Beyond the Numbers: Soft Information in Conditional Asset Pricing

Previously circulated under the title “Conditional Asset Pricing with Text-Managed Portfolios”

Jian Feng (PKU) Jiantao Huang (HKU) Shiyang Huang (HKU) Ran Shi (CU Boulder)

The CFP - SKBI - UBS Quant Investment Forum
June, 2026

Background

- Extensive literature on latent asset pricing models:

$$r_{i,t+1} = \boldsymbol{\beta}'_{i,t} \boldsymbol{f}_{t+1} + \epsilon_{i,t+1}$$

  - $\boldsymbol{f}_{t+1}$: latent common factors
  - $\boldsymbol{\beta}_{i,t}$: captures conditional information (market beta, size, valuation, technical signals, etc.)
  - When do $\boldsymbol{f}_{t+1}$ span the mean-variance efficient (MVE) space of $r_{i,t+1}$?

- In theory, $\boldsymbol{f} \approx (\boldsymbol{\beta}_t^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\beta}_t)^{-1} \boldsymbol{\beta}_t^\top \boldsymbol{\Sigma}_t^{-1} \boldsymbol{R}_{t+1}$, GLS managed portfolios
  - span MVE space of individual stocks, but impractical (large $\boldsymbol{\Sigma}_t$)

- In practice, $\boldsymbol{f} \approx (\boldsymbol{\beta}_t^\top \boldsymbol{\beta}_t)^{-1} \boldsymbol{\beta}_t^\top \boldsymbol{R}_{t+1}$, OLS managed portfolios
  - i.e., IPCA (Kelly, Pruitt, and Su, 2019)
  - more likely to span if $\boldsymbol{\beta}_t$ is rich in conditioning information (Kozak and Nagel, 2024)

$\Longrightarrow$ We need more (and better) information to model conditional stock dynamics!

Soft Information: A Gap in the Literature

- Standard firm characteristics are hard signals (accounting, past returns)
  - KPS (2019): 30+ characteristics — intensively studied
  - May miss soft, qualitative, forward-looking information
- A prominent source of soft information: earnings conference calls
  - Presentation: prepared remarks — performance, strategy, guidance example
  - Q&A: unscripted manager–analyst interaction
  - Together: soft information incremental to standard characteristics
- Research questions:
  - Construct text-managed portfolios & factors within IPCA?
  - Asset pricing implications of conditioning on soft information?
  - Incremental informational content of text?

Encoding Text: LDA Model

- $\boldsymbol{d}_{i,t}$: word counts of firm i's call at t, multinomial:

$$\boldsymbol{d}_{i,t} \sim \text{Multinomial}\left(\mathbf{1}'\boldsymbol{d}_{i,t}, \sum_{k=1}^{K} \phi_{i,t}^{(k)} \boldsymbol{\theta}_k\right)$$

  - $\boldsymbol{\theta}_k$: word distribution of topic k; $\phi_{i,t}^{(k)}$: firm i's loading on topic k

- Encode each call by loadings $[\phi_{i,t}^{(1)}, \ldots, \phi_{i,t}^{(K)}]'$ (Lopez-Lira, 2023)
  - Comparable to K firm characteristics — exposures to dominant topics

- Topic selection: perplexity, then stable-topic algorithm across seeds
  - $K = 32$ stable topics each for Q&A and PRE call examples
  - Three themes: general (e.g., "Economic Uncertainty"), industries, tech vintages
  - Common across sessions + session-specific

Stable Topics

Topics from θ; word size from $\bar{\phi}$; common/specific in red/green.

