



# Finding diamonds in the rough

An exploration of sell-side analyst coverage versus company fundamentals to identify underfollowed companies

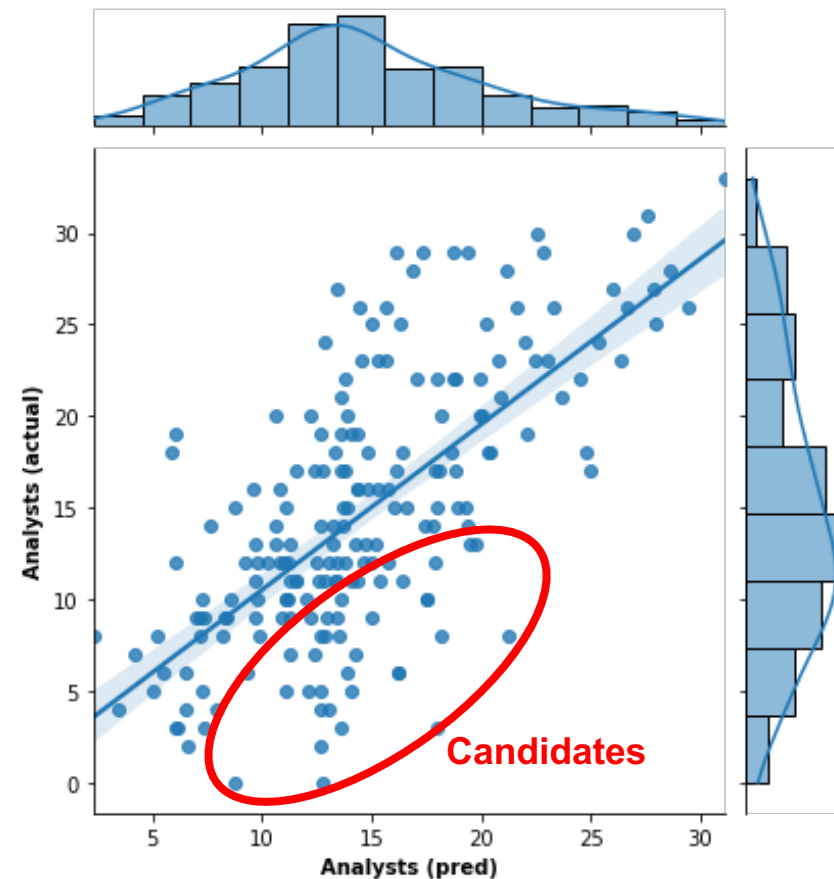
**Sam Reiff**

Aug 2021

# Executive summary

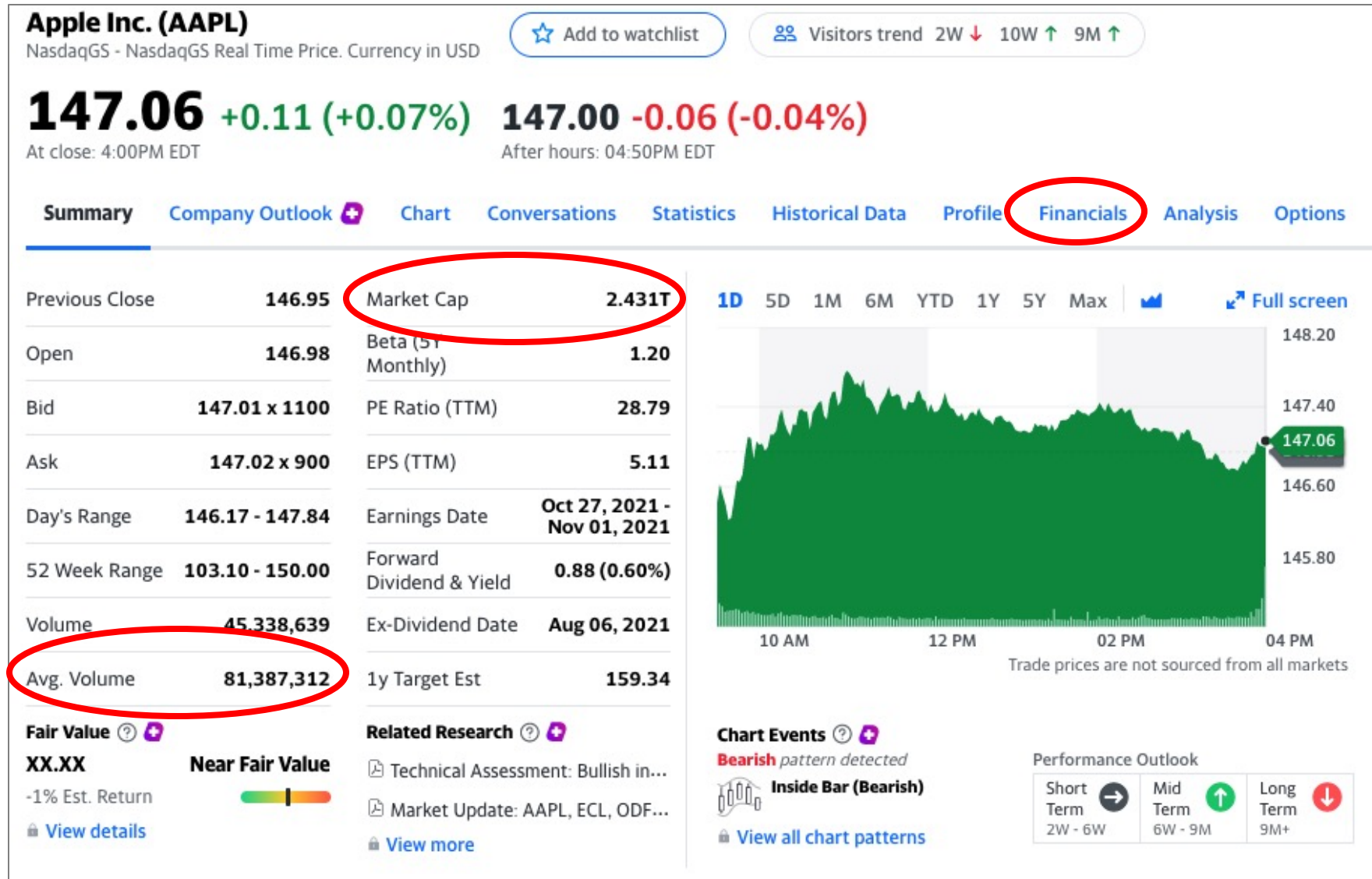
- **Motivation:** Investment opportunities could be uncovered by assessing a stock's sell-side research coverage, which proxies investor 'awareness'; equity research sample [here](#)
- **Objective:** Identify relationships between company fundamentals (e.g., revenue, profitability) and level of analyst coverage
- **Conclusion:** Fundamental features and coverage can be modeled with modest confidence
  - Target: Publishing analysts
  - Ridge  $R^2$  on test: 0.4598
  - Ridge MAE on test: 4.1630
  - Representative candidates identified

