

# **Predicting Stock Market Returns in the U.S.: Evidence from an Average Correlation Approach**

Xi Yang Li

Dongwu Business School, Department of Accounting, Soochow University, Suzhou,  
China and Department of Accounting, Finance and Economics,  
Griffith Business School, Griffith University, Nathan, Australia

Bin Li and Tarlok Singh

Department of Accounting, Finance and Economics,  
Griffith Business School, Griffith University, Nathan, Australia, and

Kan Shi

Surrey Business School, Surrey International Institute, University of Surrey,  
Guildford, UK

---

## **Abstract**

This study draws on a less explored predictor – the average correlation of pairwise returns on industry portfolios – to predict stock market returns in the United States. The model is estimated on monthly data using stock returns on a comprehensive set of 48 Fama-French industry portfolios for the period July 1963 to December 2018. The study finds that a two-month lagged average correlation predicts future stock market returns at a high level of significance and accuracy. Compared to conventional predictors, average correlation demonstrates better predictive ability in both short and long time horizons. These results are supported by multiple robustness checks. The findings of the study have important implications for assisting both financial market participants and policymakers to make more accurate predictions about future stock market movements.

---

**Keywords:** Stock Market Returns Predictability; Average Correlation; Economic Indicators; Financial Predictors; U.S. Market.

**JEL Classifications:** G10; G11; G12; G17.

# **Predicting Stock Market Returns in the U.S.: Evidence from an Average Correlation Approach**

## **1. Introduction**

The well-known efficient market theory posits that a perfect stock market adjusts each stock to its fair value in a timely manner, despite investors making potentially abnormal gains in the stock market by using various trading strategies. The kernel of these trading strategies is to form a portfolio in which the stocks are selected based on the forecast of their future performance (Hafer & Hein, 2007). Analysts and practitioners have attempted to make reliable and accurate predictions on future stock market returns (SMRs), possibly from the day when the stock market was established. Although the methods used for SMR predictions have become increasingly popular, Goyal and Welch (2003) note that the reliability and precision of the prediction itself have not yet improved much.

Indeed, long and heated debates have revolved around the issue of predicting stock market returns: sceptics even contend that future stock returns are unpredictable as they change randomly. The empirical evidence instead suggests that future stock returns are predictable with a reasonable level of precision. The conventional financial ratios and economic indicators seem to have reached their full potential as predictors and therefore researchers are shifting their attention to exploring other prospective predictors. During the last two decades, the research has centred on two issues: the selection of the most effective predictors and the construction of a predictive model with maximal predictive power. Recent studies have used predictors such as variance risk premia (Bollerslev et al., 2014), stock variance and stock pressure (Bahrami et al., 2017), price pressure and change in volume (Charles et al., 2017) and long memory in the returns (Tan et al., 2014; Lahmiri, 2015) to predict SMRs.

Some studies have argued that the trend of future SMRs could be revealed by continuously observing the co-movement of returns on individual stocks (i.e. the correlation between returns on individual stocks), as the returns on each stock share a common component that changes with the SMRs (Aneja et al., 1989; Ang & Chen, 2002; Forbes & Rigobo, 2002). Using such correlation to predict future SMRs is theoretically appealing since it is endogenous to the stock market. The correlation measures the co-movement of returns on

individual stocks, rather than measuring the return itself, and it is thus less influenced by the idiosyncratic characteristics of individual stocks. While financial ratios and economic indicators are exogenous to the stock markets, their predictive powers could be weak and conditional (Welch & Goyal, 2008; Bahrami et al., 2017).

Pollet and Wilson (2010) have recently developed a new average correlation approach to predict stock market returns. They find that the average correlation of pairwise returns on the largest 500 individual stocks in the U.S. market effectively predicts future SMRs with a one-month lag. The empirical findings of the study by Pollet and Wilson, however, suffer from several limitations. First, the use of value-weighted average correlation of returns on individual stocks in Pollet and Wilson has a significant sample selection bias, in that the sample includes only the stocks with largest market capitalisation. The industry returns encompass a substantial portion of returns on all stocks in the market with less size distortions and are therefore more effective in capturing the co-movement of all the stocks. Second, Pollet and Wilson predict stock market returns one month ahead with average correlation at time  $t$ . Such an analysis ignores the possible effects of longer time lags on the relationship between predictor and future returns. Third, Pollet and Wilson do not take into account investors' behavioural perspectives. Investors' behaviour could potentially cause co-movement of returns on all assets by shifting their level of investment (Baker & Wurgler, 2006; Barberis et al., 1998; Ho & Hung, 2009). It is important to consider the effect of investors' sentiments and thus to incorporate the behavioural aspects in the predictive model. Fourth, the Pollet and Wilson study predates the global financial crisis and, as such, the predictive power of their average correlation approach has not been interfaced with such an unusual shock. A number of researchers have established a connection between time-varying stock market predictability and disruptive events such as market crashes (Barro & Ursúa, 2017) or financial crises (Berkman et al., 2011). Financial crises, similar to investors' sentiments, could coincide with co-movements of returns on all assets, as rational investors would take similar trading strategies during a crisis period (Berkman et al., 2011).

The use of average correlation, as a measure of the co-movement of all stock returns, is conceptually appealing, despite some of the limitations of the Pollet and Wilson study. The existence of a systematic factor that is independent of individual stock self-variance, an idiosyncratic factor, and that is correlated between different stocks, is the foundation of the modern portfolio theory or even the entire modern finance theory, which has evolved overtime since the early study by Markowitz (1952). Average correlation is an instinctive

proxy of pairwise correlations between all asset returns, especially when observing “all” asset returns is virtually impossible. The average correlation approach follows the recent trends in research, exploring how the internal linkages in stock market (i.e. the linkage of individual stock returns and variances or the linkage between stock market returns and variance) could affect overall returns and predictability.

This study uses the average correlation approach of Pollet and Wilson (2010) and predicts the SMRs in the United States. The study uses the stock returns on a comprehensive set of 48 Fama-French industry portfolios that capture the industry effects and reduce the confounding size effects. The model is extended to examine the effects of a longer lag structure of one-month to four-month lags and to control for the effects of a number of variables — average variance (AV), cyclically adjusted price-to-earnings ratio (CAPE), term spread (TS), default spread (DS), risk-free rate returns ( $R_f$ ) and lagged excess market returns ( $R_S$ ).

The study contributes to the literature on three counts. First, the study uses industry portfolio returns — as compared to individual stock returns used in Pollet and Wilson (2010) — in constructing average correlation. When stock market becomes more volatile on returns, the individual stocks are more diverse on their performance; the comovement between individual stock returns might be dominated by the idiosyncratic component, which may not have any implications for future SMRs. Using the industry portfolio returns can potentially reduce such an effect by a large extent and thus can provide more reliable estimates. Second, the effects of business cycles could be better identified in a long sample period and through several sub-sample tests. This study uses a dataset which spans the period from 1963:7 to 2018:12. This long sample period covers multiple phases of business cycles. The daily data are used to compute the monthly and equally weighted average correlation of returns on 48 Fama-French industry portfolios. Third, previous studies have often ignored the use of investors’ sentiments in their prediction models, while investors’ irrational decisions could have an important impact on expected returns (Huang et al., 2015). This study extends the analysis and incorporates investors’ sentiments in the model.

The study finds that average correlation outperforms lagged excess market returns and other predictors in predicting future SMRs in both short and long time horizons. The results have important implications for market practitioners, academics and policymakers. Since average correlation is a significant predictor of future SMRs, investors should consider

including average correlation in their forecasting models, along with the traditional financial ratios and economic indicators.

The rest of the study proceeds as follows. Section 2 reviews the relevant literature on the SMR predictability. Section 3 specifies the model and discusses the data used in the study. Section 4 reports and discusses the empirical results. Section 5 concludes the study.

## **2. Stock Returns Predictability: A Literature Review**

Predicting future SMRs amounts to estimating future gains based on current information. Financial markets were once believed to be extremely efficient and all information was incorporated into price immediately (Malkiel, 2004). The current prices and returns had long been considered to contain information about future prices and returns (Fama, 1970). Fama and Schwert (1977), however, challenged this common belief, reporting a negative relationship between common stock returns and expected component of inflation. This is considered to be early evidence of market inefficiency, as the relationship failed to incorporate the known information. Research on SMR predictability has grown rapidly, spawning three strands that adopt different approaches to estimate future stock market returns.

