Predicting Market Volatility from Federal Reserve Board Meeting Minutes

Reza Bosagh Zadeh and Andreas Zollmann

Lab Advisers: Noah Smith and Bryan Routledge

Goals

- Make Money!
  - Not really.

- Find interesting patterns in meeting minutes
  - Meetings happen roughly 10 times a year
  - Interest rate changes are decided, along with other qualitative assessments of US Economy
  - Minutes freely available on the web for meetings from 1967 to 2008

- "Idea": Use established tools from NLP and ML

- Efficient market hypothesis:
  - excess returns per unit risk cannot be consistently generated using public information
  - Stock prices react on news in split-seconds
  - Automated analysis can outperform humans because of processing speed
- Strongest form of EMH: all market correction due to insiders
  - Gidofalvi and Elkan (2003): News-based prediction model has highest predictive accuracy over the 20 minutes trading window ***before*** the publication time of the respective article

Previous Attempts

	Prototype 3.1.	Prototype 3.2.	Prototype 3.3.	Prototype 3.4.	Prototype 3.5.	Prototype 3.6.	Prototype 3.7.	Prototype 3.8.
Prototype idea								
Aims to forecast…	price trends	price trends	volatilities	price trends	price trends	price trends	volatilities	price trends
Underlying	equity index	single stock	single stock	single stock	exchange rate	single stock	single stock	single stock
Forecasting horizon	24 hours	1 hour	N/A	1 hour	3 hours	1 hour	N/A	15 minutes
Text mining parameter								
Feature definition	manually	automated	manually	automated	manually	automated	automated	semi-automated
Number of features	423	N/A	145	1000	400	N/A	200	85
Feature granularity	tuple (words)	terms	tuple (terms)	single words	tuple (words)	single words	single words	tuple (terms)
Primary classifier	Naïve Bayes	Naïve Bayes	decision rules	Naïve Bayes	decision rules	linear SVM	regression	polynomial SVM
Number of categories	3	5	39	3	3	5 (training: 3)	2	4 (training: 3)
Input data								
Information age	2 - 15 hours	0 hours	0 - 24 hours	0 hours	0 - 2 hours	0 hours	0 hours	0 hours
Text analyzed	headline, body	headline, body	headline	headline, body	headline	headline, body	headline, body	headline, body
Labeling	automated	automated	manually	automated	automated	automated	automated	automated
Price frequency	daily close	10 min.	daily close	10 min.	60 min.	intraday	daily close	15 sec.
Test								
Period investigated	1997 - 1998	1999 - 2000	2001 - 2002	2001 - 2002	1993	2002 - 2003	1999 - 2002	2002
Training/Test split	3 months rolling	3 / 1.5 months	8 / 5 months	5.5 / 2 months	1 month rolling	6 / 1 month(s)	cross validation (90% / 10%)	cross validation (90% / 10%)
Prototype vs. random	44% vs. 33%	N/A	N/A	40% vs. 33%	50% vs. 33%	N/A	61% vs. 50%	45% vs. 33%
Roundtrips per year	< 600	> 100'000	(200)	< 6000	N/A	N/A	N/A	< 500
Profit per roundtrip as reported	13 bps	23 bps	(first phase: 10 bps)	10 bps	N/A	N/A	N/A	29 bps
Market	DJIA, Nikkei, FTSE, HS, ST	127 stocks (USA)	constituents Russell 3000	constituents DJIA	USD/DEM and USD/JPY	614 stocks (Hong Kong)	constituents DAX100	constituents S&P500

Source: "Text Mining Systems for Market Response to News: A Survey", Marc-André Mittermayer, Gerhard F. Knolmayer 2006

PAST WORK: FEATURES USED IN TEXT-BASED PREDICTION

- BOW, BObigrams, NPs, NNPs, NEs (frequency, TF-IDF score, or information gain)
- Lerman et al. (COLING'08):
  - News-focus features:
    - change in occurrence frequency of a word in the current day's news coverage compared to the average news coverage of the past N days
  - Dependency features

