Long-term Volatility Forecasting using Macro Economic Indicators

Xinyi Liu
Mechanical and Aerospace Engineering Department
xinyil@princeton.edu

Abstract

This project aims to evaluate the effectiveness of various machine learning models using financial data. Multiple machine learning methods for predicting the monthly volatility of $S\&P$ 500 using 114 macro economic indicators are evaluated. Methods include penalized linear regressions and non-linear tree based models. An OLS, a random walk model, and an autoregressive model are used as benchmarks. Most of the models outperform the benchmark random walk model, and achieve a similar performance compared to an autoregressive model. In this application, linear models perform better than tree based non-linear models, where elastic net and linear gradient boosting have the highest performances. The selected features by LASSO and elastic net suggest that indicators of the labor market can provide predictive power to stock market volatility forecasting.

1 Introduction

Volatility is a measure of the variability of returns of assets and is an important measure of the market. While short term S&P volatility indicates risks in the investment and allow portfolio managers make investment decisions, long term volatility can provide insights of the economic health status. A low long term volatility indicates a stable growth of the market, while a high volatility might suggest a forthcoming financial crisis. Therefore, correctly forecasting the long term volatility is critical.

A number of works have been established to study the fundamental relationships between macroeconomic indicators and long term volatility forecasting, but understanding the robustness and the correlation between volatility and macroeconomic activities remains to be an open question[1, 3, 4, 5]. Paye, B.[8], applied econometric techniques to evaluate the increase of predictive power by incorporating macroeconomic and financial variables. He shows that macroeconomic indicators is most informative during the onset of recessions and therefore provide additional prediction power to the model. However, he also suggests that comparing to the benchmark models, only limited gains is achieved, as volatility and business cycles are strongly correlated. Lagged volatility itself contains a lot of information about the business cycle[9]. Therefore, an immediate open question arises: would more sophisticated machine learning models increase the predictive power of macroeconomic variables?

In this project, the goal is to investigate various machine learning methods to forecast monthly volatility of S&P 500 using 114 macro economic indicators and compare their performance with benchmark models. The models employed in the paper includes OLS, Ridge, Lasso, Elastic Net, Random Forest, linear boosting, and PCA. We used a random walk model, an autoregressive model, and OLS as the benchmarks. The main purpose of this project is to compare multiple models and evaluate their effectiveness in financial applications.

2 Data

The financial data of S&P 500 was downloaded from Yahoo Finance from January 1959 to December 2018. The monthly realized variance, which is the square of volatility, is calculated using the daily closing price:

$$RV_t = \sum_{i=2}^{M} r_{i,t}^2 = \sum_{i=2}^{M} (log(P_{i,t}) - log(P_{i-1,t}))^2 \tag{1}$$

where M is the number of days in each month, $r_{i,t}$ is the daily return at month t. Figure1 shows the log variance from January 1960 to December 2018, which is a vector of length 708, as there are 708 months. There are two main peaks around 1989 and 2009, indicating the financial recessions around the early 90s and during 2008.

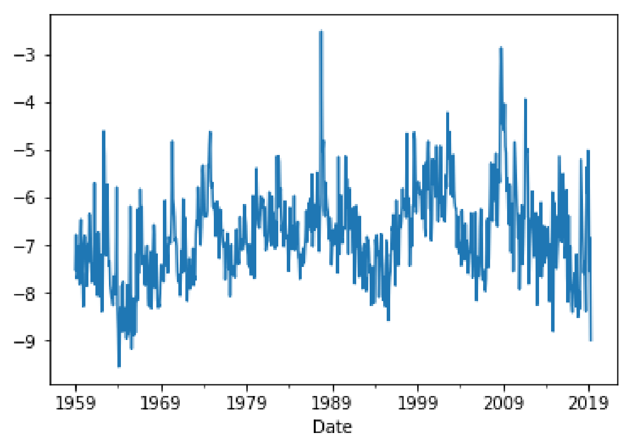

Figure 1: Log variance from January 1960 to December 2018

The macroeconomic data was downloaded from the FRED-MD database, which is a large macroeconomic database designed for economic research[6]. The FRED-MD dataset contains 114 macroeconomic indicators, which can be divided into eight subgroups: output and income; labor market; housing; consumption; money and credit; interest and exchange rate; prices; and stock market.

