Predictability of Stock Returns based on the Partial Least Squares Methodology

Qinye Lu (De Montfort University)

Abstract

Empirical evidence on the predictability of aggregate stock returns has shown that many commonly used predictor variables have little power to predict the market out-of-sample. However, a recent paper by Kelly and Pruitt (2013) find that predictors with strong out-of-sample performance can be constructed, using a partial least squares methodology, from the valuation ratios of portfolios. This paper shows that the statistical significance of this out-of-sample predictability is overstated for two reasons. Firstly, the analysis is conducted on gross returns rather than excess returns, and this raises the apparent predictability of the equity premium due to inclusion predictable movements of interest rates. Secondly, the bootstrap statistics used to assess out-of-sample significance do not account for small-sample bias in the estimated coefficients. This bias is well known to affect tests of in-sample significance (Stambaugh (1986)) and I show it is also important for out-of-sample tests of significance. Accounting for both these effects can radically change the conclusions; for example, the recursive out-of-sample $R^2$ values for the sample period 1965-2010 are insignificant for the prediction of one-year excess returns, and one-month returns, except in the case of the book-to-market ratios of six size- and value-sorted portfolios, which is significant at the 10% level.

Key words: Return predictability; Partial least squares; Bootstrap

Correspondence Details

Qinye Lu
Hugh Aston Building Room HU 3.71
Department of Accounting and Finance
Leicester Business School
Faculty of Business and Law
De Montfort University
Leicester, LE1 9BH
United Kingdom

Email: qinye.lu@dmu.ac.uk
Introduction

There is a voluminous literature providing evidence that aggregate stock market returns can be predicted using a wide variety of financial and macroeconomic variables. In particular, the dividend yield, earnings yield, book-to-market ratio, and various combinations of interest rates have been identified as key predictors. In addition, while early investigations are in favour of using in-sample (IS) tests to examine the ability of these variables to forecast returns, recent researches employ out-of-sample (OOS) tests, which simulate the ability of practitioners to predict returns using only the information set available to them at the time. The study by Goyal and Welch (2008) was particularly influential in making the case that almost all the key predictors identified in the literature performed poorly in out-of-sample tests. Subsequently various papers, including Rapach et al. (2009), Ferreira and Santa-Clara (2011) and Kelly and Pruitt (2013), deliver predictors with strong OOS performance.

My research is going to critically look at Kelly and Pruitt’s paper (hereafter referred to as KP), in which they use a partial least squares (PLS) methodology inspired by Wold (1966) and find a comparatively high OOS forecasting $R^2$ for U.S. market returns. Rather than the standard approach of using aggregate variables as predictors, their method extracts information about future market returns from the cross-section, using book-to-market ratios of portfolios. They obtain point estimates of 13% for OOS $R^2$ values at a yearly frequency (0.9% monthly), and use bootstraps to conclude that these are significant at the 1% level. The improvement in OOS performance is attributed to the new information found in the cross-section and the new method used to extract such information. In this paper I show that the statistical significance of this OOS predictability is overstated for two reasons.

Firstly, the analysis in KP is conducted on gross returns rather than excess returns. Cochrane (2005) emphasizes that the predictability literature is primarily concerned with whether there is time variation in the reward for risk. To do this, one has to establish the predictability of excess returns rather than gross returns which contain both the reward for risk and the risk-free rate. However, KP predict gross market returns (yearly and monthly) and therefore include the interest rate component, which may be relatively easy to predict. I show that the inclusion of predictable movements of interest rates overstates the predictability of the equity premium.

Secondly, the bootstrap statistics that KP used to assess OOS significance do not account for small-sample bias in the estimated coefficients. This bias is well known to affect inferences for IS predictive regressions (Stambaugh, 1986) and is particularly problematic when both the predictor variable is persistent and the innovations in the predictor variable are negatively correlated with innovations in market returns. I show that it is also important for OOS tests of significance and that, when simulating samples for the OOS bootstrap statistics, it is important to allow for non-zero correlation between the innovations of the predictor variable and market returns.

Accounting for both these effects can radically change the conclusions about whether it is possible to predict market returns OOS. KP report results for extracting predictive information from book-to-market ratios of three sets of portfolios: Fama and French’s (1993) 6, 25, and 100 size- and value-sorted portfolios. The recursive OOS $R^2$ for the sample period 1965-2010 are insignificant for predicting one-year excess returns, and for one-month returns, apart from the case of book-to-market ratios of 6 portfolios which are significant at the 10% level. I notice that book-to-market ratios of 100 portfolios always generate significant forecasting, both annually and monthly for the period 1965-2010, on gross returns at the 1% level when neglecting the small-sample bias in the bootstrap (5% significance concerning small-sample bias). In contrast, negative $R^2$ is found for forecasting one-year and one-month excess returns, and thus I conclude the gross-return effect is bigger than the bootstrap effect, though both are non-negligible.

My research contributes to two main strands of the literature. Firstly, it emphasizes the importance of maintaining the focus on the equity premium, i.e. predicting excess rather than gross returns, in order to correctly identify whether there is a time-varying reward for risk. Secondly, my results show that, although using OOS tests may help to alleviate concerns over data mining,
they still suffer from small-sample bias, similar to that which has plagued inferences drawn from IS predictive regressions (see, in particular, Cochrane, 2008). This small-sample bias considerably increases the chances of seeing high point estimates of the OOS $R^2$ under the null hypothesis of no predictability. This bias is large when the candidate predictor variables are persistent and have innovations which are negatively correlated with market return innovations, both of which conditions hold in the case of the disaggregate predictors used in KP.

My study on predictability is naturally related to a discussion of efficient market hypothesis. The basic idea of efficient market hypothesis is all information is rationally and immediately incorporated in the prices of assets. In particular, there are three levels of efficient market hypothesis, described as weak, semi-strong, and strong forms (Fama, 1970). The weak-form says prices are independent and identically distributed. This characteristic of random walk implies that excess returns cannot be systematically obtained through the study of movement of historical prices. The semi-strong form says that stock prices reflect all information publicly available so that investment strategies relying on fundamental variables, such as financial variation ratios or economic indicators, cannot help investors achieve excess returns. The strong-form says that stock prices reflect all public and private information that is available so that excess returns cannot be obtained by trading on inside information. Returns can be predictable due to two reasons. Firstly, from the behavioural theory, mispricing can cause predictability if agents are irrational. For example, investors’ decision to buy or sell is not based on rational economic behaviour (e.g. the disposition effect). Secondly, mispricing can arise in a rational model where the risk exposure varies over time, e.g. ICAPM, and therefore expected returns can vary in a predictability way (Fama and French, 1988b; Balvers et al., 1990). Therefore, mispricing could happen and cause predictability, which is evidence of market inefficiency, or the time-series predictability of returns is simply evidence of time-variation risk premium even when the market is efficient. Like most tests of predictability using lagged predictor variables my models do not allow us to distinguish between these two possible explanations.

The rest of the paper is organized as follows. Section 2 discusses relevant studies. Section 3 depicts in detail the methodologies applied in this study, particularly developing the theory and implementation of the PLS method. Data descriptions are included in Section 4. Section 5 presents the empirical results for market return predictability and bootstrap statistics. Section 6 concludes.

2 Literature Review

The literature review chronologically discusses related studies on this issue, starting from the early research on the existence of predictability in Section 2.1, followed by the findings of significant forecasting power of financial and macroeconomic variables for market stock returns in Section 2.2, which back up the acknowledged belief in mainstream academia that stock returns are predictable. Though a prominent paper convincingly showed little support for stock return predictability and thus challenged its existence, particularly in the OOS context (see, in particular, Goyal and Welch, 2008, as discussed in Section 2.3), other scholars have defended the poor OOS performance as within expectations, and new techniques have emerged that have shown improved performance OOS, as depicted in detail in Section 2.4. In each section covering the different forecasting issues, studies are linked and reviewed as much as possible based on their commonalities. Based on the review of the previous literature, I propose research objectives in Section 2.5 which is the main contribution of my study.

2.1 Early Studies on Predictability

Early research on the movement of stock prices was motivated by the belief that stock returns were unpredictable, assuming that the stock market obeyed a random walk and that available information was fully and immediately reflected by the stock prices without hindrance (Bachelier (1900) using an arithmetic form, and Osborne (1959) using a logarithmic form). A simple and colloquial explanation is that stock returns in the future are independent of their historical variations. The emergence of this idea dates back to Bachelier (1900), who looked into the return movements of
undertaking more risk, but this reward varies over time, and it is pervasive if "the dogs did not bark" (Cochrane, 2008), i.e. high prices over time fail to be followed by higher earnings or profits. Fama emphasized that investors can obtain higher returns by exploiting economically inefficient, exhibiting an unpredictable market collapse for example. Apparently, stable profits and stable risk are a rejection of informational efficiency. Fama emphasized that investors can obtain higher returns by engaging in corrective price-arbitrage because the arbitrage trade is itself not without risk (see, e.g., De Long et al., 1990).

The above studies concentrated on the weak-form predictability discussion, analysing how past

1Fama (1970, p.383):"A market in which prices always 'fully reflect' available information is called "efficient"." Fama's efficient market hypothesis forecasts that the movements of stock prices are unpredictable in such a competitive market of informational efficiency. There is a misunderstanding of informational efficiency among the public. Strictly following the definition of this terminology, it only refers to the prices reflecting available information, rather than a self-regulating market. In addition, an efficient market is not required to run efficiently, and can be economically inefficient, exhibiting an unpredictable market collapse for example. Apparently, stable profits and stable risk are a rejection of informational efficiency. Fama emphasized that investors can obtain higher returns by undertaking more risk, but this reward varies over time, and it is pervasive if "the dogs did not bark" (Cochrane, 2008), i.e. high prices over time fail to be followed by higher earnings or profits.

It was not until the latter part of the 20th century that researchers started to consider the idea of predictable market returns either due to temporary deviations from the efficient market hypothesis, or due to time variation in market risk. In the first case, it is believed that the predictability of stock returns, to some degree, reflects irrational investor behaviour that pushes the prices from the fundamental value2. Thus profit opportunities emerge for arbitrageurs and this leads to the predictable future returns as the market reverts to the fundamental value. In terms of the latter viewpoint, studies are associated with asset pricing models that test the behaviour of either time-varying betas or time-varying risk premiums. Ferson and Harvey (1991) asserted that it is the market risk premiums that capture the predictability of common stocks.

There are many studies looking at departures from the simple random walk model of unpredictable stock returns. Early work focused on whether information in past returns could forecast future returns. If realized returns are correlated with expected returns, then to some degree, past returns can be used as a proxy for the expected return. Studies examining this form of market efficiency focus on testing for mean reversion in aggregate stock market returns. The variance-ratio test is one of the most powerful tests to detect mean reversion in stock prices. Evidence based on this statistic shows that, particularly over long horizons, market excess returns are predictable. Interestingly, in the case of market indices and portfolios constructed based on size, Poterba and Summers (1988) demonstrated negative serial correlation of stock returns using monthly and yearly observations, while Lo and MacKinlay (1988) found positive correlation using weekly data, in the long run. In support of this, Lo and MacKinlay (1989) explained that the estimation could not reject the hypothesis of a random walk for weekly stock market returns until a simple volatility-based specification test was used. In addition, restrictions have been put forward regarding the mean reversion model of asset prices. For individual securities, French and Roll (1986) demonstrated statistically significant negative serial correlation in terms of daily returns, but failed to show the economic significance of this result. By examining monthly individual returns, Jegadeesh (1990) found significantly negative first-order serial correlation and significantly positive correlations for longer lags (e.g. the 12-month lag was notably strong). When looking at weekly returns of individual securities, the correlation results turn out to be insignificant, both statistically and economically, due to the large amount of idiosyncratic noise (Lo and MacKinlay, 1988).

The above studies concentrated on the weak-form predictability discussion, analysing how past

2It is worth noticing that prices may deviate from their intrinsic values in the behavioural paradigm even in the presence of rational arbitrageurs. Deviations due to noise trading can exist because risk-averse arbitrageurs refrain from engaging in corrective price-arbitrage because the arbitrage trade is itself not without risk (see, e.g., De Long et al., 1990).
returns forecast future returns. This organizing principle can be derived from Fama’s (1970) early review. The semi-strong-form test is based on the information that is publicly available while the strong-form test considers all information whether publicly available or not. My study will concentrate on the semi-strong form tests of predictability discussed in the following Section 2.2.

