Corporate Activities and the Market Risk Premium *

Erik Lie¹, Bo Meng¹, Yiming Qian¹, and Guofu Zhou²

¹Department of Finance, University of Iowa
²Olin Business School, Washington University in St. Louis

May 24, 2017

Abstract

While existing asset pricing studies focus on macroeconomic variables to predict stock market risk premium, we find that an aggregate index of corporate activities has substantially greater predictive power both in- and out-of sample, and yields much greater economic gain for a mean-variance investor. The predictive ability of the corporate index stems from its information content about future cash flows. Cross-sectionally, the corporate index performs particularly well for stocks with great information asymmetry.

Keywords: Predictability, Corporate Activities, Information Asymmetry, Economic Value.

JEL classification: G10, E44, G30, G11, G12, G15.

*We are grateful to Radhakrishnan Gopalan, Amit Goyal, Raymond Kan, Dave Rapach, Matthew Ringgenberg, Ashish Tiwari, Anand Vijh, Tong Yao, and seminar participants at University of Iowa and Washington University in St. Louis for very helpful comments. Send correspondence to Bo Meng, Henry B. Tippie College of Business, University of Iowa, Iowa City, IA 52242; email: bo-meng@uiowa.edu; phone: 319-335-0926.
1 Introduction

As emphasized by Cochrane (2008), understanding the variation in the market risk-premium has important implications in all areas of finance. Numerous studies have shown that various macroeconomic variables, such as valuation ratios, interest rates, and interest rate spreads, predict the market (see, e.g., Fama and French, 1988; Campbell and Shiller, 1988b; Fama and Schwert, 1977; and also Cochrane, 2011 and Rapach and Zhou, 2013 for recent surveys). This literature, however, pays little attention to corporate activities. But corporate activities embed important information, because corporate executives possess insider information about firm prospects that affect their decisions. For example, executives who have information that their firms are overvalued are inclined to issue equity. Indeed, the corporate finance literature finds that many corporate activities are followed by abnormal stock returns for individual stocks. Studies have documented abnormal stock returns after insider trading (Seyhun, 1986; Lakonishok and Lee, 2001; Jeng, Metrick, and Zeckhauser, 2003; Ravina and Sapienza, 2010; Alldredge and Cicero, 2015; Cohen, Malloy, and Pomorski, 2012), equity issues (Loughran and Ritter, 1995; Spiess and Affleck-Graves, 1995; Brav and Gompers, 1997; Lee, 1997; DeAngelo, DeAngelo, and Stulz, 2010), share repurchase announcements (Ikenberry, Lakonishok, and Vermaelen, 1995; Dittmar and Field, 2015), and merger announcements (Loughran and Vijh, 1997; Savor and Lu, 2009).

In this paper, we construct a comprehensive index of corporate activities and use that to predict the stock market return. We consider five major categories of corporate or managerial activities: aggregate security issues, share repurchases, corporate investments, merger activity and payments, and insider trading. Using standard measures from corporate studies, we obtain 13 corporate predictors of the stock market. To form an aggregate index that summarizes the 13 corporate predictors, we use the partial least squares (PLS) approach (Wold 1966, 1975; and Kelly and Pruitt 2013, 2015). As stressed by Huang et al. (2015), the advantage of the PLS is that it extracts only the relevant aggregate information, without retaining the common noise as the principal-component approach does.

Using quarterly data from 1986 to 2015, we find that our corporate index explains 8.5% of subsequent market returns. In comparison, a similarly constructed macroeconomic index
based on variables that have been used in past studies explains only 1.9% of market returns, while an investor sentiment index based on Baker and Wurgler (2006) explains 5.2% of market returns. We further find that the corporate variables have substantial explanatory power beyond the macroeconomic and sentiment variables. In particular, an index based on a combination of macroeconomic and sentiment variables explains 3.8% of market returns, and adding the corporate variables to the index increases this by 8.4% to a total of 12.2%. Interestingly, while macroeconomic index has little forecasting power during expansions ($R^2 = 0.8\%$), which is consistent with existing studies, our corporate index remains useful even then ($R^2 = 2.5\%$).

The superior explanatory power of the corporate index persists out of the sample. Using an initial training period of years 1986 through 1999, the out-of-sample $R^2$ for the corporate index is 12.5%. In comparison, the out-of-sample $R^2$ for the macroeconomic and sentiment indices are $-0.8\%$ and 7.12\%, respectively. We also compute the certainty equivalent gain (CER) for a mean-variance investor who employs the corporate index to optimally allocate assets between the risk-free rate and the market. Assuming a risk aversion coefficient of three, the CER gain is 8.8\%, which is considerably higher than those of the macro and sentiment indices (0.6\% and 4.1\%, respectively).

Consistent with the notion that the predictive power of the corporate index is rooted in information asymmetry between corporate executives and the general public, we find that the corporate activity index has greater explanatory power among opaque firms than among transparent firms. Using five measures of information asymmetry, including PP&E scaled by market value of the firm, analyst forecast errors, analyst forecast dispersion, size (market capitalization) and book-to-market ratio, we find that the corporate index predicts future returns better for opaque firms. These results corroborate an interpretation that executives use information that is both useful in predicting future prospects and not generally available to the public when making decisions.

Finally, we examine the economic driving source of the predictability. The predictability of returns could stem from the predictability of either cash flow or the discount rate. We use Campbell and Shiller (1988a)'s log linearization of stock returns to determine the source of predictability for the corporate index. The results strongly point to the cash flow channel. That is, the corporate index predicts aggregate earnings growth (a common proxy for cash
flow), but there is no evidence that the corporate index predicts the dividend-price ratio (a
discount rate proxy). This is consistent with the notion that corporate executives mainly
have private information about their own firms’ future profitabilities (and collectively about
aggregate cash flows), but not about the discount rate, which depends on macro conditions,
investor risk attitude or investor sentiment.

Our study builds upon the insights of numerous corporate studies that document the
impact of corporate activities on stock returns at the individual stock level, some of which
are discussed earlier. In contrast to these studies, we focus on predicting the aggregate stock
market returns. There are a few corporate studies that consider the aggregate market returns.
Seyhun (1988, 1992) and Lakonishok and Lee (2001) document the predictive ability of insider
trading; Baker and Wurgler (2000) document the predictive ability of equity issues; and Arif
and Lee (2014) document the predictive ability of the corporate investments. However, there
are several important differences between these studies and ours. First, whereas they each
focus on one type of corporate activities, we analyze the systematic effect of corporate activities
as captured by our corporate index. As such, our approach should better capture inside
information, both favorable and unfavorable, embedded in the array of executive decisions.
Second, none of the existing studies examine the out-of-sample predictive power, whereas we
document strong results both in- and out-of-sample. Third, unlike past studies, we evaluate
the economic value of corporate activities on predicting the market. Fourth, we explore the
role of asymmetric information in the predictive ability of corporate activities, which sheds
light on the economic driving force of the market return predictability.

Overall, the predictive power of our corporate activity index is strong and exceeds those of
macroeconomic and sentiment variables and their indices substantially. Nonetheless, existing
asset pricing models, such as the well known habit formation model (Campbell and Cochrane,
1999), the long-run risks (LRR) model (Bansal and Yaron, 2004), the rare disaster model
(Barro, 2006; Wachter, 2013) and their various recent extensions, are solely based on macroeco-
nomic variables.\footnote{Cochrane (1991) is an important exception who introduce a production-based model. However, there is
no separation between investors and producers in the model.} Ignoring the information contained in corporate activities clearly impedes
the ability of the asset pricing models in explaining asset returns. For example, these models are typically rejected from predictability tests (see, e.g., Ross, 2009; Huang and Zhou, 2016). Our empirical results demonstrate a role of corporate activities in asset pricing theory. Indeed, it appears that an important research direction of asset pricing is to combine corporate theory with existing models.

2 Data and Variables

We identify a set of managerial decisions and corresponding events that depend on managers’ beliefs about firms’ prospects and/or firm misvaluation. While our set might be incomplete, it captures the bulk of managerial and firm activities for which data are available.

We first consider mergers and acquisitions. Loughran and Vijh (1997) and Rau and Vermaelen (1998) find that mergers and acquisitions using cash as the method of payment experience positive long-run abnormal returns, whereas those using stock as the payment method experience negative long-run abnormal returns. This suggests that firms tend to use stock as the currency for acquisition when they believe the securities are overvalued. Aggregating information about individual firms should yield useful direction about the overall stock market. That is, if many executives have private information that their firms are overvalued, then the aggregate stock market is likely overvalued too. Thus, we hypothesize that when the aggregate stock amount used in acquisitions is high, the subsequent market return is low.

We obtain the sample of mergers from the Securities Data Company’s (SDC) U.S. Mergers and Acquisitions Database. We start with all domestic mergers and acquisitions with announcement dates between 1986 and 2015. We consider all completed mergers in which a public firm (bidder) acquires another public firm (target) using 100% common stock or 100% cash. We include deals with the following SDC merger codes: “Merger”, “Acq. of Asset”, or “Acq. Maj. Int.” (following, e.g., Vijh and Yang, 2013).

We use quarterly observations for all our corporate variables to be consistent with the quarterly reports for accounting data. For each quarter from 1986Q1 to 2015Q4, we construct

---

2 We use this time period to be consistent with the time period of Thomson Reuters insider filings database, which spans from 1986Q1 to 2015Q4.
two variables for the use of stock as the method of payment in mergers and acquisitions:

- **Percentage of stock payment**, COMPCT: the aggregate amount of stock payment divided by the sum of the aggregate amount of stock payment and cash payment (in percentage points);

- **Total stock payment (log)**, COM: the natural log of the aggregate amount of stock payment (the dollar amounts, in millions, are deflated to 1986 dollar).

Next, we consider insider trading. Possessing private information, corporate insiders have incentives to buy the stock of their own companies if they believe the stock prices will increase in the future and sell if they believe the stock prices will decrease. Consistent with this, prior studies document that corporate insiders’ net purchases of their own companies stocks are followed by positive abnormal returns at the individual stock level (e.g., Seyhun, 1986; Lakonishok and Lee, 2001; Cohen, Malloy, and Pomorski, 2012). Seyhun (1988, 1992) and Lakonishok and Lee (2001) also document evidence of positive relationship between aggregate insider trading and subsequent market returns. We expect to see aggregate net purchases positively predict market returns.

We obtain insider trading data from Thomson Reuters’ insider filings database. Corporate insiders are required to report their open market trades to the SEC within 10 days after the end of month in which these trades took place, as required by the Section 16a of the Securities and Exchange Act of 1934. In 2002, this reporting deadline was reduced to two days after the trades.

Form 4 of the SEC filing contains the main data set for insider trading, including information about each insider transaction and the insider’s position in the firm. Following the literature, we define corporate insiders as officers, managers, and beneficial owners of more than 10% of a company’s stock (e.g., Cohen, Malloy, and Pomorski, 2012). We include insider transactions with a cleanse code of R, H, L, C, or Y in Thomson Reuters’ database (e.g., Alldredge and Cicero, 2015). Also following the literature (e.g., Seyhun, 1988), we consider

---

*A cleanse code of R indicates “data verified through the cleansing process”, H indicates “cleansed with a very high level of confidence”, L indicates “Cleansed”, C indicates “a record added to nonderivative table or derivative table to correspond with a record on the opposing table”, and Y indicates “Informational.”*
only open market purchases (recorded as ‘P’) and open market sales (recorded as ‘S’). Sales
include the sale of stocks immediately after option exercising. That is, if a corporate insider
decides to exercise options and sell the stocks immediately, it is potentially a negative signal
of her view of future prospects. However, it is less of a signal when the option is approaching
the expiration date. In this case, corporate insiders exercise their options prior to that date
as long as the options are in the money, and the immediate open market sales after the option
exercise are less likely to contain private information about the stock value. Therefore, we
exclude open market sales associated with option exercises that occur within six months of
the option expiration date.\footnote{The results are unaffected if we reduce the number of months from six to one.}

According to Jeng, Metrick, and Zeckhauser (2003), after May 1991, private transactions
have the same codes (transaction codes P and S) as open market transactions. Private trans-
actions might take place with restricted securities and are more likely to be executed for
liquidity or diversification motives than are open-market transactions. Following Jeng et al.
(2003), we identify private transactions as those in which the number of shares traded exceed
the daily trading volume or have prices that fall outside the daily trading range on the open
market as recorded by CRSP. We then exclude these transactions.