Besides the 114 macroeconomic indicators, historical S&P 500 with multiple lags are also added in order to achieve a more accurate prediction. This results in an X with dimensions of 708 by 119(6 lagged values included) and 708 by 125(12 lagged values included).

2.1 Data Imputation and Transformation

Missing data were imputed using backward filling. Most of macroeconomic indicators are non stationary and tend to increase over time due to economic growth and inflation. Following recommendations in the FRED-MD database appendix, macroeconomic indicators are transformed to stationary values[6].

Figure2 shows the values of RPI(Real Personal Income) and INVEST(Securities in Bank Credit at All Commercial Banks), which have different growth rates. RPI is transformed by differentiating and taking log, $\Delta log(x_t)$, while INVEST is transformed by $\Delta^2 log(x_t)$. There are 7 modes of transformations. All the training data is normalized to have zero mean with a standard deviation of one. Details on transformations can be found in FRED-MD website.

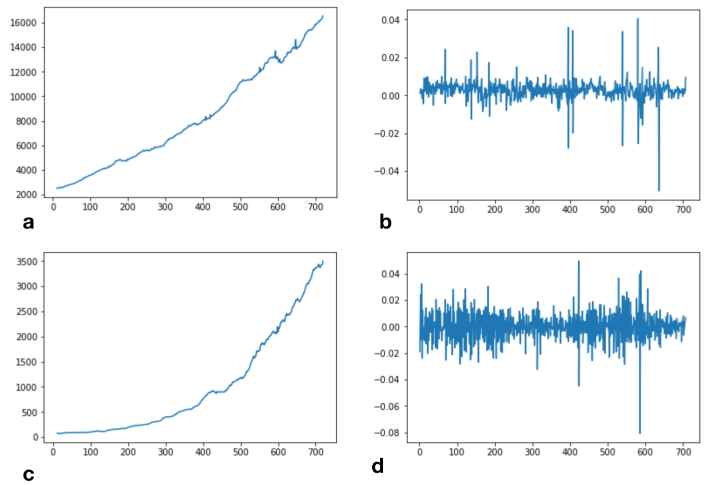

Figure 2: Transforming non-stationary data into stationary data. Real Personal Income(RPI) before transformation(a) and after transformation(b). Securities in Bank Credit at All Commercial Banks(INVEST) before transformation(c) and after transformation(d). The x axis is value series, which ranges from January 1960 to December 2018. The y axis is the absolute values before and after transformation.

3 Models

The forecasting model for realized variance can be represented as

$$log(RV_{t+1,SP500}) = g(x_t) + u_{t+1} \tag{2}$$

where x_t are features that consist past volatilities and macroeconomic indicators, $g(x_t)$ is the model, and u_{t+1} are normally distributed errors.

A rolling window with 30 years is chosen. The forecast starts in January 1990, and rolls forward to Dec 2018. For instance, data from 1960 Jan to 1989 Dec is used to forecast the volatility in January 1990. Then data from 1960 Feb to 1990 Jan is used to predict volatility in Feb 1990. To evaluate the model, mean squared error and out-of-sample R^2 are used:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3}$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{4}$$

where

$$SS_{res} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2, SS_{tot} = \sum_{i=1}^{N}(y_i - \bar{y}_i)^2 \tag{5}$$

4 Results

In this section, the model performances of various machine learning methods are compared to the three benchmark models. For each machine learning method, 1, 6, 12 lagged volatilities are included in the features with performances reported separately in Table4.1.3.

4.1 Benchmark Models

Three benchmark models are adopted, which are a random walk model, an auto regressive model, and an OLS model.