Classification vs. Regression

- Most past work: predicts increase or decrease in prices/volatility
- Kogan et al. (NAACL'09): predict indicator (stock volatility) directly using Support Vector Regression

TAPPING THE FED

- Our aim: use FOMC meeting minutes to predict financial indicators
- No previous attempts to our knowledge
- Boukus & Rosenberg: market participants do extract complex signals from these minutes
  - found correlations of e.g. Treasury yields with specific themes of the meeting minutes using Latent Semantic Analysis

Our Attempt

- Predicting Prices is too hard. Focus on Volatility:

$$\text{vol} = 1/(n-1)\sum_{i=1}^{n}[\ln \text{return}(t+i) - 1/n \sum_{j=1}^{n} \ln \text{return}(t+j)]^2 ,$$

where t is the time of the meeting and

$$\text{return}(t) = \text{price}(t)/\,\text{price}(t-1) - 1 .$$

- Predict volatility of
  - S&P 500
  - 13-week Treasury Bills
  - 10-year Treasury Notes

Machine Learning Setting

- Take meeting minutes from minutes held on day t, and predict volatility n (look-ahead) days ahead.
- Not I.I.D. training data at all. But let's hide that under the carpet.
- Features:
  - Bag of Words: unigrams, bigrams
  - Dependency fragments
  - Volatility from n days ago

Dependency features

- (S
  (NP Business inventories)
  (VP climbed
  (ADVP significantly further)
  (PP in
  (NP September))) .)

```
                                  ┌────────────prep>────────────┐
      ┌──<nn──┐──<nsubj─┼──advmod>─┐──advmod>─┐           ├pobj>┐
Business inventories climbed significantly further in September .
```

- Using seedword ‘inventories’, extract fragments:

Inventories→climbed, business→inventories→climbed, inventories→climbed←further

Features: Bag of Words

- TF-IDF: $$(\text{tf-idf})_{\mathrm{i,j}} = \text{tf}_{\mathrm{i,j}} \times \text{idf}_{\mathrm{i}}$$

$$\text{tf}_{\mathrm{i,j}} = \frac{n_{i,j}}{\sum_k n_{k,j}} \qquad \text{idf}_{\mathrm{i}} = \log \frac{|D|}{|\{d : t_i \in d\}|}$$

  - IDF dampens effect of common words

- Log1P: $$\log \left|1 + \frac{n_{i,j}}{\sum_k n_{k,j}}\right|$$

  - Don't really need IDF since we already removed stop-words

Data Mining

- Obtained meeting text from PDF and HTML files available at http://www.federalreserve.gov/monetarypolicy/fomc.htm

- Our corpus available at http://rezab.ca/useful/fomc_minutes.html

- Stemmed using Porter 2 stemmer. Removed stop-words using available online list.