The first strand of research proposes that financial ratios measure the underlying value of a company, which should be reflected in its future returns. Various financial ratios and economic indicators have been used in the literature to predict SMRs, including book-to-market ratio, dividend yield, earnings-price ratio and interest rate (Fama & Schwert, 1977; Campbell, 1987; Campbell & Shiller, 1988; Fama & French, 1988; Kothari & Shanken, 1997; Campbell & Viceira, 2002; Lewellen, 2004; Menzly et al., 2004; Campbell & Yogo, 2006). Financial ratios are firm-specific and ready-to-use predictors; they are popular among academics and market practitioners. The heterogeneity of financial ratios, due to their firm-specific characteristics, was initially regarded as an advantage, but was later considered to be too diverse to form a common trend (Campbell & Thompson, 2008; Pástor & Stambaugh, 2009). Although the use of financial ratios in SMR prediction models has been marginalised — from these ratios being main predictors to being benchmarks and control variables — these ratios are still used quite extensively in the literature (Li & Zhang, 2010; Pollet & Wilson, 2010; Jordan et al., 2014).

The second strand of research focuses on economic environment in which the companies operate. If an economy is forecast to enter a recession, the stock market price will invariably fall, and the investors will adjust their expected returns accordingly. Many

economic indicators have been associated with future SMRs: the inflation rate (Siklos & Kwok, 1999), the term and default spread on bonds (Fama & French, 1989), the money stock (Jensen & Johnson, 1993), the interest rates (Ang & Bekaert, 2007), the aggregate output (Cooper & Priestley, 2005), the investment to capital ratio (Cochrane, 1991), the unemployment rate (Boyd et al., 2005), the consumption-to-wealth ratio (Lettau & Ludvigson, 2001), and the labour income to total income ratio (Menzly et al., 2004). Whereas financial ratios are heterogeneous and firm-specific, economic indicators are important to all activities within the economy, regardless of their effects on financial markets.

The third strand of research includes investors' behaviours in the asset pricing model. It assumes irrationality and uses investors' behaviours to explain the variance in SMRs. Menzly and Ozbas (2010) argue that returns on two economically related industries, such as suppliers and customers, should have cross-predictive power. They introduce investors' expertise and knowledge in the model and find that the predictability declines if the number of informed investors increases. Investors' sentiments have attracted most attention in behavioural finance research, which investigates the effects of general emotion shared by a large group of investors (Baker & Wurgler, 2006). Barberis, Shleifer, and Vishny (1998) classify investors' attitudes into categories such as overconfidence, conservatism or representativeness. Schmeling (2009) proposes consumer confidence as a measure of investor sentiment and notes negative average correlations between sentiment and SMRs across 18 industrialised countries. Although investor sentiment is difficult to quantify, its close association with stock price and return needs to be considered in the SMR prediction model.

Another strand of literature has concentrated on intra stock market co-movements. Kallberg and Pasquariello (2008) report a significant and excess co-movement of 82 industry indexes in the U.S. stock market; although the co-movement is not directly used to predict future stock market returns, and it has a negative correlation with both market volatility and level of short-term interest rates. Liu and Mei (1992) argue that there is "a single systematic factor affecting returns on all assets" (p.415), implying a potential co-movement of returns on individual assets.

The conventional literature has thus identified the use of financial ratios, economic indicators and investors' sentiments to predict future SMRs. Pollet and Wilson (2010) have employed their average correlation approach to predict future SMRs. The economic theory of using average correlation as a variable to predict SMRs originates from the capital asset

pricing model (CAPM), the basis of many SMRs prediction studies (e.g., Beyhaghi & Hawley, 2013; Noussair & Tucker, 2013; Chaibi et al., 2015). It starts with an intuitive assumption: when systematic risk rises, risk-averse investors will demand a higher risk premium to offset the possible volatility; in other words, it gives rise to expected returns. Consistent with this principle, various models have been developed to identify the variance-in-mean relationship for SMRs (Campbell, 1987; Harvey, 2001).

Pollet and Wilson (2010) have developed a theoretical model and examined the effectiveness of average correlation and average variance in predicting SMRs (Pollet and Wilson, 2010, Equation (7), p.367). They show that average correlation is a reliable predictor of future stock market returns, while average variance has no effect on or has even been negatively correlated with future SMRs. These findings are in line with the evidence provided by Bali, Cakici, Yan and Zhang (2005), which suggests that the predictive power of average variance is conditional on datasets and measures. The use of average correlation as a prospective predictor of future SMRs has relatively remained unexplored so far.

### 3. Model Specification

This study draws on the study by Pollet and Wilson (2010) and uses the equally weighted average correlation of returns on a comprehensive set of 48 Fama-French industry portfolios to predict SMRs in the United States. While the use of value-weighted average correlation is useful, there is no evidence to date to support that the correlation of stock returns on larger companies is more likely to reveal the movement of SMRs. The study envisages that an industry portfolio shares a common component which changes with SMRs, such that the trend of future SMRs could be revealed by the co-movement of industry portfolio returns, not only of individual stock returns. There has been no consensus so far with regard to the use of optimal time lag in average correlation of industry portfolio returns to predict SMRs. The study uses up to 4-month time lag of average correlation so as to sufficiently capture the information contained in the data. The model employed to explain the effects of lagged average correlation on stock market returns is specified as follows:

$$R_{s,t} = \mu + \sum_{i=1}^4 \beta_i AC_{t-i} + \beta_s R_{s,t-1} + \varepsilon_t \quad (1)$$

The equally weighted pairwise average correlation (AC) of 48 Fama-French industry portfolios in equation (1) is calculated as  $AC_t = \sum_{i=1}^n \sum_{j=1, j \neq i}^n \rho_{ij,t} / N$ , where  $n = 48$  and  $\rho_{ij,t}$  denotes the correlation between industry  $i$  and  $j$  ( $i \neq j$ ).  $N = \frac{n(n-1)}{2} = \frac{48 \times 47}{2} = 1128$  is

the number of pairs of returns on any two industries.  $R_{s,t}$  represents excess market return at time  $t$  and it is calculated as the difference between market return and risk-free rate.  $R_{s,t-1}$  denotes one-month lagged market return.  $AC_{t-i}$  indicates average correlation from one to four lags ( $i = 1, 2, 3, 4$ ).

Model (1) is extended to control for the effects of other predictors, such as average variance (AV), CAPE, TS and DS, on future stock market returns.

$$R_{s,t} = \mu + \sum_{i=1}^4 \beta_i AC_{t-i} + \beta_k X_{k,t-1} + \varepsilon_t \quad (2)$$

where  $AC_{t-i}$  refers to the  $i$ -month lagged average correlation.  $X_{k,t-1}$  represents control variables — lagged excess market returns, risk-free rate, AV, CAPE, TS and DS. These financial ratios and economic variables are included in the model to control for their effects and to reduce the possible model mis-specification bias. All these additional predictors are used with a one-month lag.

### 3.1 Variable Description and Data

Models (1) and (2) are estimated using monthly data from 1963:7 to 2018:12. All data are recorded daily and converted to monthly frequency, except for quarterly data, which are converted to monthly data through linear interpolation. The data are sourced from four databases: Kenneth R. French's Data Library, Robert Shiller's Database, Federal Reserve Economic Database (FRED St. Louis) and the Baker and Wurgler Sentiment Index.

The dependent variable, excess market returns, is the value-weighted return of all the firms listed on the NYSE, AMEX and NASDAQ. These data are sourced from the Centre for Research in Security Prices (CRSP) database. The risk-free rate is measured in term of one-month Treasury-bill rate. The data on Fama-French 48 industries are sourced from the French Data Library, which assigns each NYSE, AMEX and NASDAQ stock to a particular industry portfolio at the end of June every year, based on its four-digit SIC code at that time. The data on excess market returns, risk-free rate, size premium (Small minus Big), value premium (High minus Low), Robust minus Weak (RMW), and Conservative minus Aggressive (CMA) are all retrieved from the French Data Library.

Average variance, which is the mean of variances in all 48 industry portfolios, is also included in the model as it has a close relationship with average correlation. It is calculated as the mean of monthly variances of all 48-industry portfolio daily returns. The industry portfolio returns are retrieved from the French Data Library. CAPE is included as one of the

predictors in the model. CAPE is retrieved from Robert Shiller’s database, and it is adjusted using consumer price index (CPI)—with a CPI base set at the first quarter of 2016. In addition to financial ratios, two economic indicators, TS and DS, are included in the model and are obtained from the Federal Reserve Economic Database (FRED St. Louis). TS measures the difference between the interest rate of the same security with different maturities; DS denotes investors’ expectation of default risk, which depends on future economic conditions. Market sentiments are considered to capture the effects brought by noise traders (De Long et al., 1990). Two indices, Market Sentiment One and Market Sentiment Two, are retrieved from the Baker and Wurgler Sentiment Indices. Details of the definition and data source for each variable are provided in Table 1.

