ANDRÁS FÜLÖP| ZALÁN KOCSIS

NEWS-BASED INDICES ON COUNTRY FUNDAMENTALS: DO THEY HELP EXPLAIN SOVEREIGN CREDIT SPREAD FLUCTUATIONS?

MNB WORKING PAPERS | 1

2018
FEBRUARY

NEWS-BASED INDICES ON COUNTRY FUNDAMENTALS: DO THEY HELP EXPLAIN SOVEREIGN CREDIT SPREAD FLUCTUATIONS?

MNB WORKING PAPERS | 1

2018

FEBRUARY

The views expressed are those of the authors' and do not necessarily reflect the official view of the central bank of Hungary (Magyar Nemzeti Bank).

MNB Working Papers 2018/1

News-Based Indices on Country Fundamentals: Do They Help Explain Sovereign Credit Spread Fluctuations?*

(Fundamentum hírindexek: Mennyit magyaráznak a szuverén felárak dinamikájából?)

Written by András Fülöp**, Zalán Kocsis***

Budapest, February 2018

Published by the Magyar Nemzeti Bank

Publisher in charge: Eszter Hergár

Szabadság tér 9., H-1054 Budapest

www.mnb.hu

ISSN 1585-5600 (online)

*We would like to thank Patrick Augustin for his insightful discussion of the paper and colleagues at ESSEC and MNB for useful comments. We are grateful for Attila Kürtösi for database programming and Varun Agrawal, Olena Chystiakova, Zsolt Szőrfi for research assistance. The views expressed in the paper are those of the authors and do not represent the official stance of Magyar Nemzeti Bank.

**ESSEC Business School. E-mail: fulop@essec.fr

***Magyar Nemzeti Bank (the Central Bank of Hungary). E-mail: kocsisz@mnb.hu.

Contents

Abstract 5

1 Introduction 6

2 Data 9

2.1 Macroeconomic and financial data 9

2.2 The news data set 10

3 News index methodology 12

3.1 Matching fundamental expressions 12

3.2 Example 1: simple expression on the housing market 13

3.3 Example 2: more complex expression on the housing market 13

3.4 Using intermediate expressions to represent more complex structures 15

3.5 Adding geographic labels to matched fundamental expressions 15

3.6 Constructing news indices from matched expressions 15

4 Properties of news indices 18

4.1 Case studies: US and Greek news indices 21

4.2 Correlations with Economic Policy Uncertainty indices 24

5 Validation: Do News Indices Extract Fundamental Information? 25

5.1 News indices vs Bloomberg economic announcements 25

5.2 News indices vs non-fundamental proxies 28

6 News indices in sovereign credit spread regressions 30

6.1 Drivers of the VIX index and implications for sovereign credit risk 35

7 Conclusions 37

References 38

Appendix A Macroeconomic and financial data 41

Appendix B Creating the news database 42

Appendix C Additional information on news indices 46

Abstract

This paper revisits the discussion about the role that fundamentals play in asset prices using sovereign credit spread data. We augment the standard macroeconomic proxy set by text-based measures of country and global fundamentals from a database of Reuters news articles between 2007 and 2016. We use a novel methodology that matches fundamental topic expressions and directly links them to tonality and geography information within the text. Our approach resolves several problems of extant text mining methods. We verify that our news indices capture fundamental information within news articles and are uncorrelated with measures of liquidity and investor sentiment. These news indices explain a large part of sovereign credit spread changes not captured by traditional fundamental proxies and thus support a significantly larger role for fundamentals. This additional information derives primarily from omitted expectations and concerns about global fundamentals. We also show that a large part of the covariance between the VIX index and sovereign spreads is related to these global fundamentals.

JEL: C8, E44, F34, G1, H63.

Keywords: financial media, textual data, regular expressions, sovereign credit risk.

Összefoglaló

Tanulmányunkban a fundamentumok eszközárakban játszott szerepét vizsgáljuk szuverén kockázati felárak adatain. A fundamentumok mérésére a hagyományos makroökonómiai változók mellett a Reuters 2007 és 2016 közötti híradatbázisa alapján létrehozott ország- és globális fundamentum hírindexeket használunk. Módszerünk újdonsága, hogy a fundamentumok tematikus kifejezéseit tonalitás- és földrajzi információkhoz kötve emeljük ki a szöveges tartalmakból. Ez a technika a meglévő szövegelemzési módszerek több problémáját kezeli. Bemutatjuk, hogy az előállított hírindexek fundamentumokra vonatkozó információkat ragadnak meg és nem korreláltak a piaci likviditás és befektetői hangulat mutatóival. A hírindexek a szuverén kockázati felárak változásának jelentős, hagyományos makrováltozók által nem megragadott részét magyarázzák. Ez a fundamentumok lényegesen nagyobb szerepére utal. A hírindexek által megragadott többletinformáció elsősorban a globális fundamentumokkal kapcsolatos kilátásokhoz, aggályokhoz köthető. Eredményeink rámutatnak arra is, hogy a VIX index és szuverén szpredek kovarianciájának nagyobb része szintén fundamentumokhoz kapcsolható.

1 Introduction

One of the basic questions of the macro-finance literature is the extent to which asset price movements can be explained by the arrival of fundamental information. Ross (1989) showed that in a no-arbitrage framework asset price volatility should be determined by fundamental news flow relating to the asset value. However, empirical work had trouble explaining asset price volatility by fundamental news only, both on the corporate and on the aggregate macroeconomic level (Officer, 1973; Roll, 1988; Schwert, 1989). This has led to considerable efforts in research to develop theories for non-fundamental sources of price fluctuations.[1]

Our paper relates to this general discussion with respect to the pricing of sovereign credit spreads: sovereign bond spreads and sovereign credit default swap (CDS) spreads. Both spreads include a default risk component, which compensates holders of risky sovereign bonds and CDS protection sellers for potential losses incurred in case the sovereign defaults. Default risk is partly rooted in domestic fundamentals as argued by theory[2] and confirmed by findings of the empirical literature beginning with Edwards (1983). However, both the default risk component and sizable risk premia in sovereign spreads have been shown to be strongly related to external shocks, liquidity and risk pricing, which have weaker ties to fundamentals.[3]

The relative role that empirical papers attribute to fundamentals has varied by the data sample and empirical methodology used. Domestic fundamentals were usually found important in studies examining the variation of levels of sovereign spreads, whereas global factors and non-fundamentals appeared to be more important in explaining their changes. Our data corroborate these findings: even at the low, quarterly, frequency the macroeconomic variables commonly used to proxy fundamentals explain a minor, less than 10 percent, share of both sovereign CDS and bond spread changes.

We claim that the small explanatory power attributed to fundamentals could be a consequence of the inadequacy of such proxies to fully account for fundamentals.[4] Some aspects of fundamentals cannot well be measured (e.g. economic policy announcements, government funding liquidity), and most available proxies of fundamentals are backward-looking in nature and do not recover the expectations elements (neither baseline expectations nor tail risks) that are central to asset pricing.

To address this claim, in this paper we investigate the effect of fundamentals on sovereign credit spread changes by augmenting the traditional proxy set with text-based measures of the fundamental information flow created directly from articles of the Reuters news archive between 2007 and 2016. Financial news agencies, competing to serve their audience of investors, arguably write about all aspects of fundamentals that they perceive relevant for asset pricing. Hence, the news flow they generate is likely to contain information about asset prices missing from traditional macro variables. News coverage extends to both real and potential policy announcements, expectations regarding the outlook for fundamentals including possible adverse scenarios.

[1] Non-fundamental price movements may originate from self-fulfilling beliefs about other investors (Diamond and Dybvig, 1983; Calvo, 2002), liquidity shocks (Allen and Gale, 2000; Brunnermeier and Pedersen, 2008) or other factors resulting in rational and speculative bubbles (see Shiller et al., 2014, for further references).

[2] A central prediction of this literature beginning with Eaton and Gersovitz (1981) is that a deterioration in the fiscal position (higher debt) and real economic growth both provide incentives for governments to choose default over debt repayment. A weak external position (low currency reserves, high current account deficits and foreign exchange debt) increases vulnerability to self-fulfulling funding crises (Calvo and Mendoza, 1996; Cole and Kehoe, 1996; Sachs et al., 1996). Fundamentals of the domestic banking system are important due to potential bailout costs (Dieckmann and Plank, 2011; Acharya et al., 2014). Political-institutional fundamentals also have an effect on fiscal conduct and sovereign credit risk as discussed by a vast literature strand (see Gaspar et al., 2017).

[3] External shocks may originate from both fundamentals or non-fundamentals (categorizations and empirical reviews are provided in Moser, 2003; Corsetti et al., 2005). Empirical findings of fundamental links due to trade spillovers and US interest rates have been mixed. Some recent papers however reported significant spillover effects from the US real economy (Dooley and Hutchison, 2009; Augustin and Tédongap, 2016). (Augustin (2014) provides further discussion and references.) Studies investigating liquidity and risk pricing (also referred to as risk appetite or investor sentiment) within spreads usually defined these factors as being distinct from fundamentals (see e.g. Beber et al. (2008) and Favero et al. (2010) in case of liquidity risk and Eichengreen and Mody (1998), Baek et al. (2005), González-Rozada and Yeyati (2008) with regard to risk pricing.)

[4] We define the fundamental component as the component relating to the future expected path of country and global fundamentals, including the distributional assumptions about this path. This definition incorporates elements of uncertainty (about fundamentals) including tail risk scenarios. Non-fundamentals are then restricted to time-variations in pricing of risks, and non-informational trading due to liquidity shocks or expectations rooted in reasons independent of fundamentals.

The idea of extracting information from financial media and using these to explain prices in various asset markets is not new (e.g. Roll, 1988, already used Dow Jones News Wire and Wall Street Journal stories between 1982 and 1987 to control for firm-specific fundamentals). In the past decade, however, increasing computational capacity and availability of text resources triggered a boom in research analyzing textual data using automated text mining techniques.

There are two key types of information extracted by textual analysis: topics (i.e. what is the text about) and tonality (whether the text reflects optimism/pessimism) both of which we need to explain asset price changes (topics to identify which country fundamentals are mentioned and tonality to identify its improvement or deterioration).[5]

It is easier to create topical indices, because most topics can be grasped by a few characteristic keywords. Consequently, the common approach runs search queries for these topical keywords and aggregates query matches into topical time series indices.[6]

Extracting the tonality of texts is more difficult. The widespread methodological approach is to use large, predefined dictionaries that classify adjectives into positive/negative categories. Documents can then be assigned a tonality score based on the particular set of words appearing within the document (the seminal paper in finance is Tetlock (2007) in this regard).[7] One problem with such full text tonality measures is that they become ambiguous if there are several relevant topics within a text (e.g. an article discussing an improving US real economy may also refer to a worsening current account or contrast the US improvement with a deteriorating economy elsewhere). Another basic and often noted problem with predefined dictionaries is that tonality words are context-specific: some words may have a positive connotation in one context but may be irrelevant in others.[8] Moreover, some of the most important verbs and adjectives that determine tonality signs depend on the specific topical expression (e.g. 'increase' and 'high' denote improvement of real economy fundamentals if the topical expressions is 'employment' but a deterioration if it is 'unemployment').

Our approach aims to provide a remedy for these problems. Namely, we define search queries that jointly match topical and tonality keywords within articles relying on a technique using so-called regular expressions. We use this method to identify fundamental topic expressions, assign each a tonality score based on tonality expressions close by and link such topic – tonality pairs to the closest geographical keyword. We explicitly define and search for topic and tonality expression pairs, which ensures taking into account the context- and expression-dependence of word tonality. Extracting expressions this way also overcomes the ambiguity problem, because it permits identifying several topic-tonality pairs within each text. The method is detailed in Section 3 and appendices.

We use this approach to find fundamental expression matches within news articles, each such match being a triplet of fundamental topic – tonality score – geography. We aggregate the tonality scores of these triplets to construct time series indices of seven different fundamental topics (real economy; external position; fiscal solvency; government funding liquidity; financial sector health; political-institutional strength; monetary policy stance) for a large cross-section of geographical units (88 countries and 11 regions).[9]

[5] We use tonality in a wider sense than sometimes seen in the literature. Instead of denoting a general sentiment of the text, we define tonality as the textual information that identifies the direction of change of fundamentals.

[6] An influential example is the Economic Policy Uncertainty index of Baker et al. (2016) extracted from US news archives, which has been used in several finance applications as an indication of uncertainty concerning the economy and policymaking. In the sovereign credit literature, several authors have used keyword searches to grasp various aspects of the eurozone crisis (Cesare et al., 2012; Mohl and Sondermann, 2013). In an application related to ours Gomes and Taamouti (2016) extract fundamental indices from Google searches.

[7] A series of papers on equity markets used such tonality indices to examine whether their predictive ability is temporary (indicative of no fundamental information) or persistent (indicating fundamental content). The findings were mixed: Tetlock et al. (2008) and Ferguson et al. (2015) found persistent effects, whereas results of Tetlock (2007) and Da et al. (2011) pointed to news mostly representing market sentiment. Looking at sovereign yields, Dergiades et al. (2014) found that social media hits on the keywords related to the Greek crisis had short-run predictive power.

[8] Realizing this caveat, several authors created dictionaries, which were more customized to their specific applications. Loughran and McDonald (2011) created a dictionary for finance contexts. Other papers used word sets designed for the areas of monetary policy and financial stability (Lucca and Trebbi, 2009; Hansen and McMahon, 2016; Correa et al., 2017).

[9] Some papers in the finance literature have simultaneously extracted tonality and topic information, but using other methods. Beetsma et al. (2013) and Ehrmann et al. (2014) manually labelled tonality after filtering news for relevant topics, which was possible as these studies worked with a relatively small news sample. Born et al. (2014) and Liu (2014) instead used automatic tonality classification methods, however, tonality in these cases was again linked to full articles. Closest to our approach are Lucca and Trebbi (2009), Hansen and McMahon (2016) and Tobback et al. (2017), which first match sentences that are relevant by topic, and then value the tonality of sentences based on learning algorithms. Although these papers do adapt tonality words that are specific to the context they analyze, these tonality words are not linked to the specific topical keywords. Our approach is more appropriate in applications (certainly our case), where tonality keywords receive their sign subject to the topical keywords they connect with.

Our news indices are by construction related to fundamentals because each expression contains topical keywords identifying one of the seven fundamental categories. The only way the indices will differ from zero is having articles in which (topical) keywords related to fundamentals are matched.[10] Nevertheless it is still an empirical issue to test, whether the indices are successful in picking up enough fundamental expressions within texts and that they correctly return tonality information. In Section 5 we document that the constructed news indices are significantly correlated with other proxies of fundamentals derived from a large data set on macroeconomic announcements and surveys available from Bloomberg. Further, to show that it is really fundamental information that our indices measure, we examine the correlations between our news indices and two proxies of non-fundamental sources of price variation identified in the literature, the noise index proposed by Hu et al. (2013) and the sentiment index of Baker and Wurgler (2006) and we find them to be insignificant.

Returning to our primary research question, we test the explanatory power of our news indices in panel regressions of sovereign credit spread changes in the 2007-2016 period on a cross-section of 58 emerging market and developed countries. Our key finding is that allowing for news dramatically increases the proportion of spread changes that can be explained by fundamentals. This suggests that the underestimation of the fundamental information flow when one only uses traditional macro variables is quantitatively significant. Further, we find that the relationship between fundamental news and CDS spread changes is mainly expressed through the global component of news. We also find that fundamental news can account for a significant fraction of the covariation between sovereign CDS spread changes and the VIX index, previously documented in the literature.[11]

The paper is structured as follows. The next section presents the data used including some basic properties of the news data set. Section 3 provides more details on our methodology through several working examples. Section 4 describes properties of the constructed news indices. Section 5 assesses the validity of the constructed news indices as proxies for fundamentals. In Section 6 we test the explanatory power of our news indices in panel regressions of sovereign credit spreads. The last section summarizes the contributions of our research.

[10] This is crucial for identifying the motives behind asset price changes. Journalists will write news articles about changing fundamentals if that is interesting for investors, i.e. if it is relevant for asset pricing. However they will also write news stories about changing asset prices even if there was no identifiable fundamental explanation for these. Even if the article mentions some background to observed price changes, these could still be non-fundamental in nature (some often seen examples are: references to general risk aversion/risk appetite: e.g 'dollar stronger on investor fears'; references to price changes in other asset classes: e.g 'increasing Greek spreads hit the euro'; 'prices reversing previous movements'; investor profit-taking; investors closing out or opening new positions; interviewing traders who refer to technical trading rules). The problem is that these non-fundamental stories can contain verbs and adjectives that relate to positive or negative tonality in a more general sense. Our method avoids (false positive) matches in these cases, because fundamental topic keywords are missing from these non-fundamental statements. Our approach only takes tonality keywords into account, when they are sufficiently close to keywords identifying fundamental topics.

[11] The VIX index displays strong correlations with the common component of sovereign spreads (e.g. McGuire and Schrijvers, 2003; Ang and Longstaff, 2011) and their coefficient estimates are usually significantly positive in sovereign spread regressions (Hilscher and Nosbusch, 2010; Alper et al., 2013; Paniagua et al., 2016, to cite just a few).

2 Data

We use several data sets in our analysis:

- news article texts (Reuters)
- traditional macroeconomic data (World Bank, IMF)
- macroeconomic announcements and economist surveys (Bloomberg)
- financial indicators (Bloomberg and Datastream)
- various other indices compiled by other researchers

We first describe macroeconomic and financial data sources and turn to news article texts in the second part of this section. Other indices used are described in later sections when they are used.

2.1 MACROECONOMIC AND FINANCIAL DATA

We use the three types of sovereign credit spread indicators that have most often been used in the empirical literature. Our primary measure are CMA sovereign CDS spreads sourced from Bloomberg. These have the benefit that they are available for a large cross-section of countries. A drawback is that, for several sovereigns including many developed countries CDS spread time series quotes began later than the start of our news sample (notably during 2007 or early 2008). For robustness purposes therefore, we also estimate each panel regression on dollar- or euro-denominated bond spreads which are based on the JP Morgan EMBI Global bond spreads (sourced from Datastream) for emerging markets and interest rate spreads compared to the German Bunds for eurozone countries (sourced from Bloomberg).

We choose (traditional) macroeconomic variables in line with the empirical sovereign risk literature beginning with the paper of Edwards (1983), which first used regressions of sovereign credit spreads on macroeconomic variables[12], and followed by a vast number of studies since. We use a quarterly frequency, because some of the most important variables are available at this frequency.

In particular, we use panel data on real GDP growth; current account to GDP ratios; reserves to GDP ratios; IMF WEO 1-year-ahead projected fiscal balance to GDP ratios; public debt ratios and per capital GDP. As a measure of the global economic outlook we use IMF WEO 1-year-ahead world real GDP growth rate projections. IMF World Economic Outlook projections are available bi-annually, per capita GDP levels are available annually: these are interpolated to quarterly frequency.

We also added Bloomberg macroeconomic announcements and survey data to our analysis (data of 'ECO' screens). This is useful to gauge the daily surprises in macroeconomic announcements in comparisons with our news index measures. To make Bloomberg published indicators comparable across countries and indicators, we use their normalized form, i.e. subtracting the 2007-2016 sample means and scaling by standard deviation. When we form the surprise component (actual value minus survey), the means drop out. Formally, for an indicator type *j*, country *i*, the surprise value on day *t*, is calculated according to:

$$X_{ijt}^{(surprise)} = SIGN_{Xj} \frac{X_{ijt}^{(actual)} - X_{ijt}^{(survey)}}{\sigma_{X_{ij,2007-2016}}}. \qquad (1)$$

[12] Earlier empirical papers on sovereign credit risk used observations of debt restructuring as dependents in logistic regressions (see McFadden et al., 1985, and references therein).

The nominal value of the surprise (actual minus survey value) is scaled by the 2007-2016 standard deviation of the given indicator of the given country. The surprises are adjusted for the sign ($SIGN_{Xj}$) to be in line with our news tonality indicators: an increase denoting improvements, a decrease denoting deterioration.

We use aggregates of the so-created daily series for two categories: REAL (relating to real economic growth) and EXTERN (external position). When no BBG publications (with surveys) are available for a given country and category, the surprise index takes the value of zero.

Regarding global financial indicators we again refer to the empirical literature and two financial market indicators, the CBOE VIX index and the CSI US corporate BBB/Baa yield to Treasury spreads (high yield spreads), which have often been used in empirical research to proxy fluctuations in risk pricing and investor sentiment.

Appendix A lists macroeconomic and financial data sources.

2.2 THE NEWS DATA SET

The news indices we create are based on the body of news article items in the Reuters news archives. At the time of writing the Reuters news archives is publicly available online[13] and spans the period between the beginning of 2007 up to the current date. The news sample we use ends on 31 October 2016. The news data set contains about 3.9 million articles in this period – after removing articles without date stamps or text bodies.

The reason that we chose the Reuters news archives is that it contains a large, publicly accessibly, edited text corpus with news items that target and reach global investors. Reuters has a large number of news items relevant in terms of containing information about country fundamentals, which is important for the construction of meaningful news indices. Google search or social media (Facebook, Twitter) data may be magnitudes larger in gross size but they have the disadvantage of a very low hit ratio of relevant information for institutional investors.[14] Among the relevant large financial media outlets, Reuters is one of the most read news agencies. Reuters – along with Bloomberg and CNBC – are more focused on the interests of non-retail, global investors (the investor group relevant in sovereign credit markets) than other popular financial media sources such as the Financial Times or The Wall Street Journal.

Even the Reuters news data set has only about one in five articles that contain relevant information about country fundamentals. Much of the remaining articles are about individual companies that are not relevant on the macroeconomic scale or about non-economics topics such as sports, entertainment, technology, etc. After preprocessing the data we remove irrelevant news items as well as duplicate versions of news as described in Appendix B. The removal of irrelevant and duplicate items reduces the number of articles to roughly a million news items or about 300 items a day, 2000 a week.

Descriptive analysis of the news data set reveals some important features (see Table 1). There is a considerable seasonality in published news items according mainly to the day of the week: daily news counts on weekends are about a fifth of those on working days (there is no considerable heterogeneity between working days or the two days of the weekend – reported in the bottom panel of the table). There are relatively more relevant items on weekdays, as sports and entertainment make up a larger share of news on weekends. A weekly moving average reduces variance significantly due to this seasonality. The moving average has a standard deviation of 61 news items in contrast to the raw series standard deviation of 161 items.

Both the number of total published news and relevant news items related are affected by holiday seasons (news counts decrease in December and in summer months), however, as Table 1 reveals, this is of a considerably smaller source of variation. News

[13] http://uk.reuters.com/resources/archive/uk/

[14] Several authors use such data sets for analyzing retail investments in corporate shares (e.g. Da et al., 2011; Joseph et al., 2011). A natural filtering choice in those papers is to only consider search/text items that mention a given corporate name or identifier. The analogous mentioning of country names to proxy sovereign risk would not be as efficient a filter in social media or Google search data (e.g. searches on Greece could reflect seasonal effects of summer holidays). Further, those studies primarily target local retail investors, which is appropriate in equity markets, however, for our purposes non-retail, global investors matter more since they constitute the relevant investor group on both sovereign CDS and international bond markets. There have been attempts to use such textual inputs in sovereign credit applications (Dergiades et al., 2014; Cesare et al., 2012), but news texts have been more popular in this area.

Table 1
Descriptive statistics of filtered news data

	Subsamples	Number of news items	Daily average of news items	Standard deviation of daily news items	Std.dev.of daily news items based on weekly moving averages[a]	Total Reuters news / filtered news
Total		1042109	290	161	61	3.75
years	2007	83952	230	119	34	4.53
	2008	120607	330	176	64	3.46
	2009	119972	329	177	53	3.73
	2010	106940	293	153	36	3.60
	2011	120454	330	171	44	3.41
	2012	130448	356	192	51	3.25
	2013	102956	282	150	39	3.49
	2014	94335	258	130	33	3.91
	2015	91265	250	136	39	4.05
	2016	71180	233	123	28	4.86
months	January	85097	275	157	66	3.67
	February	85693	303	156	55	3.84
	March	94303	304	164	58	3.64
	April	86281	288	160	57	3.89
	May	86626	279	154	50	4.07
	June	89024	297	162	58	3.61
	July	85196	275	151	51	3.91
	August	81302	262	144	46	4.02
	September	90584	302	163	56	3.55
	October	98408	317	173	67	3.75
	November	84634	313	170	57	3.73
	December	74961	269	164	78	3.33
days	Monday	173285	337	85	61	3.65
	Tuesday	194056	378	101	61	3.65
	Wednesday	204598	399	88	61	3.65
	Thursday	209248	408	95	61	3.69
	Friday	187739	366	91	61	3.63
	Saturday	34000	66	19	61	5.23
	Sunday	39183	76	22	61	4.85

Sources: Reuters news archives and authors' calculations.
Notes: Descriptive statistics of news items filtered by relevance and duplication.
[a)] Weekly moving average applied to the filtered daily news count series. This removes intra-week variation but still accounts for inter-week variations.

publication aggregates have fluctuated across the years, with news counts increasing during the crisis years, 2008-2009 and 2011-2012. Most of the increase in total news in these periods is due to the increase in relevance filtered news items (in these periods the ratio of total/filtered news decreases, see last column of Table 1). Presuming that Reuters closely follows changing investor interests, investors have spent more of their time in these periods trying to stay informed about news on macro fundamentals relative to company-specific news and news related to leisure.

3 News index methodology

To construct the fundamental news indices we proceed in two steps. In the first step (discussed next), we identify fundamental expression matches within all news articles. Each of these expressions are a triplet of a fundamental topic, a tonality score and a geographic reference. In the second step (discussed at the end of the section), we aggregate tonality scores to recover news index time series for each fundamental of each geographical unit.

3.1 MATCHING FUNDAMENTAL EXPRESSIONS

The triplets of fundamental news expressions contain a tonality score (integers in the range [-3,3]), a geographic reference (one of 88 countries and 11 regions, see Appendix table C.3 for a listing), and fundamental topic reference to one of the following seven concepts:

- REAL: real economic growth and level of development (e.g. real GDP growth, GDP per capita, industrial output, housing market, household consumption, retail sales, labor market developments, references to recession, economic crisis)
- EXTERN: external position (e.g. exports, imports, current account balance, currency reserves, external debt stock to GDP)
- FISCAL: fiscal sustainability (e.g. taxes, government spending, fiscal balance, public debt ratio)
- FUND_LIQ: government funding liquidity (e.g. demand at bond auctions, oncoming debt obligations, repayments, roll over risk, maturity structure and FX share of debt, availability of international official lending and disbursements of foreign aid)
- BANK: financial sector health (e.g. capital adequacy, ROA, ROE, NPLs, balance sheet mismatches, funding liquidity, liquidity injections, bank bailouts)
- POL_INST: political stability, institutional strength (e.g elections, minority government, government or coalition breakdowns, political crisis, coups, revolution, terrorism; strength of market institutions and democratic institutions, rule of law, transparency, corruption)
- MON_POL: monetary policy stance (e.g. central bank rate changes, hawkish-dovish stance, quantitative easing, liquidity injections, FX market interventions)

Regarding tonality, a positive sign denotes improvement in each of the first six categories. Regarding monetary policy, positive sign is defined as easing monetary conditions (increasing quantitative easing, but decreasing interest rates). Conversely, tonality is negative for deterioration of fundamentals and it is zero in case of expressions where the tone is ambiguous or neutral.[15][16]

Fundamental expression definitions specify the topic and tonality keywords and their joining rules. Each fundamental expression definition consists of 1-4 expression elements (usually one refers to tonality the others identify the topic) and an expression rule that specifies how far the expression elements can be from each other and whether or not the order that expression elements appear in matters. Definitions of all fundamental expressions are listed in Tables C.5 and C.6 of Appendix C.

We present the procedure of matching fundamental topics and tonality in news articles on a couple of working examples. We turn to adding geographic references later.

[15] Most fundamental expressions have a tonality score of either $+2$ or -2, but we use some modifiers to differentiate between less and more intensive expressions for tone (e.g. reference to a small increase in GDP would receive '+1' vs '+3' for a huge increase).

[16] Precise definitions of fundamental categories expressions and their signs are provided in our Coding Guide, which is available in the online appendix. We developed these definitions through several rounds of iterations of independently, manually labelling documents with research assistants and discussing ambiguities when checking inconsistencies between manual labels. The definitions of fundamentals also form the basis for our fundamental expression rules.

3.2 EXAMPLE 1: SIMPLE EXPRESSION ON THE HOUSING MARKET

The first fundamental expression in Table C.5 aims to match sentences that refer to a housing market bubble bursting. Whenever such an expression is found in an article, this registers a negative score for the REAL fundamental of the respective country, since housing market busts represent an adverse development in the real economy.

The expression has two elements (1) 'N_HOUSE' which is a synonym label (explained below) and (2) 'bust|burst' which represents either of the words 'bust' or 'burst'. The expression also has an expression rule that specifies that 'N_HOUSE' and 'bust|burst' can be a maximum of three words apart.[17]

The example highlights the two key ideas of our methodology with which we aim to accommodate the flexibility of language:

- First, we operate with a large set of synonym labels, such as 'N_HOUSE' in the example. These labels represent synonymous words or n-grams that are often used interchangeably in financial news. The labels are inserted into news articles in front of the synonyms that they represent. For example, the label 'N_HOUSE' would be inserted in front of the nouns 'house', 'housing', 'dwelling', 'property' and the bi-gram 'real estate' and plural forms of these, thus representing a total of 9 n-grams. This is useful because 'N_HOUSE' can then be a shorthand for 'house | houses | housing | dwelling ...' in expression rules. The complete enumeration of the 9 n-grams is impractical and all the more so, because these synonyms are part of several other expressions as well. Further, many verbal and adjective synonym lists are much longer then the list encompassed by 'N_HOUSE' and are more frequent elements of fundamental expressions.[18] All synonym labels are listed in Table C.1 and their elements in Table C.2 and Table C.3.
- The second idea is to have distance rules for expression elements, which is the key to linking fundamental topics (e.g. reference to the housing market) to their tonality expressions (e.g. bust or burst: as being an adverse development). These rules are flexible enough to capture expressions even when there are other words that wedge in between expression elements (it will match phrases 'housing (bubble has|will) burst', 'property (market) bust' and so on) without the need to specify such interim words. For this simple example explicit listing is perhaps only cumbersome. However, for expressions that involve 3-4 elements and possibly span long sentences, explicitly defining exact phrases becomes infeasible. On the other hand, expression rules provide a limit to the maximum distance that expression elements can be. This is important because the closer expression elements are to each other the more probable that they are in the proper semantic relationship: e.g that the words 'bust' or 'burst' really appear in the text referring to the housing market.

3.3 EXAMPLE 2: MORE COMPLEX EXPRESSION ON THE HOUSING MARKET

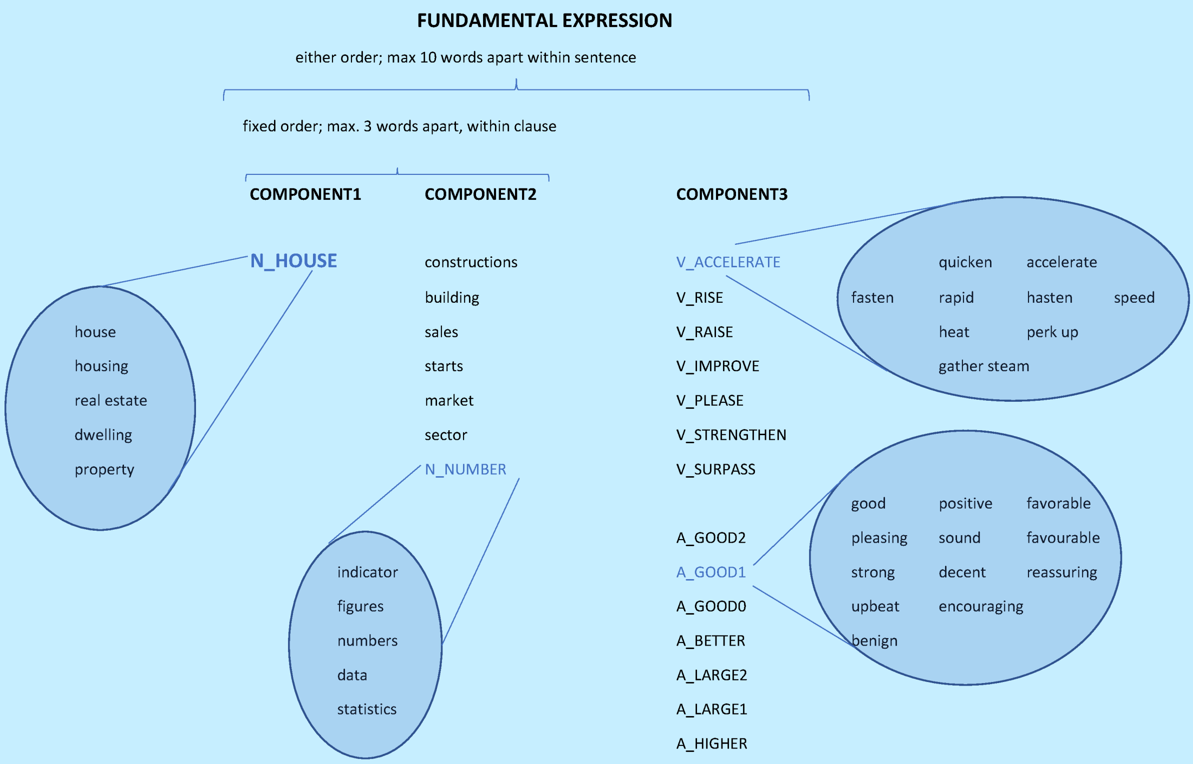
A more complex expression, and one that is more representative of most fundamental expressions that we use, is shown in Figure 1. In contrast to the previous example, this expression aims to recover more general references to the housing market or its indicators improving, increasing or generally being mentioned together with adjectives that reflect an improved or increased state. To arrive at specific statements about favorable conditions in the housing market we need references to the housing market or its indicators (accomplished by components 1 and 2) and adjectives or verbs that refer to their state (component 3).

