News and Asset Pricing: A High-Frequency Anatomy of the SDF*

Saketh Aleti† and Tim Bollerslev‡

First Draft: September 7, 2022
This Version: December 19, 2022

Abstract

We rely on a unique set of high-frequency factors to robustly estimate an intraday Stochastic Discount Factor (SDF). Exploiting the precisely timed jumps in the estimated SDF together with real-time newswire data, we identify and precise the news that is priced. We find that news related to monetary policy and finance on average account for the largest portion of the variation in the SDF and the tangency portfolio risk premium, followed by news about international affairs and macroeconomic data. Reflecting investors changing economic concerns, we also uncover significant temporal variation in the relative importance of the news that matters. Relying on a standard mimicking portfolio approach, we further document marked differences in the way in which the news, and the compensation therefor, manifests in the "factor zoo."

Keywords: SDF, high-frequency factors, jumps, news, risk premiums.

JEL classification: C58, G12, G14.

*We are grateful to Anna Cieslak, Jia Li, Andrew Patton, George Tauchen, and participants at the December 2022 EC2 Conference in Paris on the Econometrics of High-Frequency Data and Factor Models for useful comments and suggestions. We would also like to thank the Carlsberg Foundation for their financial support to help purchase the Dow Jones Newswire data.

†Department of Economics, Duke University, Durham, NC 27708; e-mail: saketh.aleti@duke.edu.

‡Department of Economics, Duke University, Durham, NC 27708, NBER, and CREATES; e-mail: boller@duke.edu.

1 Introduction

A central tenet of asset pricing holds that in a frictionless financial market there exists a unique stochastic discount factor (SDF), or pricing kernel, that prices all risky assets (e.g., Back (2010)). The classical ICAPM of Merton (1973) further stipulates that changes in this SDF should be related to changes in the state variables that determine the future investment opportunity set. But what is the economic news that accounts for the changes in these state variables and how is this news priced?

We seek to provide new insights to these fundamental questions by empirically linking large intraday changes in a robustly estimated SDF to directly observable real-time economic news and corresponding news topics. We find that news related to monetary policy and finance typically accounts for the largest portion of the variation in the SDF and the tangency portfolio risk premium, followed by news about international affairs and macroeconomic data. Even though news about monetary policy on average commands the highest news risk premium, we also find that the relative importance of different news, as manifest in our estimated topic news risk premia, varies significantly over time, reflecting investors changing economic concerns. Building on the same mimicking portfolio approach underlying these results, we further document marked differences in the way in which the different news topics, and the compensation therefor, manifest in the "factor zoo."

Our estimation of the high-frequency SDF closely follows the minimax adversarial estimation approach recently advocated by Chen, Pelger, and Zhu (2022), including the use of neural networks for flexibly approximating the unknown functional form of the SDF. This approach is directly motivated by the seminal work of Hansen and Jagannathan (1997), and the result that the approximating SDF closest to the true SDF in a mean-square-error sense may be obtained by minimizing the largest possible squared pricing errors. In contrast to Chen, Pelger, and Zhu (2022), however, who base their estimation of a monthly SDF on monthly individual stock returns, we purposely rely on the high-frequency factor zoo recently constructed by Aleti (2022) for spanning the intraday SDF. Using factor portfolios instead of individual stocks allows us to succinctly incorporate the many pricing anomalies documented in the existing asset pricing literature, while simultaneously affording a greater degree of robustness to market microstructure noise in the high-frequency-based estimation. It also heeds the concerns of Kozak and Nagel (2022), who explicitly caution against the use of "too few" factors to span the SDF.

The idea of associating volatility in the SDF with specific news topics is perhaps most closely related to the recent work by Bybee, Kelly, and Su (2022), who combine textual analysis of daily news articles with latent factor analysis to indirectly infer a set of systematic narrative news risk factors and eventually a univariate pricing kernel. However, instead of first estimating different news factors, we directly link large changes, or "jumps," in our estimate of the high-frequency SDF to specific news topics.[1] Our news topics are based on the hand-constructed categorizations previously developed by Baker et al. (2019) and Baker et al. (2021) in their news-based explanations for aggregate stock

[1] Our identification of the "jumps" in the estimated tangency portfolio formally relies on infill asymptotic arguments adapted from the high-frequency financial econometrics literature (see, e.g., Aït-Sahalia and Jacod (2014)), together with the thresholding technique of Mancini (2001). Bajgrowicz, Scaillet, and Treccani (2016) have argued that this approach may also falsely identify rapid bursts in volatility as "jumps." We purposely do not try to distinguish between these two alternative theoretical-based explanations for the large intraday changes in the SDF, simply referring to both as "SDF jumps" in the sequel.

market volatility and large market moves, respectively.[2] To allow for a more aggregated bird's-eye view of the resulting variance decompositions of the SDF and the estimated topic news risk premia, we further combine some of these previously defined news topics into a smaller set of easy-to-interpret so-called metatopics. By focusing on the SDF and the tangency portfolio returns, as opposed to the returns on the aggregate market portfolio as done in the previous work by Baker et al. (2021) and others, we thus obtain a more accurate picture of the news that actually matters from a pricing perspective.

Importantly, and in contrast to the above-cited previous studies, all of which rely on daily newspaper articles together with daily returns for deciphering the news that matters, we purposely rely on the *Dow Jones Newswires Archive*, a machine-readable collection of articles from the Dow Jones' real-time news feeds commonly used by active market participants, together with our high-frequency estimates of the jumps in the SDF. Utilizing this precisely timed financial economic news source together with the intraday SDF jumps, allows us to substantially sharpen our empirical analyses and inference, by focussing on narrow time windows around the exact time of the jumps when there is likely a singular explanation for the large estimated change in the SDF.[3] By comparison, the use of daily news articles and coarser daily returns invariably blur the distinction between large price changes, or intraday jumps, observed in response to specific news arrivals, and "smooth" intraday price moves associated with the more gradual incorporation of the varied information contained in the plethora of news articles that are published on any given day. Our high-frequency-based approach instead allows for much more precise identification of the news that matters and in turn more accurate estimation of the news-driven SDF and the resulting news risk premia.[4]

The advantage of using high-frequency data for more accurately identifying and estimating news announcement effects has also previously been emphasized in the literature. In particular, Fair (2002) explicitly points to the weak identification afforded by the use of "coarse" daily data as the reason for the apparent lack of a clear news-based explanation for many of the largest stock market changes reported in the widely cited study by Cutler, Poterba, and Summers (1989). Relatedly, there is now also a large existing, mostly empirically oriented, literature pertaining to macroeconomic news announcement effects and the way in which announcement surprises affect the intraday returns on specific assets and/or asset classes; see, e.g., Andersen et al. (2003, 2007), Faust et al. (2007), Lee and Mykland (2008), Evans (2011), Lahaye, Laurent, and Neely (2011), and Lee (2012), along with the many other references therein. There is also a more recent growing literature emphasizing the advantages of using high-frequency data for obtaining more reliable identification of various economic mechanisms through heteroskedasticity and the idea that return volatilities tend to be higher shortly after macroeconomic and other scheduled

[2] To allow for more recent news trends, we further augmented these previously defined news topics with a few additional key terms, most notably related to the COVID-19 pandemic. For completeness, we also include a few missing terms from the "automatic" news topic categorizations in the studies by Bybee et al. (2021) and Bybee, Kelly, and Su (2022).

[3] In a similar vein, Chinco, Clark-Joseph, and Ye (2018) have recently sought to link the emergence and disappearance of high-frequency cross-sectional predictive relationships to precisely timed news stories about firm fundamentals. Jeon, McCurdy, and Zhao (2022) similarly associate precisely timed firm-specific news with daily "jumps" in a large cross-section of individual stock returns, while Christensen, Timmermann, and Veliyev (2022) document jumps in after-hours prices immediately following earnings announcements.

[4] It also implicitly assumes that the most important news for the pricing of US assets occurs during regular US trading hours. This, of course, does not rule out international news, only international news that is systematically released outside US market hours.

news announcements; see, e.g, Bollerslev, Li, and Xue (2018), Nakamura and Steinsson (2018), and Bianchi, Kung, and Kind (2022), among others. The present paper adds to both of these literatures by formally characterizing the economic significance and relative importance of *all* different types of news, including unscheduled announcements, through the decomposition of the tangency portfolio risk premium into separate news risk premia associated with a set of well-defined news topics.

The remainder of the paper is organized as follows. Section 2 begins by explaining the methodology that we use for robustly estimating the SDF tangency portfolio, followed by a discussion of the high-frequency factor returns underlying the estimation, before presenting the actual tangency portfolio estimates. Section 3 discusses the real-time newswire data that we rely on for precisely timing the news, along with our procedure for linking the jumps in the tangency portfolio to specific news topics. Our main results pertaining to the relative importance for explaining the variation in the SDF and the pricing of the different news topics are presented in Sections 4 and 5, respectively. Section 6 concludes with a brief summary and a few suggestions for future research. More detailed explanations of our estimation methodology, lists of key terms underlying our text processing algorithm, along with various supportive empirical results and analyses are deferred to a series of Appendixes. Additional empirical results and robustness checks are also provided in an Online Supplemental Appendix.

2 Estimating the High-Frequency SDF

Our estimation of the high-frequency stochastic discount factor (SDF) is based on the same general adversarial method of moments procedure recently advocated by Chen, Pelger, and Zhu (2022) (henceforth CPZ) in their estimation of a monthly SDF. This approach is directly motivated by the theoretical results in Hansen and Jagannathan (1997), formally establishing that the approximate SDF closest to the true SDF in a mean-square-error sense may be obtained by minimizing the largest possible squared pricing errors. A similar approach has also previously been proposed by Bansal, Hsieh, and Viswanathan (1993).

2.1 Methodology

To fix ideas, consider the canonical conditional moment restrictions implied by the standard no-arbitrage condition,

$$E_t[M_{t+1}R^e_{i,t+1}] = 0, \tag{1}$$

where $R^e_{i,t+1}$ denotes the excess return on asset i from time t to $t+1$, and M_{t+1} refers to the change in the SDF over that same time interval. Projecting the SDF onto the space of returns, it readily follows that

$$M_{t+1} = 1 - \hat{w}_t^\intercal R^e_{t+1}, \tag{2}$$

where $\hat{w}_t^\intercal R^e_{t+1}$ equals the return on the projected tangency portfolio, defined by

$$\hat{w}_t = \left(E_t[R^{e,\intercal}_{t+1}R^e_{t+1}]\right)^{-1} E_t[R^e_{t+1}]. \tag{3}$$

While this straightforward solution for the SDF involves a simple function of estimable quantities, it is challenging to implement in practice due to the difficulties in accurately estimating the conditional risk premia $E_t[R^e_{t+1}]$, and the inversion of the potentially high-dimensional second-moment matrix $E_t[R^{e,\intercal}_{t+1} R^e_{t+1}]$.

Instead, we estimate the weights for the tangency portfolio as a function of time t information. In particular, define:

$$w_t \equiv f_w(I_t\,;\,\theta_w), \tag{4}$$

where $I_t \in \mathbb{R}^K$ refers to the time t information set, and $f_w : \mathbb{R}^K \to \mathbb{R}^N$ is a known function of I_t parameterized by θ_w. Our estimate of the tangency portfolio,

$$F_{t+1} = 1 - f_w(I_t\,;\,\theta_w)R_{t+1}, \tag{5}$$

is then constructed by choosing the θ_w parameters such that this implied SDF minimizes the resulting conditional alphas for some deliberately chosen set of test assets.

Consistently estimating the conditional alphas over fixed time intervals is, of course, impossible without additional assumptions (see, e.g., Merton (1980)). Consequently, we rely on a set instrumented unconditional moment conditions directly implied by the conditional no-arbitrage restrictions in equation (1). That is,

$$E[M_{t+1}R^e_{t+1}g_t] = 0, \tag{6}$$

where the g_t vector of instruments is determined by the $f_g : \mathbb{R}^K \to \mathbb{R}^{N\times N_g}$ function parameterized by θ_g,

$$g_t \equiv f_g(I_t;\theta_g). \tag{7}$$

The $R^e_{t+1}g_t(I_t)$ term in (6) is readily interpreted as the return on a set of N_g portfolios, with the equation thus formally stipulating that the implied SDF is able to price all of these portfolios, or equivalently that no alpha can be found by trading on the I_t information set.

Of course, both the θ_g parameters that determine the optimal set of instruments and the θ_w parameters that determine the optimal weights are unknown and must be estimated. Motivated by the aforementioned results in Hansen and Jagannathan (1997), we jointly estimate these parameters based on the following minimax objective for the weights and instruments:[5]

$$\min_w \max_g \frac{1}{N_g}\sum_{i=1}^{N_g} \left\| \underbrace{\mathbb{E}\left[\overbrace{\left(1 - w_{t+1}R^e_{t+1}\right)}^{M_{t+1}} R^e_{t+1}g_{t+1,i}\right]}_{\text{Conditional Error } \alpha_{g,i}} \right\|^2 . \tag{8}$$

The inner maximization problem, pertaining to the instruments, may naturally be interpreted as that of an arbitrageur trying to pick the θ_g parameters for the function $f_g(I_t;\theta_g)$ so as to maximize the resulting portfolios' conditional alphas, say $\alpha_{g,i}$. The outer minimization problem, pertaining to the weights, may in turn be interpreted as that of an asset pricer trying to pick the θ_w parameters for the function $f_w(I_t;\theta_w)$ such that the implied SDF minimizes the conditional alphas for the specific set of portfolios, $R^e_{t+1}g_{t+1,i}$.

[5] In addition to providing the closest approximation to the true SDF in a mean-square-error sense, Chernozhukov et al. (2020) have recently shown that under additional regularity conditions, the SDF estimated by this minimax procedure will formally converge to the true SDF at an almost parametric rate.

2.2 Practical Implementation and Functional Approximations

The unconditional expectations defining the minimax problem in (8) is, of course, not directly observable. To obtain a practically feasible version, we replace the expected conditional alphas with their full-sample averages,

$$\hat{\alpha}_{g,i} = \frac{1}{T} \sum_{t} (1 - \hat{w}_{t+1} R^e_{t+1}) R^e_{t+1} \hat{g}_{t+1,i}, \tag{9}$$

resulting in the corresponding minimax problem:

$$\min_{w} \max_{g} \frac{1}{N_g} \sum_{i=1}^{N_g} \hat{\alpha}^2_{g,i}. \tag{10}$$

The practical implementation of this problem still necessitates a choice for the $f_g(I_t; \theta_g)$ and $f_w(I_t; \theta_w)$ instrument and weight functions. Again closely following CPZ, building on techniques from the recent Machine Learning (ML) literature, we rely on neural networks for flexibily approximating both of these functions. Neural network-based approximations have also previously been used in the literature for the non-parametric estimation of the SDF in more traditional method-of-moments-based settings by Bansal and Viswanathan (1993) and Chen and Ludvigson (2009), among others.

More precisely, we rely on the functional forms implicitly defined by,

$$\begin{aligned}
h^w_t &\equiv LSTM(I_t\,;\,\theta_{w,1}), \\
h^g_t &\equiv LSTM(I_t\,;\,\theta_{g,1}), \\
w_t &\equiv f_w(I_t\,;\,\theta_w) = FFN\left(h^w_t\,;\,\theta_{w,2}\right), \\
g_t &\equiv f_g(I_t\,;\,\theta_g) = FFN\left(h^g_t\,;\,\theta_{g,2}\right),
\end{aligned}$$

where the LSTMs denote long short-term memory neural networks with parameters $\theta_{w,1}$ and $\theta_{g,1}$, respectively, and the FFNs denote feedforward neural networks with parameters $\theta_{w,2}$ and $\theta_{g,2}$, respectively (see, e.g., Hochreiter and Schmidhuber (1997) for more formal definitions of the LSTM and FFN type networks). Intuitively, the LSTMs serve to condense the high-dimensional information in I_t into the lower-dimensional state variables, h^w_t and h^g_t, by recursively updating their past values with the relevant new information. The FFNs in turn use the constructed state variables to determine the weights and the instruments. We further restrict the weights for each asset to the $[-1, 1]$ interval. Accordingly, the FFN for the weights maps from $\mathbb{R}^{\dim(h^w)}$ to $[-1, 1]^N$, while the FFN for the instruments maps from $\mathbb{R}^{\dim(h^g)}$ to $[-1, 1]^{N \times N_g}$. Further details concerning our choice of hyperparameters for the neural networks and the steps that we use for iteratively determining the θ_g and θ_w parameters in solving the minimax problem are provided in Appendix A.

Additionally, to help mitigate problems with over-fitting and improve the precision of our SDF estimate, we apply three commonly used "robustification" techniques. Firstly, we adopt a form of regularization termed "dropouts" when training the FFNs, in which we randomly drop a fraction of the units in the network. Following common practice in the ML literature, we fix this dropout fraction to 5% (see, e.g., Srivastava et al. (2014)). Secondly, we use ensemble averaging, determining our final SDF estimate as the average of ten separate SDF estimates based on different initial seeds. This procedure, which again is commonly used in the ML literature, naturally reduces the estimation error along with any dependence on the initial seeds.[6] Thirdly, we deliberately bound the Sharpe ratios for each of the individual SDF estimates to lie between 0.4 and 1.5. These particular bounds are directly motivated by Kozak, Nagel, and Santosh (2020) and the empirical analyses therein demonstrating that the imposition of similar priors on the Sharpe ratio results in SDF estimates with improved out-of-sample cross-sectional explanatory power. Each of our individually estimated SDFs also easily converges to tangency portfolios with Sharpe ratios within these bounds.[7]

The previous discussion still leaves the returns R^e_{t+1} and the conditioning information I_t used in the definition of the $f_g(I_t;\theta_g)$ and $f_w(I_t;\theta_w)$ functions to be determined. The next section discusses our specific choices for these.

2.3 Return Data and Conditioning Information

The set of risky assets underlying our estimation consists of the 272 high-frequency portfolios recently constructed by Aleti (2022). All of the portfolio returns are sampled at a 15-minute frequency and cover the sample period from January 2, 1996 to December 31, 2020, for a total of 169,965 intraday 15-minute return observations.[8] We will refer to these portfolios as $\mathbf{Z}_t$ in the sequel. The 272 portfolios are comprised of 218 factor portfolios following Chen and Zimmermann (2021) and Jensen, Kelly, and Pedersen (2022), 48 industry portfolios, as well as the commonly used six Fama-French portfolios (FF6). The long-short, or net-zero investment, portfolios are all directly compatible with the excess return format in the basic no-arbitrage condition in equation (1). Correspondingly, for the market portfolio and the 48 industry portfolios, we subtract the risk-free rate to obtain the relevant excess returns.[9] Taken as a whole, these portfolios serve as a powerful set of span and test assets. The 218 high-frequency factor portfolios, in particular, effectively capture the many "anomalies" highlighted in the asset pricing literature, while the industry portfolios account for well-documented industry-specific effects. The inclusion of the FF6 portfolios further ensures that the estimated SDF will be able to price the Fama-French workhorse factors.

Our use of portfolio returns to span the high-frequency SDF contrasts with CPZ and several other recent studies that rely on individual stock returns for spanning the SDF at lower daily or monthly frequencies.[10] Our motivation for doing so is threefold. Firstly, portfolios are always well-defined, whereas many individual stocks are not available over

[6] The empirical results reported in CPZ further suggests that when used in the present context, it also tends to improve on the cross-sectional explanatory power of the estimated SDF.

[7] Note, that as a formality, by reducing the estimation error, and in turn the in-sample return volatility, the in-sample Sharpe ratio for our ensemble average SDF estimate may actually exceed the upper bound.

