Wharton
University of Pennsylvania

JACOBS LEVY EQUITY MANAGEMENT CENTER
for Quantitative Financial Research

Dynamic Interpretation of Emerging Risks in the Financial Sector

PRESENTER
Kathleen Weiss Hanley, Lehigh University

Joint work with Gerard Hoberg, University of Southern California

National Science Foundation

- Project made feasible by grant #1449578 funded through NSF CIFRAM program .

- Understanding the economic channels of system-wide risk build-up is important in heading off future crises

Existing measures of systemic risk

Bisias, Flood, Lo and Valavanis (2012) summarize over 30 quantitative systemic risk metrics

- Liquidity mismatch (Brunnermeier, Gorton and Krishnamurthy, 2014), interconnectedness (Billio, Getmansky, Lo and Pelizzon, 2012), and bank risk (Adrian and Brunnermeier, 2016) to name only a few
- Quantitative metrics, although useful, have the following drawbacks:
  - General measures: Difficult to identify underlying source of risk
  - Specific measures: Requires a specific theory and may not be useful if source of risk is unknown

Using computational linguistics and big data, we crowd source aggregate risks across entire banking industry and present a dynamic measure that is specific about channels

Our findings

Our method can provide an early warning signal of potential financial instability, identify economic causes and determine which banks may be most affected

- Aggregate risk score becomes highly significant in 2Q2005 well in advance of the financial crisis
- Economic factors known to contribute to the financial crisis are elevated in the period leading up to Lehman's failure
- More importantly, we see significant increase in risk build-up in the current period
- Individual bank exposure to risk themes predicts crises returns, failure and volatility

Information production

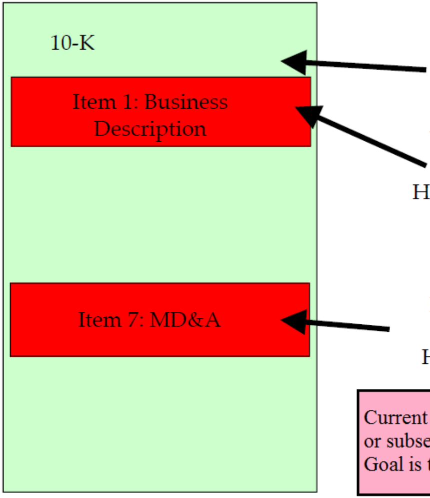
Our methodology requires that both banks and investors produce information

- Banks
  - Banks are required by SEC to disclose exposure to risks in the 10-K are high-level discussions
  - Useful to investors to determine whether the banking sector has become more risky thereby necessitating additional information production
- Investors
  - Produce and aggregate information that is manifest in stock returns (Hayek (1945), Grossman and Stiglitz (1980)
  - Use covariance of asset returns to measure commonality of risk exposure between banks

Emerging risks

Propose two methods to detect emerging risks

- Static model
  - Risks identified from manual inspection of textual data
  - Economic risks that affect the banking sector regardless of time period studied
- Dynamic model
  - Automated identification of risks
  - Allows different emerging risks to “bubble” up in each year



10-K Has many sections, MD&A, Legal are just two

Business Description: used to identify new industries as in Hoberg and Phillips (2010, 2014).

Management's Discussion and Analysis (MD&A): used in Hoberg and Maksimovic (2014).

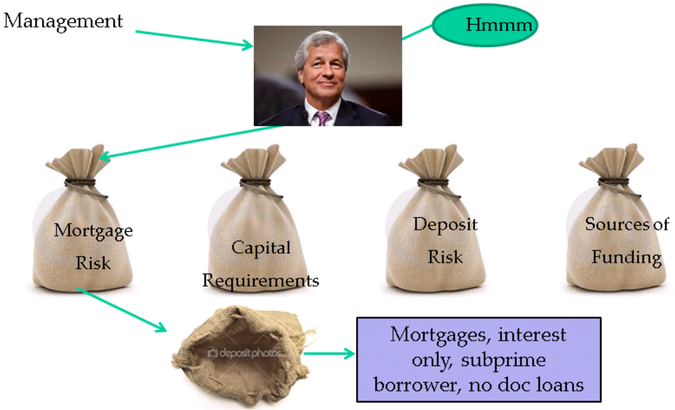
Current study searches entire 10-K for sections or subsection containing the word root ``risk". Goal is to get disclosed risk factors.