Again we use synonym labels to cut the list of synonymous ways to refer to housing indicators shorter: we insert the synonym labels 'N_HOUSE' in front of the words mentioned above and 'N_NUMBER' in front of words referring to synonyms of indicators (see figure). Seven other synonym labels for verb lists and seven more for adjective lists are used in the expression. For example, 'V_ACCELERATE' is a synonym label for 8 n-grams (see figure) and their conjugated forms (e.g. 'accelerates', 'accelerated',

[17] These rules are coded as four argument functions: the first two arguments specify the expression elements; the third represents constituent ordering (either 1: elements only in the order of function arguments or 2: reverse order possible); and the fourth argument specifies maximum distance (0: Maximum three words in between and no commas, no full stops, no paragraph breaks. 1: Maximum ten words in between and no full stops, no paragraph breaks. 2: Maximum fifteen words in between and no full stops, no paragraph breaks. 3: any number of words in between but no paragraph breaks.).
For this example the expression code is (p1,p2,1,0) which stands for p1: 'N_HOUSE', p2: 'bust|burst', 1: only this order, 0: maximum three words and no punctuation marks in between. Technically this code is converted into the regular expression:
"(N_HOUSE)(([A-OQ-z]\w*\s){0,3}?)(bust|burst)".

[18] Consider for instance synonyms of the verb 'increase' that are captured by the synonym label 'V_RISE'. 'Increase' and its synonyms are very frequent in financial news and may denote tonality for several fundamental topics from GDP growth rates to central bank rates, government debt or political instability to name just a few.

Figure 1
Example of a positive real economy expression



Notes: This specific example refers to an improvement in the housing market. The first two components of the expression encompass a variety of ways to refer to the housing market and its indicators, whereas the third component represents ways to refer to an improvement, increase (verbs) or an improved, increased state (adjectives).
This fundamental expression is matched in articles where each of the three components are found (within the specified word distance bounds and in the given word ordering). Namely, the first component (the label 'N_HOUSE') has to be a maximum three words away from any of the 7 elements of the second component and either of the first two components has to be maximum 10 words away from any elements of the third component.
As seen here, expression elements may be words (e.g. 'construction', 'market') or they may be labels (e.g. 'N_HOUSE', 'V_RISE', 'A_BETTER'). Labels are inserted into the text in front of synonymous words. Labels serve to condense such synonym lists and thus make expression rules simpler. The figure shows the synonym lists for four labels.

'accelerating'). Since these synonym labels are inserted into the text in front of the words they represent, instead of referring to original words, we can just use their synonym labels as shorthand in the expression rules.[19]

Another important benefit of using labels for synonyms is that many synonym labels are appropriate to use in many different expressions. This is especially the case for popular verbs and adjectives, examples of which are represented as component 3 in Figure 1. Because synonym groups need to be defined only once and their labels can be reused in many expression definitions, this approach substantially reduces both computer code and runtime. Notably, both are reduced from unfeasible to feasible ranges.[20]

[19] For instance the simple combination of words listed for the first two components and 'V_ACCELERATE' (also counting plural and conjugated forms) represent roughly 3000 combinations (9 x 11 x 33) of their constituent n-grams. If we account for synonyms of other verbs and adjectives in the third component, the number of combinations are boosted into the 10-100 thousand range for only this one particular type of expression about the housing market. In contrast, our approach of synonym labelling is easily implemented by a few lines of code relying on 8 synonym group definitions in Table C.3 and 2 expression definitions in Table C.6.

[20] Some notes are in place, here. First, synonyms, as we call them, are not synonyms in a strict sense, but rather loose, context-specific semantic matches. Second, labels beginning with 'N_', 'A_', 'V_' mostly represent nouns, adjectives and verbs, respectively, but this is not necessary. The key idea is that the n-grams belonging to a label should be words, phrases that are used interchangeably as constituents of specific expressions in the

Appendix tables list all synonym labels used (Table C.1) and their constituent n-grams (Table C.2 and C.3).

3.4 USING INTERMEDIATE EXPRESSIONS TO REPRESENT MORE COMPLEX STRUCTURES

Although expressions such as the ones in the previous example make up most of the fundamental expressions we match, a significant shortcoming of these is that they cannot capture information within phrases that refer to concerns, surprises, expectations. First, this information is potentially highly important for asset pricing and therefore our particular application. Second, not accounting for these increases noise in tonality identification, because verbs may be attached to these phrases instead of the fundamental topic (e.g. ’concerns increased about economic growth’).

The methodology for identifying fundamental expressions can be applied for intermediate expressions. Intermediate expressions are defined via similar expression rules (e.g one such rule would require synonyms of concern and increase, labels ’N_CONCERN’ and ’V_RISE’, close to each other). All intermediate expression definitions are listed in Appendix Table C.4.

Searching through all articles, we insert an intermediate expression label for each match, i.e where the intermediate expression rule is satisfied. Having these in the text, then, we can use intermediate expression labels just like synonym labels in fundamental expressions.

Figure 2 lists eight intermediate expressions each composed of a noun label (component 1) and an adjective or verb label (component 2). The goal of these intermediate expressions are the same as those of synonym labels, to condense the many possible ways of expressing similar semantic content into a few expression labels. These expression labels can then be used in various contexts relating to all fundamental expressions, as some examples in the figure show.

These intermediate expressions alone do not form fundamental expressions. They have to be linked to n-grams or labels that relate to fundamental concepts. For instance, increasing worries about high levels (or increases) of credit spreads or mounting concerns about Greece would not qualify as fundamental expressions without a valid reference to fundamentals. They would however qualify, if references to the banking sector, government finances or political instability were within the word distances imposed by proximity rules. This is important, because we want to extract textual information which relate to fundamentals and avoid recovering those referring to price changes induced by non-informational trading.

3.5 ADDING GEOGRAPHIC LABELS TO MATCHED FUNDAMENTAL EXPRESSIONS

Finally, fundamental expressions are completed by matching geographic labels to the extracted fundamental topic and tonality pairs using the following rule. If there is a geographic label within the expression (the text running from the first constituent to the last constituent of the fundamental expression), we use that one. If this is not available, then we check the last geographic label in the text preceding the expression. This rests on the observation that articles usually mention the reference to the country name they write about at the outset and in most cases there are rarely repeated references to the country afterwards. If there is no geographic label before the expression, we check for the first one following it. If no geographic label is found, the fundamental expression is discarded.[21]

3.6 CONSTRUCTING NEWS INDICES FROM MATCHED EXPRESSIONS

Our news indices are time series of tonality scores for seven fundamentals of 99 geographic units. For a given time period each index aggregates fundamental expression tonality scores for the respective fundamental topic and respective country/region:

particular context of the expressions. The n-grams ’gather steam’ and ’rapid’ are not verbs, but they do add the same tonality meaning when referring to economic growth, for example. Third, some word combinations represented by synonym labels do not make sense and thus almost never arise in the text, e.g. ’house’ together with ’sector’, ’house’ with ’indicator’. Nonetheless if in other contexts ’house’ is a valid substitute for ’housing’ and its several synonyms, e.g. in the case of ’starts, ’constructions’, ’building’, then it still makes sense to have it in the list of synonyms.

[21] Although not error-proof, this method is much more precise than methods which identify the most frequently cited geography reference in the text (with or without correction for their unconditional expectation).

Figure 2
Examples of intermediate expressions

COMPONENT 1

1a N_HOPE	1b N_CONCERN	1c N_TROUBLE	1d N_STRAIN	1e N_CHANCE	1f N_RISK
hope	concern	difficulties	challenge	chance	risk of
prospect	worry	problem	stress	probability	threat of
	worries	trouble	headwind	possibility	risk regarding
	anxiety	problems	strain	odds	risk relating to
	fear	troubles	pressure		risk related to
	unease				risks concerning
					risks relating to
					risks related to

COMPONENT 2

2a	A_LOWER	lower, decreased, reduced, lesser, smaller, short of, dim
2b	A_SMALL1	small, minor, insignificant, unimportant, lesser, slight, trivial, little, low, muted, subdued, tepid
2c	A_SMALL2	tiny, undersized, miniature, mini, diminutive, minuscule, smallest, bottom, lowest, least
2d	V_FALL	decrease, fall, drop, lower, reduce, slacken, decline, wane, fade, shrink, sink, dwindle, diminish, contract, moderate, narrow, subtract, dip, plunge, slide, plummet, lose, shed, shrink, halve
2e	V_CUT	reduce, cut, lower, dampen, moderate, curb, lessen, slash, scale back, drag down, halve, erode, bring down
2f	V_LIMIT	limit, restrain, constrain, curb, restrict, curtail, trim
2g	V_END	end, finish, terminate, stop, cease, interrupt, cancel, break, remove
2h	V_ALLEVIATE	soothe, alleviate, calm

INTERMEDIATE EXPRESSIONS	SAMPLE USE IN FUNDAMENTAL EXPRESSIONS	FUNDAMENTAL EXPRESSION
E_HOPE_LOW (1a + 2[a-c])	.. dim hope of financial sector liquidity conditions improving	BANK negative
E_CONCERN_LOW (1[b-d] + 2[a-c])	.. worries muted so far about the sufficiency of currency reserves	EXTERN positive
E_PROB_LOW (1e + 2[a-c])	...low probability of the fiscal deficit increasing	FISCAL positive
E_RISK_LOW (1f + 2[a-c])	...analysts see limited risks related to the housing market	REAL positive
E_HOPE_FALL (1a + 2[d-h])	... reducing the prospects of further quantitative easing	MONPOL negative
E_CONCERN_FALL (1[b-d] + 2[d-h])	... removes some of the challenges facing economic growth	REAL positive
E_PROB_FALL (1e + 2[d-h])	... decreases the odds of a UN brokered peace agreement..	POL_INST negative
E_RISK_FALL (1f + 2[d-h])	... limiting the threat of bond auction failures	FUND_LIQ positive

Notes: The figure shows the construction and use of eight sample intermediate expressions (bottom panel left column). Intermediate expression are matched in the text where relevant noun and adjective/verb synonym labels (component 1 and component 2) are within tight proximity bounds of each other. The upper part of the figure lists n-grams that the labels represent (without conjugation). The bottom panel shows the sample applications of the intermediate expressions within fundamental expressions.

$$N_{ijt} = \sum_{d \in D_t} \sum_k S_{ijdk}, \tag{2}$$

where N_{ijt} denotes the news index for fundamental j of country i in period t and S_{ijdk} are tonality scores of fundamental expressions within news items. The inner sum aggregates the tonality scores of all fundamental expressions, k, within an article d. The outer sum aggregates these article tonality scores across all articles published on the given day, D_t.

We also construct four types of subindices (for each fundamental and each geographic unit) that we use in some applications in the paper. These subindices aim to group fundamental expressions on the basis of how they relate to fundamentals. The first (CHANGE) comprises expressions that contain verbs or intermediate expressions that convey the change in fundamentals (e.g. 'unemployment rose'). The second groups expressions that contain reference to expectations, predictions of fundamentals (EXPECT: e.g. 'unemployment is expected to rise'). The third collects expressions that relate fundamentals to previous expectations (SURPRISE: e.g. 'unemployment rose more than expected'). The fourth groups expressions involving intermediate expressions relating to concerns, hopes, probabilities, risks of fundamental scenarios (CONCERN: e.g. 'decreasing concerns about the rise in unemployment'). Most occurrences belong to the first (CHANGE) and fourth (CONCERN) groups. Table C.7 reports expression occurrences in the text based on the these groupings for each fundamental category. About half of all identified fundamental expressions do not belong to any of the above groups. These usually express the state of the fundamental as they were at the time of writing or refer to the history of fundamentals.

Finally, we also create time series (for each fundamental and geographic unit) that we refer to as news counts and which aggregate the number of fundamental expressions, instead of their tonality scores. These are similar to topical indices that are frequent in the literature. They show which fundamentals and countries the media paid attention to at a given point in time, but do not reveal anything about the ”sign” of the fundamental.

4 Properties of news indices

Table 2 reports basic descriptive statistics of weekly fundamental news indices and news counts.[22] Weekly news indices were always negative for the politics-institutions category (aggregated for all countries) and most of the time for government liquidity news. Regarding these fundamentals, it could be said that ”no news is good news”. Monetary policy news tonality was tilted more toward the positive side (which denotes easy conditions by our definition). Weekly indices of other fundamentals were more evenly divided between positive and negative tone periods.

Based on news counts (right panel) most fundamental expressions were related to the real economy and politics-institutions categories, whereas we extracted the least fundamental expressions on government funding liquidity, the banking system[23] and the external position. This masks a large positive skew in the distribution for news counts of banking sector and government liquidity position fundamentals, though. The maxima reveal that there were several weeks when these latter fundamentals appeared to be highly important.

In terms of geographies, US news indices were by far the most frequent in the sample. Fundamental news counts about the UK, China and Japan followed. Regarding the eurozone, references to the fundamentals of the whole region was more frequent than references to any single member countries. Greek news stands out from the latter, single country news counts. Distribution of news indices were positively skewed for most countries: for most weeks the fundamental information flow was well below average with rare spikes of attention. This was especially true of news counts of smaller countries and crisis countries, which rarely came into the spotlight, but there was a deluge of news about them in those moments.

Table 3 reports the shares of fundamental expressions. In panel A, consistent with the weekly statistics seen in Table 2, real economy and politics-institutions comprised most expression matches. On the full sample of the Reuters archive these two categories together took up about 60 percent of all matched expressions. As mentioned in Section 2, there was a substantial difference in the structure of incoming fundamental news between weekdays and weekends. Fundamental news expressions were relatively more frequent on weekdays when published macroeconomic data and announcements are more characteristic (real economy, external position and monetary policy), whereas politics-institutions took center stage on weekends. About half of all extracted expressions belonged to this latter category on weekends, in contrast with every fourth on weekdays.

News flow on subsamples comprising the financial and sovereign crisis support intuition. In the sample of the financial crisis, relatively more fundamental expressions were matched on the health of the banking system, on the real economy, and monetary policy. By contrast, in the sovereign crisis subsample, fundamental news about the fiscal stance and government liquidity had a larger share compared with the full time sample.

The second block of Table 3 Panel B reports fundamental expression shares for news subindices. Expressions relating to changes in fundamentals appeared to be relatively more important in the case of monetary and fiscal policy stance. Deliberations about concerns, probabilities, risks of various scenarios was relatively more frequent in the government liquidity and banking sector fundamentals. This is likely a consequence of bailout measures falling into these fundamental categories, where deliberations about the timing or type of measures taken could be more important on the whole than reports about an action being taken. Fundamental expressions related to expectations and surprises were predominantly captured relating to the real economy. This is intuitive, as most indicators where surveys, polls are available fall into this category.

Regarding geographic subsamples, a key divide between countries was in terms of the share of politics-institutions. Notably, this fundamental category was more characteristic of news about emerging markets than developed countries. This may be

[22] We report aggregates at the weekly frequency, because there is considerable heterogeneity in the structure of fundamental news with respect to the weekday/weekend divide as already seen in Table 1. Also, when considering expressions broken down to fundamental categories and perhaps index types, daily aggregates are sparse for all but the largest countries.

[23] We were careful to have fundamental expressions here that only refer to the banking system as a whole and disregard information on individual banks on the basis that these should be less relevant on a macroeconomic scale.

Table 2
Descriptive statistics of weekly news indices and news counts

	News indices				News counts			
	mean	std	min	max	mean	std	min	max
	by fundamental category							
REAL	-29.0	1033.6	-4018.0	2731.0	4638.9	1771.0	1059.0	10110.0
EXTERN	44.8	127.9	-431.0	996.0	938.2	371.3	244.0	2724.0
FISCAL	3.4	243.0	-1158.0	1109.0	1519.1	1003.9	176.0	7059.0
FUND_LIQ	-131.8	185.9	-1630.0	28.0	236.0	255.0	10.0	2062.0
BANK	-64.3	251.8	-1128.0	649.0	836.7	577.2	11.0	4757.0
POL_INST	-2091.5	929.8	-7304.0	-580.0	2982.7	1156.4	896.0	8476.0
MON_POL	373.0	592.4	-1483.0	3448.0	1490.0	686.0	259.0	4664.0
	by geography							
GLOBAL	-140.7	251.2	-1981.0	495.0	884.9	530.9	116.0	4639.0
EUROZONE	59.5	254.5	-884.0	1821.0	669.7	559.4	8.0	3633.0
US	-91.8	719.7	-3822.0	2037.0	3505.2	1583.0	607.0	7966.0
CHINA	-35.4	179.2	-884.0	1664.0	691.4	437.9	31.0	3614.0
JAPAN	105.6	171.9	-466.0	751.0	498.6	296.6	15.0	1931.0
UK	-44.3	233.2	-2035.0	907.0	817.3	444.0	72.0	3517.0
GERMANY	2.8	73.8	-434.0	380.0	217.9	142.0	2.0	1090.0
SPAIN	-37.8	82.6	-888.0	97.0	180.6	273.0	0.0	2248.0
ITALY	-24.8	62.4	-972.0	107.0	141.8	177.9	0.0	1964.0
GREECE	-77.3	218.3	-2769.0	190.0	370.5	590.1	0.0	5001.0
ARGENTINA	-11.6	30.0	-262.0	46.0	47.0	48.8	0.0	434.0
ICELAND	-1.7	19.4	-275.0	64.0	10.9	31.8	0.0	367.0
CYPRUS	-1.1	15.5	-280.0	38.0	18.3	85.7	0.0	1552.0
BRAZIL	-18.2	65.1	-910.0	172.0	151.3	109.4	0.0	1098.0
RUSSIA	-112.1	145.0	-1082.0	78.0	259.2	202.6	20.0	1328.0
INDIA	-9.7	89.3	-447.0	271.0	194.4	114.4	8.0	683.0
TURKEY	-56.1	92.3	-1054.0	57.0	119.1	123.3	0.0	1446.0

Sources: Reuters news archives and authors' calculations.
Notes: News indices are constructed by summing tonality scores of all fundamental expressions matched within the period (each week). News counts are constructed by summing the number of fundamental expression matches within the period (each week). The upper panel displays the fundamental category news indices/news counts aggregated across all geographies. In the lower panel the geography category news indices/news counts are aggregated across all fundamentals.

a result of two factors. On one hand, smaller countries and emerging markets may have less macroeconomic publications, economic policy statements relative to developed countries. On the other hand, it may also suggest that incidence of political risks and their impact on asset prices may be much more important in emerging markets than macroeconomic factors given the natural assumption that Reuters endeavours to report news that is important for investors.

Most of the large developed countries had a relative higher share of real economy expressions. Again, this may be a consequence of more indicators published regularly for this group of countries. For the eurozone, monetary policy was relatively

Table 3
News count shares

	REAL	EXTERN	FISCAL	FUND_LIQ	BANK	POL_INST	MON_POL
	PANEL A: Full sample share of fundamental category (percentage)						
FULL sample	36.7	7.4	12.0	1.9	6.6	23.6	11.8
	PANEL B: Difference compared to full sample shares (percentage points)						
	REAL	EXTERN	FISCAL	FUND_LIQ	BANK	POL_INST	MON_POL
	by time sample						
weekdays	1.3	0.2	-0.0	0.0	0.2	-2.3	0.6
weekends	-16.0	-2.7	0.3	-0.2	-2.5	27.9	-6.9
FIN CRISIS[a]	3.0	-1.0	-4.3	-1.1	3.8	-1.2	0.8
SOV CRISIS[a]	-0.6	-0.6	4.0	0.9	-0.5	-1.4	-1.7
	by subindex[b]						
CHANGE	2.1	1.7	3.9	-1.2	-0.3	-13.7	7.5
CONCERN	2.1	-3.2	-3.0	3.2	3.0	-3.8	1.7
EXPECT	34.0	3.0	-2.7	-1.6	-2.4	-23.4	-6.8
SURPR	48.9	0.7	-7.5	-1.8	-5.2	-23.6	-11.4
	by geography						
GLOBAL	18.7	0.9	-2.6	-0.8	3.0	-13.6	-5.6
EUROZONE	-8.9	-4.6	-1.1	0.8	9.3	-17.9	22.5
US	16.4	-3.6	-0.3	-0.6	0.6	-16.6	4.0
CHINA	9.4	13.7	-5.9	-1.2	-3.1	-16.2	3.2
JAPAN	-0.3	3.1	3.3	-1.3	-1.3	-19.5	16.0
UK	6.4	-4.0	3.1	-1.4	3.3	-10.5	3.0
GERMANY	15.1	2.9	2.2	-0.3	0.3	-12.6	-7.6
SPAIN	-4.4	-5.1	11.6	3.3	7.9	-3.4	-9.9
ITALY	-9.0	-5.1	13.6	0.7	3.6	6.0	-9.7
GREECE	-21.0	-5.6	14.8	14.1	-0.6	8.6	-10.3
ARGENTINA	-17.8	15.5	3.7	19.0	-5.5	-5.8	-9.0
ICELAND	-5.7	-3.1	0.6	2.3	9.1	2.3	-5.6
CYPRUS	-19.6	-4.8	13.7	6.4	13.3	0.4	-9.4
BRAZIL	1.8	3.4	8.4	-1.2	-3.5	-8.5	-0.4
RUSSIA	-21.4	8.7	-4.0	-0.5	-2.9	28.2	-8.1
INDIA	-13.4	6.8	-1.0	-1.2	-3.3	-0.2	12.3
TURKEY	-27.0	0.2	-8.0	-1.4	-4.5	43.0	-2.4

Sources: Reuters news archives and authors' calculations.
Notes: Panel A reports the share of expression matches for the respective category relative to all fundamental expression matches. Panel B reports deviations from these fundamental shares in percentage points for subsamples of matched fundamental expressions. Deviations greater than 5 percentage points in absolute value are shaded for better visualization.
[a] Financial crisis sample: 1 Jul 2007 - 31 Mar 2009; Sovereign crisis sample: 1 Jan 2010 - 31 Dec 2012.
[b] See Section 3.4 and 3.6. These subindices relate to fundamental expressions which have as an expression element (i) a verb that denotes a change in fundamentals (CHANGE); (ii) an intermediate expression label that relates to concerns, hopes, risks about fundamentals (CONCERN); (iii) an intermediate expression label that relates to expectations about future fundamentals (EXPECT); (iv) an intermediate expression label that relates to changes relative to expectations about fundamentals (SURPR).

more important; a plausible result given that the references to the ECB fall into this regional category, whereas much of the real economy news expressions are reported at the country level. Additionally, the banking sector and government liquidity fundamentals had relative higher shares in the eurozone, as could be expected considering the sovereign crisis of the region. Fiscal policy and sustainability appeared to be a country-level rather than a regional phenomenon.

For smaller countries, deviations in the share of fundamental indicators reflected the country-specific fundamental stories of the sample. For China, Russia and Argentina the external position was relatively more important due to the significance of foreign trade, but also foreign currency reserves and foreign currency management in these countries. Fiscal policy was relatively more important in periphery eurozone countries as would be expected. Expressions of government funding liquidity were more frequent for Greece and Argentina, which struggled with default in the period. Fundamental expressions about the banking sector had higher shares in Spain, Cyprus, Iceland, where the banking sector was undergoing periods of severe stress. Again, as would be expected, fundamental indicators on politics and institutions had a higher than average share in Russia and Turkey.

4.1 CASE STUDIES: US AND GREEK NEWS INDICES

To provide an intuition on how our news indices look like, Figures 3 and 4 plot the time series of several US and Greek news indices and subindices.

The top and bottom panels of Figure 3 display US REAL and US MON_POL indices, respectively. The CHANGES subindex of US REAL fundamentals appears to follow general tendencies in the real economy as measured by the Markit Purchasing Managers Index and higher frequency news fluctuations in economic surprises as measured by Bloomberg. More formal tests of the relationship between economic announcements and news indices are reported in the next section.

The US REAL CONCERN subindex follows a somewhat different path. As we argued in the Introduction there are no standard proxies of perceived economic tail risks. It is suggestive of this index, however, that it reached most pessimistic levels at the turn of 2007-2008, a time when concerns about the financial crisis unfolded. At this time indicators were still at much better levels than one year later as suggested by both PMI and Bloomberg economic surprises data (as well as our own CHANGE subindex). It seems intuitive that the CONCERN subindex grasped larger fears surrounding the economy before the recession. When recession actually hit, the CONCERN index moderated (to still very pessimistic levels) plausibly because the previous tail risk scenario already materialized and this left less fear of an additional deterioration in conditions.

The US monetary policy indices highlight the differences between the two subindices even more. The CHANGE subindex reacted to actual events, announcements taking place in monetary policy. The large spikes in 2007-2008 captured interest rate cuts, liquidity injections, while later on announcements of quantitative easing (QE) phases increased the index.[24] As monetary conditions gradually turned more and more restrictive (QE ended, expected future interest rate path increased), our index turned negative after 2014.

Again, the US MON_POL CONCERN subindex takes a different path. In general it is much more symmetric than the other index. We think of this being a consequence of the CONCERN subindex identifying potential deviations from an expected consensus path of monetary policy (which is plausibly reflected by the other, CHANGE subindex), which may be of either sign whether or not policy is restrictive. Again, we do not have objective tools to test this, but the large negative spike around the tapering talk appears to support our view. (In 2013 May and June several high level speeches by the Fed signalled that QE3 purchases could be levelled off, which was commonly referred to as the ”taper talk”.) There were no immediate steps announced or taken at this point[25], but there was a sudden and widely held perception that monetary policy stance was turning more restrictive.

Figure 4 depicts evolution of three Greek news indices: fiscal position (FISCAL), government funding liquidity (FUND_LIQ) and politics-institutions (POL_INST). As seen in the top panel the main FISCAL news index had spikes into positive values that were associated with austerity package announcements. In contrast, the Greek CONCERN subindex (bottom panel) was more often negative, perhaps reflecting worries either about debt sustainability or about austerity steps implementation.

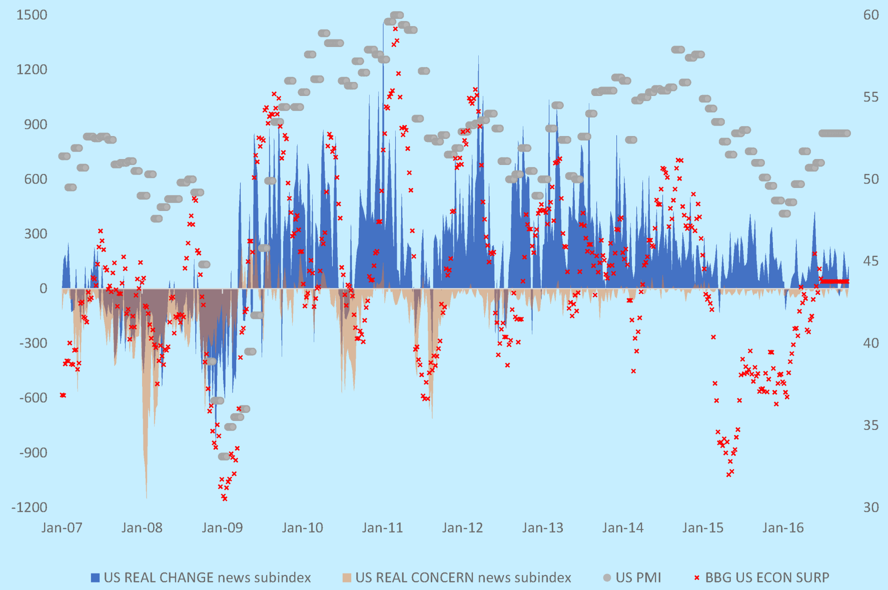
The main index and the CONCERN subindex for government funding liquidity identify the 2010 May financing crisis and worries about bailout disbursements and restructuring in 2011 and early 2012. The restructuring in March 2012 appears to grasp hopes of a better financing position, but in general most of the spikes in these indices were negative.

[24] Note that we define increases in monetary policy indices as pertaining to easing monetary conditions.

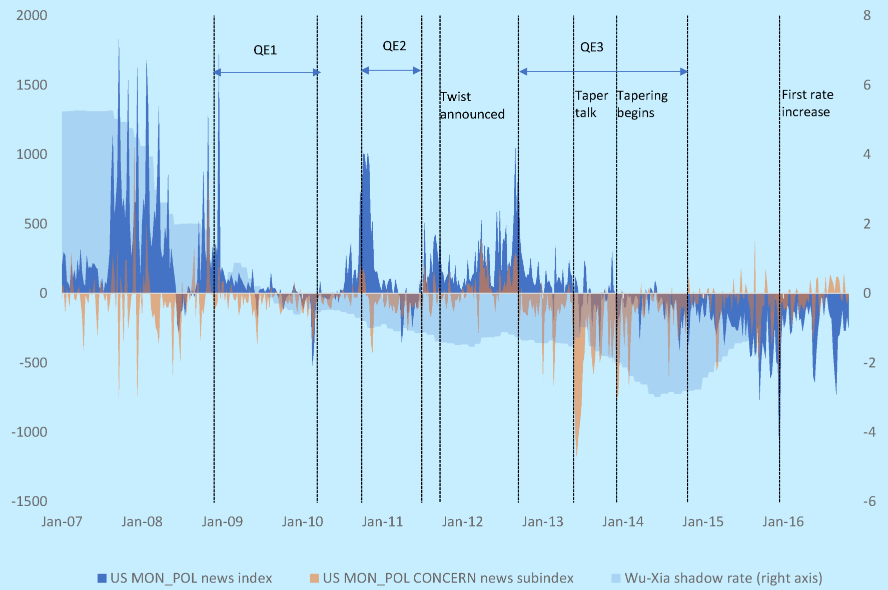
[25] Tapering was eventually announced in December that year and QE3 ended in October 2014.