Ridge model predictions versus test set actuals



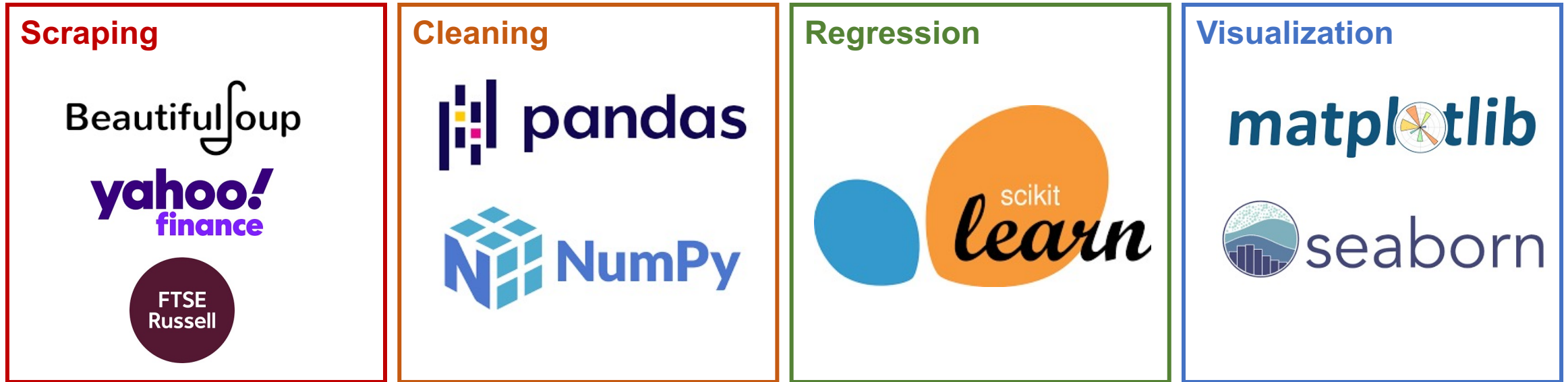
**A Ridge Regression Model may serve as a useful starting point for identifying undercovered equities**

# Primary data source: Yahoo! Finance



# Methodology

## Project workflow



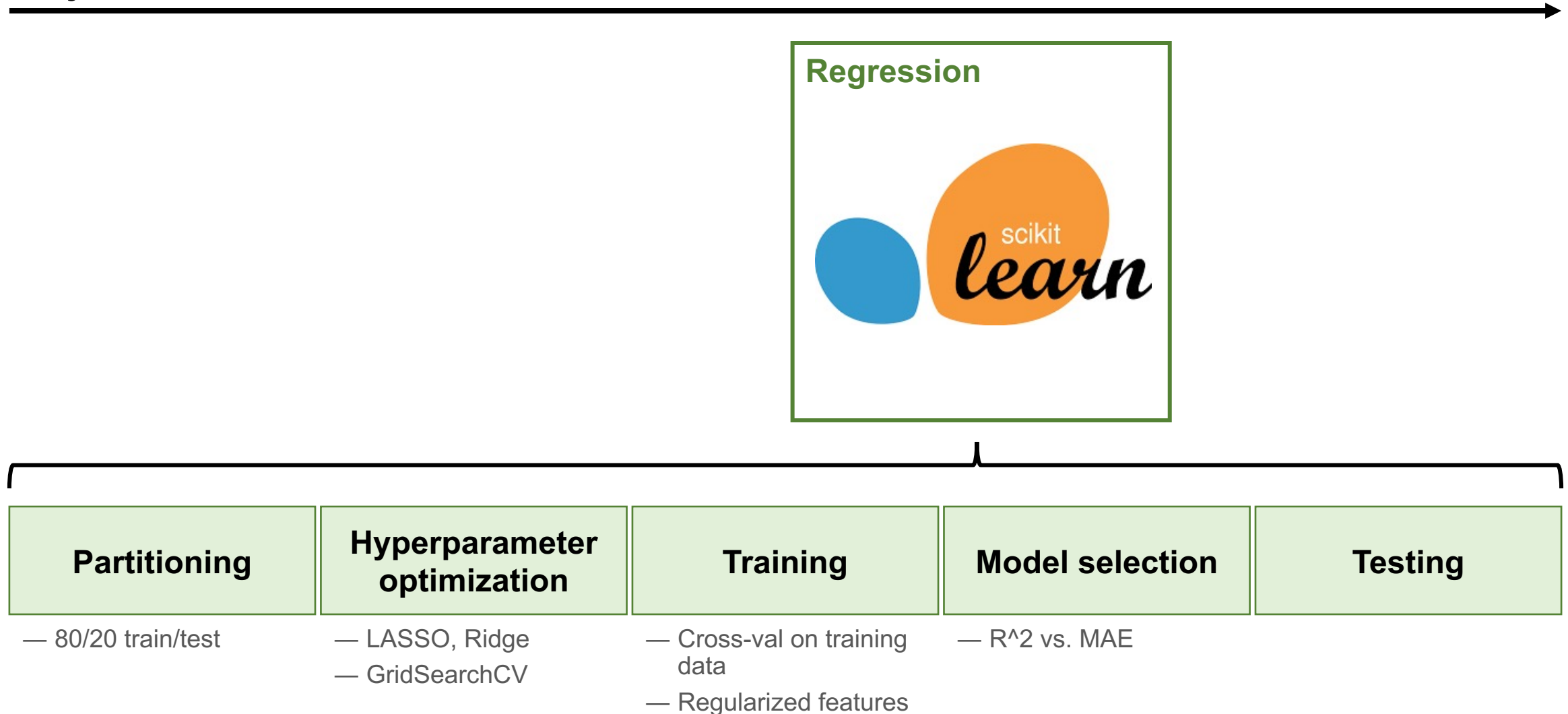
- **>1,020 equities analyzed**
- **>10 features scraped from >3,000 web pages**
  - Publishing analysts
  - Size: Enterprise value, market cap, trading volume, basic shares, float shares, revenues, assets
  - Profitability: Gross margin, operating margin, EBITDA margin, ROA, ROE
- **Supplemented with features via YF API**
  - Industry, sector, growth

Refer to glossary in the appendix for definition of financial terms

Non-scraped data was pulled via the Python yfinance package that leverages the YF API

# Methodology

## Project workflow



# Results

## Train

### ▪ Vanilla LR:

- CV score: 0.5140
- $R^2$  on train: 0.5425
- MAE on train: 3.9218 (analysts)

