

Lecture 2b: LDA Applications

EPFL Summer School in Computational Methods for Economists

Stephen Hansen
University of Oxford / Imperial College London

Procedural Transparency

How transparent should a public organization be?

Benefit of transparency: accountability.

Costs of transparency:

1. Direct costs
2. Privacy
3. Security
4. Worse behavior → “chilling effect”

Transparency and Monetary Policy

Mario Draghi (2013): “It would be wise to have a richer communication about the rationale behind the decisions that the governing council takes.”

Table: Disclosure Policies as of 2014

	Fed	BoE	ECB
Minutes?	✓	✓	X
Transcripts?	✓	X	X

Natural Experiment

FOMC meetings were recorded and transcribed from at least the mid-1970's in order to assist with the preparation of the minutes.

Committee members unaware that transcripts were stored prior to October 1993.

Greenspan then acknowledged the transcripts' existence to the Senate Banking Committee, and the Fed agreed:

1. To begin publishing them with a five-year lag.
2. To publish the back data.

"All the News
That's Fit to Print"

The New York Times

VOL. CLXIII . . . No. 56,420

© 2014 The New York Times

SATURDAY, FEBRUARY 22, 2014

Fed Misread Fiscal Crisis, Records Show

***After Caution in 2008,
Series of Bold Steps***

By BINYAMIN APPELBAUM

WASHINGTON — On the morning after Lehman Brothers filed for bankruptcy in 2008, most Federal Reserve officials still believed that the American economy would keep growing despite the metastasizing financial crisis.

The Fed's policy-making committee voted unanimously against bolstering the economy by cutting interest rates, and several officials praised what they described as the decision to let Lehman fail, saying it would help to restore a sense of accountability on Wall Street.

James Bullard, president of the Federal Reserve Bank of St. Louis, urged his colleagues "to wait for some time to assess the impact of the Lehman bankruptcy filing, if any, on the national econ-

DETROIT OUTLINES MAP TO SOLVENCY, STRESSING REPAIR

WAY OUT OF BANKRUPTCY

**Balancing Act Worries
Banks and Angers
Retirees in City**

By MONICA DAVEY
and MARY WILLIAMS WALSH

DETROIT — Seven months after this city entered bankruptcy, its leaders on Friday presented a federal judge with the first official road map to Detroit's future — documents designed to show how it aims to settle its \$18 billion debt to creditors and make itself livable again.

But the proposal is less a vision for a brand-new city than a repair estimate for the old one. It is a document designed by lawyers and bankruptcy experts to find

Deal Signed in Ukraine, but Shows St



Greenspan's View on Transparency

“A considerable amount of free discussion and probing questioning by the participants of each other and of key FOMC staff members takes place. In the wide-ranging debate, new ideas are often tested, many of which are rejected ... **The prevailing views of many participants change as evidence and insights emerge.** This process has proven to be a very effective procedure for gaining a consensus ... It could not function effectively if participants had to be concerned that their half-thought-through, but nonetheless potentially valuable, notions would soon be made public. **I fear in such a situation the public record would be a sterile set of bland pronouncements scarcely capturing the necessary debates which are required of monetary policymaking.”**

Dependent Variables: Topics

1. Herfindahl concentration index (breadth of discussion).
2. Two topics that relate to quantitative material: data, staff materials.
3. Distance from average FOMC member
 - 3.1 *Dot product similarity*
 - 3.2 *Bhattacharyya coefficient*
 - 3.3 *Kullback-Leibler divergence*
4. Predicted probability of no dissent

Difference Specification

$$y_{it} = \alpha_i + \beta D(\textit{Trans}) + \lambda X_t + \varepsilon_t \quad (\text{DIFF})$$

Proxy for Strength of Career Concerns

In career concerns models (notably Holmström 1999), signaling incentives decline in the variance of the principal's belief about an agent's type.

As time passes, principals learn more about the agent's type \Rightarrow
experience reduces career concerns.

We construct FedExp_{it} to measure i 's period t experience as (1) years spent working in the Fed system prior to FOMC + (2) years on the FOMC.