Figure 3
US real economy and monetary policy news indices

Panel A: US REAL news indices



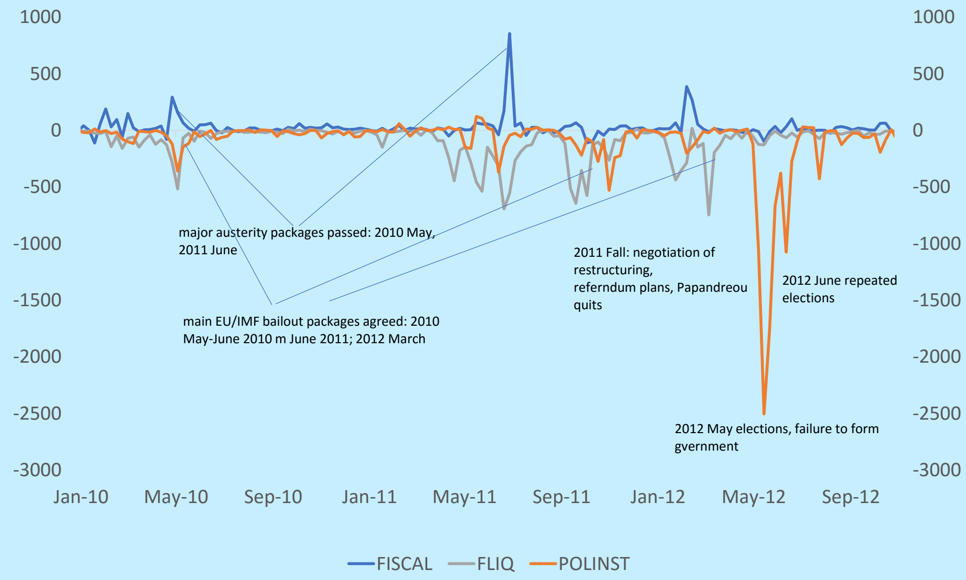
Panel B: US MON_POL news indices



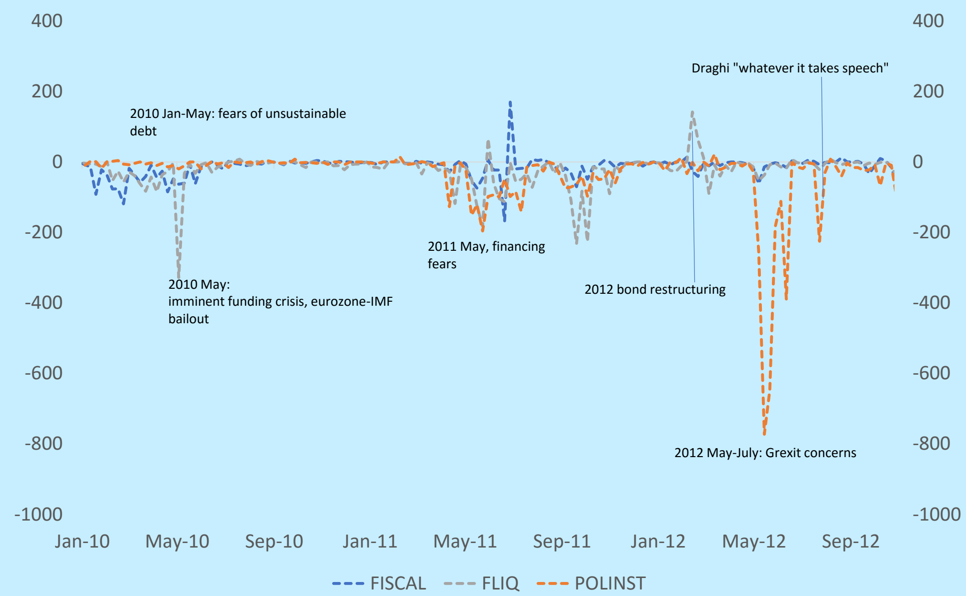
Sources: Reuters news archives, Bloomberg (Markit), US Federal Reserves and authors' calculations.
Notes: News indices reported at weekly frequency. The CHANGE subindex aggregates tonality scores of fundamental expressions referring to changes in fundamentals. The CONCERN subindex aggregates tonality scores of fundamental expressions that contain reference to concerns about fundamentals. PMI refers to Markit Purchasing Managers Index data, the Wu-Xia shadow rates estimate where short-rates would be in absence of the zero lower bound based on yield curve dynamics (Wu and Xia, 2016). Taper talk refers to a series of speeches by Fed officials that signalled QE3 bond purchases will soon be levelled off.

Figure 4
Greek fiscal and political news indices

Panel A: GREECE news indices (main indices)



Panel B: GREECE news indices (CONCERN subindices)



Sources: Reuters news archives and authors' calculations.
Notes: News indices reported at weekly frequency. Panel A aggregates tonality scores of Greek fundamental expressions referring to the fundamentals: fiscal position (FISCAL), government funding liquidity (FLIQ) and politics-institutions (POLINST). Panel B plots CONCERN subindices of the same fundamentals. These subindices aggregates tonality scores of fundamental expressions that contain reference to concerns about fundamentals.

Political-institutional indices spiked around the May 2012 elections and the repeat elections in June. At this time the more general eurozone breakup fears (defined as negative politics-institutions for both Greece and the eurozone) led to ECB Governor Mario Draghi to announce the intent of the ECB to do whatever it takes to save EMU integrity, which reduced these concerns.

4.2 CORRELATIONS WITH ECONOMIC POLICY UNCERTAINTY INDICES

The Economic Policy Uncertainty (EPU) indices of Baker et al. (2016) are probably the most popular indices in the field of economics derived from text-based (news) input. The EPU has been used in numerous economics and finance applications.[26] It is natural therefore to evaluate our news indices in relation to these indices.

The EPU indices are based on article hits, where keywords about the economy, policy and uncertainty are matched. For an article to be called a match, the article has to contain: (i) at least one word from a list of policy related keywords, (ii) either the word 'economy' or 'economic' and (iii) either the word 'uncertain' or 'uncertainty'. The number of articles that match at least one instance of all three types of keywords are then aggregated for the given time period (most indices are monthly; a few are daily) and the resulting time series are standardized.

There are several differences with respect to data, methodology and concept, which limits the correspondence between EPU indices and our news indices. Regarding data, our indices are based on the UK edition of Reuters in the period 2007-2016, whereas the EPU is based on a much larger set of textual data that ranges over many countries and collects information from a large set of periodicals (ten periodicals only in the US) and also has a longer history than our data set. A methodological difference is that we only match expressions when expression elements are close to each other within the text, whereas Baker et al. (2016) do not have such word distance restrictions. Consequently, EPU indices will be tilted towards false positives (identifying more matches than truly in the text), whereas our measures will be more balanced between false negatives (identifying less matches than truly in the text) and false positives. There are conceptual differences regarding our fundamental categories and the policy categories of the EPU indices, so indices would not align exactly even if we worked on the same data and with the same methodology. We also do not require mentions on uncertainty, although the CONCERN subindex may somewhat be related to this concept. Finally, whereas the EPU indices are topical indices only, our news indices also include information on tonality.

We can still do some comparisons based on our news count measures (number of fundamental expression matches) and calculate correlations with EPU indices. To avoid spurious correlations due to common trends we work on (monthly) changes in both our news counts and EPU time series.

Table 4
Correlations with EPU indices

	EPU indices				
News counts	US main	UK	China	US MON_POL	US FISCAL
US main	0.412	-0.018	-0.088	0.552	0.427
UK	0.273	0.460	0.152	0.171	0.122
China	0.019	0.027	0.039	0.015	-0.049
US MON_POL	0.092	-0.130	-0.179	0.274	0.050
US FISCAL	0.231	0.018	0.034	0.155	0.379

Sources: http://www.policyuncertainty.com, Reuters news archives, and authors' calculations.
Notes: Pearson's bivariate correlations between monthly changes of news counts and EPU indices. Correlations are shaded according to 5% (light) and 1% (dark) levels calculated based on 1000 bootstrap samples.

Table 4 shows correlations between five selected indices, which are thematically closest in the two data sets (aggregate indices on the US, UK and China and two policy-related indices for the US). For four of the five indicators the diagonals indicate a significant positive correlation between the EPU indices and our measures. For China, the correlation is insignificant.

Overall, given the methodological, thematic and sample differences between the two indicator sets, these weak, but positive and statistically significant correlations appear reasonable.

[26] The website of the authors, http://www.policyuncertainty.com, publishes indices for several countries and several topics along with related research and description of methodology.

5 Validation: Do News Indices Extract Fundamental Information?

We have argued that our news indices represent fundamentals by construction: news indices aggregate tonality scores of fundamental expressions and each fundamental expression has topical expression element(s) that refer to one of the seven fundamentals. This, however, does not guarantee that our expression rules will efficiently pick up many occurrences of fundamental mentions in the text, nor does it guarantee that it will correctly identify the tonality related to the fundamental topical expressions that it manages to identify. In case that expression rules did not manage to pick up enough fundamental expressions, we would recover sparse news indices with small cross-sectional and time series variation. (In the previous section we have already seen this not to be the case.) In case that tonality scores were wrongly matched with topical expressions, we would recover news indices, whose time series and cross-sectional variation had no relation to variation in other proxies of fundamentals.

To go into the possible caveats in more detail, logically, there are two things that may go wrong. One possibility is that expression rules are inefficient to extract enough fundamental information from the text. Whether expression rules seem reasonable at all may be directly checked by glancing through the list of expression definitions in Appendix Table C.5 and C.6. These tables are organized by fundamentals and thematic topics within each fundamental concept to facilitate such a review.[27] The other way to check whether there are enough such expression rules and that these rules are efficient in capturing fundamental information, is to examine whether news indices correlate with other, objective proxies of fundamentals, which is what we are set to do in this section.

The other possible caveat is that even if expression rules recovered many fundamental topical expressions and their tonality pairs, the rules for maximum word distances could be inefficient in identifying topic – tonality pairs with sufficient precision. Proximity rules between expression elements (topic and tonality keywords) may not prevent them from being unrelated to each other. In particular, a distance rule that is set too wide may confound the expression and result in false positive cases (e.g. the tonality verbs or adjectives may refer to another noun not the one representing the fundamental topic). Distance rules do help, but setting a distance rule that is too strict results in a large number of false negatives: it will not return fundamental expression matches even when there is one in the text.

This trade-off is common to all automated text mining approaches. The false negative problem is most acute in the extreme method that restricts expression elements to be neighboring each other (e.g. searching for matches of ’housing market bust’, ’housing bubble burst’). The false positive problem is most relevant in the other extreme case, which only requires that all expression elements be present in the text irrespective of their relative location (e.g. searching for ’housing’ and ’bust’ separately and calling the article a match if both are found). Our approach balances in between these extremes. Nonetheless, being a statistical approach it is still subject to these problems. Therefore our news indices will be noisy measures of fundamentals even if expression rules are valid. The question is a matter of precision: whether the magnitude of this noise is such that it swamps all information in our indices or it is relatively small and relevant information dominates the indices.

5.1 NEWS INDICES VS BLOOMBERG ECONOMIC ANNOUNCEMENTS

In the Introduction we claimed that since news agencies compete for investor attention, they will include information about fundamentals (e.g. tail risk scenarios, expectations) over and above that contained in traditional macroeconomic variables and would underweight information in published indicators that is uninteresting for investors. Because both traditional indicators and news do contain information about fundamentals, however, we would expect a positive association between them even if it is weak.

[27] The background to our fundamental definitions are available in our Coding Guide (online appendix).

The most direct test that we can think of is to look at the association between changes in our news indices and the surprise components of macroeconomic announcements. We first carry out such tests for the five largest economies (US, UK, China, Japan and Germany) in terms of news indices and pertaining to the REAL fundamental category. News is frequent for these countries and this fundamental category even on the daily level and Bloomberg also regularly publishes several indicators of these countries related to the real economy (GDP growth rates, industrial production, retail sales, to name a few) and analyst surveys on expected data prior to publication for several of these. Transforming surprises within different Bloomberg macroeconomic announcements into a comparable format we can assess whether the surprise content of these announcements is significantly and positively correlated with (daily) changes in our REAL news index.[28]

Table 5 reports correlations. The top panel of the table shows that Bloomberg surprises were significantly and positively correlated with news indices for all five countries (panel A, diagonal elements). Off-diagonal elements were not significant however, so that a positive surprise in one country did not result in improved news indices for the other four.

Table 5
Correlations between Bloomberg announcements and news indices (REAL fundamental category)

News indices /news counts	Bloomberg surprises / Bloomberg number of announcements: US	CHINA	UK	JAPAN	GERMANY
	PANEL A: Bloomberg surprises and news indices				
US	0.121	-0.08	0.023	0.013	-0.013
CHINA	0.020	0.268	-0.013	-0.010	-0.007
UK	0.019	0.07	0.218	0.019	0.050
JAPAN	0.009	0.075	-0.024	0.139	0.107
GERMANY	-0.044	0.001	0.041	-0.003	0.303
	PANEL B: Bloomberg announcements and news counts				
US	0.629	0.124	0.046	-0.018	0.119
CHINA	0.110	0.574	0.065	0.041	0.046
UK	0.108	0.109	0.525	0.009	0.029
JAPAN	0.027	-0.018	0.057	0.476	0.043
GERMANY	0.101	0.000	0.021	-0.048	0.554
	PANEL C: Bloomberg surprises and news SURPRISE subindices				
US	0.181	-0.040	-0.022	0.041	-0.035
CHINA	0.016	0.405	-0.015	0.005	-0.005
UK	0.017	0.196	0.232	0.000	0.023
JAPAN	0.030	0.070	0.049	0.288	0.089
GERMANY	-0.045	0.076	0.098	0.013	0.318

Sources: Reuters news archives, Bloomberg and authors' calculations.
Notes: Pearson's bivariate correlations between daily changes of news indices (or news counts) and Bloomberg surprises (or number of announcements). Surprises are calculated as the published data (actual) minus the analyst survey corrected for sign and variance of the series (see Section 2.2). Days without a Bloomberg announcement or fundamental expression are eliminated from the sample since these may spuriously increase correlations. Note that different number of eliminated observations across countries and fundamental category influences critical values. Correlations are shaded according to 5% (light) and 1% (dark) levels calculated based on 1000 bootstrap samples.

[28] See Appendix B and Section 2 for data used and calculation method. Note that we could have also looked at levels of published macro data relative to time series averages or compared to previous months' values. Levels data are very persistent. Taking first differences however introduces noise with respect to which macroeconomic indicator was published on different days (for example even when normalized for indicator-specific variance, the relative level of employment data could be far from the relative level of housing market data, therefore daily changes would be influenced by selection on which of these were published). Changes compared to the previous values would need to be compared to changes of news indices on the same time scale. But some indicators are weekly, some are monthly, some are quarterly, which makes such comparisons less straightforward.

Panel B examines correlations between Bloomberg announcements and news counts. Correlations appear large on the diagonal meaning that days with Bloomberg announcements was strongly associated with an increased number of REAL fundamental expressions in the respective country. The larger correlations compared to panel A could be a consequence of two factors. On one hand, it may suggest that fundamental topics are much easier to pin down than topics and tonality jointly. Obviously there is more noise in our tonality-included news indices than in simple fundamental expression news counts. An alternative explanation is that news index correlations are lower because of the extra information carried within the media discussion relative to macroeconomic announcements. Similar values of surprises may be assessed quite differently by the media depending on the concrete circumstances (e.g. what type of indicators were published, whether the index is close to perceived threshold values, whether the surprise of the day matches previous tendencies). News counts, though clearer measures, lose all information related to these deliberations.

Table 6
Correlations between Bloomberg announcements and news indices (various fundamentals)

	Bloomberg surprises				
News indices	US	CHINA	UK	JAPAN	GERMANY
	PANEL A: Bloomberg (REAL) surprises and News indices				
REAL	0.121	0.268	0.218	0.139	0.303
EXTERN	0.025	-0.057	-0.010	-0.017	0.059
FISCAL	-0.030	-0.107	0.005	0.023	-0.026
FUND_LIQ	-0.037	0.016	-0.051	-0.065	0.019
BANK	0.001	0.035	0.041	0.020	0.034
POL_INST	0.013	0.106	-0.042	0.036	0.017
MON_POL	0.018	-0.113	0.047	0.002	-0.063
	PANEL B: Bloomberg (EXTERN) surprises and News indices				
REAL	0.081	0.082	-0.048	-0.029	-0.004
EXTERN	0.413	0.110	0.367	0.115	0.154
FISCAL	0.341	0.194	0.126	0.019	0.000
FUND_LIQ	-0.017	0.035	-0.095	-0.094	-0.183
BANK	-0.009	-0.099	0.027	-0.079	-0.025
POL_INST	-0.011	0.025	0.001	0.002	-0.076
MON_POL	0.048	-0.086	-0.061	-0.053	0.146

Sources: Reuters news archives, Bloomberg and authors' calculations.
Notes: Pearson's bivariate correlations between daily changes of news indices (main index) and Bloomberg surprises of the same country. Surprises are calculated as the published data (actual) minus the analyst survey corrected for sign and variance of the series (see Section 2.2). Days without a Bloomberg announcement or fundamental expression are eliminated from the sample, since these may spuriously increase correlations. Note that different number of eliminated observations across countries and fundamental category influences critical values. Correlations are shaded according to 5% (light) and 1% (dark) levels calculated based on 1000 bootstrap samples.

The bottom panel reports correlations between Bloomberg announcement surprises and news indices similar to panel A with the difference that instead of the main news indices, the SURPRISE subindices are considered. As discussed in Section 3.4 and 3.6, this subindex aggregates fundamental expressions that refers to changes of fundamentals compared to expectations ('GDP unexpectedly increased', 'GDP increased more than anticipated', etc.). As expected, the SURPRISE subindices have larger correlations with Bloomberg surprises than the main news indices, consistent with their more immediate connection.

Table 6 expands the analysis to other fundamentals. Bloomberg macroeconomic fundamentals were available for the REAL and EXTERN categories[29] and correlations of their surprises were examined in light of all seven fundamental news indices for the same five countries.

[29] Announcements are also available for budget balances but these are so infrequent that number of observations are severely limited. Monetary policy rate announcements were also available, but were uninformative as interest rates and their expectations have been constant in most of the sample.

The results support the idea that macroeconomic announcement surprises resulted in changes in the appropriate fundamental news index. Positive surprises in REAL macroeconomic indicators were consistent with improvements in the REAL news indices of the respective country (significant at the 1% level), but did not materially change other fundamental news indices of the country. Similarly, for each country except China, unexpected improvements in Bloomberg external trade reports were associated with improvements in the EXTERN news index. Even in China however, although insignificant, the correlation coefficient was positive. Again, macro surprises related to external trade correlated primarily with the EXTERN news indices and left other fundamental indices unmoved. Exceptions in the US and UK were that trade surprises also significantly correlated with fiscal policy deliberations in the news, perhaps a consequence of media attention about twin deficits.

5.2 NEWS INDICES VS NON-FUNDAMENTAL PROXIES

Because we claim that our news indices are noisy aggregates of fundamental information, it is worth running a sort of placebo test to check whether our indices are correlated with measures of non-fundamental information.

Table 7
Correlations of news indices and non-fundamental information proxies

PANEL A: News (main) index

	HPW Noise index[a]		BW SENT index[b]	
	correl.coef.	p-value	correl.coef.	p-value
REAL	-0.018	0.390	0.093	0.338
EXTERN	0.013	0.515	-0.029	0.796
FISCAL	0.010	0.693	-0.035	0.718
FUND_LIQ	-0.027	0.415	-0.081	0.404
BANK	-0.031	0.308	-0.013	0.851
POL_INST	0.013	0.566	0.065	0.382
MON_POL	-0.007	0.790	0.002	0.937

PANEL B: News CONCERN subindex

	HPW Noise index[a]		BW SENT index[b]	
	correl.coef.	p-value	correl.coef.	p-value
REAL	-0.017	0.468	-0.121	0.162
EXTERN	0.003	0.933	0.160	0.097
FISCAL	0.042	0.163	-0.044	0.623
FUND_LIQ	-0.036	0.159	-0.032	0.736
BANK	-0.044	0.175	0.191	0.153
POL_INST	0.000	0.981	0.163	0.051
MON_POL	0.018	0.654	0.018	0.689

Sources: Websites of Jun Pan and Jeffrey Wurgler, Reuters news archives, and authors' calculations.
Notes: Pearson's bivariate correlations between daily changes of US news indices and the HPW noise index and monthly changes of US news indices and the BW sentiment index. Panel A reports correlations for the seven fundamental main indices, panel B reports correlations for seven fundamental CONCERN subindices. P-values of correlations are calculated based on 1000 bootstrap samples.
[a)] The HPW noise index (Hu et al., 2013) is a measure of US bond market liquidity.
[b)] The BW SENTIMENT index (Baker and Wurgler, 2006) is the principal component of six popular investor sentiment proxies.

We test our indices against two such measures: the Hu et al. (2013) (HPW) noise measure and the Baker and Wurgler (2006) (BW) SENTIMENT sentiment proxy. The HPW noise index is in effect a measure of funding liquidity in the US government securities market and is computed from the difference between bond yields and yields implied by fitted yield curves. Hu et al. (2013) argue and empirically demonstrate that this error is large when bond market funding liquidity is lower, because this hinders arbitrage and the yield curve will be more jagged across the maturity spectrum. Although liquidity may have a distant relation to US fundamentals, we would expect for these to play out over longer horizons, and would not expect daily changes in such a measure to be related to our indices. The BW SENTIMENT index of Baker and Wurgler (2006) is a monthly index and is the principal component of five underlying proxies for investor sentiment including average discounts on closed-end funds, number of IPOs, first-day returns of IPOs, equity-to-debt issuance ratio, market-to-book ratio difference of payers and defaulters. We consider this indicator as a proxy of a different aspect of non-fundamentals than the HPW index.

Table 7 displays correlations between these measures and our US news indices. Panel A reports correlations with the main indices, panel B with the CONCERN subindex. Bootstrapped sampling of correlation coefficients show that observed correlations are statistically insignificant regarding both the HPW and the BW non-fundamental proxies. All correlations with the CONCERN subindices are also insignificant. This is important especially in light of the BW SENTIMENT measure because one could think that a possible failure of this subindex would be for it to pick up general concerns, worries of investors and not the specific type of concerns that are only related to fundamentals.

6 News indices in sovereign credit spread regressions

Returning to our original research question, we are interested in whether our constructed news indices have additional explanatory power about sovereign credit spreads when controlling for variables customarily used as proxies for fundamentals.

The empirical literature has been versatile regarding model specifications. The literature provided evidence of local macroeconomic factors being important in the cross-sectional variation of levels of credit spreads. Such a relationship is however not obvious on the changes of spreads, where a common systemic factor appears more important (Ang and Longstaff, 2011; Longstaff et al., 2011), therefore we choose this latter, first differences, specification.[30]

Studies have also been split over choosing sovereign CDS spreads and bond spreads as the dependent variable and whether to model these in a linear or a logarithmic specification. We choose the logaritmic specification[31] and report all results on the CDS spreads. A key benefit of CDS spreads is that they allow a larger cross-section than bond spreads. We also carry out all analysis on euro- and dollar-denominated bond spreads (EMBI Global spreads and 5-year spreads between bond yields of eurozone countries and the same maturity German benchmark). The main results are reported for bond spread regressions as well. Other, not published results on these variables are in line with CDS spreads and are available to the reader upon request.

Our panel regression specification is as follows:

$$\begin{aligned} \Delta log(S_{it}) &= \alpha + \beta^{NG}\Delta(NG_t) + \beta^{NL}\Delta(NL_{it}) + \beta^{X}\Delta X_{it} + \epsilon_{it} \\ NG_t &= \sum_{i=1}^{N} NL_{it} \end{aligned} \tag{3}$$

where α represents a constant[32], β are parameters denoting sensitivities to exogenous variables. $\Delta(NL_{it})$ and ΔX_{it} are vectors that represent local news indices and macroeconomic data of country i and time period t. The vector NG_t in turn denotes news indices aggregated across all countries in the news database and are intended proxies for global fundamentals at the given point in time. Throughout the paper we use double-clustered robust standard errors (see Thompson, 2011) that allow for arbitrary forms of cross-sectional and time series dependence in the residuals, ϵ_{it}.

[30] Augustin (2014) reviews the literature with respect to global versus local determinants of sovereign CDS spreads and claims that global factors are more important on higher trading frequencies. He also asserts that the relationship is time-varying: local factors, especially those related to the sovereign-banking nexus, have become more important as of the financial crisis. Kocsis and Monostori (2016) compare the role of domestic and global factors with respect to modelling the relationship between spreads and determinants in levels or first differences.

[31] Emerging market studies initially followed the tradition of Edwards (1983) in using logarithms on the grounds of a hazard model between fundamentals and probability of default. The literature dealing with eurozone countries (convergence prior to the financial crisis and sovereign risk concerns afterwards) however chose the linear format, probably because of the occurrences of negative spreads: several countries had interest rates below the benchmark German Bunds used to proxy risk-free rates. In our data set, taking logarithm seems to result in a specification with less heteroscedasticity in the error terms as the volatility of sovereign spreads is strongly associated with spread levels.

[32] Note that there is no need for cross-section fixed effects, theoretically there is no reason to assume (heterogenous) trends in sovereign spreads.

Table 8
Regressions of sovereign CDS spread changes

Model specification:	(A): Macro only		(B) News only		(C) Macro and News		(D) add CONCERN	
Dependent variable:	Δlog(CDS)		Δlog(CDS)		Δlog(CDS)		Δlog(CDS)	
Explanatory	coef.	std.err.	coef.	std.err.	coef.	std.err.	coef.	std.err.
Main indices (Global)								
ΔREAL			-0.070*	(0.041)	-0.080	(0.063)	0.047	(0.080)
ΔEXTERN			-0.942**	(0.382)	-1.010*	(0.526)	-1.260***	(0.458)
ΔFISCAL			-0.610***	(0.154)	-0.502**	(0.198)	-0.180	(0.201)
ΔFUND_LIQ			-0.595**	(0.232)	-0.563**	(0.250)	-0.032	(0.298)
ΔBANK			-0.280	(0.407)	-0.579	(0.631)	1.080*	(0.573)
ΔPOL_INST			-0.129**	(0.059)	-0.132*	(0.076)	-0.020	(0.065)
ΔMON_POL			0.173*	(0.097)	0.260*	(0.133)	-0.096	(0.124)
Main indices (Local)								
ΔREAL			-0.184***	(0.049)	-0.150***	(0.031)	-0.147	(0.117)
ΔEXTERN			0.057	(0.259)	0.366	(0.840)	-0.190	(0.234)
ΔFISCAL			0.462	(0.394)	0.539	(0.383)	0.525	(0.344)
ΔFUND_LIQ			-0.323	(0.690)	-0.377	(0.688)	-0.881	(1.404)
ΔBANK			-0.681	(0.456)	-0.535	(0.529)	-0.049	(0.764)
ΔPOL_INST			-0.197	(0.154)	-0.074	(0.147)	-0.111	(0.106)
ΔMON_POL			0.016	(0.073)	-0.018	(0.145)	-0.098	(0.125)
CONCERNS subindices (Global)								
ΔREAL							-0.775*	(0.407)
ΔEXTERN							-1.910	(1.698)
ΔFISCAL							-3.230	(2.206)
ΔFUND_LIQ							-1.370	(1.118)
ΔBANK							-2.940*	(1.636)
ΔPOL_INST							0.292	(0.530)
ΔMON_POL							-0.674	(1.060)
CONCERNS subindices (Local)								
ΔREAL							-0.140	(0.475)
ΔEXTERN							-3.910	(7.541)
ΔFISCAL							0.807	(2.006)
ΔFUND_LIQ							2.130	(2.828)
ΔBANK							-2.770	(1.810)
ΔPOL_INST							0.536	(1.930)
ΔMON_POL							0.972	(0.847)
Traditional macro var's								
GDP growth	-1.008	(0.756)			-0.424*	(0.239)	-0.375	(0.247)
ΔCurrent Acc	-2.053	(1.268)			-0.783	(0.950)	-0.827	(0.831)
ΔReserves	-2.567**	(1.136)			-1.319**	(0.626)	-0.844**	(0.417)
ΔFiscal Bal	-1.200**	(0.489)			-0.731	(0.677)	-0.840	(0.661)
ΔGov't Debt	-0.656	(0.759)			0.281	(0.201)	0.375***	(0.121)
Global macro var's								
World GDP growth					-0.054	(3.784)	5.934	(5.775)
PC GDP growth					-0.086	(0.248)	-0.173	(0.177)
ΔPC Current Acc					2.034	(4.409)	4.742	(3.354)
ΔPC Reserves					-0.464	(0.830)	-0.805	(0.901)
ΔPC Fiscal Bal					-0.241	(0.636)	0.724	(0.621)
ΔPC Gov't Debt					-0.018	(0.184)	-0.310	(0.352)
R-squared	0.053		0.356		0.422		0.541	
Adj. R-squared	0.050		0.351		0.411		0.528	
No. time periods	31		33		30		30	
No. cross-sections	49		58		49		49	
No. observations	1463		1867		1416		1416	

Sources: News indices based on Reuters news archives and authors' calculations. Other variable sources listed in Appendix A.

Notes: Pooled regressions (constant only, no fixed effects) of sovereign CDS spread log changes on news indices (main index and CONCERN subindices) and macroeconomic variables. Global news indices aggregate news indices across all geographies, local news indices are specific to the respective country. News indices are specified in quarterly changes. Traditional macro variables: quarterly real GDP growth rates, changes in current account balance to GDP, changes in official reserves to GDP, changes in (IMF 1-year ahead projected) fiscal balance to GDP, changes in the public debt to GDP. Global macro variables: the annual world GDP growth rate and first principal components of traditional variables. For visualization purposes data are scaled: all news indices multiplied by 10000, macro variables are multiplied by 100. Double-clustered standard errors and usual significance levels are reported.

Table 9
Regressions of sovereign bond spread changes

Model specification:	(A): Macro only		(B) News only		(C) Macro and News		(D) add CONCERN	
Dependent variable[a]	Δlog(FXB+c)		Δlog(FXB+c)		Δlog(FXB+c)		Δlog(FXB+c)	
Explanatory	coef.	std.err.	coef.	std.err.	coef.	std.err.	coef.	std.err.
Main indices (Global)								
ΔREAL			-0.047	(0.033)	-0.048	(0.046)	0.075	(0.055)
ΔEXTERN			-0.662**	(0.321)	-0.570	(0.385)	-0.676***	(0.248)
ΔFISCAL			-0.373***	(0.135)	-0.333***	(0.128)	-0.128	(0.154)
ΔFUND_LIQ			-0.394**	(0.175)	-0.338*	(0.181)	0.033	(0.206)
ΔBANK			-0.178	(0.135)	-0.084	(0.166)	0.329*	(0.183)
ΔPOL_INST			-0.103**	(0.048)	-0.089	(0.057)	-0.039	(0.054)
ΔMON_POL			0.097**	(0.048)	0.074	(0.049)	-0.040	(0.046)
Main indices (Local)								
ΔREAL			-0.751	(0.529)	-3.980***	(0.914)	-3.210***	(0.909)
ΔEXTERN			-0.086	(0.502)	0.724	(1.096)	-0.055	(0.592)
ΔFISCAL			-0.086	(0.412)	0.832	(0.800)	0.076	(0.964)
ΔFUND_LIQ			-0.620	(0.450)	-0.431	(0.408)	1.830***	(0.568)
ΔBANK			-0.468	(2.557)	2.680	(2.392)	2.360	(2.161)
ΔPOL_INST			-0.134	(0.221)	0.066	(0.249)	0.227**	(0.105)
ΔMON_POL			-0.657	(0.634)	-2.030**	(0.903)	-1.370	(1.039)
CONCERNS subindices (Global)								
ΔREAL							-0.620***	(0.170)
ΔEXTERN							-0.845	(1.192)
ΔFISCAL							-2.590**	(1.033)
ΔFUND_LIQ							-0.454	(0.835)
ΔBANK							-0.853	(0.856)
ΔPOL_INST							0.062	(0.349)
ΔMON_POL							-0.151	(0.567)
CONCERNS subindices (Local)								
ΔREAL							-8.940	(5.639)
ΔEXTERN							10.250**	(4.085)
ΔFISCAL							1.870	(1.333)
ΔFUND_LIQ							-7.160**	(3.149)
ΔBANK							7.830***	(3.006)
ΔPOL_INST							-1.980	(1.209)
ΔMON_POL							-4.470	(6.797)
Traditional macro var's								
GDP growth	-0.406	(0.424)			-0.288	(0.261)	-0.345	(0.275)
ΔCurrent Acc	-2.071	(1.339)			-0.704	(0.619)	-0.692	(0.550)
ΔReserves	-4.361***	(1.004)			-3.244***	(0.642)	-2.724***	(0.437)
ΔFiscal Bal	0.170	(1.195)			0.997	(0.861)	1.370**	(0.667)
ΔGov't Debt	-0.060	(0.409)			0.324	(0.222)	0.176	(0.254)
Global macro var's								
World GDP growth					1.974	(1.917)	7.120***	(1.594)
PC GDP growth					0.110	(0.155)	-0.046	(0.143)
ΔPC Current Acc					2.487	(3.164)	3.953*	(2.225)
ΔPC Reserves					-0.326	(0.528)	-0.439	(0.563)
ΔPC Fiscal Bal					-0.666***	(0.217)	-0.056	(0.357)
ΔPC Gov't Debt					-0.022	(0.118)	-0.328*	(0.168)
R-squared	0.082		0.287		0.369		0.482	
Adj. R-squared	0.078		0.279		0.352		0.460	
No. time periods	36		37		34		34	
No. cross-sections	30		37		30		30	
No. observations	1011		1311		956		956	

Sources: News indices based on Reuters news archives and authors' calculations. Other variable sources listed in Appendix A.
Notes: Pooled regressions (constant only, no fixed effects) of euro- and dollar-denominated sovereign bond spread (FXB) log changes on news indices (main index and CONCERN subindices) and macroeconomic variables. Bond spreads are EMBI Global spreads and 5-year eurozone interest rate spreads over Bunds. Regressors are as described in Table 8. Double-clustered standard errors and usual significance levels are reported.
[a] *A correction of 50 basis points is made to all bond spreads to avoid non-positive cases.*

Table 10
Global and local determinants of sovereign CDS spread changes

Model specification:	(A): Local Macro		(B) Add Local News		(C) Global Macro		(D) add Global News	
Dependent variable:	Δlog(CDS)		Δlog(CDS)		Δlog(CDS)		Δlog(CDS)	
Explanatory	coef.	std.err.	coef.	std.err.	coef.	std.err.	coef.	std.err.
Main indices (Global)								
ΔREAL							0.032	(0.074)
ΔEXTERN							-1.170***	(0.429)
ΔFISCAL							-0.118	(0.188)
ΔFUND_LIQ							0.009	(0.284)
ΔBANK							1.010*	(0.529)
ΔPOL_INST							-0.028	(0.065)
ΔMON_POL							-0.093	(0.114)
Main indices (Local)								
ΔREAL			-0.563***	(0.100)				
ΔEXTERN			-0.661	(1.370)				
ΔFISCAL			0.282	(0.439)				
ΔFUND_LIQ			-0.927	(1.480)				
ΔBANK			0.571	(1.393)				
ΔPOL_INST			-0.177	(0.175)				
ΔMON_POL			0.283*	(0.156)				
CONCERNS subindices (Global)								
ΔREAL							-0.751**	(0.373)
ΔEXTERN							-2.000	(1.664)
ΔFISCAL							-3.210	(2.070)
ΔFUND_LIQ							-1.410	(1.090)
ΔBANK							-2.890*	(1.530)
ΔPOL_INST							0.356	(0.500)
ΔMON_POL							-0.673	(0.993)
CONCERNS subindices (Local)								
ΔREAL			0.670	(0.573)				
ΔEXTERN			-21.070	(13.210)				
ΔFISCAL			-0.614	(2.888)				
ΔFUND_LIQ			2.560**	(1.050)				
ΔBANK			-9.670***	(3.494)				
ΔPOL_INST			-0.293	(2.594)				
ΔMON_POL			3.160**	(1.489)				
Traditional macro var's								
GDP growth	-1.008	(0.756)	-0.977	(0.711)				
ΔCurrent Acc	-2.053	(1.268)	-1.875	(1.214)				
ΔReserves	-2.567**	(1.136)	-2.538**	(1.110)				
ΔFiscal Bal	-1.200**	(0.489)	-1.042*	(0.532)				
ΔGov't Debt	-0.656	(0.759)	-0.544	(0.685)				
Global macro var's								
World GDP growth					4.491	(4.805)	6.315	(5.337)
PC GDP growth					-0.280	(0.330)	-0.233	(0.166)
ΔPC Current Acc					1.110	(3.476)	4.689	(3.164)
ΔPC Reserves					-0.770	(1.033)	-0.968	(0.894)
ΔPC Fiscal Bal					-0.141	(1.322)	0.723	(0.597)
ΔPC Gov't Debt					-0.392	(0.277)	-0.305	(0.332)
R-squared	0.053		0.082		0.103		0.517	
Adj. R-squared	0.050		0.070		0.100		0.511	
No. time periods	31		31		30		30	
No. cross-sections	49		49		58		58	
No. observations	1463		1463		1699		1699	

Sources: News indices based on Reuters news archives and authors' calculations. Other variable sources listed in Appendix A.
Notes: Notes: Pooled regressions (constant only, no fixed effects) of sovereign CDS spread log changes on news indices (main index and CONCERN subindices) and macroeconomic variables. Bond spreads are EMBI Global spreads and 5-year eurozone interest rate spreads over Bunds. Regressors are as described in Table 8. Double-clustered standard errors and usual significance levels are reported.