Q&A: forward-looking ("Optimism," "Growth Outlook," "Forecasting"); PRE: operational ("Profits," "Business Operations," "Organizational Management")

(A) Q&A

(B) Presentation (PRE)

IPCA Model with Characteristics & Text

KPS IPCA model

$$r_{i,t+1} = \boldsymbol{\beta}_{i,t}' \boldsymbol{f}_{t+1} + \epsilon_{i,t+1}, \quad \text{where } \boldsymbol{\beta}_{i,t} = \boldsymbol{\Gamma} \boldsymbol{x}_{i,t}$$

- $\boldsymbol{x}_{i,t}$: $\boldsymbol{c}_{i,t}$ (35 char.) and/or $\phi_{i,t}$ (topic loadings from LDA)
- Enriching conditioning information (Kozak and Nagel, 2024):
  - $[\text{char}_{i,t}, \text{text}_{i,t}]$ more likely to span MVE space than $\text{char}_{i,t}$ alone — if text helps explain conditional expected returns or covariances
- Expanding-window ("out-of-sample") estimation:
  - Total R^2: fit for realized returns; Predictive R^2: using $\widehat{\boldsymbol{\lambda}}_{ft}$
  - OOS SR of factor tangency portfolios $\widehat{\boldsymbol{\lambda}}_{ft}^{\top} \widehat{\boldsymbol{\Sigma}}_{ft}^{-1} \widehat{\boldsymbol{f}}_{t+1}$

IPCA Estimates: Time-Series Variation in Returns

Table: Explaining time-series stock returns (R^2 in % unit)

		$p=5$	$p=10$	$p=15$	$p=20$	$p=25$	$p=30$
32 Q&A topics	total R^2	21.28	22.15	22.50	22.78	23.03	23.27
	pred R^2	0.14	0.14	0.13	0.13	0.14	0.14
32 PRE topics	total R^2	21.69	22.58	22.98	23.26	23.53	23.75
	pred R^2	0.14	0.13	0.14	0.14	0.14	0.14
64 Q&A + PRE topics	total R^2	22.05	22.99	23.41	23.72	23.99	24.24
	pred R^2	0.13	0.12	0.12	0.13	0.13	0.13
35 firm char.	total R^2	24.00	25.13	25.79	26.30	26.71	27.04
	pred R^2	0.14	0.13	0.14	0.15	0.15	0.14
64 Q&A + PRE topics + 35 firm char.	total R^2	24.46	25.86	26.58	27.03	27.42	27.76
	pred R^2	0.09	0.12	0.15	0.15	0.15	0.16

- Topic loadings: conditional information comparable to characteristics
- Text adds limited predictive power for $\mathbb{E}_t[r_{i,t+1}]$ — its value lies elsewhere

Mean-Variance Efficiency

Factor tangency portfolios: expanding-window estimates of $\widehat{\boldsymbol{\lambda}}_{ft}^{\top}\widehat{\boldsymbol{\Sigma}}_{ft}^{-1}\widehat{\boldsymbol{f}}_{t+1}$

Table: Out-of-sample Sharpe ratios of the MVE portfolios

Cross-Sectional Information Sets:	$p=5$	$p=10$	$p=15$	$p=20$	$p=25$	$p=30$
32 Q&A topics	0.34	0.67	0.47	0.74	1.15	1.04
32 PRE topics	0.79	0.92	1.03	1.17	1.07	1.16
64 Q&A + PRE topics	0.55	0.65	0.95	1.29	1.38	1.64

- No sparsity: need >20 latent factors

Mean-Variance Efficiency

Factor tangency portfolios: expanding-window estimates of $\widehat{\boldsymbol{\lambda}}_{ft}^{\top}\widehat{\boldsymbol{\Sigma}}_{ft}^{-1}\widehat{\boldsymbol{f}}_{t+1}$

Table: Out-of-sample Sharpe ratios of the MVE portfolios

Cross-Sectional Information Sets:	$p=5$	$p=10$	$p=15$	$p=20$	$p=25$	$p=30$
32 Q&A topics	0.34	0.67	0.47	0.74	1.15	1.04
32 PRE topics	0.79	0.92	1.03	1.17	1.07	1.16
64 Q&A + PRE topics	0.55	0.65	0.95	1.29	1.38	1.64
Q&A refined	0.27	0.49	0.53	0.80	0.87	1.25
(13 Q&A specific + 38 common)						
PRE refined	0.72	0.67	0.91	1.16	1.20	1.20
(13 PRE specific + 38 common)						

