Summer School Volos
July, 2017

Textual analysis in economics and finance
(a descriptive approach)

Pragidis Ioannis
As. Professor
Department of Economics, DUTH

DEMOCRITUS UNIVERSITY OF THRACE

PROFIT

A blindfolded monkey throwing darts at The Wall Street Journal could

select a portfolio that would do as well as the (stock market) experts'.

[Malkiel (2003) "The efficient market hypothesis and its critics"
Journal of Economic Perspectives 17(1): 59–82)]

**However, what if this monkey could browse the Internet
before throwing any darts?**

- online journals
- online newspapers
- dedicated blogs
- social networks
- earning conference calls

In two words "Big Data"

Thus, the Internet has changed the way in which information is delivered to investors and how investors can act on it

There is one more think:
This voluminous information cannot be processed and summarized (at least timely) by any person.

What is textual analysis all about?

Quantify the tone and content of descriptive information and learn how the market interprets such information.

In simple words this means to examine the tone and sentiment of newspaper articles, press releases, investor message boards and corporate reports.

Textual analysis is widely applied in other fields such as political science, linguistics and marketing while has been historically less used as a technique in economics

The notion of parsing text for patterns has a long history:

In the 1300's, friars of the Dominican order produced concordances of the Latin Vulgate (Biblical translations) to provide indexes of common phrases (Catholic Encyclopedia, Vol. 4, 1908)

In 1901, T.C. Mendenhall used textual analysis to examine whether some works attributed to Shakespeare might have been written by Bacon (see Williams [1975])

During the world wars, the method was increasingly adapted to political speech, where carefully scripted rhetorical choices were interpreted as signals of diplomatic trends (e.g., Burke [1939])

Some interesting questions (Loughran and McDonald, 2016 JAR) :

Can we tease out sentiment from mandated company disclosures and contextualize quantitative data in ways that might predict future valuation components?

Can we computationally read news articles and trade before humans can read and assimilate the information?

If Twitter's tweets provide the pulse of information, can we monitor these messages in real time to gain an informational edge?

Do textual artifacts provide an additional attribute that predicts bankruptcies?

Are there subtle cues in managements' earnings conference calls that computers can discern better than analysts?

Textual analysis in practice

Textual analysis is a vast topic. Myriads of methods have been proposed and still evolves.

For sake of simplicity you only need to know this:

A sample of words or phrases are in advance classified as positive or negative (or to many other categories you think that makes sense).

