A tale of sentiment driven tail events:

A dynamic quantile model for asset pricing with sentiment

Jozef Baruník
Cathy Yi-Hsuan Chen
Wolfgang Karl Härdle

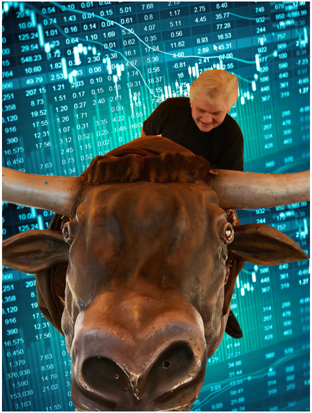
Institute of Economic Studies
Charles University in Prague
Ladislaus von Bortkiewicz Chair of Statistics
Humboldt-Universität zu Berlin

http://ies.fsv.cuni.cz
http://lvb.wiwi.hu-berlin.de

"Forget the dot-com boom with its irrational exuberance and the real estate bubble that was supposed to be invincible: Current market sentiment eclipses all of that"

Jeff Cox, CNBC, March 1 2017

Sentiment moves market



John Maynard Keynes (1936): markets can fluctuate wildly under the influence of investors' "animal spirits," which move prices in a way unrelated to fundamentals.

Sentiment can cause mispricing

Fifty years later...

De Long, Shleifer, Summers, and Waldmann (1990) formalized the role of investor sentiment in financial markets.

- uninformed noise traders base their decisions on sentiment

  ⇒ greater mispricing (Stambaugh et al., 2012)
  ⇒ excess volatility (Dumas et al., 2009)

"Now, the question is no longer, as it was a few decades ago, whether investor sentiment affects stock prices, but rather how to measure investor sentiment and quantify its effects."

(Baker and Wurgler, 2007)

We already know that news moves markets

- Baker and Wurgler (2007) show investor sentiment affects securities whose valuations are highly subjective.
- Large literature Baker and Wurgler (2007), Huang et al. (2014), Da et al. (2015), Shefrin (2007+).
- Our earlier work (Zhang et al., JBES, 2016) shows that textual sentiment provides incremental information about future stock reactions.
- All explain expected returns.

Is average enough?

- Literature quantifies effects of sentiment on cross section of returns, or volatility.
- Grand average can be good measure if we work with expected payoffs.
- But it assumes the same behavior in all points of the distribution.
- Though...
  - bear vs. bull markets
  - extreme negative vs. positive returns

Is average man enough?

Contrarians vs. Trend followers

We already know that we can measure sentiment...

but how to quantify its effect on prices?

We already know that we can measure sentiment...

but how to quantify its effect on prices?

A dynamic quantile model for asset pricing with sentiment

Contribution

- Step forward from classical asset pricing formulated in expectations.
- Provide decision-theoretic foundations of pricing in quantiles of the return distribution instead.
- Link sentiment with cross-section of quantiles of the return distributions.
- Provide simple dynamic quantile asset pricing model with sentiment.
- Confirm empirically on Panel of 100 stocks.

Classical asset pricing

- Investors do not value money directly $\Rightarrow$ utility from consumption, $\mathcal{U}(c_t)$.
- Investor's optimal decision: marginal cost of investment (price of asset) equals marginal benefit (expectation of the returns in $t+1$ times the value of the dollar invested times disctount value β.)

$$p_t = E_t\left[\beta \frac{\mathcal{U}'(c_{t+1})}{\mathcal{U}'(c_t)} R_{t+1}\right] \tag{1}$$

- Fundamental value equation: price is an expected discounted payoff

$$p_t = E_t\left[m_{t+1} R_{t+1}\right] \tag{2}$$

Classical asset pricing

- Working with excess returns $r_{t+1} = R_{t+1} - R^f$, this becomes

$$0 = E_t \left[m_{t+1} r_{t+1} \right] \tag{3}$$

- with m_{t+1} being pricing kernel (rate at which investor is willing to substitute one unit of consumption now for later.)
- Using definition of covariance, we arrive to classical factor models

$$E_t(r_{t+1}) = \beta\lambda \tag{4}$$

- Depending on m_{t+1}, we can arive to factor models as source of risk.

A route towards quantile optimization

- Previous theory works with classical von-Neumann-Morgensterns expected utility
- X is preferred to Y if there exist utility function $\mathcal{U}(.)$ such that

$$X \succeq Y \Leftrightarrow E[\mathcal{U}(X)] \geq E[\mathcal{U}(Y)] \tag{5}$$

- Manski (1988), Rostek (2010) looks at quantile preferences (τ quantile of the utility distribution)

$$X \succeq Y \Leftrightarrow Q_\tau[\mathcal{U}(X)] \geq Q_\tau[\mathcal{U}(Y)] \tag{6}$$

- Maximising lower quantile is more risk-averse than higher quantile (example of portfolio).

