

Web Appendix to “Mispricing Proxies in Factor Models for Asset Returns”

Gabriele Confalonieri
Carlo A. Favero
Ilaria Leoni

A FINBERT AND SENTIMENT CONSTRUCTION

Transformer-based financial sentiment models improve on fixed financial dictionaries by using context. FinBERT [1] adapts BERT [2] to financial text, while Mishev et al. [3] review the movement from lexicon methods toward deep contextual models. The manuscript uses this technology to construct a repeated, company-level signal from earnings calls.

A.1 BERT and financial-domain adaptation

BERT is built on the Transformer architecture [4] and uses bidirectional self-attention to represent words in context. Its pretraining combines masked-language modeling and next-sentence prediction; task-specific fine-tuning then adapts the model to labeled applications.

FinBERT is trained and fine-tuned on financial-language corpora. The Financial Phrasebank contains more than 4,800 sentences annotated by people with financial-domain knowledge as positive, negative, or neutral from an investor perspective. Domain adaptation helps the model distinguish financial meanings that a general-language classifier may miss.

A.2 Transcript processing

The transcript of each earnings call is segmented into sentences. Every sentence is passed through FinBERT and assigned positive, negative, and neutral probabilities. Sentence-level scores are aggregated with equal sentence weights to obtain a single call-level sentiment indicator. The baseline score is centered before use in panel regressions. Calls are aligned to the quarter in which they are released.

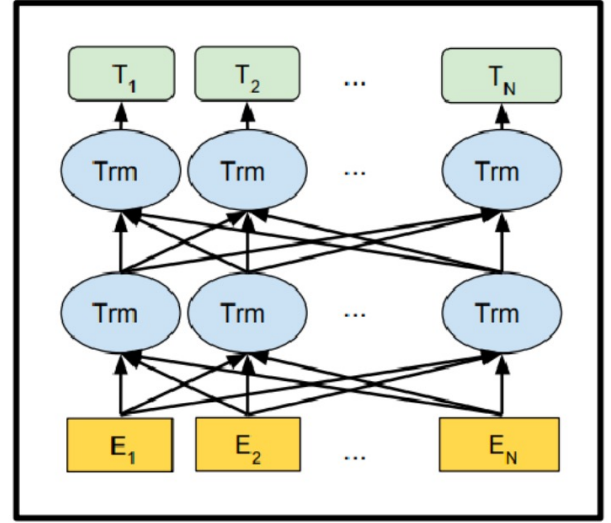


Figure 1: BERT model architecture.

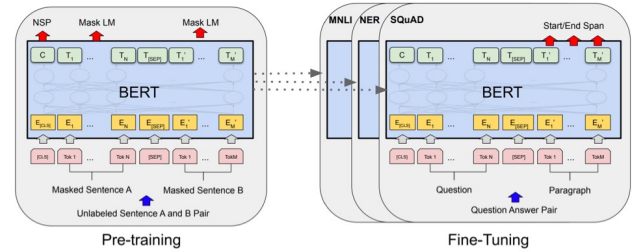


Figure 2: BERT pretraining and fine-tuning.