### ▪ LASSO:

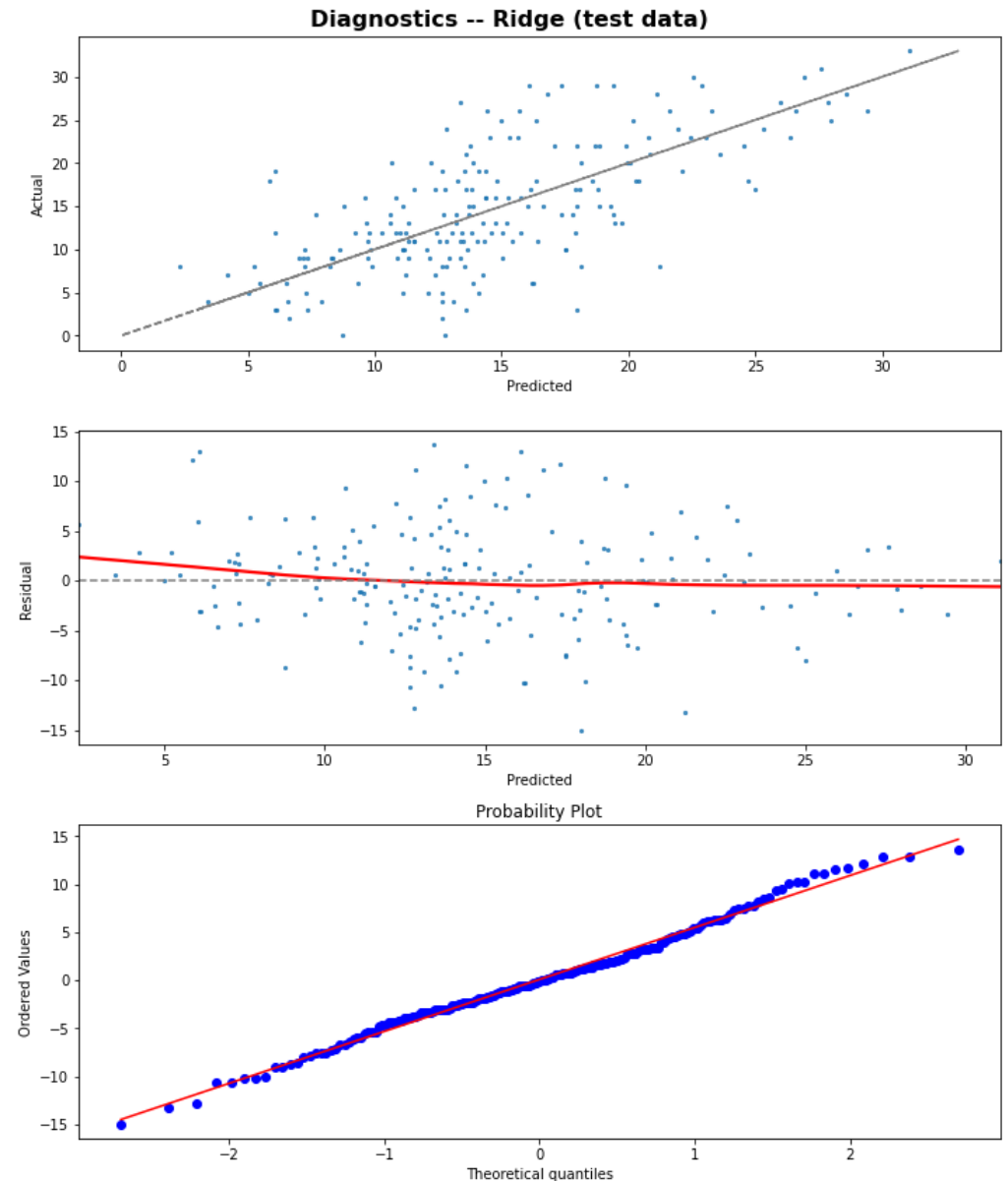
- CV score: 0.4650
- $R^2$  on train: 0.4848
- MAE on train: 4.1964 (analysts)

### ▪ Ridge:

- **CV score: 0.5148**
- $R^2$  on train: 0.5417
- MAE on train: 3.9361 (analysts)

## Test

Ridge  $R^2$ : 0.4598  
Ridge MAE: 4.1630 (analysts)



# Conclusions

## Target companies based on variance from Ridge model predictions

Equity	Name	Coverage	Pred. coverage	Industry	MC (\$)	Vol. (sh)	LTM revenue (\$)	MRQ assets (\$)	Notes
BEPC	Brookfield Renewable Corp	1	12	Utilities	7B	0.9M	3B	39B	Renewable energy rollup?
★ NUAN	Nuance Comms.	4	16	App. software	17B	5.0M	1B	4B	AI healthcare, could be very interesting.. poor ROA/ROE
PLTR	Palantir Technologies	8	21	Ent. software	42B	15.0M	1B	3B	Meme stock..
★ POOL	Pool Corporation	6	16	Leisure	20B	0.3M	5B	2B	Great ROE, COVID re-open play?
★ SCCO	Southern Copper Corp	8	18	Copper	50B	1.5M	10B	17B	Strong ROA, ROE
★ SCI	Service Corp. Int'l	4	16	Pers. services	11B	0.9M	4B	15B	Strong ROE, boring!!!

**The Ridge model isn't highly accurate, but can identify relevant companies with investment motifs that are worth considering for further financial analysis**

# Further due diligence/future work

- **Feature inclusion**

- Forward-looking estimates
- # sell-side analysts per sector
- Time since IPO
- Stock return volatility
- Affinity for M&A

- **Scraping enhancement**

- **Broaden data set**

- Russell 2000
- Global SMID-large cap equities
- Sector-specific

- **Feature engineering/modeling**

- Polynomial feature transformation
- Regime changes in investor interest



**Sam Reiff**

**reiff.sam@gmail.com**

**479.426.3700**

# Appendix

# Disclosure

The preceding analysis and commentary express my research opinions. You should assume that as of the publication date of this report, I have taken a position in any security covered herein, and therefore stand to realize significant gains in the event that the price of said securities rises or declines. Following the dissemination of this report, I intend to continue transacting in the securities covered herein, and I may be long, short, or neutral at any time hereafter regardless of my initial conclusions or recommendations. All expressions of opinion are subject to change without notice, and I do not undertake to update this report or any information contained herein. I have no obligation to advise any other party regarding their historical, current, and future trading activities.

This research presentation expresses my research opinions, which I have based upon interpretation of certain facts and observations, all of which are based upon publicly available information, and all of which are set out in this research report. Unless noted otherwise, stated financial figures represent those available through public filings. Any investment involves substantial risks, including complete loss of principal investment. Any forecasts or estimates are for illustrative purposes only and should not be taken as limitations of the maximum possible loss or gain. Any information contained in this report may include forward-looking statements, expectations, pro-forma analyses, estimates, and projections. You should assume these types of statements, expectations, pro-forma analyses, estimates, and projections may turn out to be incorrect for reasons beyond my control. This is not investment or accounting advice nor should it be construed as such. Use of my research is at your own risk. You should conduct your own research and due diligence, with assistance from financial, legal, and tax professionals, before making any investment decision with respect to the securities covered herein.

To the best of my ability and belief, as of the date hereof, all information contained herein is accurate and reliable and does not omit to state material facts necessary to make the statements herein not misleading, and all information has been obtained from public sources I believe to be accurate and reliable, and who are not insiders or connected persons of the business entity covered herein or who may otherwise owe any fiduciary duty or duty of confidentiality to the issuer or to any other person or entity that was breached by the transmission of information to me. However, I recognize that there may be non-public information in the possession of the business entity that is the subject of this report, or any insiders associated therewith that has not been publicly disclosed. Therefore, such information contained herein is presented “as is”, without warranty of any kind – whether express or implied. I make no other representations, express or implied, as to the accuracy, timeliness, or completeness of any such information or with regard to the results to be obtained from use.

This report’s estimated fundamental value only represents a best-efforts estimate of the potential fundamental valuation of a specific security, and is not expressed as, or implied as, assessments of the quality of a security, a summary of past performance, or an actionable investment strategy for an investor. This is not an offer to sell or a solicitation of an offer to buy any security, nor shall any security be offered or sold to any person, in any jurisdiction in which such offer would be unlawful under the securities laws of such jurisdiction. I am not a registered investment advisor, broker/dealer, or licensed accountant.