Latent Dirichlet Allocation (LDA)

- LDA proposed by Blei, Ng, Jordan, Michael (2003) in *Journal of Machine Learning Research*
- Proposes that writer is like a hidden Markov Chain who chooses among topics to discuss and then draws words from topic distribution
- Use Gibbs Sampling to get “most likely” topics.
- Goal is to use context to identify interpretable content
- LDA is automated, replicable and cannot be influenced by researcher bias
    - Our only input is number of topics (25) to be generated

LDA topics

Risk Factor Document Creation

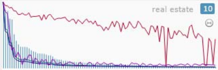


* CEO can be modeled as a hidden Markov Chain, a state is a chosen topic, and he/she draws from topics to complete the section.

MetaHeuristica Data

Interpretable topic

Topic Details



Top Terms

Term	Cnt	Prob	Cohrn	Uni-Dist
loans	7,821	0.090	0.000	0.639
real	4,659	0.054	-0.516	0.353
estate	4,502	0.052	-0.548	0.339
commercial	3,066	0.035	-0.837	0.217
loan	3,004	0.035	-0.766	0.212
residential	1,579	0.018	-1.235	0.100
portfolio	1,294	0.015	-1.260	0.079
construction	1,132	0.013	-1.726	0.067
lending	1,031	0.012	-1.737	0.060
risk	1,024	0.012	-1.548	0.060
collateral	1,021	0.012	-1.782	0.059
mortgage	1,014	0.012	-1.785	0.059
secured	959	0.011	-1.502	0.055
family	878	0.010	-2.108	0.050
value	831	0.010	-1.993	0.046
borrower	803	0.009	-2.012	0.044
borrowers	741	0.009	-1.969	0.040
business	737	0.009	-1.890	0.040
consumer	728	0.008	-2.535	0.040
generally	722	0.008	-1.753	0.039

Top Phrases

Phrase	Cnt	Prob
real estate	564	0.0556
loan portfolio	80	0.0079
real estate loans	67	0.0066
commercial real estate	56	0.0055
commercial loans	45	0.0044
secured real estate	40	0.0039
mortgage loans	39	0.0038
residential real estate	38	0.0037
commercial real estate loans	36	0.0036
real estate values	35	0.0035
consumer loans	30	0.0030
credit risk	19	0.0019
december approximately	19	0.0019
residential mortgage	19	0.0019
commercial loan	18	0.0018
real estate market	17	0.0017
estate loans	17	0.0017
residential real estate loans	17	0.0017

Documents 10

Display 10 of 4,779 Page 1 of 478

10K/2006-09-30/2006-12-21/HarleysvilleSavingsFinancialCorp/0001107160/RiskFactors/Preface/paragraph/1.3.1.8/ 99.65%

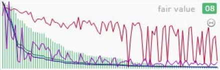
Our lending activities include loans secured by existing multi-family residential and commercial real estate. In addition, from time to time we originate loans for the construction of multi-family residential real estate and land acquisition and development loans. Multi-family residential, commercial real estate and construction lending generally is considered to involve a higher degree of risk than single-family residential lending due to a variety of factors, including generally larger loan balances, the dependency on successful completion or operation of the project for repayment, the difficulties in estimating construction costs and loan terms which often do not require full amortization of the loan over its term and, instead, provide for a balloon payment at stated maturity. Our lending activities also include commercial business loans and leases to small to medium businesses, which generally are secured by various equipment, machinery and other corporate assets, and a wide variety of consumer loans, including home improvement loans, home equity loans, education loans and loans secured by automobiles, boats, mobile homes, recreational vehicles and other personal property. Although commercial business loans and leases and consumer loans generally have shorter terms and higher interests rates than mortgage loans, they generally involve more risk than mortgage loans because of the nature of, or in certain cases the absence of, the collateral which secures such loans.