A.3 JPMorgan example

, July 2026,

Gabriele Confalonieri, Carlo A. Favero, and Ilaria Leoni

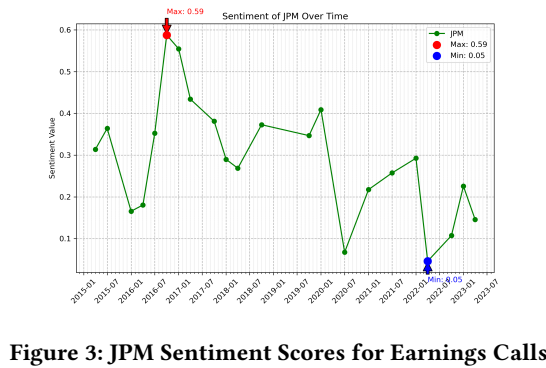


Figure 3: JPM Sentiment Scores for Earnings Calls

Regarding loan growth, we're continuing to see positive trends with loans up 8% year-on-year and 1% quarter-on-quarter as PPP with the sequential growth driven by a continued pickup in demand in our Wholesale businesses, including ongoing strength in AWH.	0.000000
On Page 2, we have some more detail on our results.	-1.40E-08
Revenue of \$31.6 billion was down \$1.5 billion or 5% year-on-year	-0.000000
Net income was up \$1.1 billion or 2% on a constant dollar basis and higher rates, partially offset by lower NII from PPP loans.	0.000000
NII or Markets was down \$2.2 billion or 17% predominantly driven by lower fee fees, lower Home Lending production revenue, losses in Credit Adjustments & Other and CIB as well as investment securities losses in corporate	-1.000000
And Markets revenue was down \$200 million or 2% against a record first quarter last year.	-0.000000
Expenses of \$19.2 billion were up approximately \$500 million or 2% predominantly on higher investments and structural expenses, largely offset by lower volume and revenue-related expenses	0.000000
Credit costs were \$1.2 billion for the quarter	-0.000000
We built \$902 million in reserves driven by increasing the probability of downside risks due to high inflation and the war in Ukraine as well as builds for Russia-associated exposures in CIB and AWH.	0.000000
Net charge-offs of \$582 million were down year-on-year and comparable to last quarter and remain historically low across our portfolios.	-0.000000
On to balance sheet and capital on Page 3.	-0.000000
Our CET1 ratio ended at 11.3%, down 130 basis points from the prior quarter.	-0.000000
As a reminder, we added the fourth quarter with an elevated buffer to absorb anticipated changes this quarter, the largest being SA-CDR adoption as well as some pickup in seasonal activity.	0.000000
In addition to those anticipated items, there were a couple of other drivers.	-0.000000
The rate sell-off led to AOCI drawdowns in our AFS portfolio.	-0.000000
But keep in mind, all else equal, these mark-to-market losses accrue back to capital through time and as securities mature.	-0.000000
And price increases across commodities resulted in higher counterparty credit and market risk RNA.	0.000000
While, of course, the environment is uncertain, many of these effects are now in the rearview mirror.	-0.000000
As a result, we believe that our current capital and future earnings profile position us well to continue supporting business growth while meeting increasing capital requirements as we look ahead.	0.000000
With that, let's go to our businesses, starting with Consumer & Community Banking on Page 4.	-0.000000
CIB reported net income of \$2.9 billion on revenue of \$12.2 billion, which was down 2% year-on-year.	-0.000000
In Consumer & Business Banking, revenue was up 8% predominantly driven by growth in deposit balances and client investment assets, partially offset by deposit margin compression.	0.000000
Deposits were up 18% year-on-year and 4% quarter-on-quarter, consistent with last quarter.	0.000000
And client investment assets were up 8% year-on-year largely driven by flows in addition to market performance.	0.000000
In Home Lending, revenue was down 20% year-on-year on lower production revenue from both lower margins and volumes against a very strong quarter last year, largely offset by higher new lending revenue.	-0.000000
Originations of \$2.7 billion declined 17% with the rise in rates.	-0.000000
And as a result, mortgage loans were down 3%.	-0.000000

