Analyzing SP 500 Index

based on time series and machine learning

Yanni Ge Ying Shi Hairong Xie

University of California,Berkeley

April 30, 2014

Data

SP500 data exacted from Yahoo Finance:

	AA	AAPL	ABC	ABT	ACE	ACN	ACT	ADBE	ADI
Date									
2010-01-05	15.38	206.05	24.95	22.90	43.80	38.89	39.89	37.70	27.96
2010-01-06	16.18	202.77	24.72	23.02	43.20	39.30	40.02	37.62	27.91
2010-01-07	15.84	202.40	24.32	23.21	43.45	39.27	39.70	36.89	27.69
2010-01-08	16.23	203.75	24.58	23.33	43.20	39.11	39.41	36.69	27.84
2010-01-11	16.64	201.95	24.86	23.45	43.67	39.07	39.77	36.21	27.69
2010-01-12	14.80	199.65	25.03	23.38	43.76	38.82	39.72	35.66	26.54
2010-01-13	15.24	202.47	25.52	23.61	43.90	39.27	40.89	36.28	26.53
2010-01-14	15.08	201.29	25.68	23.63	44.32	39.61	41.10	35.90	26.50
2010-01-15	14.91	197.93	25.40	23.69	43.85	39.33	40.90	35.87	25.61

Goal

1. Whose mind are we going to change?
    1. Provide investors a better understanding on SP500 and how to track its performance
    2. Give investors several tools to predict and visualize stock trends
2. How?
    1. LASSO, PCA
    2. Distribution approximation
    3. Time series

LASSO

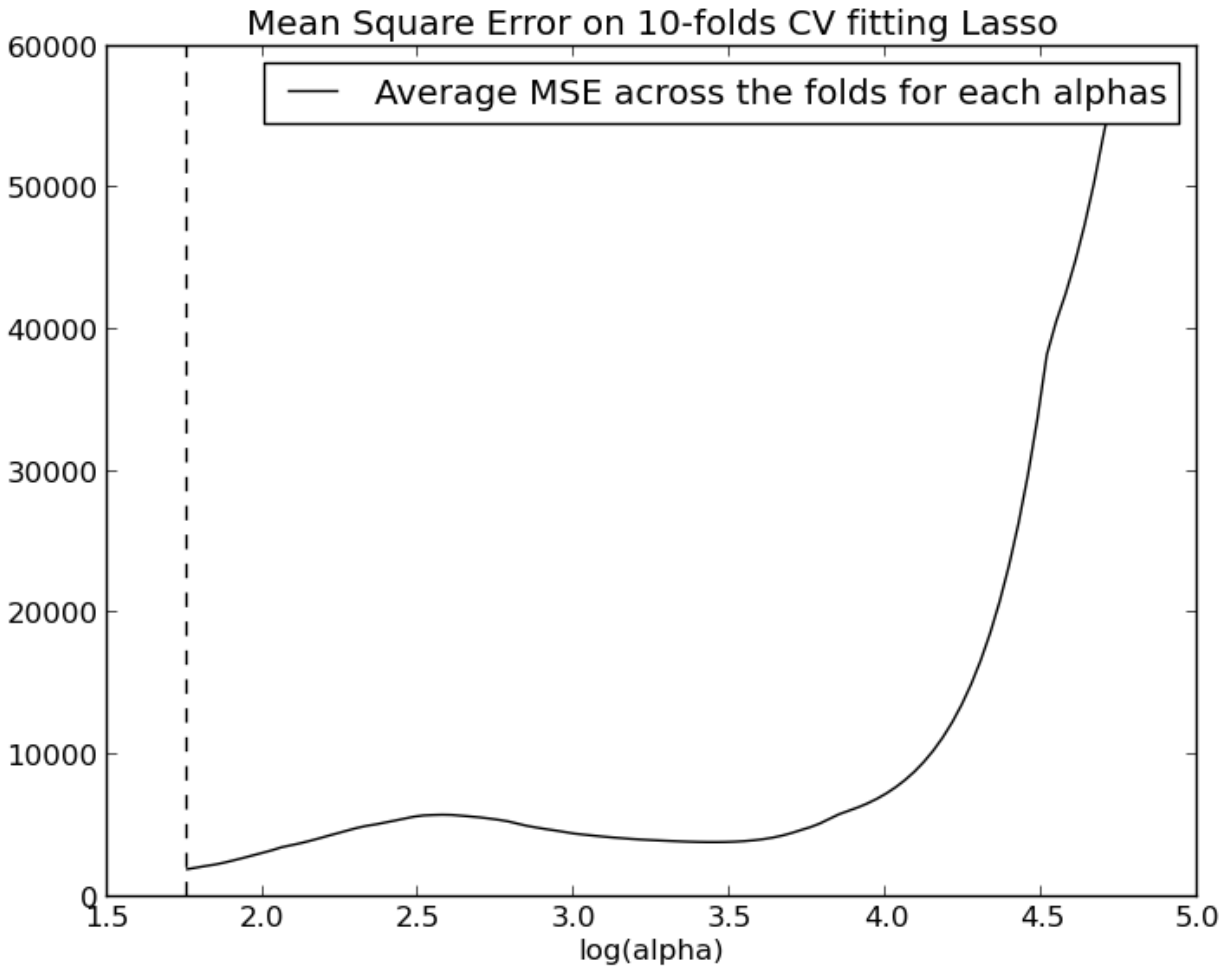

Figure 1: Penalty term of LASSO determined by using 10-fold cross-validation

LASSO

1	AAPL	Apple Inc.
2	AMZN	Amazon.com Inc.
3	BLK	BlackRock, Inc.
4	CAT	Caterpillar Inc.
5	EW	Edwards Lifesciences Corp.
6	IBM	International Business Machines Corporation
7	ISRG	Intuitive Surgical, Inc.
8	FSLR	First Solar, Inc.
9	GHC	Graham Holdings
10	GOOG	Google Inc.
11	GS	The Goldman Sachs Group, Inc.
12	GWW	W.W. Grainger, Inc.
13	NFLX	Netflix, Inc.
14	PCLN	The Priceline Group Inc.
15	RL	Ralph Lauren Corporation
16	SHW	The Sherwin-Williams Company
17	UNP	Union Pacific Corporation