10K/2006-12-31/2007-03-12/EsbFinancialCorp/0000872835/RiskFactors/Preface/paragraph/1.3.1.7/ 99.65%

Our lending activities include loans secured by existing multi-family residential and commercial real estate. In addition, from time to time we originate loans for the construction of multi-family residential real estate and land acquisition and development loans. Multi-family residential, commercial real estate and construction lending generally is considered to involve a higher degree of risk than single-family residential lending due to a variety of

Less interpretable topic

Topic Details



Top Terms

Term	Cnt	Prob	Cohrn	Uni-Dist
rate	3,460	0.040	0.000	0.254
value	3,249	0.038	-0.924	0.236
interest	2,925	0.034	-0.970	0.209
fair	2,639	0.031	-1.113	0.185
risk	2,567	0.030	-1.331	0.179
market	2,139	0.025	-1.069	0.145
instruments	1,796	0.021	-1.140	0.118
financial	1,354	0.016	-1.310	0.084
company	1,288	0.015	-1.634	0.080
derivative	1,275	0.015	-1.757	0.079
rates	1,153	0.013	-1.584	0.070
changes	1,041	0.012	-1.432	0.062
cash	979	0.011	-1.969	0.057
not	962	0.011	-1.585	0.056
hedge	805	0.009	-2.628	0.045
foreign	785	0.009	-3.120	0.044
exchange	773	0.009	-2.644	0.043
values	743	0.009	-2.608	0.041
swaps	732	0.009	-2.735	0.040
derivatives	730	0.009	-2.311	0.040

Top Phrases

Phrase	Cnt	Prob
fair value	254	0.0209
interest rate	96	0.0079
rate risk	68	0.0056
financial instruments	65	0.0053
fair values	59	0.0048
cash flows	52	0.0043
market risk	52	0.0043
fixed rate	49	0.0040
interest rates	46	0.0038
market value	46	0.0038
variable rate	39	0.0032
derivative instruments	37	0.0030
interest rate risk	34	0.0028
rate swaps	26	0.0021
foreign exchange	23	0.0019
foreign currency exchange	20	0.0016
balance sheet	20	0.0016
does not	18	0.0015

Documents 08

Display 10 of 4,952 Page 1 of 496

10K/2006-12-31/2007-03-16/UnionBanksharesCorp/0000883948/NotesToTheConsolidatedFinancialStatements/FairValueOfFinancialInstrumentsAndInterestRateRisk/paragraph/1.13.23.1/ 99.48%

The fair value of a financial instrument is the current amount that would be exchanged between willing parties, other than in a forced liquidation. Fair value is best determined based on quoted market prices. However, in many instances, there are no quoted market prices for the Company's various financial instruments. In cases where quoted market prices are not available, fair values are based on estimates using present value or other valuation techniques. Those techniques are significantly affected by the assumptions used, including the discount rate and estimates of future cash flows. Accordingly, the fair value estimates may not be realized in an immediate settlement of the instruments. SFAS No. 107, Disclosures about Fair Value of Financial Instruments ("SFAS No. 107"), excludes certain financial instruments and all non-financial instruments from its disclosure requirements. Accordingly, the aggregate fair value amounts presented may not necessarily represent the underlying fair value of the Company.

10K/2006-12-31/2007-04-02/CommunityBankSharesOfIndianaInc/0000933590/QuantitativeAndQualitativeDisclosuresAboutMarketRisk/Derivatives/paragraph/1.11.47.4/ 99.46%

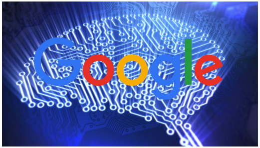
The Company formally documents the relationship between derivatives and hedged items, as well as the risk-management objective and the strategy for undertaking hedge transactions. This documentation includes linking fair value or cash flow hedges to specific assets and liabilities on the balance sheet or to specific firm commitments or forecasted transactions. The Company also formally assesses, both at the hedge's inception and on an ongoing basis, whether the derivative instruments that are used are highly effective in offsetting changes in fair values or cash flows of the hedged items. The Company discontinues hedge accounting when it determines that the derivative is no longer effective in offsetting changes in the fair value or cash flows of the hedged item, the derivative is settled or terminates, a hedged forecasted transaction is no longer probable, a

LDA limitations

- Not always interpretable
- Time-series variation in topics makes comparison difficult

Use "Semantic Vector Analysis" in second stage

- See Mikolov, Chen, Corrado, and Dean (2013) and Mikolov, Sutskever, Chen, Corrado, and Dean (2013)
- Distributional semantics: "word is characterized by the company it keeps" Firth (1957)
- Position of word matters



Semantic Vector Analysis (SVA)

