Portfolio Optimization and Corporate Networks: Extending the Black Litterman Model

Germán Creamer

Stevens Institute of Technology, Hoboken, NJ 07030
gcreamer@stevens.edu
http://www.creamer-co.com

Abstract. The Black Litterman (BL) model for portfolio optimization combines investors' expectations with the Markowitz framework. The BL model is designed for investors with private information or with knowledge of market behavior. In this paper I propose a method where investors' expectations are based on accounting variables, recommendations of financial analysts, and social network indicators of financial analysts and corporate directors. The results show promise when compared to those of an investor that only uses market price information. I also provide recommendations about trading strategies using the results of my model.

Keywords: Link mining, social network, machine learning, computational finance, portfolio optimization,boosting, Black Litterman model.

1 Introduction

Contemporary investment literature is significantly influenced by [28, 29]'s portfolio optimization approach that suggests an optimal allocation of assets that maximizes expected return and minimizes volatility. The problem is that this mean-variance portfolio optimization process may lead to the selection of few top assets, it is very sensitive to small changes in inputs, it is based on past price history, and investors can not formally input their own knowledge of the market. As a reaction to these limitations, [4] proposed a mean-variance portfolio optimization model that included investors' expectations. This methodology creates views that represent investors' market expectations with different confidence levels, and uses these views as inputs for the selection of the optimal portfolio.

A different aspect that has not been deeply explored in the literature is the application of link mining to solve finance problems. Link mining is a set of techniques that uses different types of networks and their indicators to forecast or to model a linked domain. Link mining has had several applications [31] to different areas such as money laundering [26], telephone fraud detection [16], crime detection [32], and surveillance of the NASDAQ and other markets [26, 21]. One of the most important business applications of link mining is in the area of viral marketing or network-based marketing [15, 30, 27, 22], and more

K. Glass et al. (Eds.): COMPLEX 2012, LNICST 126, pp. 83–94, 2013.

recently, in finance. [1] have applied network analysis to quantify the flow of information through financial markets. [12] have applied a link mining algorithm called CorpInterlock to integrate the metrics of an extended corporate interlock (social network of directors and financial analysts) with corporate fundamental variables and analysts' predictions (consensus). CorpInterlock used these metrics to forecast the trend of the cumulative abnormal return and earnings surprise of US companies.

In this paper, I propose PortInterlock which is a variation of the CorpInterlock algorithm that uses the return forecast to complement or substitute the investors' view of the BL approach. This methodology may help investors to incorporate qualitative and quantitative factors into their investment decisions without the intervention of financial management experts.

2 Methods

2.1 The Black Litterman Model

The BL model is one of the most extended tactical allocation models used in the investment industry. The BL model calculates the posterior excess return using a mean variance optimization model and the investors' view. Since the introduction of the BL model [4], many authors have proposed several modifications. [25] extends the BL model including a factor uncorrelated with the market; [3] and [2] substitute the investors' views by analysts' dividend forecast and by GARCH derived views, respectively.

I follow [5, 6] in describing the BL model. Additional useful references about the BL model are [23, 33].

The excess returns of n assets over the risk free rate R_f are normally distributed and are represented by the n-vector μ. Using a Bayesian framework, the prior distribution of excess returns is $\mu \sim N(\Pi, \Sigma)$ where Π is a n-vector of implied equilibrium excess return and Σ is the nxn variance covariance matrix of excess return.

Σ can be obtained by the historical excess return and the equilibrium excess return $\Pi = \lambda \Sigma w$ is the solution to the following unconstrained return maximization problem:

$$max_w w' \Pi - \frac{\lambda w' \Sigma w}{2} \tag{1}$$

where λ is the risk aversion parameter. The vector of optimal portfolio weights can be derived from the Π formula:

$$w = (\lambda \Sigma)^{-1} \Pi \tag{2}$$

In equilibrium, the market portfolio w_{mkt} derived from the capital asset pricing model (CAPM) should be the same as the mean variance optimal portfolio w. So, the prior expected excess return should be the equilibrium expected excess return:

$$\Pi = \lambda \Sigma w_{mkt} \tag{3}$$

where $w_{mkt} = \frac{M_i}{\sum_i M_i}$ and M_i is the market capitalization value of asset i.