[8] As discussed in more detail in Aleti (2022), the use of a "coarse" 15-minute sampling frequency effectively mitigates the impact of market microstructure "noise." As an aside, in the present context it also facilitates the practical estimation compared to the use of finer, say 5-minute returns, which would be prohibitively more expensive from a computational perspective.

[9] We proxy the risk-free rate by the daily returns for the one-month Treasury Bill rate from Ken French's website, "distributing" the returns equally across each of the within-day 15-minute intervals. This theoretically motivated excess return adjustment, of course, has virtually no effect on any of our estimates.

[10] Other recent lower-frequency SDF estimation procedures that explicitly rely on large cross-sections of individual stocks include the so-called "agnostic" approach of Pukthuanthong, Roll, and Wang (2020) and Kim and Korajczyk (2021), and the Bayesian approach of Kozak, Nagel, and Santosh (2020). Bryzgalova, Pelger, and Zhu (2020) similarly rely on large numbers of individual stocks in their construction of managed portfolios to span the SDF.

the full sample period. This in turn provides us with a much more manageable balanced panel of assets. Secondly, portfolios primarily capture systematic risk, while individual stocks tend to be heavily influenced by idiosyncratic risk, especially so at higher frequencies, rendering the estimation much more difficult. Lastly, and most importantly, by "averaging out" stock-specific effects, portfolio returns are generally much less susceptible to market microstructure noise than individual stock returns, thereby allowing for the more meaningful use of higher-frequency data.

Turning to the conditioning set that we use for incorporating important economic information, we again closely follow CPZ and rely on data drawn from three different sources. The first dataset, *FRED-MD*, consists of 126 monthly macroeconomic variables, as further discussed in McCracken and Ng (2016). The second dataset consists of the cross-sectional medians of the 153 firm characteristics recomputed on a monthly basis using the characteristic data from Jensen, Kelly, and Pedersen (2022).[11] The third, and last, dataset consists of the eight popular monthly equity risk premium predictor variables highlighted in the oft-cited study by Welch and Goyal (2008). We further transform each of these individual time series as necessary to render them stationarity. We also lag all of the series by one month to ensure that the combined I_t information set was actually available at time t. Additional details concerning these transformations and the interpolations that we use in the construction of the combined high-frequency dataset are provided in the Online Supplemental Appendix.

2.4 Tangency Portfolio Estimates

The key expression in equation (5) formally defines the tangency portfolio as a weighted combination of the span assets. Hence, armed with the estimated weight function and the high-frequency portfolio returns, our high-frequency tangency portfolio return estimates may simply be expressed as:

$$\hat{F}_{t+1} = 1 - \hat{f}_w(I_t\,;\,\hat{\theta}_w)\mathbf{Z}_{t+1}. \tag{11}$$

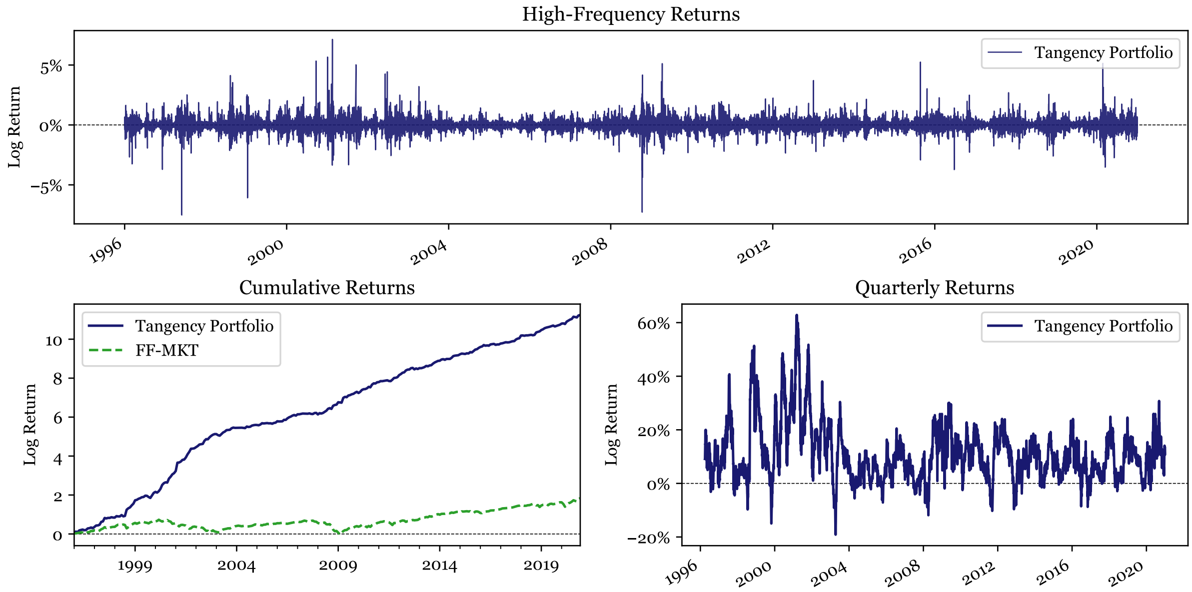
Figure 1 displays the resulting SDF returns over different frequencies, where for ease of comparison we have rescaled the returns to have the same full-sample realized volatility as the Fama-French market portfolio.

The general features of the high-frequency SDF returns depicted in the top panel directly mirror those of most other high-frequency return series. Meanwhile, looking at the cumulative returns displayed in the lower-left panel, the (rescaled) SDF seemingly performed very well as an investment portfolio, easily beating the Fama-French market portfolio over the full sample.[12] The lower-right subplot further shows that the SDF tangency portfolio only rarely delivered a negative quarterly return. Of course, any practical implementation of this SDF "paper portfolio," or any related high-frequency SDF-based trading strategy, would invariably face much higher transaction costs than simply buying and holding the market portfolio, and as such, it would also not necessarily

[11] Unlike CPZ, who rely on their own choice of 46 firm characteristics, we rely on the 153 characteristics produced by Jensen, Kelly, and Pedersen (2022). This choice is primarily motivated by data availability for our sample period.

[12] The high-frequency SDF also achieves an impressive cross-sectional R^2 for explaining the variation in the 272 high-frequency portfolios equal to 80.7%, compared to an R^2 of just 23.1% for the high-frequency Fama-French market portfolio, with the R^2's computed as one minus the full-sample squared alphas $\frac{1}{T}\sum_t Z_{t+1} - Z_{t+1}\hat{F}_{t+1}$ divided by the full-sample squared returns.

Figure 1: Tangency Portfolio Returns



Note: The figure plots the estimated tangency portfolio returns rescaled to have the same realized volatility as the Fama-French market portfolio. The upper subplot displays the 15-minute intraday and overnight returns. The lower-left subplot shows the cumulative returns on the tangency portfolio together with the cumulative return on the Fama-French market portfolio. The lower-right subplot shows the quarterly tangency portfolio returns computing over rolling 90-day windows.

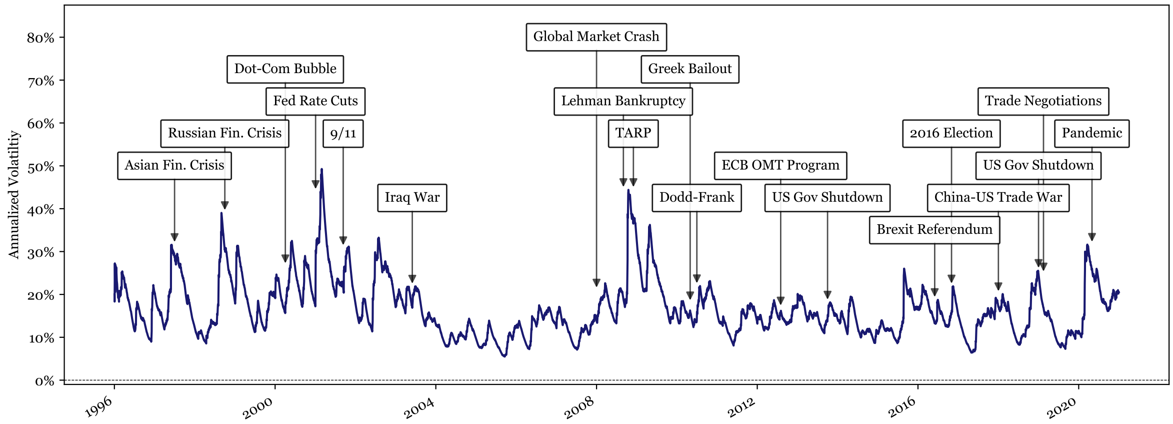
outperform a simple buy-and-hold strategy in practice.[13] However, that is also not our focus here. Instead, we seek to identify the news that drives the *true* SDF, and large changes therein in particular.

The large differences in the cumulative returns aside, the high-frequency tangency portfolio returns are, not surprisingly, on average positively correlated with the high-frequency returns on the aggregate market portfolio, with the full-sample correlation equal to 12.6%. Meanwhile, as evidenced by the additional results reported in the Online Supplemental Appendix there is also a fair amount of variation in that correlation over time. That same appendix also reports the average annual correlations between the SDF and the FF6 portfolios, 13 representative factor cluster portfolios, as well as the 48 industry portfolios that we use in spanning the SDF. As these additional results show, the correlations with the representative cluster portfolios tend to be fairly heterogeneous and also appear to vary non-trivially over the sample period. By contrast, the correlations across the different industry portfolios seem far more homogeneous, and the temporal variations therein also fairly closely match those observed for the market portfolio. We will return to this theme in our discussion of the different types of news that drive the returns on the various factors in the zoo.

Further illustrating the key features of the estimated SDF, Figure 2 reports the annualized realized volatilities computed from the summation of the 15-minute intraday and overnight squared returns. To ease interpretation the figure shows the backward-looking exponentially weighted moving average based on a half-life of 30 days. In line with the

[13] Putting this further into perspective, even though the neural network-based approximations for the weight and instrument functions that we rely on are both formally predictive, following standard procedures our estimation is still based on the full-sample data, and as such the resulting SDF effectively constitutes an "in-sample estimate." It is unlikely that the same seemingly superior investment performance could be achieved with an SDF based on a true out-of-sample estimation scheme.

Figure 2: Tangency Portfolio Volatility



Note: The figure plots the smoothed annualized realized volatility of the estimated tangency portfolio. The portfolio is rescaled to have the same unconditional realized volatility as the Fama-French market portfolio.

extensive literature on time-varying financial market volatility (see, e.g., Bollerslev et al. (2018), and the many references therein), the volatility of the SDF clearly varies over time in a strongly positively autocorrelated fashion. Regressing the daily realized volatility of the SDF on the lagged daily, weekly, and monthly realized SDF volatilities, as in the popular HAR model of Corsi (2009), also results in a fairly high R^2 of 28.9%, mirroring the degree of return volatility predictability typically observed for the aggregate market portfolio. Moreover, consistent with the idea of a traditional risk-return trade-off for the aggregate market portfolio (see e.g., the discussion in French, Schwert, and Stambaugh (1987)), regressing the daily SDF returns on the lagged daily realized SDF variance results in an R^2 of 5.1%, while that same regression produces an R^2 of 4.2% at the monthly horizon.

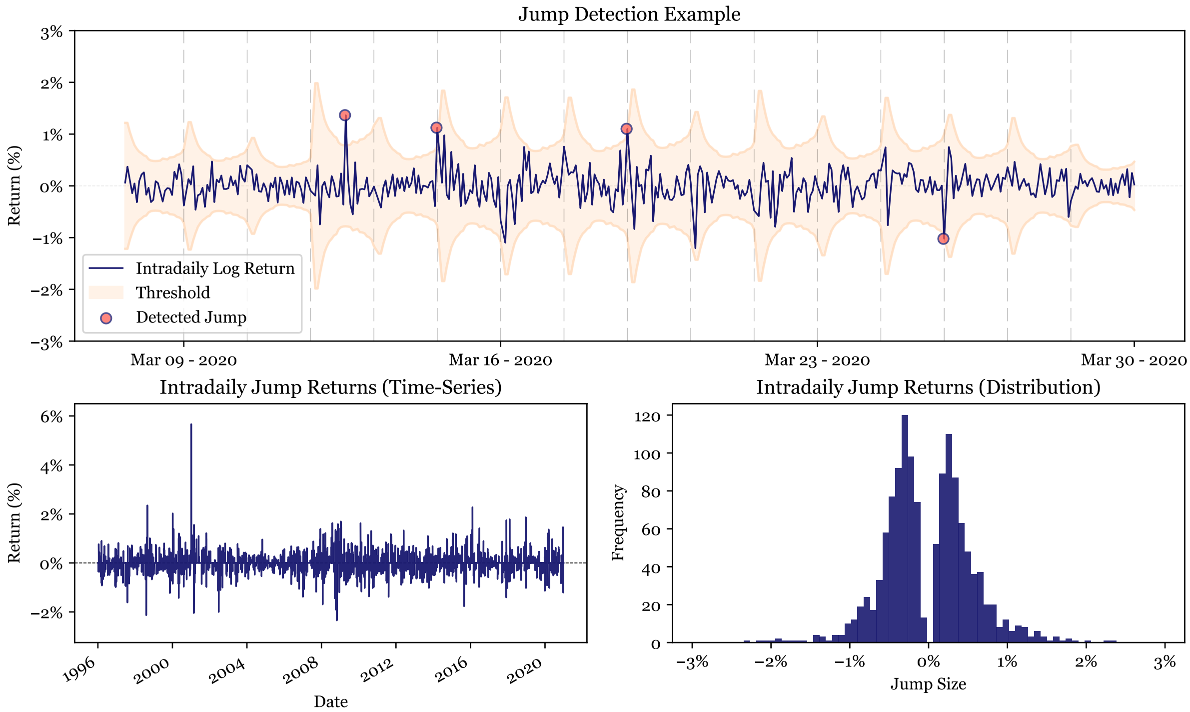
As indicated by the specific events annotated in Figure 2, there is also a clear tendency for periods associated with high economic uncertainty and/or financial crises to be accompanied by relatively high SDF volatility. Corroborating this visual impression, the correlations between the monthly realized SDF volatility and the monthly *Financial*, *Macro*, and *Real Uncertainty* indices from Jurado, Ludvigson, and Ng (2015) equal 59.2%, 37.7%, and 27.2%, respectively. This again directly mirrors existing evidence for the market portfolio, and the tendency for realized market volatility to increase during periods of market "stress" (see, e.g., Banulescu et al. (2016)). This connection has also previously been used by Manela and Moreira (2017) and Baker et al. (2019) in the construction of news-based aggregate market volatility indexes.

To help more clearly delineate these linkages and the economic news that is actually priced, following the discussion above it is instructive to further decompose the high-frequency SDF into separate continuous and more abrupt jump components.

2.5 Tangency Portfolio Jumps

Interpreting the estimated 15-minute SDF returns $\hat{F}_t$ defined in (11) as the discrete-time realization of some true underlying continuous-time SDF process, we rely on techniques from high-frequency financial econometrics to separate the continuous and discontinuous moves in the SDF. Intuitively, if the increment in $\hat{F}_t$ over a given 15-minute interval is

Figure 3: Jump Identification Example and Jump Returns



Note: The top subplot shows the intradaily returns on the estimated tangency portfolio for the last three weeks of March 2020, together with the specific jump threshold that we rely on and the the jump returns that exceed the threshold marked in red. The full-sample jump returns for the estimated tangency portfolio rescaled to have the same realized volatility as the Fama-French market portfolio are plotted as a time-series in the lower-left subplot and as a histogram in the lower-right subplot.

"too large" (in an absolute value sense) to be explained by the realization from a normal distribution with a local variance commensurate with the "normal" variation over a 15-minute interval, we classify the increment as a "jump." A more formal discussion of this thresholding procedure, including the estimation of the local variance and our specific choice of threshold, is given in Appendix B.

To illustrate the basic idea, the top panel in Figure 3 plots the intraday high-frequency returns on the estimated tangency portfolio for the last three weeks of March 2020, together with our corresponding time-varying jump thresholds. This particular time period obviously coincides with the global onset of the COVID-19 pandemic, and as indicated by the red dots in the figure, our procedure, not surprisingly, identifies several SDF jumps during this three-week period. As discussed further below, all of these jumps may also naturally be linked to specific news about the severity of the pandemic and/or statements and actions by the Federal Reserve and other policymakers intended to help mitigate its economic impact. Figure B.1 in Appendix B provides additional illustrative examples of SDF jumps during other time periods that may similarly be linked to specific economic news events.

Using this same thresholding technique, the bottom two panels in Figure 3 show the time-series and size distribution of the identified intraday jumps in the high-frequency SDF. On average, there are 52 intraday jumps per year, with a maximum of 75 jumps in 2010 and a minimum of 41 jumps in 2003. The second panel helps further visualize the jump magnitudes, showing that the jumps tend to be fat-tailed.[14] In contrast to the jump returns, the continuous returns are far smaller and are generally very difficult to associate with observable economic information, or specific news. Instead, we deliberately exploit the more abrupt changes in the SDF, as represented by the jump returns, to more clearly delineate the news that is actually priced by investors.

Correspondingly, as discussed further below, our programmatic approach for linking the intraday jumps in the SDF with the news depends critically on there being a limited number of news articles published in the temporally-adjacent 15-minute time intervals. With hundreds of news articles typically being published in the 17.5-hour time interval between the close of the market on one day and the following day's opening that same approach simply wouldn't be feasible, let alone reasonable, for "explaining" the overnight SDF returns. Accordingly, we ignore the overnight portion of the returns in all of our news-based attributions. Put differently, we implicitly assume that the most important economic news for the pricing of US assets occurs during regular US market hours. This, of course, does not rule out the pricing of international news per se, only news that is always published outside regular US market hours. Indeed, anticipating our results, news concerning *International Affairs*, as defined in the next section, emerges as our overall second most important news topic.

We turn next to a more detailed description of the news data and the approach for linking the SDF jumps with the different news topics that we rely on.

3 Linking Systematic Jumps with News

Our approach for linking the jumps in the tangency portfolio with news is based on the counts of keywords in several million precisely timed newswire articles. We begin by briefly detailing the news data, followed by a discussion of the way in which we group key news article terms into prespecified *news topics*. Aggregating the mentions of each news topic over 15-minute intervals, we in turn "explain" each of the SDF jumps with the dominant news topic over the relevant time interval.

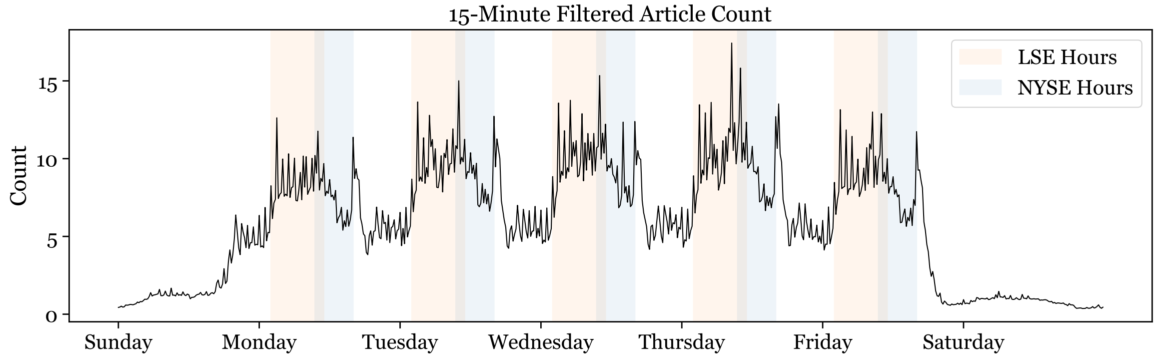
3.1 News Data

Our primary dataset consists of the *Dow Jones Newswires Archive*, a machine-readable collection of articles from the Dow Jones' real-time news feeds. Retrieving all articles from January 2, 1996, to December 31, 2020, leaves us with a total of 50,734,964 articles. Each of these articles consists of a headline, a body text, a subject/product/company code, along with additional identifiers. We deliberately exclude articles that are seemingly irrelevant to investment decisions, such as those about sports, entertainment, and lifestyle. We also deem articles that simply state open/close prices and various technical indicators as being irrelevant for capturing innovations in the state variables that drive the SDF. Similarly, we remove articles about company-specific news, which is arguably idiosyncratic and hence should not affect the SDF. A more in-depth discussion of all the exclusion-filters that we apply is provided in the Online Supplemental Appendix.