Quantile utility preference

- Hence instead of expectations,we may think of looking at agent who wants to maximize her stream of the future quantile utilities such as

$$Q_{\tau,t}\left[\beta\frac{\mathcal{U}'(c_{t+1})}{\mathcal{U}'(c_t)}r_{t+1}\right] \tag{7}$$

- Agent has quantile utility preference instead of standard expected utility.
- de Castro and Galvao (2017) shows that Euler equation then simplifies to

$$0 = Q_{\tau}\left[\beta(\tau)\frac{\mathcal{U}'(c_{t+1})}{\mathcal{U}'(c_t)}r_{t+1}\right] \tag{8}$$

hence discount factor will be τ dependent.

τ-dependent pricing kernel with sentiment

- Having agents maximizing in quantiles, and measuring discount factor via simple linear factor models,
- we can define $\tau-$ dependent pricing kernels as linear combination of factors

$$m_{t+1}(\tau) = \beta(\tau)\frac{\mathcal{U}'(c_{t+1})}{\mathcal{U}'(c_t)} = F_t^\top \beta_{FF}(\tau) \tag{9}$$

- Where one of the factors is also sentiment.
- $F_t = (F_{1,t}, F_{2,t}, \ldots, S_t)^\top$
- Factors generally represent atheoretical proxy for aggregate consumption or marginal utility growth $\Rightarrow$ source of risk

Empirical validation

To test these theoretical assertions empirically and quantify the role of sentiment, one can follow two strategies.

- Examine if sentiment is priced factor in traditional asset pricing models via construction of factor in a simple linear model.
- Estimate structural equations (mostly via GMM).

We focus on the first strategy: if sentiment is priced, linear factor model will uncover it.

Panel Quantile Regression

We use panel quantile regression where quantile function of $r_{i,t}$ return is conditioned on information in $X_{i,t}$ for $\tau \in (0,1)$ as

$$Q_{r_{i,t+1}}(\tau|X_{i,t}) = \alpha_i(\tau) + X_{i,t}^\top \beta(\tau), \quad \tau \in (0,1), \tag{10}$$

Penalized fixed effects estimator recovers parameters (Koenker (2004))

$$\min_{\alpha(\tau),\beta(\tau)} \sum_{t=1}^{n} \sum_{i=1}^{t_i} \rho_\tau(r_{i,t+1} - \alpha_i(\tau) - X_{i,t}^\top \beta(\tau)) + \lambda \sum_{i=1}^{n} |\alpha_i(\tau)|, \tag{11}$$

where $\rho_\tau(u) = u\,(\tau - I(u(<0)))$ is the quantile loss function, $\sum_{i=1}^{n} |\alpha_i|$ is l_1 penalty that controls variability introduced by the large number of estimated parameters.

Panel Quantile Regression

- Many recent advances in estimation, we still use Koenker (2004).
- Having $T >> N$, we account and to control for unobserved heterogeneity among financial assets.
- Morever, in contrast to literature, we consider individual fixed effects to have distributional effects and we concentrate on each quantile separately rather than minimizing through several quantiles.

Data

- We look at panel of 100 stocks
- Sentiment variables: distilled from Nasdaq articles

Nasdaq Articles

- Terms of Service permit web scraping
- Currently $>$ 580k articles between October 2009 and January 2017
- Data available at RDC

There is a lot of news...

Dimensions of News

- Source of news
  - Official channel: government, federal reserve bank/central bank, financial institutions
  - Internet: blog, social media, message board
- Content of news: signal vs. noise
- Type of news
  - Scheduled vs. non-scheduled
  - Expected vs. unexpected
  - Specific-event vs. continuous news flows

The Power of Words: Textual Analytics

- Sentiment analysis
  - Lexica projection : positive, neutral and negative
  - Machine learning : text classification

Unsupervised Projection

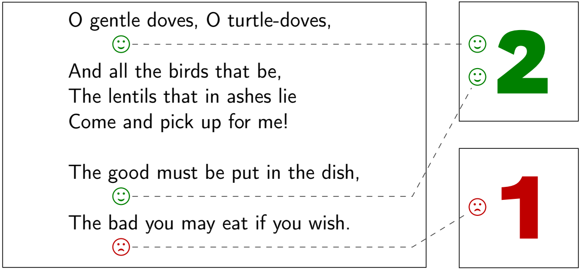


Figure 1: Example of Text Numerisization

- Many texts are numerisized via lexical projection
- Goal: Accurate values for positive and negative sentiment

