Forecasting the size premium over different time horizons

Valeriy Zakamulin*

University of Agder, Faculty of Economics and Social Sciences, Service Box 422, 4604 Kristiansand, Norway

Abstract

In this paper, we provide evidence that the small stock premium is predictable both in-sample and out-of-sample through the use of a set of lagged macroeconomic variables. We find that it is possible to forecast the size premium over time horizons that range from one month to one year. We demonstrate that the predictability of the size premium allows a portfolio manager to generate an economically and statistically significant active alpha.

JEL classification: C13; G12; G17.

Keywords: Size effect; Size premium; Stock return predictability; Active alpha

*a.k.a. Valeri Zakamouline. Tel.: +47 38 14 10 39; fax: +47 38 14 10 27.
Email address: `Valeri.Zakamouline@uia.no` (Valeriy Zakamulin)

Preprint submitted to Journal of Banking and Finance *November 5, 2012*

1. Introduction

Since the publication of the original studies by Banz (1981) and Reinganum (1981) addressing relations between firm size and risk-adjusted stock returns, numerous papers have documented that higher risk-adjusted stock returns are earned by small firms than by large firms.[1] The source of the small stock premium has been the subject of considerable debate in financial economics.[2] Several competing explanations for this size premium have been proposed. Most of the investigations regarding this topic contend that the size premium represents the investors' compensation for bearing greater systematic risk (see, among others, Chan, Chen, and Hsieh (1985), Fama and French (1993), and Vassalou and Xing (2004)). The non-risk-based explanations are as follows. Stoll and Whaley (1983) and Pastor and Stambaugh (2003) observe that relative to stocks of larger firms, small stocks demonstrate lower liquidity and therefore involve higher transaction costs. Thus, the size premium may represent the investor's compensation for the lower liquidity of small stocks. Kothari, Shanken, and Sloan (1995) and Shumway and Warther (1999) conjecture that survivorship bias (which can also be characterized as delisting bias) is the main source of the size premium. Hou and Moskowitz (2005) argue that market frictions are the source of the size premium. In particular, these authors examine the average delay with which a firm's stock price responds to information and demonstrate that this price delay captures a substantial portion of the size premium. A possible behavioral explanation for the size effect, based on disagreement and differences in investors' tastes, is presented by Fama and French (2007).

In contrast to the aforementioned rational explanations for the source of the small stock premium, Lo and MacKinlay (1990a) and Black (1993) argue that many asset pricing anomalies, including the size effect, can result from data mining. With respect to the size effect in particular, evidence exists to support the data mining argument. Specifically, many studies report that the size premium has not been robust; instead,

[1]This phenomenon is often called the "size effect".

[2]For an excellent review of the size effect in equity returns, the interested reader can consult Van Dijk (2011).

this effect has appeared only during certain specific historical periods. For example, Banz (1981) reveals that the size effect in the U.S. varies considerably over the 1926-1975 time period. Handa, Kothari, and Wasley (1989) report that the size premium was negative (although statistically insignificant) over the 1941-1954 time period. Dimson and Marsh (1999) report the presence of the size premium over the 1955-1983 time period but the absence of this phenomenon over the 1983-1997 time period. Malkiel (2003) writes that there was no small stock premium from the mid-1980s until the end of the 1990s. Moreover, Malkiel raises the concern that there is a general lack of predictable patterns in the size and value premiums across various time periods and hypothesizes that the 1963-1991 timespan examined by Fama and French (1993) was a unique period for these premiums.

Chan, Chen, and Hsieh (1985) are most likely the first researchers to suggest that the size premium varies over time. Moreover, the time variation in the size premium is related to economic conditions. In particular, these authors demonstrate that the default spread, which is computed as the difference between the yields on Baa- and Aaa-rated corporate bonds, is a significant explanatory variable for the size premium. Similarly, Vassalou and Xing (2004) establish that the size premium is closely related to the default risk, which varies with economic conditions. Perez-Quiros and Timmermann (2000) use a Markov-switching framework with two latent states and reveal that the size premium displays strong counter-cyclical variation with the latent factors. These authors conjecture that these two latent states represent periods of economic recession and expansion. In their interpretation, the size premium is high during economic recessions because small firms are more at risk during these times.

In fact, evidence of the in-sample predictability of the small stock premium has been provided. In particular, Ferson and Harvey (1999) and Cooper, Gulen, and Vassalou (2001) use a set of lagged economy-wide predictive variables to estimate a predictive regression for the small-minus-big (SMB) Fama-French factor. The predictive regression was estimated using monthly data only. The findings in these papers suggest that certain lagged macroeconomic variables can predict the SMB factor return. However, despite the existence of in-sample predictability for the small stock premium, it is well known that in-sample predictability might reflect data mining and may not hold true for out-of-sample results (see, for example, Bossaerts and Hillion (1999) and Goyal and Welch (2008)). Therefore, out-of-sample assessments are needed to counter the data mining argument.

The main contribution of this paper is to provide evidence that the small stock premium is not a result of data mining. In particular, we demonstrate that the small stock premium is predictable both in-sample and out-of-sample through the use of a set of lagged macroeconomic variables. In addition, we reveal that the predictability of the size premium allows a portfolio manager to generate an economically and statistically significant active alpha. More specifically, this paper extends prior work on the predictability of the size premium in a number of ways. First, we use a much longer time span than the periods that are examined by either Ferson and Harvey (1999) or Cooper, Gulen, and Vassalou (2001). Second, we perform predictability tests using not only monthly data but also quarterly, semi-annual, and annual data. Third, we conduct a model selection procedure to restrict the number of predictive variables and also execute robustness assessments. Finally, we estimate not only in-sample predictive regressions but also assess the performance of out-of-sample forecasts. Our findings reveal that there is in-sample predictability of the size premium at each examined sampling frequency. Moreover, we discover that there is out-of-sample predictability of the small stock premium over time horizons of one month, one quarter, and one year. To support our findings regarding the existence of out-of-sample predictability, we provide a simulation experiment. In this experiment, we demonstrate that a superior performance is generated by an active strategy that invests either in small stocks or large stocks depending on the forecast of the sign of the size premium. In particular, an active strategy is able to generate a positive alpha that is highly economically and statistically significant, even in the presence of realistic transaction costs.

The remainder of the paper is organized as follows. Section 2 presents a set of potential predictors of the small stock premium. Section 3 presents the data and summary statistics. In Section 4, we study the in-sample predictability of the size premium and discuss the implications of our findings. In Section 5, we assess the performance of out-of-sample forecasts. In Section 6, we demonstrate that the out-of-sample predictability can be exploited to generate an active alpha. Section 7 summarizes and concludes the paper.

2. Potential predictors of the small stock premium

Our main objective in this paper is to investigate whether it is possible to forecast the return on the SMB (small-minus-big) Fama-French factor through the use of a set of potential predictive variables. Because there is no theory that explains the dynamics of the small stock premium, the choice of predictive variables must be somewhat arbitrary. The set of variables that is used in our study is presented in Table 1; these variables are related to the state of both the economy and financial markets. These variables are relatively standard choices that are found in the extant literature regarding the predictability of stock returns. Below, we briefly explain the main reasons for the inclusion of each of these variables.

Variable	**Description**
MKT	The stock market return
YLD	The dividend yield
HML	The high-minus-low Fama-French factor
MOM	The momentum factor of Jegadeesh and Titman (1993)
DEF	The default spread
TBL	The T-bill rate
TRM	The term premium
INF	The inflation rate

Table 1: Potential predictors of the SMB factor return.

Our core hypothesis is that stock market returns in general and the return differentials between small and large cap stocks (i.e., the SMB factor returns) in particular are dependent on economic conditions. Changes in the variables that act as proxies for the state of the economy and for financial markets could be expected to change investors' perceptions of future cash flows and could therefore affect stock returns.

We include the stock market return (MKT) in our set of predictive variables for two reasons. First, the stock market return is commonly used as a proxy of the overall state of the economy. There is widespread agreement that stock market returns contain important information about the economic activity that will occur in the near future. For instance, the Composite Index of Leading Indicators that is published monthly by the Conference Board[3] includes the S&P 500 stock index as one of the leading indicators of economic growth. The logic for this inclusion is that in a rational expectation model, the price of a share of a stock represents the discounted value of future expected cash flows. The strength of the economy determines the magnitude of these cash flows. Thus, the price of an equity must reflect expectations of real future activities, and changes to the value of the equity in question should therefore mirror revisions in these expectations. In other words, changes in the value of an equity should predict the direction of economic changes in upcoming months. Second, Hou and Moskowitz (2005) document that stock market returns lead small stock returns.[4] Thus, the stock market return appears to be a natural predictor of the size premium.

The common explanation for the inclusion of the dividend yield (YLD) in the set of predictive variables is that this factor serves as a proxy for time variation in unobservable risk premiums. In particular, the prevalent hypothesis is that stock prices are low relative to dividends when discount rates and expected returns are high (and vice versa); thus, the dividend yield is predicted to vary with expected returns. The ability of the dividend yield to predict future stock and bond returns is documented by Fama and French (1988), Campbell and Shiller (1988), Fama and French (1989), and Lewellen (2004), among others.

The inclusion of the default spread (DEF), the term premium (TRM), the T-bill rate (TBL), and the inflation rate (INF) as predictors in this study is also motivated

[3]See `http://www.conference-board.org/`, which is a global and independent business membership and research organization that publishes information and analysis, produces economics-based forecasts, and assesses trends in the economy.