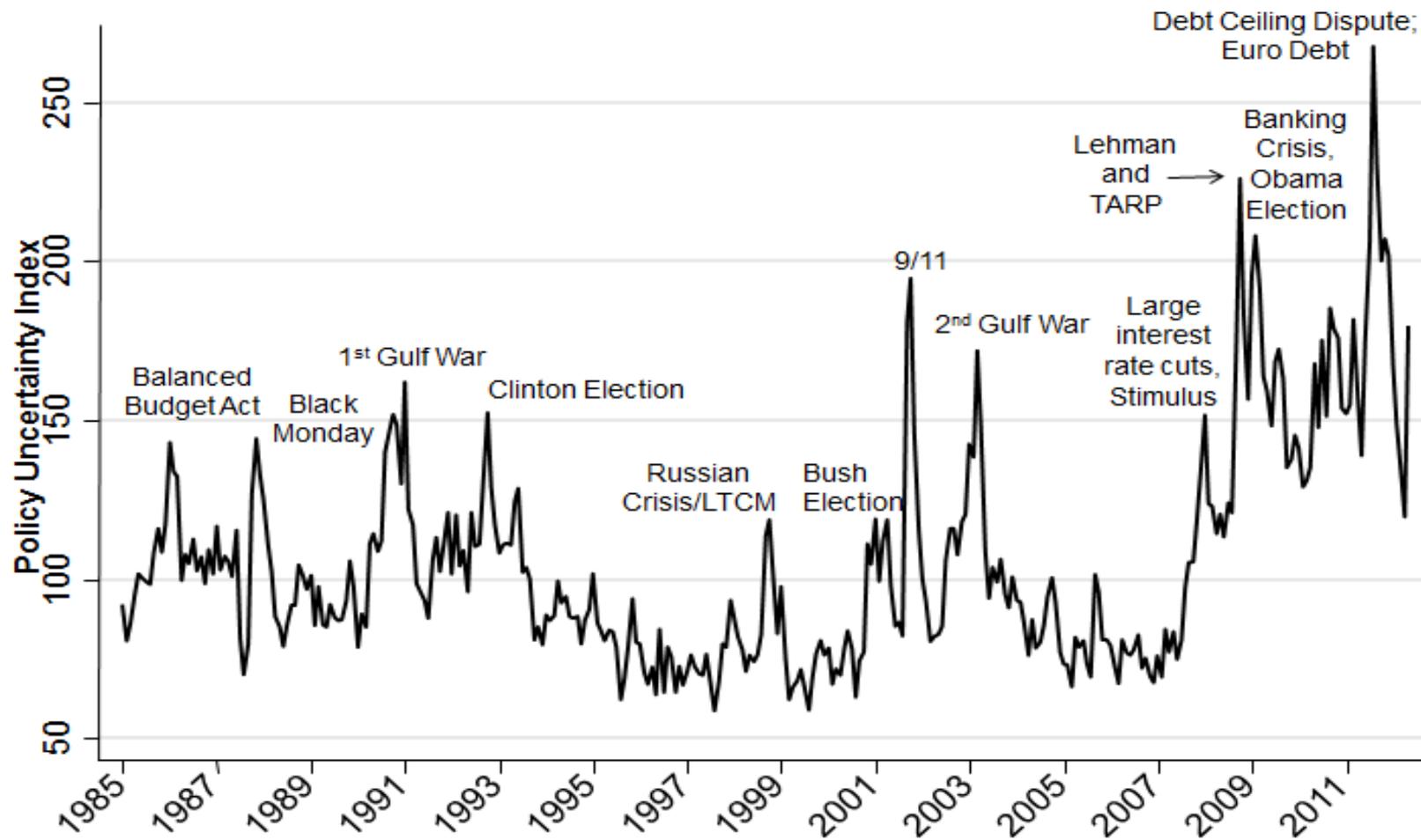These classifications are then used to computationally/mechanically classify texts.

An example

Crude drops on waning Middle East violence

Crude gains on hopes for U.S. budget resolution; falling stockpiles

A more intuitive example (Baker, Bloom and Davis, 2016 QJE)

KUWAIT CITY (AFP)--A drop in Iranian production coupled with regional tensions were pushing oil prices higher, Kuwaiti Oil Minister Hani Hussein said in remarks published Sunday. "Iranian production has dropped which has contributed to raising prices," Hussein was quoted as saying by Al-Watan newspaper. "Fears from regional tensions" and economic issues have also pushed prices higher, he added. Global oil prices rebounded sharply Friday after better-than-expected jobs data in the United States and ongoing tensions over key producer Iran. New York's main contract, West Texas Intermediate light sweet crude for September, jumped $4.27 to $91.40 a barrel. Brent North Sea crude for delivery in September soared $3.04 to $108.94 a barrel in London deals. Hussein said that despite geopolitical tensions, "oil supplies are going well and there is enough production to meet market demand which is a positive signal to the market. Iranian oil production has dropped sharply following European and U.S. sanctions on the Islamic republic over its nuclear program", according to the Organization of Petroleum Exporting Countries.

KUWAIT CITY (AFP)--A *drop* in Iranian *production* coupled with regional *tensions* were pushing oil prices higher, Kuwaiti Oil Minister Hani Hussein said in remarks published Sunday. "Iranian *production* has *dropped* which has contributed to raising prices," Hussein was quoted as saying by Al-Watan newspaper. "Fears from regional *tensions*" and economic issues have also pushed prices higher, he added. Global oil prices rebounded sharply Friday after better-than-expected jobs data in the United States and ongoing *tensions* over key producer Iran. New York's main contract, West Texas Intermediate light sweet crude for September, jumped $4.27 to $91.40 a barrel. Brent North Sea crude for delivery in September soared $3.04 to $108.94 a barrel in London deals. Hussein said that despite geopolitical *tensions*, "oil supplies are going well and there is enough production to meet market demand which is a positive signal to the market. Iranian oil *production* has *dropped* sharply following European and U.S. sanctions on the Islamic republic over its nuclear program", according to the Organization of Petroleum Exporting Countries. U.S.