# Glossary of financial terms

- **Equity:** A stock, ownership of a public or private entity
- **Publishing analyst:** Wall Street equity analyst who publishes investment research, typically by sector
- **Enterprise value** (“**EV**”): The value of a company’s operations; the value of its tangible and intangible assets
- **Market capitalization** (“**MC**”): The market value of the total equity of a company; differs from EV depending on outstanding debt of the company (equity = assets – liabilities)
- **Daily volume:** Average daily share volume traded for the company’s shares
- **MRQ assets:** “Most-recent-quarter” assets; the value of assets reported on the company’s balance sheet for the most recent quarter
- **LTM revenue:** “Latest twelve months” of revenue
- **Gross margin**: Profit after the cost of goods sold and other cost of sales
- **Operating margin**: Profit after cost of goods sold/other cost of sales and fixed costs (e.g., research and development)
- **EBITDA margin:** “Earnings before interest, taxes, depreciation, and amortization”
- **ROA**: Return on assets, a metric of profitability
- **ROE**: Return on equity, a metric of profitability

# Research sample

## COWEN

### EQUITY RESEARCH

November 18, 2018

- Semiconductors
- Semiconductor Capital Equipment
- Technology Hardware: Storage & Peripherals

**Karl Ackerman, CFA**  
415 646 7219  
karl.ackerman@cowen.com

**Krish Sankar**  
415 646 7372  
krish.sankar@cowen.com

**Steven Chin**  
415 646 7374  
steven.chin@cowen.com

**Samuel Reiff**  
415 646 7223  
samuel.reiff@cowen.com

**Robert Mertens, CFA**  
646 562 1338  
robert.mertens@cowen.com

### INDUSTRY UPDATE

## ILLUMINATING IPHONE'S IMPACT ON MEMORY DEMAND

### THE COWEN INSIGHT

iPhone production cuts are well known across the radio frequency and 3D image sensor landscapes, but we delineate the lesser-known impact on memory-exposed semis. While negative for near-term supply/demand and ASPs, we expect only \$1.2B of total industry DRAM and NAND revenue at risk across DecQ/MarQ. Expect modest ~\$0.10 impact to MU and \$0.15 impact to WDC spread across CQ4/CQ1.

### Lower iPhone Build Shipments Don't Help Memory S/D Imbalance

Over the last 8 years, Apple has grown to become one of the largest procurers of memory as a result of the proliferation of iPhone units and increased memory content per device. Indeed, our analysis herein indicates the Apple iPhone represented ~10% of all NAND demand and ~5% of all DRAM demand over the last 12 trailing months ending CQ3:18. Over the last week, our own field work across the smartphone supply chain – underscored by outlooks from SWKS, QRVO, AMS and LITE – indicates a 20MM iPhone unit cut across CQ4:18 and CQ1:19. We estimate that every 20MM unit change in current iPhone production represents a 6% reduction to mobile NAND demand and 3% reduction to mobile DRAM demand. Ergo, the recent production cut from AAPL should, unfortunately, compound an already weak supply/demand backdrop for memory producers in CQ4:18 and CQ1:19 and should be deflationary to memory ASPs. Moreover, a 20MM unit reduction in iPhone production should translate into 65k fewer NAND wafers, or 2% of total capacity in CQ3, and 70k fewer DRAM wafers, or 2% of total capacity in CQ3.

### iPhone Impact On Memory Semis

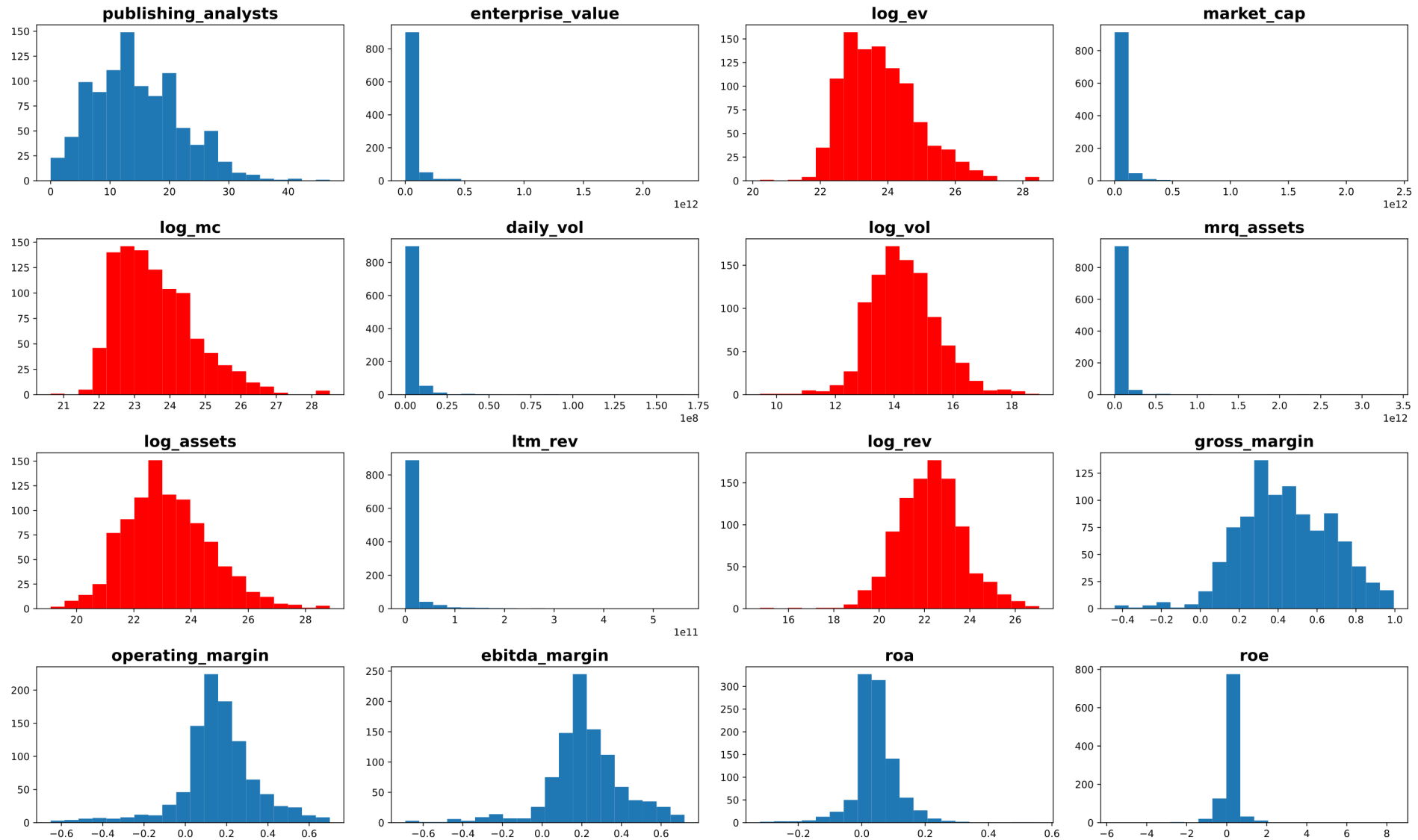
Based on our supply chain checks in Asia late last week, we now expect a 20MM reduction in iPhone builds across CQ4:18 and CQ1:19. Our analysis herein indicates these very recent supply chain cuts are more meaningful to bits than revenue for memory suppliers, and we expect a relatively modest ~\$0.10 headwind to MU and \$0.15 headwind to WDC across DecQ and MarQ from lower iPhone builds.