All in all, after applying the above filters, we are left with a total of 5,107,352 relevant articles over the full sample period. As the resulting article counts depicted in Figure 4

[14] The gap in the center of the distribution is simply an artifact of the thresholding technique which inevitably can not identify "small" jumps.

Figure 4: Article Counts



Note: The figure shows the average number of news articles in our filtered data set for each 15-minute interval and day of the week. The shaded areas indicate London Stock Exchange and New York Stock Exchange market hours.

show, the news articles tend to be posted on weekdays at the start of European market hours until slightly after American markets close. This is further underscored by the additional summary statistics and plots provided in Appendix C. Even though there has been a decline in the average number of "relevant" articles being posted per month from the beginning to the end of the sample (in part due to changes and updates to the Dow Jones dataset itself), there are typically still several hundred investment related articles being posted on a daily basis over the entire sample.

3.2 Extracting News Topics

To determine which topics are prevalent in the news, we begin by assigning topic counts for each article in our filtered dataset of relevant articles. These topic counts are computed as the number of key terms associated with each topic in a given article. For example, the topic *Monetary Policy* and its associated key terms naturally include *federal reserve*, *money supply*, *open market operations*, *fed funds rate*, among others. Counting the mentions of these key terms provides us with a direct measure of the prevalence of *Monetary Policy* as a topic in a given news article. Repeating this procedure across all articles in turn reveals what topic dominates the news at any given point in time. A similar automatic approach of aggregating news topic counts has also recently been employed by Bybee, Kelly, and Su (2022), albeit over much coarser daily time intervals.

Our topic counts are primarily based on the topics and key terms previously defined by Baker et al. (2019). These topics and terms were all "hand-selected" with the explicit purpose of studying stock market volatility.[15] They extend the news topics and associated terms previously used by Baker, Bloom, and Davis (2016) and Davis (2017) for measuring economic policy uncertainty. However, the former topic list is still not entirely up-to-date, requiring additional modifications to accommodate more recent trends in the news, most notably the COVID-19 pandemic. Additionally, to ensure that our list of key terms is comprehensive, we further augment the lists with any missing key terms from the

[15] By contrast, Bybee et al. (2021) and Larsen and Thorsrud (2019) both rely on "automatic" Latent Dirichlet Allocation (LDA) techniques for defining key terms and topics. However, LDA and other unsupervised learning procedures often produce word clusters that are difficult to interpret and/or clusters of words that seemingly have little to do with news about the state of the economy, invariably requiring some additional fine-tuning "by hand."

aforementioned Bybee, Kelly, and Su (2022) study. Altogether, this leaves us with a set of 44 distinct news topics with an average of 22 key terms each. Appendix D provides a comprehensive list of the various topics and associated key terms.

Having defined the topics and corresponding key terms, we calculate the topic counts for each article. To do so, we begin by combining the headline and body text into a single block of text. Next, we preprocess this text data using standard transformations from the Natural Language Processing (NLP) literature, consisting of the following steps: (i) convert all text to lowercase letters, (ii) remove any stop words,[16] (iii) delete multiple spaces and line breaks, (iv) remove non-alphanumeric characters, and (v) stem and lemmatize each word.[17] We then tokenize the text and extract n-grams, or groups of n adjacent words.[18] Finally, we count the number of n-grams that appear in each topic's key term list for each article, producing the requisite topic counts. This now fairly standard type of approach for automatic text-processing also underlies the related work by Ke, Kelly, and Xiu (2021) and Bybee et al. (2021).

Although uniquely identified by the key terms, many of the news topics determined by the above-defined counts are inherently related. Hence, to provide a more broad-based view on the type of news that matters, we further combine the 44 more detailed news topics into a smaller set of 8 "metatopics." The compositions of most of these metatopics are fairly self-explanatory. For instance, our *Monetary Policy and Finance* metatopic naturally combines the previously defined *Monetary Policy*, *Other Financial Indicators*, *Financial Regulation*, *Interest Rates*, and *Inflation* topics into a single topic. As another example, our *Commodities and Energy* metatopic simply combines the *Commodity Markets* and *Energy Markets* topics into a single metatopic. Of course, not all of the 44 more detailed news topics are as easily categorized and combined, and as such some of the original topics also appear in more than one of the 8 metatopics, while some do not appear in any metatopic.[19] Appendix D again provides an exact description of the relevant metatopic definitions.

3.3 Linking Topics with SDF Jumps

As discussed in Sections 2.3 and 2.5 above, to help alleviate concerns about the impact of market microstructure noise, we purposely rely on a "coarse" 15-minute sampling frequency for the estimation of the jumps in the SDF.[20] Accordingly, in order to link the estimated SDF jumps with the news, we begin by aggregating the topic counts for all of the relevant news articles across the same 15-minute time intervals used in identifying the jumps. Although the SDF in theory should respond immediately upon the release of new economic information that investors care about, this 15-minute temporal aggregation of the topic counts simultaneously serves to allow for more gradual incorporation of the

[16] Some examples of stop words are: *the*, *is*, and *are*. We obtain our list of stop words from the *Natural Language Toolkit* (NLTK), a Python library developed by Bird, Klein, and Loper (2009).

[17] This step entails converting the words such as *taxes* and *taxation* to their root word *tax*.

[18] As an example, the text *conduct monetary policy* would be tokenized into {*conduct*, *monetary*, *policy*}. The set of unigrams is simply the set itself. The set of bigrams is {*conduct monetary*, *monetary policy*}. The set of trigrams is {*conduct monetary policy*}. Since our list of key terms consists of unigrams, bigrams, and trigrams, these are also the only n-grams that we extract.

[19] More precisely, 7 of the 37 news topics that define our 8 metatopics are repeated twice, while 11 of the 44 original news topics are not included in any metatopic.

[20] The Online Supplemental Appendix provides additional robustness checks pertaining to this choice of sampling frequency.

news, consistent with the idea that it might take market participants some time, however brief, to fully digest and interpret the news, in turn resulting in "gradual" jumps as first hypothesized by Barndorff-Nielsen et al. (2009) (see also the recent discussion in Bollerlsev (2022)). Conversely, there might also be a short time delay between the actual occurrence of certain news and the write-up of said news first appearing in the form of a news article on the *Dow Jones Newswires*, thus necessitating some "slack" in the jump-news attribution.

To allow for slow-moving variation, or trends, in the importance of different types of news over the full 1996-2020 sample period, we further demean the aggregated 15-minute topic counts based on a backward-looking 30-day moving average. We then sort these demeaned aggregated topic counts to determine the dominant, or primary, topic for each of the 15-minute intervals, in turn associating each of the jumps in the tangency portfolio with the primary news topic for the particular time interval containing the jump. For example, as further detailed in Appendix E, we "explain" the jump detected over the 13:15-13:30 time interval on January 3, 2001 by the topic *Monetary Policy*, which dominated the news over that particular time interval and day. If a 15-minute jump time interval does not have a dominant topic, we simply associate that jump with the news topic *None*.[21]

The following section provides a summary of the resulting news-jump linkages for *all* of the 1,302 intradaily SDF jumps that occurred over the full sample period.

4 What Drives Systematic Risk?

We begin our analysis by assessing which of the news topics account for most of the jumps in the SDF, followed by an assessment of how much each topic contributes to the overall jump variation. We also present additional results based on aggregating the more detailed news topics into our more broadly defined, and easier-to-interpret, metatopics.

4.1 Why Does the SDF Jump?

To help assess the relative importance of the different news topics for explaining the jumps in the SDF, it is instructive to first summarize the unconditional and conditional topic frequencies. Specifically, we define the Topic Unconditional Frequency (TUF) of a given topic as the frequency by which it appears as the primary topic across all the 15-minute intervals in the sample. Formally,

$$\text{TUF}(k) \equiv \sum_{t,i}\Bigg(\underbrace{\mathbb{1}^{news}_{k,t,i}}_{\text{Topic } k \text{ in News}}\Bigg) \Big/ \Bigg(\sum_{t,i} 1\Bigg), \tag{12}$$

where $\mathbb{1}^{news}_{k,t,i}$ is an indicator variable for topic k being the primary news topic for the ith intraday time interval on day t. We similarly define and calculate the Topic Conditional Frequency (TCF) as the same average over the jump intervals only,

$$\text{TCF}(k) \equiv \sum_{t,i}\Bigg(\underbrace{\mathbb{1}^{news}_{k,t,i}}_{\text{Topic } k \text{ in News}} \times \underbrace{\mathbb{1}_{|F_{t,i}|\geq\alpha\sqrt{\tau_i BV_t}\,\Delta_n^{\varpi}}}_{\text{Jump in Tangency Portfolio}}\Bigg) \Big/ \Bigg(\sum_{t,i} \mathbb{1}_{|F_{t,i}|\geq\alpha\sqrt{\tau_i BV_t}\,\Delta_n^{\varpi}}\Bigg). \tag{13}$$

[21] This pertains almost exclusively to time-intervals with no news articles, as the vast majority of news articles contain at least one key term associated with one of our news topics.

Table 1: News and Jumps

Topic	Topic Frequency (%) Unconditional	Conditional	Difference	t(Difference)
Monetary Policy	8.96%	20.20%	11.24%	10.11
US Politics	7.23%	12.06%	4.82%	5.35
Energy Markets	6.52%	8.83%	2.31%	2.95
Middle East	5.56%	7.68%	2.12%	2.88
Labor Markets	2.82%	4.76%	1.94%	3.29
Russia	3.54%	4.53%	0.99%	1.73
Taxes	3.12%	3.30%	0.19%	0.38
National Security	3.32%	3.23%	−0.09%	−0.18
Financial Regulation	1.35%	3.23%	1.88%	3.84
China	7.58%	3.07%	−4.51%	−9.42
Inflation	2.08%	3.07%	1.00%	2.09
Commodity Markets	2.43%	2.46%	0.03%	0.07
Real Estate Markets	1.05%	2.30%	1.26%	3.02
Trade	2.16%	2.07%	−0.09%	−0.22
Elections and Political Governance	1.34%	1.77%	0.43%	1.18
Natural disasters	2.54%	1.61%	−0.93%	−2.67
North Korea	3.87%	1.61%	−2.25%	−6.45
Broad Quantity Indicators	1.28%	1.46%	0.18%	0.53
None	18.66%	1.46%	−17.20%	−51.43
Disease	1.39%	1.38%	−0.01%	−0.02

Note: The table reports the unconditional and conditional frequency counts for the different news topics, given by the $TUF(k)$ and $TCF(k)$ statistics formally defined in the main text. The last two columns report the differences in the frequencies along with the t-statistics for testing whether the differences are statistically significant. The table only reports the top 20 news topics sorted by their conditional frequencies.

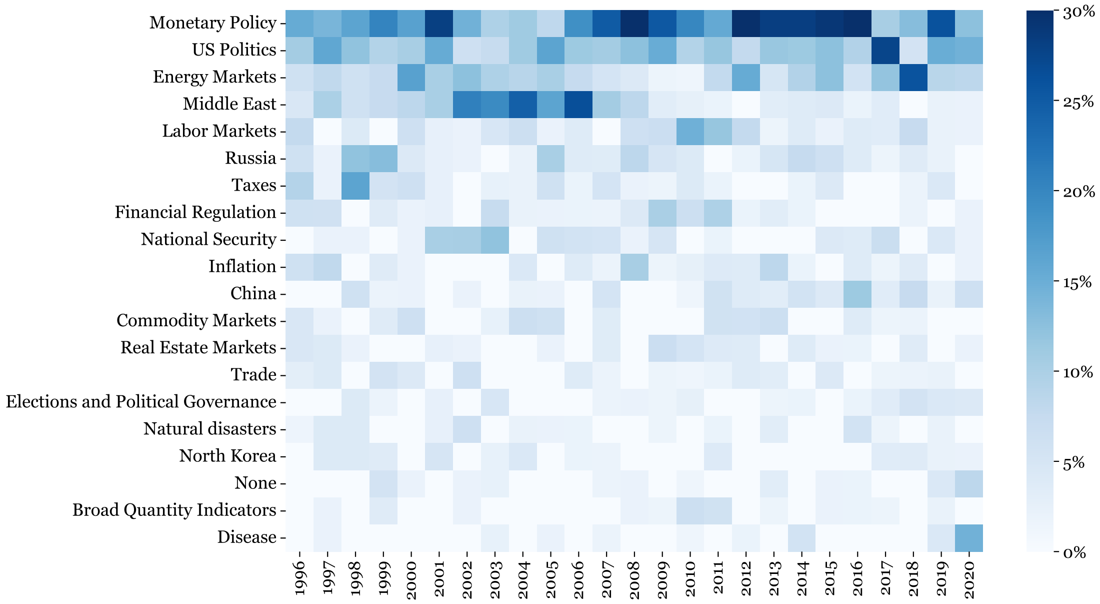
This latter statistic, of course, is simply the probability that topic k dominates the news over an interval conditional on a jump being detected. If jumps are statistically independent of the news, the $TUF(k)$ and $TCF(k)$ frequencies for any given topic k should be the same. Hence, by simply comparing the differences in the two frequencies, we obtain a simple assessment as to which of the different news topics primarily appear to be associated with systematic jumps.

Table 1 reports the resulting frequencies, along with the t-statistics for testing whether the differences in the $TUF(k)$ and $TCF(k)$ frequencies are statistically significant.[22] For brevity, we only include the top twenty topics based on the conditional frequency counts.[23] As the table shows, most of the differences are strongly statistically significant. The largest difference manifests for *Monetary Policy*, which appears with a frequency of 8.96% unconditionally compared to a conditional frequency of 20.20%. This finding is directly in line with Baker et al. (2021), who report that news associated with monetary policy seemingly triggered many of the largest (in an absolute value sense) daily stock market returns over the past half-century. It also corroborates the extensive historical

[22] The t-statistics are conveniently calculated from regressions of the form $\mathbb{1}^{news}_{k,t,i} = \beta_0 + \beta_1 \cdot \mathbb{1}_{|F_{t,i}| \geq \alpha\sqrt{\tau_i BV_t}\Delta_n^{\varpi}}$, and the significance of the β_1 coefficient associated with the indicator for the jumps.

[23] As discussed in Appendix B, these results rely on a jump threshold parameter of $\alpha = 3.0$. Additional results for other jump thresholds are reported in the Online Supplemental Appendix. The conditional frequencies for values of α in excess of 3.0, and the corresponding t-statistics for testing the differences in $TUF(k)$ and $TCF(k)$, are generally very similar to the results based on $\alpha = 3.0$ reported in Table 1.

Figure 5: Conditional Topic Frequencies



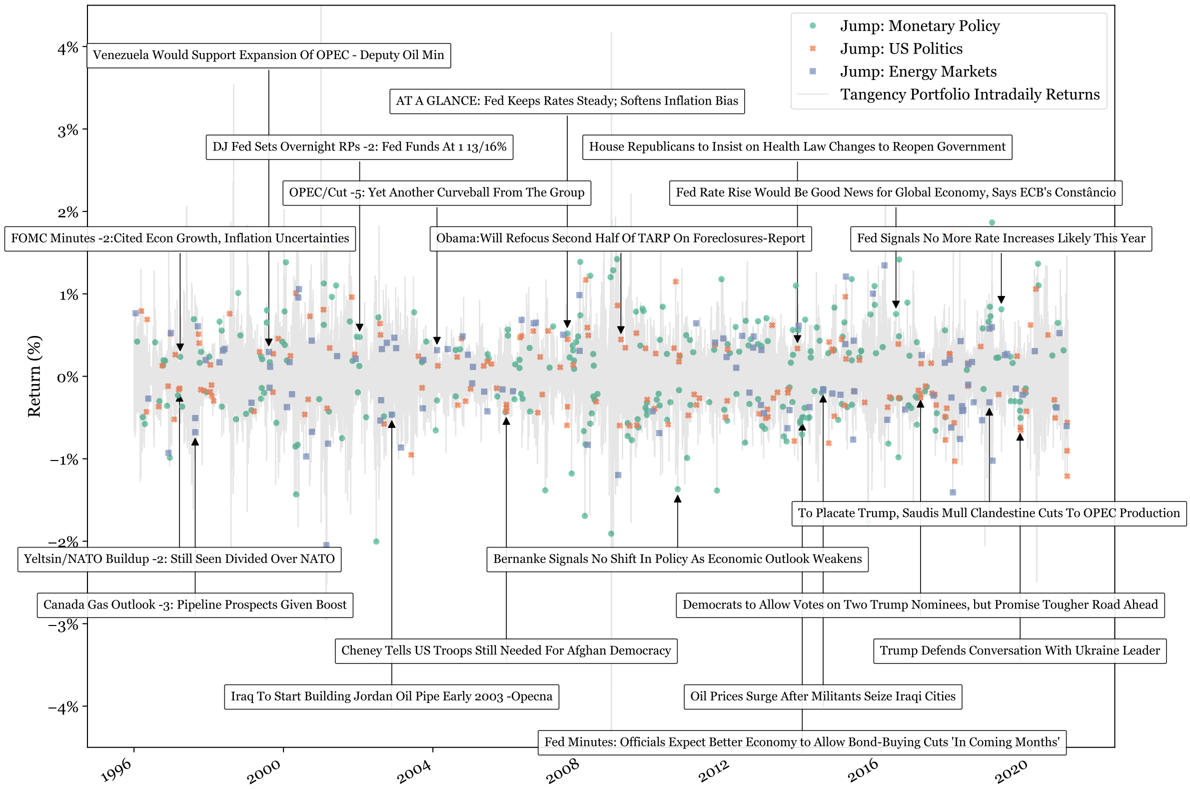
Note: The heatmap shows the fraction of jumps associated with each topic across all detected jumps for each year in the 1996-2020 sample. The display is limited to the top 20 topics by frequency over the full sample.

analysis of intraday stock market jumps in Johnson, Medeiros, and Paye (2022), which suggests that monetary policy news has become increasingly important in recent decades. Of course, the importance of Fed policy for understanding asset markets has also long been emphasized in the macroeconomics literature (see, e.g., the early studies by Kuttner (2001), Rigobon and Sack (2004) and Bernanke and Kuttner (2005), along with the more recent review by Rogers, Scotti, and Wright (2014)).

Although *Monetary Policy* stands out as the overall most important news topic, *US Politics*, *Energy Markets*, *Middle East*, and *Labor Markets*, also all exhibit highly statistically significant, although smaller, differences in their $TUF(k)$ and $TCF(k)$ frequencies. Interestingly, the topic *None*, which again refers to intervals without a primary topic, stands out as by far the most frequent topic unconditionally, being associated with 18.66% of all 15-minute intervals in the sample. However, conditional on a jump being detected, only 1.46% of the intervals lack a primary topic. In other words, the SDF almost never jumps without identifiable economic news. This finding of a news-based explanation for most of the jumps in the SDF also accords with previous studies that have successfully linked various high-frequency jumps in aggregate equity index and individual stock returns to precisely-timed market-wide and company-specific news releases (see, e.g., Lee and Mykland (2008), Lee (2012), and Jeon, McCurdy, and Zhao (2022), and the many other references therein). In contrast to most previous studies, however, which have sought to link jumps with pre-scheduled news announcements, as discussed further below, we also find a very important role for unscheduled news.

