Risk exposures from risk disclosures: What they said and how they said it

Rahul Mazumder
Massachusetts Institute of Technology
rahulmaz@mit.edu

Seth Pruitt
Arizona State University
seth.pruitt@asu.edu

Landon J. Ross*
U.S. Securities and Exchange Commission
rossla@sec.gov

April 29, 2024

Abstract

We extract information from 10-K risk disclosures: topic models measure *what* is discussed, and context models measure *how* it is discussed. We find that both contain significant predictive information about future aggregate risk exposures, even controlling for (structured) firm characteristics, and that this information is economically valuable. For market exposure, only management's choice of context is useful; for other factors, tradable and nontradable, both context and topic information is useful. Further, we show that topics and contexts are statistically distinct. We present evidence that management's discussion helps to predict some future corporate actions, which in turn can alter the firm's exposure to aggregate risk.

Keywords: Text Analysis, Asset Pricing, Factor Model, Conditional Betas, Word embedding
JEL Codes: C23, C53, G11, G12, G17

*The Securities and Exchange Commission disclaims responsibility for any private publication or statement of any SEC employee or Commissioner. This working paper expresses the author's views and does not necessarily reflect those of the Commission, the Commissioners, or other members of the staff.

1 Introduction

This paper uses forward-looking text from the risk disclosure section of 10-K filings to predict firms' aggregate risk exposure.[1] We use two methods of extracting data from text, and show that they offer distinct predictive information. The first method uses *topic* models—successfully used by recent work in Cong et al. (2020), Bybee et al. (Forthcoming), Lopez-Lira (2023), amongst others—to measure what concepts are discussed. The second method uses text embedding to measure the context in which concepts are discussed—this approach has been less widely used in the finance literature. We find that both context and topics significantly predict future betas—they are complementary information sources. Therefore, agents obtain useful information from firm managers' choice of both *what* to say and *how* to say it.

Topics—a group of words chosen and their proportion of existence in one document relative to other documents—are measured by topic models. These methods aim to measure *what* is discussed by a document. One might ask of a topic model: How much is an exchange-rate topic discussed? We find that managers' choice of topics helps predict future beta with respect to many tradable and nontradable factors, although notably not to market beta.

Contexts—how a keyword is described by the surrounding text—are measured by text-embedding models, which are less familiar in the finance literature thus far. These methods aim to map words to numeric vectors where operations affect recognizable dimensions of meaning.[2] One might ask of our text-embedding model: How is the discussion of exchange rates in one document different than its discussion in other documents? We find that managers' choice of context helps predict future exposure to all the aggregate risks we consider.

[1] From the required Section 1A section presenting management's views on the risks facing the firm.

[2] A classic example is to consider the words *queen*, *king*, *woman*, and *man*. A useful text embedding could be one that maps these to vectors where $queen - woman + man = king$. That is, these vectors would identify that *queen* and *king* are similar in terms of being royal persons, and different in gender as identified by the words *woman* and *man*.

Remarkably, topics and contexts *simultaneously* deliver important predictive information for many aggregate risk exposures. In forecast combination regressions both topic and context predictive variables are statistically significant much of the time. When not the case, it is the context-based forecast that remains strongly statistically significant. In particular, this is true for investment risk exposure—we find this interesting because investment decisions, of all the portfolio-sorting variables we implicitly consider, are the most affected by management's choice. Therefore, when 10-K text contains forward-looking information most related to managers' decisions, it comes primarily from the context of how management discusses key ideas.

It seems reasonable that a firm's relationship to aggregate risk is persistent, and consistent with this intuition we often find that current beta is a significant predictor of future beta. Importantly, the aforementioned results hold while controlling for the predictive content in current beta. We detail how forward-looking text-based information is novel relative to backward-looking market information.[3] Furthermore, we explain how context models uncover a new dimension of information that topic models cannot, and show this is valuable to investors.

Ours is the first paper to predict future exposures to familiar aggregate risks using both topic- and context-modeling of 10-K disclosures. The application is important and natural. Item 1A of a firm's 10-K must be a disclosure of the firm's relationship to risks that may affect future performance. Such text is *naturally* informative, if anything, about the importance of future risks both systematic and idiosyncratic. Our focus on aggregate risk exposure comes from two motivations. First, we contribute to the finance literature's abundant analysis of aggregate risks. Second, many important agents (investors, firm managers, risk managers, etc.) use factor models to make consequential choices. Therefore, we are showing how to

[3]In robustness checks we also consider the predictive content of other market- and accounting-based firm characteristics, many of which Kelly et al. (2021) find are significant predictors of future beta. Our qualitative results continue to hold.

extract important information about which both researchers and practitioners care.

The fact that we use *both* topic and context modeling is significant. Gentzkow et al. (2019) note that textual data use in the finance literature is on the rise, alongside the concomitant complication of mapping text to numerical quantities. Topic and context modeling represent two distinct means of such mappings; our current read of the respective literature suggest that the former is becoming common in finance and economics, while the latter is a main focus for computer science applications. Our results give concrete empirical evidence that *both* methods of extracting data from text are valuable. Moreover, we find they are often complementary, which should encourage their joint use in a variety of future research applications. We calculate the Bayesian certainty equivalent for text information and find it is about 28 basis points per annum—a reasonable yet economically-significant value.

Our paper builds on Loughran and McDonald (2011), which shows that context is crucial to effectively using financial documents for economic research. Loughran and McDonald (2011) show variation in a word's context is a potential source of significant measurement error when constructing text-based proxy variables for financial concepts. Much of the subsequent text-analytic literature uses carefully crafted dictionaries whose words are unlikely to occur in unintended contexts to reduce proxy variables' measurement error, e.g., Buehlmaier and Whited (2018), Garcia and Norli (2012), and Hoberg and Maksimovic (2015). We depart from Loughran and McDonald (2011) by viewing variation in a word's context information about firm-level variation in financial concepts.

Our results also contribute to a growing literature using firms' risk disclosures to study accounting, economic, and financial questions. Cohen et al. (2020) shows changes in the length of firms' risk disclosures predicts changes in firms' future earnings and other financial variables. Campbell et al. (2014) show that firms' risk disclosures accurately reflect pre-disclosure proxy variables for firms' risks using dictionary methods and a topic model. Tetlock et al. (2008) shows that negative fundamentals-related news predicts lower future

earnings and returns. Florackis et al. (2022) uses firms' risk disclosures to document a cybersecurity-related risk factor in the cross-section of stock returns. Dyer et al. (2017) uses a topic model to document the effect of SEC- and FASB-related reporting requirements on risk factors' contents, length, specificity, and informativeness.

The plan for the paper is as follows. The next section details the data we use, both structured (already numeric) and unstructured (requiring mapping to numbers). Section 3 explains how topic and context models distinctly extract data from the text, and the familiar forecasting methods we employ to make our case. After that we report our results demonstrating the significance of text-based information in Section 4. Section 5 provides further understanding of how text is mapping to useful information about future risk exposures. We then conclude.

2 Data

This section reports the data we use for the paper's results. First we describe the stock-market and macroeconomic data used to calculate betas. Then we detail the sample of text taken from Form 10-K filings' Item 1A.

Firms' aggregate-risk exposures are crucial—betas are used for a variety of financial applications. Anyone using the CAPM wants to know market beta: an investor to calculate expected returns, a corporate manager for capital-budgeting decisions, or a risk manager to simplify estimation of stocks' covariance matrix. Anyone instead using a multifactor model would require betas on the additional portfolio-based factors as well, and the same is true for nontradable factor models.

Firm betas are widely used—and in many applications, the beta we would like to know *pertains to the future*. It is the future aggregate covariance of the firm's return that is important. While this beta is about the future covariance of returns, it is known today

in accordance with standard no-arbitrage theory. Were betas static, then this distinction would not matter: but a host of papers (see Ferson and Harvey, 1999; Kelly et al., 2019, amongst others) point to the importance of beta dynamics. Therefore, market participants are motivated to estimate the betas that will govern future returns, and it is this point that guides our main research question.

We measure realized betas using simple regressions. The year t and firm i estimate of $\beta_{i,t}$ with respect to some factor is

$$r_{i,t}^d = \alpha_{i,t} + \beta_{i,t} f_t^d + e_{i,t}^d \tag{1}$$

where $r_{i,t}^d$ is the daily return of stock i on day d in year t and f_t^d is the daily return of some factor on day d in year t. The $\beta_{i,t}$ parameter is the beta of stock i with respect to the given factor during year t. We allow the intercept to freely vary across i and t and its estimate is not used in our analysis. If agents are rational, then this realized beta is measurement of the beta investors knew before year t.

The estimated $\beta_{i,t}$ will surely have some sampling error. So long as this is uncorrelated with information before year t, the measurement error satisfies classical assumptions. Therefore, using the $\beta_{i,t}$ estimates as dependent variables (instead of the unobservable true beta) will only reduce the R^2 of the forecasting regressions we run. Nevertheless, we want a reasonably small amount of measurement error in our dependent variables. With this in mind, we restrict ourselves to $\beta_{i,t}$ estimates coming from daily data with an abundance of available observations. If a firm has less than 230 daily observations in year t, then the firm is omitted from the sample. We start year t on the first business day in January and end on the last business day in December.

Our decision to use daily data provides virtually no limitation on what tradable factors we can consider, since portfolio returns are easily available at the daily frequency. However,

there are a host of nontradable factors that are observed only at a far lower frequency. For instance, macroeconomic series like unemployment and inflation are available only monthly, whereas consumption and gross domestic product are available only quarterly. This renders such nontradable factors unsuitable for our analysis, because we'd expect $\beta_{i,t}$ estimates to be unduly noisy due to the paucity of new observations during year t. Nevertheless, we would like to consider *at least some* macroeconomic factors that experience tells us that firm managers consider. The requirement for them to be included is that they have meaningful variation at the daily frequency.