- No sparsity: need > 20 latent factors
- Dropping Q&A or PRE specific topics $\Rightarrow$ worse performance

Mean-Variance Efficiency ⏩ online LDA

Factor tangency portfolios: expanding-window estimates of $\widehat{\boldsymbol{\lambda}}_{ft}^{\top} \widehat{\boldsymbol{\Sigma}}_{ft}^{-1} \widehat{\boldsymbol{f}}_{t+1}$

Table: Out-of-sample Sharpe ratios of the MVE portfolios

Cross-Sectional Information Sets:	$p = 5$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
32 Q&A topics	0.34	0.67	0.47	0.74	1.15	1.04
32 PRE topics	0.79	0.92	1.03	1.17	1.07	1.16
64 Q&A + PRE topics	0.55	0.65	0.95	1.29	1.38	1.64
Q&A refined (13 Q&A specific + 38 common)	0.27	0.49	0.53	0.80	0.87	1.25
PRE refined (13 PRE specific + 38 common)	0.72	0.67	0.91	1.16	1.20	1.20
35 Firm Characteristics	0.63	1.16	1.57	1.68	1.74	1.56
64 Q&A + PRE topics + 35 Firm Chars	0.23	1.49	2.30	2.33	2.53	2.65

- No sparsity: need 20+ latent factors
- Dropping session-specific topics $\Rightarrow$ worse performance
- Q&A, PRE, and characteristics: largely independent conditional information

First- vs. Second-Moment Channels of Soft Information

- Conditional MVE portfolio: $\boldsymbol{\omega}_t^{\text{mve}} \propto \boldsymbol{\Sigma}_t^{-1} \mathbb{E}_t[\boldsymbol{r}_{t+1}]$
- Textual information can help via:
  - **First-moment channel**: refine $\mathbb{E}_t[\boldsymbol{r}_{t+1}]$ via $\boldsymbol{\beta}_t \boldsymbol{\lambda}_t$
  - **Second-moment channel**: improve $\boldsymbol{\Sigma}_t$
- Empirical design — quintile sorts:
  - by $\boldsymbol{\beta}_t \boldsymbol{\lambda}_t \Rightarrow$ isolates first moment
  - by $\boldsymbol{\omega}_t^{\text{mve}} \Rightarrow$ exploits mean + covariance
- Three information sets:
  1. 35 firm characteristics
  2. 64 topic loadings (soft information)
  3. characteristics + topics

Evidence on the First-Moment Channel

Quintiles sorted by model-implied expected returns $\beta_t \lambda_t$

	Firm Char.			Topic Loadings			Firm Char. + Topic Loadings		
Portfolio rank	mean (%)	std. (%)	SR	mean (%)	std. (%)	SR	mean (%)	std. (%)	SR
1	7.8 (1.55)	20.2	0.38	9.5 (1.46)	23.2	0.41	6.6 (1.17)	20.1	0.33
2	10.3 (1.92)	20.2	0.51	11.0 (1.74)	23.0	0.48	9.4 (1.77)	20.9	0.45
3	10.8 (1.87)	22.0	0.49	11.8 (1.86)	22.7	0.52	11.4 (1.92)	22.1	0.51
4	12.6 (1.95)	24.0	0.53	12.0 (1.84)	22.9	0.52	12.3 (1.86)	24.0	0.51
5	15.1 (1.63)	29.3	0.52	12.4 (1.95)	23.7	0.52	16.9 (2.03)	28.2	0.60
5–1	7.3 (1.55)	13.4	0.55	2.9 (1.33)	9.1	0.32	10.4 (3.26)	11.8	0.88