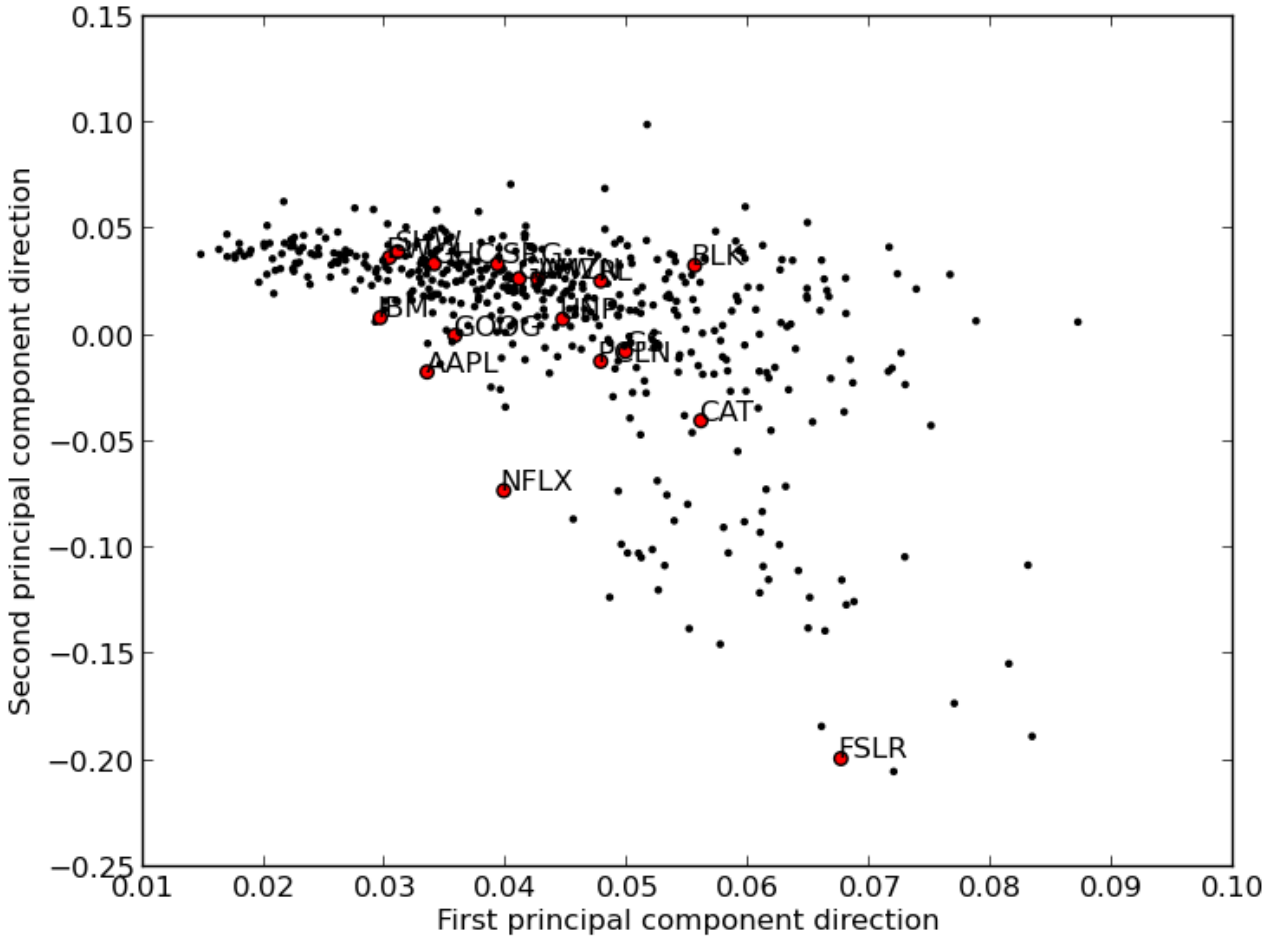

Figure 2: Visualize the 446 companies with the selected 17 companies labeled.