The main innovation of the BL model is that the investor may specify k absolute or relative scenarios or “views” about linear combinations of the expected excess return of assets. The views are independent of each other and are also independent of the CAPM. They are represented as:

$$P\mu = Q - \epsilon \tag{4}$$

P is a kxn matrix where each row represents a view. Absolute views have weights for the assets that will outperform their expected excess return and their total sum is one; relative views assign positive and negative weights to assets that over- or underperform respectively and their total sum is zero. Q is a k-vector that represents the expected excess return of each view, τ is a k-vector that represents the confidence indicator of each view, and $\epsilon \sim N(0, \Omega)$ is an error term normally distributed that represents the uncertainty of the views where Ω is a kxk diagonal covariance matrix of error terms of the views.

The posterior distribution of excess returns $\hat{\mu}$ combines the prior excess return Π and the investors' views P:

$$\hat{\mu} = N([(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\Pi + P'\Omega^{-1}P], [(\tau\Sigma)^{-1}\Pi + P'\Omega^{-1}P]^{-1}) \tag{5}$$

An alternative expression of the expected excess return is:

$$\hat{\mu} = \Pi + \tau\Sigma P'[P\tau\Sigma P' + \Omega]^{-1}[Q - P\Pi] \tag{6}$$

and the optimal portfolio weights on the unconstrained efficient frontier using the posterior distribution is:

$$\hat{w} = (\lambda\Sigma)^{-1}\hat{\mu} \tag{7}$$

2.2 Boosting

Adaboost is a machine learning algorithm invented by [19] that classifies its outputs by applying a simple learning algorithm (weak learner) to several iterations of the training set where the misclassified observations receive more weight. [18] proposed a decision tree learning algorithm called an *alternating decision tree* (ADT). In this algorithm, boosting is used to obtain the decision rules and to combine them using a weighted majority vote.

[20], followed by [8] suggested a modification of AdaBoost, called LogitBoost. LogitBoost can be interpreted as an algorithm for step-wise logistic regression. This modified version of AdaBoost–known as LogitBoost–assumes that the labels $y_i's$ were stochastically generated as a function of the $x_i's$. Then it includes $F_{t-1}(x_i)$ in the logistic function to calculate the probability of y_i, and the exponent of the logistic function becomes the weight of the training examples. Figure 1 describes Logitboost.

2.3 PortInterlock: A Link Mining Algorithm

CorpInterlock is a link mining algorithm proposed by [12] to build a bipartite social network with two partitions: one partition includes members of board of

$$
\begin{array}{l}
F_0(x) \equiv 0 \\
\text{for } t = 1 \dots T \\
\qquad w_i^t = \frac{1}{1+e^{y_i F_{t-1}(x_i)}} \\
\qquad \text{Get } h_t \text{ from } \textit{weak learner} \\
\qquad \alpha_t = \frac{1}{2} \ln \left(\frac{\sum_{i:h_t(x_i)=1, y_i=1} w_i^t}{\sum_{i:h_t(x_i)=1, y_i=-1} w_i^t} \right) \\
\qquad F_{t+1} = F_t + \alpha_t h_t
\end{array}
$$

Fig. 1. The Logitboost algorithm [20]. y_i is the binary label to be predicted, x_i corresponds to the features of an instance i, w_i^t is the weight of instance i at time t, h_t and $F_t(x)$ are the prediction rule and the prediction score at time t respectively.

directors and another partition consists of financial analysts representing companies that they cover. This social network is converted into a one-mode network where the vertices are the companies and the edges are the number of directors and analysts that every pair of companies have in common. This is the extended corporate interlock. The basic corporate interlock is calculated in the same way using only directors. The algorithm selects the largest strongly connected component of a social network and ranks its vertices using a group of investment variables presented in the appendix 1 and a group of social network statistics obtained from the basic or extended corporate interlock. Finally, the algorithm predicts the trend of a financial time series using a machine learning algorithm such as boosting. I propose the PortInterlock algorithm, an extension of CorpInterlock, to be used for portfolio optimization. This algorithm uses the trend of the financial time series predictions as the view of the investors and the prior asset excess returns to define the optimal portfolio weights (Figure 2).

