Book Section:
Zhang, H. and Su, C. (2020) A time-series bootstrapping simulation method to distinguish sell-side analysts' skill from luck. In: Lee, C.F. and Lee, J.C., (eds.) Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning. World Scientific Publishing, pp. 2011-2052. ISBN: 9789811202384.

https://doi.org/10.1142/9789811202391_0055

Chapter 55

A TIME-SERIES BOOTSTRAPPING SIMULATION METHOD TO DISTINGUISH SELL-SIDE ANALYSTS' SKILL FROM LUCK

CHAPTER OUTLINE:

Abstract

Data mining is quite common in econometric modeling when a given dataset is applied multiple times for the purpose of inference; it in turn could bias inference. Given the existence of data mining, it is likely that any reported investment performance is simply due to random chance (luck). This study develops a time-series bootstrapping simulation method to distinguish skill from luck in the investment process. Empirically, we find little evidence showing that investment strategies based on UK analyst recommendation revisions can generate statistically significant abnormal returns. Our rolling window based bootstrapping simulations confirm that the reported insignificant portfolio performance is due to sell-side analysts' lack of skill in making valuable stock recommendations, rather than their bad luck, irrespective of whether they work for more prestigious brokerage houses or not.

Keywords

data mining, time-series bootstrapping simulations, sell-side analysts, analyst recommendation revisions

55.1 INTRODUCTION

An interesting question of financial economics is whether the reported investment performance is as a result of random chance (luck), or data mining, which is considered to be a dangerous econometric practice that should be avoided (see, White, 2000; Barber, Lehavy, McNichols, and Trueman, 2001). Despite the nonexistence of a solid intrinsic relationship, extensive tests on a given set of sample frequently indicate some empirical results that seem to be superior, but actually spurious. Since the 1970s, a substantial number of methodological approaches have been proposed to test the spurious relationship generated by data mining in the context of specification searches, with particular attention being paid to the issue of inference (see, e.g., White, 2000). However, very few methods are applicable for examining the null hypothesis that the observed superior performance during a specification test has no predictive superiority over a given benchmark model.

Kosowski, Timmermann, Wermers, and White (2006) develop a cross-sectional bootstrapping simulation method on mutual funds research, which resamples the residuals from individual fund returns independently, but remains the effect of common risk factors *fixed* historically. Fama and French (2010, p. 1940), however, argue that "failure to account for the joint distribution of fund returns, and of fund and explanatory returns, biases the inferences of Kosowski

et al. (2006) toward positive performance". Extending Kosowski et al. (2006), Fama and French (2010) jointly resample both the residuals and risk factors, *ceteris paribus*. Inspired by Kosowski et al. (2006) and Fama and French (2010), this study develops a rolling window based bootstrapping simulation method in attempts to distinguish luck from skill in the investment process. Note that our method measures the performance distribution of the best performing rolling windows not only by resampling from the distribution of the *ex-post* best performing rolling windows, but using the information about luck represented by *all* rolling windows. This is a major difference between our method and those employed in previous studies, which generally ignore the possibility that luck distribution encountered by all other performance distributions also provides highly valuable and relevant information (see, e.g., White, 2000; Cuthbertson, Nitzsche, and O'Sullivan, 2008). Our rolling window based bootstrapping simulation method thus allows for a comprehensive investigation into the time-varying performance after explicitly controlling for luck and alleviating the potential bias from misspecification.

As an empirical demonstration, this study applies the time-series bootstrapping simulation method to investment strategies based on sell-side analyst recommendation revisions, though our method can be applied in evaluating other time-varying investment performance. Specifically, using a large sample of UK sell-side analyst recommendation revisions over the period January 1997 to June 2013, we construct a *long* portfolio, including all upgrades to buy-related recommendations from previous sell-related recommendations, as well as a *short* portfolio, including all downgrades to sell-related recommendations from previous buy-related recommendations. The time-varying performance of the *long* and *short* portfolios is measured by the intercept derived from the Fama and French (1993) three-factor model on a rolling window basis.

We find that the abnormal returns to the *long* and *short* portfolios are not statistically significant at the conventional level in any period of time, the results of which are robust to alternative asset pricing models, e.g., the single-factor capital assets pricing model (CAPM), the Carhart (1997) four-factor model, and the Fama and French (2015) five-factor model. Our time-series bootstrapping simulation methods further confirm that the reported insignificant portfolio performance is due to sell-side analysts' lack of skill in making valuable stock recommendation revisions, rather than their bad luck, no matter whether they work for more prestigious brokerage houses or not. From an investment perspective, it seems to be unlikely for market participants to make profits by purchasing (or short selling) stocks with upward (or downward) recommendation revisions in the UK.

The remainder of this chapter is organized as follows. The next section reviews two strands of related literature on various methods to address the problem of data mining and on sell-side analyst research. Section 55.3 describes data and methodology, while Section 55.4 presents empirical and simulated results. The final section concludes.

55.2 RELATED LITERATURE

55.2.1 An overview of various methods to address the problem of data mining

A number of tests have been conducted in assessing whether the reported investment performance is actually due to data mining, for example, (i) the out-of-sample persistence test, (ii) the conventional bootstrapping and Monte Carlo simulations, and (iii) the cross-sectional bootstrapping simulations developed by Kosowski et al. (2006) and Fama and French (2010), and so on.

