Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis BSc^2 Econometrics/Economics

Assessing Return Predictability: The Role of Multiple Predictors and Insider Trading Information

Drini Hoxha (558635)

Supervisor:	Terri van der Zwan
Second assessor:	dr. Mikhail Zhelonkin
Date final version:	28th June 2024

Abstract

This paper investigates return predictability for a plethora of variables based on short interest, options trading volume, volatility, and market sentiment. The analysis is conducted at both the market and individual equity levels, and utilises various econometric tests, as well as a factor investing framework. We do not find any consistently outperforming predictor, but the preferred choice rather depends on the return aggregation level and the evaluation method. To facilitate the understanding of our empirical results, we propose a novel theoretical model to capture insiders' trading behavior. This model provides theoretical predictions that are in line with our empirical results. In particular, we find that a short interest-based predictor performs best for market returns. Moreover, predictors based on options trading volume demonstrate superior performance in predicting individual equity returns, as well as their respective factor portfolio returns.

Contents

1 Introduction

Return predictability is one of the most widely investigated topics in the financial econometrics literature. Researchers have developed various models to explain returns in-sample, and predict them out-of-sample, and they have also developed various statistical tests to test their significance. This topic is of great interest not only to researchers, but also to practitioners, who are on an on-going pursuit to build models that yield reliable return forecasts, and convert these models into profitable trading strategies. Additionally, this topic is also relevant for policymakers, as having reliable forecasts regarding the future returns of the market, and hence also regarding the future state of the economy, would allow policy makers to adjust their actions accordingly.

To this end, this paper investigates the predictive power of a plethora of variables. Broadly speaking, these variables fall in one of the following four categories: variables utilising data on the number of stocks sold-short, variables utilising options volume trade data, future market volatility-based variables, and market sentiment-based variables. We examine the predictive power that the aforementioned variables have on future aggregate returns, which, in simple terms, are defined as the mean log returns over some time period in the future.

Our analysis begins by investigating the aggregate return predictability for the market portfolio, utilising a linear model specification. In this manner, we enrich the literature with additional evidence on return predictability, offer traders tools to improve their investment strategies, and provide policy makers with reliable models for forecasting the future market conditions. In particular, we test the in- and out-of-sample predictive power that nine predictors have on the aggregate market returns.

Moreover, we consider the predictability of aggregate returns for individual stocks. For this purpose, we test the predictive power of a linear model specification applied to only three out of the nine predictors, which can be constructed for individual equities. We perform in- and out-of-sample tests on the predictive power of these variables for each individual equity, as well as for specific sub-samples. Furthermore, for each predictor, we construct a factor using the decile approach and evaluate its predictive power on the aggregate returns of its respective portfolio. Finally, we investigate whether profitable trading strategies utilising the forecasts of these variables can be developed, from the perspective of a risk-averse mean-variance investor. This section of the paper is particularly relevant to practitioners, due its practical nature.

To enhance the understanding of our results on the predictive power of these three variables, we build on Johnson and So (2012), and introduce a theoretical framework for trading in the presence of insider information. This model aims to describe the trading behavior of a market participant who possesses insider information and explains the empirical patterns in the aggregate return predictive ability of these variables. To this end, we do not only enrich the literature with empirical findings, but also with a theoretical model that explains the actions of informed traders.

This paper is organised as follows. Section 2 presents a comprehensive literature review and explains how this paper links with previously-conducted research. Section 3 describes the data, and Section 4 outlines our methodology. Section 5 introduces the theoretical model and points out its implications. Section 6 presents the results, and Section 7 concludes.

2 Literature Review

The literature in return predictability typically uses a linear model specification, where a single predictor is used to forecast returns. To this end, the research task at hand boils down to finding variables that can predict returns in- and out-of-sample. The predictors examined in literature are typically macroeconomic variables and financial fundamentals. In particular, one of the most well-known papers is Welch and Goyal (2008), which investigates 14 potential predictors of returns based on financial fundamentals. They find that all these variables perform poorly in- and out-of-sample. These results are mostly re-confirmed in Rapach, Ringgenberg and Zhou (2016), who re-investigate these variables in a more recent setting. Moreover, Rapach, Wohar and Rangvid (2005) also find little evidence of market return predictability for a set of various macroeconomic variables.

Since macroeconomic variables and fundamentals have been documented to poorly predict stock market returns, it is pertinent to study other classes of variables. Rapach et al. (2016) propose the short interest index (SII), defined as the proportion of stocks that have been sold short over the total number of shares outstanding. They find that SII is one of the strongest predictors of market returns, both in- and out-of-sample, and they argue that its predictive ability stems from the fact that short sellers are informed traders.

In the same spirit, studies that investigate the trading actions of informed traders in the options market have been conducted. Pan and Poteshman (2006) use a data set that contains the volume of put and call trades initiated by buyers. Intuitively, insiders that possess negative signals about the future return of a stock are likely to buy puts, while those that possess positive signals are likely to buy calls. They construct the put-to-call ratio using these trading volumes, and find that stocks with low put-to-call ratios significantly outperform their high put-to-call ratio peers, on a daily and on a weekly basis. In addition, Johnson and So (2012) perform a similar analysis, however, their data set contains the total volumes of traded puts and calls, without distinguishing the initiator of the trade. They construct the put-to-call and option-to-stock ratios for each stock, and find that stocks with low option-to-stock ratios substantially outperform those with high option-to-stock ratios. On the other hand, they find that the put-to-call ratio does not convey information regarding the future returns of a stock. To motivate their findings, Johnson and So (2012) build a theoretical framework that models the behavior of informed traders that can trade in both the equity and option markets. They use this model to theoretically show that an insider who possesses negative information on a stock is more inclined to express his views in the option, rather than the underlying stock market.

These findings demonstrate that the information possessed by insiders provides crucial knowledge about future returns. Therefore, a refined approach for predicting returns involves anticipating the insights of insiders, inferred by their trading activities. This paper partly inherits the framework presented in Johnson and So (2012), but changes its assumptions, in an attempt to provide a more realistic picture of insider trading decisions. Moreover, this paper uses data on the trading volume of puts, calls and stock shares, and re-investigates the ability of put-to-call and option-to-stock variables to predict returns.

In addition to examining predictors that aim to capture the information asymmetry among market participants, this paper also investigates how future market volatility influences future market returns. The literature on this issue is quite controversial, with papers documenting positive, negative or non-significant results on the relationship between volatility and returns. Guo and Savickas (2006) find that market volatility and idiosyncratic volatility are jointly significant predictors of future market returns, with the relationship between market volatility and market returns being positive, but the one between idiosyncratic volatility and market returns being negative. On the contrary, Li, Yang, Hsiao and Chang (2005) find that market volatility negatively influences the market returns across 12 different markets.

This paper re-examines this topic, using the VIX (the Chicago Board Options Exchange Volatility Index), as a proxy for future market volatility. As advised in Bekaert and Hoerova (2014), the VIX is decomposed into two parts: future market uncertainty as measured by the conditional market variance forecast, and a premium for the risk aversion among market participants. This decomposition enables a more precise assessment of the relationship between market returns and volatility. Bekaert and Hoerova (2014) find that the risk aversion premium can significantly predict market returns, while the conditional variance forecast cannot. To this end, this paper re-assesses the predictive power of these variables in a more recent sample.

Additionally, the market sentiment could contain vital information regarding future market returns. Most studies use data on the flow of money into funds as a proxy for market sentiment. Frazzini and Lamont (2008) find that large inflows of money into mutual funds are associated with low future market returns. In this context, they refer to the money that individual investors deposit into mutual funds as 'dumb money'. Furthermore, in a more recent setting, Akbas, Armstrong, Sorescu and Subrahmanyam (2015) re-confirm this finding, and additionally find that inflows into hedge funds are a positive signal about future market returns, thereby classifying these funds as 'smart money'. These findings make intuitive sense, since the funds flowing into mutual funds come from individual investors with little investing experience, while the funds going into hedge funds are from accredited investors who thoroughly analyse the market. Consequently, this paper re-investigates the ability of 'dumb money' to predict the market returns.

To get a better grasp on the ability of individual investor beliefs to predict future market returns, this paper additionally utilises survey expectation data. These data reflect the proportion of individual investors that believe that the market will be bullish, bearish or neutral in the near future. It is to the best of our knowledge that survey expectation data have not been used in previous research to predict market returns, making this paper a pioneering research in this direction.

3 Data

This paper uses data from a broad range of sources. We start with the sources that contain equity-specific data. The first two sources are the Center for Research in Security Prices (CRSP) and Compustat. These sources are used to extract data on the number of shares sold short (known as short interest), the total number of shares outstanding, and the prices of securities of all equities traded in AMEX, NASDAQ, and NYSE. The third data source is OptionMetrics, which is used to extract data on the trading volume of all call and put options for all the aforementioned equities. Moreover, to avoid incorporating micro-caps into the analysis, we filter out stocks that are below the bottom quintile according to the market capitalisation breakpoints found in the website of Kenneth French. Moreover, we also filter out the stocks that have a share price below \$5.

Following this, we delve into the extraction of non-equity-specific data. First, we extract data on two variables that arise after decomposing the VIX index, as shown in Bekaert and Hoerova (2014). In particular, data on the conditional expected variance of the market and the market variance premium is extracted from the website of Marie Hoerova. To construct the predictors based on market sentiment, data from the AAII Investor Sentiment Survey is utilised. This survey asks individual investors whether they think that the market will be bullish, bearish, or neutral in the short-term future on a weekly basis. In this way, this survey provides data on the proportion of individual investors that believe the market will be in each of the aforementioned states in the short-term future. Moreover, another market sentiment predictor is constructed using data from the Federal Reserve Economic Data on the total market value of mutual funds in the United States.

Finally, the data set provided in the webpage of Amit Goyal is used to extract data on the market portfolio returns, as well as the risk free rate. The market portfolio refers to the Standard & Poors 500 Index, and the risk free rate refers to the one-month Treasury Bill interest rate.

4 Methodology

4.1 Aggregate Returns

A crucial observation that needs to be drawn is that this paper investigates the predictive ability of various predictor variables for the aggregate returns of some asset. To this end, it is necessary to define aggregate returns. Let us consider an asset, and denote the series of log returns realized by this asset over T periods of time by $r_1, r_2, \dots, r_T$. Then, the aggregate return generated by investing in this asset at time t, and holding the investment for the future h months, is defined as follows:

$$r_{t:t+h} = \frac{1}{h} \sum_{i=1}^{h} r_{t+i} \tag{1}$$

4.2 Predictors

This paper considers four classes of predictors: a predictor based on the proportion of short-sold shares (that is, short interest), predictors based on the options trading volume, predictors based on the VIX decomposition as done in Bekaert and Hoerova (2014), and predictors based on the market sentiment. The construction of the first two classes of predictors depends on whether the predictor is being applied to predict some portfolio returns, or the returns of a single equity. Naturally, when applying a predictor to portfolio returns, some aggregation of the predictor for each individual equity in the portfolio needs to be conducted. To avoid confusion, we start by explaining the first two classes of predictors, and then explain the other two classes.

4.2.1 Predictors of Individual Equity and Portfolio Returns

Let us first explain how the predictor construction changes depending on whether individual equity aggregate returns or portfolio aggregate returns are being predicted. In the case of individual equities, the construction involves two steps: first constructing the predictor, and then de-trending if necessary. In the case of portfolio returns, the procedure involves three steps: constructing the predictor for each equity in the portfolio, aggregating the predictors into a single predictor via taking an equal-weighted sum, and finally de-trending the aggregated predictor if necessary. In the remainder of this section, we explain how each predictor is constructed for a portfolio of equities, as that is the more general case.

In a similar fashion to Rapach et al. (2016), the predictor based on short interest is constructed as follows. First, for each equity in the portfolio, compute the ratio between the number of stocks sold-short and the total number of shares outstanding for each month. Then, take the equal-weighted sum of the predictor of each equity, to obtain a portfolio-level aggregated predictor. Finally, de-trend this predictor stochastically by subtracting its 60-month moving window mean. Again, following the convention in Rapach et al. (2016), we name this predictor SII.

The predictors based on the option trading volume are the put-to-call volume (PC) ratio and the option-to-stock (OS) volume ratio. As their names already suggest, these variables are constructed as the ratio between the number of puts and the number of calls traded each month, and the ratio between the number of options (both puts and calls) and the number of stock shares traded each month, respectively. Afterwards, an equal-weighted sum aggregation among all portfolio equities is performed for both PC and OS. Finally, stochastic de-trending using a 60-month moving window is performed for PC, but not for OS. This modelling choice is motivated by the knowledge that options are often used by institutional investors for hedging, and these investors' hedging strategies change dynamically over time depending on market conditions. On the other hand, OS is more stable, in the sense that institutional investor typically do not change the market type they invest in.

4.2.2 Predictors of Market Returns

We start this section with the crucial observation that the variables introduced here can only be applied to predict the aggregate market portfolio returns. The predictor variables introduced here are the predictors based on the VIX decomposition, and the predictors based on market sentiment.

Bekaert and Hoerova (2014) decompose the square of the monthly VIX index into two parts: the conditional market variance (CV) that reflects the future market uncertainty, and a variance risk premium (VP) that reflects the premium demanded by market participants due to their risk aversion. The decomposition is a two-step procedure. First, a model for the monthly CV is fitted, and a one-step ahead forecast is performed using the fitted model, such that $CV_{t+1|t}$ is obtained. As a general remark, variance here refers to the realized variance computed using a 22-day window. Bekaert and Hoerova (2014) experiment various model specifications, with the most general model incorporating the following explanatory variables: the squared VIX in the previous month, lagged realized variances decomposed into a continuous part and a jump part, and terms that capture the leverage effect. They choose the nested model that best fits the data in- and out-of-sample. The second step consists of simply computing the variance risk premium as:

$$VP_t = VIX_t^2 - CV_{t+1|t} \tag{2}$$

As a final remark, the two predictor variables that arise from this decomposition are $CV_{t+1|t}$ and VP_t.

In the remainder of this section, the construction of the market sentiment variables is explained. Let us begin with the predictors constructed utilising data on the individual investor sentiment regarding the future state of the market (bullish, bearish, or neutral) obtained from the AII Investor Sentiment Survey. This data is weekly, so we aggregate it into monthly via using the most recent observation corresponding to each month[1]. The predictors constructed are formally defined as follows:

$$BB_{binary} = \begin{cases} 1, & \text{if bearish sentiment} \\ 0, & \text{if neutral sentiment} \\ -1, & \text{if bullish sentiment} \end{cases} \tag{3}$$

$$BB_{ratio} = \frac{\text{percentage bearish}}{\text{percentage bullish}} \tag{4}$$

$$BB_{spread} = \text{percentage bearish} - \text{percentage bullish}, \tag{5}$$

where the binary variable is set to 1 if the majority of individual investors that participated in the survey think the market will be bearish in the short-term future, -1 if they think the market will be bullish, and zero if they think it will be neutral. Moreover, percentage bearish represents the percentage of individual investors that believe that the market will be bearish in the short-term future. Similar definition holds for percentage bullish.

The final market sentiment predictor reflects the amount of funds that flow into US mutual funds from individual investors. This predictor is created via stochastically de-trending the market value of US mutual funds using a 60-month moving window. In this way, we can capture the amount of money that flows into US mutual funds as compared to the previous five years. This predictor is named MVMF throughout this paper.

4.3 In-sample Analysis

The in-sample predictive power of each predictor introduced in Section 4.2 is tested using the following linear specification:

$$r_{t:t+h} = \alpha + \beta x_t + \epsilon_{t:t+h},\ t = 1, \ldots, T - h, \tag{6}$$

where $r_{t:t+h}$ represents the aggregate returns that occur from $t+1$ until $t+h$ for $h \in \{1, 3, 6, 12\}$,

[1]We also tried aggregating by taking the monthly average, but as the results were similar, we do not present them in the paper

and x_t denotes the value of the predictor variable at time t. The model is estimated using an expanding window with at least 60 monthly observations. To be able to compare the estimated beta coefficients across predictors, all the predictors are standardized to have a standard deviation of one before entering Equation 6.