[4]In particular, the returns on large-cap stocks lead the returns on small-cap stocks, as discussed in the subsequent section.

by two different reasons. The first of these reasons is that all of these variables are strongly related to the overall state of the economy and the financial markets. The ability of the default spread to predict future stock and bond returns is documented by Keim and Stambaugh (1986), Fama and French (1989), Fama (1990), and many others. Fama and French (1989) and Fama (1990) are among the first investigators that utilize TRM in stock return regressions. Similarly, Fama and Schwert (1977) and Campbell (1987) are among the first researchers to include TBL as a regressor in stock return equations. Fama (1981), Geske and Roll (1983), and Lee (1992) report that stock returns are negatively correlated with inflation. The second reason for the inclusion of these variables derives from the imperfect capital market theories of Bernanke and Gertler (1989) and Gertler and Gilchrist (1994).

Specifically, these imperfect capital market theories predict that the risks of small and large firms will vary in different ways over the course of the economic cycle; in particular, these theories suggest that compared with large firms, small firms with little collateral should be more strongly affected by changing monetary policy during the recession and expansion states of the business cycle. The variables TRM, TBL, and INF are explicitly related to monetary policy, and the T-bill rate is a short-term rate that is entirely determined by monetary policy. During economic expansions, monetary policy is contractionary, decreasing the total money supply by raising short-term rates. By contrast, an expansionary monetary policy that increases the total money supply by lowering short-term rates is traditionally used to combat economic recessions. Monetary policy affects the inflation rate. The term premium is the difference between the yields on long-term bonds and the short term rate. Rosenberg and Maurer (2008) demonstrate that the term premium varies with the economic cycle. During economic expansions, contractionary monetary policies typically cause the term premium to decrease, and the endpoints of these economic expansions are typically marked by an inverted yield curve (that is, the term premium becomes negative). By contrast, under undesirable economic conditions, the term premium increases; this increase is especially rapid during recessions. Estrella and Hardouvelis (1991) demonstrate that the term premium is a good predictor of economic activity in the near future. Moreover, this factor is able to predict upcoming recessions.

The default spread is closely related to the default risk. A common hypothesis is that compared with large firms, small firms with little collateral are likely to be more exposed to bankruptcy risk during recessions; therefore, one would expect that the small stock premium would vary with the default spread. In fact, this relationship is well documented in the literature (see Chan, Chen, and Hsieh (1985), Vassalou and Xing (2004), and Perez-Quiros and Timmermann (2000)). Because Vassalou and Xing (2004) and Avramov, Jostova, and Philipov (2007) demonstrate that the returns on the HML factor contain default-related information, we also include this variable in our set of potential predictors.

Finally, we include the momentum factor (MOM) in this study for the following two reasons: (a) Chordia and Shivakumar (2002) present evidence that the returns on this factor are related to macroeconomic conditions; (b) Yao (2012) documents that MOM is inversely related to the January size effect.

3. Data

In our study, we use four different data sampling frequencies: monthly, quarterly, semi-annually, and annually. Our sample period begins in January 1927 and ends in December 2010, including a total of 1008 monthly observations (as well as 252 quarterly, 168 semi-annual, and 84 annual observations). The data for our study are obtained from several sources. The original data are obtained at monthly frequencies.

Stock market return : The returns on the stock market, MKT, are obtained from the data library of Kenneth French.[5] In particular, to obtain the stock market return, we use the return on the value-weighted portfolio consisting of the top quintile (20%) of all of the firms in the aggregate stock market after these firms have been sorted by their market capitalization.[6] Note that our stock market return therefore represents the return on the stocks with the largest market capitalizations. The number of stocks in this portfolio varies from 100 to 500. Thus, our return on the

[5]See `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[6]In this study, the aggregate stock market consists of all NYSE, AMEX, and NASDAQ stocks with market equity data. The size-sorted quintile portfolios are constructed at the end of each June, using the June market capitalization and the NYSE breakpoints.

stock market roughly corresponds to the return on the Standard and Poor's Composite Stock Price Index. This specific choice for the stock market return can be justified on two grounds. First, because we use only large-cap stocks to calculate the stock market return, this return is not influenced either by the size effect or by the survivorship bias. Second, it has been documented in the literature (see, for example, Lo and MacKinlay (1990b) and McQueen, Pinegar, and Thorley (1996)) that large-cap stocks lead small-cap stocks over time horizons ranging from one week to one month. Therefore, compared with the return on the aggregate stock market, the return on large-cap stocks may represent a better predictor of the return on the SMB factor. Quarterly, semi-annual, and annual returns on the stock market are obtained by compounding the calculated monthly returns.

Dividend yield : The monthly dividend yield is obtained from the data library of Kenneth French and is calculated as the difference between the total return and the capital appreciation return on MKT. Because of the pronounced seasonality in the monthly dividend yield, a moving sum is used to smooth monthly dividend yields over the preceding year. Quarterly, semi-annual, and annual dividend yields are obtained by compounding the calculated monthly dividend yields.

SMB and HML factor returns : The SMB and HML factor returns are constructed using the 6 Fama-French value-weighted portfolios that are formed on the basis of firm size and book-to-market ratio. In particular, for each sampling frequency, SMB is defined as the average return on the three small-stock portfolios minus the average return on the three big-stock portfolios. HML, in turn, is defined as the average return on the two value portfolios minus the average return on the two growth portfolios.[7]

MOM factor return : To construct the MOM factor returns, we use the 6 Fama-French value-weighted portfolios that are formed on the basis of firm size and prior (2-12 month) returns. In particular, for each sampling frequency, MOM is defined as the average return on the two portfolios with high prior returns minus the

[7] For a complete description of the construction of the SMB and HML factor returns, see Fama and French (1993).

average return on the two portfolios with low prior returns.[8]

Default spread : The default spread, DEF, which is also called the "credit spread", is defined as the difference between the yields of Moody's Baa- and Aaa-rated corporate bonds. The monthly default spread data are obtained from the Federal Reserve Economic Data database, which is provided by the Federal Reserve Bank of St. Louis.[9] Quarterly, semi-annual, and annual default spreads are obtained by averaging the monthly default spreads.

TBill returns : The monthly TBL values are obtained from the Ibbotson SBBI 2010 Classic Yearbook. Quarterly, semi-annual, and annual T-bill returns are obtained by compounding these monthly returns.

Term premium : The term premium, TRM, is computed as the difference between the yields on long-term (30 year) government bonds and the T-bill returns. Monthly data for the annual yields on long-term government bonds are obtained from the Federal Reserve Bank of St. Louis. Monthly, quarterly, and semi-annual yields on long-term government bonds are calculated from annual yields as the geometric average return for each examined frequency.

Inflation rate : The inflation rate, INF, is computed using the Consumer Price Index (CPI) for the U.S., as calculated by Robert Shiller.[10]

Table 2 summarizes the descriptive statistics for the aforementioned data. From Table 2, it is evident that at each sampling frequency, there are positive and statistically significant correlations between SMB and the following variables: MKT, DEF, and TRM.

4. In-sample predictive regressions

Our goal in this section is to predict the value of SMB_t using a linear predictive regression. In total, we have a set of $N = 9$ lagged variables that are suspected to be relevant (the lagged values of SMB and the 8 variables that are presented in Table

[8] For a complete description of the construction of the MOM factor returns, see `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_mom_factor.html`.

[9] See `http://research.stlouisfed.org/fred2/`.

[10] See `http://www.econ.yale.edu/~shiller/data.htm`.