Similar to Seyhun (1988) and Lakonishok and Lee (2001), we define four predictors based
on insider trading as follows:

- \( \hat{\text{Net Transactions}} \), \( \text{NT} \): the aggregate number of open market purchases minus the ag-
gregate number of open market sales (in thousands);

- \( \hat{\text{Net Dollar Amount}} \), \( \text{NDA} \): the aggregate amount of open market purchases minus the
aggregate amount of open market sales (the dollar amounts, in billions, are deflated to
1986 dollars);

- \( \hat{\text{Ratio of Net Purchases}} \), \( \text{RT} \): the aggregate number of open market purchases divided by
the sum of the aggregate number of open market purchases and the aggregate number
of open market sales (in percentage points);

- \( \hat{\text{Ratio of Net Purchasing Dollar Amount}} \), \( \text{RDA} \): the aggregate amount of open market

purchases divided by the sum of the aggregate amount of open market purchases and the aggregate amount of open market sales (in percentage points).

The third type of corporate activities we consider is corporate investment. There are two hypotheses regarding the relationship between corporate investments and stock returns, and these hypotheses have opposite predictions, yet are not mutually exclusive. One hypothesis is that managers invest when they are optimistic about future prospects. Therefore, investments are likely to precede higher future profitability and, assuming slow incorporation of information, higher stock returns. Consistent with this hypothesis, Chan, Lakonishok, and Sougiannis (2001) and Eberhart, Maxwell, and Siddique (2004) document that high R&D level or growth, standardized by market capitalization, is associated with high stock returns. The other hypothesis is that firms invest more when investor sentiment is high. Consistent with this hypothesis, Titman, Wei, and Xie (2004) find a negative cross-sectional relationship between abnormal capital expenditure and stock returns. Furthermore, Arif and Lee (2014) find that high aggregate corporate investments standardized by book value of assets are associated with low market returns. Based on existing studies (e.g., Chan, Lakonishok, and Sougiannis, 2001; Eberhart, Maxwell, and Siddique, 2004), investments standardized by market capitalization might be a better measure for inside information. That is, high investment during low equity valuation is more likely due to inside information about future prospects than to investor sentiment, and vice versa. We therefore expect to see that aggregate investments standardized by market capitalization positively predict market returns, whereas aggregate investments standardized by book value of assets negatively predict market returns.

To construct the predictors based on corporate investments, we use quarterly U.S. financial statements data from Compustat and stock market data from CRSP. We exclude firms with SIC codes between 6000 and 6999 (financial firms). Similar to Arif and Lee (2014), we limit our sample firms to those with fiscal years ending in December such that all financial reports are released at roughly the same time of the year. We also require that the sample firms be traded on NYSE, AMEX, or NASDAQ with share codes 10 or 11 (common stock).

We construct two investment-based predictors based on the most direct measure of investments—capital expenditures.
- **CAPX scaled by ME**, CAPXME: aggregate capital expenditures scaled by total market capitalization (in percentage points);

- **CAPX scaled by AT**, CAPXAT: aggregate capital expenditures scaled by average total assets (in percentage points).

If we add R&D expenditures to capital expenditures, the results are similar. We do not tabulate results with the alternative measures, because R&D data in Quarterly Compustat starts in year 1989, thus reducing our time-series observations.

Similar to Arif and Lee (2014), we also measure investments as the change in net operating assets. We construct two additional predictors for aggregate corporate investment:

- **Change in net operating asset scaled by ME**, ALME: The change in net operating asset plus R&D scaled by total market capitalization (in percentage points);

- **Change in net operating asset scaled by AT**, ALAT: The change in net operating asset plus R&D scaled by average total assets (in percentage points).

The fourth type of corporate activity is equity issuance. In a world with asymmetric information (i.e., when insiders know more about the firm prospects and firm value than outsiders), firms issue more equity when their stocks are overvalued. If investors are not rational and sufficiently sophisticated to immediately adjust the stock prices, these stocks underperform in the long run. Numerous studies document that individual stocks tend to underperform following equity issuance (e.g., see Ritter, 1991; Loughran and Ritter, 1995; Spiess and Affleck-Graves, 1995; Brav and Gompers, 1997; and Jegadeesh, 2000). Baker and Wurgler (2000) show that aggregate equity issuance are positively related to future market returns. Hence, we use the aggregate level of equity issuance as another predictor of market returns and expect it to negatively predict market returns.

5Similar as Arif and Lee (2014), we construct the measure as $ALAT = \frac{\Delta NOA_{i,t} + R&D_{i,t}}{\frac{1}{2}(TA_{i,t-1} + TA_{i,t})}$, where NOA is defined as in Dechow, Richardson, and Sloan (2008): total assets (Compustat AT) less cash and short-term investments (Compustat CHE) minus non-debt liabilities; where non-debt liabilities equals total liabilities (Compustat LT) plus minority interest (MIB) less debt (Compustat DLTT plus Compustat DLC).
We obtain the equity issuance data from Jeffrey Wurgler’s website up to 2007 (http://people.stern.nyu.edu/jwurgler/), and the recent issuance data from the Federal Reserve website (http://www.federalreserve.gov/econresdata/releases/corpsecure/current.htm). Similar to Baker and Wurgler (2000), we construct the following two predictors based on equity issuances:

- **Total Equity Issuance (log), E**: the natural log of equity issuance (the dollar amounts, in millions, are deflated to 1986 dollar);

- **Ratio of Equity Issuance, S**: equity issuance scaled by the sum of equity and debt issuance (in percentage points).

The last corporate predictor is based on aggregate share repurchases. Share repurchases are the flip side of equity issues. When corporate insiders have private favorable information about the firm value, they have incentives to repurchase shares. Consistent with this logic, extant studies document firm-level evidence that repurchases are followed by high stock return (e.g., Ikenberry, Lakonishok, and Vermaelen, 1995). If many firms undertake repurchases simultaneously, the overall stock market is likely undervalued. We therefore expect that aggregate repurchases positively predict market returns.

We first calculate net repurchases for each firm, and then aggregate them. Following Fama and French (2001), we measure net repurchases as the increase in common Treasury stock if Treasury stock is not zero or missing. If Treasury stock is zero in the current and prior quarter, we measure repurchases as the difference between stock purchases and stock issuances from the statement of cash flows. If either of these estimates are negative, repurchases are set to zero. The share repurchase predictor is defined as follows:

- **Aggregate share repurchases (log), REP**: The natural log of aggregate share repurchases (in millions of 1986 dollar);

As noted earlier, each of the 13 corporate predictors spans from 1986Q1 to 2015Q4. We examine the ability of these predictors in predicting excess market returns. In addition, we compare the predictive power of the corporate variables with well-examined macroeconomic variables and investor sentiment variables. Following Welch and Goyal (2008), we use 14
well-recognized macroeconomic variables. We obtain the quarterly excess market return and the macroeconomic variables from Goyal’s web site (http://www.hec.unil.ch/agoyal/).

- **Dividend-price ratio (log), D/P**: The difference between the log of dividends paid on the S&P 500 index and the log of stock prices (S&P 500 index), where dividends are measured using a one-year moving sum;

- **Dividend yield (log), D/Y**: The difference between the log of dividends and the log of lagged stock prices;

- **Earnings-price ratio (log), E/P**: The difference between the log of earnings on the S&P 500 index and the log of stock prices, where earnings are measured using a one-year moving sum;

- **Dividend-payout ratio (log), D/E**: The difference between the log of dividends and the log of earnings;

- **Stock variance, SVAR**: The sum of squared daily returns on the S&P 500 index;

- **Book-to-market ratio, B/M**: The ratio of book value to market value for the Dow Jones Industrial Average;

- **Net equity expansion, NTIS**: The ratio of the twelve-month moving sum of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks;

- **Treasury bill rate, TBL**: The interest rate on a three-month Treasury bill (in the secondary market);

- **Long-term yield, LTY**: The long-term government bond yield;

- **Long-term return, LTR**: The return on long-term government bonds;

- **Term spread, TMS**: The difference between the long-term yield and the Treasury bill rate;

- **Default yield spread, DFY**: The difference between BAA- and AAA-rated corporate bond yields;
- Default return spread, DFR: The difference between long-term corporate bond and long-term government bond returns; and

- Inflation, INFL: Calculated from the CPI (all urban consumers); following Welch and Goyal (2008) and Rapach, Strauss, and Zhou (2010), since inflation rate data are released in the following month, we use $x_{i,t-1}$ in Equation 5 for inflation.

To measure investor sentiment, we follow Baker and Wurgler (2006, 2007). We obtain the following investor sentiment proxies from Jeffrey Wurgler’s web site http://people.stern.nyu.edu/jwurgler/.

- Close-end fund discount rate, CEFD: the value-weighted average difference between the net asset values of closed-end stock mutual fund shares and their market prices;

- Number of IPOs, NIPO: the quarterly number of initial public offerings;

- First-day returns of IPOs, RIPO: quarterly average first-day returns of initial public offerings;

- Dividend premium, PDND: The log difference of the value-weighted average market-to-book ratios of dividend payers and nonpayers; and

- Equity share in new issues, EQTI: the gross quarterly equity issuance divided by the gross quarterly equity plus debt issuance;

The data on these measures are available at monthly frequency, spanning from July 1965 through September 2015 (603 months; 201 quarters). We first convert these measures into quarterly frequency. Specifically, we use the quarter-end close-end fund discount rate (CEFD) and dividend premium (PDND). For the number of IPOs (NIPO), we calculate the sum of NIPO over the three months of each quarter. For first-day returns of IPOs (RIPO) and equity share in new issues (EQTI), we calculate the simple arithmetic averages across the three months of each quarter.

---

6Baker and Wurgler (2006, 2007) use a sixth sentiment variable, NYSE turnover, which is currently excluded from the data they provide due to institutional changes and high-frequency trading.
Following Baker and Wurgler (2006, 2007), each individual measure is first standardized, then regressed on a set of variables reflecting the fundamentals: the growth of industrial production, the growth of durable consumption, the growth of nondurable consumption, the growth of service consumption, the growth of employment, and a dummy variable for NBER-dated recessions (to remove the effect of business-cycle variation). We then use the two-quarter moving average of the regression residual to iron out idiosyncratic jumps in the individual sentiment measures. The average first-day return of IPOs and dividend premium are lagged by four quarters relative to the other three measures, because these two variables likely take more time to reveal the same sentiment.

3 Econometric methods

The partial least squares (PLS) method was developed by Wold (1966, 1975) and used in Kelly and Pruitt (2013, 2015). In this section, we outline how we use this method to construct the corporate index as an aggregation of corporate insiders’ private information.

