

The Language of Risk: Narrative Consistency and Transparency in Banks

Marialena Togia

Athens University of Economics and Business (AUEB)
Department of International and European Economic Studies (DIEES)

June 29, 2026

- Annual 10-K bank filings over 2003-2024 contain quantitative information (numbers) and qualitative information (words)
- Words in $t + 1$ are ex post narrative about data in t
- What do banks talk about after good data vs after bad data?
- Do the words correspond to the data (transparency)?

- Algorithm design for data extraction and panel creation from 10-Ks, when reporting, tables, and variables differ over time and across banks
- Creation of lexicons addressing the problem that banks change the words they use to talk about the same concepts
 - Often in response to regulatory changes

Main Findings

- "Bad" data followed by increased securitization talk
- "Good" data followed by increased hedging talk
- Regulators will not be able to understand from the words how much securitization happens when times are good or how many unrealized and non-performing losses are accumulating
- Regulators will only see securitization being blamed ex post, after bad data realizations

- **Informative vs. Strategic Disclosures:**

Qualitative text signals private, forward-looking insights (*Information Content*: Verrecchia, R. E., 1983; Diamond, D. W., 1985). Alternatively, text is deployed strategically ex-post to obfuscate failures or rationalize write-downs (*Reactive Narrative*: Schrand, C. M., & Walther, B. R., 2000; Lacker, J. M., & Weinberg, J. A., 2007)

- **Computational Linguistics in Finance:**

Shift from general dictionaries that misclassify financial sentiment (Loughran, T., & McDonald, B., 2011) to specialized taxonomies with predictive power (Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H. M., & Steele, L. B., 2014). Recent trends utilize phrase-level configurations (Li, F., 2010; Dyer, T., Lang, M., & Stice-Lawrence, L., 2017)

- **Banking Regulation and Reporting Uniformity:**

Pre-2010 rules allowed wide regulatory discretion, causing idiosyncratic reporting styles (Barth, J. R., Caprio, G., & Levine, R., 2006). Post-crisis mandates institutionalized strict uniformity, triggering a debate on true transparency vs. boilerplate text (Beatty, A., & Liao, S., 2014; Bushman, R. M., & Williams, C. D., 2015)

Data Description

- Period 2003-2024
- 11 large systemically important US banks: AIG, BoA, Capital One, Citigroup, Goldman, JPM, Morgan Stanley, PNC Fin, Truist Fin, US Bankcorp, Wells Fargo
- SEC regulatory Form 10-K filings
- Textual corpus: Balanced panel of 242 total processed docs
- Algorithm design for automated quantitative data collection
 - Net income (NI)
 - Allowance loan loss (ALL)
 - Derivative liabilities (DL)
 - Derivative assets (DA)
 - Total equity (TE)

- Textual lexicons constructed
 - Credit risk
 - Market risk
 - Interest rate risk
 - Hedging risk
 - Securitization risk
 - Basel II
 - Basel III
 - Stress test
- Corresponding annual lexicon frequencies for each bank → corresponding textual variable (time series) for each bank

Example: Securitization Lexicon

Asset backed securitization, ABS, Collateralization, Asset pooling, Structured finance, Loan repackaging, Debt securitization, Credit securitization, Financial engineering, Securities issuance, Bond creation, Securitized products, MBS, Mortgage Backed Security, Mortgage backed bond, Mortgage pass through, Pass through security, Mortgage security, Home loan backed security, Residential MBS, RMBS, Commercial MBS, CMBS, Securitized mortgage, Structured mortgage product, Mortgage pool certificate, Structured debt product, Structured finance product, Asset backed CDO, Cash CDO, Synthetic CDO, Collateralized bond obligation, CBO, Collateralized loan obligation, CLO, Credit linked note, CLN, Debt obligation security, Structured credit product, Tranching debt instrument, Repurchase agreement, Sale and repurchase agreement, Collateralized loan, Securities lending with repurchase, Buyback agreement, Short term secured borrowing, Repo transaction, Reverse repo, CD, Certificate of Deposit, Time deposit, Bank certificate, Fixed term deposit, Interest bearing deposit, Negotiable certificate, Term CD, Jumbo CD, ABCP, Asset Backed Commercial Paper, Structured commercial paper, Asset backed note, Securitized paper, Short term structured note, Collateralized commercial paper, Credit enhanced paper, Conduit issued CP, Tranche, Slice, Layer, Class, Bond class, Credit tier, Risk segment, Subordinated piece, Senior tranche, Junior tranche

- Correlations between textual and quantitative variables
- Types of correlations
 - Pearson (PC): linear relationships
 - Spearman (SC): monotonic relationships (robustness)
- Fisher combined probability: converts p-values from individual-bank correlations into one test statistic

$$\chi^2 = -2 \sum_{i=1}^k \ln(p_i)$$

Example: AIG

	2006	2008	2010
Derivative Liabilities (DL)	21M	98M	83M
Net Income (NI)	21M	4M	-2M
Securitization lexicon frequency	7.5%	13%	19%