### iPhone Impact on Semicap and Memory WFE

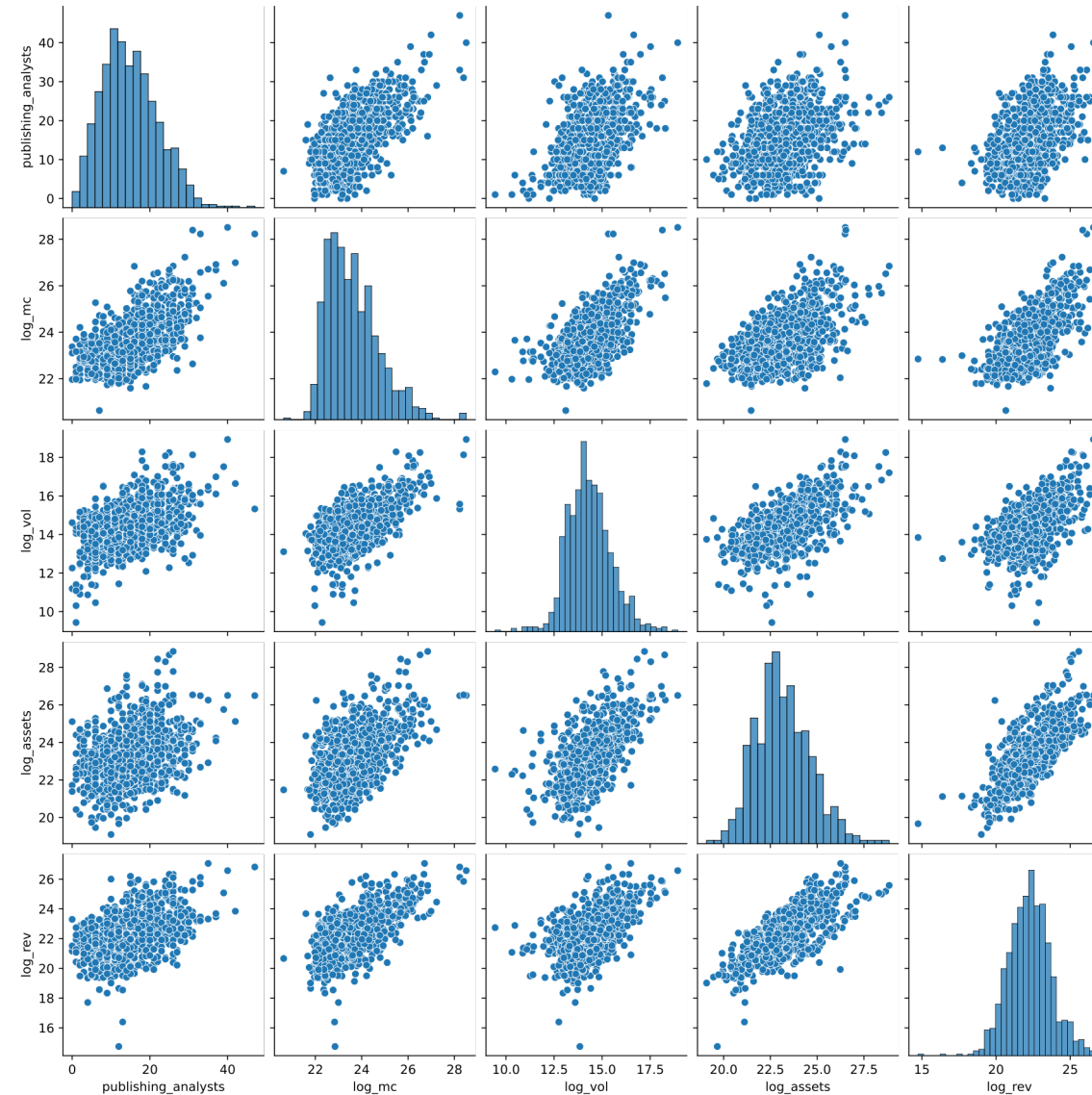
From a semicap SFE perspective, we estimate a 20M unit reduction in iPhone units could

This report is intended for samuel.reiff@cowen.com. Unauthorized redistribution of this report is prohibited.