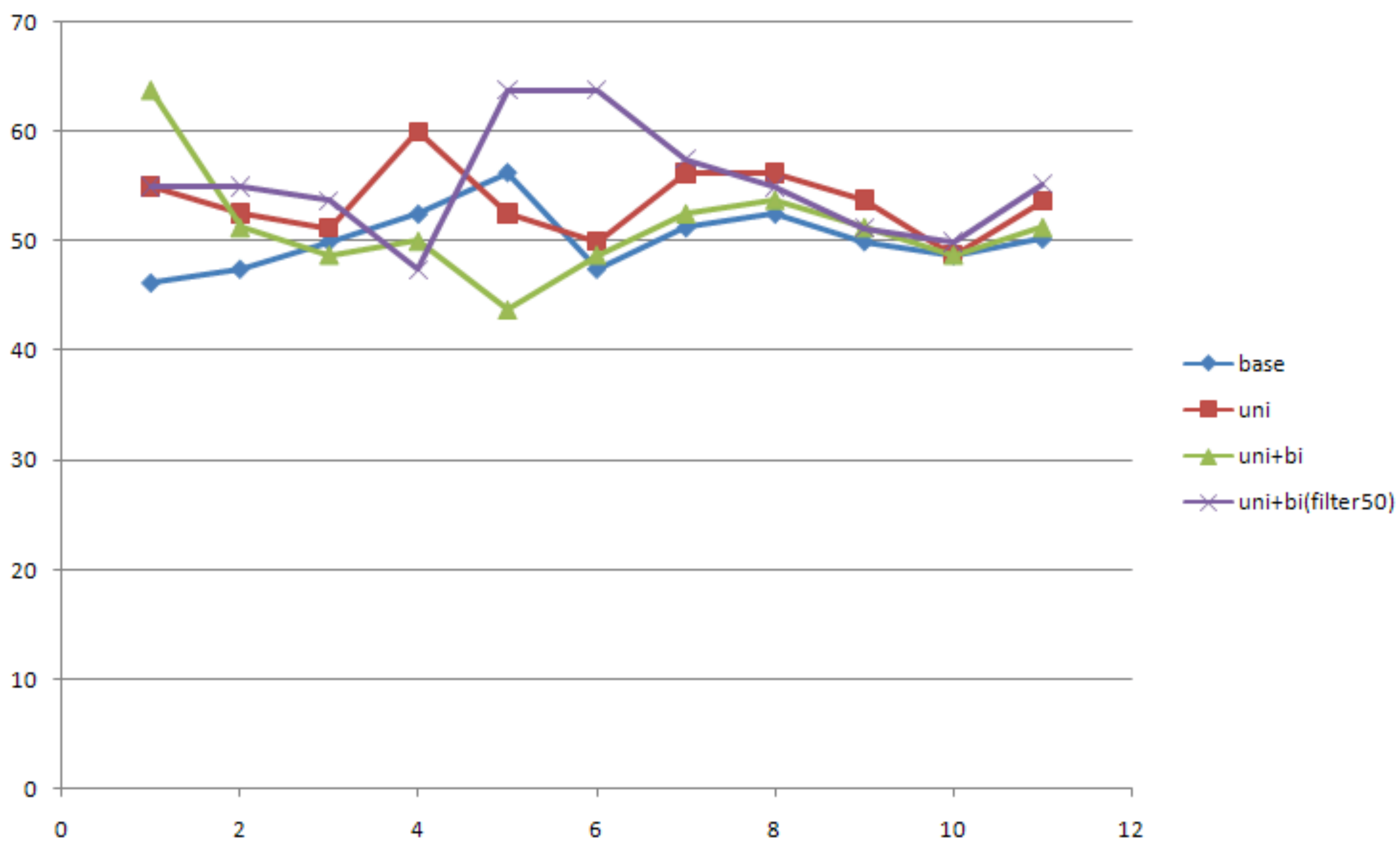
Predict what?