When predicting market return using a large number of variables, the PLS technique is superior to other commonly used statistical techniques, such as the principle component (PC) technique, because all predictors might have approximation errors to the true but unobservable predictor (e.g., the investor sentiment in Baker and Wurgler, 2006, 2007). If so, the errors are parts of their variations and the first PC potentially contains a substantial amount of common approximation errors that are not relevant for forecasting returns. Huang et al. (2015) have shown that PLS successfully extracts information in sentiment proxies that is relevant to the expected stock returns from the error or noise. By the same token, we use PLS technique to extract useful information from corporate activities.

We apply the standard predictive regression model by assuming that a corporate factor explains future stock returns as described in the following linear relationship:

\[ E_t(R_{t+1}) = \alpha + \beta C_t, \]  

where \( C_t \) is the true but unobservable corporate factor that represents corporate insiders’ private information about asset valuation and therefore is relevant for forecasting future returns.
The realized stock return is then equal to

\[ R_{t+1} = E_t(R_{t+1}) + \epsilon_{t+1} \]
\[ = \alpha + \beta C_t + \epsilon_{t+1}, \]  

where \( \epsilon_{t+1} \) is unforecastable and unrelated to \( C_t \).

Let \( x_t = (x_{1,t}, ..., x_{N,t})' \) be a \( N \times 1 \) vector of individual predictors of interest. For our purpose here, they refer to the 13 corporate predictors. We assume that the true corporate factor is not directly observable. However, each proxy predictor \( x_{i,t} (i = 1, ..., N) \) has information on it, obeying a factor structure,

\[ x_{i,t} = \eta_{i,0} + \eta_{i,1} C_t + \eta_{i,2} E_t + \epsilon_{i,t}, \]

where \( C_t \) is the corporate factor; \( \eta_{i,1} \) is the slope coefficient that measures the sensitivity of the individual proxy predictor \( x_{i,t} \) to the movement of the corporate factor; \( E_t \) is the common approximation error component of all the proxies that is irrelevant to returns; and \( \epsilon_{i,t} \) is the disturbance term. As noted in Huang et al. (2015), the PLS approach effectively extracts \( C_t \), while filtering out the irrelevant component \( E_t \).

We follow the PLS method as described in Huang et al. (2015) to estimate the corporate factor. Henceforth, we denote the estimated corporate factor, or the corporate index, as \( PLSC \), which is known as a linear combination of \( x_{i,t} \) computed from

\[ PLSC = XJ_NX'J_TR(R'J_TXJ_NX'J_TR)^{-1}R'J_TR, \]  

where \( X = (x_1', ..., x_T') \) denotes the \( T \times N \) matrix of individual predictors, and \( R \) denotes the \( T \times 1 \) vector of excess stock returns as \( (R_2, ..., R_{T+1})' \). The matrices \( J_T = I_T - \frac{1}{T}1_T1_T' \) and \( J_N = I_N - \frac{1}{N}1_N1_N' \) enter the formula because each regression is run with a constant. \( I_T \) is a T-dimensional identity matrix and \( 1_T \) is a T-vector of ones.

Intuitively, PLS extract the corporate factor from the cross-section by choosing a combination of the individual corporate predictors that is optimal for forecasting. The weight on each individual \( x_i \) is based on its covariance with future stock returns to capture the intertemporal relationship between the corporate index and the expected future stock return. As noted earlier, we, following existing practice, standardize all 13 corporate variables. We
also standardize the PLS index so that the index has a mean of zero and a standard deviation of one.

For comparison, we also use the PLS approach for the commonly used macroeconomic and sentiment variables. We end up with three separate PLS indices, which allow us to compare the predictive power of the corporate, macroeconomic, and sentiment variables. In one part, we even combine variables from several categories to create overarching PLS indices, which allow us to examine incremental predictive power.

4 Empirical Results

4.1 Summary Statistics

Table 1 shows summary statistics for each of the 13 corporate predictors, as well as the excess market return and the risk-free rate. While the corporate predictors are fairly standard in the corporate literature, it is interesting that they have varying volatilities, skewness and kurtosis, well bounding those of the market return. As a result, they are likely as a group to be able to explain the moments of the market. In addition, their first-order correlations are generally small. In contrast, though not tabulated here, the common macroeconomic predictors are highly persistent (many of them have over 95% first-order correlations).

4.2 Univariate predictive regressions of corporate variables

We start by considering the univariate predictive regression model for each corporate variable:

\[ R_{t+1}^m = \alpha_i + \beta_i x_{i,t} + \epsilon_{i,t+1}, \]

where \( R_{t+1}^m \) is the excess market return (i.e., the log of one plus the S&P 500 return, in excess of log of one plus the risk-free rate), \( x_{i,t} \) is one of the 13 corporate variables whose predictive ability is tested, and \( \epsilon_{i,t+1} \) is the idiosyncratic noise.

Finance theory suggests a prior on the sign of \( \beta \). The discussion in Section II provides the expected sign for each corporate variable. From an econometric point of view, Inoue and
Kilian (2005) also suggest the use of the one-sided alternative hypothesis. Hence, we use the one-sided test for the univariate regression models in our paper.

Table 2 reports the results of univariate predictive regressions over the sample period from 1986Q2 to 2015Q4. The signs of the coefficients are in line with theory and past empirical results. That is, the ratio and amount of equity used as merger payment negatively predict market returns; the ratio and amount of insider net purchases positively predict market returns; aggregate investment standardized by market cap (assets) positively (negatively) predict market returns; and equity issuances negatively predict market returns. However, we do not find aggregate stock repurchases to be significantly related to future market returns.

Of the 13 corporate predictors, one variable has statistically significant in-sample predictive ability at the one percent level, eight variables have statistically significant in-sample predictive ability at the five percent level, and one variable exhibits predictive ability at the ten percent level. For these ten variables, the $R^2$s range from 1.79% to 4.93%. In comparison, the macroeconomic variables exhibit much weaker in-sample predictability, with only three variables showing statistically significant predictive ability at the five percent level, and one at the ten percent level. For these four regressions, $R^2$s range from 1.79% to 2.50%. The sentiment variables do not show much predictive power for quarterly returns either. Of the five variables, only two variables shows statistical significance.

For easy interpretation and comparison, all predictors are standardized, i.e., demeaned and standardized as many other studies. Hence the predictors used in the regressions all have a mean of zero and a standard deviation of one. The regression coefficients suggest that the impact of the corporate predictors are economically significant as well. For example, a one-standard-deviation increase in $COMPCT$ is associated with a 1.33% decrease in the excess market return in the next quarter. In comparison, the mean quarterly market excess return is 1.55%. Of the ten corporate variables that exhibit statistical significance, every coefficient has a value larger than one, i.e., a one-standard-deviation increase in each variable changes the quarterly excess return by more than one percent.

Our univariate regression results are consistent with existing corporate studies that examine aggregate corporate decisions’ impact on stock returns. For example, with a sample period from 1928 to 1997, Baker and Wurgler (2000) use an annual measure of ratio of equity
issuance (S) to predict one-year-ahead annual market returns. They find that a one standard deviation increase in S leads to 7.42% decrease in the value-weighted market return in the subsequent year (their Table III). This corresponds to a quarterly return of −1.86%. Using our sample period (1986Q2-2015Q4), we find that a one standard deviation increase in S leads to 1.43% decrease in excess market return in the subsequent quarter, which is comparable to the result of Baker and Wurgler (2000).

Using annual data from 1962 to 2009, Arif and Lee (2014) document a 2.19% decrease in annual market return following a one standard deviation increase in (ALAT) based on the univariate regression (their Table 2). This corresponds to a quarterly return of −1.53%. Similar to their result, we find that a one standard deviation increase in ALAT leads to a decrease in quarterly market returns of −1.29%. In addition, we find a one standard deviation increase in CAPEX scaled by ME (CAPXME) is associated with an increase of quarterly return of 1.24%, consistent with our earlier conjecture.

Consistent with the literature, we also find strong predictive power of all four measures of aggregate insider trading. Using monthly data from January 1975 to October 1981 and an aggregate insider trading measure similar to NT, Seyhun (1988) find that insider net purchase has information about market returns two months later, and that a one standard deviation increase in insider net purchase is associated with a 1.7% increase in monthly excess market returns (corresponding to 5.1% in quarterly return). With a totally different sample period, we find that a one standard deviation change in NT is associated with a 1.52% change in excess market returns one quarter later. The magnitude of the effect is smaller than that documented by Seyhun (1988), but it is still substantial. Lakonishok and Lee (2001) also examine the relation between aggregate insider trading and market returns. Using data from January 1976 to January 1995 and a measure similar to RT, they find “a spread of 11% per year in market returns [2.75% per quarter] between the month with the NPR in the top 10 percentile (0.06) and the month with the NPR in the bottom 10 percentile (-0.46).” (p. 93). Similar to their result, we find a spread of 3.36% per quarter in market returns between the quarter with RT in the top decile (0.06) and the month with RT in the bottom decile (-0.67).

In summary, with recent quarterly data, our univariate regression results on corporate predictors match the evidence found in prior literature both in direction and in magnitude.
Although a detailed comparison is omitted, the same is true for macroeconomic and sentiment variables.

4.3 Predictive power of the corporate index

Following the PLS procedure outline earlier, our in-sample time-series estimation of the corporate index is:

\[
PLS^C = -0.15 * COMPCT - 0.19 * COM + 0.20 * NT + 0.15 * NDA + 0.17 * R Ts \\
+ 0.17 * RDA + 0.16 * CAPXME + 0.07 * CAPXAT - 0.02 * ALME \\
- 0.14 * ALAT - 0.21 * E - 0.16 * S - 0.05 * REPO.
\]

The signs of the weights are consistent with what we expect based on the literature. For example, the use of stock as the payment method in M&As contain negative information about future returns, while insider net purchase of stocks contain positive information. Hence we conjecture that higher values of the corporate index predict higher future market returns.

We now examine the predictability of the corporate events in concert, that is, the predictability of the corporate index, \(PLS^C\). Table 3 reports the univariate regression results. The regression slope \(\beta\) of \(PLS^C\) is 2.37 with a t-statistic of 3.29. That is, a one-standard-deviation increase in the corporate index leads to 2.37\% increase in quarterly return. The \(R^2\) is 8.47\%, which, from the perspective of asset pricing models, implies that the PLS index has very good predictive ability. Comparing to the results in Panel A of Table 2, both the regression coefficient and the \(R^2\) of the corporate index are much larger than those of any individual corporate variable. This shows that the corporate index has stronger predictive power than that of each individual corporate variable, using the PLS index to extract common information from individual variables adds value.

We also compare the predictive power of the corporate index with the PLS index for the 14 macroeconomic variables, \(PLS^E\), and that for the 5 sentiment proxies, \(PLS^S\). The slope coefficient for the macroeconomic index, \(PLS^E\), is 1.12, slightly lower than that of the corporate index, \(PLS^C\). The \(R^2\) of \(PLS^E\) is 1.89\%, much lower than that of \(PLS^C\). The predictive power of the sentiment index \(PLS^S\) is also lower than the \(PLS^C\), with a coefficient of -1.85 and an \(R^2\) of 5.16\%. It is worth noting that \(PLS^C\) and \(PLS^S\) have a common equity
issuance component. If we exclude this component from both indices, the results remain qualitatively the same: the regression slope of $PLS^C$ remains 2.24, with an $R^2$ of 7.58%, and the slope of $PLS^S$ is -1.84, with an $R^2$ of 5.10%.

