Risk Premia and the Conditional Tails of Stock Returns*

(Job Market Paper)

Bryan Kelly†

December 2009

Abstract

Theory suggests that the risk of infrequent yet extreme events has a large impact on asset prices. Testing models of this hypothesis remains a challenge due to the difficulty of measuring tail risk fluctuations over time. I propose a new measure of time-varying tail risk that is motivated by asset pricing theory and is directly estimable from the cross section of returns. My procedure applies Hill's (1975) tail risk estimator to the cross section of extreme events each day. It then optimally averages recent cross-sectional Hill estimates to provide conditional tail risk forecasts. Empirically, my measure has strong predictive power for aggregate market returns, outperforming all commonly studied predictor variables. I find that a one standard deviation increase in tail risk forecasts an increase in excess market returns of 4.4% over the following year. Cross-sectionally, stocks that highly positively covary with my tail risk measure earn average annual returns 6.0% lower than stocks with low tail risk covariation. I show that these results are consistent with predictions from two structural models: i) a long run risks economy with heavy-tailed consumption and dividend growth shocks, and ii) a time-varying rare disaster framework.

*I thank my thesis committee, Robert Engle (chair), Xavier Gabaix, Alexander Ljungqvist and Stijn Van Nieuwerburgh for many valuable discussions. I also thank Mikhail Chernov, Itamar Drechsler, Marcin Kacperczyk, Anthony Lynch and Seth Pruitt for insightful comments.

†Department of Finance, Stern School of Business, NYU, 44 West Fourth Street, Suite 9-197, New York, NY 10012-1126. Office Phone: 212-998-0368. Cell Phone: 646-469-4466. E-mail: bkelly@stern.nyu.edu. Homepage: http://pages.stern.nyu.edu/∼bkelly.

1 Introduction

The mere potential for infrequent events of extreme magnitude can have important effects on asset prices. Tail risk, by nature, is an elusive quantity, which presents economists with the daunting task of explaining market behavior with rarely observed phenomena. This crux has led to notions such as peso problems (Krasker 1980) and the rare disaster hypothesis (Rietz 1988; Barro 2006), as well as skepticism about these theories due to the difficulty in testing them.

The goal of this paper is to investigate the effects of time-varying extreme event risk in asset markets. The chief obstacle to this investigation is a viable measure of tail risk over time. To overcome this, I devise a panel approach to estimating economy-wide conditional tail risk. Working from standard asset pricing models, I show that tail risks of all firms are driven by a common underlying process. Because individual returns contain information about the likelihood of market-wide extremes, the cross section of firms can be used to accurately measure prevailing tail risk in the economy. I elicit a conditional tail estimate by turning to the cross section of extreme events at each point in time, rather than waiting to accumulate a sufficient number of extreme observations in univariate time series. This bypasses data limitations faced by alternative estimators, for example those relying on options prices or intra-daily data.

My framework, which fuses asset pricing theory with extreme value econometrics, distills to a central postulate for the tail distribution of returns. Define the tail as the set of return events exceeding some high threshold u. I assume that the tail of asset return i behaves according to

$$P(R_{i,t+1} > r \mid R_{i,t+1} > u \text{ and } \mathcal{F}_t) = \left(\frac{r}{u}\right)^{-a_i\zeta_t}. \tag{1}$$

Equation 1 states that extreme return events obey a power law. Since at least Mandelbrot (1963) and Fama (1963), economists have argued that unconditional tail distributions of financial returns are aptly described by a power law. The key parameter of the model, $-a_i\zeta_t$, determines the shape of the tail and is referred to as the tail exponent. High values of $-a_i\zeta_t$ correspond to "fat" tails and high probabilities of extreme returns.

In contrast to past power law research, Equation 1 is a statement about the *conditional* return tail. The exponent varies over time because ζ_t is a function of the conditioning information set $\mathcal{F}_t$. While different assets have different levels of tail risk (determined by the constant

a_i), dynamics are the same for all assets because they are driven by a common conditional process. Thus, ζ_t may be thought of as economy-wide extreme event risk in returns. I refer to the tail structure in (1) as the dynamic power law model.

The tail generating process in (1) arises naturally from at least two structural models: i) a long run risks economy (Bansal and Yaron 2004) modified to include heavy-tailed shocks, and ii) a time-varying rare disaster framework. In my long run risks modification, non-Gaussian tails in consumption and dividend growth are governed by a new tail risk state variable, Λ_t. I show that expected excess returns depend linearly on the tail risk process. I then prove that the tail distribution of returns behaves according to the dynamic power law model. In particular, tail exponent fluctuations are the same for all assets and driven by Λ_t.

An attractive feature of the dynamic power law structure is that it emerges in varied theoretical settings rather than being tied to a single modeling paradigm. I demonstrate that a second structural model with unpredictable consumption growth and time-varying rare disasters, in the spirit of Gabaix (2009b) and Wachter (2009), delivers the dynamic power law structure for the lower tail of returns.

These structural models tightly link the time-varying tail exponent to expected excess returns on risky assets since both are driven by the tail risk process Λ_t. This generates two key testable implications. First, tail risk should positively forecast excess aggregate market returns. Aggregate dividends have substantial exposure to consumption tail risk. Thus, a positive tail risk shock increases the return required by investors to hold the market. Because the tail process is persistent, its shocks have a long-lived effect. Return forecastability arises because future expected excess returns remain high until the expectations effect of a tail risk shock dies out.

The second testable prediction applies to the cross section of expected returns. High tail risk is associated with bad states of the world and high marginal utility. This implies that the price of tail risk is negative, hence assets with high betas on the tail risk process will have lower expected returns than assets with low tail risk betas. Intuitively, an asset whose return covaries highly with tail risk has a tendency to payoff in adverse states, serving as an effective tail risk hedge. As a result, it commands a high price and earns relatively low average returns.

I build an econometric estimator for the dynamic power law structure suggested by these economic frameworks. The intuition from structural models is that tail risks of individual assets are closely related to aggregate tail risk. In a sufficiently large cross section, enough stocks will experience tail events each period to provide accurate information about the prevailing level of tail risk. I use this cross-sectional extreme return information to estimate economy-wide tail risk at each point in time. This avoids having to accumulate years of tail observations from the aggregate market time series, and therefore avoids using stale observations that carry little information about current tail risk.

My procedure applies Hill's (1975) tail risk estimator to the cross section of extreme events each day. The model then optimally averages recent cross-sectional Hill estimates to provide conditional tail risk forecasts. A major obstacle in estimation is the model's potentially enormous number of nuisance parameters. I overcome this with a strategy based on quasi-maximum likelihood theory. The idea is to find a simpler version of the infeasible model that has the same maximum likelihood first order conditions. Estimation may then be based on this mis-specified, yet feasible, model. I reduce the complexity of the problem to three parameters by treating observations as though they are i.i.d. I then prove that maximizing the resulting quasi-likelihood provides consistent and asymptotically normal estimates of the data's true dynamics, and show how to calculate standard errors for inference.

I implement the dynamic power law estimator using daily returns from the cross section of CRSP stocks. I find that the cross-sectional average tail exponent is highly persistent and fluctuates between -4 and -1.5. This range is consistent with a survey by Gabaix (2009a), who finds that estimates of unconditional tail exponents consistently hover around -3 based on data for a variety of domestic and international equities. My estimates of lower and upper tail risk are positively correlated (56% monthly). There is evidence of cyclicality in lower (upper) tail risk as it shares a monthly correlation of 53% (39%) with unemployment, 15% (14%) with the aggregate log dividend-price ratio, and -10% (-7%) with the Chicago Fed National Activity Index.

Using the fitted tail series, I first test the model prediction that tail risk should forecast aggregate stock market returns. Predictive regressions show that a one standard deviation increase in the risk of negative tail events forecasts an increase in annualized excess market returns of 6.7%, 4.4%, 4.5% and 5.0% at the one month, one year, three year and five year horizons, respectively. These are all statistically significant with t-statistics of 2.9, 2.1, 2.2 and 2.3, based on Hodrick's (1992) standard error correction. At the monthly frequency, tail risk achieves an R^2 of 1.6% in-sample and 1.3% out-of-sample, outperforming the price-dividend ratio and other common predictors. Cochrane (1999) and Campbell and

Thompson (2008) argue that a monthly R^2 of this magnitude has large economic significance. A heuristic calculation suggests that a 1.3% R^2 can generate a 50% improvement in Sharpe ratio over a buy-and-hold investor.[1] My tail risk measure outperforms all commonly considered forecasting variables in terms of predictive power, including the log dividend-price ratio and fourteen other variables surveyed by Goyal and Welch (2008). Estimated tail risk coefficients and their statistical significance are robust to controlling for these alternative predictors. They are also robust to using an alternative tail series based on factor model residuals rather than raw returns.

The tail exponent has large, significant explanatory power for the cross section of returns in the direction predicted by theory. In my first test, I sort stocks into quintiles based on their estimated beta on the tail risk process. I find that stocks with the highest tail risk betas earn average annual returns 6.0% lower than the lowest tail risk beta stocks (t-statistic=2.6). This negative tail risk premium is robust to double sorts based on tail risk beta and i) market beta, ii) size or iii) book-to-market ratio. To simultaneously control for multiple alternative factors, I next test the tail risk premium with Fama and MacBeth (1973) regressions. Based on NYSE, AMEX and NASDAQ stocks, I find that a stock whose tail beta is one standard deviation above the cross-sectional mean has an annual expected return 5.6% lower than the average stock (t-statistic=2.9) after controlling for market volatility and Fama and French (1993) factors, consistent with theory and with results based on portfolio sorts. These findings are also robust to estimating the tail exponent with factor model residuals, and to using alternative sets of test assets.

My research draws on several literatures. Theoretically, I build on the long run risks literature of Bansal and Yaron (2004). My framework is most closely related to recent long run risks extensions that accommodate more sophisticated descriptions of consumption growth, particularly Eraker and Shaliastovich (2008), Bansal and Shaliastovich (2008, 2009), and Drechsler and Yaron (2009). The Rietz-Barro hypothesis and its extensions to dynamic settings by Gabaix (2009b) and Wachter (2009) are also important predecessors of the ideas developed here.

A large literature has modeled extreme returns using jump processes with time-varying

[1] Cochrane shows that the Sharpe ratio (s^*) earned by an active investor exploiting predictive information (summarized by a regression R^2) and the Sharpe ratio (s_0) earned by a buy-and-hold investor are related by $s^* = \frac{\sqrt{s_0^2+R^2}}{\sqrt{1-R^2}}$. Campbell and Thompson estimate a monthly equity buy-and-hold Sharpe ratio of 0.108 using data back to 1871. Therefore, a predictive R^2 of 1.3% implies $s^* = 0.162$, an improvement of 50% over the buy-and-hold value.

intensities that depend on observable state variables, including the widely used affine class of Duffie, Pan and Singleton (2000). My approach, which models conditional tails in discrete time and uses observable parameter updates based on the history of extreme returns, is new. Furthermore, the notion of extracting information about common, time-varying tails from the cross section of returns is novel, though similar in spirit to the identification strategy in Engle and Kelly (2009).

Recently, economists have extracted tail risk estimates from options prices or intra-daily returns to assess the relation between rare events and equity prices. Bollerslev, Tauchen and Zhou (2009) examine how the variance risk premium implicit in index option prices relates to the equity premium. Backus, Chernov and Martin (2009) use equity index options to extract higher order return cumulants and draw inferences about disaster risk premia over time. Using ultra-high frequency data for S&P 500 futures, Bollerslev and Todorov (2009) calculate realized and risk neutral tail risk measures to explain equity and variance risk premia. In these cases, researchers are bound by data limitations that restrict the sample horizon to at most twenty years. In contrast, my tail risk series is estimated using data from 1962 to 2008, and in general may be used whenever a sufficiently large cross section of returns is available.

Lastly, I contribute to a literature that attempts to jointly explain behavior of returns in the time series and cross section, including (among others) Ferson and Harvey (1991), Lettau and Ludvigson (2001a,b), Lustig and Van Nieuwerburgh (2005) and Koijen, Lustig and Van Nieuwerburgh (2009).

2 Asset Pricing Theory and Conditional Return Tails

In this section I develop two consumption-based asset pricing models. Both generate a dynamic power law structure in the tail distribution of returns and produce clear testable implications for the link between tail risk and risk premia. The first is a modification of the long run risks economy of Bansal and Yaron (2004). In addition to the standard aspects of the Bansal-Yaron formulation, I allow for non-Gaussian shocks to both consumption growth and idiosyncratic dividend growth. My specification of tail risk leads to tractable expressions both for prices and for the tail distribution of returns.

The second example economy is a version of the time-varying rare disaster model of Gabaix

(2009b) and Wachter (2009). This example is included to demonstrate the flexibility of the dynamic power law structure for encompassing a broad set of distinct economic models of returns.

2.1 Economy I: Long Run Risks and Tail Risk in Cash Flows

Investor preferences over consumption are recursive (Epstein and Zin 1989). These are summarized by the economy-wide intertemporal marginal rate of substitution, which is also the stochastic discount factor that prices assets in the economy. Written in its log form, this is

$$m_{t+1} = \theta \ln \beta - \frac{\theta}{\psi} \Delta c_{t+1} + (\theta - 1) r_{c,t+1},$$

where $\theta = \frac{1-\gamma}{1-\frac{1}{\psi}}$, γ is the risk aversion coefficient, ψ is the intertemporal elasticity of substitution (IES), Δc_{t+1} is log consumption growth, and $r_{c,t+1}$ is the log return on an asset paying aggregate consumption each period. Throughout the paper I assume $\gamma > 1$ and $\psi > 1$, which implies $\theta < 0$. These parameter restrictions ensure that risk aversion is greater than the reciprocal of IES, therefore agents have a preference for early uncertainty resolution.

Dynamics of the real economy are

$$\begin{aligned}
\Delta c_{t+1} &= \mu + x_t + \sigma_c \sigma_t z_{c,t+1} + \sqrt{\Lambda_t} W_{c,t+1} \\
x_{t+1} &= \rho_x x_t + \sigma_x \sigma_t z_{x,t+1} \\
\sigma^2_{t+1} &= \bar{\sigma}^2 (1 - \rho_\sigma) + \rho_\sigma \sigma^2_t + \sigma_\sigma z_{\sigma,t+1} \\
\Lambda_{t+1} &= \bar{\Lambda}(1 - \rho_\Lambda) + \rho_\Lambda \Lambda_t + \sigma_\Lambda z_{\Lambda,t+1} \\
\Delta d_{i,t+1} &= \mu_i + \phi_i \Delta c_{t+1} + \sigma_i \sigma_t z_{i,t+1} + q_i \sqrt{\Lambda_t} W_{i,t+1}. \qquad (2)
\end{aligned}$$

Included in this specification are standard elements of a long run risks model: log consumption growth (Δc_{t+1}), its persistent conditional mean (x_t) and volatility (σ_t), and dividend growth for asset i ($\Delta d_{i,t+1}$). The z shocks are standard normal and independent. In addition to their Gaussian shocks, consumption and dividend growth depend on non-Gaussian shocks, W. These are independent unit Laplace variables with mean zero. Their density results from splicing the densities of independent positive and negative unit exponentials together at zero,

$$f_W(w) = \frac{1}{2} \exp(-|w|), \; w \in \mathbb{R}.$$

The Laplace shocks dominate the tails of cash flow growth. To see this, consider the tail distribution of $Z = X + Y$, where $X \sim N(0, \tau)$ and Y is an independent Laplace variable with scale parameter α. It may be shown that the tail behavior of Z is determined solely by the Laplace summand, $P(Z > u + \eta | Z > u) \sim \exp(-\alpha\eta)$. The relation $f(u) \sim g(u)$, read "f is asymptotically equivalent (or tail equivalent) to g", denotes $\lim_{u\to\infty} f(u)/g(u) = 1$.

Heavy-tailed consumption growth and dividend growth shocks $W_{c,t+1}$ and $W_{i,t+1}$ are scaled by $\sqrt{\Lambda_t}$ and $q_i\sqrt{\Lambda_t}$, respectively. As the Gaussian stochastic process Λ_t evolves, the risk of extreme cash flow events fluctuates. High values of Λ_t fatten the tails of cash flow shocks while low values shrink cash flow tails; consequently, I refer to Λ_t as the tail risk process.

I solve the model with procedures commonly employed in consumption-based affine pricing models, following Bansal and Yaron (2004), Eraker and Shaliastovich (2008), and Bollerslev, Tauchen and Zhou (2009), among others.[2] The first result proves that log valuation ratios in the economy are linear in the tail risk process.

Proposition 1. *The log wealth-consumption ratio and log price-dividend ratio for asset i are linear in state variables,*

$$wc_{t+1} = A_0 + A_x x_{t+1} + A_\sigma \sigma^2_{t+1} + A_\Lambda \Lambda_{t+1} \tag{3}$$

$$pd_{i,t+1} = A_{i,0} + A_{i,x} x_{t+1} + A_{i,\sigma} \sigma^2_{t+1} + A_{i,\Lambda} \Lambda_{t+1}.$$

Proofs are relegated to Appendix A (including expressions for the A constants).

Risk premia for an asset are determined by covariation between m_{t+1} and the asset's log return. Hence, the coefficients on shocks to the stochastic discount factor take on the interpretation of risk prices. Discount factor shocks are

$$m_{t+1} - E_t[m_{t+1}] \;=\; -\lambda_c(\sigma_c \sigma_t z_{c,t+1} + \sqrt{\Lambda_t} W_{c,t+1}) - \lambda_x \sigma_x \sigma_t z_{x,t+1} - \lambda_\sigma \sigma_\sigma z_{\sigma,t+1} - \lambda_\Lambda \sigma_\Lambda z_{\Lambda,t+1}.$$

The term λ_Λ captures the price of tail risk (risk price expressions are also shown in Appendix A). The tail risk price is negative since an increase in uncertainty decreases agents' utility. Assets that covary positively with tail risk behave as hedges since they tend to pay off when marginal utility is high. Since tail risk hedges are valuable to investors, they command higher

[2] Analytical results are stated subject to linear approximations such as the log return identity of Campbell and Shiller (1988), as used in the aforementioned articles.

prices and lower expected returns, *ceteris paribus*. The next result shows that an asset's risk premium is a linear function of variance and tail risk.