We start by following Fama and French (2015) in choosing nontradable factors. We use daily excess returns on the market (Mkt-RF), the size portfolio (SMB), the value portfolio (HML), the investment portfolio (CMA), and the profitability portfolio (RMW). Of course, a host of other portfolios' returns are available daily. Due to its familiarity, we further consider the momentum factor (Jegadeesh and Titman, 1993). All of these data are available from Ken French's website.

As our three nontradable factors, we choose three macroeconomic series that reasonably affect firm performance and are available at the daily frequency—we download them from the Federal Reserve Economic Database (FRED), whose mnemonics we give. The first is the exchange rate which, given the global nature of commerce and supply chains, can have important effects on firms' performance. Since we focus on U.S. stock market returns, we use the exchange rate of the USD to a broad basket of foreign currencies constructed by the Federal Reserve (DTWEXBGS). The second nontradable factor is the credit spread, which can have differential effects on the borrowing costs for firms of different creditworthiness. We measure the credit spread using the daily return for Moody's seasoned BAA corporate bond index minus the return for Moody's seasoned AAA corporate bond index (BAA-AAA). Our final macroeconomic factor is the term spread, which can have differential effects on the borrowing costs for firms borrowing from their bank's floating-rate facility versus firms

able to issue corporate bonds at various maturities. This is measured by the daily return for the 10-year US treasury constant maturity index minus daily return for the 3-month US treasury constant maturity index (T10Y3M).

The construction of our Item 1A text sample begins with all Form 10Ks filed at the Securities and Exchange Commission (SEC) between the years 2006 and 2022. Our text sample begins in 2006 because this is the first year Item 1A is generally present in Form 10K filings. Within this sample period, there are several factors that determine the specific sample of Item 1A texts we use for the paper's results. First, we only use Item 1A texts from Form 10K filings where we can successfully extract an Item 1A text with high confidence. Second, we only use Item 1A texts when we can successfully link the Item 1A text to a CRSP/Compustat observation using the associated Form 10K's filing date and Central Index Key (CIK). Third, we drop Item 1A texts from the sample whose word counts are above the sample's 99th percentile and below the sample's 1st percentile because these texts contents is typically not the Item 1A portion of a filing.

3 Empirical Methods

Using lagged beta to forecast is simple and needs no description. However, our methods of extracting information from 10-K text are much more involved, as has been discussed in Gentzkow et al. (2019) and elsewhere. Our two methods are topic- and context-modeling, which now we describe.

3.1 Topic model

Latent Dirichlet allocation (LDA) is a Bayesian approach to factoring discrete data, here word counts, into a collection of latent topics (Blei et al., 2003). LDA assumes the following specification for generating a document's words from K latent topics:

1. Choose $N \sim \text{Poisson}(\xi)$.
2. Choose $\theta \sim \text{Dirichlet}(\alpha)$.
3. For each of the N words w_n in the document:
   (a) Choose a topic $z_n \sim \text{Multinomial}(\theta)$.
   (b) Choose a word w_n from multinomial distribution $p(w_n|z_n, B)$.

Variable N represents the number of words present in a document and is drawn from a Poisson distribution with parameter ξ. Vector $\theta \in R^K$ represents the distribution of K latent topics that generate a document's words. Vector θ is drawn from a Dirichlet distribution with parameter $\alpha \in R^K$. The expression z_n is a categorical variable representing the topic generating the word at position n in a document. The variable z_n is drawn from a multinomial distribution whose parameter θ controls the topic mixture for the document. Last, word w_n at position n in the document is drawn from a topic z_n specific multinomial distribution over all of the words occurring in the document sample. The parameter B is a vector specifying the probability each word in the sample is generated by topic z_n. The primary user-chosen parameter for LDA is the number of latent documents generating the text sample. We use an LDA specification with 25 latent topics because Lopez-Lira (2023) finds LDA factorizations of Item 1A texts with 25 topics have greater topic coherence than LDA factorizations with other numbers of topics. We use an online variational Bayes algorithm from Hoffman et al. (2010) to estimate the LDA specification's three unknown parameters: the parameter α controlling the distribution of topics within the Item 1A sample, the parameter θ controlling the distribution of topics within each document, and the parameter B controlling each topic's distribution over words present in the Item 1A sample.

3.1.1 Beta Forecasting

We use the LDA estimates of documents' topic mixtures, i.e. $\theta_{i,t} \in R^{25}$ for firm i and year t, as firm characteristics. Since LDA represents documents as relatively low-dimensional

features we use a standard ordinary least squares regression to estimate the relationship between firms' betas and Item 1A text topic mixtures:

$$\beta_{i,t+1} = a + \theta_{i,t}^T f + e_{i,t+1}. \tag{2}$$

The left-hand side variable is a firm's beta with respect to some factor. The parameter a is an intercept. The vector $f \in R^{25}$ contains the model's slopes. Scalar $e_{i,t+1}$ is an error term.

3.2 Context Model

Our main method of extracting context from 10-K text is to use text embedding methods that are very common in computer-science and statistical applications, but newer to the finance literature. Given the possible lack of familiarity to the reader, we start with simplified example of context. Suppose companies A and P both mention the term *exchange*. Company A says, "The Company uses derivative instruments, such as foreign currency forward and option contracts, to hedge certain exposures to fluctuations in foreign currency exchange rates." On the other hand, company P says, "Fluctuations in exchange rates, including as a result of currency controls or other currency exchange restrictions have had, and may continue to have, an adverse impact on our business, financial condition and results of operations." Both companies mention exchange-rate risk, but using different contexts: company A states it is actively hedging its exposure, while company P discusses adverse impacts. One might interpret the different contexts as implying (their respective management teams believe) company A has *less* future exchange-rate exposure than company P.

There is empirical support for this interpretation, it turns out. The above companies are Apple and PepsiCo, and the sentences are from the text of their respective FY2019 10-K's Section 1A. In those filings, each company mentions "exchange rate" five times—at a simplified

level, the prevalence of the exchange-rate *topic* is similar in both documents.[4] During 2019 both Apple's and PepsiCo's exchange-rate betas are insignificant (in fact, current exchange-rate beta does not predict future beta, according to results below). Thereafter during the turbulence of the COVID-19 pandemic, PepsiCo's realized beta becomes strongly significant ($t = 3.8$) while Apple's remains insignificant. While current beta and the exchange-rate topic *are not* predictive, our simple interpretation of the exchange-rate *context is* predictive. Our text-embedding approach essentially systematizes the illustration into a machine-learning method that is suitable for analyzing big data.

3.2.1 Text Embedding

We represent the context of word w in the year t Item 1A text for firm i with the vector $x^w_{i,t} \in R^L$. The definition of the context vector $x^w_{i,t}$ is

$$x^w_{i,t} = A\frac{1}{|C^w_{i,t}|}\sum_{c^j_{i,t}\in C^w_{i,t}}\sum_{w^k_{i,t}\in c^j_{i,t}} v^{w^k_{i,t}}. \tag{3}$$

Let's unpack the context vector's construction from right to left. The right-most variable is $v^{w^k_{i,t}} \in R^M$. Vector $v^{w^k_{i,t}}$ is a pre-trained vector for the word $w^k_{i,t}$ at position k of the Item 1A text for firm i in year t. The next symbol is $c^j_{i,t}$, which represents the context of the word at position j in the Item 1A text for firm i and year t. The context of word $w^j_{i,t}$ is the set $c^j_{i,t}$ of words within the K-word neighborhood of word j, i.e.

$$c^j_{i,t} = \{w^k \in W_{i,t} : |k - j| \leq K\} \tag{4}$$

[4]It turns out that topics are useful to forecasting exchange rate beta, on average. But these two documents have a similar predictive weight on informative topics—in other words, from a topic-model perspective these two observations should forecast the same. Meanwhile, the context model successfully classifies these two statements differently.

where $|k-j|$ is the distance between words $w_{i,t}^k$ and $w_{i,t}^j$ within a document and K controls the context's width. We are now able to interpret the construction's right-most summation expression;

$$\sum_{w_{i,t}^k \in c_{i,t}^j} v^{w_{i,t}^k} \tag{5}$$

sums pre-trained word vectors for words in the context of word w at position j in the Item 1A text for firm i in year t.

The same word often occurs at several different positions within a document. In this case, a word will be associated with more than one context within a document. We use the expression $C_{i,t}^w$ to represent the set of all contexts for word w within the Item 1A text for firm i and year t with the definition

$$C_{i,t}^w = \{c_{i,t}^j : w_{i,t}^j \in W_{i,t}, w_{i,t}^j = w\} \tag{6}$$

where the $w_{i,t}^j \in W_{i,t}$ restricts the set's contents only to contexts within the Item 1A text for firm i and year t and condition $w_{i,t}^j = w$ restricts the set's contents to contexts with keyword w. Additionally, we use the expression $|C_{i,t}^w|$ to represent the number of contexts for word w in an Item 1A text. We can now interpret more of the context vector's construction. The expression

$$\frac{1}{|C_{i,t}^w|} \sum_{c_{i,t}^j \in C_{i,t}^w} \sum_{w_{i,t}^k \in c_{i,t}^j} v^{w_{i,t}^k} \tag{7}$$

is the average vector representation for the context of word w in the Item 1A text for firm i and year t.