Number of positive instances: 7
Number of negative instances: 0

List of 123 keywords created from the Dow Jones Energy Service news database that should be expected to affect oil prices.

61 Keywords that should be expected to increase oil Prices (1)		18 Keywords that should be expected to decrease oil prices (2)	45 Keywords that need to be signed by a modifier (3)	
Attack	Outage	Discovered	Allotment	Reserves
Attacker	Outages	Discoveries	Allotments	Rig
Attackers	Problem	Discovery	Buying	Rigs
Attacks	Problematic	Glut	Capacities	Stock
Bomb	Problematicly	Gluts	Cargo	Stockpile

Why is textual analysis important in economics and finance?

- Sentiment indexes seem to have predictive power for financial and economic variables

- An analyst may focus on more specific events (supply driven or demand driven oil prices movements)

- Specific behavioral patterns may be tested (Do investors quickly and rationally react to the content of oil-related news articles revealing supply and demand information?)

Phrases like *output cut, production cut, hurricane,* and *demand up* in lagged news articles are associated with lower oil prices the following trading day. This may be consistent with the notion that oil traders overreact to the content of widely-read news articles.

One final example

European Central Bank's (ECB) primary goal is price stability in the Eurozone

Why?

To protect households' purchasing power, to decrease uncertainty for investors and promote confidence (https://www.ecb.europa.eu/ecb/educational/pricestab/html/index.en.html)

How?

By changing the interest rate

However....

"It takes months for rate cuts to translate into economic growth............ what an ECB rate cut – or a series of cuts – could change quickly is investor psychology" [*WSJ September, 17 2007*]