Difference-in-difference specification is

$$y_{it} = \alpha_i + \delta_t + \eta \text{FedExp}_{it} + \phi D(\text{Trans})_t \times \text{FedExp}_{it} + \epsilon_{it} \quad (\text{DinD})$$

Summary table of effects of transparency

Evidence for discipline	Evidence for conformity
<ul style="list-style-type: none">↑ numbers in FOMC1/2↑ quantitative material in FOMC1↑ distance from average in FOMC1↓ probability of agreement in FOMC1	<ul style="list-style-type: none">↓ statements in FOMC2↓ questions in FOMC2↓ topic breadth in FOMC2↓ distance from average in FOMC2↑ probability of agreement in FOMC2

Alternative Application I: Survey Data

Project to document CEO time use (with O. Bandiera, A. Prat, and R. Sadun), and its effect on firm performance.

Data on each 15-minute block of time for one week of 1,114 CEOs' time classified according to

1. type (e.g. meeting, public event, etc.)
2. duration (15m, 30m, etc.)
3. planning (planned or unplanned)
4. number of participants (one, more than one)
5. functions of participants, divided between employees of the firms or “insiders” (finance, marketing, etc.) and “outsiders” (clients, banks, etc.).

There are 4,253 unique combinations of these five features in the data.

One can summarize the data with a 1114×4253 matrix where the (i, j) th element is the number of 15-minute time blocks that CEO i spends in activities with a particular combination of features j .

Alternative Application II: Demand

Increasingly common to get detailed data on consumers' shopping behavior.

Imagine a dataset in which consumer $i \in \{1, \dots, N\}$ makes multiple visits to a store that sells M possible goods bundles.

Then we can define an $N \times M$ matrix that counts the number of times each consumer buys each bundle.

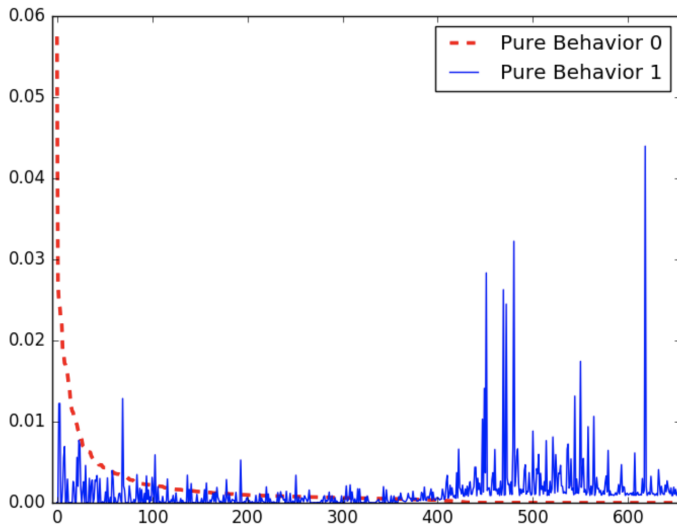
LDA on Survey Data

Since surveys are count data, we can apply LDA to response profiles (see Erosheva 2007; Munro and Ng 2019).

The β vectors have a different interpretation—‘pure behaviors’—but the statistical structure is isomorphic to that of LDA for text.

Here we present results for a model with $K = 2$.