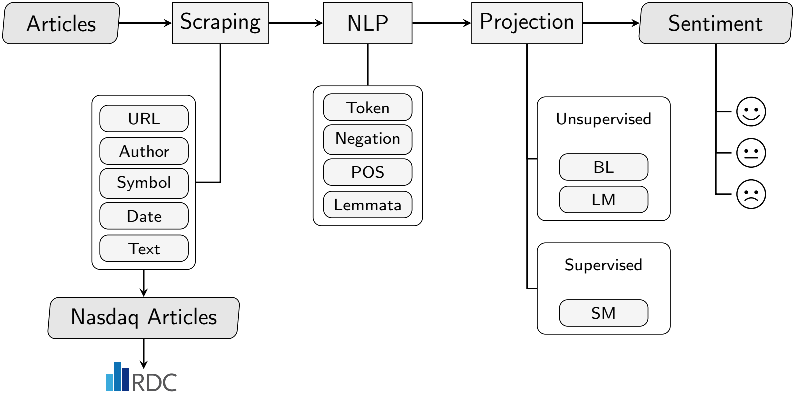
Examples

Supervised Projection

We use supervised projection (Zhang et al., JBES, 2016)

- Training data: Financial Phrase Bank by Malo et al. (2014)
  - Sentence-level annotation of financial news
  - **Manual annotation** of 5,000 sentences by 16 annotators: to incorporate human knowledge
  - Example: "profit" with different semantic orientations
    - Neutral in "profit was 1 million"
    - Positive in "profit increased from last year"

How to gather Sentiment Variables?



Lexicon-based Sentiment

Consider document i, positive sentiment Pos_i, positive lexicon entries W_j ($j = 1, \ldots, J$) and count frequency of those entries w_j:

$$Pos_i = n_i^{-1} \sum_{j=1}^{J} \mathbf{I}\left(W_j \in L\right) w_j \tag{12}$$

with n_i: number of words in document i (e.g. sentence)

Equivalent calculation of negative sentiment Neg_i

Sentence-level Polarity

$$Pol_i = \begin{cases} 1, & \text{if } Pos_i > Neg_i \\ 0, & \text{if } Pos_i = Neg_i \\ -1, & \text{if } Pos_i < Neg_i \end{cases} \tag{13}$$

for sentence i

- Measure sentiment on sentence level

Regularized Linear Models (RLM)

- Training data $(X_1, y_1) \dots (X_n, y_n)$ with $X_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$
- Linear scoring function $s(X) = \beta^\top X$ with $\beta \in \mathbb{R}^p$

Example

Regularized training error:

$$n^{-1} \sum_{i=1}^{n} \underbrace{L\{y_i, s(X)\}}_{\text{Loss Function}} + \lambda \underbrace{R(\beta)}_{\text{Regularization Term}} \tag{14}$$

with hyperparameter $\lambda \geq 0$

RLM Estimation

- Optimize via Stochastic Gradient Descent More
- 5-fold cross validation More
- Oversampling More
- Choice of: $L(\cdot), R(\cdot), \lambda$, X (n-gram range, features) ...
- Three categories: one vs. all sub-models

Bullishness

$$B = \log \left\{ \frac{1 + n^{-1} \sum_{j=1}^{n} \mathbf{I}\left(Pol_j = 1\right)}{1 + n^{-1} \sum_{j=1}^{n} \mathbf{I}\left(Pol_j = -1\right)} \right\} \tag{15}$$

by Antweiler and Frank (JF, 2004) with $j = 1, \ldots, n$ sentences in document.

- $B_{i,t}$ accounts for bullishness of company i on day t
- Consider $BN_{i,t} = \mathbf{I}\left(B_{i,t} < 0\right) B_{i,t}$

Sentiment as factor

- Aggregate market sentiment as possible risk factor.
- We control also for firm-specific sentiment and volatility
- Negative sentiment is used in literature as it captures “fear”, and can be related to VIX (Da et al., 2015)
- Following high investor sentiment, aggregate returns are low (Baker and Wurgler, 2007)
- Overly optimistic beliefs about future cash flows is not justified by fundamentals.

A dynamic quantile model for asset pricing with sentiment

A linear asset pricing model Fama-French Factors

$$Q_{r_{i,t+1}}(\tau|\Omega_t) = \alpha_i(\tau)+\beta_1(\tau)B_{i,t}+\beta_2(\tau)\sigma_{i,t}+\beta_3(\tau)|BN_t|+F_t^\top \beta_{FF}(\tau) \tag{16}$$

and in a sense of classical factor literature, we test if sentiment prices quantiles of the excess asset returns.