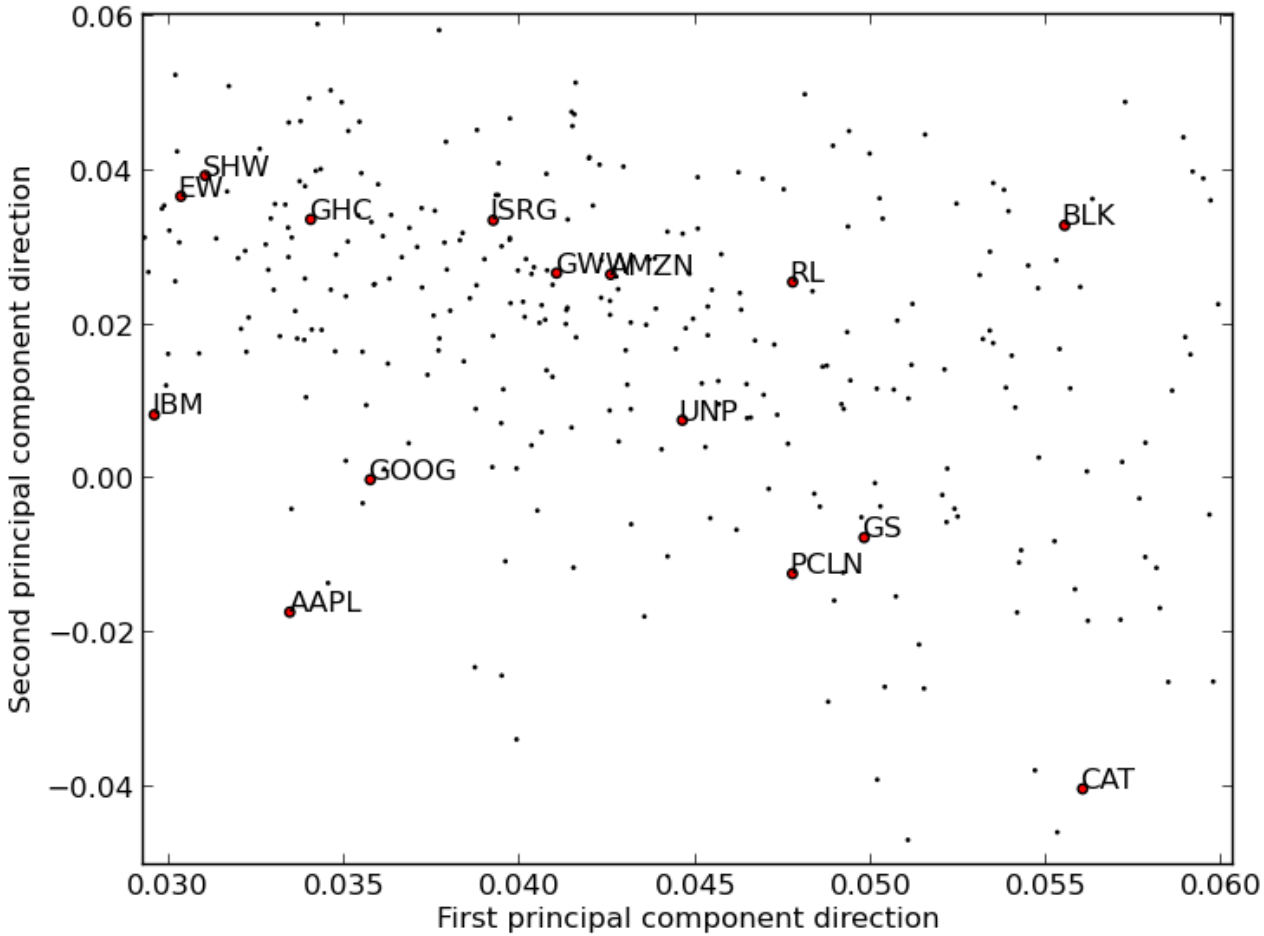

Figure 3: Zoomed-in

Scatter Plot

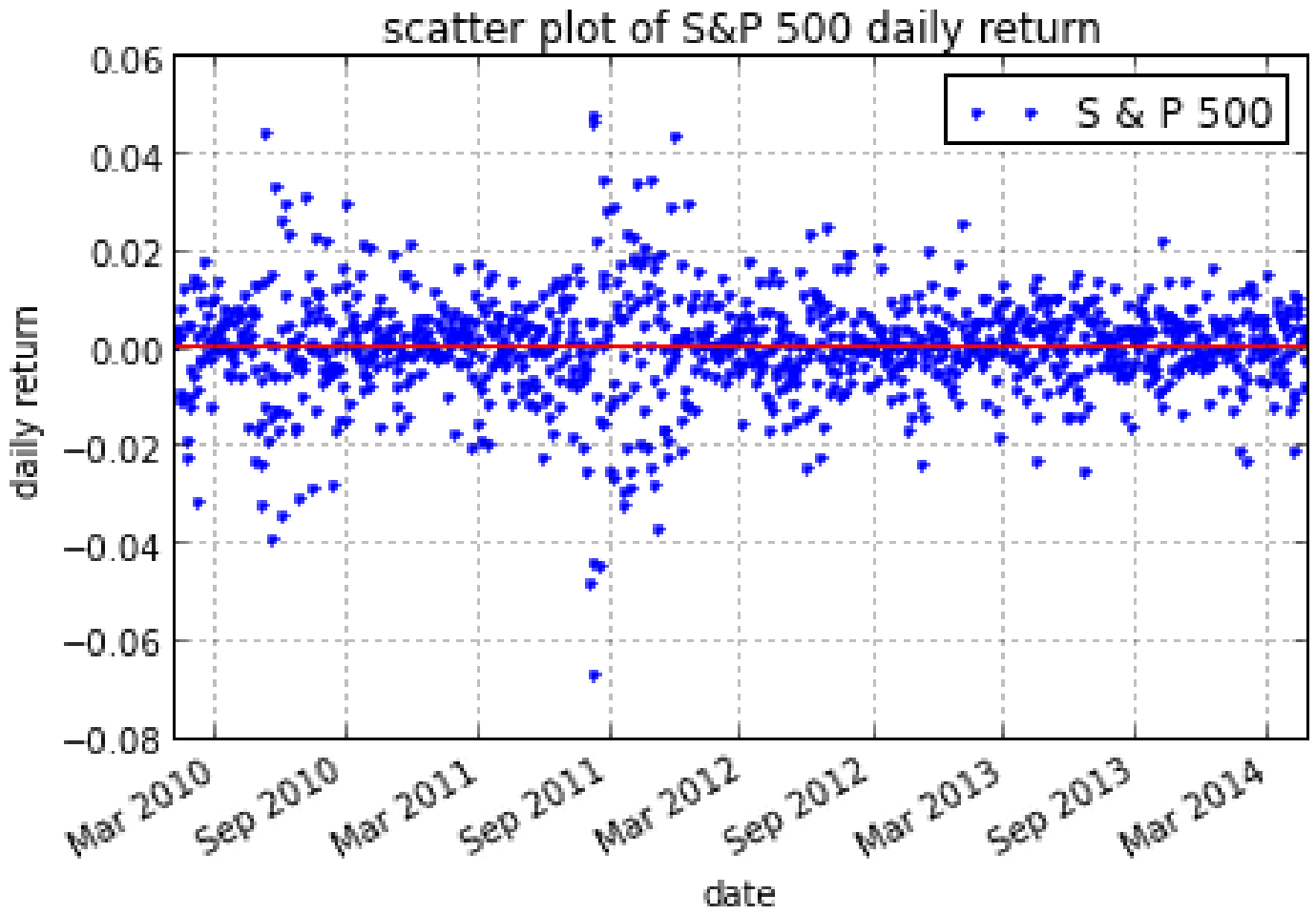

Figure 4: Scatter plot of SP500 adjusted closed price

Density Plot

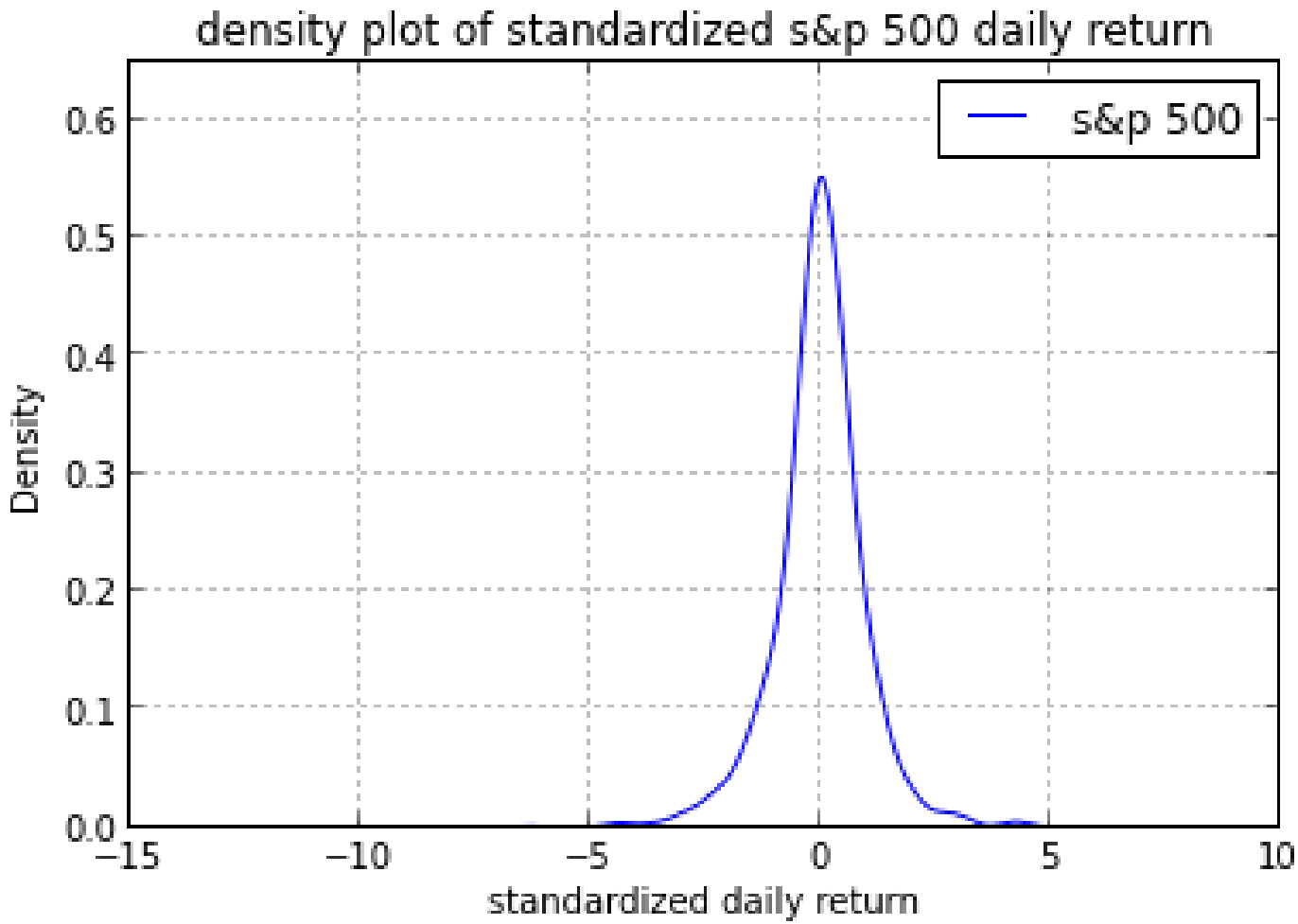