Going one step further, Figure 5 displays the $TCF(k)$ conditional topic frequencies computed on a disaggregated annual basis. Not surprisingly, *Monetary Policy* consistently ranks as one of the most frequent news topics for explaining the occurrence of jumps throughout the sample. The frequency of the *Monetary Policy* topic appears particularly high during the 2008-2009 Great Recession, and from 2012 to 2016 and the time

Figure 6: Select Topic Headlines



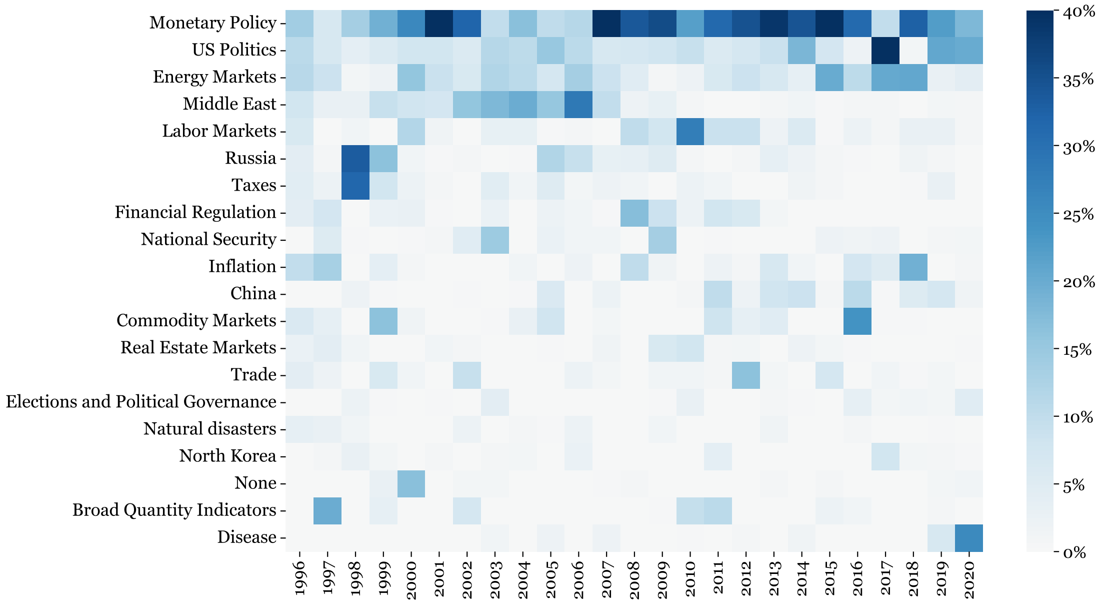
Note: The figure displays the intradaily 15-minute returns for the estimated tangency portfolio in gray, together with all the jumps associated with *Monetary Policy*, *US Politics*, and *Energy Markets*. The select headlines are drawn from articles that were published in the same time interval as the jumps and tagged as one of the three news topics. Not all jumps are annotated in order to prevent overlapping headlines.

of the European Sovereign Debt Crisis, the Taper Tantrum, and the latter rounds of Quantitative Easing. Meanwhile, the topic *US Politics* tends to play a relatively more important role in explaining the jumps in the years following US presidential elections, and especially so during the first year of the Trump presidency. The frequency of SDF jumps related to news about *Energy Markets* seemingly peaked in 2018, coincident with the steep decline in oil prices and a series of OPEC announcements, while *Middle East* and *National Security* related jumps both appeared relatively more frequent after 9/11 and near the beginning of the Iraq War in 2003. The relative frequency of jumps attributed to news about *Russia*, unsurprisingly, peaked around the time of the Russian Financial Crisis in 1998. Lastly, *Disease* naturally stands out as the overall most frequent news topic for explaining the jumps in the SDF at the height of the COVID-19 pandemic in 2020.

To more concretely convey the specific news stories that actually matter, we perform a more granular analysis by extracting the news headlines for a sample of the SDF jumps associated with our top three news topics: *Monetary Policy*, *US Politics*, and *Energy Markets*. Figure 6 displays all of the relevant detected jumps together with a select set of news headlines.[24] Mirroring the analogous display pertaining to the largest (in an absolute value sense) daily aggregate market returns in Bybee, Kelly, and Su (2022),

[24] Similar headline displays for each of the three separate news topics are given in Figures F.1-F.3 in Appendix F.

Figure 7: Topic Variance Contributions



Note: The heatmap shows the fraction of jump variation associated with each topic relative to the total jump variation for each year in the 1996-2020 sample. The display is limited to the top 20 topics by frequency over the full sample.

the economic news stories underlying the SDF jumps, although topically related, are seemingly also quite diverse. For instance, for *Monetary Policy* the underlying causes of the jumps include unscheduled announcements, official statements from Fed officials, ECB news, and various other Fed-related news. For *US Politics*, the headlines tend to be about Congress and elections. For *Energy Markets*, the news often has to do with OPEC announcements, and to a lesser extent, oil-related conflicts in the Middle East and statements by the International Energy Agency. Our text-based analysis of the news data succinctly categorizes all of these disperse news stories into a set of well-defined news topics.

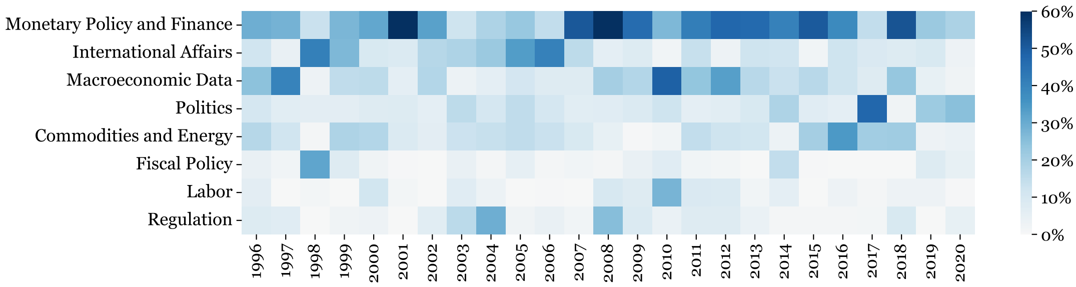
4.2 SDF Jump Variance Decompositions

In addition to the frequency counts discussed in the previous section, it is informative to also characterize how much each of the different news topics contributes to the total SDF jump variation.[25] To do so, we simply divide the sum of the squared jumps associated with a particular news topic by the total sum of all the squared jumps. Since each of the jumps is associated with a single unique topic, or the pseudo-topic *None*, these variance contributions naturally sum to one hundred percent. Seeing that the risk-return relationship directly links the variation in the tangency portfolio to its expected return, these calculations therefore also provide a simple first indicative answer as to which news topics account for most of the compensation for systematic jump risks.

The year-by-year results depicted in Figure 7, not surprisingly, fairly closely mirror the annual conditional jump frequencies in Figure 5. In particular, *Monetary Policy* again

[25] Relatedly, there is a long list of studies seeking to identify news and/or economic variables associated with the variation in aggregate stock market volatility at lower, typically monthly, frequencies; see e.g., the early oft-cited studies by Cutler, Poterba, and Summers (1989) and Schwert (1989).

Figure 8: Metatopic Variance Contributions



Note: The heatmap displays the fraction of jump variation associated with each metatopic relative to the total jump variation for each year in the 1996-2020 sample.

accounts for most of the jump variation throughout the sample, and especially so in 2001, 2007-2009, and 2012-2016. Likewise, the importance of *US Politics* again seems to peak right after the US elections and is again strikingly large in 2017. The general patterns observed for news related to *Energy Markets*, *Middle East*, and *Russia* are also all quite similar to the previously discussed patterns for the jump frequencies. Note, however, that this apparent close coherence with the results for the jump frequencies is not merely by construction, as the size of the jumps also figures importantly in the jump variance decompositions.

Further elaborating on this theme, Figure 8 shows the corresponding variance decompositions for our more general metatopics. The results reaffirm that most of the jump variation may be attributed to news about *Monetary Policy and Finance* broadly defined. In contrast to the previous results in Figure 7, however, the metatopic *International Affairs*, which is comprised of news about *China*, the *Middle East*, *North Korea*, and *Russia*, now stands out as the overall second most important type of news for explaining the jump variation, although arguably less so for the second half of the sample. By comparison, *Macroeconomic Data* on balance appears relatively more important during the second half of the sample, and especially so at the start of the European debt crisis in 2010, and more recently in 2018 during the height of the US-China trade war. Not surprisingly, the patterns and relative importance of the *Politics* and the *Commodities and Energy* metatopics fairly closely mirror the patterns previously seen in Figure 7 for the *US Politics* and *Energy Markets* topics. Foreshadowing our news risk premia estimates discussed next, the remaining three metatopics each account for relatively little of the SDF jump variation.

5 News Risk Premia

The news risk premia associated with a particular news topic is naturally defined as the return an investor would be willing to sacrifice to orthogonalize her/his portfolio with respect to the systematic variation stemming from news associated with that particular topic. Accordingly, we estimate the news risk premia based on a mimicking portfolio approach, in which we quantify the compensation for exposure to the systematic jumps linked with each of the news topics.

5.1 News Risk Premia Estimation

To begin, define the set of non-tradable *topic jump factors*,

$$F_t^{J,k} \equiv F_t^J \cdot \mathbb{1}_{k,t}^{news}, \tag{14}$$

where F_t^J denotes the time t tangency portfolio jump return, and $\mathbb{1}_{k,t}^{news}$ is an indicator function equal to one if topic k "explains" the time t jump. As such, if one were able to orthogonalize a portfolio with respect to $F_t^{J,k}$, one would effectively neutralize the portfolio's exposure to any systematic jumps associated with news topic k. Correspondingly, the risk premium associated with topic jump factor $F_t^{J,k}$, say $\lambda_t^{J,k}$, may be interpreted as the jump risk premium for news topic k. Of course, topic jump factors are non-tradable. We therefore rely on a standard mimicking portfolio approach to identify and estimate their risk premia.

In particular, it follows readily by standard asset pricing arguments that

$$\lambda_t^{J,k} = \beta_t^{J,k}\lambda_t^J, \tag{15}$$

where $\beta_t^{J,k}$ denotes the exposure of the topic jump factor to jumps in the tangency portfolio, and λ_t^J refers to the risk premium on tangency portfolio jumps. Our estimation of the jump betas $\beta_t^{J,k}$ is based on "jump regressions" of $F_t^{J,k}$ on F_t^J (see Li, Todorov, and Tauchen (2017) for a more formal discussion of jump regressions). Our approach for estimating the SDF jump risk premium λ_t^J essentially follows the continuous-time Fama-MacBeth approach recently developed by Aït-Sahalia, Jacod, and Xiu (2021) (see also Aleti (2022) for a more formal theoretical justification applicable in the present context).

Empirically, we rely on the same 272 high-frequency portfolios used for spanning the SDF as our "test assets." We re-estimate the continuous and jump betas for all of these test assets with respect to the SDF on a rolling monthly basis using one-month and one-year backward-looking windows, respectively.[26] Armed with the monthly beta estimates, we then estimate separate monthly continuous and jump risk premia for the SDF using what effectively amounts to a standard cross-sectional Fama-MacBeth regression approach. Finally, we obtain the requisite risk premium estimate for topic k by averaging $\hat{\beta}_t^{J,k}\hat{\lambda}_t^J$ across all of the months in the sample. A more extensive theoretical discussion of the approach and the practical implementation details are provided in the Online Supplemental Appendix.

The above estimation procedure relies critically on the jumps in the SDF for identifying the underlying news topics. The resulting news risk premia estimates are thus jump-risk specific, ignoring any "continuous" news-related risk compensation. To obtain additional estimates accounting for this, we simply assume that the continuous beta for topic factor k equals the jump beta for that same topic factor, or $\beta_t^{C,k} = \beta_t^{J,k}$. This assumption in turn allows for the estimation of a combined risk premium for news topic k based on the relation:

$$\beta_t^{C,k}\lambda_t^C + \beta_t^{J,k}\lambda_t^J = \beta_t^{J,k}(\lambda_t^C + \lambda_t^J) = \beta_t^{J,k}\lambda_t, \tag{16}$$

where λ_t denotes the total risk premium on the tangency portfolio. This risk premium may, of course, be estimated directly by averaging the SDF tangency portfolio returns

[26] Our use of a longer estimation window for the jump betas to account for the additional estimation error uncertainty mirrors Aït-Sahalia, Jacod, and Xiu (2021) and Aleti (2022).

Table 2: Topic Risk Premia

Topic	Jump		Total	
Monetary Policy	10.0%	(1.87)	30.4%	(10.91)
US Politics	4.3%	(2.60)	9.0%	(10.12)
Russia	3.7%	(2.37)	5.5%	(6.93)
Taxes	2.7%	(1.87)	4.4%	(5.96)
Labor Markets	2.5%	(1.88)	5.2%	(7.77)
Energy Markets	1.9%	(1.35)	8.1%	(10.41)
Inflation	1.9%	(1.95)	3.1%	(6.07)
Middle East	1.6%	(1.64)	5.8%	(8.79)
Financial Regulation	1.6%	(1.87)	2.9%	(5.87)
Commodity Markets	1.4%	(1.71)	2.7%	(5.99)
Disease	1.3%	(1.53)	0.9%	(2.44)
Trade	1.1%	(2.19)	2.1%	(6.24)
Broad Quantity Indicators	0.9%	(1.20)	1.6%	(4.51)
North Korea	0.8%	(3.32)	1.1%	(7.13)
Food and Drug Policy	0.8%	(1.93)	0.7%	(4.09)
China	0.8%	(1.47)	2.1%	(7.16)
Trade Policy	0.7%	(1.32)	0.8%	(3.31)
Elections and Political Governance	0.6%	(2.65)	0.9%	(7.19)
Entitlement and Welfare Programs	0.6%	(2.05)	0.2%	(2.18)
National Security	0.6%	(1.50)	1.7%	(6.35)

Note: The table reports the estimated jump and total risk premia for each topic factor, with the corresponding t-statistics in parentheses. The estimates are computed from the estimated tangency portfolio rescaled to a return of 100% per annum over the full sample.

over some non-trivial time interval. The resulting topic news risk premia estimates obtained by averaging $\hat{\beta}_t^{J,k}\hat{\lambda}_t$ across all of the months in the sample naturally exceed the previously defined estimates based on averaging $\hat{\beta}_t^{J,k}\hat{\lambda}_t^J$. To differentiate these more inclusive estimates from the earlier jump risk premia estimates, we will refer to the latter as the "total" news risk premia in the sequel.

To obtain a broader picture of the type of news that is priced, we also consider the risk premia associated with each of our eight metatopics. Our calculations of these more broadly defined premia rely on the exact same approach discussed above. Specifically, in parallel to equation (14), define the relevant set of *metatopic jump factors* as $F_t^{J,k'} \equiv F_t^J \cdot \mathbb{1}_{k',t}^{news}$, where the indicator variable $\mathbb{1}_{k',t}^{news}$ is simply defined as the sum over the indicator variables $\mathbb{1}_{k,t}^{news}$ associated with metatopic k'. We then use the same jump beta and risk premia estimation procedures as discussed above to quantify the corresponding metatopic risk premia.

5.2 News Risk Premia Estimates

We begin our discussion by considering the individual news topic risk premia. Table 2 reports the resulting estimates, together with the t-statistics in parentheses for testing whether the premia are statistically significantly different from zero. Since the topic risk factors are constructed from the jumps in the tangency portfolio, the magnitudes of the estimated premia are directly proportional to the average return on the tangency portfolio. Hence, to facilitate the interpretation of the results, we normalize the return on the tangency portfolio to 100% per annum, so that the numbers directly reveal the

Table 3: Metatopic Risk Premia

Metatopic	Jump		Total	
Monetary Policy and Finance	13.7%	(2.08)	37.3%	(11.05)
Macroeconomic Data	7.1%	(2.39)	15.5%	(9.56)
International Affairs	6.9%	(2.85)	14.5%	(10.84)
Politics	5.0%	(2.78)	9.9%	(10.35)
Fiscal Policy	3.4%	(2.22)	5.4%	(7.02)
Commodities and Energy	3.3%	(1.76)	10.8%	(10.35)
Labor	2.6%	(1.93)	5.3%	(7.86)
Regulation	1.9%	(1.43)	5.3%	(6.91)

Note: The table reports the estimated annualized jump and total risk premia for each of the metatopic factors, constructed by summing the topic factors for all the component topics within each metatopic. The estimates are computed from the tangency portfolio rescaled to a return of 100% per annum over the full sample. The t-statistics are reported in parentheses.

fraction associated with each of the different news topics. The table shows both the jump and the total news topic risk premia for the top twenty most important news topics sorted by their jump risk premia.

Consistent with our earlier findings pertaining to the news that causes the SDF to jump, *Monetary Policy* commands the largest (normalized) jump risk premium of 10.0%. In other words, 10.0% of the overall return earned on the jump tangency portfolio may be attributed to exposure to *Monetary Policy* related jump risk. Moreover, assuming that the continuous and jump returns are similarly exposed to the news, up to 30.4% of the return on the SDF may be traced to news about *Monetary Policy*. This echoes several other recent studies which suggest that much of the equity risk premium is earned around the time of FOMC announcements (e.g., Savor and Wilson (2013), Lucca and Moench (2015) and Cieslak, Morse, and Vissing-Jorgensen (2019)). By comparison, risks associated with news about *US Politics* and *Energy Markets* account for 9.0% and 8.1% of the overall return on the tangency portfolio, respectively.

The relative importance of the different topic risk premia also adheres fairly closely, although not perfectly, to the ordering of the jump frequencies and variance decompositions discussed earlier. This, of course, is not surprising as one would naturally expect that topics that account for most of the jumps and the jump variation in the SDF also demand the largest compensation. At the same time is noteworthy that even though the estimated risk premia for many of the less important topics are quite small, most of the estimates are still statistically significant at conventional levels.

Turning to the estimates for our more broadly defined metatopics reported in Table 3, the results imply that more than one-third of the return on the tangency portfolio comes from risk related to news about *Monetary Policy and Finance*.[27] This, of course, is entirely in line with the more nuanced results in Table 2, and the finding that *Monetary Policy* alone accounts for slightly more than 30% of the total risk premium. Interestingly, Table 3 also shows that news about *Macroeconomic Data*, broadly defined, accounts for roughly 7% of the jump risk premium and more than 15% of the total risk premium. This result also indirectly corroborates a long list of prior studies documenting large (in an absolute value sense) stock market returns in response to macroeconomic news announcements

[27] Since some of the news topics appear in more than one metatopic, the total risk premia estimates reported in Table 3 based on the topic estimates in Table 2 add up to slightly more than 100%.

(see, e.g., the early studies by Pearce and Roley (1985), French and Roll (1986), and Andersen et al. (2003, 2007), along with the more recent work by Gürkaynak, Kısacıkoğlu, and Wright (2020) among others).[28] As previously noted, news about *International Affairs* also plays a very important role on par with that of *Macroeconomic Data*, while the two metatopics *Politics* and *Commodities and Energy* both account for around 10% of the total risk premium. Even though the remaining three metatopics appear somewhat less important overall, the corresponding total news risk premia estimates are all strongly statistically significant when judged by their individual t-statistics.[29]

The metatopic risk premia estimates reported in Table 3 are all based on the full-sample averages of the monthly $\hat{\beta}_t^{J,k'}\hat{\lambda}_t^J$ estimates. This obviously masks any temporal variation in the compensation for exposure to the different news topics. Meanwhile, the metatopic variance contributions previously discussed in Figure 8 clearly suggest that the relative importance of the different types of news varies over time. To address this issue, Figure 9 plots the average annual returns, updated on a monthly basis, for the top three (based on their total risk premia estimates) metatopic mimicking portfolios. To allow for easier interpretation of the numbers, we rescale the tangency portfolio returns to 10% per annum, to make the returns comparable in magnitude to those of the market portfolio, which earned approximately 9% per annum over the full sample period. Estimating the risk premia over relatively short one-year periods obviously increases the estimation error uncertainty compared to the full-sample estimates reported in Table 3. Hence, to help assess the statistical significance of the temporal variation, we also include the corresponding 95% Confidence Intervals (CI).