It is well-known that the regression in Equation 6 suffers from two major econometric issues. First, the overlap of returns when computing $r_{t:t+h}$ leads to a high auto-correlation in the dependent variable, and in turn it also leads to auto-correlation in the error term. Second, the regressors x_t tend to be highly persistent. As shown in Kostakis, Magdalinos and Stamatogiannis (2015), the estimator of β in Equation 6 has very different statistical properties depending on the degree of persistence of x_t. As suggested in Rapach et al. (2016), to ensure reliable inferences, we bootstrap the distribution of the t-statistic associated with the β estimate using a wild bootstrap procedure. Using the bootstrapped distribution of the t-statistics, we can in turn infer the wild bootstrapped p-value for the β coefficient of each predictor. The details regarding this procedure are provided in Appendix A.

4.4 Out-of-sample Analysis: Baseline Tests

To assess the predictive power of the predictors considered in this paper, it is important to test how each predictor performs out-of-sample. To this end, the simple linear model is used again. In particular, the one-step ahead forecasts for the aggregate returns are constructed as follows:

$$\hat{r}_{t:t+h} = \hat{\alpha}_t + \hat{\beta}_t x_t, \tag{7}$$

where $\hat{r}_{t:t+h}$ represents the one-step ahead forecast for aggregate returns that occur from $t+1$ until $t+h$, for $h \in \{1, 3, 6, 12\}$; and $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the α and β estimates obtained by fitting the regression in Equation 6 using data up until time t.

The benchmark that the forecasts of Equation 7 are compared to is the mean return up until time t. Let us refer to the benchmark model by Model 1, and to the model given in Equation 7 by Model 2. To compare the out-of-sample performance across these models, we compute the mean squared forecast error (MSFE) of each model, and then compute the out-of-sample R^2 statistic, as presented in Campbell and Thompson (2008). Formally, this is done as follows:

$$R^2_{OS} = 100 \left(1 - \frac{MSFE(2)}{MSFE(1)} \right), \tag{8}$$

where R^2_{OS} represents the out-of-sample R^2 statistic, and $MSFE(1)$ and $MSFE(2)$ are the mean squared forecast errors of Model 1 and Model 2, respectively.

As argued in Rapach et al. (2016), it is crucial to note that using this benchmark simply boils down to setting $\beta = 0$ in Equation 6, since we fit the model using an expanding window. Therefore, the benchmark model is a nested model of the linear model, and hence, to test whether the linear model significantly outperforms the benchmark, we can utilise the test developed in Clark and West (2007). The testing procedure is summarised in Appendix B.

4.5 Out-of-sample Analysis: Additional Tests

In this section we describe some additional out-of-sample techniques that are specific to certain applications, rather than applied throughout the paper. First, we consider the optimal combination between the forecast of SII and that of another predictor for the aggregate return $r_{t:t+h}$. In mathematical notation, this is equivalent to finding the optimal λ coefficient in the following equation:

$$\hat{r}^{*}_{t:t+h} = (1-\lambda)\hat{r}^{other}_{t:t+h} + \lambda \hat{r}^{SII}_{t:t+h}, \tag{9}$$

where $\hat{r}^{SII}_{t:t+h}$ is the aggregate return forecast obtained using the SII predictor, while $\hat{r}^{other}_{t:t+h}$ is the aggregate return forecast obtained using any other predictor. To find the optimal combination, Harvey, Leybourne and Newbold (1998) derive a framework that estimates the optimal λ coefficient. Moreover, they also derive a test that tests $H_0 : \lambda = 0$ against $H_A : \lambda > 0$. In this paper we use their framework to find the optimal λ and also perform the test for each predictor considered (other than SII). For the sake of completeness, the framework of Harvey et al. (1998) is summarised in Appendix C.

Second, in addition to the specification provided in Equation 7, this paper also tests other specifications. In particular, we try three additional specifications, which add x_t^2, x_t^3 or both x_t^2 and x_t^3 to the specification in Equation 7. These other specifications provide the forecasting model with more flexibility, as they capture the effect of higher powers of the predictor in predicting aggregate returns, and hence could potentially lead to better forecasts.

Furthermore, we also examine the joint predictive power of all predictors via forecast combinations. This paper tests two methods for combining forecasts. The first one is based on the in-sample performance of each predictor and is formally given by:

$$w_{i,t+1} = \frac{\hat{\beta}_{i,t}}{\sum_{j=1}^{n} \hat{\beta}_{j,t}}, \tag{10}$$

where $w_{i,t+1}$ denotes the weight that predictor i gets in period $t+1$, $\hat{\beta}_{i,t}$ denotes the estimated slope coefficient as given in Equation 6 corresponding to predictor variable i using data up until time t, and n denotes the total number of predictors considered in the forecast combination procedure.

Another method for combining the forecasts of predictors is based on their recent out-of-sample performance. In particular, the weight each predictor gets in a period is directly proportional to the sum of its squared forecast errors (SFE) from the previous k periods, where $k \in \{1, 3, 6, 12\}$. Formally, the weight equation is given by:

$$w_{i,t+1} = \frac{\sum_{l=0}^{k-1} SFE_{i,t-l}}{\sum_{l=0}^{k-1} \sum_{j=1}^{n} SFE_{j,t-l}}, \tag{11}$$

where $w_{i,t+1}$ denotes the weight that predictor i gets in period $t+1$, $SFE_{i,t}$ denotes the squared forecast error of predictor i at time t, and n denotes the total number of predictors considered in the forecast combination procedure.

4.6 Factor Investing

For SII, PC, and OS we have equity-specific data, which allows us to construct a factor based on each of these predictors. To construct the factor, the so-called decile approach is utilised. This approach advocates the construction of a portfolio that takes a long position in the most desirable 10% of the stocks, and a short position in the least desirable 10%. As will become apparent later in this paper, all these predictors are measures of pessimism in future stock returns, and hence lower values of these predictors are more desirable. In addition, it is also important to note that in this context, we work with simple rather than log returns, since they allow for summation of returns across all equities contained in a portfolio.

To assess the predictive ability of each predictor, the in- and out-of-sample analysis presented in Sections 4.3 and 4.4 is performed. Note that here we use the aggregate returns of the portfolios constructed using the decile approach, and the aggregation of the predictors is done by taking the difference between the mean of the predictor for the longed stocks and the mean of the predictors for the shorted stocks.

Moreover, we also conduct an asset allocation exercise similar to the one in Rapach et al. (2016). In particular, we assume that a risk-averse investor, with risk-aversion coefficient of $\gamma = 3$, can invest his wealth in either a risky asset or the risk-less asset which yields the risk-free rate. The risky asset considered is one of the three factor portfolios. Moreover, the investor maximizes the utility function of a mean-variance investor. To this end, the weight that the investor allocates to the risky asset at time $t+1$ is given by:

$$w_{t+1} = \frac{1}{\gamma}\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}, \tag{12}$$

where $\hat{r}_{t+1}$ is the one-step ahead forecast for the aggregate portfolio excess return utilising information until time t, and $\hat{\sigma}^2_{t+1}$ is the variance estimate constructed using a five-year moving window. To forecast the returns r_{t+1} we use either the benchmark of the mean returns until time t, or the the predictive regression given in Equation 7. Moreover, as suggested in Rapach et al. (2016), we restrict the weight constructed using Equation 12 to be between -0.5 and 1.5 to take into account the leverage constraints faced by the investor in practice. Moreover, when the forecast for returns aggregated over $k > 1$ months is used to compute the weights in Equation 12, the positions are re-balanced every k months. As a final remark here, the derivation of the weight in Equation 12 is provided in Appendix D.

The remainder of this section is devoted to explaining two methods used to assess the desirability of the strategy that invests in the risky asset a fraction of wealth given by the weight in Equation 12, where the predictive regression in Equation 7 is used to forecast aggregate returns. The first method is a comparative approach, which compares the annualized gain in the Certainty Equivalent Return (CER) measure the investor attains by using the predictive regression over the benchmark when assigning weights. The annualized gain in CER is formally given by:

$$CER\ Gain = 12\ [CER(PR) - CER(BM)], \tag{13}$$

where $CER(PR)$ and $CER(BM)$ are the Certainty Equivalent Return values expressed in percentage of the strategies that use the predictive regression and the benchmark to derive the weights in Equation 12, respectively. For the sake of completeness, these Certainty Equivalent Return values are measured as:

$$CER_P = 100 \left(\bar{R}_P - \frac{1}{2}\gamma\hat{\sigma}_P^2 \right), \tag{14}$$

where $\bar{R}_P$ is the mean excess portfolio return, γ is the risk-aversion parameter which we set to $\gamma = 3$, and $\hat{\sigma}_P^2$ is the variance estimate of the portfolio excess returns.

The second performance assessment method measures the absolute desirability of the strategy via computing the annualized Sharpe Ratio of the strategy in the following way:

$$SR_P = \sqrt{12}\, \frac{\bar{R}_P}{\hat{\sigma}_P}, \tag{15}$$

where, again, $\bar{R}_P$ is the mean excess portfolio return, and $\hat{\sigma}_P$ is the estimated standard deviation of the excess portfolio returns.

5 Trading with Asymmetric Information: The ITA Model

5.1 Theoretical Framework

In this section we propose the Informed Trading Actions (ITA) Model, a novel theoretical framework that models the trading actions of an investor that possesses insider information on a stock. The model builds upon the framework introduced in Johnson and So (2012), however, the model assumptions are changed to capture a more realistic trading environment. Note, however, that for consistency purposes, the notation is kept in line with their notation.

Let us consider a stock market and only two periods in time, namely the current period and a future period, and let us denote them by $t = 1$ and $t = 2$, respectively. Moreover, let us focus on a single stock that is being traded in the market. The price of this stock at $t = 1$ is know to be μ, while its price at $t = 2$ is unknown.

However, it is common knowledge among market participants that its price at $t = 2$ is $\tilde{V} = \mu + \tilde{\epsilon} + \tilde{\eta}$, with $\tilde{\epsilon}$ and $\tilde{\eta}$ being two random variables. Moreover, the market consists of two types of traders: traders that possess private information and traders that do not. The traders that possess private information know the future value of $\tilde{\epsilon}$, but not that of $\tilde{\eta}$. On the other hand, both $\tilde{\epsilon}$ and $\tilde{\eta}$ are unknown to the traders that do not possess the private information.

Moreover, each trader in this market can buy or sell one of these three assets: the stock, a call option on the stock, or a put option on the stock. For the sake of simplicity, we let the call and put traded have the same strike, namely μ, and we consider European options that mature at time $t = 2$. Moreover, each trader can only trade in one of these assets, and the trades occur in pre-determined sizes. In particular, the size for a stock trade is denoted by γ, while the size of an option trade is denoted by θ. Since option trades have a higher leverage as compared to stock trades in practice, it holds that $\theta > \gamma$. As a final remark, when a stock is sold short, a proportion ρ of the total transaction amount is paid as an interest fee, as the trader is borrowing the stock to short-sell it.

It is now pertinent to define the distributions that the random variables $\tilde{\epsilon}$ and $\tilde{\eta}$ follow. Johnson and So (2012) simply assume that both these variables are normally distributed, with zero mean and standard deviations of $\sigma_{\tilde{\epsilon}}$ and $\sigma_{\tilde{\eta}}$, respectively. The assumption of zero mean is valid since returns typically have zero mean, and hence the expected price at time $t = 2$ should be the same as the price at $t = 1$. However, the assumption of normality might not be plausible in practice. In particular, it is a well-known stylized fact in finance that the distribution of stock returns is negatively-skewed and exhibits excess kurtosis (e.g. Officer (1972), Hagerman (1978)). The main implication of this is that large negative returns occur more often than large positive returns of the same magnitude. Moreover, this means that the distribution of $\tilde{V}$ is negatively-skewed, and it exhibits excess kurtosis. Since it holds that $\tilde{V} = \mu + \tilde{\epsilon} + \tilde{\eta}$, it follows that the distribution of $\tilde{\epsilon} + \tilde{\eta}$ must also satisfy these features.

In addition, since $\tilde{\eta}$ is the component of future price that is not known to any of the market participants, it can be assumed that it also follows a negatively-skewed distribution with fat tails. Unfortunately, little can be deduced about the distribution of $\tilde{\epsilon}$ by the information assumed so far. Furthermore, there is no literature that studies the distribution of the signal possessed by insiders. Overall, so far we have assumed that $\tilde{\epsilon} \sim D^{\tilde{\epsilon}}(0, \sigma_{\tilde{\epsilon}})$ and $\tilde{\eta} \sim D^{\tilde{\eta}}(0, \sigma_{\tilde{\eta}})$, where we do not know the shape of $D^{\tilde{\epsilon}}$, but we know that $D^{\tilde{\eta}}$ is a negatively-skewed distribution with fat tails. The fact that we miss information regarding the shape of the distribution of $\tilde{\epsilon}$ makes it impossible to assume an appropriate distribution that would match its features, and in turn this makes it impossible to derive a theoretical model equilibrium in which the traders interact with market makers that offer different bid and ask prices for each asset.

However, we can still make some inferences regarding the trades executed by traders that possess private information. Note, nonetheless, that we assume the same bid and ask prices for each asset (the stock, the call and the put). In other words, we assume that there is no bid-ask spread for the assets considered in this analysis. For informed traders, the expected value of the price of the stock at time $t = 2$, given their private signal $\tilde{\epsilon} = \epsilon$, is expressed as:

$$E[\tilde{V}|\tilde{\epsilon} = \epsilon] = \mu + \epsilon \tag{16}$$

To get a better grasp as to what asset an informed trader would invest in, Figures 1 and 2 show the profit an insider gets for each type of trade he can execute. For the sake of visualisation, we assume that $\mu = 100$ in the plots. The full derivations for the general framework can be found in Appendix E.

Figure 1 shows that, when an insider possesses a positive signal $\tilde{\epsilon} > 0$, he will sell puts if the signal is small, in an attempt to extract the premium. The insider buys the stock if the signal is medium-sized, and the insider buys calls if the signal is large, such that he exploits the leverage advantage that call options offer. Furthermore, Figure 2 illustrates the trading actions of an insider when he possesses a negative signal $\tilde{\epsilon} < 0$. In this case, the insider sells call options if the signal size is small, sells the stock if the signal is medium-sized, and buys put options if the signal size is large.

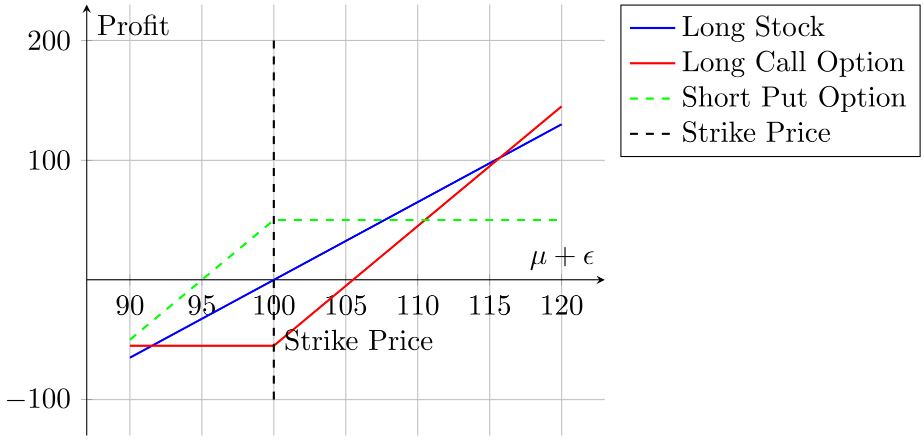

Figure 1: Expected Profit Functions for Positive Signal

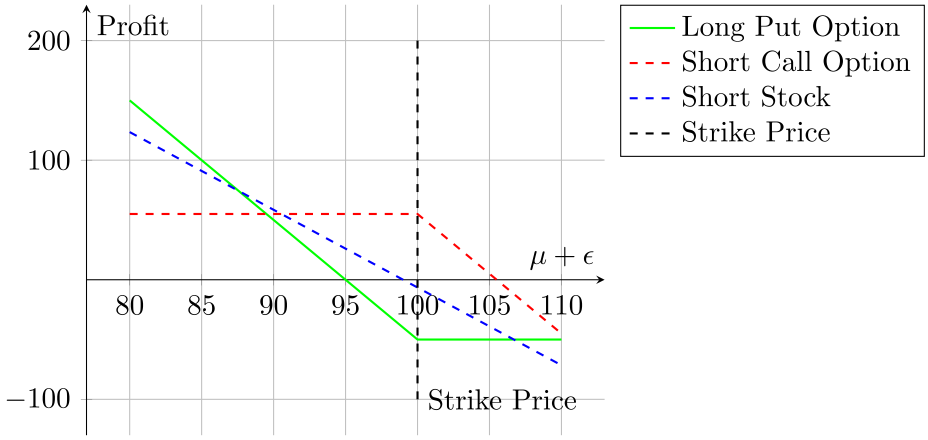

Figure 2: Expected Profit Functions for Negative Signal

Given the framework presented in this section, an informed trader would execute the following trading strategies depending on the value of the signal $\tilde{\epsilon} = \epsilon$ he possess:

$$f(\epsilon) = \begin{cases} \text{buy put} & \text{if } \epsilon < k_1 \\ \text{sell stock} & \text{if } k_1 < \epsilon \leq k_2 \\ \text{sell call} & \text{if } k_2 < \epsilon \leq k_3 < 0 \\ \text{make no trade} & \text{if } k_3 < \epsilon \leq k_4 \\ \text{sell put} & \text{if } 0 < k_4 < \epsilon \leq k_5 \\ \text{buy stock} & \text{if } k_5 < \epsilon \leq k_6 \\ \text{buy call} & \text{if } \epsilon > k_6, \end{cases} \tag{17}$$

where, as mentioned before, the values of $k_1, \ldots, k_6$ can be found in Appendix E.