Figure 5: Density plot for standardized daily return

QQ Plot

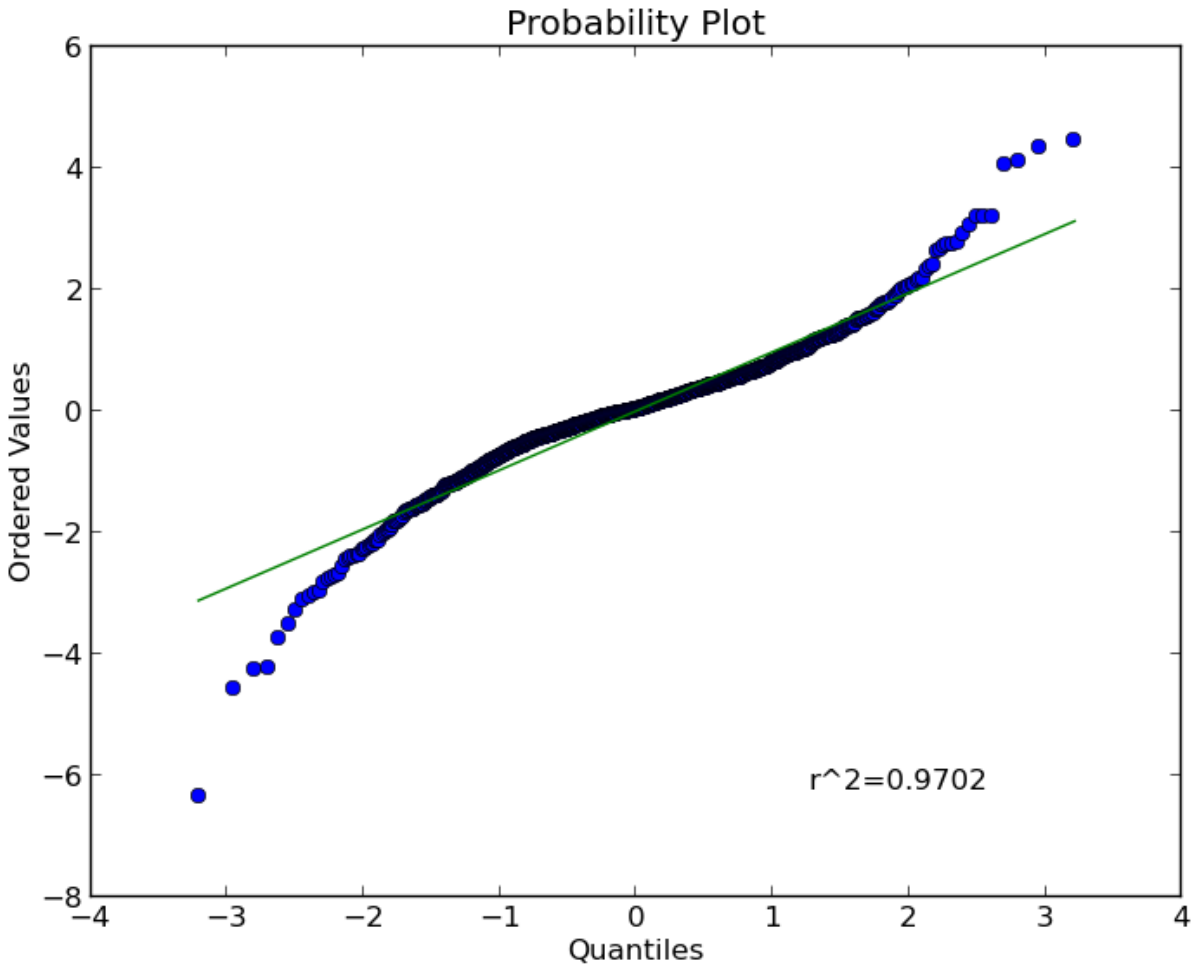

Figure 6: QQ plot for SP500 daily return

Density Plot

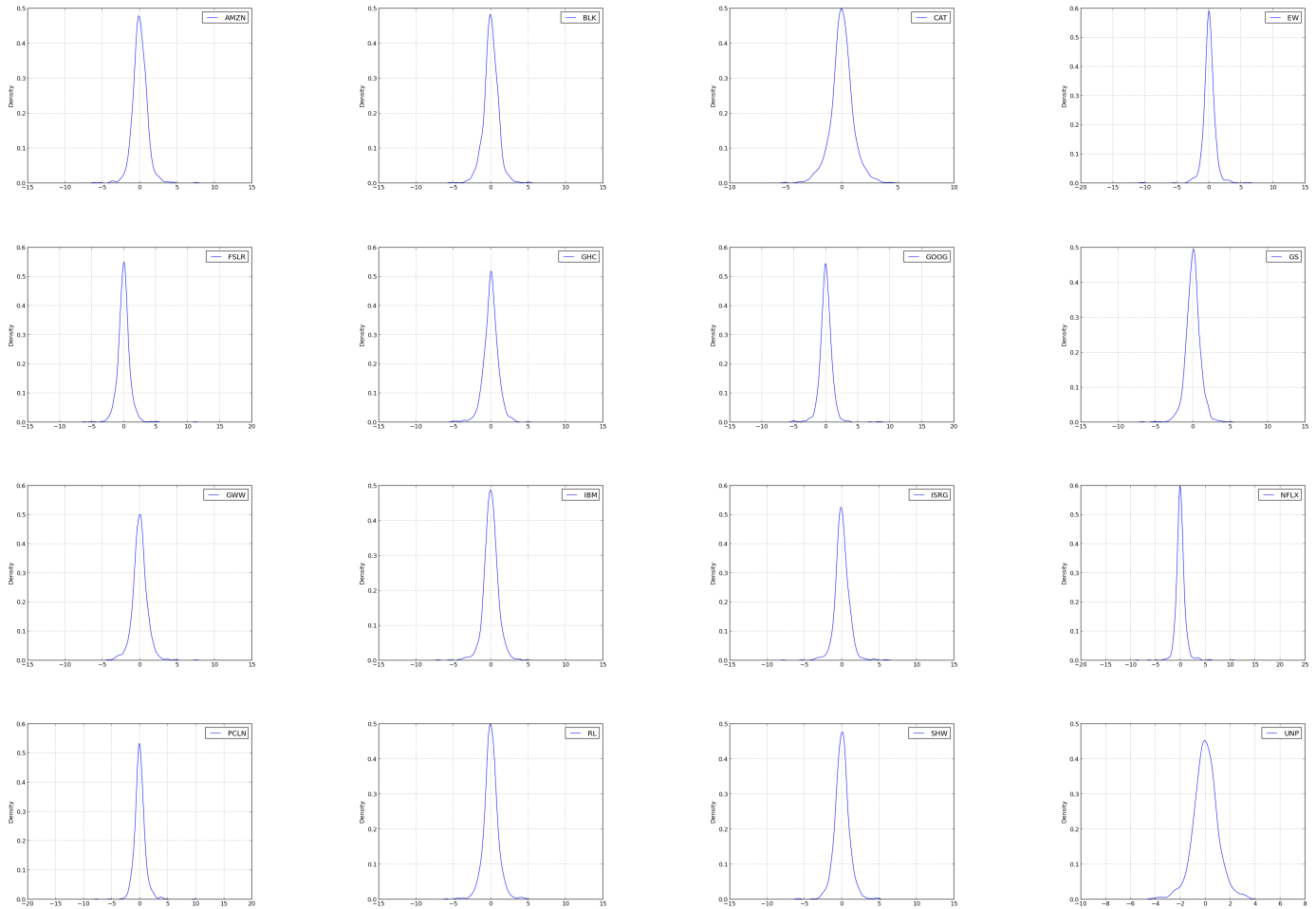

Figure 7: Density plots for selected 16 stocks