Forecasting Earnings Surprise. I used the definition of earnings surprise or forecast error proposed by [14]:

$$FE \doteq \frac{\text{CONSENSUS}_q - \text{EPS}_q}{|\text{CONSENSUS}_q| + |\text{EPS}_q|}$$

where CONSENSUS_q is the mean of earnings estimate by financial analysts for quarter q, and EPS_q is the actual earnings per share for quarter q. FE is a normalized variable with values between -1 and 1. Additionally, when CONSENSUS_q is close to zero and EPS_q is not, then the denominator will not be close to zero.

The increasing importance of organizational and corporate governance issues in the stock market suggests that the integration of indicators from the corporate interlock with more traditional economic indicators may improve the forecast of FE and CAR.

The following indicators obtained by the PortInterlock algorithm captures the power relationship among directors and financial analysts:

Input: Two disjoint nonempty sets V_{11} and V_{12}, a matrix ER of historical excess returns of each asset i, a financial time series Y to be predicted, the covariance matrix Ω of error terms of the views of investors, the vector τ that represents the confidence indicator of each view, the risk factor λ, a vector M with market capitalization values of each asset i, and additional exogenous variables.

1. Build a bipartite graph $G_1(V_1, E_1)$ where its vertex set V_1 is partitioned into two disjoint sets V_{11} and V_{12} such that every edge in E_1 links a vertex in V_{11} and a vertex in V_{12}.
2. Build a one-mode graph $G_2(V_2, E_2)$ in which there exist an edge between v_i and v_j: $v_i,\ v_j \in\ V_2$ if and only if v_i and v_j share at least a vertex $u_i\ \in\ V_{12}$. The value of the edge is equal to the total number of objects in V_{12} that they have in common.
3. Calculate the largest strongly connected component of G_2 and call it $G_3(V_3, E_3)$.
4. Calculate the adjacency matrix A and geodesic distance matrix D for G_3. a_{ij} and d_{ij} are the elements of A and D respectively.
5. For each vertex $v_i \in\ V_3$ calculate the following social network indicators:
   - Degree centrality: $deg(v_i) = \sum_j a_{ij}$
   - Closeness centrality (normalized): $C_c(v_i) \doteq \frac{n-1}{\sum_j d_{ij}}$
   - Betweenness centrality: $B_c(v_i) = \sum_i \sum_j \frac{g_{kij}}{g_{kj}}$, where g_{kij} is the number of geodesic paths between vertices k and j that include vertex i, and g_{kj} is the number of geodesic paths between k and j.
   - Clustering coefficient: $CC_i = \frac{2|\{e_{ij}\}|}{deg(v_i)(deg(v_i)-1)} : v_j\ \in\ N_i,\ e_{ij} \in\ E$
   - Normalized clustering coefficient: $CC_i' = \frac{deg(v_i)}{MaxDeg} CC_i$, where MaxDeg is the maximum degree of vertex in a network
6. Merge social network indicators with any other relevant set of variables for the population under study such as analysts' forecasts and economic variables and generate test and training samples.
7. Run a machine learning algorithm with above test and training samples to predict trends of Y.
8. Define a matrix P where each row k is the multiplication of the confidence of the prediction and the prediction of the trends of Y for each asset. P represents the absolute view of the investors.
9. Obtain Q as a k-vector that represents the expected excess return of each asset or each view k, Σ as the variance covariance matrix of ER, and $\Pi\ =\ \lambda \Sigma w_{mkt}$ as the equilibrium expected excess return where $w_{mkt} = \frac{M_i}{\sum_i M_i}$
10. Optimize the portfolio using the Black Litterman model where the expected excess return is:
$\hat{\mu} = \Pi + \tau \Sigma P'[P \tau \Sigma P' + \Omega]^{-1}[Q - P\Pi]$
and the vector of optimal portfolio weights is $\hat{w} = (\lambda \Sigma)^{-1} \hat{\mu}$.

Output: Optimal portfolio weights (w).