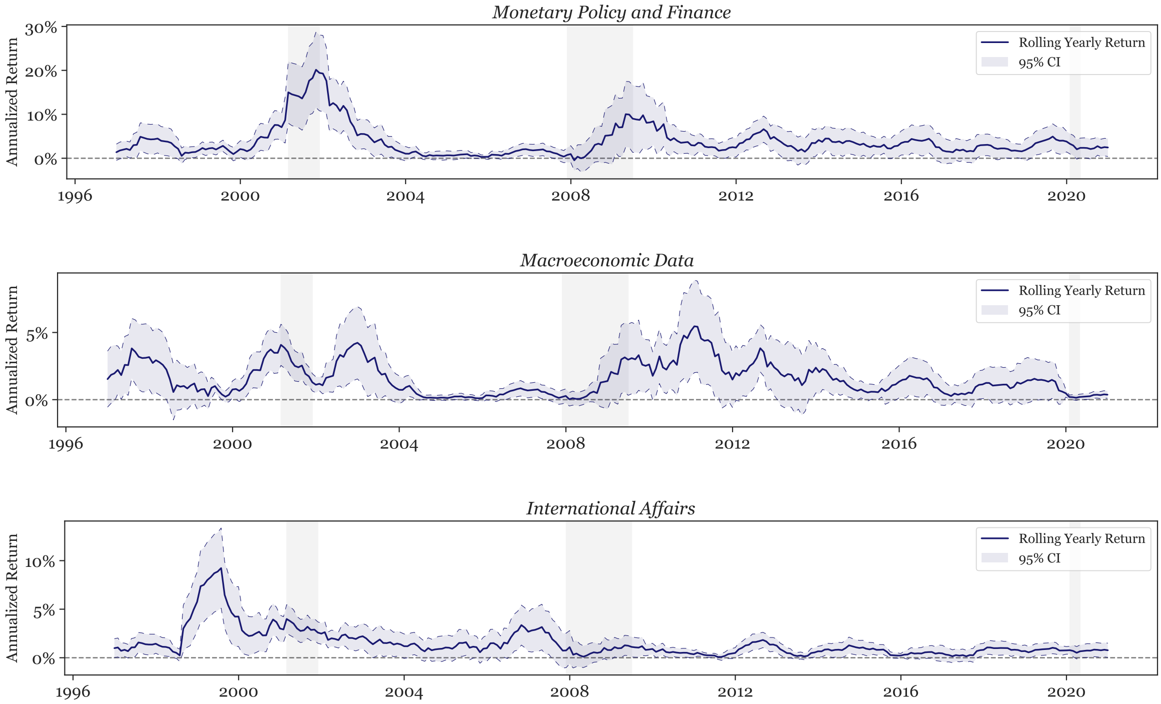
Looking at the top panel, the average returns on the *Monetary Policy and Finance* mimicking portfolio obviously remained fairly high, and statistically significantly so, throughout most of the sample, reinforcing the idea that investors generally put a high premium on the risks associated with said news. However, there is also a fair amount of variation in the average returns, with clear peaks in 2001-2002 around the time of the burst of the dot-com bubble, and in 2009-2010 at the end of the Global Financial Crisis and the early stages of the European Sovereign Debt Crisis. In other words, it appears as if investors put an even higher premium on *Monetary Policy and Finance* related news at the time of these periods of increased economic and financial market uncertainty.

The rolling returns for the *Macroeconomic Data* mimicking portfolio depicted in the second panel, not surprisingly, evidence a similar pattern, generally peaking during crisis periods. However, there are also much more marked increases in 1997 and 2011, underscoring the heightened importance of new economic data pertaining to different economic indicators and data during the Asian Financial Crisis and the European Sovereign Debt Crisis, respectively. Putting these results further into perspective of the existing literature, a number of studies have previously documented that the stock market tends to react differently to macroeconomic news announcements in recessions compared to expansions (see, e.g., the early work by McQueen and Roley (1993), along with Boyd, Hu, and Jagannathan (2005), and Andersen et al. (2007)). Law, Song, and Yaron (2020) have also recently refined that idea, arguing that the reaction to macroeconomic news

[28] A theoretical explanation for the importance of macroeconomic announcements, rooted in revealed preference theory and the idea that the announcements provide important information about the prospects of future economic growth, has also recently been developed by Ai and Bansal (2018). Relatedly, Ai et al. (2022) propose an equilibrium-based model for the cross-sectional pricing of FOMC announcements.

[29] This significance also easily "survive" a standard Bonferonni type correction for multiple testing at the 5% level.

Figure 9: Metatopic Risk Premia



Note: The figure plots the rolling annual returns on the *Monetary Policy and Finance*, *Macroeconomic Data*, and *International Affairs* metatopic mimicking portfolios, together with the corresponding 95% Confidence Intervals (CI), computed from the tangency portfolio returns rescaled to 10% per annum over the full sample. The shaded regions correspond to NBER-defined recessions.

is closely tied to investors' expectations about the likelihood of the Fed tightening its policies, while Schmeling and Wagner (2021) and Gardner, Scotti, and Vega (2022) emphasize the importance of tone and sentiment, respectively. The temporal variations in the news risk premia evident in the top two panels in Figure 9 further corroborate these ideas.

Turning to the final third panel, the average returns on the *International Affairs* mimicking portfolio peaked at the end of the Asian Financial Crisis and the beginning of the Russian Financial Crisis in 1999. The local peak in 2001, along with the higher returns throughout most of the 2000s, may naturally be ascribed to 9/11 and the War on Terror, including the Iraq war and the war in Afghanistan. Similar plots for the five other metatopic mimicking portfolios are included in Appendix G. Of these, perhaps the ones for *Commodities and Energy* and *Politics* stand out with their own most clearly distinct and easily interpretable patterns, highlighting the different economic news that investors are most concerned about, and thus carry the largest news risk premia, at different points in time.

5.3 News Risk Premia in the Factor Zoo

The significance and economic motivation for the myriad of risk factors proposed in the asset pricing literature continues to be an area of active debate (see, e.g., Harvey, Liu, and Zhu (2016), Hou, Xue, and Zhang (2020) and Jensen, Kelly, and Pedersen (2022)). There is also an older literature arguing that the success of the Fama-French model may be attributed to the ability of the corresponding characteristic-based factors to capture certain macroeconomic risks and innovations in state variables that are related to changes in the investment opportunity set (see, e.g., Vassalou (2003), Vassalou and Xing (2004), Petkova (2006), and Aretz, Bartram, and Pope (2010), among others). The importance of different news topics for explaining the factor risk premia provides an alternative new look at the risks that are actually priced in the factor zoo.

To set out the basic approach that we use for estimating the factor news risk premia, consider a zero-cost investment portfolio, or long-short factor j. Denote the time-t risk premium on the SDF tangency portfolio by λ_t, with the contributions stemming from news topic k denoted λ_t^k. The time-t risk premium for factor j may then naturally be decomposed as:

$$\mu_t^j = \beta_t^j \lambda_t = \beta_t^j \cdot \sum_k \lambda_t^k, \tag{17}$$

where β_t^j denotes the usual factor loading, or exposure, of the factor with respect to the tangency portfolio.[30] The factor loading is readily estimated by a standard time-series regression of the factor returns on the tangency portfolio returns. In the results reported below, we rely on the high-frequency 15-minute returns over rolling monthly windows for this estimation. Following the discussion in connection with equation (16) in Section 5.1 above, we similarly estimate the λ_t^k's on a rolling monthly basis by $\hat{\beta}_t^{J,k}\hat{\lambda}_t$. Combining the resulting $\hat{\beta}_t^j$ and $\hat{\lambda}_t^k$ estimates, our full-sample factor-specific topic risk premium for factor j and news topic k is in turn obtained by averaging $\hat{\beta}_t^j\hat{\beta}_t^{J,k}\hat{\lambda}_t$ over all of the months in the sample. This approach closely mirrors a traditional mimicking portfolio approach for the estimation of factor risk premia, except for the rescaling by $\hat{\beta}_t^{J,k}$. This additional rescaling stems from equation (16), and the decomposition of the total risk premium on the tangency portfolio into the different news topics.

To help more concisely convey the results, rather than reporting the estimates for all of the individual factors in the vast factor zoo, we instead focus on a set of thirteen representative factor cluster portfolios. Our definitions of the different factor clusters follow that of Jensen, Kelly, and Pedersen (2022) and Aleti (2022). The returns on the factor cluster portfolios are constructed as the average returns on the factors within each of the clusters. Table 4 reports the resulting full-sample topic risk premia estimates.[31] To help preserve space, we again restrict the display to the top ten topics based on their contributions to the risk premium earned by the tangency portfolio. For comparison, we also include the news risk premia contributions for the Fama-French market portfolio in the first row of the table.

The relative importance of the different topic risk premia for the market portfolio fairly closely mirror the total news risk premia estimates for the tangency portfolio previously reported in Table 2. In particular, *Monetary Policy* again stands out as the overall most important news topic, accounting for a large share of the market risk premium. That same topic also remains the relatively most important topic for explaining the news risk premia on most of the long-short factor cluster portfolios. The total risk premia on all of the cluster portfolios are, of course, substantially smaller (in an absolute value sense) than the risk premium on the market portfolio. Along these lines, and in contrast to the results for the market portfolio for which all of the news topics contribute positively to the overall risk premium, several of the news topics contribute negatively to the factor

[30] This implicitly assumes that $\lambda_t \equiv \sum_k \lambda_t^k$, which is guaranteed by our inclusion of the news topic *None*.

[31] The Online Supplemental Appendix reports the complete results for all of the 218 individual factor portfolios and the 48 industry portfolios that we use in the estimation of the SDF.

Table 4: Topic Risk Premia for Factor Cluster Portfolios

	Monetary Policy	US Politics	Energy Markets	Middle East	Russia	Labor Markets	Taxes	Inflation	Financial Regulation	Commodity Markets
Market	1.15	0.48	0.33	0.07	0.40	0.40	0.34	0.21	0.38	0.18
Value	0.41	−0.05	0.14	0.17	−0.11	0.17	−0.18	0.00	0.05	0.09
Investment	0.61	0.11	0.20	0.21	−0.10	0.21	−0.14	0.01	0.03	0.06
Low Risk	0.88	0.13	0.27	0.24	−0.17	0.07	−0.19	−0.01	−0.10	0.10
Profitability	0.50	0.16	0.22	0.15	0.05	0.01	0.03	0.07	−0.02	0.12
Quality	0.51	0.21	0.17	0.09	0.09	−0.08	0.12	0.04	−0.09	0.04
Leverage	−0.47	−0.01	−0.15	−0.15	0.13	−0.12	0.17	0.01	0.00	−0.06
Momentum	0.60	0.27	0.15	0.05	0.02	−0.05	0.11	−0.03	−0.11	−0.07
Size	0.16	−0.09	−0.03	0.02	−0.14	0.13	−0.16	−0.06	0.04	−0.04
Profit Growth	0.29	0.06	0.05	0.01	0.02	−0.04	0.05	0.01	−0.03	0.00
Accruals	0.19	0.07	0.01	0.05	0.11	0.04	0.10	0.00	0.03	−0.00
Debt Issuance	0.46	0.15	0.13	0.08	0.05	0.10	0.06	0.05	0.04	0.04
Skewness	0.29	0.07	0.07	0.05	−0.02	0.01	0.00	−0.00	−0.01	−0.00
Seasonality	0.28	0.02	0.05	0.06	−0.01	0.02	−0.02	0.00	−0.01	0.00

Note: The table reports the estimated percentage annualized returns for each of the thirteen representative factor cluster portfolios that are are earned from exposure to each of the different news topics. The definitions of the cluster portfolios follow Jensen, Kelly, and Pedersen (2022) and Aleti (2022). The display is limited to the top 10 news topics based on their contributions to the risk premium earned by the tangency portfolio. The first row reports the corresponding risk premia contributions for the Fama-French market portfolio.

cluster risk premia. For example, Value, Investment, Low Risk, and Size all draw large negative news risk premia from the *Taxes* topic. In contrast, the Leverage factor cluster portfolio draws a negative premium in general, but is positively affected by news about that same topic. The fact that some of the news topics contribute negatively to some of the factor risk premia is, of course, not surprising as the factors are based on equally sized long and short positions. Nonetheless, in general, most of the factors do draw a positive risk premium from each of the news topics, owing to their positive exposure to the SDF.

Table 5 reports the analogous results for our more broadly defined eight metatopics. In parallel to the previous table, the results again reveal quite heterogeneous metatopic news risk premia contributions for the different factor cluster portfolios. For instance, the Value, Investment, Low Risk, Profitability, and Quality factor cluster portfolios all draw a fairly large positive portion of their risk premia from news about *Commodities and Energy* generally defined, while that same metatopic continues to negatively affect the Size factor cluster portfolio. These estimates for the Value and Investment cluster portfolios also echo the findings of Lopez-Lira (2020), who reports that the Fama-French HML value and CMA investment factors both load positively on an "Oil risk" factor. For many of the clusters, news about *International Affairs* also carries a relatively large positive risk premium. At the same time, the risk premium for Size draws its largest (in an absolute value sense) negative contribution from news about *International Affairs*, closely followed

Table 5: Metatopic Risk Premia for Factor Cluster Portfolios

	Monetary Policy and Finance	Macroeconomic Data	International Affairs	Politics	Fiscal Policy	Commodities and Energy	Regulation	Labor
Market	1.61	1.05	0.56	0.53	0.40	0.44	0.39	0.59
Value	0.45	0.17	0.01	−0.13	−0.19	0.24	0.15	0.02
Investment	0.77	0.32	0.11	0.08	−0.08	0.28	0.22	0.05
Low Risk	0.94	0.11	0.17	0.12	−0.14	0.42	0.07	−0.16
Profitability	0.57	0.24	0.28	0.18	0.05	0.34	0.02	0.01
Quality	0.57	0.13	0.32	0.29	0.14	0.23	−0.05	−0.04
Leverage	−0.55	−0.13	−0.02	0.02	0.15	−0.24	−0.11	0.04
Momentum	0.75	0.13	0.20	0.40	0.19	0.15	−0.01	−0.11
Size	0.17	0.04	−0.20	−0.12	−0.18	−0.08	0.14	0.03
Profit Growth	0.23	0.02	0.05	0.08	0.06	0.05	−0.04	−0.03
Accruals	0.24	0.07	0.17	0.10	0.11	−0.01	0.05	0.07
Debt Issuance	0.58	0.30	0.21	0.17	0.11	0.17	0.11	0.09
Skewness	0.23	0.03	0.03	0.06	0.01	0.06	0.01	−0.04
Seasonality	0.27	0.01	0.06	0.02	−0.01	0.05	0.02	−0.01

Note: The table reports the estimated percentage annualized returns for each of the thirteen representative factor cluster portfolios that are are earned from exposure to each of the eight metatopics. The definitions of the cluster portfolios follow Jensen, Kelly, and Pedersen (2022) and Aleti (2022). The first row reports the corresponding risk premia contributions for the Fama-French market portfolio.

by news related to *Fiscal Policy*. The finding that *Fiscal Policy* plays a comparatively large role for the Size factor risk premium is consistent with the idea that smaller firms tend to be more strongly affected by credit market conditions. Meanwhile, news about *Macroeconomic Data* seems especially important for the Investment, Profitability, and Debt Issuance factors. Perhaps not surprisingly, aside from news about *Monetary Policy and Finance*, the risk premia for the Seasonality and Skewness factor cluster portfolios are both largely unaffected by any of the other economic news.

6 Conclusion

We exploit high-frequency data and real-time economic news to provide a novel characterization of systematic financial market risks. Our approach sidesteps the need for an explicit model for the stochastic discount factor, relying instead on a large panel of high-frequency portfolio returns combined with a minimax method of moments approach to robustly recover the tangency portfolio under minimal assumptions. By directly linking the jumps in the estimated tangency portfolio with the textual information in a comprehensive collection of precisely timed news articles, we are able to explicitly identify the type of news that matters to investors. Grouping the news articles into intuitive and interpretable categories of news topics, we find that *Monetary Policy*, *US Politics*, and *Energy Markets* stand out as the overall most important news topics for explaining the variation in the tangency portfolio returns.

To further address the economic significance of the news, we employ a mimicking portfolio approach, allowing us to decompose the risk premium on the tangency portfolio into separate components associated with each of the different news topics. Consistent with other recent studies emphasizing the importance of FOMC announcements for explaining the return on the market portfolio, news about *Monetary Policy* again emerge as the overall most salient news topic, explaining more than 30% of the tangency portfolio risk premium. Further combining the news articles into more broadly defined metatopics, we find that news related to *Monetary Policy and Finance* explains close to 40% of the tangency portfolio returns. Extending our procedure to allow for the decomposition of the news risk premia on other assets, we also shed new light on the type of news that accounts for the risk premia earned by the myriad of risk factors proposed in the asset pricing literature. In parallel to the results for the tangency portfolio, news related to *Monetary Policy and Finance* again emerges as the overall most important news topic for explaining most of the factor risk premia. At the same time, however, our results also reveal quite distinct news risk premia contributions for different factor cluster portfolios, with news about *International Affairs* and *Commodities and Energy* playing a comparatively larger role for many of the factors in the zoo.

The approach developed here could similarly be used to help illuminate the news that drive the systematic risks and returns for other financial assets. It may also be used for risk management purpose to help devise portfolios immune to certain types of news. Given the overall importance of the *Monetary Policy* topic, it would also be interesting to conduct a more thorough investigation into the type of monetary news that is actual priced. We leave further work along these lines for future research.

A Appendix: SDF Estimation Details

This appendix provides additional information about our practical implementation of the minimax optimization problem and our choice of hyperparameters.

A.1 Hyperparameters

Table A.1 list the hyperparameters used to define the weight and instrument functions, f_w and f_g.

Table A.1: Implementation – Hyperparameters for Main Model

Variable	Neural Net	Hyperparameter	Hyperparameter Value
Weights	LSTM	State variable dimension	4
Weights	LSTM	Activation	Tanh
Weights	FNN	Layer Structure	[4,2]
Weights	FNN	Intermediate Layer Activations	ReLu
Weights	FNN	Final Layer Activation	Tanh
Weights	FNN	Dropout Fraction	5%
Instruments	LSTM	State variable dimension	4
Instruments	LSTM	Activation	Tanh
Instruments	FNN	Layer Structure	[]
Instruments	FNN	Intermediate Layer Activations	ReLu
Instruments	FNN	Final Layer Activation	Tanh
Instruments	FNN	Dropout Fraction	5%
Instruments	FNN	Output Dimension	1000

The variable column indicates whether the specific hyperparameter is being defined for f_w or f_g. As stated in the main text, both of these functions consist of two components: a LSTM network that produces a low-dimensional set of state variables, and a FFN network that transforms the state variables to produce the weights/instruments. We purposely strive for simplicity when defining the hyperparameters, generally using the same values for both the instruments and weights. In particular, following CPZ, we fix the dimension of both state variables to be four. We also rely on the same type of hyperbolic tangent activation functions for the LSTM and the last layer of the FFN. This choice of activation is quite standard for LSTMs. In regards to the FFNs, the final layer activation functions ensure that the portfolio weights and instruments are bounded between reasonable values. We also use two hidden layers with four and two neurons each (as indicated by "$[4, 2]$" in the table) for the weight-generating FFN, combined with a trivial neural net (generalized linear model) for the instrument-generating FFN. This represents a fairly parsimonious choice motivated by the relative simplicity of our problem. It also matches the optimal layer structure suggested by CPZ, who formally tune the hyperparameters using a training/validation split. Both FFNs also use ReLu activations and a 5% dropout fraction, both of which are standard in the literature, and again mirror the choices of CPZ. Lastly, we fix the output dimension for the instruments to be 1000. This reflects a fairly large set of generated assets, while still respecting computational constraints.