- Regression:
  - Actual value of the volatility

- Classification:
  - Two classes, volatility goes UP or DOWN

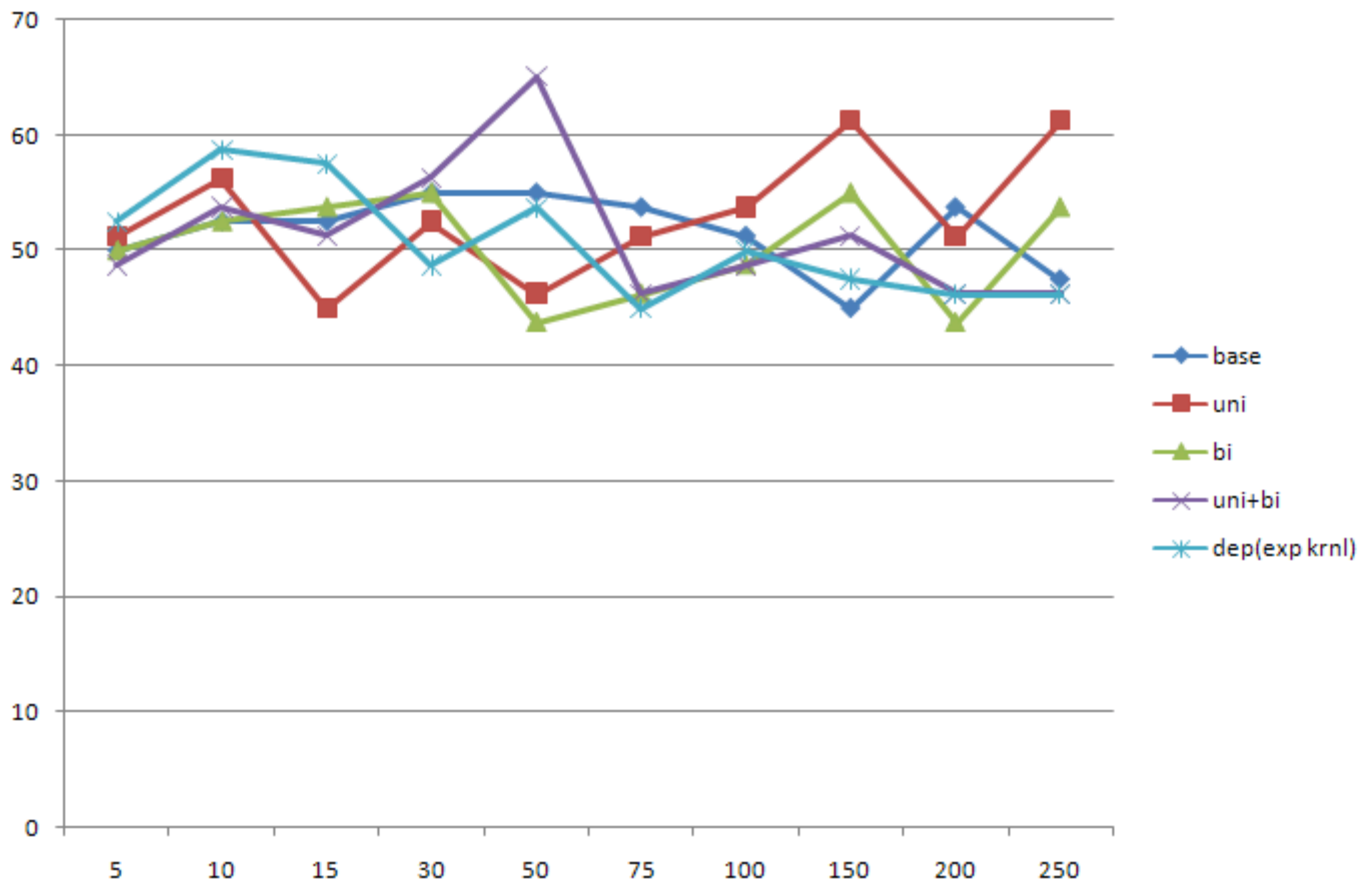
First set of experiments:
Classification for different indices.

