Narrative Augmented Bubble Detection and Prediction

Yuting Chen[1,2] Valerio Potì[1] Don Bredin[1]

[1]Universities College Dublin

[2]Rennes School of Business

September 18, 2022

Outline

Outline

Research Questions

- Can market narratives help predict financial market bubbles and their bust?
- Can market narratives help classify financial market bubbles?

Findings

- Narrative information can help explain the explosiveness of price-to-dividend ratio.
- Narrative features explosiveness can help predict the explosiveness of price-to-dividend ratio.
  - Because the former causes the latter?
  - Or because narrative features explosiveness can be detected earlier than the explosiveness of the price-to-dividend ratio due to the statistical properties of each?
- Narrative information can help predict market drops.
  - Narrative features become explosive during extreme-market event periods.
  - Since they do so in a predictable way (persistence of narrative features), it can be used to augment predictions of bubble bursts.
- Clustering the explosive episodes with narrative features provides some insights and evidence of rational and irrational bubbles.

Outline

Phillips Shi & Yu (2015) method

- PSY on the price-to-dividend ratio of S&P 500

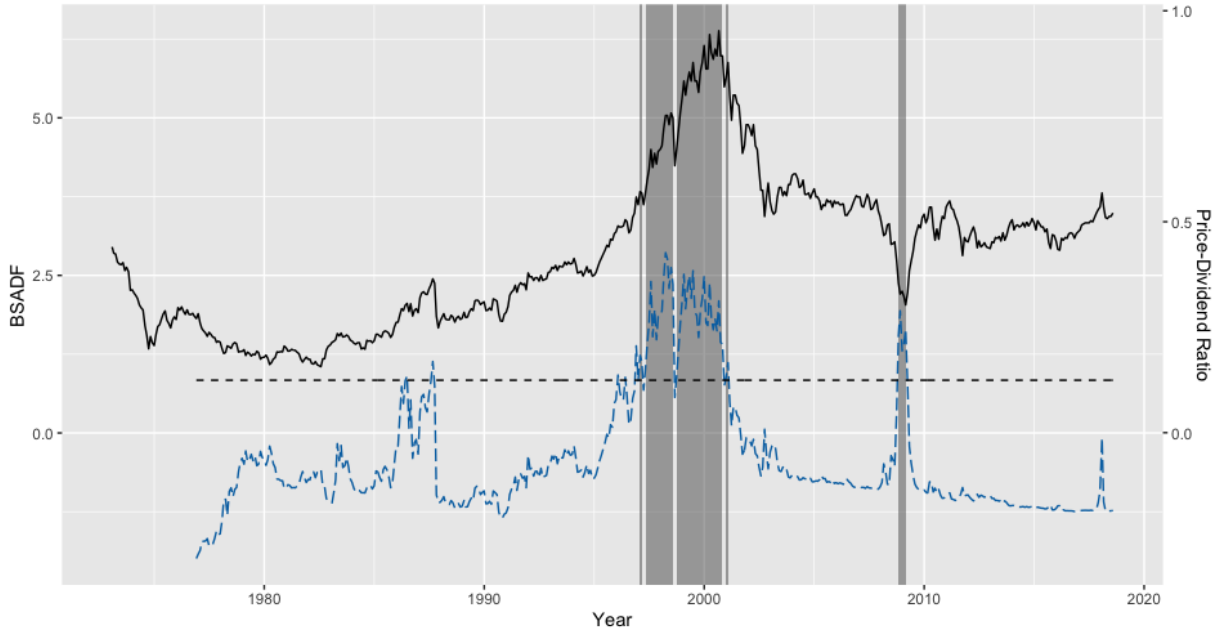


Figure: PSY test results for the S&P 500 price to dividend ratio

Notes: The dark line represents the S&P 500 price to dividend ratio, and the grey shaded regions indicate the detected explosive periods, in which the BSADF statistics (blue dashed line) exceed the bootstrapped critical values (dark dashed line).