2.2 The Relationship between Stock Returns and Financial/ Macroeconomic Variables

A major strand of the literature examines whether aggregate market variables can forecast stock market returns. Various studies have examined the relationship between aggregate stock market excess returns and lagged financial ratios, such as the dividend-price ratio, and book-to-market ratio (Fama and French, 1988a; Campbell and Shiller, 1988a,b; Kothari and Shanken, 1997). These studies generally model a linear relationship by regressing market excess returns on the lagged predictor variable in a univariate regression. Collectively these studies find evidence to reject the simple random walk null hypothesis of no predictability (Roeff, 1984). Graham and Dodd (1934) pointed out that higher valuation ratios, such as the dividend yield, could indicate an under-valued stock market, and would forecast greater subsequent returns. This standpoint was taken up by Rozeff (1984), Fama and French (1988a), and Hodrick (1992), among others. Using present value relationships, Campbell and Shiller (1988a,b) elaborate on the simple mechanism by which valuation ratios will reflect expectations of future returns in a dynamic setting where both expected returns and dividend growth can vary over time. They show that in general the dividend-price ratio is negatively related to future expectations of dividend growth and positively related to market expected returns; this holds regardless of the underlying reason for the predictability, i.e. whether predictability due to the rational time-variation of risk or predictability due to mispricing/behavioural reasons. They found that the dividend-price ratio is positively correlated with subsequent aggregate market returns and that its explanatory power (R²) rises with the length of the time horizon. It should also be noted that it is an old tradition initiated by Dow (1920) to use the dividend-price ratio, which has stronger explanatory power than the earnings yield, to predict returns (as explained by Goyal and Welch (2003)). The work by Fama and French (1992a,b, 1993, 1996, 1998) revealed a strong cross-sectional link between the book-to-market ratio and the average returns of individual stocks. Following this, a series of studies have been conducted on whether aggregate market returns are predictable using the aggregate book-to-market ratio, for example Kothari and Shanken (1997), Pontiff and Schall (1998)⁴, and Kheradyar et al. (2011). The basic form of the predictive regression of excess stock market return (R) on the aggregate book-to-market ratio (B/M) is expressed as follows:

\[
R_{t+1} = \alpha + \beta B/M_t + \epsilon_{t+1}
\]

where \(\alpha\) is a constant, \(\epsilon\) is the contemporaneous residual term, and the predictor variable, the book-to-market ratio, is the lagged value, implying the market information known at time \(t\). The aim is to see whether this variable forecasts the next period’s returns, and the test is whether the coefficient \(\beta\) is significantly different from zero. Values of R-square (\(R^2 = \sum_{t=0}^{T}(\hat{r}_t - \bar{r}_t)^2 / \sum_{t=0}^{T}(r_t - \bar{r}_t)^2\)) rises with the length of the time horizon. This organizing principle can be derived from Fama’s (1970) early review. The semi-strong-form test is based on the information that is publicly available while the strong-form test considers all information whether publicly available or not.

---

³The dividend-price ratio is the dividend divided by the price, where the dividend is the 12-months of dividend over a year, and the price is the price at the end of this period. Fama and French (1988a) also called it dividend yield and viewed these two terminologies as the same concept. Goyal and Welch (2008) defined them slightly differently, with dividend yield being the dividend divided by the price at the beginning of the dividend period. In this study, I follow Goyal and Welch’s definition although there is no theoretical motivation for this choice.

⁴Pontiff and Schall (1998, pp.158-159) showed the detailed inference of the book-to-market ratio’s ability to forecast returns. The current book value \(B\) is expressed as \(B = ECF + \eta\), where \(ECF\) is the cash flow expected to be distributed in the next time period, and \(\eta\) is the white noise with zero mean. The cash flow in the next period is \(CF = ECF + \varepsilon\). The current market price level \(M\) is estimated as \(M = \frac{ECF}{\varepsilon}\), where \(\kappa\) is the predetermined discount rate. Then, the book-to-market ratio \(B/M = \frac{CF}{\varepsilon} = \kappa + \frac{\kappa}{\varepsilon}\), and the return \(R = \frac{CF}{M} - 1 = \kappa + \frac{\kappa}{\varepsilon} - 1\). Therefore, the correlation between the book-to-market ratio and the return is

\[
\text{Corr}(B/M, R) = \frac{\text{Var}(\kappa)}{\sqrt{(\text{Var}(\kappa) + \text{Var}(\eta)\text{Var}(\frac{ECF}{\varepsilon})/\text{Var}(\varepsilon))\text{Var}(\kappa) + \text{Var}(\varepsilon)\text{Var}(\frac{ECF}{\varepsilon})}}
\]

An increase in either \(\text{Var}(\eta)\) or \(\text{Var}(\varepsilon)\) will result in a decreased ability of \(B/M\) to forecast returns.
average returns) are reported for market excess returns over the period.

It is important to notice that the persistence of financial ratios widely used as predictive variables in the literature, such as B/M and dividend-price ratio, is very high (e.g. B/M yearly values have an AR(1) estimate of 0.68 (monthly 0.97) (Pontiff and Schall, 1998)), implying predictability over long horizons\(^5\). This high persistence can cause econometric issues in small samples (Stambaugh, 1986). If the persistence is high and the innovations in returns and the financial ratios are negatively correlated then the value of \(\beta\) in the predictive regression is biased upwards. There are various approaches to correcting for this bias in the predictive regressions (Stambaugh, 1999; Lewellen, 1999, 2004; Campbell and Yogo, 2006), and one of the most common approaches in empirical work is to simulate bootstrap statistics under the null hypothesis (Goetzmann and Jorion, 1993; Kothari and Shanken, 1997; Rapach and Wohar, 2005, 2006; Lettau and Van Nieuwerburgh, 2008; Goyal and Welch, 2008). One of the main points of my paper is to show that this kind of small sample bias is also present in OOS tests of predictability and needs to be accounted for in the bootstrap tests of significance.

Apart from the financial ratios, interest rates are also commonly used as predictors. Fama and Schwert (1977), Keim and Stambaugh (1986), Campbell (1986), and Fama and French (1989) all asserted that yields on short-term and long-term treasury and corporate bonds are correlated with subsequent stock returns. More recently, new predictor variables have been taken into account, such as corporate payout and financing activity (Lamont, 1998; Baker and Wurgler, 2000), the ex-post consumption-wealth-income ratio (Lettau and Ludvigson, 2001), relative valuations of high/low-beta stocks (Polk et al., 2006) and others. As for macroeconomic variables, Chen et al. (1986), applying efficient market theory and rational expectations intertemporal asset pricing theory, asserted that asset prices should rely on their exposure to the state variables (e.g. industrial production) which explain the economy. Moreover, Lettau and Ludvigson (2001) explained the relationship between financial and macroeconomic forecast variables by stating that the former (e.g. dividend yield and payout ratio) perform better in predicting over long time horizons, while the latter (e.g. consumption and wealth ratio, and detrended T-bill rates) are more suitable for forecasting over short and medium horizons. In sum, it would be better to consider both financial and macroeconomic variables in forecasting. Thus, the basic predictive regression with only one predictor can be extended to contain two predictors or more, as follows:

\[
R_{t+1} = \alpha + \sum \beta_i X_{it} + \epsilon_{t+1}
\]

where \(X_i\) could be any potential predictor variable (e.g. dividend-price ratio, book-to-market ratio, earnings price ratio, and others). However, problems arise when too many predictor variables are considered. In other words, there is an issue with degrees of freedom and over-fitting, particularly when the number of predictor variables is large and/or the sample size is small.

### 2.3 Out-of-Sample Performance of Key Predictors

An influential study by Goyal and Welch (2008) used OOS tests of predictability to question whether the apparent predictability identified by previous research is spurious. They comprehensively reviewed the empirical evidence on equity premium prediction, by testing the ability of a set of the most commonly used predictor variables to predict market returns using an updated common time period. They considered predicting returns over a variety of horizons, including five-year, annual and monthly horizons. Furthermore, while most of the previous literature considered either IS or OOS tests, this paper employed both\(^6\). The outcomes demonstrated poor prediction

\(^5\)Cochrane (2005) argued that a scenario of increasing regression coefficients and \(R^2\) as the forecast horizon increases is a result of persistent predictor variables.

\(^6\)Lettau and Van Nieuwerburgh (2008) pointed out the ambiguity of the term “forecast” appearing in some research. Conventionally, it refers to IS regressions using the entire sample of data. When only available data are used to predict, it is normally termed an “out-of-sample forecast”. In this project, I follow this convention to avoid ambiguity, or more intuitively, I call “in-sample test” or “out-of-sample test”.

6
both IS and OOS for thirty years. Most variables were not significant even IS, using their up-to-date sample. The main focus was on the OOS predictive test. This type of test differs from the in-sample regression described by the regression (1) which uses all the data in the sample to assess the level of predictability. In contrast, the OOS test takes the perspective of an investor who has access only to the past data in order to form his forecast at a point in time. Therefore, at time $t$, the predictive model (1) is fit to past data $(t - 1 : t - n)$ only and the estimates of $\beta$ are used to make the next period’s forecast. Comparing the forecasts of the model to a simple forecast based on the rolling historical mean, one can define an OOS pseudo $R^2$ value. The OOS $R^2$ is similar to the traditional $R^2$, however, the value can be negative, implying less accurate predictions by the model than by the simple mean of historical returns. Goyal and Welch (2008) found that standard predictor variables performed particularly poorly, with negative adjusted $R^2$. For example, the adjusted $R^2$ for the dividend-price ratio and book-to-market ratio were -2.06% and -1.72% respectively at an annual frequency, and -0.30% and -3.28% respectively at a monthly frequency (Goyal and Welch, 2008). Other related studies include Bossaerts and Hillion (1999), who failed to find OOS forecasting power from the dividend yield, even though the model selection criteria implied external validity provided by the OOS test, and by Ang and Bekaert (2007), who found weak predictive power from the dividend yield through a nonlinear present value model.

Two papers have dealt with this criticism of the poor OOS ability of commonly used predictor variables. Campbell and Thompson (2008) argued that predictors with significant IS performance usually had better OOS performance than predictions based on the historical average returns. They showed that given the predictability implied by IS regressions, the poor OOS performance indicated in Goyal and Welch’s paper could occur for extended periods of time. This is due to the estimation error inherent in forming the OOS forecasts and not necessarily because returns are not predictable. The main contribution of their empirical section was that they imposed relatively weak restrictions on variables’ signs in the theoretically based forecast regressions and obtained considerably better OOS results. What is more, they showed that OOS forecasts could be further improved by restricting the coefficients’ magnitudes to the values suggested by steady-state models. Cochrane (2008) also supported return predictability, based on an inspection of papers demonstrating poor forecasting power over long-horizons and those providing regressions of return forecast with poor OOS $R^2$. He confirmed Goyal and Welch’s findings of poor OOS tests, but argued that such results were not unlikely considering the persistence of the dividend yield and the comparatively small set of usable sample data. In order to construct a powerful test for predictability, Cochrane established a null hypothesis in which return predictions could explain all dividend-yield volatility, and provided the same statistics as Goyal and Welch’s, which is the difference between the root mean squared error from the sample-mean return prediction and the root mean squared error from the fitted variation-ratio prediction. As we know, it should be changes in expectations of either returns or the dividend growth, or both, that drives the observable movement in dividend yields. Cochrane (2008)’s results showed that the OOS performance of dividend growth prediction was as bad as or even worse than the outcomes of return predictability in the data 30-40% of the time. Therefore, Goyal and Welch did not really provide a statistical rejection of predictable returns.

2.4 Emergence of New Techniques

Most recently, two papers have shown strong empirical evidence of OOS predictability by employing new predictor variables. First, the sum-of-the-parts (SOP) method, proposed by Ferreira and Santa-Clara (2011), predicts stock market returns OOS. This methodology decomposes market returns into three components, i.e. dividend-price ratio, earnings growth rate, and price-earnings ratio growth rate. They then analysed the time series properties and accordingly construct forecasts for each component, thereby constructing overall OOS forecast of market returns. The results demonstrated better OOS results than those of the historical mean or predictive regressions. Besides this, the paper also showed that average stock returns mainly depend on the dividend-price ratio and earnings growth components, while earnings growth and price-earnings ratio growth lead to most of the return volatility.
Another new technique has been proposed by Kelly and Pruitt (2013). They used the PLS methodology to forecast equity returns and explain the market price behaviour of U.S. capital stocks. The approach is a version of the three-pass regression filter (3PRF) method discussed in another paper by the authors (Kelly and Pruitt, 2015). Retaining the cross-sectional combination of information through least squares, PLS solves the potential computational problem of increasing the cross-sectional dimensions in the Kalman filter, induced by the fast growth of the state space and the number of parameters. Furthermore, PLS outstrips the widely applied principal components (PC) benchmark, particularly reflected through a reduction of dimensions. To illustrate, PC extracts predictive information according to the covariance within predictors that would generate suboptimal forecasts. By contrast, PLS condenses the cross-section depending on the forecast target, and consequently the selected linear combination of predictors is optimal for forecasting (Kelly and Pruitt, 2013).

A big advance made by KP’s study was the use of disaggregate valuation ratios as predictors, which provide more changeability than their aggregate counterparts that were used in most of the previous studies. To illustrate, assume that the expected market returns ($\mu_t$) and cash flow growth rate, also referred as return on equity ($g_t$), are economically important common factors. The individual valuation ratio, $B/M_i$, and the aggregate market valuation ratio, $B/M_t$, can be assumed to be exposed to these factors (Kelly and Pruitt, 2013):

\[ B/M_{i,t} = \beta_i + \beta_{i,\mu}\mu_t + \beta_{i,g}g_t + \omega_{i,t} \]  
\[ B/M_t = \beta_0 + \beta_{\mu}\mu_t + \beta_{g}g_t \]  

where $\mu$ is the expected market return over a unit time period such that the realized market returns are given by $r_{m,t} = \mu_{t-1} + \eta_{m,t}$, $g$ represents the expected cash flow growth rate such that the dividend growth is given by $\Delta cf = g_{t-1} + \eta_{d,t}$, and $\eta^r$ and $\eta^d$ are residual terms for $r_{m}$ and $\Delta cf$ respectively. Therefore, the linear forecast regressions of returns on the predictor $B/M_t$ can be expressed as follows:

\[ E_t[r_{t+1}|B/M_t] = \tilde{\alpha} + \tilde{\gamma}B/M_t \]
\[ = \tilde{\alpha} + \gamma(\beta_{\mu}\mu_t + \beta_g g_t), \]  

where both $\tilde{\alpha}$ and $\tilde{\alpha}$ are constants. Clearly, from (5), the regressions are affected by the information given by the expected return on equity. Since both $\mu_t$ and $g_t$ are important latent factors, the noise from each could potentially generate accuracy problems in the information extraction process. Even though Lettau and Van Nieuwerburgh (2008) use the present value system to deal with this difficulty of a joint relationship, the regression still wholly depends on the aggregate variable. By contrast, using the disaggregate scheme, $B/M_{i,t}$, takes more information regarding these two latent factors into account, which improves the precision of the final estimations, as long as no redundant information is encompassed in $B/M_{i,t}$. As a matter of fact, this paper was not the first to model valuation ratios for individual assets. However, it was the first to investigate the use of a factor structure in valuation ratios to predict market returns.