The last undefined expression in the context vectors' construction is the matrix $A \in R^{L \times M}$. The matrix A serves two purposes. The first purpose of A is to adapt the pre-trained word vectors to better represent words meanings' within Item 1A texts. Pretrained word vectors are estimated using very large text samples taken from a variety of domains. The

matrix A tilts the pre-trained vectors' meanings away from their original multiple-domain setting and towards the paper's finance-specific domain. The second purpose of matrix A is to dimension reduction. The standard dimension of pretrained word vectors is 300 (Mikolov et al., 2018). The matrix A maps high-dimensional pre-trained word vectors into lower-dimensional vectors for the paper's specific task.

The paper's context vector construction requires specifying several details for the paper's results. We use the fasttext word vectors from Mikolov et al. (2018) as pre-trained word vectors. We use a context window size of $K = 20$. We estimate the matrix A using a modified form of the specification from Khodak et al. (2018), which we describe in appendix B. We report results for context vectors with seven dimensions. We determine seven components are sufficient for the context vectors using a test from Gavish and Donoho (2014), which we also describe in appendix B. We estimate context vectors for all nouns that occur in at least 5% of the Item 1A texts and that are not stop words. We restrict our sample of contexts to nouns because risks are typically nouns.

3.2.2 Beta forecasting

We use a group lasso specification to estimate the relationship between our embedding for Item 1A texts and firms' aggregate risk exposures. A few factors make the group lasso an appropriate specification. First, our text embedding uses approximately 18,000 firm characteristics to represent Item 1A texts' information about future returns. Since the text embedding is high-dimension, we consider some manner of regularization natural to ensure our empirical estimates plausibly generalize to new data. Second, we are equally interested in producing empirically useful descriptions of firms' aggregate risk exposures and interpreting our empirical estimates of firms' aggregate risk exposures. Since one of our aims is interpretation, we consider sparsity an appropriate form of regularization because sparse empirical solutions are relatively amenable to inspection. Third, the natural unit of sparsity

for our text embedding is at the level of the context vector for one word, which contains more than one variable. So, group sparsity, where each group contains the context vector for one word, is the appropriate form of sparsity to consider.

The linear model associated with the group lasso specification is

$$\beta_{i,t+1} = a + \sum_{w} x_{i,t}^{w^T} f_w + e_{i,t+1} \tag{8}$$

where a is an intercept, $x_{i,t}^{w}$ is the context vector for word w, $f_w \in R^7$ is the slopes for components of the word w context vector, and $e_{i,t+1}$ is an error term. We estimate the linear model's slopes via the regularized least squares problem

$$\hat{a}, \hat{f}_1, \ldots, \hat{f}_W = \underset{a, f_1, \ldots, f_W}{\arg\min} \ \frac{1}{2} \sum_{i,t} \left(a + \sum_{w} x_{i,t}^{w^T} f_w - \beta_{i,t+1} \right)^2 + \lambda \sum_{w} \|f_w\|_2. \tag{9}$$

The minimization problem's first term is the standard least squares term. The problem's second term is a group lasso penalty which shrinks each context vector's slopes towards zero. The λ is a scaling parameter controlling the magnitude of the group lasso penalty. We use five-fold cross-validation and a grid search to estimate this parameter.

3.3 Out-of-sample analysis

We take seriously the potential for overfitting. The large dimension of our empirical analysis stems from the large dimensionality inherent in text information (see Gentzkow et al., 2019). Therefore, an important concern is the possibility of overfitting when estimating the text-embedding mapping and the beta-forecasting parameters. We have detailed how those parameters are estimated by a variety of dimension-reduction techniques, and these constitute our first safeguard against overfitting. Our second, important safeguard is sample-splitting. We estimate those text-related parameters using only the first half of the sample,

2006–2014. Then we evaluate their performance only on the second half of the sample, 2015–2022. Therefore our main results are out-of-sample, which is crucial to delivering conclusions that are not driven by overfitting.

4 Results

We start by showing the out-of-sample predictive performance of forecasts based on past betas, topics, or contexts. Text information, of both types, are significant predictors of future risk exposure. These conclusions are robust to controlling for a wealth of structured firm-characteristic data.

4.1 Out-of-sample performance

Table 1 reports out-of-sample R^2 of forecasting realized beta, using forecasts errors from the 2015–2022 sample. That is, the statistic is constructed as

$$1 - \left(\sum_{2015-2022} (\beta_{i,t} - b_{i,t-1})^2 \right) / \left(\sum_{2015-2022} \left(\beta_{i,t} - \bar{b}_i\right)^2 \right).$$

We have abused notation and use "2015-2022" to denote the set of firm-year observations in the 2015–2022. In the numerator we forecast $\beta_{i,t}$ using a prediction $b_{i,t-1}$ based on year $t-1$ information—the forecast comes from year $t-1$ beta, topics, or contexts. In the denominator, the R^2 uses as its benchmark $\bar{b}_i$ the firm's beta on the 2006–2014 sample, or if the firm does not exist during 2006–2014 we use the average across firms. This benchmark forecast seems imminently reasonable, as it is simply the long-run average beta the agent saw during 2006–2014 (or else the average across firms). Note that this benchmark allows each firm i to have a different benchmark forecast, somewhat analogous to if one constructed a panel R^2 by using

Table 1: Out-of-sample performance

Notes– Reports the out-of-sample R^2 of forecasting realized beta, using forecasts errors from the 2015–2022 sample. All parameters are estimated on the 2006–2014 sample. The realized beta for each is estimated from a univariate regression of daily stock returns on daily factor realizations, as described in the text. The R^2 uses as its benchmark the firm's beta on the 2006-2014 sample, and if the firm does not exist during that period we use the average across firms.

Factor	Lagged beta	Topic	Context
Panel A: Portfolios			
Market	−0.054	−0.034	0.072
Size	0.156	0.016	0.082
Value	0.294	0.151	0.159
Investment	0.140	0.292	0.320
Profitability	0.599	0.103	0.104
Momentum	−0.584	0.230	0.236
Panel B: Macroeconomic			
Exchange Rate	−0.169	0.112	0.098
Credit Spread	−0.304	0.015	0.028
Term Spread	−0.165	0.006	0.038

forecast errors from firm fixed effects as the benchmark for the denominator.[5] We prefer this statistic as it embodies the idea that different firms have different betas, and so our R^2 is positive only if the predictions $b_{i,t-1}$ improve upon this basic fact.

Our first important result is that only contexts provide positive out-of-sample R^2 for market beta. The 7.2% is considerably strong—for context, Kelly et al. (2021) found an *in-sample* R^2 of 25.6% using thirty-six firm characteristics over a longer period. Strikingly, lagged beta gives negative forecast performance, implying it is usually a worse predictor of market beta than the longer-run average beta for that firm. Putting these observations together, we see evidence that market beta is dynamic, but its own past value is too noisy

[5]If we instead take every firm's benchmark forecast as simply the average beta in 2006–2014 across firms, this serves to increase all the R^2s relative to what we report—hence our statistic is a conservative approach. Additionally, for the market beta it could make sense to take a value of 1 as the benchmark for firms born after 2015; but for other factors there is no such sensible assumption. Therefore we use the average across firms for all factor betas.

to provide a good forecast. On the other hand, management's risk disclosures can provide valuable risk-exposure information that backward-looking market information does not.

Topics provide negative forecast performance for market beta. This could follow because management's information about the firm's exposure to the market, an aggregation of everything that is going in all firms, is not related to their choice of topics. We see a great variety of topics discussed in 10-Ks (we will see this below), but a particular filing's choice of topic tells us nothing about their future market beta. Yet, *the context of how* management describes topical keywords does reveal insight into market beta. This is a first glance at our second important result, that topic and context models reveal complementary and distinct information.

While only contexts are useful in forecasting market beta, there are several other aggregate risks for which both lagged betas and topics also perform well. Continuing with the portfolio factors, we see that lagged beta delivers a solid 15.6% for size exposure. The choice of topics delivers a modest 1.6%, indicating that management's choice of topics is only slightly connected to their future size beta. Meanwhile, the contexts chosen by management deliver 8.2% of the variation in future size beta, meaning that their disclosures do provide useful information to investors if extracted properly.

When forecasting value beta, lagged beta is once again a good forecaster, delivering R^2 of 29.4%. For value risk, both topic and context information are very similarly useful and deliver R^2s of 15.1% and 15.9%, respectively. Qualitatively similar, lagged beta provides a very strong prediction of future profitability beta with an out-of-sample R^2 of 60%, while topics and contexts are moderately useful with R^2s of 10.3-10.4%. Hence, value and profitability exposures are predicted quite well by their own recent past values, while text-based information is less useful.

Meanwhile investment risk has a flipped result: text-based information is *more* informative of future exposure than its own lag. Whereas 14% of the out-of-sample variation in

investment beta is explained by its lagged value, text-based information explains more than twice as much: topic-based forecasts deliver a R^2 of 29% and context-based forecasts deliver 32%.

Finally, momentum risk provides a stark contrast between backward-looking-market and forward-looking-text information. Its lagged value provides extremely negative forecasting power for momentum beta, with an out-of-sample R^2 of -58%. Nonetheless, there is significant information about future momentum beta coming from 10-K risk exposures. Both topic and context information deliver an out-of-sample R^2 of around 23-24%. The contrast here highlights the novel richness of forward-looking text-based information extracted by both topic and context models. Daniel and Moskowitz (2016) show that momentum strategies have experienced significant crashes, including during our 2006-2014 training sample. Kelly et al. (2021) show that other firm characteristics are useful for identifying large shifts in a firm's momentum exposures and avoiding crashes, and here we show a similar story for 10-K text information.