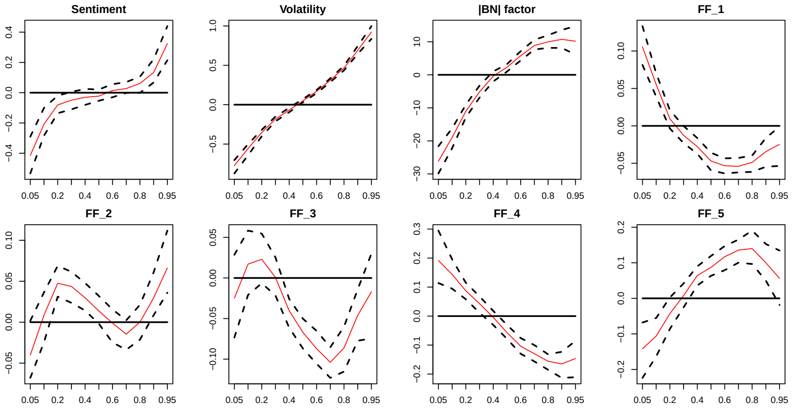
- Coefficients then capture marginal effects of pricing factors on the τ quantile of future returns.
- Coefficient varying across τ implies marginal effect of factor on returns vary along conditional distribution of returns.
- Coefficients constant over τ $\Rightarrow$ classical expectations should work well.

Results

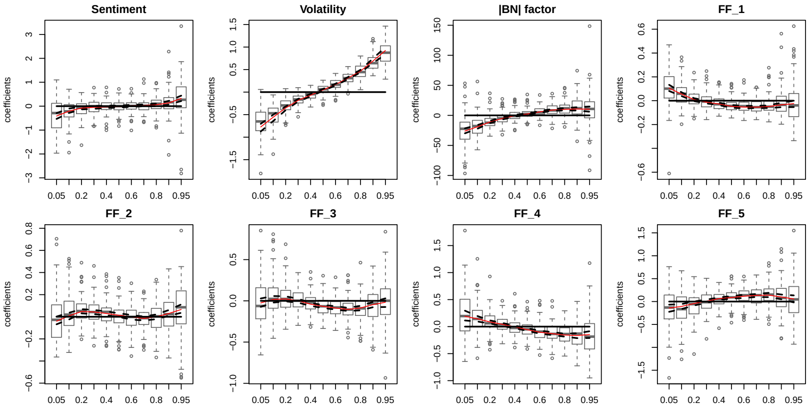
We estimate the model on:

- Panel of 100 stocks.
- 10 main sectors Details

Results: Panel of 100 stocks



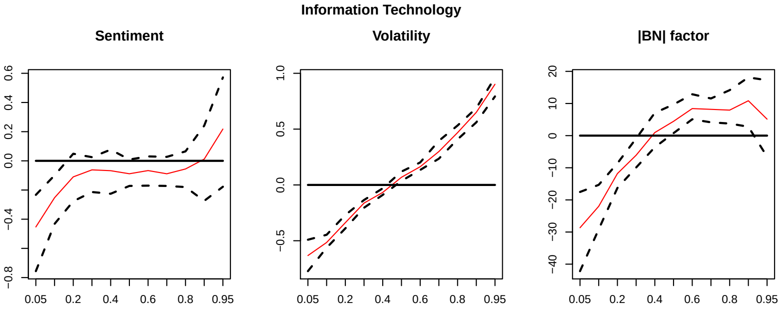
Results: Panel of 100 stocks



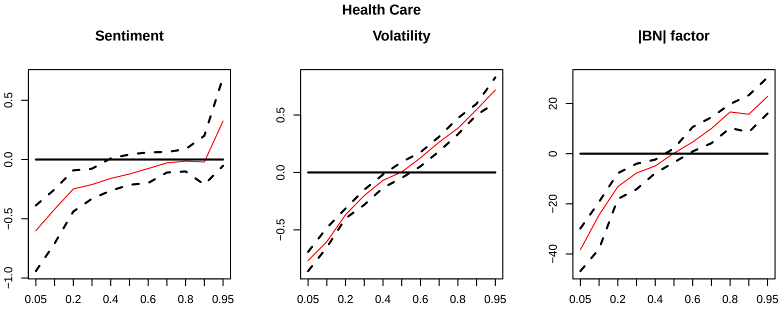
Empirical Results

- Tails are heavily influenced.
- Sentiment and volatility effects similarly.
- $\beta(\tau) \neq 0$ for most of the τs.
- Asymmetric impact of market sentiment.
- Holds even after we control for firm specific sentiment and volatility.
- An increase in negative bullishness have positive effect on right tail, and negative effect on left tail.
- Contrary to literature, factors explain daily data in quantiles too.