In fact, PLS is a modification based on the two-pass regression model initially used by Fama and MacBeth (1973) in risk-return tests. It is comprised of three regressions, all of which are carried out via ordinary least squares (OLS). Details of the process will be presented and explained in the methodology part of this paper. Through this method, a large number of predictors can be exploited for a relatively short time series, which used to represent a difficulty in the empirical asset pricing area. The outcomes show good predictions of market returns both IS and OOS. Particularly, the OOS $R^2$ for stock returns is approximately 13% on a one-year horizon when using 100 portfolios of book-to-market ratios as predictors.

2.5 Research Objectives

In this paper I propose to re-examine the extent to which market returns are predictable using the PLS approach pioneered by KP. Particularly, I focus on the OOS results which suggest strong levels of predictability, for example, KP report the one-year OOS $R^2$ of 5.81%, 3.49%, and 13.07% for book-to-market ratios of 6, 25, and 100 portfolios respectively.
I am particularly concerned about two aspects which may affect their analysis. Firstly, they report results for predicting gross market returns rather than the excess market return. Although the risk-free rate is a relatively small component of the total return, this can potentially affect the apparent predictability of the equity premium by including a component due to the predictable variation of the risk-free rate. It is a common practise to focus on predicting excess returns, particularly when dealing with longer, yearly horizons. I therefore conduct my analysis on excess returns, and in order to show the important difference this can make I repeat the analysis on gross returns.

Secondly, although the OOS $R^2$ values that come out of a PLS approach are very high, particularly for yearly returns, an important question is how frequently such high $R^2$ occur under the null hypothesis of no predictability. In order to assess this one needs to construct a bootstrap which accounts for the effects of small sample bias in the OOS test statistic. In particular, I show that it is important to maintain the correlation between innovations of the market returns and the innovations of the predictor variables. In KP’s bootstrap analysis these correlations are set to zero and this gives rise to critical values which are substantially lower. In order to show the magnitude of this effect I present results for bootstrap significance for both cases.

My main final set of bootstrap results account for both effects simultaneously, i.e. predicting excess returns and using a bootstrap technique which accounts for small sample bias. I think this gives the best picture of whether it is possible to predict returns OOS using the PLS technique. In the subsequent sections I describe the methodology including the details of my bootstrap technique.

3 Methodology

3.1 In-Sample Fit and Out-of-Sample Forecast

A. In-Sample Fit

The framework of the conventional linear regression forecasting model is as follows:

$$ r_t = \alpha + \sum_{i=1}^{N} \beta_i x_{i,t-1} + \varepsilon_t, $$

where $r_t$ is the log stock return in excess of the risk-free rate for the period $(t-1, t)$, measured as $r_t = log(R_t) - log(R_{f,t})$ where $R_t$ is the stock return and $R_{f,t}$ the risk-free rate. $x_{i,t-1}$ denotes the log value of the lagged predictor at the end of $t-1$, which could be a vector ($T \times 1$) of one variable or a matrix ($T \times N$) containing a combination of $N$ variables that are believed to have potential forecasting power for future returns. For example, $x_t$ could be book-to-market ratios of 6 size- and value-sorted portfolios, $[sl_t sm_t sh_t bl_t bm_t bh_t]$. $\varepsilon_t$ is a disturbance term with mean zero. Suppose there are observations of $r_t$ and $x_t$ for $t = 1, 2, ..., T$. Then, there are $T$ usable observations when using the entire data sample to conduct an IS forecast regression based on (6), giving

$$ \mathbb{E}[r_t] = \alpha + \sum_{i=1}^{N} \beta_i x_{i,t-1}. $$

I use $\hat{\alpha}$ and $\hat{\beta}_i$ as OLS estimates for $\alpha$ and $\beta_i$ respectively. Small-sample bias means that $\hat{\beta}_i$ is biased upward. This is because $x_t$ is not an exogenous regressor in (6), and could be correlated with the disturbance, i.e. $\mathbb{E}(\varepsilon_t x_{i,t-1}) \neq 0$, or more generally, $\mathbb{E}(\varepsilon_t x_m x_n) \neq 0, m < t \leq n$ (Stambaugh, 1986, 1999; Nelson and Kim, 1993). Another issue worth noting is the overlapping property of multi-period returns. The benefit of using overlapping observations rather than non-overlapping observations is acknowledged that it can improve the power of test statistics, and to account for overlapping, it is important to consider the autocorrelation (Richardson and Stock, 1986, 1999) also noted that this bias is a decreasing function of the sample size and an increasing function of the persistence in the predictor and of the contemporaneous correlation between innovations in the return and predictor.

---

7sl: small size and low book-to-market ratio (bm), sm: small size and medium bm, sh: small size and high bm; bl: big size and low bm, bm: big size and medium bm, bh: big size and high bm.
8Stambaugh (1986, 1999) also noted that this bias is a decreasing function of the sample size and an increasing function of the persistence in the predictor and of the contemporaneous correlation between innovations in the return and predictor.
B. Out-of-Sample Forecast

The pure OOS prediction is a recursive scheme consisting of a series of OLS regressions. The total sample of $T$ is divided into two parts, i.e., a training sample (from 0 to $\tau - 1$) and a forecast sample (from $\tau$ to $T$), with the former used to build the model which is then used to make a forecast on the latter. Following Goyal and Welch (2003, 2008), the one-step-ahead OOS prediction of returns is generated using an expanding window for evaluating $\alpha$ and $\beta_i$ in (6). To explain, I start by using the training sample of $\{x_{i,k}\}_{k=0}^{\tau - 1}$ to estimate alpha and beta, and use Eq.(7) to forecast the next period’s return $r_{\tau}$. The next return, $r_{\tau+1}$, is forecasted using an expanded training sample of $\{x_{i,k}\}_{k=0}^{\tau + 1}$. I repeat this regression process until the forecasting sample is exhausted.

3.2 Partial Least Squares Methodology Applied to Forecasting Returns

PLS provides a framework for connecting aggregate market return expectations to disaggregate valuation ratios in a dynamic latent factor system. It consists of two kinds of inputs, the forecast target and a set of predictors, and uses the three-step filter to extract factors driving the expected stock market returns from the panel of predictors.

3.2.1 Present Value Structure

First I describe the present value structure that allows KP to link the book-to-market ratios to the same latent factors which drive expected market returns. Kelly and Pruitt (2013) focused on aggregate market gross returns ($\mu_m$) and cash flow growth ($g_m$) as the inputs for the forecast target, and the book-to-market ratios of size- and value-sorted portfolios as predictors. I follow their idea but concentrate on forecasting the gross and excess market returns. To build the model, I start from the cross-sectional present value scheme. Similar to (3) and (4), it is assumed that the expected one-period-ahead log returns and log cash flow growth respond to the underlying latent factors, and $\zeta_i$ is the residual term of cash flow growth. Hereafter, I use the superscript $'$ as the mathematical notation for the transpose. For the aggregate market, denoted by the subscript $m$ for $r$ and $\Delta cf$, the structures are the same but the residuals are removed, such that

$$
\mu_t = E_t[r_{m,t+1}] = \kappa_0 + \kappa' F_t \tag{8}
$$

$$
g_t = E_t[\Delta cf_{m,t+1}] = \varphi_0 + \varphi' F_t \tag{9}
$$

where $\kappa$ and $\varphi$ are the loading factors demonstrating how the expected asset returns and cash flow growth respond to the underlying latent factors, and $\zeta_i$ is the residual term of cash flow growth. Hereafter, I use the superscript $'$ as the mathematical notation for the transpose. For the aggregate market, denoted by the subscript $m$ for $r$ and $\Delta cf$, the structures are the same but the residuals are removed, such that

$$
\mu_t = E_t[r_{m,t+1}] = \kappa_0 + \kappa' F_t \tag{10}
$$

$$
g_t = E_t[\Delta cf_{m,t+1}] = \varphi_0 + \varphi' F_t \tag{11}
$$

I assume that there are $J_o$ factors relevant to the market expectations, and the $(J - J_o)$ irrelevant factors will be filtered through the PLS no matter how large $J$ is. To further explain, in Section (2.4), we have already seen that the realized aggregate stock returns and the market cash flow growth rates are equivalent to their expected values plus a noisy term. The structures for the individual equity $i$ are the same in that

$$
r_{i,t+1} = \mu_{i,t} + \zeta^r_{i,t+1} \tag{12}
$$

$$
\Delta cf_{i,t+1} = g_{i,t} + \zeta^d_{i,t+1} \tag{13}
$$

9 Other than using an expanding (i.e. recursive) window, prediction can also be conducted by using a rolling window, which leaves out some earlier observations when later observations are added in. However, this may cause structural instability. Thus, I use the expanding window in the forecasting process.

10 It is noticeable that the last repetition of the OOS forecasting looks like the IS process. However, IS forecasting carries out the regression on the full series, while OOS forecasting always leaves one return for forecasting. Therefore, the beta obtained in the last OOS repetition work can be very close to the IS result but they are not the same.

11 There is no idiosyncratic component for individual expected returns, the assumption being that they are completely explained by systematic market elements. However, for the cash flow rates (or expected dividend growth), the noise term cannot be ignored.
where \( \xi_t^r \) and \( \xi_t^d \) are the residual with respect to returns and the growth rate respectively. We can also link the individual returns, \( r_i \), to the expected market returns, \( \mu \), and link the individual growth rate, \( \Delta cf_i \), to the expected return on equity, \( g \), such that

\[
\begin{align*}
    r_{i,t+1} &= C_{i,\mu} \mu_t + \eta_{i,t+1}^r \\
    \Delta cf_{i,t+1} &= C_{i,g} g_t + \eta_{i,t+1}^d
\end{align*}
\]

(14)

(15)

where \( C_{i,\mu} \) and \( C_{i,g} \) are the coefficient constants of individual asset \( i \) with respect to the market returns and the growth rate respectively. \( \eta_{i}^r \) and \( \eta_{i}^d \) are the residual terms of asset \( i \) with respect to the returns and the growth rate respectively.

KP assume that the factor vector, \( \mathbf{F} \), has the property of a first-order vector autoregression, in that

\[
\mathbf{F}_{t+1} = \mathbf{\Omega F}_t + \mathbf{v}_{t+1},
\]

(16)

where \( \mathbf{\Omega} \) is the autoregression coefficient and \( \mathbf{v} \) is the residual. Then, combining Eq.(8), (16) and applying Vuolteenaho’s (2002) linearized present value system, we can derive the cross-section of the log disaggregate valuation ratios, \( bm_i \) (I use the log value of book-to-market ratios as a proxy here), such that

\[
\begin{align*}
    bm_{i,t} &= \frac{a_i}{1-b_i} + \sum_{j=1}^{\infty} b_j \mathbb{E}[\epsilon_t \Delta f_{i,t+j} + \Delta cf_{i,t+j}] \\
    &= \frac{a_i}{1-b_i} + \sum_{j=0}^{\infty} b_j \mathbb{E}\left[\left(\kappa_{t,0} + \kappa'_{t,0} F_{t+j} + \varphi_{t,0} + \varphi'_{t,0} F_{t+j} + \nu_{t,0} + \nu'_{t,0} F_{t+j}\right) + \epsilon_{i,t}ight]
\end{align*}
\]

(17)

where \( a_i \) and \( b_i \) are constants. Let \( t = (-1,1)' \), \( \mathbf{H}_t = (\kappa'_t, \varphi'_t) \), \( \mathbf{I} \) is the identity matrix, and given that \( \mathbf{\Omega} \) is the autoregression coefficient explained in Eq.(16), Eq.(17) can be rewritten as

\[
\begin{align*}
    bm_{i,t} &= \frac{a_i - \kappa_{i,0} + \varphi_{i,0}}{1-b_i} + \sum_{j=0}^{\infty} b_j \mathbb{E}\left[\mathbf{H}_t' \mathbf{F}_{t+j} + \epsilon_{i,t}\right] \\
    &= \frac{a_i - \kappa_{i,0} + \varphi_{i,0}}{1-b_i} + \sum_{j=0}^{\infty} b_j \mathbb{E}\left[\mathbf{H}_t' \mathbf{F}_{t+j} + \epsilon_{i,t}\right] \\
    &= \frac{a_i - \kappa_{i,0} + \varphi_{i,0}}{1-b_i} + \mathbf{H}_t' \mathbf{(I - b_i \mathbf{\Omega})^{-1}} \mathbf{F}_{t+j} + \epsilon_{i,t}
\end{align*}
\]

(18)

(19)

I define \( \psi_{i,0} = \frac{a_i - \kappa_{i,0} + \varphi_{i,0}}{1-b_i} \), and \( \psi' = \mathbf{H}_t' \mathbf{(I - b_i \mathbf{\Omega})^{-1}} \), then can get

\[
\begin{align*}
    bm_{i,t} &= \psi_{i,0} + \psi' \mathbf{F}_t + \epsilon_{i,t}
\end{align*}
\]

(19)

where I use the superscript \(-1\) as the mathematical notation for the inverse matrix. Therefore, referring to Eq.(10), Eq.(11), and Eq.(19), we can see that the disaggregate variation ratios are linked to the aggregate market expectations through the latent common factors.