	Lag 1-month		Lag 6-months		Lag 12-months	
	R-square	MSE	R-square	MSE	R-square	MSE
Benchmark_RW	0.3420	0.6233	NA	NA	NA	NA
Benchmark_AR	0.4775	0.4949	NA	NA	NA	NA
Benchmark_OLS	0.1807	0.7762	0.2206	0.7384	0.2127	0.7458
Ridge	0.3232	0.6411	0.3671	0.5996	0.3794	0.5879
LASSO	0.4304	0.5396	0.4697	0.5024	0.4661	0.5058
Elastic Net	0.4354	0.5349	0.4698	0.5023	0.47	0.5021
PCA	0.1836	0.7734	0.3795	0.5878	0.3760	0.5911
Boosting	0.4400	0.53050	0.4691	0.5029	0.4695	0.5026
RF	0.3001	0.6630	0.3552	0.6108	0.3842	0.5833

Table 1: Out-of-sample R^2 and mean squared errors for different models, with 1-month lag, 6-month lags, and 12-months lags included.

4.1.1 Random Walk

The most simple model would be a random walk model where the volatility forecast for the current month is the volatility from the previous month. The mean squared error and out-of-sample R^2 are presented in Table4.1.3.

4.1.2 Autoregressive Model

An autoregressive model is one of the time series models that is often used in economic forecasting[8]. The autoregressive model consists lagged values with Pth order, as defined in the following equation:

$$y_t = w_0 + \sum_{i=1}^{P} w_i y_{t-i} + \epsilon_t \tag{6}$$

where $w_0, w_1, ..., w_p$ are model coefficients and P is the number of lags. P can be selected using Akaike's information criterion(Akaike, 1969), which is a measure of model fit and model complexity.

$$AIC(p) = Nlog\hat{\sigma}^2(p) + 2p \tag{7}$$

where $\hat{\sigma}$ is the mean squared error, and p is the degree of freedom. The optimal order p is chosen when AIC(p) reaches its minimum. Here we limit the maximum lags to be 6 and 12, which corresponds to 6 months and a year. For volatilities that have strong dependencies on business cycles, a lot of information is actually contained in its past volatilities. Therefore, it is not surprising that by only including past volatilities with different lags, a high accuracy can usually be achieved(Table4.1.3.

4.1.3 OLS

OLS is also used as a benchmark model for comparing different machine learning methods. Since the number of features is quite large and those macroeconomic indicators do not have strong predictions, therefore, OLS does not perform very well, which only has an R^2 of 0.1807 as shown in Table4.1.3. By adding more lagged volatilities(later referred to as Lag-6 model and Lag-12 model), the out-of-sample R^2 increases from 0.1807(Lag-1), to 0.2206(Lag-6) and 0.2127(Lag-12). The slight performance decrease from Lag-6 to Lag-12 suggests that adding more lagged terms beyond 6 months no longer add additional predictive power to the model.

4.2 Ridge

Ridge is a penalized linear method that adds an L2 penalty to the objective function to reduce overfitting. For models that only include one past volatility(referred later as Lag-1 Model), the out-of-sample R^2 is only 0.3232 which is less than the random walk(0.3420) and AR benchmarks(0.4775), but is much higher than the Lag-1, Lag-6, and Lag-12 OLS benchmarks as shown in Table4.1.3. This suggests that compared to OLS, Ridge reduces variance by adding regularization to the system and therefore leads to an overall better performance.

4.3 Lasso

Lasso adds an L1 penalized term to the OLS objective function and can select relevant variables by setting the coefficients of irrelevant variables to zero. Compared to Ridge, Lasso has a much stronger performance due to its feature selection nature. Since financial data is usually quite noisy and sometimes collinear, by only selecting the relevant features, Lasso can significantly reduce the feature size.

Figure3 compares the number of selected variables for Ridge, LASSO, and elastic net using the Lag-6 model. Since ridge does not setting any coefficients to zero, the number of relevant variables are constant. Both LASSO and elastic net tend to select more variables before 1994 and after 2009. Lasso reduces relevant features from over 100 to less than 20. Elastic net selects more features than LASSO, but still less than 20.

Table4 shows the selected features by Lasso. The first column shows the selected features, and they are sorted by the number of times a feature is selected from a total number of 348 rolling windows. For instance, the top 5 selected features are lag values. 'lag1' and 'SP500RVyahoo' are selected in all the windows. Lag4, lag2 and lag5 of past SP500RV are also highly selected. Besides those lagged values, 'Personal Consumption', 'Materials', 'Real Money Stock', 'All Employees: Wholesale Trade' are also selected quite often.

