Text Mining for the Social Sciences
Lecture 7: Using Topic Models in Social Science

Stephen Hansen

Introduction

Last time we discussed inference in the LDA graphical model.

Formally topics are distributions over words, and documents are distributions over topics.

In many applications, researchers treat these distributions as inherently meaningful, and then use them to evaluate social science models.

Important to realize that neither objective is intrinsic to LDA, which is simply a clustering model.

Outline

1. Topic meaning
   1.1 Human interpretation
   1.2 External validity
2. Case study on transparency and career concerns

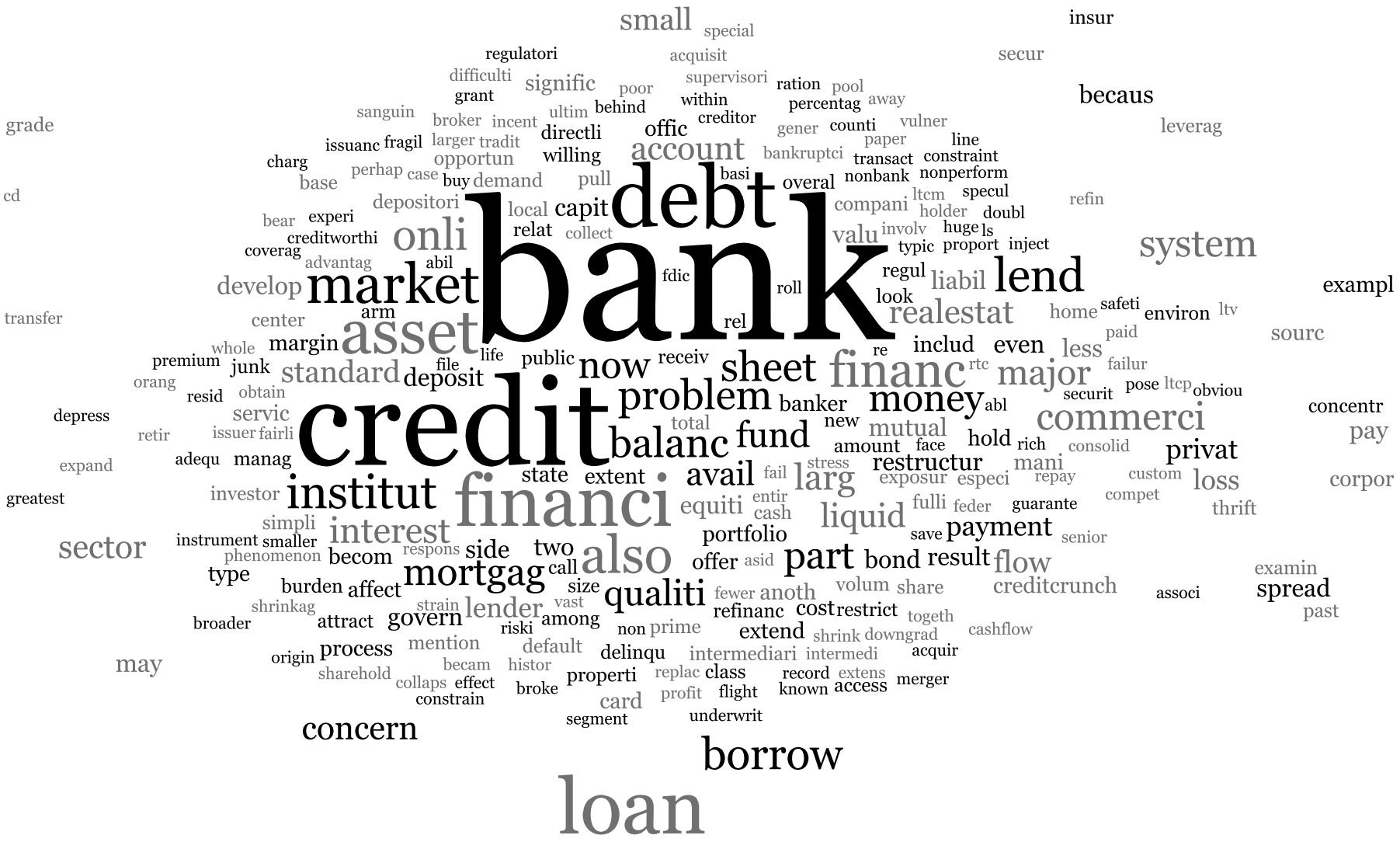
Topics and Meaning

Recall that the LSA representation of a document was useful for information retrieval, but not so much for interpreting content.

On the other hand, LDA consistently produces topics that seem to "mean" something.

But important to realize that any meaning we put on topics is an ex-post subjective judgment.

Topic 38


bank
credit
debt
financi
market
asset
institut
interest
mortgag
also
qualiti
balanc
problem
sheet
financ
money
fund
lend
major
commerci
system
account
onli
standard
deposit
now
realestat
liquid
part
payment
flow
loss
privat
loan
borrow
concern
sector
small
insur
secur
becaus
leverag
grade
cd
transfer
develop
center
margin
premium
junk
orang
resid
obtain
depress
servic
retir
issuer
fairli
expand
adequ
manag
greatest
investor
simpli
instrument
smaller
phenomenon
type
burden
affect
shrinkag
broader
attract
govern
mention
process
origin
may
sharehold
becam
collaps
effect
constrain
histor
broke
properti
card
segment
underwrit
profit
flight
replac
class
intermediari
intermedi
record
extens
known
access
merger
acquir
cashflow
shrink
downgrad
extend
non
prime
default
delinqu
refinanc
cost
restrict
togeth
fewer
anoth
volum
share
creditcrunch
associ
spread
past
examin
exampl
sourc
concentr
pay
corpor
thrift
custom
compet
guarante
feder
fulli
exposur
especi
repay
mani
consolid
rich
hold
face
amount
mutual
new
total
banker
abl
securit
pose
ltcp
obviou
failur
less
paid
even
includ
re
rtc
home
safeti
environ
ltv
regul
look
liabil
roll
fdic
rel
collect
relat
local
capit
depositori
bear
experi
creditworthi
coverag
advantag
abil
arm
life
public
receiv
file
stress
extent
state
avail
fail
larg
equiti
entir
cash
portfolio
offer
asid
save
senior
bond
result
call
size
vast
strain
lender
among
riski
default
non
fewer
cash
fund
sheet
regulatori
difficulti
grant
sanguin
broker
incent
ultim
behind
directli
willing
larger
tradit
opportun
issuanc
fragil
charg
perhap
case
base
buy
demand
pull
basi
overal
nonbank
nonperform
transact
constraint
paper
line
bankruptci
counti
vulner
gener
creditor
within
percentag
away
pool
ration
supervisori
acquisit
special
poor
significic
ltcm
specul
holder
doubl
compani
refin
involv
huge
ls
typic
proport
inject
valu

Topic 23

inventori
invest
capit
incom
busi
spend
stockmkt
profit
high
demand
earn
sector
consumpt
level
wealth
stock
effect
final
corpor
equiti
mortgag
expenditur
adjust
tax
basic
tech
rel
lag
fix
accumul
hold
fundament
purchas
equip
declin
firm
still
factor
swing
neg
see
fall
sale
part
gain
return
person
ratio
share
dispos
good
cashflow
exist
signific
strong
rapidli
though
analyst
real
play
overhang
relat
current
rise
famili
margin
big
valu
histor
saverate
substanti
import
low
cyclic
start
quit
show

