}\sigma_Q^2, \sigma_Q^2\right), \tag{19}$$

$$Prob(dN_t = 1) = \lambda_t dt, \text{ with } \lambda_t = Z_t^2. \tag{20}$$

The Brownian motions $W^{(1)}$, $W^{(2)}$, and $W^{(3)}$ have constant correlations (ρ_{12}, ρ_{13}, and ρ_{23}). N is a Poisson process and Q is the jump size, which follows an independent displaced log-normal distribution. Details on the calculation of the risk premium ϕ_t can be found in Section A2.3 of the Online Appendix.

We use the parameters from Table 2 of Santa-Clara and Yan (2010). That is, $\mu_Y = 2.841$, $\kappa_Y = -18.079$, $\sigma_Y = 0.334$, $\mu_Z = 7.745$, $\kappa_Z = -9.436$, $\sigma_Z = 1.529$, $\mu_Q = -0.098$, $\sigma_Q = 0.160$, $\rho_{12} = -0.495$, $\rho_{13} = -0.597$, $\rho_{23} = 0.168$, and $\gamma = 1.917$. As starting values we use $X_0 = 1$ and the unconditional averages from Table 3 of Santa-Clara and Yan (2010). The other steps of the simulation are similar to the ones for the Pan (2002) model. The option prices are also computed based on a Fourier transform and we apply the same measurement error specification.[24]

[24] Since the Fourier transform of the Santa-Clara and Yan (2010) model directly yields a solution for the call price rather than the distribution of the index price at maturity, we use the approach of Breeden and Litzenberger (1978), with 901 option prices between moneyness 0.2 and 2 to approximate the density.

We present the return predictability results in Panels C and D of Table 12. Panel C shows that the overall predictability by all infeasible measures in the Santa-Clara and Yan (2010) model is clearly weaker in terms of the partial R^2s and more heterogeneous (median correlation of 0.76) than in the Pan (2002) model. Again, almost all the infeasible measures taken directly from the model state have significant return predictability. Only the jump variation with a time-varying threshold based on ten times the at-the-money implied volatility performs very poorly, as in the Pan (2002) model. In the Santa-Clara and Yan (2010) model, the jump variation measures underlying the $BT11Q$ measure of Bollerslev and Todorov (2011b) has a higher return predictability than other jump variation measures, this time markedly so. Panel D shows that $ATprob$, $BT11Q$, and RIX are the best tail risk measures, again consistent with the empirical analysis. Most other measures do not perform as well as their infeasible counterparts.[25]

Finally, we analyze the mechanisms of return predictability in the two models in more detail. In the Pan (2002) model, the risk premium component of the index price X depends on two components: one that captures premia due to "Brownian" return risks (σ_t^2) and one that is associated with jump risk ($\lambda\sigma_t^2$). Thus, in this model, the tail risk measures must capture the main state variable σ_t^2 well in order to predict returns. In the Santa-Clara and Yan (2010) model, on the other hand, the risk premium depends on Y_t, Y_t^2, $\lambda_t = Z_t^2$, and the interaction $Y_t Z_t$ (see Equation (A.30) in Section A2.3 of the Online Appendix).

Thus, the return predictability of the tail risk measures depends on the extent to which they can capture the risk premia in the model. This may fail along two dimensions: (i) some of the infeasible measures that the tail risk measures proxy may predict returns better than others, and (ii) the approaches may be susceptible to measurement error and noise induced by the way they are estimated. We have already noted that most of the infeasible measures have strong return predictability, although there are subtle differences, particularly in the Santa-Clara and Yan (2010) model.

We examine dimension (ii) in Panel E of Table 12. The table reports the correlations of the tail risk measures with their corresponding infeasible counterparts (according to the measure definitions; see Table 1), as well as with the model state variables that drive the risk premia. We find that the $BT11Q$ and RIX measures work so well in the two models because they consistently approximate their infeasible counterparts with high precision. Correspondingly, these measures are also highly correlated with the model state variables, in particular the jump intensity. Other tail risk measures, especially those based on parametric or nonparametric optimization (e.g., $ATprob$, $ATvol$, $BT14Q$, $BTX15Q$, and TLM), fail to consistently capture their corresponding infeasible counterparts very well. Interestingly, some tail risk measures, such as $ATprob$, $BT14Q$, and TLM, do a much better job of capturing their infeasible counterparts in the simpler Pan (2002) model than in the more complex and non-affine model of Santa-Clara and Yan (2010).

Thus, the simulation analysis shows that measurement error is a primary driver of differences in the relative performance of the tail risk measures. The fact that the tail risk measures capture slightly different aspects of tail risk also plays a role, but the accuracy with which the tail risk measures capture the aspect of tail risk they are trying to capture appears to be more important.

