

Finding S&P 500 Predictors using Lasso Feature Selection

Nicholas Trieu

Math189r Fall 2016, Harvey Mudd College

The Churning Problem

- Diversifying assets: less risk, same expected return
- Mutual funds: pool money to invest diversely
- Hidden transaction fees: rebalancing, “churning”

Solutions?

- Don't rebalance?
 - Bad idea
- Instead, maximize variance by selecting subset of stocks

Dow Theory

- Transportation sector moves with the general market
- Efficient market hypothesis: market updates with full information
 - Plausible for visible indexes traded with high volume
- Question of interest:
 - What stocks in the Dow Jones Transportation sector matter most?
(Delivery services > ground transport > air transport)
 - Do any individual components perform better than the DJ Transport Avg Index?
(Only UPS)

Table 1: Dow Jones Transports Components (Dec 5, 2016). Excluded: MATX (too few data points).

Corporation	Ticker	Industry
Alaska Air Group, Inc.	ALK	Airlines
American Airlines Group Inc.	AAL	Airlines
Avis Budget Group, Inc.	CAR	Rental and Leasing Services
C.H. Robinson Worldwide, Inc.	CHRW	Trucking
CSX Corp.	CSX	Railroads
Delta Air Lines	DAL	Airlines
Expeditors International	EXPD	Delivery Services
FedEx Corporation	FDX	Delivery Services
JB Hunt Transport Services, Inc.	JBHT	Trucking
JetBlue Airways Corp.	JBLU	Airlines
Kansas City	KSU	Railroads
Kirby Corp.	KEX	Marine Transportation
Landstar System, Inc.	LSTR	Trucking
Norfolk Southern Corp.	NSC	Railroads
Ryder System, Inc.	R	Transportation Services
Southwest Airlines, Inc.	LUV	Airlines
Union Pacific Corp.	UNP	Railroads
United Continental Holdings	UAL	Airlines
United Parcel Service, Inc.	UPS	Delivery Services

Data Processing

Given data has the following features:

- *Date*: The year, month, and day, given as Month=Day=Y ear.
- *Time*: The time of day, given as HourMinutes concatenated in a single string. Trading hours are from 9:30 AM to 3:59 PM.
- *Open*: The opening price at the start of the minute.
- *High*: The highest transaction price during the minute.
- *Low*: The lowest transaction price during the minute.
- *Close*: The closing price at the end of the minute.
- *Volume*: The trading volume, i.e. the number of shares transacted during that minute.
- Note: No information about best bid or best ask.

Data Processing

- Steps:
 - Conform data to common time range
 - Fill in time gaps (reuse old values)
 - Account for stock (reverse) splits
 - Derive percentage change: $(Close - PreviousClose) / Open * 100$
 - Some data points had the same *Open*, *Close*, *High*, and *Low* while the prices across minutes were changing...

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Date	Time	Open	High	Low	Close	Volume	OC_Mag	HL_Mag	OC_Dir	OC_Chain	Per_Change	Per_Chain
2	8/19/2004	1155	50.06	50.31	49	49.05	2563134	1.01	1.31	-1	-1	2.017579	-2.01758
3	8/19/2004	1156	49.05	49.42	48.4	48.43	1819175	0.62	1.02	-1	-2	1.264016	-3.2816
4	8/19/2004	1157	48.43	48.48	48.03	48.46	831342	0.03	0.45	1	1	0.061945	0.061945
5	8/19/2004	1158	48.4	49.18	48.17	49.16	764878	0.76	1.01	1	2	1.570248	1.632193