Topic 29

deceler
averag
return
import
shift
offset
ad
name
pass
anecdot
anoth
capit
suppli
rise
acceler
wage
compens
markup
bonus
collect
pattern
live
ulc
may
part
recent
job
labor
product
trend
increas
cost
price
structur
demand
labmkt
pressur
technolog
employ
profit
output
worker
benefit
firm
gain
rate
per
enough
contribut
productgrowth
also
improv
higher
develop
boom
occur
much
remark
annual
surg
exert
consumprice
bargain
chain
soon
element
innov
perform
ultim
analysi
decad
declin
popul
combin
found
individu
difficult
payrol
slow
stori
enjoy
competit
resourc
medic
earn
unit
seen
result
fall
servic
insur
capu
workforc
clear
simpli
side
run
old
turn
larg
effect
pay
risen
new
upsid
power
shortag
escal
introduct
compon
total
associ
cyclic
attribut
potentioutput
behind
premium
avail
perman
know
onlin
welfar
longer
even
evid
faster
outputgrowth
despit
unusu
plenti
explan
upward
multifactor
squeez
help
experienc
deepen
previous
nonfarm
lower
effort
qualifi
dispar
latter
short
back
inflationari
match
transitori
stem
busicycl
skill
statist
lack
relat
employgrowth
reconcil
harder
health
hourli
produc
prevent
optimist
inde
account
particip
significantli
eci
work
becom
catch
train
materi
hypothesi
possibl
forc
workweek
input
unemploy
pickup
hike
translat
creation
advanc
virtual
share
less
slower
line
still
manag
demandgrowth
rapid
qualiti
extraordinari
mfp
ue
effici
estim
past
factor
asid
deriv
normal
util
allow
late
rais
salari
given
led
lf
tight
margin
growth
rapidli
attract
sum
hire
focu
rather
overal
eventu
onlin
measur
onlin
age
minimum
believ
prone
enhanc
employe
payrol
basic
prell
difficult
unit
output
profit
effect
turn
larg
risen
new
part
may
suppli
rise
acceler
wage
compens
markup
bonus
collect
pattern
live
ulc
shift
offset
ad
name
pass
anecdot
anoth
capit
averag
return
import
deceler
bargain
consumprice
chain
soon
element
innov
annual
surg
exert
develop
boom
perform
ultim
remark
much
occur
higher
improv
analysi
decad
declin
popul
combin
found
individu
difficult
payrol
slow
stori
enjoy
competit
resourc
medic
earn
unit
seen
result
fall
servic
insur
capu
workforc
clear
simpli
side
run
old
turn
larg
effect
pay
risen
new
upsid
power
shortag
escal
introduct
compon
total
associ
cyclic
attribut
potentioutput
behind
premium
avail
perman
know
onlin
welfar
longer
even
evid
faster
outputgrowth
despit
unusu
plenti
explan
upward
multifactor
squeez
help
experienc
deepen
previous
nonfarm
lower
effort
qualifi
dispar
latter
short
back
inflationari
match
transitori
stem
busicycl
skill
statist
lack
relat
employgrowth
reconcil
harder
health
hourli
produc
prevent
optimist
inde
account
particip
significantli
eci
work
becom
catch
train
materi
hypothesi
possibl
forc
workweek
input
unemploy
pickup
hike
translat
creation
advanc
virtual
share
less
slower
line
still
manag
demandgrowth
rapid
qualiti
extraordinari
mfp
ue
effici
estim
past
factor
asid
deriv
normal
util
allow
late
rais
salari
given
led
lf
tight
margin
growth
rapidli
attract
sum
hire
focu
rather
overal
eventu
onlin
measur
onlin
age
minimum
believ
prone
enhanc
employe
payrol
basic
prell
difficult
unit
output
profit
effect
turn
larg
risen
new
part
may
suppli
rise
acceler
wage
compens
markup
bonus
collect
pattern
live
ulc
shift
offset
ad
name
pass
anecdot
anoth
capit
averag
return
import
contribut
per
enough
productgrowth
also
improv
higher
develop
boom
occur
much
remark
annual
surg
exert
consumprice
bargain
chain
soon
element
innov
perform
ultim

Quinn et al (2010)

Table 3 Topic Keywords for 42-Topic Model

Topic (Short Label)	Keys
1. Judicial Nominations	*nomine, confirm, nomin, circuit, hear, court, judg, judici, case, vacanc*
2. Constitutional	*case, court, attornei, supreme, justic, nomin, judg, m, decis, constitut*
3. Campaign Finance	*campaign, candid, elect, monei, contribut, polit, soft, ad, parti, limit*
4. Abortion	*procedur, abort, babi, thi, life, doctor, human, ban, decis, or*
5. Crime 1 [Violent]	*enforc, act, crime, gun, law, victim, violenc, abus, prevent, juvenil*
6. Child Protection	*gun, tobacco, smoke, kid, show, firearm, crime, kill, law, school*
7. Health 1 [Medical]	*diseas, cancer, research, health, prevent, patient, treatment, devic, food*
8. Social Welfare	*care, health, act, home, hospit, support, children, educ, student, nurs*
9. Education	*school, teacher, educ, student, children, test, local, learn, district, class*
10. Military 1 [Manpower]	*veteran, va, forc, militari, care, reserv, serv, men, guard, member*
11. Military 2 [Infrastructure]	*appropri, defens, forc, report, request, confer, guard, depart, fund, project*
12. Intelligence	*intellig, homeland, commiss, depart, agenc, director, secur, base, defens*

Formalizing Interpretablility

Chang et. al. (2009) propose an objective way of determining whether topics are indeed interpretable.

Two tests:

1. *Word intrusion*. Form set of words consisting of top five words from topic k + word with low probability in topic k. Ask subjects to identify inserted word.

2. *Topic intrusion*. Show subjects a snippet of a document + top three topics associated to it + randomly drawn other topic. Ask to identify inserted topic.

Estimate LDA and other topic models on NYT and Wikipedia articles for $K = 50, 100, 150$.