[Inset Table 1 here]

#### **4. Empirical Results and Analysis**

The empirical analysis begins by presenting summary statistics for all the variables included in the model (Table 2). The mean of the average correlation between all pairs of the returns on 48 industry portfolios is 0.005, with a standard deviation of 0.002. The mean of excess market returns is 0.005, with a standard deviation of 0.044 and the minimum and maximum values of -0.02 and 0.16, respectively, indicating that excess market returns are relatively volatile. In terms of benchmark variables, CAPE and DS are less volatile, while TS has a standard deviation of 1.68, which is larger than its mean of 1.06. The skewness and kurtosis of these variables are quite high. The highest values of skewness and kurtosis are 7.30 and 72.36, respectively, for the average variance. Although the values of skewness and kurtosis are high, these features are prevalent in the financial literature and are shared by many studies (e.g., Goyal & Welch, 2003; Lewellen, 2004). The autocorrelation of each variable is also reported in Table 2. Average correlation is fairly persistent, with a first-order autocorrelation of 0.61, whereas the excess market returns are much less persistent, with a first-order autocorrelation of 0.07.

[Insert Table 2 here]

To examine the possible problem of multicollinearity among the model variables, the pairwise correlation matrix of variables is constructed (Table 3). The results show that the pair-wise correlation between lagged average correlation and SMRs is 0.09 and the one between lagged average correlation and lagged average variance is 0.37. TS, among all predictors, is the only one that has a higher correlation with future SMRs. CAPE has a

moderate correlation (-0.37) with average correlation, indicating that a higher average correlation is associated with a lower price-earnings ratio. Risk-free rate and CAPE negatively correlate with excess market returns, with correlations being -0.05 and -0.04, respectively. The highest correlation, between TS and risk-free rate, is -0.615: these two are control variables and have low correlation with future SMRs and average correlation. The results from the correlation matrix suggest that there is no multicollinearity problem in general.

[Insert Table 3 here]

The ordinary least squares (OLS) estimates of the regression of excess market returns on various combinations of one-month to four-month lagged average correlations (Model 1 to Model 4), average variance (Model 1\* to Model 4\*) and one-month lagged excess market returns are reported in Table 4. Model (1) reports the estimates of excess market returns on two-month lagged average correlation. The two-month lagged average correlation is a significant predictor of excess market, with a  $t$ -statistic of 2.63. Model (2) includes the lagged excess market returns in the estimation. The lagged excess market returns statistically fail to predict excess market returns; however, the adjusted  $R^2$  of this regression increases from 0.7% to 1.1%. Model (3) includes all the lagged average correlations (from one-month to four-month) to estimate the excess market returns. Once again, the two-month lagged average correlation remains the most significant predictor, although the  $t$ -statistic of its coefficient declines to 2.41. The adjusted  $R^2$  of the regression is slightly higher than the adjusted  $R^2$  in Model (1), indicating that the inclusion of two-month lagged average correlations increases the explanatory power of the model. In Model (4), which includes all the lagged average correlations and the lagged excess market returns, the two-month lagged average correlation is the only significant predictor, with a  $t$ -statistic of 2.01. Among the four models, only the two-month lagged average correlation produces the significant results ( $t$ -statistic = 2.69)<sup>1</sup>. While adjusted  $R^2$  of the model is considerably low, this is consistent with prior literature (e.g., Lettau and Ludvigson, 2001; Pollet and Wilson, 2010).

The average correlation of industry portfolio returns captures the trend of future SMRs with a longer time lag (two-month lag) than the average correlation of individual stocks used in Pollet and Wilson (2010) (one-month lag). The average correlation of industry

---

<sup>1</sup> Two-month lagged average correlation ( $AC_{t-2}$ ) is the most significant predictor among lagged average correlations. This paper therefore uses only  $AC_{t-2}$  in further tests; other lagged average correlations are not considered.

portfolios, however, seems more useful, even though it predicts stock market returns with a longer time lag. The average correlation of industry portfolios tends to smooth out the noise that could be contained in the random and transitory co-movements of individual stock returns.

Models (1\*) to (4\*) are re-estimated by replacing average correlation by average variance. Although Models (1\*) and (2\*) report only the results for two-month lagged average variance, the study has estimated the model with all one- to four-month lags (of average variance). The study found that no individual lag (of average variance) is statistically significant when tested separately. However, as presented in Models (3\*) and (4\*), one-month and two-month lagged average variances are significant in predicting future SMRs. The adjusted  $R^2 = 1.8\%$  is relatively high — adjusted  $R^2$  of around 1% is quite normal in such type of studies (see, for example, Huang et al., 2015). Consistent with average correlation, the two-month lagged average variance is used in the prediction models and the results show the insignificant effects of average variance.

[Insert Table 4 here]

Table 5 analyses the predictive ability of average correlation by controlling for average variance, CAPE, TS, DS, risk-free rate and lagged excess market returns. These results are obtained using Equation (2). Model (1) to Model (7) present the results of the estimation of excess market returns on each individual variable. Among all the predictors, benchmarks and control variables, average correlation and TS are the only two predictors, which are statistically significant with  $t$ -statistics of 2.63 and 2.25, respectively. Model (8) to Model (13) estimate excess market returns on various combinations of average correlation and control variables. Average correlation is consistently significant, with the highest  $t$ -statistic of 2.88 and the lowest  $t$ -statistic of 2.07. Model (14) includes all the variables in the estimation. Average correlation is consistently significant, with a  $t$ -statistic of 2.64 in the model with all the variables included. These results indicate that average correlation remains the most significant predictor among all predictors (with  $t$ -statistics of at least 2.61). The average correlation and term spreads in Model (11) are the two variables that are statistically significant. The results reported in Tables 4 and 5 suggest that the average correlation with two-month lag significantly predict the future excess market returns.

[Insert Table 5 here]

This study also follows Lettau and Ludvigson (2001) to perform predictions for excess market returns over different time horizons (Table 6). When lagged average correlation, average variance (at time  $t-2$ ) and other lagged predictors (at time  $t-1$ ) are the individual predictors, the dependent variable is the excess market returns for the holding periods of one to 36 months (time  $t$  to time  $t+k$ ,  $k = 1$  to 36), with overlapping observations. The predictive ability of each predictor is therefore examined for time spanning up to 36 months. The results for the estimation of long-horizon excess market returns on average correlation are reported in row 1 (Table 6). Up to 6 months, average correlation remains significant, with  $t$ -statistics ranging between 1.72 and 4.65. Average correlation can significantly predict 36-month-ahead excess market returns ( $t$ -statistic = 3.29). The adjusted  $R^2$  of the regression is almost five times as high as the adjusted  $R^2$  for the model estimated with one-month-ahead excess market returns. Other independent variables that could predict long-horizon excess market returns are AV, TS, and CAPE, which retain their predictive ability of estimating future excess market returns for up to 36 months. AV could predict SMRs in  $k = 12$  and 24 time horizons, while TS could predict SMRs in  $k = 1, 2, 12, 24$  and 36 time horizons. Considering the significance level ( $t$ -statistic) and explanatory power (adjusted  $R^2$ ), average correlation outperforms all other predictors in a long time horizon.

[Insert Table 6 here]

Many SMR prediction models have ignored investors' sentiments, notwithstanding the fact that investors' behaviours could affect the estimation of and the expectations for future SMRs. To bridge this research gap and examine the effects of the investor behaviours and market downsides, the market sentimental index and recession period are included in the model. Table 7 reports the results of the regression of excess market returns on average correlation and its interaction with market sentiments. Row 1 indicates that the interaction of average correlation with recession period is significant but negative, with a  $t$ -statistic of -2.40. This negative interaction suggests that lower average correlation in the recession period could predict higher excess market returns. In contrast, there is no evidence that average correlation could interact with market sentiments in predicting future excess market returns. The interaction of market sentiments and average correlation is insignificant. This finding underpins other recent research findings that the business cycle has an impact on the predictive ability of average correlation and on the expected returns (e.g., Kenourgios et al., 2009; Samitas & Tsakalos, 2013; Velázquez & Smith, 2013).