Table 8 reports our main estimation results. In the first column (specification A) we only include traditional macro variables (real GDP growth, current account and expected fiscal balance changes, and changes in key stock measures: the government debt ratio and central bank reserves). Most variables have the expected negative sign, but overall the estimates confirm that these variables only explain a small portion of spread variation, with an R^2 of only around 5 percent. In contrast, the second column (specification B) that includes only changes in our main news (both global and local) indices explains around 35 percent of variation in the data. Global news indices have the expected negative sign (increase in news index denotes improvement) with the exception of monetary policy (recall that a decrease here denotes monetary expansion), while local news indices are mostly insignificant.

The third column (specification C) includes traditional and global macro variables as well as news indices (global macroeconomic variables are the world real GDP growth rate and first principal components of domestic macro variables).[33] Together, these variables explain around 40 percent of the variance. Again, most global news indices appear statistically important and have an intuitive sign, while local news indices are largely insignificant. These results suggest that once we use our fundamental news indices, a substantial amount of spread variation can indeed be linked to fundamental information. Further, the effect of news indices mainly seems to be exerted through its global component.

Standard finance theory asserts that asset prices weigh the distributions of future expected outcomes. A particular strength of our method compared to traditional macro news or survey expectations is that we can create news indices that are likely to identify concerns of investors about fundamentals, which are likely to be related to tail risks. A line of research (Barro, 2006; Gabaix, 2012; Wachter, 2013) suggests that low probability high impact scenarios, rare disasters, must be important in asset pricing. As discussed in Section 3 our CONCERN subindices try to pick up such concerns. Hence, the fourth column (Specification D) conveys the additional explanatory power of these subindices, which we think of as proxies for tail risk scenarios regarding future fundamentals. Overall the results seem to support the importance of tail risk considerations for sovereign spreads. The overall explanatory power of the regression goes up substantially, from 42 to over 54 percent. The significance levels of individual regressors however decrease due to collinearity between the main news index and subindices.

To ensure that our results are not specific to the CDS market, Table 9 reports similar specifications for bond spread regressions. Note that here we have added 50 basis points to all spreads before taking logarithms to deal with negative spreads. The results are qualitatively similar to the CDS regressions, with the only notable difference that traditional macro variables explain a bit more of the bond spread variation compared to CDSs.

Table 10 investigates whether global or local fundamentals appear more important in sovereign spread variation. The first two columns explore the effects of local fundamentals, while the last two columns assess the effects of global fundamentals. As already seen in Table 8, traditional domestic fundamental variables explain only a marginal part of the information in CDS spread changes. Our local news indices (both the main index and CONCERNS subindex series) do not add much additional information and only marginally lift the share of explained variance.

In contrast, global fundamentals appear much more important especially when our news indices are included in the specification. Global proxies based on macroeconomic variables explain 10 percent of the CDS spread variation, already double the amount grasped by local macro factors. However this is still just a fraction of the overall variation that we may attribute to global fundamentals once we include global news indices. Together with our news indices, the R^2 statistic jumps four-fold to over 50 percent.

These results speak to the debate in the sovereign credit risk literature about whether global or local factors are more important in explaining sovereign spread variation (see Augustin, 2014). Our results strongly support the majority view that within changes of sovereign credit spreads, global factors appear much more important than local factors.

We view our key addition to this debate in that we relate directly to the fundamental component of these spreads and find that the global factors still keep their dominance over local factors. The empirical literature that studied the variance shares attributable to global versus local factors customarily used financial indicators to proxy the global component (the VIX index, corporate spreads and composite indices or the principal components of sovereign spreads themselves). Because financial

[33] Since there are no commonly used proxies for common trends in external and fiscal balances, we simply extract the first principal components of the traditional macroeconomic variable series.

indicators have both a fundamentals-related component and a component related to risk pricing and liquidity, it is not clear from other papers, which of these two components is responsible for explaining the large systemic co-movement found in sovereign credit spread dynamics. Notably, it could be that non-fundamentals, such as general investor sentiment are behind all these movements. Our analysis suggests that much of the common variation is due to factors related to fundamentals. In the following, we estimate a more explicit decomposition of two key systemically important financial variables along the fundamentals/non-fundamentals dimension.

6.1 DRIVERS OF THE VIX INDEX AND IMPLICATIONS FOR SOVEREIGN CREDIT RISK

Last, we turn to a common specification in the empirical literature, which includes both traditional macroeconomic variables and financial variables. Namely, we include two such US financial variables, the CBOE VIX index (the 3-month ATM implied volatility of options on the US S&P-500 stock index) and the CSI US corporate high yield index, which is the spread between the yield of a portfolio of BBB/Baa-rated corporate bonds and the 10-year US Treasury bond.

Table 11
Sovereign CDS Regressions with US financial variables and its components

Model spec	(A) Financials		(B) add Macro		(C1) Decomp1		(C2) Decomp2		(C3) Decomp3		(D) Fundam. content	
Dependent	Δlog(CDS)		Δlog(CDS)		Δlog(CDS)		Δlog(CDS)		Δlog(CDS)		Δlog(CDS)	
Exploratory	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
US financials												
ΔVIX	0.864***	(0.227)	0.863***	(0.260)								
ΔHY	32.374***	(2.298)	33.173***	(2.587)								
US financials: fundamental content												
$\widehat{\Delta VIX}$					1.618***	(0.465)	1.657**	(0.665)			1.782***	(0.663)
$\widehat{\Delta HY}$					27.471***	(7.065)	24.010**	(11.820)			23.455**	(11.060)
US financials: non-fundamental content												
$\Delta VIX - \widehat{\Delta VIX}$					0.305	(0.307)			0.360	(0.899)		
$\Delta HY - \widehat{\Delta HY}$					31.194***	(1.998)			29.502***	(8.476)		
Traditional macro var's												
GDP growth			-0.293	(0.216)							-1.008*	(0.547)
ΔCurrent Acc			0.064	(1.202)							-1.126	(1.052)
ΔReserves			-0.454	(0.348)							-2.020**	(0.799)
ΔFiscal Bal			-0.437*	(0.257)							-0.975**	(0.445)
ΔGov't Debt			0.167	(0.312)							0.054	(0.486)
R-squared	0.509		0.530		0.531		0.299		0.200		0.355	
Adj. R-squared	0.509		0.528		0.530		0.298		0.199		0.351	
No. time periods	31		29		31		32		31		30	
No. cross-sections	58		49		58		58		58		49	
No. observations	1753		1367		1753		1810		1753		1415	

Sources: News indices are based on Reuters news archives and authors' calculations. Other variable sources are listed in Appendix A.
Notes: Pooled regressions (constant only, no fixed effects) of sovereign CDS spread log changes on US financial indicators (CBOE VIX index, CSI High yield index) and traditional macroeconomic variables. In specifications C1–D components of US financial indicators are used: financial indicators are regressed on CONCERN news subindices. Linear projections based on such regressions are referred to as the fundamental content of financial indicators, whereas regression residuals are referred to as their non-fundamental content. Macroeconomic variables included are quarterly real GDP growth rates, changes in the current account balance to GDP, changes in central bank reserves to GDP, changes in (IMF 1-year ahead projected) fiscal balance to GDP and changes in the public debt to GDP ratio.

The interpretation of global (or globally important US) financial variables are somewhat vague in the literature and in the media, though they are generally thought of as related to global investor concerns. (In business parlance the VIX index is often referred to as the "fear gauge", for instance.) Whether this does include concerns about future possible paths of fundamental variables or it rather refers to investor sentiment unrelated to fundamentals is not clear. Our news index measures provide a tool to investigate this issue.

Table 11 shows results of regressions of changes of sovereign CDS spreads with changes in the VIX index and the US high-yield spread (or its components) in the regressor list. The two variables explain about 50 percent of the variation (specification A). In specification B, we add macroeconomic variables, which barely raises explanatory power. Recall that in Tables 8 and 9 we estimated a specification with macroeconomic variables only (specification A) and we found that macroeconomic variables alone accounted for a meagre 5 and 8 percent of the variation in sovereign CDS and bond spreads, respectively. All this supports the usual result of the empirical literature that global fluctuations rooted in financial markets are responsible for much of the time series variation in spreads.

Next, we decompose US financial indicators into a fundamental and a non-fundamental component. We regress both the VIX index and the high yield spread on our seven CONCERN subindices. We use the projections based on news indices as the fundamental component and the regression residuals as the non-fundamental component.[34]

Specifications C1–C3 includes these components in sovereign CDS spread regressions. The estimates convey the message that more than half of the explanatory power of US financial indicators derive from expectations and concerns about the future of global fundamentals. Non-fundamentals are still statistically and economically important in CDS spreads comprising about 20 percent of the variation. Specification D collects all explanatory variables of fundamental content (traditional macroeconomic variables and the fundamental components of financial indicators) and finds that these explain 35 percent of the sovereign spread variation.

In sum, a significant part of sovereign credit spread changes can be explained by fundamentals as reflected by the explanatory power of specifications with news-based measures of fundamentals even though much of this explained variance relates to global and not local fundamentals. US financial variables often used as explanatory variables of systematic movements in sovereign spreads derive a larger part of their significance from the outlook and risks surrounding global fundamentals, whereas a smaller part of their significance may reflect general investor sentiment unrelated to fundamentals.

[34] This method underestimates the fundamental content in these financial indicators due to the noise in the news indices. Regression residuals understood here as non-fundamentals may still have further fundamental-related content.

7 Conclusions

Estimates on the role that fundamentals play in asset prices depend on the quality of proxies used for fundamentals. We argued that traditional macroeconomic indicators are imperfect, because they are backward-looking in nature, are not available for many aspects of country fundamentals including the possibility of tail risk scenarios that may be important for pricing. We also claimed that news articles contain information on such omitted factors, because journalists gather and summarize available information on these matters.

The first contribution of our research is a novel method that enables better extraction of such information. The extant literature proposes adequate methods to extract topical information from texts, which can gauge the importance of various fundamentals or countries at a given point in time. However topical expressions do not measure tonality, which is crucial if one wants to investigate how changing market perceptions about fundamentals affect asset prices. Popular methods of tonality extraction, on the other hand, are restricted to gauging tones of full documents, leading to difficulties when there are more topics within a document. Moreover, tonality of words depend on the particular topical expression, which current methods cannot handle.

Our method uses regular expressions to jointly extract topical and related tonality information from textual data. This allows us to create indices from news articles which can tell us information about the level of fundamentals, the direction of their changes and possibly the risks surrounding their future. We think the general idea underlying our method is a useful addition that could be used in many future applications relying on textual input in economics and finance.

The second contribution of the paper relates to an ongoing discussion in finance about the role of fundamentals within asset pricing and, in particular, sovereign credit spreads. To add to this discussion we extract news indices on seven fundamentals of a large cross-section of countries and provide empirical evidence that these indices do indeed recover information about country fundamentals.

These constructed news indices are then shown to explain a significant variation in sovereign credit spreads that are not captured by traditional indicators of fundamentals. Using traditional macro variables we find a very limited role for fundamentals with respect to changes of sovereign spreads: explaining less than a tenth of total variation. Together with our news indices, however, this share increases to the 40-50 percent range.

Fundamentals of globally important countries explain most of this additional variation in sovereign spreads. This provides empirical support for the view that common fluctuations in sovereign credit spreads are partly rooted in fundamentals or their expectations (e.g. Benzoni et al., 2015; Augustin and Tédongap, 2016). It also shows that the often found explanatory power of globally important financial variables (such as the VIX index and corporate high yield spreads) in sovereign spread regressions is partly a result of fundamentals and should not all be attributed to non-fundamental factors.

Because a significant part of systemic fluctuations in asset prices are common not only across countries but also across asset classes, our results may generalize to assets outside of sovereign credit. Hence an interesting avenue for future research would be to revisit the existing evidence on the sources of asset price fluctuations more generally using our news indices.

References

Acharya, V., I. Drechsler and P. Schnabl (2014), 'A pyrrhic victory? Bank bailouts and sovereign credit risk', *The Journal of Finance*, vol. 69 no. 6, 2689–2739. p.

Allen, F. and D. Gale (2000), 'Financial Contagion', *Journal of Political Economy*, vol. 108 no. 1 (Feb.), 1–33. p.

Alper, C. E., L. Forni and M. Gerard (2013), 'Pricing of sovereign credit risk: Evidence from advanced economies during the financial crisis', *International Finance*, vol. 16 no. 2, 161–188. p.

Ang, A. and F. A. Longstaff (2011), *Systemic Sovereign Credit Risk: Lessons from the U.S. and Europe*, NBER Working Papers 16982, National Bureau of Economic Research, Inc.

Augustin, P. (2014), 'Sovereign Credit Default Swap Premia', *Journal of Investment Management*, vol. 12 no. 2 (June),

Augustin, P. and R. Tédongap (2016), 'Real economic shocks and sovereign credit risk', *Journal of Financial and Quantitative Analysis*, vol. 51 no. 2, 541–587. p.

Baek, I.-M., A. Bandopadhyaya and C. Du (2005), 'Determinants of market-assessed sovereign risk: Economic fundamentals or market risk appetite?', *Journal of International Money and Finance*, vol. 24 no. 4, 533–548. p.

Baker, M. and J. Wurgler (2006), 'Investor sentiment and the cross-section of stock returns', *The Journal of Finance*, vol. 61 no. 4, 1645–1680. p.

Baker, S. R., N. Bloom and S. J. Davis (2016), 'Measuring economic policy uncertainty', *The Quarterly Journal of Economics*, vol. 131 no. 4, 1593–1636. p.

Barro, R. J. (2006), 'Rare disasters and asset markets in the twentieth century', *The Quarterly Journal of Economics*, vol. 121 no. 3, 823–866. p.

Beber, A., M. W. Brandt and K. A. Kavajecz (2008), 'Flight-to-quality or flight-to-liquidity? Evidence from the euro-area bond market', *The Review of Financial Studies*, vol. 22 no. 3, 925–957. p.

Beetsma, R., M. Giuliodori, F. De Jong and D. Widijanto (2013), 'Spread the news: The impact of news on the European sovereign bond markets during the crisis', *Journal of International Money and Finance*, vol. 34, 83–101. p.

Benzoni, L., P. Collin-Dufresne, R. S. Goldstein and J. Helwege (2015), 'Modeling credit contagion via the updating of fragile beliefs', *The Review of Financial Studies*, vol. 28 no. 7, 1960–2008. p.

Born, B., M. Ehrmann and M. Fratzscher (2014), 'Central bank communication on financial stability', *The Economic Journal*, vol. 124 no. 577, 701–734. p.

Brunnermeier, M. K. and L. H. Pedersen (2008), 'Market liquidity and funding liquidity', *The review of financial studies*, vol. 22 no. 6, 2201–2238. p.

Calvo, G. A. (2002), *Contagion in Emerging Markets: When Wall Street is a Carrier*,

Calvo, G. A. and E. G. Mendoza (1996), 'Mexico's balance-of-payments crisis: a chronicle of a death foretold', *Journal of International Economics*, vol. 41 no. 3, 235–264. p.

Cesare, A. D., G. Grande, M. Manna and M. Taboga (2012), *Recent estimates of sovereign risk premia for euro-area countries*, Questioni di Economia e Finanza (Occasional Papers) 128, Bank of Italy, Economic Research and International Relations Area.

Cole, H. L. and T. J. Kehoe (1996), 'A self-fulfilling model of Mexico's 1994–1995 debt crisis', *Journal of international Economics*, vol. 41 no. 3, 309–330. p.

Correa, R., K. Garud, J. M. Londono and N. Mislang (2017), *Sentiment in Central Bank's Financial Stability Reports*, tech. rep. 1203, Board of Governors of the Federal Reserve System.

Corsetti, G., M. Pericoli and M. Sbracia (2005), "Some contagion, some interdependence': More pitfalls in tests of financial contagion', *Journal of International Money and Finance*, vol. 24 no. 8 (Dec.), 1177–1199. p.

Da, Z., J. Engelberg and P. Gao (2011), 'In search of attention', *The Journal of Finance*, vol. 66 no. 5, 1461–1499. p.

Dergiades, T., C. Milas and T. Panagiotidis (2014), 'Tweets, Google trends, and sovereign spreads in the GIIPS', *Oxford Economic Papers*, vol. 67 no. 2, 406–432. p.

Diamond, D. W. and P. H. Dybvig (1983), 'Bank runs, deposit insurance, and liquidity', *Journal of political economy*, vol. 91 no. 3, 401–419. p.

Dieckmann, S. and T. Plank (2011), 'Default risk of advanced economies: An empirical analysis of credit default swaps during the financial crisis', *Review of Finance*, vol. 16 no. 4, 903–934. p.

Dooley, M. and M. Hutchison (2009), 'Transmission of the US subprime crisis to emerging markets: Evidence on the decoupling–recoupling hypothesis', *Journal of International Money and Finance*, vol. 28 no. 8, 1331–1349. p.

Eaton, J. and M. Gersovitz (1981), 'Debt with potential repudiation: Theoretical and empirical analysis', *The Review of Economic Studies*, vol. 48 (Jan.), 289–309. p.

Edwards, S. (1983), *LDC's foreign borrowing and default risk: An empirical investigation*, NBER Working Papers 1172, National Bureau of Economic Research, Inc.

Ehrmann, M., C. Osbat, J. Stráskỳ and L. Uusküla (2014), 'The euro exchange rate during the European sovereign debt crisis–Dancing to its own tune?', *Journal of International Money and Finance*, vol. 49, 319–339. p.

Eichengreen, B. and A. Mody (1998), *What Explains Changing Spreads on Emerging-Market Debt: Fundamentals or Market Sentiment?*, NBER Working Papers 6408, National Bureau of Economic Research, Inc.

Favero, C., M. Pagano and E.-L. Von Thadden (2010), 'How does liquidity affect government bond yields?', *Journal of financial and quantitative analysis*, vol. 45 no. 1, 107–134. p.

Ferguson, N. J., D. Philip, H. Y. T. Lam and J. M. Guo (2015), 'Media Content and Stock Returns: The Predictive Power of Press', *Multinational Finance Journal*, vol. 19 no. 1, 1–31. p.

Gabaix, X. (2012), 'Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance', *The Quarterly Journal of Economics*, vol. 127 no. 2, 645–700. p.

Gaspar, V., M. S. Gupta and M. C. Mulas-Granados (2017), *Fiscal Politics*, International Monetary Fund.

Gomes, P. and A. Taamouti (2016), 'In search of the determinants of European asset market comovements', *International Review of Economics & Finance*, vol. 44, 103–117. p.

González-Rozada, M. and E. L. Yeyati (2008), 'Global factors and emerging market spreads', *The Economic Journal*, vol. 118 no. 533, 1917–1936. p.

Hansen, S. and M. McMahon (2016), 'Shocking language: Understanding the macroeconomic effects of central bank communication', *Journal of International Economics*, vol. 99 no. Supplement 1, S114 –S133. p.

Hilscher, J. and Y. Nosbusch (2010), 'Determinants of sovereign risk: Macroeconomic fundamentals and the pricing of sovereign debt', *Review of Finance*, vol. 14 no. 2, 235–262. p.

Hu, G. X., J. Pan and J. Wang (2013), 'Noise as information for illiquidity', *The Journal of Finance*, vol. 68 no. 6, 2341–2382. p.

Joseph, K., M. B. Wintoki and Z. Zhang (2011), 'Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search', *International Journal of Forecasting*, vol. 27 no. 4, 1116–1127. p.

Kocsis, Z. and Z. Monostori (2016), 'The role of country-specific fundamentals in sovereign CDS spreads: Eastern European experiences', *Emerging Markets Review*, vol. 27, 140–168. p.

Liu, S. (2014), 'The Impact of Textual Sentiment on Sovereign Bond Yield Spreads: Evidence from the Eurozone Crisis', *Multinational Finance Journal*, vol. 18 no. 3-4, 215–248. p.

Longstaff, F. A., J. Pan, L. H. Pedersen and K. J. Singleton (2011), 'How Sovereign Is Sovereign Credit Risk?', *American Economic Journal: Macroeconomics*, vol. 3 no. 2 (Apr.), 75–103. p.

Loughran, T. and B. McDonald (2011), 'When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks', *The Journal of Finance*, vol. 66 no. 1, 35–65. p.

Lucca, D. O. and F. Trebbi (2009), *Measuring central bank communication: an automated approach with application to FOMC statements*, NBER Working Paper 15367, National Bureau of Economic Research.

McFadden, D, R Eckaus, G Feder, A. Hajivassiliou, S O'Connell, G. Smith and J. Cuddington (1985), 'Is There Life after Debt? An Econometric Analysis of the Creditworthiness of Developing Countries', , 179–209. p.

McGuire, P. and M. A. Schrijvers (2003), 'Common factors in emerging market spreads', *BIS Quarterly Review*, vol. 4 (Dec.),

Mohl, P. and D. Sondermann (2013), 'Has political communication during the crisis impacted sovereign bond spreads in the euro area?', *Applied Economics Letters*, vol. 20 no. 1, 48–61. p.

Moser, T. (2003), 'What is international financial contagion?', *International Finance*, vol. 6 no. 2, 157–178. p.

Officer, R. R. (1973), 'The variability of the market factor of the New York Stock Exchange', *The Journal of Business*, vol. 46 no. 3, 434–453. p.

Paniagua, J., J. Sapena and C. Tamarit (2016), 'Sovereign debt spreads in EMU: The time-varying role of fundamentals and market distrust', *Journal of Financial Stability*,

Roll, R. (1988), 'The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis.' *The Journal of Finance*, vol. 43 no. 3, 541–566. p.

Ross, S. A. (1989), 'Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy', *The Journal of Finance*, vol. 44 no. 1, 1–17. p.

Sachs, J., A. Tornell and A. Velasco (1996), 'The Mexican peso crisis: Sudden death or death foretold?', *Journal of international economics*, vol. 41 no. 3, 265–283. p.

Schwert, G. W. (1989), 'Why does stock market volatility change over time?', *The Journal of Finance*, vol. 44 no. 5, 1115–1153. p.

Shiller, R. J. (2014), 'Speculative asset prices', *American Economic Review*, vol. 104 no. 6, 1486–1517. p.

Tetlock, P. C. (2007), 'Giving content to investor sentiment: The role of media in the stock market', *The Journal of Finance*, vol. 62 no. 3, 1139–1168. p.

Tetlock, P. C., M. Saar-Tsechansky and S. Macskassy (2008), 'More than words: Quantifying language to measure firms' fundamentals', *The Journal of Finance*, vol. 63 no. 3, 1437–1467. p.

Thompson, S. B. (2011), 'Simple formulas for standard errors that cluster by both firm and time', *Journal of financial Economics*, vol. 99 no. 1, 1–10. p.

Tobback, E., S. Nardelli and D. Martens (2017), *Between hawks and doves: measuring central bank communication*, Working Paper Series 2085, European Central Bank.

Wachter, J. A. (2013), 'Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?', *The Journal of Finance*, vol. 68 no. 3, 987–1035. p.

Wu, J. C. and F. D. Xia (2016), 'Measuring the macroeconomic impact of monetary policy at the zero lower bound', *Journal of Money, Credit and Banking*, vol. 48 no. 2-3, 253–291. p.

Appendix A Macroeconomic and financial data

Table A1
Data sources

Data	Download Source (original source/MNEMONIC)
	Sovereign credit risk pricing data[a]
Sovereign 5-year Credit Default Swap premia	Bloomberg (CMA)
(Eurozone) benchmark 5-year bond yields	Bloomberg (generic rates)
EMBI Global spreads	Datastream (JP Morgan)
	(Traditional) Macroeconomic and financial data[b]
World real GDP annual growth rate	IMF WEO (NGDP_RPCH)
Real GDP annual growth rate	WB WDI (NY_GDP_MKTP_KD_ZG)
GDP constant prices, national curr., seas.adj.	WB GEM (NYGDPMKTPSAKN)
GDP current prices, USD, seas.adj.	WB GEM (NYGDPMKTPSACD)
Current account balance, current USD	IMF IFS (BGS_BP6_USD)
Official reserves, current USD	WB GEM (TOTRESV)
Fiscal balance/GDP	IMF WEO (GGXCNL_NGDP)
Gross public debt/GDP	IMF HIST (GGXWDG_GDP)
VIX index (3-mo ATM implied vol, S&P500)	Bloomberg (CBOE, VIX index)
US Corp.spec.grade bond spread over 10-year Treasury notes	Bloomberg (CBOE, VIX index)
	Other macroeconomic data
US macroeconomic announcements and surveys	Bloomerg ECO[c]
UK macroeconomic announcements and surveys	Bloomerg ECO[d]
China macroeconomic announcements and surveys	Bloomerg ECO[e]
Germany macroeconomic announcements and surveys	Bloomerg ECO[f]
Japan macroeconomic announcements and surveys	Bloomerg ECO[g]

Notes: Descriptive statistics of news items filtered by relevance and duplication.

[a)] Country lists: CDS spreads were available for: ARGE (Argentina), AUSL (Australia), AUT (Austria), BELG (Belgium), BRAZ (Brazil), BULG (Bulgaria), CHIL (Chile), CHIN (China), COLO (Colorado), CROA (Croatia), CZEC (Czech Republic), DENM (Denmark), EGYP (Egypt), ESTO (Estonia), FINL (Finland), FRAN (France), GERM (Germany), GREE (Greece), HONG (Hong Kong), HUNG (Hungary), ICEL (Iceland), INDO (Indonesia), IREL (Ireland), ISRA (Israel), ITAL (Italy), JAPA (Japan), KAZA (Kazakhstan), KORE (South Korea), LATV (Latvia), LEBA (Lebanon), LITH (Lithuania), MALA (Malaysia), MEXI (Mexico), NETH (Netherlands), NZ (New Zealand), NORW (Norway), PAKI (Pakistan), PANA (Panama), PERU (Peru), PHIL (Phillippines), POLA (Poland), PORT (Portugal), ROMA (Romania), RUSS (Russia), SOAF (South Africa), SPAI (Spain), SRIL (Sri Lanka), SWED (Sweden), SWI (Switzerland), THAI (Thailand), TUNE (Tunesia), TURK (Turkey), UK (United Kingdom), UKRA (Ukraine), URUG (Uruguay), US (United States), VENE (Venezuela), VIET (Vietnam). EMBI Global spreads available for: ARGE, BRAZ, BULG, CHIL, CHIN, COLO, EGYP, HUNG, INDO, KAZA, LEBA, MALA, MEXI, PAKI, PANA, PERU, PHIL, POLA, ROMA, RUSS, SOAF, SRIL, TUNE, TURK, UKRA, URUG, VENE, VIET. Eurozone benchmark yields: AUT, BELG, ESTO, FINL, FRAN, GERM, GREE, IREL, ITAL, NETH, PORT, SPAI.

[b)] IMF WEO: International Monetary Fund World Economic Outlook; IFS: International Financial Statistics; HIST: Historical Public debt; WB WDI: World Bank World Development Indicators; GEM: Global Economic Monitor; CBOE: Chicago Board of Exchange.

[c)] US Bloomberg tickers: REAL: ADP CHNG Index, AWH TOTL Index, CFNAI Index, CGNOXAI% Index, CHPMINDX Index, CNSTTMOM Index, CONCCONF Index, CONSCURR Index, CONSEXP Index, CONSPXMD Index, CONSSENT Index, CPTICHNG Index, DFEDGBA Index, DGNOCHNG Index, ECI SA% Index, ECONGECC Index, EMPRGBCI Index, ETSLMOM Index, ETSLTOTL Index, GDP CQOQ Index, GDP PIQQ Index, GDPCTOT% Index, HPI PURQ Index, HPIMMOM% Index, INJCJC Index, INJCSP Index, IP CHNG Index, IPMGCHNG Index, KCLSSACI Index, LEI CHNG Index, MPMIUSCA Index, MPMIUSMA Index, MPMIUSSA Index, MTIBCHNG Index, MWSLCHNG Index, NAPMPMI Index, NFP PCH Index, NFP TCH Index, NHCHATCH Index, NHCHSTCH Index, NHSLCHNG Index, NHSLTOT Index, NHSPATOT Index, NHSPSTOT Index, NWORCHNG Index, OUTFGAF Index, PCE CHNC Index, PCE CRCH Index, PITLCHNG Index, PRUSTOT Index, RSTAMOM Index, RSTAXAG% Index, RSTAXAGM Index, RSTAXMOM Index, SAARDTOT Index, SAARTOTL Index, SBOITOTL Index, TMNOCHNG Index, TMNOXTM% Index, USEMNCHG Index, USHBMIDX Index, USMMMNCH Index, USPHTMOM Index, USPHTYOY Index, USUDMAER Index, USURTOT Index; EXTERN: IMP1CHNG Index, IMP1YOY% Index, USCABAL Index, USTBTOT Index, USTGTTCB Index.