Figure 4: Minimum-sentiment call

Starting on page 1 and taking a look at the quarter we had strong performance in each of our businesses despite the continuation of reasonably challenging conditions.	0.000000
Bringing it all together this quarter's result was clean with no significant items and with the firm reporting net income of \$6.3 billion, EPS of \$1.36 and the return on tangible common equity of 13% on \$25.1 billion of revenue.	0.000000
Highlights of the quarter include the highest reported revenue for a third quarter in CIB with IBUs up 58% and markets revenues up 33% with strong performance across the board.	0.000000
Robust core loan growth for the Company of 18% on the back of sustained demand across businesses.	0.000000
And the continuation of strong credit performance including a net release for Oil & Gas.	0.000000
Card sales are back to double-digit growth year on year and we saw a strong positive market reaction to new proprietary products.	0.000000
Finally, we had record consumer deposit growth up 12%.	0.000000
Before I move on to our 2026 resolution filing.	-0.000000
The Board and management believes that we submitted a credible plan and more than met the requirements for the October submission.	0.000000
It was a tremendous effort across the Company involving all businesses and functions and we took many significant actions, perhaps most notably improving the firm's overall liquidity and pre-positioning our material legal entities for both liquidity and capital.	0.000000
We determined this to be in the best interest of the Company, albeit at some cost.	0.000000
We took many other important actions which hopefully you've had the chance to review in our public filings.	0.000000
Moving back to the quarter and moving on to page 2 revenue of \$25.5 billion was up \$2 billion year on year or up 8%.	0.000000
On the back of continued strong growth in core loans net interest income was up \$700 million and is trending for the full year to be above the \$2.3 billion guided last quarter.	0.000000
Non-interest revenue was up \$1.3 billion driven by strong performance in the CIB.	0.000000
Adjusted expense of \$14.5 billion was up \$500 million both year on year and quarter on quarter, largely driven by two notable expense items in Consumer which I'll talk about later as well as the increase in DCC, a charge which took effect this quarter and some higher marketing expenses.	-0.000000
Credit cost of \$1.3 billion in the quarter includes Consumer reserve builds of \$225 million primarily Card.	-0.000000
But against that we have a net reserve release in Wholesale for Oil & Gas of about \$50 million.	-0.000000
So as a net result income was \$6.3 billion.	-0.000000
And while down 8% year on year you will recall that there were a number of significant items in last year's results, most notably significant tax benefits.	0.000000
If you adjust for tax, legal expense and credit reserves net income is up over \$800 million year on year.	0.000000
Dealing with Oil & Gas here, we are encouraged by how quickly investor sentiment and risk appetite for the sector returned as the outlook for both oil and gas prices continued to improve.	0.000000

Figure 5: Maximum-sentiment call

A.4 Preliminary analysis of mis-pricing proxies

We report the cointegration-based and the sentiment based measures of mispricing in Figure 6.

Figure 6 concentrates on the subset of 16 stocks cointegrated with factor prices and reports the cross-sectional average and the upper and lower bounds of the two measures. Both measures feature stationarity with correlation that varies across the cross section of stocks. Figure 6 shows a small positive correlation between the cross-sectional averages of the two measures. In fact, there is rather remarkable heterogeneity in the cross-section as witnessed by Figure 7 that reports the correlation for the two different measures for

Mean Mispricing and Sentiment with Upper and Lower Bounds Over Time

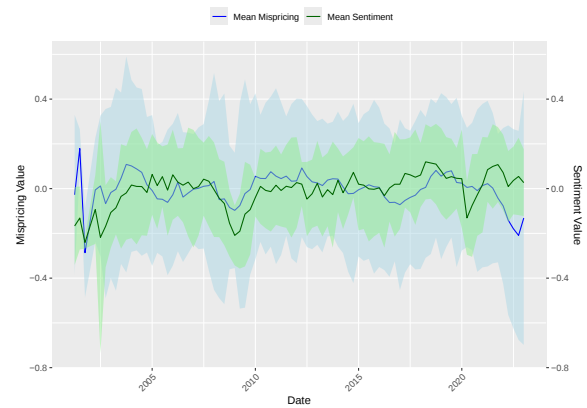


Figure 6: Sentiment-Based and Cointegration-Based Measures of Mispricing

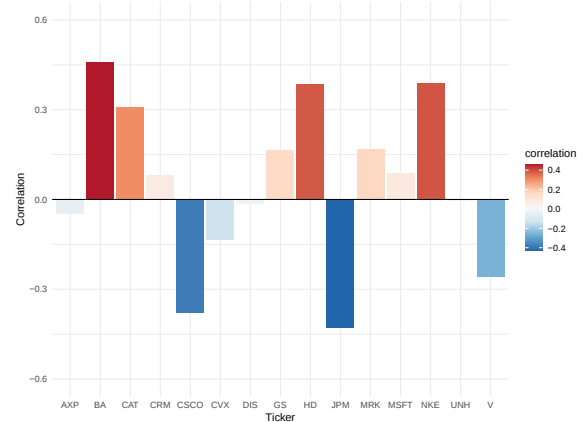


Figure 7: Correlation Between Sentiment-Based and Cointegration-Based Measures of Mispricing

each stocks. Correlation ranges from -0.4 for CSCO and V to 0.7 for BA, and it is positive for 12 out of the 16 considered stocks. Figure 8 illustrates the pattern of correlation for the most negatively and the most positively correlated stocks.