Lasso Feature Selection

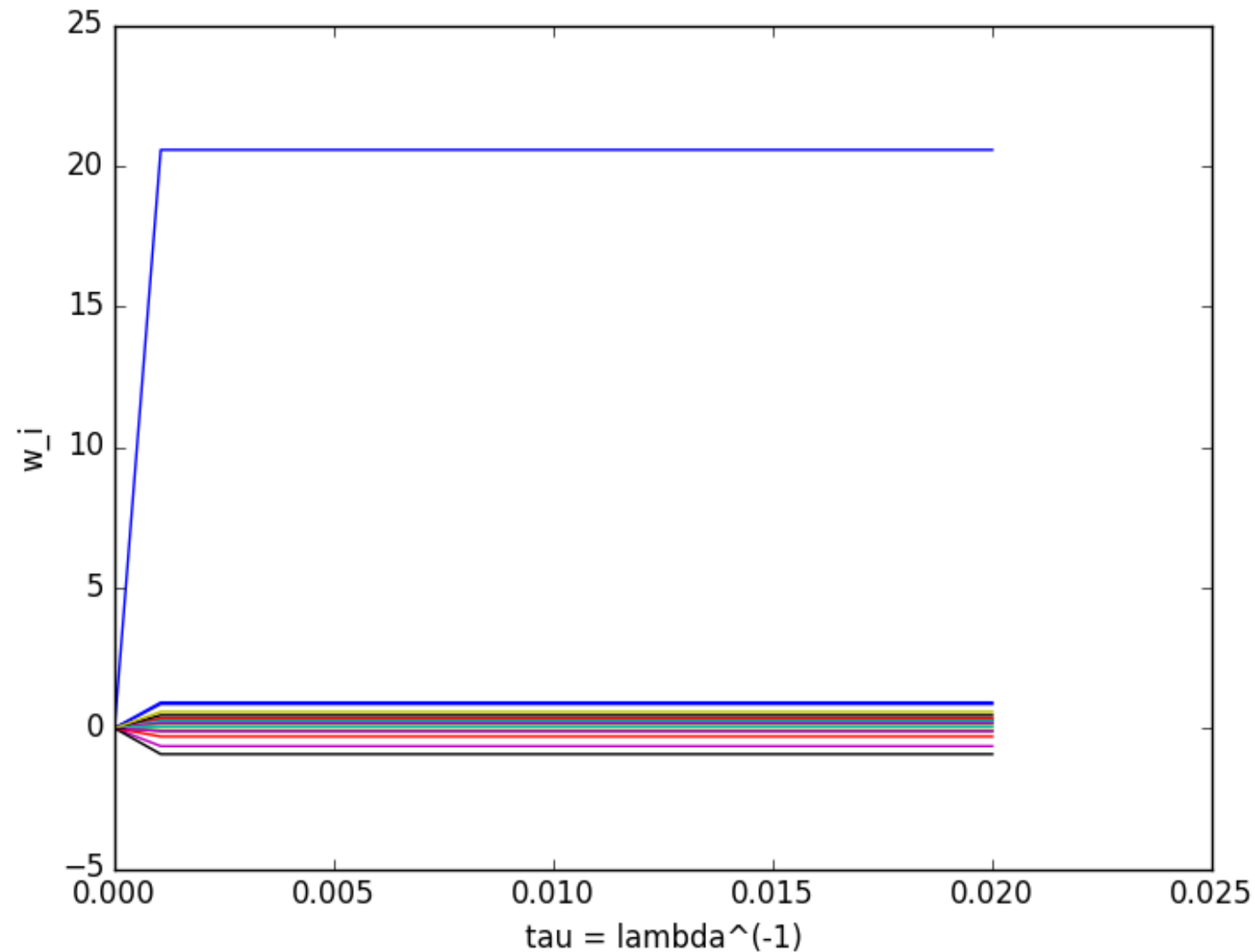
- Rationale: L1 Regularization to encourage sparse solutions
- Optimization problem:

$$\min_w \|Xw - y\|_2^2 + \lambda \|w\|_1$$

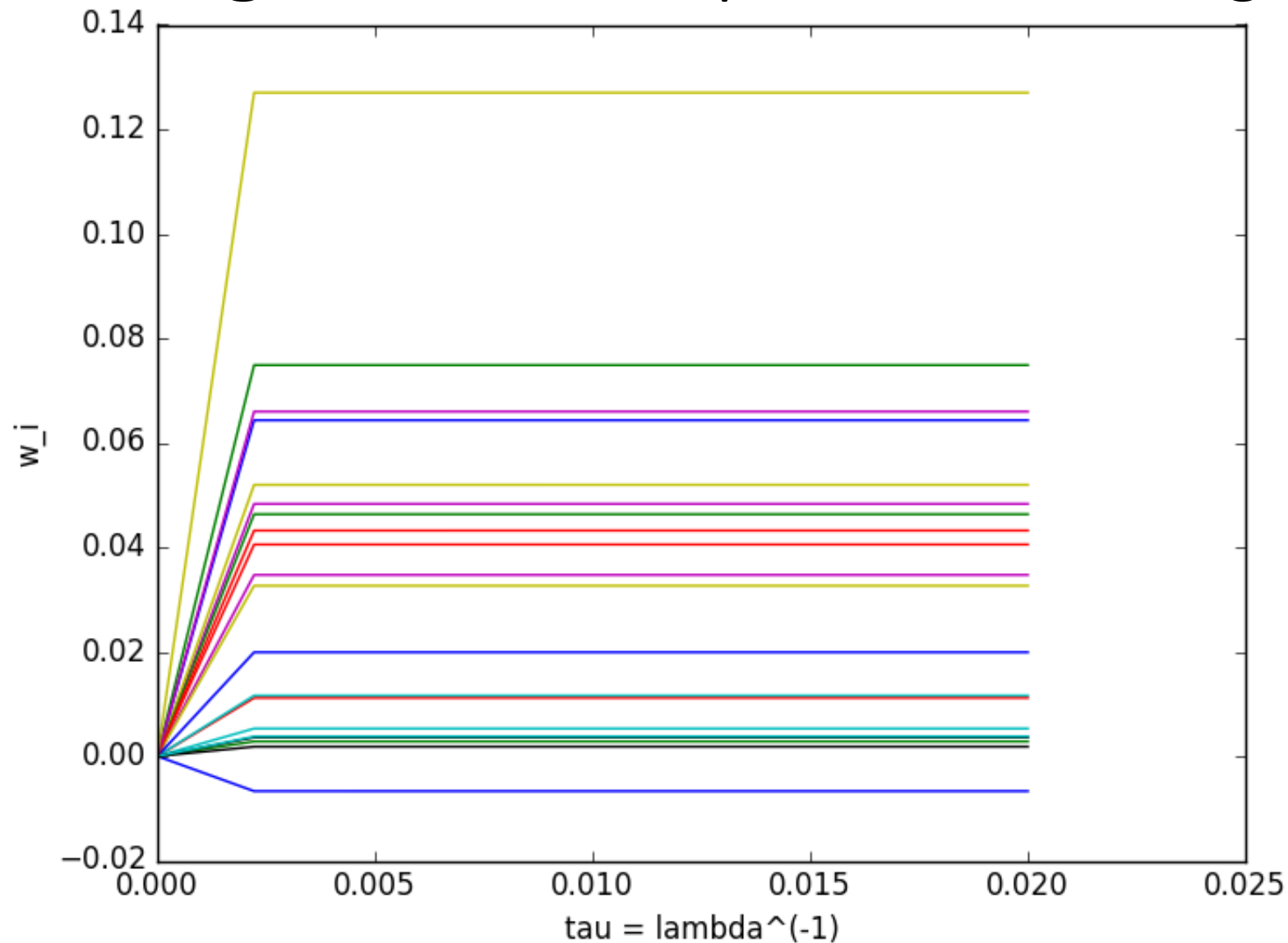
- Solve using proximal gradient descent

$$\text{prox}_r(x_i) = \begin{cases} x_i - r, & x_i > r \\ 0, & |x_i| \leq r \\ x_i + r, & x_i < -r \end{cases}$$