[Insert Table 7 here]

The analysis is carried further, with sample period divided into three sub-periods: (1) the recession period (Recession = 1) and the rest (Recession = 0); (2) the above median (Market sentiment one = 1) and the below median (Market sentiment one = 0) of investor sentiment; and (3) the above median (Market sentiment two = 1) and the below median (Market sentiment two = 0) of investor sentiment. The means of the excess market returns in each period are compared to assess whether these phases make any difference to the analysis. Table 8 reports that the excess market returns in the recession period, on average, are 0.01 lower than those in the rest of the period, with a *t*-statistic of -2.55. The excess market returns are not statistically different in the different sentiment periods, no matter which measure is used.

The predictive power of average correlation is also tested for both recession and high investor sentiment periods. The results shed more light on the relationship between average correlation and investor sentiments. Average correlation is relatively low in recession periods, reflecting that stock performances differ during the recession period. The possible scenario is that good stocks could maintain their higher level of returns, while other stocks suffer a fall in their returns. In other words, having a higher or lower return depends more on the idiosyncrasy of any particular stock than on the movement of the systematic factor. Table 8 also shows that average correlation is significantly lower in a high investor sentiment period. Lower average correlation tends to lead to lower future SMRs as suggested by the main finding of this study. This could potentially explain the findings of previous studies that there is a negative relationship between investor sentiment and future SMRs. However, when controlled for average correlation, investor sentiment is no longer related to future SMRs as shown in Table 7.

[Insert Table 8 here]

Overall, the results provide evidence that average correlation effectively predicts the trend of future excess market returns. After controlling for the effects of financial ratios, economic indicators and investors' sentiments, the average correlation still predicts future excess market returns, in both short and long time (i.e. two year) horizons. The results thus lend empirical support to the hypothesis that the co-movement of returns on industry portfolios predicts future excess market returns.

#### 4.1 Robustness Check and Bootstrapping

Several robustness checks are performed to compute the critical values for the test statistics and to cross-examine the statistical significance of the model variables. Three types of tests are performed: first, using other measures of the co-movement of returns on industry portfolios, such as median and the 25 and 75 percentiles of the correlation; second, using returns other than excess market returns, such as individual returns on industry portfolios, returns on Fama-French 5 factors portfolios, and the 10 deciles and 4 quantiles of the excess market returns; third, using sub-sample periods instead of the entire sample period. The first set of tests substitutes the average correlation with the median, the 25 and the 75 percentiles of correlations of returns on industry portfolios and the difference between the average correlations of two adjacent periods.

Table 9 reports the results for the use of alternative measures of correlation with two-month lags to predict future excess market returns. The median, the 25 and the 75 percentiles of correlations are all significant predictors of future market returns, with  $t$ -statistics ranging from 2.41 to 2.63. The results show that the use of the co-movement of returns on industry portfolios to predict future SMRs is not sensitive to the selection of the measure. Still, average correlation has the highest coefficient and  $t$ -statistic ( $t$ -statistic = 2.63) and thus explains most of the changes in excess stock market returns among all the predictive variables.

[Insert Table 9 here]

Table 10 presents the outcomes of using average correlation to estimate the returns on 48 industry portfolios individually. In more than half the cases, average correlation predicts the individual industry return quite reliably, with  $t$ -statistics of the coefficients remaining more than 1.96. In particular, the coefficient of average correlation in the case of the “shipping containers” industry is statistically significant at the 1% level, with its  $t$ -statistic (3.71) being highest among all the industries. These results in general suggest that the average correlation predicts individual returns on industry portfolios.

[Insert Table 10 here]

Table 11 reports the results of the regression of returns on portfolios constructed using various factors. Average correlation could predict returns on portfolios constructed using single factors, such as size, book-to-market ratio, operating profitability, and investment. For

instance, average correlation estimates the returns on prior low return portfolio (Loser) with a  $t$ -statistic of 3.34. In contrast, the differences of returns on two portfolios constructed using a single factor (e.g., HML is the difference of the returns on the portfolios constructed using high and low book-to-market ratios) could not be estimated by average correlation, except for the difference on the returns on portfolios constructed using momentum ( $t$ -statistic = -2.60, adjusted  $R^2 = 1.2\%$ ). The results provide evidence that the average correlation could potentially be extended to predict the returns on various factor portfolios.

[Insert Table 11 here]

Table 12 estimates the model regressing excess market returns of intervals of 4 quantiles and 10 deciles on average correlation. Average correlation is a significant predictor for the lower half and the 40th decile of excess market returns, with  $t$ -statistics higher than 1.96 (equivalent to a significance level of 5%). This result indicates that lower excess market returns are more likely to be predicted by average correlation. To sum up, Tables 10, 11 and 12 demonstrate that average correlation still predicts the returns on the factor portfolios and on the lower quantile of the excess market returns, with a reasonable level of reliability and accuracy.

[Insert Table 12 here]

Pollet and Wilson (2010) have focussed only on the full-sample analysis, and have ignored the sub-sample period analysis, which may yield different results. To address this issue, the study investigates the predictive ability of average correlation during a number of sub-sample periods. Table 13 reports the estimates of the model for a 10-year and a 20-year period as well as for an earlier half and a later half of the sample period. In the period 1973:1–1982:12, the average correlation has good predictive ability for excess market returns, with a  $t$ -statistic of 3.77, and adjusted  $R^2$  of 4.6%. The results also provide significant support to the predictive ability of average correlation in the period from 1963:7 to 1992:12, with a  $t$ -statistic of 2.42 and an adjusted  $R^2$  of 0.8%. These results are similar to the result of the full sample ( $t$ -statistic = 2.63, adjusted  $R^2 = 0.7\%$ , Table 5). Welch and Goyal (2008) find that many conventional predictors have little predictive ability during the period from 1976 to 2007, due to the first oil price shock in the early 1970s.

This study finds that the significance of the predictor disappears when the model is estimated on a decade-by-decade basis, except for the period 1973:1–1982:12. This drop in significance could be ascribed to the loss of power of the model, given the relatively small-

sample period of only one decade. To make sure that the results obtained for the full-sample period are not solely driven by 1973:1–1982:12, the model is re-estimated by excluding this period from the full sample. The results suggest that the coefficient of two-month lagged average correlation (L2AC) of 0.024 (*t*-statistic 2.43) remains statistically significant at the 1% level (Table 13). These results are consistent with the results obtained from the model estimated for the full-sample period 1963:7–2018:12 (Table 5).

The results obtained from the model estimated for a longer sample period of 1963:7–2018:12 (with period 1973:1–1982:12 excluded) instead suggest that the coefficient of two-month lagged average correlation (L2AC) of 0.024 (*t*-statistic 2.43) continues to remain statistically significant at nearly 1% level of significance (Table 13). This suggests that the results are consistent with the results obtained from the model estimated for the entire period from 1963:7 to 2018:12 as reported in Table 5.

[Insert Table 13 here]

Bootstrapping, a common technique to address the issues such as heteroscedasticity of standard errors and possible modelling failure (Chernick, 2007), is performed to assess the small sample properties of the test statistics. The results from the bootstrapping simulations are presented in Table 14. The simulations are performed for a number of draws (up to 20,000 draws). The results indicate that the z-statistics converge after 2000 draws and the average correlation remains statistically significant.

[Insert Table 14 here]

## 5. Conclusions

The study has attempted to extend the previous literature by using the average correlation approach suggested by Pollet and Wilson (2010) to predict stock market returns in the United States. The model is estimated using monthly data for a long time horizon, from 1963:7 to 2018:12, for the portfolios comprising 48 Fama-French industries. The study finds that the two-month lagged average correlation of returns on individual industry portfolios, used individually and collectively with financial predictors and economic factors, predicts excess returns on the stock market in an effective manner. These results provide useful insights on how average correlation could reveal the changes in stock market returns. The results remain consistent across full-sample and some of the sub-sample and long-horizon tests as well as across various robustness checks.

The findings of the study have important implications for market practitioners, academics and policymakers. Market practitioners could include average correlation in their asset pricing models to improve the predictions for future trend in stock market returns. Investors could consider including average correlation in their forecasting models, along with financial ratios and economic factors. The academics could further explore the use of average correlation to improve the predictive powers of their models. The finding that recession periods have effects on the stock market returns would be useful for the policy makers. Investors could adjust their expected returns to a lower level when average correlation increases during recession. The understanding of the co-movement of returns on industry portfolios during recession would be useful for the formulation of policies aimed at ensuring stability of the financial markets.