Narrative information

- Textual data selection: searching query design
  1. Contain the keyword - "market"
  2. Major media news (from LexisNexis)
  3. "Economy Economic indicators" and "Financial Market Updates" subject tags
  4. January 1st 1975 - December 31st 2021
  5. 765,645 articles
- Language processing
  1. Use n grams to add country tags to each article
  2. 10 countries considered: the United States, the United Kingdom, Canada, Japan, Spain, Italy, France, Germany, Ireland and China
  3. Break down the articles into sentences

Narrative information

- Narrative features
  1. Market narrative intensity (count of sentences)
  2. Textual sentiment (both L&M and BERT)
  3. Textual sentiment dispersion (sentiment standard deviation)
  4. Topic consensus (Shannon entropy on LDA topic probabilities)
  5. Intensity of "risk narratives", "opportunity narratives" and "bubble narratives"

Narratives in forecasting bubbles

- Narrative features help predicting the change of price explosiveness

$$PD_{BSADF,c,t+1} = \alpha + \alpha_c + \beta_1 \cdot PD_{c,t} + \beta_2 \cdot PD_{BSADF,c,t} + \gamma \cdot \boldsymbol{Narrative_{c,t}} + \boldsymbol{\theta} \cdot \boldsymbol{X_{c,t}} + \epsilon_{c,t} \quad (1)$$

- Narrative features help predicting the scale of future market drops

$$Y_{c,t+12|t} = \alpha + \alpha_c + \beta_1 \cdot PD_{c,t} + \beta_2 \cdot PD_{BSADF,c,t} + \gamma \cdot \boldsymbol{Narrative_{c,t}} + \boldsymbol{\theta} \cdot \boldsymbol{X_{c,t}} + \epsilon_{c,t} \quad (2)$$

- Narrative features for predicting the PD ratio being explosive
  - Up-sampling (SMOTE) to deal with imbalanced data
  - Logistic regression, decision trees and kNN models
  - Feature selection with Information-Gains and the Sequential Feature Selection method