To gain further insight into the incremental predictive power of the corporate variables, we construct (i) $PLS^{EC}$ based on the combination of the macroeconomic and corporate variables, (ii) $PLS^{SC}$ based on the combination of the sentiment and corporate variables, (iii) $PLS^{ES}$ based on the combination of the macroeconomic and sentiment variables, and (iv) $PLS^{ESC}$ based on the combination of the macroeconomic, sentiment, and corporate variables. The $R^2$ for $PLS^{EC}$ is 11.58%, which is an improvement of the predictability of $PLS^C$ (with an $R^2$ of 8.47%) and a greater improvement yet of the predictability of $PLS^E$ (with an $R^2$ of 1.89%). The same can be said about $PLS^{SC}$, relative to $PLS^C$ and $PLS^S$. Combining both macroeconomic and sentiment variables, $PLS^{ES}$ has an $R^2$ of 3.82%. Lastly, if we combine all three types of variables, the $R^2$ for $PLS^{ESC}$ increases to 12.19%, which is a substantial improvement of the predictability of $PLS^{ES}$. Thus, the corporate variables significantly enhance our ability to predict excess market returns beyond the use of just the macroeconomic and sentiment variables.

We also study the predictability of market returns for periods in different stages of the business cycle. Following Rapach, Strauss, and Zhou (2010) and Huang, Jiang, Tu, and Zhou (2015), we compute the $R^2$ statistics separately for periods of economic booms ($R^2_{up}$) and economic troughs ($R^2_{down}$),

$$
R^2_c = 1 - \frac{\sum_{t=1}^{T} I^c_t (\hat{\epsilon}_{i,t})^2}{\sum_{t=1}^{T} I^c_t (R^m_t - \bar{R}^m)^2}, \quad c = \text{up, down},
$$

where $I^c_t$ is an indicator that takes a value of one when quarter $t$ is in an NBER expansion (recession) quarter and zero otherwise; $\hat{\epsilon}_{i,t}$ is difference between $R^m_t$ and the fitted value of excess market return $\hat{R}^m_t$ based on the in-sample estimates of the predictive regression model in Equation 5; $\bar{R}^m$ is the full-sample average of $R^m_t$; and $T$ is the number of quarters in our sample. While the full-sample $R^2$ is always positive, the $R^2_{up}$ and $R^2_{down}$ can be positive or negative.

Columns 5 and 6 of Table 3 report the $R^2_{up}$ and $R^2_{down}$ statistics. Consistent with the existing literature (e.g., Huang, Jiang, Tu, and Zhou, 2015), we find that the return predictability
is higher during recessions than during expansions for both $PLS^E$ and $PLS^S$. This is also true for the corporate index $PLS^C$. More importantly, $PLS^C$ exhibits strong in-sample predictive abilities during both periods. During recessions, the predictive power of $PLS^C$ is higher than those of $PLS^E$ and $PLS^S$. During expansions, its predictive power is similar to that of $PLS^S$ and higher than that of $PLS^E$. Consistent with Rapach, Strauss, and Zhou (2010), and others, the predictability measured by any of the $R^2$s is concentrated in recessions. Recently, Cujean and Hasler (2017) explain theoretically that such a concentration is caused by countercyclical investors’ disagreement.

In the last two columns of Table 3, we divide the sample into high- and low-sentiment periods according to the quarterly sentiment index $PLS^S$. Following Stambaugh, Yu, and Yuan (2012), we classify a quarter as high (low) sentiment if the sentiment level ($PLS^S$) in the previous quarter is above (below) its median value for the sample period, and compute the $R^2_{high}$ and $R^2_{low}$ statistics for the high- and low-sentiment periods, respectively, in a manner similar to Equation 7. The results show that $PLS^C$ has a higher R-squared during high-sentiment periods than in low-sentiment periods. For example, during high-sentiment periods, $PLS^C$ has an $R^2_{high}$ of 12.82. In contrast, during low-sentiment periods, $PLS^C$ has an $R^2_{low}$ of 3.91. This is consistent with the notion that insider information are more valuable and/or managers are more likely to take advantages of their private information during those periods. The sentiment index also has higher R-squared during high-sentiment periods, suggesting that investors are more likely to overlook macro information albeit its wide availability, and thus, sentiment likely influences asset pricing during those periods. Our results are consistent with Huang, Jiang, Tu, and Zhou (2015) and Shen, Yu, and Zhao (2016), who document that investor sentiment’s predictive power is stronger during high-sentiment periods. In short, we find that the predictive power of all PLS indices, including our new corporate index $PLS^C$, mainly stems from high-sentiment periods.

4.4 Bivariate regressions with macroeconomics predictors

We further compare the forecasting power of the corporate index $PLS^C$ with macroeconomic predictors by investigating whether the forecasting power of $PLS^C$ remains significant after
controlling for economic predictors.

To analyze the incremental forecasting power of $PLS_C$, we conduct the following bivariate predictive regressions on $E_t^{k-1}$ and $PLS_C$,

$$ R_t^m = \alpha + \phi E_t^{k-1} + \beta PLS_t^{C-1} + \epsilon_t, \quad k = 1, \ldots, 14, \quad (8) $$

where $E_t^{k-1}$ is either one of the macroeconomic predictors or the aggregate macro index. We are interested to see whether the regression slope $\beta$ of $PLS_C$ is significant and $R^2$ improves.

Panel B of Table 4 reports the results of the bivariate regressions. For easy comparison, we provide the univariate regressions of the macroeconomic variables in Panel A (which are part of Table 2). We observe that for each regression, the slope $\beta$ of $PLS_C$ remains statistically significant when augmented by the economic predictors. The value of the coefficient ranges from 2.31 to 3.37, in line with the coefficient of 2.44 in the univariate regression on the corporate index reported in Table 3. Each $R^2$ of the bivariate regressions in Panel B is substantially larger than the corresponding $R^2$ of the univariate regression in Panel A of Table 4 when an macroeconomic variable is the only predictor. For the first 14 rows, the $R^2$s in Panel A range from 0.07% to 2.50%, whereas the $R^2$s in Panel B range from 8.55% to 13.36%.

In the 15th row, we replace an individual macroeconomic variable with the macro index, $PLS^E$. As reported earlier, the univariate regression has an $R^2$ of 1.89%. Adding the corporate index $PLS_C$ as another predictor, the $R^2$ increases significantly to 9.93%. These results demonstrate that $PLS_C$ contains sizable additional forecasting information beyond what is contained in the macro predictors.

### 4.5 Out-of-sample forecasts

Welch and Goyal (2008), among others, argue that out-of-sample tests are more relevant for assessing true return predictability for real world investors. In addition, out-of-sample tests are less susceptible to the small-sample size distortions such as the Stambaugh bias and the look-ahead bias of the PLS approach (Kelly and Pruitt, 2013). We therefore examine the out-of-sample return predictability of the corporate index.
Following the literature (e.g., Welch and Goyal, 2008), we run the out-of-sample tests by estimating the predictive regression model recursively:

\[ \hat{R}_{t+1}^m = \hat{\alpha}_t + \hat{\beta}_t PLS_t^C, \]  

(9)

where \( \hat{\alpha}_t \) and \( \hat{\beta}_t \) are the OLS estimates from regressing \( \{ R_{s+1}^m \}_{s=1}^{t-1} \) on a constant and a predictor \( \{ PLS_s^C \}_{s=1}^{t-1} \).

Let \( p \) be a fixed number chosen for the initial training periods. The future periods can be expressed as \( t = p + 1, p + 2, \ldots, T \). This will give us \( T - p \) out-of-sample periods. In order to balance the need for a relatively long out-of-sample period for forecast evaluation with the need for enough observations to accurately estimate the initial parameters, our training period is from 1986Q2 through 1999Q4, and our forecast evaluation period is from 2000Q1 through 2015Q4.

We evaluate the out-of-sample forecasting performance based on the widely-used Campbell and Thompson (2008) \( R^2_{OS} \) statistic. The \( R^2_{OS} \) statistic captures the proportional reduction in mean squared forecast error (MSFE) for the predictive regression forecast relative to the historical average benchmark,

\[ R^2_{OS} = 1 - \frac{\sum_{t=p}^{T-1} (R_{t+1}^m - \hat{R}_{t+1}^m)^2}{\sum_{t=p}^{T-1} (R_{t+1}^m - \bar{R}_{t+1}^m)^2}, \]  

(10)

where \( \bar{R}_{t+1}^m \) is the historical average market return up to time \( t \):

\[ \bar{R}_{t+1}^m = \frac{1}{t} \sum_{s=1}^{t} R_{s}^m, \]  

(11)

which is the standard benchmark for assessing predictability. The \( R^2_{OS} \) statistics lie in the range \((-\infty, 1]\), and \( R^2_{OS} > 0 \) implies predictability.

Welch and Goyal (2008) show that it is very hard for an individual economic variable to beat the historical average benchmark. To test whether \( R^2_{OS} > 0 \), we apply Diebold and Mariano (1995) statistic modified by McCracken (2007) (DM-test hereafter). This test is standard. It tests in general for the equality of the mean squared forecast errors (MSFE) of one forecast relative to another. Here our null hypothesis is that the historical average has a MSFE that is less than, or equal to, that of the predictive regression model. Comparing a predictive regression forecast to the historical average entails comparing nested models, as
the predictive regression model reduces to the historical average under the null hypothesis. McCracken (2007) shows that the modified DM-test statistic follows a nonstandard normal distribution when testing nested models, and provides bootstrapped critical values for the nonstandard distribution.

Table 5 reports the $R^2_{OS}$ for $PLS^C$, and for comparison purposes, various other PLS indices. Columns 2 and 3 of the table show the $R^2_{OS}$s in market booms and market troughs. The three PLS indices ($PLS^C, PLS^E, PLS^S$) all generate positive $R^2_{OS}$ statistics, and they perform better in market downturns. Nonetheless, the corporate index $PLS^C$ performs much better than the other two PLS indices, both for the full sample and for up and down markets separately. For example, the out-of-sample $R^2_{OS}$ of $PLS^C$ is much greater than those of $PLS^E$ and $PLS^S$ (12.48% vs. −0.80% and 7.12%).

We also look at $R^2_{OS}$s for indices based on the aggregation of variables of different categories, such as $PLS^{EC}, PLS^{SC}$, and $PLS^{ESC}$. Consistent with the in-sample tests, the $R^2_{OS}$s of both $PLS^{EC}$ and $PLS^{SC}$, after incorporating corporate information, are substantial improvements of those of $PLS^E$ and $PLS^S$, and the $R^2_{OS}$ of $PLS^{ESC}$ is much larger than that of $PLS^{ES}$ (14.13% vs. 2.77%). These results suggest that taking into account corporate information can substantially improve the market return predictability.

In summary, this subsection shows that the corporate index $PLS^C$ displays strong out-of-sample forecasting power for the aggregate stock market. Its $R^2_{OS}$ is as high as 12.48%, exceeding substantially those of $PLS^E$ and $PLS^S$. The DM-test statistic of $PLS^C$ is 2.93, suggesting that $PLS^C$’s MSFE is significantly smaller than that of the historical average at 5% level. In addition, corporate variables substantially improve the predictive performance of macroeconomic variables and sentiment variables in light of the combined PLS indices, consistent with the in-sample results.

4.6 Asset allocation implications

In this subsection, we examine the economic value of stock return forecasts based on the corporate index. Following the literature (e.g., Kandel, Ofer, and Sarig, 1996; Campbell and Thompson, 2008; Huang, Jiang, Tu, and Zhou, 2015), we compute the certainty equivalent
return (CER) gain for a mean-variance investor who optimally allocates across equities and
the risk-free asset by using the out-of-sample predictive regression forecasts.