5.2 Implications for SII, OS, and PC

In this section we discuss the theoretical implications that the ITA model presented in Section 5.1 has for the power of SII, OS and PC in predicting aggregate returns. Let us start with SII.

If informed investors possess a medium-sized negative signal for some stock, they will short the stock. So, high levels of SII are associated with negative expected returns for that stock.

The implications for OS and PC are a bit more convoluted. In particular, OS compares the volume of options to the volume of stocks being traded. From Equation 17 it follows that an insider would trade options if he possess a weak or strong signal (either negative or positive), and the insider would trade the stock if he possesses a medium signal (either negative or positive). This means that OS will be high either when there is a strong or weak signal, and OS will be low when there is a medium signal. Hence, OS does not convey any clear-cut information regarding the information possessed by the insiders, and hence we expect it to predict future aggregate returns poorly.

On the other hand, PC is a better-constructed variable. To see this, consider the following. Insiders trade puts either when they possess a strong negative signal, or when they possess a weak positive signal. Moreover, insiders trade calls either when they possess a strong positive signal, or a weak negative signal. To this end, high PC values are associated with either high negative or low positive expected returns. In addition, low PC values are associated with either high positive or low negative expected returns. Due to this feature of PC, it is likely that PC can explain extreme values of returns to some extent. Hence, we expect PC to contain some predictive power on aggregate returns. For illustration purposes, consider Figure 3.

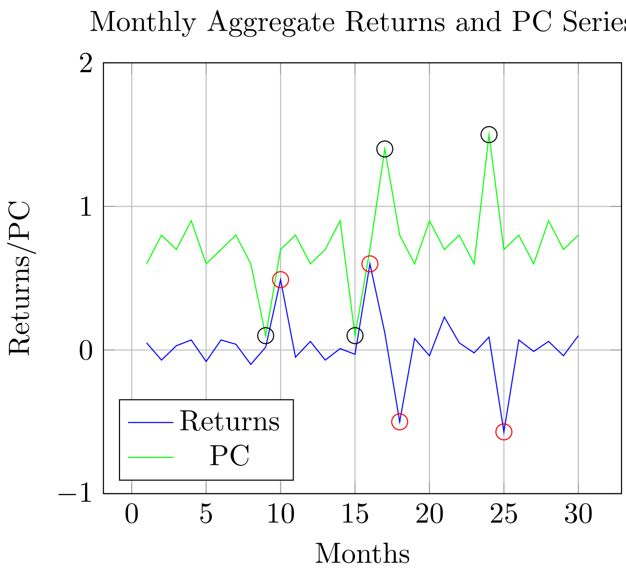

Figure 3: Illustration of Monthly Returns and PC Series with Extreme Values Highlighted

Figure 3 plots some randomly generated monthly aggregate returns. The return series fluctuates around the mean of zero, with two extremely positive and two extremely negative aggregate returns occurring. On the other hand, the PC series fluctuates around a mean of 0.7. The crucial observation is that right before the two extremely positive returns, PC attains extremely low values. On the contrary, right before the extremely low returns, PC attains extremely high values. In theory, this should occur for each stock traders have insider information on, as well as for portfolios that contain such stocks. Overall, due to the aforementioned reasoning, we expect PC to have some predictive power over aggregate returns of stocks traded by insiders, and portfolios consisting at least in part by such stocks.

6 Results

In this section we present and interpret the results of this paper. We begin with the findings for the market portfolio, and then proceed to the results for individual equities. Finally, we discuss the outcomes of the factor investing framework.

6.1 Market Portfolio Results

6.1.1 In-sample Results

We start by presenting our findings on the in-sample predictive power of all variables on the aggregate market returns. The results are summarised in Table 1.

Table 1: In-sample Results for Aggregate Market Returns

Predictor	$\hat{\beta}$	R^2 (%)	$\hat{\beta}$	R^2 (%)	$\hat{\beta}$	R^2 (%)	$\hat{\beta}$	R^2 (%)
	$h = 1$		$h = 3$		$h = 6$		$h = 12$	
SII	-0.68	2.26	-0.72	7.70	-0.69	12.43	-0.57	16.05
	$[-1.92]^{**}$		$[-1.90]^{*}$		$[-1.83]^{*}$		$[-2.13]^{*}$	
OS	0.13	0.08	-0.04	0.02	-0.03	0.03	-0.15	1.05
	[0.49]		$[-0.17]$		$[-0.16]$		$[-0.87]$	
PC	-0.47	1.09	-0.42	2.68	-0.40	4.18	-0.15	1.76
	$[-1.37]$		$[-1.84]^{*}$		$[-2.25]^{**}$		$[-1.03]$	
CV	0.22	0.25	-0.00	0.00	0.21	1.14	0.28	3.80
	[0.45]		$[-0.00]$		[1.52]		[3.15]	
VP	0.18	0.17	0.41	2.44	0.34	3.11	0.28	4.02
	[0.46]		$[1.67]^{*}$		$[2.30]^{*}$		[2.35]	
BB_{binary}	0.00	0.00	-0.01	0.00	0.07	0.14	0.17	1.53
	[0.03]		$[-0.06]$		[0.71]		$[2.51]^{***}$	
BB_{ratio}	-0.15	0.11	-0.00	0.00	0.12	0.41	0.25	3.17
	$[-0.58]$		$[-0.01]$		[0.77]		$[1.95]^{*}$	
BB_{spread}	-0.18	0.16	-0.05	0.04	0.10	0.26	0.24	2.84
	$[-0.78]$		$[-0.34]$		[0.68]		$[2.07]^{*}$	
MVMF	-0.01	0.00	-0.27	1.07	-0.39	4.10	-0.53	13.88
	$[-0.02]$		$[-0.95]$		$[-1.95]^{*}$		$[-3.07]^{**}$	

Note: The results presented in this table are for the period Jan 2001 - Jan 2022. This table reports the results of estimating the equation $r_{t:t+h} = \alpha + \beta x_t + \epsilon_{t:t+h}$ using ordinary least squares, where $r_{t:t+h}$ represents the market returns aggregated from $t+1$ until $t+h$, with $h \in \{1, 3, 6, 12\}$, and x_t denotes the value of the predictor variable at time t. Note that all the predictors are standardized to have a standard deviation of one for comparison purposes. The table reports the estimated slope coefficient $\hat{\beta}$, the regression R^2, the Newey-West Standard Error (given inside square brackets below the $\hat{\beta}$). The significance of $\hat{\beta}$ as measured by the bootstrapped p-value is presented next to the Newey-West Standard Error. The null hypothesis being tested is $H_0 : \beta = 0$, and the alternative is one-sided, meaning that $H_A : \beta < 0$ if $\hat{\beta} < 0$ or $H_A : \beta > 0$ if $\hat{\beta} > 0$. * represents a p-value between 0.05 and 0.10, ** represents a p-value between 0.01 and 0.05, and *** represents a p-value smaller than 0.01.

The first observation that can be drawn from Table 1 is that the only predictor that significantly explains the aggregate market returns at all return aggregation levels considered is SII. Moreover, the estimated β coefficients corresponding to this predictor are larger than those corresponding to all other predictors at all aggregation levels. All $\hat{\beta}$-s are negative, which is in line with the theory that a higher level of stock shorting indicates lower future returns in the market. Moreover, SII also has the highest R^2 values, meaning that it is the predictor that is best capable of fitting aggregate market returns in-sample. Overall, the in-sample tests seem to lead to a clear-cut preferred predictor, which is SII.

However, it is worth noting that there are also other predictors that can significantly explain aggregate market returns at certain aggregation levels. In particular, PC is able to deliver sizeable β estimates for all aggregation levels, but statistically significant estimates only for the three- and six-month aggregation levels. Note that the signs of these estimates are in line with the implications of the ITA model. On the other hand, as indicated by the model, OS cannot explain the market returns at any aggregation level.

Moreover, even though CV has positive sizeable $\hat{\beta}$ coefficients (except for $h = 3$), it cannot significantly explain aggregate market returns. On the other hand, as indicated in Bekaert and Hoerova (2014), VP has some predictive power over aggregate market returns. In particular, a higher VP is significantly associated with higher aggregate returns over the next three and six months.

Finally, we restrict our attention to the performance of market sentiment predictors. BB_{binary}, BB_{ratio}, and BB_{spread} can only significantly predict the aggregate market return over the next 12 months. Interestingly, $\hat{\beta} > 0$ for all these predictors, indicating that the more investors believe that market will be bearish rather than bullish, the higher the market returns are in the future. Similarly, higher inflow of funds into US mutual funds, as measured by MVMF, leads to significantly lower market aggregate returns over the next six and 12 months. Both these results are in line with the findings of Frazzini and Lamont (2008), which indicate that individual investors lack the skill of timing their investments properly.

6.1.2 Out-of-sample Results

In-sample tests provide an understanding of the ability of predictors to explain historical returns. However, from the perspective of investors and policy-makers, being able to predict future aggregate market returns is crucial. Hence, out-of-sample tests are vital to assess the performance of predictors. The summarised results of these tests are presented in Table 2, where a linear specification[2] is used to construct the forecasts for each predictor.

Table 2: Out-of-sample Results for Aggregate Market Returns

Predictor	**Out-of-sample R^2 statistic (%)**				**Encompassing λ statistic**			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
SII	0.43^{**}	5.56^{***}	-0.99^{***}	16.04^{***}	–	–	–	–
OS	-3.28	-6.31	-9.58	-7.40	0.67^{***}	0.65^{***}	0.56^{***}	0.73^{***}
PC	-0.03	1.78^{*}	4.39^{**}	5.91^{*}	0.55^{**}	0.58^{***}	0.44^{***}	0.64^{***}
CV	-2.63	-3.34	-16.22	-3.57	0.70^{***}	0.63^{***}	0.61^{***}	0.72^{***}
VP	-3.66	0.57	-5.24	-2.82	0.84^{***}	0.58^{***}	0.53^{***}	0.71^{***}
BB_{binary}	-1.88	-2.94	-1.04	2.41	0.65^{***}	0.62^{***}	0.50^{***}	0.68^{***}
BB_{ratio}	-1.63	-0.94	0.52	6.22^{**}	0.63^{***}	0.60^{***}	0.49^{***}	0.65^{***}
BB_{spread}	-1.30	-0.98	0.31	6.16^{**}	0.61^{***}	0.60^{***}	0.49^{***}	0.64^{***}

Continued on next page

[2] We tried to add a quadratic term, a cubic term, or both; to capture possible non-linear relationships between returns and predictors; but forecast errors increased substantially. Hence, we do not present these results here.

Table 2 – continued from previous page

Predictor	**Out-of-sample R^2 statistic (%)**				**Encompassing λ statistic**			
	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
MVMF	-1.49	-3.44	-1.58	3.01	0.66^{***}	0.65^{***}	0.51^{***}	0.62^{***}
FC_1	0.24	6.59^{***}	10.27^{***}	15.73^{***}	–	–	–	–
FC_2[3]	-1.30	-3.40	-4.61	-6.25	–	–	–	–

Note: The results presented in this table are for the period Jan 2006 - Jan 2022. The table presents the out-of-sample R^2 statistics (R^2_{OS}), as defined in Section 4.4, for each predictor, which compares the one-step ahead aggregate market return forecasts using the linear specification $\hat{r}_{t:t+h} = \hat{\alpha}_t + \hat{\beta}_t x_t$ to the forecasts of the mean return benchmark. In the aforementioned specification, $\hat{r}_{t:t+h}$ represents the one-step ahead forecasted return, and $\hat{\alpha}$ and $\hat{\beta}$ represent the coefficients of the lienar model fitted using data up until time t. Next to the R^2_{OS} statistic, the significance of the Clark and West (2007) test statistic that tests the null hypothesis $H_0 : R^2_{OS} \leq 0$ against $H_A : R^2_{OS} > 0$ is reported; where $*$ indicates a p-value between 0.05 and 0.10, $**$ indicates a p-value between 0.01 and 0.05, and $***$ indicates a p-value lower than 0.01. Moreover, the table also presents the optimal coefficient (referred to as λ) allocated to the SII predictor when combining the forecasts of SII and another predictor using a convex combination of forecasts, as developed in Harvey et al. (1998). Next to each λ coefficient, we also report the significance of the test developed in Harvey et al. (1998) that tests $H_0 : \lambda = 0$ against $H_A : \lambda > 0$. The meaning of the symbols $*$, $**$, and $***$ remains the same as for the Clark and West (2007) test. FC_1 and FC_2 stand for the combination of forecasts as defined in Equations 10 and 11, respectively.

The main takeaway from Table 2 is that the predictors with the best out-of-sample performance are SII and PC. SII achieves the highest R^2_{OS} compared to all other predictors for the aggregation levels of one, three and 12 months, however, it performs poorly in predicting six-month aggregate returns. In particular, its R^2_{OS} is negative, indicating that the predictions of SII are outperformed by those of the mean return benchmark. However, one should note that the Clark and West (2007) test results still indicate that the SII predictions for the six-month aggregate returns significantly outperform those of the benchmark. This is probably due to the large differences between the forecasts of SII and those of the benchmark, which drive the test statistic up, and in turn, the p-value of the test down.

The out-of-sample performance of PC, on the other hand, is generally worse than that of SII, but it is more stable. In particular, the R^2_{OS} of PC is increasing with the level of aggregation. It is about zero for $h = 1$, and about 5.91 for $h = 12$. Moreover, PC outperforms SII for $h = 6$. This is reflected in a higher R^2_{OS} as well as the fact that $\lambda = 0.44 < 0.50$, indicating that the optimal forecast combination between the forecasts of SII and PC allocates a larger weight on PC.

While SII and PC beat the benchmark in three out of four cases, neither CV nor VP can significantly outperform the benchmark for any aggregation level of market returns. Apart from VP for $h = 3$, they actually always attain negative R^2_{OS} statistics. On the other hand, the market sentiment variables seem to perform well out-of-sample for high levels of market return aggregation. In particular, they all realise sizeable R^2_{OS} statistics for $h = 12$, with BB_{ratio} and BB_{spread} realising higher statistics than PC, and statistically higher values than zero at the 5% level. Another interesting result is that the optimal weight allocated to all the market sentiment variables when combining their forecasts with those of SII is typically about 0.4, indicating that their forecasts provide considerable information on top of those of SII.

Next, we comment on the performance of forecast combinations. FC_1 gives positive R^2_{OS} statistics for all aggregation levels, which are sizeable and statistically significant for all aggreg-

[3]Here we only present the results for $k = 6$, as the past information incorporated using this choice is neither too long, nor too short. The results for $k \in \{1, 3, 12\}$ are similar.

ations levels except for $h = 1$. On the other hand, FC_2 fails to outperform the benchmark at all aggregation levels, with its relative performance becoming increasingly worse with the increase in the return aggregation level h. This can possibly be explained by its backward-looking structure, and the intuition that if a predictor performed well out-of-sample in the near past, it will continue to do so in the near future. On the contrary, FC_1 utilises data on the entire past in-sample performance of each predictor, thereby testing the predictors' effectiveness in capturing various patterns in the past aggregate market returns.