### 3.2.2 Implementation of PLS

I describe PLS in a general framework first following Kelly and Pruitt (2015), and apply it to predicting returns at the end of this section. I have three sets of variables in implementing the PLS. The first set is the target variable I wish to predict: \( y \), which is a \( T \times 1 \) vector with time equal to \( 2, 3, ..., T + 1 \). The second is the predictor: \( \mathbf{W} = (w_1', w_2', ..., w_T')' = (w_1, w_2, ..., w_N) \), which is a \( T \times N \) matrix of variance-standardized predictor variables. The third is a matrix of proxy data, which are variables driven by target-related factors: \( \mathbf{Z} = (z_1', z_2', ..., z_T') \) with dimensions \( T \times L \), where \( L < \min(N, T) \). These proxy data are important because, through them, it is possible to reduce the amount of information used in the prediction. There are three steps in conducting the PLS approach. The first step is to run a set of time-series regressions of each predictor variable, \( w_i \), \( i = 1, ..., N \), on the proxy, \( \mathbf{Z} \), such that

\[
\begin{align*}
    w_{i,t} &= \phi_{i,0} + z_i \phi'_t + \epsilon_{i,t},
\end{align*}
\]

(20)
In this way, the factors irrelevant to the market returns are removed, and only the relevant latent factors are used in the second step regression to show the forecasting power of the predictor variables. The methodology simplifies considerably for the case of predicting the single target of aggregate market returns. In the second step, for each time point \( t = 1, 2, ..., T \), cross-sectional regressions of the predictors on the lagged fitted factors driving the target variable, which were derived in the second step, such that

\[
\begin{align*}
\phi_{i,t} & = \phi_{i,0} + F_t \beta + \zeta_{i,t}, \quad (i = 1, 2, ..., N), \quad F_t = (F_{1,t}, F_{2,t}, ..., F_{L,t})
\end{align*}
\]

are the coefficients with respect to the loading factors estimated in the first step, and their estimated values are presented as \( F_t \). The second step regression effectively looks for the best predictors using univariate regression over the full sample, and then the Step 2 regressions use this information to construct an optimal predictor. The PLS three-step process can also be applied in an OOS way, for which the data mining bias is mitigated. I discuss the IS and OOS methods in the next section.

### 3.3 Measurement of Forecast Evaluation

In the IS test, the null hypothesis of no predictability, \( \beta = 0 \), implies constant expected returns (see Eq.(22)). Since the direction of the impact of the lag predictor at time \( t \) on \( r_{t+1} \) is theoretically suggested to be positive, my test looks at whether \( \beta \) is significantly different from zero. The forecasting ability of the predictor is typically measured using \( R^2 \), which is evaluated as

\[
R^2 = \frac{\sum_{t=0}^{T} (\hat{r}_t - \bar{r}_t)^2}{\sum_{t=0}^{T} (r_t - \bar{r}_t)^2}
\]

where \( \{w_{i,t}\} = (w_{i,1}, w_{i,2}, ..., w_{i,T})' \) (\( t = 1, 2, ..., T \)), \( \phi_i' = (z_{i,1}, z_{i,2}, ..., z_{i,T})' \), and assuming, as in KP, that one latent factor drives the expected market returns. The potential predictors are the set of disaggregate \( bm_i \), and the proxy for the expected return \( \mathbb{E}[r_{m,t+1}] \) is the realised market return over that period \( r_{m,t+1} \). This leads to the following three regressions:

- **Step 1**: \( bm_{i,t} = \phi_{i,0} + r_{m,t+1} \phi_i + \epsilon_{i,t} \)
- **Step 2**: \( bm_{i,t} = \phi_{i,0} + \hat{\phi}_i F_t + \zeta_{i,t} \)
- **Step 3**: \( \hat{r}_{m,t+1} = \beta_0 + \hat{F}_t \beta_i \)

Intuitively, the first step uses univariate regressions to assess how much of the future market returns are in each candidate predictor variables \( bm_i \); those with large \( |\phi_i| \) are likely to be good predictors. The second step is a cross-sectional regression which effectively constructs a latent variable, \( F_t \), which is a weighted composite of the potential predictors \( bm_i \), where those with higher \( |\phi_i| \) from the first pass regression will be given more weight in the composite predictor. The final pass regression is a time-series regression to assess how well the latent variable \( F \) predicts returns. When this procedure is applied IS it is clear that data mining will be an issue. This is because the first step regression effectively looks for the best predictors using univariate regression over the full sample, and then the Step 2 regressions use this information to construct an optimal predictor. My bootstrap statistics will account for this data mining bias, as well as the inherent small-sample bias.
where \( \hat{r}_t \) denotes the forecasted return based on the predictor variables, and \( \bar{r}_t \) denotes the historical average returns. This indicator estimates the amount of the stock returns that is explained by the time-variation in the expected returns.

In the OOS test, I use Campbell and Thompson’s (2008) \( R^2_{oos} \) to assess the forecasting ability. In this stage, two strategies for making return predictions are performed and compared to build the OOS \( R^2 \). The first is to predict the return by implementing the three-step PLS in terms of OOS based on Eq.(23), that is, run time-series regression for the time 0 to \( \tau \), do cross-sectional regression at the time just before the period to predict, and then predict the return at time \( \tau + 1 \). The second is to calculate the mean of the sample return at \( \tau + 1 \). Then the OOS \( R^2 \) is estimated by comparing the mean squared forecast errors (MSFEs) of the above two strategies:

\[
R^2_{oos} = 1 - \frac{MSFE_O}{MSFE_H} = 1 - \frac{\sum_{t=0}^{\tau}(r_t - \hat{r}_t)^2}{\sum_{t=0}^{\tau}(\bar{r}_t - \bar{r}_t)^2}
\]

where \( MSFE_O \) and \( MSFE_H \) represent the MSFEs from the OOS PLS model and the historical mean model respectively. The OOS \( R^2 \) is different from the traditional \( R^2 \), with a range of \((-\infty, 1]\). When \( R^2 \leq 0 \), it indicates that the predictor variables do not outperform the target’s historical mean in providing an accurate forecast in terms of the MSFE.

3.4 Bootstrap under the Null

The log-linear vector autoregression (VAR) can be applied to generate artificial data that satisfy either the null hypothesis of no predictability or the alternative hypothesis. Since the PLS technique effectively searches over \( N \) potential predictors in order to produce an estimate of the expected return, it is conceivable that this collective search over the \( N \) potential predictors will cause the IS \( R^2 \) values to be biased upwards. Firstly, the PLS estimate of the conditional expected return is formed as a weighted average of the \( N \) predictor variables, with higher weights given to those predictors that predict well in the first-stage regression. As \( N \) increases, the possibility of weighting a spurious predictor from the first-stage highly becomes more likely, which would induce data mining. There is also the small-sample bias, that stems from persistent predictors and correlated innovations. To be more specific, take a simple OLS system, for example, which consists of a forecast target, \( r \), and a predictor, \( x \). \( r \) can be forecasted using lag \( x \) when \( x \) has a first-order autocorrelation, such that

\[
\begin{align*}
    r_t & = \alpha + p \cdot x_{t-1} + \nu_t; \quad \nu \sim i.i.d.(0, \sigma^2_{\nu}) \\
    x_t & = C + \beta \cdot x_{t-1} + \varepsilon_t; \quad \varepsilon \sim i.i.d.(0, \sigma^2_{\varepsilon})
\end{align*}
\]

where \( \nu \) and \( \varepsilon \) are shocks which are contemporaneously correlated with their covariance \( \sigma_{\nu, \varepsilon} \). \( \alpha \) and \( C \) are constants, and \( p \) and \( \beta \) are the AR(1) coefficients with respect to return \( r \) and predictor \( x \). Even though the predictive regression of \( r_t \) on \( x_{t-1} \) is consistent asymptotically, the regression coefficient \( p \) would be biased upward in finite samples due to the unixed \( x \) in the repeated samples, and this would further induce \( R^2 \) to be upwardly biased. Also, the bias in the least squares estimate of \( p \) will be proportional to the bias in the least squares estimate of \( \beta \) (Stambaugh, 1986).

In this section, I am going to simulate under the null hypothesis of no predictability and examine whether the \( R^2 \) values are biased upwards, and how the bias depends on \( N \).

3.4.1 Bootstrap A: Ignoring the Correlation Structure of Innovations

In this section, following Kelly and Pruitt’s (2013) idea of placebo tests, I perform bootstrap A under the null of predictability. The basic idea is that I use simulated book-to-market ratios of \( N \) portfolios \((N=6, 25, 100)\), which remove the true predicting power, to predict the simulated market returns, and thus produce the critical values examining on the predictability results. In each simulation, I conduct the first-order autocorrelation processes on book-to-market ratios, and
kept the mean, variance and autocorrelation of book-to-market ratios the same\textsuperscript{12}. Innovations of simulated market returns and innovations of book-to-market ratios are constructed by a random-number generator. The procedure of bootstrap B is similar to bootstrap A under the null, which will be described in the next section. The only difference is that bootstrap B does not set the correlation of the innovations of returns and predictors to zero, while bootstrap A sets it to zero.

To explain in detail, bootstrap A starts under the null hypothesis of no predictability of stock returns:

$$r_t = \alpha + \varepsilon_{0,t},$$

where \( r \) is the stock market returns, \( \alpha = \mathbb{E}[r] \) is a constant, subscript \( t = 1, 2, ..., T \), and

$$\{\varepsilon_{0,t}\} = [\varepsilon_{0,1}, \varepsilon_{0,2}, ..., \varepsilon_{0,T}]^T$$

are innovations of returns. I consider the first-order autocorrelation (AR(1)) of book-to-market ratios to get innovations. Suppose the time-series length is \( T \) (\( 0 \leq t \leq T - 1 \)) and I have \( N \) book-to-market ratios (\( N = 6, 25, 100 \)); then I have a matrix of innovations of size \([T \times N]\). To illustrate, I consider the vector of book-to-market ratios, \( bm_{i,t} \), such that\textsuperscript{13}

$$bm_{i,t} = [x_{1,t}, x_{2,t}, ..., x_{N,t}]$$

$$x_{1,t} = C_1 + \beta_1 \cdot x_{1,t-1} + \varepsilon_{1,t}$$

$$x_{N,t} = C_N + \beta_N \cdot x_{N,t-1} + \varepsilon_{N,t}$$

where \( x \) represents the predictors (i.e., the book-to-market ratios of size- and value-sorted portfolios), \( C_{i,0} (i = 1, ..., N) \) are constants, and \( \beta_i \) are the AR(1) coefficients. The innovations, \( \varepsilon_{i,1}, \varepsilon_{i,2}, ..., \varepsilon_{i,T} \) form a matrix of size \([T \times N]\):

$$\{\varepsilon_{i,j}\}_{bm} = \begin{pmatrix}
\varepsilon_{1,1} & \varepsilon_{1,2} & \cdots & \varepsilon_{1,N} \\
\varepsilon_{2,1} & \varepsilon_{2,2} & \cdots & \varepsilon_{2,N} \\
\vdots & \vdots & \ddots & \vdots \\
\varepsilon_{T,1} & \varepsilon_{T,2} & \cdots & \varepsilon_{T,N}
\end{pmatrix}$$

\textsuperscript{12}AR(1) obeys the difference equation for the predictor:

$$x_t = C + \beta \cdot x_{t-1} + \varepsilon_t$$

where \( C \) is a constant, and \( \varepsilon_t \) is a white noise sequence with mean zero (\( \mathbb{E}(\varepsilon_t) = 0 \)) and variance \( \sigma^2 \) (\( \mathbb{E}(\varepsilon_t^2) = \sigma^2 \)). These \( \varepsilon \) are uncorrelated across time. When \( |\beta| < 1 \), there exists a covariance-stationary process for \( x_t \) such that

$$x_t = (C + \varepsilon_t) + \beta \cdot (C + \varepsilon_{t-1}) + \beta^2 \cdot (C + \varepsilon_{t-2}) + \beta^3 \cdot (C + \varepsilon_{t-3}) + \cdots$$

$$= \frac{C}{1 - \beta} + \varepsilon_t + \beta \varepsilon_{t-1} + \beta^2 \varepsilon_{t-2} + \beta^3 \varepsilon_{t-3} + \cdots.$$

The mean of a stationary AR(1) is \( \mu = \mathbb{E}(x_t) = \frac{C}{1 - \beta} \), and the variance is

$$\mathbb{E}(X_t - \mu)^2 = \mathbb{E}(\varepsilon_t + \beta \varepsilon_{t-1} + \beta^2 \varepsilon_{t-2} + \beta^3 \varepsilon_{t-3} + \cdots)$$

$$= (1 + \beta^2 + \beta^4 + \beta^6 + \cdots) \cdot \sigma^2$$

$$= \frac{\sigma^2}{1 - \beta^2}$$

Therefore, the mean, variance and autocorrelation are related to \( C, \beta, \sigma^2 \). \( C \) and \( \beta \) are estimated by OLS from the AR(1) equation with the full sample of observations. I store the residuals, \( \varepsilon_t \), and then generate 10,000 bootstrapped samples by drawing with replacement from the residuals. Since simulated residuals are drawn randomly from the original set, if we keep the values of \( C \) and \( \beta \) unchanged, then we keep the same mean, variance and AR(1) in the synthetic sample of book-to-market ratios.\textsuperscript{13}