At the end of quarter $t$, the investor optimally allocates

$$w_t = \frac{1}{\gamma} \frac{\hat{R}_{t+1}^{ms}}{\hat{\sigma}_{t+1}^2}$$

(12)
of the portfolio to equities during quarter $t + 1$, where $\gamma$ is the risk aversion coefficient, $\hat{R}_{t+1}^{ms}$
is the out-of-sample forecast of the simple excess market return, and $\hat{\sigma}_{t+1}^2$ is the variance
forecast. The investor then allocates $1 - w_t$ of the portfolio to the risk-free asset. Thus, the
realized portfolio return in quarter $t + 1$ is

$$R_{t+1}^p = w_t R_{t+1}^{ms} + R_{t+1}^f,$$

(13)
where $R_{t+1}^f$ is the risk-free return. Following Campbell and Thompson (2008), we use a five-
year moving window of past quarterly returns to estimate the variance of the excess market
return, and limit $w_t$ to lie between 0 and 1.5 to exclude short sales and allow up to 50%
leverage.

The CER of the portfolio is

$$CER_p = \mu_p - 0.5\gamma \hat{\sigma}_p^2,$$

(14)
where $\mu_p$ and $\hat{\sigma}_p$ are the mean and variance, respectively, for the investor’s portfolio over the
$(T - p)$ forecasting evaluation periods, where $p$ is the initial training period $p$. The CER
gain reflects the difference between the CER for the investor who uses a predictive regression
forecast of market return generated by Equation 9 and the CER for an investor who uses the
historical average forecast, Equation 11. We multiply this difference by four so that it can be
interpreted as the annual portfolio management fee that an investor would be willing to pay
to have access to the predictive regression forecast instead of the historical average forecast.

Table 6 reports the CER values under three commonly used risk-aversion coefficient values
of 1, 3 and 5. The results show that $PLS^C$ generates significant economic gains for a mean-
variance investor, and consistently provides positive CERs in both up and down markets. We
compute CERs for $PLS^E$ and $PLS^S$ as well for comparison. Under all risk-aversion coefficient
values, the CERs of $PLS^C$ are consistently and substantially higher than those of $PLS^E$ and
\( PLS^S \), both for the full sample and for up and down markets respectively. For example, assuming a risk-aversion coefficient of 3, the CER gain for a mean-variance investor is 8.78\% per annum, which is more than that of \( PLS^E \) of 0.57\%. The CER gain of \( PLS^C \) is also larger than that of \( PLS^S \), which is 4.14\%.

5 Economic Explanations

5.1 Forecasting characteristics portfolios based on information asymmetry

The premise to use corporate variables to predict market returns is that corporate executives have private information. The greater the information asymmetry between corporate executives and the general public, the greater explanatory power we expect the corporate predictors to have. In this subsection, we investigate the impact of \( PLS^C \) on portfolios constructed based on asymmetric information measures.

We use several different measures of asymmetric information that are all well accepted in the literature. The literature suggests that asymmetric information decreases with firm size (e.g., Vermaelen, 1981; Diamond and Verrecchia, 1991), decreases with asset tangibility (e.g., Aboody and Lev, 2000), and increases with growth opportunity (e.g., Smith and Watts, 1992; McLaughlin, Safieddine, and Vasudevan, 1998). Firms with greater asymmetric information also tend to have higher analyst forecast errors and dispersions (e.g., Krishnaswami and Subramaniam, 1999).

We measure asset tangibility as property, plant and equipment (PP&EqE) scaled by the market value of firm assets. We also use book assets as an alternative standardization and find that the results are robust. Firm size is measured as market capitalization. For growth opportunity, we use the book-to-market ratio as an inverse measure. To measure analyst forecast errors and dispersion, we use earnings forecast data from the Institutional Brokers Estimate System (I/B/E/S). Analyst forecast error is the ratio of the absolute difference between the average earnings forecast and the actual earnings per share to the price per share. The analyst forecast dispersion measure is the standard deviation of earnings forecasts.
At the end of each June from 1985 to 2015, we sort firms into quintiles with an information asymmetry measure. We then calculate the value-weighted portfolio returns for these portfolios from July of that year up to June of the next year. We calculate five time series of quarterly portfolio returns.

We estimate Equation 5, but replace excess market returns with excess portfolio returns. Table 7 reports the results. For ease of interpretation, we present portfolios in the ascending order of information asymmetry for all information asymmetry measures. That is, portfolios in Panel A decrease in PP&E ratio, moving from bottom to top quintile, because firms with more tangible assets are more transparent. Similarly, portfolios in Panel B and C have increasing analyst forecast error (dispersion). On the other hand, portfolios in Panel D and E have decreasing market cap (or book-to-market).

We find that under all information asymmetry measures, the corporate index $PLS_C$ have predictive ability across all five portfolios. More interestingly, the regression coefficient steadily increases from the bottom quintile (most transparent) to the top quintile (most opaque). For example, in Panel A, the coefficient increases from 1.54 to 3.40. This suggests that with a one-standard-deviation increase in $PLS_C$, the portfolio return increases by 1.54% for the most transparent portfolio, and by 3.40% for the most opaque portfolio. The greater sensitivity of the returns of opaque firms to the corporate index is consistent with our conjecture that the corporate index contains more valuable information about these firms. In contrast, $PLS_E$ and $PLS_S$ can barely predict the portfolio returns sorted on asymmetry measures.

Under various measures of information asymmetry, the regression $R^2$ does not show a consistent directional change when moving from transparent to opaque portfolios. This is because opaque firms have larger stock volatilities. We report the standard deviation of the dependent variable in a separate column. This statistic steadily increases from transparent to opaque portfolios in each panel of the table. We also compute the explained standard deviation, which is the regression coefficient times the standard deviation of the predictor $PLS_C$. As expected, explained variation steadily increases from transparent to opaque portfolios.\(^7\)

\(^7\)Because the standard deviation of $PLS_C$ is one, the explained standard deviation is the same as the regression coefficient.
5.2 Cash flow and discount rate predictability

The predictability of returns could stem from predictability of cash flows or discount rates, or both. Fama and French (1989) and Cochrane (2008, 2011), among many others, argue that stock market predictability stems from the discount rate channel. Under the discount rate channel, high $PLS^C$ predicts high future return, because it predicts discount rates to be lower than the market consensus. Under the cash flow channel, high $PLS^C$ predicts high future return, because it predicts cash flow to be higher than the market consensus.

To test whether the predictability of $PLS^C$ is from either or both of the channels, we use several proxies for future discount rates and cash flows. Because Cochrane (2008, 2011) find that the time variation in aggregate dividend-price ratio is mainly attributable to discount rates, we use the aggregate dividend-price ratio as our discount rate proxy. For the cash-flow proxies, following the literature (e.g., Campbell and Shiller, 1988a; Fama and French, 2000; Menzly, Santos, and Veronesi, 2004; Lettau and Ludvigson, 2005; Cochrane, 2008, 2011; van Binsbergen and Koiwen, 2010; Kelly and Pruitt, 2013), we use aggregate dividend growth. To circumvent dividend smoothing (see Fama and French, 2000), we examine aggregate earnings growth and real GDP growth as two alternative cash-flow proxies.

We use Campbell and Shiller (1988a)’s log linearization of stock return to determine the source of predictability of $PLS^C$. Stock returns can be broken down into three parts,

$$ R_{t+1} = k + DG_{t+1} - \rho D/P_{t+1} + D/P_t, $$

where $R_{t+1}$ is the aggregate stock market return from $t$ to $t+1$, $DG_{t+1}$ is the log aggregate dividend-growth rate, $D/P_{t+1}$ is the log aggregate dividend-price ratio, and $\rho$ is a positive log-linearization constant. Equation 15 implies that if $PLS^C$ predicts next period market return $R_{t+1}$ beyond the information contained in $D/P_t$, it must predict either $DG_{t+1}$ or $D/P_{t+1}$ (or both). Because $DG_{t+1}$ and $D/P_{t+1}$ represent cash flows and discount rates, respectively, the predictive power of $PLS^C$ for $DG_{t+1}$ and $D/P_{t+1}$ will reveal whether the predictability is through the cash-flow channel or the discount rate channel, or both.

Therefore, our study focuses on the following bivariate predictive regressions,

$$ Y_{t+1} = \alpha + \beta PLS_t^C + \phi D/P_t + \epsilon_{t+1}, \quad Y = D/P, DG, EG, GDPG, $$

(16)
where $D/P_{t+1}$ is the log dividend-price ratio on the S&P 500 index at the end of year $t+1$, $DG_{t+1}$ is the annual log dividend-growth rate on the S&P 500 index from year $t$ to $t+1$, $EG_{t+1}$ is the annual log earning growth rate on the S&P 500 index from year $t$ to $t+1$, $GDPG_{t+1}$ is the annual log real GDP growth rate from year $t$ to $t+1$. The latter three variables are all cash-flow proxies. Following the literature, we use annual data in the above regressions to avoid spurious predictability arising from within-year seasonality. We construct $D/P_{t+1}$ and $DG_{t+1}$ following Cochrane (2008, 2011). The sample period is from 1987 to 2015.

Panel A of Table 8 shows the regression results. In the regression of $D/P_{t+1}$, the slope of $PLS_C$ is virtually zero and statistically insignificant, suggesting that $PLS_C$ does not predict discount rates. In contrast, $PLS_C$ is able to predict two out of three cash flow proxies; the slope coefficient is statistically significant for both the regressions of $EG_{t+1}$ and $GDPG_{t+1}$. The evidence suggests that the corporate index predicts future market returns by predicting future cash flows.

Panel B of Table 8 reports the corresponding results of using the macroeconomic index $PLS_E$ in place of $PLS_C$. The slope of $PLS_E$ is statistically significant in the regression of $EG_{t+1}$, but not in the regressions of $D/P_{t+1}$, $DG_{t+1}$, or $GDPG_{t+1}$. Panel C of Table 8 shows the results for $PLS_S$. Like $PLS_E$, $PLS_S$ can predict earnings growth and dividend growth, but not the regressions of GDP growth, or the dividend yield.

The evidence suggests that that the predictive power of the corporate index is due to its ability to predict cash flows, rather than discount rates. This is consistent with the notion that corporate executives have private information about the cash flows of their firms (and in turn about the aggregate cash flows). Their private information is not about discount rates which depend on the economy wide factors, investor risk aversion, or investor sentiment.

Our results are consistent with Huang, Jiang, Tu, and Zhou (2015), who document that negative-return predictability of $PLS_S$ for aggregate stock market return is coming from the cash-flow channel. Both papers’ results are in contrast to the popular time-varying discount-rate interpretation of market-return predictability in extant literature.
5.3 The cross-section of cash-flow channel

To further illuminate the economic source of the predictive ability of the corporate index, we extend our analysis to the cross-section at the portfolio level. If the predictive ability of $PLS^C$ comes from the cash flow channel, it should have stronger forecasting power for the cash flows of opaque firms as well.

We examine the predictive ability of $PLS^C$ for the cross-section of cash flows using the predictive regression,

$$EG^i_{t+1} = \alpha_j + \phi_j PLS^C_t + \epsilon^i_{t+1},$$  \hspace{1cm} (17)

where $EG^i_{t+1}$ is the annual log earning-growth rate for one of the characteristic portfolios based on asymmetric information, examined in Table 7, where $j=1$ to 5. We are interested in the predictive regression slope $\phi_j$ on $PLS^C$ in Equation 17, which measures the ability of the corporate index to forecast cash flows in the cross-section. We do not use the other two cash flow proxies because we do not find predictability for $DG^i_{t+1}$ in Table 8, and there is no cross-sectional variation for $GDPG_{t+1}$.