Proposition 2. *The expected return on asset i in excess of the risk free rate is*

$$E_t[r_{i,t+1} - r_{f,t}] = \beta_{i,c}\lambda_c(\sigma_c^2\sigma_t^2 + 2\Lambda_t) + \beta_{i,x}\lambda_x\sigma_x^2\sigma_t^2 + \beta_{i,\sigma}\lambda_\sigma\sigma_\sigma^2 + \beta_{i,\Lambda}\lambda_\Lambda\sigma_\Lambda^2 - \frac{1}{2}Var(r_{i,t+1}). \quad (4)$$

Proposition 2 describes both the time series and cross-sectional relation between tail risk and expected returns. First, Equation 4 is a predictive regression that implies returns are forecastable by variance and tail risk in equilibrium. This is a natural consequence of predictable changes in compensation that investors require in order to bear these risks, and is not an arbitrage opportunity or violation of efficient markets. Let subscript m denote the asset that pays aggregate dividends (i.e., the market portfolio). In the modified long run risks model, the predictive regression coefficient on tail risk is $2\beta_{m,c}\lambda_c$. The constant $\beta_{m,c}$ is positive as long as the exposure of aggregate dividend growth to consumption growth is greater than the reciprocal of IES ($\phi_m > 1/\psi$). This is typically assumed in calibrations of long run risks models, where aggregate dividends are treated as levered claims on consumption ($\phi_m > 1$, as in Bansal and Yaron 2004). The price of transitory consumption risk λ_c is also positive, which delivers the intuitive implication that the predictive coefficient is positive. Higher tail risk increases the return investors require to hold the market portfolio going forward.[3]

The term $\beta_{i,\Lambda}\lambda_\Lambda\sigma_\Lambda^2$ governs cross-sectional differences in expected return due to differential exposure to the tail risk process. Because tail risk carries a negative price of risk, Equation 4 generates the prediction that stocks with high tail risk betas ($\beta_{i,\Lambda}$) earn a negative risk premium over low tail risk beta stocks. High tail risk beta stocks perform well when tail risk is high. Because tail risk is utility decreasing, these stocks serve as effective hedges. They therefore command a high price and earn low expected returns.

Using the price expressions in Proposition 1, I show that the tail distribution of arithmetic returns for each stock satisfies the dynamic power law structure in Equation 1.

Proposition 3. *The lower and upper tail distributions of arithmetic returns are asymptoti-*

[3] The predictive regression will also be affected by the Jensen term $\frac{1}{2}Var(r_{i,t+1})$ since it is a function of Λ_t. As long as the Jensen effect is small relative to the risk premium effect ($2\beta_{m,c}\lambda_c$), the positive sign of the predictive coefficient will still hold.

cally equivalent to a power law,

$$P_t(R_{i,t+1} < r \mid R_{i,t+1} < u) \sim \left(\frac{r}{u}\right)^{a_i\zeta_t}$$

$$P_t(R_{i,t+1} > r \mid R_{i,t+1} > u) \sim \left(\frac{r}{u}\right)^{-a_i\zeta_t},$$

where $a_i = \max(\phi_i, q_i)^{-1}$ *and* $\zeta_t = 1/\sqrt{\Lambda_t}$.

Here the relation ($\sim$) describes tail equivalence at the lower and upper support boundaries of $R_{i,t+1}$ (i.e., $\lim_{u\downarrow 0} f(u)/g(u) = 1$ for the lower tail or $\lim_{u\to\infty} f(u)/g(u) = 1$ for the upper tail). The value r is assumed to vary in fixed proportion with u (i.e., $r = u\eta$, $\eta > 0$). As $a_i\zeta_t$ decreases, both the upper and lower tails become fatter.[4] This proposition demonstrates the link between tail risk in cash flows and returns. The process governing consumption and dividend growth tail risk, Λ_t, drives time variation in the parameter governing the tail distribution of returns, ζ_t. The key insight of this result is that extreme event risk in the real economy may be estimated from the tail distribution of returns. It also means that the model implications stated above may be tested based on parameter estimates for the conditional tail distribution of returns.

2.2 Economy II: Time-Varying Rare Disasters

In this subsection I develop an economy with variable rare disasters in consumption growth and idiosyncratic dividend growth. I am brief in my exposition of the rare disaster model as much of the intuition carries over from the first model economy.

Investor preferences again take the Epstein-Zin form. In contrast to the long run risks model, consumption growth is unpredictable in a rare disaster economy. In most periods, consumption growth is Gaussian. Upon the rare occurrence of a disaster, consumption growth experiences a heavy-tailed negative shock. The severity of disasters varies through time, so that consumption growth takes the form

$$\Delta c_{t+1} = \mu + \sigma_c\sigma_t z_{c,t+1} - \iota_{c,t+1}\Lambda_t V_{c,t+1}.$$

[4]Note that with a trivial reformulation, the lower tail of the distribution can be written identically to the upper tail distribution. This is done by reversing the sign of the lower tail of log returns before exponentiating, which is clear from the proof in Appendix A.

The first shock $z_{c,t+1}$ is standard normal. In non-disaster times, variations in the consumption growth distribution arise only from heteroskedasticity (σ^2_{t+1}, which is unchanged from system 2). The second shock, $V_{c,t+1}$, is the disaster shock. It is drawn from the relatively heavy-tailed unit exponential distribution. $V_{c,t+1}$ is first multiplied by a *Bernoulli*(δ) random variable, $\iota_{c,t+1}$, that determines whether or not a disaster occurs at $t+1$.[5] The occurrence of a disaster therefore follows the distribution

$$\iota_{c,t+1} = \begin{cases} 1 & \textit{with probability} \quad \delta \\ 0 & \textit{with probability} \quad 1-\delta. \end{cases}$$

Second, $V_{c,t+1}$ is multiplied by Λ_t, which determines the severity of a disaster. In this context, I refer to Λ_t as the disaster risk process. It is stochastic and evolves as in system 2.

Dividend growth of stock i is given by

$$\Delta d_{i,t+1} = \mu_i + \phi_i \Delta c_{t+1} + \sigma_i \sigma_t z_{i,t+1} - q_i \iota_{i,t+1} \Lambda_t V_{i,t+1}.$$

As in the time-varying rare disaster model of Gabaix (2009b), individual stock dividends have exposure to aggregate consumption, and hence aggregate disaster risk. I also allow for the possibility of idiosyncratic dividend disasters. These are associated with severe negative shocks to firms' idiosyncratic payoffs that are independent of the broader economy. Rare idiosyncratic disasters occur through the shock $q_i \iota_{i,t+1} \Lambda_t V_{i,t+1}$, where $\iota_{i,t+1}$ is an i.i.d. *Bernoulli*(δ) variable and q_i determines the magnitude of stock i's idiosyncratic disasters relative to consumption disasters.

The following result demonstrates that log valuation ratios are linear in disaster risk.

Proposition 4. *The log wealth-consumption ratio and log price-dividend ratio of asset i are linear in state variables,*

$$wc_{t+1} = A_0 + A_\sigma \sigma^2_{t+1} + A_\Lambda \Lambda_{t+1} \tag{5}$$

$$pd_{i,t+1} = A_{i,0} + A_{i,\sigma} \sigma^2_{t+1} + A_{i,\Lambda} \Lambda_{t+1}.$$

Expressions for the A coefficients are found in Appendix A. As in the long run risks model, $(1-\theta)\kappa_1 A_\Lambda$ may be interpreted as the price of disaster risk. The proof of Proposition 4 shows that $A_\Lambda < 0$, which implies that disaster risk has a negative price. The next result shows that the risk premium for each asset is linear in disaster risk.

[5]It is straight-forward to allow for a time-varying disaster probability δ_t.

Proposition 5. *The expected return on asset i in excess of the risk free rate is*

$$E_t[r_{i,t+1} - r_{f,t}] = r_{i,0} + b_{i,\sigma}\sigma_t^2 + b_{i,\Lambda}\Lambda_t. \quad (6)$$

Lastly, I prove that the lower tail structure of returns generated by the time-varying rare disaster model is consistent with the dynamic power law model of Equation 1.

Proposition 6. *The lower tail distribution of arithmetic returns is asymptotically equivalent to a power law,*[6]

$$P_t(R_{i,t+1} < r \mid R_{i,t+1} < u) \sim \left(\frac{r}{u}\right)^{a_i\zeta_t}$$

where $a_i = \max(\phi_i, q_i)^{-1}$ *and* $\zeta_t = 1/\Lambda_t$.

Because the disaster shock is strictly negative, only the lower tail of returns exhibits power law behavior. The upper tail is lognormal since large upside moves result only from Gaussian shocks. The model can be easily reformulated to accommodate time-varying rare booms alongside rare disasters. The qualitative effects of tail risk in this case are similar to those from the disaster and long run risks models I've presented.

2.3 Testable Implications

To summarize, structural economic models predict a close link between the risk of extreme events in the real economy, Λ_t, and risk premia across assets and over time. Direct estimation of conditional tail risk from consumption and dividend data is essentially infeasible due to their infrequent observation and poor measurement. The two structural models presented here highlight the path to an alternative estimation strategy since the power law tail exponent of stock returns, $-\zeta_t$, is also driven by Λ_t. Because returns are frequently and precisely observed, estimates of their tail distribution identify the Λ_t process. Most importantly, the tight structure that these models place on the return tail distribution implies that the cross section can be exploited to extract conditional tail risk estimates at high frequencies.

The model-implied pricing effects of tail risk can be tested with estimates of the time-varying component in return power law exponents. The exponent series $-\zeta_t$ is an increasing function

[6]As in Propostion 3, the value r is assumed to vary in fixed proportion with u ($r = u\eta$, $\eta > 0$).

of Λ_t, so that when cash flow tail risk rises, return tails become fatter.[7] A preliminary assumption of the economic model is that economy-wide tail risk Λ_t varies persistently through time, which implies that $-\zeta_t$ should as well. This assumption is testable.

Testable Implication 1. *The dynamic power law exponent $-\zeta_t$ is time-varying and persistent.*

The next implication applies to the equity premium time series. Equations 4 and 6 imply that aggregate tail risk forecasts excess returns on the market portfolio. Because aggregate dividends have substantial exposure to consumption tail risk, a positive tail risk shock increases the return required by investors to hold the market. Since the tail process is persistent, future expected excess returns remain high until the expectations effect of a tail risk shock dies out.

Testable Implication 2. *The tail risk series $-\zeta_t$ positively forecasts excess market returns.*

Next, because tail risk detracts from utility, it carries a negative price of risk. This generates the cross-sectional prediction that stocks with positive tail risk betas earn a negative risk premium. This should also be true of the tail risk proxy $-\zeta_t$.

Testable Implication 3. *Stock with high betas on the tail risk process $-\zeta_t$ earn a negative risk premium in relation to those with low tail risk betas.*

3 Empirical Methodology

3.1 The Dynamic Power Law Model

In this section I propose a procedure for estimating the dynamic power law model. My approach exploits the comparatively rich information about tail risk in the cross section of

[7] Different models imply different relations between the tail exponent, $-\zeta_t$, and tail risk in fundamentals, Λ_t. In the long run risks model $\Lambda_t = 1/\zeta_t^2$, while in the rare disaster model $\Lambda_t = 1/\zeta_t$. Specifications of the Λ_t process are flexible and can generate a wide range of functional forms linking Λ_t and $-\zeta$. Under any specification, however, it is the case that $-\zeta_t$ is increasing in Λ_t. Rather than rely on a model-specific functional link between these two, I test model implications using the estimated tail exponent $-\zeta_t$ without further transformation. My empirical results are robust to transformations based on the economic models here.

returns, as opposed to relying, for example, on short samples of high frequency univariate data or options prices.

Estimating fully-specified versions of the tail models from Section 2 is extremely difficult, and essentially infeasible without multi-step estimation. It requires specifying a dependence structure among return tails and estimating stock-specific a_i parameters. Incorporating both considerations adds an enormous number of parameters: Estimating the a_i constants adds n parameters while imposing dependence structures like those implied by the models in Section 2 adds another nK parameters, where K is the number of factors in a given model.[8] Furthermore, these parameters are nuisances since the goal is to measure the common element of tail risk, not univariate distributions. The stochastic nature of the tail risk process further complicates estimation. Contemporaneous shocks to the tail exponent can be thought of as the extreme value equivalent of stochastic volatility. It rules out simple likelihood methods, instead requiring computation-intensive procedures like simulation-based estimation.

I propose several simplifications of the dynamic power law model that isolate the common component of tail risk with a tractable, accurate procedure. My simplification requires estimating only three parameters instead of several thousand, and reduces estimation time to under one minute despite working with a daily cross section of several thousand stocks. To begin, I more fully specify the statistical model, including an evolution equation for the tail exponent (which was left unspecified in Equation 1).

Assumption 1 (Dynamic Power Law Model)**.** *Let* $R_t = (R_{1,t}, ..., R_{n,t})'$ *denote the cross section of returns in period* t*.*[9] *Let* K_t *denote the number of* R_t *elements exceeding threshold* u *in period* t*.*[10] *The tail of individual returns for stock* i *(*$i = 1, ..., n$*), conditional upon*

[8] In Section 2, returns obey a tail factor structure due to the factor structure in heavy-tailed cash flow shocks. The common heavy-tailed shock enters via each firm's exposure to consumption growth while the heavy-tailed idiosyncrasy comes from firm-specific dividend shocks. In that case, $K = 1$.

[9] R denotes arithmetic return, which directly maps the tail distribution here with theoretical results in Section 2. This is without loss of generality as the model is equally applicable to log returns. In estimation, I work with daily returns. Because of the small scale of daily returns, the approximation $\ln(1 + x) \approx x$ is accurate to a high order and the distinction between arithmetic and log returns is negligible.

[10] I assume for notational simplicity that these are the first K_t elements of R_t. This is immaterial since, in the treatment here, elements of R_t are exchangeable.

exceeding u and given information $\mathcal{F}_t$, obeys the probability distribution[11]

$$F_{u,i,t}(r) = P(R_{i,t+1} > r | R_{i,t+1} > u, \mathcal{F}_t) = \left(\frac{r}{u}\right)^{-a_i\zeta_t}$$

with corresponding density

$$f_{u,i,t}(r) = \frac{a_i\zeta_t}{u}\left(\frac{r}{u}\right)^{-(1+a_i\zeta_t)}.$$

The common element of exponent processes, ζ_{t+1}, evolves according to

$$\frac{1}{\zeta_{t+1}} = \pi_0 + \pi_1\frac{1}{\zeta_t^{upd}} + \pi_2\frac{1}{\zeta_t} \tag{7}$$

and the observable update of ζ_{t+1} is

$$\frac{1}{\zeta_t^{upd}} = \frac{1}{K_t}\sum_{k=1}^{K_t}\ln\frac{R_{k,t}}{u}.$$

The evolution of ζ_t is designed to capture autoregressive time series behavior with a parsimonious parameterization. The update term $1/\zeta_t^{upd}$ is a summary statistic calculated from the cross section of tail observations on date t, which I discuss in more detail below. Recursively substituting for ζ_t shows that

$$\frac{1}{\zeta_{t+1}} = \frac{\pi_0}{1-\pi_2} + \pi_1\sum_{j=0}^{\infty}\pi_2^j\frac{1}{\zeta_{t-j}^{upd}}.$$

Thus, $1/\zeta_{t+1}$ is simply an exponentially-weighted moving average of daily updates based on observed extreme returns. When the π coefficients are estimated with maximum likelihood, this moving average is an optimal forecast of future tail risk.

The role of the update is to summarize information about prevailing tail risk from recent extreme return observations. This conditioning information enters the evolution of $1/\zeta_{t+1}$ via the summary statistic to refresh the conditional tail measure. I calculate a summary of tail risk from the cross section each period using Hill's (1975) estimator. The Hill estimator

[11]This formulation applies similarly to the lower tail of returns. When estimating the lower tail, I reverse the sign of log returns and estimate the upper tail of the model. This is customary in extreme value statistics, as it streamlines exposition of models as well as computer code used in estimation.

is a maximum likelihood estimator of the cross-sectional tail distribution. It takes the form

$$\frac{1}{\zeta_t^{Hill}} = \frac{1}{K_t} \sum_{k=1}^{K_t} \ln \frac{R_{k,t}}{u}.$$

To see why this makes sense as an update, note that when u-exceedences (i.e., $R_{i,t+1}/u$) obey a power law with exponent $-a_i\zeta_t$, the *log* exceedence is exponentially distributed with scale parameter $a_i\zeta_t$. By the properties of an exponential random variable, $E[\ln(R_{i,t+1}/u)] = 1/(a_i\zeta_t)$. As a consequence, the expected value of update $1/\zeta_{t+1}^{upd}$ is the cross-sectional harmonic average tail exponent,

$$E_t\left[\frac{1}{K_{t+1}} \sum_{k=1}^{K_{t+1}} \ln \frac{R_{k,t+1}}{u}\right] = \frac{1}{\bar{a}\zeta_t}, \quad \text{where} \quad \frac{1}{\bar{a}} \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{1}{a_i}.$$

This important property will be used to establish consistency and asymptotic normality of the dynamic power law estimation procedure that follows.