Lastly, we compare the R^2_{OS} statistics of SII and FC_1, to investigate whether utilising all predictors at hand is more beneficial than only using SII. The results in Table 2 do not suggest a definitive winner, as the best-performing predictor depends on the aggregation level used. In particular, SII performs exceptionally well for the short-term future ($h = 1$), being the only predictor that significantly outperforms the benchmark. Moreover, SII also performs slightly better for the long-term aggregate returns ($h = 12$). However, for the market returns aggregated over three and six months, which represent the medium-term, FC_1 performs substantially better, with the most noticeable difference occurring for $h = 6$. Overall, investors and policy-makers should choose between SII and FC_1 depending on the time horizon of their strategies or policies.

6.2 Individual Equity Results

In this section we restrict the attention to analysing the predictive power of SII, OS, and PC for the aggregate individual equity returns. Recall that the analysis is only performed for these three predictors because they are the only ones that can be constructed for each equity individually. On the other hand, the other variables can only be used to predict the aggregate market returns.

To analyse the predictive performance of SII, OS, and PC on aggregate individual equity returns, we utilise the in-sample and out-of-sample tests presented in Sections 4.3 and 4.4, respectively. These tests are conducted for each individual equity in the cross-section, and the statistics corresponding to each equity are then aggregated into a single statistic for ease of interpretation. Finally, for each time period we rank the stocks according to the value of the predictor they attain in that period, and evaluate the accuracy of the one-step ahead forecasts of the aggregate returns of the top and bottom decile stocks only.

6.2.1 In-sample Results

Table 3 presents the results of the in-sample predictive power of SII, OS, and PC for aggregate individual equity returns. The details of the tests conducted can be found in Section 4.3. The analysis is conducted for each equity individually to obtain the $\hat{\beta}$ coefficient corresponding to each predictor and its associated p-value using the bootstrap procedure given in Appendix A. To obtain interpretable statistics, we report the mean $\hat{\beta}$ and mean p-value, as well as the proportion of stocks that have a p-value smaller than 10%, 5%, and 1% for each predictor.

Table 3: In-sample Results for the Full Cross-Section of Equities

			$h = 1$		
Predictor	$\hat{\beta}$	**P-value**	% **P-value** < 0.10	% **P-value** < 0.05	% **P-value** < 0.01
SII	0.04	0.26	18.06	8.88	1.63
OS	−0.71	0.23	25.81	15.35	4.47
PC	−0.01	0.25	19.71	9.54	2.28
			$h = 3$		
Predictor	$\hat{\beta}$	**P-value**	% **P-value** < 0.10	% **P-value** < 0.05	% **P-value** < 0.01
SII	0.01	0.26	17.88	9.56	1.75
OS	−0.66	0.22	27.74	15.65	5.18
PC	−0.01	0.25	21.16	9.96	1.66
			$h = 6$		
Predictor	$\hat{\beta}$	**P-value**	% **P-value** < 0.10	% **P-value** < 0.05	% **P-value** < 0.01
SII	0.02	0.26	18.00	9.63	2.31
OS	−0.54	0.23	26.12	16.06	4.17
PC	−0.05	0.25	21.16	9.54	2.07
			$h = 12$		
Predictor	$\hat{\beta}$	**P-value**	% **P-value** < 0.10	% **P-value** < 0.05	% **P-value** < 0.01
SII	0.06	0.26	18.69	10.88	2.81
OS	−0.44	0.24	24.59	13.52	3.96
PC	0.01	0.27	18.26	9.34	3.11

Note: The results presented in this table are for the period Jan 2001 - Jan 2022. The table presents the results of fitting the linear model $r_{t:t+h} = \alpha + \beta x_t + \epsilon_{t:t+h}$, for $t = 1, \ldots, T$ and $h \in \{1, 3, 6, 12\}$, for each individual equity contained in the sample. Here $r_{t:t+h}$ represents the aggregated return as defined in Section 4.1 for an individual equity, and x_t refers to the value of the predictor variable for the equity under analysis at time t. Note that for comparison purposes, x_t is standardized to have a standard deviation of one. Since the analysis is performed for multiple equities, the statistics obtained for each equity are aggregated into a single measure. In particular, $\hat{\beta}$ represents the mean fitted β coefficient for all equities in the sample. P-value also refers to the mean p-value associated with $\hat{\beta}$ for all stocks in the sample, where the p-value for each equity is constructed using the wild bootstrap procedure explained in Appendix A. Finally, the other three statistics presented refer to the percentage of stocks in the sample that attain a p-value smaller than 0.10, 0.05 or 0.01, respectively.

The results in Table 3 show that all the predictors perform poorly for individual equities. In particular, the mean p-value for the β coefficient is above 0.20, suggesting that these predictors cannot significantly explain aggregate equity returns in-sample for the majority of equities. Moreover, the proportion of equities that attain a p-value below the 10%, 5%, and 1% threshold is relatively small. The highest such statistic is attained by using the OS predictor for the three-month aggregate returns at the 10% threshold, and it is still below 30%. Note, however, that this result makes sense intuitively. The predictive power of SII, OS, and PC should in theory arise due to the fact that these variables incorporate information regarding the trades executed by insiders. However, insiders do not possess information regarding every equity traded, but rather a portion of them. Therefore, SII, OS, and PC cannot significantly explain aggregate returns for the complete cross-section of stocks.

Another important result is that OS is the best-performing predictor for individual equities in-sample. This predictor attains a sizeable mean $\hat{\beta}$. The negative sign of the mean $\hat{\beta}$ suggests that a higher volume of options traded as compared to stocks typically leads to negative future returns. Furthermore, OS achieves the lowest mean p-value, and the highest proportions of significant $\hat{\beta}$-s at all thresholds and return aggregation levels. This is an interesting result, in the sense that it seems to contradict the theoretical implication of the ITA model that OS should be performing poorly in predicting aggregate returns. However, diving deeper into this, leads to two possible rationales.

The first one is related to the fact that the cross-section of stocks used in the analysis of each predictor is different. First, the data set used to construct SII is different from the one used to construct OS and PC. Moreover, when stochastically de-trending SII and PC, stocks that have less than 60 observations are dropped out of the analysis. So, the first possibility is that OS performs better than SII and PC simply due to chance.

The second rationale is a theoretical implication of the ITA model when the distribution of the signal $\tilde{\epsilon}$ is negatively-skewed. To facilitate the explanation, we provide an illutration for the distribution of $\tilde{\epsilon}$ below.

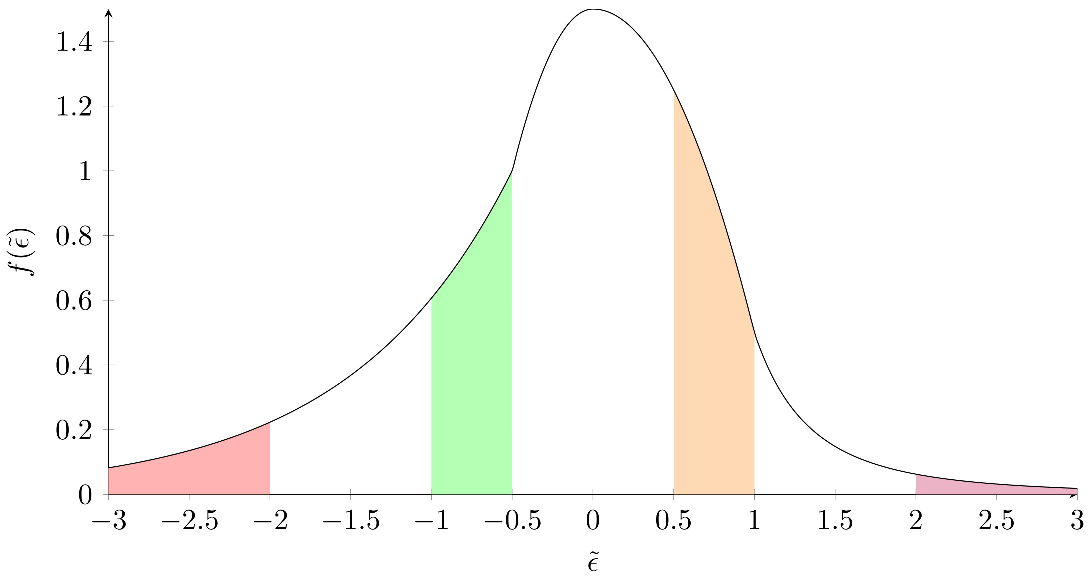

Figure 4: Illustration of a negatively-skewed probability density function for $\tilde{\epsilon}$

Note: This figure provides an illustration of a negatively-skewed probability density function (PDF) for the signal $\tilde{\epsilon}$. Under this distribution, large negative occurrences of $\tilde{\epsilon}$ take place more often than large positive ones of the same magnitude. This is illustrated by the area under the PDF in the interval $(-3, -2)$ being larger than the one in the interval $(2, 3)$. Moreover, small positive occurrences take place more often than small negative ones of the same magnitude. This is illustrated by the area under the PDF in the interval $(0.5, 1)$ being larger than the one in the interval $(-1, -0.5)$.

As illustrated in Figure 4, small positive signals occur more often than small negative ones of the same magnitude. Recall that, when a small positive signal occurs, the insider sells puts, whereas they sell calls in response to a small negative signal. So, a higher OS value is more often associated to small positive expected returns than to small negative ones.

Conversely, large negative signals occur more often than large positive ones of the same magnitude. In case of a large negative signal, the insider buys puts, while they buy calls in response to a large positive signal. Using a similar logic, a higher OS value is associated with large negative expected returns more often than large positive ones.

Overall, the assumption of a negatively-skewed $\tilde{\epsilon}$ introduces the following trade-off. High OS values are followed by large negative returns more often than large positive returns. However, high OS values are also followed by small positive returns more often than small negative returns. As argued in Section 5.2, it is the ability to anticipate large-magnitude returns that matters the most, and in turn, it is this ability that provides the OS variable with some predictive power.

The assumption of a negatively-skewed distribution for $\tilde{\epsilon}$ does not have any effect on PC. In particular, it leads to a higher mean for PC, but since we stochastically de-trend PC, there is no real effect. Overall, under this new assumption, OS should gain some predictive power, nevertheless, PC still remains a more direct measure to incorporate information regarding the trades of insiders.

However, when the analysis is conducted at equity level, PC is more prone to noise than OS. This is because changes in hedging strategies by market participant influence PC more than OS. To see this, recall that the accredited investors, which drive the market with their trades, tend to dynamically update their strategies within a market, but rarely change their strategies across markets. To provide a concrete example, when traders switch from trading calls to trading puts of the stock due to a change in their hedging strategies, PC increases, while OS remains unchanged. This higher sensitivity to noise causes PC to be outperformed by OS in predicting individual equities. As a final remark, recall that PC performed better than OS for market returns. That is because aggregation across stocks removes the noise contained in PC, and therefore, PC gains superior performance in predicting aggregate market returns.

6.2.2 Out-of-sample Results: Full Cross-Section

Table 4 presents the results of the out-of-sample predictive power of SII, OS, and PC for aggregate individual equity returns of the full cross-section. To this end, the methodology introduced in Section 4.4 is utilised. In particular, we use a linear specification with each predictor as the explanatory variable to construct one-step ahead forecasts on aggregate returns. After that, we compute the R^2_{OS} statistic for each equity, and its associated p-value as measured by the Clark and West (2007) test. Finally, to make interpretation easier, we report the mean R^2_{OS} statistic and the mean p-value, as well as the proportion of stocks that have a p-value smaller than 10%, 5%, and 1%.

Table 4: Out-of-sample Results for the Full Cross-Section of Equities

	$h = 1$				
Predictor	R^2_{OS}	**P-value**	% **P-value** < 0.10	% **P-value** < 0.05	% **P-value** < 0.01
SII	-1.20	0.56	3.26	1.71	0.38
OS	-3.73	0.53	5.99	2.53	0.53
PC	-1.22	0.53	2.58	1.55	0.78

Continued on next page

Table 4 – continued from previous page

			$h = 3$		
Predictor	R^2_{OS}	**P-value**	% **P-value** < 0.10	% **P-value** < 0.05	% **P-value** < 0.01
SII	-3.93	0.55	4.81	2.72	0.78
OS	-8.04	0.51	7.56	4.53	1.33
PC	-1.29	0.47	5.94	3.10	1.29
			$h = 6$		
Predictor	R^2_{OS}	**P-value**	% **P-value** < 0.10	% **P-value** < 0.05	% **P-value** < 0.01
SII	-4.53	0.53	6.60	4.66	1.48
OS	-15.56	0.48	10.65	5.59	2.00
PC	-1.46	0.47	10.33	6.98	1.81
			$h = 12$		
Predictor	R^2_{OS}	**P-value**	% **P-value** < 0.10	% **P-value** < 0.05	% **P-value** < 0.01
SII	-12.47	0.50	9.70	6.44	2.88
OS	-32.85	0.46	14.51	9.59	3.60
PC	-1.96	0.43	13.70	8.01	3.62

Note: The results presented in this table are for the period Jan 2006 - Jan 2022. The table presents the results of using the linear model $r_{t:t+h} = \alpha + \beta x_t + \epsilon_{t:t+h}$, for $t = 1, \ldots, T$ and $h \in \{1, 3, 6, 12\}$, and where x_t is the value of the predictor at time t, to construct one-step ahead forecasts for the aggregate returns $r_{t:t+h}$, defined in Section 4.1. The one-step ahead forecasts are constructed as $\hat{r}_{t:t+h} = \hat{\alpha} + \hat{\beta} x_t$, where $\hat{\alpha}$ and $\hat{\beta}$ are the ordinary least squares parameter estimates of the linear model using data until time t. R^2_{OS} denotes the mean out-of-sample R^2 statistic as defined in Section 4.4, across all the stocks contained in the sample. Similarly, P-value denotes the mean p-value of the Clark and West (2007) test, as explained in detail in Appendix B. The remaining three columns represent the percentage of stocks in sample that significantly outperform the benchmark at the 10%, 5%, and 1% significance levels, respectively, as indicated by the p-value of the Clark and West (2007) test statistic.

The results presented in Table 4 show that none of the predictors is able to provide forecasts for the returns of individual equities that beat the mean return benchmark at any aggregation level. This is in line with the poor in-sample performance of all predictors, as reported in Section 6.2.1. The poor out-of-sample performance is reflected in the negative R^2_{OS} statistic for all predictors at all aggregation levels h, as well as in the very high mean p-value for the Clark and West (2007) test. In a similar line of reasoning as for the poor in-sample performance, this result makes intuitive sense since the predictive power of SII, OS, and PC can be theoretically attributed to the reflection of information regarding the trades executed by insiders. Since insiders generally possess information on a limited number of stocks, the out-of-sample performance of these predictors across the full cross-section is poor.

Another crucial observation that should be drawn from Table 4 is regarding the predictor that performs the best out-of-sample for individual equities. Overall, there is no clear-cut answer to this, however, it can be said that SII is outperformed by OS and PC across all statistics. We first discuss the advantages of using PC, and then explain the advantages of using OS.

The mean R^2_{OS} statistic obtained when using PC is substantially lower as compared to using SII or OS, especially at higher aggregation levels. In particular, PC generates a mean R^2_{OS} between -2% and -1%, depending on the return aggregation level, while SII and OS, on the other hand, attain R^2_{OS} statistics as high as about -13% and -33%, respectively. Moreover, PC also has the lowest mean p-value for the Clark and West (2007) test at all return aggregation levels, although this statistic is sizeable.

On the contrary, using OS results in the highest proportion of stocks that significantly outperform the benchmark. In particular, the proportion of stocks that attains a p-value lower than 10% for the Clark and West (2007) test is the highest for OS at all return aggregation levels. Furthermore, OS attains the highest proportion of stocks with p-value lower than 5% for three out of four return aggregation levels, and the highest proportion of stocks with p-value lower than 1% for two of the return aggregation levels.

Even though these findings initially seem convoluted, they align well with intuitive expectations if the distribution of the signal possessed by insider traders $\tilde{\epsilon}$ exhibits negative skewness. In particular, recall the second explanation provided in the previous section. Under this assumption, PC still remains a more direct measure of the trades executed by insiders, however, it is more prone to noise for individual equities. To this end, when considering the mean R^2_{OS} statistic, which is an aggregate measure of out-of-sample performance, and therefore should partially remove the noise contained in PC, PC outperforms OS. On the other hand, due its relatively higher stability, OS can predict the future returns of a higher number of stocks that were traded by insiders. This is reflected in the higher percentage of stocks whose returns OS predicts significantly better than the benchmark at various return aggregation levels. As a final remark, note that OS and PC outperform the benchmark at the 1% significance level for a similar proportion of stocks. This again makes sense in this framework, since for stocks heavily traded by insiders, the noise contained in PC causes less erosion in its aggregate return predictive ability.