We test whether the ability of the corporate index to forecast stock returns is positively associated with its ability to forecast cash flows in the cross-section. We use a cross-sectional regression to statistically test the cash-flow channel, in the spirit of Hong, Torous, and Valkanov (2007). We run the cross-section regression

$$\beta_j = \alpha + g\phi_j + e_j,$$  \hspace{1cm} (18)

where $\phi_j$ is from Equation 17 that measures the ability of the corporate index to forecast the cash flows for a characteristic portfolio $j$, and $\beta_j$ is from Equation 16 measuring the ability of the corporate index to forecast the returns of portfolio $j$ (annualized by multiplying 4). If the cash-flow channel hypothesis holds, we expect a positive relationship between $\beta_j$ and $\phi_j$; that is, $g > 0$.

Panel A of Table 9 presents the results of Regression 17. Under each of the information asymmetry measures, $\phi_j$ increases when moving from transparent to opaque portfolios of firms. For example, for the five portfolios sorted based on PP&E ratio, $\phi_j$ increases from 2.77 to
7.12. This suggests that the cash-flow predictability by the corporate index is higher for more opaque firms, which is consistent with the higher predictability of the firms’ stock returns.

Panel B presents the results of Regression 18. Consistent with our hypothesis, the slope coefficient, $g$, is significantly positive, under all measures of information asymmetry. This confirms that the corporate index predict returns through the cash flow channel.

### 5.4 Corporate predictors and economic growth

In this section, we examine whether corporate predictors can forecast real economic activity. Cochrane (2007) argues that return forecasts are more plausibly related to macroeconomic risk if the return predictors can forecast business cycles. Then the predictability can be more credibly attributed to time-varying risk premia due to changing risks or risk aversion. The macro variables are likely to have better predicting power for future macro conditions. But they do not have good power in forecasting return as shown by numerous studies, including ours. The question is whether corporate predictors have any relation to future economic activities.

To examine whether the PLS indices can forecast real economic activity, following Lin, Wu, and Zhou (2015) and others, we consider the following predictive regression:

$$\Delta Y_{t+1} = \alpha + \beta PLs^C_t + \epsilon_t,$$

where $\Delta Y_{t+1}$ is the change in macroeconomic conditions in the next period.

We employ the following measures of $Y_{t+1}$ in the predictive regression, following Lin, Wu, and Zhou (2015):

- **Smooth recession probability**, SRP: The data of smooth recession probability is obtained from the Federal Reserve Bank of St. Louis. This recession probability is estimated by the dynamic-factor Markov-switching model of Chauvet (1998) using four monthly coincident variables: non-farm payroll employment, the index of industrial production, real personal income excluding transfer payments, and real manufacturing and trade sales;
- *Industrial production growth*, IPG: The production growth rate data are also obtained from the Federal Reserve Bank of St. Louis;

- *Treasury bill rates*, TBL: Interest rate on a three-month Treasury bill (secondary market);

- *Default yield spreads*, DFY: Difference between BAA- and AAA-rated corporate bond yields;

- *Implied volatility index*, VIX: VIX is the implied volatility of S&P 500 index options, which reflects the expectation of stock market volatility over the next 30-day period. It is widely used as a gauge of fear (Remolona, Scatigna, and Wu, 2008). The data are downloaded from Chicago Board Options Exchange (CBOE);

- *Chicago Fed National Activity Index*, CFNAI: The CFNAI is a monthly index designed to capture economic activity and inflationary pressure. The CFNAI data are downloaded from the Federal Reserve Bank of Chicago. We use quarter end data in our study.

Table 10 reports results of the predictive regression in Equation 19 at quarterly frequency. It shows that our corporate index can forecast macroeconomic activity. The regression slopes are all significant at the 10% level except one on TBL. While the macro index $PLS^E$ can predict most of the macro activities better than $PLS^C$, it does much worse than $PLS^C$ in predicting the market volatility (with $R^2$'s 1.08% vs. 12.06%). Taken together with the earlier findings that the corporate index can predict the market return better than the macro index, it suggests that the corporate index contains more private information while the information contained in the macro index is mostly public and well understood by investors.

6 Conclusion

In this study, we provide the first comprehensive examination of the predictive power of corporate activities for aggregate stock market returns. We use the partial least square method to construct a corporate index out of 13 variables that measure five categories of corporate activities. We find that the corporate index can better predict stock market returns than
individual corporate variables, popular macroeconomic predictors, as well as sentiment predictors. A mean-variance investor who follows predictions based on this index could earn an additional 8.78% per annum.

Cross-sectionally, the corporate index has greater predictive power for portfolios of firms with greater information asymmetry, consistent with the notion that its predictive power is rooted in corporate insiders’ private information. We also find evidence that the corporate index predicts market returns via predicting future cash flows, but not the discount rate.
References


Table 1: Summary statistics

This table provides summary statistics for the quarterly excess market return \(R^m\), the log return on the S&P 500 index in excess of the risk-free rate, risk-free rate \(R^f\), percentage of stock payment \(\text{COMPCT}\), log total stock payment \(\text{COM}\), net transaction of insider trading \(\text{NT}\), net dollar amount of insider trading \(\text{NDA}\), ratio of net purchases \(\text{RT}\), ratio of net purchasing dollar amount \(\text{RDA}\), capital expenditure scaled by market capitalization \(\text{CAPXME}\), capital expenditure scaled by average asset \(\text{CAPXAT}\), change in net operating asset scaled by market capitalization \(\text{ALME}\), change in net operating asset scaled by average asset \(\text{ALAT}\), log total equity issuance \(\text{E}\), ratio of equity issuance \(\text{S}\), aggregate share repurchases \(\text{REPO}\). For each variable, the time-series average \(\text{Mean}\), standard deviation \(\text{Std}\), skewness \(\text{Skew}\), kurtosis \(\text{Kurt}\), minimum \(\text{Min}\), maximum \(\text{Max}\), and first-order autocorrelation \(\rho_1\) are reported. The quarterly Sharpe ratio \(\text{SR}\) is the mean excess market return divided by its standard deviation. All variables are at quarterly frequency. The sample period is from 1986Q2 through 2015Q4.

<table>
<thead>
<tr>
<th>Quarterly Excess market return and risk-free rate</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
<th>Min</th>
<th>Max</th>
<th>(\rho_1)</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R^m) (%)</td>
<td>1.55</td>
<td>8.11</td>
<td>-0.96</td>
<td>4.57</td>
<td>-27.31</td>
<td>18.39</td>
<td>0.04</td>
<td>0.19</td>
</tr>
<tr>
<td>(R^f) (%)</td>
<td>0.84</td>
<td>0.63</td>
<td>0.04</td>
<td>1.82</td>
<td>0.00</td>
<td>2.21</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>13 corporate predictors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPCT (%)</td>
<td>60.25</td>
<td>30.48</td>
<td>-0.41</td>
<td>1.81</td>
<td>0.06</td>
<td>99.09</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>COM</td>
<td>8.51</td>
<td>1.60</td>
<td>-0.72</td>
<td>4.74</td>
<td>2.14</td>
<td>12.30</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>NT</td>
<td>-7.84</td>
<td>11.09</td>
<td>-2.07</td>
<td>7.52</td>
<td>-52.40</td>
<td>4.13</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>NDA</td>
<td>-2.09</td>
<td>1.81</td>
<td>-0.61</td>
<td>2.36</td>
<td>-7.24</td>
<td>0.04</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>RT (%)</td>
<td>-34.94</td>
<td>28.50</td>
<td>0.63</td>
<td>2.93</td>
<td>-82.73</td>
<td>54.31</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>RDA (%)</td>
<td>-50.34</td>
<td>17.55</td>
<td>1.38</td>
<td>5.06</td>
<td>-79.56</td>
<td>11.86</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>CAPXME (%)</td>
<td>1.78</td>
<td>0.51</td>
<td>1.06</td>
<td>4.59</td>
<td>0.88</td>
<td>3.51</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>CAPXAT (%)</td>
<td>0.87</td>
<td>0.23</td>
<td>0.57</td>
<td>3.06</td>
<td>0.45</td>
<td>1.57</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>ALME (%)</td>
<td>1.10</td>
<td>0.65</td>
<td>-1.71</td>
<td>9.95</td>
<td>-2.52</td>
<td>2.34</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>ALAT (%)</td>
<td>1.71</td>
<td>0.77</td>
<td>1.02</td>
<td>7.15</td>
<td>-0.75</td>
<td>5.12</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>9.67</td>
<td>0.50</td>
<td>-0.93</td>
<td>4.61</td>
<td>7.97</td>
<td>10.99</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>S (%)</td>
<td>12.43</td>
<td>5.89</td>
<td>1.30</td>
<td>5.68</td>
<td>4.13</td>
<td>34.21</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>REPO</td>
<td>9.18</td>
<td>1.19</td>
<td>-0.13</td>
<td>1.98</td>
<td>6.67</td>
<td>11.06</td>
<td>0.97</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Forecasting market return with individual predictors

This table provides in-sample estimation results for the predictive regression \( R_{mt} = \alpha + \beta x_{t-1} + \epsilon_t \), where \( R_{mt} \) denotes the quarterly excess market return (%). \( x_{t-1} \) is a predictor. Panel A considers 13 corporate variables; Panel B considers 14 macroeconomic variables; and Panel C considers 5 sentiment variables. All variables are normalized by the sample mean and standard deviation. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

<table>
<thead>
<tr>
<th>Panel A: Corporate variables</th>
<th>( \beta )</th>
<th>t-stat</th>
<th>( R^2 ) (%)</th>
<th>Panel B: Macroeconomic variables</th>
<th>( \beta )</th>
<th>t-stat</th>
<th>( R^2 ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPCT</td>
<td>-1.33***</td>
<td>-1.79</td>
<td>2.67</td>
<td>DP</td>
<td>1.23**</td>
<td>1.66</td>
<td>2.29</td>
</tr>
<tr>
<td>COM</td>
<td>-1.63***</td>
<td>-2.21</td>
<td>4.02</td>
<td>DY</td>
<td>1.29**</td>
<td>1.73</td>
<td>2.50</td>
</tr>
<tr>
<td>NT</td>
<td>1.52**</td>
<td>2.06</td>
<td>3.50</td>
<td>EP</td>
<td>0.73</td>
<td>0.97</td>
<td>0.80</td>
</tr>
<tr>
<td>NDA</td>
<td>1.09*</td>
<td>1.46</td>
<td>1.79</td>
<td>DE</td>
<td>0.26</td>
<td>0.34</td>
<td>0.10</td>
</tr>
<tr>
<td>RT</td>
<td>1.30**</td>
<td>1.76</td>
<td>2.57</td>
<td>SVAR</td>
<td>-0.41</td>
<td>-0.54</td>
<td>0.25</td>
</tr>
<tr>
<td>RDA</td>
<td>1.25**</td>
<td>1.68</td>
<td>2.36</td>
<td>BM</td>
<td>1.26**</td>
<td>1.69</td>
<td>2.39</td>
</tr>
<tr>
<td>CAPXME</td>
<td>1.24**</td>
<td>1.66</td>
<td>2.31</td>
<td>NTIS</td>
<td>0.80</td>
<td>1.07</td>
<td>0.97</td>
</tr>
<tr>
<td>CAPXAT</td>
<td>0.55</td>
<td>0.74</td>
<td>0.46</td>
<td>TBL</td>
<td>-0.22</td>
<td>-0.29</td>
<td>0.07</td>
</tr>
<tr>
<td>ALME</td>
<td>-0.28</td>
<td>-0.38</td>
<td>0.12</td>
<td>LTY</td>
<td>-0.42</td>
<td>-0.56</td>
<td>0.27</td>
</tr>
<tr>
<td>ALAT</td>
<td>-1.29***</td>
<td>-1.73</td>
<td>2.50</td>
<td>LTR</td>
<td>0.66</td>
<td>0.88</td>
<td>0.66</td>
</tr>
<tr>
<td>E</td>
<td>-1.81***</td>
<td>-2.46</td>
<td>4.93</td>
<td>TMS</td>
<td>-0.22</td>
<td>-0.30</td>
<td>0.08</td>
</tr>
<tr>
<td>S</td>
<td>-1.43**</td>
<td>-1.93</td>
<td>3.10</td>
<td>DFY</td>
<td>-0.33</td>
<td>-0.44</td>
<td>0.17</td>
</tr>
<tr>
<td>REPO</td>
<td>-0.48</td>
<td>-0.64</td>
<td>0.35</td>
<td>DFR</td>
<td>0.78</td>
<td>1.04</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>INFL</td>
<td>-1.09*</td>
<td>-1.46</td>
<td>1.79</td>
</tr>
<tr>
<td>Panel C: Sentiment variables</td>
<td></td>
<td></td>
<td></td>
<td>CEFD</td>
<td>1.95*</td>
<td>1.62</td>
<td>2.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NIPO</td>
<td>0.39</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RIPO</td>
<td>-1.80**</td>
<td>-1.98</td>
<td>3.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PDND</td>
<td>1.14</td>
<td>0.79</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EQTI</td>
<td>0.78</td>
<td>0.65</td>
<td>0.36</td>
</tr>
</tbody>
</table>
Table 3: Forecasting market return with PLS indices (in-sample)