Results

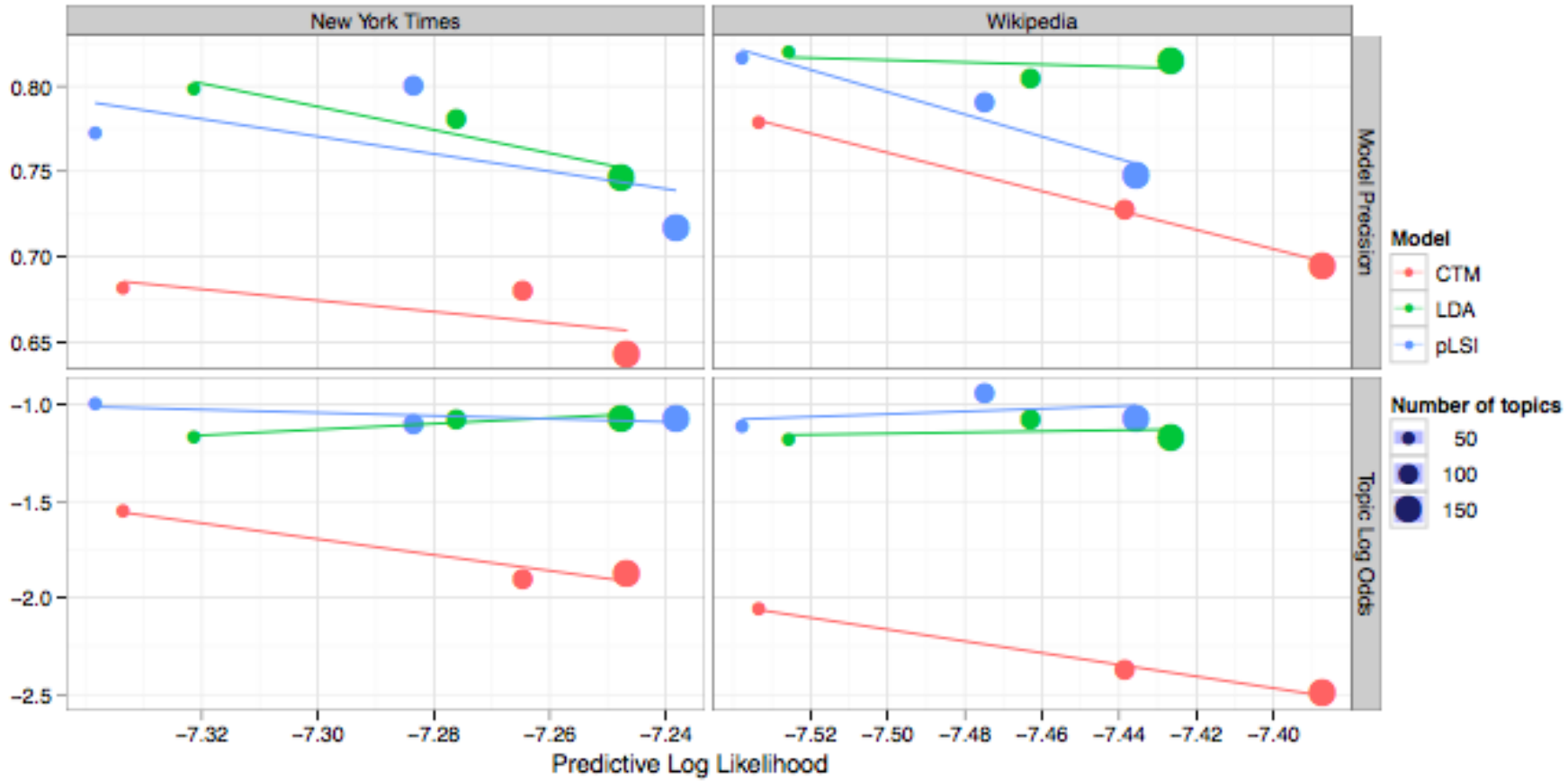


Figure 5: A scatter plot of model precision (top row) and topic log odds (bottom row) vs. predictive log likelihood. Each point is colored by model and sized according to the number of topics used to fit the model. Each model is accompanied by a regression line. Increasing likelihood does not increase the agreement between human subjects and the model for either task (as shown by the downward-sloping regression lines).

Implications

Topics seem objectively interpretable, at least in this context.

Tradeoff between goodness-of-fit and interpretablility, which is generally more important in social science.

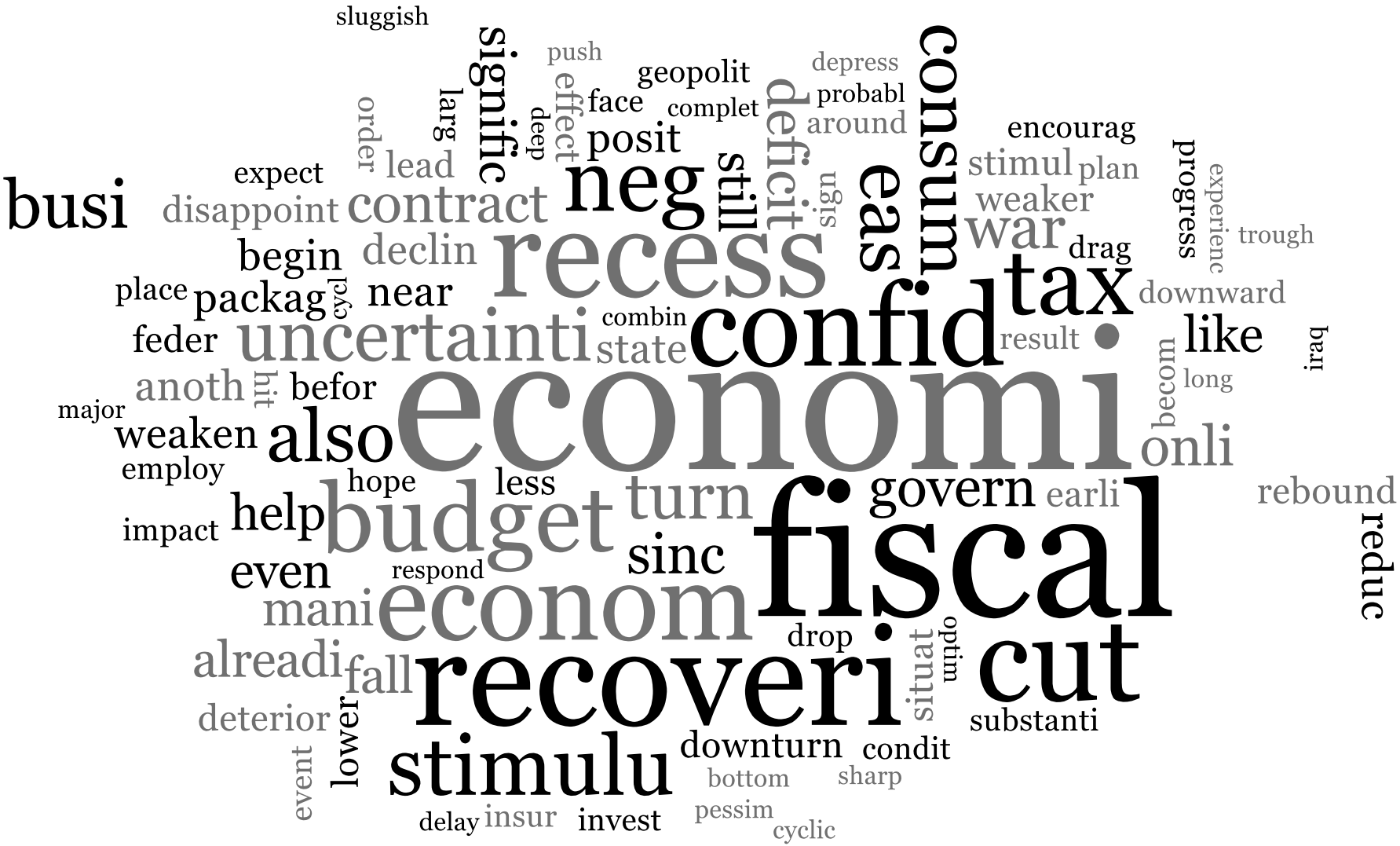
Potential development of statistical models in future to explicitly maximize interpretablility.