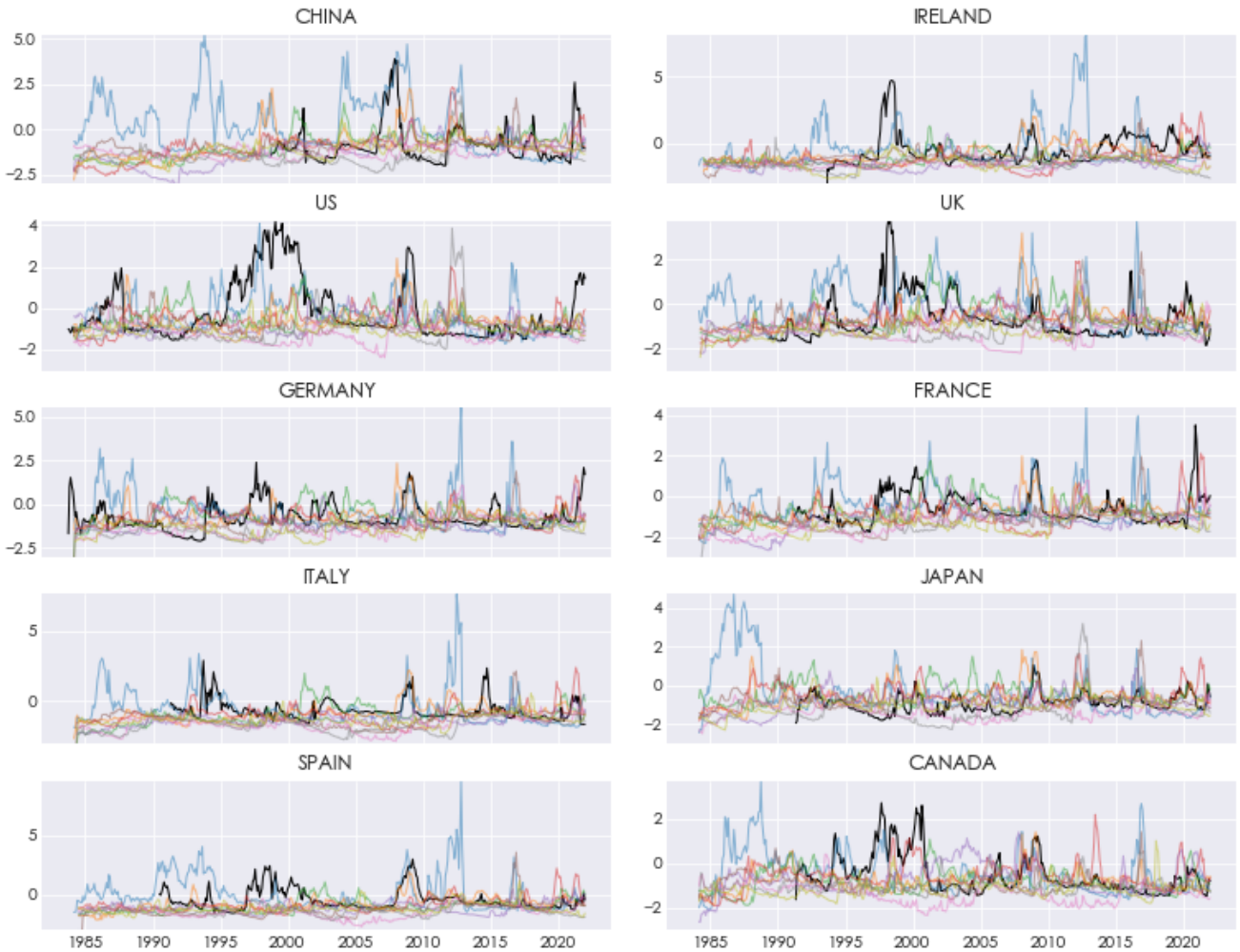
Narratives bubble classification

- Unsupervised learning with k-Means with $k = 2$
- Use narrative features only
- 166 observations with explosive price-to-dividend ratio

Outline

Narrative explosiveness

- BSADF statistics



Narrative features help predicting the change of price explosiveness

$$PD_{BSADF,c,t+1} = \alpha + \alpha_c + \beta_1 \cdot PD_{c,t} + \beta_2 \cdot PD_{BSADF,c,t} + \gamma \cdot \boldsymbol{Narrative_{c,t}} + \boldsymbol{\theta} \cdot \boldsymbol{X_{c,t}} + \epsilon_{c,t}$$

Narratives help predicting the change of price explosiveness

	Change of PD_B											
	model_1		model_2		model_3		model_4		model_5		model_6	
	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
Const	0.000	0.000***	0.000	0.000***	0.000	0.000***	0.000	0.000***	0.000	0.000***	0.000	0.000***
PD_B	0.026	0.080**			0.033	0.075**			0.025	0.074**	0.025	0.066*
PD	-0.073***	-0.037			-0.090***	-0.025			-0.076***	-0.034	-0.096***	-0.026
Bubble			-0.020	0.064***	-0.007	0.064***					0.025	0.056**
Opportunity			-0.027	-0.018	-0.028	-0.019					-0.041*	-0.008
Risk			0.003	0.011	-0.006	0.013					-0.003	0.041
Consensus			-0.054**	-0.014	-0.055**	-0.014					-0.058**	-0.021
# of sentence			0.057***	0.037	0.061***	0.036					0.072***	0.059
Dispersion (BERT)			0.023	-0.003	0.018	-0.002					0.009	-0.013
Dispersion (LM)			-0.016	-0.064**	-0.010	-0.067***					-0.029	-0.077***
Sentiment (BERT)			0.090***	-0.034	0.093***	-0.033					0.099***	-0.043
Sentiment (LM)			0.169***	-0.147***	0.172***	-0.142***					0.183***	-0.151***
Bubble_B							0.044**	-0.045**	0.041*	-0.044**	0.048**	-0.031
Opportunity_B							-0.018	0.015	-0.019	0.016	-0.034	0.016
Risk_B							0.015	0.040*	0.016	0.036*	0.014	0.048**
Consensus_B							0.004	-0.015	0.006	-0.012	0.002	-0.017
# of sentence_B							-0.036*	-0.016	-0.034*	-0.017	-0.047**	-0.038
Dispersion (BERT)_B							-0.014	-0.027	-0.015	-0.030	-0.012	-0.035
Dispersion (LM)_B							-0.068***	-0.029	-0.071***	-0.026	-0.085***	-0.026
Sentiment (BERT)_B							-0.035	0.036	-0.033	0.033	0.006	-0.009
Sentiment (LM)_B							-0.096***	-0.070***	-0.100***	-0.064**	-0.008	-0.096***
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Country	Country	Country	Country	Country	Country	Country	Country	Country	Country	Country	Country
No. observations	2354	1946	2361	1947	2354	1946	2361	1947	2354	1946	2354	1946
R-squared	0.84%	0.70%	5.17%	2.66%	5.84%	3.21%	1.88%	1.05%	2.39%	1.63%	7.03%	4.44%