This table provides in-sample estimation results for the predictive regression \( R_{mt} = \alpha + \beta \text{PLS}_{k,t-1} + \epsilon_t \), where \( R_{mt} \) is the quarterly market excess return. \( \text{PLS}_{k,t-1} \) (k = C, E, S, ES, EC, SC, ESC) is one of the PLS indices. \( \text{PLS}^C \) is the PLS index of 13 corporate variables; \( \text{PLS}^E \) is the PLS index of 14 macroeconomic variables; \( \text{PLS}^S \) is the PLS index of 5 sentiment variables; \( \text{PLS}^{ES} \) is the PLS index of 14 macroeconomic variables and 5 sentiment variables; \( \text{PLS}^{EC} \) is the PLS index of 14 macroeconomic variables and 13 corporate variables; \( \text{PLS}^{SC} \) is the PLS index of 5 sentiment variables and 13 corporate variables, and \( \text{PLS}^{ESC} \) is the PLS index of 14 macroeconomic variables, 5 sentiment variables and 13 corporate variables. \( R^2_{\text{up}} \) (\( R^2_{\text{down}} \)) statistics are calculated over NBER-dated business-cycle expansions (recessions). \( R^2_{\text{high}} \) (\( R^2_{\text{low}} \)) are calculated over high (low) sentiment periods, respectively. Both indicators are corresponding to time t. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from 1986Q2 through 2015Q4.

<table>
<thead>
<tr>
<th></th>
<th>( \beta )</th>
<th>t-stat</th>
<th>( R^2(%) )</th>
<th>( R^2_{\text{up}}(%) )</th>
<th>( R^2_{\text{down}}(%) )</th>
<th>( R^2_{\text{high}}(%) )</th>
<th>( R^2_{\text{low}}(%) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{PLS}^C )</td>
<td>2.37***</td>
<td>3.29</td>
<td>8.47</td>
<td>2.53</td>
<td>21.29</td>
<td>12.82</td>
<td>3.91</td>
</tr>
<tr>
<td>( \text{PLS}^E )</td>
<td>1.12*</td>
<td>1.50</td>
<td>1.89</td>
<td>0.84</td>
<td>4.16</td>
<td>0.03</td>
<td>3.84</td>
</tr>
<tr>
<td>( \text{PLS}^S )</td>
<td>-1.85***</td>
<td>-2.52</td>
<td>5.16</td>
<td>3.30</td>
<td>9.15</td>
<td>7.02</td>
<td>3.19</td>
</tr>
<tr>
<td>( \text{PLS}^{EC} )</td>
<td>2.77***</td>
<td>3.91</td>
<td>11.58</td>
<td>2.22</td>
<td>31.75</td>
<td>19.76</td>
<td>2.99</td>
</tr>
<tr>
<td>( \text{PLS}^{SC} )</td>
<td>-2.52***</td>
<td>-3.52</td>
<td>9.55</td>
<td>3.07</td>
<td>23.52</td>
<td>13.50</td>
<td>5.89</td>
</tr>
<tr>
<td>( \text{PLS}^{ES} )</td>
<td>-1.59**</td>
<td>-2.16</td>
<td>3.82</td>
<td>2.18</td>
<td>7.36</td>
<td>2.38</td>
<td>5.35</td>
</tr>
<tr>
<td>( \text{PLS}^{ESC} )</td>
<td>-2.84***</td>
<td>-4.03</td>
<td>12.19</td>
<td>2.77</td>
<td>32.51</td>
<td>20.57</td>
<td>3.40</td>
</tr>
</tbody>
</table>
Table 4: Comparison with economic return predictors

Panel A reports the in-sample estimation for the predictive regression model of regressing quarterly market excess return on one of the 14 macroeconomic predictors $x_i$, and on both the lagged corporate index $PLS^C$ and $x_i$, respectively. The first column of the first 14 rows are the individual macroeconomic variables. $PLS^E$ is the PLS index based on the 14 macroeconomic variables. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from 1986Q2 through 2015Q4.

<table>
<thead>
<tr>
<th>Panel A: Univariate predictive regression</th>
<th>Panel B: Bivariate predictive regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_t^m = \alpha + \phi E_{t-1}^E + \epsilon_t$</td>
<td>$R_t^m = \alpha + \phi E_{t-1}^E + \beta PLSC_{t-1}^C + \epsilon_t$</td>
</tr>
<tr>
<td>$\phi$</td>
<td>$\phi$</td>
</tr>
<tr>
<td>DP</td>
<td>1.23**</td>
</tr>
<tr>
<td>DY</td>
<td>1.29**</td>
</tr>
<tr>
<td>EP</td>
<td>0.73</td>
</tr>
<tr>
<td>DE</td>
<td>0.26</td>
</tr>
<tr>
<td>SVAR</td>
<td>-0.41</td>
</tr>
<tr>
<td>BM</td>
<td>1.26**</td>
</tr>
<tr>
<td>NTIS</td>
<td>0.80</td>
</tr>
<tr>
<td>TBL</td>
<td>-0.22</td>
</tr>
<tr>
<td>LTY</td>
<td>-0.42</td>
</tr>
<tr>
<td>LTR</td>
<td>0.66</td>
</tr>
<tr>
<td>TMS</td>
<td>-0.22</td>
</tr>
<tr>
<td>DFY</td>
<td>-0.33</td>
</tr>
<tr>
<td>DFR</td>
<td>0.78</td>
</tr>
<tr>
<td>INFL</td>
<td>-1.09*</td>
</tr>
<tr>
<td>$PLS^E$</td>
<td>1.12*</td>
</tr>
</tbody>
</table>
Table 5: Out-of-sample forecasting results

This table reports the out-of-sample performance of various PLS indices in predicting the quarterly excess market returns. All of the predictors and regression slopes are estimated recursively using the data available at the forecast formation time t. $R^2_{OS}$ is the out-of-sample $R^2$, calculated using Equation 10. DM-test is the modified Diebold and Mariano (1995) $t$-statistics. $R^2_{OS,up}$ ($R^2_{OS,down}$) statistics are calculated over NBER-dated business-cycle expansions (recessions). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The full sample period is from 1986Q2 to 2015Q4. The initial training period is from 1986Q2 through 1999Q4, and the out-of-sample evaluation period is from 2000Q1 through 2015Q4.

<table>
<thead>
<tr>
<th>PLS Index</th>
<th>$R^2_{OS}$ (%)</th>
<th>DM-test</th>
<th>$R^2_{OS,up}$ (%)</th>
<th>$R^2_{OS,down}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PLS^C$</td>
<td>12.48***</td>
<td>2.93***</td>
<td>7.67**</td>
<td>19.01***</td>
</tr>
<tr>
<td>$PLS^E$</td>
<td>−0.80</td>
<td>−0.20</td>
<td>−0.52</td>
<td>−1.19</td>
</tr>
<tr>
<td>$PLS^S$</td>
<td>7.12**</td>
<td>1.77*</td>
<td>5.99*</td>
<td>8.66**</td>
</tr>
<tr>
<td>$PLS^{EC}$</td>
<td>13.90***</td>
<td>2.88***</td>
<td>8.69***</td>
<td>20.97***</td>
</tr>
<tr>
<td>$PLS^{SC}$</td>
<td>13.08***</td>
<td>3.01***</td>
<td>8.70***</td>
<td>19.03**</td>
</tr>
<tr>
<td>$PLS^{ES}$</td>
<td>2.77</td>
<td>0.62</td>
<td>4.89*</td>
<td>−0.10</td>
</tr>
<tr>
<td>$PLS^{ESC}$</td>
<td>14.13***</td>
<td>2.90***</td>
<td>9.44***</td>
<td>20.49***</td>
</tr>
</tbody>
</table>
Table 6: Asset allocation results

This table reports the portfolio performance measures for a mean-variance investor with a risk-aversion coefficient ($\gamma$) of 1, 3, or 5 who allocates quarterly between equities and risk-free bills using the out-of-sample forecasts of the excess market returns based on a lagged PLS indices. All of the predictors and regression slopes are estimated recursively using the data available at the forecast formation time $t$. CER gain is the annulized certainty equivalent return gain for the investor. The portfolio weights are estimated recursively, using the data available at the forecast formation time $t$. The sample period is from 1986Q2 to 2015Q4. The initial training period is from 1986Q2 to 1999Q4 and the out-of-sample evaluation period is from 2000Q1 through 2015Q4.

<table>
<thead>
<tr>
<th></th>
<th>CER gain (%)</th>
<th>CER gain in up market (%)</th>
<th>CER gain in down market (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Risk aversion $\gamma = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$PLSC$</td>
<td>3.52</td>
<td>1.93</td>
<td>16.21</td>
</tr>
<tr>
<td>$PLSE$</td>
<td>-1.38</td>
<td>-9.77</td>
<td>31.03</td>
</tr>
<tr>
<td>$PLSS$</td>
<td>2.48</td>
<td>-0.91</td>
<td>18.62</td>
</tr>
<tr>
<td>Panel B: Risk aversion $\gamma = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$PLSC$</td>
<td>8.78</td>
<td>2.03</td>
<td>10.31</td>
</tr>
<tr>
<td>$PLSE$</td>
<td>0.57</td>
<td>-13.2</td>
<td>14.69</td>
</tr>
<tr>
<td>$PLSS$</td>
<td>4.14</td>
<td>-4.02</td>
<td>7.28</td>
</tr>
<tr>
<td>Panel C: Risk aversion $\gamma = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$PLSC$</td>
<td>7.91</td>
<td>1.80</td>
<td>6.62</td>
</tr>
<tr>
<td>$PLSE$</td>
<td>0.19</td>
<td>-11.6</td>
<td>9.25</td>
</tr>
<tr>
<td>$PLSS$</td>
<td>4.40</td>
<td>-3.03</td>
<td>4.80</td>
</tr>
</tbody>
</table>
This table provides in-sample estimation results for the predictive regression \( R_i^j = \alpha_j + \beta_j PLS_{t-1}^k + \epsilon_i^j \), where \( R_i^j \) is the quarterly portfolio excess returns. \( PLS_{t-1}^k \) (k = C, E, S) is one of the PLS indices. Panel A to E report the results of predictive regressions for five portfolios formed by PP&E scaled by market value, analyst forecast error, analyst forecast dispersion, size, and book-to-market ratio, respectively. Std denotes the standard deviation of the portfolio quarterly returns. Exp Std denotes \( \sqrt{R^2 \times \text{Std}^2} \). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from 1986Q2 through 2015Q4.