Variable	Mean	Std	$\mathbf{AC_1}$	Correlation coefficients								
				SMB	MKT	YLD	HML	MOM	DEF	TBL	TRM	INF
Monthly sampling frequency												
SMB	0.0025	0.0334	**0.07**	1.00								
MKT	0.0090	0.0529	**0.09**	**0.25**	1.00							
$YIELD$	0.0392	0.0145	**0.99**	0.05	**0.07**	1.00						
HML	0.0039	0.0359	**0.19**	**0.10**	**0.22**	0.04	1.00					
MOM	0.0070	0.0482	**0.08**	**-0.16**	**-0.34**	-0.04	**-0.40**	1.00				
DEF	1.1402	0.7162	**0.98**	**0.09**	-0.02	**0.34**	0.02	**-0.10**	1.00			
$TBILL$	0.0030	0.0025	**0.97**	**-0.06**	-0.01	**-0.12**	0.02	0.05	**-0.09**	1.00		
$TERM$	0.0013	0.0011	**0.86**	**0.11**	0.02	-0.06	-0.01	-0.05	**0.27**	**-0.49**	1.00	
INF	0.0025	0.0053	**0.55**	-0.01	-0.01	0.01	**0.09**	0.02	**-0.25**	**0.26**	**-0.13**	1.00
Quarterly sampling frequency												
SMB	0.0101	0.0777	-0.04	1.00								
MKT	0.0283	0.1082	-0.06	**0.51**	1.00							
$YIELD$	0.0098	0.0040	**0.78**	-0.00	-0.01	1.00						
HML	0.0132	0.0796	-0.04	**0.32**	**0.34**	0.01	1.00					
MOM	0.0188	0.1055	-0.09	**-0.30**	**-0.48**	0.01	**-0.54**	1.00				
DEF	1.1402	0.7082	**0.93**	**0.18**	-0.02	**0.35**	0.03	**-0.17**	1.00			
$TBILL$	0.0090	0.0076	**0.97**	-0.09	-0.02	-0.10	-0.01	0.10	-0.09	1.00		
$TERM$	0.0040	0.0032	**0.86**	**0.15**	0.05	-0.07	0.02	**-0.13**	**0.29**	**-0.47**	1.00	
INF	0.0076	0.0134	**0.52**	0.01	0.00	-0.04	0.08	0.02	**-0.29**	**0.32**	**-0.17**	1.00
Semi-annual sampling frequency												
SMB	0.0193	0.1084	-0.09	1.00								
MKT	0.0563	0.1428	0.11	**0.40**	1.00							
$YIELD$	0.0197	0.0077	**0.86**	0.06	0.02	1.00						
HML	0.0252	0.1086	-0.03	**0.35**	**0.19**	0.03	1.00					
MOM	0.0439	0.1235	-0.02	-0.09	**-0.27**	-0.04	**-0.16**	1.00				
DEF	1.1402	0.6956	**0.91**	**0.24**	-0.06	**0.35**	-0.02	**-0.22**	1.00			
$TBILL$	0.0181	0.0153	**0.96**	-0.11	-0.03	-0.11	0.00	0.14	-0.09	1.00		
$TERM$	0.0081	0.0064	**0.81**	**0.21**	0.08	-0.08	0.03	**-0.20**	**0.30**	**-0.45**	1.00	
INF	0.0154	0.0237	**0.54**	-0.07	-0.04	0.03	0.07	0.07	**-0.30**	**0.36**	**-0.18**	1.00
Annual sampling frequency												
SMB	0.0375	0.1425	**0.29**	1.00								
MKT	0.1140	0.1982	0.04	**0.31**	1.00							
$YIELD$	0.0399	0.0151	**0.90**	0.04	-0.00	1.00						
HML	0.0489	0.1394	-0.00	0.07	0.10	0.04	1.00					
MOM	0.0973	0.1597	-0.03	-0.09	-0.05	-0.00	-0.18	1.00				
DEF	1.1402	0.6768	**0.82**	**0.27**	-0.10	**0.35**	-0.12	**-0.21**	1.00			
$TBILL$	0.0367	0.0311	**0.91**	-0.15	-0.01	-0.11	0.05	0.20	-0.08	1.00		
$TERM$	0.0162	0.0137	**0.56**	**0.24**	0.12	-0.06	0.08	**-0.23**	**0.32**	**-0.45**	1.00	
INF	0.0313	0.0417	**0.63**	-0.03	0.01	0.04	0.14	0.03	**-0.33**	**0.41**	**-0.28**	1.00

Table 2: Descriptive statistics for the study data, including means, standard deviations, first-order autocorrelations (denoted by $\mathbf{AC}_1$), and correlation coefficients. The variable MKT represents the stock market return. YLD is the dividend yield on the stock market. SMB and HML are Fama-French risk factors. MOM is the Jegadeesh-Titman momentum factor. DEF is the default spread. TBL is the T-bill return. TRM is the term premium. Finally, INF is the inflation rate. The bold text indicates values of first-order autocorrelations and (contemporaneous) correlations that are statistically significant at the 5% level.

1). The traditional approach to estimating a predictive regression is to use all of these potential predictors:

$$SMB_t = \beta_0 + \sum_{j=1}^{N} \beta_j X_{j,t-1} + \varepsilon_t,$$

where X_j, $1 \leq j \leq N$, is a potential predictor. After this regression is estimated, the importance of the predictors is typically evaluated by assessing the degree of statistical significance of the coefficient for each predictor.

However, because no theory exists that explains the dynamics of the SMB factor, there are high levels of uncertainty regarding which variables should enter the predictive regression. Consequently, one potential problem with the traditional approach is the possible inclusion of irrelevant predictive variables. Although irrelevant variables should generate no bias, their inclusion would reduce the precision of the regression. In particular, the inclusion of irrelevant variables increases the variances of the estimated coefficients (which can result in the omission of an important predictor) and typically decreases the adjusted R^2 values of regressions. Therefore, to avoid efficiency losses that could be caused by the inclusion of irrelevant variables, we perform a model selection procedure that is similar to the procedure used by Bossaerts and Hillion (1999).

In particular, in the presence of model uncertainty, there are 2^N competing regression specifications of the following form:

$$SMB_t = \begin{cases} \beta_0 + \sum_{j=1}^{n} \beta_j x_{j,t-1} + \varepsilon_t & \text{if } n > 0, \\ \beta_0 + \varepsilon_t & \text{if } n = 0, \end{cases}$$

where $0 \leq n \leq N$ and $\{x_1, \ldots, x_n\}$ is a model-unique subset of $\{X_1, \ldots, X_N\}$. This subset represents a possible combination of n selected variables out of the N potential predictors. For example, when $n = 0$, the returns on the SMB factor are independent and identically distributed. The all-inclusive model corresponds to $n = N$. We estimate each competing model and rank each model according to a specific model selection criterion. The competing model with the highest rank is then selected as the final model for this study. In other words, using a model selection criterion, we select the "best" model from among a set of candidate models. In this study on the predictability of the small stock premium, we use several popular model selection criteria: the adjusted R^2 value, the Akaike information criterion (AIC, suggested by Akaike (1974)), the Bayesian information criterion (BIC, presented by Schwarz (1978)), and the Hannan-Quinn information criterion (HQC, see Hannan and Quinn (1979)). We obtain similar results from the use of the adjusted R^2 value, the AIC, or the HQC as the model selection criterion; by contrast, the use of the BIC as the model selection criterion produces notably weaker results with respect to predictability. We believe that these findings reflect the fact that the BIC penalizes free parameters much more strongly than the other criteria that are utilized. As a result, a model that is selected by the BIC is prone to underfitting (which appears to produce worse results than overfitting). The results that are reported throughout the remainder of the paper are based on the use of the AIC as the model selection criterion.

To assess the robustness of our findings, we also perform the best model selection procedure using subsamples of our data. In particular, in addition to the total sample, which spans 84 years, we consider two overlapping subsamples; each of these subsamples incorporates a timespan of 50 years. The timespan of the first subsample begins in January 1927 and ends in December 1976, and the timespan of the second subsample begins in January 1961 and ends in December 2010. The primary reason for using overlapping subsamples is to include sufficient quantities of data to permit reliable regression estimates to be obtained for the semi-annual and annual sampling frequencies.

The results of the model selection procedure for the in-sample (IS) predictive regressions at different sampling frequencies and for different subsamples are reported in Table 3. First, it may be observed that the SMB factor return is predictable not only at the monthly frequency (as indicated by Ferson and Harvey (1999) and Cooper, Gulen, and Vassalou (2001)) but also at the other three examined frequencies. The explanatory power of each regression, as measured by the adjusted R^2 value, increases as the time horizon expands.[11] Second, note that for the second subsample, the

[11]This phenomenon is not new; see, for instance, Fama and French (1988), who report that the power of the dividend yield to predict returns increases as time horizons expand. However, this phenomenon is due in part to the fact that a reduced number of observations are considered as the sampling frequency decreases.

Period	Coefficient for									Adjusted R^2
	SMB	MKT	YLD	HML	MOM	DEF	TBL	TRM	INF	
Monthly sampling frequency										
1927-2010	-	**0.153**	-	**0.066**	-	**0.007**	-	-	-	0.091
		(0.000)		(0.021)		(0.000)				
1927-1976	-	**0.176**	-	**0.116**	**0.084**	**0.008**	-	-	-	0.130
		(0.000)		(0.005)	(0.011)	(0.000)				
1961-2010	-	**0.148**	-	-	-	**0.006**	-0.918	-	-	0.049
		(0.000)				(0.029)	(0.094)			
Quarterly sampling frequency										
1927-2010	-	-	-	-	**0.123**	**0.028**	-	-	**-0.642**	0.096
					(0.002)	(0.000)			(0.043)	
1927-1976	-0.120	-	-	**0.205**	**0.226**	**0.031**	-	-	**-0.891**	0.136
	(0.114)			(0.030)	(0.001)	(0.000)			(0.040)	
1961-2010	-	-	**5.452**	-	0.092	**0.023**	**-3.268**	-2.407	-	0.060
			(0.004)		(0.080)	(0.032)	(0.002)	(0.115)		
Semi-annual sampling frequency										
1927-2010	**-0.173**	-	-	-	**-0.299**	**0.035**	-	-	-	0.187
	(0.018)				(0.000)	(0.002)				
1927-1976	**-0.231**	-	-	-	**-0.526**	**0.032**	-	-	-	0.345
	(0.008)				(0.000)	(0.017)				
1961-2010	-	-0.140	**4.637**	-	-	-	-1.322	-	-	0.053
		(0.057)	(0.019)				(0.126)			
Annual sampling frequency										
1927-2010	**0.300**	**-0.196**	-	-	-	**0.057**	-	-	-	0.218
	(0.007)	(0.011)				(0.010)				
1927-1976	**0.480**	**-0.272**	-	-	**-0.506**	-	-	-	-	0.273
	(0.001)	(0.011)			(0.006)					
1961-2010	**0.307**	**-0.272**	-	-	-	-	-	-	-	0.143
	(0.028)	(0.016)								

Table 3: The results of the model selection procedures and estimations for the in-sample predictive regressions. The variable MKT represents the stock market return. YLD is the dividend yield on the stock market. SMB and HML are Fama-French risk factors. MOM is the Jegadeesh-Titman momentum factor. DEF is the default spread. TBL is the T-bill return. TRM is the term premium. Finally, INF is the inflation rate. To obtain these results, we estimate 2^N (for $N = 9$) competing models of the form $SMB_t = \beta_0 + \sum_{j=1}^{n} \beta_j x_{j,t-1} + \epsilon_t$, where $0 \leq n \leq N$ and $\{x_1, \ldots, x_n\}$ is a subset of N predictive variables. We rank each of these models in accordance with its AIC. The final model that is selected is the model with the lowest AIC. A hyphen for the value of a coefficient indicates that the AIC does not include this variable in the model. The p-value for each coefficient is given in brackets. Bold text is used to indicate values that are statistically significant at the 5% level.