Narrative features help predicting the maximum future drops

$$Y_{c,t+12|t} = \alpha + \alpha_c + \beta_1 \cdot PD_{c,t} + \beta_2 \cdot PD_{BSADF,c,t} + \gamma \cdot \boldsymbol{Narrative_{c,t}} + \boldsymbol{\theta} \cdot \boldsymbol{X_{c,t}} + \epsilon_{c,t}$$

Narratives help predicting the maximum market drop in 12 months

	Scale of Market Drop											
	model_1		model_2		model_3		model_4		model_5		model_6	
	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
Const	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PD_B	0.014	0.096***			0.003	0.088***			0.011	0.081***	0.004	0.071***
PD	0.038	-0.189***			0.038	-0.168***			0.040	-0.186***	0.039	-0.162***
Bubble			0.026*	0.020	0.038	0.023					0.035	0.022
Opportunity			0.024	0.020	0.033	0.014					0.032	0.040*
Risk			-0.035***	-0.042*	-0.069	-0.045**					-0.075***	-0.008
Consensus			-0.021	-0.016***	-0.030	-0.018***					-0.031	-0.033***
# of sentence			0.119***	-0.086***	0.122***	-0.091***					0.120***	-0.103***
Dispersion (BERT)			-0.045**	-0.101***	-0.044*	-0.093***					-0.068**	-0.106***
Dispersion (LM)			0.017	0.081	0.025	0.085					0.038	0.076
Sentiment (BERT)			-0.031	0.018	-0.030	0.033					-0.039	0.016
Sentiment (LM)			-0.091***	-0.164***	-0.105***	-0.150***					-0.088***	-0.185***
Bubble_B							0.000	-0.037*	0.002	-0.030	-0.001	-0.009
Opportunity_B							0.005	0.063***	0.005	0.067***	0.008	0.067***
Risk_B							0.010	0.052**	0.010	0.030	-0.025	0.040
Consensus_B							0.002	-0.023	0.003	-0.019	-0.002	-0.041**
# of sentence_B							0.037**	-0.019	0.038**	-0.027	0.006	0.019
Dispersion (BERT)_B							-0.039*	-0.012	-0.036	-0.025	-0.062**	-0.059***
Dispersion (LM)_B							0.026	-0.013	0.027	-0.010	0.042	0.009
Sentiment (BERT)_B							-0.014	0.045*	-0.016	0.044*	-0.040	0.024
Sentiment (LM)_B							0.121***	-0.197***	0.124***	-0.198***	0.057*	-0.208***
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Country	Country	Country	Country	Country	Country	Country	Country	Country	Country	Country	Country
No. observations	2211	1924	2327	2022	2211	1924	2327	2022	2211	1924	2211	1924
R-squared	0.50%	4.11%	4.01%	4.35%	4.73%	7.81%	1.92%	4.87%	1.96%	8.48%	5.36%	12.86%

Narrative information helps predicting the PD ratio being explosive (kNN)