<table>
<thead>
<tr>
<th>Panel A: Property, plant, and equipment scaled by market value of the firm</th>
<th>Corporate index</th>
<th>Macroeconomic index</th>
<th>Sentiment index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLS(^C)</td>
<td>t-value</td>
<td>( R^2(%) )</td>
</tr>
<tr>
<td>Transparent</td>
<td>1.54***</td>
<td>2.14</td>
<td>3.78</td>
</tr>
<tr>
<td>2</td>
<td>2.19***</td>
<td>3.20</td>
<td>8.03</td>
</tr>
<tr>
<td>3</td>
<td>2.55***</td>
<td>3.29</td>
<td>8.47</td>
</tr>
<tr>
<td>4</td>
<td>2.60***</td>
<td>3.21</td>
<td>8.10</td>
</tr>
<tr>
<td>Opaque</td>
<td>3.40***</td>
<td>3.37</td>
<td>8.84</td>
</tr>
<tr>
<td>Panel B: Analyst forecast error</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparent</td>
<td>2.25***</td>
<td>3.42</td>
<td>9.11</td>
</tr>
<tr>
<td>2</td>
<td>2.65***</td>
<td>3.41</td>
<td>9.06</td>
</tr>
<tr>
<td>3</td>
<td>2.41***</td>
<td>3.04</td>
<td>7.30</td>
</tr>
<tr>
<td>4</td>
<td>2.59***</td>
<td>2.82</td>
<td>6.35</td>
</tr>
<tr>
<td>Opaque</td>
<td>3.06***</td>
<td>2.56</td>
<td>5.30</td>
</tr>
<tr>
<td>Panel C: Analyst forecast dispersion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparent</td>
<td>2.27***</td>
<td>3.15</td>
<td>7.81</td>
</tr>
<tr>
<td>2</td>
<td>2.57***</td>
<td>3.60</td>
<td>9.98</td>
</tr>
<tr>
<td>3</td>
<td>2.65***</td>
<td>3.40</td>
<td>8.99</td>
</tr>
<tr>
<td>4</td>
<td>2.39***</td>
<td>2.81</td>
<td>6.31</td>
</tr>
<tr>
<td>Opaque</td>
<td>2.74***</td>
<td>2.22</td>
<td>4.06</td>
</tr>
<tr>
<td>Panel D: Size portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparent</td>
<td>2.68***</td>
<td>3.80</td>
<td>10.97</td>
</tr>
<tr>
<td>2</td>
<td>2.87***</td>
<td>3.38</td>
<td>8.91</td>
</tr>
<tr>
<td>3</td>
<td>2.97***</td>
<td>3.34</td>
<td>8.72</td>
</tr>
<tr>
<td>4</td>
<td>3.13***</td>
<td>3.25</td>
<td>8.29</td>
</tr>
<tr>
<td>Opaque</td>
<td>3.22***</td>
<td>3.01</td>
<td>7.17</td>
</tr>
<tr>
<td>Panel E: Book-to-market portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparent</td>
<td>2.25***</td>
<td>2.66</td>
<td>5.71</td>
</tr>
<tr>
<td>2</td>
<td>2.16***</td>
<td>2.88</td>
<td>6.61</td>
</tr>
<tr>
<td>3</td>
<td>2.36***</td>
<td>3.26</td>
<td>8.34</td>
</tr>
<tr>
<td>4</td>
<td>2.33***</td>
<td>3.13</td>
<td>7.74</td>
</tr>
<tr>
<td>Opaque</td>
<td>3.02***</td>
<td>3.81</td>
<td>11.05</td>
</tr>
</tbody>
</table>
Table 8: Forecasting cash flows and discount rates with PLS indices

This table reports in-sample estimation results for the bivariate predictive regressions

\[ Y_t = \alpha + \beta PLS_{t-1}^k + \phi D/P_{t-1} + \epsilon_t, \quad Y = D/P, DG, EG, GDPG, \quad k = C, E, S \]

where \( D/P_t \) is the log dividend-price ratio on the S&P 500 index at the end of year \( t \), \( DG_t \) is the annual log dividend-growth rate on the S&P 500 index from year \( t-1 \) to \( t \), \( EG_t \) is the annual log earning growth rate on the S&P 500 index from year \( t-1 \) to \( t \), \( GDPG_t \) is the annual log real GDP growth rate from year \( t-1 \) to \( t \), \( PLS_{t-1}^k \) (\( k = C, E, S \)) is one of the PLS indices at the end of year \( t-1 \), and \( \epsilon_t \) is an error term. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, based on one-sided p-values. The sample period is from 1987 through 2015.

<table>
<thead>
<tr>
<th>( Y_t )</th>
<th>( \beta )</th>
<th>t-stat</th>
<th>( \phi )</th>
<th>t-stat</th>
<th>( R^2 ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Corporate PLS index ( PLS^C )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D/P</td>
<td>-0.03</td>
<td>-0.76</td>
<td>0.97***</td>
<td>7.81</td>
<td>85.65</td>
</tr>
<tr>
<td>DG (%)</td>
<td>-0.75</td>
<td>-0.29</td>
<td>4.61</td>
<td>0.60</td>
<td>1.83</td>
</tr>
<tr>
<td>EG (%)</td>
<td>33.89***</td>
<td>2.62</td>
<td>-59.57*</td>
<td>-1.56</td>
<td>23.23</td>
</tr>
<tr>
<td>GDPG (%)</td>
<td>1.15**</td>
<td>2.32</td>
<td>-2.67**</td>
<td>-1.82</td>
<td>17.16</td>
</tr>
<tr>
<td><strong>Panel B: Macroeconomic PLS index ( PLS^E )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D/P</td>
<td>0.00</td>
<td>0.15</td>
<td>0.88***</td>
<td>10.26</td>
<td>85.34</td>
</tr>
<tr>
<td>DG (%)</td>
<td>3.33**</td>
<td>2.01</td>
<td>-2.53</td>
<td>-0.52</td>
<td>14.79</td>
</tr>
<tr>
<td>EG (%)</td>
<td>11.14</td>
<td>1.14</td>
<td>4.40</td>
<td>0.15</td>
<td>7.60</td>
</tr>
<tr>
<td>GDPG (%)</td>
<td>0.20</td>
<td>0.55</td>
<td>-0.22</td>
<td>-0.20</td>
<td>1.19</td>
</tr>
<tr>
<td><strong>Panel C: Sentiment PLS index ( PLS^S )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D/P</td>
<td>0.03</td>
<td>1.04</td>
<td>0.93***</td>
<td>11.69</td>
<td>85.91</td>
</tr>
<tr>
<td>DG (%)</td>
<td>-2.79**</td>
<td>-1.77</td>
<td>-0.89</td>
<td>-0.19</td>
<td>12.06</td>
</tr>
<tr>
<td>EG (%)</td>
<td>-17.51**</td>
<td>-2.00</td>
<td>-0.94</td>
<td>-0.04</td>
<td>15.90</td>
</tr>
<tr>
<td>GDPG (%)</td>
<td>-0.20</td>
<td>-0.58</td>
<td>-0.17</td>
<td>-0.16</td>
<td>1.33</td>
</tr>
</tbody>
</table>
Table 9: Cross-sectional relation between stock-return predictability and earning-growth predictability with corporate predictors

Panel A reports the predictive regression results for $EG_t^j = \alpha_j + \phi_j PLS_{t-1}^C + \epsilon_t^j$, $j = 1, ..., 5$, where $EG_t^j$ is the annual earning-growth rate for asymmetric information portfolio $j$ constructed following Cochrane (2008, 2011). Panel B reports the estimation results for the cross-sectional linear regression $\beta_j = \alpha + g\phi_j + \epsilon_j$, where $\beta_j$ is the predictive regression slope coefficient of information asymmetry portfolio $j$’s annualized excess return on $PLS^C$ (in Table 7). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period is from 1987 through 2015.

<table>
<thead>
<tr>
<th>Portfolios constructed based on PPE scaled by Market value</th>
<th>Analyst forecast error</th>
<th>Analyst forecast dispersion</th>
<th>Size</th>
<th>Book-to-market ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Forecasting earning-growth of portfolios</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\phi_j$, t-stat, $R^2$(%)</td>
<td>$\phi_j$, t-stat, $R^2$(%)</td>
<td>$\phi_j$, t-stat, $R^2$(%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparent</td>
<td>2.77, 1.07, 4.09</td>
<td>0.96, 0.96, 3.22</td>
<td></td>
<td>2.78, 1.11, 4.34</td>
</tr>
<tr>
<td>2</td>
<td>2.17*, 1.49, 7.61</td>
<td>2.46*, 1.43, 7.06</td>
<td></td>
<td>3.56***, 2.58, 19.80</td>
</tr>
<tr>
<td>3</td>
<td>4.06***, 3.33, 29.12</td>
<td>4.73**, 2.31, 16.56</td>
<td></td>
<td>4.50**, 2.24, 15.63</td>
</tr>
<tr>
<td>4</td>
<td>4.80***, 2.99, 24.87</td>
<td>6.67**, 2.34, 16.82</td>
<td></td>
<td>5.61**, 1.92, 12.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Panel B: Cross-sectional regression, $\beta_j = \alpha + g\phi_j + \epsilon_j$</strong></th>
<th>g, t-stat, $R^2$(%)</th>
<th>g, t-stat, $R^2$(%)</th>
<th>g, t-stat, $R^2$(%)</th>
<th>g, t-stat, $R^2$(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g$, t-stat, $R^2$(%)</td>
<td>1.26**, 3.46, 79.95</td>
<td>0.29**, 2.19, 61.60</td>
<td>0.22*, 1.82, 54.49</td>
<td>0.28**, 3.50, 80.30</td>
</tr>
<tr>
<td></td>
<td>0.62***, 5.17, 89.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10: Forecasting future macroeconomic conditions with PLS indices

This table reports results of the predictive regression $\Delta Y_t = \alpha + \beta PLS^k_{t-1} + \epsilon_t$, where $\Delta Y_t$ is the change in the recession probability (SRP), industrial production growth (IPG), Treasury bill rate (TBL), default yield spread (DFY), implied volatility index (VIX), or Chicago Fed National Activity Index (CFNAI) from quarter $t-1$ to $t$; $PLS^k_{t-1}$ ($k = C, E$) is one of the PLS indices in quarter $t-1$. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample period $t$ is from 1986Q2 through 2015Q4.

<table>
<thead>
<tr>
<th>$\Delta Y_t$</th>
<th>$PLS^C$</th>
<th>$PLS^E$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$</td>
<td>\beta</td>
</tr>
<tr>
<td>SRP</td>
<td>-1.98**</td>
<td>-2.13</td>
</tr>
<tr>
<td>IPG</td>
<td>0.03**</td>
<td>2.15</td>
</tr>
<tr>
<td>TBL</td>
<td>0.03</td>
<td>0.79</td>
</tr>
<tr>
<td>DFY</td>
<td>-0.06***</td>
<td>-2.77</td>
</tr>
<tr>
<td>VIX</td>
<td>-2.51***</td>
<td>-3.99</td>
</tr>
<tr>
<td>CFNAI</td>
<td>0.13***</td>
<td>2.10</td>
</tr>
</tbody>
</table>