7. Conclusion

We contribute to the literature by conducting a comprehensive empirical analysis of a wide range of tail risk measures that have been proposed over the last decades. We find a large heterogeneity across different tail risk measures. The first two principal components explain only 58% of their total variation, while some tail risk measures are even negatively correlated. This finding is a clear warning to researchers and practitioners not to treat different tail risk measures as interchangeable.

We find that the option-implied measure of Bollerslev and Todorov (2011b), $BT11Q$, performs best overall. Other measures perform even better for specialized tasks, notably the $JumpRisk$ measure of Maheu et al. (2013) for predicting future tail events and the $ADBear$ measure of Lu and Murray (2019) for predicting future market excess returns. However, only $BT11Q$ consistently excels at all tasks: It can predict the occurrence and the magnitude of future tail events, as well as the variation caused by them. The measure also predicts market excess returns at several horizons up to one year. Moreover, it is priced in the cross-section of stock returns and affects real economic activity. A simulation analysis shows that low measurement error is an important driver of the good performance of $BT11Q$.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2024.105769.

[25] The most extreme case is the TLM measure, which inversely predicts returns in the Santa-Clara and Yan (2010) model. We note that this is related to the shape of the distribution implied by the model. Empirically and in the Pan (2002) model, the tail shape parameter ξ (see Section A1 of the Online Appendix for more details) is generally positive, while in the Santa-Clara and Yan (2010) model, it is almost always negative. Since the tail risk measure is defined as $\frac{\beta}{1-\xi}$, where β is the scale parameter, $\xi > 0$ and $\xi < 0$ have fundamentally different implications for the TLM measure.

