Dynamic Interpretation of Emerging Risks in the Financial Sector

PRESENTER
Kathleen Weiss Hanley, Lehigh University

Joint work with Gerard Hoberg, University of Southern California
• 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.

Hanley and Hoberg (2018)  
Jacobs Levy Equity Management Center Conference
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
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
Corpus of 10-K Bank Risk Factors

10-K

Item 1: Business Description

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).

Item 7: MD&A

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
Risk Factor Document Creation

Management

Hmmm

Mortgage Risk

Capital Requirements

Deposit Risk

Sources of Funding

Mortgages, interest only, subprime borrower, no doc loans

* 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.
Interpretable topic

Hanley and Hoberg (2018)

Jacobs Levy Equity Management Center Conference
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 instrument.

SFAS No. 107, **Disclosures about Fair Value of Financial Instruments**, 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.

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 the fair value of cash flows hedged 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 of 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 of or cash flows of the hedged item. The derivative is settled or terminates, a hedged forecasted transaction is no longer probable, or
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
<table>
<thead>
<tr>
<th>Row</th>
<th>Word</th>
<th>Cosine Dist</th>
<th>Word</th>
<th>Cosine Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>real</td>
<td>0.7875</td>
<td>deposits</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>estate</td>
<td>0.7875</td>
<td>deposit</td>
<td>0.7046</td>
</tr>
<tr>
<td>3</td>
<td>foreclosure</td>
<td>0.4898</td>
<td>brokered deposits</td>
<td>0.593</td>
</tr>
<tr>
<td>4</td>
<td>property</td>
<td>0.4619</td>
<td>cdars</td>
<td>0.5864</td>
</tr>
<tr>
<td>5</td>
<td>personal</td>
<td>0.4563</td>
<td>account registry</td>
<td>0.5712</td>
</tr>
<tr>
<td>6</td>
<td>physical possession</td>
<td>0.4539</td>
<td>brokered certificates</td>
<td>0.568</td>
</tr>
<tr>
<td>7</td>
<td>foreclosed real</td>
<td>0.4503</td>
<td>bearing checking</td>
<td>0.5657</td>
</tr>
<tr>
<td>8</td>
<td>foreclosed</td>
<td>0.4423</td>
<td>bearing deposits</td>
<td>0.565</td>
</tr>
<tr>
<td>9</td>
<td>deed</td>
<td>0.4323</td>
<td>certificates</td>
<td>0.5632</td>
</tr>
<tr>
<td>10</td>
<td>beneficiary</td>
<td>0.4283</td>
<td>negotiable order</td>
<td>0.5154</td>
</tr>
<tr>
<td>11</td>
<td>real estate</td>
<td>0.4262</td>
<td>promontory interfinancial</td>
<td>0.5129</td>
</tr>
<tr>
<td>12</td>
<td>possession</td>
<td>0.4147</td>
<td>cdars program</td>
<td>0.5067</td>
</tr>
<tr>
<td>13</td>
<td>oreo</td>
<td>0.4063</td>
<td>sweep ics</td>
<td>0.495</td>
</tr>
<tr>
<td>14</td>
<td>lien</td>
<td>0.4044</td>
<td>brokered</td>
<td>0.4943</td>
</tr>
<tr>
<td>15</td>
<td>securing</td>
<td>0.4039</td>
<td>withdrawal</td>
<td>0.4804</td>
</tr>
<tr>
<td>16</td>
<td>h2c</td>
<td>0.4014</td>
<td>overdrafts</td>
<td>0.4738</td>
</tr>
<tr>
<td>17</td>
<td>owned</td>
<td>0.3996</td>
<td>sweep accounts</td>
<td>0.4726</td>
</tr>
<tr>
<td>18</td>
<td>repossessed</td>
<td>0.3981</td>
<td>bearing</td>
<td>0.4591</td>
</tr>
<tr>
<td>19</td>
<td>death</td>
<td>0.3974</td>
<td>cdars network</td>
<td>0.4547</td>
</tr>
<tr>
<td>20</td>
<td>owner</td>
<td>0.3949</td>
<td>fdic insured</td>
<td>0.4505</td>
</tr>
</tbody>
</table>
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

\[ \text{Covariance}_{i,j,t} = \alpha_0 + \gamma X_{i,j,t} + \varepsilon_{i,j,t}, \]  

(1)

\[ \text{Covariance}_{i,j,t} = \alpha_0 + \beta_1 S_{i,j,t,1} + \beta_2 S_{i,j,t,2} + \beta_3 S_{i,j,t,3} + \ldots + \beta_T S_{i,j,t,31} + \gamma X_{i,j,t} + \varepsilon_{i,j,t}, \]  

(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

Hanley and Hoberg (2018) Jacobs Levy Equity Management Center Conference
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
Run regression once per quarter with one observation bank-pair \((i \text{ and } j)\).
Dependent variable is quarterly return covariance of bank \(i\) and \(j\) measured using daily returns
Semantic theme of pair is the product \(S_{i,j} = S_i \cdot S_j\)
\(X\) 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