Before proceeding to the estimation approach, note that Equation 7 is a stochastic process because ζ_{t+1}^{upd} is a function of time t returns. However, ζ_{t+1} is deterministic conditional upon time-t information. This is different than the specification in the structural models of Section 2, which imply that ζ_{t+1} is subject to a $t+1$ shock. I argue that this discrepancy is largely innocuous. In the limit of small time intervals, tail risk processes in the structural models and the exponent process in the econometric model can be specified to line up exactly. A conditionally deterministic tail exponent process, then, can be thought of as a discrete time approximation to a continuous time stochastic process. The advantage of the approximation is that straight-forward likelihood maximization procedures can be used for estimation. This property is the tail analogue to the relation between GARCH models (in which volatility is conditionally deterministic) and stochastic volatility models. Nelson (1990) shows that a discrete GARCH(1,1) return process converges to a stochastic volatility process as the time interval shrinks to zero. An important result of Drost and Werker (1996) proves that estimates of a GARCH model at *any* discrete frequency completely characterize the parameters of its continuous time stochastic volatility equivalent. The same notion lies behind treating the process in (7) as a discrete time approximation to a continuous time stochastic process for the tail exponent.[12]

[12] I conduct a Monte Carlo experiment to examine the tail analogue of the Drost and Werker result. I find that the dynamic power law estimator based on the model in Assumption 1 continues to provide accurate estimates of the tail exponent process when the exponent follows a Gaussian autoregression. I discuss this

3.2 Estimating the Dynamic Power Law Model

My estimation strategy uses a quasi-likelihood technique, and is an example of a widely used econometric method with early examples dating at least back to Neyman and Scott (1948), Berk (1966), and the in-depth development of White (1982). The general idea is to use a partial or even mis-specified likelihood to consistently estimate an otherwise intractable model. The proofs that I present can also be thought of as a special case of Hansen's (1982) GMM theory. To avoid the nuisance parameter problem, I treat assets as though they are independent with identical tail distributions each period. The independence assumption avoids the need to estimate factor loadings for each stock, and the identical assumption avoids having to estimate each a_i coefficient. These simplifications, however, alter the likelihood from the true likelihood associated with Assumption 1, to a "quasi"-likelihood, written below. I show that maximizing the quasi-likelihood produces consistent and asymptotically normal estimates for the parameters that govern tail dynamics, π_1 and π_2. Ultimately, the estimated ζ_t series is shown to be the fitted cross-sectional harmonic average tail exponent. Since the average exponent series differs from ζ_t only by the multiplicative factor $\bar{a}$, the two are perfectly correlated.

Before stating the main proposition I discuss two important objects, the log quasi-likelihood and the "score" function (the derivative of the log quasi-likelihood with respect to model parameters). I refer to the tail model in Assumption 1 as the "true" model. Suppose, counterfactually, that all returns in the cross section share the same exponent, which is equal to the cross-sectional harmonic average exponent. Then the tail distribution of all assets becomes

$$\tilde{F}_{u,i,t}(R_{i,t+1}) = \left(\frac{r}{u}\right)^{-\bar{a}\zeta_t}$$

with corresponding density

$$\tilde{f}_{u,i,t}(R_{i,t+1}) = \frac{\bar{a}\zeta_t}{u}\left(\frac{r}{u}\right)^{-(1+\bar{a}\zeta_t)}.$$

Tildes signify that this distribution is different than the true marginal, $F_{u,i,t}$. Under cross-

more at the end of the section, and provide a detailed description of the experiment and its results in Appendix B.

sectional independence, the corresponding (scaled) log quasi-likelihood is

$$\mathcal{L}(\{R_t\}_{t=1}^T;\pi) = \frac{1}{T}\sum_{t=0}^{T-1} \ln \tilde{f}(R_{t+1};\pi,\mathcal{F}_t) = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{k=1}^{K_{t+1}} \left(\frac{\bar{a}}{\zeta_t} - \ln\frac{R_{k,t+1}}{u}\right), \tag{8}$$

where u-exceedences are included in the likelihood and non-exceedences are discarded. Define the gradient of $\ln \tilde{f}_t(R_{t+1};\pi)$ with respect to π (the time-t element of the score function) as

$$s_t(R_{t+1};\pi) \equiv \nabla_\pi \ln \tilde{f}_t(R_{t+1};\pi) = \frac{\bar{a}K_{t+1}}{\zeta_t} - \sum_{k=1}^{K_{t+1}} \ln\frac{R_{k,t+1}}{u}.$$

With these expressions in place, I present my central econometric result.

Proposition 7. *Let the true data generating process of $\{R_t\}_{t=1}^T$ satisfy the dynamic power law model of Assumption 1 with parameter vector π^*. Define the quasi-likelihood estimator $\hat{\pi}_{QL}$ as*

$$\hat{\pi}_{QL} = \arg\max_{\pi\in\Pi} \mathcal{L}(\{R_t\}_{t=1}^T;\pi).$$

If the following conditions are satisfied

i. π^* is interior to the parameter space Π over which maximization occurs;

ii. for $\pi \neq \pi^*$, $E[s_t(R_{t+1};\pi)] \neq 0$, and

iii. $E[\sup_{\pi\in\Pi} \|s_t(R_{t+1};\pi)\|] < \infty$,

Then $\hat{\pi}_{QL} \xrightarrow{p} \pi^*$.

Furthermore, if

iv. $E[\sup_{\pi\in\Pi} \|\nabla_\pi s_t(R_{t+1};\pi)\|] < \infty$,

v. $\frac{1}{\sqrt{T}}\sum_{t=0}^{T-1} s_t(R_{t+1};\pi^*) \xrightarrow{d} N(0,G)$ and

vi. $E[\nabla_\pi s_t(R_{t+1};\pi^*)]$ is full column rank,

Then $\sqrt{T}(\hat{\pi}_{QL} - \pi^*) \xrightarrow{d} N(0,\Psi)$, where $\Psi = S^{-1}GS^{-1}$, $S = E[\nabla_\pi s_t(R_{t+1};\pi^*)]$, and $G = E[s_t(R_{t+1};\pi^*)s_t(R_{t+1};\pi^*)']$.

Proof. The proof follows Newey and McFadden (1994). Before proceeding, I establish a key lemma upon which the remainder of the proposition relies. It shows that $s_t(R_{t+1};\pi)$ (which is based on the mis-specified model $\tilde{F}_{u,t}$) has expectation equal to zero given that the true data generating process satisfies Assumption 1.

Lemma 1. *Under Assumption 1,* $E[s_t(R_{t+1};\pi)] = 0$.

By the law of iterated expectations,

$$\begin{aligned} E[s_t(R_{t+1};\pi)] &= E\big[E_t[s_t(R_{t+1};\pi)]\big] \\ &= E\left[E_t\left[\frac{K_{t+1}}{\bar{a}\zeta_t} - \sum_{k=1}^{K_{t+1}} \ln\frac{R_{k,t+1}}{u}\right]\right] \\ &= E\left[\frac{K_{t+1}}{\bar{a}\zeta_t} - \frac{K_{t+1}}{\bar{a}\zeta_t}\right] \\ &= 0. \end{aligned}$$

The third equality follows from the fact that the identities of the K_{t+1} exceedences are unknown at time t, thus the expectation reduces to the cross-sectional average of expected exceedences, proving the lemma.[13]

Observe that the first order condition for maximization of (8) is $\frac{1}{T}\sum_{t=0}^{T-1} s_t(R_{t+1};\pi) = 0$. That is, maximization of the log quasi-likelihood produces a valid moment condition upon which estimation may be based. With this insight in hand, the approach of Newey and McFadden may be employed to establish asymptotic properties of the dynamic power law quasi-likelihood estimator. By condition (*i*) and the fact that the true generating process satisfies Assumption 1, Lemma 1 shows that the moment condition arising from maximization of $\mathcal{L}(\{R_t\}_{t=1}^T;\pi)$ is satisfied. Adding condition (*ii*), π^* is uniquely identified. By the dominance condition (*iii*), the uniform law of large numbers may be invoked to establish convergence in probability.

To establish convergence to normality, I use a mean value expansion of the moment condition sample analogue around $\bar{\pi}$ (a value between $\hat{\pi}$ and π^*), which gives

$$\frac{1}{\sqrt{T}}(\hat{\pi} - \pi) = -\Big[\frac{1}{T}\sum_t \nabla_\pi s_t(R_{t+1};\bar{\pi})\Big]^{-1}\frac{1}{\sqrt{T}} s_t(R_{t+1};\pi^*).$$

This expansion is performed noting that π^* is interior to the parameter space and, by the functional form of the quasi-likelihood, $s_t(R_{t+1};\bar{\pi})$ is continuously differentiable over Π. Be-

[13]While the identities of the exceedences are unknown, the number of exceedences *is* known since the tail is defined by a fixed fraction of the cross section size.

cause $\bar{\pi}$ is between $\hat{\pi}$ and π^*, $\bar{\pi}$ is also consistent for π^* by the convergence in probability result just shown. Using this fact together with condition (iv) delivers $\frac{1}{T}\sum_t \nabla_\pi s_t(R_{t+1};\bar{\pi}) \xrightarrow{p} S$. By condition (v), $\frac{1}{\sqrt{T}}\sum_{t=0}^{T-1} s_t(R_{t+1};\pi^*) \xrightarrow{d} N(0,G)$. This, together with Slutzky's theorem and assumption (vi), proves the result. □

3.3 Volatility and Heterogeneous Exceedence Probabilities

Implicit in the formulation above is that each element of the vector R_t has an equal probability of exceeding threshold u. However, heterogeneity in individual stock volatilities affects the likelihood that a particular stock will experience an exceedence. Let X be a power law variable such that $P(X > u) = bu^{-\zeta}$. The u-exceedence distribution of X is $P(X > x|X > u) = \left(\frac{x}{u}\right)^{-\zeta}$. Now consider a volatility rescaled version of this variable, $Y = \sigma X$. The exceedence probability of Y equals $b\left(\frac{u}{\sigma}\right)^{-\zeta}$, different than that of X. When $\sigma > 1$, Y has a higher exceedence probability than X. However, the shape of Y's u-exceedence distribution is identical to that of X.

A reformulation of the estimator to allow for heterogeneous volatilities is easily established. Let each stock have unique u-exceedence probability p_i, and consider the effect of this heterogeneity on the expectation of the tail exponent update. In this case, the expectation is no longer the harmonic average tail exponent, but is instead the *exceedence probability-weighted* average exponent,

$$E_t\left[\frac{1}{K_{t+1}}\sum_{k=1}^{K_{t+1}} \ln\frac{R_{k,t+1}}{u}\right] = \frac{1}{\zeta_t}\sum_{i=1}^{n}\frac{\omega_i}{a_i},$$

where $\omega_i = p_i/\sum_j p_j$. The entire estimation approach and consistency argument outlined above proceeds identically after establishing this point. The ultimate result is that the fitted ζ_t series is no longer an estimate of the equal-weighted average exponent, but takes on a volatility-weighted character due to the effect that volatility has on the probability of tail occurrences.

Another potential concern is contamination of tail estimates due to time-variation in volatility. I address this by allowing the threshold u to vary over time. My procedure selects u as a fixed $q\%$ quantile,

$$\hat{u}_t(q) = \inf\left\{R_{(i),t} \in R_t : \frac{q}{100} \leq \frac{(i)}{n}\right\}$$

where (i) denotes the i^{th} order statistic of $(n \times 1)$ vector R_t. In this case, u expands and contracts with volatility so that a fixed fraction of the most extreme observations are used for estimation each period, nullifying the effect of volatility dynamics on tail estimates. My estimates are based on $q = 5$ (or 95 for the upper tail).[14]

3.4 Monte Carlo Evidence

Appendix B describes a series of Monte Carlo experiments designed to assess finite sample properties of the dynamic power law estimator. Table 11 shows results confirming that the asymptotic properties derived above serve as accurate approximations in finite samples. They also demonstrate the estimator's robustness to dependence among tail observations and volatility heterogeneity across stocks, both of which are suggested by the structural models. Table 12 explores the estimator's performance when the true tail exponent is conditionally stochastic. Even though the estimator presented here relies on a conditionally deterministic exponent process, its estimates achieve over 80% correlation with the true tail series on average.

4 Empirical Results

4.1 Tail Risk Estimates

Estimates for the dynamic power law model use daily CRSP data from August 1962 to December 2008 for NYSE/AMEX/NASDAQ stocks with share codes 10 and 11. Accuracy of extreme value estimators typically requires very large data sets because only a small fraction of data is informative about the tail distribution. Since the dynamic power law estimator relies on the cross section of returns, I require a large panel of stocks in order to gather sufficient information about the tail at each point in time. Figure 1 plots the number

[14]Threshold choice can have important effects on results. An inappropriately low threshold will contaminate tail exponent estimates by using data from the center of the distribution, whose behavior can vary markedly from tail data. A very high threshold can result in noisy estimates resulting from too few data points. While sophisticated methods for threshold selection have been developed (Dupuis 1999; Matthys and Beirlant 2000; among others), these often require estimation of additional parameters. In light of this, Gabaix et al. (2006) advocate a simple rule fixing the u-exceedence probability at 5% for unconditional power law estimation. I follow these authors by applying a similar simple rule in the dynamic setting. Unreported simulations suggest that $q = 1$ to 5 (or 95 to 99 for the upper tail) is an effective quantile choice in my dynamic setting.

of stocks in CRSP each month. The sample begins with just under 500 stocks in 1926, and has fewer than 1,000 stocks for the next 25 years. In July 1962, the sample size roughly doubles to almost 2,000 stocks with the addition of AMEX. In December 1972, NASDAQ stocks enter the sample and the stock count leaps above 5,000, fluctuating around this size through 2008.[15] The dramatic cross-sectional expansion of CRSP beginning in August 1962 leads to my focus on the 1963 to 2008 sample.

Other data used in my analysis are daily Fama-French return factors, monthly risk free rates and size/value-sorted portfolio returns from Ken French's Data Library[16], market return predictor variables from Ivo Welch's website[17], variance risk premium estimates from Hao Zhou's website[18], and macroeconomic data from the Federal Reserve.

I focus my empirical analysis on the tails of raw returns. In each test I consider tail risk estimated using data from the lower tail only, from the upper tail only, and from combining lower and upper tail data. I refer to the latter case as "both" tails, which in the tests that follow should *not* be understood as simultaneously including separate estimates of the lower and upper tail in regressions.

For robustness, I also explore how results change when residuals from the Fama-French three-factor model are used to estimate tail risk. Factor model residuals offer a means of mitigating the effects of dependence on the estimator's efficiency.[19] Threshold u_t is chosen to be the 5% cross-sectional quantile each period. Standard errors are estimated based on the form of the asymptotic covariance matrix Ψ derived in Proposition 7. In particular, $\hat{S}$ is the log quasi-likelihood Hessian evaluated at $\hat{\pi}$ parameters and $\hat{G}$ is the sample average log quasi-likelihood gradient outer product evaluated at $\hat{\pi}$. Further details on standard error estimation via likelihood Hessians and gradient outer products may be found in Hayashi (2000).

Table 1 reports estimates for the dynamic power law evolution (7). The lower tail exponent of raw returns ζ_t varies around a mean of 2.20, with $\hat{\pi}_1 = 0.072$ and $\hat{\pi}_2 = 0.923$. The p-values

[15] The dynamic power law estimator in Section 3 accommodates changes in size of the cross section over time, highlighting another attractive feature of the estimator.

[16] URL: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[17] URL: http://welch.econ.brown.edu/.

[18] URL: http://sites.google.com/site/haozhouspersonalhomepage/.

[19] As my asymptotic theory results and Monte Carlo evidence show, abstracting from dependence does not affect the estimator's consistency. It may, however, affect the variance of estimates. The asymptotic covariance derived in Proposition 7 accounts for this decreased efficiency, hence test statistics maintain appropriate size.

for all statistics reported in Table 2 are below 0.001, rejecting the null hypothesis of constant tail risk and supporting Testable Implication 1. The fact that $\hat{\pi}_1 + \hat{\pi}_2 > 0.99$ implies that the tail exponent is highly persistent. The upper tail is slightly fatter (consistent with the results of Jansen and de Vries 1991) and less persistent, though with more time series variability (given the higher value of $\hat{\pi}_1$). When stock returns are converted to factor model residuals, estimation results are qualitatively unchanged.

I plot the fitted tail series ($-\zeta_t$) based on raw returns in Figure 2. The estimates from both tails together are shown in Figure 2a, and separate estimates for the lower and upper tails are shown in Figure 2b. These are plotted alongside the log price-dividend ratio for the aggregate market. The estimated tail risk series appears moderately countercyclical, sharing a monthly correlation of -14%, -15% and -14% with the log price-dividend ratio based on estimates from both tails, the lower tail and the upper tail, respectively. The beginning of the sample sees high tail risk, immediately following a drop in the value-weighted index of 28% in the first half of 1962 (the first major market decline during the post-war period). Tail risk declines steadily until December of 1968, when it reaches its lowest levels in the sample. This corresponds to a late 1960's bull market peak, the level of which is not reached again until the mid-1970's. Tail risk rises throughout the 1970's, accelerating its ascent during the oil crisis. It remains around its mean value of -2 to -2.5 for most of the remaining sample. Tail risk recedes in the four bull market years leading up to the 1987 crash, rising quickly in the following months. During the technology boom tail risk retreats sharply but briefly, spiking to its highest post-2000 level amid the early 2003 market trough. At this time the value-weighted index was down 49% from its 2000 high and NASDAQ was 78% of its peak. Finally, during the late 2008 financial crisis, tail risk sees a modest climb after falling in the first half of 2008. Figure 3 plots the tail series estimated from Fama-French three-factor model residuals. The broad dynamic patterns are the same as those in Figure 2.

Figure 4 shows the threshold series u_t for raw returns alongside monthly realized volatility of the CRSP value-weighted index. The threshold based on both tails has a 60% correlation with volatility. As the figure and correlation show, the threshold appears to successfully absorb volatility changes.

In Table 2, I report monthly correlations between tail risk estimates and macroeconomic variables. Lower (upper) tail estimates have correlation with unemployment of 53% (39%) and -10% (-7%) with the Chicago Fed National Activity Index (CFNAI), once again suggesting some cyclicality in tail risk.

As a brief exploration into the determinants of tail risk, I estimate a regression of the tail process on its own lag and lags of a collection of macroeconomic variables, including realized equity volatility, unemployment, inflation, growth in industrial production, CFNAI and the aggregate stock market return. I show the estimated impacts of these variables on future tail risk in Table 3. Coefficients have been scaled to be interpreted as the response of tail risk ($-\zeta_t$, in number of standard deviations) to a one standard deviation increase in the dependent variable. The most important variable for determining tail risk is the past market return. A return one standard deviation above its mean predicts that the lower tail becomes thinner by 0.175 standard deviations (Newey-West t=7.1), while a high past return increases the weight of the upper tail by 0.227 standard deviations (t=4.6). Other potentially important determinants of tail risk are past unemployment and equity volatility.

4.2 Predicting Stock Market Returns

Testable Implication 2 predicts that the estimated tail risk process ($-\zeta_t$) forecasts risk premia over short and long horizons. I investigate this hypothesis with a series of predictive regressions.