Fig. 2. The PortInterlock algorithm

1. Degree centrality: directors and analysts of a company characterized by a high degree or degree centrality coefficient are connected through several companies.
2. Closeness centrality: directors and analysts of a company characterized by a high closeness centrality coefficient are connected through several companies that are linked through short paths.
3. Betweenness centrality: directors and analysts of a reference company characterized by a high betweenness centrality coefficient are connected through several companies. Additionally, the reference company mentioned above has a central role because it lies between several other companies, and no other company lies between this reference company and the rest of the companies.
4. Clustering coefficient: directors and analysts of a company characterized by a high clustering coefficient are probably as connected amongst themselves as is possible through several companies.

Each of the measures above show a different perspective of the relationship between directors and analysts. Hence, I could include them as features in a decision system to forecast FE and CAR. Because the importance of these features combined with a group of financial variables to predict FE may change significantly in different periods of time, I decided to use boosting, specifically Logitboost, as the learning algorithm. Boosting is well-known for its feature selection capability, its error bound proofs [19], its interpretability, and its capacity to combine continuous and discrete variables. [9, 10, 11] have already applied boosting to forecast equity prices and corporate performance showing that Logitboost performs significantly better than logistic regression, the baseline algorithm. [14] have also compared tree-induction algorithms, neural networks, naive Bayesian learning, and genetic algorithms to classify the earnings surprise before announcement.

3 Experiments

The asset price and return series are restricted to the US stock market. They are from the Center for Research in Security Prices (CRSP), the accounting variables from COMPUSTAT[1], the list of financial analysts and earnings forecast or consensus from IBES, and the annual list of directors for the period 1996 - 2005 is from the Investor Responsibility Research Center. The number of companies under study changes every year. The minimum and maximum number of companies included in my study are 3,043 for 2005 and 4,215 for 1998.

I implemented the PortInterlock algorithm (Figure 2) with the software Pajek [13] to obtain the basic (social network of directors) and extended corporate interlock. I computed the investment signals as described in appendix 1 and the social network statistics introduced in the previous section of the basic and extended corporate interlock. I merged the accounting information, analysts'

[1] COMPUSTAT is an accounting database managed by Standard & Poor's.

predictions (consensus) and social networks statistics using quarterly data and selected the last quarter available for every year.[2] I forecasted the trend of FE and CAR. CAR is calculated using the cumulative abnormal return of the month following the earnings announcement. Every instance has the label 1 if the trend was positive and -1 otherwise. CAR is calculated as the return of a specific asset minus the value weighted average return of all assets in its risk-level portfolio according to CRSP. FE is based on the predictions of the analysts available 20 days before the earnings announcement as fund managers may suggest [14]. Fund managers take a position, short or long [3], a certain number of days before the earnings announcement and, according to their strategy, they will liquidate the position a given number of days after the earnings announcement. Investors profit when the market moves in the direction expected and above a certain threshold, even though the market movement might not be in the exact amount forecasted.

I restricted my analysis to trading strategies using FE because the prediction of FE (test error of 19.09%) outperformed the prediction of CAR (test error of 47.56%). According to [12], the long-only portfolio is the most profitable strategy when it is compared with a long-short, a long-short for the most precise decile, and a long only strategy when analysts predict that earnings will be larger than consensus. Based on these results, the weights of the long-only portfolio multiplied by the confidence of the prediction are used as the investors' views of the BL model. This portfolio is compared against a market portfolio where the weight of each asset is based on its market capitalization.

4 Results

Table 1 compares the result of several views based on a portfolio completely generated by the PortInterlock algorithm with an equally weighted portfolio and the market portfolio. The PortInterlock portfolio and the investors' view based on the PortInterlock show the largest Sharpe ratio (risk-adjusted return). When the confidence in this view decreases (lower Ω and τ), the Sharpe ratio deteriorates. The Sharpe ratio decreases even more when the risk parameter (λ) increases. The difference of Sharpe ratios between these scenarios and the market portfolio is significant according to the heteroskedasticity and autocorrelation robust (HAC) estimation test.

In the simulations, a portfolio based on social networks and fundamental indicators with high confidence in the investors' perspective has an annual Sharpe ratio of 6.56, while the market portfolio with 20% confidence has an annual Sharpe ratio of 1.415.