A.2 Sharpe Ratio Bound

As discussed in the main text, we constrain the Sharpe ratios of our individually estimated tangency portfolios to lie in a pre-specified band. We enforce this constraint by adding a penalty to the objective function defined by (9) and (10):

$$\left\| \frac{1}{N_g \cdot T} \sum_t \hat{\alpha}_{g,i}^2 \right\|^2 + \underbrace{c \cdot \max\{S_{lower} - \hat{S},\ 0,\ \hat{S} - S_{upper}\}}_{\text{Sharpe Penalty}}. \tag{A.1}$$

We scale the penalty by setting $c = 10\%$. Since all of our fitted models achieve a Sharpe ratio within the $S_{lower} = 0.4$ to $S_{upper} = 1.5$ bounds that we impose, the penalty essentially works as a hard constraint. Alternatively, the penalty may also be interpreted as enforcing an uninformed prior on the true Sharpe ratio.

A.3 Numerical Optimization

With our hyperparameters and functional approximations defined above, we proceed to solve the minimax objective in equation (A.1) in three steps. We begin by determining the initial set of weights $\{w_t\}$ that minimize the unconditional and conditional alphas. By targeting both the monthly and full sample alphas, we obtain a sensible first guess for the tangency portfolio that should work reasonably well both conditionally and unconditionally. In total, this step consists of 1024 iterations, optimized using the well-known "Adam" algorithm (Kingma and Ba (2015)). In the second step, we then spend 64 iterations maximizing the objective function by updating the instruments $\{g_t\}$. The final third step then spends 1024 iterations minimizing the objective function by updating the weights $\{w_t\}$ using the instruments from the second step. Unlike the first step, this last step thus targets the conditional alphas generated by the instruments. Moreover, since the first step converges quite rapidly, we employ a "learning rate" of 0.01 to form our initial guess and then use a learning rate of 0.001 for the second and third steps. Finally, we repeat the last two maximization and minimization steps until convergence. Based on experimentation, we find that repeating the last two steps for a total of five iterations is generally sufficient for convergence.

In order to address potential concerns about our specific choices of hyperparameters, we also construct a series of additional SDF estimates based on a large grid of alternative hyperparameters.[32] The additional parameter values that we investigate are detailed in the Online Supplemental Appendix. Underscoring the robustness of our results, we find that the estimated tangency portfolios appear quite insensitive to the choice of hyperparameters. Correspondingly, the topic risk premia estimates reproduced using the alternative SDF estimates also closely adhere to those produced by our main estimate of the SDF. Again, these additional robustness results are further detailed in the Online Supplemental Appendix.

[32] Optimally selecting the hyperparameters, as sometimes done in the literature, would necessitate further partitioning of the sample into training and validation subsets. However, given our limited sample period spanning "just" 25 years, such an approach would also result in fairly unreliable alpha estimates.

B Appendix: SDF Jump Identification

To begin, recall that the tangency portfolio is defined as a weighted combination of the span assets. Hence, under the maintained assumptions that all of the span assets follow no-arbitrage Itô semimartingale processes, and the weights $w_t \equiv f_w(I_t; \theta_w)$ form a bounded predictable process, the tangency portfolio will itself be an Itô semimartingale. Succinctly expressing this process as

$$F_t \;=\; F_0 + \int_0^t \mu_s ds \;+\; \int_0^t \sigma_s dW_s \;+\; J_t, \qquad t \geq 0, \tag{B.1}$$

where μ_s defines the drift in the SDF, σ_s defines the SDF diffusive volatility, W_s is a standard Brownian motion, our goal is to identify the realization of the J_t jump process that accounts for large discontinuous changes in the SDF. To do so, we rely on the now standard thresholding approach originally proposed by Mancini (2001).

In particular, following Bollerslev and Todorov (2011) we classify a SDF return in the ith intraday time-interval on day t as a jump if the following condition holds,

$$|F_{t,i}| \geq \alpha\sqrt{\tau_i BV_t}\,\Delta_n^{\varpi}, \tag{B.2}$$

where Δ_n denotes the sampling frequency corresponding to n intraday observations per day, and the τ_i time-of-day indicator, and the BV_t bipower variation measure (Barndorff-Nielsen and Shephard (2006)) are defined as follows:

$$\tau_i = \left(\sum_t F_{t,i}^2\right) / \left(\sum_{t,j} F_{t,j}^2\right), \tag{B.3}$$

and

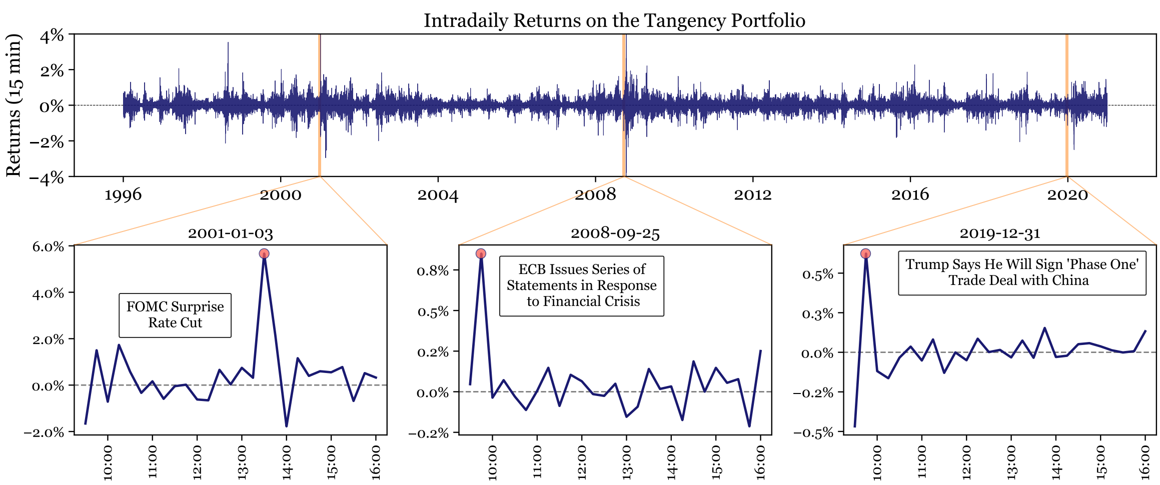
$$BV_t = \frac{\pi}{2} \cdot \frac{n}{n-1} \cdot \sum_{i=2}^{n} |F_{t,i}||F_{t,i-1}|. \tag{B.4}$$

Following now common choices in the literature (e.g., Todorov and Bollerslev (2010), Bollerslev and Todorov (2011) and Aït-Sahalia, Jacod, and Xiu (2021)), we further fix the two tuning parameters at $\alpha = 3.0$ and $\omega = 0.49$. Our procedure thus effectively amounts to classifying a SDF return that exceeds 3.0 local standard deviations as a jump. The Online Supplemental Appendix reports additional robustness checks for larger, and more conservative, values of α. All of our qualitative findings remain intact to these other choices of thresholds.

B.1 Illustration of News-Driven SDF Jumps

Figure 3 in the main text shows the full-sample continuous and jump returns in our estimated SDF identified by the above-discussed procedure. To further illustrate the idea, Figure B.1 shows three specific examples of SDF jumps, along with readily identifiable economic news occurring in the same 15-minute time interval as the identified jumps.

Figure B.1: Examples of News-Driven Jumps

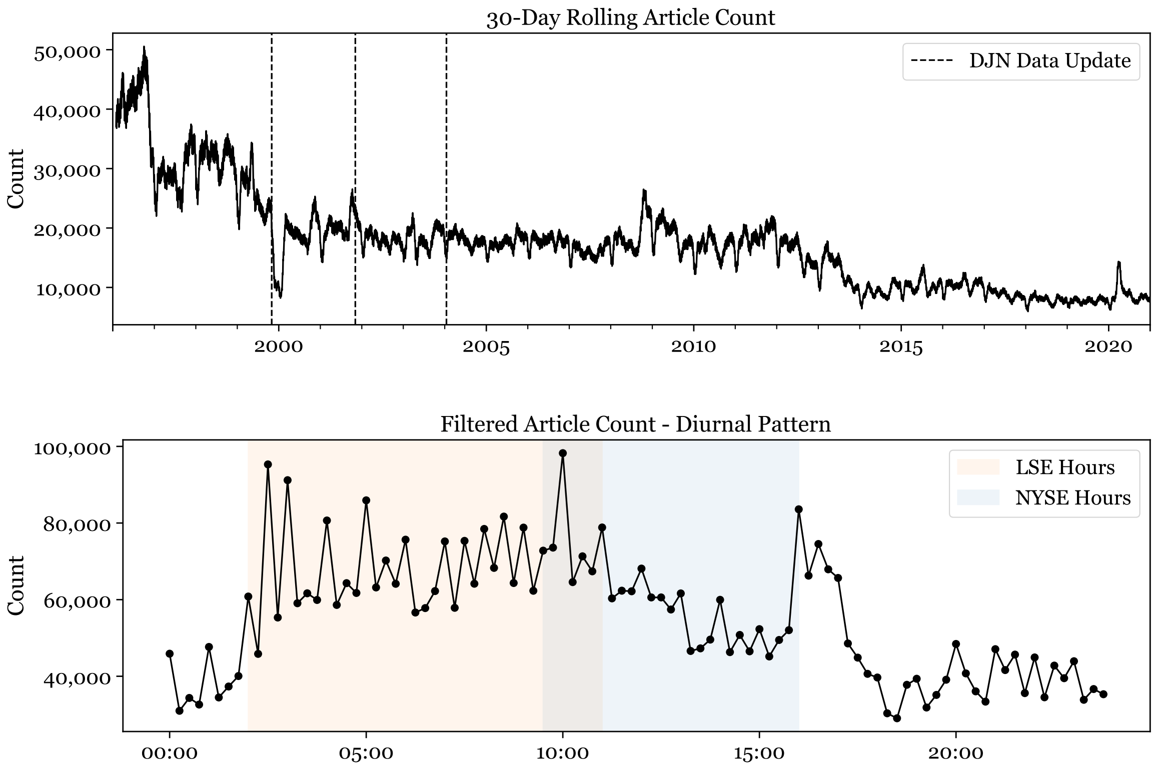


Note: The upper subplot shows the intradaily returns on the estimated tangency portfolio. The lower three subplots show three specific days with large jump returns that are readily associated with specific economic news.

C Appendix: News Article Counts

The first subplot in Figure C.1 shows the 30-day rolling count of "relevant" Dow Jones Newswires articles over our full sample. The plot evidence clear breaks associated with alterations in the underlying Dow Jones Newswires dataset, which itself was originally constructed from multiple sources by Dow Jones & Company. A document containing a full description of the breaks in the dataset is available upon request. The second subplot shows the filtered article counts over 15-minute intervals totalled across all of the days in the sample. As noted in the main text, the article counts are generally the highest during European and American market hours.

Figure C.1: Filtered Article Counts



Note: The first subplot shows the number of filtered news articles over 30-day rolling windows. The dashed lines indicate dates when there were major updates to the Dow Jones Newswires dataset. The second subplot shows the total article count for each 15-minute interval over the full sample period. The shaded areas indicate London Stock Exchange and New York Stock Exchange market hours.

D Appendix: News Topic Lists and Key Terms

Table D.1 reports the news topics and their associated key terms that we rely on. As discussed in the main text, these topics and key terms are based on Baker, Bloom, and Davis (2016), Davis (2017), Baker et al. (2019), and Bybee et al. (2021).

Table D.1: Topics and Key Terms

Topic	Associated Key Terms
Broad Quantity Indicators	*broad quantity indicator; depression; economic crisis; economic growth; economic slowdown; gdp; industrial production; ism report; macroeconomic indicator; macroeconomic news; macroeconomic outlook; manufacturing index; rail loading; railroad loading; recession; wholesale inventory*
Inflation	*consumer price; cpi; deflation; gold; inflation; ppi; producer price; silver*
Interest Rates	*bill rate; bond rate; bond yield; fed fund rate; interest rate; overnight rate; repo rate; yield curve*
Other Financial Indicators	*bank loan; business borrowing; business credit; business debt; consumer credit; credit spread; financial indicator; household borrowing; household credit; household debt; household saving; mortgage loan*
Labor Markets	*employment; hire; job; job data; job loss; job report; jobless claim; jobless rate; labor earnings; labor force; labor income; labor market; labor strike; nonfarm payroll; payroll; payroll data; payroll number; quits; ui claim; underemployment; unemployment; unemployment insurance; wage; weekly hour; work hour; workforce*
Real Estate Markets	*building permit; commercial construction; commercial real estate; home price; home sale; homebuilder; homebuilding; housing price; housing start; mortgage; real estate; real estate market; residential construction; residential sale*
Trade	*national export; national import; trade; trade balance; trade deficit; trade gap; trade news; trade surplus*
Business Investment and Sentiment	*business confidence; business inventory; business investment; business investment sentiment; business sentiment*
Consumer Spending and Sentiment	*automotive sale; consumer confidence; consumer purchase; consumer sentiment; consumer sentiment index; consumer spending; consumer spending sentiment; durable good; retail sale*
Commodity Markets	*aluminum; beef; board trade; cbot; cme; commodity exchange; copper; corn; cotton; gold; intercontinental exchange; lme; london metal exchange; mercantile exchange; metal; nymex; platinum; pork; rare earth metal; silver; soy; steel; sugar; tin; wheat; zinc*
Energy Markets	*alaska pipeline; biofuel; brent crude; coal; crude price; ethanol; export country; gas pipeline; high oil; higher oil; iea chief; international energy; international energy agency; keystone pipeline; natural gas; nonopec; oil; oil demand; oil export; oil exporter; oil market; oil minister; oil price; oil production; oil refiner; oil supply; opec; opec member; petroleum; petroleum export; refiner; world oil*
Healthcare Matters	*affordable care act; drug policy; fda; food drug administration; health insurance; healthcare; medicaid; medical liability; medical malpractice; medicare; national institute health; obamacare; prescription drug; va healthcare; va hospital; veteran affair healthcare; veteran affair hospital; veteran health administration*
Litigation Matters	*class action; copyright infringement; lawsuit; litigation; medical malpractice; patent infringement; punitive damage; supreme court; tort; trademark infringement*

continued...

Topics & Key Terms (continued)

Topic	Associated Key Terms
Competition Matters	*antitrust; cartel; clayton act; competition law; competition policy; european commission; federal trade commission; ftc; hartscott rodino; monopolization; monopoly; price conspiracy; price fixing; robinson patman act; sherman act; unfair business practice*
Labor Disputes	*employee discrimination; labor class action; labor dispute; labor litigation; labor unrest; strike; wage hour litigation*
Intellectual Property Matters	*copyright; federal trade commission; ftc; hatch waxman; intellectual property; international trade commission; new drug application; patent; patent trademark office; trademark*
Natural disasters	*baton rouge; beach; boat; coast guard; dam; debris; earthquake; emergency management; fire; fisherman; fishery; flood; flooding; floodwaters; hurricane; hurricane katrina; katrina; mississippi river; monsoon; natural disaster; quake; rita; salmon; storm; tsunami; volcano*
Disease	*acquire immune; aid virus; breast; breast cancer; cdc; corona; coronavirus; covid; deficiency syndrome; diagnosed; disease; disease control; ebola; epidemic; flu; heart disease; immune deficiency; immune system; implant; infected; infection; infectious; infectious disease; symptom; virus; world health*
Taxes	*401 k; accelerated depreciation; alcohol fuel credit; biofuel producer tax credit; biofuel tax credit; black liquor credit; black liquor tax credit; business tax; capital gain tax; carbon tax; corporate tax; deductibility state local tax; deduction mortgage interest; dividend tax; energy tax; ethanol credit; ethanol tax rebate; excise tax; fica; fiscal cliff; fuel excise tax rebate; fuel tax credit; futa; good service tax; gross receipt tax; income tax; internal revenue service; investment tax credit; ira account; low income housing credit; medicare tax; mortgage interest deduction; payroll tax; personal tax; profit tax; property tax; r tax credit; research development tax credit; roth ira; sale tax; social security contribution; social security tax; state local tax deduction; tax; tax capital gain; tax credit low income housing; tax cut; tax individual; taxation; taxed; traditional ira; unemployment tax; value added tax; vat*
Government Spending, Deficits and Debt	*balance budget; balanced budget; budget battle; budget deficit; budget office; budget sequestration; budget surplus; cbo; congressional budget; current debt; debt ceiling; debt gdp; debt limit; debt sustainability; debt war; defense appropriation; defense purchase; defense spending; deficit debt; deficit estimate; ecofin council; entitlement spending; euro area debt; extend debt; farm bill; federal budget; fin min; finance minister; fiscal challenge; fiscal cliff; fiscal stimulus; government appropriation; government budget; government debt; government deficit; government outlay; government purchase; government sequester; government shutdown; government spending; government subsidy; gramm rudman; military purchase; military spending; sovereign debt; stability growth pact*
Entitlement and Welfare Programs	*afdc; affordable housing; aid tfamilies dependent child; disability insurance; early childhood development program; earned income tax credit; eitc; entitlement program; entitlement spending; entitlement welfare program; food stamp; government entitlement; government subsidized housing; head start program; housing assistance; medicaid; medicare; public assistance; section 8; social security; ssi; supplemental nutrition assistance program; supplemental security income; taa program; tanf; temporary assistance needy family; unemployment benefit; unemployment insurance; welfare reform; wic program*

continued...