S&P 500 – Classification – Short Periods

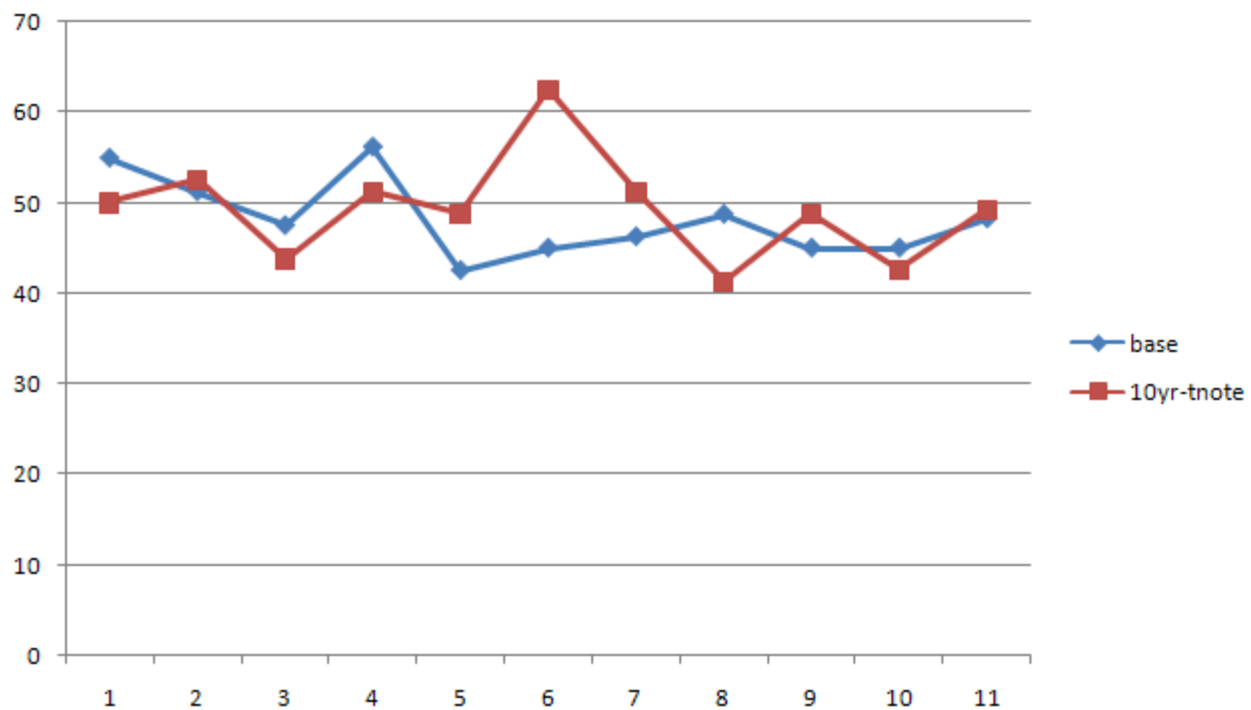


Higher is better.