External Validity

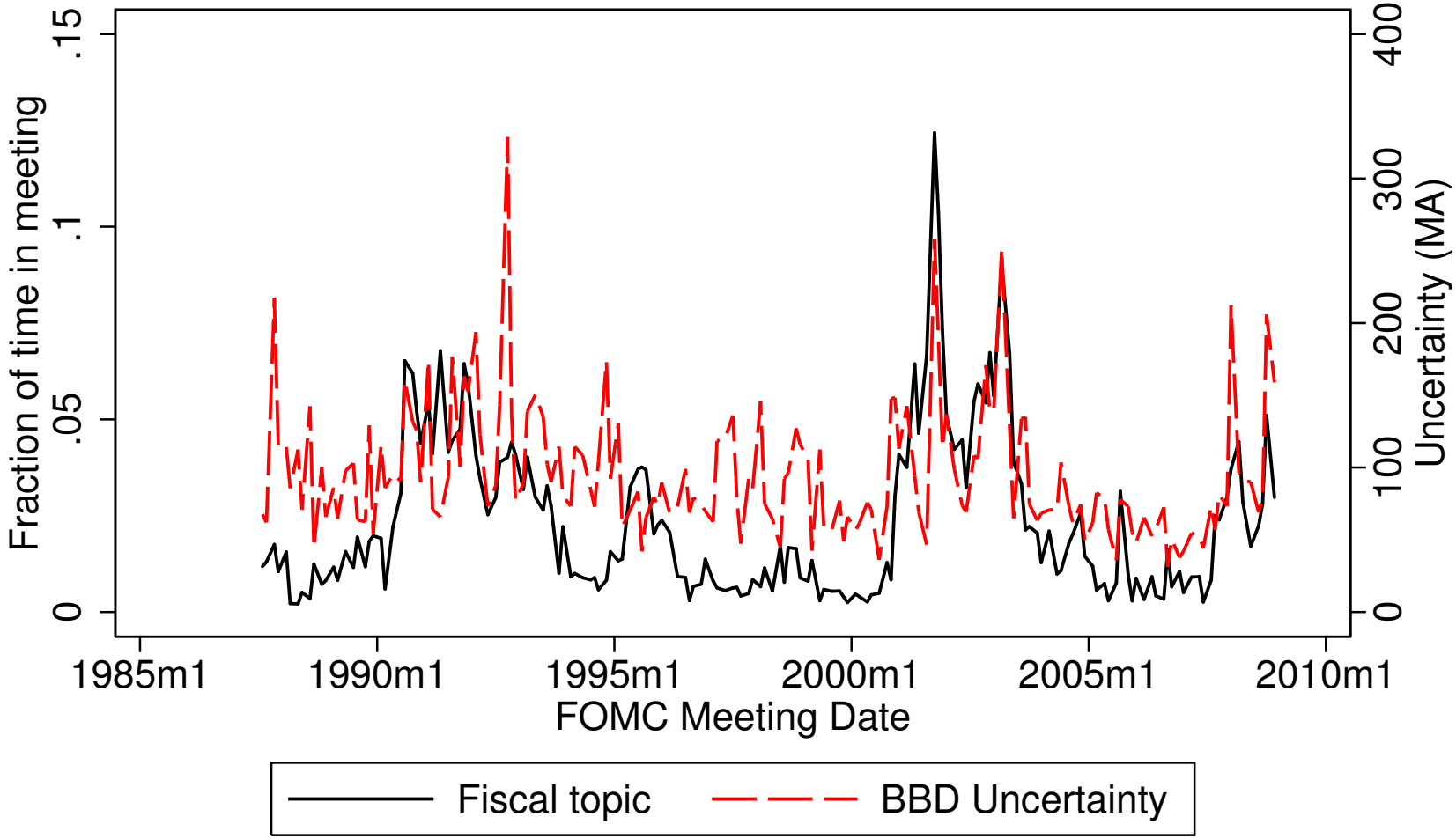
One common way of validating the subjective labels placed on topics is to see whether they relate to objective measures in ways we expect.

In FOMC context, can compare how attention changes with market conditions and external events.

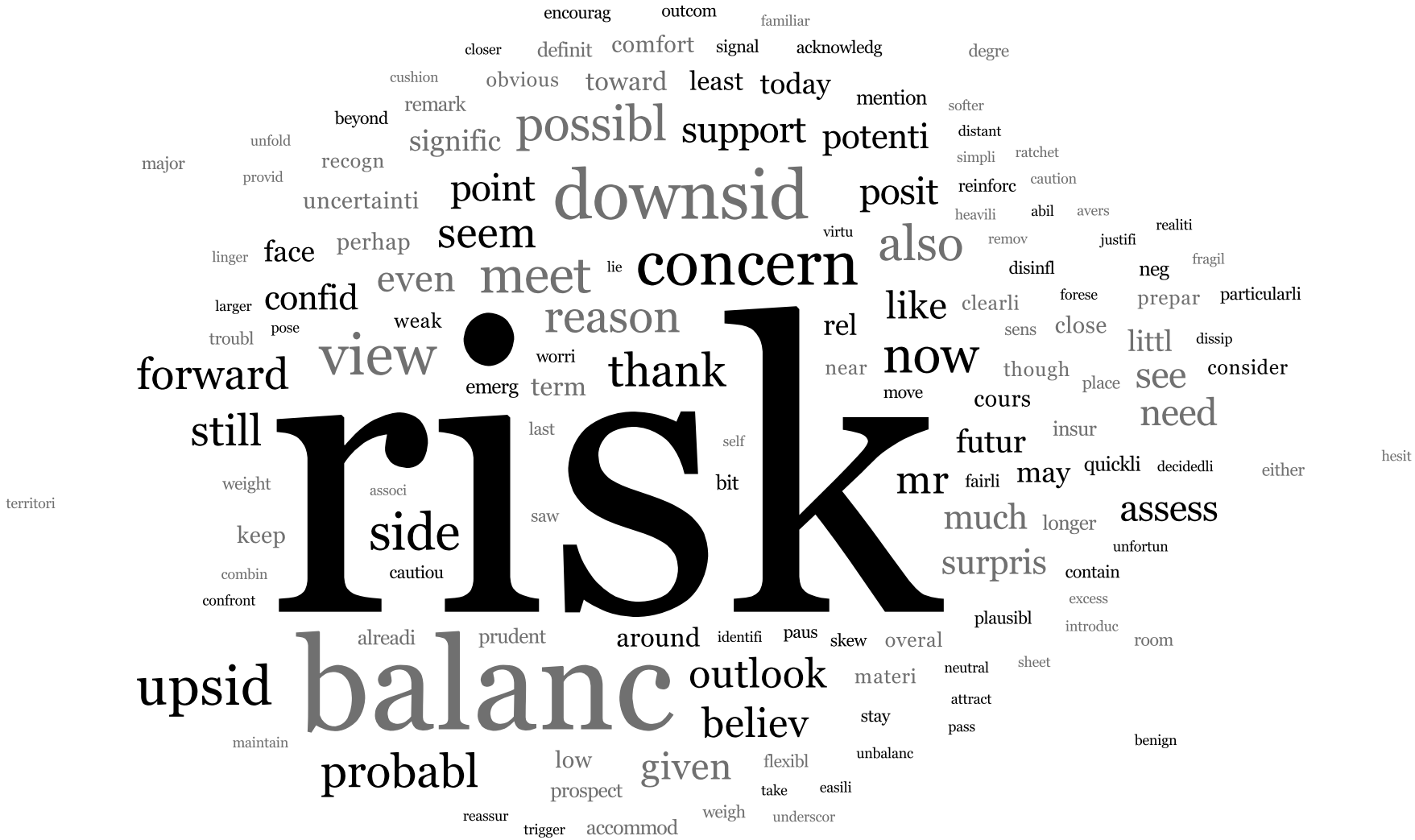
Topic 45


sluggish
push
geopolit
depress
consum
signific
effect
face
complet
probabl
around
order
larg
deep
posit
deficit
encourag
expect
lead
neg
still
sign
eas
stimul
plan
progress
experienc
busi
disappoint
contract
weaker
war
trough
begin
declin
recess
drag
place
packag
cycl
near
tax
downward
combin
confid
feder
uncertainti
state
result
like
iraq
anoth
hit
befor
becom
long
major
economi
weaken
also
onli
employ
hope
less
turn
govern
earli
rebound
impact
help
budget
fiscal
reduc
even
respond
sinc
mani
econom
drop
alreadi
fall
recoveri
situat
optim
cut
deterior
lower
substanti
event
stimulu
downturn
condit
bottom
sharp
delay
insur
invest
pessim
cyclic