The bootstrap process under the null is similar to the procedures in Nelson and Kim (1993), and Berkowitz and Kilian (2000), in that the data are generated as follows:

$$r_t = \alpha_0 + \varepsilon_{0,t}$$

$$x_t = C + \beta \cdot x_{t-1} + \varepsilon_{1,t}$$

where \( \varepsilon = [\varepsilon_{0,t}, \varepsilon_{1,t}] \) is independently and identically distributed (i.i.d.) with covariance matrix \( \Sigma \). However, this procedure implicitly means that each predictor variable is analysed in isolation. In fact, there are many candidate predictors, which would lead to a size distortion in this case. My process explicitly takes this point into account by putting \( N \) different variables as potential predictors into the predictive regression function.
where \( i \in (1, 2, \ldots, N) \) and \( j \in (1, 2, \ldots, T) \). Now there are two sets of innovations: innovations of returns, \( \{ \epsilon^{r}_{i,j} \} \), and innovations of predictors, \( \{ \epsilon_{i,j} \} \). I use a random-number generator to create a synthetic sample of innovations with replacement. This process is conducted separately for \( \{ \epsilon^{r}_{i,j} \} \) and \( \{ \epsilon_{i,j} \} \), and thus the correlation between \( \epsilon^{r}_{0,t} \) and \( \epsilon^{r}_{1,t} \) is set to zero. Then, the original innovations of returns and predictors become the new ones, \( \{ \epsilon^{*r}_{i,j} \} \) and \( \{ \epsilon^{*}_{i,j} \} \), as follows:

\[
\{ \epsilon^{r}_{0,j} \} = [\epsilon^{r}_{0,1}, \epsilon^{r}_{0,2}, \ldots, \epsilon^{r}_{0,T}]
\]

\[
\{ \epsilon^{*r}_{i,j} \} = \begin{pmatrix}
\epsilon^{r}_{1,1} & \epsilon^{r}_{2,1} & \cdots & \epsilon^{r}_{N,1} \\
\epsilon^{r}_{1,2} & \epsilon^{r}_{2,2} & \cdots & \epsilon^{r}_{N,2} \\
\cdots & \cdots & \cdots & \cdots \\
\epsilon^{r}_{1,T} & \epsilon^{r}_{2,T} & \cdots & \epsilon^{r}_{N,T}
\end{pmatrix}.
\]

I carry out this process because I want to keep the constant, \( C_{i} \), unchanged, the persistence related to \( x_{i,t-1} \), \( \dot{p}_{i} \), unchanged, and the distribution of \( \epsilon_{i,j} \) unchanged. I arrange the matrix of innovations in Eq.(31) in this new order and recreate the synthetic book-to-market ratios, as follows:

\[
\begin{align*}
\tilde{b}_{m,t,t} & = [\tilde{x}_{1,t} \tilde{x}_{2,t} \cdots \tilde{x}_{N,t}] \\
\tilde{x}_{1,t} & = C_{1} + \beta_{1} \cdot \tilde{x}_{1,t-1} + \epsilon^{*}_{1,t} \\
\vdots & \vdots \\
\tilde{x}_{N,t} & = C_{N} + \beta_{N} \cdot \tilde{x}_{N,t-1} + \epsilon^{*}_{N,t}
\end{align*}
\]

where the star denotes the new innovations after replacement, and the tilde the new set of book-to-market ratios. Similarly, I build the synthetic stock returns with rearranged innovations as follows:

\[
\tilde{r}_{t} = \alpha + \epsilon^{*r}_{0,t}
\]

I use each 10,000 bootstrap sample of replaced forecast target, \( \tilde{r} \), and replaced predictors, \( \tilde{b}_{m} \), to run the PLS IS process, and store the \( R^2 \) value. From these, critical values at 1%, 5%, and 10% levels can be gained by selecting the 9,000th, 9,500th and 9,900th value based on the size of bootstrap \( R^2 \) values (low to high). Similarly, I run the PLS OOS regression for each 1,000 bootstrap sample to produce an OOS \( R^2 \), and further generate critical values at 1%, 5%, and 10% levels.

### 3.4.2 Bootstrap B: Maintaining the Correlation Structure of Innovations

Bootstrap B is similar to bootstrap A. The only difference is when bootstrap A stores the innovations of market returns and innovations of book-to-market ratios separately in Eq.(27) and Eq.(29), bootstrap B stores them in one matrix as follows:

\[
\begin{pmatrix}
\epsilon^{r}_{0,1} & \epsilon_{1,1} & \epsilon_{2,1} & \cdots & \epsilon_{N,1} \\
\epsilon^{r}_{0,2} & \epsilon_{1,2} & \epsilon_{2,2} & \cdots & \epsilon_{N,2} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\epsilon^{r}_{0,T} & \epsilon_{1,T} & \epsilon_{2,T} & \cdots & \epsilon_{N,T}
\end{pmatrix}
\]

for \( i = 0, 1, \ldots, N \) and \( j = 1, 2, \ldots, T \).

After that, a random number generator is used to generate a synthetic sample of innovations. The innovations are randomly drawn in tandem, thus maintaining the correlation between the market return innovations and the predictor variable innovations. Therefore, I obtain synthetic
innovations as follows:
\[
\{ \varepsilon_{i,j}^* \} = \begin{pmatrix}
\varepsilon_{0,1}^* & \varepsilon_{1,1}^* & \cdots & \varepsilon_{N,1}^* \\
\varepsilon_{0,2}^* & \varepsilon_{1,2}^* & \cdots & \varepsilon_{N,2}^* \\
\vdots & \vdots & \ddots & \vdots \\
\varepsilon_{0,T}^* & \varepsilon_{1,T}^* & \cdots & \varepsilon_{N,T}^*
\end{pmatrix},
\]
(35)

The process after innovation replacement is the same as bootstrap A. I form synthetic market returns and book-to-market ratios with these constructed innovations. Then I use each bootstrap sample of replaced variables to run the PLS, both IS and OOS, with the values of $R^2$ stored, which would finally provide critical values.

4 Data

The primary data used in this study includes the forecast target, i.e. the aggregate market returns, and the predictor inputs of book-to-market ratios. The corresponding raw data are collected from the Kenneth R. French Data Library17. The alternative commonly used predictor variables are collected from Amit Goyal’s website18. Details about alternative predictor variables can be found in Appendix 7.1.

Log Gross Returns: Monthly market returns are taken from the Kenneth R. French Data Library for the period of January 1930 - December 2010. The “Fama/French factors” contain the data for $R_m - R_f$ and $R_f$. The former is the gross market return for the CRSP NYSE/AMEX/NASDAQ value-weighted index19 at the beginning of month $t$ minus the one-month Treasury bill rate taken from Ibbotson Associates. These data are transformed into monthly log returns and overlapping annual continuously compounded gross returns, as follows:

\[
\begin{align*}
\log r_a &= \log r_t + \log r_{t-1} + \ldots + \log r_{t-11} \\
&= \log \left[ (1 + R_t)(1 + R_{t-1}) \ldots (1 + R_{t-11}) \right] \\
&= \log \left( \prod_{j=0}^{11} (1 + R_{t-j}) \right)
\end{align*}
\]
(36)

where $r_j$ ($j = 0, 1, \ldots, 11$) is the continuously compounded monthly return, $R_j$ is the simple monthly return, and $r_a$ is the continuously compounded overlapping annual return.

Log Excess Returns: The one-month T-bill rates used for generating monthly excess returns ($R_f$) are downloaded from the Kenneth R. French Data Library. Then, the log excess returns, also called the equity premium ($EQP$), can be calculated as follows:

\[
\begin{align*}
EQP_{(m)} &= \log(1 + R_{m,t}) - \log(1 + R_{f,t}) \\
&= \log \left( \frac{1 + R_{m,t}}{1 + R_{f,t}} \right) \quad \text{for monthly;}
\end{align*}
\]

\[
\begin{align*}
EQP_{(a)} &= \log \left( \prod_{j=0}^{11} (1 + R_{m,t-j}) \right) - \log \left( \prod_{j=0}^{11} (1 + R_{f,t-j}) \right) \\
&= \sum_{j=0}^{11} \log \left( \frac{1 + R_{m,t-j}}{1 + R_{f,t-j}} \right) \quad \text{for annual.}
\end{align*}
\]
(37)

For monthly data, the period runs from January 1930 to December 2010, and for annual data it runs from December 1930 to December 2010.

17Source: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
18Source: http://www.hec.unil.ch/agoyal/
19Fama and French have revised this proxy, i.e. $R_m - R_f$ in the U.S. market, to include all the CRSP companies incorporated in the U.S. and listed on the NYSE, AMEX or NASDAQ with a CRSP share code of 10 or 11 (common stock).
20In this paper, $\log x$ refers to the natural logarithm of $x$
**Book-to-Market Ratios:** 6/25/100 portfolios of the monthly average size of firms, the monthly number of firms, and the annual book equity (BE) used for generating the book-to-market ratios of size- and value-sorted portfolios are available from the Kenneth R. French Data Library\(^{21}\). The product of the former two variables is the market capitalization (ME). The log book-to-market ratio, \( \log(bm) \), equals to the log value of BE divided by ME, assuming that the BE of the last fiscal year ending in calendar year \( Y \) becomes observable after June of the year \( Y + 1 \).\(^{22}\) I build the monthly book-to-market ratios of 6/25/100 size- and valued-sorted portfolios from December 1929 to December 2009 for overlapping annual forecasting, and those from December 1929 to November 2010 for monthly forecasting. Figure 1 demonstrates the year-end book-to-market ratios of 6 portfolios. The disaggregated data for the book-to-market ratios is more volatile than the aggregate market data, with a standard deviation of 0.49 on average across the book-to-market ratios of 6 portfolios for the year-end data of 1930-2010, versus 0.33 for the aggregate book-to-market ratios.

*** Insert Figure 1 about here ***

5 **Empirical Results**

5.1 **In-Sample Prediction of Market Returns**

In this section, I analyse the IS predictability of aggregate gross market returns and excess market returns, by using the cross-section of book-to-market ratios as predictors. I use the IS PLS methodology described in Section 3.2. I compare my results with those of Kelly and Pruitt (2013), who, in fact, focus on the gross returns forecast. Table 1 shows the results of the market return predictions based on the book-to-market ratios of 6, 25 and 100 portfolios of the U.S. stocks. I predict returns over two horizons, one-year and one-month.

*** Insert Table 1 about here ***

Panel A in Table 1 uses gross returns as the forecast target, and panel B uses excess returns. In both panels, the \( R^2 \) results indicate that a single factor extracted through PLS can generate the predictability for one-year returns with high \( R^2 \) values, approaching 8.94\%, 14.92\% and 19.89\% based on book-to-market ratios of 6, 25 and 100 portfolios respectively for gross returns, and 7.01\%, 15.02\% and 17.12\% respectively for excess returns. For the monthly forecasts, prediction of gross returns in panel A shows the monthly IS \( R^2 \) values of 0.54\%, 0.89\% and 2.01\% from the PLS factor extracted from book-to-market ratios of 6, 25, and 100 portfolios, and prediction of excess returns in panel B shows similar \( R^2 \) values but a bit smaller. I produce \( R^2 \) values for gross returns which are very close to those of Kelly and Pruitt (2013), showed in brackets. I do not get exactly same \( R^2 \) values as KP’s, because although I use the same data period, the data updates in the database and thus is a slightly later version of KP’s data, and therefore leads to the slightly changes of result values, but basically, I get very similar results, except they are actually slightly better in annual forecast with higher \( R^2 \).

*** Insert Table 2 about here ***

\(^{21}\)Fama and French (1992a) used all the nonfinancial firms in the intersection of (a) the NYSE, AMEX, and NASDAQ return files (represented by 1, 2 and 3 respectively in terms of exchange code) and (b) the merged COMPUSTAT annual industrial files of income-statement and balance-sheet data. Nonfinancial firms refer to firms with ordinary common equity (as classified by CRSP share code 10 or 11). Thus, American depository receipts, real estate investment trusts, and units of beneficial interest are excluded (Fama and French (1993, p.9)).