[2] Most of the fundamental and accounting variables used are well-known in the finance literature and [24] demonstrated that these variables are good predictors of cross-sectional returns.

[3] Long or short positions refer to buy a specific asset or to sell a borrowed asset based on the expectation that price of the asset will increase or decrease respectively.

Table 1. Annual Sharpe ratio, risk and return by portfolio.
Ω is the covariance and τ is the confidence indicator in a particular view (according to equation 6). λ is a risk factor. BL is the Black Litterman model that includes investors' views. Sharpe ratio is the ratio of mean and standard deviation of excess return over the risk free rate. *,**: 95% & 99% confidence level of the Sharpe ratio difference between each scenario and the market portfolio.

Portfolios/Views	Sharpe	Risk	Return
BL, PI, τ=1,Ω=0.000001, λ=0.00001	6.563 **	0.49	31.36% *
BL, CI, τ=1,Ω=0.0001,λ=0.0001	6.563 **	0.49	31.36% *
BL, CI, τ=1,Ω=0.001,λ=0.001	6.561 **	0.49	31.36% *
BL, CI, τ=1,Ω=0.001,λ=0.0025	6.558 **	0.49	31.35% *
BL, CI, τ=1,Ω=0.001,λ=0.005	6.552 **	0.49	31.34% *
BL, CI, τ=1,Ω=0.001,λ=0.01	6.542 **	0.49	31.32% *
BL, CI, τ=1,Ω=0.001,λ=0.5	5.438 **	0.51	29.26% *
BL, CI, τ=1,Ω=0.001,λ=1	4.402 **	0.53	27.53%
BL, CI, τ=0.5,Ω=0.001,λ=1	4.063 **	0.55	27.34%
BL, CI, τ=0.01,Ω=0.001,λ=1	1.604 *	0.61	14.06%
BL, CI, τ=0.005,Ω=0.001,λ=1	1.517 *	0.62	13.59%
BL, CI, τ=0.0025,Ω=0.001,λ=1	1.468 *	0.62	13.31%
BL, CI, τ=0.001,Ω=0.001,λ=1	1.437 *	0.63	13.13%
PortInterlock (PI): soc.network	6.563 **	0.49	31.36%
Equally weighted	2.840 *	0.47	14.01%
Market portfolio	1.415	0.63	13.00%

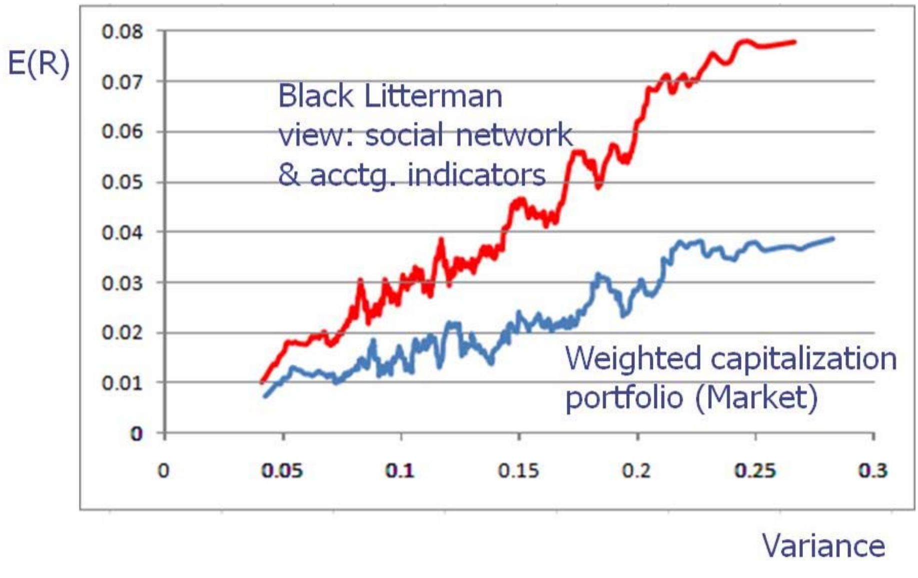

Fig. 3. Abnormal return and risk by portfolio type

Graph 3 indicates that the inclusion of social network and accounting indicators (red line) generates a portfolio with a higher level of accumulated expected return than a portfolio that uses the current market capitalization (blue line) as input. The inclusion of corporate social network indicators might capture interactions among directors and financial analysts that improve the prediction of earnings surprise. This effect, combined with the predictive capacity of selected accounting indicators, explains why a portfolio with a social network perspective outperforms the market portfolio.