To get more insights about the categories of those selected macroeconomic indicators, I calculate the percentage of times a certain group is selected using the count in Figure4. The result is presented in Figure5. Lagged volatilities are selected 60% of time, followed by indicators in Labor Market(9.72%), indicators in Output and Income(8.76%), indicators in Prices(6.46%). Housing and Interests/Exchange Rates are less relevant, which makes intuitively sense, since housing price is less sensitive to economic fluctuations, and might take a long time to reflect the changes. On the other hand, labor market and Output/Income are most sensitive to the economic activities, and therefore can provide predictive power to the stock market volatilities.

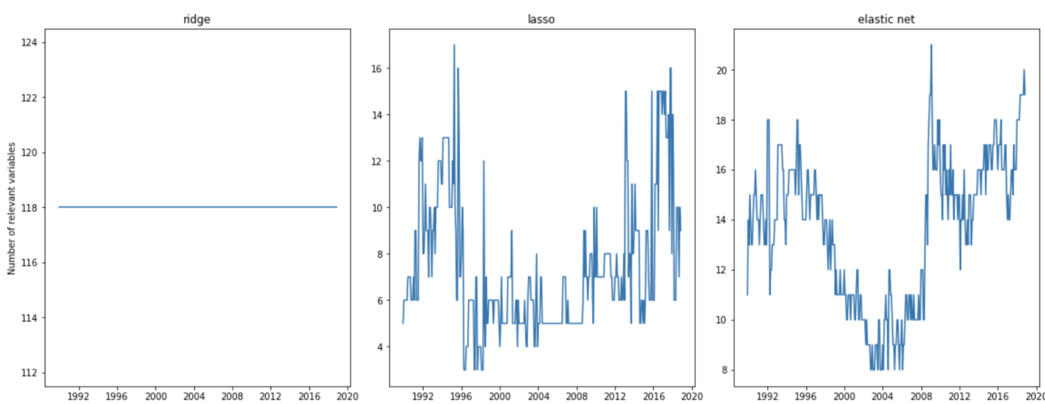

Figure 3: The number of relevant features selected by Ridge, LASSO, and elastic net using Lag 6 model for each rolling window.

4.4 Elastic Net

Elastic net combines Lasso and Ridge by including both L1 and L2 regularization terms, and therefore achieves the highest performance among models in the Lasso family. Furthermore, by adding an L2 term, elastic net has a strictly convex objective function, and becomes more stable and easier

	count	Title	Group	Description
lag1	348	lag1	9	Laged Volatilities
SP500_RV_yahoo	348	Lag0	9	Laged Volatilities
lag4	320	lag4	9	Laged Volatilities
lag2	252	lag2	9	Laged Volatilities
lag5	242	lag5	9	Laged Volatilities
DDURRG3M086SBEA	136	Personal Cons. Exp: Durable goods	7	Prices
IPMAT	112	IP: Materials	1	Output and Income
M2REAL	100	Real MÂ£ Money Stock	5	Money and Credit
USWTRADE	73	All Employees: Wholesale Trade	2	Labor Market
S&P 500	67	S&PÃ s Common Stock Price Index: Composite	8	Stock Market
RPI	61	Real Personal Income	1	Output and Income
RETAILx	53	Retail and Food Services Sales	4	Consumption
CMRMTSPLx	44	Real Manu. and Trade Industries Sales	4	Consumption
REALLN	40	Real Estate Loans at All Commercial Banks	5	Money and Credit
USGOVT	37	All Employees: Government	2	Labor Market
IPBUSEQ	35	IP: Business Equipment	1	Output and Income
USTPU	33	All Employees: Trade, Transportation & Utilities	2	Labor Market
CLAIMSx	32	Initial Claims	2	Labor Market
CUSR0000SAS	22	CPI : Services	7	Prices
PERMITMW	20	New Private Housing Permits, Midwest (SAAR)	3	Housing

Figure 4: LASSO: Number of times a certain feature is selected over a total of 348 windows. The table presents the first 20 relevant features, sorted from the high to low. The title and the macroeconomic groups can be found on the right.