Hanley and Hoberg (2018)
Other emerging risk metrics

VIX Level

Std Dev Returns (Financials)

EPU USA

Hanley and Hoberg (2018)  Jacobs Levy Equity Management Center Conference
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

---

Hanley and Hoberg (2018) Jacobs Levy Equity Management Center Conference
2015 major risks

Mergers Acquisition

Cash

Real Estate

Lawsuit

Taxes

Counterparty

Operational Risk

Hanley and Hoberg (2018) Jacobs Levy Equity Management Center Conference
Drill-down model: Real estate

Hanley and Hoberg (2018)
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

<table>
<thead>
<tr>
<th>Emerging Risk</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>related litigation</td>
<td>200401</td>
</tr>
<tr>
<td>deposits borrowings</td>
<td>200401</td>
</tr>
<tr>
<td>mortgage banking</td>
<td>200403</td>
</tr>
<tr>
<td>operational risk</td>
<td>200403</td>
</tr>
<tr>
<td>charged off</td>
<td>200403</td>
</tr>
<tr>
<td>origination fees</td>
<td>200404</td>
</tr>
<tr>
<td>backed securities</td>
<td>200404</td>
</tr>
<tr>
<td>off balance</td>
<td>200502</td>
</tr>
<tr>
<td>rate environment</td>
<td>200502</td>
</tr>
<tr>
<td>real estate</td>
<td>200503</td>
</tr>
<tr>
<td>rate swap</td>
<td>200504</td>
</tr>
<tr>
<td>recruiting hiring</td>
<td>200601</td>
</tr>
<tr>
<td>board directors</td>
<td>200602</td>
</tr>
<tr>
<td>interest bearing</td>
<td>200602</td>
</tr>
<tr>
<td>underwriting standards</td>
<td>200603</td>
</tr>
<tr>
<td>time deposits</td>
<td>200604</td>
</tr>
<tr>
<td>brokered deposits</td>
<td>200604</td>
</tr>
<tr>
<td>investment securities</td>
<td>200604</td>
</tr>
<tr>
<td>senior notes</td>
<td>200701</td>
</tr>
<tr>
<td>board directors</td>
<td>200702</td>
</tr>
<tr>
<td>prevent fraud</td>
<td>200703</td>
</tr>
<tr>
<td>damage reputation</td>
<td>200704</td>
</tr>
<tr>
<td>extend credit</td>
<td>200704</td>
</tr>
<tr>
<td>cost funds</td>
<td>200801</td>
</tr>
<tr>
<td>rate risk</td>
<td>200802</td>
</tr>
<tr>
<td>real property</td>
<td>200803</td>
</tr>
<tr>
<td>legal proceedings</td>
<td>200804</td>
</tr>
<tr>
<td>mergers acquisitions</td>
<td>200901</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emerging Risk</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>economic downturn</td>
<td>201103</td>
</tr>
<tr>
<td>education loans</td>
<td>201103</td>
</tr>
<tr>
<td>identity theft</td>
<td>201104</td>
</tr>
<tr>
<td>secondary mortgage</td>
<td>201201</td>
</tr>
<tr>
<td>deposit insurance</td>
<td>201202</td>
</tr>
<tr>
<td>foreclosure process</td>
<td>201202</td>
</tr>
<tr>
<td>operational risk</td>
<td>201204</td>
</tr>
<tr>
<td>trust preferred</td>
<td>201302</td>
</tr>
<tr>
<td>extend credit</td>
<td>201302</td>
</tr>
<tr>
<td>executive compensation</td>
<td>201303</td>
</tr>
<tr>
<td>supervision regulation</td>
<td>201304</td>
</tr>
<tr>
<td>regulatory requirements</td>
<td>201304</td>
</tr>
<tr>
<td>basel iii</td>
<td>201401</td>
</tr>
<tr>
<td>negative publicity</td>
<td>201402</td>
</tr>
<tr>
<td>supervision regulation</td>
<td>201402</td>
</tr>
<tr>
<td>capital levels</td>
<td>201403</td>
</tr>
<tr>
<td>regulatory authorities</td>
<td>201403</td>
</tr>
<tr>
<td>brokered deposits</td>
<td>201404</td>
</tr>
<tr>
<td>senior management</td>
<td>201501</td>
</tr>
<tr>
<td>legal proceedings</td>
<td>201601</td>
</tr>
<tr>
<td>servicing rights</td>
<td>201601</td>
</tr>
<tr>
<td>institution failures</td>
<td>201601</td>
</tr>
<tr>
<td>merger agreement</td>
<td>201603</td>
</tr>
<tr>
<td>credit risk</td>
<td>201603</td>
</tr>
<tr>
<td>data processing</td>
<td>201604</td>
</tr>
</tbody>
</table>