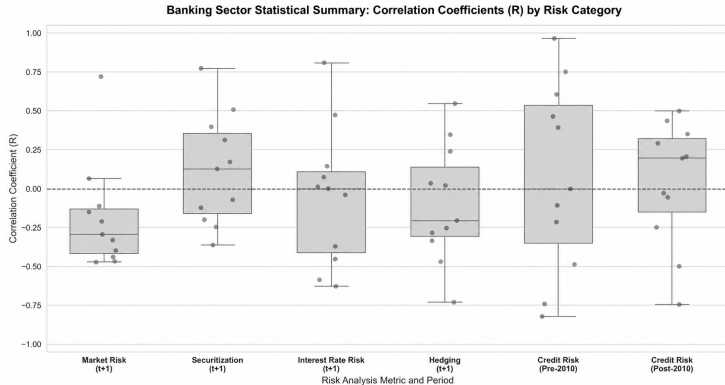
- In fall 2008, AIG's CDS portfolio collapsed
- DL $\uparrow$ from 21M in 2006 to almost 100M in 2008
- NI $\downarrow$ from 21M in 2006 to 4M in 2008, turns negative in 2010
- As DL $\uparrow$ and NI $\downarrow$, frequency of securitization lexicon almost tripped
- Textual narrative in $t + 1$ turned to securitization, likely as explanation for negative data shocks ("bad" data) observed in t
- Securitization talked about more in bad times (2008, 2010) than in good times (2006)

Example all banks: Fisher of Lexicons with DL (Table)

Lexicon	Window	Fisher	P-Value	Significant
Credit Risk	Pre-2010	0.07	0.001	Y
Credit Risk	Post-2010	0.03	0.01	Y
Interest Rate Risk	all	-0.05	0.000007	Y
Market Risk	all	-0.18	0.0003	Y
Hedging	all	-0.09	0.0006	Y
Securitization	all	0.11	0.001	Y

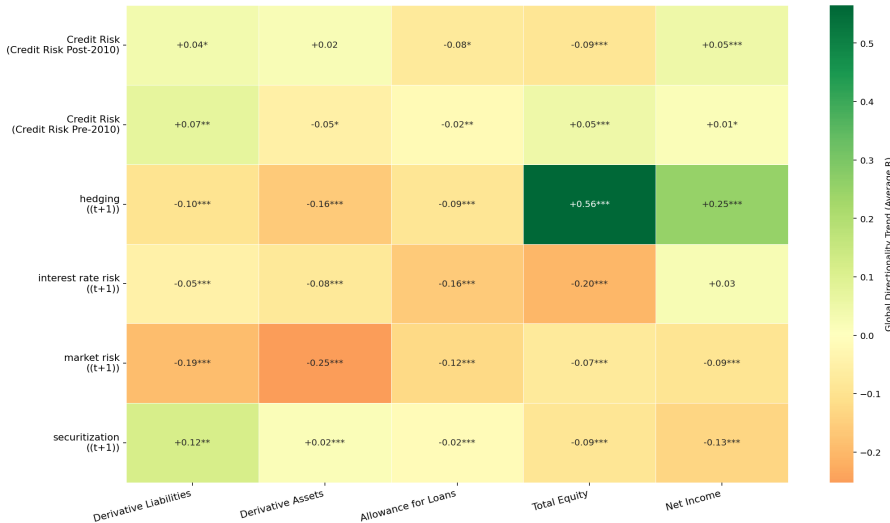
- Textual lexicons of $t + 1$ refer to quantitative DL data of t
- DL increase in $t \rightarrow$ in $t + 1$ banks talk more about credit risk and securitization, and less about market risk, hedging, and rate risk
- Securitization lexicon freq increases more than credit-risk lexicon freq
- Banks seem to attribute contract ((un)realized) losses to securitization and credit defaults
 - And not to mistakes they may have made in hedging or in forecasting rates and/or markets

Example all banks: Fisher of Lexicons with DL (Graph)



Fisher Heatmap: All Textual-Quantitative Correlations

Global Meta-Analysis Heatmap (Fisher's Combined Results)
Asterisks indicate Sector-Wide Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



Fisher Heatmap: Detailed Results

- DL $\uparrow$ in $t \rightarrow$ in $t + 1$ banks talk more about securitization and credit risk, and less about market risk, hedging, and rate risk
- ALL $\uparrow$ in t (consistent with higher expected or realized credit losses) $\rightarrow$ in $t + 1$ banks talk less about all risks, including credit risk (!)
- NI $\downarrow$ in $t \rightarrow$ in $t + 1$ banks talk more about securitization (as strong as in DL) and market risk, and less about hedging, credit risk and rate risk
- DA $\uparrow$ in $t \rightarrow$ in $t + 1$ banks talk more about securitization (but less strongly than in DL), and less about market risk, hedging (more strongly than in DL), rate risk, credit risk
- TE $\uparrow$ in $t \rightarrow$ in $t + 1$ banks talk more about hedging, and less about all risks

Fisher Heatmap: Interpretation of Results

- Bad data followed by increased securitization talk
- Good data followed by increased hedging talk
- Scant securitization data (FRED, not shown) indicate almost monotonic increase during the period, with exception 2008-2010
 - Implying that securitization consistently brings profits
 - Yet only talked about ex post, in case of losses
- Banks do not change lexicons (narrative) in response to increased ALL
 - Tied to non-performing losses (NPLs)
 - Scant quantitative data on NPLs, and cannot tell from words → hidden risk

Conclusions

- Regulators will not be able to understand from risk disclosure narratives how much securitization happens when times are good or how many NPLs and/or unrealized losses are accumulating
 - Bad news for forecasting risk accumulation and banking crises
- They will only see securitization being blamed ex post, after bad data realizations
- After good data, they will see that the banks likely praise their hedging skills

Thank you!