predictive power of the model, as measured by the regression adjusted R^2 value, is notably lower than the predictive power of the first subsample. This result might suggest that the predictability of the SMB factor return has diminished over time. In other words, this finding might indicate that stock market efficiency has increased over time. Third, observe that at the semi-annual frequency, the best predictive model for the second subsample is completely different from the best predictive model for the first subsample. We conclude[12] that at this frequency, there are no robust predictors[13] of the SMB factor return. However, at the monthly, quarterly, and annual frequencies, robust predictors of the SMB factor return do exist. At the monthly frequency, these robust predictors are the return on the stock market and the default spread. At the quarterly frequency, the robust predictor of the SMB factor return is the default spread. Observe that at this frequency, the momentum factor is also considered by the best predictive model for both the total sample and each subsample. However, for the second subsample, MOM is not statistically significant at the 5% level. Finally, at the annual frequency, the robust predictors of the SMB factor return are the market return and the lagged SMB factor return.[14]

To conclude this section, we would like to briefly discuss the implications of our results for the ongoing debate regarding the source of the size effect and emphasize a controversy that these results have revealed. Recall that Chan, Chen, and Hsieh (1985) and Vassalou and Xing (2004) find that the default spread is a significant explanatory variable for the size premium. This result is also supported by Table 2, which reports that at every examined sampling frequency, the correlation coefficient between the default spread and the SMB factor return is positive and statistically significant. Our findings also reveal that if we focus on the results of the best model selection procedure over the complete data sample, an increase in the default spread

[12]This conclusion is also supported by the results of the out-of-sample predictability tests, as demonstrated in the subsequent section.

[13]By a robust predictor, we mean a statistically significant variable that is selected by the best model selection procedure for both the full sample and each subsample.

[14]In fact, this finding naturally derives from the fact that the SMB factor return is a persistent variable at the annual frequency (see Table 2; the value of the autocorrelation of SMB over a one-year horizon is positive and statistically significant).

predicts a higher SMB factor return at every examined sampling frequency. Thus, our results provide several implications. First, the default spread appears to be not only an explanatory variable for the size premium but also a predictive variable for this premium. Secondly, because an increase in the default spread signals a deterioration of economic conditions, the effects of the default spread on the size premium suggest that this size premium tends to appear during adverse economic conditions. This interpretation is consistent with the hypothesis that the size premium represents compensation to investors in exchange for bearing greater systematic risk. Furthermore, note that at the annual sampling frequency, a negative return on the stock market predicts a positive size premium. This result is also consistent with the risk-based explanation for the size effect. The logic underlying this explanation is that a negative return on the stock market signals a deterioration in economic conditions. As an additional argument in favor of the risk based explanation, observe from Table 2 that at each sampling frequency, the correlation coefficient between the SMB factor return and the term premium is positive and statistically significant. In other words, a positive return on the size factor tends to coincide with an increase in the term premium. Because the term premium typically increases as a result of the expansionary monetary policy that is implemented during difficult economic situations, this result also supports the risk-based explanation of the size effect.

However, the study results also provide evidence that contradicts the risk-based explanation of the size effect. Observe that our findings reveal the existence of a lead-lag relationship between the return on the stock market and the return on the size premium at both the annual and monthly sampling frequencies. Unexpectedly, opposite signs are found for this relationship at these two sampling frequencies. In particular, at the one-month horizon, a positive return on the market predicts a positive SMB factor return. This result is not consistent with the risk-based explanation of the size effect but is consistent with the price delay hypothesis of Hou and Moskowitz (2005). Moreover, observe from Table 2 that at each sampling frequency, the correlation coefficient between the SMB factor return and the stock market return is positive and statistically significant. This finding implies that a positive return on the size factor tends to coincide with a positive return on the stock market, a result that also appears to be inconsistent with the risk-based explanation. We believe that the resolution of this controversy would constitute an interesting topic for further research on the size premium.

5. The out-of-sample performance of predictive regressions

The results that are reported in the previous section indicate the existence of IS predictability for the return on the SMB factor over horizons ranging from one month to one year. However, it is well known that IS predictability can be spurious (for instance, this predictability can appear as a result of data mining) and not hold for out-of-sample (OOS) observations (see, for example, Bossaerts and Hillion (1999) and Goyal and Welch (2008)). In this section, to ensure the absence of data mining, we assess the performance of OOS forecasts.

Our OOS recursive forecasting procedure is as follows. Denote the number of observations in the total sample by T. The initial IS period $[1, k]$ (for $k < T$) is used to complete the procedure of selecting the optimal model, as described in the previous section. Subsequently, the estimated coefficients from the best predictive regression up to time k are used to compute the first one-step-ahead return forecast, which refers to time $k+1$. We then expand our in-sample period by one observation, perform the best model selection procedure once again using the new IS period of $[1, k+1]$, and compute the OOS forecast for time $k+2$. We repeat this procedure, pushing the endpoint of our IS period ahead by one observation with each iteration of this process, until we compute the one-step ahead return forecast for the last time T. Note that our OOS forecasting procedure is free from look-ahead bias because we forecast the SMB factor return for time $k+j$ using only information that is available at time $k+j-1$.

To assess the performance of OOS forecast, we employ two test statistics. The first of these test statistics is the ratio of the mean squared forecasting error (MSFE) of the historical mean model to the MSFE of a predictive regression. In particular, the historical mean model is the reduced version of a predictive regression and may be expressed in the following form:

$$SMB_t = \beta_0 + \varepsilon_t. \tag{1}$$

We will refer to the predictive regression and the historical mean model as the unrestricted and restricted models, respectively. The test statistic, which we refer to as the MSFE Ratio, is calculated as follows:

$$\text{MSFE-R} = \frac{MSFE_R}{MSFE_U}, \tag{2}$$

where $MSFE_R$ and $MSFE_U$ are the MSFEs of the restricted and unrestricted models, respectively. The comparison of the mean squared forecasting errors of two alternative models is a well-established approach for evaluating which of the two alternative models provides superior forecasting ability, as discussed by McCracken (2007) and references therein. The historical mean model is a version of the random walk hypothesis and uses the historical average as the prediction for the next period. Consequently, the computation of the p-values of the MSFE-R are conducted under the null hypothesis that the SMB factor return cannot be predicted and that the SMB factor returns are therefore independent and identically distributed. Under the null hypothesis, the restricted model has an MSFE that is less than or equal to the MSFE of the unrestricted model. In other words, under the null hypothesis, $MSFE_R \leq MSFE_U$, and the expected MSFE-R values should therefore be less than or equal to unity. We reject the null hypothesis if the actual MSFE-R estimates are significantly greater than the expected value.

If two alternative forecasting errors are assumed to be Gaussian, serially uncorrelated, and contemporaneously uncorrelated, then the MSFE ratio should exhibit the typical F-distribution if the null hypothesis holds.[15] However, in our case, not all of the aforementioned assumptions are satisfied. In particular, because the restricted model is the reduced version of the unrestricted model (in other words, the best predictive model *encompasses* the random walk model), the forecasting errors of both of these models are contemporaneously correlated. One possibility for obtaining correct statistical inferences in this case is to perform asymptotically valid tests that are similar to the seminal tests that were performed by Diebold and Mariano (1995). However, because we use relatively small samples when we perform the forecasting

[15] In this case, testing the null hypothesis largely corresponds to the standard F-test for examining equal forecast error variances.

at the semi-annual and annual frequencies, to compute the p-values of the MSFE-R statistics, we employ a nonparametric bootstrap method.

More specifically, the computation of the MSFE-R statistic and the bootstrap procedure to compute the p-values of this statistic are performed as follows. First, using the original time series of the SMB and the set of potential predictive variables, we recursively estimate the best predictive regression and the historical mean model to obtain the OOS forecasts of the unrestricted and restricted models, respectively. We then compute the mean squared forecasting errors $MSFE_U$ and $MSFE_R$, and subsequently calculate MSFE-R. After completing these calculations, we bootstrap the original series to obtain random resamples of the SMB and the set of potential predictive variables. Because the potential predictive variables and the SMB typically exhibit statistically significant correlations (see Table 2), we account for these correlations; in particular, to address these correlations, we resample the entire vector of variables instead of resampling each variable separately. As a result, the bootstrapped observation of the vector $\{X^*_{1,\tau}, \ldots, X^*_{N,\tau}\}$ for time τ represents[16] the actual historical observation of the vector $\{X_{1,t}, \ldots, X_{N,t}\}$ for a time t. Through this process, we retain the historical correlations among the examined variables. We repeat this bootstrap procedure a large number of times; in each iteration, the best predictive regression and the historical mean model are recursively estimated, and an estimate is obtained for MSFE-R*. In this manner, we estimate the sampling distribution of MSFE-R under the conditions specified by the null hypothesis. Finally, to estimate the significance level of the results, we count how many times the computed value for MSFE-R* after bootstrapping happens to be above the value of the actual estimate for MSFE-R. In other words, under the null hypothesis, we compute the probability of obtaining a more extreme value for the MSFE Ratio than the actual estimate. In this manner, we compute one-tailed p-values.