It is noteworthy that management's disclosures provide the most predictive information for exposure to investment risk. Of all the portfolios considered, this factor has the most to do with management's own choices. The investment factor is constructed as a spread portfolio between firms with conservative (low) and aggressive (high) investment ratios. Fama and French (2015) measure this using the firm's asset growth rate, which is exactly the type of financial ratio management's decisions can directly affect. Our results suggest that firm managers have forward-looking information about their future investment decisions, that they reveal it in 10-K disclosures, and it has significant power for understanding their future investment-risk exposure. This fact has clear value for investors viewing investment risk as an important source of returns' first and second conditional moments.

Turning to momentum, we see very poor performance in lagged beta. As mentioned, this is in comparison to a benchmark of the firm's average beta. It turns out that these average

momentum betas are not very disperse amongst stocks—over a longer term, firms have a similar exposure to momentum, in keeping with that portfolio's great deal of turnover. Hence, the benchmark for all stocks is relatively similar, and in part it is the excessive cross-sectional variation of lagged momentum betas that contributes to the negative R^2. In turn, this excessive volatility does not forecast anything, and therefore must come from unforecastable shocks. One might have taken this to suggest that expected momentum betas are not time-varying—but the text-based forecasts argue against this. There is year $t-1$ information that usefully predicts future momentum exposure: that is, its conditional expectation is dynamic. Management's choice of topics and their context reveal that this forward-looking information exists.

Panel B of Table 1 reports results for nontradable factors. Across the board, lagged beta provides negative forecasting power. Meanwhile, text-based forecasts both provide moderately strong forecast performance for exchange-rate beta, and more mild performance for credit-spread and term-spread betas. Overall, the out-of-sample R^2 for macroeconomic factors are lower than for portfolio factors, consistent with the well-known observation that nontradable factors can be somewhat weak (e.g. Giglio and Xiu, 2021). Nevertheless, forward-looking text information uniformly provides better predictions than backward-looking market information.

4.2 Marginal forecast significance

To investigate whether or not the information in lagged betas and text is complementary, now we estimate forecast combinations. If one forecast is statistically significant while controlling for other forecasts, this is evidence that the former provides important predictive content. Since we are running these regressions over the 2015–2022 sample, the next set of results involve an in-sample aspect. But note that the construction of the forecasts—in par-

ticular, the text-extraction parameters—continue to be estimated on the 2006–2014 sample, to protect against possible overfitting. Moreover, the regressions we run have at most three regressors and hence there are no high-dimensionality concerns. The R^2s reported in Table 2 are not directly comparable to the out-of-sample R^2s reported in Table 1 because the former uses the standard in-sample R^2 formula.

Panel A of Table 2 reports the market-beta results for several combinations of the context, lagged-beta, and topic forecasts. The first three columns essentially provide Mincer-Zarnowitz regressions of each forecast. For the joint null hypothesis that the intercept equals 0 and the slope equals 1, only the topic-based forecasts fail to reject. The context-based forecasts reject the Mincer-Zarnowitz null because the slope is 0.90 and significantly different than 1, not because the intercept is far from 0. Meanwhile, the lagged-beta forecasts reject the Mincer-Zarnowitz null because the slope is 0.54 and the intercept is 0.51, and both are far from the hypothesized values. In fact, these qualities of the lagged-beta forecast led the out-of-sample R^2 to be negative in Table 1. Based on the Mincer-Zarnowitz test we'd say that the topic-based forecasts are systematically unbiased—but that null hypothesis is not the last word on economic significance. Looking at the R^2s, though the topic-based forecasts are unbiased they also are the least powerful: the context-based and lagged-beta forecasts provide about double the predictive content, proving their value. To ease exposition, from here on we will often drop the word "forecasts" and discuss the estimates for context, lagged beta, and topic—hopefully there is no confusion.

Column 4 shows that context remains strongly significant (at the 0.1% level) even when controlling for lagged beta. The context slope drops from 0.90 to 0.53 and the lagged-beta slope drops from 0.54 to 0.41. Hence, context appears to contai some backward-looking information that is captured in lagged beta—controlling for the latter suggests that about 60% ($\approx 0.53/0.90$) of the context information is forward-looking. Column 6 reports a similar finding for topic. Column 5 includes just the two text-based forecasts, and both remain

Table 2: Marginal forecast significance

Notes– From regressions over 2015–2022 of future beta on out-of-sample forecasts coming from contexts, lagged betas, or topics. Standard errors are clustered by firm, and t-statistics in parentheses. We denote 5% significance by *, 1% significance by **, and 0.1% significance by ***. The number of observations for each panel is more than 10,300.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: Market			
Context	0.903***			0.527***	0.781***		0.454***
	(34.03)			(23.26)	(23.90)		(17.93)
Lagged beta		0.542***		0.410***		0.474***	0.406***
		(41.85)		(28.72)		(34.82)	(28.24)
Topic			1.011***		0.284***	0.532***	0.179***
			(21.76)		(5.34)	(17.00)	(5.12)
intercept	0.0464	0.507***	-0.0852	0.0301	-0.147**	-0.0506	-0.0915**
	(1.47)	(31.25)	(-1.54)	(1.45)	(-2.98)	(-1.61)	(-2.89)
R^2	0.217	0.286	0.126	0.343	0.223	0.316	0.345
				Panel B: Size			
Context	0.830***			0.532***	0.701***		0.450***
	(37.82)			(28.85)	(25.08)		(20.77)
Lagged beta		0.517***		0.367***		0.437***	0.360***
		(47.15)		(33.66)		(39.52)	(32.79)
Topic			1.049***		0.339***	0.609***	0.231***
			(26.73)		(7.49)	(22.31)	(7.20)
intercept	-0.181***	0.481***	-0.481***	-0.111***	-0.470***	-0.281***	-0.309***
	(-5.68)	(33.37)	(-8.68)	(-4.90)	(-9.75)	(-8.16)	(-9.04)
R^2	0.212	0.236	0.127	0.304	0.220	0.274	0.307
				Panel C: Value			
Context	0.973***			0.696***	0.563***		0.403***
	(48.48)			(39.27)	(14.30)		(12.27)
Lagged Beta		0.468***		0.272***		0.280***	0.258***
		(52.26)		(30.23)		(30.93)	(28.85)
Topic			1.169***		0.570***	0.829***	0.428***
			(46.37)		(12.48)	(39.69)	(11.41)
intercept	-0.626***	0.0902***	-0.777***	-0.433***	-0.750***	-0.537***	-0.537***
	(-34.04)	(10.63)	(-35.28)	(-28.21)	(-36.48)	(-29.85)	(-30.92)
R^2	0.266	0.223	0.260	0.320	0.281	0.318	0.328

continued on next page

Table 2 – continued from previous

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel D: Investment			
Context	0.752***			0.424***	0.812***		0.471***
	(17.62)			(13.21)	(12.55)		(11.28)
Lagged beta		0.439***		0.381***		0.409***	0.381***
		(41.21)		(32.40)		(32.71)	(32.98)
Topic			0.821***		-0.112	0.426***	-0.0882
			(8.38)		(-0.84)	(6.47)	(-1.02)
intercept	-0.146***	-0.172***	-0.120***	-0.0485***	-0.165***	-0.0318*	-0.0638***
	(-10.81)	(-15.41)	(-4.59)	(-4.96)	(-6.44)	(-2.03)	(-3.92)
R^2	0.102	0.219	0.056	0.247	0.102	0.233	0.247
				Panel E: Profitability			
Context	0.631***			0.301***	0.525***		0.254***
	(21.98)			(18.80)	(14.55)		(11.40)
Lagged beta		0.511***		0.463***		0.480***	0.462***
		(48.19)		(46.23)		(47.50)	(46.33)
Topic			0.763***		0.222***	0.347***	0.101***
			(18.39)		(4.68)	(16.59)	(3.52)
intercept	0.729***	-0.115***	0.964***	0.392***	0.934***	0.483***	0.487***
	(16.32)	(-14.52)	(14.37)	(15.45)	(14.96)	(13.90)	(14.23)
R^2	0.096	0.284	0.065	0.303	0.099	0.296	0.304
				Panel F: Momentum			
Context	0.620***			0.671***	0.558***		0.598***
	(13.07)			(14.31)	(6.03)		(6.37)
Lagged beta		-0.0235*		-0.0506***		-0.0453***	-0.0509***
		(-2.18)		(-4.94)		(-4.32)	(-4.95)
Topic			0.531***		0.0718	0.572***	0.0853
			(12.32)		(0.86)	(13.42)	(1.00)
intercept	-0.170***	-0.168***	-0.170***	-0.170***	-0.170***	-0.170***	-0.170***
	(-23.91)	(-21.89)	(-23.63)	(-22.93)	(-23.94)	(-22.73)	(-22.95)
R^2	0.016	0.001	0.012	0.019	0.016	0.015	0.019