[d)] UK Bloomberg tickers: REAL: DTSDD1RB Index, DTSRR1RB Index, ITSR1B Index, KPRSLFLS Index, LTSBBSBX Index, MPMIGBCA Index, MPMIGBMA Index, MPMIGBSA Index, MPMIGBXA Index, MTEF1C Index, UKBINPEQ Index, UKBINPEY Index, UKCCI Index, UKCNALSM Index, UKCNALSY Index, UKDHUKY Index, UKEGESTG Index, UKGEABRQ Index, UKGENMYQ Index, UKGRABIQ Index, UKGRABIY Index, UKHB3MYR Index, UKHBSAMM Index, UKIPIMOM Index, UKIPIYOY Index, UKLFEMCH Index, UKMLMNHP Index, UKMPIMOM Index, UKMPIYOY Index, UKNBAAMM Index, UKNBANYY Index, UKRMNAPM Index, UKRMNAPY Index, UKRVAMOM Index, UKRVAYOY Index, UKRVINFM Index, UKRVINFY Index, UKRXPBAL Index, UKUEILOR Index, UKUEMOM Index, UKUER Index, UKVHRYY Index; EXTERN: UKCA Index, UKCR Index, UKGEIKKQ Index, UKGEIKLQ Index, UKTBALEE Index, UKTBLGDT Index, UKTBTTBA Index

[e)] China Bloomberg tickers: REAL: CHBNINDX Index, CHVAICY Index, CHVAIOY Index, CNCILI Index, CNDIINRY Index, CNGDPC$Y Index, CNGDPQOQ Index, CNGDPYOY Index, CNPRETLY Index, CNRSACMY Index, CNRSCYOY Index, CPMINDX Index, CPMINMAN Index, MNCCINDX Index, MPMICNCA Index, MPMICNMA Index, MPMICNSA Index; EXTERN: CNFRBAL$ Index, CNFREXPY Index, CNFRIMPY Index, CNGFOREX Index, CNTSECNY Index, CNTSICNY Index, CNTSTCN Index.

[f)] Germany Bloomberg tickers: REAL: ECO1GFKC Index, GDPB95YY Index, GEINYY Index, GEIOYY Index, GRFIFINB Index, GRFRIAMM Index, GRFRINYY Index, GRGDARCL Index, GRGDGCIQ Index, GRGDGCQ Index, GRGDICQ Index, GRGDPCQ Index, GRGDPPGQ Index, GRGDPPGY Index, GRIFPBUS Index, GRIFPCA Index, GRIFPEX Index, GRIORTMM Index, GRIPIMOM Index, GRUECHNG Index, GRUEPR Index, GRZECURR Index, GRZEWI Index, MPMIDECA Index, MPMIDEMA Index, MPMIDERA Index, MPMIDESA Index, MPMIDEXA Index; EXTERN: GRBTEXMM Index, GRBTIMMM Index, GRCAEU Index, GRGDEXQ Index, GRGDIMQ Index, GRIMP95M Index, GRIMP95Y Index, GRTBALE Index.

[g)] Japan Bloomberg tickers: REAL: JBLDHFOY Index, JBSIBCLA Index, JBSIBCLM Index, JCOMSHCF Index, JGDOQOQ Index, JGDPAGDP Index, JGDPCIQ Index, JGDPPCQ Index, JGDPQGDP Index, JHHSLERY Index, JNC SALE Index, JNCAPMOM Index, JNCICLEI Index, JNCVSSY Index, JNCSTOTY Index, JNDSNYOY Index, JNDSTYOY Index, JNHSAN Index, JNHSYOY Index, JNIPMOM Index, JNIPYOY Index, JNMOCHNG Index, JNMOYOY Index, JNMTOY Index, JNNETYOY Index, JNRETMOM Index, JNRSYOY Index, JNSASYOY Index, JNSBALLI Index, JNSMTYOY Index, JNTEMFG Index, JNTENMFG Index, JNTIAIAM Index, JNTIAMOM Index, JNTSMFG Index, JNTSNMFG Index, JNUE Index, JNVHPYOY Index, JNVHSYOY Index, JNVNIYOY Index, JNVNYOYS Index, JPTFLMFG Index, JPTFLNMF Index, JPTFSMFG Index, JPTFSNMF Index, JTFIFILA Index, MPMIJPCA Index, MPMIJPMA Index, MPMIJPSA Index; EXTERN: JNBPAB Index, JNBPABA Index, JNBPTRD Index, JNFRTOTL Index, JNTBAL Index, JNTBALA Index, JNTBEXPY Index, JNTBIMPY Index.

Appendix B Creating the news database

Preparing the news data for analysis consisted of the following steps:

1. downloading and parsing html content to obtain article title, text bodies, date stamps;
2. formatting and cleaning text from html tags, company tickers, etc., inserting paragraph identifiers;
3. removing irrelevant news items;
4. removing duplications of news items;
5. inserting labels representing synonyms;
6. inserting labels representing simple expressions;
7. inserting labels representing fundamental expressions;

One advantage of the Reuters data set is that article html sources codes have a standardized structure, therefore the same extraction and parsing code can be used for all items in the data set. After parsing the html code for article id, title, date stamp, article text, we used regular expressions to clean the text from html tags left within the article and to remove meta-information items at the beginning and end of articles. Text was transformed to lowercase (before this action keywords were inserted into the text where lowering case would lose information, e.g keyw_us before 'US', keyw_cds before 'CDS', keyw_fed before 'Fed'), so that we could easily differentiate later inserted identifiers which would enter with upper case characters. Most non-alphanumeric characters were either removed or replaced with remaining separator characters: comma and period and the identifier 'P' inserted to represent paragraph shift. These separators were left in the text, which could later be used as proximity criteria to restrict elements of fundamental expressions to belong into the same unit of text (clause, sentence, paragraph).

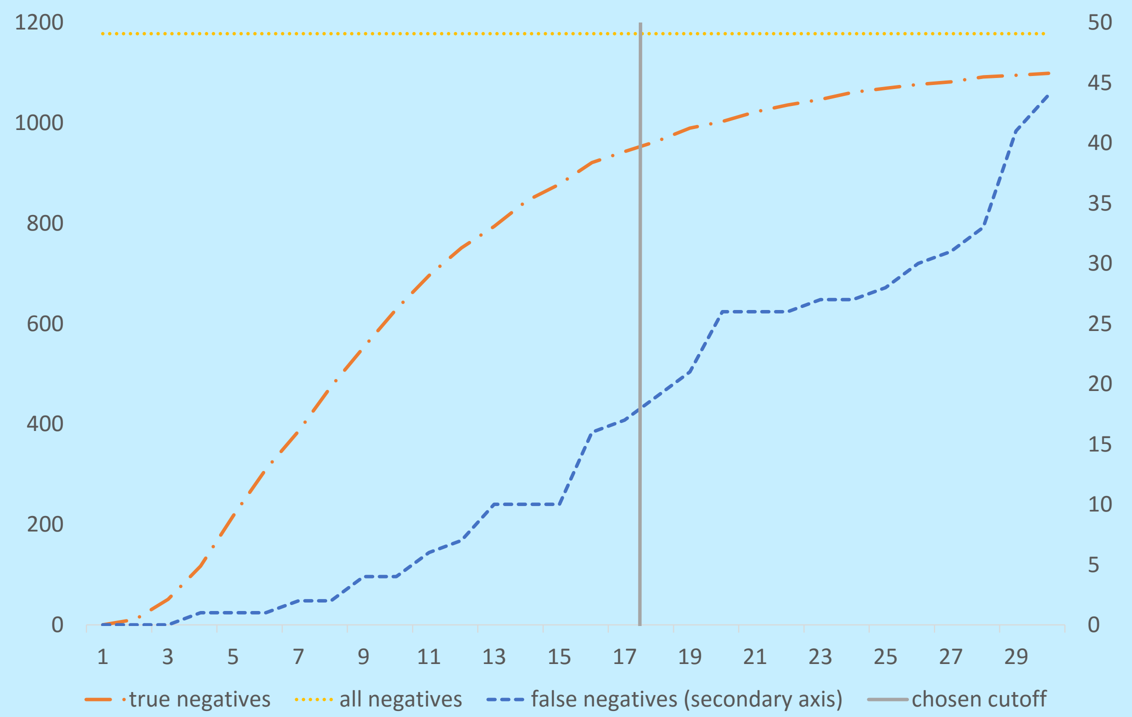
Removing irrelevant news at this stage served to reduce the size of the data set to make later calculations more computationally efficient. We rather wanted to err on the side of caution, so that we aimed to keep more of the relevant articles at the expense of throwing away less of the likely irrelevant articles. We randomly selected 6000 news articles and labelled them based only on the article titles into the classes: 'not relevant', 'relevant', 'not obvious'.

The support vector machine (SVM) supplied by Oracle Text was then trained using the standard bag-of words approach on the binary 'relevant' – 'not relevant' cases (not obvious cases were not included in the estimation). First, we trained the SVM on two-thirds of the sample and used the other one-third, 1415 news items, to test the method and choose an appropriate posterior probability cutoff value for discriminating between relevant and irrelevant classes. Figure 1 shows how increasing the posterior probability cutoff for relevance (positives) cutoff increases predictions of belonging into the irrelevant (negative) class. True negative predictions rise much faster than false negatives demonstrating the efficiency of the learning algorithm. The probability value of 0.28 (SVM score 28) maximizes the Matthews correlation coefficient, but we chose a lower cutoff of 18 that had only 19 false negatives compared with 33 false negatives at the 28 cutoff. The chosen cutoff still identified 964 true negative cases (about 100 less than the 28 cutoff).

The last step in data preparation that further reduced our news data set is a filtering of duplicate or close to duplicate news. This is important, because the number and ratio of duplicate news is seen to vary greatly across different time periods, which would distort our measure of daily relevant news. Even a quick inspection of lists of daily titles on the Reuters website calls attention to duplicates by observing consecutive items with the same titles. For instance, the first date in the archive[35] lists 173 items of which 30 have the same title as the subsequent news item. While perfect matches in the title are a good indication of duplicates, it is neither a sufficient nor a necessary condition for matching text bodies. Computationally, the perfect matches

[35] http://uk.reuters.com/resources/archive/uk/20070101.html

Figure B.1
True and false negatives at chosen SVM cutoff



Source: authors' calculations.
Notes: In a test sample of 1415 news items 1178 news items were labelled as not having relevant fundamental information. The figure shows the results of predictions of a support vector machine (SVM) trained on a separate set of news articles. As the posterior probability cutoff increases (X axis) more items are classified as irrelevant either correctly (true negatives) or incorrectly (false negatives). It shows the efficiency of the algorithm that at any cutoff correct filtering (primary axis) significantly exceeds Type II errors (secondary axis).

on the same date are easy to find. However, this leaves out many more articles that are only close matches, such as article pairs of which one is an update, expansion or correction of a previously published news item. Spotting these requires more computational effort.

The methodology we found to be accurate and computationally feasible is the following. For computational feasibility we needed to first reduce the number of news item pairs to investigate. We achieved this by inspecting all pairs of news titles on a given day checking whether the longer title of the two contained 60 percent of words (rounded up to the nearest integer) in the shorter title. This produced a lot of false negatives, but it also drastically reduced the number of article pairs to compare. All permutations were then checked within each group whether the longer article of a pair largely encompassed the shorter article or not. We defined the encompassing rule as having at least 2 matched paragraphs (perfect string match after removing preceding and trailing non-alphanumeric characters). When so defined duplicates were found, the shorter article was flagged for removal.

Several types of keywords were inserted into the text that later formed parts of the fundamental expressions that we were looking for. At the beginning of the section, we referred to these as synonyms for simplicity, but in fact the groups of words or simple expressions identified often had wider differences in meaning than what could be labelled as synonyms. Our idea was to identify words or expressions that referred to similar concepts and which would be close substitutes within fundamental expressions.

Table B.1
Filtering news items by relevance

SVM relevance score	title	url
	LEAST RELEVANTS (SVM 0-3)	
0	tennis-adelaide international men's singles round robin results	http://UK.reuters.com/article/UK_TENNIS/idUKISS66050620070101
1	press digest - new york times - jan 1	http://UK.reuters.com/article/governmentFilingsNews/idUKBNG13770420070101
1	text-internet gold completes 012 golden lines purchase	http://UK.reuters.com/article/governmentFilingsNews/idUKL0189775620070101
2	brisa says to invest 393 mln euros in 2007	http://UK.reuters.com/article/basicIndustries/idUKL2983318020070101
2	golf-revamped tour seeks excitement to last tee	http://UK.reuters.com/article/golfNews/idUKL0187483820070101
2	update 4-tennis-auckland open women's singles round 1 results	http://UK.reuters.com/article/UK_TENNIS/idUKISS65674420070101
2	update 1-tennis-hopman cup singles results	http://UK.reuters.com/article/UK_TENNIS/idUKISS65702220070101
2	nissan to build 200,000-unit plant in india-paper	http://UK.reuters.com/article/governmentFilingsNews/idUKT14803520070101
2	soccer-israeli championship results and standings	http://UK.reuters.com/article/UK_SOCCER/idUKISS65887920070101
2	pope says peace depends on respecting human rights	http://UK.reuters.com/article/worldNews/idUKL0189313020070101
2	factbox-golf-inaugural fedexcup cup	http://UK.reuters.com/article/golfNews/idUKL0188151720070101
2	update 1-soccer-buchwald completes double with reds in send-off	http://UK.reuters.com/article/UK_SOCCER/idUKSP13728020070101
3	update 1-tennis-chennai open men's singles round 1 results	http://UK.reuters.com/article/UK_TENNIS/idUKISS65930520070101
3	update 4-tennis-qatar open men's singles round 1 results	http://UK.reuters.com/article/UK_TENNIS/idUKISS65928820070101
3	update 4-tennis-australian women's hardcourts women's singles round 1 results	http://UK.reuters.com/article/UK_TENNIS/idUKISS65700520070101
3	cricket-rain delays start of final ashes test	http://UK.reuters.com/article/UK_CRICKET/idUKSP14824120070101
3	update 1-tennis-australia's luczak stuns hrbaty in adelaide	http://UK.reuters.com/article/UK_TENNIS/idUKSP13109820070101
3	japan tv apologises for "topless" new year's eve shock	http://UK.reuters.com/article/oddlyEnoughNews/idUKT13300820070103
3	gene-engineered cattle resist mad cow disease: study	http://UK.reuters.com/article/scienceNews/idUKN3126493620070104
3	diary - global environment	http://UK.reuters.com/article/oilRpt/idUKENVIRO20070101
3	tennis-myskina loses in auckland, may miss australian open	http://UK.reuters.com/article/UK_TENNIS/idUKSP14629720070101
	CLOSE TO RELEVANCE (SVM 13-18)	
13	soccer-results/standings from australian a-league	http://UK.reuters.com/article/UK_WORLDFOOTBALL/idUKSP14587420070101
13	latin balladeers, reality stars eye breakthrough	http://UK.reuters.com/article/musicNews/idUKN3146991620070102
14	iraqis ponder lessons of history after saddam hangs	http://UK.reuters.com/article/featuresNews/idUKPAR15182820070103
14	repeat-cricket-one-day international series new zealand v sri lanka line-ups	http://UK.reuters.com/article/UK_CRICKET/idUKISS66382220070101
14	eu newcomers hopeful, anxious about membership	http://UK.reuters.com/article/worldNews/idUKL0185119420070101
14	farewells fuel bid for ashes clean sweep	http://UK.reuters.com/article/UKNews1/idUKSP13321520070101
14	somali government vows to pursue fleeing islamists	http://UK.reuters.com/article/worldNews/idUKL0186366220070101
14	chrysler signs small-car deal with china's chery	http://UK.reuters.com/article/basicIndustries/idUKN2925438120070101
15	update 1-china auto exports hit record in 2006	http://UK.reuters.com/article/governmentFilingsNews/idUKPEK13212620070101
15	update 1-tennis-baghdatis defeats kohlschreiber in qatar	http://UK.reuters.com/article/UK_TENNIS/idUKL0189818420070101
15	tv shows restraint with limited saddam footage	http://UK.reuters.com/article/televisionNews/idUKN2821679420070101
16	romanian and bulgaria celebrate eu entry	http://UK.reuters.com/article/worldNews/idUKL2977273320070101
	RELEVANTS (SVM 18-100)	
18	storms wash out new year parties across britain	http://UK.reuters.com/article/UKNews1/idUKL0189036120070101
18	minimum cigarette buying age to rise to 18	http://UK.reuters.com/article/UKNews1/idUKL3080853820070101
20	stay off bangkok streets -foreign governments	http://UK.reuters.com/article/worldNews/idUKBKK14142020070101
21	soccer-australian championship results and standings	http://UK.reuters.com/article/UK_WORLDFOOTBALL/idUKISS65889520070101
21	golf-world ranking standings	http://UK.reuters.com/article/golfNews/idUKISS66242520070101
22	horse meat, kale and crickets on menu at bronx zoo	http://UK.reuters.com/article/featuresNews/idUKN2821130020070102
22	thai pm blames ex-politicians for bombs	http://UK.reuters.com/article/worldNews/idUKBKK14460620070101
22	saddam's daughter attends protest in jordan	http://UK.reuters.com/article/breakingNews/idUKL0186083720070101
24	eu newcomers hopeful, anxious about membership	http://UK.reuters.com/article/worldNews/idUKL0185119420070101
28	hyundai heavy 2006 sales up 22 pct on orders	http://UK.reuters.com/article/basicIndustries/idUKSEO14750020070101
30	mild jobs may lift stocks as '07 starts	http://UK.reuters.com/article/pressReleases/idUKN2925725620070102
31	rpt-wall st week ahead: mild jobs may lift stocks as '07 starts	http://UK.reuters.com/article/oilRpt/idUKN3126282720070101
32	russia, belarus sign gas deal	http://UK.reuters.com/article/businessNews/idUKL2978817720070101
40	iraq to probe filming of saddam hanging	http://UK.reuters.com/article/worldNews/idUKPAR96599620070101
41	thailand says thaksin backers may be behind blasts	http://UK.reuters.com/article/worldNews/idUKBKK14775120070101
47	chronology of somalia's collapse, conflict	http://UK.reuters.com/article/worldNews/idUKL0188726320070101
48	dead leaders drive lebanese political life	http://UK.reuters.com/article/featuresNews/idUKL2215259520070103
49	somali islamists flee towards kenya and to the hills	http://UK.reuters.com/article/worldNews/idUKL2859346720070101
56	web wishes show contrasting french election styles	http://UK.reuters.com/article/technologyNews/idUKL0178602420070102
57	brazil's lula pledges economic growth in 2nd term	http://UK.reuters.com/article/governmentFilingsNews/idUKN2838881720070101
58	somalis stroll mogadishu under eye of govt victors	http://UK.reuters.com/article/worldNews/idUKL0189682420070101
70	weekahead-emerging debt to start 2007 eyeing brazil, keyw_us data	http://UK.reuters.com/article/marketsNewsUS/idUKN3134666220070101
73	dollar a shade softer, yen stays subdued	http://UK.reuters.com/article/hotStocksNewsUS/idUKN2941694320070101
73	slovenia joins euro to crown ex-communist reform	http://UK.reuters.com/article/businessNews/idUKL3185272820070101
90	update 1-peru consumer prices rose 1.14 percent in 2006	http://UK.reuters.com/article/marketsNewsUS/idUKN0116566320070101

Sources: Reuters news archives and authors' calculations.
Notes: News items are considered relevant in terms of fundamental information if the SVM score is at least 18. Based on the news sample of January 1, 2007.

Table B.2
Filtering news items – the first 40 items of the Jan 1, 2007 sample

title	url	SVM relevance score	is_relevant	is_duplicate	is_included
Newcomer a ”Rae” of light at Grammys	http://UK.reuters.com/article/entertainmentNews/idUKN0126427220070101	5	0		0
physical abuse leads to adult depression -study	http://UK.reuters.com/article/UKNews1/idUKN2924492920070101	5	0		0
repeat-cricket-one-day international series new zealand v sri lanka line-ups	http://UK.reuters.com/article/UK_CRICKET/idUKISS66382220070101	14	0		0
somalis stroll mogadishu under eye of govt victors	http://UK.reuters.com/article/worldNews/idUKL0189682420070101	58	1	0	1
hoggard ruled of fifth test	http://UK.reuters.com/article/sportsNews/idUKSP13634320070101	10	0		0
airasia no comment on easyjet, virgin tie-up report	http://UK.reuters.com/article/businessNews/idUKL0174299720070101	7	0		0
cricket-rain delays start of new zealand v sri lanka one-dayer	http://UK.reuters.com/article/UK_CRICKET/idUKSP13543120070101	8	0		0
dollar a shade softer, yen stays subdued	http://UK.reuters.com/article/hotStocksNewsUS/idUKN2941694320070101	73	1	0	1
romanian and bulgaria celebrate eu entry	http://UK.reuters.com/article/worldNews/idUKL2977273320070101	16	0		0
iraq to probe filming of saddam hanging	http://UK.reuters.com/article/worldNews/idUKPAR96599620070101	40	1	0	1
eu newcomers hopeful, anxious about membership	http://UK.reuters.com/article/worldNews/idUKL0185119420070101	14	0		0
eu newcomers hopeful, anxious about membership	http://UK.reuters.com/article/featuresNews/idUKL0185119420070102	24	1	0	1
eu newcomers hopeful, anxious about membership	http://UK.reuters.com/article/featuresNews/idUKL0185119420070103	24	1	1	0
australian lexicon can leave you a few roos loose	http://UK.reuters.com/article/featuresNews/idUKSYD13951520070101	8	0		0
cricket-rain delays start of final ashes test	http://UK.reuters.com/article/UK_CRICKET/idUKSP14824120070101	3	0		0
hyundai motor says missed sales target amid strike	http://UK.reuters.com/article/basicIndustries/idUKSEO17075520070101	8	0		0
delta loses $49 mln in november	http://UK.reuters.com/article/basicIndustries/idUKN2923590820070101	7	0		0
goodyear workers ratify three-year contract	http://UK.reuters.com/article/basicIndustries/idUKN2923964120070101	7	0		0
brisa says to invest 393 mln euros in 2007	http://UK.reuters.com/article/basicIndustries/idUKL2983318020070101	2	0		0
italy opens for bidding for unprofitable alitalia	http://UK.reuters.com/article/basicIndustries/idUKL2928824420070101	12	0		0
ace says initial aeroplan payout worth c$856 mln	http://UK.reuters.com/article/basicIndustries/idUKN2817860220070101	6	0		0
hyundai heavy 2006 sales up 22 pct on orders	http://UK.reuters.com/article/basicIndustries/idUKSEO14750020070101	28	1	0	1
chrysler signs small-car deal with china’s chery	http://UK.reuters.com/article/basicIndustries/idUKN2925438120070101	14	0		0
vw brand head bernhard set to leave - paper	http://UK.reuters.com/article/basicIndustries/idUKL2982699320070101	8	0		0
in kidnapping, finesse works best	http://UK.reuters.com/article/featuresNews/idUKN1130230420070104	6	0		0
in kidnapping, finesse works best	http://UK.reuters.com/article/featuresNews/idUKN1130230420070103	6	0		0
in kidnapping, finesse works best	http://UK.reuters.com/article/featuresNews/idUKN1130230420070102	6	0		0
in kidnapping, finesse works best	http://UK.reuters.com/article/featuresNews/idUKN1130230420070101	6	0		0
india’s forgotten tribes gain rights over forests	http://UK.reuters.com/article/featuresNews/idUKDEL25463820070101	12	0		0
greying workers wanted for hire in aging japan	http://UK.reuters.com/article/featuresNews/idUKT13946420070102	11	0		0
greying workers wanted for hire in aging japan	http://UK.reuters.com/article/featuresNews/idUKT13946420070101	11	0		0
photographer, palestinian gunmen abducted in gaza	http://UK.reuters.com/article/worldNews/idUKL0188700420070101	10	0		0
priest’s death shows russia’s rural rot	http://UK.reuters.com/article/featuresNews/idUKL2733377520070102	4	0		0
priest’s death shows russia’s rural rot	http://UK.reuters.com/article/featuresNews/idUKL2733377520070101	4	0		0
horse meat, kale and crickets on menu at bronx zoo	http://UK.reuters.com/article/featuresNews/idUKN2821130020070102	22	1	0	1
horse meat, kale and crickets on menu at bronx zoo	http://UK.reuters.com/article/featuresNews/idUKN2821130020070101	22	1	1	0
dead leaders drive lebanese political life	http://UK.reuters.com/article/featuresNews/idUKL2215259520070103	48	1	0	1
dead leaders drive lebanese political life	http://UK.reuters.com/article/featuresNews/idUKL2215259520070102	48	1	1	0
dead leaders drive lebanese political life	http://UK.reuters.com/article/featuresNews/idUKL2215259520070101	48	1	1	0
celebrating new year in deadly safrican hotspot	http://UK.reuters.com/article/featuresNews/idUKL0160780920070101	11	0		0

Sources: Reuters news archives and authors’ calculations.
Notes: News items considered relevant in terms of fundamental information (SVM score>=18) are flagged in column 4, duplicate filtering is then carried out on these items.

Our list therefore was not intended to be a general-purpose thesaurus. It is specific to the context and language of the economic-financial media that we are dealing with and it is restricted to the expressions and phrases of fundamentals that are of interest for this specific research project.

Appendix C Additional information on news indices

Table C.1
Synonym labels

geography	geography	negation / adjectives	currency names	nouns	nouns	nouns	verbs
G1_ARGE	G2_AFGH	NEG	N_FX N_USD	N_CB G1_UK N_BOE	N_DEBT	N_PANIC	V_ACCELERATE
G1_AUSL	G2_AFR	NEG2	N_FX N_EUR	N_CB G1_EZ N_ECB	N_AID	N_TROUBLE	V_DECELERATE
G1_AUT	G2_ALGE	A_GOOD2	N_FX N_GBP	N_CB G1_US N_FED	N_GOVT	N_RISK	V_RISE
G1_BELG	G2_ASIA	A_GOOD1	N_FX N_CHF	N_CB G1_SWED N_RIKSBANK	N_FISCAL	N_HOPE	V_FALL
G1_BRAZ	G2_BOLI	A_GOOD0	N_FX N_JPY	N_CB G1_GERM N_BUNDESBANK	N_DEFICIT	N_CONCERN	V_RAISE
G1_BULG	G2_CHIL	A_BAD2	N_FX N_CNY	N_CB G1_JAPA N_BOJ	N_SURPLUS	N_CHANCE	V_CUT
G1_CANA	G2_COLO	A_BAD1	N_FX N_ARS	G1_UK N_SLS	N_BALANCE	N_FORECAST	V_IMPROVE
G1_CHIN	G2_DEV	A_BAD0	N_FX N_AUD	N_BRATE	N_REVENUE	N_THAN	V_WORSEN
G1_CROA	G2_ECUA	A_LARGE2	N_FX N_BRL	N_ZLB	N_SPENDING	N_UPTURN	V_TIGHTEN
G1_CYPR	G2_EGYP	A_LARGE1	N_FX N_CAD	N_QE	N_INVEST	N_ACTUAL	V_EASE
G1_CZEC	G2_EM	A_SMALL1	N_FX N_CLP	G1_US N_TAPER	N_SOCIAL	N_NUMBER	V_STRENGTHEN
G1_DENM	G2_GEOR	A_SMALL2	N_FX N_HRK	G1_EZ N_OMT	N_DEFENSE	N_IMPROVEMENT	V_WEAKEN
G1_EMEA	G2_HONG	A_BETTER	N_FX N_CZK	G1_EZ N_SMP	N_GUARANTEE	N_DETERIORATION	V_BEGIN
G1_ESTO	G2_INDO	A_WORSE	N_FX N_HKD	G1_EZ N_ELA	N_BANKS	N_INCREASE	V_END
G1_EU	G2_IRAN	A_HIGHER	N_FX N_HUF	G1_EZ N_LTRO	N_BAILOUT	N_DECREASE	V_CRUSH
G1_EZ	G2_IRAQ	A_LOWER	N_FX N_KRW	G1_US N_TAF	N_RECAPITAL	N_LABORM	V_SURPASS
G1_FINL	G2_ISRA	A_INSTABLE	N_FX N_LVL	G1_US N_TALF	N_NPL	N_RULELAW	V_TRAIL
G1_FRAN	G2_JAMA	A_STABLE	N_FX N_LTL	N_CB	N_PORTFOLIO	N_PROPRIGHTS	V_ACHIEVE
G1_GERM	G2_KAZA	A_AMPLE	N_FX N_EEK	N_REQRESERVES	N_CAPADEQ	N_FREEDOM	V_BECOME
G1_GLOB	G2_KORE	A_SCARCE	N_FX N_MYR	N_GDP	N_CAPITAL	N_CONTROLS	V_HISTORY
G1_GREE	G2_LEBA	A_FAILED	N_FX N_IDR	N_HHI	N_TOXIC	N_REGULATIONS	V_GO
G1_HUNG	G2_LYBI	A_SEVERE	N_FX N_INR	N_CONS	N_PROFITS	N_INSTITUTIONS	V_PERCEIVE
G1_ICEL	G2_MALA	A_MATURING	N_FX N_ISK	N_BCONF	N_BFUNDING	N_STRUCTURES	V_PREDICT
G1_INDI	G2_MORO	A_RIGID	N_FX N_MXN	N_CCONF	G1_EZ N_ESM	N_FLEXIBILITY	V_THINK
G1_IREL	G2_NIGE	A_FLEXIBLE	N_FX N_NZD	N_PMI	G1_EZ N_BANKUNION	N_RIGIDITY	V_CONVEY
G1_ITAL	G2_NKOR	A_UNSUSTAIN	N_FX N_NOK	N_INDU	N_LIQCRUNCH	N_RULING	V_ANNOUNCE
G1_JAPA	G2_PAKI	A_PROLONGED	N_FX N_SEK	N_MANUF	N_MACROPRUD	N_STABILITY	V_SUSTAIN
G1_LATAM	G2_PALE	A_RECURRING	N_FX N_DKK	N_CONSTR	N_ELECT	N_INSTABILITY	V_BLOCK
G1_LATV	G2_PANA	A_CONCERNED	N_FX N_PLN	N_EARN	N_PROTEST	G1_UK N_BREXIT	V_CHANGE
G1_LITH	G2_PERU		N_FX N_RON	N_UNEMP	N_CONFVOTE	G1_GREE N_GREXIT	V_EXIT
G1_LUXE	G2_PHIL		N_FX N_RUB	N_EMPL	N_PEACEK	N_CONFLICT	V_ENTER
G1_MALT	G2_SAUD		N_FX N_SGD	N_CPI	N_COUP	N_SUSTAIN	V_BREAKUP
G1_MEXI	G2_SERB		N_FX N_SKK	N_PPI	N_REBEL	G1_US N_FISCALCLIFF	V_ADOPT
G1_NETH	G2_SING		N_FX N_SIT	N_HOUSE	N_REVOL	N_DEBTCEIL	V_WITHDRAW
G1_NORW	G2_SOAF		N_FX N_ZAR	N_PRICE	N_PRIVATIZE	G1_US N_AUTOCUTS	V_RELAX
G1_NZ	G2_SRIL		N_FX N_TWD	N_BREAKUP	N_NATIONALIZE		V_WIDEN
G1_PIIGS	G2_SYRI		N_FX N_THB	N_ETRADE	N_WAR		V_LIMIT
G1_POLA	G2_TAIW		N_FX N_TRY	N_EXPORTS	N_ASSASS		V_REVALUE
G1_PORT	G2_THAI		N_FX N_UAH	N_IMPORTS	N_TERROR		V_DEVALUE
G1_ROMA	G2_TUNE		N_FX N_VND	N_EDEBT	N_CORRUPT		V_MISS
G1_RUSS	G2_UAE		N_FX N_VEF	N_FDI	N_POPULISM		V_DEPLETE
G1_SLOVAK	G2_URUG		N_FX N_COP	N_RES	N_CRISIS		V_REGAIN
G1_SLOVEN	G2_VENE		N_FX N_BGN	N_LIQUIDITY	N_PEACE		V_REJECT
G1_SPAI	G2_VIET		N_FX N_EGP	N_LENDING	N_ACCESSION		V_AGREE
G1_SWED	G2_YEME		N_FX N_ILS	N_PSI	N_COMMUNICATION		V_FAIL
G1_SWI	G3_AFR		N_FX N_KZT	N_INTLEND	N_TALKS		V_RECAPITAL
G1_TURK	G3_ASI		N_FX N_PEN	N_AUCTION	N_CHANGE		V_SAVE
G1_UK	G3_EUR		N_FX N_TND	N_CDS	N_AGREEMENT		V_PROTECT
G1_UKRA	G3_LAT		N_FX N_PAB	N_DEFAULT	N_STRAIN		V_EXPROP
G1_US			N_FX N_LKR	N_BONDS	N_FAILURE		V_IMPEACH
			N_FX N_UYU				V_IMPOSE
							V_PREVENT
							V_RESOLVE
							V_IMPLEMENT
							V_BREACH
							V_PLEDGE
							V_NEED
							V_DISAPPEAR
							V_REQUEST
							V_RECEIVE
							V_UNLOCK
							V_ALLEVIATE
							V_DEFAULT

Notes: Synonym labels that are inserted into the text where instances of tokens (or n-grams) are found, which belong to the given synonym group. For a detailed list of tokens and n-grams for each group, see Tables C.2 and C.3.