- Char. only: 5–1 mean **7.3%**, SR **0.55** (pred. R^2 only 0.15%)
- Text only: weak predictor — 5–1 mean **2.9%**, SR **0.32**
- Char. + text: 5–1 mean **10.4%** (t=3.26), SR **0.88**

  $\Rightarrow$ first-moment improvement remains modest

Second-Moment Channel and Takeaways

Quintiles sorted by MVE weights ω_t^{mve}

	Firm Char.			Topic Loadings			Firm Char. + Topic Loadings		
Portfolio rank	mean (%)	std. (%)	SR	mean (%)	std. (%)	SR	mean (%)	std. (%)	SR
1	7.9 (1.13)	25.0	0.32	7.6 (1.10)	24.2	0.32	6.3 (0.89)	24.9	0.25
2	10.1 (1.60)	23.7	0.43	10.4 (1.59)	23.3	0.45	9.4 (1.44)	23.5	0.40
3	11.5 (1.88)	22.6	0.51	12.5 (1.94)	23.1	0.54	11.4 (1.90)	22.4	0.51
4	12.2 (1.98)	21.9	0.56	12.6 (2.13)	22.3	0.57	12.9 (2.15)	22.5	0.57
5	14.9 (2.39)	21.5	0.69	13.3 (2.28)	21.7	0.62	16.6 (2.70)	21.3	0.78
5–1	7.0 (5.43)	6.3	1.11	5.7 (3.72)	5.0	1.14	10.3 (7.43)	5.9	1.74

- MVE-sorted quintiles: means rise, volatilities fall, SRs rise monotonically

  ω_t^{mve} uses Σ_t to build better-diversified portfolios

- 5–1 SR: chars **1.11**; topics **1.14** (covariance benefits despite weak means); chars + topics **1.74**

  $\Rightarrow$ second-moment channel is the primary driver of gains

Interpretation

Table: Correlation coefficients of MVE portfolios

	$p = 25$			$p = 30$		
	32 Q&A topics	32 PRE topics	35 firm char.	32 Q&A topics	32 PRE topics	35 firm char.
32 Q&A topics	1.00			1.00		
32 PRE topics	0.22	1.00		0.17	1.00	
35 Firm Chars	0.24	−0.04	1.00	0.25	0.07	1.00

Soft information captures covariance dimensions not spanned by characteristics

- Text- and char.-based MVE portfolios only weakly correlated
- Soft information not spanned by standard firm characteristics

Main conclusion: conference call topics enhance MVE efficiency mainly by improving $\boldsymbol{\Sigma}_t$, not expected-return forecasts

- Echoes Kozak and Nagel (2024): characteristics valuable even if they add little to expected returns, if they capture main covariance sources

Textual vs. Characteristic Information: Return Decomposition

Step 1: Construct signals with IPCA

- Estimate IPCA separately on: characteristics $\Rightarrow \widehat{R}^{char}_{i,t+1}$; text $\Rightarrow \widehat{R}^{text}_{i,t+1}$
- Summary statistics for info. in char. and text

Step 2: Combine signals in returns

$$R_{i,t+1} = b_0 + b_1 \widehat{R}^{char}_{i,t+1} + b_2 \widehat{R}^{text}_{i,t+1} + e_{i,t+1}$$

- Best linear combination of characteristic- and text-based signals

Step 3: Variance decomposition of total explained share

$$\underbrace{\frac{\mathrm{Cov}(R_{i,t+1}, b_1 \widehat{R}^{char}_{i,t+1} + b_2 \widehat{R}^{text}_{i,t+1})}{\mathrm{Var}(R_{i,t+1})}}_{R^2_{total} \approx 27.8\%} = \underbrace{b_1 \frac{\mathrm{Cov}(R_{i,t+1}, \widehat{R}^{char}_{i,t+1})}{\mathrm{Var}(R_{i,t+1})}}_{R^2_{char} \approx 18.7\%} + \underbrace{b_2 \frac{\mathrm{Cov}(R_{i,t+1}, \widehat{R}^{text}_{i,t+1})}{\mathrm{Var}(R_{i,t+1})}}_{R^2_{text} \approx 9.1\%}$$