Topics & Key Terms (continued)

Topic	Associated Key Terms
Government-Sponsored Enterprises and Related Agencies	*fannie mae; farmer mac; federal agricultural mortgage corporation; federal farm credit bank; federal home loan bank; federal home loan mortgage association; federal housing agency; federal housing finance agency; federal national mortgage association; freddie mac; ginnie mae; government national mortgage association; refcorp; resolution funding corporation; sallie mae; student loan marketing association*
Monetary Policy	*bank england; bank italy; bank japan; bernanke; bond purchase; bundesbank; central bank; central bank china; discount window; draghi; ecb; ecb weber; emergency lending; euro currency zone; european central bank; fed; fed chair; fed chairman; fed fund; fed fund rate; fed official; fed treasury; federal reserve; federal reserve chairman; fomc; forward guidance; greece budget; greenspan; interest reserve; kaplan; kashkari; kuroda; lender last resort; lending program; lockhart; mester; monetary policy; money supply; ny fed; open market committee; open market operation; pbc; pboc; people bank china; plosser; powell; quantitative easing; rate cut; rate rise; rosengren; taper tantrum; tapering; trichet; volcker; volcker rule; volker; weber; yellen*
Financial Regulation	*bail; bank supervision; banking sector; banking system; basel; bureau consumer financial protection; capital requirement; cfpb; cftc; commodity future trading commission; comptroller currency; consumer financial protection bureau; deposit insurance; dodd frank; fdic; federal saving loan insurance corporation; financial crisis; financial reform; financial regulation; financial stability oversight council; financial system; fincl crisis; fincl sector; firrea; fslic; glass steagall; hedge fund regulation; high risk trade; house financial service committee; occ; office thrift supervision; ots; proprietary trading; sarbanes oxley; sba loan program; sec; security exchange commission; stress test; tarp; taxpayer insured bank; thrift supervision; troubled asset relief program; truth lending; volcker rule*
Competition Policy	*antitrust policy; clayton act; competition law; competition policy; european commission; federal trade commission; ftc; hartscott rodino; robinson patman act; sherman act*
Intellectual Property Policy	*copyright law; intellectual property policy; international trade commission; patent law; patent policy; patent trademark office; trademark law; trademark policy*
Labor Regulations	*advance notice requirement; affirmative action; card check; closed shop; davisbacon; department labor; eeoc; employment; equal employment opportunity; erisa; labor regulation; living wage; mine safety health administration; minimum wage; national labor relation board; nlrb; occupational safety health administration; osha; overtime requirement; pbgc; pension benefit guaranty corporation; right twork; trade adjustment assistance; union right; wage hour; worker compensation*
Immigration	*farm worker job program; farm worker program farm worker program; farmworker program; guest worker program; guestworker program; h 1b program; h 1b visa; h 2a program; h 2b program; immigrant labor; immigrant worker; immigration; immigration custom enforcement; immigration naturalization service; immigration policy; immigration reform; migration reform; refugee crisis; schengen*

continued…

Topics & Key Terms (continued)

Topic	Associated Key Terms
Energy and Environmental Regulation	*alaska oil pipeline; biofuel producer tax credit; biofuel tax credit; cafe standard; cap tax; cap trade; carbon tax; clean air act; clean water act; climate change regulation; corporate average fuel economy; drilling restriction; endangered specie; energy environmental regulation; energy policy; energy tax; environmental protection agency; environmental restriction; epa; ethanol credit; ethanol mandate; ethanol subsidy; ethanol tax credit; ethanol tax rebate; federal energy regulatory commission; ferc; greenhouse gas regulation; keystone pipeline; nuclear regulatory commission; offshore drilling; pipeline hazardous material safety administration; pollution control; transalaska pipeline; wetland protection*
Legal Reforms and Supreme Court	*class action reform; lawsuit reform; legal reform supreme court; medical malpractice reform; punitive damage reform; supreme court; tort reform*
Housing and Land Management	*bureau land management; department housing urban development; department interior; endangered specie; federal housing administration; federal housing finance agency; fheo; housing land management; hud; office fair housing equal opportunity; section 8 housing; u forest service; united state forest service; zoning law; zoning regulation*
Other Regulation	*consumer product safety commission; department education; fcc; federal communication commission; fish wildlife service; regulation; small business administration*
National Security	*9 11; airstrike; armed force; army; base closure; civil war; coup; darpa; defense advanced research project agency; defense policy; defense spending; department defense; department homeland security; enrichment; esuicide bomber; fly zone; gulf war; international atomic; iran nuclear; isi; islamic jihad; islamic militant; islamic state; kurd; laden; military action; military conflict; military embargo; military exercise; military procurement; military spending; military takeover; national security; nato; naval blockade; nonproliferation; nuclear program; nuclearweapons; nuclearweapons program; oil embargo; osama; palestinian bomber; saddam; saddam hussein; taliban; terror; terrorism; u gen; u general; war; warfare*
Trade Policy	*antidumping; bilateral trade; doha round; dumping; export duty; export restriction; export tax; federal maritime commission; freetrade; freetrade agreement; gatt; general agreement; import barrier; import duty; import restriction; international trade commission; investment restriction; jones act; nafta; north american free trade agreement; protectionism; protectionist; tariff; trade act; trade adjustment assistance; trade agreement; trade barrier; trade deal; trade dispute; trade gap; trade official; trade pact; trade partner; trade policy; trade quota; trade representative; trade talk; trade treaty; trade war; trans pacific partnership; transpacific partnership; unfair trade; uruguay round; world trade organization; wto*
Healthcare Policy	*affordable care act; health insurance; healthcare policy; malpractice reform; malpractice tort reform; medicaid; medicare; national institute health; obamacare; va healthcare; va hospital; veteran affair healthcare; veteran affair hospital; veteran health administration*
Food and Drug Policy	*drug policy; fda; food drug administration; prescription drug act*
Transportation, Infrastructure and Public Utilities	*amtrak; bonneville power administration; corp engineer; department transportation; faa; federal aviation administration; federal highway administration; federal highway fund; federal maritime commission; los angeles department water power; nasa; national aeronautics space administration; national highway traffic safety administration; national railroad passenger corporation; new york public power authority; pipeline hazardous material safety administration; salt river project; santee cooper; south carolina public service authority; southeastern power administration; tennessee valley authority; u surface transportation board*

continued…

Topics & Key Terms (continued)

Topic	Associated Key Terms
Elections and Political Governance	*brexit; brexit deal; congressional election; eurozone breakup; eurozone exit; greek exit; grexit; military revolt; parliamentary election; peace treaty; presidential election; presidential impeachment; scottish referendum; u n; united nation*
US Politics	*appropriation committee; assistant secretary; baker; barack; barack obama; baucus; bill clinton; bush; bush administration; bush plan; carter administration; cheney; clinton; clinton administration; clinton plan; committee chairman; contra; daschle; deputy assistant; dick cheney; george shultz; gop leadership; hastert; hillary rodham; house bill; house counsel; house floor; house gop; house press; impeachment; independent counsel; irancontra; james baker; janet reno; leader harry; leader trent; lewinsky; lott; majority leader; meese; michael flynn; minority leader; nicaraguan; nicholas brady; obama; obama administration; paula; president barack; president bush; president clinton; president dick; president donald; president george; president joe; president obama; president reagan; presidentelect; reagan; reagan administration; regan; reno; robert rubin; rodham; rodham clinton; rubin; sandinista; secretary james; secretary nicholas; senate appropriation; senate bill; senate democrat; senate floor; senate majority; senate version; senior administration; shultz; starr; state george; trump; trump administration; u house; u senate; watergate; white house; whitewater*
Agricultural Policy	*biofuel producer tax credit; biofuel tax credit; department agriculture; ethanol credit; ethanol mandate; ethanol subsidy; ethanol tax credit; ethanol tax rebate; usda*
Middle East	*abu dhabi; ankara; arafat; assad; baghdad; bahrain; damascus; dhabi; erdogan; gaza; gaza strip; hamas; hezbollah; hussein; iraq war; iraqi; iraqi government; iraqi leader; iraqi official; israel; israeli; israeli official; kurd; lebanese; libyan war; middle east; milosevic; mubarak; netanyahu; northern iraq; palestinian; palestinian authority; palestinian leader; persian; plo; recep; recep tayyip; saddam; saddam hussein; serb; syrian war; tayyip; turkish; west bank*
Russia	*boris yeltsin; crimea; crimean annexation; crimean invasion; gazprom; gorbachev; kiev; kremlin; medvedev; mikhail; mikhail gorbachev; moscow; president vladimir; putin; ruble; russia; russian; russian president; sergei; ukraine; ukraine conflict; ukraine invasion; ukrainian; vladimir; vladimir putin; yeltsin*
North Korea	*daewoo; jong; kim; kim jong; korea; korean; korean government; north korea; north korean; pyongyang; seoul south; south korean*
China	*beijing; beijing china; china; china central; china sea; chinese; chinese authority; chinese bank; chinese company; chinese government; chinese leader; chinese official; deng; guangzhou; jiang; south china sea conflict; wen; xinhua; xinhua news; yuan; zhang; zhu*

This table reports each of our topics and their associated key terms. The majority of the topics and key terms are based on Baker et al. (2019), with a few additional key terms adopted from Bybee et al. (2021).

D.1 Metatopics

The metatopics *Macroeconomic Data*, *Fiscal Policy*, and *Regulation* are based on the topic categories previously defined by Baker et al. (2019). The remaining metatopics are based on our own definitions as detailed in Table D.2. The topics *Interest Rates*, *Labor Markets*, *Financial Regulation*, *Other Financial Indicators*, *Inflation*, *Labor Regulations*, and *Immigration* all appear twice. All other topics only appear once. As such, there are 37 associated topics in total, of which 30 are unique. Correspondingly, 14 of the 44 news topics in Table D.1 are not associated with any of our eight metatopics.

Table D.2: Metatopic Compositions

Metatopic	Associated Topics
Monetary Policy and Finance	*Interest Rates; Other Financial Indicators; Financial Regulation; Monetary Policy; Inflation*
International Affairs	*Middle East; Russia; China; North Korea*
Macroeconomic Data	*Broad Broad Quantity Indicators; Inflation; Interest Rates; Other Financial Indicators; Labor Markets; Real Estate Markets; Trade; Business Investment and Sentiment; Consumer Spending and Sentiment*
Politics	*Elections and Political Governance; US Politics*
Commodities and Energy	*Commodity Markets; Energy Markets*
Fiscal Policy	*Taxes; Government Spending, Deficits and Debt; Entitlement and Welfare Programs*
Labor	*Labor Regulations; Labor Markets; Immigration*
Regulation	*Financial Regulation; Competition Policy; Intellectual Property Policy; Labor Regulations; Immigration; Energy and Environmental Regulation; Legal Reforms and Supreme Court; Housing and Land Management; Other Regulation*

Note: This table reports the topics we associate with each metatopic. Note that certain topics appear more than once.

E Appendix: Illustrating the Topic-Jump Linkage

To help better understand our topic-jump linkage procedure, this appendix provides a step-by-step illustration of our approach based on a specific sample of news from January 3, 2001, when the Fed issued a surprise rate cut. The return on the estimated tangency portfolio also experienced a significant jump shortly after the announced rate cut. Hence, the major news topic associated with that particular jump should naturally be "Monetary Policy." As it turns out, this is indeed the topic that our procedure detects as the primary news topic for the 13:15-13:30 time interval on that day.

To begin, Table E.1 reports the news headlines in the 13:00-13:30 time window. To facilitate the presentation and keep the table manageable, we only report the topic counts for two of our 44 different topics. As the table shows, within this time interval, there are seven articles with non-zero topic counts for *Monetary Policy*, six of which are directly related to the rate cut. Conversely, all the headlines associated with the Fed rate cut are also flagged with non-zero topic counts for *Monetary Policy*.

Aggregating the topic counts across articles over a slightly wider set of 15-minute intervals produces the results shown in Table E.2. The raw aggregated topic counts reported in the first two columns reveal which of the different topics are the most prevalent over each of the time intervals. However, as discussed in the main text, we purposely focus on topics that appear abnormally high, as opposed to topic counts which may simply be high on average. Accordingly, the last two columns in the table reports the corresponding demeaned aggregated topic counts.

Finally, in order to assign a topic to the tangency portfolio jump that occurred in the 13:15-13:30 time interval, we sort the demeaned counts and select the topic with the highest value as the one that "explains" the jump. Table E.3 reports the resulting aggregate demeaned counts for the top ten topics on that particular day sorted by the values in the 13:15-13:30 time interval. Since *Monetary Policy* stands out as the topic with the highest aggregate demeaned count in that interval, it therefore emerges as the topic associated with the observed jump.

Table E.1: News Headlines and Topic Counts

Date	Headline	Topic Counts: Monetary Policy	...	US Politics
2001-01-03 13:00:31	News Highlights:Ford Dec US Tota...	4	...	2
2001-01-03 13:01:00	News Highlights:Reserved Mideast...	0	...	5
2001-01-03 13:05:07	Investors' Intelligence Poll: Bu...	0	...	0
2001-01-03 13:06:15	Magnitude 5.8 Earthquake Hits Ku...	0	...	0
2001-01-03 13:13:06	Fed Cuts Funds Rate 0.50-Pt; Cut...	1	...	0
2001-01-03 13:14:41	Fed Still Sees Risks Weighted To...	1	...	0
2001-01-03 13:15:16	News Highlights:Reserved Mideast...	0	...	7
2001-01-03 13:15:57	Fed:"Stands Ready" To OK Further...	1	...	0
2001-01-03 13:16:47	FOMC: Actions Taken In Light Of ...	1	...	0
2001-01-03 13:17:34	FOMC:Lower Consumer Confidence, ...	1	...	0
2001-01-03 13:17:36	News Highlights:Fed Cuts Funds R...	3	...	5
2001-01-03 13:17:50	NY Stks Rally On Fed Rate Cut; D...	2	...	0
2001-01-03 13:18:24	FOMC: "Little Evidence" Producti...	1	...	0
2001-01-03 13:21:30	NY Stks Continue Surge;DJIA Up 3...	0	...	0
2001-01-03 13:25:04	Verbatim Text Of Fed Rate Cut An...	7	...	0
2001-01-03 13:26:17	ACA Financial Guaranty Rtgs Plac...	0	...	0
2001-01-03 13:29:36	US Crude Futures Leap 40C To Day...	2	...	0

Note: The table reports the article topic counts for *Monetary Policy* and *US Politics* based on the number of key terms associated with each of the two topics in each of the different articles. All of the articles were published on January 3, 2001 between 13:00 and 13:30.

Table E.2: Demeaned Aggregate Topic Counts

	Aggregated Topic Counts			Demeaned Aggregated Topic Counts		
Date	Monetary Policy	...	US Politics	Monetary Policy	...	US Politics
2001-01-03 12:30:00	6	...	3	2.40	...	-2.41
2001-01-03 12:45:00	0	...	28	-3.60	...	22.59
2001-01-03 13:00:00	1	...	21	-2.60	...	15.59
2001-01-03 13:15:00	6	...	7	2.41	...	1.59
2001-01-03 13:30:00	18	...	12	14.40	...	6.59
2001-01-03 13:45:00	21	...	6	17.40	...	0.58

Note: The table reports the aggregate and demeaned aggregate topic counts for *Monetary Policy* and *US Politics* over 15-minute intervals based on the number of key terms associated with each of the two topics. The demeaned aggregated counts are constructed by subtracting the aggregated topic counts over 30-day backward-looking rolling windows.

Table E.3: Sorted Demeaned Aggregated Topic Counts

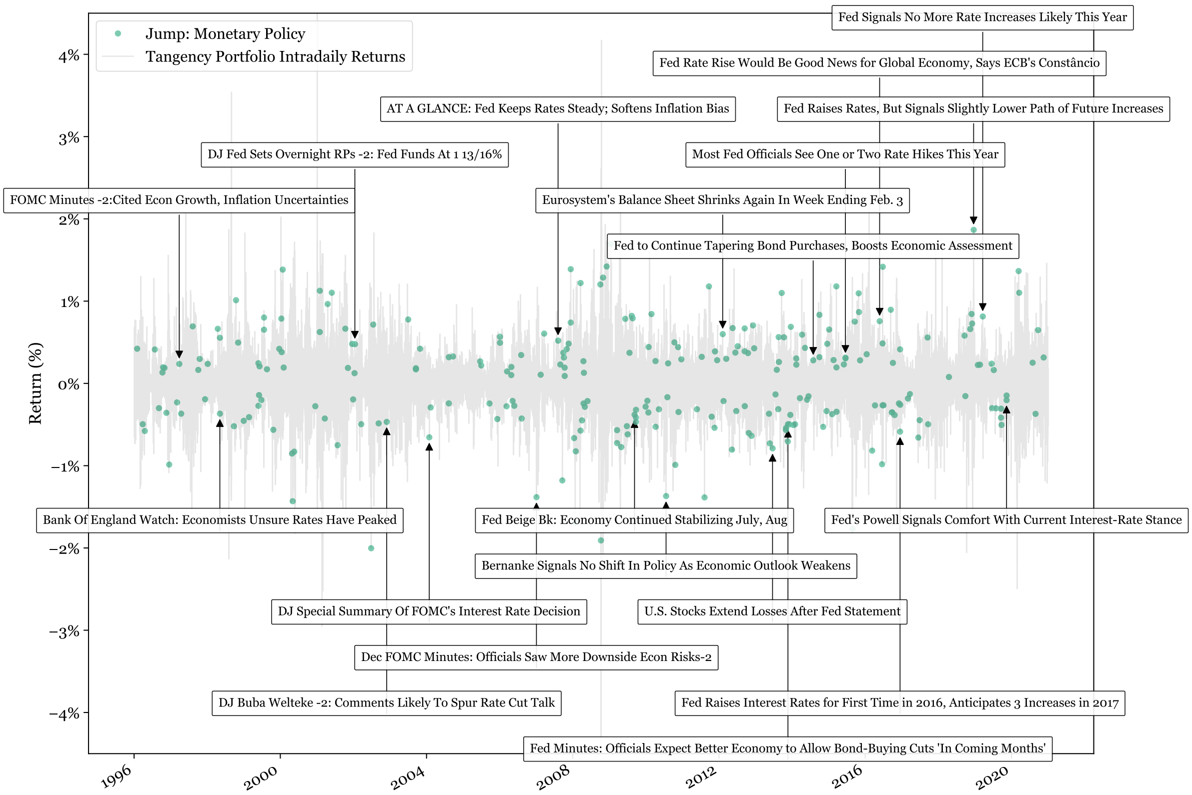
Topic	Time 12:30	12:45	13:00	13:15	13:30	13:45
Monetary Policy	2.40	−3.60	−2.60	2.41	14.40	17.40
Taxes	6.73	−1.27	14.72	6.72	6.72	−1.28
US Politics	−2.41	22.59	15.59	1.59	6.59	0.58
Consumer Spending and Sentiment	1.68	−0.32	0.68	1.68	3.68	0.68
Broad Quantity Indicators	2.09	−0.91	2.09	2.09	3.09	−0.91
Labor Markets	0.94	−1.06	−0.06	0.94	0.94	−1.06
Russia	−1.37	−1.36	0.64	0.63	0.63	−1.37
Trade	−0.51	0.49	−0.51	2.49	0.49	−1.51
Intellectual Property Policy	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
Housing and Land Management	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01

Note: The table reports the demeaned aggregate topic counts for the top ten topics on January 3, 2001. The demeaned aggregated counts are constructed by subtracting the aggregated topic counts over 30-day backward-looking rolling windows. The table is sorted by the topic counts at 13:30:00.

F Appendix: Select Topic Headlines

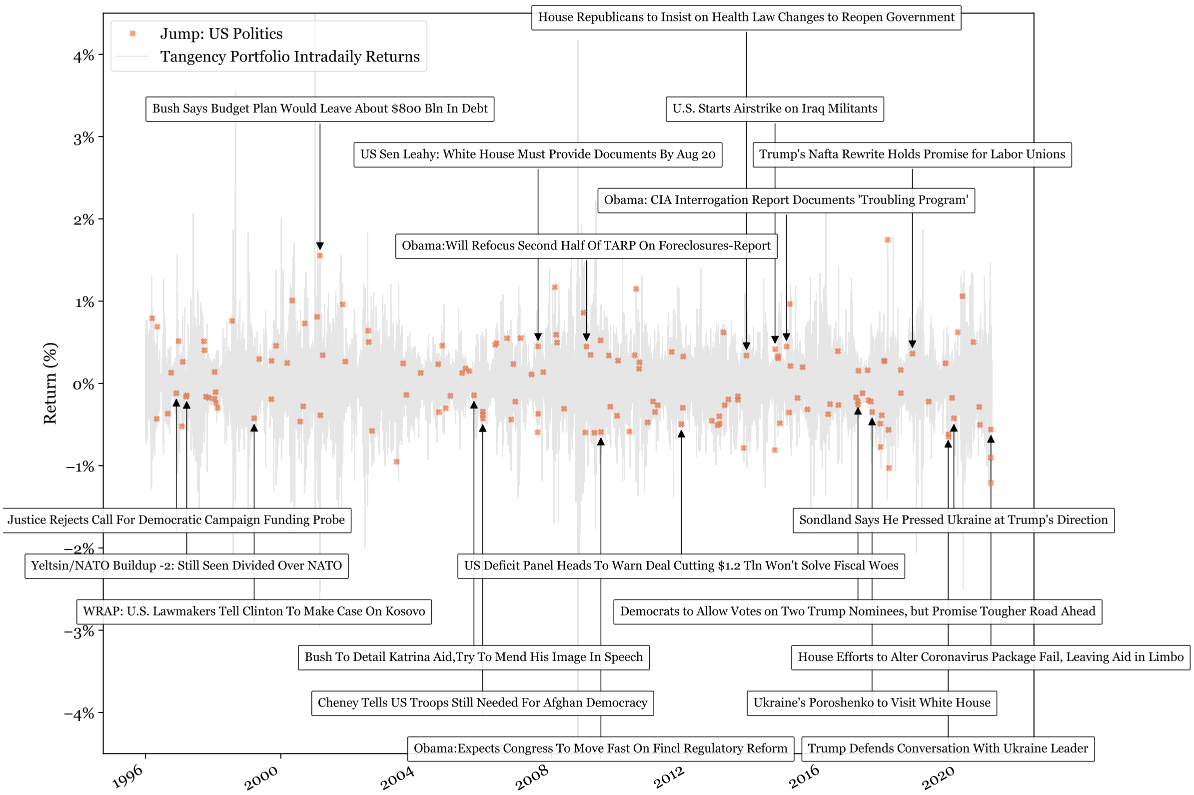
Complementing Figure 6 in Section 4.1, this appendix displays a set of additional select news headlines for the SDF jumps attributed to news about *Monetary Policy*, *US Politics*, and *Energy Markets*, respectively.