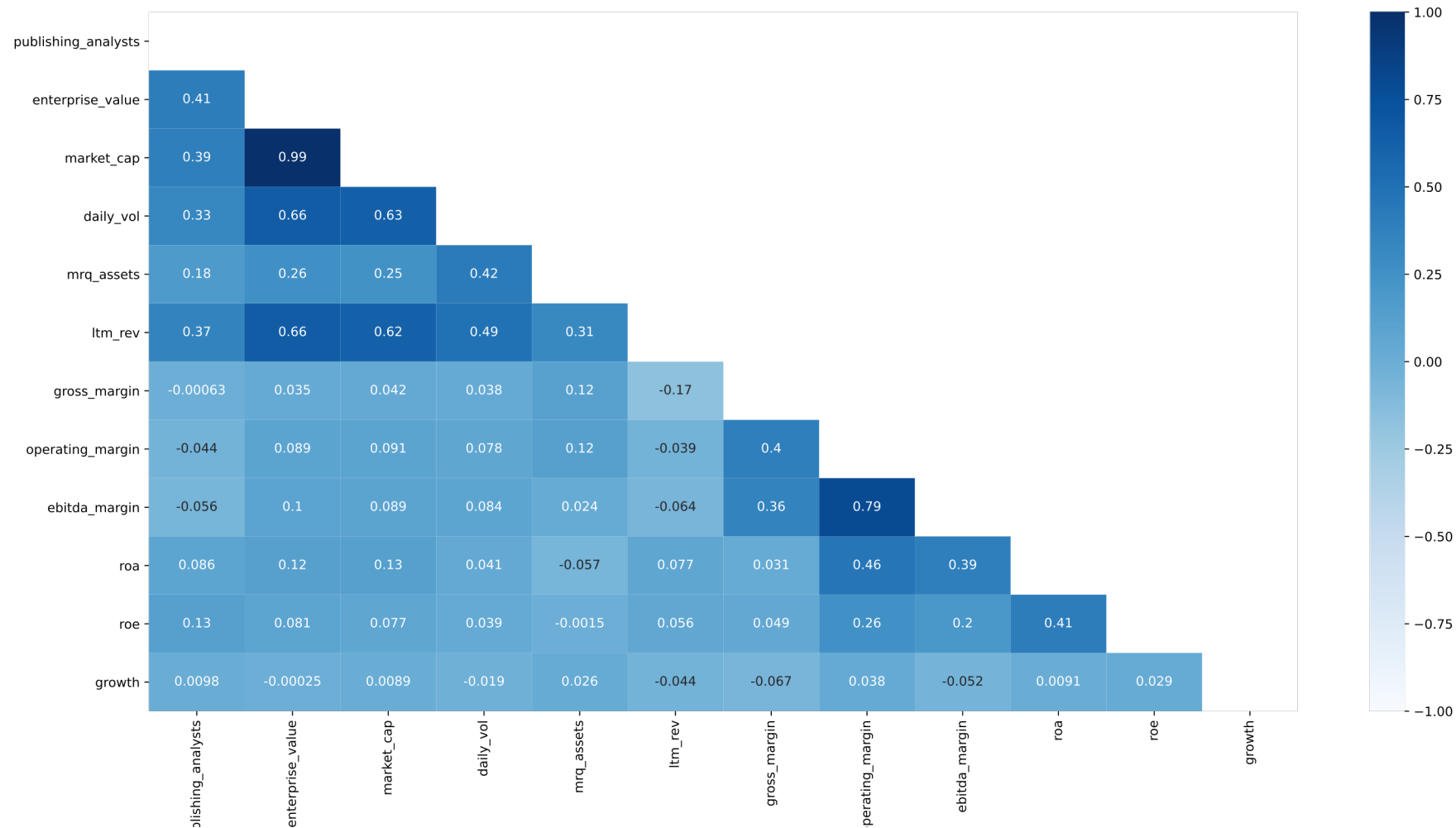
# Quantitative features considered



# Pairplot with major log features

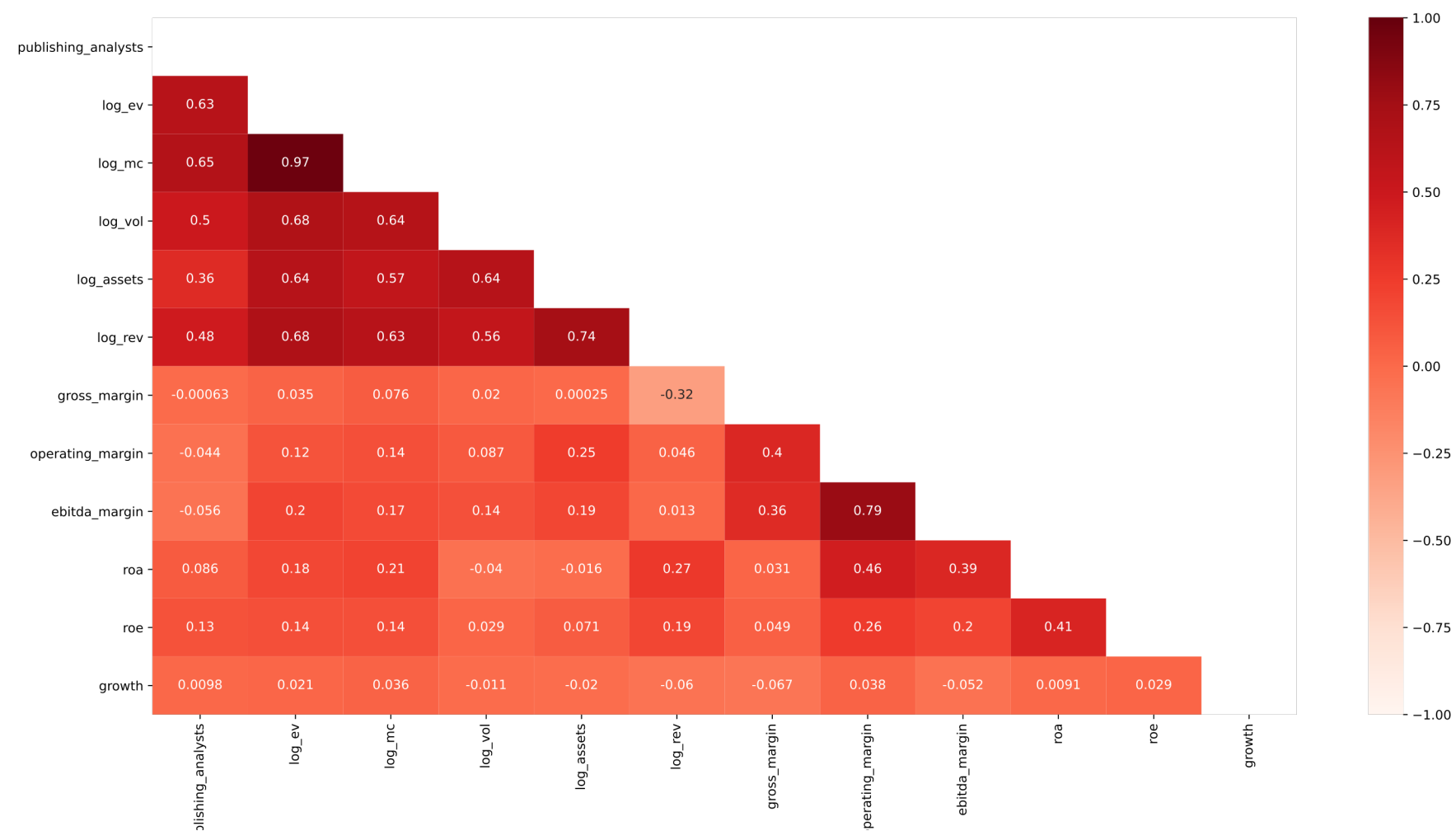


# Nominal feature correlation heatmap

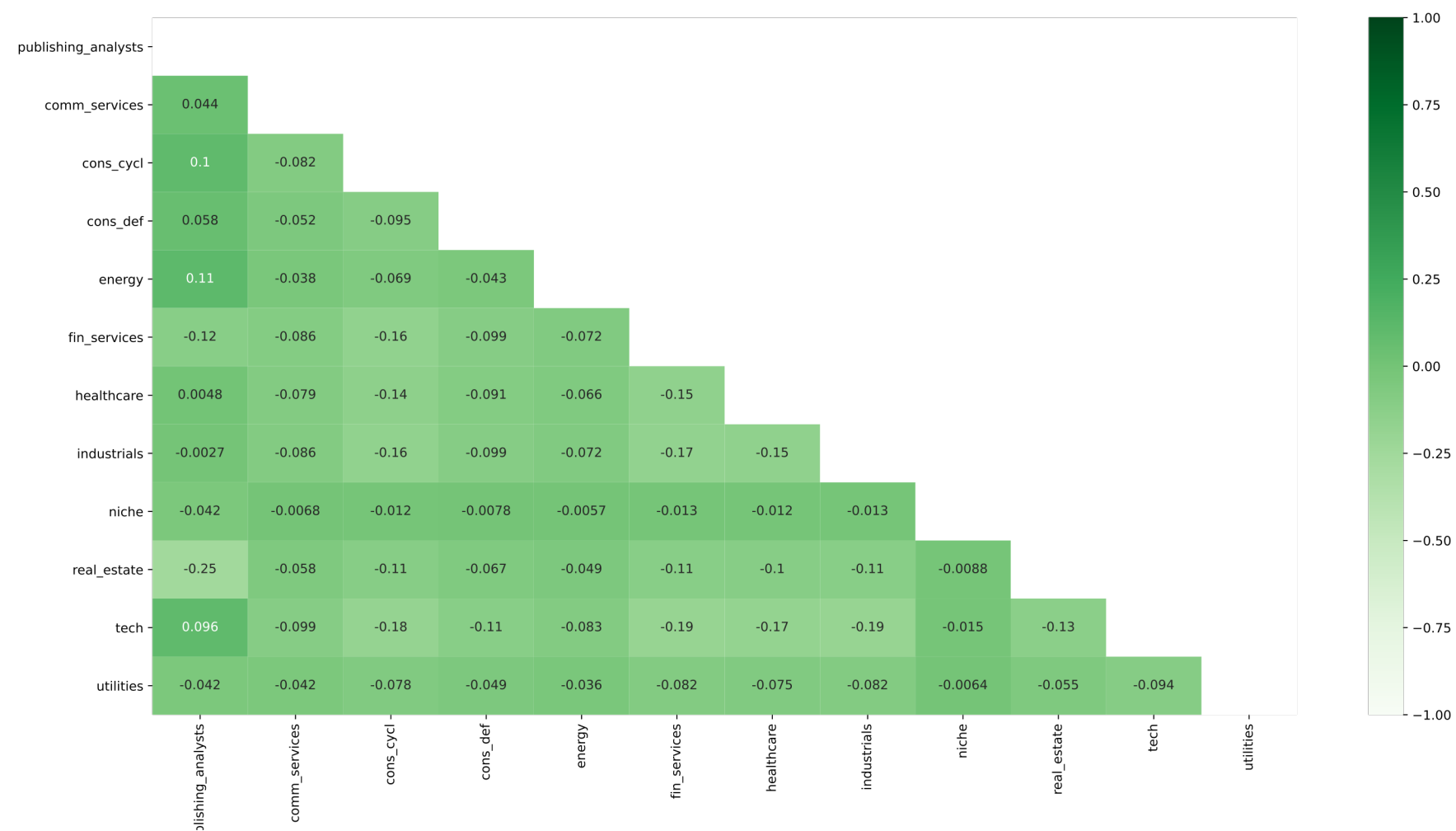




# Log feature correlation heatmap



# Sector feature correlation heatmap

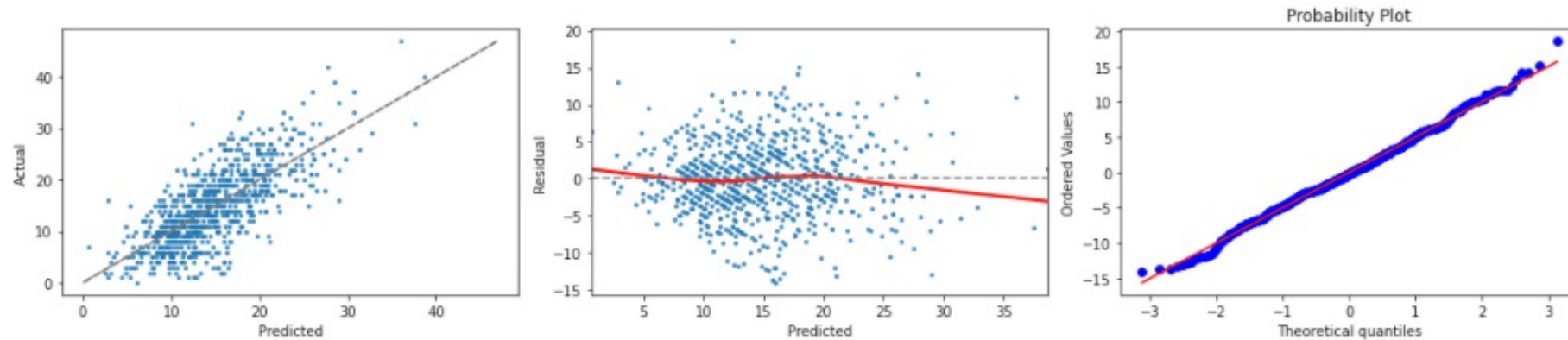


# Vanilla LR results on training data

## Vanilla linear model:

```
-----  
Vanilla LR cross-validation score: 0.514  
Vanilla R^2 on training data of: 0.5425  
Vanilla LR MAE on training data of: 3.9218  
-----
```

## Diagnostics -- vanilla LR (training data)

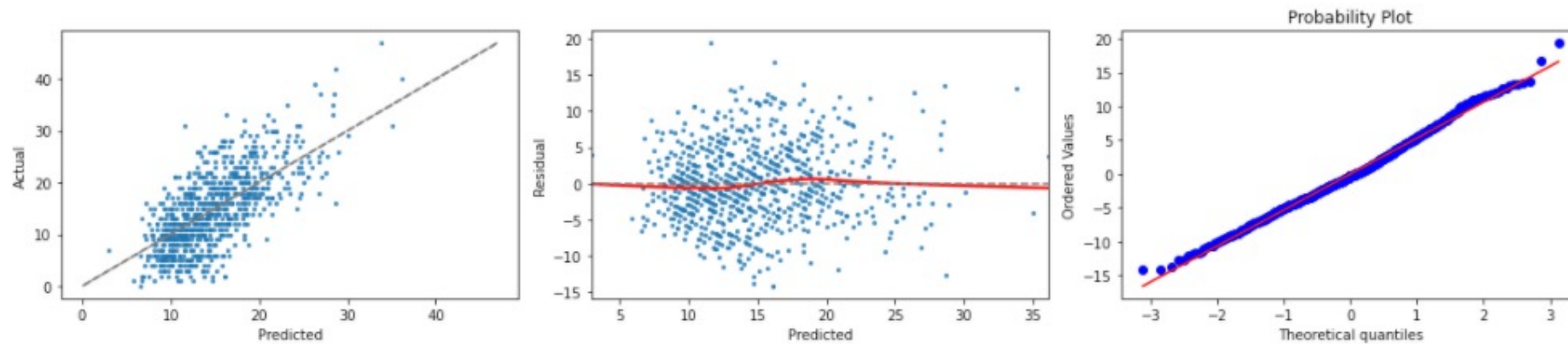


# LASSO results on training data

## Lasso model:

```
-----  
Optimal alpha: Lasso(alpha=0.250626566416031)  
Lasso cross-validation score: 0.465  
Lasso R^2 on training data of: 0.4848  
Lasso MAE on training data of: 4.1964  
-----
```

## Diagnostics -- Lasso (training data)



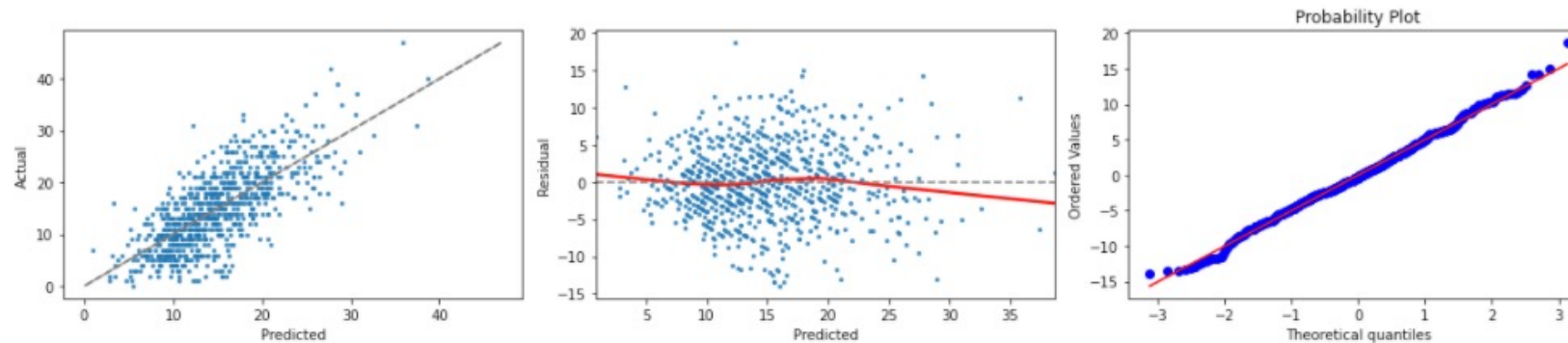
# Ridge results

## Ridge model:

```
-----  
Optimal alpha: Ridge(alpha=3.002001334222815)  
Ridge cross-validation score: 0.5148  
Ridge R^2 on training data of: 0.5417  
Ridge MAE on training data of: 3.9361
```

```
Ridge R^2 on test data of: 0.4598  