continued on next page

Table 2 – continued from previous

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel G: Exchange Rate			
Context	0.595***			0.413***	0.398***		0.277***
	(23.27)			(19.99)	(9.81)		(8.69)
Lagged Beta		0.337***		0.293***		0.298***	0.290***
		(37.99)		(34.42)		(34.06)	(33.75)
Topic			0.734***		0.328***	0.506***	0.230***
			(22.28)		(6.72)	(19.20)	(5.86)
intercept	-0.00205***	-0.00757***	-0.000116	-0.00222***	-0.000230	-0.000889*	-0.000947**
	(-5.88)	(-74.01)	(-0.26)	(-8.14)	(-0.55)	(-2.57)	(-2.82)
R^2	0.064	0.119	0.058	0.148	0.069	0.145	0.150
				Panel H: Credit Spread			
Context	1.316***			0.991***	0.802***		0.594***
	(23.51)			(20.92)	(9.98)		(8.82)
Lagged beta		0.292***		0.234***		0.237***	0.223***
		(28.71)		(25.36)		(24.74)	(24.20)
Topic			2.084***		1.242***	1.590***	0.996***
			(27.39)		(10.98)	(24.39)	(10.34)
intercept	0.00199	0.0416***	-0.0322***	0.00125	-0.0304***	-0.0256***	-0.0247***
	(0.79)	(46.43)	(-9.73)	(0.60)	(-9.62)	(-9.39)	(-9.25)
R^2	0.059	0.070	0.059	0.101	0.071	0.102	0.108
				Panel I: Term Spread			
Context	1.116***			0.659***	0.812***		0.466***
	(26.44)			(19.46)	(12.76)		(10.26)
Lagged beta		0.459***		0.371***		0.392***	0.361***
		(36.05)		(28.20)		(27.66)	(26.58)
Topic			1.578***		0.722***	0.925***	0.485***
			(25.45)		(7.85)	(18.31)	(7.16)
intercept	-0.0133***	0.0349***	-0.0503***	-0.00995***	-0.0468***	-0.0332***	-0.0325***
	(-3.97)	(26.27)	(-10.44)	(-4.24)	(-10.60)	(-9.78)	(-9.86)
R^2	0.101	0.157	0.081	0.187	0.111	0.182	0.191

significant, with the topic slope falling by about 75% but the context slope falling by only 15%. Therefore we find two important results: first, text-based forecasts provide valuable forward-looking information even while controlling for backward-looking market information; second, topic and context are complementary to each other.

The final column reports that all three forecasts are statistically significant when evaluated together—the clearest sign yet that topics, contexts, and lagged betas provide distinct predictive content. Since we are regressing onto fitted values, it is somewhat natural that the slope coefficients for the eventual forecast (what we'll call the fitted value in column 7) almost sum up to 1. Roughly speaking, about two-fifths of the eventual forecast comes from both context and lagged betas, and one-fifth from topic.[6] The R^2 in column 7 shows that these distinct information sources explain 34.5% of the variation in market betas, an economically significant amount as we'll later quantify.

Panel B shows the results for size betas. Once again, all three forecasts enter the combination significantly and the R^2 is an economically-significant 30.7%. As for market beta, from the slope coefficients we see that context is the biggest part of the eventual forecast, with lagged beta and topic following. Panel C shows that all three forecasts are significant, with the two text-based forecasts providing similar amounts to the eventual forecast whose R^2 is 32.8%. Jumping to profitability, panel E column 7 shows a similar R^2 of 30.4%, but this time lagged beta is responsible for most of the eventual forecast, with context and topic coming next in that order.

As highlighted in the previous section, results for investment and momentum betas show differences. For investment betas, panel D column 4 shows us that topic information is basically nested in context information. That is, in the bivariate regression we see that the context is strongly significant while topic is not. While both text-based forecasts are

[6] If anything this is generous to topic, as the context and lagged-beta R^2s are similar implying their fitted values are about equally variable. A formal variance decomposition would need to take this into account—we opt to instead speak in rough terms.

significant while controlling for backward-looking lagged-beta information, we see evidence that context is extracting the relevant predictive information for future investment beta. This shines through in column 7 where context and lagged beta are strongly significant, but topic is not.

Meanwhile, momentum betas provide more nuanced differences. Notably, while lagged beta is significant in Panel F column 2, it enters with a negative sign—this holds true even in column 7. Meanwhile, topic is significant on its own (column 3) but is driven out by context (columns 5 and 7). Notably, the context slopes barely change between columns 1, 5, and 7. This indicates that context uniquely contains the 10-K's forward-looking predictive information, and it is virtually uncorrelated with lagged beta's backward-looking information.

Panels G–I of Table 2 show that lagged beta and the text-based forecasts are all statistically significant. For these macroeconomic risks, topic and context are uniformly complementary. A modest amount of their informativeness is correlated with lagged beta, as the slope coefficients vary with the latter's inclusion. But they consistently are important contributors to the eventual forecast in column 7, where R^2s range respectably from 11% to 19%. Therefore, text-based information is useful for predicting future nontradeable-factor risk exposures.

A broad takeaway from Table 2 is that context and lagged beta are always significant, complementary predictors of future risk exposures.[7] Most often, but not always, topic is significant as well. For most factor risks, backward-looking market information is enhanced by both types of forward-looking text information we extracted.

[7]We have also calculated standard errors clustered by year—while this serves to lower the t-statistics nearly every qualitative conclusion remains. But we note that issues with small numbers of clusters (we have 7 when clustering by year) are well known.

Table 3: Marginal forecast significance, robustness to firm-characteristic forecasts

Notes– From regressions over 2015–2022 of future beta on out-of-sample forecasts coming from contexts, firm characteristics, or topics. Standard errors are clustered by firm, and t-statistics in parentheses. We denote 5% significance by *, 1% significance by **, and 0.1% significance by ***. The number of observations for each column is more than 10,300.

	Market	Size	Value	Investment	Profitability	Momentum
Context	0.592***	0.491***	0.389***	0.694***	0.331***	0.354***
	(18.84)	(17.41)	(10.13)	(11.05)	(9.42)	(4.46)
Characteristic	0.534***	0.563***	0.535***	0.526***	0.478***	1.192***
	(19.27)	(14.84)	(12.78)	(4.89)	(12.49)	(24.32)
Topic	0.248***	0.294***	0.518***	-0.230	0.195***	-0.225**
	(5.52)	(7.43)	(11.95)	(-1.75)	(4.49)	(-3.03)
intercept	-0.514***	-0.892***	-0.989***	-0.0581	1.392***	-0.173***
	(-11.07)	(-17.25)	(-37.27)	(-1.73)	(19.51)	(-28.66)
R^2	0.286	0.292	0.322	0.137	0.139	0.121
	Exchange Rate	Credit Spread	Term Spread			
Context	0.331***	0.594***	0.651***			
	(8.37)	(7.23)	(9.86)			
Characteristic	0.320***	1.197***	0.701***			
	(9.33)	(14.14)	(10.88)			
Topic	0.282***	1.143***	0.685***			
	(5.96)	(10.65)	(7.73)			
intercept	0.00264***	-0.0701***	-0.0871***			
	(5.10)	(-17.11)	(-15.92)			
R^2	0.080	0.095	0.132			

4.3 Firm characteristic information

So far we have used lagged beta to capture structured, non-textual information that agents have available to forecast future risk exposure. Yet Kelly et al. (2021) gives us reason to believe that other firm characteristics may also be valuable.

Table 3 shows that our main results continue to hold with respect to this larger structured information source. Comparing to column 7 of Table 2, we see the same pattern of topic and context statistical significance. In fact, in numerous cases the context and topic coefficient estimates are negligibly changed.