Table 1: Variable description and data sources

Variable (Acronym)	Full Description	Summary Description	Data Source
Excess Market Returns ( $R_{s,t}$ )	Value-weighted return of all the firms listed on the NYSE, AMEX and NASDAQ (that have a Centre for Research in Security Prices share code of 10 or 11) adjusted by risk-free rate, which is measured by 1-month Treasury bill rate.	$R_{s,t} = R_{m,t} - R_f$	Kenneth R. French's Data Library
Cyclically Adjusted Price Earnings Ratio (CAPE)	The aggregate PE ratio of the stock market adjusted by the fraction of the CPI of the quarterly DP ratio. The base quarter is the first quarter of 2016.	$CAPE = \frac{earnings}{price} * \frac{CPI_t}{CPI_{16}}$	Robert Shiller's Database
Term Spread (TS)	The difference between the return on 10-year maturity government security and the return on 3-month maturity government security.	$TS = R_{Short} - R_{Long}$	Federal Reserve Economic Database (FRED St. Louis)
Default Spread (DS)	The difference between the average yields on a (closest to 20-year maturity) portfolio of corporate bonds rated Baa and a portfolio of bonds rated AAA by Moody's.	$DS = R_{AAA} - R_{Baa}$	Federal Reserve Economic Database (FRED St. Louis)
Market Sentiment Index: Market Sentiment One and Market Sentiment Two	Market sentiment index one is computed from six raw proxies: CEF, TURN, NIPO, RIPO, S and $P^{D-ND}$ . Market sentiment index two is computed based on the residuals from the regression of these raw proxies on business cycle variations.	$Market\ Sentiment\ One_t = -0.241CEFD_t + 0.242TURN_{t-1} + 0.253NIPO_t + 0.257RIPO_{t-1} + 0.112S_t - 0.283P_{t-1}^{D-ND}$ $Market\ Sentiment\ Two_t = -0.198CEFD_t^* + 0.225TURN_{t-1}^* + 0.234NIPO_t^* + 0.263RIPO_{t-1}^* + 0.211S_t^* - 0.243P_{t-1}^{D-ND*}$	Baker and Wurgler Sentiment Index

Table 2: Summary statistics for excess market returns and their predictors: 1963:7–2018:12

Variable	Mean	SD	Min	Max	Skewness	Kurtosis	Autocorrelation
$R_{s,t}$	0.005	0.044	-0.023	0.161	-0.545	4.980	0.070
$R_f$	0.004	0.003	0.000	0.014	0.618	3.702	0.972
AC	0.005	0.002	0.001	0.009	-0.084	2.542	0.605
AV	0.017	0.023	0.003	0.289	7.295	72.356	0.670
TS	1.059	1.680	-3.070	3.850	-1.139	5.129	0.954
DS	2.066	0.797	0.290	6.010	0.853	5.655	0.969
CAPE	20.325	8.133	6.639	44.198	0.526	3.177	0.995
IS1	0.000	1	-2.42	3.20	0.155	3.862	0.982
IS2	-0.000	1	-2.32	3.10	0.458	3.671	0.986

Note:  $R_{s,t}$  denotes stock market returns minus risk-free rate (excess market returns);  $R_f$ : risk-free rate; AC: average correlation; AV: average variance; TS: term spread; DS: default spread; CAPE: cyclically adjusted price earnings ratio; IS: investor sentiment; SD: standard deviation; and Autocorrelation: first-order autocorrelation.

Table 3: Pair-wise correlation matrix for the model variables

Variable	$R_{s,t}$	$R_{f,t-1}$	$R_{s,t-1}$	$AC_{t-2}$	$AV_{t-2}$	$TS_{t-1}$	$DS_{t-1}$	$CAPE_{t-1}$
$R_{s,t}$	1							
$R_{f,t-1}$	-0.048	1						
$R_{s,t-1}$	0.071	-0.083	1					
$AC_{t-2}$	0.094	0.012	0.029	1				
$AV_{t-2}$	0.029	-0.220	-0.072	0.373	1			
$TS_{t-1}$	0.107	-0.615	0.121	-0.065	0.094	1		
$DS_{t-1}$	0.067	-0.342	0.060	0.355	0.551	0.365	1	
$CAPE_{t-1}$	-0.037	-0.533	-0.022	-0.368	-0.065	0.100	0.010	1

Table 4: OLS estimates of excess stock market returns

$$R_{s,t} = \mu + \sum_{i=1}^4 \beta_i AC(AV)_{t-i} + \beta_5 R_{s,t-1} + \varepsilon_t$$

Model	Constant	$R_{s,t-1}$	$AC_{t-1}$	$AC_{t-2}$	$AC_{t-3}$	$AC_{t-4}$	Adj. $R^2$
Model 1	-0.007 (-1.49)			0.029*** (2.63)			0.007
Model 2	-0.008 (-1.58)	0.001 (1.58)		0.025*** (2.69)			0.011
Model 3	-0.012 (-1.69)		-0.014 (-0.89)	0.029*** (2.41)	0.000 (0.02)	0.017 (1.16)	0.008
Model 4	-0.013 (-1.91)	0.075 (1.66)	-0.003 (-0.19)	0.025** (2.01)	-0.004 (-0.30)	0.015 (1.00)	0.011
	Constant	$R_{s,t-1}$	$AV_{t-1}$	$AV_{t-2}$	$AV_{t-3}$	$AV_{t-4}$	Adj. $R^2$
Model 1*	0.007*** (3.96)			-0.001 (-1.66)			0.004
Model 2*	-0.007*** (3.34)	0.001 (1.42)		-0.001 (-1.31)			0.005
Model 3*	0.005** (2.27)		-0.003*** (-3.25)	0.003*** (2.54)	-0.002 (-0.89)	0.002** (1.96)	0.018
Model 4*	0.005** (2.12)	0.000 (1.01)	-0.003** (-2.55)	0.003*** (2.39)	-0.002 (-0.95)	0.002 (1.89)	0.018

Note:  $AC_{t,i}$  indicates the lagged average correlation ( $i = 1, 2, 3, 4$ ).  $AV$  indicates the lagged average variance ( $i = 1, 2, 3, 4$ );  $t$ -statistics are adjusted using the Newey-West (1987) standard errors. Figures in parentheses are the  $t$ -statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 5: OLS estimates of excess market returns with lagged average correlations and other predictors

Model	$R_{s,t} = \mu + \beta_1 AC_{t-2} + \beta_2 X_{k,t-1} + \varepsilon_t$								Adj. $R^2$
	Constant	$AC_{t-2}$	$AV_{t-2}$	$R_{s,t-1}$	$R_{f,t-1}$	$TS_{t-1}$	$DS_{t-1}$	$CAPE_{t-1}$	
Model 1	-0.007 (-1.49)	0.025*** (2.63)							0.007
Model 2	0.004 (1.70)		0.001 (0.43)						-0.001
Model 3	0.005*** (2.68)			0.071 (1.72)					0.004
Model 4	0.009*** (2.87)				0.008 (1.25)				0.001
Model 5	0.002 (0.96)					0.003** (2.25)			0.010
Model 6	-0.002 (-0.41)						0.004 (1.17)		0.003
Model 7	0.009 (1.82)							-0.000 (-0.88)	-0.001
Model 8	-0.008 (1.52)	0.026*** (2.50)	-0.000 (-0.09)						0.006
Model 9	-0.008 (-1.58)	0.025*** (2.69)		0.001 (1.58)					0.008
Model 10	-0.004 (-0.81)	0.026*** (2.63)			-0.008 (-1.22)				0.008
Model 11	-0.016** (-2.20)	0.028*** (2.88)				0.003** (2.32)			0.019
Model 12	-0.010 (-1.46)	0.022** (2.07)					0.002 (0.68)		0.007
Model 13	-0.007 (-0.89)	0.025*** (2.57)						-0.000 (-0.07)	0.006
Model 14	-0.002 (-0.98)	0.003*** (2.64)	-0.000 (-0.15)	0.001 (1.21)	0.006 (0.47)	0.003 (1.78)	-0.000 (-0.02)	0.000 (0.13)	0.014