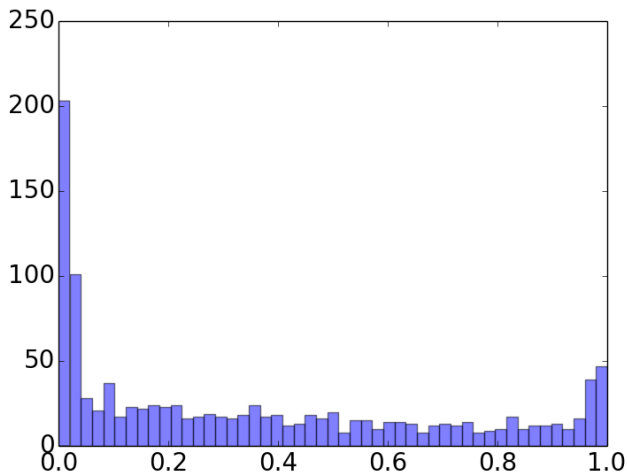
Pure Behaviors are Sharply Distinct



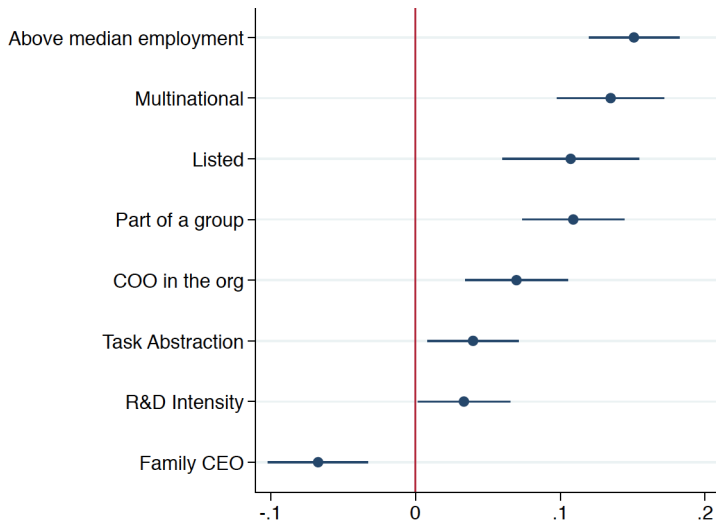
Differences across Pure Behaviors

Feature	X times less likely in Behavior 1	Feature	X times more likely in Behavior 1
Plant Visits	0.11	Communications	1.9
Just Outsiders	0.5	Outsiders + Insiders	1.9
Production	0.5	C-suite	34
Suppliers	0.3	Multifunction	1.5

Estimated Behavior Indices



Correlates of Behavior Index



Managers vs. Leaders

Kotter (1999) emphasizes a behavioral distinction between “management” and “leadership”.

Management involves monitoring and implementing tasks, i.e. “setting up systems to ensure that plans are implemented precisely and efficiently.”

Leadership aims primarily at the creation of organizational alignment, and involves significant investments in interpersonal communication.

The knowledge worker makes much greater time demands than the manual worker on his superiors as well as on his associates...One has to sit down with a knowledge worker and think through with him what should be done and why, before when knowing whether he is doing a satisfactory job or not (Drucker 1967).

Forecasting

Forecasting with unstructured data is key area of interest for central banks.

In recent survey of 52 central banks, 'forecasting' and 'nowcasting' are two most-cited potential applications of Big Data.

Jyry Hokkanen, the head of Statistics at Sveriges Riksbank was asked "what are the key big data research questions—what needs to be answered?" His answer was:

"I like text-mining techniques. I'd like that to become stable, because there's huge potential in it. We do communicate and central banks produce text, but we also receive a lot of text, and you can monitor society and the economy via tweets, speeches, articles, and company statements."

Forecasting

Forecasting with unstructured data is key area of interest for central banks.

In recent survey of 52 central banks, 'forecasting' and 'nowcasting' are two most-cited potential applications of Big Data.

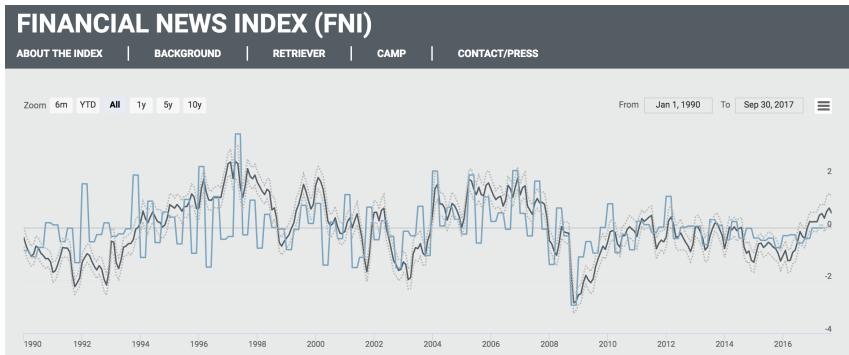
Jyry Hokkanen, the head of Statistics at Sveriges Riksbank was asked "what are the key big data research questions—what needs to be answered?" His answer was:

"I like text-mining techniques. I'd like that to become stable, because there's huge potential in it. We do communicate and central banks produce text, but we also receive a lot of text, and you can monitor society and the economy via tweets, speeches, articles, and company statements."

One strategy growing in popularity is to reduce the dimensionality of text with LDA, then use the topic shares as covariates in a standard time series model:

Examples include Thorsrud (2018) and Larsen and Thorsrud (2018) in macroeconomic forecasting; and Mueller and Rauh (2018) in conflict forecasting.

Financial News Index



Derived State-of-the-World