$\Rightarrow$ text accounts for roughly one-third of the common explained variation

Heterogeneous Role of Soft Information Across Industries

Ex ante intuition

- Text matters more where fundamentals are intangible-intensive, rapidly evolving, cyclical
- Text matters less in mature / regulated sectors (e.g., utilities)

Industry	N_{firms}	R^2_{total} (%)	R^2_{char} (%)	R^2_{text} (%)	$\frac{R^2_{text}}{R^2_{total}}$ (%)
All Firms	2338	27.80	18.67	9.14	32.87
Measuring and Control Equipment	45	31.00	16.32	14.68	47.36
Mines	19	29.96	15.76	14.20	47.40
Chips	132	28.80	16.59	12.21	42.40
Restaurants, Hotels, Motels	46	37.81	22.28	15.53	41.08
Business Services	123	28.39	17.01	11.38	40.07
Computer Hardware	43	19.27	11.63	7.64	39.63
Medical Equipment	80	20.06	12.18	7.88	39.29
	⋮	⋮	⋮	⋮	⋮
Consumer Goods	32	30.83	23.82	7.01	22.74
Chemicals	64	32.95	24.35	8.60	26.11
Retail	124	23.42	17.77	5.65	24.13
Personal Services	31	26.33	20.76	5.57	21.15
Entertainment	29	34.30	29.05	5.25	15.31
Utilities	69	29.90	28.77	1.13	3.79

$\Rightarrow$ text share largest in tech / equipment / services; smallest in utilities

MVE Portfolio Weights and Conditional Information: Setup

Idea: how strongly does the MVE portfolio load on each instrument over time?

Method

- Monthly cross-sectional regression:

$$w_{i,t}^{\text{MVE}} = a_t + \boldsymbol{b}_t^\top \boldsymbol{X}_{i,t} + \varepsilon_{i,t}$$

  $w_{i,t}^{\text{MVE}}$: MVE weights; $\boldsymbol{X}_{i,t}$: conditional instruments

- Obtain time series $\{\boldsymbol{b}_t\}$ and R_t^2
- To mitigate collinearity:
  - Topics in both sessions (e.g., "Cloud Service"): average of Q&A and PRE loadings
  - Characteristics with pairwise corr. > 0.7: grouped, equally-weighted average
- All instruments standardized monthly to cross-sectional st. dev. of 1%

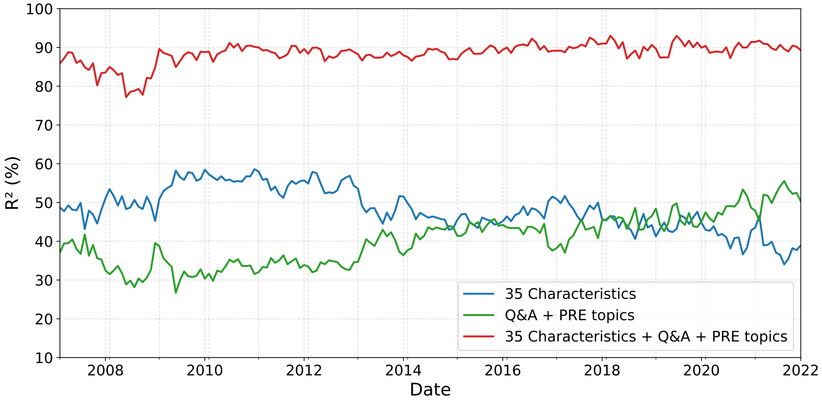
Joint Role of Soft Information and Characteristics in MVE Weights

Conditional on text + characteristics — regress MVE weights on combined instruments