Note:  $AC_{t-2}$  refers to two-month lagged average correlation,  $AV_{t-2}$  indicates the two-month lagged average variance;  $X_{k,t-1}$  represents the other predictors with one-month lag ( $AV_{t-2}$ ,  $R_{s,t-1}$ ,  $R_{f,t-1}$ ,  $TS_{t-1}$ ,  $DS_{t-1}$ ,  $CAPE_{t-1}$ ). Figures in parentheses are the  $t$ -statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 6: Excess market returns predictions over different time horizons ( $k$  months)

Predictors	$R_{S,t \rightarrow t+k} = \mu + \beta X + \varepsilon_t$							
	k=1	k=2	k=3	k=4	k=6	k=12	k=24	k=36
$AC_{t-2}$	0.018 (1.72)	0.035** (2.26)	0.047*** (2.46)	0.058*** (2.63)	0.072*** (2.64)	0.106*** (2.73)	0.247*** (4.65)	0.204*** (3.29)
Adj. $R^2$	0.003	0.006	0.008	0.009	0.009	0.010	0.031	0.015
$AV_{t-2}$	0.001 (0.42)	0.000 (0.19)	0.002 (0.46)	0.003 (0.94)	0.005 (1.87)	0.008** (2.09)	0.016** (2.15)	0.014 (1.36)
Adj. $R^2$	-0.001	-0.001	0.001	0.000	0.010	0.012	0.026	0.015
$R_{s,t-1}$	-0.031 (-0.75)	-0.019 (-0.29)	-0.007 (-0.08)	0.049 (0.53)	-0.004 (-0.04)	-0.014 (-0.11)	-0.276 (-1.58)	-0.270 (-1.92)
Adj. $R^2$	-0.001	-0.001	-0.002	-0.001	-0.002	-0.002	0.002	0.001
$R_{f,t-1}$	-0.009 (-1.34)	-0.017 (-1.28)	-0.024 (-1.19)	-0.030 (-1.13)	-0.042 (-1.10)	-0.074 (-1.50)	-0.107 (-1.08)	-0.142 (-0.93)
Adj. $R^2$	0.002	0.003	0.005	0.006	0.008	0.014	0.015	0.020
$TS_{t-1}$	0.003** (2.19)	0.005** (2.09)	0.006 (1.94)	0.008 (1.87)	0.012 (1.89)	0.026*** (3.03)	0.034** (2.17)	0.047** (2.30)
Adj. $R^2$	0.009	0.013	0.016	0.020	0.029	0.077	0.066	0.100
$DS_{t-1}$	0.003 (1.14)	0.007 (1.28)	0.011 (1.56)	0.016 (1.78)	0.021 (1.76)	0.034 (1.83)	0.054 (1.67)	0.075 (1.89)
Adj. $R^2$	0.002	0.005	0.012	0.018	0.021	0.027	0.038	0.055
$CAPE_{t-1}$	-0.000 (-0.99)	-0.000 (-0.98)	-0.001 (-1.00)	-0.001 (-0.99)	-0.001 (-0.99)	-0.003 (-1.18)	-0.006 (-1.57)	-0.009** (-2.05)
Adj. $R^2$	0.000	0.001	0.003	0.004	0.007	0.020	0.053	0.079

Note: The sample period of monthly excess market returns is from 1963:7 to 2018:12. Figures in parentheses are the  $t$ -statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 7: Market sentiments and the estimates of excess market returns

Market State	$R_{s,t} = \mu + \beta_1 AC_{t-2} + \beta_2 AC_{t-2} * MS + \varepsilon_t$						
	Constant	$t$ -statistic	$AC_{t-2}$	$t$ -statistic	$AC_{t-2} * MS$	$t$ -statistic	Adj. $R^2$
Recession	-0.008	(-1.65)	0.027***	(3.19)	-0.027**	(-2.40)	0.019
Market sentiment one	-0.011	(-1.96)	0.027***	(2.59)	-0.002	(-0.47)	0.007
Market sentiment two	-0.011	(-1.93)	0.027***	(2.56)	-0.002	(-0.58)	0.007

Note: MS denotes market state, which is measured by the following three variables: (1) Recession — a dummy variable — indicates the period of economic downturns; (2) Market sentiment one indicates the investor sentiment index one provided in Baker and Wurgler (2006); and (3) Market sentiment two indicates the investor sentiment index two. Independent variable is average correlation with 2 lags ( $AC_{t-2}$ ). Figures in parentheses are the  $t$ -statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 8: Means difference test ( $t$ -test) for different sentiment periods

Market State	Above Median		Below Median		Difference	$t$ -statistic
	Mean	SD	Mean	SD		
Recession	-0.006	0.066	0.007	0.040	-0.013***	(-2.55)
Market sentiment one	0.004	0.043	0.007	0.045	-0.003	(-0.83)
Market sentiment two	0.003	0.043	0.007	0.045	-0.004	(-1.15)
AC in recession	0.590	0.131	0.476	0.160	0.113***	(6.16)
AC in market sentiment one	0.460	0.168	0.523	0.148	-0.063***	(-5.15)
AC in market sentiment two	0.453	0.164	0.529	0.149	-0.075***	(-6.19)

Note: The means of excess market returns and of average correlation are compared based on three measures of market state: (1) the recession period and the rest; (2) the above-median market sentiment one period and the rest; and (3) the above-median market sentiment two period and the rest. Figures in parentheses are the  $t$ -statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 9: OLS estimates of excess market returns with different correlation measures

Predictors	$R_{s,t} = \mu + \beta X_{t-2} + \varepsilon_t$				Adj. $R^2$
	Constant	$t$ -statistic	Coefficient	$t$ -statistic	
ACt-2	-0.007	(-1.49)	0.025***	(2.63)	0.007
MCt-2	-0.009	(-1.71)	0.022***	(2.47)	0.006
QCt-2	-0.006	(-1.45)	0.024**	(2.27)	0.005
TQCt-2	-0.014	(-1.62)	0.021***	(2.41)	0.006
$\Delta ACt-2$	0.002	(1.72)	0.003	(0.89)	0.001

*Note:* Average correlation is the mean of pairwise correlations between any two industry returns in all the 48 industries.  $X_{t-2}$  represents Median Correlation (MC), Quantile Correlation (QC) and Three Quantile Correlation (TQC), which are the median, the 25 percentiles and the 75 percentiles of these correlations, respectively.  $\Delta AC_{t-2}$  is the difference between average correlations of months  $t-1$  and  $t-2$ . Figures in parentheses are the  $t$ -statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 10: Estimates of returns on 48 industry portfolios

Industry	$R_{i,t} = \mu + \beta AC_{t-2} + \varepsilon_t$				Adj. $R^2$
	Constant	$t$ -statistic	$\beta$	$t$ -statistic	
Agriculture	-0.317	(-0.44)	2.666	(1.87)	0.003
Food	-0.309	(-0.60)	2.699***	(2.82)	0.008
Soda	0.280	(0.35)	1.761	(1.11)	0.001
Liquor	0.388	(0.62)	1.480	(1.21)	0.001
Tobacco	0.424	(0.48)	1.831	(1.16)	0.001
Recreation	-0.534	(-0.55)	2.690	(1.50)	0.002
Entertainment	-0.513	(-0.67)	3.817**	(2.31)	0.005
Printing and Publishing	-0.609	(-0.91)	3.066**	(2.18)	0.006
Consumer Goods	0.468	(0.90)	0.840	(0.77)	0.001
Apparel	-0.316	(-0.48)	2.867**	(2.26)	0.004
Healthcare	-33.35***	(-2.54)	48.42***	(2.41)	0.058
Medical Equipment	0.604	(1.00)	1.098	(0.97)	0.000
Pharmaceutical Products	0.236	(0.43)	1.696	(1.75)	0.002
Chemicals	-0.878	(-1.57)	3.696***	(3.29)	0.010
Rubber and Plastic Products	-0.216	(-0.29)	2.588	(1.72)	0.003
Textiles	-1.174	(-1.38)	4.383***	(2.59)	0.008
Construction Materials	-1.193	(-1.74)	4.417***	(3.20)	0.012
Construction	-0.288	(-0.35)	2.545	(1.54)	0.002
Steel works Etc	-0.907	(-1.21)	3.274***	(2.33)	0.004
Fabricated Products	-0.732	(-1.05)	3.022**	(2.22)	0.003
Machinery	-0.250	(-0.38)	2.485	(1.89)	0.003
Electrical Equipment	-0.082	(-0.12)	2.462	(1.85)	0.003
Automobiles and Trucks	-1.201	(-1.64)	4.064***	(2.73)	0.008
Aircraft	-0.176	(-0.22)	2.085	(1.36)	0.001
Ship Building, Railroad Equipment	-0.991	(-1.27)	4.176***	(2.65)	0.007
Defence	-0.431	(-0.52)	3.280**	(2.11)	0.005
Precious Metals	0.709	(0.62)	0.496	(0.20)	0.000
Non-Metallic and Industrial Metal	-0.338	(-0.40)	2.821	(1.66)	0.002
Mining Coal	-0.186	(-0.15)	2.571	(1.02)	0.001
Petroleum and Natural Gas	0.104	(0.20)	1.731	(1.62)	0.001
Utilities	-0.121	(-0.24)	1.916**	(2.04)	0.005
Telecom Communication	-0.980	(-1.45)	3.718***	(3.07)	0.016
Personal Services	-0.751	(-0.92)	2.981**	(2.04)	0.004
Business Services	-0.475	(-0.59)	3.288**	(2.25)	0.005
Computers	-0.095	(-0.09)	2.164	(1.15)	0.001
Electronic Equipment	-0.128	(-0.12)	2.388	(1.27)	0.001
Measuring and Control Equipment	-0.061	(-0.07)	2.366	(1.51)	0.002
Business Supplies	-0.458	(-0.88)	2.865***	(2.46)	0.006
Shipping Containers	-1.089	(-1.78)	4.198***	(3.71)	0.013
Transportation	-0.758	(-1.05)	3.470***	(2.41)	0.008
Wholesale	-0.506	(-0.78)	3.067***	(2.48)	0.006
Retail	-0.600	(-1.07)	3.374***	(3.08)	0.009
Restaurants, Hotel, Motels	0.199	(0.30)	2.027	(1.62)	0.001
Banking	-0.076	(-0.10)	2.047	(1.25)	0.002
Insurance	-0.513	(-0.80)	3.135***	(2.36)	0.007
Real Estate	-1.019	(-1.21)	3.419	(1.88)	0.004
Trading	-0.220	(-0.32)	2.638	(1.95)	0.003
Others	-0.566	(-0.73)	2.303	(1.39)	0.001

