AGGREGATE TAIL RISK, ECONOMIC DISASTERS,
AND THE CROSS-SECTION OF EXPECTED RETURNS

DAVID A. CHAPMAN∗†
University of Virginia

MICHAEL GALLMEYER
University of Virginia

NOVEMBER 26, 2014

Abstract

The aggregate tail risk measure constructed in Kelly and Jiang (2014) has no additional explanatory power when added to a standard model of the unconditional cross-section of expected returns. In conditional predictive regression systems and vector-autoregressions of the market portfolio and the long- and short-sides of the SMB and HML portfolios, the tail measure appears to forecast expected returns — and not cash flows — on both small and large stocks and on growth stocks. This is inconsistent with a rare disaster model of asset prices which should operate primarily through a cash flow channel. The tail measure appears to be related to decades long movements in expected returns which may be more consistent with a long-run risks model.

∗Mailing Address (both authors): McIntire School of Commerce, University of Virginia, Rouss & Robertson Halls, East Lawn, P.O. Box 400173, Charlottesville, VA 22904-4173, e-mail: david.chapman@virginia.edu and mgallmeyer@virginia.edu. We would like to thank Alex Chinko, Andrew Ellul, Tim Johnson, Dana Kiku, Neil Pearson, Josh Pollett, and seminar participants at the Darden School FEBB seminar at the University of Virginia, the University of Illinois at Urbana-Champaign, and Indiana University for helpful comments. We would like to thank Bryan Kelly and Hao Jiang for providing the time series of their tail risk factor. The current version of this paper (along with supporting programs and public domain data) is available online at https://sites.google.com/site/davidchapmanswebsite/research/cg_tailrisk.

†Corresponding author
1 Introduction

A possible connection between the behavior of the tail of the distribution of market returns and the expected return on the market and on a cross-section of U.S. equities can be motivated in a number of ways, including the preference-based approaches of Harvey and Siddique (2000) or the long-run risks model of Bansal and Yaron (2004). There is also an established line of research interested in the connection between extreme economic (quantity) events and average stock returns. Reitz (1988) is among the earliest examples of this research. He uses large but infrequent drops in aggregate consumption as a resolution of the equity premium puzzle posed in Mehra and Prescott (1985). The immediate response to a rare disaster theory of asset prices was provided by Mehra and Prescott (1988): In order to resolve the equity premium puzzle (in a simple consumption-based model) the required extreme realizations are three times larger than the largest declines observed in U.S. consumption; so they are, well, really rare in the time series.

The panel estimator of aggregate tail risk constructed in Kelly and Jiang (2014) (hereafter ‘KJ’) is motivated – in part – by precisely this problem. KJ show how to construct a measure of aggregate tail risk from the monthly cross-section of stock returns assuming that individual stock returns are well-described by a dynamic power law with a common component. A fundamental advantage of this approach is that it does not rely on the realizations of quantity disasters. Beliefs about a time-varying probability of disaster can be potentially quite important for realized stock returns. KJ establish some basic properties of this tail risk measure: (1) Exposure to the tail factor results in a return spread that is not explained by existing cross-sectional empirical factors. (2) Tail risk is persistent, and it forecasts the market return – both in univariate and bivariate regressions – at horizons from 1-month to 5-years, and (3) the tail index is related to movements in real macroeconomic quantities.

We ask two basic sets of questions about the KJ measure: Is it compatible with a rare disaster explanation for stock returns, and what does the KJ tail index say about the cross-section of expected returns? These are distinct but related questions: If the tail index explains aspects of the cross-section through a rare disaster story, then stocks and industries that are more exposed, ex ante, to disaster risk should load more heavily on this measure.

We begin by adding an excess return formed by sorting on exposure to the tail risk factor to a standard unconditional empirical factor model. This point is fundamentally different from point (1) above (in KJ): there may be many cross-sectional characteristics-sorted port-

---

1This line of inquiry has been extended by Barro (2006), Gabaix (2012), and Wachter (2013).
2Our research question draws on several strands of literatures ... (R)esearchers have hypothesized that heavy-tailed shocks to economic fundamentals help explain certain asset pricing behavior that has proved otherwise difficult to reconcile with traditional macrofinance theory. Examples include the Rietz (1988) and Barro (2006) rare disaster hypothesis and its extensions to dynamic settings by Gabaix (2012), Gourio (2012), and Wachter (2013) ...” (KJ, page 2844).
folios that cannot be priced by a parsimonious factor model, but that does not imply that these return differences represent priced risks in the cross-section. Using monthly return data from 1973 to 2011, we find that the tail risk factor adds nothing to an unconditional empirical model based on the market portfolio, size, value-growth, and momentum factors. We believe that this result is not surprising since it seems unlikely that a tail risk factor would necessarily have a large and linear effect on the unconditional mean of stock returns. This approach, while empirically common, may have very little power to detect a tail effect that is clustered (infrequently) in time and (possibly) nonlinear.

An examination of the univariate betas formed from the regression of the value-weighted portfolio returns to the 17 industries classified by Fama and French on the tail risk also supports the conclusion that the this measure has no information about the average cross-section of stock returns. The full sample average betas constructed from 5-year overlapping samples are virtually all positive and clustered in a narrow range between 0.18 and 0.27; i.e., average expected returns of all industry portfolios increase with a fattening of aggregate tail risk and they increase mostly in the same way. An examination of the time-series of the univariate betas for the two extreme industries (‘oil’ and ‘clothing’) do not reveal any interesting dynamic behavior. Again, we interpret this as additional evidence that the tail measure does not have valuable information about the unconditional cross-section of average returns.

In order to examine the conditional implications of aggregate tail risk, we use the predictive regression and vector autoregression (VAR) approach in Cochrane (2008, 2011). The tail risk factor, along with the log dividend yield, are combined to form bivariate predictive regressions and VARs for the market portfolio and the separate long- and short-portions of the size (SMB) and value-growth (HML) factors from Fama-French. Following the argument in Cochrane (2011), we use these factor portfolios as “data reduction tools” for examining the cross-section of expected returns. The data are annual from 1965 to 2011. Our results confirm the finding in KJ that the tail risk factor has predictive power for the market return.

We extend their results to demonstrate that tail risk predicts future discount rates rather than future changes in cash flows. This suggests that it is unlikely that the tail risk measure extracted from the cross-section of returns is operating through a consumption crash mechanism. We go on to show that the tail risk factor appears to have predictive content for both the long- and short-sides of the size factor (SMB) and for the short-side of HML (i.e., growth firms). However, since it forecasts both the long- and the short-sides of both portfolios with positive coefficients, its effect on the overall factor portfolios is negligible.

---

3 A natural extension would be to add the long- and short-sides of the momentum (UMD) portfolios. However, constructing the dividend growth and dividend yield series for this specific version of a momentum strategy is non-trivial. We defer this exercise to later work.
Interestingly, the tail factor seems to have implications for long-run expected returns at a
very low frequency; i.e., on the order of a decade or more.

Our findings suggest that the information about extreme events extracted from the tails
of stock returns are not likely to be related to a standard consumption-crash risk definition
of a rare disaster. That does not, however, imply that the measure is not interesting. As we
mentioned at the beginning of this section, there are other mechanisms through which tail
events might have an economically significant impact on asset returns and many of these
mechanisms may not be detected with reasonable power using standard linear modeling
techniques based on either aggregate consumption or aggregate dividends as measures of
cash flow. The implications of the tail risk measure for long-run expected returns may have
significant implications for a model based on long-run risks. Alternatively, it may prove to
be a useful mechanism for identifying the extreme states that affect asset prices through a
Campbell and Cochrane (1999) or generalized disappointment aversion (see Routledge and
Zin, 2010) model. These implications, however, remain to be seen, since powerful tests of
these non-nested models requires exploiting their specific economic structures.