It is also interesting to look at the correlation across tickers for the two measures to evaluate the idiosyncratic versus common nature of mispricing. Figure 9 reports correlation heatmaps across tickers for cointegration based and sentiment based measures of mispricing.

Correlation across tickers of the econometric based measure of mispricing is moderate and predominantly negative, while correlation across tickers of the sentiment-based measures for mispricing is also moderate but predominantly positive. Mispricing based by sentiment features positive comovement potentially related to a common sentiment, while the same does not hold for the econometric-based measure. Interestingly, the heatmaps reveal

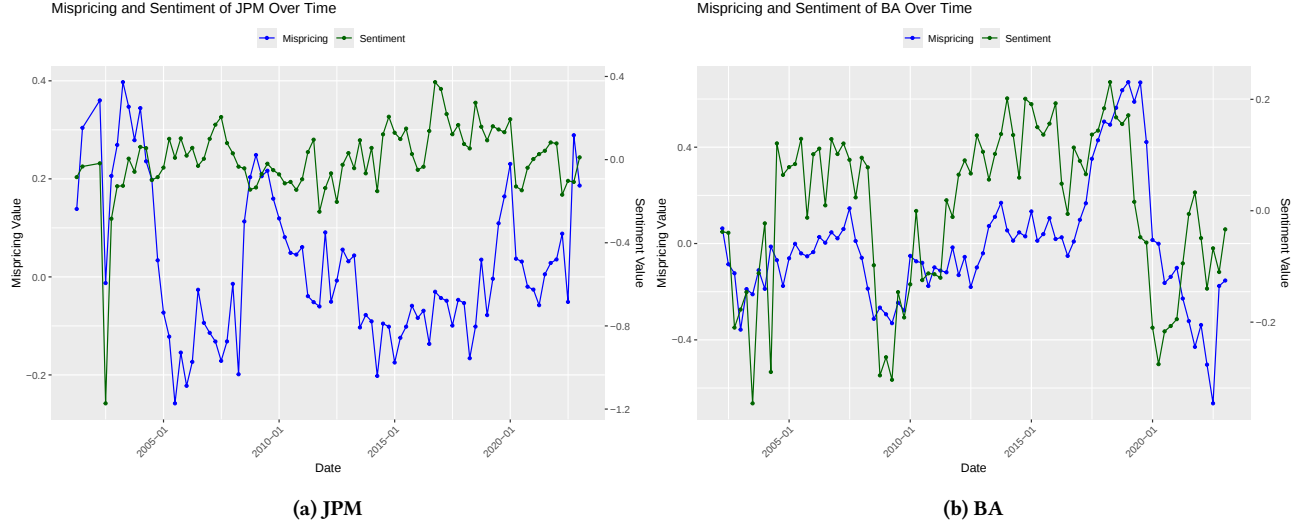


Figure 8: Sentiment-Based and Cointegration-Based Measures of Mispricing for JPM and BA