6.2.3 Out-of-sample Results: Top & Bottom Deciles

As argued in the previous sections, it makes sense that none of the predictors can explain aggregate returns of the majority of stocks in- and out-of-sample, since insiders only trade in a subset of these stocks. This raises the natural question of whether SII, OS, and PC can predict the future aggregate returns of the stocks that have been traded by insiders. To this end, we need some mechanism that determines which stocks are traded by insiders. For this purpose, we evaluate the out-of-sample performance of these predictors for stocks that attain extreme values on them, as the best proxy in our possession for stocks traded by insiders. The intuition behind this proxy is that if insiders possess information regarding a stock, they will execute trades in assets related to this stock (in line with Equation 17). Therefore, this in turn will be reflected in the SII, OS, or PC values of that stock, and depending on the trade executed, push this stock towards the upper or lower ranking extremes.

The procedure we conduct is the following. For every time period, we rank all the stocks traded in that period in ascending order of the predictor considered, and select the top and bottom deciles according to this ranking. For each selected stock, we make sure there is a return history of at least 60 months to obtain reliable ordinary least squares estimates. If that is not the case, the stock is dropped out of the analysis. For each of the remaining stocks, a one-step ahead forecast of the stock's aggregate returns is computed and stored. This procedure is conducted for every time period, and the forecasts are compared to the mean return benchmark, to obtain the R^2_{OS} statistic, as defined in Section 4.4. To compute $R^2_{OS,top}$, only the forecasts of the stocks in the top decile are used. Similarly, to compute $R^2_{OS,bottom}$, only the forecasts of the stocks in the bottom decile are used. For $R^2_{OS,total}$, the forecasts of stocks in both the top and bottom deciles are used. The results are summarised in Table 5.

Table 5: Out-of-sample Results for the Top and Bottom Decile Equities

	$h=1$			$h=3$		
Predictor	$R^2_{OS,total}$	$R^2_{OS,top}$	$R^2_{OS,bottom}$	$R^2_{OS,total}$	$R^2_{OS,top}$	$R^2_{OS,bottom}$
SII	-3.21	-3.97	-2.68	-9.79	-10.94	-8.99
OS	-5.47	-0.12^{*}	-5.76	-11.67	-0.19	-12.22^{*}
PC	-4.00	-4.49	-3.56	-5.39	-7.47	-3.86
	$h=6$			$h=12$		
Predictor	$R^2_{OS,total}$	$R^2_{OS,top}$	$R^2_{OS,bottom}$	$R^2_{OS,total}$	$R^2_{OS,top}$	$R^2_{OS,bottom}$
SII	-16.66	-20.19	-14.28	-30.61	-40.54	-23.35
OS	-20.90	0.67^{**}	-21.83	-37.28^{***}	3.21	-39.15^{***}
PC	-5.01	-7.62	-2.99	-6.17	-11.52	-2.60

Note: The results presented in this table are for the period Jan 2006 - Jan 2022. The table presents the results of using the linear model $r_{t:t+h} = \alpha + \beta x_t + \epsilon_{t:t+h}$, for $t = 1, \ldots, T$ and $h \in \{1, 3, 6, 12\}$, and where x_t is the value of the predictor at time t, to construct one-step ahead forecasts for the aggregate returns $r_{t:t+h}$ defined in Section 4.1. The one-step ahead forecasts are constructed as $\hat{r}_{t:t+h} = \hat{\alpha}_t + \hat{\beta}_t x_t$, where $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the ordinary least squares parameter estimates of the linear model using data until time t. $R^2_{OS,total}$ denotes the out-of-sample R^2 statistic, as defined in Section 4.4, for the series of one-step ahead forecasts of all stocks that are in the top or bottom decile of the predictor, at each time period t. Similarly, $R^2_{OS,top}$ and $R^2_{OS,bottom}$ denote the out-of-sample R^2 statistic for the series of one-step ahead forecasts of all stocks that are only in the top predictor decile and only in the bottom predictor decile, respectively, at each time period t. Next to each R^2_{OS} statistic, its significance according to the Clark and West (2007) that tests the null hypothesis $H_0 : R^2_{OS} \leq 0$ against $H_A : R^2_{OS} > 0$ is reported; where $*$ indicates a p-value between 0.05 and 0.10, $**$ indicates a p-value between 0.01 and 0.05, and $***$ indicates a p-value lower than 0.01.

The results presented in Table 5 still show a rather pessimistic view on the predictive power of SII, OS, and PC for aggregate equity returns. While this could be a result of the poor predictive power of the variables, it could also be a mere result of the proxy poorly representing stocks traded by insiders. Nevertheless, it is worth noting that it makes sense to assume that the proportion of stocks that insiders trade is higher in this sub-sample as compared to the full cross-section. To this end, we can still provide intuitive interpretations for the results observed in Table 5.

Let us begin by interpreting the results for SII. This predictor delivers negative R^2_{OS} statistics for all return aggregation levels, with the statistics becoming increasingly negative with the increase of the return aggregation level. Moreover, it is worth noting that the performance of SII for this particular sub-sample is substantially worse than its performance for the full cross-section. This indicates that the predictive power of SII might not be mainly driven by the information on insider trades it incorporates, but rather by some other mechanism. These results are re-confirmed by the results of the VAR decomposition framework, presented in Table 12 of Appendix G. The results presented in this table suggest that the predictive power of SII for future aggregate market returns is not driven by its ability to anticipate future cash flows, or in other words, it is not driven by insider trading.

Now we restrict our attention to OS and PC. PC yields a negative R^2_{OS} statistic for all return aggregation levels, and its performance for this sub-sample is worse than its performance for the full cross-section. Nonetheless, as also argued previously, PC is prone to a lot of noise when used with individual equities. On the other hand, OS is able to outperform the benchmark at certain return aggregation levels. Its overall and bottom decile R^2_{OS} statistic are rather low, however, OS has high predictive power for the stocks in the top decile. These results align with our expectations when the insider signal $\tilde{\epsilon}$ follows a negatively-skewed distribution. In such case, large negative signals occur more often that positive ones of the same magnitude, meaning that a high OS should typically be associated with low future returns. Moreover, as also argued before, OS is more stable than PC for individual equities, and hence its higher predictive power in this case should come as no surprise.

6.3 Factor Investing Results

To get some additional insights on the predictive power of SII, OS, and PC, as well as to get a better understanding on how incorporating information regarding the trades executed by insiders influences their predictive power, this section presents the results of the factor investing framework introduced in Section 4.6.

Recall that, in the end of the previous section, we explored the predictive power of these variables on the returns of individual equities that exhibit extreme values of each predictor variable. This was done based on the argument that stocks with extreme values attained by predictors are likely to be stocks that are traded by insiders. To this end, it is interesting to explore how (an aggregate version of) each predictor can predict the returns of a portfolio that contains all these stocks. We first construct a portfolio for each predictor variable using the decile approach, and then perform in- and out-of-sample tests. In the remainder of this section, we first present the results of the in- and out-of-sample tests, and afterwards we present the results of the asset allocation analysis.

6.3.1 In- and Out-of-sample Results

Table 6 presents the results of the in- and out-of-sample tests for the predictive power of SII, OS, and PC on the respective aggregate portfolio returns.

Table 6: In- and Out-of-sample Results for the Portfolios Constructed Using the Decile Approach

	$h=1$			
Predictor	$\hat{\beta}$	$R^2(\%)$	R^2_{OS}	**Encompassing** λ
SII	0.21	0.53	-3.26	–
OS	0.41^{*}	2.21	3.77^{***}	0.35
PC	0.08	0.25	-2.69	0.51
	$h=3$			
Predictor	$\hat{\beta}$	$R^2(\%)$	R^2_{OS}	**Encompassing** λ
SII	0.42	0.79	-9.25	–
OS	1.04^{*}	4.10	-3.46	0.12
PC	0.60^{**}	3.62	3.83^{***}	0.23

Continued on next page

Table 6 – continued from previous page

	$h = 6$			
Predictor	$\hat{\beta}$	$R^2(\%)$	R^2_{OS}	**Encompassing** λ
SII	0.74	1.68	-3.31^{**}	–
OS	1.22	2.93	-10.57	0.01
PC	0.42	1.23	9.53^{***}	0.26

	$h = 12$			
Predictor	$\hat{\beta}$	$R^2(\%)$	R^2_{OS}	**Encompassing** λ
SII	1.13	3.86	-1.07^{***}	–
OS	2.98	8.05	-18.15	0.46^{**}
PC	2.16^{**}	14.02	28.15^{***}	0.04

Note: The in-sample results presented in this table are for the period Jan 2001 - Jan 2022, while the out-of-sample results are for the period Jan 2006 - Jan 2022. This table presents the results of using the linear specification $r_{t:t+h} = \alpha + \beta x_t + \epsilon_{t:t+h}$, with $t = 1, \ldots, T-h$ and $h \in \{1, 3, 6, 12\}$, for predicting in- and out-of-sample the aggregate returns of the portfolios constructed using the decile approach based on each predictor. In this specification, $r_{t:t+h}$ refers to the aggregate return of the portfolio constructed using the decile approach based on the values of x_t, and x_t corresponds to the value of the aggregated predictor at time t. $\hat{\beta}$ represents the ordinary least squares estimate of β, and $R^2(\%)$ the regression R^2 statistic expressed in percentage. R^2_{OS} presents the values of the out-of-sample R^2 statistic, as defined in Equation 8 in Section 4.4. Finally, Encompassing λ shows the optimal weight that the forecast obtained using SII gets in a convex combination of this forecast and the forecast of another predictor. Next to $\hat{\beta}$, the significance as measured by the wild bootstrapped p-values (details in Appendix A) is presented. Similarly, next to R^2_{OS} and Encompassing λ the significance as measured by the tests of Clark and West (2007) and Harvey et al. (1998) is presented. For the details of these tests, please look at Appendix B and Appendix C, respectively. Finally, note that $*$ indicates a p-value between 0.05 and 0.10, $**$ indicates a p-value between 0.01 and 0.05, and $***$ indicates a p-value lower than 0.01.

Let us begin by interpreting the in-sample test results. The first observation to be drawn is that OS and PC both outperform SII in terms of fitting the aggregate returns of their respective factors in-sample. In particular, none of the fitted $\hat{\beta}$ coefficients corresponding to the SII predictor is statistically significant. On the contrary, $\hat{\beta}$ is significant at the 10% level for OS at two return aggregation levels. Moreover, $\hat{\beta}$ is significant at the 5% level for PC at two return aggregation levels. In addition, OS and PC can fit the aggregate returns of their corresponding portfolios better, attaining superior in-sample R^2-s compared to SII. While OS and PC outperform SII in-sample, determining a winner among them is not so straightforward. Overall, it seems that both these predictors perform similarly in predicting their respective aggregate factor portfolio returns in-sample.

In the remainder of this section, we focus on the results of the out-of-sample tests. The main conclusion that can be drawn from these results is that OS is the best-performing predictor out-of-sample for $h = 1$, while PC is the best-performing predictor for $h \in \{3, 6, 12\}$, achieving a substantially high R^2_{OS} statistic of about 28% for $h = 12$. Put differently, OS can predict its portfolio's short-term returns well relative to the benchmark. On the other hand, PC can predict the future returns of its corresponding portfolio well for the medium- and long-term. SII, on the contrary, yields a negative R^2_{OS} statistic at all return aggregation levels. Note that according to the Clark and West (2007) test SII still statistically outperforms the benchmark for $h = 6$ and $h = 12$. However, PC yields substantially higher and significantly better than the benchmark R^2_{OS} statistics, and therefore it should be preferred over SII at these return aggregation levels.

Finally, we consider combining the forecasts of OS with those of SII for the factor portfolio returns of OS. We also perform the same procedure for PC. As shown in Table 6, we find no significant evidence as shown by the Harvey et al. (1998) test that utilising data on the SII predictor adds additional predictive power on top of that of PC. Do note, nonetheless, that the most sizeable λ coefficient for PC is attained for $h = 1$, that is, for the aggregation level it has the least predictive power on. Similar results hold for OS, however, note that for the return aggregation level $h = 12$, SII obtains a statistically significant weight in the forecast combination. Overall, these results re-confirm the strong out-of-sample predictive power of OS and PC for $h = 1$ and $h \in \{3, 6, 12\}$, respectively.

These results and those in the previous section are in line with the implications of the ITA model (with negatively-skewed $\tilde{\epsilon}$) and the insight that PC is more prone to noise than OS. In the previous section we documented some predictive power for OS on the returns of stocks in the extreme deciles, however, no such predictive power was found for SII. As previously discussed, the lack of predictive power of PC can be attributed to its susceptibility to noise. On the contrary, when the stocks in extreme deciles are aggregated into a portfolio, the noise in PC is (at least partially) removed, and therefore it gains substantial predictive power. In particular, as also implied by the ITA model, PC seems to outperform OS in predicting its portfolio's aggregate returns.

6.3.2 Asset Allocation

Table 7 presents the results of using a linear predictive regression for each predictor variable to determine the proportion of wealth invested by a mean-variance investor in the factor corresponding to the predictor. In particular, we present the gain in Certainty Equivalent Return (CER) from using the forecasts of the linear model with each predictor as explanatory variable, as compared to simply using the mean return benchmark, to determine the weight the mean-variance investor allocates to the risky asset. Furthermore, we also present the Sharpe Ratio of the strategy that utilises the linear model with each predictor as explanatory variable to determine the weight the mean-variance investor allocates to the risky asset.

Table 7: Certainty Equivalent Return Gain and Sharpe Ratio of Using a Predictive Regression to Determine the Factor Portfolio Weights

Predictor	**CER Gain**				**Sharpe Ratio**			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
				SII and BH				
SII	−2.03	−4.35	0.00	0.00	0.24	0.02	0.73	0.50
BH	−11.59	−12.60	−11.30	−7.27	−0.65	−0.80	−0.73	−0.50
				OS and BH				
OS	3.34	−1.05	−0.04	0.31	0.17	−0.16	−0.27	−0.07
BH	−1.80	−4.38	−1.62	−4.23	−0.32	−0.42	−0.29	−0.49
				PC and BH				
PC	0.00	0.00	0.00	−2.08	0.94	0.51	0.65	0.03
BH	−7.91	−4.89	−6.26	−6.42	−0.94	−0.51	−0.65	−0.66

Note: The results presented in this table are for the period Jan 2006 - Jan 2022. SII, OS, and PC refer to using a linear prediction with the appropriate variable. BH refers to a buy-and-hold strategy that simply allocates the entire wealth of the investor to the portfolio under consideration. CER Gain refers to the annualised Certainty Equivalent gain (in percentage) of using either the predictive regression or the BH strategy, as compared the mean return benchmark to determine the weights in the asset allocation framework. For the predictors SII, OS, and PC; Sharpe Ratio refers to the annualised Sharpe Ratio of the strategy that invests in the corresponding portfolio with weights as determined by Equation 12, where $\hat{r}_{t+1}$ is determined using the corresponding predictive regression. For BH, Sharpe Ratio refers to the annualised Sharpe Ratio of the strategy that fully invests in the corresponding portfolio. As a final remark, note that the weight invested in the factor portfolio is restricted to be between -0.5 and 1.5[4]. For further details, please look at Section 4.6.

Recall that Table 6 from Section 6.3.1 showed promising results related to the in-sample fits of OS and PC, and exceptional results for the out-of-sample performance of PC for predicting the aggregate returns of factor portfolios. However, as shown in Table 7, neither of these predictors can consistently yield better weights than the mean return benchmark in the factor investing analysis. This is reflected in the relatively poor performance in terms of CER Gain for both predictors. A similar poor performance is also observed for the SII predictor.

However, it is worth noting that investing using any of the predictors is better than simply buying and holding the associated portfolio. This is reflected in the substantially better CER Gain of each predictor as compared to the buy-and-hold strategy. Moreover, using the buy-and-hold strategy gives a negative Sharpe Ratio for each factor portfolio at each return aggregation level. On the other hand, using SII and PC always yields investment strategies with positive Sharpe Ratios. This is, nonetheless, not the case for OS.