2 The Definition of the Aggregate Tail Risk Measure

KJ employ a dynamic generalization of the power law tail measure developed in the statistics
literature beginning with Hill (1975). The critical assumptions underlying their measure
are that tail dynamics are similar across firms and that these dynamics are well-described
by a power law defined as

$$P(R_{i,t+1} < r \mid R_{i,t+1} < u_t \text{ and } F_t) = \left( \frac{r}{u_t} \right)^{-a_i/\lambda_t},$$  \hspace{1cm} (1)

where $r < u_t < 0$, $R_{i,t+1}$ is the return on asset $i$ from $t$ to $t+1$, $u_t$ is a time-varying tail cutoff
value that is common across firms, $a_i$ is a time invariant firm-specific tail parameter, and
$\lambda_t$ is the common determinant of tail dynamics; i.e., the KJ definition of “tail risk.” “High
values of $\lambda_t$ correspond to ‘fat’ tails and high probabilities of extreme returns.” (KJ, page
2842).

$\lambda_t$ is estimated in KJ as follows:

- $u_t$ is chosen each month to represent the 5th-percentile of the cross-section of realized
  residuals from the 3-factor empirical model of Fama and French (1993).

- Daily returns on all NYSE/AMEX/NASDAQ stocks with share codes 10 and 11 from
  January 1963 to December 2010 are pooled into monthly estimates.

---

4The introduction to KJ and their Section 1 describe the estimator and its underlying assumptions very
clearly. This brief section is included in order to make this paper self-contained. Our notation is taken
directly from KJ.
Before we summarize the findings in KJ, we note that power law distributions are also called Pareto-type distributions. The power law assumption in eq. (1) places significant restrictions on the rate at which the tail of the distribution dies out, beyond some specified value.\(^5\) The issue of the class(es) of distributions that fail to satisfy a power law and their relevance for financial data is potentially important.\(^6\) Whether or not this matters for the performance of the tail measure is not clear.

Figure 1 is a time series plot of aggregate tail risk, \(\lambda_t\), over time along with the monthly volatility of the CRSP value-weighted index.\(^7\) The first piece of casual evidence against a disaster risk explanation of the tail measure is that spikes in volatility, which presumably also reflect changes in beliefs about the probability of bad states of the world, are not reflected in significant changes in the tail index measure. Indeed, the largest sustained increase in aggregate tail risk occurred in the decade between 1968 and 1977.

\[\text{Figure 1: Market Volatility and the Tail Index Level}\]

The figure plots the time series of monthly observations on the KJ aggregate tail index series (\(\lambda_t\)) (measured using the scale on the right vertical axis) versus the level of the monthly volatility in the CRSP value-weighted index (measured using the scale on the left vertical axis) computed from daily returns as in Schwert (1989), page 1117. The shaded regions are NBER recessions. The data are from January 1963 to December 2011.

\(^5\)Clauset, Shalizi, and Newman (2009) examine the problem of fitting a power law to a small data set.

\(^6\)We conjecture that some stochastic volatility and/or jump models will result in tails that are not well described by a power law.

\(^7\)We would like to thank Bryan Kelly and Hao Jiang for providing their tail index series. The monthly volatility series is constructed from daily returns to the CRSP value-weighted index as in Schwert (1989), page 1117.
A few additional observations: (1) As KJ emphasize, the tail index series is highly persistent and this persistence is not mechanical since each monthly estimate of $\lambda_t$ is based on a separate cross-section. (2) The unconditional correlation between the tail index and market volatility is $-0.08$. (3) In NBER recessions, this correlation decreases (slightly) to $-0.12$. So, tail risk and volatility are distinct components of aggregate returns.

KJ document the following facts regarding the tail index measure:

**Fact #1:** If stocks are sorted into portfolios based on exposure to tail risk, then the subsequent return spread between a high tail beta portfolio and a low tail beta portfolio averages a (marginally) statistically significant 0.36 percent per month.\(^8\) This spread is not explained by the FF plus momentum factors.

**Fact #2:** The tail risk factor forecasts the CRSP value-weighted market returns at horizons from 1-month to 5-years both in- and out-of-sample in both a univariate predictive regression and in bivariate predictive regressions that include standard predictor variables. For example, at the 1-year horizon, the tail index positively predicts market returns (i.e., if return tails fatten, then expected market returns increase) with an $R^2$ of 6.1 percent and at the 5-year horizon the tail index again positively predicts market returns with an $R^2$ of 20.9 percent.

**Fact #3:** The tail index is related to movements in real macroeconomic quantities. Specifically, when the monthly tail index series is included in a vector autoregression (VAR) with macroeconomic prices and quantities, a one standard deviation shock to the tail risk measure is associated with a short-term (1-year) drop in investment of “2.5% to 4% ... followed by a recovery by year three.” (KJ, page 2866 and Figure 5.)

Fact #3, on its face, also seems inconsistent with asset pricing models based on “heavy-tailed shocks to economic fundamentals” – that is the “rare disasters” literature of Reitz (1988), Barro (2006), and Gabaix (2012) – alluded to in KJ. The real cash flow shocks in the VAR are simply too small and too quick to dissipate relative to the conventional heuristic of a “disaster.” For example, the tail index shock of 2.5% to 4% documented in KJ is well below the 8.82% per year standard deviation in the annual percentage change in gross private domestic investment from 1998 to 2013. But if the aggregate tail risk measure is not a conventional disaster risk, what is it? Facts #1 and #2 still raise important and intriguing questions. If tail risk appears to be compensated by cross-sectional return differences and not spanned by existing empirical factors (Fact #1), is the tail risk measure a priced risk? More generally, what is the relationship between tail risk in the cross-section and tail risk in the time-series?

\(^8\)This is the one-month average spread return in the value-weighted beta quintile portfolios reported in Table 4 of KJ.
3 Does Tail Risk Help to Explain the Cross-Section of Unconditional Expected Returns?

The (by now) standard four-factor unconditional empirical factor model is the Fama-French factors augmented with a momentum factor; i.e. for any (monthly) return $R_{i,t}$ in excess of a proxy for the monthly risk-free rate $R_{ft}$:

$$R_{i,t} - R_{ft} = \alpha + \beta_{i1}MKT_t + \beta_{i2}HML_t + \beta_{i3}SMB_t + \beta_{i4}UMD_t + \varepsilon_{it},$$  \hspace{1cm} (2)

where the first three factors are the market excess return, value, and size factors of Fama-French, and the last factor is a return spread based on a measure of return momentum. The simplest way to answer the question posed in the title of this section is to consider eq. (2) in comparison to an augmented five-factor model

$$R_{i,t} - R_{ft} = \alpha + \beta_{i1}MKT_t + \beta_{i2}HML_t + \beta_{i3}SMB_t + \beta_{i4}UMD_t + \beta_{i5}TAIL_t + \varepsilon_{it},$$  \hspace{1cm} (3)

where “TAIL” is defined as the spread between the return to an equally weighted portfolio of high tail beta stocks and an equally weighted portfolio of low tail beta stocks.\(^9\)

The results of estimating these two models on the monthly returns to the standard 25 size- and B/M-sorted portfolios and 10 sector portfolios over the period from January 1973 to December 2011 ($T = 468$) are in Tables 1 and 2.\(^{10}\) All alphas are estimated in a single time-series regression, and t-statistics are OLS corrected for serial correlation and heteroskedasticity. Alphas are reported in percent per month.