Table C.2
Synonym labels and associated tokens, n-grams

SYN_KEYS	TOKENS, N-GRAMS	N (000s)
NEG	doesnt, not, cant, didnt, wont, cannot, shouldnt, couldnt, no, wouldnt, nor, isnt, wasnt	1765.178
NEG2	despite, in spite of, regardless, although, albeit, notwithstanding	282.786
N_FX N_USD	usd, keyw_us us dollar, keyw_us dollar, keyw_us us currency, keyw_us currency	92.550
N_FX N_EUR	eur, single currency, european currency	27.980
N_FX N_GBP	gbp, pound sterling, british pound, poundsterling, british currency, keyw_uk currency	9.265
N_FX N_CHF	chf, swiss franc, swiss currency	25.104
N_FX N_JPY	jpy, japanese yen, yen, japans currency, japanese currency	294.168
N_FX N_CNY	cny, yuan, renminbi, chinese currency	111.789
N_FX N_ARS	ars, argentinas currency, argentin peso, argentinian peso	2.443
N_FX N_AUD	aud, australias currency, australian dollar, australian currency	20.825
N_FX N_BRL	brl, brazils currency, brazilian currency, brazilian peso, brazil peso	3.653
N_FX N_CAD	canadas currency, canadian currency, canadian dollar, canada dollar	32.841
N_FX N_CLP	clp, chiles currency, chilean currency, chilean peso, chile peso	3.058
N_FX N_HRK	hrk, croatias currency, croatian currency, croatian kuna, kuna	1.952
N_FX N_CZK	czk, czechs currency, czech koruna, czech krona	0.296
N_FX N_HKD	hkd, hong kongs currency, hong kong dollar	1.356
N_FX N_HUF	huf, hungarys currency, hungarian forint, hungarian currency, forint	8.629
N_FX N_KRW	krw, koreas currency, korean won	4.880
N_FX N_LVL	lvl, latvias currency, latvian lat, latvian currency, lat	0.673
N_FX N_LTL	ltl, lithuanias currency, lithuanian lita, lithuanian currency, lita	0.316
N_FX N_EEK	estonias currency, estonian kroon, estonian currency	0.023
N_FX N_MYR	myr, malaysias currency, malaysian ringgit, malaysian currency, ringgit	11.919
N_FX N_IDR	idr, indonesias currency, indonesian rupiah, indonesian currency, rupiah	53.502
N_FX N_INR	inr, indias currency, indian rupee, indian currency	4.907
N_FX N_ISK	isk, icelands currency, icelandic krona, icelandic currency	0.090
N_FX N_MXN	mxn, mexicos currency, mexican currency, mexican peso	4.030
N_FX N_NZD	nzd, new zealands currency, new zealand dollar	8.107
N_FX N_NOK	nrw, norways currency, norwegian krone, norwegian currency	0.802
N_FX N_SEK	sek, swedens currency, swedish krona, swedish currency	1.011
N_FX N_DKK	dkk, denmarks currency, danish krone, danish currency	0.302
N_FX N_PLN	pln, polands currency, polish zloty, polish currency, zloty	10.354
N_FX N_RON	ron, romanias currency, romanian leu, romanian currency, leu	5.909
N_FX N_RUB	russias currency, russian rubel, russian currency, rubel	0.479
N_FX N_SGD	sgd, singapores currency, singapore dollar	1.824
N_FX N_SKK	skk, slovakias currency, slovakian koruna, slovak koruna	0.305
N_FX N_SIT	slovenias currency, slovenian tolar, tolar	0.014
N_FX N_ZAR	zar, south africas currency, south african rand, south african currency, rand	14.988
N_FX N_TWD	twd, taiwans currency, taiwanese dollar, taiwanese currency	0.065
N_FX N_THB	thb, thailands currency, thai baht, thai currency, baht	17.169
N_FX N_TRY	turkeys currency, turkish lira, turkish currency	1.804
N_FX N_UAH	uah, ukrains currency, ukrainian hryvnia, ukrainian currency, hrivnya, hryvnia	2.427
N_FX N_VND	vnd, vietnams currency, vietnamese dong, vietnamese currency	0.183
N_FX N_VEF	vef, venezuelas currency, venezuelan bolivar, venezuelan currency	0.169
N_FX N_COP	colombian peso, colombias peso, colombian currency, colombias currency	0.973
N_FX N_BGN	bulgarian lev, bulgarias lev, bulgarian currency, bulgarias currency	0.177
N_FX N_EGP	egyptian pound, egypts pound, egyptian currency, egypts currency	3.164
N_FX N_ILS	israeli sekel, israels sekel, israeli currency, israels currency	0.245
N_FX N_KZT	kazakhstani tenge, kazakh tenge, kazakhstani currency, kazakh currency	0.175
N_FX N_PEN	peruvian peso, perus peso, peruvian currency, perus currency	0.057
N_FX N_TND	tnd, tunisian dinar, tunisias dinar, tunisian currency, tunisias currency	0.072
N_FX N_PAB	panamanian balboa	0.000
N_FX N_LKR	sri lankan rupee, sri lankas rupee, sri lankan currency, sri lankas currency	2.138
N_FX N_UYU	uruguayi peso, uruguays peso, uruguayi currency, uruguays currency	0.006
N_CB G1_UK N_BOE	boe, bank of england, mervyn king, mark carney	91.259
N_CB G1_EZ N_ECB	ecb, european central bank, trichet, draghi	206.359
N_CB G1_US N_FED	keyw_fed, fomc, federal reserve, yellen, bernanke, feds	512.164
N_CB G1_SWED N_RIKSBANK	riksbank	2.998
N_CB G1_GERM N_BUNDESBANK	bundesbank	7.670
N_CB G1_JAPA N_BOJ	boj, bank of japan	88.625
G1_EZ N_BANKUNION	banking union, single resolution, bank resolution, single supervisory mechanism, european deposit insurance	7.117
G1_EZ N_ELA	ela, emergency liquidity assistance	1.727
G1_EZ N_ESM	efsf, esm, european stability mechanism, european financial stability facility	20.444
G1_EZ N_LTRO	ltro, long term refinancing, longterm refinancing, targeted longterm, targeted long term, tltro	4.058
G1_EZ N_OMT	omt, outright monetary transaction, whatever it takes	3.470
G1_EZ N_SMP	smp, securities markets programme, securities markets program, securities market program, securities market programme	0.957
G1_GREE N_GREXIT	grexit	1.204
G1_UK N_BREXIT	brexit	10.755
G1_UK N_SLS	special liquidity scheme, sls	0.279
G1_US N_AUTOCUTS	sequester, automatic spending cuts	3.100
G1_US N_FISCALCLIFF	fiscal cliff	10.341
G1_US N_TAF	term auction facility, taf	0.776
G1_US N_TALF	term asset-backed securities, term assetbacked securities, talf	1.622
G1_US N_TAPER	tapering, taper tantrum	8.889
N_ACTUAL	actual, published, announced, announcement	202.472
N_AGREEMENT	agreement, approval, deal, accord	458.681
N_AID	aid, financial support, financial assistance, help	297.960
N_ASSASS	assassin, assassination	4.834
N_AUCTION	bond auction, debt auction, debt sale, bond sale, bond issuance, debt issuance	30.412
N_BAILOUT	bailout, bail out, financial help, financial assistance, rescue	171.300
N_BALANCE	balance, position	227.182
N_BANKS	banks, banking system, financial institutions, financial intermediaries, banking sector, financial sector, financial system, banking system, financial industry	795.424
N_BCONF	economic confidence, business confidence, business survey, investor confidence, investors confidence, business sentiment, business climate index, economic confidence	25.969
N_BONDS	bonds, securities	411.405
N_BRATE	policy rate, base rate, central bank rate, refinancing rate, repo rate	26.937
N_CAPADEQ	capital adequacy, capital position	7.337
N_CAPITAL	capital, equity	513.418
N_CB	central bank, monetary authorities	1346.342
N_CCONF	consumer confidence, consumer survey, consumer sentiment	35.607
N_CDS	keyw_cds, credit default swap, protection against default, insure against default, protect against default	18.775
N_CHANCE	chance, probability, possibility, likelihood, odds	133.292
N_CHANGE	change, modification, alteration, shift, adjustment, revision, adaptation, adopt	304.565
N_COMMUNICATION	signal, communication, statement, message, stance, rhetoric	310.954
N_CONCERN	concern, worry, worries, anxiety, fear, unease	483.362
N_CONFLICT	conflict, standoff, tension, clash, struggle, impasse, deadlock, stalemate, faceoff, row	214.548
N_CONFVOTE	vote of confidence, confidence vote	5.463
N_CONS	consumption, consumer demand, personal expenditure, household expenditure, durable goods, retail sale, consumer spending, household spending	94.231
N_CONSTR	constructions, construction ouptut, construction activity, construction sector	4.458
N_CONTROLS	controls, hurdles, restrictions, constraints, curbs, limits	83.560
N_CORRUPT	corruption, corrupt, nepotism, crony, cronies	39.114

Table C.2 (continued)
Synonym labels and associated tokens, n-grams

SYN_KEYS	TOKENS, N-GRAMS	N (000s)
N_COUP	coup, overthrow, rebellion, government takeover	35.451
N_CPI	consumer price index, cpi	27.423
N_CRISIS	crisis, turmoil, turbulence, chaos, disorder, disarray, mayhem, meltdown, mess, distress	426.381
N_DEBT	debt, liabilities, obligations	646.659
N_DEBTCEIL	debt ceiling	10.137
N_DECREASE	reduction, shrinkage, loss, cutback, waning, descent, deceleration	142.362
N_DEFAULT	credit event, debt swap, default, restructuring, moratorium, arrear, bankruptcy	168.655
N_DEFENSE	defense, military	161.646
N_DEFICIT	deficit, shortfall, gap	190.216
N_DETERIORATION	deterioration, worsening, fading, weakening, disappointment	67.062
N_EARN	earnings season, corporate earnings, earning announcements, earning season, earnings announcements	19.964
N_EDEBT	external debt, external liabilities, foreign liabilities, foreign debt	3.087
N_ELECT	election, referendum, presidential campaign	304.383
N_EMPL	employment, payrolls, payroll count, payroll number, payroll figure, employment figure, employment number, employment count, jobs creation, job growth, jobs growth	93.533
N_ETRADE	trade, current account, balance of payment, bop, balanceofpayment	318.005
N_EXPORTS	exports, export growth, export number, export figure, export	181.067
N_FAILURE	failure, shutdown, breakdown, collapse	105.523
N_FDI	foreign direct investment, fdi, direct investment	8.280
N_FORECAST	outlook, forecast, expectation, prediction, prognosis, prognoses, projection, estimate, consensus, speculation	778.244
N_FREEDOM	free, liberalize, liberalise, liberalization, liberalisation, freedom, deregulate, deregulation	77.118
N_GDP	gdp, gross domestic product, gni, nni, national income, national output, economic growth, economic output, economic activity, economic conditions, economic indicators, real growth, potential output	260.856
N_GOVT	public, fiscal, budget, budgetary, government, sovereign, state	2129.147
N_HHI	disposable income, personal income, household income	7.631
N_HOPE	hope, prospect	223.750
N_HOUSE	house, housing, real estate, home, dwelling, property	682.725
N_IMPORTS	imports, import growth, import number, import figure, import	87.955
N_IMPROVEMENT	improvement, enhancement, advance, progress, strengthening	98.640
N_INCREASE	upsurge, escalation, expansion, quickening, acceleration	342.893
N_INDU	industrial output, industrial production, industrial activity, factory output, factory orders, capacity utilisation, capacity utilization, industrial orders, factory activity, factory sector	52.660
N_INSTABILITY	instability, weakness, fragility, uncertainty, vulnerability	151.580
N_INTLEND	troika, international lenders, official lenders, imf, international monetary fund, world bank, international creditors, official creditors	146.752
N_LABORM	labor market, employment, working hours, labour market, hiring, firing, sacking, laying off, lay off, severance, notice period	109.152
N_LENDING	disbursement, facility, tranch, instalment	583.232
N_LIQCRUNCH	credit crunch, liquidity crunch, liquidity squeeze, credit squeeze	21.147
N_LIQUIDITY	liquidity, financing, funding, cash reserves	279.693
N_MACROPRUD	macroprudential, macro prudential	1.805
N_MANUF	manufacturing output, manufacturing production, manufacturing activity, manufacturing sector	21.363
N_NATIONALIZE	nationalise, nationalised, nationalisation, nationalize, nationalized, nationalization	11.192
N_NPL	npl, nonperforming, non performing, arrear	18.548
N_NUMBER	figure, number, data, statistics	884.022
N_PANIC	panic, alarm, terror, horror, fright, shock	40.268
N_PEACE	peace, truce, ceasefire	79.490
N_PEACEK	peace keeping, peacekeeping, peace keeper, peacekeeper	11.986
N_PMI	purchasing manager, pmi	42.077
N_POPULISM	populism, populist	5.718
N_PORTFOLIO	portfolio, balance sheet, asset quality	114.728
N_PPI	producer price index, ppi	7.661
N_PRICE	value, valuation	145.163
N_PRIVATIZE	privatise, privatised, privatisation, privatize, privatized, privatization	16.503
N_PROFITS	profitability, profits, earnings, income, roe, roa	346.737
N_PROPRIGHTS	property rights, private property, private ownership, ownership rights	1.874
N_PROTEST	protests, demonstrations, general strike, mass demonstration, protester, demonstrator	126.787
N_PSI	psi, private sector involvement	2.195
N_QE	qe, quantitative easing, largescale asset purchase, large scale asset purchase, qe1, qe2, qe3, qe4, operation twist, bond buying programme, bond buying program, asset purchase programme, asset purchase program, asset purchases, bond purchases, bondbuying	92.583
N_REBEL	rebel, militant, separatist, insurgent	200.573
N_RECAPITAL	recapitalization, recapitalisation	7.693
N_REGULATIONS	rules, regulations, directives, laws	127.733
N_REQRESERVES	reserve requirement, required reserve	9.169
N_RES	currency reserve, official reserve, central bank reserve, international reserve, foreign exchange reserve, fx reserve	12.729
N_REVENUE	revenue, income	236.122
N_REVOL	revolution, uprising, civil war, civil conflict, anarchy, hostilities, insurgency, civil unrest	65.097
N_RIGIDITY	rigidity, stiffness, bureaucracy	2.955
N_RISK	risk, threat	390.709
N_RULELAW	rule of law, legal system, judicial system, regulatory framework, legal framework, judicial framework	7.131
N_SOCIAL	safety net, social net, pension, health	127.054
N_SPENDING	expenditure, outlay, spending	217.329
N_STABILITY	stability, strength, certainty, firmness	132.018
N_STRAIN	challenge, stress, headwind, strain, pressure	317.427
N_STRUCTURES	pension, labor market, labour market, health care, tax system	98.740
N_SUSTAIN	sustainability, sustainable	27.122
N_TALKS	negotiation, talks, diplomatic effort, diplomacy	240.238
N_TERROR	terrorist attack, bomb attack, bombing, terrorists, terrorist incident	23.649
N_THAN	than, compared to, compared with, relative to	935.455
N_TOXIC	toxic asset, illiquid asset, troubled asset, toxic mortgage asset	7.981
N_TROUBLE	difficulties, problem, trouble	206.393
N_UNEMP	unemployment, jobless claim, continuing claims, initial claims, jobless rate, new jobless, jobs claims	128.875
N_WAR	military conflict, hostilities, warfare	11.433
N_ZLB	zlb, zero bound, zero lower bound	0.313
V_ACCELERATE	quicken, accelerate, fasten, rapid, hasten, speed, heat, perk up, gather steam	107.426
V_ACHIEVE	achieve, accomplish, arrive at, reach, broker, restore	295.952
V_ADOPT	adopt	26.273
V_AGREE	agree, approve, authorize, authorise	299.637
V_ALLEVIATE	soothe, alleviate, calm	8.288
V_ANNOUNCE	announce, reveal, publish, broadcast, distribute, issue, print, post, disclose	682.545
V_BECOME	become, get, grow, turn out	608.393
V_BEGIN	begin, initiate, start, commence, instigate, create, open, launch, embark, prompt, rebuild, set off, introduce, create	856.585
V_BLOCK	bar, block, obstruct, obstruct, impede, thwart	69.928
V_BREACH	breach, violate, renege	25.907
V_BREAKUP	break up, disintegrate, dissolve	10.127
V_CHANGE	change, alter, modify, shift, adjust, amend, transform, revise, overhaul	476.796
V_CONVEY	say, speak, mention, declare, articulate, convey, communicate, answer, reply, express, voice, state, confirm, affirm, insist, acknowledge, tell	5291.492
V_CRUSH	abolish, terminate, extinguish, obliterate, devastate, wipe out, break, wreck, crush, subdue, defeat	159.422
V_CUT	reduce, cut, lower, dampen, moderate, curb, lessen, slash, scale back, drag down, halve, erode, bring down	1099.152
V_DECELERATE	decelerate, slow, brake, cool	167.504
V_DEFAULT	default, restructure, reschedule	152.293
V_DEPLETE	deplete, drain, exhaust	13.678
V_DEVALUE	devaluation, devalue	11.676
V_DISAPPEAR	disappear, evaporate, vanish	9.345
V_EASE	ease, cut	673.057

Table C.2 (continued)
Synonym labels and associated tokens, n-grams

SYN_KEYS	TOKENS, N-GRAMS	N (000s)
V_END	end, finish, terminate, stop, cease, interrupt, cancel, break, remove	708.229
V_ENTER	enter, join, accede, accession, entrance	99.779
V_EXIT	exit, leave, secede	134.787
V_EXPROP	expropriate, seize, confiscate	27.494
V_FAIL	break down, fail, collapse, disappoint	208.368
V_FALL	decrease, fall, drop, lower, reduce, slacken, decline, wane, fade, shrink, sink, dwindle, diminish, contract, moderate, narrow, subtract, dip, plunge, slide, plummet, lose, shed, shrink, halve	2472.326
V_GO	drift, pull, push, go, move, shift, step, trend, edge	1118.052
V_HISTORY	used to, had been, historically, in the past, past year, past years, past decade, past decades, long ago, last year, chronology	411.142
V_IMPEACH	impeach	1.184
V_IMPLEMENT	implement, carry out, fulfill, execute, undertake, accomplish	69.477
V_IMPOSE	impose, enforce, enact, levy	78.849
V_IMPROVE	improve, better, upgrade, recover, mend	332.139
V_LIMIT	limit, restrain, constrain, curb, restrict, curtail, trim	239.790
V_MISS	miss, fail	152.391
V_NEED	need, require	422.763
V_PERCEIVE	perceive, feel, sense	95.696
V_PLEDGE	pledge, promise, vow	134.274
V_PREDICT	predict, forecast, foresee, envisage, calculate, foretell, anticipate, expect, estimate, project, speculate	1278.910
V_PREVENT	prevent, avert, avoid, offset	183.397
V_PROTECT	protect, defend, guard, safeguard, preserve, support, endorse	506.940
V_RAISE	increase, raise, boost, lift, hike, advance, intensify, double	1225.850
V_RECAPITAL	recapitalize, recapitalise	7.677
V_RECEIVE	receive, acquire, obtain, clinch, secure	176.271
V_REGAIN	regain, gain, return	446.540
V_REJECT	reject, deny, refuse	119.345
V_RELAX	relax, slacken, loosen, unwind	31.613
V_REQUEST	request, turn to, ask for, seek	180.964
V_RESOLVE	resolve, solve, tackle, address	127.061
V_REVALUE	revaluation, revalue	3.373
V_RISE	increase, rise, lift, boost, elevate, augment, expand, soar, swell, pick up, add, gain, climb, rebound, surge, intensify, jump, double, triple	2636.574
V_SAVE	save, bail, rescue	105.085
V_STRENGTHEN	strengthen, bolster, boost, reinforce, support, aid, assist, promote, prop up, encourage, shore up	982.240
V_SURPASS	surpass, exceed, beat, outshine, outstrip, top, transcend, trounce, above	510.007
V_SUSTAIN	sustain, maintain, stay, hold, keep	758.408
V_THINK	think, believe, assume, presume, guess, reckon, suspect, suppose, imagine	534.059
V_TIGHTEN	tighten, hike	163.944
V_TRAIL	trail, lag, below, lag	202.378
V_UNLOCK	unlock, release, disburse, pay out	175.678
V_WEAKEN	weaken, impair, undermine, dent, exhaust, sap, damage, harm, injure, wane, fade, sway	222.325
V_WIDEN	widen, extend, expand, broaden, add to, spread, deepen	400.074
V_WITHDRAW	withdraw	22.180
V_WORSEN	worsen, deteriorate, downgrade, crumble	132.393
A_AMPLE	plenty, sufficient, abundant, ample	34.310
A_BAD0	mediocre, middling, unexceptional, modest	32.931
A_BAD1	bad, negative, disappointing, adverse, unsatisfactory, poor, inadequate, unfavorable, unfavourable, troubling, worrying, discouraging, sour, unpleasant, meagre, meager, gloomy, woeful, dark, pessimistic, weak, ailing, struggling, troubled, dim	497.892
A_BAD2	terrible, horrible, awful, dismal, abysmal, dreadful, appalling, horrifying, horrific, frightful, harrowing, depressing, upsetting, disillusioning, disheartening, frustrating, disenchanting, disconcerting, shocking, distressing, disturbing, worst	87.106
A_BETTER	better, nicer, sounder, safer, superior, stronger, brighter, more normal	155.873
A_CONCERNED	afraid, worried, concerned	78.650
A_FAILED	failed, vain, unsuccessful, abortive, fruitless, futile, ineffective	75.287
A_FLEXIBLE	flexible	0.000
A_GOOD0	adequate, reasonable, suitable, appropriate, satisfactory, acceptable	52.758
A_GOOD1	good, positive, decent, upbeat, favorable, promising, encouraging, reassuring, benign, pleasing, sound, favourable, strong	529.792
A_GOOD2	excellent, brilliant, outstanding, superb, exceptional, splendid, ideal, perfect, astonishing, fantastic, amazing, breathtaking, best, top	359.174
A_HIGHER	larger, higher, increased, greater, elevated, excessive, bigger	516.944
A_INSTABLE	shaky, wobbly, instable, fragile, delicate, flimsy, breakable, brittle, unstable, uneven, unsteady, volatile, erratic, weak, feeble, vulnerable, uncertain	246.788
A_LARGE1	large, sizable, big, great, considerable, significant, substantial, sizeable, high, major, mounting	987.641
A_LARGE2	huge, enormous, extreme, intense, excessive, vast, colossal, gigantic, massive, oversized, soaring, highest, largest, immense, most	717.769
A_LOWER	lower, decreased, reduced, lesser, smaller, short of, dim	389.732
A_MATURING	maturing, oncoming, coming due	15.781
A_PROLONGED	continuing, prolonged, protracted, lingering, lasting, persistent, recurring, frequent, remaining, persisting, returning, reappearing, relapsing, periodic	115.026
A_RECURRING	recurring, frequent, remaining, returning, reappearing, relapsing, periodic	48.280
A_RIGID	rigid, bureaucratic, stiff	6.243
A_SCARCE	scarce, inadequate, lacking, short supply, scarcity, shortage, starved of	20.533
A_SEVERE	severe, serious, grave, harsh, stark, critical, acute, sharp	167.896
A_SMALL1	small, minor, insignificant, unimportant, lesser, slight, trivial, little, low, muted, subdued, tepid	600.337
A_SMALL2	tiny, undersized, miniature, mini, diminutive, minuscule, smallest, bottom, lowest, least	261.374
A_STABLE	stable, strong, longstanding, unwavering, steady, enduring, balanced, certain, resilient, solid	461.030
A_UNSUSTAIN	unsustainable, unmanageable, unmaintainable	6.460
A_WORSE	worse, inferior, poorer, weaker, gloomier, darker	85.811

Notes: Number of matches (last column; thousands) are based on the relevance filtered news data set aggregated across tokens and n-grams for each synonym group.

Table C.3
Geographic group labels and associated tokens, n-grams

GEO_KEYS	TOKENS, N-GRAMS	N (000s)
G1_ARGE	argentina,argentine,argentinian,buenos aires	61.515
G1_AUSL	sydney,melbourne,australia,australian,canberra	151.876
G1_AUT	austria,vienna	29.590
G1_BELG	belgium,belgian,brussels	80.904
G1_BRAZ	brazil,brazilian,brasília,brasilia,rio de janeiro,sao paulo	188.422
G1_BULG	bulgaria,bulgarian,sofia	21.048
G1_CANA	canada,canadian,ottawa,toronto,montreal,quebec	328.488
G1_CHIN	china,chinese,beijing,shanghai,shenzhen,guangzhou	880.419
G1_CROA	croatia,croatian,zagreb	14.276
G1_CYPR	cyprus,cypriot,nicosia	34.423
G1_CZEC	czech,prague	42.140

Table C.3 (continued)
Geographic group labels and associated tokens, n-grams

GEO_KEYS	TOKENS, N-GRAMS	N (000s)
G1_DENM	denmark,danish,copenhagen	29.994
G1_EMEA	emea,cee,ceemea,eastern europe,eastern european,postsocialist,post socialist,postcommunist,post communist,transition countries	21.490
G1_ESTO	estonia,estonian,tallinn	6.649
G1_EU	eu,european union,keyw_eu	360.588
G1_EZ	eurozone,euro zone,euro area,euroarea,eurobloc,european monetary union,euro bloc,emu	797.020
G1_FINL	finland,finnish,helsinki	19.089
G1_FRAN	france,french,paris,marseille	318.527
G1_GERM	german,berlin,munich,hamburg,stuttgart,germany	377.956
G1_GLOB	global,world	963.378
G1_GREE	greece,greek,athens	458.670
G1_HUNG	hungary,hungarian,budapest	52.571
G1_ICEL	iceland,icelandic,reykjavik	18.146
G1_INDI	india,indian,mumbai,delhi	206.309
G1_IREL	ireland,irish,dublin	131.547
G1_ITAL	italy,italian,rome,milan	242.246
G1_JAPA	japan,japanese,tokyo,kyoto	581.207
G1_LATAM	latin america,south america,latin american,south american	55.139
G1_LATV	latvia,latvian,riga	10.596
G1_LITH	lithuania,lithuanian,vilnius	8.666
G1_LUXE	luxembourg	11.553
G1_MALT	malta,maltese,valletta	3.881
G1_MEXI	mexico,mexican	120.766
G1_NETH	netherlands,dutch,amsterdam	51.727
G1_NORW	norway,norwegian,oslo	35.500
G1_NZ	new zealand,new zealander,oakland	38.781
G1_PIIGS	piigs,giips,gips,periphery countries,euro periphery,zone periphery,euro peripheral,zone peripheral	2.894
G1_POLA	poland,polish,warsaw	63.765
G1_PORT	portugal,portuguese,lisbon	82.590
G1_ROMA	romania,romanian	23.563
G1_RUSS	russia,russian,moscow,saint petersburg	502.342
G1_SLOVAK	slovakia,slovakian,bratislava	18.633
G1_SLOVEN	slovenia,slovenian,ljubljana	16.080
G1_SPAI	spain,spanish,madrid,barcelona	247.283
G1_SWED	sweden,swedish,stockholm	45.721
G1_SWI	switzerland,swiss,bern	99.506
G1_TURK	turkey,turkish,ankara,istanbul	169.543
G1_UK	keyw_uk,united kingdom,britain,british,wales,scotland,england,english,london,scotch,scottish,welsh	967.911
G1_UKRA	ukraine,ukrainian,kiev	178.469
G1_US	united states,keyw_us,usa,washington,new york,chicago,san francisco,los angeles,boston,miami,houston,philadelphia	3532.952
G2_AFGH	afghanistan,afghan,kabul	72.624
G2_AFR	africa,african	191.799
G2_ALGE	algeria,algerian,algiers	17.317
G2_ASIA	asia,asian	210.368
G2_BOLI	bolivia,bolivian,la paz	7.412
G2_CHIL	chile,chilean,santiago	49.385
G2_COLO	colombia,colombian,bogota	45.241
G2_DEV	advanced economies,oecd countries,developed countries	9.958
G2_ECUA	ecuador,ecuadorian,quito	11.534
G2_EGYP	egypt,egyptian,cairo	115.858
G2_EM	emerging market,emerging world,third world,developing country,developing countries,developing world,emerging economies	90.344
G2_GEOR	georgia,georgian,tbilisi	37.603
G2_HONG	hong kong	51.050
G2_INDO	indonesia,indonesian,jakarta	69.237
G2_IRAN	iran,iranian,tehran	153.997
G2_IRAQ	iraq,iraqi,baghdad	95.007
G2_ISRA	israel,israeli,jerusalem,tel aviv	35.369
G2_JAMA	jamaica,jamaican,kingston	2.612
G2_KAZA	kazakhstan,kazakh,astana	21.087
G2_KORE	south korea,south korean,seoul	83.711
G2_LEBA	lebanon,lebanese,beirut	29.243
G2_LYBI	libya,libyan,tripoli	119.487
G2_MALA	malaysia,malaysian,malay,kuala lumpur	55.715
G2_MORO	morocco,moroccan,rabat	10.183
G2_NIGE	nigeria,nigerian,abuja,lagos	57.160
G2_NKOR	north korea,north korean,pyongyang	22.951
G2_PAKI	pakistan,pakistani,islamabad,karachi	79.664
G2_PALE	palestinian,palestine,gaza,ramallah	22.613
G2_PANA	panama	6.084
G2_PERU	peru,peruvian,lima	24.960
G2_PHIL	philippines,manila	26.942
G2_SAUD	saudi,saudi arabia,riyadh	116.649
G2_SERB	serbia,serbian,belgrade	26.079
G2_SING	singaporean,singapore	64.368
G2_SOAF	south africa,south african,pretoria,cape town,johannesburg	58.213
G2_SRIL	sri lanka,colombo	27.859
G2_SYRI	syria,syrian,damascus	193.976
G2_TAIW	taiwan,taiwanese,taipei	32.904
G2_THAI	thailand,thai,bangkok	100.592
G2_TUNE	tunisia,tunisian,tunis	25.021
G2_UAE	uae ,dubai,abu dhabi,arab emirates	72.079
G2_URUG	uruguay,montevideo	3.862
G2_VENE	venezuela,venezuelan,caracas	51.888
G2_VIET	vietnam,vietnamese,hanoi	31.045
G2_YEME	yemen,yemeni,sanaa	43.090
G3_AFR	ethiopia, ethiopian, addis ababa, congo, congolese, kinshasa, tanzania, tanzanian, kenya, kenyan, nairobi, uganda, ugandan, kampala, sudan, sudanese, khartoum, ghana, accra, mozambique, maputo, madagascar, antananarivo, cote divoire, abidjan, cameroon, burkina faso, ouagadougou, niamey, malawi, lilongwe, senegal, dakar, angola, luanda, mali , bamako, zambia, lusaka, zimbabwe, harare, rwanda, kigali, chad, guinea, conakry, somalia, mogadishu, burundi, bujumbura, sierra leone, eritrea, asmara, bangui, liberia, monrovia, mauritania, nouakchott, lesotho, namibia, windhoek, botswana, gaborone, gambia, bissau, gabon, libreville, mauritius, port louis, swaziland	341.565
G3_ASI	bangladesh, dhaka, burma, naypyidaw, nepal, kathmandu, uzbek, tashkent, cambodia, phnom penh, azerbaijan, baku, tajik, dushanbe, laos, vientiane, jordan, amman, kyrgyz, bishkek, turkmen, ashgabat, mongolia, muscat, armenia, yerevan, kuwait, qatar, doha, bahrain, manama, east timor, bhutan, brunei, bandar seri begawan, maldives	129.531
G3_EUR	albania, tirana, belarus, minsk, bosnia, sarajevo, gibraltar, guernsey, jersey, saint helier, kosovo, pristina, liechtenstein, vaduz, macedonia, skopje, moldova, chisinau, monaco, montenegro, podgorica, transnistria, tiraspol, vatican	76.196
G3_LAT	antigua, bahamas, nassau, barbados, costa rica, costa rican, cuba, cuban, havana, dominica, dominican republic, santo domingo, el salvador, san salvador, grenada, guatemala, guyana, georgetown, haiti, portauprince, honduras, tegucigalpa, nicaragua, managua, paraguay, asuncion, suriname, trinidad	64.082

Notes: Number of matches (last column; thousands) are based on the relevance filtered news data set aggregated across tokens and n-grams for each geography group.