The dependent variable is the excess return on the CRSP value-weighted index at frequencies of one month, one year, three years and five years. The regressor of interest is the tail exponent.[20] Since my estimator is available daily, forecasts are based on $-\zeta_t$ on the last day prior to the forecast horizon.

For comparison, I consider two additional predictor variables motivated by my theory model. (Then, in robustness checks, I consider a much larger set of alternative predictors.) The first is monthly realized volatility of the aggregate stock market. Along with the tail risk process, my modified long run risks model suggests that conditional volatility drives expected excess

[20]To avoid look ahead bias, I do not use the tail series estimated in Section 4.1 for forecasting since its construction relies on parameters estimated from the full sample. Instead, I use an exponentially-weighted moving average of the period-by-period update ζ_t^{upd}. The weighting parameter is fixed *ex ante* at 0.94, the value used by RiskMetrics for its "exponential smoother". As pointed out in Section 3, the dynamic power law estimator is simply an optimal exponentially-weighted moving average of updates. Thus, the RiskMetrics exponential smoother applied to the daily cross-sectional Hill estimate is a special case of my model. The RiskMetrics smoothing parameter is analogous to $\pi_2 = 0.94$ in the tail exponent evolution (7). For comparison, my maximum likelihood estimates of π_2 are 0.923, 0.683 and 0.798 for the lower tail, upper tail and both tails, respectively. My predictive results are quite insensitive to the choice of the weighting parameter. Using values ranging from 0.70 to 0.98, including my estimated values, produces only small differences in results, in many cases improving upon the results reported here.

returns. Second, I consider the log dividend-price ratio. The Campbell-Shiller identity implies that variables driving risk premia necessarily enter into the log dividend-price ratio, hence it too should forecast future excess returns as long as one of the state variables does.[21]

All regressions are conducted at the monthly frequency, meaning that observations are overlapping for the one, three and five year analyses. Richardson and Smith (1991), Hodrick (1992) and Boudoukh and Richardson (1993) (among others) have noted the inferential problems concomitant with overlapping horizon predictive regressions. Overlapping return observations induce a moving average structure in prediction errors, distorting the size of tests based on OLS, and even Newey-West, standard errors. Ang and Bekaert (2007) demonstrate in a Monte Carlo study that the standard error correction of Hodrick (1992) provides the most conservative test statistics relative to other commonly employed standard error estimators, and maintains appropriate test size over horizons as long as five years. I therefore calculate all statistics with Hodrick standard errors.[22]

Predictive regressions based on tail estimates from raw returns are reported in Table 4. To illustrate economic magnitudes, coefficients are scaled to be interpreted as the effect of a one standard deviation increase in the regressor on future annualized returns. The lower tail risk series has large, significant forecasting power over all horizons. A one standard deviation increase in lower tail risk predicts an increase in future excess returns of 6.7%, 4.4%, 4.5% and 5.0% per annum, based on data for one month, one year, three year and five year horizons, respectively. The corresponding t-statistics are 2.9, 2.1, 2.2 and 2.3. The effect of lower tail risk retains the same economic and statistical magnitudes when included alongside realized volatility and the log dividend-price ratio. Similarly, a one standard deviation increase in risk measured from both tails forecasts an increase in excess market returns of between 3.7% and 5.6%. Upper tail risk carries the same sign premium as lower tail risk, which is consistent with the long run risks model of Section 2. The forecasting power of the upper tail is weaker

[21]While true in theory, it may be the case in practice that some state variables have forecasting power while valuation ratios do not. A failure to find return predictability by valuation ratios despite predictability by state variables may arise from poor measurement of aggregate fundamentals such as dividends or other payouts, while purely price-based measures of state variables like tail risk may be more accurate.

[22]A second statistical consideration is high persistence in the daily tail exponent series. While the daily series appears nearly integrated, regressions are run at the monthly frequency, and ζ-based forecasts only use its value on the last day of each month. Persistence in the month-end series is substantially weaker, with an AR(1) coefficient of 0.890. While this value is mild compared to a monthly AR(1) coefficient of 0.996 for the log dividend-price ratio, it nonetheless calls for cautious inference. I find that Stambaugh bias has little effect on my results. The lower tail risk regression using a one month horizon is a representative case. Stambaugh bias accounts for no more than 3% of the predictive coefficient magnitude in that regression, thus the estimate remains statistically significant at the 1% level.

than the lower tail, though its predictive power is impressive at long horizons.

I also evaluate the out-of-sample predictive ability of tail risk. Using data only through month t (beginning with $t = 60$ to provide a sufficiently large initial estimation period), I run univariate predictive regressions of market returns on lower tail risk. This coefficient is used to forecast the $t + 1$ return. Because the coefficient is based only on data through t, this procedure mimics the information set an investor would work with in real-time. I plot the resulting sequence of estimated predictive coefficients in Figure 5. Estimates are remarkably stable and are significant at the 95% level in most months, despite having few observations in the early part of the analysis. Moreover, the out-of-sample R^2 is 1.30%, only slightly lower than the in-sample value.

In Table 5, I explore robustness of these results to estimating tail risk from Fama-French three-factor model residuals rather than raw returns. This mitigates dependence among tail observations, which can potentially improve the finite sample properties of the estimator. The analysis proceeds as in Table 4, and conclusions are unchanged. Lower tail risk demonstrates the strongest ability to forecast market returns, and upper tail risk predicts returns in the same direction as lower tail risk. Residual tails provide even more explanatory power than raw returns over one, three and five year horizons, and upper tail risk becomes a statistically successful predictor at the one month horizon.

In a second robustness check, I compare the forecasting power of my tail risk measure to a large set of alternative forecasting variables that have been shown to successfully predict returns. Goyal and Welch (2008) conduct a comprehensive analysis of forecasting performance using an extensive set of predictors from the literature. Table 6 reports results of univariate forecasting regressions using fifteen Goyal and Welch variables that are available at the monthly frequency over the 1963-2007 sample. No variable forecasts returns as strongly or consistently over all horizons as the lower tail risk process. Inflation strongly predicts one month returns, but its effect dies out at longer horizons. After tail risk, the most successful long horizon univariate predictors are the cross-sectional premium (Polk, Thompson and Vuolteenaho 2006, data available through 2002), the term spread, and the long term yield (Campbell 1987; Fama and French 1989). I also run bivariate regressions using lower tail risk alongside each Goyal and Welch variable from Table 6. Table 7 presents these results. The predictive ability of tail risk is broadly unaffected by including alternative regressors.[23]

[23] In one of the fifteen cases, coefficient estimates suggest a potential multicollinearity problem: This is when tail risk and the cross-sectional premium are included together. The monthly correlation between

In addition to the Goyal and Welch predictors, I compare the performance of my tail risk measure to the variance risk premium (VRP). Bollerslev, Tauchen and Zhou (2009) demonstrate the outstanding predictive power of VRP, which is the difference between risk neutral variance extracted from options data and realized variance calculated from ultra-high frequency data (see the last row of Table 6 for univariate VRP regressions). These authors, as well as Drechsler and Yaron (2009), demonstrate how VRP is fundamentally linked to tail risk in the real economy and thus naturally related to my measure. The shortcoming of VRP, however, is that it is only available starting in 1990 due to intra-daily data limitations. I compare the predictive power of my tail risk estimate and VRP in bivariate predictive regressions over the 1990-2008 subsample. Results are reported in the last row of Table 7. The monthly R^2 increases from 3.5% and 2.1% when VRP and tail risk, respectively, are included alone, to 5.2% when they are included together. At the annual horizon, the univariate VRP and tail risk R^2 increases from 3.8% and 18.5%, respectively, to 22.8% when they are included together. The coefficient on $-\zeta_t$ is slightly larger than its estimate from the full sample, and is stable over all horizons. These results suggest that my tail risk measure and VRP are complementary descriptors of tail risk. In summary, I conclude that increases in my tail risk measure significantly predict increases in excess returns on the market over short and long horizons, consistent with Testable Implication 2.

4.3 Tail Risk and the Cross Section of Expected Returns

In Section 2, I show that the heavy-tailed long run risks model and time-varying rare disaster model predict a negative price of tail risk (Testable Implication 3). As a first test of this hypothesis, I examine average returns of portfolios formed on the basis of stocks' exposure to my estimated tail risk process. I proceed as follows. In each month $t + 1$, I estimate tail risk betas by regressing monthly excess returns over the 60 months ending at t on contemporaneous innovations in the month-end tail risk ($-\zeta_t$, estimated from both tails). I then sort stocks into tail risk beta quintile portfolios, including all NYSE/AMEX/NASDAQ stocks with share codes 10 and 11 and at least 36 months of non-missing returns out of the previous 60 months. Table 8, Panel A reports time series average post-formation portfolio

these two variables is 64%. Multicollinearity presents an inferential problem for individual regressors, but *not* for the total explanatory power of the model. The monthly predictive R^2 for the two variables together is 7.0%, which is an extremely high degree of predictability (see Campbell and Thompson 2008). At the five year horizon, the cross section premium effect dies out, but the tail risk effect remains large and significant (scaled coefficient of 4.2% with t-statistic 2.0).

returns for tail risk beta-sorted portfolios. Stocks that covary least with tail risk (quintile 1) earn average annualized returns of 6.40%, while stocks with the highest tail risk betas (quintile 5) earn only 0.36% per annum. The difference in average annualized returns between quintile 5 and quintile 1 is -6.03% with t-statistic 2.6, supporting the hypothesis of a negative tail risk premium.

To check whether this result is driven by a particular subset of stocks, I construct double-sorted portfolios. In Panel B of Table 8, I first sort stocks into quintiles by market beta (estimated in the same rolling manner as tail risk beta). Then, within market beta quintiles, I sort again by tail risk beta. The negative tail risk premium appears in each quintile, with a difference in average returns across high and low tail risk beta quintiles ranging from -3.06% to -4.43% per annum. Similarly, I find a negative tail risk premium across size quintiles (Panel C) and book-to-market quintiles (Panel D), with the exception of the very smallest stocks and very highest book-to-market stocks.

To test the negative tail risk price hypothesis while controlling for multiple alternative explanatory characteristics, I run a series of Fama-MacBeth regressions. Table 9 shows results from second stage regressions using NYSE/AMEX/NASDAQ stocks as test assets. I scale coefficients so that they may be interpreted as the change in annual expected excess returns for a one (cross-sectional) standard deviation change in risk factor exposure. When regressions include only tail risk, the risk price estimate is -6.2% (t=2.9) based on the lower tail and -5.3% (t=2.5) using both tails. After controlling for realized equity market volatility (as prescribed by the long run risks model) and Fama-French factors, lower tail and both tail risk price estimates are -5.6% and -4.4%, respectively (t=2.9 and 2.4).

Repeating the Fama-MacBeth analysis with tail risk estimated from Fama-French factor model residuals (Panel B of Table 9) leads to the same conclusions. When all four additional explanatory variables are included, lower tail and both tail risk prices are estimated at -3.9% and -3.0% (t=2.2 and 1.9). Finally, I repeat the analysis using three alternative sets of test assets. First, to determine if results are being driven by smaller NASDAQ stocks, I consider NYSE stocks separately. Results are presented in Table 10. I find similar tail risk price magnitudes as in the full CRSP cross section (-5.1% and -4.5% for the lower tail and both tails) with stronger significance (t=3.5 and 3.3). Table 10 also includes test results using 100 size and value-sorted portfolios as test assets (from Ken French's website) and 25 market beta and tail risk beta-sorted portfolios. In both cases, the estimated price of lower tail risk is significantly negative. More generally, the results of this section support the prediction of

Testable Implication 3 that the price of tail risk is negative, hence high tail risk betas stocks earn comparatively low average returns.

5 Conclusion

A measure of extreme event risk is essential to understanding the behavior of asset prices. If this risk changes through time, extreme value techniques based on aggregate data will be incapable of providing conditional tail measures. I present a new dynamic tail risk model that overcomes this difficulty. The model uses the cross section of individual stock returns to inform estimates of conditional tail risk at each point in time. Tests show that extremal risk is significantly time varying and persistent. This conclusion holds for raw returns as well as factor model residuals.

Evidence suggests that tail risk has large predictive power for excess aggregate stock market returns over horizons of one month to five years, outperforming all alternative predictors commonly consider in the literature. A one standard deviation increase in tail risk forecasts a 4.4% increase in excess market returns over the following year. Furthermore, individual stock tails have large, significant explanatory power for the cross section of average stock returns. Stocks that covary highly with my estimated tail risk series earn average annual returns 6.0% lower than stocks with low tail risk covariation.

These results can be understood from the perspective of (at least) two structural models: i) a long run risks model modified to include time-varying heavy-tailed shocks and ii) a time-varying rare disaster model. I show that in both of these models the price of tail risk is negative. Because tail risk is detrimental to investors' utility, assets that pay off in high tail risk states are valuable hedges and thus earn low average returns, consistent with empirical results in the cross section. In the time series, high tail risk increases the return required by investors to hold the market portfolio, which results in return forecastability, also consistent with my empirical findings.

A Model Derivations

Proof of Proposition 1

Proof. The proof proceeds by the method of undetermined coefficients. The equilibrium solution is obtained by demonstrating that the Euler equation $E_t[\exp(m_{t+1} + r_{c,t+1})] = 1$ is always satisfied for the consumption claim. First, substitute the Campbell-Shiller return approximation,

$$r_{c,t+1} = \kappa_0 + \kappa_1 wc_{t+1} - wc_t + \Delta c_{t+1},$$

into the pricing kernel in the Euler equation to obtain

$$E_t\left[\exp\left(\theta \ln \beta + \theta(1 - \frac{1}{\psi})\Delta c_{t+1} + \theta[\kappa_0 + \kappa_1 wc_{t+1} - wc_t]\right)\right] = 1.$$

Next, substitute for the log wealth-consumption ratio using the hypothesized form in (3). Evaluating the expectation involves computing the cumulant generating function of Laplace variables. The following property of the Laplace distribution is informative for this end.

Lemma 2. *The cumulant generating function of a unit Laplace variable W_{t+1} evaluated at s is*

$$\ln E_t\big[\exp(sW_{t+1})\big] = \ln\left(\frac{1}{1-s^2}\right).$$

The cumulant generating function of the cash flow growth shock is therefore

$$\ln E_t\left[\exp\left(s\sqrt{\Lambda_t}W_{t+1}\right)\right] = \ln\left(\frac{1}{1-s^2\Lambda_t}\right). \tag{9}$$

To obtain log prices that are linear in Λ_t, I use a first order Taylor expansion of (9) around zero: $\ln E_t\left[\exp\left(s\sqrt{\Lambda_t}W_{t+1}\right)\right] \approx s^2\Lambda_t$.

Based on this, the equilibrium restriction used to determine state variable coefficients A_x, A_σ

and A_Λ is

$$\begin{aligned}1 &= \exp\Bigg\{\theta\ln\beta + (1-\gamma)\,\mu + \theta\Big(\kappa_0 + A_0[\kappa_1 - 1] + \kappa_1[A_\sigma\bar{\sigma}^2(1-\rho_\sigma) + A_\Lambda\bar{\Lambda}(1-\rho_\Lambda)]\Big)\\ &\quad + \frac{(\theta\kappa_1)^2}{2}\left[A_\sigma^2\sigma_\sigma^2 + A_\Lambda^2\sigma_\Lambda^2\right] + x_t\Big(\theta A_x[\kappa_1\rho_x - 1] + 1 - \gamma\Big)\\ &\quad + \sigma_t^2\Big(\theta A_\sigma[\kappa_1\rho_\sigma - 1] + \frac{(1-\gamma)^2\sigma_c^2}{2} + \frac{(\theta\kappa_1)^2}{2}A_x^2\sigma_x^2\Big)\\ &\quad + \Lambda_t\Big(\theta A_\Lambda[\kappa_1\rho_\Lambda - 1] + (1-\gamma)^2\Big)\Bigg\}.\end{aligned}$$

In equilibrium, this restriction must hold for any realization of state variables. This is satisfied when coefficients on state variables and the constant term are exactly zero, yielding implicit solutions for A_x, A_σ and A_Λ.