5 Conclusions

This paper shows that a modified BL model that includes a forecast based on social networks and fundamental indicators as investors' view outperforms the market portfolio. Even though the BL model includes the investors' subjective views, these views can be substituted or enriched by forecasts based on the optimal combination of social networks and accounting indicators.

Acknowledgment. The author would like to thank Ionut Florescu, Maria Christina Mariani, Frederi G. Viens and participants of the Financial Mathematics session of the American Mathematical Society meeting 2010, the Workshop on Information in Networks (WIN)-NYU 2010, the IEEE CEC2011 Workshop on Agent-Based Economics and Finance, the Sixth Rutgers-Stevens Workshop Optimization of Stochastic Systems 2011, and the Eastern Economics Association meeting 2012 for their valuable comments. The opinions presented are the exclusive responsibility of the author.

References

[1] Adamic, L., Brunetti, C., Harris, J., Kirilenko, A.: Information flow in trading networks. Paper presented at 1st Workshop on Information in Networks, NYU, NY (2009)
[2] Beach, S., Orlov, A.: An application of the Black-Litterman model with EGARCH-M derived views for international portfolio management. Financial Markets Portfolio Management 21, 147–166 (2007)
[3] Becker, F., Gurtler, M.: Quantitative forecast model for the application of the Black-Litterman approach. In: 2009 Finance International Meeting AFFI - EUROFIDAI (December 2010)
[4] Black, F., Litterman, R.: Asset allocation: Combining investor views with market equilibrium. Fixed Income Research. Goldman Sachs & Co. (September 1990)
[5] Black, F., Litterman, R.: Global portfolio optimization. Financial Analysts Journal 48(5), 28–43 (1992)
[6] Black, F., Litterman, R.: The intuition behind Black-Litterman model portfolios. Investment Management Research. Goldman Sachs & Co. (December 1999)
[7] Borgatti, S.P., Everett, M.: A graph-theoretic perspective on centrality. Social Networks 28(4), 466–484 (2006)
[8] Collins, M., Schapire, R.E., Singer, Y.: Logistic regression, AdaBoost and Bregman distances. Machine Learning 48(1-3), 253–285 (2004)

[9] Creamer, G., Freund, Y.: Predicting performance and quantifying corporate governance risk for Latin American ADRs and banks. In: Proceedings of the Second IASTED International Conference Financial Engineering and Applications, pp. 91–101. Acta Press, Cambridge (2004)
[10] Creamer, G., Freund, Y.: Using AdaBoost for equity investment scorecards. In: Machine Learning in Finance Workshop in NIPS 2005, Whistler, B.C. (2005)
[11] Creamer, G., Freund, Y.: A boosting approach for automated trading. The Journal of Trading 2(3), 84–95 (2007)
[12] Creamer, G., Stolfo, S.: A link mining algorithm for earnings forecast and trading. Data Mining and Knowledge Discovery 18(3), 419–445 (2009)
[13] de Nooy, W., Mrvar, A., Batagelj, V.: Exploratory social network analysis with Pajek. Cambridge University Press, New York (2005)
[14] Dhar, V., Chou, D.: A comparison of nonlinear methods for predicting earnings surprises and returns. IEEE Transactions on Neural Networks 12(4), 907–921 (2001)
[15] Domingos, P., Richardson, M.: Mining the network value of customers. In: KDD 2001: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 57–66. ACM, New York (2001)
[16] Fawcett, T., Provost, F.: Activity monitoring: Noticing interesting changes in behavior. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 1999), San Diego, CA, USA, pp. 53–62. ACM, New York (1999)
[17] Freeman, L.: Centrality in social networks conceptual clarification. Social Networks 1(3), 215–239 (1979)
[18] Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: Machine Learning: Proceedings of the Sixteenth International Conference, pp. 124–133. Morgan Kaufmann Publishers Inc., San Francisco (1999)
[19] Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), 119–139 (1997)
[20] Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: A statistical view of boosting. The Annals of Statistics 38(2), 337–374 (2000)
[21] Goldberg, H.G., Kirkland, J.D., Lee, D., Shyr, P., Thakker, D.: The NASD securities observation, news analysis and regulation system (SONAR). In: IAAI 2003, Acapulco, Mexico (2003)
[22] Hill, S., Provost, F., Volinsky, C.: Network-based marketing: Identifying likely adopters via consumer networks. Statistical Science 21(2), 256–276 (2006)
[23] Idzorek, T.: A step-by-step guide to the Black-Litterman model. Working paper (January 2009)
[24] Jegadeesh, N., Kim, J., Krische, S.D., Lee, C.M.C.: Analyzing the analysts: When do recommendations add value? Journal of Finance 59(3), 1083–1124 (2004)
[25] Khrishnan, H., Mains, N.: The two-factor Black-Litterman model. Risk, 69–73 (2005)
[26] Kirkland, J.D., Senator, T.E., Hayden, J.J., Dybala, T., Goldberg, H.G., Shyr, P.: The NASD Regulation Advanced Detection System (ads). AI Magazine 20(1), 55–67 (1999)
[27] Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. In: EC 2006: Proceedings of the 7th ACM Conference on Electronic Commerce, pp. 228–237. ACM, New York (2006)
[28] Markowitz, H.: Portfolio selection. The Journal of Finance 7(1), 77–91 (1952)