Observe that the MSFE measures the statistical goodness of fit, which can be regarded as how close the forecasted values of $\widehat{SMB}_t$ for $t = k+1, \ldots, T$ are to

[16]The superscripted asterisk is the typical symbol for denoting a bootstrapped observation. Note that the SMB series is an X_i series and is therefore also bootstrapped.

the actual historical realizations of SMB_t. However, it can be argued that investors are typically less concerned about the ability of the forecast to correctly predict the magnitude of SMB_t but are instead focused on the ability of the forecast to provide the correct direction for trading activity. In other words, $\widehat{SMB}_t$ may be a poor forecast by the MSFE standard but may nonetheless consistently determine the profitable direction for trading activity. Our second test statistic for assessing the quality (the usefulness) of the OOS forecast is based on the well-known Henriksson and Merton test of directional accuracy (see Henriksson and Merton (1981) and the subsequent formalization of this test by Pesaran and Timmermann (1992)). It is natural to use the sign of the SMB factor return as a numerical characterization of the direction of the SMB factor. Consequently, we compute the Henriksson-Merton test statistics in the following manner:

$$\text{HM-S} = Prob\left(\widehat{SMB}_t > 0 \middle| SMB_t > 0\right) + Prob\left(\widehat{SMB}_t \leq 0 \middle| SMB_t \leq 0\right), \tag{3}$$

In the equation above, $Prob\left(\widehat{SMB}_t > 0 \middle| SMB_t > 0\right)$ is the conditional probability of a correct forecast of a positive return on the SMB factor given that the realized return on the SMB factor is positive, whereas $Prob\left(\widehat{SMB}_t \leq 0 \middle| SMB_t \leq 0\right)$ is the conditional probability of a correct forecast of a negative return on the SMB factor given that the realized return on the SMB factor is negative. The idea underlying this approach is as follows. If a model is able to forecast the sign of the SMB factor return, then HM-S > 1. If a model always correctly predicts the sign of the SMB factor return, then HM-S $= 2$. Under the null hypothesis that the SMB factor return cannot be predicted, the value of the HM statistic should be unity (HM-S $= 1$). Consequently, using the same nonparametric bootstrap method that was described above for the MSFE-R statistic, we compute the p-values of the HM statistic and reject the null hypothesis if the actual estimate for the HM statistic is significantly greater than unity.

In our tests we use two different specifications for the initial IS and OOS periods. In the first specification, adopting the approach of Goyal and Welch (2008), the OOS forecast begins in January 1947, 20 years after the first sample data are available. In the second specification, the OOS forecast begins in January 1976. The motivation for this second specification is the findings of Goyal and Welch (2008), who examine the performance of different predictive models and report that after the occurrence of the oil shock of 1973-1975, the examined models have produced poor predictive performances. Table 4 reports the computed values of the MSFE ratio and HM statistics and their estimated p-values from the performance of 1000 bootstraps. From the statistical significance of the MSFE ratio statistic, we can reject the null hypothesis of no OOS predictability only at the monthly predictive frequency. If we adopt the approach of Goyal and Welch (2008) and rely only on the goodness of fit as the sole measure of OOS predictability, we would conclude that there is no predictability for the SMB factor returns at the other frequencies. However, the values of the Henriksson-Merton test statistics are significant not only at the monthly frequency but also at the quarterly and annual frequencies. Consequently, we conclude that there is OOS predictability of the sign of the SMB factor return at the monthly, quarterly, and annual frequencies. In contrast to the findings of Goyal and Welch (2008), our results indicate that OOS predictability persists after the 1973-1975 oil shock, although the values of all of the test statistics that are statistically significant for the 1976-2010 time period are less than the corresponding values of each of these statistics for the 1947-2010 period. This result agrees with our findings regarding IS predictability and may suggest that the predictability of SMB factor returns has been diminishing over time.

6. Exploiting the predictability of SMB factor returns through active portfolio management

The results reported in the previous section demonstrate the existence of the OOS predictability of SMB factor returns at the monthly, quarterly, and annual frequencies. Because the SMB factor returns are predictable, it would be natural to explore the usefulness of this predictability for active portfolio management. One of the possible simple and realistic trading strategies that can be implemented in practice is to invest in small size firms if the predicted return on the SMB factor is positive and to invest in large size firms otherwise. To evaluate the performance of this strategy, we conduct the following experiment.

Frequency	**MSFE Ratio**		**HM Statistic**	
	1947-2010	1976-2010	1947-2010	1976-2010
Monthly	**1.038** (0.000)	**1.027** (0.000)	**1.159** (0.000)	**1.095** (0.003)
Quarterly	0.861 (0.981)	0.876 (0.974)	**1.121** (0.010)	**1.103** (0.032)
Semi-annual	0.651 (0.993)	0.642 (0.999)	1.042 (0.213)	1.112 (0.068)
Annual	0.833 (0.536)	0.898 (0.572)	**1.369** (0.001)	**1.280** (0.017)

Table 4: The performance of the out-of-sample forecasts for the SMB factor return. This table reports the computed values of the mean squared forecasting error ratio (MSFE Ratio, as specified by equation (2)) and the Henriksson-Merton test statistic (HM Statistic, as specified by equation (3)). The values in the brackets report the p-values of one-tailed tests of the null hypothesis that SMB factor returns are unpredictable. In particular, a p-value reports the probability of observing a value of a test statistic that is above the computed value of this statistic under the assumptions of the null hypothesis. The bold text indicates values that are statistically significant at the 5% level.

The additional data for this experiment are the monthly returns on 10 value-weighted portfolios that are formed on the basis of size (market equity) and monthly returns on a broad market index. The data are obtained from the data library of Kenneth French. The return on the market index is the value-weighted return on all NYSE, AMEX, and NASDAQ stocks (from the Center for Research in Security Prices). The size decile portfolios include all NYSE, AMEX, and NASDAQ stocks for which market equity data exist. These decile portfolios are denoted by S1, S2, ..., S10; S1 refers to the firms with the lowest 10 percent of market capitalization, and S10 refers to the largest 10 percent of firms. We employ the same OOS recursive forecasting procedure for the SMB factor return that was detailed in the preceding section. The active portfolio is constructed in the following manner. If the predictive return on the SMB factor is positive, funds are allocated to decile portfolio Si (where $i = 1, 2, \dots, 5$) for the next period. Otherwise, funds are allocated to decile portfolio S10. As before, the OOS forecast begins 20 years after the initial sample data are available, that is, an actively managed portfolio is constructed for the first time in January 1947. We also report the performances of the actively managed portfolios beginning in January 1976, after the occurrence of the 1973-1975 oil shock.

To measure portfolio performance, we use the following five measures: (1) the

Sharpe ratio; (2) the alpha of the standard one-factor market model (CAPM); (3) the alpha of the Fama-French three-factor model (FF); (4) the alpha of the Fama-French-Carhart four-factor model (FFC); and (5) the Treynor-Mazuy gamma (as devised by Treynor and Mazuy (1966)). Regardless of the forecasting horizon, all of these performance measures are computed at a monthly frequency. We remind the reader that the Sharpe ratio is the appropriate risk-adjusted performance measure if a risky portfolio represents an investor's exclusive investment vehicle. The Sharpe ratio is computed as follows:

$$\text{Sharpe ratio} = \frac{E[R_p]}{Std[R_p]},$$

where R_p is the time series of the excess return on a portfolio, $E[.]$ is the expectation operator, and $Std[.]$ is the standard deviation. An alpha, on the other hand, is the appropriate risk-adjusted performance measure if a risky portfolio is one of many portfolios that are combined into a large investment fund. To estimate the alphas for the CAPM, FF, and FFC, we run the following regressions:

$$\begin{aligned} \text{CAPM alpha: } & R_{pt} = \alpha_p + \beta_p R_{Mt} + \varepsilon_{pt}, \\ \text{FF alpha: } & R_{pt} = \alpha_p + \beta_p R_{Mt} + s_p SMB_t + h_p HML_t + \varepsilon_{pt}, \\ \text{FFC alpha: } & R_{pt} = \alpha_p + \beta_p R_{Mt} + s_p SMB_t + h_p HML_t + m_p MOM_t + \varepsilon_{pt}, \end{aligned}$$

where R_{pt} and R_{Mt} are the excess returns on the examined portfolio and the market index at time t, respectively, and ε_{pt} is the disturbance term for this portfolio at time t. If alpha is positive and statistically significant, we have evidence that a portfolio manager has an ability to generate alpha (in other word, genuine stock-picking ability). The Treynor-Mazuy gamma measures the market timing ability of a portfolio manager and is estimated by performing the following regression:

$$\text{TM gamma: } R_{pt} = \alpha_p + \beta_p R_{Mt} + \gamma_p R_{Mt}^2 + \varepsilon_{pt}.$$

A result for the Treynor-Mazuy gamma, γ_p, that is positive and statistically significant provides evidence that a portfolio manager possesses market timing abilities. The reasoning underlying this idea is that compared with a situation in which $\gamma_p = 0$, the return on a portfolio tends to decline less in a falling market and to increase more in