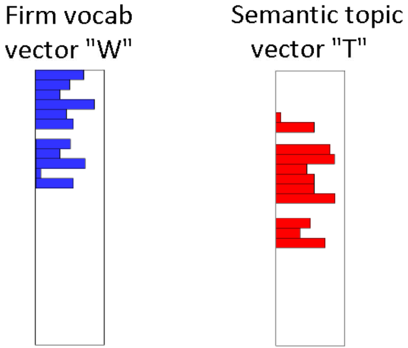
Two stages

1. All 10-Ks are loaded and distributional information about proximity of each word to other words is determined
   - Uses a two layer neural network to
     - Predict a single word given its immediate surrounding words
     - Predict words surrounding a single word
2. Input any word or commongram and the application returns a vector of words with weights indicating importance that best describe that token

Semantic theme content

	Real Estate		Deposits	
Row	Word	Cosine Dist	Word	Cosine Dist
1	real	0.7875	deposits	1
2	estate	0.7875	deposit	0.7046
3	foreclosure	0.4898	brokered deposits	0.593
4	property	0.4619	cdars	0.5864
5	personal	0.4563	account registry	0.5712
6	physical possession	0.4539	brokered certificates	0.568
7	foreclosed real	0.4503	bearing checking	0.5657
8	foreclosed	0.4423	bearing deposits	0.565
9	deed	0.4323	certificates	0.5632
10	beneficiary	0.4283	negotiable order	0.5154
11	real estate	0.4262	promontory interfinancial	0.5129
12	possession	0.4147	cdars program	0.5067
13	oreo	0.4063	sweep ics	0.495
14	lien	0.4044	brokered	0.4943
15	securing	0.4039	withdrawal	0.4804
16	h2c	0.4014	overdrafts	0.4738
17	owned	0.3996	sweep accounts	0.4726
18	repossessed	0.3981	bearing	0.4591
19	death	0.3974	cdars network	0.4547
20	owner	0.3949	fdic insured	0.4505

Mapping semantic themes to bank-years



Firm i's loading on semantic theme k is thus the cosine similarity $S_{i,k,t}$:

$$S_{i,k,t} = \frac{W_{i,t}}{||W_{i,t}||} \cdot \frac{T_{k,t}}{||T_{k,t}||}$$

Result: A firm-year panel database of semantic theme loadings.

Emerging risk model

$$Covariance_{i,j,t} = \alpha_0 + \gamma \mathbf{X_{i,j,t}} + \varepsilon_{i,j,t}, \tag{1}$$

$$Covariance_{i,j,t} = \alpha_0 + \beta_1 S_{i,j,t,1} + \beta_2 S_{i,j,t,2} + \beta_3 S_{i,j,t,3} + ... + \beta_T S_{i,j,t,31}$$
$$+ \gamma \mathbf{X_{i,j,t}} + \varepsilon_{i,j,t}, \tag{2}$$

Aggregate risk score

- Take difference in R^2 from Eq. (1) and (2)
- Scale differential R^2 using its mean and standard deviation from baseline period to get t-statistic in each quarter
- Elevated t-statistic indicates importance of risk themes and hence, emerging risk

Data sources

- CRSP (stock returns), Compustat (accounting variables)
- FDIC Failures and Assistance Transactions List
- VIX data.
- Call Reports for bank-specific characteristics
- metaHeuristica used to extract risk factor discussions from bank 10-Ks from 1997 to 2014
- Include banks defined as having SIC codes from 6000 to 6199
- Require machine readable 10-K, with some non-empty discussion of risk factors

Static risk method

Determining static themes

Examine LDA output and feed prevalent (most frequent) key phrases (tokens) from LDA to SVA

- These are high-level risk factors that remain constant over time
- Remove any boilerplate such as “balance sheet” or “million December”
- Group the remaining individual terms into broad categories of risks fundamental to the banking sector aided by a review of the literature e.g. “Credit Card” or “Regulatory Capital”
- For our static model, we choose 61 initial semantic themes upon reviewing the LDA output for key phrases and reduce this to 31 themes due to multicollinearity