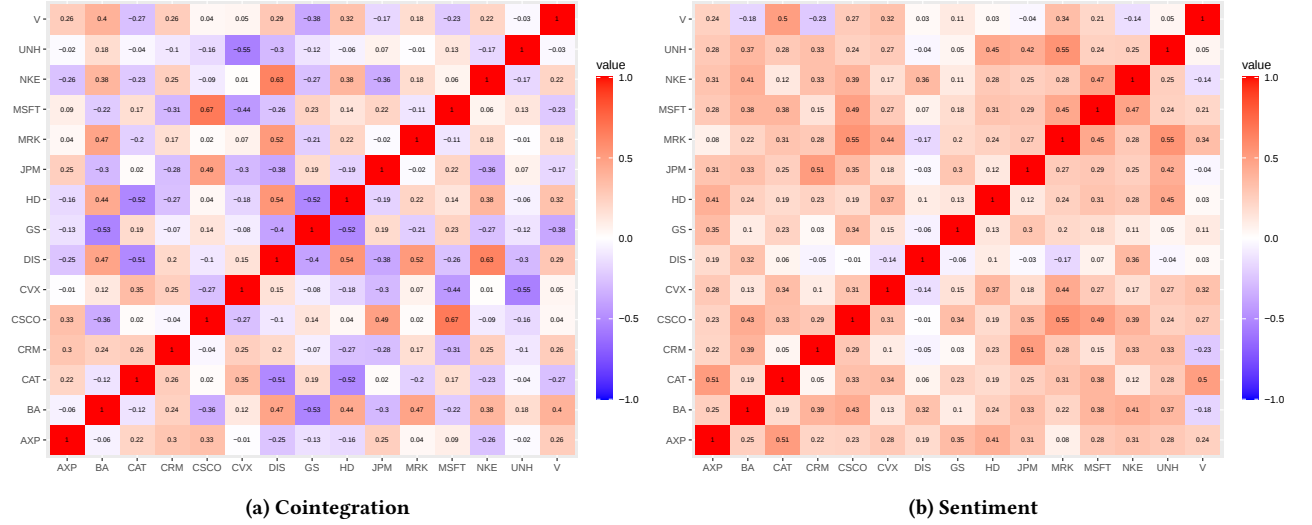


Figure 9: Correlation Across Tickers of Sentiment-Based and Cointegration-Based Measures of Mispricing

a different pattern of correlation across tickers for the two mispricing measures, suggesting different natures for the underlying phenomena captured by the two differently constructed measures.

On the basis of this initial evidence we proceed to unrestricted system estimation of 16 equations projecting the cointegration based mispricing on the sentiment based mispricing and to a restricted panel estimation. That is, we estimate

$$u_{i,t}^{SB} = \beta_i u_{i,t-1}^{CB} + \epsilon_{i,t}, \quad (1)$$

where $u_{i,t-1}^{CB}$ is the ECT for company i in quarter t , while $u_{i,t}^{SB}$ is the sentiment extracted with FinBERT from the earnings call transcript at time t . No constant is included in the regression as the ECT for each company has zero mean by construction, we also estimate a restricted version of the system imposing the restriction $\beta_i = \beta$. No

constant is included as both variables have zero mean. We have an unbalanced panel, due to the availability of the sentiment measure. Our data is at the quarterly frequency, and our estimation window goes from 2001 : Q1 to 2023 : Q2 (although the sample is not balanced across all equations included in the system). We report our restricted and unrestricted system estimates in Tables 1. The estimate of coefficient β from Equation (1) is 0.035, positive and statistically different from zero at the ten per cent level. Overall, the evidence from the estimation confirms that on average ECT and sentiment are mildly positively correlated, with wild variation across different tickers.

Table 1: System and Panel Estimation. This table reports the coefficient estimates for the estimation of the system $u_{i,t}^{SB} = \beta_i u_{i,t-1}^{CB} + \epsilon_{i,t} \cdot u_{i,t}^{CB}$ is the ECT for company i in quarter t , the residual from a regression of the log-price of asset i on risk-drivers. $u_{i,t}^{SB}$ is the FinBERT sentiment extracted from the earnings call transcripts released by company i at time t . No constant is included as both variables have zero mean. The table also reports the coefficient estimates for the panel regression $u_{i,t}^{SB} = \beta u_{i,t-1}^{CB} + \epsilon_{i,t}$ where the restrictions $\beta_i = \beta$, are imposed. Number of observations is not balanced across equations. Standard errors are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively.