References

Adrian, T., Boyarchenko, N., Giannone, D., 2019. Vulnerable growth. Amer. Econ. Rev. 109 (4), 1263–1289.
Agarwal, V., Ruenzi, S., Weigert, F., 2017. Tail risk in hedge funds: A unique view from portfolio holdings. J. Financ. Econ. 125 (3), 610–636.
Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Internat. Econom. Rev. 39 (4), 885–905.
Andersen, T.G., Fusari, N., Todorov, V., 2015. The risk premia embedded in index options. J. Financ. Econ. 117 (3), 558–584.
Andersen, T.G., Fusari, N., Todorov, V., 2017. Short-term market risks implied by weekly options. J. Finance 72 (3), 1335–1386.
Andersen, T.G., Fusari, N., Todorov, V., 2020. The pricing of tail risk and the equity premium: Evidence from international option markets. J. Bus. Econom. Statist. 38 (3), 662–678.
Andersen, T.G., Todorov, V., Ubukata, M., 2021. Tail risk and return predictability for the Japanese equity market. J. Econometrics 222 (1), 344–363.
Azen, R., Budescu, D.V., 2003. The dominance analysis approach for comparing predictors in multiple regression. Psychol. Methods 8 (2), 129–148.
Bakshi, G., Kapadia, N., Madan, D., 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. Rev. Financ. Stud. 16 (1), 101–143.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes. Econom. J. 12 (3), 1–32.
Baron, M., Xiong, W., 2017. Credit expansion and neglected crash risk. Q. J. Econ. 132 (2), 713–764.
Barro, R.J., 2006. Rare disasters and asset markets in the twentieth century. Q. J. Econ. 121 (3), 823–866.
Bekaert, G., Harvey, C.R., Lundblad, C.T., Siegel, S., 2011. What segments equity markets? Rev. Financ. Stud. 24 (12), 3841–3890.
Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Rev. Financ. Stud. 22 (11), 4463–4492.
Bollerslev, T., Todorov, V., 2011a. Estimation of jump tails. Econometrica 79 (6), 1727–1783.
Bollerslev, T., Todorov, V., 2011b. Tails, fears, and risk premia. J. Finance 66 (6), 2165–2211.
Bollerslev, T., Todorov, V., 2014. Time-varying jump tails. J. Econometrics 183 (2), 168–180.
Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. J. Financ. Econ. 118 (1), 113–134.
Bondarenko, O., 2014. Why are put options so expensive? Q. J. Finance 4 (3), 1450015.
Breeden, D.T., Litzenberger, R.H., 1978. Prices of state-contingent claims implicit in option prices. J. Bus. 51 (4), 621–651.
Britten-Jones, M., Neuberger, A., 2000. Option prices, implied price processes, and stochastic volatility. J. Finance 55 (2), 839–866.
Brownlees, C.T., Gallo, G.M., 2010. Comparison of volatility measures: A risk management perspective. J. Financ. Econom. 8 (1), 29–56.
Carr, P., Madan, D., 1999. Option valuation using the fast Fourier transform. J. Comput. Finance 2 (4), 61–73.
Carr, P., Wu, L., 2003. What type of process underlies options? A simple robust test. J. Finance 58 (6), 2581–2610.
Carr, P., Wu, L., 2009. Stock options and credit default swaps: A joint framework for valuation and estimation. J. Financ. Econom. 8 (4), 409–449.
Chong, C.H., Todorov, V., 2023. Volatility of volatility and leverage effect from options. Northwestern University Working Paper.
Christoffersen, P., Jacobs, K., Ornthanalai, C., 2012. Dynamic jump intensities and risk premiums: Evidence from S&P 500 returns and options. J. Financ. Econ. 106 (3), 447–472.
Cremers, M., Halling, M., Weinbaum, D., 2015. Aggregate jump and volatility risk in the cross-section of stock returns. J. Finance 70 (2), 577–614.
Dew-Becker, I., Giglio, S., Kelly, B., 2021. Hedging macroeconomic and financial uncertainty and volatility. J. Financ. Econ. 142 (1), 23–45.
Du, J., Kapadia, N., 2013. The tail in the volatility index. EFA Meetings 2013 Working Paper.
Faff, R.W., Hillier, D., Hillier, J., 2000. Time varying beta risk: An analysis of alternative modelling techniques. J. Bus. Finance Account. 27 (5–6), 523–554.
Gao, G.P., Gao, P., Song, Z., 2018. Do hedge funds exploit rare disaster concerns? Rev. Financ. Stud. 31 (7), 2650–2692.
Gao, G.P., Lu, X., Song, Z., 2019. Tail risk concerns everywhere. Manage. Sci. 65 (7), 3111–3130.
Gormsen, N.J., Jensen, C.S., 2022. Higher-moment risk. University of Chicago Working Paper.
Gourio, F., 2012. Disaster risk and business cycles. Amer. Econ. Rev. 102 (6), 2734–2766.
Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)? J. Appl. Econometrics 20 (7), 873–889.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 3 (5), 1163–1174.
Hollstein, F., Prokopczuk, M., 2016. Estimating beta. J. Financ. Quant. Anal. 51 (4), 1437–1466.
Hollstein, F., Prokopczuk, M., Wese Simen, C., 2019. Estimating beta: Forecast adjustments and the impact of stock characteristics for a broad cross-section. J. Financial Mark. 44, 91–118.
Jackwerth, J.C., 2000. Recovering risk aversion from option prices and realized returns. Rev. Financ. Stud. 13 (2), 433–451.
Jiang, G.J., Tian, Y.S., 2005. The model-free implied volatility and its information content. Rev. Financ. Stud. 18 (4), 1305–1342.
Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Rev. Financ. Stud. 27 (10), 2841–2871.
Lazarus, E., Lewis, D.J., Stock, J.H., Watson, M.W., 2018. HAR inference: Recommendations for practice. J. Bus. Econom. Statist. 36 (4), 541–559.
Lindeman, R.H., Merenda, P., Gold, R.Z., 1980. Introduction to Bivariate and Multivariate Analysis. Scott, Foresman & Co, New York.
Lu, Z., Murray, S., 2019. Bear beta. J. Financ. Econ. 131 (3), 736–760.
Maheu, J.M., McCurdy, T.H., Zhao, X., 2013. Do jumps contribute to the dynamics of the equity premium? J. Financ. Econ. 110 (2), 457–477.
Mancini, C., 2001. Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. Giornale dell'Istituto Italiano degli Attuari 64 (19–47), 44.
Muir, T., 2017. Financial crises and risk premia. Q. J. Econ. 132 (2), 765–809.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Pan, J., 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. J. Financ. Econ. 63 (1), 3–50.
Rapach, D.E., Strauss, J.K., Zhou, G., 2013. International stock return predictability: What is the role of the United States? J. Finance 68 (4), 1633–1662.
Rietz, T.A., 1988. The equity risk premium a solution. J. Monetary Econ. 22 (1), 117–131.
Santa-Clara, P., Yan, S., 2010. Crashes, volatility, and the equity premium: Lessons from S&P 500 options. Rev. Econ. Stat. 92 (2), 435–451.
Seo, S.B., Wachter, J.A., 2018. Option prices in a model with stochastic disaster risk. Manage. Sci. 65 (8), 3449–3469.
Stambaugh, R.F., 1999. Predictive regressions. J. Financ. Econ. 54 (3), 375–421.
Symitsi, E., Symeonidis, L., Kourtis, A., Markellos, R., 2018. Covariance forecasting in equity markets. J. Bank. Financ. 96, 153–168.
Vilkov, G., Xiao, Y., 2015. Option-implied information and predictability of extreme returns. SAFE Working Paper.
Weller, B., 2018. Measuring tail risks at high frequency. Rev. Financ. Stud. 32 (9), 3571–3616.