Table C.4
Intermediate expressions

EXPR_KEYS	EXPR CODE	p1	p2	p3	N (000s)
E_EXPECT	(p1,p2,1,0)	N_THAN	N_FORECAST | V_PREDICT | thought | perceived | assumed | presumed | believed | guessed | reckoned | suspected | supposed | imagined | hoped		90.478
E_SURP	p1	(surprisingly | unexpectedly | shockingly)			31.824
E_BETTER1	(p1,p2,1,0)	A_BETTER | A_GOOD1	E_EXPECT		79.321
E_BETTER2	(p1,p2,1,0)	A_GOOD2	E_EXPECT		0.013
E_BETTER1	(p1,p2,1,0)	V_SURPASS | overshot | overshooting | overshoots	N_FORECAST		79.321
E_WORSE1	(p1,p2,1,0)	A_WORSE | A_BAD1	E_EXPECT		37.258
E_WORSE2	(p1,p2,1,0)	A_BAD2	E_EXPECT		0.023
E_WORSE1	(p1,p2,1,0)	V_TRAIL | undershot | undershooting | undershoots	N_FORECAST		37.258
E_HIGHER1	(p1,p2,1,0)	A_HIGHER | A_LARGE1	E_EXPECT		38.002
E_HIGHER2	(p1,p2,1,0)	A_LARGE2	E_EXPECT		0.116
E_LOWER1	(p1,p2,1,0)	A_LOWER | A_SMALL1	E_EXPECT		30.287
E_LOWER2	(p1,p2,1,0)	A_SMALL2	E_EXPECT		0.060
E_IMPROVEMENT0	(p1,p2,1,0)	A_SMALL1 | A_SMALL2	N_IMPROVEMENT		1.676
E_IMPROVEMENT2	(p1,p2,1,0)	A_LARGE2	N_IMPROVEMENT		0.650
E_DETERIORATION2	(p1,p2,1,0)	A_LARGE2	N_DETERIORATION		0.240
E_DETERIORATION0	(p1,p2,1,0)	A_SMALL1 | A_SMALL2	N_DETERIORATION		0.363
E_INCREASE2	(p1,p2,1,0)	A_SMALL1 | A_SMALL2	N_INCREASE		2.236
E_INCREASE0	(p1,p2,1,0)	A_LARGE2	N_INCREASE		1.816
E_DECREASE2	(p1,p2,1,0)	A_LARGE2	N_DECREASE		1.292
E_DECREASE0	(p1,p2,1,0)	A_SMALL1 | A_SMALL2	N_DECREASE		0.769
V_RISE	(p1,p2,1,0)	V_GO	higher | up		2636.574
V_FALL	(p1,p2,1,0)	V_GO	lower | down		2472.326
V_WORSEN	(p1,p2,1,0)	V_BECOME	A_WORSE		132.393
V_IMPROVE	(p1,p2,1,0)	V_BECOME	A_BETTER		332.139
E_BETTER1	p1	(betterthanexpected | strongerthanexpected)			79.321
E_WORSE1	p1	(worsethanexpected | weakerthanexpected)			37.258
E_HIGHER1	p1	(higherthanexpected | largerthanexpected | morethanexpected | revised up | revise up | revising up | revises up)			38.002
E_LOWER1	p1	(lowerthanexpected | smallerthanexpected | lessthanexpected | revised down | revise down | revising down | revises down)			30.287
E_HIGHER1	((p1,p2,1,0),p3,1,0)	V_RISE	more	E_EXPECT	38.002
E_LOWER1	((p1,p2,1,0),p3,1,0)	V_FALL	more	E_EXPECT	30.287
E_HIGHER0	((p1,p2,1,0),p3,1,0)	V_RISE	less	E_EXPECT	1.868
E_LOWER0	((p1,p2,1,0),p3,1,0)	V_FALL	less	E_EXPECT	2.077
E_BETTER1	((p1,p2,1,0),p3,1,0)	V_STRENGTHEN | V_IMPROVE	more	E_EXPECT	79.321
E_WORSE1	((p1,p2,1,0),p3,1,0)	V_WEAKEN | V_WORSEN	more	E_EXPECT	37.258
E_BETTER1	(p1,p2,2,0)	E_SURP	A_GOOD1 | A_BETTER		79.321
E_WORSE1	(p1,p2,2,0)	E_SURP	A_BAD1 | A_WORSE		37.258
E_BETTER2	(p1,p2,2,0)	E_SURP	A_GOOD2		0.013
E_WORSE1	(p1,p2,2,0)	E_SURP	A_BAD2		37.258
E_HIGHER2	(p1,p2,2,0)	E_SURP	A_LARGE2		0.116
E_LOWER2	(p1,p2,2,0)	E_SURP	A_SMALL2		0.060
E_HIGHER1	(p1,p2,2,0)	E_SURP	A_LARGE1		38.002
E_LOWER1	(p1,p2,2,0)	E_SURP	A_SMALL1		30.287
E_HIGHER1	(p1,p2,1,0)	short of | below | under	N_FORECAST		38.002
E_LOWER1	(p1,p2,1,0)	in excess of | over | above	N_FORECAST		30.287
E_PRED_RISE1	(p1,p2,2,0)	N_FORECAST | V_PREDICT	E_UP | N_INCREASE | E_INCREASE0 | E_INCREASE2 | V_RISE		57.400
E_PRED_FALL1	(p1,p2,2,0)	N_FORECAST | V_PREDICT	E_DOWN | E_DECREASE0 | N_DECREASE | E_DECREASE2 | V_FALL | V_CUT		73.461
E_PRED_IMPROVE	(p1,p2,2,0)	N_FORECAST | V_PREDICT	V_STRENGTHEN | V_IMPROVE		25.638
E_PRED_WORSEN	(p1,p2,2,0)	N_FORECAST | V_PREDICT	V_WEAKEN | V_WORSEN		8.734
E_PRED_HIGH	(p1,p2,2,0)	N_FORECAST | V_PREDICT	A_LARGE2		11.440
E_PRED_LOW	(p1,p2,2,0)	N_FORECAST | V_PREDICT	A_SMALL2		2.198
E_PRED_HIGH2	(p1,p2,2,0)	N_FORECAST | V_PREDICT	A_LARGE2		0.002
E_PRED_LOW2	(p1,p2,2,0)	N_FORECAST | V_PREDICT	A_SMALL2		0.000
E_PRED_GOOD	(p1,p2,2,0)	N_FORECAST | V_PREDICT	A_GOOD1		9.503
E_PRED_BAD	(p1,p2,2,0)	N_FORECAST | V_PREDICT	A_BAD1		20.348
E_PRED_GOOD2	(p1,p2,2,0)	N_FORECAST | V_PREDICT	A_GOOD2		1.913
E_PRED_BAD2	(p1,p2,2,0)	N_FORECAST | V_PREDICT	A_BAD2		0.932
E_PROB_HIGH	p1	(probably | in all probability | likely | almost certainly | doubtless | undoubtedly | no doubt | without a doubt | definitely)			38.272
E_PROB_MED	p1	(perhaps | maybe | uncertain | possibly | uncertain | questionable)			42.863
E_PROB_LOW	p1	(unlikely | doubtful | improbable)			7.083
E_PROB_HIGH	(p1,p2,2,0)	N_CHANCE	A_HIGHER | A_LARGE1 | A_LARGE2		38.272
E_PROB_LOW	(p1,p2,2,0)	N_CHANCE	A_LOWER | A_SMALL1 | A_SMALL2		7.083
E_PROB_RISE	(p1,p2,2,0)	N_CHANCE	E_UP | V_RISE | V | RAISE		4.617
E_PROB_FALL	(p1,p2,2,0)	N_CHANCE	E_DOWN | V_FALL | V_CUT | V_LIMIT		5.148
E_RISK_HIGH	(p1,p2,1,0)	A_HIGHER | A_LARGE1 | A_LARGE2	risk of | threat of | risk regarding | risk concerning | risk relating to | risk related to | risks regarding | risks concerning | risks relating to | risks related to		2.588
E_RISK_LOW	(p1,p2,1,0)	A_LOWER | A_SMALL1 | A_SMALL2	risk of | threat of | fear of | risk that | fears that | risk regarding | threat regarding | fear regarding | risk concerning | threat concerning | fear concerning		1.203
E_RISK_RISE	(p1,p2,1,0)	E_UP | V_RISE | V | RAISE	risk of | threat of | risk regarding | risk concerning | risk relating to | risk related to | risks regarding | risks concerning | risks relating to | risks related to		0.495
E_RISK_FALL	(p1,p2,1,0)	E_DOWN | V_FALL | V_CUT | V_LIMIT | V_ALLEVIATE	risk of | threat of | fear of | risk that | fears that | risk regarding | threat regarding | fear regarding | risk concerning | threat concerning | fear concerning		2.561
E_CONCERN_CONT	(p1,p2,2,0)	N_CONCERN | N_TROUBLE | N_STRAIN	A_PROLONGED | A_RECURRING		84.475
E_CONCERN_HIGH	(p1,p2,2,0)	N_CONCERN | N_TROUBLE | N_STRAIN	A_HIGHER | A_LARGE1 | A_LARGE2		37.644
E_CONCERN_LOW	(p1,p2,2,0)	N_CONCERN | N_TROUBLE | N_STRAIN	A_LOWER | A_SMALL1 | A_SMALL2		5.848
E_CONCERN_RISE	(p1,p2,2,0)	N_CONCERN | N_TROUBLE | N_STRAIN	E_UP | V_RISE | V | RAISE | V_BEGIN		40.595
E_CONCERN_FALL	(p1,p2,2,0)	N_CONCERN | N_TROUBLE | N_STRAIN	E_DOWN | V_FALL | V_CUT | V_EASE | V_LIMIT | V_END | V_ALLEVIATE		47.615
E_CONCERN_FALL	(p1,p2,2,0)	N_TROUBLE	V_RESOLVE		47.615
E_HOPE_CONT	(p1,p2,2,0)	N_HOPE	A_PROLONGED | A_RECURRING		0.283
E_HOPE_HIGH	(p1,p2,2,0)	N_HOPE	A_HIGHER | A_LARGE1 | A_LARGE2		3.341
E_HOPE_LOW	(p1,p2,2,0)	N_HOPE	A_LOWER | A_SMALL1 | A_SMALL2		3.394
E_HOPE_RISE	(p1,p2,2,0)	N_HOPE	E_UP | V_RISE | V | RAISE | V_BEGIN		6.577
E_HOPE_FALL	(p1,p2,2,0)	N_HOPE	E_DOWN | V_FALL | V_CUT | V_EASE | V_LIMIT | V_END		7.556
V_AGREE	(p1,p2,2,0)	sign | signing | signed | signs | V_ACHIEVE	N_AGREEMENT		299.637
V_REJECT	(p1,p2,1,0)	fail to | failure to | fails to | failing to	V_AGREE | V_IMPOSE | V_IMPLEMENT		119.345
E_CONCERN_CONT	p1	A_CONCERNED			84.475

Notes: EXPR_KEYS are intermediate expression labels inserted into the text. EXPR CODE defines the proximity and ordering rules of expression elements and are functions with four arguments: the first two arguments are expression elements (can be tokens, n-grams or intermediate expressions or a list of these), the third is a binary indicating if ordering should (value of 1) be preserved or is flexible (value of 2), the fourth identifies proximity (0: distance of maximum three tokens no punctuation marks allowed; 1: distance of maximum 10 tokens comma allowed; 2: distance of maximum 15 words, comma allowed 3: distance restricted only by full stop or paragraph break). Number of matches (last column; thousands) are based on the relevance filtered news data set aggregated across tokens and n-grams for each geography group.

Table C.5
Fundamental expression structures

EXPR_CODE	MOD	SIGN	p1	p2	p3	p4
REAL(+)						
(p1,p2,1,0)	CH	–	N_HOUSE	bust | burst		
(p1,p2,2,1)	CH	+	recovery	V_WIDEN		
p1	CH	-	into recession			
p1	CH	+	out of recession | out from recession			
(p1,p2,2,1)	CH	-	recession	enter | enters | entered | entering | V_BEGIN | V_WIDEN		
(p1,p2,2,1)	CH	+	recession	exit | exits | V_END | V_PREVENT		
(p1,p2,2,1)	CH	-	recession	V_WIDEN		
p1		-	depression | depressed economy			
((p1,p2,1,0),p3,2,1)	CH	+	economic	N_CRISIS	V_EASE | V_LIMIT | V_END	
((p1,p2,1,0),p3,2,1)	CH	-	economic	N_CRISIS	V_WIDEN | V_PREVENT	
EXTERN(+)						
(p1,p2,2,1)	CH	0	capital inflows	V_ACCELERATE | V_DECELERATE | V_FALL | V_RISE | V_BEGIN | V_END | A_LARGE2 | A_SMALL2 | A_LARGE1 | A_SMALL1 | E_GO_DOWN | E_GO_UP		
(p1,p2,2,1)	CH	0	capital outflows	V_ACCELERATE | V_DECELERATE | V_FALL | V_RISE | V_BEGIN | V_END | A_LARGE2 | A_SMALL2 | A_LARGE1 | A_SMALL1 | E_GO_DOWN | E_GO_UP		
(p1,(p2,p3,2,0),1,1)	CH	-	sanctions	impose | imposed | imposing | V_WIDEN	G/d/w+	
(p1,(p2,p3,2,0),1,1)	CH	+	sanctions	lift | lifting | lifted | remove	G/d/w+	
(p1,p2,2,1)	CH	+	currency N_CRISIS	V_EASE | V_LIMIT | V_END		
(p1,p2,2,1)	CH	-	currency N_CRISIS	V_WIDEN | V_PREVENT		
EXTERN,MONPOL(+,+)						
((p1,p2,1,0),p3,2,2)		0	currency | fx | verbal	intervention	N_CB | official	
(p1,(p2,p3,2,0),1,2)	CH	0	N_CB	intervene | intervenes | intervened	currency market | FX market	
((p1,p2,1,1),p3,1,1)	CH	+	N_CB	talk down | talked down | talks down | talking down | weaken | weakens | weakened | weakening	N_FX	
((p1,p2,1,0),p3,2,1)	CH	0	currency | FX | exchange rate	regime	V_CHANGE | N_CHANGE	
(p1,p2,2,1)	CH	-	N_FX | currency	V_REVALUE		
(p1,p2,2,1)	CH	+	N_FX | currency	V_DEVALUE		
MONPOL(+)						
((p1,p2,1,1),p3,2,1)		0	N_CB	V_SUSTAIN	rate | rates	
(p1,p2,2,1)	CH	-	policy | cycle | monetary | N_CB | N_BRATE	V_TIGHTEN | tight | tighter		
(p1,p2,2,1)	CH	+	policy | cycle | monetary | N_CB | N_BRATE	V_EASE | V_RELAX | looser | accommodative | loose | expansionary | accommodation		
(p1,p2,2,0)	CH	-	N_CB	V_TIGHTEN | V_RAISE		
(p1,p2,2,0)	CH	+	N_CB	V_EASE | V_CUT		
(p1,(p2,p3,1,0),2,1)	CH	-	N_CB | monetary	V_LIMIT | V_WITHDRAW	stimulus	
(p1,(p2,p3,1,0),2,1)	CH	+	N_CB | monetary	V_RAISE | V_STRENGTHEN	stimulus	
(p1,p2,2,1)	CH	-	N_REQRESERVES	V_LIMIT | V_CUT		
(p1,p2,2,1)	CH	+	N_REQRESERVES	V_TIGHTEN | V_RAISE		
(p1,p2,2,1)	CH	-	hawk | hawkish	N_COMMUNICATION | monetary | policy | N_CB		
(p1,p2,2,1)	CH	+	dove | dovish	N_COMMUNICATION | monetary | policy | N_CB		
(p1,(p2,p3,2,0),2,1)	CH	-	N_CB	print | printing | prints | create | creates | creating | creation	money	
POLINST,MONPOL(+,0)						
(p1,(p2,p3,2,1),1,1)	CH	-	G/d/w+	V_EXIT	G1_EZ | euro	
(p1,p2,2,0)	CH	-	V_BREAKUP | disintegration | dissolution	G1_EZ		
(p1,(p2,p3,2,1),1,1)	CH	+	G/d/w+	V_ENTER | N_ACCESSION	G1_EZ | euro	
((p1,p2,2,1),p3,1,1)	CH	+	G/d/w+	V_ADOPT	euro | N_EUR	
MONPOL,BANK(+,+)						
(p1,p2,2,1)	CH	+	N_SLS | N_QE | N_OMT | N_SMP | N_ELA | N_LTRO | N_TAF | N_TALF	V_WIDEN | V_AGREE | V_PLEDGE		
(p1,p2,2,1)	CH	-	N_SLS | N_QE | N_OMT | N_SMP | N_ELA | N_LTRO | N_TAF | N_TALF	V_LIMIT | V_DISAPPOINT		
(p1,(p2,p3,2,1),1,1)	CH	-	N_CB	V_LIMIT	collateral	
(p1,(p2,p3,2,1),1,1)	CH	+	N_CB	V_EASE | V_RELAX | V_WIDEN | wider | broader	collateral	
(p1,(p2,p3,2,1),1,1)	CH	+	N_CB	inject | injects | injecting | injection | provision | provide | providing | provides | pump | pumps | pumped	liquidity | cash	
(p1,(p2,p3,1,1),1,1)	CH	+	N_CB	V_STRENGTHEN | N_BAILOUT | V_SAVE	N_BANKS	
((p1,p2,1,0),(p3,p4,2,0),1,1)	CH	+	N_BANKS	V_RECEIVE	N_CB	liquidity | cash | N_AID | money
((p1,p2,1,1),(p3,p4,2,1),1,1)	CH	+	N_CB	V_AGREE | V_PLEDGE | V_UNLOCK | V_IMPLEMENT	N_BAILOUT | N_AID | V_SAVE | N_LIQUIDITY	N_BANKS
((p1,p2,1,1),(p3,p4,2,1),1,1)	CH	-	N_CB	V_REJECT | V_FAIL | V_BLOCK	N_BAILOUT | N_AID | V_SAVE | N_LIQUIDITY	N_BANKS
FISCAL,BANK(+,-)						
((p1,p2,1,0),p3,1,1)	CH	-	G/d/w+	V_STRENGTHEN | N_BAILOUT | V_SAVE | N_AID	N_BANKS	
(p1,(p2,p3,1,1),1,1)	CH	-	N_GOVT	V_STRENGTHEN | N_BAILOUT | V_SAVE | N_AID	N_BANKS	
((p1,(p2,p3,2,0),1,1),p4,1,2)	CH	-	N_GOVT	inject | injects | injecting | injection | provision | provide | providing | provides | pump | pumps | pumped	liquidity | cash | capital | guarantee | guarantees	N_BANKS
((p1,p2,1,0),(p3,p4,2,0),1,2)	CH	-	N_BANKS	V_RECEIVE	N_GOVT | taxpayer | taxpayers	liquidity | cash | capital | guarantee | guarantees | N_AID | money
((p1,p2,1,1),(p3,p4,2,1),1,2)	CH	-	N_GOVT	V_AGREE | V_PLEDGE | V_UNLOCK | V_IMPLEMENT	N_BAILOUT | N_AID | V_SAVE	N_BANKS
((p1,p2,1,1),(p3,p4,2,1),1,2)	CH	+	N_GOVT	V_REJECT | V_FAIL | V_BLOCK	N_BAILOUT | N_AID | V_SAVE	N_BANKS
(p1,p2,2,1)	CH	-	V_RECAPITAL | N_RECAPITAL	N_BANKS		
(p1,p2,2,1)	CH	-	bad bank	V_BEGIN | V_AGREE		
(p1,p2,2,1)	CH	-	deposits	guarantee | guarantees | guaranteeing | guaranteed		
FISCAL(+)						
((p1,p2,2,1),p3,2,1)	CH	-	N_GOVT	stimulus	V_LIMIT | V_WITHDRAW	
((p1,p2,2,1),p3,2,1)	CH	+	N_GOVT	stimulus	V_RAISE | V_STRENGTHEN	
(p1,p2,2,1)		-	N_GOVT | G/d/w+	debt trap | debt spiral		
(p1,p2,2,1)		-	N_GOVT	insolvent | insolvency		
(p1,p2,2,1)		+	N_GOVT	solvent		
((p1,p2,1,0),p3,2,1)		-	G/d/w+	N_DEBT	A_UNSUSTAIN	
((p1,p2,1,0),p3,2,1)		+	G/d/w+	N_DEBT	sustainable	
(p1,(p2,p3,2,1),2,1)		-	N_GOVT	N_DEBT	A_UNSUSTAIN	
(p1,(p2,p3,2,1),2,1)		+	N_GOVT	N_DEBT	sustainable	
(p1,p2,2,1)		-	fiscal | budget	A_UNSUSTAIN		
(p1,p2,2,1)		+	fiscal | budget	sustainable		
(p1,p2,2,1)	CH	+	austerity	V_PLEDGE | V_AGREE | V_IMPOSE | V_IMPLEMENT		
(p1,p2,2,1)	CH	-	austerity	V_LIMIT | V_REJECT | V_WITHDRAW		
((p1,p2,1,0),p3,2,1)	CH	+	fiscal | budget	rules | reforms | consolidation	V_AGREE | V_PLEDGE | V_IMPLEMENT | V_IMPOSE	
((p1,p2,1,0),p3,2,1)	CH	-	fiscal | budget	rules | reforms | consolidation	V_REJECT | V_LIMIT | V_BREACH	

Table C.5 (continued)
Fundamental expression structures

<table>
<tr><th>EXPR_CODE</th><th>MOD</th><th>SIGN</th><th>p1</th><th>p2</th><th>p3</th><th>p4</th></tr>
<tr><td colspan="7">FISCAL(+)</td></tr>
<tr><td>((p1,p2,1,0),p3,2,1)</td><td>CH</td><td>-</td><td>fiscal | budget | deficit</td><td>targets</td><td>V_MISS | V_BREACH</td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>+</td><td>fiscal | budget</td><td>policy</td><td>V_TIGHTEN | tight | tighter | strict | prudent | stringent</td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>-</td><td>fiscal | budget</td><td>policy</td><td>loose | looser | accommodative | expansionary | V_EASE | V_RELAX</td><td></td></tr>
<tr><td colspan="7">FUNDLIQ(+)</td></tr>
<tr><td>(p1,(p2,p3,1,0),2,1)</td><td></td><td>-</td><td>N_GOVT</td><td>funding | financing</td><td>need | requirements | requirements | needs</td><td></td></tr>
<tr><td>(p1,(p2,p3,1,0),2,0)</td><td></td><td>-</td><td>G/d/w+</td><td>funding | financing</td><td>need | requirements | requirements | needs</td><td></td></tr>
<tr><td>(p1,(p2,p3,1,0),2,1)</td><td>CH</td><td>-</td><td>N_GOVT</td><td>V_MISS | V_REJECT</td><td>payment | obligation | repayment | payments | obligations | repayments | repay | repaying</td><td></td></tr>
<tr><td>(p1,(p2,p3,2,1),2,1)</td><td></td><td>-</td><td>N_GOVT</td><td>N_DEBT</td><td>N_DEFAULT | V_DEFAULT</td><td></td></tr>
<tr><td>(p1,p2,2,0)</td><td></td><td>-</td><td>N_GOVT</td><td>N_DEFAULT | V_DEFAULT</td><td></td><td></td></tr>
<tr><td>(p1,(p2,p3,1,0),1,0)</td><td>CH</td><td>-</td><td>G/d/w+</td><td>V_MISS | V_REJECT</td><td>payment | obligation | repayment | payments | obligations | repayments | repay | repaying</td><td></td></tr>
<tr><td>(p1,(p2,p3,2,1),2,0)</td><td></td><td>-</td><td>G/d/w+</td><td>N_DEBT</td><td>N_DEFAULT | V_DEFAULT</td><td></td></tr>
<tr><td>(p1,(p2,p3,2,1),2,0)</td><td></td><td>-</td><td>G/d/w+</td><td>N_GOVT</td><td>N_DEFAULT | V_DEFAULT</td><td></td></tr>
<tr><td>((p1,(p2,p3,2,1),2,1),p4,2,1)</td><td></td><td>+</td><td>N_GOVT | G/d/w+</td><td>N_AUCTION</td><td>demand</td><td>A_AMPLE | A_STABLE</td></tr>
<tr><td>((p1,(p2,p3,2,1),2,1),p4,2,1)</td><td></td><td>-</td><td>N_GOVT | G/d/w+</td><td>N_AUCTION</td><td>demand</td><td>A_SCARCE | V_DISAPPEAR | A_INSTABLE</td></tr>
<tr><td>(p1,(p2,p3,2,0),2,1)</td><td></td><td>+</td><td>N_GOVT | G/d/w+</td><td>N_AUCTION</td><td>successful</td><td></td></tr>
<tr><td>(p1,(p2,p3,2,0),2,1)</td><td></td><td>-</td><td>N_GOVT | G/d/w+</td><td>N_AUCTION</td><td>N_FAILURE | V_FAIL | A_FAILED</td><td></td></tr>
<tr><td>(p1,((p2,p3,2,0),p4,2,1),1,1)</td><td>CH</td><td>-</td><td>N_GOVT | G/d/w+</td><td>access</td><td>market</td><td>lost | loses | losing</td></tr>
<tr><td>(p1,((p2,p3,2,0),p4,2,1),1,1)</td><td>CH</td><td>+</td><td>N_GOVT | G/d/w+</td><td>access</td><td>market</td><td>V_REGAIN</td></tr>
<tr><td>(p1,(p2,p3,1,0),1,1)</td><td>CH</td><td>+</td><td>N_GOVT | G/d/w+</td><td>returns to | returning to | returned to | return to</td><td>market</td><td></td></tr>
<tr><td>(p1,((p2,p3,2,0),p4,2,1),1,1)</td><td>CH</td><td>-</td><td>N_GOVT | G/d/w+</td><td>access</td><td>official | N_AID | N_BAILOUT | N_LENDING</td><td>lost | loses | losing</td></tr>
<tr><td>(p1,((p2,p3,2,0),p4,2,1),1,1)</td><td>CH</td><td>+</td><td>N_GOVT | G/d/w+</td><td>access</td><td>official | N_AID | N_BAILOUT | N_LENDING</td><td>V_REGAIN</td></tr>
<tr><td>(p1,(p2,p3,1,0),2,1)</td><td>CH</td><td>-</td><td>V_DEPLETE</td><td>treasury | N_GOVT</td><td>coffers | funds | reserves</td><td></td></tr>
<tr><td>(p1,(p2,p3,2,0),1,1)</td><td></td><td>+</td><td>N_GOVT</td><td>N_LIQUIDITY</td><td>A_AMPLE</td><td></td></tr>
<tr><td>(p1,(p2,p3,2,0),1,1)</td><td></td><td>-</td><td>N_GOVT</td><td>N_LIQUIDITY</td><td>A_SCARCE</td><td></td></tr>
<tr><td>((p1,p2,1,1),(p3,p4,2,0),1,1)</td><td>CH</td><td>+</td><td>N_ECB | N_INTLEND</td><td>V_AGREE | V_PLEDGE | V_UNLOCK | breakthrough | N_AGREEMENT</td><td>N_BAILOUT | N_LENDING | N_AID</td><td>G/d/w+</td></tr>
<tr><td>((p1,p2,1,1),(p3,p4,2,0),1,1)</td><td>CH</td><td>-</td><td>N_ECB | N_INTLEND</td><td>V_REJECT | V_FAIL | V_BREACH | V_BLOCK | delays | delayed | delaying | delay</td><td>N_BAILOUT | N_LENDING | N_AID</td><td>G/d/w+</td></tr>
<tr><td>((p1,p2,1,1),(p3,p4,2,1),1,1)</td><td></td><td>+</td><td>G/d/w+</td><td>V_RECEIVE | secure | secures | securing | secured | V_ACHIEVE</td><td>N_BAILOUT | N_LENDING | N_AID</td><td>N_ECB | N_INTLEND</td></tr>
<tr><td>((p1,p2,1,1),(p3,p4,2,1),1,1)</td><td>CH</td><td>-</td><td>G/d/w+</td><td>V_REQUEST | V_NEED</td><td>N_ECB | N_INTLEND</td><td>N_BAILOUT | N_LENDING | N_AID</td></tr>
<tr><td>BANK(+)</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td></td><td>-</td><td>funding markets | funding market | shadow banking | credit market | credit markets | funding liquidity | N_BANKS</td><td>N_CRISIS</td><td></td><td></td></tr>
<tr><td>((p1,p2,2,0),p3,2,1)</td><td>CH</td><td>+</td><td>funding markets | funding market | shadow banking | credit market | credit markets | funding liquidity | N_BANKS</td><td>N_CRISIS</td><td>V_EASE | V_LIMIT | V_END</td><td></td></tr>
<tr><td>((p1,p2,2,0),p3,2,1)</td><td>CH</td><td>-</td><td>funding markets | funding market | shadow banking | credit market | credit markets | funding liquidity | N_BANKS</td><td>N_CRISIS</td><td>V_WIDEN | V_PREVENT</td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>N_LIQCRUNCH</td><td></td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td></td><td>-</td><td>funding markets | funding market | shadow banking | credit market | credit markets | funding liquidity</td><td>freeze | freezes | locks up | lockup</td><td></td><td></td></tr>
<tr><td>((p1,p2,2,1),p3,2,1)</td><td>CH</td><td>-</td><td>N_BANKS</td><td>liquidity | funding</td><td>V_DISAPPEAR</td><td></td></tr>
<tr><td>((p1,p2,2,1),p3,2,1)</td><td></td><td>+</td><td>N_BANKS</td><td>A_AMPLE</td><td>N_LIQUIDITY</td><td></td></tr>
<tr><td>((p1,p2,2,1),p3,2,1)</td><td></td><td>-</td><td>N_BANKS</td><td>A_SCARCE</td><td>N_LIQUIDITY</td><td></td></tr>
<tr><td>((p1,p2,1,1),p3,1,1)</td><td>CH</td><td>-</td><td>N_BANKS</td><td>V_REQUEST | V_NEED</td><td>liquidity | cash | capital | guarantee | guarantees | N_AID | money</td><td></td></tr>
<tr><td>((p1,p2,2,1),p3,2,1)</td><td></td><td>+</td><td>N_INTLEND</td><td>N_BAILOUT | N_AID</td><td>N_BANKS</td><td></td></tr>
<tr><td>((p1,p2,2,1),p3,2,1)</td><td>CH</td><td>+</td><td>N_BANKS</td><td>V_RECEIVE</td><td>N_BAILOUT | N_AID</td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>+</td><td>N_BANKUNION</td><td>V_BEGIN | V_AGREE</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>+</td><td>N_MACROPRUD</td><td>V_IMPLEMENT</td><td></td><td></td></tr>
<tr><td colspan="7">POLINST(+)</td></tr>
<tr><td>(p1,p2,2,0)</td><td></td><td>-</td><td>market | price | prices | trade | investment</td><td>A_RIGID</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,0)</td><td></td><td>+</td><td>market | price | prices | trade | investment</td><td>N_FREEDOM</td><td></td><td></td></tr>
<tr><td>((p1,p2,2,0),p3,2,1)</td><td>CH</td><td>+</td><td>market | price | prices | trade | investment</td><td>N_CONTROLS</td><td>V_RELAX | V_LIMIT | V_END</td><td></td></tr>
<tr><td>((p1,p2,2,0),p3,2,1)</td><td>CH</td><td>-</td><td>market | price | prices | trade | investment</td><td>N_CONTROLS</td><td>V_WIDEN | V_STRENGTHEN</td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>+</td><td>market institutions</td><td>V_STRENGTHEN | V_PROTECT</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>-</td><td>market institutions</td><td>V_WEAKEN | V_LIMIT | V_FAIL</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>+</td><td>N_PROPRIGHTS</td><td>V_STRENGTHEN | V_PROTECT</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>-</td><td>N_PROPRIGHTS</td><td>V_WEAKEN | V_LIMIT</td><td></td><td></td></tr>
<tr><td>p1</td><td>CH</td><td>-</td><td>N_NATIONALIZE</td><td></td><td></td><td></td></tr>
<tr><td>p1</td><td>CH</td><td>+</td><td>N_PRIVATIZE</td><td></td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>-</td><td>N_GOVT</td><td>V_EXPROP</td><td></td><td></td></tr>
<tr><td>((p1,p2,1,0),p3,2,1)</td><td></td><td>-</td><td>N_LABORM</td><td>N_REGULATIONS | N_CONTROLS</td><td>A_RIGID</td><td></td></tr>
<tr><td>((p1,p2,1,0),p3,2,1)</td><td></td><td>+</td><td>N_LABORM</td><td>N_REGULATIONS | N_CONTROLS</td><td>flexible | V_RELAX | V_LIMIT</td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>+</td><td>N_RULELAW</td><td>V_STRENGTHEN | V_PROTECT</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>-</td><td>N_RULELAW</td><td>V_WEAKEN | V_LIMIT | V_FAIL</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>+</td><td>democratic institutions | democracy</td><td>V_STRENGTHEN | V_PROTECT</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>-</td><td>democratic institutions | democracy</td><td>V_WEAKEN | V_LIMIT | V_FAIL</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,0)</td><td></td><td>+</td><td>government | N_ELECT</td><td>transparent</td><td></td><td></td></tr>
<tr><td>p1</td><td></td><td>+</td><td>structural reform | structural reforms</td><td></td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>+</td><td>structural reform | structural reforms</td><td>V_AGREE | V_IMPLEMENT | PLEDGE | V_ACHIEVE</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>-</td><td>structural reform | structural reforms</td><td>V_FAIL | V_REJECT</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td></td><td>+</td><td>N_STRUCTURES</td><td>reform | reforms | reformed | reforming | overhaul | overhauled | overhauls</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td></td><td>-</td><td>N_STRUCTURES</td><td>reform | reforms | reformed | reforming | overhaul | overhauled | overhauls</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,0)</td><td>CH</td><td>+</td><td>N_ELECT</td><td>landslide</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,0)</td><td>CH</td><td>-</td><td>N_ELECT</td><td>inconclusive</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,0)</td><td>CH</td><td>-</td><td>government | coalition | ruling party | governments | coalitions</td><td>N_FAILURE | V_FAIL | A_FAILED</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,0)</td><td></td><td>-</td><td>political | government</td><td>N_CRISIS</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,0)</td><td></td><td>-</td><td>political</td><td>N_CONFLICT</td><td></td><td></td></tr>
<tr><td>((p1,p2,2,0),p3,2,1)</td><td>CH</td><td>+</td><td>political | government</td><td>N_CRISIS</td><td>V_EASE | V_LIMIT | V_END</td><td></td></tr>
<tr><td>((p1,p2,2,0),p3,2,1)</td><td>CH</td><td>-</td><td>political | government</td><td>N_CRISIS</td><td>V_WIDEN | V_PREVENT</td><td></td></tr>
<tr><td>((p1,p2,2,0),p3,2,1)</td><td>CH</td><td>+</td><td>political | government</td><td>N_CONFLICT</td><td>V_EASE | V_LIMIT | V_END</td><td></td></tr>
<tr><td>((p1,p2,2,0),p3,2,1)</td><td>CH</td><td>-</td><td>political | government</td><td>N_CONFLICT</td><td>V_WIDEN | V_PREVENT</td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>minority government</td><td></td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td></td><td>+</td><td>majority government | clear majority</td><td>N_ELECT</td><td></td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>fragmented coalition</td><td></td><td></td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>N_PROTEST</td><td></td><td></td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>N_COUP</td><td></td><td></td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>N_REBEL</td><td></td><td></td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>N_REVOL</td><td></td><td></td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>N_WAR</td><td></td><td></td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>N_ASSASS</td><td></td><td></td><td></td></tr>
<tr><td>p1</td><td></td><td>-</td><td>N_TERROR</td><td></td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>+</td><td>N_PEACE</td><td>V_AGREE | V_IMPLEMENT | PLEDGE</td><td></td><td></td></tr>
<tr><td>(p1,p2,2,1)</td><td>CH</td><td>-</td><td>N_PEACE</td><td>V_FAIL | V_BREACH</td><td></td><td></td></tr>
<tr><td>(p1,(p2,p3,2,0),2,1)</td><td>CH</td><td>-</td><td>G/d/w+</td><td>V_EXIT</td><td>G1_EU</td><td></td></tr>
<tr><td>(p1,(p2,p3,2,0),2,1)</td><td>CH</td><td>+</td><td>G/d/w+</td><td>V_ENTER | accession</td><td>G1_EU</td><td></td></tr>
</table>

Notes: EXPR CODE defines the proximity and ordering rules of expression elements (see Table C.4 notes). MOD CH: denotes whether expression belongs to CHANGE subindex. SIGN: denotes improvement(+) or deterioration (-). p1-p4 are the expression elements.