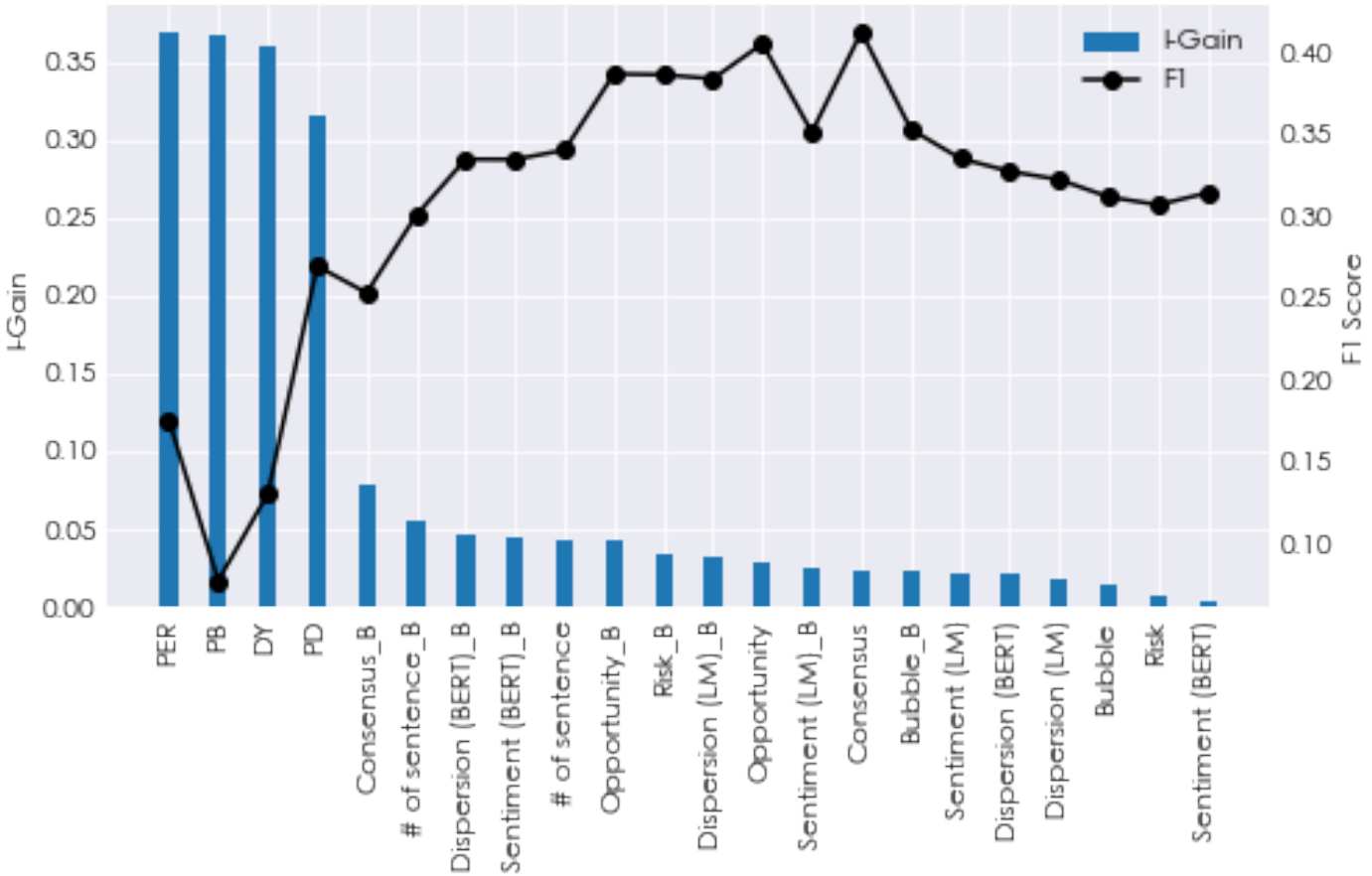


Figure: Information Gain and F1 Scores

Narrative information helps predicting the PD ratio being explosive

Table: Validation performance metrics of the sequential feature selection method

	TNR	TPR	Accuracy	F1
PER	0.91	0.13	0.88	0.07
PD	0.82	0.59	0.81	0.18
Consensus_B	0.84	0.59	0.83	0.20
# of sentence_B	0.87	0.62	0.86	0.24
Sentiment (BERT)_B	0.9	0.72	0.89	0.32
# of sentence	0.92	0.59	0.91	0.32
Opportunity_B	0.93	0.62	0.92	0.35
Sentiment (LM)_B	0.92	0.51	0.91	0.29
Consensus	0.93	0.54	0.91	0.31
Bubble_B	0.93	0.56	0.91	0.32
Sentiment (LM)	0.93	0.49	0.91	0.28
Dispersion (BERT)	0.93	0.56	0.92	0.34
Dispersion (LM)	0.93	0.59	0.92	0.35
Risk	0.93	0.54	0.92	0.32
Sentiment (BERT)	0.93	0.51	0.91	0.30