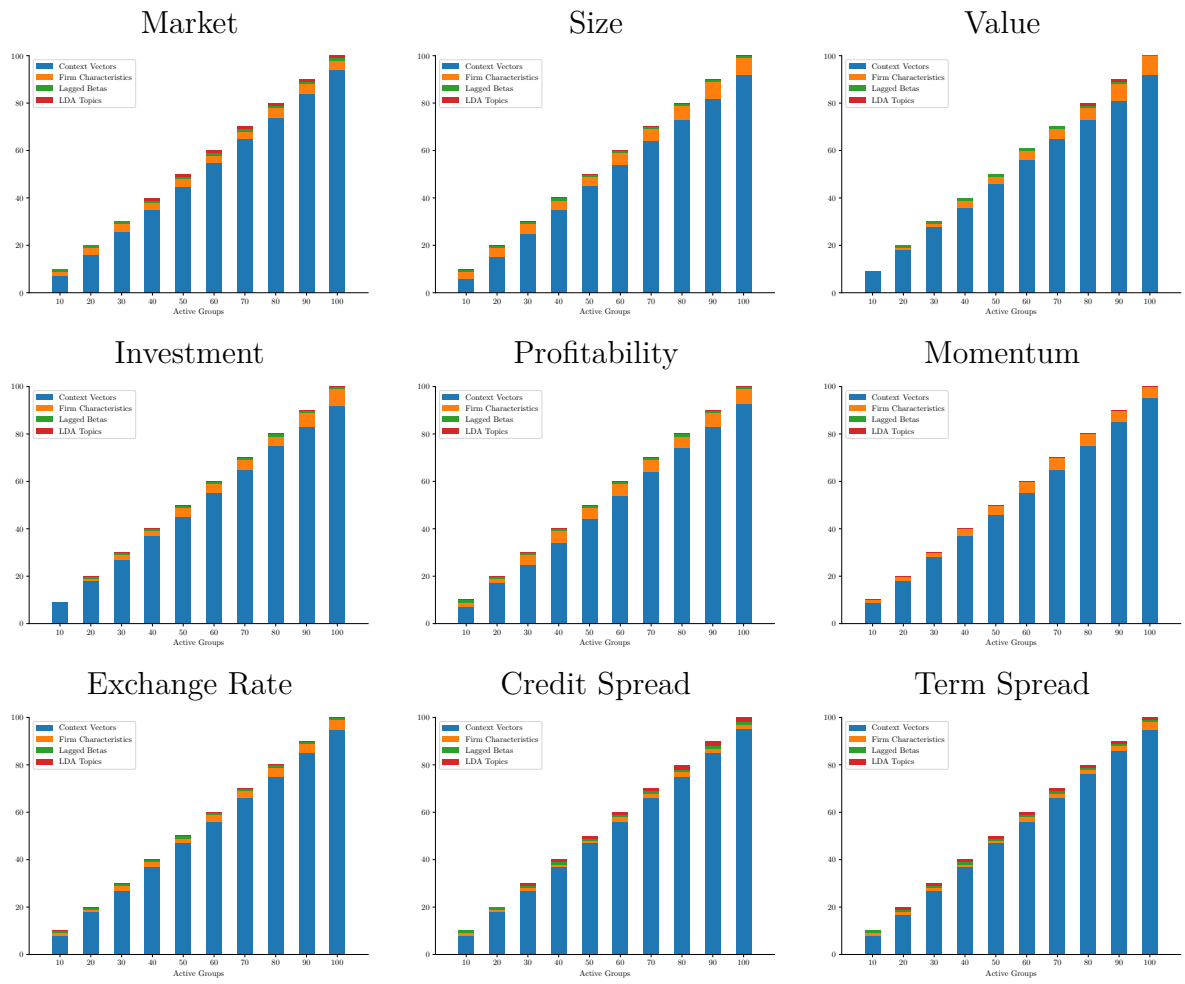

Figure 1: Information selected by group LASSO

Notes– How many groups of each information type are chosen. The types are context vectors, firm characteristics, lagged beta, and LDA topics. Going horizontally is the number of groups allowed to be active, from 10 to 100 by 10. The vertical axis measures how many groups of each type are chosen, as the height of the bar of each color.

4.4 Statistical selection of information

We have explored two types of structured information (lagged beta and firm characteristics) and two types of textual information (topic and context). We have found that these information sources provide forecasts that are often complementary to one another, using forecast-combination regressions. We now provide another means of understanding the value of these different information sets. We use the group LASSO with varying levels of regularization and ask: what type of information is chosen? If we allow the LASSO to only select 10 information groups, which are the ones it picks up? How does this change as we allow more groups to be selected? Answering these questions tell us which types of information are selected by purely statistical methods.

The striking conclusion from Figure 1 is that context is the earliest and most prevalent information statistically selected. While it is true there is only one lagged beta variable to be chosen, there are dozens of topics and firm characteristics that could be chosen. Uniformly, context variables constitute the majority of groups selected for every factor and every total number of active groups. Visually this is apparent by the prevalence of blue in Figure 1, denoting the number of context groups chosen.

Let us consider in detail the top-right panel for market beta. When ten groups are allowed active, seven of them are context variables—two firm characteristics and the lagged betas compose the remaining three. Only when forty groups are allowed active does a single topic get selected—no more are ever chosen. By time one hundred groups are allowed active, only five firm characteristics are included, alongside lagged beta and the lone topic. The remaining 93 active groups are context variables. For every factor considered, context information comprises more than ninety percent of what the group LASSO eventually selects.

Scanning across the other panels, it is only the topic, firm characteristic, and lagged beta selection that really varies across factor betas. Topic groups are never selected for size,

investment, profitability, and momentum betas. For value and investment betas, the first ten groups selected are purely context.

Turning to nontradable factors in the last row, the prominence of context information is unchanged. Exchange-rate betas never choose topic information, term-spread betas choose a single topic, and credit-spread betas choose two topics. Despite there being thirty-six firm characteristics available, never more than handful are selected.

There are a few ways to interpret Figure 1 in light of results we've seen before. Recall that forecasts based on topics and firm characteristics are very often statistically significant in forecast-combination regressions. Since the group LASSO chooses so few of them, it is natural to suppose that those valuable topic and firm-characteristics forecasts were driven by a small number of topics and firm characteristics, respectively. Meanwhile, it is perhaps the case that context variables have a more complicated relationship to one another, such that a greater number of them are consistently chosen as we loosen the regularization. One could say that the value of topic and firm-characteristic information tends to rest in just a few variables. At the same time, the broad pattern of selection indicates that when LASSO can choose a new variable to include in beta forecasting, it almost always chooses context information. This suggests that context information is quite rich.

5 Interpretation

In this section we present simple reduced-form regressions which indicate how 10-K contexts are forward-looking via managerial decisions or business-condition predictions. We then calculate the economic significance of this textual information using a Bayesian certainty equivalent.

5.1 Context and future firm characteristics

We do not believe that management simply *writes* into existence its future aggregate beta. Instead, our view is that management's risk disclosures are revealing their information about future managerial decisions or business conditions, and that *these* are signicant drivers of the realized covariance the firm experiences. With this in mind, we run simple reduced-form regressions to suggest why the text information is predictive. For simplicity and since it was statistically significant for all them, we focus on context information and the five Fama and French (2015) factor betas; as above, we consider firm characteristics from Kelly et al. (2021).

According to classic corporate theory, a firm increases its measured (levered) market beta by increasing its leverage. Perhaps this is one of the few financial-managerial decisions that drives the CAPM. Table 4's first row provides supporting evidence—first look at the column labeled Market. There we see that book leverage during year t is significantly and positively related to the realized market beta during year t, with a t-statistic of 4.93, in a *contemporaneous* regression. Hence, a firm's increased leverage explains a part of the increase in market beta. And the first column of that row (labeled Context) shows that context is *predictive* of future leverage. From managerial risk disclosures we get forward-looking information of the firm's future indebtedness, which of course carries significant effects for investors.

Market capitalization is not something managers can directly control, yet it should (at some point) be connected to firm performance. As a measure of performance, earnings over assets gives a sense of how productively the firm is employing its capital, and this productivity should result in firm growth. The second row shows a strongly negative relationship between between earnings/assets and size beta ($t = -15.09$), which is consistent with this idea because size exposure is smaller for bigger firms. Meanwhile, context strongly predicts

Table 4: Context, firm characteristics, and risk exposures

Notes– All regressions have more than 6,200 observations and employ standard errors clustered by firm. In the *Predictive* Context column we report the point estimate and t-statistic of forecasting the firm characteristic named in that row using the context-based beta forecast for the factor whose column in the *Contemporaneous* panel is not blank. For example, the *Predictive* Context numbers next to Book/Market tell us that the context-based-value-beta-forecast in year $t-1$ predicts book-to-market in year t with a point estimate of 0.392 and t-statistic of 15.74. The *Contemporaneous* panel reports regressions of the year t realized beta of the factor in the column label (e.g. Value) on the year t characteristic (e.g. Book/Market). All predictive regressions are univariate; the Profitability contemporaneous regression is bivariate while the remaining are univariate. Standard errors are clusted by firm.

	Predictive	*Contemporaneous*				
	Context	Market	Size	Value	Investment	Profitability
Book Leverage	0.294	0.003				
	(2.21)	(4.93)				
Earnings/Assets	−0.059		−1.959			
	(−8.12)		(−15.09)			
Book/Market	0.392			0.335		
	(15.74)			(3.11)		
Asset growth	−0.095				−0.374	
	(−8.90)				(−2.33)	
Profitability	1.910					0.0002
	(1.67)					(2.76)
Earnings/Assets	0.065					3.576
	(11.05)					(15.52)
R^2		0.002	0.107	0.092	0.029	0.131

future earnings/assets ($t = -8.12$). Hence, size exposure is somewhat explained by its contemporaneous asset productivity, which in turn is predicted by risk-disclosure text.

The story for value and investment exposures is more direct. The book-to-market ratio explains value beta ($t = 3.11$) and asset growth explains investment beta ($t = -2.33$) with the expected signs. And the context of firm disclosures significantly forecast the future book-to-market ($t = 15.74$) and asset growth ($t = -8.90$). In the case of the former, we think this has more to do with management's inside information about the firm's future valuation, since they cannot directly control stock prices. In the case of the latter, context may reveal

management's investment plans which manifests as as asset growth.

Context does not forecast future (gross) profitability as strongly as the other characteristics we've considered ($t = 1.67$), though profitability is contemporaneously related to profitability-risk exposure ($t = 2.76$) as one expects. Because of this we considered earnings/assets as well, because it too is a firm productivity measure and it worked significantly in the size-exposure case.[8] We see that context robustly forecasts earnings/assets ($t = 11.05$) which itself is contemporaneously related to profitability exposure ($t = 15.52$).[9] Therefore risk disclosures are more informative of future bottom-line profit measures, and these relate strongly to aggregate risk exposure.

The straightforward story presented here is that 10-K disclosures are forward looking because managers have useful information about future business conditions or future corporate plans. Future research might explore related questions, such as: which type of information is more precise? (we might expect the latter); or, which type is more useful to investors? (does a firm's business-conditions forecast spillover to information about related firms?); or, does management have incentives to reveal some plans but not others?

5.2 Economic significance

We measure the economic significance of risk-disclosure information by calculating a certainty equivalent, answering the question "What value do investors ascribe to this text data?" A modestly challenging aspect to this question is that the reduction in uncertainty occurs for *parameters* employed by the investor. That is, a typical certainty-equivalent calculation takes a risky payoff and calculates the riskless payoff yielding the same utility. But in our setup it is uncertainty *about the perceived risk* that is at play. Therefore, we adopt a Bayesian

[8]Hence the contemporaneous regression for profitability beta is run with both profitability and earnings/assets as regressors.

[9]Context predicts earnings/assets differently in the size and profitability cases—the estimates are -0.059 and 0.065, respectively. This is because the forecast of size beta is a different linear combination of context variables than the forecast of profitability beta.

perspective, to jointly capture both the uncertainty about the payoff (unaffected by the text information) and the uncertainty about the model parameters (what the text information *does* affect).