Q&A + PRE topics + Characteristics			
Instrument	mean	5th	95th
(prc, dolvol_126d, market_equity)	-0.47	-0.65	-0.32
(ebit_bev, ni_be, niq_be, niq_at)	0.40	0.19	0.74
(at_me, be_me)	0.41	0.21	0.66
gp_at	0.40	0.28	0.54
ret_1_0	-0.38	-0.55	-0.26
prc_highprc_252d	0.33	0.12	0.55
	⋮	⋮	⋮
Business Operation (PRE-specific)	0.48	0.24	0.64
Organizational Management (PRE-specific)	-0.35	-0.45	-0.16
International Finance (PRE-specific)	-0.31	-0.41	-0.12
Financial Reporting (PRE-specific)	0.24	0.17	0.37
Semiconductor and Electronics (Common)	-0.24	-0.36	-0.13
Human Capital (QA-specific)	-0.17	-0.33	-0.05
Capital Structure (QA-specific)	0.15	0.08	0.27
	⋮	⋮	⋮

⇒ several topics remain top instruments, with stable coefficients, after controlling for characteristics

Decomposition of MVE Weights: Text vs. Characteristics



- Characteristics: $\approx$ **48%** of cross-sectional variation in MVE weights
- Text alone: $\approx$ **40%**
- Text share rises over time, exceeding **50%** in recent years: soft information now at least as important as characteristics

Conclusion

Soft, qualitative information matters for conditional asset pricing

- **Text vs. characteristics**
  - Conference-call topics as crucial as standard characteristics
  - Incremental value within an IPCA framework
- **Main channel**
  - Text substantially improves mean–variance efficiency
  - Gains primarily covariance-driven; expected-return power modest
- **Cross-section and portfolios**
  - Text $\approx$ one-third of common variation; larger for high-growth, intangible-intensive, hard-to-value firms
  - In MVE weights: text nearly as much as characteristics; importance grown over time

Appendix: Example of soft information in presentation

Fortive Corporation, summary of presentation session, July 2022

Back

Quantitative Summary (Hard Metrics)

Q2 core revenue growth	9%
Q2 adj. op margin expansion	190 basis points
Q2 adj. EPS growth	18%
Hardware orders growth (Q2)	9%
Hardware backlog vs YE21	21%
COVID/China risk mitigated	$60 million
FX headwind on revenue	$100 million
FX headwind on EPS	$0.08
IOS revenue growth	16%
IOS core revenue growth	12%
IOS core margin expansion	205 basis points
iNet bookings growth	55%
PT total revenue growth	6%
PT core revenue growth	9%
PT adj. op margin expansion	90 basis points
AHS total revenue growth	9%

Soft, Qualitative, and Forward-Looking Narrative

We had an excellent second quarter with strong broad-based execution across the portfolio, contributing to revenue, margins and earnings all above the high end of our guidance, resulting in our raised outlook for the year.

Starting on the left with the current environment, demand and orders remained strong in the second quarter, driven by accelerated innovation, continued share gains and leverage to favorable secular drivers.

While ongoing COVID lockdowns in China remain a risk, we substantially mitigated the headwind from the lockdowns that commenced in late March in Shanghai and continued through most of May, shifting most of the $60 million of risk we previously highlighted in the first half.

Moving to the right side of the slide, we expect higher core growth for the remainder of the year as our more resilient product portfolio positions us to benefit from continued customer demand.

We also continue to build momentum in our software businesses, with upsell and cross-sell bookings, new logo generation and lower churn all contributing to double-digit ARR growth for the full year.

IOS continued its strong momentum, with revenues up 16% and core revenue growth of 12% in the second quarter, with strong double-digit growth in North America and Western Europe and high single-digit growth in China.