Dependent variable: $u_{i,t}^{SB}$								
	AXP	BA	CAT	CRM	CSCO	CVX	DIS	GS
$u_{i,t-1}^{CB}$	-0.049 (0.111)	0.259*** (0.055)	0.285*** (0.099)	0.027 (0.039)	-0.096*** (0.027)	-0.082 (0.067)	-0.008 (0.058)	0.158 (0.106)
First Obs.	2002:2	2002:2	2001:2	2004:3	2003:1	2002:2	2001:2	2002:2
Last Obs.	2023:1	2023:1	2023:1	2023:1	2023:1	2023:1	2023:1	2023:1
R ²	0.002	0.211	0.090	0.007	0.143	0.018	0.0002	0.027
Adjusted R ²	-0.010	0.202	0.079	-0.007	0.131	0.006	-0.012	0.015
	HD	JPM	MRK	MSFT	NKE	UNH	V	Panel
$u_{i,t-1}^{CB}$	0.240*** (0.065)	-0.497*** (0.115)	0.149 (0.096)	0.044 (0.054)	0.193*** (0.051)	-0.002 (0.058)	-0.229** (0.112)	0.035* (0.018)
First Obs.	2001:3	2001:2	2001:2	2001:2	2001:2	2001:2	2008:2	
Last Obs.	2023:1	2023:1	2023:1	2023:1	2023:1	2023:1	2023:1	
R ²	0.144	0.180	0.029	0.008	0.152	0.00002	0.066	0.03
Adjusted R ²	0.134	0.171	0.017	-0.004	0.141	-0.014	0.05	-0.001

Note:

*p<0.1; **p<0.05; ***p<0.01

B ADDITIONAL NONLINEAR RESULTS

The nonlinear specification allows sentiment to modify the speed of adjustment separately for positive and negative ECTs. The estimated interaction is positive and significant only for positive ECTs. Thus, favorable sentiment slows correction when the stock is above its factor-implied trend but has no detectable effect on correction from below.

REFERENCES

- [1] Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063* (2019).
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [3] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov. 2021. Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access* 9 (2021), 29583–29597. <https://doi.org/10.1109/ACCESS.2021.3051823>
- [4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

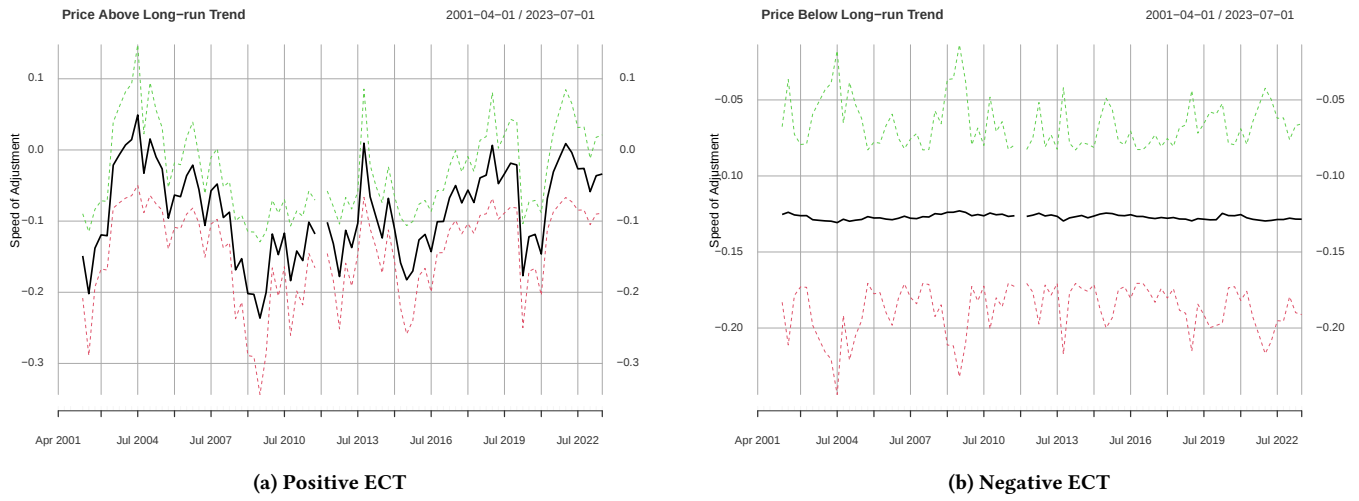


Figure 10: American Express adjustment speeds. Dotted lines show 90% confidence intervals.

, July 2026,

Gabriele Confalonieri, Carlo A. Favero, and Ilaria Leoni