To ease exposition, we put the details in appendix C and here give an overview of the problem. Agents have mean-variance preferences and use an APT-consistent factor model, à la Stambaugh (1983), to form expectations of future returns—our twist is that the agent is Bayesian and uncertain of the β parameters determining conditional expected returns. Employing conjugate priors, the typical mean-variance solution obtains, albeit with the agent viewing returns as given by a multivariate non-central t-distribution. Intuitively, the perceived variance of returns is increasing in the variance of investors' prior on β—hence Bayesian mean-variance investors dislike parameter uncertainty.

Let $MV^*(r_f, \tilde{\kappa})$ denote the agent's optimized utility as a function of the risk-free rate r_f and hyperparameter $\tilde{\kappa}$. The certainty equivalent calculation solves for s in

$$MV^*(r_f, \tilde{\kappa}_1) = MV^*(r_f + s, \tilde{\kappa}_2), \quad \text{with } \tilde{\kappa}_1 > \tilde{\kappa}_2, s \geq 0, \tag{10}$$

where $\tilde{\kappa}_1, \tilde{\kappa}_2$ govern prior-distribution precision (of the beta parameters) in the case of excluding text information ($\tilde{\kappa}_2$) or including text information ($\tilde{\kappa}_1$). Then s is the non-negative increase in the risk-free rate that compensates the investor for excluding text information. That is, s is the Bayesian certainty-equivalent value of that information.

Using Fama and French (2015) data for our sample period, a moderately high risk aversion of $\gamma = 10$, and simplifying assumptions detailed in the appendix, we solve (10) for s. We find that the certainty equivalent is 28 basis points per annum. With the average risk-free rate around 14 basis points per annum, this says that investors require a risk-free rate twice as high to compensate for losing valuable information.[10] Even when simply judging the certainty

[10]Of course, risk-free rates at the end of our sample are much higher than this average. We have found that a higher risk-free rate serves to increase the certainty equivalent calculation, all else equal, so believe

equivalent at face value, we view 28 basis points as a reasonable yet economically-significant value of the text information we've extracted.

6 Conclusion

This paper uses the risk disclosure text of 10-Ks to predict future aggregate risk exposures. We find notable success, even controlling for structured (firm characterstic) data that agents already have in hand. As an accurate measure of beta is valuable to any agent using a factor model to understand expected returns, cost of capital, or covariance—this shows that text provides significant information for an important problem. Furthermore, we expand upon the value of different text models. Both topics and contexts provide significant information to forecasting future betas, and this information is distinct.

our 28 basis point conclusion is conservative.

References

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003) "Latent dirichlet allocation," *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022.

Buehlmaier, Matthias MM and Toni M Whited (2018) "Are financial constraints priced? Evidence from textual analysis," *The Review of Financial Studies*, Vol. 31, No. 7, pp. 2693–2728.

Bybee, Leland, Bryan Kelly, Asaf Manela, and Dacheng Xiu (Forthcoming) "Business News and Business Cycles," *Journal of Finance*.

Campbell, John L, Hsinchun Chen, Dan S Dhaliwal, Hsin-min Lu, and Logan B Steele (2014) "The information content of mandatory risk factor disclosures in corporate filings," *Review of Accounting Studies*, Vol. 19, pp. 396–455.

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen (2020) "Lazy prices," *The Journal of Finance*, Vol. 75, No. 3, pp. 1371–1415.

Cong, William, Tangyuan Liang, and Xiao Zhang (2020) "Textual Factors: A Scalable, Interpretable, and Data-driven Approach to Analyzing Unstructured Information,"Technical report, Cornell University.

Daniel, Kent and Tobias J. Moskowitz (2016) "Momentum crashes," *Journal of Financial Economics*, Vol. 122, No. 2, pp. 221–247.

Dyer, Travis, Mark Lang, and Lorien Stice-Lawrence (2017) "The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation," *Journal of Accounting and Economics*, Vol. 64, No. 2-3, pp. 221–245.

Fama, Eugene F and Kenneth R French (2015) "A five-factor asset pricing model," *Journal of financial economics*, Vol. 116, No. 1, pp. 1–22.

Ferson, Wayne E. and Campbell Harvey (1999) "Conditioning Variables and the Cross Section of Stock Returns," *Journal of Finance*, Vol. 54, No. 4, pp. 1325–1360.

Florackis, Chris, Christodoulos Louca, Roni Michaely, and Michael Weber (2022) "Cybersecurity Risk," *The Review of Financial Studies*, Vol. 36, No. 1, pp. 351–407, URL: `https://doi.org/10.1093/rfs/hhac024`, DOI: 10.1093/rfs/hhac024.

Garcia, Diego and Øyvind Norli (2012) "Geographic dispersion and stock returns," *Journal of Financial Economics*, Vol. 106, No. 3, pp. 547–565.

Gavish, Matan and David L Donoho (2014) "The optimal hard threshold for singular values is 4/*sqrt*{3}," *IEEE Transactions on Information Theory*, Vol. 60, No. 8, pp. 5040–5053.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019) "Text as Data," *Journal of Economic Literature*, Vol. 57, No. 3, pp. 535–574.

Giglio, Stefano and Dacheng Xiu (2021) "Asset Pricing with Omitted Factors," *Journal of Political Economy*, Vol. 129, No. 7, pp. 1947 – 1990.

Hoberg, Gerard and Vojislav Maksimovic (2015) "Redefining financial constraints: A text-based analysis," *The Review of Financial Studies*, Vol. 28, No. 5, pp. 1312–1352.

Hoffman, Matthew, Francis Bach, and David Blei (2010) "Online learning for latent dirichlet allocation," *advances in neural information processing systems*, Vol. 23.

Jegadeesh, Narasimhan and Sheridan Titman (1993) "Returns to buying winners and selling losers: Implications for stock market efficiency," *The Journal of finance*, Vol. 48, No. 1, pp. 65–91.

Kelly, Bryan T, Tobias J Moskowitz, and Seth Pruitt (2021) "Understanding momentum and reversal," *Journal of financial economics*, Vol. 140, No. 3, pp. 726–743.

Kelly, Bryan T., Seth Pruitt, and Yinan Su (2019) "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics*, Vol. 134, No. 3, pp. 501–524, URL: `https://www.sciencedirect.com/science/article/pii/S0304405X19301151`, DOI: https://doi.org/10.1016/j.jfineco.2019.05.001.

Khodak, Mikhail, Nikunj Saunshi, Yingyu Liang, Tengyu Ma, Brandon M Stewart, and Sanjeev Arora (2018) "A La Carte Embedding: Cheap but Effective Induction of Semantic Feature Vectors," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12–22.

Lopez-Lira, Alejandro (2023) "Risk Factors that Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns,"Technical report, University of Florida.

Loughran, Tim and Bill McDonald (2011) "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of finance*, Vol. 66, No. 1, pp. 35–65.

Mikolov, Tomas, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin (2018) "Advances in Pre-Training Distributed Word Representations," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Stambaugh, Robert F. (1983) "Arbitrage pricing with information," *Journal of Financial Economics*, Vol. 12, No. 3, pp. 357–369, URL: `https://ideas.repec.org/a/eee/jfinec/v12y1983i3p357-369.html`.

Tetlock, Paul C, Maytal Saar-Tsechansky, and Sofus Macskassy (2008) "More than words: Quantifying language to measure firms' fundamentals," *The journal of finance*, Vol. 63, No. 3, pp. 1437–1467.

A Text data

This appendix summarizes our procedure for extracting and processing Item 1A sections from Form 10-K documents. We first discard plain text 10-K filings from the sample. We discard plain text filings because they make a negligible fraction of the sample, occur only in the first few years of the paper's sample, and have difficult to identity Item 1A sections. Next, we use two programs to extract Item 1A from the sample of HTML 10-K documents.

The first program we use to extract Item 1A sections from Form 10-K filings has three subroutines. The first subroutine identifies the table of contents at the beginning of a Form 10-K. The routine finds a document's table of contents by searching for anchor tags within the document's HTML whose text is similar to the phrase "table of contents." The routine then locates the HTML element with an id attribute matching the id given by the document's links to its table of contents. The second subroutine parses a Form 10K document's table of contents into individual sections. The subroutine first finds the beginning of the Item 1A section by searching for a section title similar to "Item 1A. Risk Factors." Then, the subroutine finds the section immediately after the document's risk factors section. Next, the subroutine finds links to the beginning of each section's contents within both sections' table of contents entries. The third subroutine extracts the Item 1A text from the Form 10-K document using the table of contents links to the Item 1A section and the section immediately after the Item 1A section. The subroutine extracts all of the text between the link to the Item 1A section and the link to the next section.

The second program we use to extract Item 1A sections from Form 10-K filings has three subroutines. The first subroutine directly searches the text of a Form 10-K section for phrases similar to "Item 1A" at the beginning of an HTML tag. When the subroutine finds multiple matches for this condition, the subroutine assumes the last match is the beginning of the Item 1A section. This assumption works well because practically all other uses of the phrase "Item 1A" within a Form 10-K occur in the Management Discussion and Analysis portion of the Form 10-K, which is earlier in a Form 10-K than the risk factors section. The second subroutine searches for the occurrence of the phrase "Item" at the beginning of an HTML tag after the start of the Item 1A section's text. This second subroutine usually matches several different positions in the text. We assume the first match is the conclusion of the Item 1A section and the beginning of the filing's next section. The third subroutine extracts the Item 1A section from the filing using the section beginning and section conclusion identified by the previous two routines. The third routine also discards sections with less than ten words. The purpose of this last condition is to discard Item 1A sections where the text is "Not required" or another similar phrase.