Density Plot

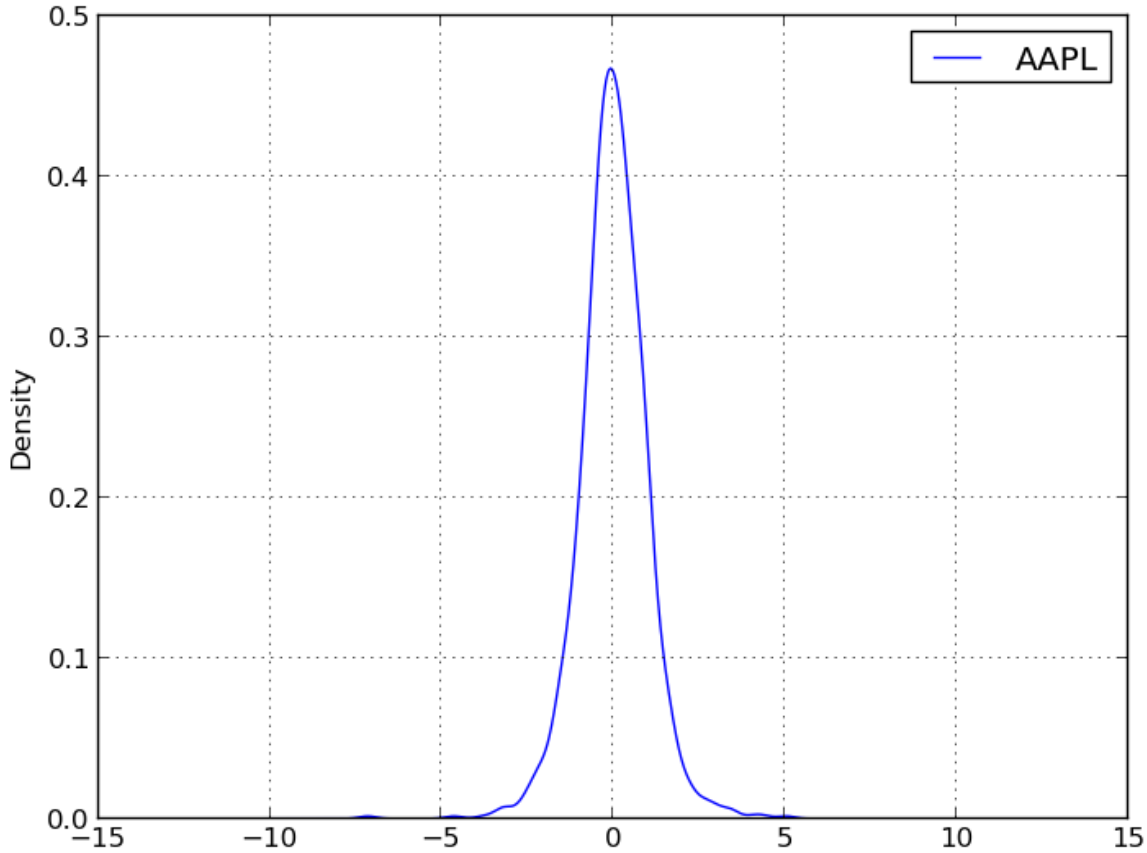

Figure 8: On Mar 14, the adjusted closed price is 524.69 and the mean daily return is 0.104%. Then the best guess for the adjusted closed price would be $524.69 * (1 + 0.104\%) = 525.24$.