In sum, the main conclusion here is that an investment strategy that utilises a predictive regression based on SII, OS, or PC to trade the corresponding factor portfolio, cannot consistently outperform the benchmark. However, the predictive power these predictors have on the future aggregate returns of their associated factor portfolio can still be exploited by a mean-variance investor. In particular, some of these investment strategies yield positive and sizeable Sharpe Ratios, with the best strategy utilising PC for a return aggregation level of one month.

7 Conclusion

In this section we summarise the research conducted in this paper, outline the main results, and we conclude with some final remarks. This paper investigates the performance of a plethora of variables in predicting aggregate returns. The first variable is the measure based on short interest introduced in Rapach et al. (2016), which is stochastically, rather than linearly, de-trended here (SII). Moreover, we also examine the ratio between the volume of options and stocks traded (OS), as well as the (stochastically de-trended) ratio between the volume of puts and calls traded (PC). Finally, we examine two variables based on the VIX decomposition, and four market sentiment predictor variables. The details on the construction of these variables are presented in Section 4.2.

The paper introduces a novel theoretical framework, named the ITA model, which is used to argue that the predictive power of SII, OS, and PC theoretically comes, at least in part, from their ability to incorporate information regarding the trades executed by insiders. To this end,

[4]Results are robust to various weight lower and upper bounds. In particular, we obtain similar results when using a lower bound of -1 and upper bound of 2, as well as a lower bound of -1.5 and an upper bound of 2.5.

we investigate whether the implications of the ITA model also hold empirically, while at the same time also testing the predictive power of these three predictors.

The empirical investigation of the variable's predictive power is first done for the aggregate market returns, and then for the aggregate returns of individual equities. For aggregate market returns, the predictive power of all predictors is examined. We find that SII performs the best overall, however, we also document satisfying performance for PC. In particular, PC significantly outperforms the benchmark out-of-sample for three return aggregation levels, as well as outperforms SII both in- and out-of-sample in predicting six-month aggregate returns.

Regarding the predictive power for individual equity returns, we only consider SII, OS, and PC, as these are the only predictors for which we possess equity-specific data. We find that when the whole cross-section of stocks in considered, all these predicts perform poorly in- and out-of-sample. However, when we only analyse the stocks in the top and bottom deciles at each time period, we find that OS gains some predictive power. On the other hand, the predictive power of SII becomes substantially worse. This sub-sample is special because the stocks contained in it are more likely to be stock traded by insiders as compared to the full cross-section. Finally, when considering the predictive power of SII, OS, and PC on the returns of factor portfolios constructed using the decile approach, we find that PC gains a substantial predictive power. In particular, we find that OS has high predictive power for the short-term future returns of its portfolio, while PC has high predictive power for the medium- and long-term future returns of its portfolio.

To be able to intuitively interpret all the empirical findings, it must hold that the signal possessed by insider traders on the future asset returns follows a negatively-skewed distribution. Under this assumption, the ITA model makes the following predictions.

First, SII should have predictive power on future returns, since medium-sized negative signals are associated with high values of SII. Second, high values of PC are associated either with large negative or small positive signals; while low values of PC are associated either with large positive or small negative signals. Since being able to capture the patterns in large absolute magnitude signals is of most importance, PC gains some predictive power, where high values of PC indicate low future returns, and low values of PC indicate high future returns. Finally, OS should also contain some predictive power, however, this should be smaller than the one of PC. High OS values are associated with either small- or large-sized future returns. However, the additional assumption allows us to deduce that high OS values will be associated with large negative returns more often than positive ones of the same magnitude; and with small negative returns less often than positive ones of the same magnitude. Since it is the ability to capture patterns in large-sized returns that matters the most, OS gains some predictive power over future returns.

Overall, the ITA model predicts that all three variables should have some predictive power over future returns. Unfortunately, the power of SII cannot be compared to that of PC and OS. However, we can still infer that PC will have more power than OS. As mentioned before, these theoretical predictions also hold empirically. There is, however, one additional insight that needs to be taken into account. In particular, PC is a much more powerful predictor than OS for portfolio returns, nonetheless, OS performs better than PC for individual equities. Since large institutional investors dynamically change their hedging positions within the options market, but rarely change them between the equity and options markets, OS is less susceptible to noise than PC, making it a better predictor for individual equity returns.

We conclude this paper by suggesting directions for future research. First, it could be beneficial to combine SII and OS, as they convey complementary information. In particular, recall that the ITA model suggests that high values of SII should be observed when insiders possess a medium-sized negative signal. However, any other type of signal could be associated with low values of SII. To this end, the information contained in OS can be utilised. To see why this is the case, consider the following observations, which are implied by the ITA model. First, if SII is high and OS is low, then insiders possess a medium-sized negative signal. Second, if SII is low and OS is low, then insiders possess a medium-sized positive signal. To this end, investigating the performance of a predictor that takes high values when SII is high and OS is low, and low values when SII is low and OS is low could be interesting. To provide some additional insights into this, Appendix F explains one way to construct a predictor that reflects this intuition, as well as some results on its performance.

Another direction for future research, which falls beyond the scope of this paper, is to build on the implications of the ITA model (with negatively-skewed $\tilde{\epsilon}$), particularly those corresponding to PC. According to the model, when PC is high, returns are expected to be either small positive or large negative, whereas low PC corresponds to either small negative or large positive expected returns. Based on these insights, two research pathways emerge. First, developing options trading strategies that utilise these 'interval predictions' could be highly beneficial. For instance, in scenarios with a high PC, an effective strategy involves selling slightly out-of-the-money (OOM) calls and buying deep OOM puts. Conversely, in low PC scenarios, selling slightly OOM puts and purchasing deep OOM calls could be effective. Additionally, more complex structures, such as combinations of iron condors, strangles, or spreads might be worth investigating. Second, a more involved approach is to model the returns distribution. Recall that PC indicates where the returns are likely to concentrate, showing two distinct areas of higher probability, which justifies the use of a bimodal distribution. As an example, one can model the distribution of future returns as a mixture of normal distributions where the locations of the peaks (means) are based on the value of PC.[5] This distribution modelling approach can be further used as an extension to the standard Black-Scholes assumption of log-normally-distributed security prices in an options pricing framework. Pursuing these research directions, not only promises to deepen our understanding of return predictability, but also opens up practical applications in options trading.

[5]This model is known as a mixture of regressions model, or a mixture of normals with covariate-dependent means (McLachlan & Peel, 2000).

References

Akbas, F., Armstrong, W. J., Sorescu, S. & Subrahmanyam, A. (2015). Smart money, dumb money, and capital market anomalies. *Journal of Financial Economics*, *118*(2), 355–382.

Bekaert, G. & Hoerova, M. (2014). The vix, the variance premium and stock market volatility. *Journal of econometrics*, *183*(2), 181–192.

Campbell, J. Y. (1991). A variance decomposition for stock returns. *The economic journal*, *101*(405), 157–179.

Campbell, J. Y. & Ammer, J. (1993). What moves the stock and bond markets? a variance decomposition for long-term asset returns. *The journal of finance*, *48*(1), 3–37.

Campbell, J. Y. & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509–1531.

Clark, T. E. & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, *138*(1), 291–311.

Frazzini, A. & Lamont, O. A. (2008). Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of financial economics*, *88*(2), 299–322.

Guo, H. & Savickas, R. (2006). Idiosyncratic volatility, stock market volatility, and expected stock returns. *Journal of Business & Economic Statistics*, *24*(1), 43–56.

Hagerman, R. L. (1978). More evidence on the distribution of security returns. *The journal of finance*, *33*(4), 1213–1221.

Harvey, D. I., Leybourne, S. J. & Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business & Economic Statistics*, *16*(2), 254–259.

Johnson, T. L. & So, E. C. (2012). The option to stock volume ratio and future returns. *Journal of Financial Economics*, *106*(2), 262–286.

Kostakis, A., Magdalinos, T. & Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, *28*(5), 1506–1553.

Li, Q., Yang, J., Hsiao, C. & Chang, Y.-J. (2005). The relationship between stock returns and volatility in international stock markets. *Journal of Empirical Finance*, *12*(5), 650–665.

McLachlan, G. J. & Peel, D. (2000). *Finite mixture models* (Vol. 299). John Wiley & Sons.

Officer, R. R. (1972). The distribution of stock returns. *Journal of the american statistical association*, *67*(340), 807–812.

Pan, J. & Poteshman, A. M. (2006). The information in option volume for future stock prices. *The Review of Financial Studies*, *19*(3), 871–908.

Rapach, D. E., Ringgenberg, M. C. & Zhou, G. (2016). Short interest and aggregate stock returns. *Journal of Financial Economics*, *121*(1), 46–65.

Rapach, D. E., Wohar, M. E. & Rangvid, J. (2005). Macro variables and international stock return predictability. *International journal of forecasting*, *21*(1), 137–166.

Welch, I. & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

A Wild Bootstrap Algorithm

In this appendix we present the pseudo-code for performing wild bootstrapping to derive the distribution of the t-statistics.

Algorithm 1 Wild Bootstrap Procedure

procedure WILD BOOTSTRAP(n)
    Run the regression $r_{t:t+h} = \alpha + \beta x_t + \epsilon_{t:t+h}$ for $t = 1, \ldots, T-h$, and get the residuals
    $\hat{\epsilon}_{t:t+h}$.
    Define t_{distr} as a vector that stores all the wild bootstrapped t-statistics.
    for $i \leftarrow 1$ to n **do**
        Draw k random numbers from a Standard Normal distribution, with k equal to the
        number of aggregate returns in the sample. Call this sequence u^*.
        Define the resampled returns as $r_t^* = \bar{r} + \hat{\epsilon}_{t:t+h} u_t^*$ for all t, with $\bar{r}$ being the mean
        return.
        Aggregate the resampled returns r^* into $r^*_{t:t+h}$.
        Regress $r^*_{t:t+h}$ on the predictor variable. Use the beta coefficient from this regression
        and the Newey-West Standard Errors to compute the t-statistic. Name it t_i.
        Update $t_{distr}[i] = t_i$.
    end for
    return t_{distr}
end procedure

Note that, for each predictor variable, we already know the sign of its β coefficient from theory. Therefore, the alternative hypothesis being tested is one-sided. For instance, if theory suggests for a predictor that $\beta < 0$, then the alternative hypothesis considered in the test is $H_A : \beta > 0$. So, after having computed t_{distr} using the algorithm presented above, one can simply compute the bootstrapped p-value as the fraction of t-statistics in t_{distr} that are smaller than the t-statistic of the original regression computed using Newey-West Standard Errors.

B Clark and West (2007) Testing Procedure

In this section we summarise the testing procedure introduced in Clark and West (2007). This procedure is used to test whether the forecasts of a model are more accurate than those of one of its nested models, where accuracy is measured by the mean squared forecast error (MSFE). Let us refer to the general model by Model 2 and to the nested model used as benchmark by Model 1. The null hypothesis tested here is models have equal MSFE-s, and the on-sided alternative hypothesis is that Model 2 has a smaller MSFE than Model 1.

In addition, denote the aggregate return during periods $t+1$ until $t+h$ by $r_{t:t+h}$, and the one-step ahead forecasts of the aggregate returns of Model 1 and Model 2 by $\hat{r}_{1,t:t+h}$ and $\hat{r}_{2,t:t+h}$, respectively. Finally, let us assume that there are T time periods. Then, the first step of the test requires computing:

$$\hat{f} = \frac{1}{T-h} \sum_{t=1}^{T-h} \left[(r_{t:t+h} - \hat{r}_{1,t:t+h})^2 - \left[(r_{t:t+h} - \hat{r}_{2,t:t+h}) - \hat{r}_{1,t:t+h} - \hat{r}_{2,t:t+h}) \right] \right] \tag{18}$$

Furthermore, the second step requires regressing $\hat{f}$ on a constant. Clark and West (2007) show that the t-statistic of the constant approximately follows a standard normal distribution. However, it is crucial to note that a heteroskedasticity- and autocorrelation-robust standard error estimator needs to be used for the standard deviation of the constant. Therefore, we use the Newey-West standard errors to compute the t-statistic.

C Harvey, Leybourne, and Newbold (1998) Encompassing λ and Test

In this appendix we summarise the procedure of Harvey et al. (1998) to obtain the optimal combination between the forecasts of two competing models, as well as test whether the combination of their forecasts is statistically equivalent to the forecast of only one of the models.

To this end, we begin by defining some notation. The first thing to note is that the test can be generally applied to predicting any variable, however, we are specifically applying it in the context of predicting aggregate returns $r_{t:t+h}$ for $t = 1, \ldots, T-h$, as defined in Section 4.1. In addition, let us denote the first forecasting model by Model 1, and the second model by Model 2. In this paper, Model 1 refers to the linear predictive model using SII as predictor, and Model 2 refers to the linear predictive model using some other predictor. Now we restrict our attention to the following equation:

$$f_{t:t+h} = (1-\lambda) f_{1,t:t+h} + \lambda f_{2,t:t+h}, \tag{19}$$

where $f_{1,t:t+h}$ denotes the forecast of Model 1, f_2 denotes the forecast of Model 2, and f denotes the combined forecast for the aggregate returns during $t+1$ till $t+h$.

Moreover, let us denote the forecast errors of Model 1 and Model 2 by $e_{1,t:t+h}$ and $e_{2,t:t+h}$, respectively. The errors of each model are defined as the difference between the actual realised aggregate returns and the model's forecasts. Furthermore, define the difference between errors as $\Delta e_{t:t+h} = e_{1,t:t+h} - e_{2,t:t+h}$.

To obtain the optimal forecast combination, Harvey et al. (1998) show that one needs to run the following regression:

$$e_{1,t:t+h} = \lambda \Delta e_{t:t+h} + \epsilon_{t:t+h}, \tag{20}$$

and simply set the optimal λ to $\lambda^* = \hat{\lambda}$, where $\hat{\lambda}$ is the ordinary least square estimate of λ in Equation 20.

Now we move forward to explaining the procedure to test the null hypothesis $H_0 : \lambda = 0$ against the one-sided alternative $H_A : \lambda > 0$. To this end, we need to further define some variables before presenting the final test statistic. First, denote the residuals of the regression in Equation 20 by $\hat{\epsilon}_t$. Second, define the Q_1 statistics as follows:

$$Q_1 = \lim_{n \to \infty} V[\frac{1}{T-h} \sum_{t=1}^{T-h} \Delta e_{t:t+h} \epsilon_{t:t+h}], \tag{21}$$

which is in turn estimated as follows for finite samples:

$$\hat{Q}_1 = \begin{cases} \frac{1}{T-h}\sum_{t=2}^{T-h}\Delta e_{t:t+h}^2 \hat{\epsilon}_{t:t+h} & \text{if } h=1 \\ \frac{1}{T-h}\sum_{\tau=-(h-1)}^{\tau=h+1}\sum_{t=|\tau|+1}^{T-h}\Delta e_{t:t+h}\ \hat{\epsilon}_{t:t+h}\ \Delta e_{t-|\tau|:t+h-|\tau|}\ \hat{\epsilon}_{t-|\tau|:t+h-|\tau|} & \text{if } h>1 \end{cases} \tag{22}$$

The final variable that needs to be defined is $\bar{d} = \frac{1}{T-h}\sum_{t=1}^{T-h}\Delta e_{t:t+h} e_{1,t:t+h}$. Then the Harvey et al. (1998) test statistic for testing $H_0 : \lambda = 0$ against $H_A : \lambda > 0$ is given by:

$$R_1 = \frac{1}{T-h}\hat{Q}_1^{-\frac{1}{2}}\bar{d} \tag{23}$$

They show that asymptotically this statistic follows a standard normal distribution, that is $R_1(T) \sim N(0,1)$ as $T \to \infty$.

D Derivation of Weight Equation 12

In this section we provide the derivation of the weights presented in Equation 12. As mentioned in Section 4.6, this weight is derived for a mean-variance investor that chooses to allocate his wealth between a risky asset and the risk-free asset. Denote the returns corresponding to the risky asset by r_{ra}, and those to the risk-free asset by r_{rf}. For the sake of completeness, note that the risky asset considered in Section 4.6 is the portfolio constructed using on the decile approach based on either SII, OS, or PC.

Let us consider a time period t, on which the investor decides on the optimal weights for the next period. Since we are considering a mean-variance investor, the investor's utility function is:

$$U = E_t[R_P] - \frac{1}{2}\gamma\hat{\sigma}_{P,t+1}^2, \tag{24}$$

where $E_t[R_P]$ is the expected value of the portfolio returns at $t+1$ given the information set up until t, $\hat{\sigma}_{P,t+1}^2$ is the variance of the portfolio returns at $t+1$, and γ is the risk-aversion parameter, which is fixed at $\gamma = 3$ in this paper.