Appendix: examples in conference call

Broadwing Communications, 2002Q4 Earnings call

Bardone, Analyst (Question)	Hey, guys, congratulations on securing the restructuring. Just wanted to ask three quick questions. First, I think Tom, you were talking about some deleveraging targets through 2005, which looked pretty substantial. Could you be more specific about how the deleveraging is going to take place? Are you talking about operational cash flows being used to buy back debt early in the public markets, or are you talking about potentially an equity deal here and whether you've made commitments to the banks on that front? The second question I have was on the Oak Hill paper...
Schilling, CFO (Answer)	Hey, David. Tom. I'll take the first couple questions then probably turn it over to Jack for the last one. On the senior debt deleveraging I was talking about I was specifically talking about our senior debt which is our senior secured debt which is our bank credit facility and our capital leases, as well as the 7% notes. And that -- the deleveraging there is coming from the junior capital we're raising arranged by Goldman Sachs of $350m, along with the operational cash flows during the next three years that will be generated from the business. On the convert terms of Oak Hill...

Topic Loadings:
Capital Structure, 0.403 (Highest)

Appendix: examples in conference call

Back

Dun & Bradstreet Corp, 2004Q3 Earnings Call	
Alesio, COO	... Our team members' focused leadership drove our risk management results in this quarter ... Let me show an example of how our team members' leadership benefited our e-business results in the quarter. In this other example, we have a customer which is a large company in the high tech industry that has used D&B Solutions for many years. In some cases they needed access to deeper and timelier information than their D&B current solution offered. And they wanted to offer access to this business insight more broadly through their organization. So, members of our sales and marketing solutions team partnered with our Hoover's team to meet our customers' needs. Through teamwork and results leadership they showed how our customer could achieve deeper, more accurate, and more timely business insight by adding Hoover's information to their current D&B Solution. As a result, our customer purchased an enterprise wide license and a full subscription to Hoover's data. This is another great example of how our team members are leading with the interest of our customers first in order to drive superb results...
Topic Loadings: Organizational Management, 0.270 (Highest)	

Appendix: Out-of-Sample Sharpe Ratios From 2013

Cross-Sectional Information Sets:	$p = 5$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
	Panel A. Full-Sample LDA					
32 Q&A topics	0.67	0.89	0.36	0.85	1.31	1.17
32 PRE topics	1.36	1.33	1.29	1.22	1.33	1.35
64 Q&A + PRE topics	1.08	0.91	1.33	1.79	1.71	1.79
35 Firm Characteristics	1.09	1.32	1.74	1.79	1.89	1.91
64 Q&A + PRE topics + 35 Firm Chars	0.42	1.60	2.46	2.46	2.83	2.95
	Panel B. Online LDA					
32 Q&A topics	0.55	0.23	0.84	0.98	1.21	1.20
32 PRE topics	1.00	1.16	1.19	1.32	1.34	1.37
64 Q&A + PRE topics	0.85	0.71	1.37	1.69	1.83	1.85
64 Q&A + PRE topics + 35 Firm Chars	0.37	1.41	2.27	2.23	2.38	2.60

Real-Time Strategies with Online LDA ⏩ Back

- Lookahead bias is universal in pretrained language models (Sarkar and Vafa, 2024)
  - Examples: GPT and BERT related models
  - Full-sample LDA: $\boldsymbol{\Theta}$ relies on full sample; loadings from time-t calls only
- Solution: Online LDA (Hoffman et al., 2010)
  - Use information up to time t to estimate $\boldsymbol{\Theta}$ and loadings
  - Pros: real-time, no information leakage; Cons: time-varying, harder to interpret

	$p = 5$	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$
32 Q&A topics	0.20	0.02	0.73	0.91	1.03	1.00
32 PRE topics	0.52	0.69	0.95	1.05	1.08	1.16
64 Q&A + PRE topics	0.43	0.42	1.08	1.38	1.46	1.47
64 Q&A + PRE topics + 35 Firm Chars	0.21	1.27	1.99	2.14	2.19	2.30