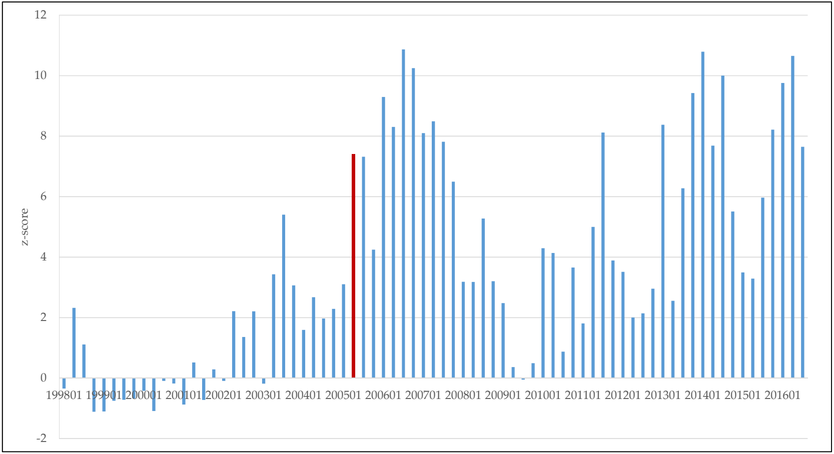
Static semantic themes

- Accounting
- Cash
- Certificate Deposit
- Commercial Paper
- Compensation
- Competition
- Counterparty
- Credit Card
- Currency Exchange
- Data Security
- Deposits
- Derivative
- Dividends
- Fees
- Funding Sources
- Governance
- Growth Strategy
- Insurance
- Internal Controls
- Lawsuit
- Mergers Acquisitions
- Off Balance Sheet
- Operational Risk
- Prepayment
- Rating Agency
- Real Estate
- Regulatory Capital
- Reputation
- Securitization
- Student Loans
- Taxes

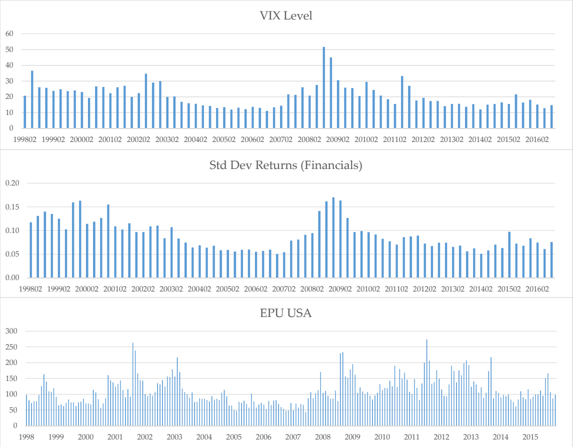
Aggregate risk metric

- Run regression once per quarter with one observation bank-pair (*i* and *j*).
- Dependent variable is quarterly return covariance of bank *i* and *j* measured using daily returns
- Semantic theme of pair is the product $S_{i,j} = S_i \, S_j$
- X is a set of pairwise controls including size, age, profitability, leverage, and industry controls
- Aggregate risk score is the contribution of SVA themes to R^2

Aggregate emerging risk score



Other emerging risk metrics

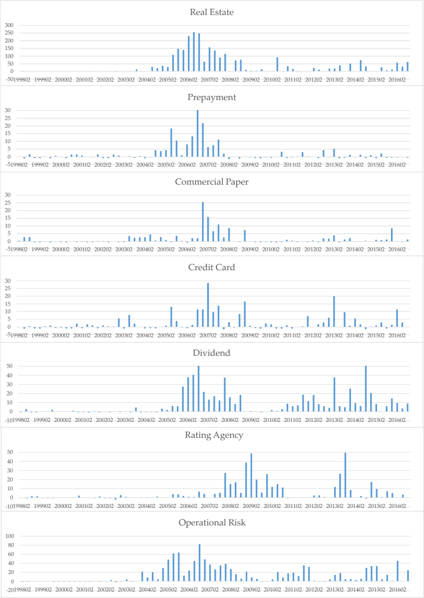


Identifying individual risks

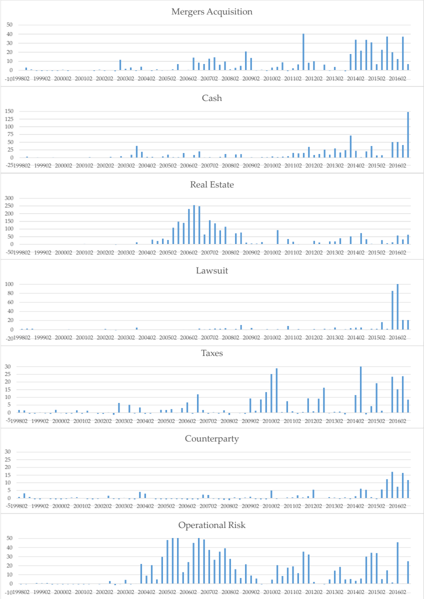
- Use each of 31 semantic themes from SVA
- We compute the individual contribution to R^2 of each theme in explaining pairwise return covariance in each quarter
- Standardize each marginal R^2 by its mean and standard deviation from the baseline period 1998 to 2003
- Resulting t-statistics illustrate how strong each individual risk factor is in explaining comovement
- Importantly, individual risk factors are interpretable

This has important ramifications both for understanding the crisis and monitoring emerging risk in the current period.