We perform additional processing after extracting Item 1A sections from Form 10-K documents. The purpose of this further processing is to standardize the documents' words. We first remove all HTML-related tags from the texts. We also remove anchor tags' text because practically all links in risk factors sections are table of contents links. We convert Unicode characters to their closest ASCII equivalents. We remove page numbers. Last, we remove punctuation and other non-alphanumeric characters.

B A Matrix Estimation

This section describes our estimate of the matrix A from equation (3). We use a methodology similar to the one proposed in Khodak et al. (2018) with an additional step that reduces the context vectors' dimension. The matrix A is the solution to the multivariate least squares regression problem

$$V = UA + E \tag{11}$$

$$V \in R^{W\times K} \quad U \in R^{W\times D} \quad A \in R^{D\times K} \quad E \in R^{W\times D}. \tag{12}$$

The matrix V contains vector representations for the W words in the paper's vocabulary taken from the FastText embedding after a dimension reduction step. Let $\tilde{V} \in R^{W\times D}$ be the matrix where each row is the FastText vector representation of some word occurring in the Item 1A sample. Let the singular value decomposition of the matrix $\tilde{V}$ be

$$\tilde{V} = FGH^T \tag{13}$$

where we use F, G, H to represent the singular value decomposition of $\tilde{V}$ to avoid collisions with the paper's other notation. The matrix V is the seven largest left singular vectors of $\tilde{V}$ scaled by the seven largest singular values of $\tilde{V}$:

$$V = \sum_{j=1,\ldots,7} F_{:,j}G_{j,j}. \tag{14}$$

We used cross-validation to select the number of left singular vectors we use to construct V.

The matrix U contains context vectors for the W words in the paper's vocabulary. Row w of the matrix U is the vector $u_w \in R^D$ with the definition

$$u_w = \frac{1}{|C^w|} \sum_{c^j_{i,t}\in C^w} \sum_{w^k_{i,t}} v^{w^k_{i,t}}. \tag{15}$$

The set C^w contains the context for every occurrence of the word w in our sample of Item 1A texts; i.e., $C^w = \cup_{i,t} C^w_{i,t}$. Note that C^w does not contain duplicate contexts because C^w is a set. So, each context occurs only once in the set C^w even though the context may belong to several $C^w_{i,t}$ sets.

C Bayesian Certainty Equivalent

Consider an investor with mean-variance preferences and information—they solve the problem

$$\max_{w\in\mathbb{R}^N} \mathbb{E}\left(\left[\begin{array}{c} 1-w'\iota \\ w \end{array}\right]' \left[\begin{array}{c} r_f \\ r \end{array}\right] \middle| r_f, \mathcal{I}\right) - \frac{\gamma}{2}\mathbb{V}\left(\left[\begin{array}{c} 1-w'\iota \\ w \end{array}\right]' \left[\begin{array}{c} r_f \\ r \end{array}\right] \middle| r_f, \mathcal{I}\right) \tag{16}$$

where risk aversion $\gamma > 0$, risky returns $r \in \mathbb{R}^N$, and risk-free $r_f \geq 0$. We are emphasizing that $\{r_f, \mathcal{I}\}$ is the investor's information set, because later we will alter both on our way

to calculating a certainty equivalent accounting for parameter uncertainty—we call this a *Bayesian certainty equivalent*.

We focus solely on uncertainty about the parameters $\beta \in \mathbb{R}^{N\times K}$ that appear in a factor model, following standard APT assumptions as in Stambaugh (1983). That is, agents believe returns' mean is given by $\beta\lambda$ and covariance is given by $\beta\Delta_F\beta'+\Delta_\epsilon$. For simplicity we assume $\lambda \in \mathbb{R}^K, \Delta_F \in \mathbb{R}^{K\times K}, \Delta_\epsilon \in \mathbb{R}^{N\times N}$ are known ($\Delta_F, \Delta_\epsilon$ are positive definite with the latter diagonal).

We assume $r|\mathbb{E}(r), \mathbb{V}(r) \sim N(\mathbb{E}(r), \mathbb{V}(r))$ and therefore use the conjugate, normal-inverse-Wishart prior to express uncertainty about $\mathbb{E}(r)$ and $\mathbb{V}(r)$. A wrinkle in our formulation is that we wish to express uncertainty about β and this parameter appears *in both* the mean and variance. We have not seen this particular challenge tackled before, so to simplify analysis we represent beta uncertainty solely via the hyperparameter governing the conditional distribution of returns' mean.[11] That is, we assume (with abuse of notation) that

$$\mathbb{E}(r)|\beta_0, \kappa, \Delta_F, \Delta_\epsilon \sim N\left(\beta_0\lambda, \frac{1}{\kappa}\left[\beta_0\Delta_F\beta_0' + \Delta_\epsilon\right]\right) \tag{17}$$

$$\mathbb{V}(r)|\beta_0, \upsilon, \Delta_F, \Delta_\epsilon \sim W^{-1}\left(\beta_0\Delta_F\beta_0' + \Delta_\epsilon, \upsilon\right) \tag{18}$$

where $W^{-1}(\cdot)$ is the inverse-Wishart distribution. Note that by $\mathbb{E}(r)$ we mean $\beta\lambda$. Given these assumptions, we have

$$r \sim t_{\upsilon-N+1}\left(\beta_0\lambda, \frac{\kappa+1}{\kappa(\upsilon-N+1)}\left[\beta_0\Delta_F\beta_0' + \Delta_\epsilon\right]\right) \tag{19}$$

which is the multivarate non-central t-distribution—this is the distribution that the parameter-uncertain investor perceives as driving returns.

The usual mean-variance solution holds where the risky-asset portfolio is

$$w^* = \frac{1}{\gamma}\left[\mathbb{V}(r)\right]^{-1}\mathbb{E}(r)$$

and using this alongside the mean and variance of the multivariate t given in (19), we find that the optimized utility can be written

$$\begin{aligned}
MV^* &= (w^*)'\mathbb{E}(r) + (1-\iota'w^*)r_f - \frac{\gamma}{2}(w^*)'\mathbb{V}(r)w^* \\
&= (\frac{1}{\gamma}\left[\mathbb{V}(r)\right]^{-1}\mathbb{E}(r))'\mathbb{E}(r) + (1-\iota'\frac{1}{\gamma}\left[\mathbb{V}(r)\right]^{-1}\mathbb{E}(r))r_f - \frac{\gamma}{2}(\frac{1}{\gamma}\left[\mathbb{V}(r)\right]^{-1}\mathbb{E}(r))'\mathbb{V}(r)\frac{1}{\gamma}\left[\mathbb{V}(r)\right]^{-1}\mathbb{E}(r) \\
&= r_f + \left(\frac{1}{2}\mathbb{E}(r) - r_f\iota\right)'\left[\mathbb{V}(r)\right]^{-1}\mathbb{E}(r)\frac{1}{\gamma} \\
&= r_f + \frac{\tilde{\upsilon}\tilde{\kappa}}{\gamma}\left(\frac{1}{2}\beta_0\lambda - r_f\iota\right)'\left[\beta_0\Delta_F\beta_0' + \Delta_\epsilon\right]^{-1}\beta_0\lambda
\end{aligned} \tag{20}$$

[11] Future work can explore the viability of this assumption, in a fully-formed analysis

for $\tilde{\upsilon} \equiv \upsilon - N - 1$ and $\tilde{\kappa} \equiv \kappa(\kappa+1)^{-1}$.

Now we can simply state our Bayesian certainty equivalent calculation. Write $MV^*(r_f, \tilde{\kappa})$ because we will hold constant the other parameters in (20). We find the s solving the equation we repeat from the main text

$$MV^*(r_f, \tilde{\kappa}_1) = MV^*(r_f + s, \tilde{\kappa}_2), \quad \text{with } \tilde{\kappa}_1 > \tilde{\kappa}_2, s \geq 0 \tag{10}$$

for some given $\tilde{\kappa}_1, \tilde{\kappa}_2$. Note that $\tilde{\kappa}$ is an increasing transformation of κ: therefore their increase represent an increase in the precision of the prior distribution by (17). Note that s is a non-negative *increase* to the risk-free rate the agent receives. Intuitively, this risk-free-rate bump is positive when the smaller $\tilde{\kappa}_2$ means that the agent is facing *more* parameter uncertainty, which flows into the perceived variance of returns by (19) which the agent dislikes. Therefore we call s the *Bayesian certainty equivalent* for the increase in parameter precision (and therefore the agent's posterior return precision) implied by increasing $\tilde{\kappa}_2$ to $\tilde{\kappa}_1$.

Generally (20) requires stock-specific values owing to the presence of β_0 and Δ_ϵ. To empirically ground our calculations as simply as possible, we now assume that the $N = K$ and the assets available to the investor are the K factors themselves. Of course, literally these means that $\beta_0 = I$ and it appears on the surface that we've assumed away β_0 uncertainty entirely. However this is not an accurate depiction of what is going on in reality, because in the real world investors need to mimic the factors using the available assets, and *these* mimicking weights require β_0. Our conjecture is that a fully-formed analysis would show that β uncertainty translates into "something like" uncertainty about the factors in the large N limit, but we leave this analysis for future work. With our assumption, we can represent

$$MV^*(r_f, \tilde{\kappa}) = r_f + \tilde{\kappa}\frac{\tilde{\upsilon}}{\gamma}\left(\frac{1}{2}\lambda - r_f\iota\right)' \Delta_F^{-1}\lambda \tag{21}$$

Plugging in the mean and covariance of the Fama and French (2015) factors as well as the average r_f from in Ken French's data over our sample period, we set $\gamma = 10$ and $\tilde{\upsilon} = 20$, we find s by solving (10).