Let us denote the weight allocated to the risky asset at time t by w_t. Then, the expected portfolio returns at $t+1$ will be $E_t[R_P] = w_t E_t[r_{ra}] + (1-w_t)r_{rf}$. Moreover, the portfolio variance at time $t+1$ is given by $\hat{\sigma}_{P,t+1}^2 = w_t^2\hat{\sigma}_{ra,t+1}^2$, with $\hat{\sigma}_{ra,t+1}^2$ being the variance at time $t+1$ of the risky asset.

Therefore, the problem boils down to maximizing the investor's utility function as given below, with respect to the weight w_t:

$$U = w_t E_t[r_{ra}] + (1-w_t)r_{rf} - \frac{\gamma}{2}w_t^2\hat{\sigma}_{ra,t+1}^2 \tag{25}$$

The first-order condition of this optimisation reads:

$$0 = \frac{\partial U}{\partial w_t} = E_t[r_{ra}] - r_{rf} - \gamma w_t\hat{\sigma}_{ra,t+1}^2 \tag{26}$$

$$\Rightarrow w_t = \frac{1}{\gamma}\frac{E_t[r_{ra}] - r_{rf}}{\hat{\sigma}_{ra,t+1}^2} \tag{27}$$

Note that this expression is the same as the one presented in Equation 12, with the only difference being that the variance estimate here is for the returns of the risky asset, rather than the excess returns of the risky asset. Note, however, that these two are equal to each other, as shown below:

$$V_{era,t+1} = V[r_{ra,t+1} - r_{rf,t+1}] = V[r_{ra,t+1}] + V[r_{rf,t+1}] - 2Cov[r_{ra,t+1}, r_{rf,t+1}] = \tag{28}$$

$$\Rightarrow V_{era,t+1} = V[r_{ra,t+1}] = \hat{\sigma}^2_{ra,t+1}, \tag{29}$$

where V denotes the variance operator, Cov denotes the covariance operator, and $V_{era,t+1}$ is the variance of the excess returns of the risky asset at $t+1$. That is because the variance of the risk-free rate is a scalar, and hence its variance and covariance with other random variables is zero.

E Derivation of Intervals for Theoretical Model

In this appendix we derive the theoretical intervals of the signal ϵ possessed by the informed trader for which such a trader would execute each trade type. Put differently, in this section we derive the theoretical values of $k_1, \ldots, k_6$ in Equation 17.

Before starting with the actual derivations, it is necessary to define some notation that will be used throughout the derivations. First, we denote the profit function of the investor by π. Second, given a signal $\tilde{\epsilon} = \epsilon$, the expected price of the stock at time $t = 2$ is $E[\tilde{V}] = \mu + \epsilon$. Moreover, recall that we assume that the stock price at time $t = 1$ (μ) is the same as the strike price of the call and put options (K), such that it holds that $\mu = K$. Recall that stock trades occur in lots of γ, while options trades in lots of θ, with $\theta > \gamma > 0$. Also, recall that when shorting the stock, a fraction ρ of the total transaction is paid as an interest fee. Finally, denote the premium of the call option by fee, and the premium of the put option by fee^*.

Moreover, it must hold that $fee \geq fee^*$ for the put-call parity $fee - fee^* = \mu - Ke^{-rT}$ to hold under the theoretical model assumptions, where r is the risk-free rate, and T stands for the time to maturity of both options. The proof is given below:

$$\begin{aligned}
&-rT \leq 0 \\
&e^{-rT} \leq 1 \\
&Ke^{-rT} \leq K \\
&Ke^{-rT} - \mu \leq K - \mu \\
&Ke^{-rT} - \mu \leq 0 \\
&fee - fee^* = \mu - Ke^{-rT} \geq 0 \\
&fee \geq fee^*
\end{aligned} \tag{30}$$

Now we fully restrict the attention to deriving $k_1, \ldots, k_6$. First, note that when the signal is positive, that is $\epsilon > 0$, then the insider will either sell the put, buy the stock, or buy the call.

Moreover, if the signal is negative, that is $\epsilon < 0$, then the insider will either buy the put, sell the stock, or sell the call. To this end, the expected profit of the informed trader as a function of the strategy he executes is given as follows:

$$E[\pi(S)] = \begin{cases} -\theta\epsilon - fee^* & \text{if } S = BP \ (\epsilon < 0) \\ -(1-\rho)\gamma\epsilon & \text{if } S = SS \ (\epsilon < 0) \\ fee & \text{if } S = SC \ (\epsilon < 0) \\ 0 & \text{if } S = NT \\ fee^* & \text{if } S = SP \ (\epsilon > 0) \\ \gamma\epsilon & \text{if } S = BS \ (\epsilon > 0) \\ \theta\epsilon - fee & \text{if } S = BC \ (\epsilon > 0), \end{cases} \tag{31}$$

where S stands for strategy, BP for buy put, SS for sell stock, SC for sell call, NT for make no trade, SP for sell put, BS for buy stock, and BC for buy call. Also, brackets indicate the sign of the signal ϵ for which it makes sense for the insider to execute the trade.

It is easy to derive from Equation 31 the values of the signal ϵ that make the insider indifferent between choosing various investment strategies. For instance, the insider is indifferent between buying the put or selling the stock if $-\theta\epsilon - fee^* = -(1-\rho)\gamma\epsilon$, which holds for $\epsilon = -\frac{fee^*}{\theta-(1-\rho)\gamma}$. One can solve the other indifference equations in a similar fashion, and arrive at the following optimal trading strategy function for an insider as a function of the signal ϵ:

$$S(\epsilon) = \begin{cases} BP & \text{if } \epsilon < -\frac{fee^*}{\theta-(1-\rho)\gamma} \\ SS & \text{if } -\frac{fee^*}{\theta-(1-\rho)\gamma} \leq \epsilon < -\frac{fee}{(1-\rho)\gamma} \\ SC & \text{if } -\frac{fee}{(1-\rho)\gamma} \leq \epsilon < 0 \\ NT & \text{if } \epsilon = 0 \\ SP & \text{if } 0 < \epsilon \leq \frac{fee^*}{\gamma} \\ BS & \text{if } \frac{fee^*}{\gamma} < \epsilon \leq \frac{fee}{\theta-\gamma} \\ BC & \text{if } \epsilon > \frac{fee}{\theta-\gamma}, \end{cases} \tag{32}$$

Equation 32 shows the optimal trading actions of an insider depending on the signal that he possesses. This equation is analogous to Equation 17, but here the values of $k_1, \ldots, k_6$ are explicitly stated.

F Predictor That Combines SII and OS

In this appendix we present one method to combine SII and OS into a single predictor, named COMB onwards. The first step involves dynamically standardizing SII and OS. Since these variables possibly have different scales, combining the un-standardised versions of these variables would lead to a predictor that gives a higher importance to the variable that has a larger scale. To overcome this, we dynamically standardise SII and OS to have zero mean and standard deviation of one at each point in time.

Using the standardised versions of SII and OS, we can construct the predictor. Formally, at each time period t, COMB is constructed as:

$$COMB_t = \begin{cases} 1 & \text{if } SII_t > \frac{1}{2}SD_{SII,1:t},\ OS_t < -\frac{1}{2}SD_{OS,1:t} \\ -1 & \text{if } SII_t < -\frac{1}{2}SD_{SII,1:t},\ OS_t < -\frac{1}{2}SD_{OS,1:t} \\ 0 & \text{otherwise,} \end{cases} \tag{33}$$

where $SD_{SII,1:t}$ is the standard deviation of the standardised SII variable, computed using data from the start of the sample until time t. Similarly, $SD_{OS,1:t}$ is the standard deviation of the standardised OS variable, computed using data from the start of the sample until time t.

To provide the reader with an idea regarding the performance of this predictor, we perform the in- and out-of-sample tests for the predictive power of this predictor on the aggregate market returns. The results of these tests are presented in Table 8, where we also include the results for SII in the table for comparison purposes.

Table 8: In- and Out-of-sample Test Results for the Predictive power of COMB on Aggregate Market Returns

Predictor	$\hat{\beta}$	R^2 (%)	$\hat{\beta}$	R^2 (%)	$\hat{\beta}$	R^2 (%)	$\hat{\beta}$	R^2 (%)
	$h=1$		$h=3$		$h=6$		$h=12$	
SII	-0.68	2.26	-0.72	7.70	-0.69	12.43	-0.57	16.05
	$[-1.92]^{**}$		$[-1.90]^{*}$		$[-1.83]^{*}$		$[-2.13]^{*}$	
COMB	-0.65	2.07	-0.90	12.11	-0.99	25.96	-0.59	17.20
	$[-1.75]^{*}$		$[-3.29]^{***}$		$[-3.57]^{***}$		$[-4.17]^{***}$	

Predictor	**Out-of-sample R^2 statistic** (%)				**Encompassing λ statistic**			
	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
SII	0.43^{**}	5.56^{***}	-0.99^{***}	16.04^{***}	–	–	–	–
COMB	1.10	8.53^{**}	24.31^{**}	20.20^{***}	0.44^{**}	0.44^{***}	0.22	0.45^{**}

Note: The in-sample results presented in this table are for the period Jan 2001 - Jan 2022, while the out-of-sample results are for the period Jan 2006 - Jan 2022. This table presents the results of the same tests as the ones performed in Tables 1 and 2, for the COMB variable. Hence, for the details of these tests, we refer you to the notes of Tables 1 and 2.

The results presented in Table 8 indicate strong in- and out-of-sample performance for the COMB predictor. These results are in line with the implications of the theoretical model, that predicts that COMB should have a higher predictive power than SII. In particular, COMB outperforms SII in predicting aggregate market returns over 3, 6 and twelve months, with the outperformance being the most noticeable for $h = 6$. On the other hand, SII seems to perform slightly better than COMB for $h = 1$. In the remainder of this appendix we elaborate on the results of the in- and out-of-sample tests.

The in-sample tests indicate that the combined variable COMB substantially outperforms SII for $h = 3$ and $h = 6$. This is reflected in the sizeable increases in the magnitude of the beta estimates, their significance, as well as the regression R^2 statistics. An improvement is also observed for $h = 12$, however, this improvement is substantial only for the significance of the significance of $\hat{\beta}$. While COMB outperforms SII for $h \in \{3, 6, 12\}$, it fails to do so for $h = 1$. It is worth noting, however, that COMB performs only marginally worse than SII in-sample.

The results of the out-of-sample tests are mostly in line with those of the in-sample tests. Again, COMB substantially outperforms SII for $h = 3$ and $h = 6$, and it slightly outperforms SII for $h = 12$. The main difference as compared to the in-sample results is for $h = 1$. In particular, COMB attains a higher R^2_{OS} statistic than SII, however, this statistic is not significant according to the Clark and West (2007) test. Moreover, when combining the forecasts of SII and COMB, the optimal weight that should be allocated to COMB, as measured by the Harvey et al. (1998) λ statistic is higher than that of SII. Overall, the out-of-sample test results re-confirm that COMB outperforms SII for $h \in \{3, 6, 12\}$, but they do not determine a clear-cut winner among COMB and SII for $h = 1$.

G Replication

In this appendix we present the results of replicating the Rapach et al. (2016) paper. It is crucial to note that, for the sake of consistency, the sample period used when presenting the replication results will be the same as the one used throughout this paper, that is, January 2001 until January 2022. The replications analysis is performed on SII as well as the 14 Welch and Goyal (2008) predictor variables. SII in the replication also refer to the stochastically de-trended SII variable. The other 14 predictors are constructed using data from the website of Amit Goyal, and for their definitions we refer you to the Welch and Goyal (2008) paper.

This paper already re-utilises the in-sample and out-of-sample analysis of Rapach et al. (2016). However, we use these tests for SII and some additional predictor variables. Therefore, for the sake of completeness, we also present the in- and out-of-sample test results for the 14 Welch and Goyal (2008) variables. For the interested readers, the details of these tests are presented in Sections 4.3 and 4.4.

In addition, in this paper we also re-utilise the asset allocation exercise presented in Rapach et al. (2016). Nonetheless, this is done in the context of factor portfolios in our paper, rather than the market portfolio. Therefore, we also present the results of the asset allocation exercise on the market portfolio for SII and the 14 Welch and Goyal (2008) variables. The details of the asset allocation exercise are explained in Section 4.6.

Finally, we also present the results of performing a VAR decomposition of returns, and investigating the source of the predictive power of SII, as suggested in Rapach et al. (2016). An overview of this methodology will be presented in Appendix G.4.

G.1 In-sample Results

Table 9: In-sample Results for Aggregate Market Returns

Predictor	$\hat{\beta}$ $h=1$	R^2 (%)	$\hat{\beta}$ $h=3$	R^2 (%)	$\hat{\beta}$ $h=6$	R^2 (%)	$\hat{\beta}$ $h=12$	R^2 (%)
SII	−0.68	2.26	−0.72	7.70	−0.69	12.43	−0.57	16.05
	[−1.92]**		[−1.90]*		[−1.83]*		[−2.13]*	
SII|PC3	−0.98	2.99	0.68	11.89	1.96	26.04	0.89	39.13
	[−0.39]		[0.38]		[1.08]		[0.71]	
DP	0.55	1.58	0.49	3.52	0.56	8.23	0.59	17.32
	[1.09]		[1.18]		[1.88]		[3.32]*	
DY	0.65	2.21	0.52	4.02	0.58	9.02	0.61	18.51
	[1.52]*		[1.47]		[2.30]*		[4.06]**	
EP	0.12	0.07	−0.02	0.01	−0.07	0.15	−0.02	0.02
	[0.26]		[−0.06]		[−0.22]		[−0.08]	
DE	0.11	0.07	0.20	0.62	0.27	1.94	0.23	2.53
	[0.28]		[0.62]		[1.21]		[1.74]	
RVOL	0.38	0.77	0.41	2.50	0.37	3.69	0.33	5.50
	[1.52]*		[1.88]*		[1.76]*		[1.79]*	
BM	0.71	2.63	0.68	6.84	0.76	15.40	0.72	25.61
	[2.65]***		[3.18]***		[3.87]***		[4.27]***	
NTIS	0.30	0.46	0.38	2.11	0.36	3.51	0.26	3.38
	[0.66]		[0.84]		[0.88]		[0.81]	
TBL	−0.67	2.33	−0.57	4.88	−0.62	10.15	−0.66	21.86
	[−2.55]**		[−2.61]***		[−2.61]**		[−3.00]**	
LTY	−0.89	4.11	−0.83	10.28	−0.81	17.45	−0.71	25.20
	[−3.70]***		[−3.71]***		[−3.65]***		[−3.85]***	
LTR	0.20	0.20	−0.01	0.00	0.10	0.26	0.07	0.26
	[1.00]		[−0.05]		[0.84]		[0.95]	
TMS	−0.16	0.13	−0.20	0.59	−0.12	0.37	0.04	0.10
	[−0.53]		[−0.77]		[−0.51]		[0.23]	
DFY	−0.18	0.17	−0.08	0.09	0.12	0.40	0.25	3.17
	[−0.37]		[−0.19]		[0.45]		[2.02]	
DFR	0.29	0.43	0.06	0.06	0.13	0.44	0.07	0.23
	[0.89]		[0.21]		[0.59]		[0.55]	
INFL	0.39	0.80	−0.09	0.13	−0.27	2.01	−0.24	2.89
	[0.96]		[−0.35]		[−1.46]		[−1.89]**	

Note: The results presented in this table are for the period Jan 2001 - Jan 2022. This table reports the result of estimating the linear specification $r_{t:t+h} = \alpha + \beta x_t + \epsilon_{t:t+h}$ using ordinary least squares, where $r_{t:t+h}$ represent the market returns aggregated from $t+1$ until $t+h$, with $h \in \{1, 3, 6, 12\}$, and x_t denoting the value of the predictor variable at time t. Note that all the predictors are standardized to have a standard deviation of one for comparison purposes. The table reports the estimated slope coefficient $\hat{\beta}$, the regression R^2, the Newey-West Standard Error (given inside square brackets below the $\hat{\beta}$). The significance of $\hat{\beta}$ as measured by the bootstrapped p-value is presented next to the Newey-West Standard Error. The null hypothesis being tested is $H_0 : \beta = 0$, and the alternative is one-sided, meaning that $H_A : \beta < 0$ if $\hat{\beta} < 0$ or $H_A : \beta > 0$ if $\hat{\beta} > 0$. Moreover, * represents a p-value between 0.05 and 0.10, ** represents a p-value between 0.01 and 0.05, and *** represents a p-value smaller than 0.01. Finally, note that SII|PC3 refers to using the aforementioned linear specification, but also adding the first three principal components of the other variables as regressors. R^2 for SII|PC3 refers to the partial R^2 corresponding to SII.