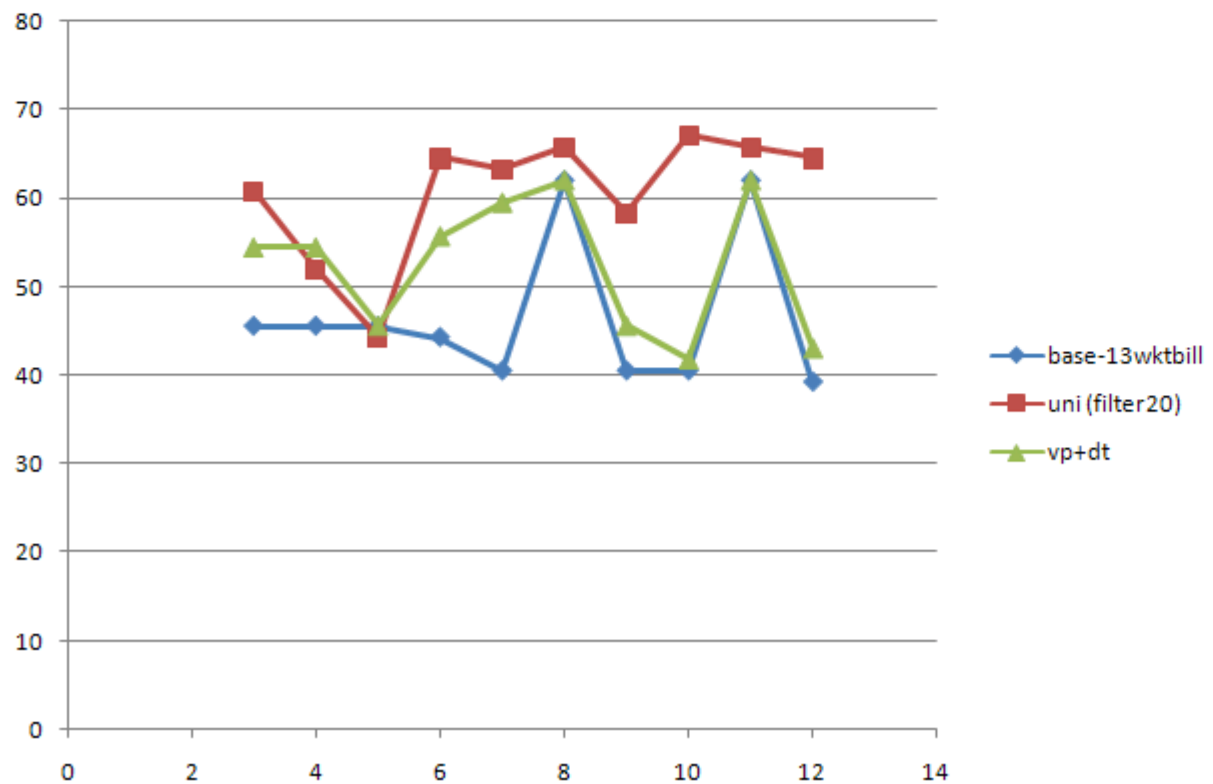
S&P 500 – Classification – Long Periods



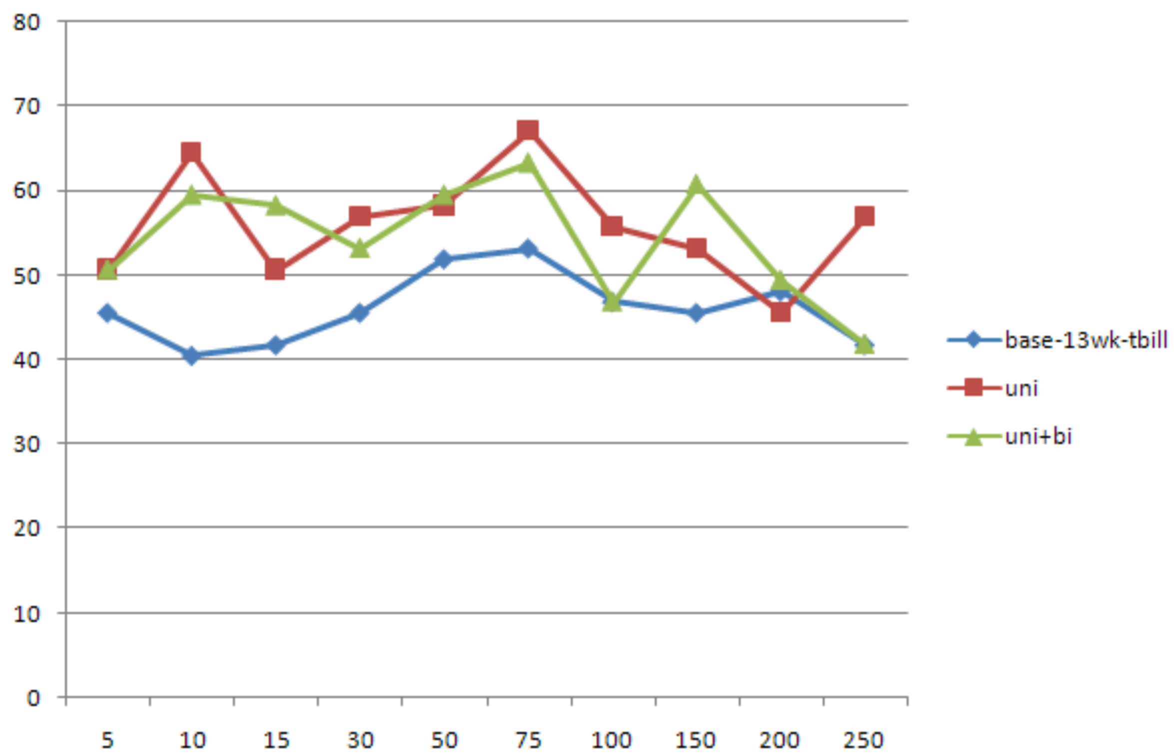
10 Year Treasury Note– Classification – Short Periods



13 Week Treasury Bills – Classification – Short Periods



13 Week Treasury Bills – Classification – Long Periods



Example decision Tree

```
w_stimulu < 6.2E-4
|   w_capit < 0.00253
|   |   w_statist < 0.00146: DOWN(80.0/45.0)
|   |   w_statist >= 0.00146
|   |   |   w_specifi < 0.00173: UP(120.0/79.0)
|   |   |   w_specifi >= 0.00173: DOWN(17.0/3.0)
|   w_capit >= 0.00253: UP(19.0/1.0)
w_stimulu >= 6.2E-4: DOWN(28.0/4.0)
```