Thus, after each meeting of BoG an announcement from ECB is released describing the economic outlook of the Eurozone.
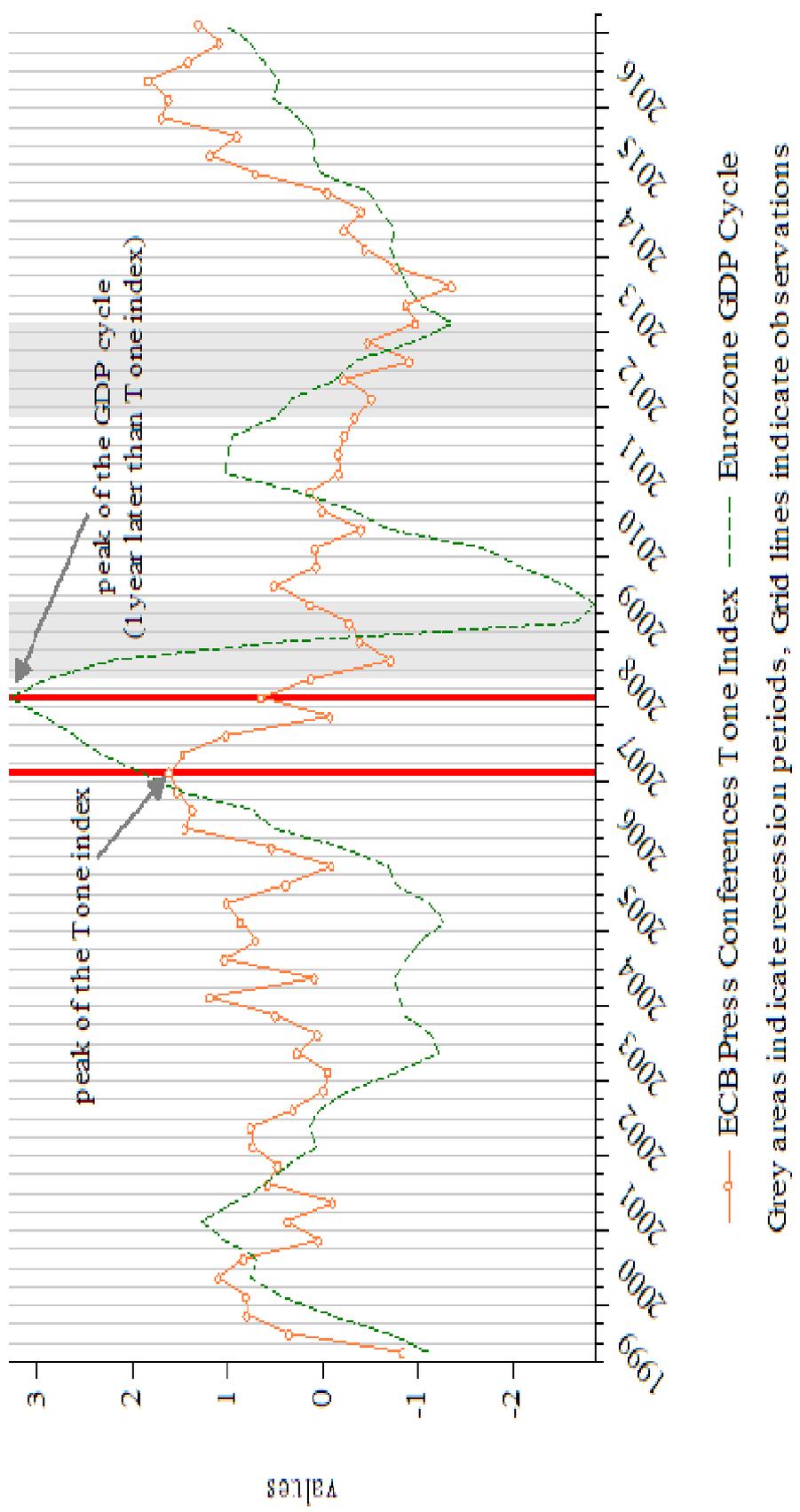
These announcements contain significant information. Textual analysis helps to visualize these information in few seconds

ECB press conference announcement 08/09/2016

Based on our regular economic and monetary analyses, we decided to keep the key ECB interest rates unchanged. We continue to expect them to remain at present or lower levels for an extended period of time. Today, we assessed the economic and monetary data which had become available since our last meeting and discussed the new ECB staff macroeconomic projections. Overall, while the available evidence so far suggests resilience of the euro area economy to the continuing global economic and political uncertainty, our baseline scenario remains subject to downside risks. Our comprehensive policy measures continue to ensure supportive financing conditions and underpin the momentum of the euro area economic recovery.

A sample of concept words

Concept words		
ADMINISTRED PRICES	EMPLOYMENT	INFLATION PROJECTIONS
AGGREGATE DEMAND	ENERGY PRICE INFLATION	INFLATION RATES
BORROWING GROWTH	ENERGY PRICES	INFLATIONARY PRESSURE
BROAD MONEY	EQUITY PRICES	INFLATIONARY PRESSURES
CAPACITY UTILISATION	EURO AREA DEMAND	INVESTMENT GROWTH
CAPITAL	EURO AREA ECONOMY	LABOUR MARKET CONDITIONS
COMMODITY PRICES	EURO AREA EXPORTS	LABOUR MARKET PROSPECTS
CONFIDENCE	EURO AREA GROWTH	LIQUIDITY
CONSUMER PRICE INFLATION	EXTERNAL DEMAND	LOAN
CONSUMER PRICES	FINANCING CONDITIONS	LOANS
CONSUMPTION	FOOD PRICES	MACRO. ENV.
COST EXTERNAL FINANCING	FOREIGN DEMAND	MONETARY DYNAMICS
CREDIT DEMAND	FOREIGN TRADE	MONETARY EXPANSION

peak of the GDP cycle
(1 year later than Tone index)

peak of the Tone index

— ECB Press Conferences Tone Index   — — Eurozone GDP Cycle

Grey areas indicate recession periods. Grid lines indicate observations

Values

1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016

Thank you!