Prediction

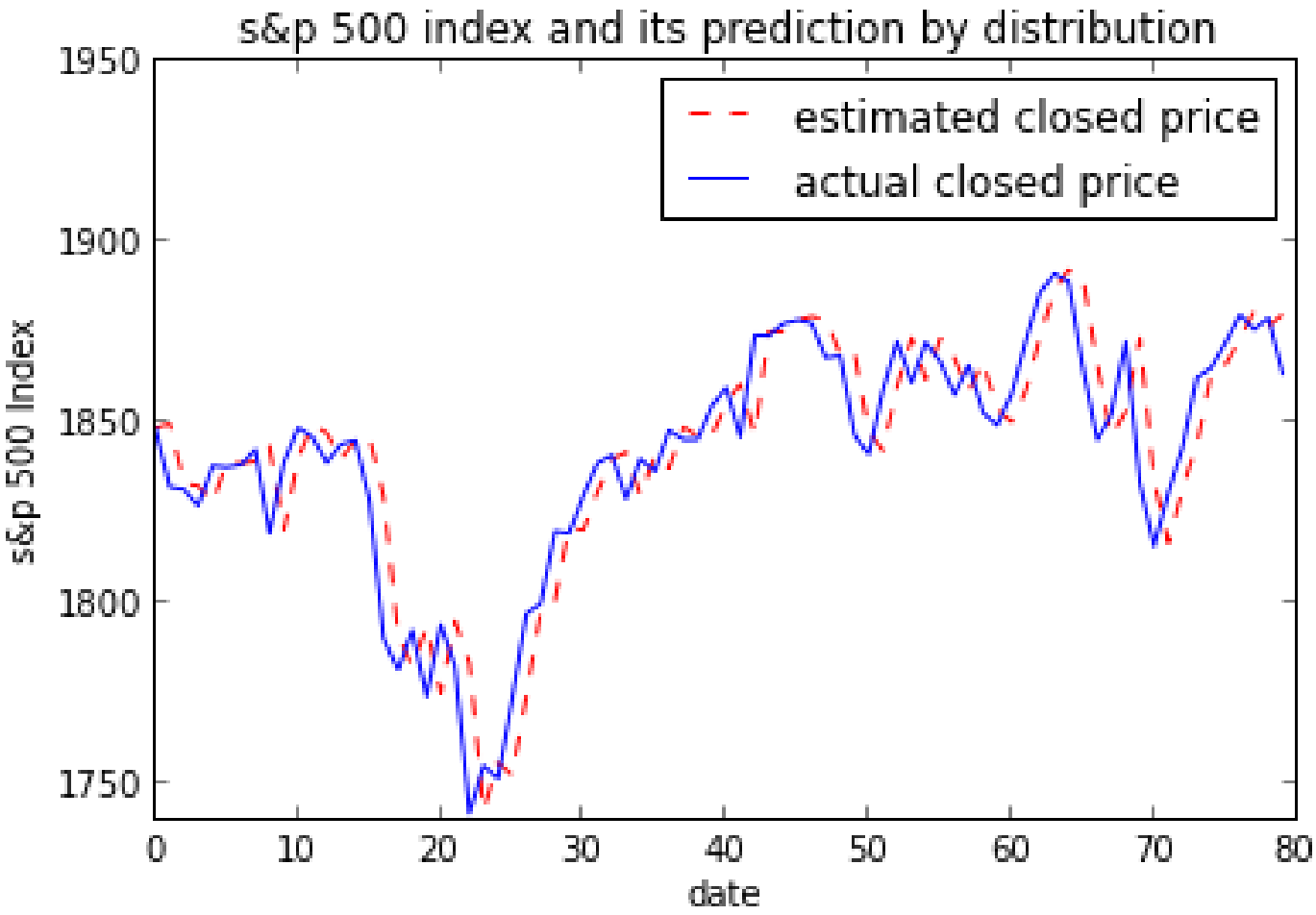

Figure 9: Prediction for SP500 index by daily return mean

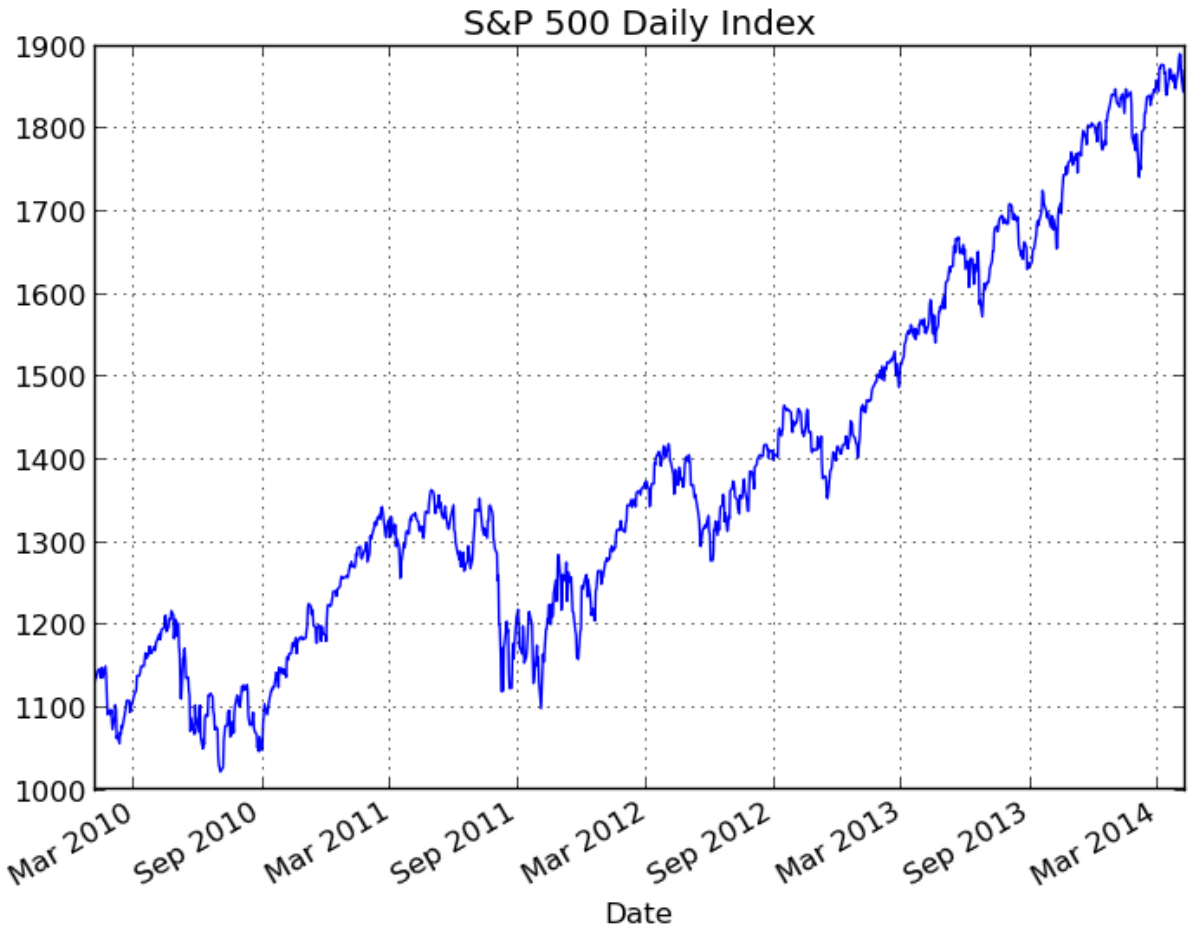

Figure 10: SP500 daily prices from Jan 4, 2010 to now

AR(p) process:

$$Y_t = \beta_0 + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t,$$

where

$$\beta_0 = \{1 - (\phi_1 + \ldots + \phi_p)\}\mu$$

MA(q) process:

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}$$

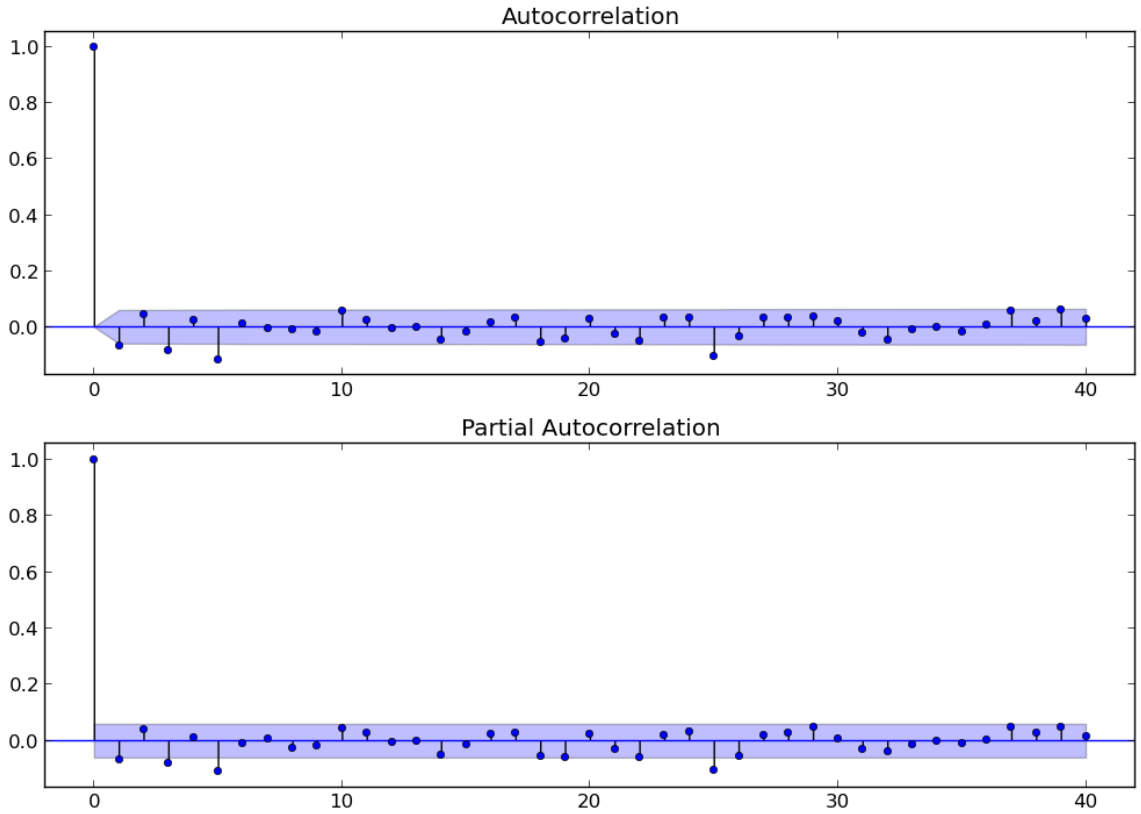

Figure 11: Autocorrelation and Partial Autocorrelation function shows that SP500 daily return is a stationary process