This had 64% Accuracy

SVM Prominent Terms

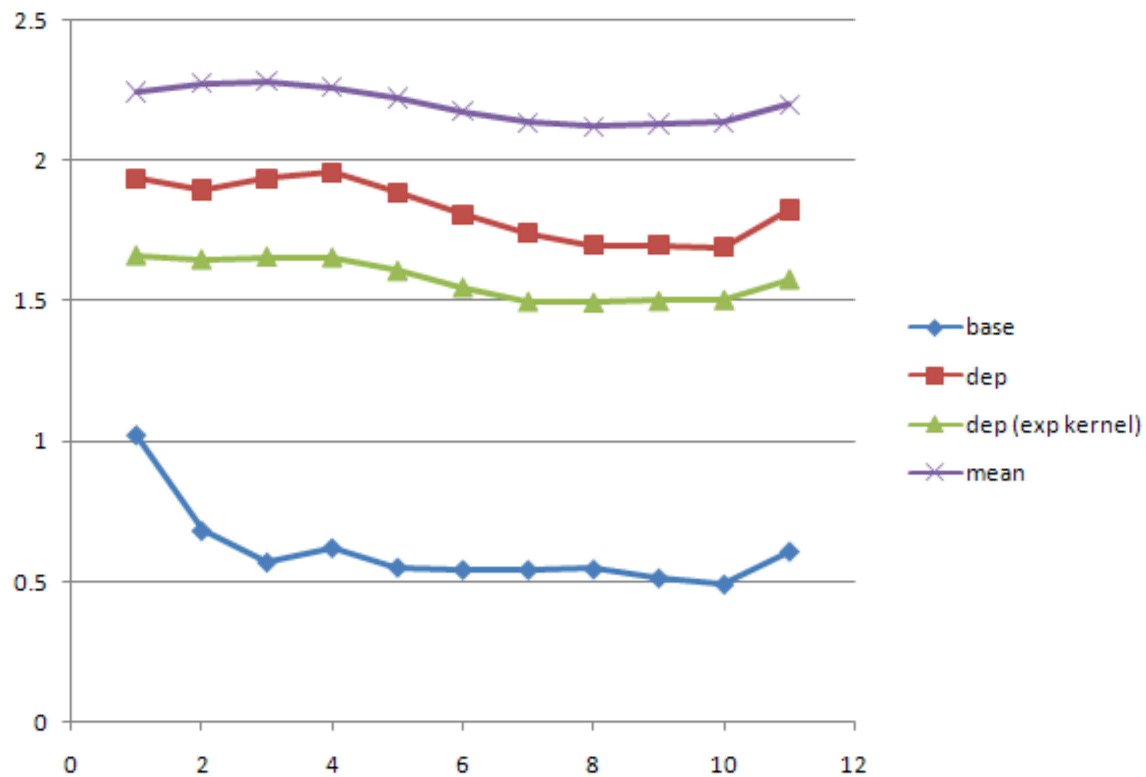
```
negative:
      -0.5677 * (normalized) w_action
      -0.554  * (normalized) w_manufactur
      -0.4998 * (normalized) w_slowli
      -0.4965 * (normalized) w_craven
      -0.4947 * (normalized) w_recent
      -0.4771 * (normalized) w_outcom
      -0.3755 * (normalized) w_crude
      -0.3732 * (normalized) w_institut
      -0.3718 * (normalized) w_affect
      -0.3694 * (normalized) w_climb
      -0.3551 * (normalized) w_canadian
      -0.3533 * (normalized) w_cumul
positive:
       0.3664 * (normalized) w_surg
       0.3679 * (normalized) w_polici
       0.3735 * (normalized) w_warehous
       0.375  * (normalized) w_resum
       0.3864 * (normalized) w_job
       0.4039 * (normalized) w_implement
       0.4054 * (normalized) w_outlook
       0.4059 * (normalized) w_struckmey
       0.4068 * (normalized) w_cutback
       0.6298 * (normalized) w_downward
       0.6536 * (normalized) w_curtail
```