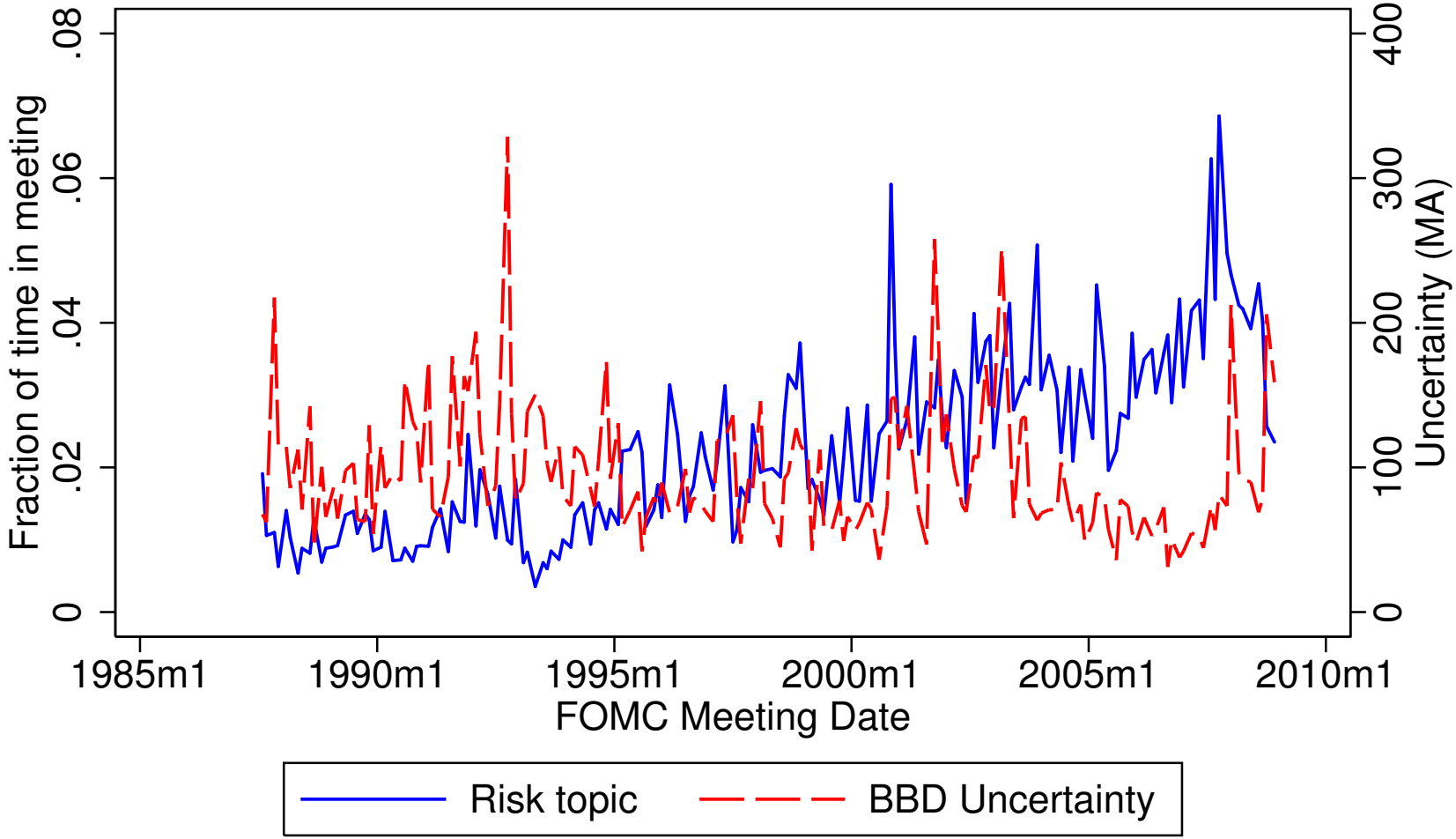
FOMC and Market Attention to "Recession/Fiscal"



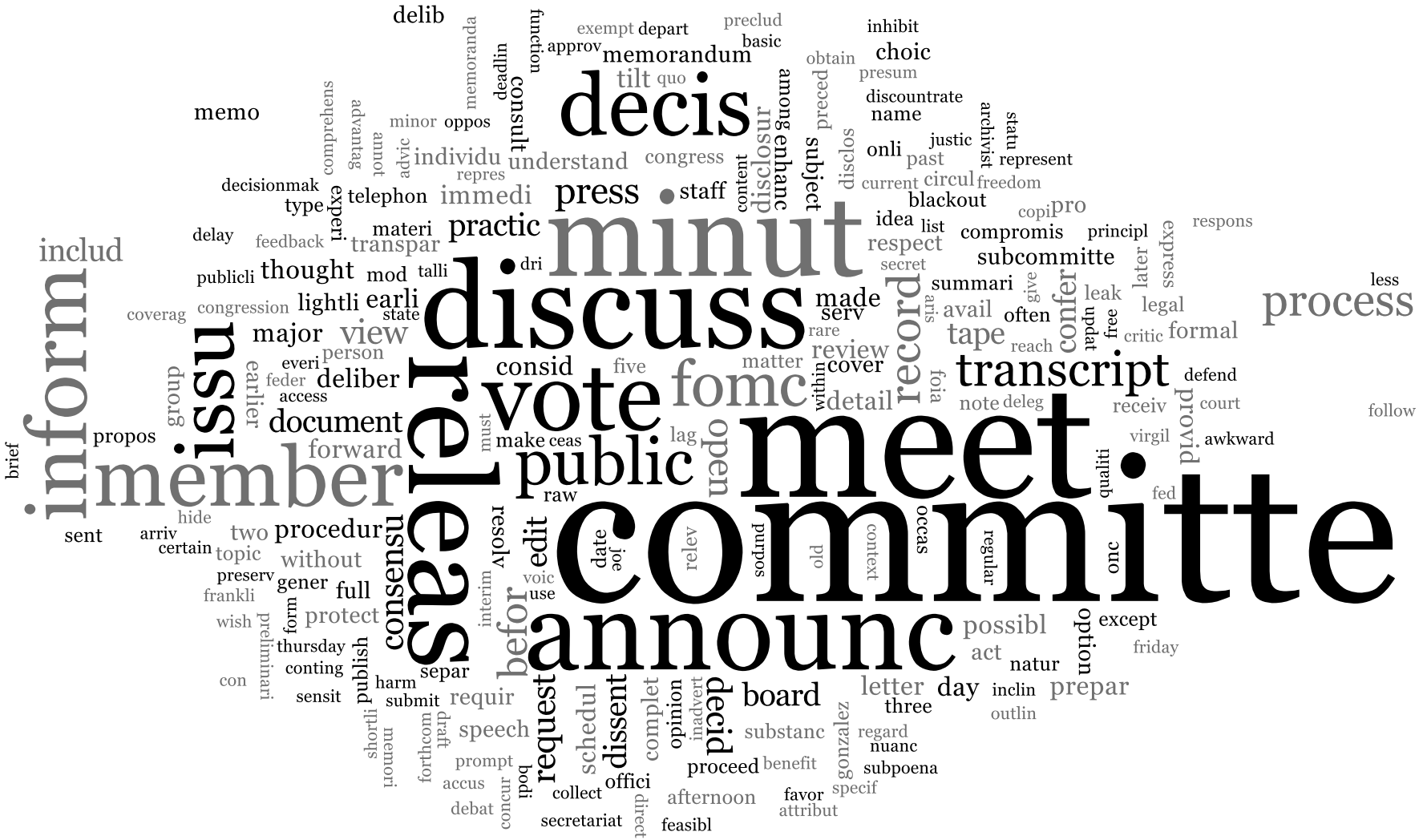
Topic 40


encourag
outcom
familiar
closer
definit
comfort
signal
acknowledg
degre
cushion
obvious
toward
least
today
mention
softer
remark
beyond
possibl
support
potenti
distant
unfold
significc
simpli
ratchet
major
recogn
provid
point
downsid
posit
reinforc
caution
uncertainti
heavili
abil
avers
realiti
seem
virtu
also
remov
justifi
linger
face
perhap
lie
concern
disinfl
neg
fragil
even
meet
larger
confid
like
clearli
forese
prepar
particularli
weak
reason
rel
sens
close
troubl
pose
littl
dissip
view
worri
now
forward
thank
near
though
place
see
consider
emerg
term
move
cours
need
still
last
self
insur
futur
risk
hesit
weight
associ
bit
mr
fairli
may
quickli
decidedli
either
territori
saw
much
longer
assess
keep
side
unfortun
surpris
combin
cautiou
contain
confront
excess
plausibl
introduc
alreadi
prudent
around
identifi
paus
skew
overal
room
balanc
neutral
sheet
upsid
outlook
materi
attract
stay
believ
pass
maintain
benign
unbalanc
low
given
flexibl
probabl
prospect
take
easili
reassur
weigh
underscor
trigger
accommod