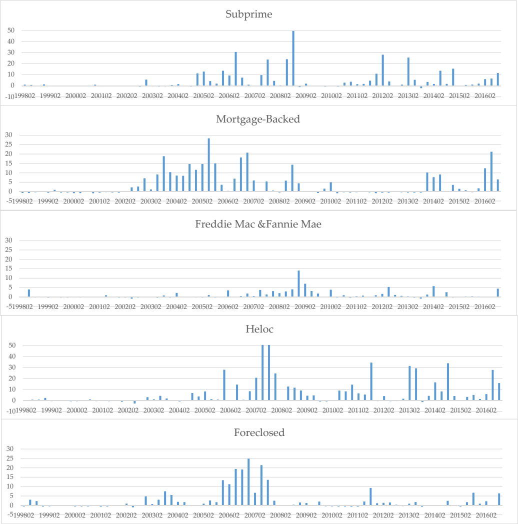
2008 major risks



2015 major risks



Drill-down model: Real estate



Dynamic methodology

- Extract top 25 terms from each of the 25 LDA topics per year (625 possible topics per year)
- Limit to bigrams (400 possible topics per year)
- Remove boilerplate (150 possible topics per year)
- Use covariance model and stepwise regression to maximize R^2
- Baseline R^2 measured using four year moving window of adjusted R^2 ending in the year being tested

Dynamic emerging risks

Emerging Risk	Year	Emerging Risk	Year
related litigation	200401	economic downturn	201103
deposits borrowings	200401	education loans	201103
mortgage banking	200403	identity theft	201103
operational risk	200403	customer deposits	201104
charged off	200403	secondary mortgage	201201
origination fees	200404	deposit insurance	201202
backed securities	200404	foreclosure process	201202
off balance	200502	commercial real	201203
rate environment	200502	operational risk	201204
real estate	200503	trust preferred	201302
rate swap	200504	extend credit	201302
recruiting hiring	200601	weather events	201303
board directors	200602	executive compensation	201303
interest bearing	200602	supervision regulation	201304
underwriting standards	200603	regulatory requirements	201304
time deposits	200604	basel iii	201401
brokered deposits	200604	negative publicity	201402
investment securities	200604	supervision regulation	201402
senior notes	200701	capital levels	201403
board directors	200702	regulatory authorities	201403
prevent fraud	200703	brokered deposits	201404
damage reputation	200704	senior management	201501
extend credit	200704	legal proceedings	201601
cost funds	200801	servicing rights	201601
rate risk	200802	institution failures	201601
real property	200803	merger agreement	201603
legal proceedings	200804	credit risk	201603
mergers acquisitions	200901	data processing	201604

Individual bank exposure to emerging risk

Create *Emerging Risk Exposure* as average quarterly predicted covariance bank *i* has with all other banks *j* using the main covariance model in Equation (2)

Uses the following procedure:

1. Take product of fitted coefficients for each SVA theme (β_1 to β_{31}) from the baseline covariance model and multiply by the given bank-pair's SVA theme loading
2. Sum the resulting 31 products for each bank-pair to get the total predicted covariance of bank *i* with each bank *j*
3. Average predicted covariances over banks *j* to get the total *Emerging Risk Exposure* for bank *i* in quarter *t*

Cross-sectional tests using static model

- In each quarter, run ***single*** cross sectional regression
- Dependent variable is one of the following:
  - Bank's stock return from 9/2008 to 12/2012
  - Bank's stock return from 12/2015 to 2/2016
  - Dummy variable indicating whether the given bank failed in the 3 year period beginning with the Lehman bankruptcy
- Also run monthly Fama-McBeth regressions where dependent variable is the ex post monthly stock return volatility computed using daily stock returns.
- Main independent variable of interest is *Emerging Risk Exposure*

Predicting post-2008 crisis returns (9/2008-12/2012)