---

\(^9\)The results are the same if a value-weighted spread portfolio is used.

\(^{10}\)All test portfolio returns and the Fama-French factors are from Ken French’s website at Dartmouth. The “TAIL” measure was constructed using the KJ tail index, return data from WRDS, and the Matlab programs \texttt{rets\_process.m}, \texttt{betaform.m}, and \texttt{portform.m} which are all available for download at the paper’s website at https://sites.google.com/site/davidchapmanswebsite/research/cg\_tailrisk. The beta estimates and portfolio construction are designed to replicate the data constructed in KJ.
<table>
<thead>
<tr>
<th>Alpha Estimates</th>
<th>t-Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>B/M1</td>
<td>B/M2</td>
</tr>
<tr>
<td>S1</td>
<td>-0.52</td>
</tr>
<tr>
<td>S2</td>
<td>0.03</td>
</tr>
<tr>
<td>S3</td>
<td>0.05</td>
</tr>
<tr>
<td>S4</td>
<td>0.15</td>
</tr>
<tr>
<td>S5</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Panel B: 5-Factor Model

<table>
<thead>
<tr>
<th>Alpha Estimates</th>
<th>t-Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>B/M1</td>
<td>B/M2</td>
</tr>
<tr>
<td>S1</td>
<td>-0.56</td>
</tr>
<tr>
<td>S2</td>
<td>0.03</td>
</tr>
<tr>
<td>S3</td>
<td>0.04</td>
</tr>
<tr>
<td>S4</td>
<td>0.14</td>
</tr>
<tr>
<td>S5</td>
<td>0.14</td>
</tr>
</tbody>
</table>

The four factor model consists of the $MKT$, $SMB$, $HML$, and $UMD$ factors. These factors, along with the 25 size- and B/M-sorted portfolios are from Ken French’s website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The 5th factor is the spread portfolio formed from the equally weighted quintile of high exposure to the KJ tail factor minus the equally weighted quintile of low exposure to the KJ tail factor. The betas underlying these portfolios are estimated and the portfolios are formed in the Matlab program betaform.m (available online at https://sites.google.com/site/davidchapmanswebsite/research/cg_tailrisk) in a manner consistent with the construction in KJ. Alphas are in units of percent per month and are calculated from a single time-series regression of excess returns on the 4- or 5-factors. $t$-statistics are constructed using robust errors based on the Newey-West estimator with the asymptotically optimal bandwidth parameter. The data are for the $T = 468$ months from January 1973 to December 2011.
The four factor model consists of the MKT, SMB, HML, and UMD factors. These factors, along with the 10 sector portfolios are from Ken French’s website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The 5th factor is the spread portfolio formed from the equally weighted quintile of high exposure to the KJ tail factor minus the equally weighted quintile of low exposure to the KJ tail factor. The betas underlying these portfolios are estimated and the portfolios are formed in the Matlab program betaform.m (available online at https://sites.google.com/site/davidchapmanswebsite/research/CGtailrisk) in a manner consistent with the construction in KJ. Alphas are in units of percent per month and are calculated from a single time-series regression of excess returns on the 4- or 5-factors. t-statistics are constructed using robust errors based on the Newey-West estimator with the asymptotically optimal bandwidth parameter. The data are for the \( T = 468 \) months from January 1973 to December 2011.

In Table 1, 21 of 25 alphas are not statistically significantly different from zero and of the four that are statistically significant, only the smallest size- and lowest B/M-portfolio is economically significant, with an unexplained excess return of more than 6 percent per year (in absolute value). In Table 2, three of the ten sector portfolios have alphas that are significantly different from zero, with the largest (the healthcare sector) generating unexplained excess returns of 3.6 percent per year. This simply documents the fact that the 4-factor model fits the excess return to the test assets well. The new – and important point for our analysis – is the comparison of the 4- and 5-factor models. A casual examination of Tables 1 and 2 documents that the TAIL factor adds no additional explanatory power to the cross-section of average returns. The measured alphas for the 5-factor model are virtually identical.
to those of the 4-factor model. Furthermore, as Table 3 below shows, the fitted multivariate betas for the tail factor are not statistically significant in 31 of the 35 test portfolios, and they are economically small even when they are distinguishable from zero.

### Table 3: Fitted Betas for the TAIL Factor

#### Panel A: Size- & B/M-Sorted Portfolios

<table>
<thead>
<tr>
<th>Sector</th>
<th>Beta Estimate</th>
<th>t-Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>B/M1</td>
<td>0.07</td>
<td>0.98</td>
</tr>
<tr>
<td>B/M2</td>
<td>0.07</td>
<td>2.21</td>
</tr>
<tr>
<td>B/M3</td>
<td>0.04</td>
<td>1.81</td>
</tr>
<tr>
<td>B/M4</td>
<td>−0.01</td>
<td>−0.66</td>
</tr>
<tr>
<td>B/M5</td>
<td>−0.00</td>
<td>−0.04</td>
</tr>
</tbody>
</table>

#### Panel B: Sector Portfolios

<table>
<thead>
<tr>
<th>Sector</th>
<th>Beta Estimate</th>
<th>t-Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durable Goods</td>
<td>0.13</td>
<td>2.50</td>
</tr>
<tr>
<td>Energy</td>
<td>−0.08</td>
<td>−0.88</td>
</tr>
<tr>
<td>High-Tech</td>
<td>−0.04</td>
<td>−0.92</td>
</tr>
<tr>
<td>Healthcare</td>
<td>0.05</td>
<td>0.85</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.05</td>
<td>1.18</td>
</tr>
<tr>
<td>NonDurable Goods</td>
<td>0.01</td>
<td>0.18</td>
</tr>
<tr>
<td>Other</td>
<td>0.08</td>
<td>2.15</td>
</tr>
<tr>
<td>Shops</td>
<td>0.07</td>
<td>0.83</td>
</tr>
<tr>
<td>Telecom</td>
<td>0.04</td>
<td>0.56</td>
</tr>
<tr>
<td>Utilities</td>
<td>−0.12</td>
<td>−2.03</td>
</tr>
</tbody>
</table>

The five factor model consists of the MKT, SMB, HML, and UMD factors. These factors, along with the 25 size- and B/M-sorted portfolios and 10 sector portfolios are from Ken French’s website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The 5th factor is the spread portfolio formed from the equally weighted quintile of high exposure to the KJ tail factor minus the equally weighted quintile of low exposure to the KJ tail factor. The betas underlying these portfolios are estimated and the portfolios are formed in the Matlab program `betaform.m` (available online at https://sites.google.com/site/davidchapmanswebsite/research/cg_tailrisk) in a manner consistent with the construction in KJ. The betas are the multivariate betas from a single time-series regression of excess returns on the 5-factors. t-statistics are constructed using robust errors based on the Newey-West estimator with the asymptotically optimal bandwidth parameter. The data are for the $T = 468$ months from January 1973 to December 2011.
In summary, there is no evidence that the spread portfolio based on the \textit{TAIL} measure provides any additional explanatory power for the unconditional cross-section of average returns. This is not terribly surprising, however. The existing empirical factor models have been engineered to explain \textit{most} of the variation in measured returns, subject to the principle of parsimony. Furthermore, tail events – by definition – are rare and not likely to have a large weight in unconditional average returns. If anything, they are likely to have an effect on expected returns at certain concentrated point in time and possibly in a nonlinear way. Fact \#2 in Section 2 suggests that the more likely channel for connecting $\lambda_t$ to the time-series and cross-section of expected returns is through a conditional pricing model, and we pursue this intuition later in the paper.