Narratives help classifying explosive episodes

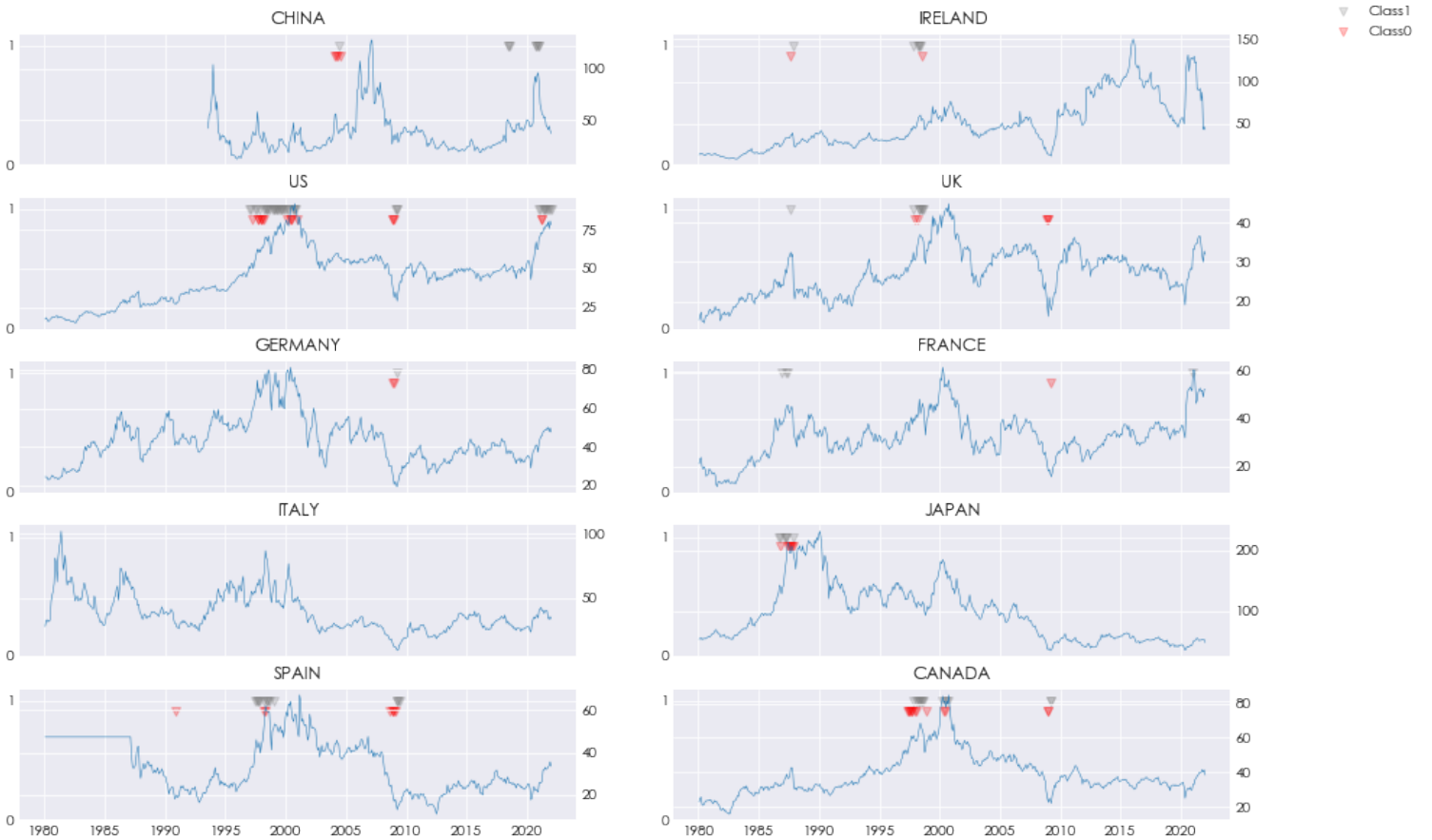


Figure: Classification of the Explosive Episodes

Narratives help classifying explosive episodes

Table: Means of narrative features for the two classes of explosive episodes

	Level		BSADF	
	Class 0	Class 1	Class 0	Class 1
# of sentence	0.886	-0.375	1.387	-0.208
Sentiment (LM)	-0.829	0.318	0.247	-0.018
Sentiment (BERT)	-0.082	0.025	0.142	0.089
Dispersion (LM)	0.539	-0.341	0.219	0.172
Dispersion (BERT)	0.366	-0.167	-0.079	0.126
Consensus	0.3	-0.149	-0.17	0.339
Risk	0.121	-0.105	0.203	0.014
Opportunity	0.097	0.039	-0.182	0.17
Bubble	0.404	-0.078	0.267	0.293

UCD Michael Smurfit
Graduate Business School