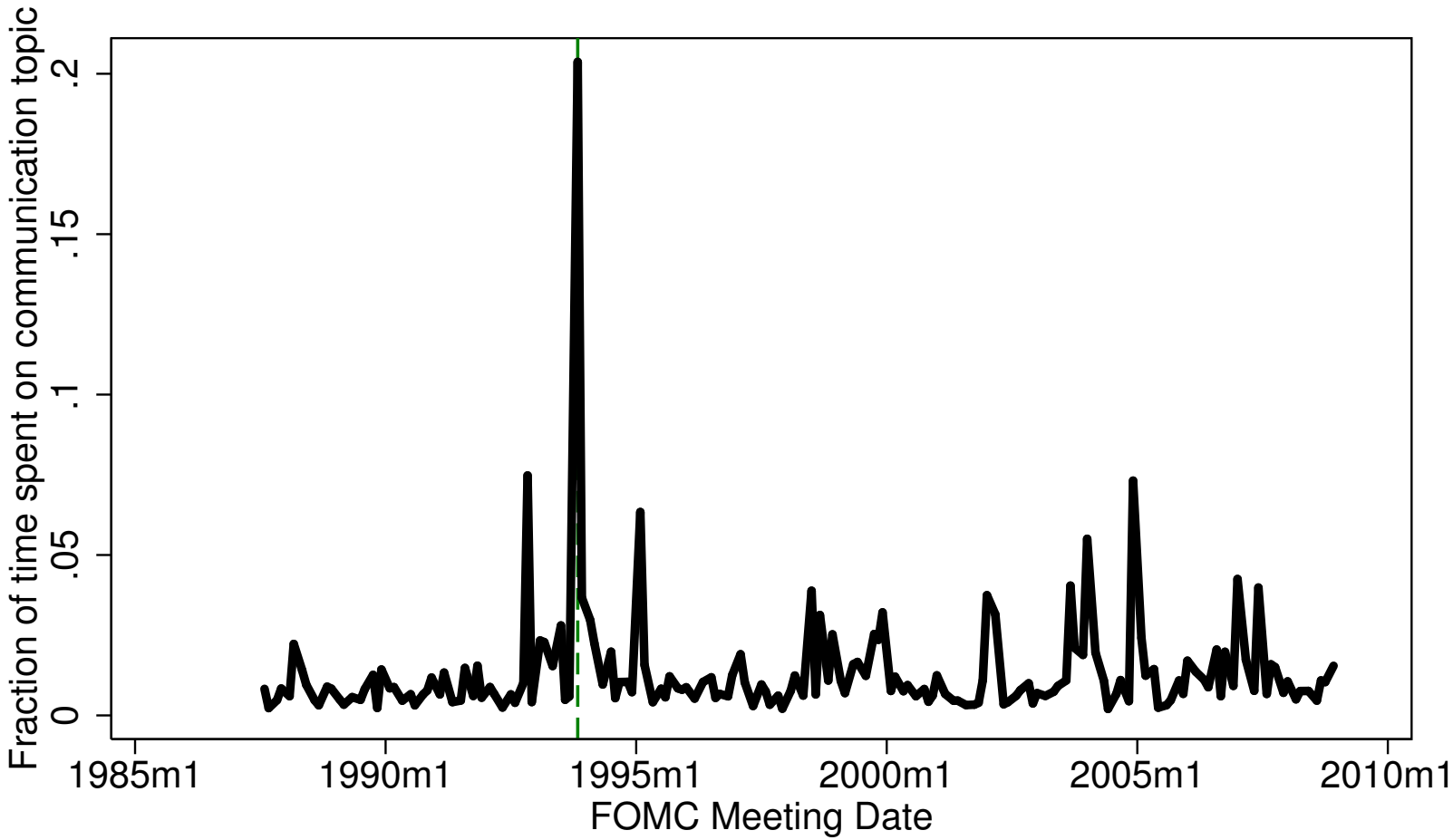
FOMC and Market Attention to "Risk"



Topic 6


committe
meet
announc
minut
discuss
decis
releas
vote
fomc
transcript
public
member
inform
issu
record
process
press
consensu
befor
request
decid
board
schedul
dissent
complet
opinion
substanc
letter
day
prepar
possibl
except
option
act
natur
inclin
outlin
three
regard
nuanc
subpoena
benefit
proceed
afternoon
favor
attribut
specif
gonzalez
secretariat
feasibl
collect
offici
debat
accus
prompt
speech
requir
submit
separ
harm
publish
thursday
conting
sensit
con
wish
preliminari
form
protect
full
gener
preserv
frankli
without
topic
certain
arriv
sent
hide
two
procedur
edit
date
joe
relev
sodpunc
old
context
occas
onc
regular
qualiti
fed
virgil
awkward
provid
court
receiv
note
deleg
defend
foia
detail
within
cover
review
matter
five
consid
tape
reach
critic
formal
legal
avail
often
confer
leak
later
express
respons
less
follow
summari
secret
subcommitte
compromis
principl
respect
idea
list
pro
copi
blackout
current
circul
freedom
represent
past
justic
onli
archivist
statu
name
discountrate
presum
obtain
choic
inhibit
memorandum
basic
preclud
exempt
depart
approv
function
delib
tilt
quo
among
enhanc
disclosur
subject
disclos
preced
consult
deadlin
memoranda
minor
oppos
individu
understand
congress
content
repres
staff
immedi
telephon
type
decisionmak
experi
memo
comprehens
advantag
annot
advic
delay
feedback
materi
practic
transpar
dri
publicli
thought
mod
talli
lightli
earli
state
congression
coverag
major
view
person
everi
feder
deliber
access
earlier
group
document
forward
propos
brief
make
ceas
must
lag
open
raw
resolv
interim
use
voic
frankli
wish
thursday
publish
harm
submit
requir
speech
prompt
accus
debat
concur
bodi
collect
offici
secretariat
feasibl
afternoon
favor
attribut
specif
gonzalez
regard
nuanc
subpoena
benefit
proceed
substanc
board
decid
opinion
inadvert
complet
schedul
dissent
request
befor
consensu
releas
issu
member
inform
includ
propos
brief
sent
arriv
certain
topic
preserv
frankli
without
two
procedur
hide
edit
date
joe
relev
sodpunc
old
context
occas
onc
regular
qualiti
fed
virgil
awkward
provid
court
receiv
note
deleg
defend
foia
detail
within
cover
review
matter
five
consid
tape
reach
critic
formal
legal
avail
often
confer
leak
later
express
respons
less
follow
summari
secret
subcommitte
compromis
principl
respect
idea
list
pro
copi
blackout
current
circul
freedom
represent
past
justic
onli
archivist
statu
name
discountrate
presum
obtain
choic
inhibit
memorandum
basic
preclud
exempt
depart
approv
function
delib
tilt
quo
among
enhanc
disclosur
subject
disclos
preced
consult
deadlin
memoranda
minor
oppos
individu
understand
congress
content
repres
staff
immedi
telephon
type
decisionmak
experi
memo
comprehens
advantag
annot
advic
delay
feedback
materi
practic
transpar
dri
publicli
thought
mod
talli
lightli
earli
state
congression
coverag
major
view
person
everi
feder
deliber
access
earlier
group
document
forward
propos
brief
make
ceas
must
lag
open
raw
resolv
interim
use
voic
frankli
wish
thursday
publish
harm
submit
requir
speech
prompt
accus
debat
concur
bodi
collect
offici
secretariat
feasibl
afternoon
favor
attribut
specif
gonzalez
regard
nuanc
subpoena
benefit
proceed
substanc
board
decid
opinion
inadvert
complet
schedul
dissent
request
befor
consensu
releas
issu
member
inform
includ

Surprised by Transparency?



Topics and Model Testing

Once we have estimated topics, what can we do with them?

One option is to simply describe and index text.