\section{Industry Betas with Respect to the Tail Index Measure}

Intuitively, if the aggregate tail index captures the market’s beliefs about the likelihood of an economic disaster, then exposure to this risk should be larger for firms and industries that are particularly sensitive to aggregate crashes. The problem for the cross-section, though, is that it is hard to imagine an industry that is expected to do “well” in an aggregate crash. Figure 2 shows the full-sample averages of overlapping 5-year (univariate) beta estimates for each of the 17 (value-weighted) Fama & French industry portfolios. The graph shows that betas for almost all industries lie in the narrow range of 0.15 to 0.25. So, the expected returns of most industries seem to increase with increases in tail risk. Along with Table 3, this figure makes it clear that exposure to tail risk cannot plausibly generate cross-sectional differences in industry or size- and B/M-sorted portfolio returns.
Figure 2: Tail Factor Betas of the 38 Fama-French Industry Portfolios

The graph shows the average across the full sample of the value of the tail index beta estimated with overlapping 5-year intervals for each of the 38 Fama-French value-weighted industry portfolios. The data are monthly from 1963 to 2011.

Is there any possible information contained in the time series variation in tail betas? Figure 3 shows the time series of the overlapping 5-year beta estimates for the two extreme industries.
Figure 3: Tail Factor Betas over Time for Two Industries

The graph shows the time-series plots of the tail factor betas for the industries with the lowest (‘Oil’ in Panel (a)) and highest (‘Clothing’ in Panel (b)) average betas based on rolling 5-years of monthly data. The data are monthly from 1963 to 2011.

There seems to be little or no economic intuition in the time-series plots of the (rolling) tail betas on the two extreme tail beta industry portfolios. Both industries have beta estimates that are relatively low in the beginning of the sample and relatively high at the very end of the sample during aftermath of the financial crisis. In summary, there is little intuition about the nature of tail index risk from examining the betas in the cross-section of either industry sorted or characteristics sorted portfolios.

5 What Does Tail Risk Say About Cash Flows, Discount Rates, and the Term Structure of Risk Premiums?

All existing attempts to disentangle variation in future cash flows versus variation in expected discount rates start with the Campbell-Shiller decomposition of dividend yields:

\[
dp_t(i) \approx \sum_{j=1}^{k} \rho^{j-1} \Delta r_{t+j}^{(i)} - \sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j}^{(i)} + \rho^k dp_{t+k}^{(i)},
\]  

(4)
where \( dp_t^{(i)} \) is the log dividend yield on asset (portfolio) \( i \), \( r_{t+1}^{(i)} \) is the continuously compounded return on asset \( i \) from \( t \) to \( t+1 \), \( \triangle d_{t+1}^{(i)} \) is the growth rate in dividends on asset \( i \) from \( t \) to \( t+1 \), and \( \rho \approx 0.96 \) (at an annual frequency) is a constant of approximation. Equation (4) is valid for a long-only (dividend paying) portfolio. It can easily be extended to a long-short portfolio by evaluating the long and the short portions of the portfolio separately. This is necessary, below, because we examine the Fama-French size and value factors as “data reduction tools” for understanding the cross-section of expected returns.\(^{11}\)

The log dividend yield and the dividend growth rate series for \( MKT \) and the long and short sides of \( SMB \) and \( HML \) can be constructed from the with- and without-dividend return series.\(^{12}\) Figure 4 plots the annual observations of the five (log) dividend yield and dividend growth rate series. It shows that there is a high correlation in the time series among the log dividend yields. There is also substantial cross-sectional variation in actual dividend yields. For example, a log dividend yield of \(-3.5\) for \( HML \)-short in 1980 implies a dividend yield of \(3.02\) percent while the log dividend yield on \( HML \)-long at the same time is roughly \(-2.9\) which implies a dividend yield of \(5.5\) percent. The cross-sectional variation at many dates in the sample is almost as large as the time series variation of an individual series. The cross-sectional differences are intuitively reasonable, e.g., \( dp \) for \( HML \)-short is the lowest and \( dp \) for \( HML \)-long and \( SMB \)-short are the highest. It is important to emphasize two points: (1) As Cochrane (2008) notes, the logic of eq. (4) implies that variation in the log dividend yield over time must track time variation in discount rates or time variation in cash flows. (2) A comparison of Panels (a) and (b) of Figure 4 illustrate why it is consistently difficult to find evidence that dividend yields track future variation in dividend growth.

---

\(^{11}\)This language is from Cochrane (2011).

\(^{12}\)The standard computation for any portfolio, \( i \), is

\[
\frac{D_t^{(i)}}{P_t^{(i)}} = R_{w,i,t} - R_{x,i,t}^{(i)},
\]

where \( R_{w,i,t} \) is the with-dividend simple return to asset \( i \) and \( R_{x,i,t}^{(i)} \) is the ex-dividend simple return to asset \( i \). The dividend growth series can be constructed as

\[
\frac{D_{t+1}^{(i)}}{D_t^{(i)}} = (1 + R_{x,i,t}) \frac{R_{w,i,t+1} - R_{x,i,t+1}^{(i)}}{R_{w,i,t} - R_{x,i,t}^{(i)}}.
\]

Cochrane (2011) uses the Campbell-Shiller approximation to construct a dividend growth series that is exact in the log-linear approximation. He notes that the results are not significantly different with this alternate measure. We retain the simpler definition for transparency.
Figure 4: The Log Dividend Yields and Dividend Growth Rates on MKT, SMB and HML Factors

The plot shows the log $D/P$ (Panel (a)) and $\Delta d$ (Panel (b)) the for the MKT factor and the long- and short-sides of the SMB and HML factors constructed as described in Section 5. The data are annual from 1965 to 2012.

We first use simple forecasting regressions with returns, dividend growth, dividend yield, or $\lambda_t$ as dependent variables and lagged dividend yield and $\lambda_t$ as independent variables. These models are estimated using OLS and robust covariance matrix estimators. In order to summarize the long-run impact of dividend yields and $\lambda_t$ on returns and dividend growth, we use the first-order VAR and multivariate regression from Cochrane (2011):

\[
\begin{bmatrix}
    dp_{t+1}^{(i)} \\
    \lambda_{t+1}
\end{bmatrix} = \Phi
\begin{bmatrix}
    dp_t^{(i)} \\
    \lambda_t
\end{bmatrix} + \begin{bmatrix}
    \varepsilon_{dp}^{t+1} \\
    \varepsilon_{\lambda}^{t+1}
\end{bmatrix},
\]

\[
\begin{bmatrix}
    r_{t+1}^{(i)} \\
    \Delta d_{t+1}^{(i)}
\end{bmatrix} = B
\begin{bmatrix}
    dp_t^{(i)} \\
    \lambda_t
\end{bmatrix} + \begin{bmatrix}
    \varepsilon_r^{t+1} \\
    \varepsilon_{\Delta d}^{t+1}
\end{bmatrix},
\]

where $\Phi$ and $B$ are $2 \times 2$ coefficient matrices. The forecast of long-horizon returns are then calculated as

\[
\begin{bmatrix}
    r_{t}^{dr} \\
    \Delta d_{t}^{dr}
\end{bmatrix} \equiv \begin{bmatrix}
    \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} \\
    \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}
\end{bmatrix} = B (I - \rho \Phi)^{-1}
\begin{bmatrix}
    dp_{t}^{(i)} \\
    \lambda_{t}
\end{bmatrix},
\]
for $\rho = 0.96$ and with the long-run regression coefficients defined as $B (I - \rho \Phi)^{-1}$. The asymptotic covariance matrix of the long-run regression coefficients can be computed using the delta method, and we also bootstrap the distribution of the $t$-statistics on the long-run regression coefficients under the null hypothesis of no short or long run predictability from $\lambda_t$.\textsuperscript{13}