Performance measure	1947-2010						1976-2010					
	S1	S2	S3	S4	S5	S10	S1	S2	S3	S4	S5	S10
Sharpe ratio	0.13	0.13	0.14	0.14	0.15	0.13	0.14	0.13	0.15	0.14	0.16	0.11
CAPM alpha	0.17	0.08	0.13	0.10	0.12	-0.01	0.26	0.19	0.22	0.19	**0.23**	-0.04
	(0.23)	(0.48)	(0.17)	(0.25)	(0.11)	(0.71)	(0.18)	(0.26)	(0.11)	(0.15)	(0.04)	(0.41)
FF alpha	-0.06	**-0.09**	-0.02	-0.03	0.01	**0.04**	-0.08	-0.11	-0.04	-0.06	0.03	**0.05**
	(0.37)	(0.03)	(0.48)	(0.39)	(0.77)	(0.05)	(0.39)	(0.06)	(0.37)	(0.26)	(0.55)	(0.05)
FFC alpha	-0.06	**-0.10**	-0.00	-0.03	0.03	**0.04**	-0.08	-0.11	-0.02	-0.06	0.05	0.05
	(0.37)	(0.02)	(0.89)	(0.40)	(0.43)	(0.04)	(0.39)	(0.06)	(0.61)	(0.29)	(0.37)	(0.06)
TM gamma	**-1.19**	**-0.77**	**-0.66**	**-0.60**	**-0.53**	**0.30**	**-1.62**	**-1.09**	**-0.93**	**-0.77**	**-0.63**	**0.40**
	(0.00)	(0.02)	(0.02)	(0.02)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)

Table 5: The performance measures of the benchmark portfolios. S1 refers to a portfolio of firms that are in the lowest 10 percent with respect to market capitalization, whereas S10 refers to a portfolio of firms that are in the largest 10 percent with respect to market capitalization. CAPM alpha is the intercept in the standard one-factor market model. FF alpha is the intercept in the Fama-French three-factor model (which includes the SMB and HML factors). FFC alpha is the intercept in the Fama-French-Carhart four-factor model (which includes the SMB, HML, and MOM factors). The alpha values are estimated from monthly data. The alpha values are given as percentages, and the corresponding p-values are reported in brackets. The TM gamma is the Treynor-Mazuy measure of market timing ability, and the p-value of each gamma is reported in brackets. The bold text indicates values that are statistically significant at the 5% level.

a rising market if $\gamma_p > 0$. By contrast, compared with a situation in which $\gamma_p = 0$, the return on a portfolio tends to decline more in a falling market and to increase less in a rising market if $\gamma_p < 0$.

The benchmarks for our active strategies are passive buy-and-hold portfolios[17] for smaller capitalization firms (portfolios S1 to S5) and portfolio S10, which include the firms with the largest market capitalization. Table 5 reports the performance measures of the benchmark portfolios. Observe that among the benchmark portfolios, only portfolio S5 has a positive and statistically significant alpha in the standard one-factor market model, and this significance is only observed for the 1976-2010 time

[17]Note, however, that the size decile portfolios are not truly passive but are instead dynamic in the sense that their compositions change slightly over time because these portfolios necessarily encompass a degree of turnover. For example, a stock in the smallest decile does not constantly remain in this decile; after the stock in question has experienced a sufficiently large price increase, this stock will exit from the smallest decile. Therefore, annual rebalancing is required within each decile; in particular, the portfolios are reconfigured at the end of each June using the June market equity values and NYSE breakpoints, as discussed in Fama and French (1993). We assume that these rebalancing costs are relatively insignificant.

period. The passive portfolios of smaller capitalization firms do not deliver positive and statistically significant alphas in either the Fama-French three-factor model or the Fama-French-Carhart four-factor model. On the contrary, the majority of the measured alphas are negative. Furthermore, if a portfolio manager allocates funds to a passive portfolio of smaller capitalization firms, then by the sign and statistical significance of the Treynor-Mazuy metric, this manager is demonstrating a clear evidence of negative market timing ability. Surprisingly, the passive buy-and-hold portfolio of the largest firms, S10, delivers positive and statistically significant alphas in both the Fama-French three-factor model and the Fama-French-Carhart four-factor model (over the 1976-2010 time period, the FFC alpha is statistically significant at the 6% level). In addition, this passive portfolio demonstrates a positive and statistically significant TM gamma; thus, investing in the decile of largest firms allows a portfolio manager without any particular skills to produce evidence of superior stock picking and market timing abilities!

Table 6 reports the performance measures of five actively managed portfolios in the absence of transaction costs. First, it is notable that regardless of the length of the predictive horizon that is examined (one month, one quarter, or one year), the Sharpe ratio of an active portfolio is higher than the Sharpe ratio of each passive component. At the monthly predictive horizon, all of the active portfolios deliver positive and statistically significant alphas in the standard one-factor model, in the Fama-French three-factor model, and in the Fama-French-Carhart four-factor model. The Treynor-Mazuy gammas are also positive, and certain of these values are statistically significant. The performances of the actively managed portfolios are less impressive for the one-year predictive horizon than for the one-month horizon. In particular, for the one-year predictive horizon, the active portfolios always deliver positive and statistically significant alphas in the standard one-factor model, frequently produce these types of alphas in the Fama-French three-factor model, and sometimes produce positive and significant alphas in the Fama-French-Carhart four-factor model. The Treynor-Mazuy gammas are almost all positive, and three out of five of the gamma values during the 1947-2010 time period are statistically significant. The use of the one-quarter predictive horizon for forecasting the sign of the SMB factor return pro-

Performance measure	1947-2010					1976-2010				
	S1-S10	S2-S10	S3-S10	S4-S10	S5-S10	S1-S10	S2-S10	S3-S10	S4-S10	S5-S10
					Monthly forecast					
Sharpe ratio	0.21	0.19	0.19	0.18	0.19	0.21	0.18	0.18	0.18	0.18
CAPM alpha	**0.50**	**0.34**	**0.32**	**0.28**	**0.29**	**0.55**	**0.39**	**0.36**	**0.33**	**0.31**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
FF alpha	**0.39**	**0.25**	**0.26**	**0.23**	**0.24**	**0.38**	**0.24**	**0.25**	**0.23**	**0.23**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.01)	(0.02)	(0.01)
FFC alpha	**0.46**	**0.33**	**0.33**	**0.28**	**0.30**	**0.47**	**0.32**	**0.33**	**0.29**	**0.29**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
TM gamma	0.03	0.36	0.39	**0.37**	**0.39**	-0.12	0.34	0.42	0.41	**0.46**
	(0.91)	(0.14)	(0.07)	(0.05)	(0.02)	(0.75)	(0.28)	(0.13)	(0.10)	(0.04)
					Quarterly forecast					
Sharpe ratio	0.16	0.15	0.16	0.15	0.16	0.16	0.15	0.16	0.16	0.16
CAPM alpha	**0.20**	0.14	**0.16**	0.12	**0.13**	**0.32**	0.23	**0.26**	**0.23**	**0.23**
	(0.04)	(0.10)	(0.03)	(0.07)	(0.03)	(0.03)	(0.07)	(0.01)	(0.02)	(0.01)
FF alpha	0.14	0.09	0.12	0.10	**0.11**	0.17	0.10	0.16	0.15	**0.17**
	(0.13)	(0.25)	(0.06)	(0.12)	(0.05)	(0.19)	(0.37)	(0.09)	(0.10)	(0.05)
FFC alpha	0.13	0.10	**0.15**	0.12	**0.14**	0.16	0.11	0.17	0.16	**0.19**
	(0.16)	(0.20)	(0.03)	(0.07)	(0.02)	(0.25)	(0.36)	(0.08)	(0.09)	(0.03)
TM gamma	-0.48	-0.15	-0.13	-0.14	-0.10	-0.24	0.12	0.12	0.18	0.20
	(0.09)	(0.51)	(0.51)	(0.46)	(0.56)	(0.50)	(0.70)	(0.64)	(0.47)	(0.36)
					Annual forecast					
Sharpe ratio	0.17	0.16	0.17	0.17	0.17	0.19	0.17	0.18	0.17	0.18
CAPM alpha	**0.35**	**0.25**	**0.26**	**0.22**	**0.23**	**0.52**	**0.37**	**0.39**	**0.32**	**0.34**
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
FF alpha	**0.22**	**0.15**	**0.17**	**0.16**	**0.17**	**0.30**	0.17	**0.22**	0.17	**0.21**
	(0.01)	(0.05)	(0.01)	(0.01)	(0.00)	(0.02)	(0.13)	(0.01)	(0.06)	(0.01)
FFC alpha	0.13	0.07	0.11	0.09	**0.13**	0.24	0.11	**0.18**	0.12	**0.18**
	(0.17)	(0.39)	(0.10)	(0.16)	(0.03)	(0.07)	(0.30)	(0.05)	(0.17)	(0.03)
TM gamma	0.22	0.44	**0.51**	**0.41**	**0.43**	-0.12	0.29	0.36	0.31	0.38
	(0.51)	(0.12)	(0.03)	(0.05)	(0.02)	(0.77)	(0.42)	(0.23)	(0.26)	(0.12)

Table 6: The performance measures of the actively managed portfolios *in the absence of transaction costs*. S1 refers to a portfolio of firms that are in the smallest 10 percent with respect to market capitalization, S10 refers to a portfolio of firms that are in the largest 10 percent with respect to market capitalization. The active strategy, which is denoted by Si-S10 (where i takes values from 1 to 5), allocates funds for the next period to portfolio Si if the OOS forecasted return on the SMB factor is positive but otherwise allocates these funds to portfolio S10. CAPM alpha is the intercept in the standard one-factor market model. FF alpha is the intercept in the Fama-French three-factor model (which includes the SMB and HML factors). FFC alpha is the intercept in the Fama-French-Carhart four-factor model (which includes the SMB, HML, and MOM factors). The alphas are estimated from monthly data. The alpha values are provided in percentages, and the corresponding p-values are reported in brackets. TM gamma is the Treynor-Mazuy measure of the market timing ability, and the p-value of each gamma is reported in brackets. The bold text indicates values that are statistically significant at the 5% level.

duces the least impressive performance measures of the actively managed portfolios. With the quarterly predictive frequency, although all of the alphas are positive, only a handful of these alphas are statistically significant. The majority of the Treynor-Mazuy gammas are negative, but none of these values are statistically significant.