Row	Quarter	Emerging Risk Exposure	# Obs	Predictive Timing
(1)	2004 1Q	2.410 (2.16)	352	Predictive
(2)	2004 2Q	2.489 (3.69)	352	Predictive
(3)	2004 3Q	0.319 (0.18)	368	Predictive
(4)	2004 4Q	0.415 (0.28)	368	Predictive
(5)	2005 1Q	-0.670 (-0.31)	388	Predictive
(6)	2005 2Q	-0.519 (-0.28)	388	Predictive
(7)	2005 3Q	-1.006 (-0.36)	418	Predictive
(8)	2005 4Q	1.147 (0.40)	418	Predictive
(9)	2006 1Q	0.918 (0.65)	407	Predictive
(10)	2006 2Q	-2.462 (-1.44)	407	Predictive
(11)	2006 3Q	-2.656 (-1.06)	430	Predictive
(12)	2006 4Q	-3.374 (-1.09)	430	Predictive
(13)	2007 1Q	-4.268 (-2.01)	444	Predictive
(14)	2007 2Q	-3.436 (-2.01)	444	Predictive
(15)	2007 3Q	-3.908 (-3.04)	469	Predictive
(16)	2007 4Q	-3.406 (-3.27)	469	Predictive
(17)	2008 1Q	-3.970 (-3.65)	468	Predictive
(18)	2008 2Q	-4.943 (-7.80)	468	Predictive
(19)	2008 3Q	-3.113 (-2.21)	489	Non Predictive
(20)	2008 4Q	-1.778 (-1.02)	491	Non Predictive
(21)	2009 1Q	-1.823 (-1.15)	518	Non Predictive
(22)	2009 2Q	-2.471 (-1.55)	518	Non Predictive
(23)	2009 3Q	-2.942 (-9.97)	529	Non Predictive
(24)	2009 4Q	-2.107 (-2.88)	522	Non Predictive

Predicting current period returns (12/2015-2/2016)

Row	Quarter	Emerging Risk Exposure	# Obs	Predictive Timing
(1)	2010 1Q	-0.928 (-3.25)	334	Predictive
(2)	2010 2Q	-0.657 (-3.27)	334	Predictive
(3)	2010 3Q	-0.738 (-4.44)	341	Predictive
(4)	2010 4Q	-0.282 (-1.53)	341	Predictive
(5)	2011 1Q	-0.746 (-3.33)	351	Predictive
(6)	2011 2Q	-0.758 (-4.22)	350	Predictive
(7)	2011 3Q	-0.941 (-11.7)	356	Predictive
(8)	2011 4Q	-0.671 (-4.30)	356	Predictive
(9)	2012 1Q	-0.778 (-2.40)	349	Predictive
(10)	2012 2Q	-0.660 (-1.40)	349	Predictive
(11)	2012 3Q	-0.916 (-3.73)	360	Predictive
(12)	2012 4Q	-0.798 (-1.77)	360	Predictive
(13)	2013 1Q	-0.121 (-1.45)	351	Predictive
(14)	2013 2Q	-0.228 (-1.92)	351	Predictive
(15)	2013 3Q	0.198 (0.95)	368	Predictive
(16)	2013 4Q	-0.375 (-2.54)	368	Predictive
(17)	2014 1Q	-0.024 (-0.17)	356	Predictive
(18)	2014 2Q	-0.222 (-3.00)	356	Predictive
(19)	2014 3Q	-0.832 (-2.42)	367	Predictive
(20)	2014 4Q	-0.681 (-2.30)	367	Predictive
(21)	2015 1Q	-0.440 (-1.53)	358	Predictive
(22)	2015 2Q	-0.505 (-1.47)	358	Predictive
(23)	2015 3Q	-1.015 (-2.33)	387	Predictive
(24)	2015 4Q	-0.500 (-1.49)	386	Non Predictive