### 5.1 Tail Risk and the Market Portfolio

The results of the univariate and multivariate regressions for $MKT$ are shown in Table 4.\textsuperscript{14}

<table>
<thead>
<tr>
<th>Equation</th>
<th>$dp_t$</th>
<th>$\lambda_t$</th>
<th>$dp_t$</th>
<th>$\lambda_t$</th>
<th>$R^2$</th>
<th>$\sigma \left[ \hat{y}_{t+1} \right] %$</th>
<th>$\sigma' \left[ \hat{y}_{t+1} \right] %$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.105</td>
<td>0.045</td>
<td>1.703</td>
<td>1.836</td>
<td>0.15</td>
<td>6.834</td>
<td>1.630</td>
</tr>
<tr>
<td>(2)</td>
<td>0.040</td>
<td>0.023</td>
<td>0.795</td>
<td>1.144</td>
<td>0.05</td>
<td>3.094</td>
<td>0.979</td>
</tr>
<tr>
<td>(3)</td>
<td>0.961</td>
<td>-0.022</td>
<td>19.26</td>
<td>-1.128</td>
<td>0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>0.121</td>
<td>0.813</td>
<td>0.569</td>
<td>9.612</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$dp_t$</th>
<th>$\lambda_t$</th>
<th>$dp_t$</th>
<th>$\lambda_t$</th>
<th>$dp_t$</th>
<th>$\lambda_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{t+1}^{lr}$</td>
<td>0.655</td>
<td>0.234</td>
<td>0.304</td>
<td>0.329</td>
<td>0.131</td>
<td>0.341</td>
</tr>
<tr>
<td>$\Delta d_{t+1}^{lr}$</td>
<td>0.310</td>
<td>0.016</td>
<td>0.292</td>
<td>0.474</td>
<td>0.067</td>
<td>0.187</td>
</tr>
</tbody>
</table>

This table reports the forecasting regression coefficients and long-run forecasting coefficients for $MKT$ from a model that includes both dividend yield, $dp$, and tail risk, $\lambda$, as predictors. $MKT$ is defined as the annual return to the CRSP value-weighted portfolio of all stocks. $\sigma \left[ \hat{y}_{t+1} \right] \%$ is the volatility of the fitted value from the forecast, and $\sigma' \left[ \hat{y}_{t+1} \right] \%$ is the additional volatility of the fitted value explained by tail risk alone. The long-run regression coefficients reported below the individual equation results are defined as $r_{t+1}^{lr} \equiv \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$ and $\Delta d_{t+1}^{lr} \equiv \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}$. The asymptotic $p$-values for the long-run coefficients are constructed from the individual $t$-statistics using the delta method as described in the Appendix 1. The bootstrap $p$-values are constructed from the $t$-statistics under the null hypothesis that tail has no short-run or long-run predictability for either cash flows or discount rates as described in Appendix 2. The data are annual from 1965 to 2011.

The tail measure has marginal forecasting power for one year market returns. The co-

\textsuperscript{13}See the appendices for details.

\textsuperscript{14}This is a slight abuse of notation since we are examining the return to the value-weighted CRSP portfolio and not the excess return above a proxy for the risk-free rate.
efficient of 0.045 is marginally significant at conventional levels, and since returns here are measured annually not in percent, the coefficient of 0.045 translates into 4.5 percent predicted change in expected returns for a one standard deviation change in the tail measure (holding the log dividend yield constant). Tail risk also explains an additional 1.63 percent of the volatility of annual returns to the market portfolio. The level of predictability reported in Table 4 is roughly consistent with the bivariate results in Table 2 of KJ. Line (2) in Table 4 shows that $\lambda_t$ does not forecast 1-year dividend growth. The coefficient of $\lambda_t$ for dividend growth is less than half of the value for future returns, and it is not statistically significantly different from zero.

The point estimates of the long-run response to shocks to $\lambda_t$, reported at the bottom of Table 4, suggest that tail risk forecasts expected returns at long horizons, although its importance relative to the log dividend yield appears to be slightly less. Inference in long-horizon regressions with a relatively short sample period is problematic. The long-run coefficients are not statistically significant using either the asymptotic distribution or the bootstrap distribution under the null of no short- or long-run predictability from tail risk. However, the logic from Cochrane (2008, 2011) still binds: If dividend yields vary over time, then they must – by construction – forecast either discount rates or cash flows. The evidence (as is common) appears more consistent with forecasts of market discount rates.

The importance of the tail risk factor for long-run return predictability based on point estimates is shown in Figure 5, which shows the long-run expected return based on a forecasting model that uses dividend yield alone and a forecast based on the fitted VAR coefficients using both dividend yield and $\lambda_t$. The two forecasts are almost perfectly correlated. Including the tail risk measure tends to moderate the forecast of long-run returns at both the peak in the early eighties and at the trough of the end of the tech boom, but it provides a more extreme forecast in the late 1960s and early 1970s.

\[15\] In the regressions in Tables 6 through 8, $\lambda_t$ has been normalized to have a standard deviation of 1.0.

\[16\] The scale of the variables is slightly different and KJ use monthly overlapping annual regressions, but the $R^2$ and significance levels for the coefficients are qualitatively similar.
The plot shows the forecast (fitted value) of \( r_{t}^{lr} = \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} \) using only the log dividend yield, \( dp \), and using both \( dp \) and the tail risk, \( \lambda \). The long-run coefficient in the \( dp \) only case is equal to \( b_{lr} = \frac{b_{r}}{1-\rho \phi} \), where \( b_{r} \) is the coefficient in a regression of next year’s return on \( MKT \) on a constant and \( dp \), \( \rho \) is the Campbell-Shiller approximation coefficient, and \( \phi \) is the slope coefficient in the \( dp \) first-order autoregression. See Cochrane (2008) for further discussion. The long-run coefficients in the case of both \( dp \) and \( \lambda \) are defined in eq. (6). The data are annual from 1965 to 2011.