To reinforce our findings regarding the usefulness of the SMB factor return predictability for active portfolio management, we perform an additional test. In this assessment, we analyze the joint significance of all of the intercepts of the actively managed portfolios. In particular, we denote the vector of intercepts as $\boldsymbol{\alpha}$, and $\boldsymbol{\alpha} = [\alpha_{1-10}\, \alpha_{2-10}\, \alpha_{3-10}\, \alpha_{4-10}\, \alpha_{5-10}]$, where α_{i-10} is the alpha of the active strategy Si-S10. Formally, we test the hypothesis $H_0 : \boldsymbol{\alpha} = 0$. For this purpose, we perform the F test devised by Gibbons, Ross, and Shanken (1989) (also known as the GRS test). Table 7 reports the results of this test. Observe that at the monthly predictive horizon, we can reject the null hypothesis at the 5% level regardless of the model that is employed (CAPM, FF, or FFC) and the period that is examined (1947-2010 or 1976-2010). At the annual predictive horizon, we can reject the null hypothesis at the 5% level for all of the models and periods except for the Fama-French-Carhart model during the 1947-2010 time period. At the quarterly predictive horizon, we can reject the null hypothesis in only two out of the six cases that are examined. On the whole, the results of the GRS test of the joint significance of intercepts agree with the conclusions that were suggested by the results for the t-test of the individual significance of intercepts. Specifically, we have solid evidence that by forecasting the sign of the size premium for the upcoming month, a portfolio manager can "beat the market". Similarly, we have less solid but nonetheless convincing evidence that a portfolio manager can generate alpha by forecasting the sign of the size premium for the upcoming year. By contrast, we have only sporadic evidence that by forecasting the sign of the size premium for the upcoming quarter, one can construct an active portfolio that generates alpha.

Thus far, we have examined the performance of active strategies in the absence of transaction costs. In practice, the rebalancing of an active portfolio incurs transaction costs; therefore, it is interesting to assess whether the active strategies can deliver superior performance in the presence of realistic transaction costs. Berkowitz, Logue,

Model	**1947-2010**	**1976-2010**
	Monthly forecast	
CAPM	**6.83** (0.00)	**3.82** (0.00)
Fama-French	**5.42** (0.00)	**2.57** (0.03)
Fama-French-Carhart	**7.04** (0.00)	**3.43** (0.00)
	Quarterly forecast	
CAPM	1.88 (0.10)	**2.25** (0.05)
Fama-French	1.66 (0.14)	1.90 (0.09)
Fama-French-Carhart	**2.36** (0.04)	2.16 (0.06)
	Annual forecast	
CAPM	**3.58** (0.00)	**4.16** (0.00)
Fama-French	**2.70** (0.02)	**3.06** (0.01)
Fama-French-Carhart	1.89 (0.09)	**2.55** (0.03)

Table 7: The results of the Gibbons, Ross, and Shanken (1989) test of the hypothesis that all of the intercepts of the actively managed portfolios *in the absence of transaction costs* are jointly equal to zero. The table reports the values of the GRS test statistic for each situation, and the corresponding p-values are reported in brackets. The bold text indicates values that are statistically significant at the 5% level (implying the rejection of the null hypothesis).

and Noser (1988), Chan and Lakonishok (1993), and Knez and Ready (1996) estimate the average one-way transaction costs for institutional investors to be in the range of 0.23% to 0.25%. We therefore assume that round-trip transaction costs are 0.5% and perform a simulation analysis of active portfolio management in the presence of these transaction costs. The results of these simulations are reported in Tables 8 and 9. It appears that the active strategies based on either monthly or quarterly forecasts of the sign of the SMB factor return are highly transaction intensive[18] and lose their superior performance under conditions involving realistic transaction costs. Although the individual active alphas for these frequencies remain positive, practically none of these alpha values are statistically significant at the 5% level. Similarly, in the majority of cases, we fail to reject the hypothesis that all the intercepts of the actively managed portfolios are jointly zero. By contrast, an active strategy based on annual forecasts delivers superior performance even in the presence of realistic transaction costs.[19] In particular, the values and statistical significances of virtually all of the performance measures of the active strategies for the one-year predictive horizon are similar between the scenario without transaction costs and the scenario that includes these costs.

7. Summary and conclusions

In this paper, we use monthly, quarterly, semi-annual, and annual data to perform tests of the predictability of the small stock premium. Our methodology examines a set of potential macroeconomic predictive variables to determine the best linear predictive regression, as assessed by a model selection criterion. We conclude that there is both in-sample and out-of-sample predictability of the size premium at the monthly, quarterly, and annual frequencies. We reinforce the evidence regarding the out-of-sample predictability by a simulation experiment. In this experiment, we

[18]On average, an active portfolio is rebalanced every 2.4 months if one uses the monthly predictive horizon. If the quarterly predictive horizon is used, then an active portfolio is rebalanced every 3.1 quarters.

[19]On average, an active portfolio is rebalanced every 2.9 years if one uses the annual predictive horizon.

Performance	1947-2010					1976-2010				
measure	S1-S10	S2-S10	S3-S10	S4-S10	S5-S10	S1-S10	S2-S10	S3-S10	S4-S10	S5-S10
					Monthly forecast					
Sharpe ratio	0.17	0.15	0.15	0.14	0.14	0.16	0.14	0.14	0.13	0.13
CAPM alpha	**0.29**	0.13	0.11	0.07	0.08	**0.33**	0.16	0.14	0.10	0.09
	(0.01)	(0.13)	(0.14)	(0.33)	(0.23)	(0.03)	(0.20)	(0.22)	(0.31)	(0.35)
FF alpha	0.18	0.04	0.05	0.02	0.03	0.16	0.01	0.03	0.00	-0.00
	(0.06)	(0.60)	(0.48)	(0.80)	(0.66)	(0.25)	(0.92)	(0.79)	(0.98)	(1.00)
FFC alpha	**0.25**	0.11	0.11	0.06	0.08	0.24	0.09	0.10	0.06	0.06
	(0.01)	(0.16)	(0.11)	(0.33)	(0.18)	(0.08)	(0.41)	(0.33)	(0.53)	(0.50)
TM gamma	0.01	0.34	0.37	0.35	**0.37**	-0.17	0.28	0.36	0.35	0.40
	(0.96)	(0.16)	(0.08)	(0.07)	(0.03)	(0.63)	(0.37)	(0.19)	(0.17)	(0.08)
					Quarterly forecast					
Sharpe ratio	0.15	0.14	0.15	0.14	0.15	0.15	0.14	0.15	0.15	0.15
CAPM alpha	0.15	0.08	0.10	0.07	0.08	0.27	0.18	**0.21**	0.18	**0.19**
	(0.13)	(0.32)	(0.15)	(0.32)	(0.19)	(0.06)	(0.15)	(0.05)	(0.06)	(0.04)
FF alpha	0.09	0.04	0.07	0.05	0.06	0.13	0.06	0.12	0.11	0.12
	(0.34)	(0.61)	(0.26)	(0.46)	(0.26)	(0.34)	(0.62)	(0.22)	(0.25)	(0.15)
FFC alpha	0.08	0.05	0.10	0.07	0.09	0.11	0.06	0.13	0.11	0.15
	(0.39)	(0.53)	(0.15)	(0.31)	(0.12)	(0.41)	(0.60)	(0.19)	(0.22)	(0.09)
TM gamma	-0.48	-0.16	-0.13	-0.14	-0.10	-0.23	0.13	0.13	0.19	0.21
	(0.08)	(0.51)	(0.51)	(0.45)	(0.55)	(0.51)	(0.68)	(0.61)	(0.44)	(0.33)
					Annual forecast					
Sharpe ratio	0.17	0.16	0.17	0.17	0.17	0.19	0.17	0.18	0.17	0.18
CAPM alpha	**0.33**	**0.23**	**0.24**	**0.21**	**0.22**	**0.51**	**0.35**	**0.38**	**0.31**	**0.32**
	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)
FF alpha	**0.21**	0.14	**0.16**	**0.14**	**0.16**	**0.29**	0.15	**0.21**	0.15	**0.20**
	(0.02)	(0.07)	(0.01)	(0.02)	(0.00)	(0.02)	(0.15)	(0.02)	(0.08)	(0.02)
FFC alpha	0.11	0.05	0.09	0.07	**0.11**	0.22	0.10	0.17	0.11	**0.16**
	(0.23)	(0.51)	(0.16)	(0.24)	(0.05)	(0.08)	(0.36)	(0.07)	(0.22)	(0.05)
TM gamma	0.22	0.44	**0.51**	0.41	**0.42**	-0.13	0.29	0.36	0.30	0.37
	(0.51)	(0.12)	(0.03)	(0.06)	(0.02)	(0.76)	(0.43)	(0.23)	(0.27)	(0.13)

Table 8: The performance measures of actively managed portfolios *in the presence of round-trip transaction costs of 0.5%*. S1 refers to a portfolio of firms that are in the smallest 10 percent with respect to market capitalization, S10 refers to a portfolio of firms that are in the largest 10 percent with respect to market capitalization. The active strategy, which is denoted by Si-S10 (where i takes values from 1 to 5), allocates funds for the next period to portfolio Si if the OOS forecasted return on the SMB factor is positive and otherwise allocates these funds to portfolio S10. CAPM alpha is the intercept in the standard one-factor market model. FF alpha is the intercept in the Fama-French three-factor model (which includes the SMB and HML factors). FFC alpha is the intercept in the Fama-French-Carhart four-factor model (which includes the SMB, HML, and MOM factors). The alphas are estimated from monthly data. The alpha values are provided in percentages, and the corresponding p-values are reported in brackets. TM gamma is the Treynor-Mazuy measure of the market timing ability, and the p-value of each gamma is reported in brackets. The bold text indicates values that are statistically significant at the 5% level.