LASSO

Group	Fraction	Description
1	0.087550	Output and Income
2	0.097189	Labor Market
3	0.008032	Housing
4	0.041767	Consumption
5	0.057430	Money and Credit
6	0.010040	Interests and Exchange Rates
7	0.064659	Prices
8	0.026908	Stock Market
9	0.606426	Laged Volatilities

EN

Group	Fraction	Description
1	0.101071	Output and Income
2	0.195895	Labor Market
3	0.014949	Housing
4	0.050647	Consumption
5	0.078536	Money and Credit
6	0.027220	Interests and Exchange Rates
7	0.088130	Prices
8	0.069612	Stock Market
9	0.373940	Laged Volatilities

Figure 5: The relative importance of macroeconomic indicators from the 9 groups, for LASSO and elastic net. The fraction is calculated using the number of times a group indicator is selected.

to solve compared to Lasso. The out-of-sample R^2 are 0.4354, 0.4698, and 0.47 for Lag-1, Lag-6, and Lag-12 respectively. The performance increases as more lagged terms are included. Since elastic net can select features, the overfitting problem due to increased number of features becomes less a problem compared to Ridge.

As shown in Figure3, elastic net selects more features than Lasso. Figure6 presents the top 20 relevant features. Similar to Lasso, the top 5 relevant features are lagged variances, followed by 'Personal consumption' in Prices, 'Real Personal Income' in Output and Income, 'Wholesale Trade' in Labor Market, 'Materials' in Output and Income. Many of selected variables using elastic net are the same as the ones selected using Lasso. Again, Figure5 shows the relative importance of group indicators. Lagged volatilities are most relevant, followed by labor market and output/income.

	count	Title	Group	Description
SP500_RV_yahoo	348	Lag0	9	Laged Volatilities
lag1	348	lag1	9	Laged Volatilities
lag4	347	lag4	9	Laged Volatilities
lag2	341	lag2	9	Laged Volatilities
S&P 500	312	S&PÃ s Common Stock Price Index: Composite	8	Stock Market
lag5	292	lag5	9	Laged Volatilities
DDURRG3M086SBEA	252	Personal Cons. Exp: Durable goods	7	Prices
RPI	194	Real Personal Income	1	Output and Income
USWTRADE	194	All Employees: Wholesale Trade	2	Labor Market
IPMAT	142	IP: Materials	1	Output and Income
M2REAL	134	Real MÂ£ Money Stock	5	Money and Credit
USGOVT	125	All Employees: Government	2	Labor Market
REALLN	124	Real Estate Loans at All Commercial Banks	5	Money and Credit
CE16OV	103	Civilian Employment	2	Labor Market
GS10	93	10-Year Treasury Rate	6	Interests and Exchange Rates
CUSR0000SAS	90	CPI : Services	7	Prices
RETAILx	88	Retail and Food Services Sales	4	Consumption
CMRMTSPLx	77	Real Manu. and Trade Industries Sales	4	Consumption
CLAIMSx	70	Initial Claims	2	Labor Market
USTPU	66	All Employees: Trade, Transportation & Utilities	2	Labor Market

Figure 6: Elastic net: Number of times a certain feature is selected over a total of 348 windows. The table presents the first 20 relevant features, sorted from the high to low. The title and the macroeconomic groups can be found on the right.

4.5 Linear Gradient Boosting

Boosting and random forests are two ensemble methods where a group of simple models are combined in parallel or sequentially to form a strong model. There are many types of boosting, such as gradient boosting, adaboosting, linear factor boosting[2] and etc, but they all share the same idea of using sequential weak learners to improve based on the feedback from the previous learner. In this project, linear gradient boosting is used[2], which sequentially chooses the best feature at each time step. At each step, the model loops over all the features and picks the one that fits the best to the residual until reaching an early stopping threshold. With a fixed learning rate, the model is updated every time, and the calculated residual is then used to fit the next model. Detailed algorithms can be found in [7].

Linear gradient boosting and elastic net have the highest performance among all the machine learning models. Linear boosting achieves an out-of-sample R^2 of 0.4400, 0.4691, and 0.4695 for Lag-1, Lag-6, and Lag-12 months models respectively as shown in Table4.1.3. It is interesting to notice that linear boosting outperforms elastic net at Lag-1 model, but loses slightly at Lag-6 and Lag-12 models. The reason is that linear boosting aggressively exploits all the information in the training set, and can benefit the most when there are only a limited amount of information available. If too much information is provided, those information brings noise to the system and may reduce the performance.