The stated wealth-consumption ratio may be used within the Campbell-Shiller approximation to deduce returns on the consumption asset and the stochastic discount factor. This, in turn, is used to derive the log price-dividend ratio for each asset in the economy, which is also linear in tail risk. The proof of the log price-dividend ratio function proceeds in the same manner. The following equilibrium restriction delivers the result and determines state variable coefficients:

$$\begin{aligned}1 &= \exp\Bigg\{\theta\ln\beta + (\phi_i-\gamma)\,\mu + \mu_i + (\theta-1)\Big(\kappa_0 + A_0[\kappa_1 - 1] + \kappa_1[A_\sigma\bar{\sigma}^2(1-\rho_\sigma) + A_\Lambda\bar{\Lambda}(1-\rho_\Lambda)]\Big)\\ &\quad + \kappa_{i,1}[A_{i,\sigma}\bar{\sigma}^2(1-\rho_\sigma) + \bar{\Lambda}A_{i,\Lambda}(1-\rho_\Lambda)] + \kappa_{i,0} + A_{i,0}[\kappa_{i,1} - 1]\\ &\quad + \frac{\sigma_\sigma^2}{2}\big([\theta-1]\kappa_1 A_\sigma + \kappa_{i,1}A_{i,\sigma}\big)^2 + \frac{\sigma_\Lambda^2}{2}\big([\theta-1]\kappa_1 A_\Lambda + \kappa_{i,1}A_{i,\Lambda}\big)^2\\ &\quad + x_t\Big[(\theta-1)A_x(\kappa_1\rho_x - 1) + A_{i,x}(\kappa_{i,1}\rho_x - 1) + \phi_i - \gamma\Big]\\ &\quad + \sigma_t^2\Big[(\theta-1)A_\sigma(\kappa_1\rho_\sigma - 1) + A_{i,\sigma}(\kappa_{i,1}\rho_\sigma - 1) + \frac{\sigma_x^2}{2}\big([\theta-1]\kappa_1 A_x + \kappa_{i,1}A_{i,x}\big)^2 + \frac{\sigma_c^2}{2}(\phi_i-\gamma)^2 + \frac{\sigma_i^2}{2}\Big]\\ &\quad + \Lambda_t\Big[(\theta-1)A_\Lambda(\kappa_1\rho_\Lambda - 1) + A_{i,\Lambda}(\kappa_{i,1}\rho_\Lambda - 1) + (\phi_i-\gamma)^2 + q_i^2\Big]\Bigg\}.\end{aligned}$$

Of particular interest to this paper are the coefficients on tail risk. For the wealth-consumption ratio, this is[24]

$$A_\Lambda = \frac{(1-\gamma)(1-\frac{1}{\psi})}{1-\kappa_1\rho_\Lambda} < 0$$

while for the price-dividend ratio it is

$$A_{i,\Lambda} = \frac{1}{1-\kappa_{i,1}\rho_\Lambda}\left[A_\Lambda(\theta-1)(\kappa_1\rho_\Lambda-1) + (\phi_i-\gamma)^2 + q_i^2\right].$$

For future reference, I also present the following expressions:

$$A_x = \frac{1-\frac{1}{\psi}}{1-\kappa_1\rho_x} \text{ and } A_{i,x} = \frac{\phi_i-\frac{1}{\psi}}{1-\kappa_{i,1}\rho_x}.$$

□

Proof of Proposition 2

Proof. Based on the wealth-consumption ratio derived in Proposition 1 and the Campbell-Shiller identity, shocks to the discount factor are

$$\begin{aligned} m_{t+1} - E_t[m_{t+1}] &= -\gamma\sigma_c\sigma_t z_{c,t+1} - \gamma\sqrt{\Lambda_t}W_{c,t+1} - (1-\theta)\kappa_1 A_x\sigma_x\sigma_t z_{x,t+1} \\ &\quad -(1-\theta)\kappa_1 A_\sigma\sigma_\sigma z_{\sigma,t+1} - (1-\theta)\kappa_1 A_\Lambda\sigma_\Lambda z_{\Lambda,t+1}. \end{aligned}$$

The prices of risk in the economy are the coefficients on the shocks. I use λ_j to denote the price of risk for shock j, so that $\lambda_c = \gamma$, $\lambda_x = (1-\theta)\kappa_1 A_x$, $\lambda_\sigma = (1-\theta)\kappa_1 A_\sigma$ and $\lambda_\Lambda = (1-\theta)\kappa_1 A_\Lambda$ are the prices for exposure to transitory consumption growth, long run consumption growth, variance and tail risks, respectively.

Shocks to log returns have a similar representation,

$$\begin{aligned} r_{i,t+1} - E_t[r_{i,t+1}] &= \beta_{i,c}(\sigma_c\sigma_t z_{c,t+1} + \sqrt{\Lambda_t}W_{c,t+1}) + \beta_{i,x}\sigma_x\sigma_t z_{x,t+1} \\ &\quad +\beta_{i,\sigma}\sigma_\sigma z_{\sigma,t+1} + \beta_{i,\Lambda}\sigma_\Lambda z_{\Lambda,t+1} + \sigma_i\sigma_t z_{i,t+1} + q_i\sqrt{\Lambda_t}W_{i,t+1}. \end{aligned}$$

where $\beta_{i,c} = \phi_i$, $\beta_{i,x} = \kappa_{i,1}A_{i,x}$, $\beta_{i,\sigma} = \kappa_{i,1}A_{i,\sigma}$ and $\beta_{i,\Lambda} = \kappa_{i,1}A_{i,\Lambda}$.

[24] While these expressions are written in an implicit form, the characterization is sufficient for signing A_Λ since the Campbell-Shiller linearization ensures κ_1, which is a function of A coefficients, falls in the interval $(0,1)$.

Evaluating the Euler condition shows that

$$\begin{aligned} 0 &= E_t[m_{t+1}] + E_t[r_{i,t+1}] + \ln\left\{E_t\Big[\exp\left(m_{t+1} - E_t[m_{t+1}] + r_{i,t+1} - E_t[r_{i,t+1}]\right)\Big]\right\} \\ &= E_t[r_{i,t+1}] - r_{f,t} + \frac{1}{2}Var(r_{i,t+1}) - \beta_{i,c}\lambda_c(\sigma_c^2\sigma_t^2 + 2\Lambda_t) - \beta_{i,x}\lambda_x\sigma_x^2\sigma_t^2 - \beta_{i,\sigma}\lambda_\sigma\sigma_\sigma^2 - \beta_{i,\Lambda}\lambda_\Lambda\sigma_\Lambda^2. \end{aligned}$$

□

Proof of Proposition 3

Proof. From the proof of the previous proposition, return shocks are

$$\begin{aligned} r_{i,t+1} - E_t[r_{i,t+1}] &= \beta_{i,c}(\sigma_c\sigma_t z_{c,t+1} + \sqrt{\Lambda_t}W_{c,t+1}) + \beta_{i,x}\sigma_x\sigma_t z_{x,t+1} \\ &\quad + \beta_{i,\sigma}\sigma_\sigma z_{\sigma,t+1} + \beta_{i,\Lambda}\sigma_\Lambda z_{\Lambda,t+1} + \sigma_i\sigma_t z_{i,t+1} + q_i\sqrt{\Lambda_t}W_{i,t+1}. \end{aligned}$$

Using these shocks I derive the tail distribution of returns.

Let $S = \phi_i\sqrt{\Lambda_t}W_{c,t+1}$ and let $Y = \exp(S)$. The density of S is

$$g_S(s) = \frac{1}{2\phi_i\sqrt{\Lambda_t}}\exp\left(\frac{-|s|}{\phi_i\sqrt{\Lambda_t}}\right).$$

The derivative of S with respect to Y is $1/Y$. The conservation of probability law, $g_S(s)ds = g_Y(y)dy$, therefore implies that the density of Y is

$$G_Y(y) = \begin{cases} y^{1/(\phi_i\sqrt{\Lambda_t})} & \text{if } y < 1 \\ y^{-1/(\phi_i\sqrt{\Lambda_t})} & \text{if } y > 1, \end{cases}$$

showing that Y obeys a power law in both tails. Repeating this argument for $\exp(q_i\sqrt{\Lambda_t}W_{i,t+1})$ shows that it also obeys a power law in both tails with exponents $\pm 1/(q_i\sqrt{\Lambda_t})$. The asymptotic aggregation properties of power law distributions (i.e., in the limit as u approaches the end of the support, in this case 0 for the lower tail and ∞ for the upper tail) dictate that if two variables have power law tails with exponents ξ_1 and ξ_2, their product also behaves as a power law in its tail with exponent $\min(\xi_1, \xi_2)$; that is, the heavier-tailed power law dominates.[25] Since the remaining shocks are Gaussian (or lognormal when exponentiated),

[25] These properties are summarized in the appendix of Gabaix et. al (2006). It is this asymptotic equivalence

they have no asymptotic effect on the tail distribution and the result follows. $\square$

Proof of Proposition 4

Proof. The proof follows the same logic as in Proposition 1. The following property of exponential random variables is used to evaluate the expectation in the Euler condition.

Lemma 3. *The cumulant generating function of a unit exponential variable V_{t+1} evaluated at s is*

$$\ln E_t\big[\exp(sV_{t+1})\big] = \ln\left(\frac{1}{1-s}\right).$$

The cumulant generating function of the disaster shock is therefore

$$\ln E_t\left[\exp\left(s\iota_{c,t+1}\Lambda_t V_{t+1}\right)\right] = \ln\left(1-\delta+\frac{\delta}{1-s\Lambda_t}\right). \tag{10}$$

Because $\iota_{c,t+1}$ is a $Bernoulli(\delta)$ variable,

$$E_t\left[\exp\left(s\iota_{c,t+1}\Lambda_t V_{t+1}\right)\right] = (1-\delta)(1)+\delta E_t\left[\exp\left(s\Lambda_t V_{t+1}\right)\right].$$

Employing the cumulant generating function to the second term proves the lemma. To obtain log prices that are linear in Λ_t, I use a first order Taylor expansion of (10) around $\Lambda_t = \bar{\Lambda}$:

$$\ln\left(1-\delta+\frac{\delta}{1-s\Lambda_t}\right) \approx d(s)+c(s)\Lambda_t,$$

where $d(s) = \ln\left(1+\frac{\delta s\bar{\Lambda}}{1-s\bar{\Lambda}}\right) - c(s)\bar{\Lambda}$ and $c(s) = \delta s\big((1-s\bar{\Lambda})^2+\delta s\bar{\Lambda}(1-s\bar{\Lambda})\big)^{-1}$. Finally, evaluating the expectation results in the condition

$$\begin{aligned} 1 \;=\; \exp\Bigg(&\theta\ln\beta+(1-\gamma)\mu+\theta\big(\kappa_0+A_0[\kappa_1-1]+\kappa_1[A_\sigma\bar{\sigma}^2(1-\rho_\sigma)+A_\Lambda\bar{\Lambda}(1-\rho_\Lambda)]\big)+d(\gamma-1)\\ &+\frac{1}{2}(\theta\kappa_1)^2(A_\sigma^2\sigma_\sigma^2+A_\Lambda^2\sigma_\Lambda^2)+\sigma_t^2\big[\theta A_\sigma(\kappa_1\rho_\sigma-1)+\frac{1}{2}(1-\gamma)^2\sigma_c^2\big]\\ &+\Lambda_t\big[\theta A_\Lambda(\kappa_1\rho_\Lambda-1)+c(\gamma-1)\big]\Bigg). \end{aligned}$$

This equation is an equilibrium restriction yielding implicit solutions for A_σ and A_Λ. Of

that gives rise to the $\sim$ notation shown in the proposition. Rigorous proofs of asymptotic aggregation rules may be found in Gnedenko and Kolmogorov (1968).

particular interest is the coefficient on Λ_t,

$$A_\Lambda = \frac{-\delta c(\gamma - 1)}{\theta(1 - \kappa_1 \rho_\Lambda)} < 0.$$

(The sign of A_Λ may be established numerically.)

In the case of the price-dividend ratio for asset i, the result follows from the next equilibrium condition.

$$\begin{aligned} 1 \quad = \quad \exp\Bigg\{ & \theta \ln \beta + (\phi_i - \gamma)\mu + \mu_i + (\theta - 1)\Big[\kappa_0 + A_0(\kappa_1 - 1) + \kappa_1\big(A_\sigma \bar{\sigma}^2(1 - \rho_\sigma) + A_\Lambda \bar{\Lambda}(1 - \rho_\Lambda)\big)\Big] \\ & + \kappa_{i,0} + A_{i,0}[\kappa_{i,1} - 1] + \kappa_{i,1}[A_{i,\sigma}\bar{\sigma}^2(1 - \rho_\sigma) + A_{i,\Lambda}\bar{\Lambda}(1 - \rho_\Lambda)] + d(\gamma - \phi_i) \\ & + d(-q_i) + \frac{1}{2}\big([\theta - 1]\kappa_1 A_\sigma + \kappa_{i,1} A_{i,\sigma}\big)^2 \sigma_\sigma^2 + \frac{1}{2}\big([\theta - 1]\kappa_1 A_\Lambda + \kappa_{i,1} A_{i,\Lambda}\big)^2 \sigma_\Lambda^2 \\ & + \sigma_t^2 \Big[(\theta - 1) A_\sigma(\kappa_1 \rho_\sigma - 1) + A_{i,\sigma}(\kappa_{i,1}\rho_\sigma - 1) + \frac{1}{2}(\phi_i - \gamma)^2 \sigma_c^2 + \frac{1}{2}\sigma_i^2\Big] \\ & + \Lambda_t \Big[(\theta - 1) A_\Lambda(\kappa_1 \rho_\Lambda - 1) + A_{i,\Lambda}(\kappa_{i,1}\rho_\Lambda - 1) + c(\gamma - \phi_i) + c(-q_i)\Big]\Bigg\}, \end{aligned}$$

which delivers the result. □

Proof of Proposition 5

Proof. The expected return on asset i is

$$E_t[r_{i,t+1}] = \kappa_{i,0} + \phi_i \mu + \mu_i + A_{i,0}(\kappa_{i,1} - 1) + A_{i,\sigma}(\kappa_{i,1}\rho_\sigma - 1)\sigma_t^2 + [A_{i,\Lambda}(\kappa_{i,1}\rho_\Lambda - 1) - \delta(\phi_i + q_i)]\Lambda_t.$$

Substituting the wealth-consumption ratio from Proposition 4 into the stochastic discount factor (via the Campbell-Shiller identity) and evaluating the expectation gives

$$r_{f,t} = r_{f,0} + b_{f,\sigma}\sigma_t^2 + b_{f,\Lambda}\Lambda_t$$

where $b_{f,\Lambda} = A_\Lambda(1 - \theta)(\kappa_1 \rho_\Lambda - 1) - c(-\gamma)$. Assembling the preceding expressions produces the result. □

Proof of Proposition 6

Proof. The proof proceeds as in Proposition 3. Let $S = -\phi_i \Lambda_t V_{c,t+1}$ and let $Y = \exp(S)$. The density of S is

$$g_S(s) = \frac{1}{\phi_i \Lambda_t} \exp\left(\frac{s}{\phi_i \Lambda_t}\right), \quad s \leq 0.$$

The derivative of S with respect to Y is $1/Y$, which implies that the density of Y is

$$g_Y(y) = g_S(s)\frac{ds}{dy} = \frac{1}{\phi_i \Lambda_t} y^{1/(\phi_i \Lambda_t)-1}$$

with corresponding cumulative distribution function

$$G_Y(y) = y^{1/(\phi_i \Lambda_t)}.$$

Accounting for the interaction between S and the Bernoulli variable $\iota_{c,t+1}$ amounts to deriving the distribution of $Y^{\iota_{c,t+1}}$, which is

$$G_{(Y^{\iota})}(y) = \delta 1_{y=1} + (1-\delta) y^{1/(\phi_i \Lambda_t)} \sim (1-\delta) y^{1/(\phi_i \Lambda_t)}.$$

Repeating this argument for $\exp(-q_i \iota_{i,t+1} \Lambda_t V_{i,t+1})$ shows that its lower tail is also power law-distributed with exponent $1/(q_i \Lambda_t)$. Applying the asymptotic tail aggregation properties referenced in Proposition 3 and noting that the remaining shocks to $r_{i,t+1}$ are Gaussian (thus they have no asymptotic contribution to the tail distribution) delivers the result. □

B Monte Carlo Evidence

B.1 Correct Specification of Tail Parameter Evolution

The first Monte Carlo experiment I run is designed to assess the finite sample properties of the dynamic power law quasi-maximum likelihood estimator under different dependence and heterogeneity conditions. In all cases, the evolution of the parameter governing tail risk follows Equation 7, and therefore the statistical model's specification of this process is correct. I allow for mis-specification in terms of dependence and in the level of the tail exponent across stock. In particular, data is generated by the following process:

$$R_{i,t} = b_i R_{m,t} + e_{i,t}$$

where $R_{m,t}$ and $e_{i,t}, i = 1, ..., n$, are independent Student t variates with $a_i \zeta_t$ degrees of freedom. A well-known property of the Student t is that its tail distribution is asymptotically equivalent to a power law with tail exponent equal to (minus) the degrees of freedom. In generating data I therefore set the degrees of freedom equal to ζ_t, whose transition is described by Equation 7. The b_i coefficients control cross section dependence and heterogeneity in volatility. The a_i coefficients control the tail risk heterogeneity across observations. I consider four cases:

1. Independent and identically distributed observations: $b_i = 0$ and $a_i = 1$ for all i,
2. Dependent and identically distributed observations: $b_i \sim N(1, .5^2)$ and $a_i = 1$ for all i,[26]
3. Independent and heterogeneously distributed observations: $b_i = 0$ and $a_i \sim N(1, .2^2)$ for all i,
4. Dependent and heterogeneously distributed observations: $b_i \sim N(1, .5^2)$ and $a_i \sim N(1, .2^2)$ for all i.

The cross section size is n=1000 or 2500, and the time series length is T=1000 or 5000. Parameters used to generate data are fixed at $\pi_1 = 0.05$ and $\pi_2 = 0.93$, with an intercept that ensures the mean value of ζ_t is three.

In each simulation, the quasi-maximum likelihood procedure described in Section 3 is used to estimate the model and its asymptotic standard errors. Summary statistics for parameter and asymptotic standard error estimates are reported in Table 11. Also reported is the time series correlation and mean absolute deviation between the fitted tail series and the true ζ_t series, averaged across simulations.

The general conclusion of the experiment is that the asymptotic theory of Section 3 is a good approximation for the finite sample behavior of the dynamic power law estimator. This is true not only when data are i.i.d., but also when observations are dependent and heterogenous. In all cases, the fitted tail series achieves a correlation of at least 97% with the true tail series.

[26] Observations in this case are identical only in terms of their tail exponent. Differences in b_i across stocks introduces volatility and dependence heterogeneity.

B.2 Correct Specification of Tail Parameter Evolution

The next experiment proceeds as in the i.i.d. case above, but the true tail diverges from that assumed in the statistical model. In particular, the true degrees of freedom parameter ζ_t is conditionally stochastic and follows a first order Gaussian autoregression,

$$\zeta_{t+1} = \bar{\zeta}(1-\rho) + \rho\zeta_t + \sigma\eta_{t+1},\ \eta_{t+1} \sim N(0,1). \tag{11}$$

I fix $\rho = 0.99$, n=1000 and T=1000, and let σ =0.005 or 0.010. The parameter σ governs the variability of the process, and thus the range of tail risk values that the data can experience.

In each simulation, the quasi-maximum likelihood procedure described in Section 3 is used to estimate the model and its asymptotic standard errors. Summary statistics for the true process and the fitted process are reported in Table 12, as well as summary statistics for π parameter estimates and their asymptotic standard errors.

The deterministic tail process provides accurate estimates even when the true tail parameter is stochastic. The mean absolute error between the fitted and true series ranges from 0.173 to 0.307, and their correlation ranges from 81.8% to 87.5%.

References

Ang, A. and Bekaert, G. (2007), "Stock return predictability: Is it there?" *Review of Financial Studies*, 20, 651.

Backus, D., Chernov, M., and Martin, I. (2009), "Disasters implied by equity index options," *NYU Working Paper*.