**Figure 5: Forecast of Long-Run Returns to \( MKT \)**

*In summary*, consistent with the results in KJ, we find that the aggregate tail risk measure has significant forecasting power at an annual horizon for short-run market returns. However, when we extend their findings using the methods in Cochrane (2011), we find \( \lambda_{t} \) seems to forecast discount rates and not cash flows. This result is an important finding in that it is inconsistent with a disaster risks interpretation of aggregate tail risk. It suggests that the tail risk measure may be picking up a component of the stochastic discount factor that is related to the relative valuation of extreme states; for example, the extreme state valuation in rank-dependent expected utility or generalized disappointment aversion.
5.2 Tail Risk and the Size Factor

The results of estimating the forecasting regressions and VARs – using OLS and robust covariance matrix estimation – for the long and short components of the $SMB$ factor are in Table 5 and Figure 4.
Table 5: Forecasting Regressions for SMB

**Panel A: The Long Portfolio of SMB**

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>t-statistics</th>
<th>Other Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$dp_t$</td>
<td>$\lambda_t$</td>
<td>$dp_t$</td>
</tr>
<tr>
<td>(1) $r_{t+1}$</td>
<td>0.090</td>
<td>0.054</td>
</tr>
<tr>
<td>(2) $\Delta d_{t+1}$</td>
<td>0.036</td>
<td>0.013</td>
</tr>
<tr>
<td>(3) $dp_{t+1}$</td>
<td>0.965</td>
<td>-0.042</td>
</tr>
<tr>
<td>(4) $\lambda_t$</td>
<td>0.060</td>
<td>0.822</td>
</tr>
</tbody>
</table>

$p$-Values for the Long-Run Coeffs.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Asymptotic</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{t}^{lr}$</td>
<td>0.423</td>
<td>0.314</td>
</tr>
<tr>
<td>$\Delta d_{t}^{lr}$</td>
<td>0.080</td>
<td>-0.038</td>
</tr>
</tbody>
</table>

**Panel B: The Short Portfolio of SMB**

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>t-statistics</th>
<th>Other Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$dp_t$</td>
<td>$\lambda_t$</td>
<td>$dp_t$</td>
</tr>
<tr>
<td>(1) $r_{t+1}$</td>
<td>0.098</td>
<td>0.043</td>
</tr>
<tr>
<td>(2) $\Delta d_{t+1}$</td>
<td>0.040</td>
<td>0.001</td>
</tr>
<tr>
<td>(3) $dp_{t+1}$</td>
<td>0.976</td>
<td>-0.039</td>
</tr>
<tr>
<td>(4) $\lambda_t$</td>
<td>0.192</td>
<td>0.809</td>
</tr>
</tbody>
</table>

$p$-Values for the Long-Run Coeffs.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Asymptotic</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{t}^{lr}$</td>
<td>0.849</td>
<td>0.308</td>
</tr>
<tr>
<td>$\Delta d_{t}^{lr}$</td>
<td>0.326</td>
<td>-0.050</td>
</tr>
</tbody>
</table>

This table reports the forecasting regression coefficients and long-run forecasting coefficients from a model that includes both dividend yield, $dp$, and tail risk, $\lambda$, as predictors. SMB is the size factor constructed as in Fama-French. $\sigma[\hat{y}_{t+1}]\%$ is the volatility of the fitted value from the forecast, and $\sigma'[\hat{y}_{t+1}]\%$ is the additional volatility of the fitted value explained by tail risk alone. The long-run regression coefficients reported below the individual equation results are defined as $r_{t}^{lr} \equiv \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$ and $\Delta d_{t}^{lr} \equiv \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}$. The asymptotic $p$-values for the long-run coefficients are constructed from the individual $t$-statistics using the delta method as described in the Appendix 1. The bootstrap $p$-values are constructed from the $t$-statistics under the null hypothesis that tail has no short-run or long-run predictability for either cash flows or discount rates as described in Appendix 2. The data are annual from 1965 to 2011. Panel A shows the results for the long portion of SMB, and Panel B shows the results for the short portion of SMB.
The direct evidence for either dividend yields or tail risk providing significant predictive power for the long side of the SMB portfolio (SMBL) is weak. The conventional t-statistics fail to reject the null of zero coefficient estimates for either predictor when applied to either future returns or future dividend growth (lines (1) and (2) in Panel A). However, given the evidence in Figure 4 showing time variation in the dividend yield on SMBL, the logic of equation (4) suggests that there is time variation in discount rates rather than cash flows. The t-statistics suggest that tail risk is marginally more important in forecasting discount rates. The heuristic measures advocated in Cochrane (2008, 2011) also suggest that predictability from tail risk is important. The combined predictors generate a volatility of fitted returns (σ[\hat{y}_{t+1}] % in Table 5) that is roughly as important as the model for the overall market return.

The results for the short side of the SMB portfolio (SMBS) are more consistent with those of the market return in Table 4. This follows from the relative importance of large market capitalization firms in both portfolios. Here, the tail risk factor appears to be a slightly stronger predictor of short term large stock returns than measured dividend yield. The volatility of fitted returns compared to total volatility is, again, comparable to that of the market portfolio. The evidence in favor of forecasts for discount rates is that there is no direct evidence that tail risk forecasts future cash flows to the SMBS portfolio.

As is common in this literature, there is no direct statistical support for the hypothesis that tail risk – combined with dividend yield – forecasts long-term variation in either the discount rate variation or the cash flow variation in either SMBL or SMBS. This is true whether we use asymptotic or bootstrap inference on the long-run coefficients. However, when we plot the fitted values of the long-run predictions based on point estimates in Figure 6, there are interesting differences between the short and long sides of SMB. Adding tail risk to the long-horizon results leads to economically substantial differences between the long-run expected return series. In the first half of the sample, a period dominated by the macroeconomic uncertainty of the 1970s and of the Fed experiment period through 1982, adding tail risk lowers expected returns on SMBL by as much as 1% per year. In the last half of the sample, dominated by the tech bubble and the housing crisis, adding tail risk to the forecasting model raises long-run expected returns by roughly 0.5% per year. For the large firms that make up the SMBL portfolio in Panel (b) of Table 6, there is relatively little difference between the long-run forecasts of the two models.

Tail risk forecasts both short- and long-run variation in SMBL and SMBS with positive coefficients. Therefore, it is important to note that its implications for both short- and long-run forecasts of the spread portfolio SMB are small.
Figure 6: Forecast of Long-Run Returns to SMB-Long & SMB-Short

The plots show the forecast (fitted value) of $r_{lt}^L = \sum_{j=1}^{\infty} \rho^{j-1}r_{t+j}$ for SMB-Long (Panel (a)) and SMB-Short (Panel (b)) using only the log dividend yield, $dp$, and using both $dp$ and the tail risk, $\lambda$. The long-run coefficient in the $dp$ only case is equal to $b'_{r} = \frac{b_{r}}{1-\rho_{d}}$, where $b_{r}$ is the coefficient in a regression of next year’s return on $MKT$ on a constant and $dp$, $\rho$ is the Campbell-Shiller approximation coefficient, and $\phi$ is the slope coefficient in the $dp$ first-order autoregression. See Cochrane (2008) for further discussion. The long-run coefficients in the case of both $dp$ and $\lambda$ are defined in eq. (6). The data are annual from 1965 to 2011.