[29] Markowitz, H.: Portfolio Selection: Efficient Diversification of Investments. John Wiley & Sons, New York (1959)
[30] Richardson, M., Domingos, P.: Markov logic networks. Machine Learning 62(1-2), 107–136 (2006)
[31] Senator, T.E.: Link mining applications: Progress and challenges. SIGKDD Explorations 7(2), 76–83 (2005)
[32] Sparrow, M.K.: The application of network analysis to criminal intelligence: An assessment of the prospects. Social Networks 13(3), 251–274 (1991)
[33] Walters, J.: The Black-Litterman model in detail. Working paper (February 2009)
[34] Watts, D., Strogatz, S.: Collective dynamics of small world networks. Nature 393, 440–442 (1998)

Appendix 1. Investment Signals Used for Prediction

I do not include firm-specific subscripts in order to clarify the presentation. Subscript q refers to the most recent quarter for which an earnings announcement was made. The fundamental variables are calculated using the information of the previous quarter (SUE,SG,TA,and CAPEX). My notation is similar to the notation used by [24].

Variable	**Description**	**Calculation detail**
SECTOR	Two-digit sector classification according to the Global Industrial Classification Standards (GICS) code.	Energy 10, Materials 15, Industrials 20, Consumer Discretionary 25, Consumer Staples 30, Health Care 35, Financials 40, Information Technology 45 Telecommunication Services 50, Utilities 55
Price momentum:		
CAR1	Cumulative abnormal return for the preceding six months since the earnings announcement day	$[\Pi_{t=m-6}^{m-1}(1+R_t)-1]-[\Pi_{t=m-6}^{m-1}(1+R_{tw})-1]$, where R_t is return in month t, R_{tw} is value weighted market return in month t, and m is last month of quarter
CAR2	Cumulative abnormal return for the second preceding six months since the earnings announcement day	$[\Pi_{t=m-12}^{m-7}(1+R_t)-1]-[\Pi_{t=m-6}^{m-1}(1+R_{tw})-1]$
Analysts variables:		
ANFOR (ANFORLAG)	Number of analysts predicting that earnings surprise increase (lagged value)	
CONSENSUS	Mean of earnings estimate by financial analysts	
FELAG	Lagged forecast error	$\frac{\text{CONSENSUS}_q-\text{EPS}_q}{\lvert\text{CONSENSUS}_q\rvert+\lvert\text{EPS}_q\rvert}$ [14] where EPS is earnings per share
Earnings momentum:		
FREV	Analysts earnings forecast revisions to price	$\sum_{i=0}^{5}\frac{\text{CONSENSUS}_{m-i}-\text{CONSENSUS}_{m-i-1}}{P_{m-i-1}}$ where P_{m-1} is price at end of month $m-1$, and i refers to the previous earnings revisions