Bansal, R. and Shaliastovich, I. (2008), "Confidence Risk and Asset Prices," *Duke University Working Paper*.

— (2009), "Learning and Asset-Price Jumps," *Duke University Working Paper*.

Bansal, R. and Yaron, A. (2004), "Risks for the long run: A potential resolution of asset pricing puzzles," *Journal of Finance*, 1481–1509.

Barro, R. (2006), "Rare disasters and asset markets in the twentieth century," *The Quarterly Journal of Economics*, 121, 823–866.

Berk, R. (1966), "Limiting behavior of posterior distributions when the model is incorrect," *The Annals of Mathematical Statistics*, 51–58.

Bollerslev, T., Tauchen, G., and Zhou, H. (Forthcoming), "Expected stock returns and variance risk premia," *Review of Financial Studies*.

Bollerslev, T. and Todorov, V. (2009), "Tails, Fears and Risk Premia," *Duke University Working Paper*.

Boudoukh, J. and Richardson, M. (1993), "Stock returns and inflation: A long-horizon perspective," *The American Economic Review*, 83, 1346–1355.

Campbell, J. (1987), "Stock returns and the term structure," *Journal of Financial Economics*, 18, 373–399.

Campbell, J. and Shiller, R. (1988), "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies*, 1, 195–228.

Campbell, J. and Thompson, S. (2008), "Predicting excess stock returns out of sample: Can anything beat the historical average?" *Review of Financial Studies*, 21, 1509–1531.

Cochrane, J. (1999), "Portfolio advice for a multifactor world." *Economic Perspectives*, 23, 59–61.

Drechsler, I. and Yaron, A. (2009), "What's vol got to do with it?" *Wharton Working Paper*.

Drost, F. and Werker, B. (1996), "Closing the GARCH gap: Continuous time GARCH modeling," *Journal of Econometrics*, 74, 31–57.

Duffie, D., Pan, J., and Singleton, K. (2000), "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, 1343–1376.

Dupuis, D. (1999), "Exceedances over high thresholds: A guide to threshold selection," *Extremes*, 1, 251–261.

Engle, R. and Kelly, B. (2009), "Dynamic equicorrelation," *NYU Working Paper*.

Epstein, L. and Zin, S. (1989), "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework," *Econometrica*, 57, 937–969.

Eraker, B. and Shaliastovich, I. (2008), "An equilibrium guide to designing affine pricing models," *Mathematical Finance*, 18, 519–543.

Fama, E. (1963), "Mandelbrot and the stable Paretian hypothesis," *Journal of Business*, 36, 420–429.

Fama, E. and French, K. (1989), "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics*, 25, 23–49.

— (1993), “Common risk factors in the returns on stocks and bonds,” *Journal of Financial Economics*, 33, 3–56.

Fama, E. and MacBeth, J. (1973), “Risk, return, and equilibrium: Empirical tests,” *Journal of Political Economy*, 81, 607.

Ferson, W. and Harvey, C. (1991), “The variation of economic risk premiums,” *Journal of Political Economy*, 385–415.

Gabaix, X. (2009a), “Power laws in economics and finance,” *Annual Review of Economics*, 1.

— (2009b), “Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance,” *NYU Working Paper.*

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. (2006), “Institutional investors and stock market volatility,” *Quarterly Journal of Economics*, 121, 461–504.

Glosten, L., Jagannathan, R., and Runkle, D. (1993), “On the relation between the expected value and the volatility of the nominal excess return on stocks,” *Journal of Finance*, 48, 1779–1801.

Gnedenko, B. and Kolmogorov, A. (1968), *Limit distributions for sums of independent random variables*, Addison-Wesley.

Goyal, A. and Welch, I. (2008), “A comprehensive look at the empirical performance of equity premium prediction,” *Review of Financial Studies*, 21, 1455–1508.

Hansen, L. (1982), “Large sample properties of generalized method of moments estimators,” *Econometrica*, 1029–1054.

Hayashi, F. (2000), *Econometrics*, Princeton University Press.

Hill, B. (1975), “A simple general approach to inference about the tail of a distribution,” *The Annals of Statistics*, 3, 1163–1174.

Hodrick, R. (1992), “Dividend yields and expected stock returns: Alternative procedures for inference and measurement,” *Review of Financial Studies*, 5, 357–386.

Jansen, D. and De Vries, C. (1991), “On the frequency of large stock returns: putting booms and busts into perspective,” *The Review of Economics and Statistics*, 73, 18–24.

Koijen, R., Lustig, H., and Van Nieuwerburgh, S. (2009), “The Cross-Section and Time-Series of Stock and Bond Returns,” *NYU Working Paper.*

Krasker, W. (1980), “The “peso problem” in testing the efficiency of forward exchange markets,” *Journal of Monetary Economics*, 6, 269–276.

Lettau, M. and Ludvigson, S. (2001a), "Consumption, aggregate wealth, and expected stock returns," *Journal of Finance*, 815–849.

— (2001b), "Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying," *Journal of Political Economy*, 109, 1238–1287.

Lustig, H. and Van Nieuwerburgh, S. (2005), "Housing collateral, consumption insurance, and risk premia: An empirical perspective," *The Journal of Finance*, 60, 1167–1219.

Mandelbrot, B. (1963), "The variation of certain speculative prices," *Journal of business*, 36, 394.

Matthys, G. and Beirlant, J. (2000), "Adaptive threshold selection in tail index estimation," *Extremes and Integrated Risk Management*, 37–49.

McNeil, A. and Frey, R. (2000), "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach," *Journal of Empirical Finance*, 7, 271–300.

Nelson, D. (1990), "ARCH models as diffusion approximations," *Journal of Econometrics*, 45, 7–38.

Newey, W. and McFadden, D. (1994), "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

Newey, W. and West, K. (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Neyman, J. and Scott, E. (1948), "Consistent estimates based on partially consistent observations," *Econometrica*, 16, 1–32.

Polk, C., Thompson, S., and Vuolteenaho, T. (2006), "Cross-sectional forecasts of the equity premium," *Journal of Financial Economics*, 81, 101–141.

Richardson, M. and Smith, T. (1991), "Tests of financial models in the presence of overlapping observations," *Review of Financial Studies*, 227–254.

Rietz, T. (1988), "The equity risk premium: A solution," *Journal of Monetary Economics*, 22, 117–131.

Wachter, J. (2009), "Can time-varying risk of rare disasters explain aggregate stock market volatility?" *Wharton Working Paper*.

White, H. (1982), "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–25.

Table 1. Dynamic Power Law Estimates for the NYSE/AMEX/NASDAQ Panel.
The table reports estimates for the dynamic power law model using the panel of NYSE/AMEX/NASDAQ stocks from August 1962 to December 2008. Estimation follows the quasi-maximum likelihood procedure described in Section 3. The model is estimated separately using data from both tails together, the lower tail alone and the upper tail alone. Results in Panel A are based on raw returns and results in Panel B are based on residuals from the Fama-French three-factor model. Standard errors of coefficients are reported in parentheses, and are calculated from estimates of the quasi-maximum likelihood asymptotic covariance structure derived in Section 3. The parameter ζ is the implied unconditional mean tail exponent based on estimates of π_0, π_1 and π_2 (the reported standard error estimate for ζ has been appropriately transformed based on standard errors of π_0, π_1 and π_2).

	Both Tails	Lower Tail	Upper Tail
Panel A: Raw Returns			
ζ	2.110	2.201	1.872
	(0.021)	(0.044)	(0.018)
π_1	0.188	0.072	0.239
	(0.014)	(0.010)	(0.058)
π_2	0.798	0.923	0.683
	(0.015)	(0.011)	(0.092)
Panel B: Factor Model Residuals			
ζ	2.090	2.145	2.055
	(0.024)	(0.052)	(0.017)
π_1	0.211	0.118	0.182
	(0.013)	(0.007)	(0.014)
π_2	0.780	0.879	0.801
	(0.014)	(0.007)	(0.017)

Table 2. Tail Risk Correlation with Macroeconomic Variables.
The table reports correlations between tail risk estimates and macroeconomic variables. Tail risk ($-\zeta_t$) is estimated for the upper and lower tail separately and for both tails together by the dynamic power law estimator on raw return data for NYSE/AMEX/NASDAQ stocks. Macroeconomic variables included are the log dividend-price ratio, unemployment rate, inflation rate, growth rate of industrial production, Chicago Fed National Activity Index and the variance risk premium (VIX^2 minus realized S&P 500 variance). Since tail risk is measured daily, correlations are calculated based on month-end values. The sample horizon is 1963 to 2008 (1990-2008 for the variance risk premium).

	Lower $-\zeta$	Upper $-\zeta$	Both $-\zeta$	ln D/P	Unemp.	Infl.	Ind. Prod. Growth	CFNAI	VRP
Lower $-\zeta$	1.00								
Upper $-\zeta$	0.56	1.00							
Both $-\zeta$	0.91	0.64	1.00						
ln D/P	0.15	0.14	0.14	1.00					
Unemp.	0.53	0.39	0.47	0.57	1.00				
Inflation	-0.05	0.02	-0.07	0.36	0.08	1.00			
Ind. Prod. Gr.	-0.04	0.02	0.01	-0.07	-0.07	-0.01	1.00		
CFNAI	-0.10	-0.07	-0.06	-0.05	-0.14	0.03	0.46	1.00	
VRP	0.03	0.29	0.14	-0.11	-0.10	0.15	0.06	0.07	1.00

Table 3. Macroeconomic Determinants of Tail Risk.
The table reports results from predictive regressions for lower and upper tail risk ($-\zeta_t$) estimated from raw returns on the NYSE/AMEX/NASDAQ cross section. Regressions are run using monthly data. In addition to the lagged dependent variable, regressors include lagged realized equity market volatility (calculated as the realized daily volatility for the CRSP value-weighted index each month), unemployment rate, inflation rate, growth rate of industrial production, Chicago Fed National Activity Index and return on the aggregate market. Test statistics are reported below coefficients in italics and use Newey-West standard errors with twelve lags. The sample horizon is 1963 to 2008.

	$-\zeta_t$ (lower)							$-\zeta_t$ (upper)						
R. Vol_{t-1}	0.027	0.107						0.098	-0.009					
	1.1	*3.1*						*3.3*	*0.4*					
$Unemploy_{t-1}$	0.041		0.055					0.053		0.083				
	3.2		*3.4*					*2.6*		*3.3*				
$Inflation_{t-1}$	0.002			0.016				0.010			-0.022			
	0.2			*1.5*				*0.5*			*1.1*			
Ind. Pro. $Gr._{t-1}$	-0.002				-0.017			-0.027				-0.020		
	0.1				*0.9*			*0.7*				*0.7*		
$CFNAI_{t-1}$	0.008					-0.017		0.046					0.000	
	0.4					*1.2*		*1.0*					*0.0*	
Market $Ret._{t-1}$	-0.175						-0.188	0.227						0.204
	7.1						*8.0*	*4.6*						*4.6*
Intercept	-0.239	0.004	-0.316	0.003	0.003	0.006	0.003	-0.312	0.005	-0.479	0.006	0.005	0.014	0.006
	2.8	*0.2*	*3.0*	*0.1*	*0.1*	*0.2*	*0.2*	*2.4*	*0.1*	*2.9*	*0.1*	*0.1*	*0.3*	*0.2*
Obs.	501	565	565	565	565	501	565	501	565	565	565	565	501	565
R^2	0.842	0.808	0.802	0.797	0.797	0.804	0.829	0.684	0.607	0.619	0.607	0.607	0.628	0.648

Table 4. Predicting Excess Aggregate Stock Market Returns (Raw Return Tails).

The table reports predictive regressions of CRSP value-weighted market index excess returns over one month, one year, three year and five year horizons. The first three predictors are the dynamic power law model tail risk process ($-\zeta_t$) estimated using both tails, the lower tail and the upper tail of raw returns for NYSE/AMEX/NASDAQ stocks (using RiskMetrics moving average weighting parameters). Since the tail process is estimated daily, forecasts use the value of $-\zeta_t$ on the last day before the forecast period. Other regressors are the log dividend-price ratio and monthly realized volatility of the CRSP value-weighted market index. Test statistics are calculated using Hodrick (1992) standard errors with lag length equal to the number of months in each horizon. The sample horizon is 1963 to 2008.

	One month horizon							One year horizon						
$-\zeta$ (both)	5.63				4.92			3.72				3.67		
	2.4				*2.0*			*1.8*				*1.7*		
$-\zeta$ (lower)		6.69				6.14			4.44				4.24	
		2.9				*2.6*			*2.1*				*2.0*	
$-\zeta$ (upper)			3.11				2.94			3.56				3.15
			1.4				*1.3*			*1.7*				*1.6*
ln D/P				2.60	2.21	2.01	2.21				3.57	3.29	3.16	3.14
				1.0	*0.9*	*0.8*	*0.9*				*1.4*	*1.3*	*1.2*	*1.2*
R. Volatility				-3.05	-1.94	-2.28	-3.21				1.06	1.92	1.63	0.78
				1.2	*0.7*	*0.9*	*1.2*				*0.8*	*1.3*	*1.2*	*0.6*
Intercept	4.36	4.35	4.36	4.33	4.34	4.34	4.34	5.54	5.57	5.58	5.61	5.63	5.64	5.60
	1.9	*1.9*	*1.9*	*1.9*	*1.9*	*1.9*	*1.9*	*2.3*	*2.3*	*2.3*	*2.3*	*2.4*	*2.4*	*2.3*
Obs.	565	565	565	565	565	565	565	556	556	556	556	556	556	556
R^2	0.011	0.016	0.003	0.006	0.014	0.019	0.009	0.048	0.069	0.044	0.042	0.086	0.103	0.076

	Three year horizon							Five year horizon						
$-\zeta$ (both)	3.96				3.93			4.76				4.81		
	1.9				*1.8*			*2.1*				*2.2*		
$-\zeta$ (lower)		4.52				4.37			5.02				4.81	
		2.2				*2.1*			*2.3*				*2.2*	
$-\zeta$ (upper)			3.87				3.70			4.73				4.46
			1.9				*1.8*			*2.1*				*2.0*
ln D/P				2.31	2.08	1.86	1.85				3.31	3.05	2.71	2.74
				1.0	*1.0*	*0.9*	*0.8*				*1.6*	*1.7*	*1.5*	*1.5*
R. Volatility				0.16	1.05	0.72	-0.23				0.87	1.96	1.44	0.38
				0.1	*0.8*	*0.6*	*0.2*				*0.7*	*2.2*	*1.6*	*0.4*
Intercept	6.46	6.57	6.50	6.46	6.40	6.51	6.38	7.51	7.71	7.55	7.34	7.24	7.46	7.25
	2.7	*2.8*	*2.7*	*2.5*	*2.7*	*2.8*	*2.7*	*2.7*	*2.8*	*2.6*	*2.3*	*2.7*	*2.8*	*2.6*
Obs.	532	532	532	532	532	532	532	508	508	508	508	508	508	508
R^2	0.148	0.198	0.143	0.048	0.188	0.229	0.176	0.238	0.272	0.236	0.097	0.329	0.339	0.302

Table 5. Predicting Excess Aggregate Stock Market Returns (Factor Model Residual Tails).

The table reports predictive regressions of CRSP value-weighted market index excess returns over one month, one year, three year and five year horizons. The first three predictors are the dynamic power law model tail risk process ($-\zeta_t$) estimated using both tails, the lower tail and the upper tail of Fama-French three-factor model residuals for NYSE/AMEX/NASDAQ stocks (using RiskMetrics moving average weighting parameters). Since the tail process is estimated daily, forecasts use the value of $-\zeta_t$ on the last day before the forecast period. Other regressors are the log dividend-price ratio and monthly realized volatility of the CRSP value-weighted market index. Test statistics are calculated using Hodrick (1992) standard errors with lag length equal to the number of months in each horizon. The sample horizon is 1963 to 2008.

	One month horizon							One year horizon						
$-\zeta$ (both)	4.23				3.79			4.25				3.97		
	1.9				*1.7*			*1.7*				*1.7*		
$-\zeta$ (lower)		4.29				4.28			4.61				4.46	
		2.0				*2.0*			*2.0*				*2.0*	
$-\zeta$ (upper)			6.74				6.46			3.50				3.10
			2.8				*2.7*			*1.4*				*1.4*
ln D/P				2.60	2.03	2.21	1.50				3.57	3.01	3.18	2.97
				1.0	*0.8*	*0.9*	*0.6*				*1.4*	*1.2*	*1.3*	*1.1*
R. Volatility				-3.05	-3.09	-3.43	-3.33				1.06	1.10	0.66	0.81
				1.2	*1.2*	*1.3*	*1.3*				*0.8*	*0.8*	*0.5*	*0.6*
Intercept	4.56	4.56	4.55	4.33	4.58	4.58	4.57	5.32	5.34	5.31	5.61	5.38	5.38	5.36
	2.0	*2.0*	*2.0*	*1.9*	*2.0*	*2.0*	*2.0*	*2.1*	*2.2*	*2.1*	*2.3*	*2.2*	*2.2*	*2.2*
Obs.	565	565	565	565	565	565	565	556	556	556	556	556	556	556
R^2	0.006	0.006	0.016	0.006	0.012	0.013	0.021	0.062	0.073	0.042	0.042	0.093	0.108	0.072

	Three year horizon							Five year horizon						
$-\zeta$ (both)	4.84				4.67			5.45				5.19		
	2.1				*2.1*			*2.4*				*2.4*		
$-\zeta$ (lower)		5.27				5.19			5.65				5.49	
		2.2				*2.2*			*2.5*				*2.4*	
$-\zeta$ (upper)			4.04				3.82			4.83				4.47
			1.8				*1.7*			*2.2*				*2.1*
ln D/P				2.31	1.74	1.93	1.69				3.31	2.63	2.86	2.52
				1.0	*0.8*	*1.0*	*0.8*				*1.6*	*1.5*	*1.6*	*1.4*
R. Volatility				0.16	0.29	-0.20	-0.05				0.87	1.03	0.51	0.63
				0.1	*0.2*	*0.2*	*0.0*				*0.7*	*1.2*	*0.6*	*0.7*
Intercept	6.30	6.35	6.27	6.46	6.23	6.25	6.18	7.56	7.63	7.53	7.34	7.31	7.32	7.27
	2.6	*2.7*	*2.5*	*2.5*	*2.6*	*2.7*	*2.5*	*2.6*	*2.7*	*2.5*	*2.3*	*2.7*	*2.8*	*2.5*
Obs.	532	532	532	532	532	532	532	508	508	508	508	508	508	508
R^2	0.225	0.268	0.155	0.048	0.252	0.304	0.181	0.313	0.339	0.242	0.097	0.374	0.412	0.297

Table 6. Market Return Predictability: Alternative Predictor Performance.
The table reports predictive regressions of CRSP value-weighted market index excess returns over one month, one year, three year and five year horizons. The first row repeats forecasting results for the lower tail risk series from Table 4. Results in all other rows but the last are from univariate regressions for the 1963-2007 sample using predictors considered in Goyal and Welch (2007) (data from Ivo Welch's website). The last row reports results using the variance risk premium (Bollerslev, Tauchen and Zhou 2009, data from Hao Zhou's website). In this case, regressions use data from 1990-2008 due to the unavailability of the variance risk premium prior to 1990. Test statistics are calculated using Hodrick (1992) standard errors with lag length equal to the number of months in each horizon.