Another goal is to test behavioral models by using topics as dependent variables in regressions motivated by theory.

Example is Hansen, McMahon, and Prat (2015).

The Environment

Recall discussion of benefits and costs of transparency and natural experiment at FOMC.

How does economic theory predict FOMC members should react to increase in transparency?

Career Concerns

Dynamic models in which only short-term contracts are possible.

Agent's type is unknown, and observed behavior/performance today determines future principals' beliefs.

Agents today maximizes future principals' beliefs (which may be different from today's productivity).

Payoff to good reputation can be:

1. Higher wages
2. Non-monetary perks
3. Psychological benefit

Career Concerns and Transparency

What are the predicted effect of transparency?

Career Concerns and Transparency

What are the predicted effect of transparency?

Positive
Discipline: Holmström (1999).

Career Concerns and Transparency

What are the predicted effect of transparency?

Positive
Discipline: Holmström (1999).

Negative
Conformity: Scharfstein and Stein (1990), Prat (2005)
Non-conformity: Ottaviani and Sørensen (2000), Levy (2007)

Career Concerns and Transparency

What are the predicted effect of transparency?

Positive
Discipline: Holmström (1999).

Negative
Conformity: Scharfstein and Stein (1990), Prat (2005)
Non-conformity: Ottaviani and Sørensen (2000), Levy (2007)

These effects predicted to decline with experience.

FOMC Background

The Federal Open Market Committee meets eight times per year.

19 members: seven Governors, and twelve regional Fed Presidents.

Meeting Structure:

1. **Economic Situation Discussion (FOMC 1)**
   - 1.1 Board staff present the economic situation (including forecast).
   - 1.2 FOMC members present their views of the economic outlook.
2. **Policy Strategy Debate (FOMC 2)**
   - 2.1 Staff presents a variety of monetary policy alternatives (without a recommendation).
   - 2.2 First the Chairman, and then other FOMC members discuss their policy preferences.

FOMC takes Government in Sunshine Act seriously.

Testing through Differences

One option is to run

$$y_{ts} = \alpha + \beta D(\mathit{Trans}) + \lambda X_t + \varepsilon_t \qquad \text{(DIFF)}$$

Main issue is that other changes may have coincided with increase in transparency.

With high frequency data, for example tick-level equity prices, one can conduct event study.

FOMC only meets eight times per year, so event study not feasible.

Difference in Differences

Difference-in-differences controls for (common) time trends by comparing how one group's reaction to transparency differs from another's.

In a pure RCT, there would be a treatment and control group.

Instead, HMP use experience in Fed system to proxy for strength of career concerns, which should correlate with behavioral response to transparency:

$$FedExp_{it} = \text{years in Fed system prior to FOMC} + \text{years on the FOMC}$$

$$y_{its} = \alpha_i + \delta_t + \beta D(Trans) + \eta FedExp_{it} + \phi D(Trans) \times FedExp_{it} + \epsilon_{it} \quad \text{(DinD)}$$

Test ϕ coefficient on interaction term.

Dependent Variables

Counts

1. How much someone speaks—words
2. How many statements they make
3. How many questions they ask
4. How many numbers they refer to

Dependent Variables

Counts

1. How much someone speaks—words
2. How many statements they make
3. How many questions they ask
4. How many numbers they refer to

Functions of Topic Content

Dependent Variables

Counts

1. How much someone speaks—words
2. How many statements they make
3. How many questions they ask
4. How many numbers they refer to

Functions of Topic Content

Influence

Counts—Diff Regression

Main Regressors	(1) Total Words	(2) Statements	(3) Questions	(4) Numbers	(5) Total Words	(6) Statements	(7) Questions	(8) Numbers
D(Trans)	52.3*	-0.36	-0.00093	3.56***	35.3	-1.98**	-0.67***	1.60***
	[0.081]	[0.324]	[0.990]	[0.004]	[0.137]	[0.011]	[0.006]	[0.001]
D(Short)	-71.9***	-0.30**	-0.11	-1.74**	-178***	-1.18***	-0.40***	-1.83***
	[0.000]	[0.014]	[0.327]	[0.049]	[0.000]	[0.000]	[0.000]	[0.000]
D(NBER)	14.4	-0.38	-0.075	-1.00	-30.6	0.46	-0.068	-0.64
	[0.571]	[0.208]	[0.665]	[0.197]	[0.259]	[0.482]	[0.752]	[0.175]
BBD uncertainty	0.22	-0.0032	0.00021	0.0033	-0.042	-0.0093**	-0.0042***	0.00018
	[0.252]	[0.506]	[0.870]	[0.527]	[0.872]	[0.035]	[0.000]	[0.969]
D(2 day)	34.5	1.39**	0.57**	1.44**	59.5	-0.13	0.071	1.08*
	[0.115]	[0.048]	[0.020]	[0.044]	[0.219]	[0.841]	[0.733]	[0.079]
Constant	655***	4.59***	1.13***	7.93***	329***	6.31***	1.69***	2.20***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.004]
Unique Members	36	36	36	36	36	36	36	36
Member FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	No	No	No	No
Within Meeting	FOMC1	FOMC1	FOMC1	FOMC1	FOMC2	FOMC2	FOMC2	FOMC2
Sample	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09
Obs	1148	1148	1148	1148	1138	1138	1138	1138
Transparency effect	7	-	-	14	-	-49	-49	14

Counts—DiD Regression

Main Regressors	(1) Total Words	(2) Statements	(3) Questions	(4) Numbers	(5) Total Words	(6) Statements	(7) Questions	(8) Numbers
D(Trans)	-486***	-5.85**	-2.72	-8.43***	-2,940	82.3	38.9	-20.2
	[0.000]	[0.010]	[0.139]	[0.002]	[0.268]	[0.293]	[0.117]	[0.706]
Fed Experience	973***	6.38	5.04	21.6***	232	-5.49	-2.62	1.83
	[0.000]	[0.142]	[0.163]	[0.001]	[0.200]	[0.305]	[0.124]	[0.618]
D(Trans) x Fed Experience	0.42	0.026	0.0047	-0.21***	-0.68	0.11***	0.037***	-0.078***
	[0.798]	[0.298]	[0.667]	[0.004]	[0.738]	[0.010]	[0.007]	[0.005]
Constant	-10,240***	-66.7	-55.7	-234***	0	0	0	0
	[0.000]	[0.175]	[0.172]	[0.002]	[.]	[.]	[.]	[.]
Unique Members	36	36	36	36	36	36	36	36
Member FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Within Meeting	FOMC1	FOMC1	FOMC1	FOMC1	FOMC2	FOMC2	FOMC2	FOMC2
Sample	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09
Obs	1148	1148	1148	1148	1138	1138	1138	1138
Rookie effect	-	-	-	14	-	-49	-49	14

Dependent Variables: Topics

Statement $\xrightarrow{LDA}$ K-dimensional probability vector $\xrightarrow{Function}$ dependent variable.

Dependent Variables: Topics

Statement $\xrightarrow{LDA}$ K-dimensional probability vector $\xrightarrow{Function}$ dependent variable.

1. Percentage of statement about economic topics (preparedness):
   Define conditional topic distribution $\chi_{i,t,s}$

Dependent Variables: Topics

Statement $\xrightarrow{LDA}$ K-dimensional probability vector $\xrightarrow{Function}$ dependent variable.

1. Percentage of statement about economic topics (preparedness):
   Define conditional topic distribution $\chi_{i,t,s}$
2. Herfindahl concentration index (breadth of discussion).

Dependent Variables: Topics

Statement $\xrightarrow{LDA}$ K-dimensional probability vector $\xrightarrow{Function}$ dependent variable.

1. Percentage of statement about economic topics (preparedness):
   Define conditional topic distribution $\chi_{i,t,s}$
2. Herfindahl concentration index (breadth of discussion).
3. Percentage of time on quantitative topics (information acquisition).