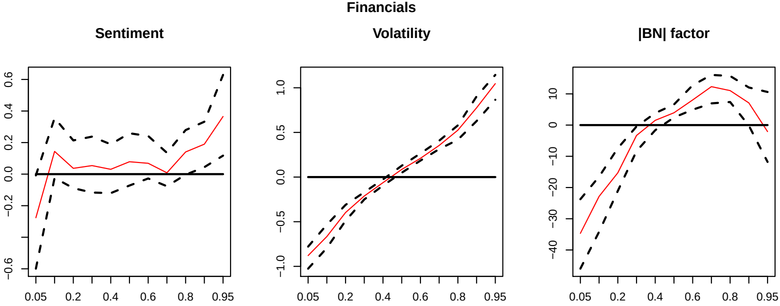
Results: Sectors



Results: Sectors

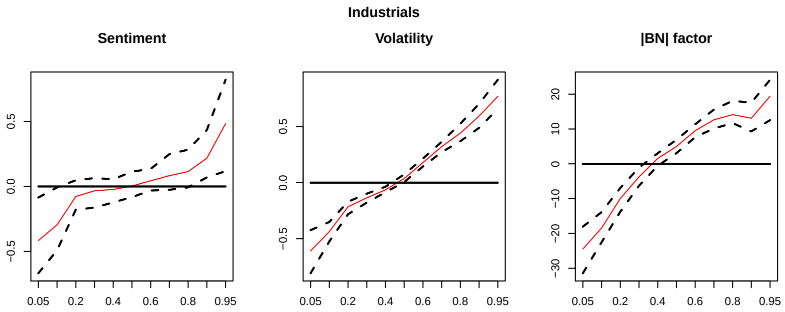


Results: Sectors

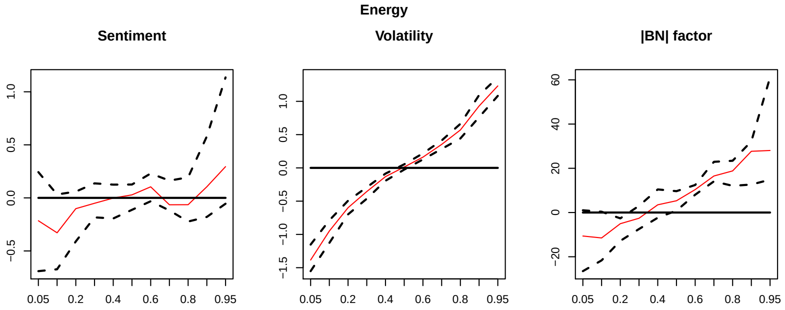


Results: Sectors

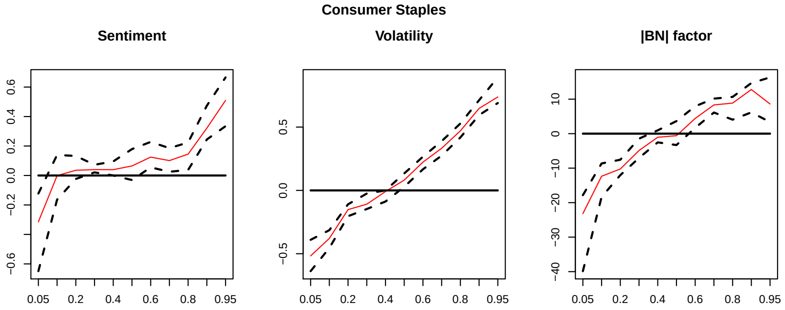
Results: Sectors



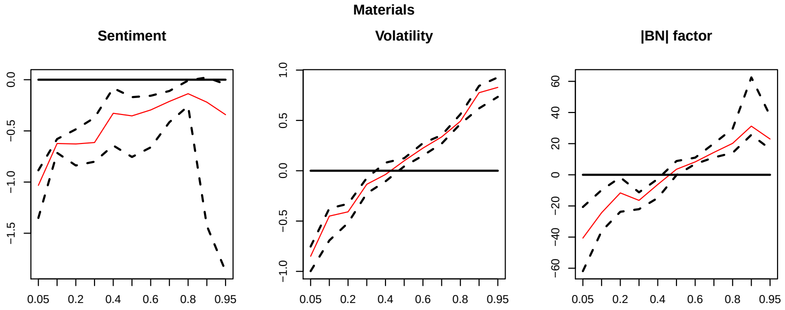
Results: Sectors



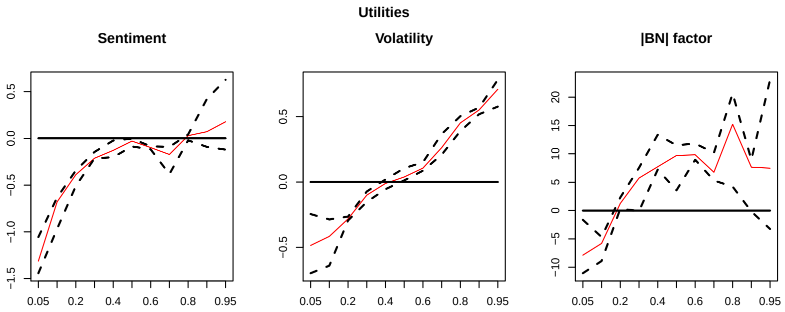
Results: Sectors



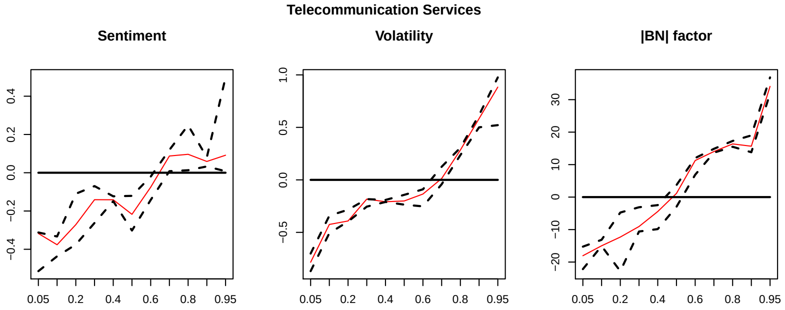
Results: Sectors



Results: Sectors



Results: Sectors



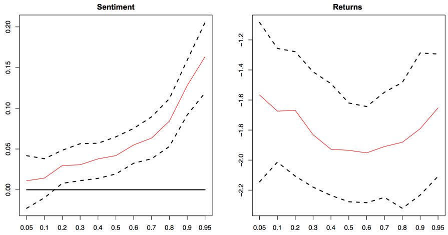
Results: Sectors

- Effects are uniform also at sector level.
- More pronounced asymmetry and impact on left tail.
- Right tail less significantly explained except Utilities.

Results: Volatility

In addition, we explain firm-specific volatility by firm-specific sentiment as

$$Q_{\sigma_{i,t+1}}(\tau|\Omega_t) = \alpha_i(\tau) + \beta_1(\tau)B_{i,t} + \beta_2(\tau)r_{i,t} \qquad (17)$$



- Increase in sentiment increases volatility.
- Right tail (high risk) is largely impacted

Summary

- We provide new dynamic quantile model for asset pricing with sentiment
- We connect investor sentiment distilled from public news with cross-section of future return's quantiles.

A tale of sentiment driven tail events:

A dynamic quantile model for asset pricing with sentiment

Jozef Baruník
Cathy Yi-Hsuan Chen
Wolfgang Karl Härdle

Institute of Economic Studies
Charles University in Prague