SUE	Standardized unexpected earnings	$\frac{(EPS_q - EPS_{q-4})}{\sigma_t}$ where EPS is earnings per share, and σ_t is standard deviation of EPS for previous seven quarters
Growth indicators:		
LTG	Mean of analysts' long-term growth forecast	
SG	Sales growth	$\frac{\sum_{t=0}^{3} \text{Sales}_{q-t}}{\sum_{t=0}^{3} \text{Sales}_{q-4-t}}$
Firm size:		
SIZE	Market cap (natural log)	$ln(P_q\ shares_q)$ where $shares_q$ are outstanding shares at end of quarter q
Fundamentals:		
TA	Total accruals to total assets	$\frac{\triangle \text{C.As.}_q - \triangle Cash_q - (\triangle \text{C.Lb.}_q - \triangle \text{C.Lb.D}_q) - \triangle \text{T}_q - \text{D\&A}_q}{\frac{(\text{T.As.}_q - \text{T.As.}_{q-4})}{2}}$ where $\triangle\ X_q = X_q - X_{q-1}$ and C.As., C.Lb.,C.Lb.D.,T,D&A,and T.As. stands for current assets, current liabilities, debt in current liabilities, deferred taxes, depreciation and amortization, and total assets respectively.
CAPEX	Rolling sum of capital expenditures to total assets	$\frac{\sum_{t=0}^{3} \text{capital expenditures}_{q-t}}{(T.As._q - T.As._{q-4})/2}$
Valuation multiples:		
BP	Book to price ratio	$\frac{\text{book value of common equity}_q}{\text{market cap}_q}$, where market $\text{cap}_q = P_q\ \text{shares}_q$
EP	Earnings to price ratio (rolling sum of EPS of the previous four quarters deflated by prices)	$\frac{\sum_{t=0}^{3} EPS_{q-t}}{P_q}$
Social networks:		
$deg(v_i)$	Degree centrality or degree: number of edges incidents in vertex v_i	$\sum_j a_{ij}$, where a_{ij} is an element of the adjacent matrix A
$C_c(v_i)$	Closeness centrality (normalized): inverse of the average geodesic distance from vertex v_i to all other vertices	$\frac{n-1}{\sum_j d_{ij}}$, where d_{ij} is an element of the geodesic distance matrix D [17, 7]
$B_c(v_i)$	Betweenness centrality: proportion of all geodesic distances of all other vertices that include vertex v_i	$\sum_i \sum_j \frac{g_{kij}}{g_{kj}}$, where g_{kij} is the number of geodesic paths between vertices k and j that include vertex i, and g_{kj} is the number of geodesic paths between k and j [17]
CC_i	Clustering coefficient: cliquishness of a particular neighborhood or the proportion of edges between vertices in the neighborhood of v_i divided by the number of edges that could exist between them [34]	$\frac{2\lvert\{e_{ij}\}\rvert}{deg(v_i)(deg(v_i)-1)} : v_j \in N_i,\ e_{ij} \in E$, where each vertex v_i has a neighborhood N defined by its immediately connected neighbors: $N_i = \{v_j\} :\ e_{ij} \in E$.
CC'_i	Normalized clustering coefficient	$\frac{deg(v_i)}{MaxDeg} CC_i$, where MaxDeg is the maximum degree of vertex in a network [13]
C (not used for forecasting)	Mean of all the clustering coefficients	$\frac{1}{n} \sum_{i=1}^{n} CC_i$
SW (not used for forecasting)	"Small world" ratio [34].	$\frac{C}{L} \frac{L_{random}}{C_{random}}$, where $L_{random} \approx \frac{ln(n)}{ln(k)}$ and $C_{random} \approx \frac{k}{n}$
		Labels:
LABELFE	Label of forecast error (FE)	1 if $CONSENSUS \geq EPS$ (current quarter) , -1 otherwise
LABELCAR	Label of cumulative abnormal return (CAR)	1 if $CAR_{m+1} \geq 0$, -1 otherwise, where CAR_{m+1} refers to the CAR of the month that follows the earnings announcement