5.3 Tail Risk and the Value Factor

The presentation of the results for the two components of the value factor, $HML$, follow the structure of the $SMB$ results, and they are shown in Table 6 and Figure 5.
Table 6: Forecasting Regressions for HML

**Panel A: The Long Portfolio of HML**

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>$t$-statistics</th>
<th>Other Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$dp_t$</td>
<td>$\lambda_t$</td>
<td>$dp_t$</td>
</tr>
<tr>
<td>(1) $r_{t+1}$</td>
<td>0.113</td>
<td>0.038</td>
</tr>
<tr>
<td>(2) $\Delta d_{t+1}$</td>
<td>0.031</td>
<td>0.001</td>
</tr>
<tr>
<td>(3) $dp_{t+1}$</td>
<td>0.961</td>
<td>−0.037</td>
</tr>
<tr>
<td>(4) $\lambda_{t+1}$</td>
<td>0.151</td>
<td>0.814</td>
</tr>
</tbody>
</table>

**Coefficients** | **$p$-Values for the Long-Run Coeffs.**
| $dp_t$ | $\lambda_t$ | Asymptotic | Bootstrap |
| $r_{t+1}^{lr}$ | 0.909 | 0.266 | 0.237 | 0.291 | 0.054 | 0.265 |
| $\Delta d_{t+1}^{lr}$ | 0.219 | −0.060 | 0.324 | 0.669 | 0.220 | 0.715 |

**Panel B: The Short Portfolio of HML**

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>$t$-statistics</th>
<th>Other Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$dp_t$</td>
<td>$\lambda_t$</td>
<td>$dp_t$</td>
</tr>
<tr>
<td>(1) $r_{t+1}$</td>
<td>0.130</td>
<td>0.052</td>
</tr>
<tr>
<td>(2) $\Delta d_{t+1}$</td>
<td>0.041</td>
<td>0.037</td>
</tr>
<tr>
<td>(3) $dp_{t+1}$</td>
<td>0.933</td>
<td>−0.015</td>
</tr>
<tr>
<td>(4) $\lambda_t$</td>
<td>−0.044</td>
<td>0.827</td>
</tr>
</tbody>
</table>

| Coefficients | **$p$-Values for the Long-Run Coeffs.**
| $dp_t$ | $\lambda_t$ | Asymptotic | Bootstrap |
| $r_{t+1}^{lr}$ | 0.340 | 0.333 | 0.301 | 0.193 | 0.267 | 0.202 |
| $\Delta d_{t+1}^{lr}$ | −0.023 | 0.089 | 0.532 | 0.226 | 0.271 | 0.009 |

This table reports the forecasting regression coefficients and long-run forecasting coefficients from a model that includes both dividend yield, $dp$, and tail risk, $\lambda$, as predictors. HML is the value factor constructed as in Fama-French. $\sigma \hat{y}_{t+1}$ is the volatility of the fitted value from the forecast, and $\sigma' \hat{y}_{t+1}$ is the additional volatility of the fitted value explained by tail risk alone. The long-run regression coefficients reported below the individual equation results are defined as $r_{t+1}^{lr} = \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$ and $\Delta d_{t+1}^{lr} = \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}$. The asymptotic $p$-values for the long-run coefficients are constructed from the individual $t$-statistics using the delta method as described in the Appendix 1. The bootstrap $p$-values are constructed from the $t$-statistics under the null hypothesis that tail has no short-run or long-run predictability for either cash flows or discount rates as described in Appendix 2. The data are annual from 1965 to 2011. Panel A shows the results for the long portion of HML, and Panel B shows the results for the short portion of HML.
Short-term predictability of “value” firms (the long side of $HML$, or “$HMLL$”) comes primarily from the log dividend yield. There is little statistical evidence in support of the tail measure forecasting future discount rates and even less evidence that it forecasts future dividend growth. A comparison of the volatility of the fitted value of the model based on dividend yield alone and the model that includes tail risk suggests that close to 85 percent of the variation in expected discount rates comes from dividend yield with the remainder accounted for by variation in tail risk. This is indirect evidence of the relative importance of the tail risk measure. The long-run predictability coefficients inferred from the VAR for $HMLL$ do not suggest that tail risk plays an important role in forecasting either discount rates or cash flows. The picture in Panel (a) of Figure 5 suggests that the only difference in forecasts between the model that includes tail risk and the model based on dividend yield alone is during the very early years of the sample from the late 1960s to the early 1970s.

Growth firms – which make up the short side of the $HML$ portfolio ($HMLS$) – exhibit short run predictability in discount rates but not dividend growth. This predictability is through dividend yields more than tail risk, which is not a statistically significant predictor of discount rate variation at conventional confidence levels. In the long-run, the heuristic evidence is more consistent with forecasting power of both dividend yield and tail risk for long-run expected returns. Indeed, Panel (b) of Figure 7 has a similar pattern in expected returns to Panel (a) in Table 6 (for $SMBL$). Adding tail risk to the forecasting model leads to substantially lower estimates of expected returns in the first half of the sample and substantially higher estimates of expected returns in the second half of the sample.
Figure 7: Forecast of Long-Run Returns to HML-Long & HML-Short

The plots show the forecast (fitted value) of \( r_{t}^{Lr} = \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} \) for HML-Long (Panel (a)) and HML-Short (Panel (b)) using only the log dividend yield, \( dp \), and using both \( dp \) and the tail risk, \( \lambda \). The long-run coefficient in the \( dp \) only case is equal to \( b_{r}^L = \frac{b_{r}}{1-\rho \phi} \), where \( b_{r} \) is the coefficient in a regression of next year’s return on \( MKT \) on a constant and \( dp \), \( \rho \) is the Campbell-Shiller approximation coefficient, and \( \phi \) is the slope coefficient in the \( dp \) first-order autoregression. See Cochrane (2008) for further discussion. The long-run coefficients in the case of both \( dp \) and \( \lambda \) are defined in eq. (6). The data are annual from 1965 to 2011.

In summary, the nature of the tail risk captured by \( \lambda_{t} \) is not clear from a careful examination of its conditional predictive implications. Since it has no apparent effect on cash flows – in either the short- or the long-term – it is not directly connected to a “fundamentals” or crash-based story of predictability. The tail risk measure seems to be forecasting components in the discount rates of both small- and large-cap firms and growth firms at both short- and long-horizons.

Of course this statement depends on the validity of the decomposition in equation (4) and measured cash flows. If discount rates also reflect market beliefs about long-run cash flows in different states in the future, then apparent discount rate movement might actually be indirect evidence of cash flow predictability.
6 Conclusions

Is there robust information about the cross-section of expected stock returns that can be extracted from the cross-section of the tails of stock returns? Kelly and Jiang (2014) propose a measure of aggregate tail risk and show that it forecasts future market returns at horizons of 1-month to 5-years. They also show that the return spread earned from differential exposure to their measure of aggregate tail risk is not priced by existing empirical factors.

What exactly do we learn from this aggregate tail risk measure? Kelly and Jiang (2014) motivate their measure, in part, by an appeal to existing models of “rare disasters” as explaining expected returns, but these models are based on large movements in the quantity of aggregate consumption. The cross-section of individual stock returns, however, may reflect information from a variety of sources and is not obviously connected with aggregate consumption – outside of a specific model.

In this paper, we have shown that a spread portfolio formed on exposure to the tail risk measure does not add any explanatory power to existing unconditional empirical models of the cross-section of expected returns. The univariate betas of industry and characteristics sorted portfolio does not offer any obvious intuition about what the tail index is measuring.

We also show, using the market portfolio and the Fama-French factor portfolios for size and value-growth, that the tail risk measure has short- and long-run predictive power for variation in the discount rates of small- and large-cap firms and growth firms. The KJ tail risk measure is not likely to be related to the rare disasters explanation of asset returns, which should have its primary impact through realized cash flows. Finally, since the coefficients on both sides of the long-short factor portfolios are positive, tail risk has limited conditional predictive power for $SMB$ and $HML$. This is consistent with its limited ability to price stocks (linearly) in the cross-section.