Hanley and Hoberg (2018) | Jacobs Levy Equity Management Center Conference
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 ($\beta_1$ to $\beta_{31}$) 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 $t$
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)

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

*Hanley and Hoberg (2018) Jacobs Levy Equity Management Center Conference*
<table>
<thead>
<tr>
<th>Row</th>
<th>Quarter</th>
<th>Emerging Risk Exposure</th>
<th># Obs</th>
<th>Predictive Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>2010 1Q</td>
<td>-0.928 (-3.25)</td>
<td>334</td>
<td>Predictive</td>
</tr>
<tr>
<td>(2)</td>
<td>2010 2Q</td>
<td>-0.657 (-3.27)</td>
<td>334</td>
<td>Predictive</td>
</tr>
<tr>
<td>(3)</td>
<td>2010 3Q</td>
<td>-0.738 (-4.44)</td>
<td>341</td>
<td>Predictive</td>
</tr>
<tr>
<td>(4)</td>
<td>2010 4Q</td>
<td>-0.282 (-1.53)</td>
<td>341</td>
<td>Predictive</td>
</tr>
<tr>
<td>(5)</td>
<td>2011 1Q</td>
<td>-0.746 (-3.33)</td>
<td>351</td>
<td>Predictive</td>
</tr>
<tr>
<td>(6)</td>
<td>2011 2Q</td>
<td>-0.758 (-4.22)</td>
<td>350</td>
<td>Predictive</td>
</tr>
<tr>
<td>(7)</td>
<td>2011 3Q</td>
<td>-0.941 (-11.7)</td>
<td>356</td>
<td>Predictive</td>
</tr>
<tr>
<td>(8)</td>
<td>2011 4Q</td>
<td>-0.671 (-4.30)</td>
<td>356</td>
<td>Predictive</td>
</tr>
<tr>
<td>(9)</td>
<td>2012 1Q</td>
<td>-0.778 (-2.40)</td>
<td>349</td>
<td>Predictive</td>
</tr>
<tr>
<td>(10)</td>
<td>2012 2Q</td>
<td>-0.660 (-1.40)</td>
<td>349</td>
<td>Predictive</td>
</tr>
<tr>
<td>(11)</td>
<td>2012 3Q</td>
<td>-0.916 (-3.73)</td>
<td>360</td>
<td>Predictive</td>
</tr>
<tr>
<td>(12)</td>
<td>2012 4Q</td>
<td>-0.798 (-1.77)</td>
<td>360</td>
<td>Predictive</td>
</tr>
<tr>
<td>(13)</td>
<td>2013 1Q</td>
<td>-0.121 (-1.45)</td>
<td>351</td>
<td>Predictive</td>
</tr>
<tr>
<td>(14)</td>
<td>2013 2Q</td>
<td>-0.228 (-1.92)</td>
<td>351</td>
<td>Predictive</td>
</tr>
<tr>
<td>(15)</td>
<td>2013 3Q</td>
<td>0.198 (0.95)</td>
<td>368</td>
<td>Predictive</td>
</tr>
<tr>
<td>(16)</td>
<td>2013 4Q</td>
<td>-0.375 (-2.54)</td>
<td>368</td>
<td>Predictive</td>
</tr>
<tr>
<td>(17)</td>
<td>2014 1Q</td>
<td>-0.024 (-0.17)</td>
<td>356</td>
<td>Predictive</td>
</tr>
<tr>
<td>(18)</td>
<td>2014 2Q</td>
<td>-0.222 (-3.00)</td>
<td>356</td>
<td>Predictive</td>
</tr>
<tr>
<td>(19)</td>
<td>2014 3Q</td>
<td>-0.832 (-2.42)</td>
<td>367</td>
<td>Predictive</td>
</tr>
<tr>
<td>(20)</td>
<td>2014 4Q</td>
<td>-0.681 (-2.30)</td>
<td>367</td>
<td>Predictive</td>
</tr>
<tr>
<td>(21)</td>
<td>2015 1Q</td>
<td>-0.440 (-1.53)</td>
<td>358</td>
<td>Predictive</td>
</tr>
<tr>
<td>(22)</td>
<td>2015 2Q</td>
<td>-0.505 (-1.47)</td>
<td>358</td>
<td>Predictive</td>
</tr>
<tr>
<td>(23)</td>
<td>2015 3Q</td>
<td>-1.015 (-2.33)</td>
<td>387</td>
<td>Predictive</td>
</tr>
<tr>
<td>(24)</td>
<td>2015 4Q</td>
<td>-0.500 (-1.49)</td>
<td>386</td>
<td>Non Predictive</td>
</tr>
</tbody>
</table>

Hanley and Hoberg (2018) Jacobs Levy Equity Management Center Conference
Predicting bank failures

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

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

Hanley and Hoberg (2018)