The results presented in Table 9 show that SII delivers sizeable and statistically significant β estimates. Moreover, these estimates are in line with theory: the higher the level of shorting in the market, the lower its future aggregate returns. However, it is worth noting that while $\hat{\beta}$ coefficients here are more sizable than those in Rapach et al. (2016), they are less statistically significant. Moreover, the R^2 statistics associated with SII are higher here.

Table 9 also shows that the 14 Welch and Goyal (2008) predictors seem to perform better in-sample, as compared to their performance in Rapach et al. (2016). The predictor that exhibits the most notable difference is BM. Rapach et al. (2016) report insignificant $\hat{\beta}$-s at all return aggregation levels, however, for our sample period we find highly significant $\hat{\beta}$-s, as well as large R^2 statistics associated with this predictor. Similar results also hold for the TBL and LTY predictors.

The fact that some of the Welch and Goyal (2008) significantly predict aggregate market returns in-sample, also changes the results for SII|PC3. In particular, when controlling for the effect of the first three principal components from the set of the 14 Welch and Goyal (2008) predictors, SII becomes statistically insignificant in explain market aggregate returns. This shows that SII does not have any significant predictive power above that of the 14 Welch and Goyal (2008) predictors.

In general, Table 9 shows that the in-sample predictive power of SII for aggregate market returns, even though still statistically significant, is weaker for the sample considered in this study as compared to the original Rapach et al. (2016) paper.

G.2 Out-of-sample Results

Table 10: Out-of-sample Results for Aggregate Market Returns

Predictor	**Out-of-sample R^2 statistic (%)**				**Encompassing λ statistic**			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
SII	0.43^{**}	5.56^{***}	-0.99^{***}	16.04^{***}	-	-	-	-
DP	-6.30	-23.51	-37.60	-35.95	0.88^{***}	0.90^{***}	0.78^{***}	0.87^{***}
DY	-4.82	-18.26	-29.46	-20.74	0.82^{***}	0.86^{***}	0.74^{***}	0.80^{***}
EP	-4.86	-21.02	-39.90	-35.78	0.75^{***}	0.78^{***}	0.71^{***}	0.90^{***}
DE	-5.10	37.15	-77.11	-36.40	0.71^{**}	0.82^{***}	0.82^{***}	0.90^{***}
RVOL	-0.17	-0.92	-1.97	0.87	0.55^{**}	0.61^{***}	0.51^{***}	0.69^{***}
BM	1.18^{*}	3.18^{*}	9.97^{**}	15.43^{**}	0.44^{**}	0.55^{***}	0.36^{*}	0.51^{**}
NTIS	-1.31	-4.49	-10.65	-19.43	0.62^{***}	0.65^{***}	0.58^{***}	0.79^{***}
TBL	-0.04	0.79	1.14^{*}	11.18^{***}	0.58^{*}	0.61^{***}	0.47^{**}	0.61^{**}
LTY	-1.50^{**}	-3.01^{**}	-2.26^{***}	3.76^{**}	0.69^{**}	0.76^{***}	0.52^{**}	0.73^{**}
LTR	-1.28	-1.68	-0.40	2.00	0.64^{***}	0.62^{***}	0.49^{***}	0.68^{***}
TMS	-0.96	-0.90	-2.70	-9.09	0.63^{**}	0.61^{***}	0.52^{***}	0.81^{***}
DFY	-3.81	-28.36	-86.04	-16.93	0.66^{***}	0.74^{***}	0.84^{***}	0.82^{***}
DFR	-6.78	-7.23	-4.35	-4.73	1.00^{***}	0.71^{***}	0.53^{***}	0.76^{***}
INFL	-0.89	-2.32	0.87	2.14	0.59^{**}	0.64^{***}	0.48^{***}	0.68^{***}

Note: The results presented in this table are for the period Jan 2006 - Jan 2022. The table presents the out-of-sample R^2 statistics (R^2_{OS}), as defined in Section 4.4, for each predictor, which compares the one-step ahead aggregate market return forecasts using the linear specification $\hat{r}_{t:t+h} = \hat{\alpha}_t + \hat{\beta}_t x_t$ to the forecasts of the mean return benchmark. In the aforementioned specification, $\hat{r}_{t:t+h}$ represents the one-step ahead forecasted return, and $\hat{\alpha}$ and $\hat{\beta}$ represent the coefficients of the lienar model fitted using data up until time t. Next to the R^2_{OS} statistic, the significance of the Clark and West (2007) test statistic that tests the null hypothesis $H_0 : R^2_{OS} \leq 0$ against $H_A : R^2_{OS} > 0$ is reported; where $*$ indicates a p-value between 0.05 and 0.10, $**$ indicates a p-value between 0.01 and 0.05, and $***$ indicates a p-value lower than 0.01. Moreover, the table also presents the optimal coefficient (referred to as λ) allocated to the SII predictor when combining the forecasts of SII and another predictor using a convex combination of forecasts, as developed in Harvey et al. (1998). Next to each λ coefficient, we also report the significance of the test developed in Harvey et al. (1998) that tests $H_0 : \lambda = 0$ against $H_A : \lambda > 0$. The meaning of the symbols $*$, $**$, and $***$ remains the same as for the Clark and West (2007) test.

Let us start by interpreting the out-of-sample results for SII. Table 10 shows sizeable and statistically significant R^2_{OS} statistics for the SII predictor at all return aggregation levels. However, as also noted before, for $h = 6$, SII delivers a negative R^2_{OS} statistic. On the contrary, Rapach et al. (2016) document positive and significant R^2_{OS} statistics for SII at all return aggregation levels.

Moreover, the results summarised in Table 10 are mostly in line with those of the in-sample tests of Table 9. In particular, the predictors that perform well in-sample tend to continue doing so out-of-sample. The best performing predictor out-of-sample from the set of 14 Welch and Goyal (2008) predictors is BM, with its performance closely matching that of SII. In particular, BM achieves positive (and sizable) and statistically significant β coefficient estimates at all return aggregation levels. Furthermore, when its forecasts are combined with those of SII, the Harvey et al. (1998) shows that the forecast of BM should be given a larger weight than that of SII for two out of four return aggregation levels.

Another important difference from the results of Rapach et al. (2016) lies in the encompassing λ statistics and their statistical significance. In particular, in Rapach et al. (2016) most estimated λ-s equal 1, indicating that the forecasts of the 14 Welch and Goyal (2008) predictors are typically encompassed by the forecasts of SII. However, this is not the case for the sample considered in our paper. We find $\hat{\lambda}$-s that range from 0.36 to 1, with most $\hat{\lambda}$-s being around $0.60 - 0.80$. This once again indicates that the performance of SII in our sample is worse than the one reported in Rapach et al. (2016), and the performance of the 14 Welch and Goyal (2008) predictors is better than the one reported in Rapach et al. (2016).

G.3 Asset Allocation

Table 11: CER Gain and Sharpe Ratio of Using a Predictive Regression to Determine the Factor Portfolio Weights

Predictor	**CER Gain %**				**Sharpe Ratio**			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
SII	6.46	6.01	6.13	7.59	1.04	1.04	1.05	1.18
DP	0.72	2.24	2.22	2.69	0.81	0.88	0.87	0.94
DY	2.83	3.56	5.64	7.59	0.98	0.94	1.05	1.18
EP	−0.34	3.28	4.26	4.67	0.77	0.94	0.99	1.04
DE	−0.65	3.04	3.68	4.67	0.74	0.93	0.96	1.04
RVOL	−1.05	−1.40	−1.10	3.83	0.70	0.69	0.69	1.00
BM	−1.84	−4.27	−2.97	−2.01	0.68	0.54	0.63	0.79
NTIS	−2.91	−3.80	−4.76	−3.12	0.61	0.56	0.50	0.64
TBL	4.78	6.01	6.13	7.59	1.04	1.04	1.05	1.18
LTY	6.50	6.01	6.13	7.59	1.03	1.04	1.05	1.18
LTR	−0.31	3.82	4.89	7.59	0.75	0.97	0.99	1.18
TMS	-0.73	4.08	5.19	−1.37	0.73	0.98	1.01	0.86
DFY	-0.51	4.33	4.81	4.95	0.74	1.00	1.02	1.05
DFR	1.26	2.37	3.20	4.67	0.97	0.95	0.95	1.04
INFL	−4.02	4.28	6.05	7.59	0.51	0.99	1.08	1.18
BH	−0.79	−1.51	−1.52	−2.43	0.78	0.82	0.81	0.84

Note: The results presented in this table are for the period Jan 2006 - Jan 2022. This table presents the results of using the forecasts of a predictive regression when constructing the weights a mean-variance investor invests in the market portfolio at each time period. CER Gain denotes the percentage gain in the annualized certainty equivalent measure of using a predictive regression with the appropriate predictor over using the simple mean return benchmark. Sharpe Ratio refers to the annualized Sharpe Ratio measure of the investment strategy that invests a portion of wealth given by Equation 12 in the market portfolio. BH refers to the buy-and-hold strategy that simply invests the full wealth in the market portfolio, that is, always uses a weight of 1 in Equation 12. As a final remark, note that the weight invested in the market portfolio is restricted to be between -0.5 and 1.5. For further details, we refer you to Section 4.6.

Table 11 summarises the results of the asset allocation framework, where an investor invests a portion of his wealth given by Equation 12 to the market portfolio, and invests the remainder of his wealth in the risk-free asset. These results, re-confirm the benefit of using a predictive regression with SII as predictor over the benchmark to determine the weights. The gain in the Certainty Equivalent measure is positive and sizeable, with its magnitude being higher than the one reported in Rapach et al. (2016). In addition, using the SII predictive regression to determine the weights of the investment strategy also generates large annualised Sharpe Ratios, exceeding 1 at all return aggregation levels. Again, note that we find Sharpe Ratio statistics that are substantially higher than those reported in Rapach et al. (2016).

Regarding the 14 Welch and Goyal (2008) predictors, we observe that some of these predictors deliver positive and sizeable gains in the Certainty Equivalent measure, as well as positive and sizeable Sharpe Ratio statistics. TBL and LTY perform well in particular, closely matching the performance of SII. On the contrary, it comes as a surprise that BM performs poorly in this asset allocation exercise, especially since its in- and out-of-sample tests indicate strong predictive power.

Overall, the results of this asset allocation exercise are similar to the ones reported in Rapach et al. (2016), with SII being the predictor that gives the highest gains in Certainty Equivalent, as well as the highest Sharpe Ratios. However, contrary to Rapach et al. (2016), we find that some of the Welch and Goyal (2008) predictors closely match SII's performance. Moreover, the statistics reported here tend to be more sizeable than those presented in Rapach et al. (2016), suggesting that investing in the market portfolio according to a linear predictive model has become more beneficial recently.

G.4 VAR Return Decomposition

We begin this appendix by describing the VAR return decomposition framework used in Rapach et al. (2016), inherited from Campbell (1991) and Campbell and Ammer (1993). The purpose of this decomposition is to determine where the predictive power of SII arises from. To this end, returns are decomposed into three components, and the predictive power of SII on each of these components is investigated. These three components are expected returns, cash flow news, and discount rate news. The decomposition is done using a Vector Autoregression (VAR) framework, and it leads to three variables, with each variable representing one of the return components. As a final step, a regression of each of these variables on SII is performed, and the significance of the coefficient estimate is assessed. For the details of this procedure, please look at Section 5 of Rapach et al. (2016).

Table 12: Predictive Power of SII on Each Component of Returns Obtained From the VAR Decomposition

VAR variables	$\hat{\beta}_E$	$\hat{\beta}_{CF}$	$\hat{\beta}_{DR}$	VAR variables	$\hat{\beta}_E$	$\hat{\beta}_{CF}$	$\hat{\beta}_{DR}$
r, DP	−0.17	−0.27	0.20	r, DP, LTY	−0.66	−0.11	−0.05
	[−3.55]***	[−1.70]*	[1.30]		[−12.38]***	[−0.57]	[−0.24]
r, DP, DY	−0.20	−0.27	0.19	r, DP, LTR	−0.17	−0.27	0.20
	[−3.84]***	[−1.72]*	[1.25]		[−3.43]***	[−1.70]*	[1.32]
r, DP, EP	−0.22	−0.51	−0.09	r, DP, TMS	−0.22	−0.27	0.15
	[−3.84]***	[−1.94]*	[−0.81]		[−3.95]***	[−1.56]	[0.85]
r, DP, DE	−0.22	−0.51	−0.09	r, DP, DFY	−0.42	−0.22	0.02
	[−3.84]***	[−1.94]*	[−0.81]		[−5.15]***	[−0.92]	[0.10]
r, DP, RVOL	−0.19	−0.33	0.12	r, DP, DFR	−0.17	−0.27	0.20
	[−4.13]***	[−1.65]*	[1.04]		[−3.64]***	[−1.70]*	[1.33]
r, DP, BM	−0.23	−0.20	0.21	r, DP, INFL	−0.15	−0.29	0.17
	[−4.15]***	[−1.23]	[1.14]		[−2.22]*	[−1.76]*	[1.10]
r, DP, NTIS	−0.26	−0.24	0.15	r, DP, PC3	−0.46	−0.24	−0.04
	[−3.66]***	[−1.60]	[0.78]		[−9.83]***	[−1.29]	[−0.25]
r, DP, TBL	−0.34	−0.20	0.12				
	[−6.89]***	[−1.24]	[0.74]				

Note: The results presented in this table are for the period Jan 2001 - Jan 2022. This table presents the results of the VAR decomposition framework performed in Rapach et al. (2016), inherited from Campbell (1991) and Campbell and Ammer (1993). For the details of this procedure, please look at ... This table particularly presents the ordinary least square estimates of β_y in the following specification: $y_{t+1} = \alpha_y + \beta_y SII_t + \epsilon_{t+1}$, for $t = 1, \ldots, T-1$, where y_t denotes the value attained at time t by one of the components of returns derived from the VAR decomposition, and SII_t denotes the value of the SII index at time t. The three return components derived from the decomposition are the expected return component, the cash flow news component, and the discount rate news component. These are denoted by the subscripts E, CF, and DR, respectively. VAR Variables denotes the variables that are included in the VAR model specification when decomposing the returns. Finally, note that PC3 refers to the first three principal components of all the variables considered here (that is, SII and the 14 Welch and Goyal (2008) variables).

The results presented in Table 12 differ substantially from the ones documented in Rapach et al. (2016). In particular, recall that in Rapach et al. (2016), $\hat{\beta}_E$ is typically highly significant, but its magnitude was relatively small; while $\hat{\beta}_{CF}$ is usually less significant, but it is substantially more sizeable. To this end, Rapach et al. (2016) associate the predictive power of SII with its ability to anticipate future cash flow news, and therefore they argue that in turn SII can anticipate the trades executed by insiders.

However, the results in Table 12 do not suggest this. In particular, $\hat{\beta}_E$ coefficients are more sizeable and highly significant. In many cases, $\hat{\beta}_E$ coefficients are even more sizeable than the respective $\hat{\beta}_{CF}$ coefficients. Furthermore, we find that $\hat{\beta}_{CF}$ coefficients are rarely statistically significant, and even when they are, that is only at the 10% level. Overall, the findings in our paper do not seem to support the idea that the predictive power of SII is driven by its ability to anticipate future cash flow news. On the contrary, we find that the predictive power of SII is primarily induced by its ability to anticipate the future expected returns.

H Programming Code

The code used to conduct the analysis in this paper is written in R. In particular, we use three R scripts. The first one is named Replication, and it is used to replicate all the results of the Rapach et al. (2016) paper, which are presented in Appendix G.

The other two R scripts, named Extensions and Extensions2, are used to conduct the complete analysis presented in the main part of this paper. Extensions contains the code that performs the complete market-level analysis, presented in Section 6.1, as well as the code that performs the factor analysis, presented in Section 6.3. Extensions2 contains the code that performs the entire equity-by-equity analysis, presented in Section 6.2.