Ladislaus von Bortkiewicz Chair of Statistics
Humboldt-Universität zu Berlin

http://ies.fsv.cuni.cz
http://lvb.wiwi.hu-berlin.de

Bibliography

Antweiler, W. and Frank, M. Z.
Is All That Talk Just Noise?
J. Finance, 2004

Baker, M., and J. Wurgler.
Investor sentiment and the cross-section of stock returns
Journal of Finance, 2006

de Castro, L. I. and A. F. Galvao
Dynamic quantile models of rational behavior
2017

Da, Z., Engelberg J. and Gao, P.
The Sum of All FEARS Investor Sentiment and Asset Prices
Review of Financial Studies, 2015

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann.
Noise trader risk in financial markets.
Journal of Political Economy, 1990

Dumas, B., Kurshev, A., Uppal, R.
Equilibrium Portfolio Strategies in the Presence of Sentiment Risk and Excess Volatility
Journal of Finance, 2009

Fama, E. and K. French.
A Five-Factor Asset Pricing Model
J. Financial Econom., 2015

Huang, D., Jun Tu, J., Jiang, F., and Zhou, G.
Investor Sentiment Aligned: A Powerful Predictor of Stock Returns
Journal of Finance, 2014

Keynes, J. M.
The general theory of employment, interest and money.
London: Macmillan, 1936

Koenker, R.
Quantile regression for longitudinal data.
Journal of Multivariate Analysis, 2004

Manski, C.F.
Ordinal utility models of decision making under uncertainty.
Theory and Decision, 1988

Rostek, M.
Quantile maximization in decision theory.
The Review of Economic Studies, 2010