Ridge MAE on test data of: 4.163  
-----
```

Diagnostics -- Ridge (training data)



# Coefficients for LASSO and Ridge on train

Vanilla LR model coefficient assignment by feature:

```
Out[30]: [('log_mc', 4.035320181170686),
('log_vol', 1.1849758300298086),
('log_assets', 0.25304860053911193),
('log_rev', -0.10955040615800127),
('comm_services', -2.347199878348221),
('cons_cycl', 2.4644331094124827),
('cons_def', -0.2154348018241661),
('energy', 2.7068679993013487),
('fin_services', -2.579272229452887),
('healthcare', -1.3146057640936581),
('industrials', -0.194077073175095),
('niche', -3.977322048576674),
('real_estate', -6.9912878861159955),
('tech', 0.45117394248471565),
('utilities', -3.7815499302896964)]
```

Lasso model coefficient assignment by feature:

```
Out[57]: [('log_mc', 3.8743903984552825),
('log_vol', 0.9957844373421654),
('log_assets', 0.0),
('log_rev', 0.0998782070932138),
('comm_services', -0.02556271481189812),
('cons_cycl', 0.785469724456464),
('cons_def', 0.0),
('energy', 0.3820618175848931),
('fin_services', -0.3917166976973273),
('healthcare', -0.0),
('industrials', 0.0),
('niche', -0.0),
('real_estate', -1.296532756417945),
('tech', 0.20995032551268145),
('utilities', -0.2943377109710476)]
```