If the tail measure does not reflect a traditional disaster risk story, what is it capturing? The answer to this question is the subject of ongoing research. If it reflects the way that the pricing kernel values extreme states of nature – and surely it must – then there is no alternative to using a specific structural model (or set of structural models) to evaluate this claim. There have been a number of important recent advances in models of preferences that place differential value on extreme states; for example, rank-dependent expected utility and generalized disappointment aversion. The intriguing question is whether tail behavior has any power to distinguish between these alternative models.
Appendix 1: The Distribution of the Long-Run Regression Coefficients Constructed from the VAR

We rewrite the predictive VAR from eq. (5) in Section 5 in slightly different notation for convenience:

\[ x_{t+1} = \Phi x_t + \varepsilon_t^x, \]

\[ y_{t+1} = B x_t + \varepsilon_t^y \]

where \( x_t \equiv \begin{pmatrix} d p_t^{(i)} & \lambda_t \end{pmatrix}', y_t \equiv \begin{pmatrix} r_t^{(i)} & \Delta d_t^{(i)} \end{pmatrix}', \varepsilon_t^x = \begin{pmatrix} \varepsilon_t^{dp} & \varepsilon_t^{\lambda} \end{pmatrix}', \) and \( \varepsilon_t^y = \begin{pmatrix} \varepsilon_t^r & \varepsilon_t^{\Delta d} \end{pmatrix}'. \)

The standard OLS results imply

\[ \sqrt{T} \left( \text{vec} \left( \hat{\Phi}_T \right) - \text{vec} \left( \Phi \right) \right) \xrightarrow{d} \mathcal{N} \left( 0_{4 \times 1}, \Omega_x \otimes Q^{-1} \right), \]

and

\[ \sqrt{T} \left( \text{vec} \left( \hat{B}_T \right) - \text{vec} \left( B \right) \right) \xrightarrow{d} \mathcal{N} \left( 0_{4 \times 1}, \Omega_y \otimes Q^{-1} \right), \]

where \( \text{vec} (\bullet) \) is the operator that converts a matrix to a vector by stacking the columns of the matrix, \( \Omega_x = E \left( \varepsilon_t^x \varepsilon_t^x' \right), \Omega_y = E \left( \varepsilon_t^y \varepsilon_t^y' \right) Q = E \left( x_t x_t' \right), \otimes \) is the Kronecker product, \( \xrightarrow{d} \) denotes convergence in distribution, and \( \hat{\bullet} \) denotes an estimate based on a sample of size \( T. \)

Equation (6) can be written as

\[ \begin{bmatrix} \text{vec} \left( \hat{r}_T^{rl} \right) \\ \text{vec} \left( \hat{\Delta} d_T^{rl} \right) \end{bmatrix} = \text{vec} \left( B \left( I - \rho \Phi \right)^{-1} \right) = f \left( \begin{bmatrix} \text{vec} (B)' & \text{vec} (\Phi)' \end{bmatrix}' \right), \]

where we explicitly note that the long-run parameters are functions of the underlying VAR parameters. Applying the delta method, yields

\[ \begin{bmatrix} \text{vec} \left( \hat{r}_T^{rl} \right) \\ \text{vec} \left( \hat{\Delta} d_T^{rl} \right) \end{bmatrix} \xrightarrow{d} \mathcal{N} \left( \begin{bmatrix} \text{vec} \left( B \left( I - \rho \Phi \right)^{-1} \right), \Gamma \Sigma_{B\Phi} \Gamma' \end{bmatrix}_{4 \times 1} \right), \]

where \( \Gamma_{4 \times 8} \equiv \nabla f \) and

\[ \begin{bmatrix} \Sigma_B & : & 0_4 \\ \vdots & \cdot & \cdot & \cdot \\ 0_4 & : & \Sigma_{\Phi} \end{bmatrix} \]

with \( \Sigma_B \) the covariance matrix of the \( B \) coefficients, \( \Sigma_{\Phi} \) the covariance matrix of the \( \Phi \) coefficients, and \( 0_4 \) a \( 4 \times 4 \) matrix of zeros. So, the sample estimator of the asymptotic covariance matrix of the long-run coefficients is \( \hat{\Sigma} \hat{\Sigma}_{B\Phi} \hat{\Gamma}', \) and the asymptotic \( t \)-statistics are

---

18See, for example, Proposition 11.1 (and the regularity conditions stated there) from Hamilton (1994).
constructed using the elements on the main diagonal.

**Appendix 2: Bootstrapping the Long-Run $t$-Statistics Under the Null of No Predictability from $\lambda$**

In order to bootstrap the $t$-distribution on the long-run predictability coefficients in eq. (6), we assume that the tail measure does not predict either discount rates or cash flows at either the short or the long horizon. No short-horizon predictability from $\lambda$ is imposed by setting $b_{12} = b_{22} = 0$, where

$$B \equiv \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$  

Now, in order to impose the no predictability null on long-run predictability, begin with eq. (6)

$$\begin{bmatrix} r_{t} & \Delta d_{t}^{lr} \end{bmatrix} = B (I - \rho \Phi)^{-1} \begin{bmatrix} d_{t}^{(i)} \\ \lambda_{t} \end{bmatrix}.$$  

So, the null requires that the second column of $B (I - \rho \Phi)^{-1}$ consists of zeros. Let $Q \equiv (I - \rho \Phi)^{-1}$, then

$$B (I - \rho \Phi)^{-1} = BQ = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} = \begin{bmatrix} b_{11}q_{11} + b_{12}q_{21} & b_{11}q_{12} + b_{12}q_{22} \\ b_{21}q_{11} + b_{22}q_{21} & b_{21}q_{12} + b_{22}q_{22} \end{bmatrix}.$$  

So, imposing the no long-run predictability restriction requires

$$b_{11}q_{12} + b_{12}q_{22} = 0,$$

$$b_{21}q_{12} + b_{22}q_{22} = 0,$$

and using the short-run restriction this implies

$$b_{11}q_{12} = 0,$$

$$b_{21}q_{12} = 0.$$  

Using the definition of $Q$ implies that

$$q_{12} = \frac{\rho \phi_{12}}{(1 - \rho \phi_{11})(1 - \rho \phi_{22}) - \rho^2 \phi_{12} \phi_{21}}.$$  

One solution to this pair of equations is to set $b_{11} = b_{21} = 0$, but this is a short-run restriction on the predictability of discount rates and cash flows by dividend yield is counter-intuitive. Alternatively, the long-run restriction is satisfied if $\phi_{12} = 0$ which requires that the tail risk measure not forecast dividend yield in the short run. We impose this second restriction.
under the null.

For a given portfolio return, the bootstrap algorithm we implement the bootstrap as follows:

1. Estimate the full system in eq. (5) in order to generate the fitted residuals from the actual data.

2. Simulate 10,000 bootstrap samples (of the same length as the actual data) of the four residual series using a stationary (block) bootstrap.\(^{19}\)

3. Construct 10,000 samples under the null of no predictability by constructing the returns, dividend yields, dividend growth, and tail risk series using the VAR and the parameter restrictions described above.

4. Estimate the VAR model for each of the simulated series and construct the \(t\)-statistic for the long-run predictability coefficients. Save these simulated values to construct the distribution of the \(t\)-statistic for each test.

5. Compare the value of the \(t\)-statistic to the bootstrap distributions in order to calculate the bootstrap \(p\)-value.

\(^{19}\)We use the \texttt{stationary_bootstrap.m} program from the “MFE Toolbox” in Matlab written by Kevin Sheppard.
References