Stambaugh, R.F., Yu, J.F., Yuan, Y.
The short of it: Investor sentiment and anomalies.
Journal of Financial Economics, 2012

Zhang, J., Chen C. Y., Härdle, W. K. and Bommes, E.
Distillation of News into Analysis of Stock Reactions
J. Bus. Econom. Statist., 2016

Appendix

Tagging Example - BL

... McDonald's has an obesity **problem** that continues to get **worse**. And that's nothing to do with the food itself, but rather the huge menus that can now double as medieval fortification. For perspective, the chain's menu has grown 70% since 2007. And while more offerings might seem **like** a **good** thing, large menus result in **slower** service and more flare-ups between franchisees and the corporation.
Bloated menus raise inventory costs for smaller franchisees and **lead** to lower profit margins. The McDonald's corporate franchise fee is based upon sales instead of profits, making it a smaller **concern** for the company overall. ...

3 **positive words** and 5 **negative words**

TXTMcDbm
Article source

Tagging Example - LM

... McDonald's has an obesity **problem** that continues to get **worse**. And that's nothing to do with the food itself, but rather the huge menus that can now double as medieval fortification. For perspective, the chain's menu has grown 70% since 2007. And while more offerings might seem like a **good** thing, large menus result in **slower** service and more flare-ups between franchisees and the corporation.
Bloated menus raise inventory costs for smaller franchisees and lead to lower profit margins. The McDonald's corporate franchise fee is based upon sales instead of profits, making it a smaller **concern** for the company overall. ...

1 **positive word** and 4 **negative words**

TXTMcDlm

Back

Web Scraping

- Databases to buy?
- Automatically extract information from web pages
- Transform unstructured data (HTML) to structured data
- Use HTML tree structure to parse web page
- Legal issues
  - Websites protected by copyright law
  - Prohibition of web scraping possible
  - Comply to Terms of Service (TOS)

Back

Natural Language Processing (NLP)

- Text is unstructured data with implicit structure
  - Text, sentences, words, characters
  - Nouns, verbs, adjectives, ..
  - Grammar
- Transform implicit text structure into explicit structure
- Reduce text variation for further analysis
- Python Natural Language Toolkit (NLTK)
- TXTnlp

Back

Tokenization

- String

  ``'McDonald's has its work cut out for it. Not only are sales falling in the U.S., but the company is now experiencing problems abroad.''

- Sentences

  ``'McDonald's has its work cut out for it.'',
  ``'Not only are sales falling in the U.S., but the company is now experiencing problems abroad.''

- Words

  ``'McDonald'', '''s'', ''has'', ''its'', ''work'', ''cut'', ''out'' ...

Negation Handling

- "not good" $\neq$ "good"
- Reverse polarity of word if negation word is nearby
- Negation words
  `"n't", "not", "never", "no", "neither", "nor", "none"`

Part of Speech Tagging (POS)

- Grammatical tagging of words
  - `dogs - noun, plural (NNS)`
  - `saw - verb, past tense (VBD) or noun, singular (NN)`
- Penn Treebank POS tags
- Stochastic model or rule-based

Lemmatization

- Determine canonical form of word
  - `dogs - dog`
  - `saw (verb) - see and saw (noun) - saw`
- Reduces dimension of text
- Takes POS into account
  - Porter stemmer: `saw (verb and noun) - saw`

Back

Loss Functions for Classification

- Logistic: Logit

$$L\{y, s(X)\} = \log(2)^{-1} \log[1 + \exp\{-s(X)y\}] \tag{18}$$

- Hinge: Support Vector Machines

$$L\{y, s(X)\} = \max\{0,\ 1 - s(X)y\} \tag{19}$$

Back

Regularization Term

- L2 norm

$$R(\beta) = 2^{-1} \sum_{i=1}^{p} \beta_i^2 \tag{20}$$

- L1 norm

$$R(\beta) = \sum_{i=1}^{p} |\beta_i| \tag{21}$$

Back

RLM Example

Sentence 1: "The profit of Apple increased."
Sentence 2: "The profit of the company decreased."

$$y = (1, -1) \quad (22)$$

$$X = \begin{array}{c} \\ \textit{the} \\ \textit{profit} \\ \textit{of} \\ \textit{Apple} \\ \textit{increased} \\ \textit{company} \\ \textit{decreased} \end{array} \begin{array}{c} \begin{array}{cc} X_1 & X_2 \end{array} \\ \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \end{array} \quad (23)$$

Back

k-fold Cross Validation (CV)

- Partition data into k complementary subsets
- No loss of information as in conventional validation
- Stratified CV: equally distributed response variable in each fold