	One month horizon			One year horizon			Three year horizon			Five year horizon		
	Coeff.	t-stat	R^2	Coeff.	t-stat	R^2	Coeff.	t-stat	R^2	Coeff.	t-stat	R^2
Tail Risk (-ζ lower)	**6.70**	**2.9**	**0.016**	**4.44**	**2.1**	**0.069**	**4.53**	**2.2**	**0.198**	**5.02**	**2.3**	**0.272**
Book-to-market	0.81	0.3	0.000	1.76	0.7	0.011	0.10	0.0	0.000	1.15	0.4	0.013
Cross section premium	5.53	2.4	0.010	-4.19	1.8	0.061	-4.29	2.5	0.168	-4.28	1.9	0.179
Default return spread	1.73	0.7	0.001	-0.08	0.1	0.000	-0.02	0.1	0.000	-0.12	0.6	0.000
Default yield spread	4.65	2.0	0.008	2.26	0.9	0.018	0.71	0.3	0.005	2.45	1.3	0.063
Dividend payout ratio	-0.27	0.1	0.000	0.88	0.4	0.003	0.87	0.4	0.006	1.90	0.7	0.028
Dividend price ratio	2.55	1.0	0.002	3.19	1.2	0.036	2.30	1.0	0.048	3.15	1.4	0.092
Dividend yield	2.65	1.1	0.003	3.18	1.2	0.036	2.25	1.0	0.046	3.10	1.4	0.089
Earnings price ratio	2.80	1.1	0.003	2.86	1.1	0.029	1.89	0.8	0.034	2.17	0.9	0.049
Inflation	-6.72	2.6	0.016	-1.99	1.1	0.014	-0.77	0.6	0.005	-0.21	0.1	0.000
Long term return	5.35	2.3	0.010	2.21	3.0	0.017	0.81	2.4	0.006	1.18	2.5	0.014
Long term yield	-0.70	0.3	0.000	1.50	0.6	0.008	2.41	1.4	0.053	3.93	2.0	0.151
Net equity expansion	-4.34	2.1	0.007	-2.11	0.9	0.016	-0.40	0.2	0.001	-1.04	0.5	0.010
Stock volatility	0.15	0.1	0.000	0.44	0.2	0.001	-0.35	0.2	0.001	-0.03	0.0	0.000
Term Spread	4.49	2.0	0.007	3.95	1.9	0.056	3.50	1.8	0.114	4.00	1.8	0.159
Treasury bill rate	-3.08	1.3	0.003	-0.81	0.3	0.002	0.23	0.1	0.001	1.26	0.6	0.016
Variance risk premium	9.73	3.3	0.035	4.34	3.1	0.038	-1.41	0.6	0.007	-4.64	2.9	0.090

Table 7. Market Return Predictability: Bivariate Alternative Predictor Performance.
The table repeats the analysis of Table 6, but instead uses bivariate regressions that include each alternative predictor alongside the lower tail risk process estimated with the dynamic power law model (using raw returns of NYSE/AMEX/NASDAQ stocks). Test statistics are calculated using Hodrick (1992) standard errors with lag length equal to the number of months in each horizon. For each horizon, the first two columns are the coefficient estimate and *t*-statistic for the alternative predictor, while the third and fourth columns are the coefficient and *t*-statistic for the tail risk process.

	One month horizon					One year horizon				
	Coeff.	*t*-stat	Tail Coeff.	Tail *t*-stat	R^2	Coeff.	*t*-stat	Tail Coeff.	Tail *t*-stat	R^2
Book-to-market	0.73	*0.3*	6.47	*2.8*	0.015	1.70	*0.7*	4.41	*2.1*	0.080
Cross section premium	16.9	*4.7*	16.7	*4.6*	0.070	-2.17	*0.9*	2.97	*1.3*	0.081
Default return spread	1.45	*0.6*	6.42	*2.8*	0.016	-0.28	*0.5*	4.44	*2.1*	0.070
Default yield spread	3.36	*1.5*	5.72	*2.5*	0.019	1.33	*0.5*	4.13	*2.0*	0.076
Dividend payout ratio	0.49	*0.2*	6.54	*2.8*	0.015	1.42	*0.7*	4.59	*2.2*	0.077
Dividend price ratio	1.80	*0.7*	6.27	*2.7*	0.016	2.70	*1.1*	4.11	*2.0*	0.096
Dividend yield	1.78	*0.7*	6.23	*2.7*	0.016	2.61	*1.0*	4.06	*1.9*	0.094
Earnings price ratio	1.66	*0.6*	6.18	*2.6*	0.016	2.11	*0.8*	4.04	*1.9*	0.085
Inflation	-6.31	*2.5*	6.05	*2.6*	0.030	-1.70	*1.0*	4.31	*2.0*	0.080
Long term return	4.56	*2.0*	5.86	*2.5*	0.023	1.64	*2.5*	4.21	*2.0*	0.079
Long term yield	-3.45	*1.5*	7.71	*3.2*	0.019	-0.10	*0.0*	4.46	*2.0*	0.070
Net equity expansion	-2.34	*1.1*	5.66	*2.3*	0.017	-0.63	*0.3*	4.21	*2.0*	0.071
Stock volatility	1.02	*0.4*	6.62	*2.9*	0.016	1.04	*0.6*	4.56	*2.1*	0.074
Term Spread	2.46	*1.0*	5.59	*2.3*	0.017	2.70	*1.3*	3.45	*1.6*	0.092
Treasury bill rate	-3.92	*1.7*	6.96	*3.0*	0.021	-1.37	*0.5*	4.59	*2.2*	0.077
Variance risk premium	9.19	*2.9*	6.73	*2.0*	0.052	4.63	*3.3*	7.42	*2.6*	0.228

	Three year horizon					Five year horizon				
	Coeff.	*t*-stat	Tail Coeff.	Tail *t*-stat	R^2	Coeff.	*t*-stat	Tail Coeff.	Tail *t*-stat	R^2
Book-to-market	0.04	*0.0*	4.56	*2.2*	0.198	1.05	*0.4*	5.04	*2.3*	0.283
Cross section premium	-2.04	*1.2*	3.30	*1.6*	0.231	-1.45	*0.9*	4.15	*2.0*	0.286
Default return spread	-0.20	*0.8*	4.57	*2.2*	0.199	-0.32	*1.4*	5.07	*2.3*	0.273
Default yield spread	-0.34	*0.2*	4.64	*2.1*	0.199	1.35	*0.7*	4.75	*2.0*	0.290
Dividend payout ratio	1.50	*0.8*	4.74	*2.3*	0.216	2.67	*1.2*	5.34	*2.3*	0.325
Dividend price ratio	1.75	*0.8*	4.35	*2.1*	0.225	2.44	*1.2*	4.74	*2.2*	0.325
Dividend yield	1.61	*0.7*	4.34	*2.0*	0.221	2.29	*1.2*	4.72	*2.1*	0.319
Earnings price ratio	1.08	*0.5*	4.36	*2.0*	0.209	1.25	*0.5*	4.83	*2.1*	0.287
Inflation	-0.43	*0.4*	4.53	*2.2*	0.200	0.17	*0.1*	5.07	*2.3*	0.272
Long term return	0.16	*0.5*	4.54	*2.2*	0.199	0.44	*1.4*	5.00	*2.3*	0.273
Long term yield	0.83	*0.5*	4.27	*2.0*	0.204	2.24	*1.3*	4.24	*2.0*	0.313
Net equity expansion	1.60	*0.8*	5.13	*2.3*	0.217	1.08	*0.6*	5.44	*2.2*	0.281
Stock volatility	0.23	*0.2*	4.59	*2.2*	0.199	0.62	*0.6*	5.14	*2.4*	0.275
Term Spread	2.07	*1.2*	3.81	*1.8*	0.233	2.36	*1.3*	4.20	*2.1*	0.319
Treasury bill rate	-0.33	*0.2*	4.60	*2.2*	0.199	0.57	*0.3*	4.99	*2.3*	0.275
Variance risk premium	-0.79	*0.5*	6.57	*2.8*	0.289	-3.90	*2.9*	5.45	*1.9*	0.313

Table 8. Returns on Tail Risk Beta-Sorted Portfolios.

The table reports equal-weighted monthly returns (in annualized percentages) for portfolios of NYSE/AMEX/NASDAQ stocks from 1963-2008. At the beginning of month t+1, I form portfolios based on tail risk beta, market beta, size, and book-to-market ratio, estimated using data from the 60 months beginning at date t-59 and ending at date t. In Panel A, stocks are sorted into quintile portfolios by their beta on the tail risk measure, estimated by regressing monthly portfolio returns on monthly innovations from an AR(1) model for the tail risk series $-\zeta_t$ (which is itself estimated in a preliminary step with the dynamic power law model using both tails). In Panels B, C and D, stocks are first sorted into quintiles based on market beta (β_{MKT}), market equity and book-to-market, respectively. Then, within each quintile, stocks are further sorted into tail risk beta quintiles. Also, within each quintile (or, in the case of Panel A, among all stocks), I calculate the difference in average returns between high tail risk beta stocks and low tail risk beta stocks, as well as the associated t-statistic. For a stock to be included in a portfolio at t+1, I require that it has at least 36 months of non-missing return data out of the previous 60 months.

		Tail Risk Beta						
		Low 1	2	3	4	High 5	Diff. (5-1)	Diff. t-stat
Panel A: Tail Risk Beta Only								
All		6.40	7.13	6.23	4.44	0.36	-6.03	*2.55*
Panel B: Market Beta / Tail Risk Beta								
Low β_{MKT}	1	6.71	7.41	7.11	6.40	3.65	-3.06	*1.76*
	2	5.91	6.19	5.97	4.50	2.27	-3.64	*2.01*
	3	4.32	5.12	4.34	3.25	0.44	-3.88	*2.08*
	4	2.54	3.36	2.53	1.06	-1.01	-3.55	*1.87*
High β_{MKT}	5	-0.02	1.78	-0.35	-1.57	-4.45	-4.43	*2.20*
Panel C: Market Equity / Tail Risk Beta								
Small	1	10.16	9.27	10.15	10.67	13.98	3.82	*1.72*
	2	2.76	4.48	3.46	0.51	-4.33	-7.10	*2.81*
	3	4.81	5.81	4.99	0.86	-5.32	-10.13	*4.18*
	4	6.71	7.72	6.45	4.79	-2.18	-8.89	*3.69*
Big	5	6.76	6.82	6.80	5.68	1.43	-5.33	*2.31*
Panel D: Book-to-Market / Tail Risk Beta								
Growth	1	5.75	6.16	5.23	3.05	-1.88	-7.63	*3.07*
	2	7.50	7.34	7.07	5.23	1.94	-5.56	*2.37*
	3	9.14	8.78	8.11	7.22	4.71	-4.43	*1.97*
	4	10.35	9.93	9.00	8.98	7.11	-3.24	*1.52*
Value	5	11.09	10.66	11.64	12.01	14.06	2.96	*1.42*

Table 9. Cross-Sectional Return Regressions.

The table reports slope and intercept estimates from second-stage regressions of NYSE/AMEX/NASDAQ stock returns on factor exposures estimated in first-stage regressions. The first three factors are innovations from AR(1) models of tail risk series estimated using the dynamic power law model for both tails, the lower tail and the upper tail of raw returns (Panel A) and of Fama-French three-factor model residuals (Panel B). Other factors included are AR(1) innovations to the monthly realized volatility of the CRSP value-weighted market portfolio, the excess market return and Fama and French's (1993) SMB and HML returns. Coefficients are standardized to represent the change in average annualized percentage excess returns resulting from a one standard deviation increase in the regressor. The *t*-statistic is reported below each coefficient in italics. The sample horizon is 1963-2008.

Panel A: Raw Return Tails									
$-\zeta$ (both)	-5.303			-5.582			-4.399		
	2.5			*2.3*			*2.4*		
$-\zeta$ (lower)		-6.197			-5.782			-5.586	
		2.9			*2.7*			*2.9*	
$-\zeta$ (upper)			-0.112			-0.005			-1.617
			0.1			*0.0*			*1.2*
R. Volatility				6.001	6.309	6.863	5.209	4.663	4.978
				2.9	*3.1*	*3.6*	*2.8*	*2.6*	*3.0*
R_{mkt}^{e}							-1.249	-1.150	-1.087
							0.7	*0.7*	*0.6*
SMB							-3.787	-3.596	-3.583
							2.4	*2.3*	*2.3*
HML							-0.940	-0.930	-0.588
							0.6	*0.6*	*0.4*
Intercept	2.420	2.268	-0.645	4.082	4.023	4.561	5.264	4.883	4.647
	1.1	*1.0*	*-0.2*	*2.2*	*2.2*	*2.3*	*6.0*	*5.6*	*5.3*
R^2	0.030	0.029	0.020	0.053	0.056	0.031	0.143	0.151	0.135
Panel B: Factor Model Residual Tails									
$-\zeta$ (both)	-4.688			-3.847			-3.016		
	2.5			*2.0*			*1.9*		
$-\zeta$ (lower)		-3.199			-2.605			-3.881	
		2.1			*1.6*			*2.2*	
$-\zeta$ (upper)			-5.339			-4.479			-2.577
			2.9			*2.3*			*1.8*
R. Volatility				6.379	6.801	7.196	4.830	4.435	5.305
				3.2	*3.4*	*3.5*	*2.8*	*2.6*	*3.0*
R_{mkt}^{e}							-1.116	-1.075	-1.214
							0.7	*0.6*	*0.7*
SMB							-3.596	-3.506	-3.591
							2.3	*2.3*	*2.3*
HML							-0.923	-0.975	-0.582
							0.6	*0.6*	*0.4*
Intercept	0.406	-1.171	0.616	4.042	3.928	4.295	4.897	4.713	4.906
	0.2	*-0.4*	*0.2*	*2.0*	*1.9*	*2.1*	*5.6*	*5.4*	*5.6*
R^2	0.024	0.024	0.025	0.049	0.048	0.049	0.136	0.139	0.137

Table 10. Cross-Sectional Return Regressions: Alternative Test Assets (Raw Return Tails).

The table reports slope and intercept estimates from second-stage regressions of portfolio returns on factor exposures estimated in first-stage regressions. The first three factors are innovations from AR(1) models of tail risk series estimated using the dynamic power law model for both tails, the lower tail and the upper tail of raw returns. Other factors included are AR(1) innovations to the monthly realized volatility of the CRSP value-weighted market portfolio, the excess market return and Fama and French's (1993) SMB and HML returns. I consider three alternative sets of test assets: 1) NYSE-listed stocks, 2) 100 size and value-sorted portfolios (from Ken French's Data Library) and 3) 25 portfolios sorted on market beta and tail risk beta. Coefficients are standardized to represent the change in average annualized percentage excess returns resulting from a one standard deviation increase in the regressor. The *t*-statistic is reported below each coefficient in italics. The sample horizon is 1963-2008.

	NYSE Stocks						Size / Value-Sorted Portfolios						Market Beta / Tail Beta-Sorted Portfolios					
$-\zeta$ (both)	-4.367			-4.471			-1.200			-0.529			-3.176			-0.321		
	2.4			*3.3*			*1.6*			*3.2*			*3.0*			*0.8*		
$-\zeta$ (lower)		-5.273			-5.090			-1.802			-0.394			-4.238			-0.517	
		3.0			*3.5*			*3.3*			*2.4*			*3.7*			*1.9*	
$-\zeta$ (upper)			-0.856			-1.400			-0.360			-0.117			-0.441			-0.409
			0.8			*1.4*			*1.4*			*0.7*			*1.1*			*1.5*
R. Vol.	5.785	5.913	6.348	4.201	3.738	4.137	-0.453	-0.701	0.724	0.321	0.372	0.322	0.966	-0.100	4.266	0.288	0.252	0.233
	3.7	*3.8*	*4.3*	*3.4*	*3.0*	*3.5*	*1.2*	*1.2*	*1.0*	*1.6*	*1.9*	*1.6*	*1.7*	*0.2*	*4.3*	*1.5*	*1.4*	*1.3*
R_{mkt}^{e}				-1.458	-1.596	-1.193				-1.307	-1.233	-1.232				-2.185	-2.114	-2.218
				1.2	*1.2*	*1.0*				*4.2*	*4.0*	*4.1*				*3.3*	*3.2*	*3.2*
SMB				-3.338	-3.039	-3.118				0.716	0.665	0.692				0.070	0.004	-0.030
				2.8	*2.5*	*2.6*				*0.9*	*0.9*	*0.9*				*0.1*	*0.0*	*0.0*
HML				-1.936	-1.982	-1.772				2.365	2.321	2.353				2.223	2.011	1.967
				1.7	*1.8*	*1.6*				*3.7*	*3.6*	*3.7*				*3.5*	*3.2*	*3.2*
Intercept	7.196	7.473	7.400	8.378	8.386	7.598	13.127	16.977	12.943	23.886	23.021	23.085	20.545	19.674	26.604	14.583	14.829	14.918
	4.1	*4.4*	*4.1*	*7.6*	*7.6*	*7.0*	*4.3*	*4.7*	*4.0*	*8.9*	*8.5*	*8.6*	*9.0*	*8.7*	*7.1*	*5.2*	*5.3*	*5.3*
R^2	0.060	0.058	0.059	0.137	0.137	0.134	0.200	0.199	0.180	0.367	0.369	0.365	0.523	0.522	0.527	0.702	0.699	0.708

Table 11. Dynamic Power Law Monte Carlo Results.
The table reports simulation results based on the Monte Carlo experiment described in Appendix B. In all cases, data is generated by the process $R_{i,t}=b_iR_{m,t}+e_{i,t}$ where $R_{m,t}$ and $e_{i,t}$, i=1,...,n are independent Student t variates with $a_i\zeta_t$ degrees of freedom. I consider four cases: i) Independent and identically distributed observations: b_i=0 and a_i =1 for all i, ii) dependent and identically distributed observations: $b_i \sim N(1,.5^2)$ and a_i=1 for all i, iii) independent and heterogeneously distributed observations: b_i =0 and $a_i \sim N(1,.2^2)$ for all i, and iv) dependent and heterogeneously distributed observations: $b_i \sim N(1,.5^2)$ and $a_i \sim N(1,.2^2)$ for all i. The cross section size is n=1,000 or 2,500 and the time series length is T=1,000 or 5,000. Parameters used to generate the data are shown in the "True Value" row. I report the mean, median and standard deviation of parameter estimates across simulations, as well as the mean asymptotic standard error estimate. In the last row of each set of results, I report the mean absolute error and the correlation between the fitted and true ζ_t series. The column heading *d.o.f.* denotes estimates of the intercept parameter (transformed to be interpreted as the time series mean of ζ_t). Results are based on 1,000 replications.