Lasso Reg Path: Transports *Open* predicting SPY



Lasso Reg Path: Transports % *Change* predicting SPY



Stocks & IYT % *Change* predicting SPY

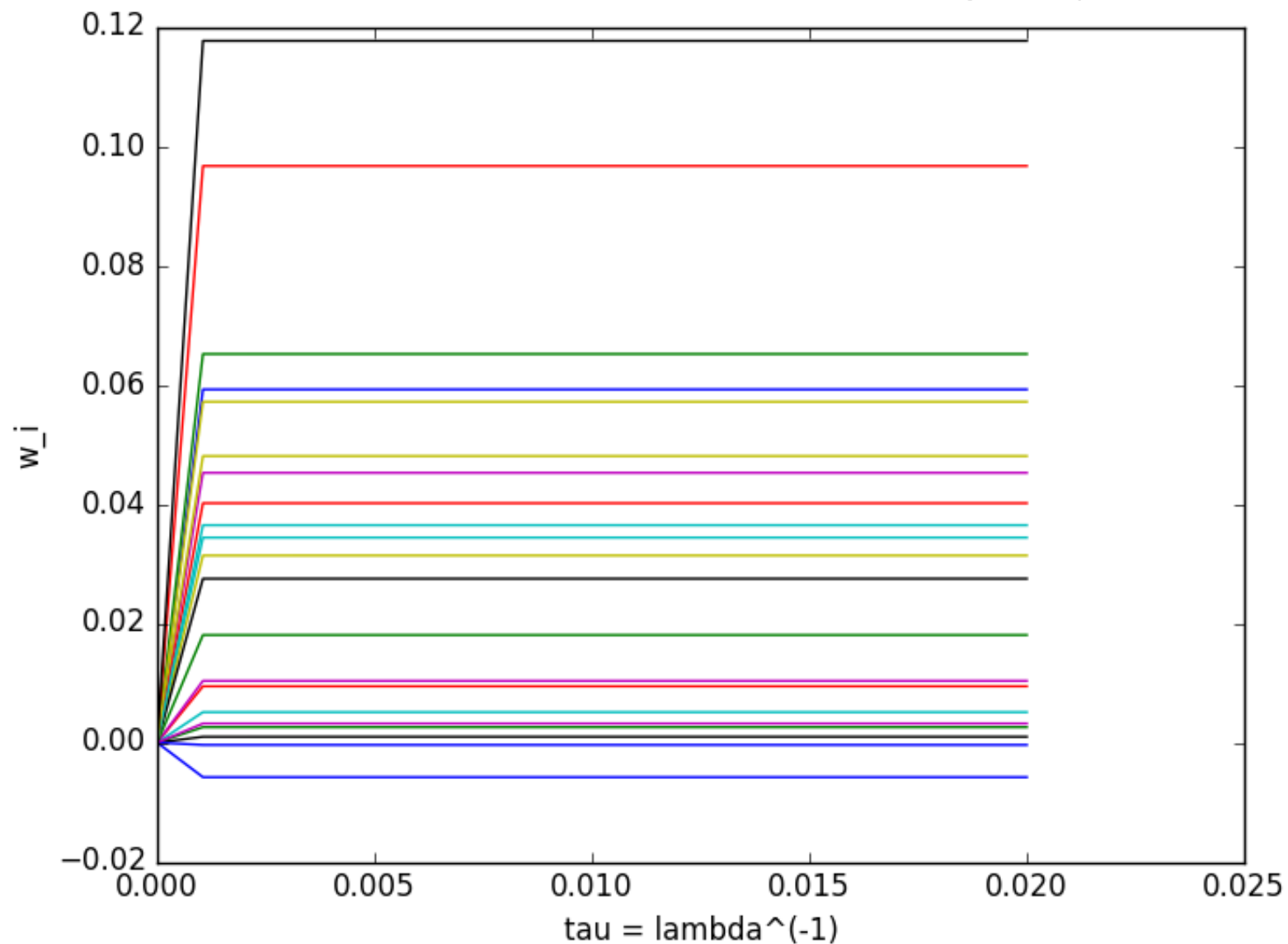


Table 2: Dow Jones Transports Components *Per_Change* Weights.

Corporation	Ticker	Industry	Weight (w_i)
United Parcel Service, Inc.	UPS	Delivery Services	0.126998024737
FedEx Corporation	FDX	Delivery Services	0.0748592406015
Union Pacific Corp.	UNP	Railroads	0.0659628826315
Expeditors International	EXPD	Delivery Services	0.06430810349
CSX Corp.	CSX	Railroads	0.0519410568611
C.H. Robinson Worldwide, Inc.	CHRW	Trucking	0.0482897806767
Norfolk Southern Corp.	NSC	Railroads	0.0463079329717
Ryder System, Inc.	R	Transportation Services	0.0431943007092
JB Hunt Transport Services, Inc.	JBHT	Trucking	0.0405205437987
Kirby Corp.	KEX	Marine Transportation	0.0346942025747
Kansas City	KSU	Railroads	0.0326286114762
Southwest Airlines, Inc.	LUV	Airlines	0.0198953182821
JetBlue Airways Corp.	JBLU	Airlines	0.0116079269819
Alaska Air Group, Inc.	ALK	Airlines	0.0111592409286
Avis Budget Group, Inc.	CAR	Rental and Leasing Services	0.0052952656847
United Continental Holdings	UAL	Airlines	0.00377363673377
Landstar System, Inc.	LSTR	Trucking	0.00359611247851
American Airlines Group Inc.	AAL	Airlines	0.0028201725868
Delta Air Lines	DAL	Airlines	0.00182567356804

Table 3: Dow Jones Transports Components and IYT *Per_Change* Weights.

Corporation	Ticker	Industry	Weight (w_i)
United Parcel Service, Inc.	UPS	Delivery Services	0.117660544381
iShares Dow Jones Transport. Avg. (ETF)	IYT	Exchange Traded Product	0.0966779643449
FedEx Corporation	FDX	Delivery Services	0.0651563154648
Expeditors International	EXPD	Delivery Services	0.0592153834167
Union Pacific Corp.	UNP	Railroads	0.0571421595973
CSX Corp.	CSX	Railroads	0.0480401072606
C.H. Robinson Worldwide, Inc.	CHRW	Trucking	0.0452393253124
Norfolk Southern Corp.	NSC	Railroads	0.040143297956
Ryder System, Inc.	R	Transportation Services	0.036453092742
JB Hunt Transport Services, Inc.	JBHT	Trucking	0.0343717495153
Kirby Corp.	KEX	Marine Transportation	0.0313717091875
Kansas City	KSU	Railroads	0.0274816703841
Southwest Airlines, Inc.	LUV	Airlines	0.0180378182726
JetBlue Airways Corp.	JBLU	Airlines	0.0103502320302
Alaska Air Group, Inc.	ALK	Airlines	0.00943822842662
Avis Budget Group, Inc.	CAR	Rental and Leasing Services	0.00510799412822
United Continental Holdings	UAL	Airlines	0.00319945764963
American Airlines Group Inc.	AAL	Airlines	0.0025910135552
Delta Air Lines	DAL	Airlines	0.000977802693483
Landstar System, Inc.	LSTR	Trucking	-0.000391734850494

Conclusions

- Delivery services > Ground transportation > Air transportation
- Index > most stocks (diversity rules)