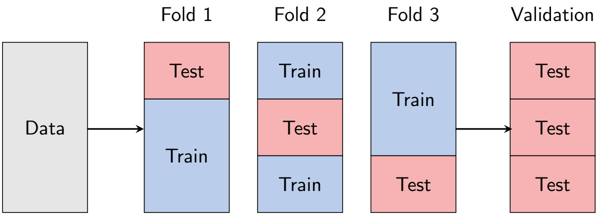


Figure 2: 3-fold Cross Validation

Back

Oversampling

- Härdle (2009) Trade-off between Type I and Type 2 error in classification Error types
- Balance size of neutral sentences and ones with polarity in sample
- Duplicate sentences within folds of stratified cross validation until the sample is balanced

Back

Classification Error Rates

- Type I error rate $= \text{FP}/(\text{FP} + \text{TN})$
- Type II error rate $= \text{FN}/(\text{FN} + \text{TP})$
- Total error rate $= (\text{FN} + \text{FP})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$

with TP as true positive, TN as true negative, FP as false positive and FN as false negative.

Back

Stochastic Gradient Descent (SGD)

- Approximately minimize loss function

$$L(\theta) = \sum_{i=1}^{n} L_i(\theta) \tag{24}$$

- Iteratively update

$$\theta_i = \theta_{i-1} - \eta \, \frac{\partial L_i(\theta)}{\partial \theta} \tag{25}$$

SGD Algorithm

1. Choose learning rate η
2. Shuffle data
3. For $i = 1, \ldots, n$, do:

$$\theta_i = \theta_{i-1} - \eta \, \frac{\partial L_i(\theta)}{\partial \theta}$$

Repeat 2 and 3 until approximate minimum obtained.

SGD Example

$X \sim \mathrm{N}(\mu, \sigma)$ and $x_1, ..., x_n$ as randomly drawn sample

$$\min_{\theta} \; n^{-1} \sum_{i=1}^{n} (\theta - x_i)^2$$

Update step

$$\theta_i = \theta_{i-1} - 2\eta(\theta_{i-1} - x_i)$$

Optimal gain

Set $2\eta = 1/i$ and obtain $\theta_n = \bar{x}$ with $\bar{x}$ as sample mean.

SGD Example ctd

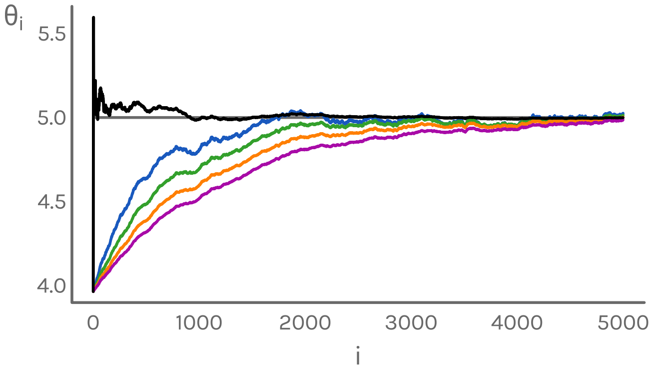


Figure 3: Estimate Mean via SGD, $x_t \sim N(5,1)$

$\eta \in \{1/t,\ 1/1000,\ 1/1500,\ 1/2000,\ 1/2500\}$ TXTSGD

Back

Evaluation Supervised Learning

True \ Pred	-1	0	1	Total
-1	**1,983**	298	254	2,535
0	96	**2,134**	305	2,535
1	105	469	**1,961**	2,535
Total	2,184	2,901	2,520	7,605

Table 1: Confusion Matrix - Supervised Learning with Oversampling

Back

Abbreviations

Sector	Abbreviation
Consumer Discretionary	CD
Consumer Staples	CS
Energy	EN
Financials	FI
Health Care	HC
Industrials	IN
Information Technology	IT
Materials	MA
Telecommunication	TE
Utilities	UT

Table 2: Sector Abbreviations

back

Fama-French 5 factors

$FF1$ - the Mkt factor: excess return on the market index

$FF2$ - the SMB factor: (Small Minus Big) the average return on the nine small-stock portfolios minus that on the nine big-stock portfolios.

$FF3$ - the HML factor: (High Minus Low) the average return on the two value-stock portfolios minus that on the two growth-stock portfolios

Back

Fama-French 5 factors cont.

$FF4$ - the RMW factor: (Robust Minus Weak) the average return on the two robust operating profitability portfolios minus that on the two weak operating profitability portfolios

$FF5$ - the CMA factor: (Conservative Minus Aggressive) the average return on the two conservative investment portfolios minus that on the two aggressive investment portfolios

Back