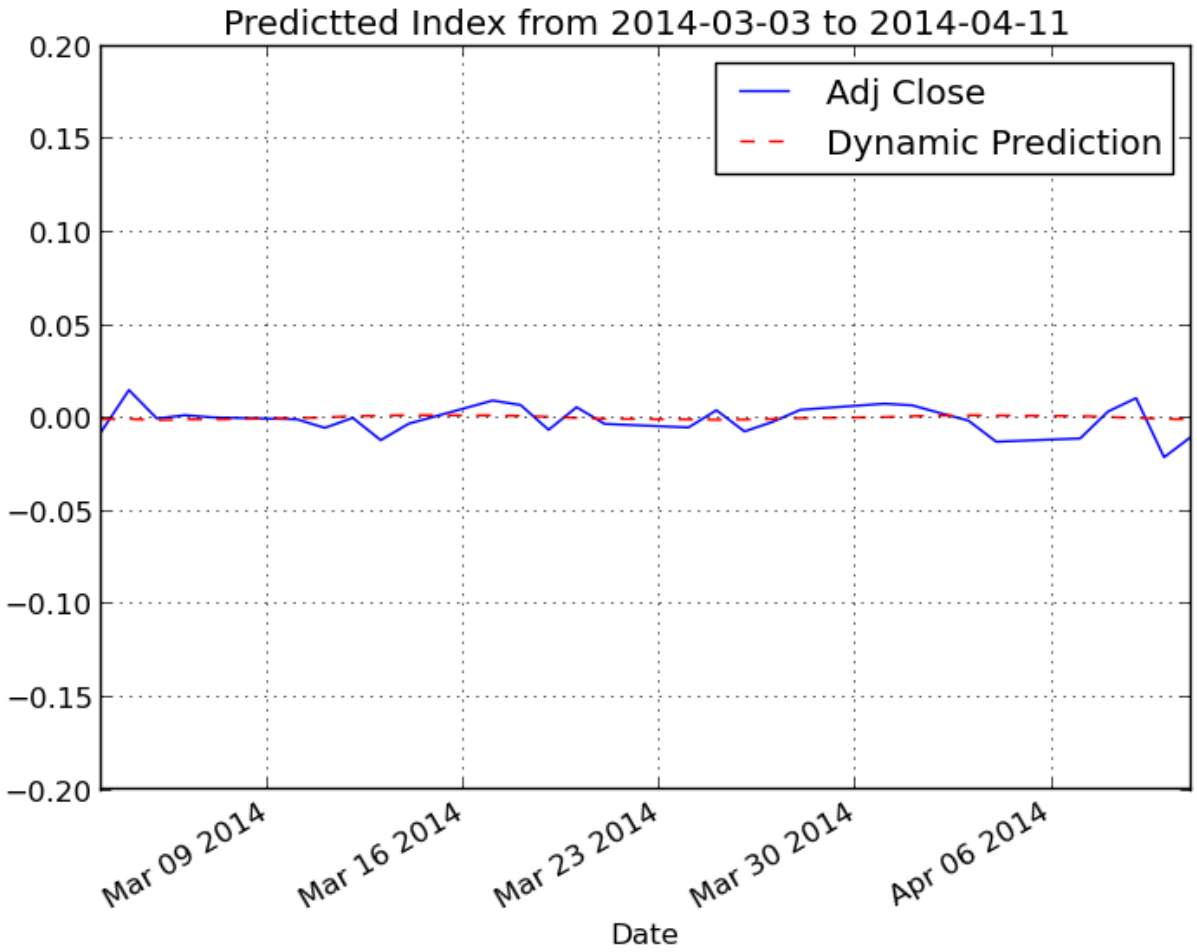

Figure 12: ARIMA(3,0,3) to predict more than 1 month daily returns

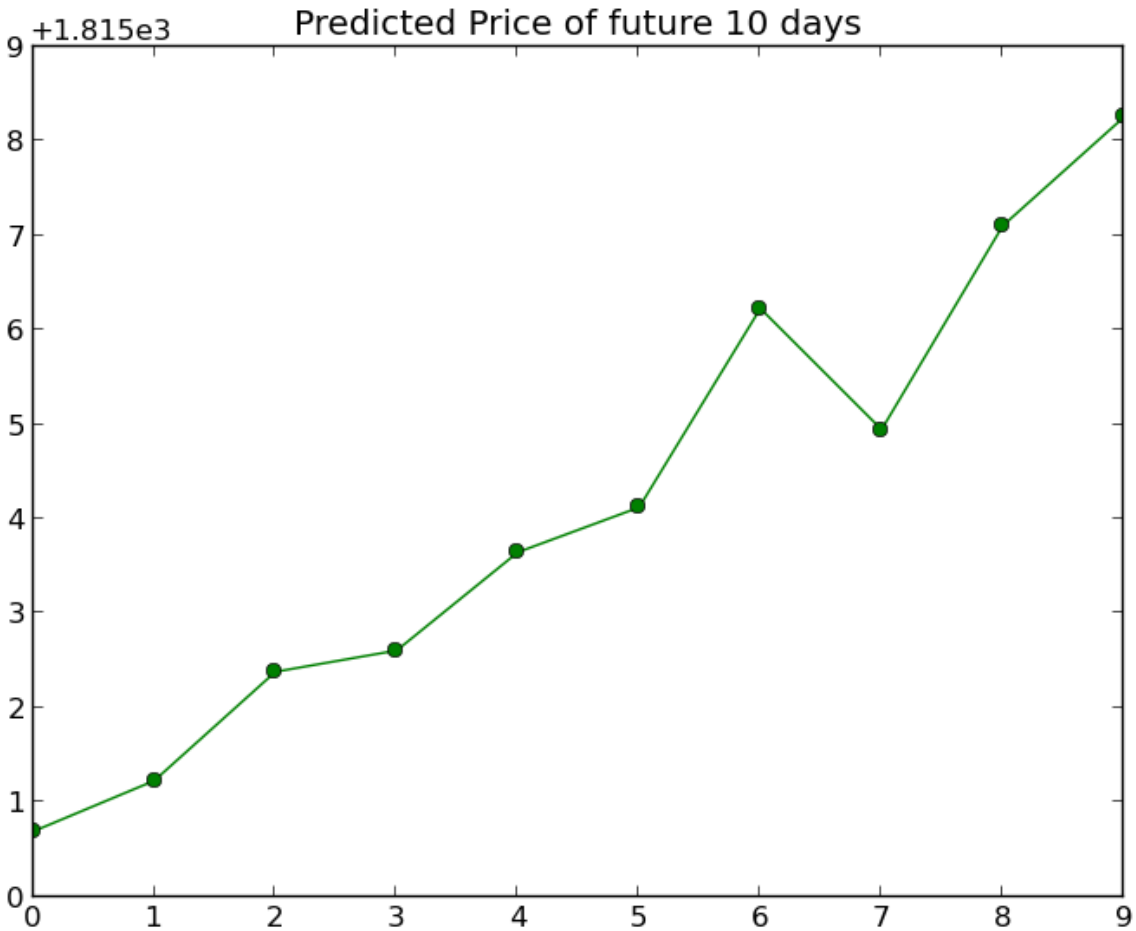

Figure 13: Predicted prices of the next 10 days from 04/11/2014

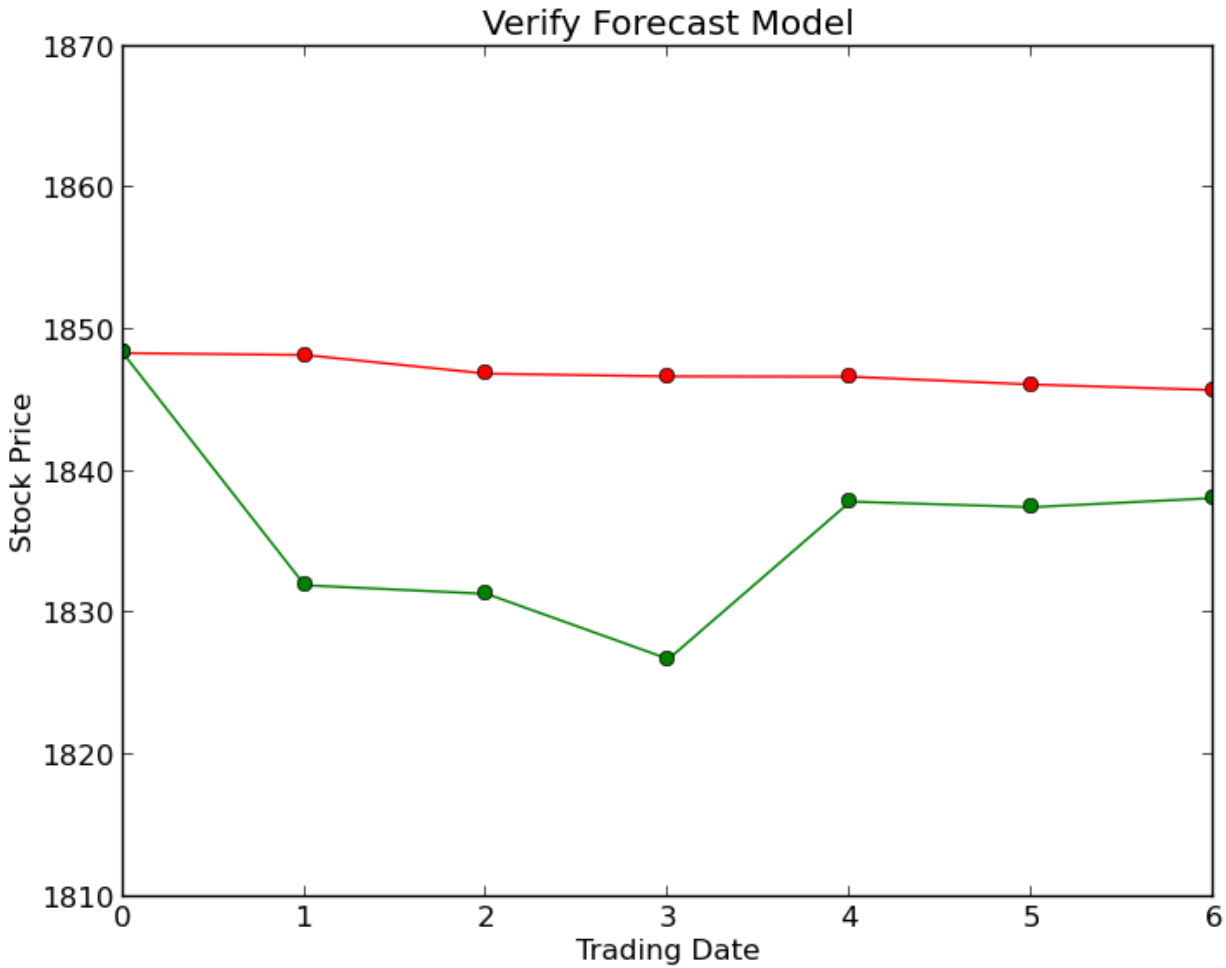

Figure 14: Cross-validation. The red line is the predicted prices, and the green line is the actual prices.

Conclusion

Offer investors:

- A portfolio of 17 stocks that well represents the SP 500 index by LASSO
- Two robust prediction models based on distribution approximation and time series

Thank you!