		Independent, Identical			Dependent, Identical			Independent, Heterogeneous			Dependent, Heterogeneous		
		π_1	π_2	*d.o.f.*	π_1	π_2	*d.o.f.*	π_1	π_2	*d.o.f.*	π_1	π_2	*d.o.f.*
							T=1,000						
	True Value	0.050	0.930	3.000	0.050	0.930	3.000	0.050	0.930	3.000	0.050	0.930	3.000
n=1,000	Mean	0.042	0.914	2.910	0.042	0.907	3.248	0.050	0.922	2.358	0.049	0.920	2.640
	Median	0.041	0.925	2.909	0.041	0.919	3.255	0.050	0.925	2.343	0.048	0.926	2.637
	Mean ASE	0.015	0.040	0.069	0.015	0.049	0.085	0.016	0.029	0.172	0.014	0.029	0.091
	Std. Dev.	0.014	0.049	0.067	0.015	0.074	0.100	0.014	0.029	0.157	0.015	0.031	0.115
	MAE, Corr.	0.053	0.980		0.086	0.980		0.166	0.969		0.104	0.981	
n=2,500	Mean	0.044	0.920	2.700	0.043	0.912	3.019	0.057	0.923	2.257	0.052	0.925	2.460
	Median	0.043	0.926	2.699	0.042	0.923	3.023	0.057	0.924	2.153	0.051	0.927	2.401
	Mean ASE	0.014	0.032	0.048	0.015	0.037	0.051	0.014	0.023	0.125	0.014	0.025	0.144
	Std. Dev.	0.015	0.041	0.057	0.014	0.046	0.078	0.016	0.030	0.230	0.015	0.028	0.178
	MAE, Corr.	0.085	0.973		0.033	0.977		0.202	0.960		0.153	0.955	
							T=5,000						
n=1,000	Mean	0.041	0.927	2.900	0.040	0.927	3.255	0.049	0.929	2.297	0.049	0.928	2.605
	Median	0.042	0.928	2.901	0.040	0.928	3.255	0.049	0.930	2.299	0.049	0.929	2.609
	Mean ASE	0.006	0.013	0.031	0.006	0.014	0.035	0.006	0.010	0.036	0.006	0.010	0.041
	Std. Dev.	0.006	0.013	0.029	0.007	0.013	0.037	0.006	0.010	0.053	0.006	0.010	0.065
	MAE, Corr.	0.043	0.996		0.076	0.996		0.170	0.994		0.105	0.997	
n=2,500	Mean	0.044	0.927	2.684	0.043	0.927	3.021	0.054	0.933	2.232	0.053	0.927	2.352
	Median	0.044	0.929	2.685	0.043	0.927	3.016	0.054	0.931	2.083	0.052	0.928	2.345
	Mean ASE	0.006	0.012	0.019	0.006	0.012	0.024	0.006	0.008	0.186	0.006	0.009	0.033
	Std. Dev.	0.007	0.013	0.021	0.006	0.012	0.024	0.005	0.013	0.267	0.005	0.008	0.078
	MAE, Corr.	0.085	0.995		0.020	0.997		0.203	0.980		0.158	0.990	

Table 12. Stochastic Tail Exponent Monte Carlo Results.

The table reports simulation results based on the Monte Carlo experiment described in Appendix B. In all cases, data is generated as a vector of n i.i.d. Student t variates with ζ_t degrees of freedom over T periods, where $\zeta_{t+1} = \zeta(1-\rho)+\rho\zeta_t+\sigma\eta_{t+1}$, η_{t+1} is standard normal, n=1,000, T=1,000, and ρ=0.999. The standard deviation of the tail risk process is σ=0.005 or 0.010 (Panels A and B, respectively). I report summary statistics for the true and fitted tail processes, as well as their mean absolute error and correlation, averaged over all simulations. I also report summary statistics of parameter estimates and the mean asymptotic standard error estimate. The column heading *d.o.f.* denotes estimates of the intercept parameter (transformed to be interpreted as the time series mean of ζ_t). Results are based on 1,000 replications.

Panel A: σ=0.005

	Mean	Std. Dev.	Max	Min
True ζ_t	3.194	0.517	4.688	2.346
Fitted ζ_t	3.045	0.311	3.805	2.428
True/Fitted MAE	0.264			
True/Fitted Correlation	0.818			

Parameter Estimates	π_1	π_2	*d.o.f.*
Mean	0.036	0.954	2.968
Median	0.036	0.957	2.950
Mean ASE	0.010	0.015	0.239
Std. Dev.	0.010	0.020	0.209

Panel B: σ=0.010

	Mean	Std. Dev.	Max	Min
True ζ_t	3.464	1.179	7.390	1.909
Fitted ζ_t	3.103	0.634	4.752	1.972
True/Fitted MAE	0.525			
True/Fitted Correlation	0.875			

Parameter Estimates	π_1	π_2	*d.o.f.*
Mean	0.070	0.923	2.955
Median	0.070	0.924	2.948
Mean ASE	0.013	0.015	0.340
Std. Dev.	0.015	0.017	0.303

Figure 1. CRSP Cross Section Size, 1926-2008.
This figure plots the number of NYSE/AMEX/NASDAQ stocks in the CRSP database each month from 1926 to 2008. The count jumps in July 1962 with the addition of AMEX and in December 1972 with the addition of NASDAQ.

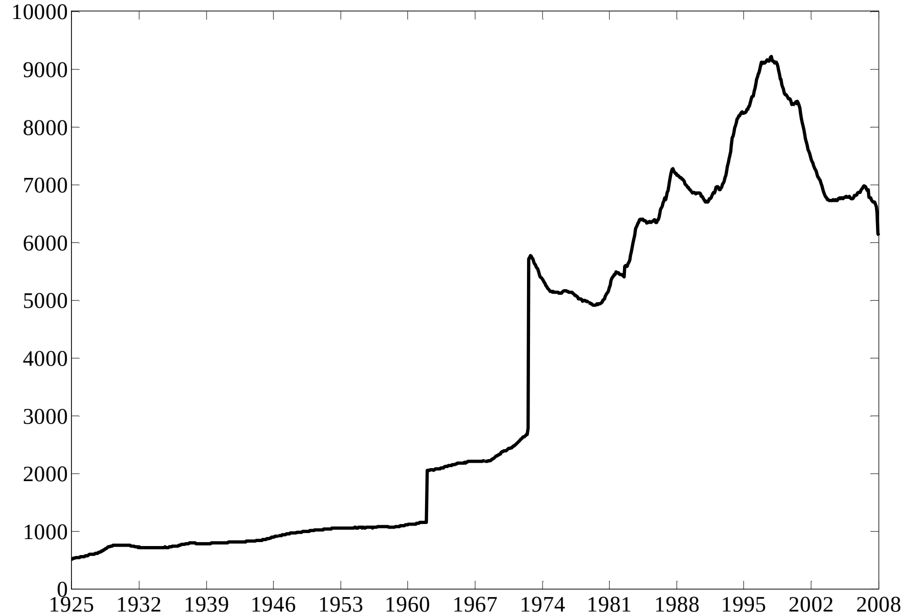

Figure 2. Tail Exponent Estimates (Raw Returns).
Plotted is the fitted tail risk time series ($-\zeta_t$) on the last day of each month. Estimates are found using the dynamic power law model and raw returns for the cross section of NYSE/AMEX/NASDAQ stocks from 1963 to 2008. The tail series is estimated using both tails together (Figure 2a) and the lower and upper separately (Figure 2b). Tail exponent values are shown on the right vertical axis. Shown in the grey shaded region is the aggregate log price-dividend ratio, whose scale is shown on the left vertical axis.

Figure 2a. Exponent Estimated from Both Tails.

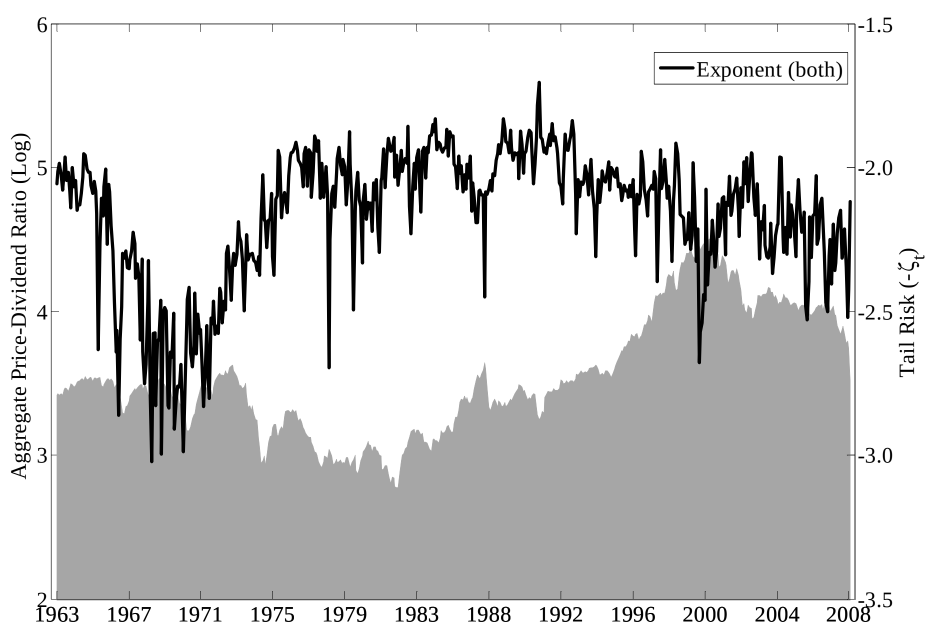

Figure 2. Continued.

Figure 2b. Exponent Estimated from Lower and Upper Tails Separately.

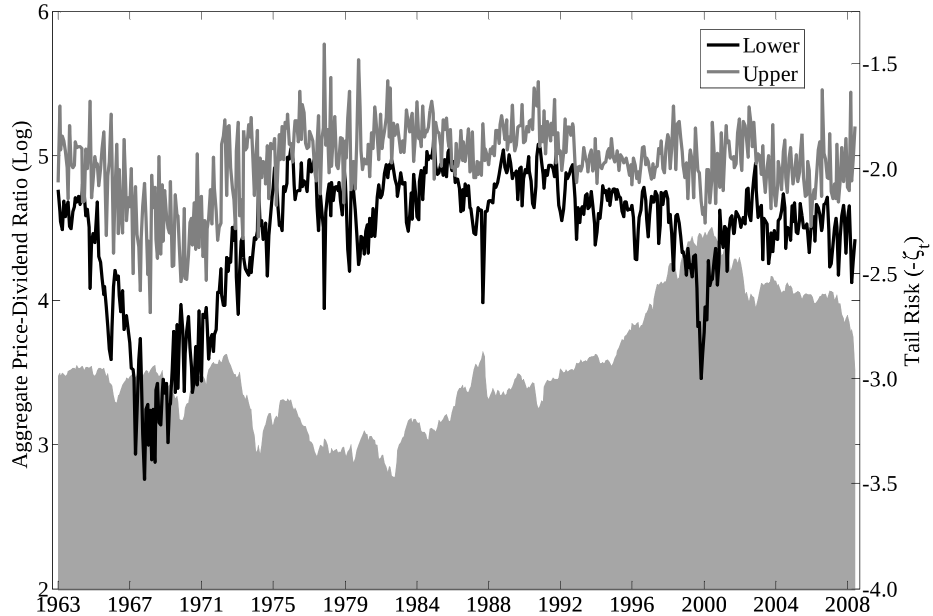

Figure 3. Daily Tail Exponent Estimates (Factor Model Residuals).

Plotted is the fitted tail risk time series ($-\zeta_t$) on the last day of each month. Estimates are found using the dynamic power law model and Fama-French three-factor model residuals for the cross section of NYSE/AMEX/NASDAQ stocks from 1963 to 2008. The tail series is estimated using both tails. Tail exponent values are shown on the right vertical axis. Shown in the grey shaded region is the aggregate log price-dividend ratio, whose scale is shown on the left vertical axis.

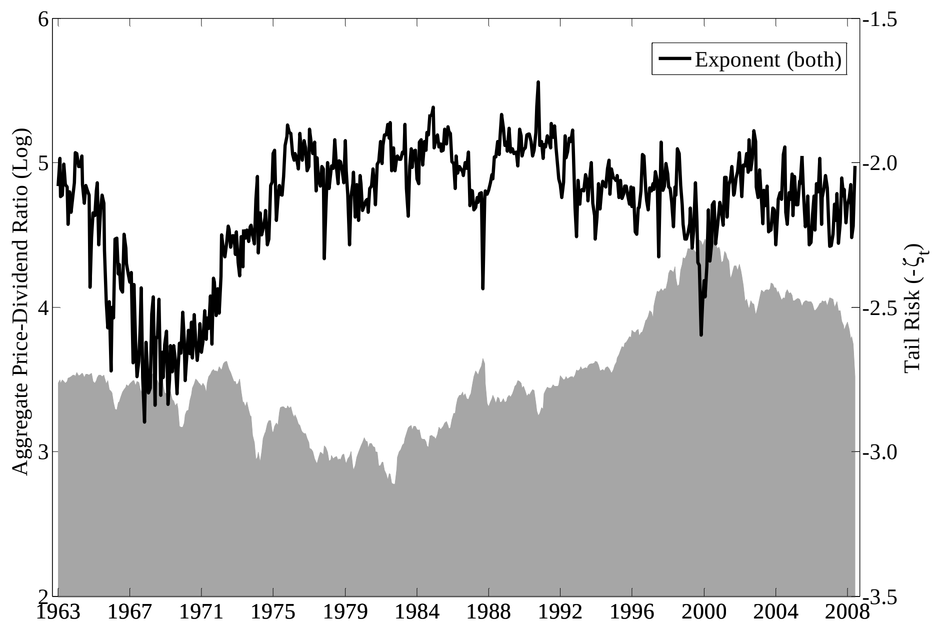

Figure 4. Tail Threshold and Aggregate Market Volatility (Raw Returns).
Plotted is the tail threshold series on the last day of each month. Estimates are found using the dynamic power law model and raw returns for the cross section of NYSE/AMEX/NASDAQ stocks from 1963 to 2008. The tail series is estimated using both tails together. Also shown is the monthly realized volatility of the CRSP value-weighted index.

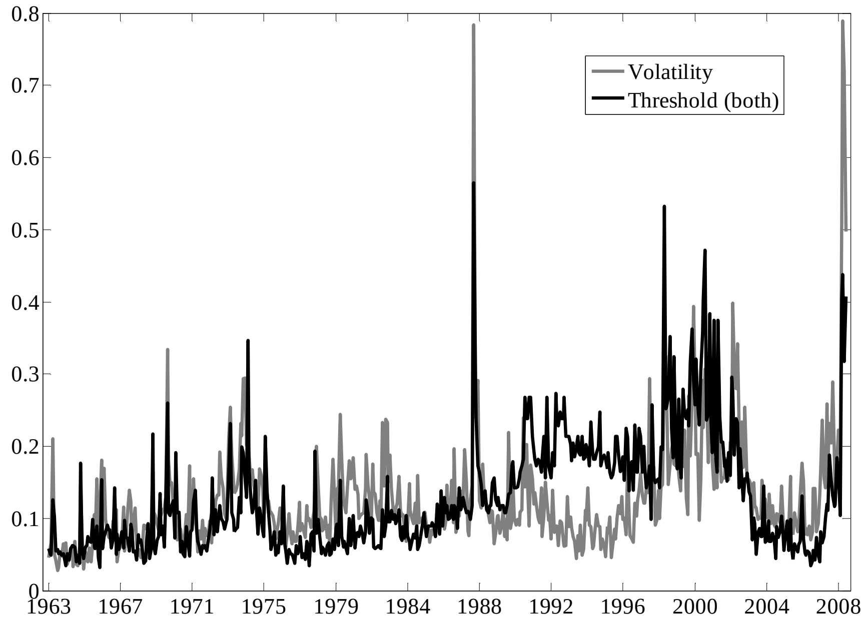

Figure 5. Out-of-Sample Aggregate Stock Market Predictive Regressions.
In each month t (beginning at t=60 to allow for a sufficiently large initial estimation period), I estimate rolling univariate forecasting regressions of monthly CRSP value-weighted index returns on tail risk ($-\zeta_t$, estimated from the lower tail of raw returns). Estimates only use data through date t; these are then used to forecast returns at t+1. The thick black line shows the sequence of estimated coefficients, and dotted lines represent 95% confidence intervals. The out-of-sample R^2 from this procedure is 1.30%.