Note:  $R_{i,t}$  is the return on the individual industry portfolio  $i$ . Figures in parentheses are the  $t$ -statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 11: Estimates of returns on different portfolios

Portfolio	$R_{p,t} = \mu + \beta AC_{t-2} + \varepsilon_t$				
	Constant	<i>t</i> -statistic	$\beta$	<i>t</i> -statistic	Adj. <i>R</i> <sup>2</sup>
SMB	-0.098	(-0.25)	0.694	(0.95)	0.000
Small	0.102	(0.13)	2.012	(1.31)	0.001
Big	-0.589	(-1.10)	2.873***	(2.82)	0.010
HML	0.829	(-1.20)	-1.031	(-1.20)	0.002
Value	-0.242	(-0.42)	2.880***	(2.49)	0.007
Growth	-0.748	(-1.24)	3.193***	(2.92)	0.009
RMW	0.577	(1.36)	-0.655	(-0.89)	0.001
Robust	-0.320	(-0.62)	2.648***	(2.71)	0.007
Weak	-1.104	(-1.38)	3.772***	(2.60)	0.010
CMA	0.346	(1.04)	-0.125	(-0.20)	0.000
Conservative	-0.803	(-1.43)	3.894***	(3.47)	0.015
Aggressive	-0.855	(-1.06)	3.358***	(2.33)	0.007
Mom	2.262***	(3.95)	-3.096***	(-2.60)	0.012
Winner	0.814	(0.89)	1.365	(0.79)	0.000
Loser	-2.954***	(-3.10)	6.164***	(3.34)	0.013

Note: Small denotes the average return on the nine small stock portfolios; Big: the average return on nine large portfolios. SMB: Small minus Big; High: companies with high book-to-market ratio, known as value stocks; Low: companies with lower book-to-market values, known as growth stocks. HML: High minus Low; Robust: the average return on the two robust portfolios; Weak: the average return on the two weak portfolios. RMW: Robust minus Weak; Conservative: the average return on the two conservative portfolios; Aggressive: the average return on the two aggressive portfolios. CMA: Conservative minus Aggressive; Mom: the equally weighted average of the highest performing companies minus equally weighed average of the lowest performing companies; Winner: the average return on two prior high return portfolios; Loser: the average return on two prior low return portfolios. Figures in parentheses are the *t*-statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 12: Estimates of excess market returns by quintiles and deciles

Quintile (p)	$R_{s,t} = \mu + \beta AC_{t-2} + \varepsilon_t$				
	Constant	<i>t</i> -statistic	$\beta$	<i>t</i> -statistic	Adj. <i>R</i> <sup>2</sup>
$\geq 0$	-0.025	(-1.57)	0.054**	(1.97)	-0.004
$\geq 25$	-0.023***	(-2.41)	0.048***	(2.74)	-0.004
$\geq 50$	-0.005	(-0.55)	0.017	(0.80)	-0.004
$\geq 75$	0.003	(0.26)	0.005	(0.22)	-0.004
Decile (p)	Constant	<i>t</i> -statistic	$\beta$	<i>t</i> -statistic	Adj. <i>R</i> <sup>2</sup>
$\geq 0$	-0.011	(-0.34)	0.036	(0.70)	-0.004
$\geq 10$	-0.018	(-1.04)	0.043	(1.36)	-0.004
$\geq 20$	-0.046**	(-2.06)	0.078	(1.82)	-0.004
$\geq 30$	-0.018	(-1.13)	0.038	(1.37)	-0.004
$\geq 40$	-0.023	(-1.99)	0.059***	(2.65)	-0.004
$\geq 50$	-0.001	(-0.11)	0.010	(0.33)	-0.004
$\geq 60$	0.003	(0.24)	0.019	(0.52)	-0.004
$\geq 70$	-0.003	(-0.24)	-0.009	(-0.26)	-0.004
$\geq 80$	0.001	(0.08)	-0.001	(-0.03)	-0.004
$\geq 90$	-0.005	(-0.26)	0.030	(0.82)	-0.004

Note: Figures in parentheses are the *t*-statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 13: Estimates of excess market returns by sub-sample periods

Period	$R_{s,t} = \mu + \beta AC_{t-2} + \varepsilon_t$				
	Constant	<i>t</i> -statistic	$\beta$	<i>t</i> -statistic	Adj. $R^2$
1963:7-1972:12	-0.006	(-0.94)	0.023	(1.71)	0.003
1973:1-1982:12	-0.064***	(-3.58)	0.111***	(3.77)	0.046
1983:1-1992:12	-0.070	(-0.70)	0.030	(1.64)	0.000
1993:1-2002:12	-0.009	(-0.82)	0.038	(1.40)	0.003
2003:1-2012:12	-0.012	(-0.93)	0.030	(1.40)	0.000
2013:1-2018:12	-0.003	(-0.19)	0.024	(1.00)	0.000
1963:7-1982:12	-0.015	(-1.76)	0.034**	(2.14)	0.011
1983:1-2002:12	-0.007	(-0.90)	0.032**	(1.97)	0.005
2003:1-2018:12	-0.005	(-0.49)	0.022	(1.22)	0.001
1963:7-1992:12	-0.012	(-1.74)	0.031***	(2.42)	0.008
1993:1-2018:12	-0.041	(-0.56)	0.022	(1.51)	0.005
1963:7-2018:12	-0.519	(-1.06)	0.024***	(2.43)	0.007

[1973:1-1982:12  
Excluded]

Note: Figures in parentheses are the *t*-statistics. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

Table 14: Bootstrapping with different draws

Number of draws	Constant	z-value	$\beta$	z-value	Adj. $R^2$
50	-0.007	(-1.82)	0.026***	(2.85)	0.007
100	-0.007	(-1.63)	0.026***	(2.76)	0.007
200	-0.007	(-1.64)	0.026***	(2.67)	0.007
500	-0.007	(-1.51)	0.026***	(2.51)	0.007
1000	-0.007	(-1.49)	0.026***	(2.47)	0.007
2000	-0.007	(-1.49)	0.026***	(2.51)	0.007
5000	-0.007	(-1.52)	0.026***	(2.54)	0.007
10000	-0.007	(-1.52)	0.026***	(2.53)	0.007
20000	-0.007	(-1.52)	0.026***	(2.55)	0.007

Note: Samples are randomly drawn in bootstrap simulations, and the estimation of subsequent market returns by lagged average correlation is performed based on a number of draws. The average correlation significantly predicts subsequent market returns in all the random draws of samples. Figures in parentheses are the z-values. \*\*\*, \*\* indicate statistical significance at the 1% and 5% levels, respectively.