Model	**1947-2010**	**1976-2010**
	Monthly forecast	
CAPM	**2.86** (0.01)	1.37 (0.23)
Fama-French	1.96 (0.08)	0.86 (0.51)
Fama-French-Carhart	**2.39** (0.04)	1.02 (0.41)
	Quarterly forecast	
CAPM	1.23 (0.29)	1.67 (0.14)
Fama-French	1.04 (0.39)	1.44 (0.21)
Fama-French-Carhart	1.54 (0.17)	1.65 (0.15)
	Annual forecast	
CAPM	**3.28** (0.01)	**3.98** (0.00)
Fama-French	**2.42** (0.03)	**2.90** (0.01)
Fama-French-Carhart	1.66 (0.14)	**2.41** (0.04)

Table 9: The results of the Gibbons, Ross, and Shanken (1989) test of the hypothesis that all of the intercepts of the actively managed portfolios *in the presence of the round-trip transaction costs of 0.5%* are jointly equal to zero. The table reports the values of the GRS test statistic, and the corresponding p-values are reported in brackets. The bold text indicates values that are statistically significant at the 5% level (which imply the rejection of the null hypothesis).

demonstrate that an active strategy that involves investing either in small stocks or large stocks depending on the forecasted sign of the size premium can deliver a superior performance. Notably, even in the presence of realistic transaction costs, an active strategy is able to generate a positive alpha that is highly economically and statistically significant. All of these findings allow us to conclude that the size premium is not an artifact of data mining but instead represents a real time-varying phenomenon. Moreover, the time variation in the size premium is related to economic conditions.

Acknowledgments

The author is grateful to the following individuals for their helpful comments and suggestions regarding earlier drafts of this paper: an anonymous referee; Øyvind Eitrheim; Jochen Jungeilges; Steen Koekebakker; and the participants in the 2011 European Conference of the Financial Management Association (Porto, Portugal), the 2012 Fagkonferansen i Bedriftsøkonomiske Emner (FIBE Conference, Bergen, Norway), the 2012 Forecasting Financial Markets Conference (Marseille, France), and the 18th International Business Research Conference (Las Vegas, USA). Any remaining errors in the manuscript are the author's responsibility.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Avramov, D., Jostova, G., Philipov, A., 2007. Understanding changes in corporate credit spreads. Financial Analysts Journal 63 (2), 90–105.

Banz, R. W., 1981. The relationship between return and market value of common stocks. Journal of Financial Economics 9, 3–18.

Berkowitz, S. A., Logue, D. E., Noser, E. A., 1988. The total costs of transactions on the NYSE. Journal of Finance 43 (1), 97–112.

Bernanke, B., Gertler, M., 1989. Agency costs, net worth, and business fluctuations. American Economic Review 79 (1), 14–31.

Black, F., 1993. Beta and return. Journal of Portfolio Management 20 (1), 8–18.

Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: What do we learn? Review of Financial Studies 12 (2), 405–428.

Campbell, J. Y., 1987. Stock returns and the term structure. Journal of Financial Economics 118, 373–399.

Campbell, J. Y., Shiller, R. J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1 (3), 195–227.

Chan, C., Chen, N.-F., Hsieh, D. A., 1985. An exploratory investigation of the firm size effect. Journal of Financial Economics 14, 451–471.

Chan, L. K. C., Lakonishok, J., 1993. Institutional trades and intraday stock price behavior. Journal of Financial Economics 33 (2), 173–199.

Chordia, T., Shivakumar, L., 2002. Momentum, business cycle, and time-varying expected returns. Journal of Finance 57 (2), 985–1019.

Cooper, M., Gulen, H., Vassalou, M., 2001. Investing in size and book-to-market portfolios using information about the macroeconomy: Some new trading rules. Working paper, Graduate School of Business, Columbia University and Krannert Graduate School of Management, Purdue University.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13 (3), 253–263.

Dimson, E., Marsh, P., 1999. Murphy's law and market anomalies. Journal of Portfolio Management 25, 53–69.

Estrella, A., Hardouvelis, G. A., 1991. The term structure as a predictor of real economic activity. Journal of Finance 46 (2), 555–576.

Fama, E. F., 1981. Stock returns, real activity, inflation, and money. American Economic Review 71 (4), 545–565.

Fama, E. F., 1990. Stock returns, expected returns, and real activity. Journal of Finance 45 (4), 1089–1108.

Fama, E. F., French, K. R., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–25.

Fama, E. F., French, K. R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25 (1), 23–49.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Fama, E. F., French, K. R., 2007. Disagreement, tastes, and asset prices. Journal of Financial Economics 83 (3), 667–689.

Fama, E. F., Schwert, G. W., 1977. Assets returns and inflation. Journal of Financial Economics 5, 115–146.

Ferson, W. E., Harvey, C. R., 1999. Conditioning variables and the cross section of stock returns. Journal of Finance 54 (4), 1325–1360.

Gertler, M., Gilchrist, S., 1994. Monetary policy, business cycles, and the behavior of small manufacturing firms. Quarterly Journal of Economics 109 (2), 309–340.

Geske, R., Roll, R., 1983. The monetary and fiscal linkage between stock returns and inflation. Journal of Finance 38 (1), 1–33.

Gibbons, M. R., Ross, S. A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57 (5), 1121–1152.

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

Handa, P., Kothari, S., Wasley, C., 1989. The relation between the return interval and betas: Implications for the size effect. Journal of Financial Economics 23, 79–100.

Hannan, E. J., Quinn, B. G., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society. Series B (Methodological) 41 (2), 190–195.

Henriksson, R. D., Merton, R. C., 1981. On market timing and investment performance part II: Statistical procedures for evaluating forecasting skills. Journal of Business 54 (4), 513–533.

Hou, K., Moskowitz, T. J., 2005. Market frictions, price delay, and the cross-section of expected returns. Review of Financial Studies 18 (3), 981–1020.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance 48 (1), 65–91.

Keim, D. B., Stambaugh, R. F., 1986. Predicting returns in the stock and bond markets. Journal of Financial Economics 17, 357–390.

Knez, P. J., Ready, M. J., 1996. Estimating the profits from trading strategies. Review of Financial Studies 9 (4), 1121–1163.

Kothari, S. P., Shanken, J., Sloan, R. G., 1995. Another look at the cross-section of expected stock returns. Journal of Finance 50 (1), 185–224.

Lee, B.-S., 1992. Causal relations among stock returns, interest rates, real activity, and inflation. Journal of Finance 47 (4), 1591–1603.

Lewellen, J., 2004. Predicting returns with financial ratios. Journal of Financial Economics 74 (2), 209–235.

Lo, A. W., MacKinlay, A. G., 1990a. Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies 3 (3), 431–467.

Lo, A. W., MacKinlay, A. G., 1990b. When are contrarian profits due to stock market overreaction? Review of Financial Studies 3 (2), 175–205.

Malkiel, B., 2003. The efficient market hypothesis and its critics. Journal of Economic Perspectives 17 (1), 59–82.

McCracken, M. W., 2007. Asymptotics for out of sample tests of Granger causality. Journal of Econometrics 140 (2), 719–752.

McQueen, G., Pinegar, M., Thorley, S., 1996. Delayed reaction to good news and the cross-autocorrelation of portfolio returns. Journal of Finance 51 (3), 889–919.

Pastor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111 (3), 642–685.

Perez-Quiros, G., Timmermann, A., 2000. Firm size and cyclical variations in stock returns. Journal of Finance 55 (3), 1229–1262.

Pesaran, M. H., Timmermann, A., 1992. Is value riskier than growth? Journal of Business & Economic Statistics 10 (4), 461–465.

Reinganum, M. R., 1981. Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. Journal of Financial Economics 9, 19–46.

Rosenberg, J. V., Maurer, S., 2008. Signal or noise? Implications of the term premium for recession forecasting. Economic Policy Review 14 (1), 1–11.

Schwarz, G. E., 1978. Estimating the dimension of a model. Annals of Statistics 6 (2), 461–464.

Shumway, T., Warther, V. A., 1999. The delisting bias in CRSP's Nasdaq data and its implications for the size effect. Journal of Finance 54 (6), 2361–2379.

Stoll, H. R., Whaley, R. E., 1983. Transaction costs and the small firm effect. Journal of Financial Economics 12, 57–79.

Treynor, J. L., Mazuy, K., 1966. Can mutual funds outguess the market? Harvard Business Review 44, 131–136.

Van Dijk, M. A., 2011. Is size dead? A review of the size effect in equity returns. Journal of Banking and Finance 35 (12), 3263–3274.

Vassalou, M., Xing, Y., 2004. Default risk in equity returns. Journal of Finance 59 (4), 831–868.

Yao, Y., 2012. Momentum, contrarian, and the January seasonality. Journal of Banking and Finance 36 (10), 2757–2769.