Figure F.1: Monetary Policy



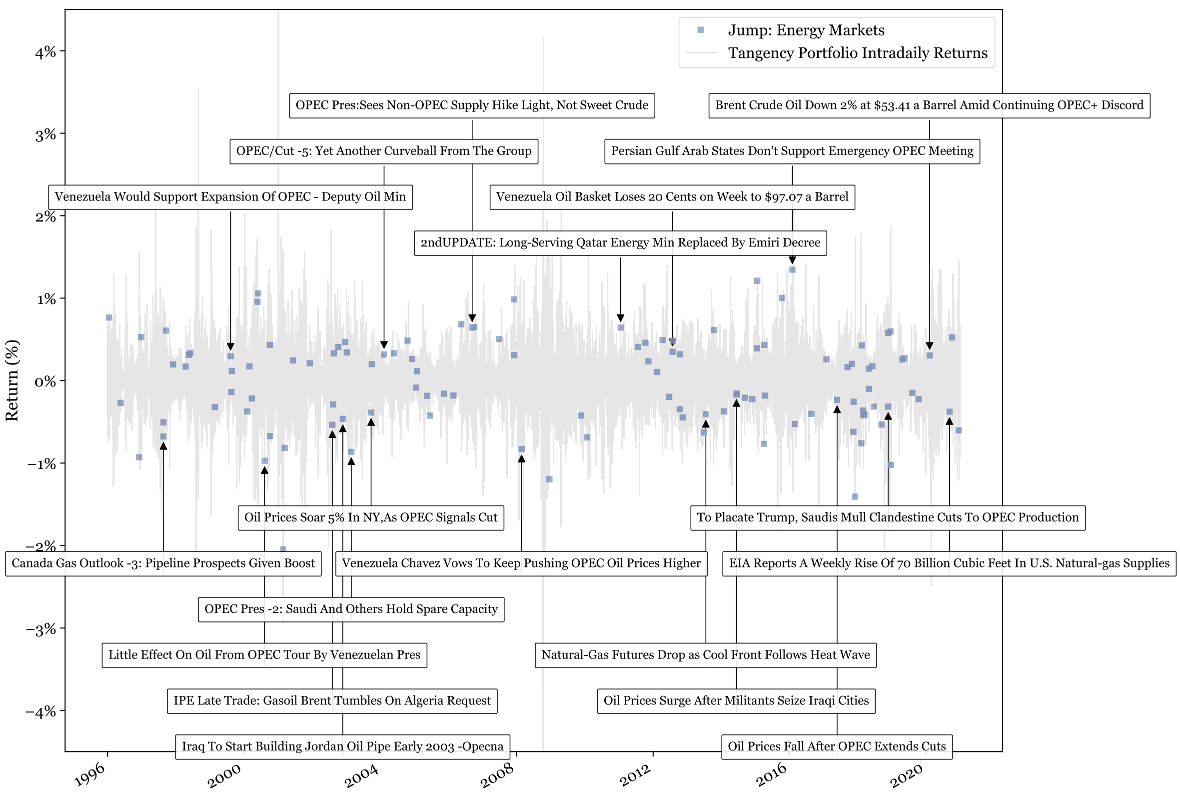
Note: The figure displays the intradaily 15-minute returns for the estimated tangency portfolio in gray, together with all the jumps associated with *Monetary Policy*. The select headlines are drawn from articles that were published in the same time interval as the jumps. Not all jumps are annotated in order to prevent overlapping headlines.

Figure F.2: Topic Headlines – US Politics



Note: The figure displays the intradaily 15-minute returns for the estimated tangency portfolio in gray, together with all the jumps associated with *US Politics*. The select headlines are drawn from articles that were published in the same time interval as the jumps. Not all jumps are annotated in order to prevent overlapping headlines.

Figure F.3: Topic Headlines – Energy Markets

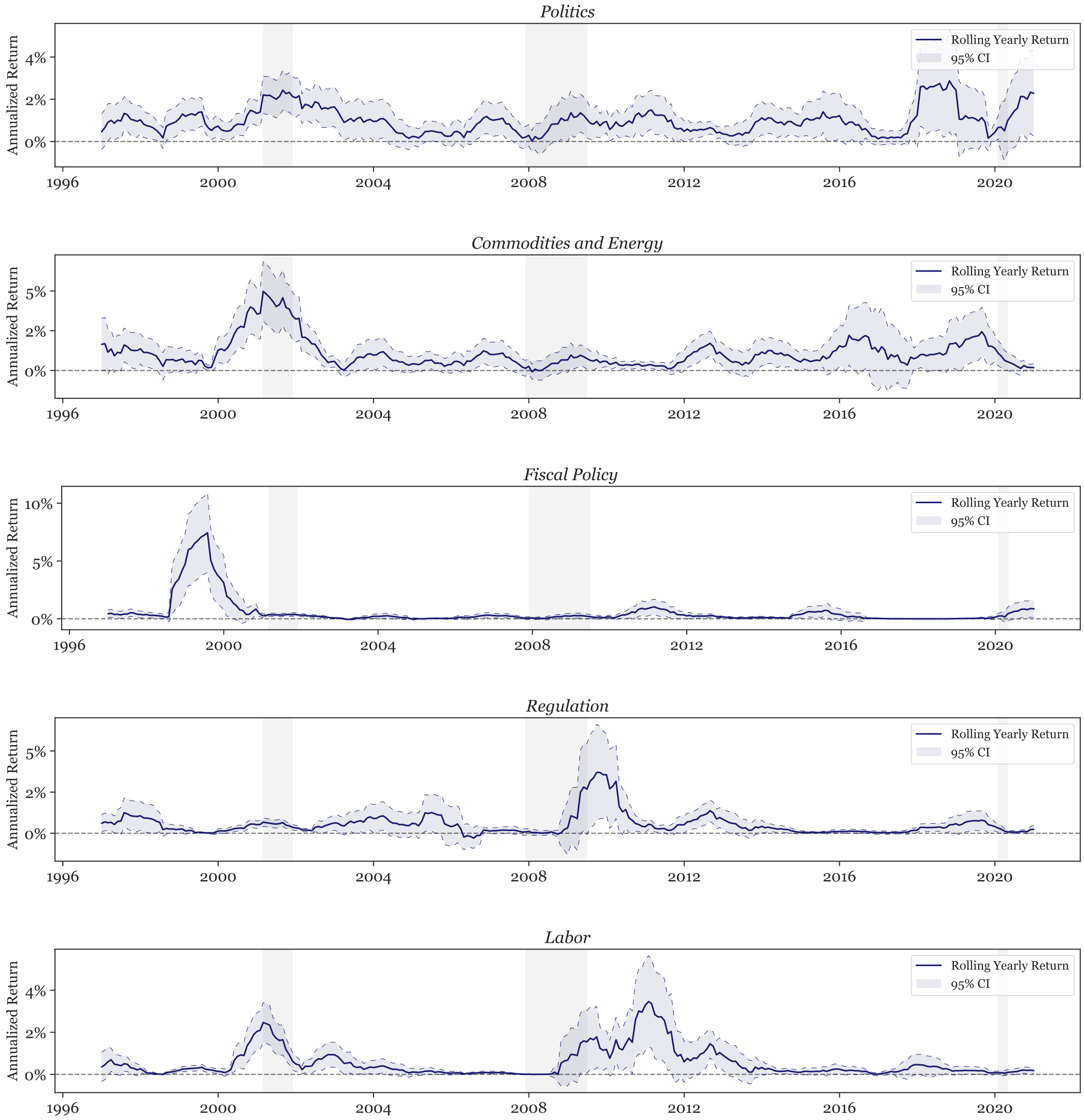


Note: The figure displays the intradaily 15-minute returns for the estimated tangency portfolio in gray, together with all the jumps associated with *Energy Markets*. The select headlines are drawn from articles that were published in the same time interval as the jumps. Not all jumps are annotated in order to prevent overlapping headlines.

G Appendix: Metatopic Risk Premia

The following figure shows the time series risk premium estimates for the additional five metatopics, not included in Figure 9 in the main text.

Figure G.1: Metatopic Risk Premia



Note: The figure plots the rolling annual returns on the *International Affairs*, *Commodities and Energy*, *Fiscal Policy*, *Regulation*, and *Labor* metatopic mimicking portfolios, together with the corresponding 95% Confidence Intervals (CI). The estimates are calculated from the tangency portfolio returns rescaled to 10% per annum. The shaded regions mark NBER-defined recessions.

References

Ai, H., & Bansal, R. (2018). Risk preferences and the macroeconomic announcement premium. *Econometrica*, *86*(4), 1383–1430.

Ai, H., Jianyu, L., Pan, X. N., & Xu, L. (2022). Missing events in event studies: Identifying the effects of partially measured news surprises. *Journal of Financial Economics*, *143*(1), 247–276.

Aït-Sahalia, Y., & Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y., Jacod, J., & Xiu, D. (2021). Inference on risk premia in continuous-time asset pricing models. *NBER Working Paper*.

Aleti, S. (2022). The high-frequency factor zoo. *Duke University, Working Paper*.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review*, *93*(1), 38–62.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, *73*, 251–277.

Aretz, K., Bartram, S. M., & Pope, P. F. (2010). Macroeconomic risks and characteristic-based factor models. *Journal of Banking and Finance*, *34*(6), 1383–1399.

Back, K. E. (2010). *Asset Pricing and Portfolio Choice*. Oxford University Press.

Bajgrowicz, P., Scaillet, O., & Treccani, A. (2016). Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science*, *62*(8), 2198–2217.

Baker, S., Bloom, N., Davis, S. J., & Sammon, M. (2021). What triggers stock market jumps? *NBER Working Paper*.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, *131*(4), 1593–1636.

Baker, S. R., Bloom, N., Davis, S. J., & Kost, K. J. (2019). Policy news and stock market volatility. *NBER Working Paper*.

Bansal, R., Hsieh, D. A., & Viswanathan, V. (1993). A new approach to international arbitrage pricing. *Journal of Finance*, *48*(5), 1719–1747.

Bansal, R., & Viswanathan, V. (1993). No arbitrage and arbitrage pricing: A new approach. *Journal of Finance*, *48*(4), 1231–1262.

Banulescu, D., Hansen, P. R., Huang, Z., & Matei, M. (2016). Volatility during the financial crisis through the lens of high frequency data: A realized GARCH approach. *University of North Carolina at Chapel Hill, Working Paper*.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal*, *12*(3), C1–C32.

Barndorff-Nielsen, O. E., & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, *4*(1), 1–30.

Bernanke, B., & Kuttner, K. N. (2005). What explains the stock market's reaction to federal reserve policy? *Journal of Finance*, *60*(3), 1221–1257.

Bianchi, F., Kung, H., & Kind, T. (2022). Threats to central bank independence: High-frequency identification with Twitter. *London Business School, Working Paper*.

Bird, S., Klein, E., & Loper, E. (2009). *Processing Raw Text*. O'Reilly.

Bollerlsev, T. (2022). Realized semi(co)variation: Signs that all volatilities are not created equal. *Journal of Financial Econometrics*, *20*(2), 219–252.

Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *Review of Financial Studies*, *31*(7), 2730–2773.

Bollerslev, T., Li, J., & Xue, Y. (2018). Volume, volatility, and public news announcements. *Review of Economic Studies*, *85*(4), 2005–2041.

Bollerslev, T., & Todorov, V. (2011). Tails, Fears, and Risk Premia. *Journal of Finance*, *66*(6), 2165–2211.

Boyd, J. H., Hu, J., & Jagannathan, R. (2005). The stock markets reaction to unemployment news: Why bad news is usually good for stocks. *Journal of Finance*, *60*(2), 649–672.

Bryzgalova, S., Pelger, M., & Zhu, J. (2020). Forest through the trees: Building cross-sections of stock returns. *Stanford University, Working Paper.*

Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2021). Business news and business cycles. *NBER Working Paper.*

Bybee, L., Kelly, B. T., & Su, Y. (2022). Narrative asset pricing: Interpretable systematic risk factors from news text. *NBER Working Paper.*

Chen, A. Y., & Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Board of Governors of the Federal Reserve System, Working Paper.*

Chen, L., Pelger, M., & Zhu, J. (2022). Deep Learning in Asset Pricing. *Management Science, forthcoming.*

Chen, X., & Ludvigson, S. C. (2009). Land of addicts? An empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics*, *24*(7), 1057–1093.

Chernozhukov, V., Newey, W., Singh, R., & Syrgkanis, V. (2020). Adversarial Estimation of Riesz Representers. *MIT, Working Paper.*

Chinco, A., Clark-Joseph, A. D., & Ye, M. (2018). Sparse signals in the cross-section of returns. *Journal of Finance*, *74*(1), 449–492.

Christensen, K., Timmermann, A., & Veliyev, B. (2022). Prices moving at warp speed: Jumps after earnings announcements. *Aarhus University, Working Paper.*

Cieslak, A., Morse, A., & Vissing-Jorgensen, A. (2019). Stock returns over the FOMC cycle. *Journal of Finance*, *74*(5), 2201–2248.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Cutler, D. M., Poterba, J. M., & Summers, L. H. (1989). What moves stock prices? *Journal of Portfolio Management*, *15*, 4–12.

Davis, S. J. (2017). Regulatory complexity and policy uncertainty: Headwinds of our own making. *University of Chicago, Working Paper.*

Evans, K. P. (2011). Intraday jumps and US macroeconomic news announcements. *Journal of Banking and Finance*, *35*(10), 2511–2527.

Fair, R. C. (2002). Events that shook the market. *Journal of Business*, *75*(4), 713–731.

Faust, J., Rogers, J. H., Wang, S.-Y. B., & Wright, H., Jonathan. (2007). The high-frequency response of exchange rates and interest rates to macroeconomic annuncements. *Journal of Monetary Economics*, *54*(4), 1051–1068.

French, K. R., & Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, *17*(1), 5–26.

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, *19*(1), 3–29.

Gardner, B., Scotti, C., & Vega, C. (2022). Words speak as loudly as actions: Central bank communication and the response of equity prices to macroeconomic announcements. *Journal of Econometrics*, *231*(1), 387–409.

Gürkaynak, R. S., Kısacıkoğlu, B., & Wright, J. H. (2020). Missing events in event studies: Identifying the effects of partially measured news surprises. *American Economic Review*, *110*(12), 3871–3912.

Hansen, L. P., & Jagannathan. (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance*, *52*(2), 557–590.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the Cross-Section of Expected Returns. *Review of Financial Studies*, *29*(1), 5–68.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating Anomalies. *Review of Financial Studies*, *33*(5), 2019–2133.

Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2022). Is there a replication crisis in finance? *Journal of Finance, forthcoming.*

Jeon, Y., McCurdy, T. H., & Zhao, X. (2022). News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics*, *145*(1), 1–17.

Johnson, J. A., Medeiros, M. C., & Paye, B. S. (2022). Jumps in stock prices: New insights from old data. *Journal of Financial Markets*, *forthcoming.*

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216.

Ke, Z., Kelly, B. T., & Xiu, D. (2021). Predicting returns with text data. *University of Chicago, Working Paper.*

Kim, S., & Korajczyk, R. A. (2021). Large sample estimators of the stochastic discount factor. *Northwestern University, Working Paper.*

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *ICLR Conference Proceedings.*

Kozak, S., & Nagel, S. (2022). When do cross-sectional asset pricing factors span the stochastic discount factor. *University of Chicago, Working Paper.*

Kozak, S., Nagel, S., & Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, *135*(2), 271–292.

Kuttner, K. N. (2001). Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of Monetary Economics*, *47*(3), 523–544.

Lahaye, J., Laurent, S., & Neely, C. J. (2011). Jumps, cojumps and macro announcements. *Journal of Applied Econometrics*, *26*(6), 893–921.

Larsen, V., & Thorsrud, L. A. (2019). Business cycle narratives. *Norges Bank, Working Paper.*

Law, T.-H., Song, D., & Yaron, A. (2020). Fearing the fed: How wall street reads main street. *Boston College, Working Paper.*

Lee, S. S. (2012). Jumps and information flow in financial markets. *Review of Financial Studies*, *25*(2), 439–479.

Lee, S., & Mykland, P. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, *21*(6), 2535–2563.

Li, J., Todorov, V., & Tauchen, G. (2017). Jump regressions. *Econometrica*, *85*(1), 173–195.

Lopez-Lira, A. (2020). Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns. *University of Florida, Working Paper.*

Lucca, D. O., & Moench, E. (2015). The pre-FOMC announcement drift. *Journal of Finance*, *70*(1), 329–371.

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari*, *64*, 19–47.

Manela, A., & Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, *123*(1), 137–162.

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics*, *34*(4), 574–589.

McQueen, G., & Roley, V. V. (1993). Stock prices, news and business conditions. *Review of Financial Studies*, *6*(3), 683–707.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, *8*(4), 323–361.

Nakamura, E., & Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: The information effect. *Quarterly Journal of Economics*, *133*(3), 1283–1330.

Pearce, D. K., & Roley, V. V. (1985). Stock prices and economic news. *Journal of Business*, *58*(1), 49–67.

Petkova, R. (2006). Do the Fama-French factors proxy for innovations in predictive variables? *Journal of Finance*, *61*(2), 581–612.

Pukthuanthong, K., Roll, R., & Wang, J. (2020). An agnostic and practically useful estimator of the stochastic discount factor. *University of Missouri, Working Paper.*

Rigobon, R., & Sack, B. (2004). The impact of monetary policy on asset prices. *Journal of Monetary Economics*, *51*(8), 1553–1575.

Rogers, J. H., Scotti, C., & Wright, J. H. (2014). Evaluating asset-market effects of uncnventional monetary policy: A multi-country review. *Economic Policy*, *29*(80), 249–799.

Savor, P., & Wilson, M. (2013). How much do investors care about macroeconomic risk? Evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis*, *48*(02), 343–375.

Schmeling, M., & Wagner, C. (2021). Does central bank tone move asset prices? *University of Vienna, Working Paper.*

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance*, *44*(5), 1115–1153.

Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Todorov, V., & Bollerslev, T. (2010). Jumps and betas: A new framework for disentangling and estimating systematic risks. *Journal of Econometrics*, *157*(1), 220–235.

Vassalou, M. (2003). News related to future GDP growth as a risk factor in equity returns. *Journal of Financial Economics*, *68*(1), 47–73.

Vassalou, M., & Xing, Y. (2004). Default risk in equity returns. *Journal of Finance*, *59*(2), 831–868.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508.