Ridge model coefficient assignment by feature:

```
Out[25]: [('log_mc', 3.7415949497316903),
('log_vol', 1.2275877904258856),
('log_assets', -0.2182229699540663),
('log_rev', 0.5577523740496148),
('comm_services', -0.6328251075179409),
('cons_cycl', 2.5084645488919857),
('cons_def', 0.2793728294105315),
('energy', 1.6855664542218896),
('fin_services', -1.0325872062268353),
('healthcare', -0.22296758758375318),
('industrials', 0.4822496955366807),
('niche', -0.10745274763003876),
('real_estate', -3.7112313575337454),
('tech', 1.2480083003480595),
('utilities', -1.3732253965319212)]
```

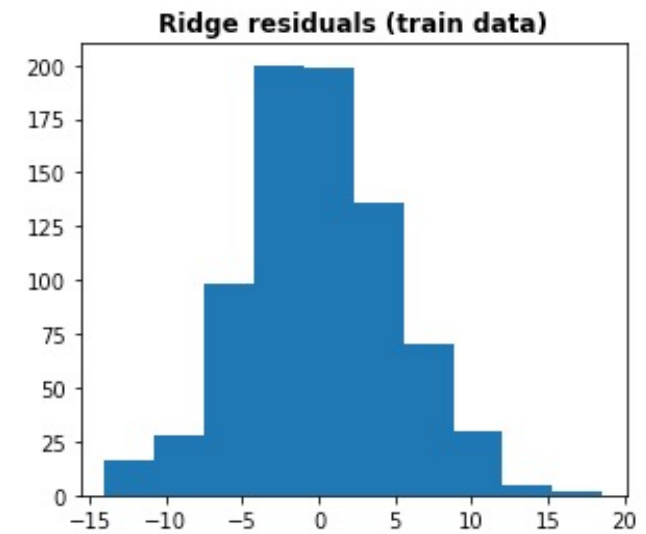
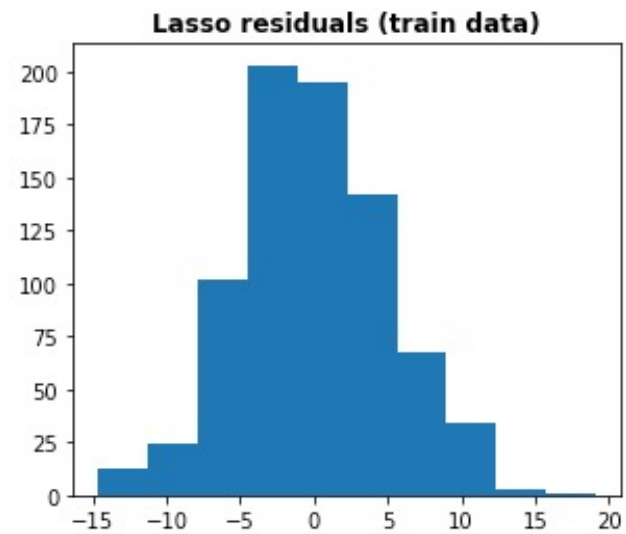
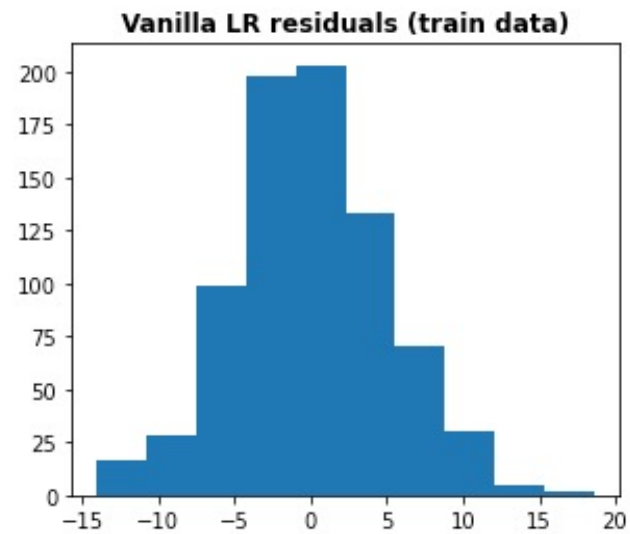
**Both regularized models de-emphasized assets and select sectors, and emphasized MC and volume; the standard linear model's treatment of revenue coefficient was negative**

# Model residual distributions

Vanilla LR resid. mean of:  $3.3805158219198644e-14$

Lasso resid. mean of:  $1.812609019796174e-16$

Ridge resid. mean of:  $5.890979314337565e-16$



# Regression data set sample

Out[8]:

	equity	publishing_analysts	industry	log_ev	log_mc	log_vol	log_assets	log_rev	comm_services	cons_cycl	cons_def	energy	fin_servic
0	A	17	Diagnostics & Research	24.595924	24.559056	14.922356	22.987837	22.484566	0	0	0	0	
1	AA	10	Aluminum	22.981899	22.719326	14.436839	23.421939	23.070824	0	0	0	0	
2	AAL	21	Airlines	24.508591	23.341391	15.675505	24.850529	23.651254	0	0	0	0	
3	AAP	26	Specialty Retail	23.524199	23.351551	13.379328	23.194719	23.097241	0	1	0	0	
4	AAPL	40	Consumer Electronics	28.485436	28.510648	18.931706	26.503664	26.573052	0	0	0	0	
5	ABBV	21	Drug Manufacturers—General	26.378348	26.070135	16.671981	25.737661	24.707237	0	0	0	0	
6	ABC	10	Medical Distribution	23.984051	23.959982	14.198253	24.513682	26.002808	0	0	0	0	
7	ABT	21	Medical Devices	26.141364	26.094648	16.683409	25.007514	24.417879	0	0	0	0	
8	ACC	3	REIT—Residential	23.135602	22.656235	14.131634	22.742315	20.591881	0	0	0	0	
9	ACGL	11	Insurance—Diversified	23.703377	23.462815	15.086170	24.491010	22.978759	0	0	0	0	
10	ACHC	10	Medical Care Facilities	22.679126	22.433454	13.691647	22.594970	21.479388	0	0	0	0	
11	ACI	19	Grocery Stores	23.898817	23.033819	13.604049	24.004102	24.945857	0	0	1	0	
12	ACM	11	Engineering & Construction	23.163001	22.953280	14.129445	23.288134	23.314033	0	0	0	0	
13	ACN	24	Information Technology Services	26.000666	26.030939	15.661158	24.336306	24.593425	0	0	0	0	
14	ADBE	25	Software—Infrastructure	26.410580	26.414233	15.373529	23.913084	23.389799	0	0	0	0	
15	ADI	24	Semiconductors	24.903788	24.841001	15.117064	23.789857	22.547815	0	0	0	0	
16	ADM	15	Farm Products	24.490494	24.230024	15.531540	24.629653	25.039687	0	0	1	0	
17	ADP	21	Staffing & Employment Services	25.214819	25.206042	15.261865	24.391062	23.431982	0	0	0	0	
18	ADPT	7	Biotechnology	22.225119	22.360319	13.726023	20.833388	18.568325	0	0	0	0	
19	ADS	12	Credit Services	23.672954	22.300181	13.086529	23.838872	22.172535	0	0	0	0	