\(^{22}\)The alignment of the book-to-market structure is as follows:

\[
\begin{pmatrix}
192907bm = \log(\frac{1929BE}{1929ME}) & 193007bm = \log(\frac{1930BE}{1930ME}) & \cdots & 201007bm = \log(\frac{2010BE}{2010ME}) \\
192908bm = \log(\frac{1929BE}{1929ME}) & 193008bm = \log(\frac{1930BE}{1930ME}) & \cdots & 201008bm = \log(\frac{2010BE}{2010ME}) \\
\vdots & \vdots & \ddots & \vdots \\
193006bm = \log(\frac{1930BE}{1930ME}) & 193106bm = \log(\frac{1931BE}{1931ME}) & \cdots & 201016bm = \log(\frac{2010BE}{2010ME})
\end{pmatrix},
\]

with yearly BE used from July of the BM-corresponding year to June of the next year.
To compare the forecasting power of the PLS methodology, which uses disaggregated variables as predictors, Table 2 shows the predictability from a series of alternative predictors that have been used and studied widely in the earlier literature. The table considers IS forecasts under a one-year horizon in panel A and a one-month horizon in panel B for each predictor variable. In both panels, the $R^2$ results using the gross return and the excess return as the forecast target are demonstrated. I investigate 13 representative aggregate predictor variables surveyed by Goyal and Welch (2008). The full sample covers the 1930-2010 period. The OOS forecasting starts in January 1980 for a one-month horizon and December 1980 for a one-year horizon. Among the alternative predictors of the aggregate valuation ratios, the most significant predictions over the one-year horizon (panel A) come from the aggregate book-to-market ratio (bm) and the net equity expansion (ntis) with the IS $R^2$ values of 9.24% and 9.18% respectively for gross return forecast, and 8.39% and 8.36% respectively for excess return forecast. The most significant forecasts over the one-month horizon (panel B) come from the bm and the cross-sectional premium (csp) with the IS $R^2$ values of 0.69% and 0.61% respectively for gross return forecast, and 0.62% and 1.23% respectively for excess return forecast. However, none of these are as high as the $R^2$ values produced by disaggregate book-to-market ratios of 100 portfolios.

*** Insert Figure 2 about here ***

I also show the one-year fitted excess return series obtained using the PLS method based on book-to-market ratios of 100 portfolios, along with the forecasts based on typical aggregate valuation ratios, including the aggregate price-dividend ratio (dp) and bm in Figure 2. Inspecting the estimated line of expected returns in comparison to the realized returns, I can intuitively judge how good the fit is. Clearly, the fits of IS and OOS PLS estimates are consistent and very close to each other, particularly for the later years with an identical tendency. In addition, compared with the fits of other aggregate variables, the fits of the PLS estimates indicate a more qualitatively sound series for expected returns both IS and OOS, mainly due to the properties of cross-section valuation ratios. Other OOS results will be explained in detail in Section 5.3.

### 5.2 Bootstrap for In-Sample Forecasting

The results in Section 5.1 generally suggest that better IS predictability performance can be achieved through an PLS procedure based on disaggregated valuation ratios than through the commonly used aggregate predictor variables. To check whether the IS PLS procedure can mechanically accrue return predictability, I run two bootstraps under the null of no predictability, as described and explained in detail in Section 3.4. I consider the first-order autocorrelation of the lagged cross-sectional valuation ratios. Bootstrap A assumes the explained variable to be a constant, i.e. the average value of realized returns, and replaces the innovations to returns and the innovations to the regressors separately so that these artificial predictors have no true predictive power and the underlying relations between the innovations to the returns and to the predictors are destroyed. The simulated innovations are generated through a random number generator and are therefore independent of the real market data. If the artificial predictors still produce a positive $R^2$, this pseudo predictability will imply there is data mining inherent in a pure PLS procedure, coming from the higher weights allocated to the predictors across the portfolio that perform better in terms of forecasting power in the first stage of the PLS, further amplifying the bias in the second and third stages. Bootstrap B also assumes the constant explained variable, but collects the innovations to the returns and innovations to the regressors in a set so as to randomize simultaneously in each loop before putting them back into the PLS procedure, thus locking the underlying correlation structure of innovations in the returns and in the innovations from the predictor variables. Therefore, bootstrap B considers the small-sample bias as well as data mining.

*** Insert Table 3 about here ***

*** Insert Table 4 about here ***

I run both bootstraps for the one-year-ahead and one-month-ahead IS PLS procedure based on book-to-market ratios of 6, 25 and 100 portfolios respectively and record the $R^2$. I repeat the
process 10,000 times, and thus receive an empirical distribution of $R^2$ statistics estimated from the bootstrap samples. Table 3 and Table 4 present the critical values computed using the bootstraps under the null. The forecast target is the gross return in Table 3 and the excess return in Table 4. In both tables, panel A shows the result of bootstrap A and panel B shows the result of bootstrap B. The result in panel A of Table 3 is the replication of KP’s result, which forecasts the gross return with the bootstrap procedure ignoring the correlation structure. We can see that the IS PLS $R^2$ of annual forecast is high and the statistics is very good at the 1% significance level using book-to-market ratios of 100 portfolios as predictors (5% for monthly forecast). However, I expect the high IS PLS $R^2$ values with high significance are due to the overlook of small-sample bias and the use of gross returns as the predicted target. When taking both the data mining and the small-sample bias into account, my forecast results on excess returns in panel B of Table 4 demonstrate that the $R^2$ values for both the one-year and the one-month forecast are insignificant using book-to-market ratios of 100 portfolios as predictors. Though book-to-market ratios of 6 and 25 portfolios show significant $R^2$ values for annual forecasts, it is not as significant as KP’s at the 1% level.

5.3 Out-of-Sample Prediction of Market Returns

In this section, I study the predictability of aggregate market returns by using the OOS PLS method extracting the information from predictors of disaggregate book-to-market ratios. Table 5 shows the OOS PLS result by using the sample prior to 1980 as the training sample and the OOS forecast starts in December 1980 for annual forecasts and in January 1980 for monthly forecasts. The annual forecasts show high $R^2$ values, with 14.07% for the gross return and 8.03% for the excess return using book-to-market ratios of 100 portfolios. According to Cochrane (2005), Campbell and Thompson (2008), and Kelly and Pruitt (2013), an OOS $R^2$ of 0.90% indicates high economic significance for monthly forecasts. Panel A of Table 5, using gross returns as the forecast target, shows a monthly OOS $R^2$ very close to or even slightly over 0.90% based on predictors of book-to-market ratios of 6, 25 and 100 portfolios. However, panel B, in which the forecast target is excess returns, presents smaller monthly OOS $R^2$ values (e.g. 0.55% based on book-to-market ratios of 6 portfolios).

The OOS forecasts demonstrate significant economic magnitudes compared with commonly used predictors in earlier studies, the results of which are demonstrated in Table 6. I have shown that the most significant IS predictions over the one-year horizon come from the bm and the ntis in Table 2. However, their OOS $R^2$ values are, remarkably negative, showing no predictability. Only the default yield spread (dfy) and the term spread (tms) present positive $R^2$ values of annual OOS forecasts for both gross returns and excess returns, but the values are small. Turning to the one-month horizon, none of the alternatives give positive OOS $R^2$ values except the inflation rate (infl) ($R^2$ of 0.11% forecasting on excess returns). By contrast, using book-to-market ratios of 100 portfolios, the PLS approach gives the OOS $R^2$ value of 14.07% for the one-year forecast, and 0.64% for one-month forecast on the gross return. However, this $R^2$ value of one-month forecast turns to be negative predicting on the excess return (-0.13%) though it still shows positive values using book-to-market ratios of 6 or 25 portfolios.

5.4 Bootstrap for Out-of-Sample Forecasting

In Section 5.3, I have shown the recursive OOS $R^2$ for the sample period 1980-2010. In this section, I will provide relative critical values by running two bootstraps with 1,000 loops, and then I will present a whole picture of varying recursive OOS $R^2$ for the sample period from $\tau$ to 2010, where $\tau$ varies from 1945 to 1995. I define $\tau$ as the sample split date. For OOS, the data mining might not be there, but the small-sample bias still has the effect, and that is what I am analysing. Table 7 and Table 8 show the OOS prediction with statistics on the gross return and the excess return respectively. We can observe that the OOS PLS $R^2$ is significant at the prominent 1% level forecasting on the gross return when only the data mining is taken into account (panel A of Table 7). However, when predicting on the excess return and the bootstrap considers both the data mining and small-sample bias, the $R^2$ statistics are insignificant, except that book-to-market
ratios of 100 portfolios show 10% significant $R^2$ for one-year forecast, and book-to-market ratios of 6 portfolios show 10% significant $R^2$ for one-month forecast (panel B of Table 8). Therefore, we should overlook the remarkable predictability performance of the OOS PLS as it might be driven by the gross-return effect and the neglect of small-sample bias effect.

Figures (3, 4), (5, 6), and (7, 8) in pairs plot the varying OOS $R^2$ of the gross and excess return forecasts, each using book-to-market ratios of 6, 25 and 100 portfolios as predictors respectively. In each plot, subplot (a) follows bootstrap A, both annually and monthly, and subplot (b) follows bootstrap B. The sample split dates begin from December 1945 for the annual forecast and from January 1945 for the monthly forecast, and end in December 1995, based on the full sample of 1930-2010. Again, I record the 10%, 5% and 1% critical values, also called 90%, 95% and 99% confidence intervals, at each individual sample split point, which eventually creates the OOS null region clouds. I also plot the actual OOS $R^2$ line using the corresponding book-to-market ratios.

Generally, under the null hypothesis of no predictability, I expect a zero $R^2$ as the asymptotic value. If the pseudo samples broadly produce negative OOS $R^2$, or if the significant $R^2$ clouds are consistently underneath the actual OOS line, it will suggest that the OOS PLS outcomes are unlikely to be driven by a mechanical bias inherent in the PLS method, that is, even if there is no predicting power, the PLS can still generate high $R^2$ values. It is worth mentioning that, since my bootstraps use finite samples and the estimated forecast factors are generally non-zero, it is possible that the OOS PLS forecast will perform less well than a regular historical mean forecast for particular time periods.

From panel A of Figures 3, 5 and 7, in which the forecast target is the gross return and the bootstrap ignores the small-sample bias, we can see that, generally, the clouds widen with the expanding estimation window. This is because, when the training window grows, the OOS test sample becomes smaller. Observably, book-to-market ratios of 100 portfolios perform well for the yearly OOS PLS approach (panel A of Figure 7). To be specific, the actual OOS $R^2$ is always above zero. Although the 99% null region cloud is slightly above zero, the clouds are clearly far below the actual OOS line. Additionally, the monthly OOS forecast is also significant. The majority of the actual OOS line is positive and above the clouds. However, when I use bootstrap B considering the small-sample bias (panel B of Figure 7), the consistent ascending clouds, showing the 1% significance level, surpass the actual OOS line and remain above the actual line for most of the time, though still implying significance at the 5% or 10% level. Turning to panel B of Figures 4, 6 and 8, in which the forecast target is the excess return and the bootstrap considers the small-sample bias, it is clear that book-to-market ratios of 6 portfolios occasionally provide significant OOS $R^2$ values at the 10% level, while 25 portfolios show insignificant predictions, both annually and monthly. To judge which effect is bigger, gross-return effect or small-sample bias effect, I inspect the forecasting performance using book-to-market ratios of 100 portfolios. The gross-return effect shifts the actual OOS line from significant positive $R^2$ to the poor performance of negative $R^2$ for monthly prediction. Ignoring the small-sample bias, we can still observe the yearly excess return forecast at the 1% significance when the OOS forecast starts after 1970. The consideration of the small-sample bias only changes this significance level to 5%. Therefore, the gross-return effect is bigger than the small-sample bias effect. Overall, while Table 8 shows the result for one forecast
period of 1980-2010. Figure 4, 6 and 8 show the whole picture for varying OOS $R^2$, and thus we can see the predicting performance for different forecast periods. For example, the OOS $R^2$ for the sample period 1965-2010 are insignificant for predicting both one-year and one-month excess returns, except the case of book-to-market ratios of 6 portfolios which is significant at the 10% level.

From Figure 8(a), clear upward jumps occur in the period around the crises, in particular the 1970s energy crisis and the early 1980s global economic recession\textsuperscript{23}. The distinct sign of improvement in annual forecasting post-1970 may economically due to the fact that stock markets are involved in an evolutionary process of globalisation. In retrospect, lots of financial techniques and innovations emerged in the financial markets, deeply affecting the competition structure and risk characteristics of corporations, market regulation, and relative monetary policy. Many economic rules in fields such as freight transport and banking are dismantled in the late 1970s in the U.S. More institute investors are involved and new markets emerge. All these events indicate that markets are becoming more integrated and efficient from the 1970s.

The implication of predictive results is useful in a number of contexts. Firstly, we can use these models to test asset pricing models that utilise time-varying expected returns, e.g. ICAPM, and require a model for expected returns (see, for example, Campbell and Vuolteenaho, 2004). Secondly, predictive models can be used when modelling optimal portfolio choice. In particular, investors can formulate portfolio choice model when allocating among different asset classes based on the models incorporating time varying expected returns (DeMiguel et al., 2009\textsuperscript{24}). My results suggest that the ability to predict returns OOS is not statistically significant and would not improve portfolio choice.

6 Conclusion

The behaviour of market prices and the predictability of stock returns have long been of interest to finance researchers. While previous studies have documented abundant evidence on the relation between aggregate valuation ratios and the movements of stock returns, particularly with an IS fit, recent papers such as Goyal and Welch (2008) show very poor OOS forecasts. By applying the PLS approach, Kelly and Pruitt (2013) used disaggregated valuation ratios to forecast market returns and showed significant OOS predictability. This study reports remarkable OOS performance with a high level of significance. However, their study focused on gross returns forecasting, and their statistical test did not consider the small-sample bias in the OOS statistics. These two effects change radically the significance of OOS results. My IS results show some degree of predictability for one-year returns. However, my models are unable to distinguish between the possible reasons for time variation in expected returns. It would be interesting to shed light on whether the predictability stems from rational time variation in risk or from more behavioural explanations involving mispricing. However, my results do show that there is very little OOS predictability which an investor could exploit in real time suggesting that if there is mispricing in the market returns it is relatively small.