4.6 Random Forest

Random forest is widely used in many applications and usually has robust results. The randomness in RF comes from selecting a subset of samples as well as from selecting a subset of features. In this project, block bootstrap is used to select a subset of samples, since it can preserve the order in time series data, and prevent forward looking bias. The number of features is set as the square root of the total features. Altogether there are 100 regression tress. By averaging the trees, randomness is introduced in the fitting and therefore reduce the variance. Unfortunately, even though random forest outperforms the random walk model at Lag-6 and Lag-12, the model performance is less compared to linear models.

4.7 PCA

Principal component analysis is a dimension reduction technique and is especially useful if the feature space is larger than the sample space. PCA projects the data onto a set of orthogonal axis where the maximum amount of variance in the data is preserved. By choosing only the first k principal directions, the data can be transferred into a lower dimensional feature space.

After PCA, the feature size is reduced from 114 to k, where k is the number of principle components (k <114). The number of principle components is a hyper parameter, which can be chosen by specifying the percentage of variance explained. Usually a percentage of 90, 95, or 99 is used. Here I look into the effects of pre-specified percent variance on model performance. Figure7 shows out-of-sample R^2 as a function of the fraction of explained variance that was specified. The performance peaks around 0.66 and start to decrease significantly after 0.95. This result suggests that most of the macroeconomic indicators are weak predictors. Including more directions brings in noise and does not improve the overall performance.

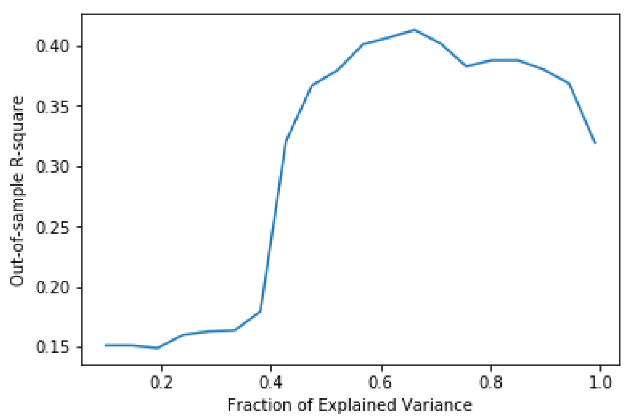

Figure 7: Out-of-sample R^2 as a function of number of principle components.

5 Conclusion

In this project, performances of multiple machine learning models in using macroeconomic indicators to forecast monthly S&P volatilities are evaluated. By comparing the out-of-sample R^2 and mean squared errors to a random walk and an autoregressive benchmark models, I find it challenging to compete with the autoregressive model, as a lot of financial information is self-contained in the lagged terms. Therefore, it is challenging to use machine learning to provide additional insights using macroeconomic indicators.

Nevertheless, elastic net and linear gradient boosting demonstrate strong predictive power. In this application, linear models tend to perform better than non-linear tree based models. It is also shown that macroeconomic indicators associate with labor market can provide additional predictive power to the model.

References

[1] Tobias Adrian and Hyun Song Shin. Liquidity and leverage. *Journal of financial intermediation*, 19(3):418–437, 2010.

[2] Jushan Bai and Serena Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008.

[3] Alexander David and Pietro Veronesi. Macroeconomic uncertainty and fear measures extracted from index options. *SSRN eLibrary*, 2009.

[4] Robert F Engle, Eric Ghysels, and Bumjean Sohn. On the economic sources of stock market volatility. In *AFA 2008 New Orleans Meetings Paper*, 2008.

[5] Sydney C Ludvigson and Serena Ng. The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222, 2007.

[6] Michael W McCracken and Serena Ng. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016.

[7] Marcelo C Medeiros, Gabriel Vasconcelos, Alvaro Veiga, and Eduardo Zilberman. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. 2018.

[8] Bradley S Paye. déjà vol: Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, 106(3):527–546, 2012.

[9] G William Schwert. Why does stock market volatility change over time? *The journal of finance*, 44(5):1115–1153, 1989.