## References

- Aneja, Y. P., Chandra, R. & Gunay, E., 1989. A portfolio approach to estimating the average correlation coefficient for the constant correlation model. *The Journal of Finance*, 44(5), pp. 1435-1438.
- Ang, A. & Bekaert, G., 2007. Stock return predictability: Is it there? *Review of Financial Studies*, 20(3), pp. 651-707.
- Ang, A. & Chen, J., 2002. Asymmetric correlation of equity portfolios. *Journal of Financial Economics*, 63, pp. 443-493.
- Bahrami, A., Shamsuddin, A. & Uylangco, K., 2017. Out-of-sample stock return predictability in emerging markets. *Accounting & Finance*, 58, pp. 1-24.
- Baker, M. & Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4), pp. 1645-1680.
- Bali, T. G., Cakici, N., Yan, X., and Zhang, Z. 2005. Does idiosyncratic risk really matter?. *Journal of Finance*, 60, pp. 905-929.
- Barberis, N., Shleifer, A. & Vishny, R., 1998. A model of investor sentiment. *Journal of Financial Economics*, 49(3), pp. 307-343.
- Barro, R. & Ursúa, J., 2017. Stock-market crashes and depressions. *Research in Economics*, 71(3), pp. 384-398.
- Berkman, H., Jacobsen, B. & Lee, J., 2011. Time-varying rare disaster risk and stock returns. *Journal of Financial Economics*, 101(2), pp. 313-332.
- Beyhaghi, M. & Hawley, J., 2013. Modern portfolio theory and risk management: Assumptions and unintended consequences. *Journal of Sustainable Finance & Investment*, 3(1), pp. 17-37.
- Bollerslev, T., Marrone, J., Xu, L. & Zhou, H., 2014. Stock return predictability and variance risk premia: Statistical inference and international evidence. *Journal of Financial and Quantitative Analysis*, 49(3), pp. 633-661.
- Boyd, J., Hu, J. & Jagannathan, R., 2005. The stock market's reaction to unemployment news: Why bad news is usually good for stocks. *The Journal of Finance*, 60(2), pp. 649-672.
- Campbell, J. Y., 1987. Stock returns and the term structure. *Journal of Financial Economics*, 18(2), pp. 373-399.
- Campbell, J. Y. & Shiller, R. J., 1988. Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3), pp. 661-676.
- Campbell, J. Y. & Thompson, S. B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4), pp. 1509-1531.

- Campbell, J. Y. & Viceira, L. M., 2002. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. USA: Oxford University Press.
- Campbell, J. Y. & Yogo, M., 2006. Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1), pp. 27-60.
- Chaibi, A., Alioui, S. & Xiao, B., 2015. On the impact of firm size on risk and return: Fresh evidence from the American stock market over the recent years. *Journal of Applied Business Research*, 31(1), pp. 29-40.
- Charles, A., Darné, O. & Kim, J., 2017. International stock return predictability: Evidence from new statistical tests. *International Review of Financial Analysis*, 54, pp. 97-113.
- Chernick, M. R., 2007. What is bootstrapping? In: *Bootstrap Methods: A Guide for Practitioners and Researchers*. 2 ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 1-25.
- Cochrane, J., 1991. Production-based asset pricing and the link between stock returns and economic fluctuations. *The Journal of Finance*, 46(1), pp. 209-237.
- Cooper, I. & Priestley, R., 2005. *Stock Return Predictability in a Production Economy*. Boston, AFA 2006 Boston Meetings Paper.
- De Long, J. B., Shleifer, A., Summers, L. H. & Waldmann, R. J., 1990. Noise trader risk in financial markets. *Journal of Political Economy*, 98(4), pp. 703-738.
- Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), pp. 383-417.
- Fama, E. F. & French, K. R., 1988. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1), pp. 3-25.
- Fama, E. F. & French, K. R., 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1), pp. 23-49.
- Fama, E. F. & Schwert, G. W., 1977. Asset returns and inflation. *Journal of Financial Economics*, 5(2), pp. 115-146.
- Forbes, K. J. & Rigobo, R., 2002. No contagion, only interdependence: Measuring stock market comovements. *The Journal of Finance*, 57, pp. 2223-2261.
- Goyal, A. & Welch, I., 2003. Predicting the equity premium with dividend ratios. *Management Science*, 49(5), pp. 639-654.
- Hafer, R. W. & Hein, S. E., 2007. *The Stock Market*. Greenwood Publishing Group.
- Harvey, C. R. 2001. The specification of conditional expectations. *Journal of Empirical Finance*, 8, pp. 573-637.
- Ho, C. W. & Hung, C. H., 2009. Investor sentiment as conditioning information in asset pricing. *Journal of Banking & Finance*, 33(5), pp. 892-903.

- Huang, D., Jiang, F., Tu, J. & Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies*, 28(3), pp. 791-837.
- Jensen, G. R. & Johnson, R. R., 1993. An examination of stock price reactions to discount rate changes under alternative monetary policy regimes. *Quarterly Journal of Business and Economics*, 32(2), pp. 26-51.
- Jordan, S. J., Vivian, A. J. & Wohar, M. E., 2014. Forecasting returns: New European evidence. *Journal of Empirical Finance*, 26, pp. 76-95.
- Kallberg, J. & Pasquariello, P., 2008. Time-series and cross-sectional excess comovement in stock indexes. *Journal of Empirical Finance*, 15(3), pp. 481-502.
- Kenourgios, D., Samitas, A. & Paltalidis, N., 2009. Financial market dynamics in an enlarged European Union. *Journal of Economic Integration*, 24(2), pp. 197-221.
- Kothari, S. P. & Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics*, 44(2), pp. 169-203.
- Lahmiri, S., 2015. Long memory in international financial markets trends and short movements during 2008 financial crisis based on variational mode decomposition and detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, 437, pp. 130-138.
- Lettau, M. & Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance*, 56(3), pp. 815-849.
- Lewellen, J., 2004. Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2), pp. 209-235.
- Li, D. & Zhang, L., 2010. Does q-theory with investment frictions explain anomalies in the cross section of returns? *Journal of Financial Economics*, 98(2), pp. 297-314.
- Liu, C. & Mei, J., 1992. The predictability of returns on equity REITs and their co-movement with other assets. *The Journal of Real Estate Finance and Economics*, 5, pp. 401-418.
- Malkiel, B. G., 2004. Models of stock market predictability. *The Journal of Financial Research*, 27(4), pp. 449-459.
- Markowitz, H. 1952. Portfolio selection. *Journal of Finance*, 7, pp. 77-91.
- Menzly, L. & Ozbas, O., 2010. Market segmentation and cross-predictability of returns. *The Journal of Finance*, 65(4), pp. 1555-1580.
- Menzly, L., Santos, T. & Veronesi, P., 2004. Understanding predictability. *Journal of Political Economy*, 112(1), pp. 1-47.
- Newey, W. & West, K., 1987. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28(3), pp. 777-787.

- Noussair, C. & Tucker, S., 2013. Experimental research on asset pricing. *Journal of Economic Surveys*, 27(3), pp. 554-569.
- Pástor, L. & Stambaugh, R. F., 2009. Predictive systems: Living with imperfect predictors. *The Journal of Finance*, 64(4), pp. 1583-1628.
- Pollet, J. M. & Wilson, M., 2010. Average correlation and stock market returns. *Journal of Financial Economics*, 96(3), pp. 364-380.
- Samitas, A. & Tsakalos, I., 2013. How can a small country affect the European economy? The Greek contagion phenomenon. *Journal of International Financial Markets, Institutions and Money*, 25, pp. 18-32.
- Schmeling, M., 2009. Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, 16(3), pp. 394-408.
- Siklos, P. L. & Kwok, B., 1999. Stock returns and inflation: A new test of competing hypotheses. *Applied Financial Economics*, 9(6), pp. 567-581.
- Tan, P., Chin, C. & Galagedera, D., 2014. A wavelet-based evaluation of time-varying long memory of equity markets: A paradigm in crisis. *Physica A: Statistical Mechanics and its Applications*, 410, pp. 345-358.
- Velázquez, A. & Smith, P., 2013. Equity returns and the business cycle: The role of supply and demand shocks. *The Manchester School*, 81(S1), pp. 100-124.
- Welch, I. & Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), pp. 1455-1508.