My main findings can be summarized as follows. First, using disaggregate book-to-market ratios as predictors, I provide prediction performance with high $R^2$ values both IS and OOS. Usually, when the number of book-to-market ratios of portfolios increases from 6 to 100, I get bigger $R^2$ values, indicating much better prediction performance achieved by the disaggregated predictors than by the conventional historical mean of the return target. Second, I look into the gross-return effect.

\textsuperscript{23}To further investigate forecasting performance during crisis periods versus “tranquil” periods, we can refer to Goyal and Welch’s (2008) study by outputting the cumulative SSE difference against year. To illustrate, the cumulative SSE difference denotes the difference between the cumulative squared prediction errors of the null and the cumulative squared prediction error of the alternative. The null is the average equity premium and the alternative is a predictive model based on predictor variables. An increase in the line suggests better performance of the alternative predictive model than the performance of the null and vice versa. By presenting such plots, Goyal and Welch (2008) found that many predictors show superior forecasting performance during the Oil Shock of 1973-1975. We can plot similar plots in future study and see whether the crisis jumps the $R^2$.

\textsuperscript{24}DeMiguel et al. (2009) find that $1/N$ portfolio strategy outperforms those more sophisticated models OOS which is similar to the finding of Goyal and Welch (2008) that most predictors fail in OOS tests.
For the IS fit, both annual and monthly returns predictability is statistically significant. Although the predictability of the excess return is a bit weaker than that of the gross return, the OOS $R^2$ of the yearly forecast is still as high as 17% when using 100 book-to-market ratio portfolios, and this performance is absolutely higher than the classic, commonly used alternative predictors. For the OOS forecast, the predictability of gross returns is still good, but that of excess returns is not, particularly for the monthly return forecast that produces only a -0.13% $R^2$ when using 100 book-to-market portfolios. Third, I shed light on the bootstrap effect regarding the statistical test, particularly given statistical clouds with rolling split dates for the OOS forecast. Two types of bootstraps under the null hypothesis of no predictability are performed, with bootstrap A ignoring the correlation structure as KP did, and bootstrap B maintaining the correlation structure, thus taking the small-sample bias into account. We indeed observe the predictability, though less amazing, that KP showed when using bootstrap A, with the OOS split line always far above the bootstrap clouds. However, this is when predicting the gross return with the risk-free rate. The predictability power fades when I test the excess returns and use bootstrap B. Only for particular times when the OOS starts in the later sample period is the OOS’ yearly predictability significant, but only at the 5% or 10% significance levels, and not the 1% level as in the former case. The monthly predictability, meanwhile, is insignificant most of the time. Therefore, I conclude that the remarkable OOS predictive power found by Kelly and Pruitt changes radically when both the gross-return effect and the small-sample bias effect are taken into account.

7 Appendix

7.1 Alternative Predictors

I employ a series of alternative predictors, i.e. the commonly used aggregated predictor variables, from Goyal and Welch (2008), as follows. Details refer to their paper. Except the variable $csp$ (1937/5-2002/12) for the limited data source, other alternative predictor variables are collected for the sample of 1930-2010, from Amit Goyal’s website.

**Default yield spread (dfy)** is the returns difference between the yields of BAA-rated corporate bond and AAA-rated corporate bond. The data of corporate bond yields are collected from database of Federal Reserve Economic Data (FRED).

**Inflation rate (infl)** is a continuous rising price level on the general goods and services over time in the economy. I use the consumer price index as the proxy of inflation rate collected from the Bureau of Labor Statistics.

**Stock variance (svar)** is defined as the sum of squared daily returns on the index of S&P 500, collected from CRSP.

**Cross-sectional beta premium (csp)** is initially proposed by Polk et al. (2006) capturing the relative valuations between high-beta and low-beta stocks. The variable construction has mainly three steps. Firstly, estimate the valuation ratios of all stocks and transform them into a yearly sequential corporation-level measure. Second, calculate the market beta for each individual stock as the measure of risk. Lastly, estimate the association of valuation grade and beta which measures the cross-sectional price of risk. The data is collected from Samuel Thompson.

**Dividend payout ratio (de)** is the log value of dividends ($D$) minus the log value of earnings ($E$) that $de = \log(D_t) - \log(E_t) = d_t - e_t$, where dividends are the rolling sum of twelve months’ dividends on the index of S&P 500 and earnings are the rolling sum of twelve months’ earnings on the index of S&P 500. I use the corresponding lower cases as the log value. Both data are from the S&P Corporation.

**Long term yield (lty)** is the long-term yield based on the government bond from Ibbotson’s *Stocks, Bonds, Bills and Inflation Yearbook*. 

22
Treasury bill rate (tbl) is viewed as the risk-free rate and I select the three-month Treasury bill from the Federal Reserve Economic Data.

Term spread (tms) is the long-term government bond yield minus the Treasury bill rate. I use the U.S. yield on long-term U.S. bonds as the long-term government bond yield, collected from the macro-history database of National Bureau of Economic Research (NBER).

Dividend price ratio (dp) is the log value of dividends minus the log value of prices ($P_t$), from the S&P Corporation, so that \( dp_t = \log(D_t) - \log(P_t) = d_t - p_t \).

Dividend yield (dy) is the log value of dividends ($D_t$) minus the log value of lagged prices ($P_t$) so that \( dp_t = \log(D_t) - \log(P_{t-1}) = d_t - p_{t-1} \).

Earnings price ratio (ep) is the log value of earnings minus the log value of prices so that \( ep = \log(E_t) - \log(P_t) = e_t - p_t \).

Book-to-market ratio (bm) is the ratio of book value divided by the market value with respect to the Dow Jones Industrial Average. The data are from Value Line’s website.

Net equity expansion (ntis) is the ratio of twelve-month sums of net of NYSE stocks divided by the relative total market cap at the year end. In particular, the amount of net equity issuing activity ($NetIssue_t$) is calculated as \( NetIssue_t = Mcap_t - Mcap_{t-1}(1 + \text{vwret}_x) \), where $Mcap$ represents the total market cap and $\text{vwret}_x$ represents the value-weighted return excluding dividends on the NYSE index. The data of $Mcap$ and $\text{vwret}$ are from the CRSP.
References


Figure 1: Book-to-Market Ratios of 6 Portfolios

Note: This figure shows log values of book-to-market ratios for 6 size- and value-sorted portfolios of the U.S. stocks from 1930-2010 (book value and market capitalization (number of firms × average firm size) are available on Kenneth French’s website). The National Bureau of Economic Research (NBER) recession dates are shaded in the figure. Lines with symbols represent the year-end book-to-market ratios.
Figure 2: One-Year Excess Return Predictions

Note: This figure shows the comparison between the actual U.S. market returns and a series of estimated market returns through forecasting models, which include in-sample and out-of-sample forecasts from the PLS method using book-to-market ratios of 100 size- and value-sorted portfolios, in-sample and out-of-sample forecasts from predictive regressions using the aggregate book-to-market ratios, and in-sample forecasts using the aggregate price-dividend ratio. We use excess return (EQP) from 1930-2010 as forecast target. For the out-of-sample forecast, the sample is split in 1945, which uses the pre-1945 period as a training window, and recursively forecasts returns beginning in December 1945.
Figure 3: OOS Prediction of Gross Returns by Book-to-Market Ratios of 6 Portfolios

Note: This figure shows the clouds of 90%, 95% and 99% null regions implying the confidence intervals, from 1,000 loops of bootstraps. We use gross returns (Rm) as the predicted target, and book-to-market ratios of 6 size- and value-sorted portfolios as predictors. The start date of out-of-sample forecasts, also called sample split date (τ), is rolling from 1945 to 1995. The value of book-to-market ratio in month t equals the latest observable annual book equity of the portfolio divided by its capitalization at the end of month t. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.
Figure 4: OOS Prediction of Excess Returns by Book-to-Market Ratios of 6 Portfolios

Note: This figure shows the clouds of 90%, 95% and 99% null regions implying the confidence intervals, from 1,000 loops of bootstraps. We use excess returns (EQP) as the predicted target, and book-to-market ratios of 6 size- and value-sorted portfolios as predictors. The start date of out-of-sample forecasts, also called sample split date ($\tau$), is rolling from 1945 to 1995. The value of book-to-market ratio in month $t$ equals the latest observable annual book equity of the portfolio divided by its capitalization at the end of month $t$. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.
Figure 5: OOS Prediction of Gross Returns by Book-to-Market Ratios of 25 Portfolios

Note: This figure shows the clouds of 90%, 95% and 99% null regions implying the confidence intervals, from 1,000 loops of bootstraps. We use gross returns (Rm) as the predicted target, and book-to-market ratios of 25 size- and value-sorted portfolios as predictors. The start date of out-of-sample forecasts, also called sample split date (τ), is rolling from 1945 to 1995. The value of book-to-market ratio in month t equals the latest observable annual book equity of the portfolio divided by its capitalization at the end of month t. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.
Figure 6: OOS Prediction of Excess Returns by Book-to-Market Ratios of 25 Portfolios

Note: This figure shows the clouds of 90%, 95% and 99% null regions implying the confidence intervals, from 1,000 loops of bootstraps. We use excess returns (EQP) as the predicted target, and book-to-market ratios of 25 size- and value-sorted portfolios as predictors. The start date of out-of-sample forecasts, also called sample split date ($\tau$), is rolling from 1945 to 1995. The value of book-to-market ratio in month $t$ equals the latest observable annual book equity of the portfolio divided by its capitalization at the end of month $t$. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.
Figure 7: OOS Prediction of Gross Returns by Book-to-Market Ratios of 100 Portfolios

Note: This figure shows the clouds of 90%, 95% and 99% null regions implying the confidence intervals, from 1,000 loops of bootstraps. We use gross returns ($R_m$) as the predicted target, and book-to-market ratios of 100 size- and value-sorted portfolios as predictors. The start date of out-of-sample forecasts, also called sample split date ($\tau$), is rolling from 1945 to 1995. The value of book-to-market ratio in month $t$ equals the latest observable annual book equity of the portfolio divided by its capitalization at the end of month $t$. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.
Figure 8: OOS Prediction of Excess Returns by Book-to-Market Ratios of 100 Portfolios

Note: This figure shows the clouds of 90%, 95% and 99% null regions implying the confidence intervals, from 1,000 loops of bootstraps. We use excess returns (EQP) as the predicted target, and book-to-market ratios of 100 size- and value-sorted portfolios as predictors. The start date of out-of-sample forecasts, also called sample split date ($\tau$), is rolling from 1945 to 1995. The value of book-to-market ratio in month $t$ equals the latest observable annual book equity of the portfolio divided by its capitalization at the end of month $t$. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.
**Tables**

Table 1: Market Return Predictions (1930-2010): IS

<table>
<thead>
<tr>
<th>Panel A: Gross Return (Rm)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>In-Sample $R^2$ (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annual</td>
<td>Monthly</td>
</tr>
<tr>
<td>BMs of 6 portfolios</td>
<td>8.94 (7.72)</td>
<td>0.54 (0.60)</td>
</tr>
<tr>
<td>BMs of 25 portfolios</td>
<td>14.92 (13.50)</td>
<td>0.89 (1.12)</td>
</tr>
<tr>
<td>BMs of 100 portfolios</td>
<td>19.89 (18.05)</td>
<td>2.01 (2.38)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Excess Return (EQP)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>In-Sample $R^2$ (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annual</td>
<td>Monthly</td>
</tr>
<tr>
<td>BMs of 6 portfolios</td>
<td>7.01</td>
<td>0.45</td>
</tr>
<tr>
<td>BMs of 25 portfolios</td>
<td>15.02</td>
<td>0.82</td>
</tr>
<tr>
<td>BMs of 100 portfolios</td>
<td>17.12</td>
<td>1.95</td>
</tr>
</tbody>
</table>

*Note:* Outcomes of annual and monthly predictions of aggregate market returns by using in-sample partial least squares (PLS). We use 6, 25 or 100 portfolios, generated from the Fama-French data library, as predictors to forecast returns. We use natural logarithm values. The $R^2$ is shown as a percentage. The figures in brackets are Kelly and Pruitt’s (2013), for comparison.
Table 2: Market Return Predictions with Common Alternative Predictors: IS

<table>
<thead>
<tr>
<th></th>
<th>Panel A: One Year</th>
<th>Panel B: One Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rm $R^2$ (%)</td>
<td>EQP $R^2$ (%)</td>
<td>Rm $R^2$ (%)</td>
</tr>
<tr>
<td>dfy 0.49</td>
<td>0.81</td>
<td>0.05</td>
</tr>
<tr>
<td>infl 0.00</td>
<td>0.17</td>
<td>0.05</td>
</tr>
<tr>
<td>svar 0.01</td>
<td>0.15</td>
<td>0.09</td>
</tr>
<tr>
<td>csp 0.36</td>
<td>0.23</td>
<td>0.61</td>
</tr>
<tr>
<td>de 0.13</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>lty 0.73</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>tms 1.40</td>
<td>2.86</td>
<td>0.11</td>
</tr>
<tr>
<td>tbl 0.09</td>
<td>1.16</td>
<td>0.00</td>
</tr>
<tr>
<td>dp 3.36</td>
<td>4.19</td>
<td>0.18</td>
</tr>
<tr>
<td>dy 3.61</td>
<td>4.45</td>
<td>0.30</td>
</tr>
<tr>
<td>ep 4.83</td>
<td>3.47</td>
<td>0.48</td>
</tr>
<tr>
<td>bm 9.24</td>
<td>8.39</td>
<td>0.69</td>
</tr>
<tr>
<td>ntis 9.18</td>
<td>8.36</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Note: The table reports in-sample percentage $R^2$ values for overlapping yearly forecasts and non-overlapping monthly forecasts from 1930-2010. We use gross returns (Rm) and excess returns (EQP) as the forecast target respectively. We use alternative predictors from Goyal and Welch (2008) updated to 2010/12, including the default yield spread (dfy), the inflation rate (infl), stock variance (svar), the cross-sectional premium (csp), the dividend payout ratio (de), the long term yield (lty), the term spread (tms), the Treasury bill rate (tbl), the dividend price ratio (dp), the dividend yield (dy), the earnings price ratio (ep), the book-to-market ratio (bm), and the net equity expansion (ntis). Alternative predictors are available for the full sample of 1930-2010 (except csp: 1937/5-2002/12).
### Table 3: IS Prediction Regression: Gross Market Return (1930-2010)