Dependent Variables: Topics

Statement $\xrightarrow{LDA}$ K-dimensional probability vector $\xrightarrow{Function}$ dependent variable.

1. Percentage of statement about economic topics (preparedness):
   Define conditional topic distribution $\boldsymbol{\chi}_{i,t,s}$
2. Herfindahl concentration index (breadth of discussion).
3. Percentage of time on quantitative topics (information acquisition).
4. Proximity to Chairman in FOMC2:
   - 4.1 *Dot product similarity*: $DP_{it} = \sum_k \chi^k_{G,t}\chi^k_{i,t}$.
   - 4.2 *Bhattacharyya coefficient*: $BH_{it} = \sum_k \sqrt{\chi^k_{G,t}\chi^k_{i,t}}$.
   - 4.3 *Kullback-Leibler divergence*: $KL_{it} = \sum_k \chi^k_{G,t} \ln\left(\frac{\chi^k_{G,t}}{\chi^k_{i,t}}\right)$.

Topic 0

certainli
come
even
look
particular
stronger
better
see
shape
board
situat
princip
bit
mike
moment
anecdot
seem
talk
fairli
perform
case
fair
attitud
number
hard
find
cautiou
perhap
particularli
lot
far
clearli
ani
side
feel
realli
good
part
comment
modest
improv
concern
beyond
obvious
gener
quit
hear
pictur
place
sens
anticip
develop
still
optimist
pretti
cours
bad
economi
term
impress
anyth
peopl
least
posit
surpris
realgrowth
regard
although
might

Topic 7—Data Topic

Transparency and Deliberation

Main Regressors	(1) Economics	(2) Economics	(3) Herfindahl	(4) Herfindahl
D(Trans)	0.16***	0.064	-0.0038	0.34*
	[0.002]	[0.933]	[0.674]	[0.055]
Fed Experience	-0.28***	0.036	-0.0035	-0.018
	[0.008]	[0.484]	[0.852]	[0.134]
D(Trans) x Fed Experience	0.00019	-0.0014**	0.00061*	-0.00028***
	[0.655]	[0.018]	[0.060]	[0.003]
Constant	3.70***	0	0.15	0
	[0.002]	[.]	[0.483]	[.]
Unique Members	36	36	36	36
Member FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Within Meeting	FOMC1	FOMC2	FOMC1	FOMC2
Sample	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09
Obs	1148	1138	1148	1138
Rookie effect	-	21	-20	12

Transparency and Quantitative Discussion

Main Regressors	(1) Numbers	(2) Numbers	(3) Data Topics (7&11)	(4) Data Topics (7&11)
D(Trans)	-8.43***	-20.2	-0.032**	-0.10
	[0.002]	[0.706]	[0.019]	[0.631]
Fed Experience	21.6***	1.83	0.066**	0.010
	[0.001]	[0.618]	[0.032]	[0.485]
D(Trans) x Fed Experience	-0.21***	-0.078***	-0.00071***	-0.00027**
	[0.004]	[0.005]	[0.003]	[0.035]
Constant	-234***	0	-0.69**	0
	[0.002]	[.]	[0.045]	[.]
Unique Members	36	36	36	36
Member FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Within Meeting	FOMC1	FOMC1	FOMC1	FOMC2
Sample	89:11-97:09	89:11-97:09	89:11-97:09	89:11-97:09
Obs	1148	1138	1148	1138
Rookie effect	14	14	17	16

Proximity and Preferences

Main Regressors	(1) D(Voice Dissent)	(2) D(Voice Dissent)	(3) D(Voice Dissent)
D(Non-Voter)	0.0060	0.0072	0.0083
	[0.802]	[0.764]	[0.727]
DP	-1.03*		
	[0.059]		
BH		-0.46***	
		[0.005]	
KL			0.11***
			[0.003]
Constant	0.49***	0.81***	0.34***
	[0.000]	[0.000]	[0.000]
R-squared	0.226	0.229	0.229
Unique Members	35	35	35
Member FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Obs	1194	1194	1194
Type of measure	Similarity	Similarity	Distance

Transparency and Proximity

Main Regressors	(1) DP	(2) BH	(3) KL
D(Trans)	0.013***	0.78	1.70
	[0.000]	[0.253]	[0.576]
Fed Experience	-0.011	0.0035	-0.065
	[0.453]	[0.940]	[0.754]
D(Trans) x Fed Experience	-0.00021**	-0.00058**	0.0023*
	[0.037]	[0.033]	[0.055]
Constant	0.17	0	0
	[0.307]	[.]	[.]
Unique Members	35	36	36
Member FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Within Meeting	FOMC2	FOMC2	FOMC2
Sample	89:11-97:09	89:11-97:09	89:11-97:09
Obs	1074	1138	1138
Type of measure	Similarity	Similarity	Distance
Rookie effect	12	11	-9

Summary

Table: Evidence for career concerns

Discipline	**Conformity**
↑ use of numbers in FOMC1	↓ statements in FOMC2
↑ topic breadth in FOMC1	↓ questions in FOMC2
↑ references to data topics in FOMC1	↓ distance from Greenspan in FOMC2
	↓ topic breadth in FOMC2
↑ economics topic percentage in FOMC2	

Influence as a Measure of Informativeness

Discipline tends to increase informativeness, conformity to decrease it. Which effect dominates?

If member's statement more informative, it should drive the debate more.

Influence

When what member i talks about affects what member j talks about, call this influence. Use a PageRank-like algorithm.

Influence as a Measure of Informativeness

Discipline tends to increase informativeness, conformity to decrease it. Which effect dominates?

If member's statement more informative, it should drive the debate more.

Influence

When what member i talks about affects what member j talks about, call this influence. Use a PageRank-like algorithm.

Measure influence:

1. Within meeting
2. Within meeting on Greenspan
3. Across meetings
4. Across meetings on Greenspan

Overall Influence

Main Regressors	(1) W_D	(2) A_D	(3) W_{BH}	(4) A_{BH}
D(Trans)	0.0028***	0.0019**	0.0030***	-0.000057
	[0.000]	[0.036]	[0.000]	[0.810]
Fed Experience	-0.0039	-0.010*	-0.0012	0.00039
	[0.380]	[0.057]	[0.534]	[0.830]
D(Trans) x Fed Experience	-0.000015	-0.00019***	-0.000042**	-0.000012
	[0.732]	[0.009]	[0.041]	[0.439]
Constant	0.096*	0.17***	0.064***	0.051**
	[0.068]	[0.006]	[0.005]	[0.018]
Unique Members	35	32	35	32
Within Meeting	Intra	Inter	Intra	Inter
Obs	1074	1039	1074	1039
Rookie effect	-	17	7	-

Influence on Greenspan

Main Regressors	(1) W_D^G	(2) A_D^G	(3) W_{BH}^G	(4) A_{BH}^G
D(Trans)	0.00031***	0.000032	0.00030***	-0.000012
	[0.009]	[0.814]	[0.000]	[0.711]
Fed Experience	0.00023	-0.0017*	-0.000079	-0.000092
	[0.748]	[0.054]	[0.785]	[0.637]
D(Trans) x Fed Experience	-6.6e-06	-0.000022***	-5.3e-06***	-1.4e-06
	[0.198]	[0.004]	[0.000]	[0.319]
Constant	0.00012	0.022**	0.0034	0.0040*
	[0.988]	[0.030]	[0.309]	[0.080]
Unique Members	35	32	35	32
Within Meeting	Intra	Inter	Intra	Inter
Obs	1074	1039	1074	1039
Rookie effect	-	18	8	-

Conclusion

Topic models are a growth area in social science because they allow us to measure text in an interpretable way, and to quantitatively test behavioral models with qualitative data.

This process has been underway for a few years in political science, but essentially brand new in economics.

In future, scope for developing structural topic models whose latent variables explicitly form part of an economic model's data generating process.