Table C.6
Fundamental expression structures with complex endings

EXPR_CODE	TYPE	p1	p2	p3	(-): PROB_DOWN[a] (+): PROB_UP[a]	(-): RISK_UP[a] (+): RISK_DOWN[a]	(-): CONCERN_UP[a] (+): CONCERN_DOWN[a]	(-): V_END (+): V_BEGIN	(-): A_INSTABLE (+): A_STABLE	(-): V_DECELERATE (+): V_ACCELERATE	(-): V_FALL[b] (+): V_RISE[b]	(-): V_CUT (+): V_RAISE	(-): V_WORSEN[c] (+): V_IMPROVE[c]	(-): V_DISAPPOINT (+): V_PLEASE	(-): V_WEAKEN (+): V_STRENGTHEN
REAL(+)															
(p1,p2,2,1)	DATA	N_GDP				+	+		+	+	+	+	+	+	+
(p1,p2,2,1)	DATA	N_HHI							+	+	+	+	+	+	+
(p1,p2,2,1)	DATA	N_CONS							+	+	+	+	+	+	+
(p1,p2,2,1)	DATA	N_BCONF									+	+	+	+	+
(p1,p2,2,1)	DATA	N_CCONF									+	+	+	+	+
(p1,p2,2,1)	DATA	N_PMI							+	+	+	+	+	+	+
(p1,p2,2,1)	DATA	N_INDU							+	+	+	+	+	+	+
(p1,p2,2,1)	DATA	N_MANUF							+	+	+	+	+	+	+
((p1,p2,1,0),p3,2,1)	DATA	car | auto | vehicle	sales | registrations						+	+	+	+	+	+	+
((p1,p2,1,0),p3,2,1)	DATA	durable goods	sales | orders						+	+	+	+	+	+	+
((p1,p2,1,0),p3,2,1)	DATA	manufacturing	N_NUMBER							+	+	+	+	+	+
(p1,p2,2,1)	DATA	N_CONSTR							+	+	+	+	+	+	+
((p1,p2,1,0),p3,2,1)	DATA	construction	N_NUMBER							+	+	+	+	+	+
(p1,p2,2,1)	DATA	productivity							+	+	+	+	+	+	+
(p1,p2,2,1)	DATA	N_EARN							+	+	+	+	+	+	+
(p1,p2,2,1)	DATA	N_UNEMP				+				-	-	-	+	+	
((p1,p2,1,0),p3,2,1)	DATA	job	cuts | losses			+				-	-	-		+	
(p1,p2,2,1)	DATA	N_EMPL							+	+	+		+	+	+
(p1,p2,2,1)	DATA	vacancy rate									-	-	+		
((p1,p2,1,0),p3,2,1)	DATA	N_HOUSE	constructions | building | sales | starts						+	+	+	+	+	+	+
((p1,p2,1,0),p3,2,1)	DATA	N_HOUSE	market | sector | markets | sectors | N_NUMBER			+	+		+	+	+	+	+	+	+
((p1,p2,1,0),p3,2,1)		N_HOUSE	bubble				+								
((p1,p2,1,0),p3,2,1)		N_HOUSE	bust | burst		-	+	+								
(p1,p2,2,1)	DATA	economy				+	+		+	+			+	+	+
((p1,p2,1,0),p3,2,1)		economic	upturn | expansion		+			+	+	+			+		+
(p1,p2,2,1)		recovery			+	+	+	+	+	+			+		+
((p1,p2,2,1),p3,2,1)		G1_CHIN	hard landing		-	+	+	-		-					
((p1,p2,1,0),p3,2,1)		economic	downturn | slowdown		-	+	+	-		-					
(p1,p2,2,1)		recession			-	+	+	-							
((p1,p2,1,0),p3,2,1)		economic	N_CRISIS		-	+	+	-							
EXTERN(+)															
(p1,p2,2,1)	DATA	N_EXPORTS				+	+		+	+	+	+	+	+	+
(p1,p2,2,1)	DATA	N_IMPORTS								-	-	-			
((p1,p2,2,0),p3,2,1)	DATA	N_ETRADE	surplus				+		+		+	+	+	+	+
((p1,p2,2,0),p3,2,1)	DATA	N_ETRADE	N_DEFICIT				+				-	-			
((p1,p2,2,1),p3,2,1)	DATA	N_ETRADE	N_BALANCE			+	+		+				+	+	+
(p1,p2,2,1)	DATA	remittances				+	+		+		+	+	+	+	+
(p1,p2,2,1)	DATA	N_RES				+	+		+		+	+	+	+	
(p1,p2,2,1)	DATA	N_FDI				+	+		+		+	+	+	+	
(p1,p2,2,1)	DATA	foreign assets | foreign investments									+	+			
(p1,p2,2,1)	DATA	N_EDEBT				+	+				-	-			
(p1,(p2,p3,2,1),2,1)		G/d/w+	sanctions		-	+	+	-							
(p1,(p2,p3,2,1),2,1)		N_RES	V_DEPLETE | A_SCARCE		-	+	+								
(p1,p2,2,2)		currency N_CRISIS			-	+	+	-							
(p1,p2,2,1)		sudden V_END stop | capital flight			-	+	+	-			-	-			-
(p1,p2,2,1)		dollarization			-	+	+	-		-	-				-
(p1,p2,2,1)	DATA	external N_BALANCE				+	+		+				+	+	+
(p1,p2,2,1)		competitiveness				+	+				+		+	+	+
(p1,p2,2,1)		terms of E_TRADE trade | termsoftrade				+	+						+		+
EXTERN,POLINST(0,+)															
(p1,p2,2,1)		protectionism			-	+	+	-			-	-			-
EXTERN,MONPOL(+,+)															
((p1,p2,1,1),p3,1,1)		N_FX | currency	V_REVALUE		-			-							
((p1,p2,1,1),p3,1,1)		N_FX | currency	V_DEVALUE		+			+							
MONPOL(+)															
((p1,p2,1,0),p3,2,1)	DATA	N_CB	rate | rates							-		-			
(p1,p2,2,1)	DATA	N_BRATE								-		-			
((p1,p2,1,0),p3,2,1)		rate | rates	V_TIGHTEN		-	+	+	-							
((p1,p2,1,0),p3,2,1)		rate | rates	V_EASE		+	-	+	+							
((p1,p2,1,0),p3,1,0),2,2)		N_CB	rate	N_INCREASE	-	+	+	-							
((p1,p2,1,0),p3,1,0),2,2)		N_CB	rate	N_DECREASE	+	-	+	+							
((p1,p2,2,0),p3,2,1)		N_BRATE	N_INCREASE		-	+	+	-							
((p1,p2,2,0),p3,2,1)		N_BRATE	N_DECREASE		+	-	+	+							
((p1,p2,2,1),p3,2,1)		policy | cycle | monetary | N_CB | N_BRATE	V_TIGHTEN | tight | tighter		-	+	+	-							
((p1,p2,2,1),p3,2,1)		policy | cycle | monetary | N_CB | N_BRATE	V_EASE | V_RELAX | looser | accommodative | loose | expansionary | accommodation		+	-	+	+							
((p1,p2,2,1),p3,2,1)		N_CB	V_TIGHTEN | V_RAISE		-	+	+	-							
((p1,p2,2,1),p3,2,1)		N_CB	V_EASE | V_CUT		+	-	+	+							
(((p1,p2,2,0),p3,2,1),p4,2,2)		N_CB | monetary	V_LIMIT | V_WITHDRAW	stimulus	-	+	+	-							
(((p1,p2,2,0),p3,2,1),p4,2,2)		N_CB | monetary	V_RAISE | V_STRENGTHEN	stimulus	+	-	+	+							
((p1,p2,2,1),p3,2,1)		N_REQRESERVES	V_LIMIT | V_CUT		-	+	+	-							
((p1,p2,2,1),p3,2,1)		N_REQRESERVES	V_TIGHTEN | V_RAISE		+	-	+	+							
POLINST,MONPOL(+,0)															
((p1,(p2,p3,2,1),1,1),p4,2,2)		G/d/w+	V_EXIT	G1_EZ | euro	-	+	+								
((p1,p2,2,0),p3,2,2)		V_BREAKUP | disintegration | dissolution	G1_EZ		-	+	+								
((p1,(p2,p3,2,1),1,1),p4,2,2)		G/d/w+	V_ENTER | N_ACCESSION	G1_EZ | euro	+	+	+								
((p1,(p2,p3,2,1),1,1),p4,2,2)		G/d/w+	V_ADOPT	euro | N_EUR	+	+	+								
(p1,p2,2,1)		N_GREXIT			-	+	+								
MONPOL,BANK(+,+)															
((p1,(p2,p3,2,0),1,1),p4,2,1)	DATA	N_CB	inject | injects | injecting | injection | provision | provide | providing | provides | pump | pumps | pumped	liquidity | cash	+	+	+	+		+		+			
(p1,p2,2,1)	DATA	N_SLS | N_QE | N_OMT | N_SMP | N_ELA | N_LTRO | N_TAF | N_TALF			+	+	+	+		+		+			
(p1,(p2,p3,2,1),2,1)		N_CB	collateral			+	+								
(((p1,p2,1,1),p3,1,1),p4,2,2)		N_CB	V_STRENGTHEN | N_BAILOUT | V_SAVE	N_BANKS	+	+	+								
(p1,p2,2,1)		lender of last resort | lenderoflastresort				+	+								
(p1,p2,2,1)		N_TAPER			-	+	+	-							

Table C.6 (continued)
Fundamental expression structures with complex endings

EXPR_CODE	TYPE	p1	p2	p3	(-): PROB_DOWN[a] (+): PROB_UP[a]	RISK_UP[a] RISK_DOWN[a]	CONCERN_UP[a] CONCERN_DOWN[a]	V_END V_BEGIN	A_INSTABLE A_STABLE	V_DECELERATE V_ACCELERATE	V_FALL[b] V_RISE[b]	V_CUT V_RAISE	V_WORSEN[c] V_IMPROVE[c]	V_DISAPPOINT V_PLEASE	V_WEAKEN V_STRENGTHEN
FISCAL(+)															
((p1,p2,2,0),p3,2,1)	DATA	N_GOVT	N_BALANCE | finances			+	+		+				+	+	+
((p1,p2,2,0),p3,2,1)	DATA	N_GOVT	surplus			+	+		+		+	+			
((p1,p2,2,0),p3,2,1)	DATA	N_GOVT | G/d/w+	N_DEFICIT			+	+				-	-		+	
((p1,p2,2,0),p3,2,1)	DATA	N_GOVT	N_REVENUE			+	+		+		+	+	+	+	+
(p1,p2,2,1)	DATA	tax | taxes | levy | levies							+		+	+			
((p1,p2,2,0),p3,2,1)	DATA	N_GOVT | N_DEFENSE | N_SOCIAL	N_SPENDING			+	+			-	-	-			
((p1,p2,2,0),p3,2,1)	DATA	N_GOVT	investment | investments | constructions | construction			+	+			-	-	-			
((p1,p2,1,0),p3,2,1)	DATA	N_GOVT	N_DEBT			+	+		+		-	-			
(p1,p2,2,1)	DATA	debttogdp | debt to N_GDP				+	+		+	-	-	-	+	+	
((p1,p2,1,1),p3,2,2)		N_GOVT | G/d/w+	debt trap | debt spiral		-	+	+								
((p1,p2,1,1),p3,2,2)		N_GOVT	insolvent		-	+	+								
((p1,p2,1,1),p3,2,2)		N_GOVT	solvent		+	+	+								
((p1,p2,1,1),p3,2,2)		N_GOVT	solvency | insolvency			+	+								
((p1,p2,2,1),p3,2,1)	DATA	N_GOVT	wage | wages								-	-			
(((p1,p2,1,0),p3,2,1),p4,2,2)		G/d/w+	N_DEBT	A_UNSUSTAIN	-	+	+								
(((p1,p2,1,0),p3,2,1),p4,2,2)		G/d/w+	N_DEBT	sustainable	+	+	+								
((p1,(p2,p3,2,1),2,1),p4,2,2)		N_GOVT	N_DEBT	A_UNSUSTAIN	-	+	+								
((p1,(p2,p3,2,1),2,1),p4,2,2)		N_GOVT	N_DEBT	sustainable	+	+	+								
((p1,(p2,p3,2,1),2,1),p4,2,2)		fiscal | budget	sustainable | stability | instability | A_UNSUSTAIN			+	+								
(p1,p2,2,2)		austerity			+	+	+	+				+			
((p1,p2,1,0),p3,2,2)		fiscal | budget	rules | reforms | consolidation		+	+	+					+			
((p1,p2,1,0),p3,2,2)		fiscal | budget | deficit	targets		+	+	+					+			
FISCAL,BANK(+,-)															
(((p1,p2,1,0),p3,1,1),p4,2,2)		G/d/w+	V_STRENGTHEN | N_BAILOUT | V_SAVE | N_AID	N_BANKS	-		-								
(((p1,p2,1,1),p3,1,1),p4,2,2)		N_GOVT	V_STRENGTHEN | N_BAILOUT | V_SAVE | N_AID	N_BANKS	-		-								
((p1,p2,2,1),p3,2,2)		V_RECAPITAL | N_RECAPITAL	N_BANKS		-		-								
((p1,p2,2,1),p3,2,2)		bad bank	V_BEGIN | V_AGREE		-		-								
((p1,p2,2,1),p3,2,2)		deposits	guarantee | guarantees | guaranteeing | guaranteed		-		-	-							
FUNDLIQ(+)															
((p1,(p2,p3,2,0),2,2),p4,2,2)		N_GOVT	N_DEBT	A_MATURING		+	+				-	-			
((p1,p2,2,2),p3,2,2)		N_GOVT	maturity | redemption | maturities | redemptions			+	+			-	-				
((p1,(p2,p3,2,0),2,2),p4,2,2)		N_GOVT	maturity | redemption | redemptions | interest payments	schedule | profile | structure		+	+						+		+
((p1,(p2,p3,2,0),2,0),p4,2,2)		G/d/w+	N_DEBT	A_MATURING		+	+				-	-			
((p1,p2,2,1),p3,2,2)		G/d/w+	maturity | redemption | maturities | redemptions			+	+			-	-				
((p1,(p2,p3,2,0),2,1),p4,2,2)		G/d/w+	maturity | redemption | redemptions | interest payments	schedule | profile | structure		+	+						+		+
((p1,(p2,p3,1,0),2,2),p4,2,2)		N_GOVT	funding | financing	need | requirements | requirements | needs		+	+			-	-	-			
((p1,(p2,p3,1,0),2,0),p4,2,2)		G/d/w+	funding | financing	need | requirements | requirements | needs		+	+			-	-	-			
((p1,(p2,p3,1,0),2,1),p4,2,2)		N_GOVT	V_MISS | V_REJECT	payment | obligation | repayment | payments | obligations | repayments | repay | repaying	-	+	+								
((p1,(p2,p3,1,0),2,1),p4,2,2)		N_GOVT	honor | honour | honoring | honouring	debt | payment | obligation | repayment | payments | obligations | repayments | repay | repaying		+	+								
((p1,(p2,p3,2,1),2,1),p4,2,2)		N_GOVT	N_DEBT	N_DEFAULT | V_DEFAULT	-	+	+	-							
((p1,p2,2,0),p3,2,1)		N_GOVT	N_DEFAULT | V_DEFAULT		-	+	+	-							
((p1,(p2,p3,1,0),1,1),p4,2,2)		G/d/w+	V_MISS | V_REJECT	payment | obligation | repayment | payments | obligations | repayments | repay | repaying	-	+	+								
((p1,(p2,p3,1,0),1,1),p4,2,2)		G/d/w+	honor | honour | honoring | honouring	debt | payment | obligation | repayment | payments | obligations | repayments | repay | repaying		+	+								
((p1,(p2,p3,2,1),2,0),p4,2,1)		G/d/w+	N_DEBT	N_DEFAULT | V_DEFAULT	-	+	+	-							
((p1,(p2,p3,2,1),2,0),p4,2,2)		G/d/w+	N_GOVT	N_DEFAULT | V_DEFAULT	-	+	+	-							
((p1,(p2,p3,2,1),2,1),p4,2,1)	DATA	N_GOVT | G/d/w+	N_AUCTION	demand | bid to cover | bidtocover | turnout		+	+		+	+	+	+	+	+	+
((p1,p2,2,0),2,1)		N_GOVT | G/d/w+	N_AUCTION			+	+								
((p1,(p2,p3,2,0),1,1),p4,2,1)		N_GOVT	access	market	+	+	+		+				+		+
((p1,(p2,p3,1,0),1,1),p4,2,2)		N_GOVT | G/d/w+	returns to | returning to | returned to | return to	market	+	+	+								
((p1,(p2,p3,2,0),1,1),p4,2,1)		N_GOVT | G/d/w+	access	official | N_AID | N_BAILOUT | N_LENDING	+	+	+		+				+		+
(p1,(p2,p3,2,0),1,1)		N_GOVT	N_LIQUIDITY			+	+		+		+		+		
((p1,(p2,p3,2,0),2,1),p4,2,2)		N_ECB | N_INTLEND	N_BAILOUT | N_LENDING | N_AID	G/d/w+	+	+	+	+			+				
(((p1,p2,1,0),p3,1,1),p4,2,2)		G/d/w+	V_RECEIVE | secure | secures | securing | secured | V_ACHIEVE	N_BAILOUT | N_LENDING | N_AID	+	+	+								
(((p1,p2,1,0),p3,1,1),p4,2,2)		G/d/w+	V_REQUEST | V_NEED	N_BAILOUT | N_LENDING | N_AID	-	+	+								
(((p1,p2,1,1),p3,1,1),p4,2,2)		N_GOVT	V_RECEIVE | secure | secures | securing | secured | V_ACHIEVE	N_BAILOUT | N_LENDING | N_AID	+	+	+								
(((p1,p2,1,1),p3,1,1),p4,2,2)		N_GOVT	V_REQUEST | V_NEED	N_BAILOUT | N_LENDING | N_AID	-	+	+								
((p1,p2,2,0),p3,2,1)		G1_GREE | G1_PORT | G1_IREL	N_BAILOUT | N_LENDING | N_AID		-	+	+								
BANK(+)															
(p1,(p2,p3,2,1),2,1)	DATA	N_BANKS	N_CAPADEQ			+	+		+				+		+
(p1,(p2,p3,2,1),2,1)	DATA	N_BANKS	N_NPL | arrears | delinquent | delinquencies | delinquency			+	+				-	-			
(p1,(p2,p3,2,1),2,1)	DATA	N_BANKS	N_CAPITAL | N_PROFITS			+	+		+		+	+	+	+	+
(p1,(p2,p3,2,1),2,1)	DATA	N_BANKS	writedowns			+	+			-	-	-			
((p1,p2,1,1),p3,2,1)		N_BANKS	bankrupt | bankruptcy			+	+								
((p1,p2,2,1),p3,2,1)		funding markets | funding market | shadow banking | credit market | credit markets | funding liquidity | N_BANKS	N_CRISIS		-	+	+				-				
((p1,p2,2,0),p3,2,1)		N_BANKS	run on		-	+	+	+			-				
(p1,p2,2,1)		bankrun			-	+	+	+			-				
((p1,p2,2,1),p3,2,1)		N_BANKS	N_TOXIC			+	+				-				
(p1,p2,2,1)		N_LIQCRUNCH			-	+	+				-				
((p1,p2,2,1),p3,2,1)		funding markets | funding market | shadow banking | credit market | credit markets | funding liquidity	freeze | freezes | locks up | lockup		-	+	+				-				
((p1,(p2,p3,2,0),1,1),2,1)		N_BANKS	access	N_LIQUIDITY		+	+				+		+		+
((p1,p2,2,1),p3,2,1)		N_BANKS	N_PORTFOLIO			+	+						+		+
((p1,p2,2,1),p3,2,1)		N_BANKS	liquidity | funding			+	+						+		+
((p1,p2,2,1),p3,2,1)		N_BANKS	N_BAILOUT | N_AID		-	+	+								

Table C.6 (continued)

Fundamental expression structures with complex endings

					(-):	PROB_DOWN[a]	RISK_UP[a]	CONCERN_UP[a]	V_END	A_INSTABLE	V_DECELERATE	V_FALL[b]	V_CUT	V_WORSEN[c]	V_DISAPPOINT	V_WEAKEN
EXPR_CODE	TYPE	p1	p2	p3	(+):	PROB_UP[a]	RISK_DOWN[a]	CONCERN_DOWN[a]	V_BEGIN	A_STABLE	V_ACCELERATE	V_RISE[b]	V_RAISE	V_IMPROVE[c]	V_PLEASE	V_STRENGTHEN
FUNDLIQ(+)																
(p1,p2,2,1)		N_BANKUNION						+								
(p1,p2,2,1)		N_MACROPRUD					+	+								
(p1,p2,2,1)		financial stability					+	+						+		+
(p1,(p2,p3,2,1),2,2)		N_BANKS	stability				+	+						+		+
(p1,(p2,p3,2,1),2,2)		N_BANKS	N_INSTABILITY				+	+				-	-			
POLINST(+)																
((p1,p2,1,0),p3,2,1)		market | price | prices | trade | investment	N_RIGIDITY					+				-				
((p1,p2,1,0),p3,2,1)		N_LABORM	N_RIGIDITY					+				-				
((p1,p2,1,0),p3,2,1)		market | price | prices | trade | investment	flexibility					+				+		+		
((p1,p2,1,0),p3,2,1)		N_LABORM	flexibility					+				+		+		
(p1,p2,2,1)		market institutions						+								
(p1,p2,2,1)		market institutions						+								
(p1,p2,2,1)		N_PROPRIGHTS						+								
(p1,p2,2,1)		N_PROPRIGHTS						+								
p1		N_NATIONALIZE						+								
p1		N_PRIVATIZE						+								
((p1,p2,1,0),p3,2,1)		N_GOVT	V_EXPROP					+								
(p1,p2,2,1)		democratic institutions | democracy						+								
(p1,p2,2,1)		democratic institutions | democracy						+								
(p1,p2,2,1)		N_RULELAW						+								+
(p1,p2,2,1)		N_RULELAW						+								+
((p1,p2,2,1),p3,2,2)		government | N_ELECT	transparency				+	+				+		+		
(p1,p2,2,1)		N_ELECT					+	+								
(p1,p2,2,1)		N_POPULISM						+				-				-
(p1,p2,2,1)		N_CORRUPT						+				-				-
(p1,p2,2,2)		structural reform | structural reforms				+	+	+	+							
((p1,p2,1,0),p3,2,2)		N_STRUCTURES	reform | reforms | reformed | reforming | overhaul | overhauled | overhauls					+	+							
((p1,p2,2,0),p3,2,1)		government | coalition | ruling party | governments | coalitions	N_FAILURE | A_FAILED			-	+	+								
((p1,p2,2,0),p3,2,1)		government | coalition | ruling party | governments | coalitions	N_INSTABILITY				+	+	-			-	-			
((p1,p2,2,0),p3,2,1)		government | coalition | ruling party | governments | coalitions	N_STABILITY				+	+				+	+	+		+
((p1,p2,2,0),p3,2,1)		political | government	N_CRISIS			-	+	+	-			-	-			
((p1,p2,2,0),p3,2,1)		political	N_CONFLICT			-	+	+	-			-	-			-
(p1,p2,2,1)		N_CONFVOTE				-	+	+								
(p1,p2,2,1)		V_IMPEACH				-	+	+								
(p1,p2,2,1)		N_PROTEST					+	+	+							
(p1,p2,2,1)		N_COUP				-	+	+	+							
(p1,p2,2,1)		N_REBEL					+	+	+							
(p1,p2,2,1)		N_REVOL					+	+	+							
(p1,p2,2,1)		N_WAR				-	+	+	+							
(p1,p2,2,1)		N_PEACE				+	+	+	+							
(p1,p2,2,1)		N_ASSASS					+	+	+							
(p1,p2,2,1)		N_TERROR					+	+	+							
((p1,(p2,p3,2,0),2,0),p4,2,2)		G/d/w+	V_EXIT	G1_EU		-	+	+								
((p1,(p2,p3,2,0),2,0),p4,2,2)		G/d/w+	V_ENTER | accession	G1_EU		+	+	+								
(p1,p2,2,1)		N_BREXIT				-	+	+	-							
((p1,p2,1,0),p3,2,2)		hard	N_BREXIT			-	+	+								
FISCAL,POLINST(+,-)																
(p1,p2,2,2)		N_FISCALCLIFF				+	-	-								
(p1,p2,2,2)		N_DEBTCEIL				+	-	-								
((p1,p2,2,1),p3,2,2)		N_AUTOCUTS	N_FISCALCLIFF			+	-	-								

Notes: EXPR CODE defines the proximity and ordering rules of expression elements (see Table C.4 notes). TYPE DATA: expressions that represent published data. p1-p3 are expression elements. The last 11 columns denote whether the column header verbs/adjectives can be used as last elements in the expression (see also following footnotes for similar words) a (+) sign in cells means that the top column header word denotes deterioration (e.g PROB_DOWN, RISK_UP) and second word (PROB_UP, RISK_DOWN, etc) are improvement; a (-) sign means reverse signs (e.g PROB_DOWN is improvement); blank cells mean the given word is not applicable in the expression.

a) PROB_DOWN: E_PROB_LOW| E_PROB_FALL; RISK_UP: E_RISK_HIGH| E_RISK_RISE*; CONCERN_UP: N_CONCERN| N_TROUBLE| N_STRAIN| E_CONCERN_CONT| E_CONCERN_HIGH |E_CONCERN_RISE* |E_HOPE_LOW |E_HOPE_FALL*.*

*b) same signs as V_FALL for V_TRAIL| A_SMALL2| A_SMALL1| A_LOWER| E_DECREASE2| E_DECREASE0 and only in case of where TYPE='DATA' E_LOWER0**| E_LOWER1**| E_LOWER2**| E_PRED_FALL1***| E_PRED_HIGH***| E_PRED_HIGH2***; same signs as V_RISE for V_SURPASS| A_LARGE2| A_LARGE1| A_HIGHER| E_INCREASE2| E_INCREASE0 and only in case of where TYPE='DATA' E_HIGHER0**| E_HIGHER1**| E_HIGHER2**| E_PRED_RISE1***| E_PRED_LOW***| E_PRED_LOW2***.*

c) same signs as V_WORSEN for A_BAD2| A_BAD1| A_BAD0| A_WORSE| N_DETERIORATION| E_DETERIORATION2| E_DETERIORATION0| E_WORSE1| E_WORSE2 and only in case of where TYPE='DATA' E_PRED_WORSEN| E_PRED_GOOD| E_PRED_GOOD2 same signs as V_IMPROVE for A_GOOD2| A_GOOD1| A_GOOD0| A_BETTER| N_IMPROVEMENT| E_IMPROVEMENT2| E_IMPROVEMENT0| E_BETTER1| E_BETTER2 and only in case of where TYPE='DATA' E_PRED_IMPROVE| E_PRED_BAD| E_PRED_BAD2.

**) having these synonym keys also assigns to subindex of change ('CH')*

***) having these synonym keys also assigns to subindex of surprise ('SURP')*

****) having these synonym keys also assigns to subindex of expectation ('EXP')*

Table C.7
Number of fundamental expressions by category (000s)

	sign	REAL	EXTERN	FISCAL	FUND_LIQ	BANK	POL_INST	MON_POL
ALL	positive	570.920	120.533	202.081	15.052	94.804	111.817	236.119
	neutral	9.512	4.012	0.231	0.045	1.126	0.001	43.624
	negative	574.670	111.189	203.257	57.792	113.048	639.401	141.549
CHANGE	positive	347.594	80.280	137.928	5.697	68.879	67.168	208.545
	neutral	0.507	2.740	0.073	0.001	0.044	0.000	11.126
	negative	289.029	69.824	124.722	4.942	33.852	84.907	109.511
EXPECT	positive	10.152	1.423	1.258	0.026	0.506	0.034	0.707
	neutral	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	negative	8.183	1.299	1.181	0.031	0.586	0.019	0.589
SURPRISE	positive	21.626	1.853	0.988	0.009	0.464	0.000	0.085
	neutral	1.426	0.193	0.028	0.000	0.001	0.000	0.000
	negative	13.746	1.487	0.885	0.002	0.131	0.000	0.059
CONCERN	positive	36.361	3.135	9.937	4.419	8.639	18.082	17.686
	neutral	0.000	0.000	0.000	0.000	0.000	0.000	2.301
	negative	103.020	11.808	22.328	13.583	25.688	52.762	30.480

Notes: The category ALL comprises total matches; CHANGE: fundamental expressions that refer to changing state of fundamentals; EXPECT: fundamental expressions that refer to expected/predicted state of fundamentals; SURPRISE: fundamental expressions that refer to the state of fundamentals related to previous expectations; CONCERN: fundamental expressions that refer to concerns, threats, risks, hopes, probabilities regarding the future state of fundamentals;

MNB Working Papers 2018/1

News-Based Indices on Country Fundamentals: Do They Help Explain Sovereign Credit Spread Fluctuations?

Budapest, February 2018

mnb.hu

1054 BUDAPEST, SZABADSÁG TÉR 9.