#### Panel A: Bootstrap A

<table>
<thead>
<tr>
<th></th>
<th>One-Year $R^2$ (%)</th>
<th>One-Month $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 bm</td>
<td>25 bm</td>
</tr>
<tr>
<td>PLS</td>
<td>8.94***</td>
<td>14.92**</td>
</tr>
<tr>
<td>Bootstrap under null</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% critical value</td>
<td>9.19</td>
<td>15.27</td>
</tr>
<tr>
<td>5% critical value</td>
<td>6.84</td>
<td>11.47</td>
</tr>
<tr>
<td>10% critical value</td>
<td>5.74</td>
<td>9.85</td>
</tr>
</tbody>
</table>

#### Panel B: Bootstrap B

<table>
<thead>
<tr>
<th></th>
<th>One-Year $R^2$ (%)</th>
<th>One-Month $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 bm</td>
<td>25 bm</td>
</tr>
<tr>
<td>PLS</td>
<td>8.94**</td>
<td>14.92**</td>
</tr>
<tr>
<td>Bootstrap under null</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% critical value</td>
<td>10.93</td>
<td>17.16</td>
</tr>
<tr>
<td>5% critical value</td>
<td>7.94</td>
<td>13.26</td>
</tr>
<tr>
<td>10% critical value</td>
<td>6.43</td>
<td>11.65</td>
</tr>
</tbody>
</table>

**Note:** Critical values of $R^2$ statistics are obtained via bootstraps of 10,000 loops under the null. They are presented for in-sample partial least squares (PLS) predictions of gross market returns ($R_m$). There are three levels of critical values: 1%, 5%, and 10%, and the superscripts ***, **, and * denote the corresponding 99%, 95%, and 90% confidence levels. We use book-to-market ratios of 6/25/100 size- and value-sorted portfolios, generated from the Fama-French data library, as predictors to forecast returns. We use natural logarithm values. The $R^2$ is shown as a percentage. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.
Table 4: IS Prediction Regression: Excess Market Return (1930-2010)

Panel A: Bootstrap A

<table>
<thead>
<tr>
<th>Variable</th>
<th>One-Year $R^2$ (%)</th>
<th>One-Month $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 bm</td>
<td>25 bm</td>
</tr>
<tr>
<td>PLS</td>
<td>7.01**</td>
<td>15.02**</td>
</tr>
<tr>
<td>Bootstrap under null</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% critical value</td>
<td>8.60</td>
<td>16.89</td>
</tr>
<tr>
<td>5% critical value</td>
<td>6.40</td>
<td>11.88</td>
</tr>
<tr>
<td>10% critical value</td>
<td>5.37</td>
<td>10.02</td>
</tr>
</tbody>
</table>

Panel B: Bootstrap B

<table>
<thead>
<tr>
<th>Variable</th>
<th>One-Year $R^2$ (%)</th>
<th>One-Month $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 bm</td>
<td>25 bm</td>
</tr>
<tr>
<td>PLS</td>
<td>7.01*</td>
<td>15.02**</td>
</tr>
<tr>
<td>Bootstrap under null</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% critical value</td>
<td>11.23</td>
<td>17.47</td>
</tr>
<tr>
<td>5% critical value</td>
<td>7.94</td>
<td>13.41</td>
</tr>
<tr>
<td>10% critical value</td>
<td>6.56</td>
<td>11.75</td>
</tr>
</tbody>
</table>

Note: Critical values of $R^2$ statistics are obtained via bootstraps of 10,000 loops under the null. They are presented for in-sample partial least squares (PLS) predictions of excess market returns (EQP). There are three levels of critical values: 1%, 5%, and 10%, and the superscripts ***, **, and * denote the corresponding 99%, 95%, and 90% confidence levels. We use book-to-market ratios of 6/25/100 size- and value-sorted portfolios, generated from the Fama-French data library, as predictors to forecast returns. We use natural logarithm values. The $R^2$ is shown as a percentage. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.

Table 5: Market Return Predictions (1930-2010): OOS

Panel A: Gross Return (Rm)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Out-of-Sample $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Annual</td>
</tr>
<tr>
<td>BMs of 6 portfolios</td>
<td>9.22 (5.81)</td>
</tr>
<tr>
<td>BMs of 25 portfolios</td>
<td>7.62 (3.49)</td>
</tr>
<tr>
<td>BMs of 100 portfolios</td>
<td>14.07 (13.07)</td>
</tr>
</tbody>
</table>

Panel B: Excess Return (EQP)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Out-of-Sample $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Annual</td>
</tr>
<tr>
<td>BMs of 6 portfolios</td>
<td>2.97</td>
</tr>
<tr>
<td>BMs of 25 portfolios</td>
<td>1.51</td>
</tr>
<tr>
<td>BMs of 100 portfolios</td>
<td>8.03</td>
</tr>
</tbody>
</table>

Note: Outcomes of one-year-ahead and one-month-ahead predictions of aggregate market returns by using out-of-sample partial least squares (PLS). We use 6, 25 or 100 portfolios, generated from the Fama-French data library, as predictors to forecast returns. We use natural logarithm values. The $R^2$ is shown as a percentage. To implement the out-of-sample procedure, the sample is split in 1980, with the pre-1980 period as a training window, and recursively predicts returns starting in December 1980 for the overlapping annual forecast and in January 1980 for the monthly forecast. The figures in brackets are Kelly and Pruitt’s (2013), for comparison.
Table 6: Market Return Predictions with Common Alternative Predictors: OOS

<table>
<thead>
<tr>
<th></th>
<th>Panel A: One Year</th>
<th></th>
<th>Panel B: One Month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rm $R^2$ (%)</td>
<td>EQP $R^2$ (%)</td>
<td>Rm $R^2$ (%)</td>
</tr>
<tr>
<td>dfy</td>
<td>0.76</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>infl</td>
<td>-0.33</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>svar</td>
<td>-0.36</td>
<td>-0.18</td>
<td></td>
</tr>
<tr>
<td>csp</td>
<td>1.72</td>
<td>-2.63</td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>-9.16</td>
<td>-4.41</td>
<td></td>
</tr>
<tr>
<td>lty</td>
<td>-0.68</td>
<td>-7.95</td>
<td></td>
</tr>
<tr>
<td>tms</td>
<td>1.28</td>
<td>2.70</td>
<td></td>
</tr>
<tr>
<td>tbl</td>
<td>-3.45</td>
<td>-6.33</td>
<td></td>
</tr>
<tr>
<td>dp</td>
<td>-6.49</td>
<td>-10.24</td>
<td></td>
</tr>
<tr>
<td>dy</td>
<td>-7.13</td>
<td>-11.33</td>
<td></td>
</tr>
<tr>
<td>ep</td>
<td>-15.88</td>
<td>-20.50</td>
<td></td>
</tr>
<tr>
<td>bm</td>
<td>-30.06</td>
<td>-32.50</td>
<td></td>
</tr>
<tr>
<td>ntit</td>
<td>-51.32</td>
<td>-57.43</td>
<td></td>
</tr>
</tbody>
</table>

Note: The table reports out-of-sample percentage $R^2$ values for overlapping yearly forecasts and non-overlapping monthly forecasts from 1930-2010. We use gross returns (Rm) and excess returns (EQP) as the forecast target respectively. We use alternative predictors from Goyal and Welch (2008) updated to 2010/12, including the default yield spread (dfy), the inflation rate (infl), stock variance (svar), the cross-sectional premium (csp), the dividend payout ratio (de), the long term yield (lty), the term spread (tms), the Treasury bill rate (tbl), the dividend price ratio (dp), the dividend yield (dy), the earnings price ratio (ep), the book-to-market ratio (bm), and the net equity expansion (ntit). Out-of-sample forecasts begin in January 1980 for a one-month horizon and December 1980 for a one-year horizon. Alternative predictors are available for the full sample of 1930-2010 (except csp: 1937/3-2002/12).
Table 7: OOS Prediction Regression: Gross Market Return (1930-2010)

### Panel A: Bootstrap A

<table>
<thead>
<tr>
<th></th>
<th>One-Year $R^2$ (%)</th>
<th>One-Month $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 bm</td>
<td>25 bm</td>
</tr>
<tr>
<td>PLS</td>
<td>9.22***</td>
<td>7.62***</td>
</tr>
</tbody>
</table>

**Bootstrap under null**

<table>
<thead>
<tr>
<th>1% critical value</th>
<th>5% critical value</th>
<th>10% critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.79</td>
<td>6.82</td>
<td>4.15</td>
</tr>
<tr>
<td>3.94</td>
<td>2.96</td>
<td>0.83</td>
</tr>
<tr>
<td>1.95</td>
<td>1.15</td>
<td>-0.38</td>
</tr>
</tbody>
</table>

### Panel B: Bootstrap B

<table>
<thead>
<tr>
<th></th>
<th>One-Year $R^2$ (%)</th>
<th>One-Month $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 bm</td>
<td>25 bm</td>
</tr>
<tr>
<td>PLS</td>
<td>9.22**</td>
<td>7.62**</td>
</tr>
</tbody>
</table>

**Bootstrap under null**

<table>
<thead>
<tr>
<th>1% critical value</th>
<th>5% critical value</th>
<th>10% critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.52</td>
<td>12.15</td>
<td>13.94</td>
</tr>
<tr>
<td>5.62</td>
<td>6.04</td>
<td>6.48</td>
</tr>
<tr>
<td>3.36</td>
<td>3.10</td>
<td>2.89</td>
</tr>
</tbody>
</table>

**Note:** Critical values of $R^2$ statistics are obtained via bootstraps of 1,000 loops under the null. They are presented for one-year-ahead and one-month-ahead predictions of gross market returns ($R_m$) by using out-of-sample partial least squares (PLS). To implement the out-of-sample procedure, the sample is split in 1980, with the pre-1980 period as a training window, and recursively predicts returns starting in December 1980 for the overlapping annual forecast and in January 1980 for the monthly forecast. There are three levels of critical values: 1%, 5%, and 10%, and the superscripts ***, **, and * denote the corresponding 99%, 95%, and 90% confidence levels. We use book-to-market ratios of 6/25/100 size- and value-sorted portfolios, generated from the Fama-French data library, as predictors to forecast returns. We use natural logarithm values. The $R^2$ is shown as a percentage. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.
### Table 8: OOS Prediction Regression: Excess Market Return (1930-2010)

#### Panel A: Bootstrap A

<table>
<thead>
<tr>
<th></th>
<th>One-Year $R^2$ (%)</th>
<th>One-Month $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 bm</td>
<td>25 bm</td>
</tr>
<tr>
<td>PLS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.97***</td>
<td>1.51**</td>
</tr>
<tr>
<td>Bootstrap under null</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% critical value</td>
<td>4.14</td>
<td>4.02</td>
</tr>
<tr>
<td>5% critical value</td>
<td>0.95</td>
<td>-0.04</td>
</tr>
<tr>
<td>10% critical value</td>
<td>-0.51</td>
<td>-1.28</td>
</tr>
</tbody>
</table>

#### Panel B: Bootstrap B

<table>
<thead>
<tr>
<th></th>
<th>One-Year $R^2$ (%)</th>
<th>One-Month $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 bm</td>
<td>25 bm</td>
</tr>
<tr>
<td>PLS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.97</td>
<td>1.51</td>
</tr>
<tr>
<td>Bootstrap under null</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% critical value</td>
<td>12.87</td>
<td>14.63</td>
</tr>
<tr>
<td>5% critical value</td>
<td>6.27</td>
<td>6.29</td>
</tr>
<tr>
<td>10% critical value</td>
<td>3.66</td>
<td>3.55</td>
</tr>
</tbody>
</table>

*Note:* Critical values of $R^2$ statistics are obtained via bootstraps of 1,000 loops under the null. They are presented for one-year-ahead and one-month-ahead predictions of excess market returns (EQP) by using out-of-sample partial least squares (PLS). To implement the out-of-sample procedure, the sample is split in 1980, with the pre-1980 period as a training window, and recursively predicts returns starting in December 1980 for the overlapping annual forecast and in January 1980 for the monthly forecast. There are three levels of critical values: 1%, 5%, and 10%, and the superscripts ***, **, and * denote the corresponding 99%, 95%, and 90% confidence levels. We use book-to-market ratios of 6/25/100 size- and value-sorted portfolios, generated from the Fama-French data library, as predictors to forecast returns. We use natural logarithm values. The $R^2$ is shown as a percentage. Under the null hypothesis, bootstrap A ignores the correlation structure of innovations, and bootstrap B maintains the correlation structure of innovations of returns and book-to-market ratios.