Conclusions from Classification

- Shorter Periods are easier to predict than longer periods
- 13 Week Treasury bills are easier to predict than S&P 500 and 10 Year Treasury notes
- Bigrams don't help in our case
- Dependency fragments don't help either

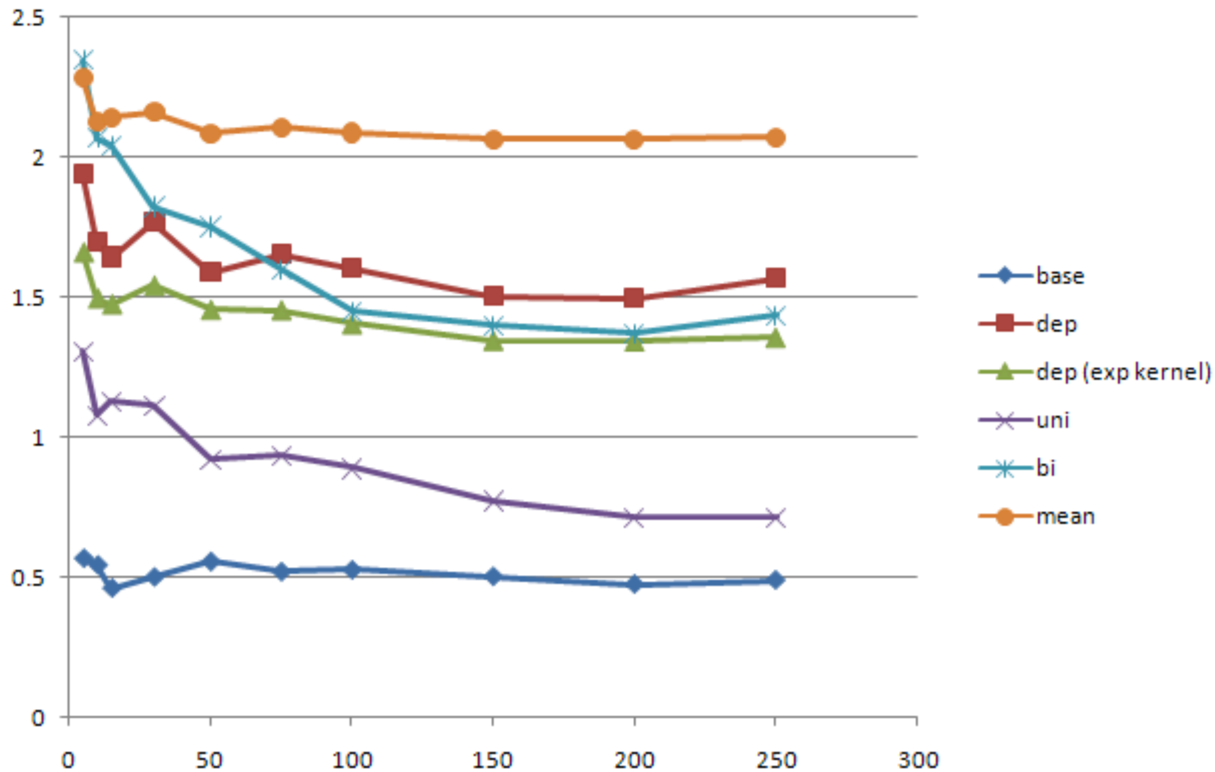
Now onto Regression…

S&P 500 – Regression – Short Periods



Now lower is better.

S&P 500 – Regression – Long Periods



Conclusions from Regression

- Regression for S&P 500 is hard – can't beat simple straw man baseline using only words

- Oddly enough, training on the previous volatility does worse than just predicting the previous volatility.
  - Over-fitting happening with just two dimensions – very surprising, a testament to the difficulty of the problem.