Predicting bank failures

Quarter	Emerging Risk Exposure s	Obs	Predictive Timing
2004 1Q	0.004 (0.80)	625	Predictive
2004 2Q	0.004 (0.94)	625	Predictive
2004 3Q	-0.005 (-1.03)	625	Predictive
2004 4Q	-0.004 (-0.79)	625	Predictive
2005 1Q	-0.002 (-1.33)	615	Predictive
2005 2Q	-0.001 (-1.36)	615	Predictive
2005 3Q	0.008 (3.56)	615	Predictive
2005 4Q	0.006 (2.55)	615	Predictive
2006 1Q	-0.002 (-0.14)	578	Predictive
2006 2Q	-0.001 (-0.08)	578	Predictive
2006 3Q	0.003 (0.58)	578	Predictive
2006 4Q	0.008 (3.97)	578	Predictive
2007 1Q	0.009 (3.96)	588	Predictive
2007 2Q	0.011 (7.36)	588	Predictive
2007 3Q	0.010 (2.31)	588	Predictive
2007 4Q	0.014 (4.37)	588	Predictive
2008 1Q	0.014 (4.42)	562	Predictive
2008 2Q	0.015 (3.89)	562	Predictive
2008 3Q	0.015 (3.72)	562	Predictive
2008 4Q	0.004 (0.63)	562	Non Predictive
2009 1Q	0.024 (8.54)	564	Non Predictive
2009 2Q	0.010 (3.87)	564	Non Predictive
2009 3Q	-0.001 (-0.27)	564	Non Predictive
2009 4Q	0.007 (1.96)	564	Non Predictive

Unconditional Fama-MacBeth volatility regressions

Lag	1 Quarter Exposure	2 Quarter Exposure	3 Quarter Exposure	Obs.
1	0.086 (8.94)	0.105 (10.26)	0.112 (11.35)	52641
2	0.084 (8.72)	0.104 (10.22)	0.108 (11.13)	52476
3	0.086 (9.18)	0.099 (10.53)	0.104 (11.38)	52312
4	0.086 (9.13)	0.098 (10.81)	0.102 (11.43)	52148
5	0.085 (9.13)	0.093 (10.42)	0.097 (11.32)	51786
6	0.079 (8.96)	0.088 (10.40)	0.088 (11.09)	51410
7	0.076 (9.52)	0.083 (10.66)	0.081 (10.52)	51035
8	0.069 (8.66)	0.077 (10.04)	0.074 (9.60)	50660
9	0.064 (8.59)	0.069 (9.39)	0.071 (9.09)	50284
10	0.062 (8.65)	0.064 (8.62)	0.066 (8.82)	49908
11	0.058 (8.38)	0.060 (8.28)	0.063 (8.51)	49569
12	0.053 (7.51)	0.057 (7.74)	0.060 (8.06)	49230
13	0.045 (6.84)	0.049 (7.40)	0.054 (7.43)	48891
14	0.041 (6.29)	0.046 (6.79)	0.051 (6.95)	48541
15	0.037 (5.81)	0.044 (6.49)	0.047 (6.56)	48191
16	0.032 (5.09)	0.040 (5.54)	0.043 (5.83)	47841
17	0.031 (4.63)	0.040 (5.40)	0.042 (5.61)	47490
18	0.032 (4.73)	0.039 (5.25)	0.042 (5.60)	47139
19	0.030 (4.02)	0.036 (4.73)	0.041 (5.27)	46788
20	0.033 (4.62)	0.036 (5.00)	0.041 (5.30)	46438
21	0.029 (4.26)	0.035 (4.99)	0.039 (5.12)	46088
22	0.028 (4.16)	0.036 (5.24)	0.039 (5.25)	45738
23	0.024 (3.80)	0.034 (4.68)	0.036 (4.86)	45404
24	0.028 (4.23)	0.034 (4.59)	0.035 (4.72)	45071
25	0.030 (4.24)	0.035 (4.34)	0.035 (4.50)	44738
26	0.028 (3.60)	0.031 (3.80)	0.033 (4.14)	44397
27	0.027 (3.43)	0.029 (3.65)	0.033 (4.10)	44056
28	0.027 (3.36)	0.030 (3.85)	0.033 (4.20)	43716
29	0.025 (3.17)	0.030 (3.95)	0.034 (4.46)	43376
30	0.021 (2.65)	0.027 (3.53)	0.029 (3.78)	43035
31	0.019 (2.61)	0.024 (3.19)	0.026 (3.46)	42694

Conclusions

- We propose a model of emerging risks in the financial sector based on computational linguistic analysis of firm disclosures and return covariances
- Method is flexible, dynamic, timely, allowing the prediction of **interpretable** emerging risks for which a researcher might not even be aware
- Allows for high-level (aggregate) to granular level (theme and bank) determination of risk build-up
- Can be used by researchers and regulators alike to monitor threats to financial stability