55.2.1.1 Out-of-sample persistence tests

Numerous studies on portfolio performance evaluation account for luck by using the out-of-sample persistence test in the spirit of Carhart (1997). Carhart (1997) sorts mutual funds into the winner and loser portfolios based on the lagged one-year returns and examine the short-term persistence as shown in Lakonishok et al. (1992), Hendricks, Patel, and Zeckhauser (1993), Goetzmann and Ibbotson (1994), and Elton, Gruber, and Blake (1996). Carhart (1997) attributes the short-term persistence to price momentum (Jegadeesh and Titman, 1993), while little evidence is consistent with skilled or informed mutual fund managers. The rationale behind the out-of-sample persistence test is that, if investors really possess skill or information advantages over the market, the superior performance should be persistent and reflected in the out-of-sample period as well (see, also, Neely, Weller, and Ulrich, 2009); otherwise, the superior performance could disappear in the out-of-sample period, if investors are lucky. The out-of-sample persistence test generally rejects the hypothesis of data mining in the financial literature on the profitability of the simple technical trading rules in the stock market (see, Brock, Lakonishok, and LeBaron, 1992; Hudson, Dempsey, and Keasey, 1996) and in the foreign currency market (see, Okunev and White, 2003; Olson, 2004; Neely et al., 2009).

Although the out-of-sample persistence test is quite popular, it underestimates the likelihood that luck (either good luck or bad luck) can also be persistent, as the allocation of subsamples (such as the winner and loser portfolios) is largely based on noises (see, Kosowski et al., 2006; Fama and French, 2010). In addition, the subsamples may not be directly comparable, as the separation of the

whole sample is somewhat arbitrary and thus lacks the expected objectivity (see, Hsu and Kuan, 2005). Several theoretical hypotheses challenge the rationale of the out-of-sample persistence test. For example, Berk and Green (2004) make a theoretical argument that the performance persistence should not exist, that is, the historical performance should not predict the future performance. Focusing on rational investors who optimally invest among passive and active assets, Berk and Green (2004) argue that successful fund managers will capture abnormal returns by raising fees; alternatively, the size of their funds will increase and abnormal returns will disappear and deteriorate predictability, due to the diseconomies of scale, such as the higher transaction costs, or due to the needs of including the poorer-performing assets. Elton, Gruber, and Blake (2012), however, reject most of the predictions made by Berk and Green (2004). Specifically, they show that the best-performing funds reduce fees and the poor-performing funds raise fees, since management fee schedules generally decrease with the size of funds and the administrative costs have a high fixed component. Elton et al. (2012) document the performance persistence, that is, the historical performance predicts the future performance up to a three-year holding period. They also find that expense ratios are relatively lower for large funds and decrease as the funds become larger and/or perform better.

In the recent years, the literature on the performance persistence for hedge funds has emerged, mainly focusing on the US market where the industry is quite mature, while the results remain mixed and inconclusive. For example, using a sample of 399 offshore funds over the period 1989 to 1995, Brown, Goetzmann, and Ibbotson (1999) are among the first to report the positive risk-adjusted returns, though they find little evidence of the performance persistence in the raw returns or in the style-adjusted returns (see, also, Kat and Menexe, 2003). However, taking into account non-surviving hedge funds over the period 1995 to 1997, Kouwenberg (2003) provides clear evidence of the performance persistence (see, also, Harri and Brorsen, 2004). Agarwal and Naik (2001) test the performance persistence for hedge funds with the use of various measurement periods, i.e., the quarterly, semi-annual, and annual intervals, showing the maximum performance persistence at the quarterly intervals. Baquero, Horst, and Verbeek (2005) also find the performance persistence for a total of 1,797 hedge funds in the quarterly returns; further evidence shows that the performance can be persistent in the long-term period. Consistently, based on the monthly returns of 1,665 hedge funds, Edwards and Caglayan (2001) report that over a quarter of these funds generate positive excess returns, and, most notably, such significant performance persists over the one- and two-year horizons.

Mikhail, Walther, and Willis (2004) show that sell-side analysts exhibit relative persistence in their stock recommendations, that is, sell-side analysts making more (less) profitable stock recommendations in the past tend to make more (less) profitable stock recommendations in the

future. In particular, the market recognizes the performance persistence in the five days surrounding sell-side analyst recommendation revisions. However, the market reaction is incomplete, as trading strategies taking the long (short) positions in upgrades (downgrades) based on sell-side analysts' prior performance become unprofitable after accounting for transaction costs and trading delays. In contrast, Li (2005) documents that sell-side analysts with the above-median risk adjusted performance in the estimation period (i.e., the past winners) consistently outperform their counterparts with below-median performance (i.e., the past losers) in the subsequent holding periods. Specifically, the annualized risk adjusted returns of trading strategies based on the performance persistence are statistically and economically significant, even after taking transaction costs and trading delays into account. For example, the annualized abnormal returns, with an approximate magnitude of 10%, are substantially higher than those typically reported in mutual funds research, presenting a clear violation against the semi-strong form of market efficiency. The significantly positive relationship between the historical performance and the future performance is also supported by Loh and Mian (2006), Ertimur, Sunder, and Sunder (2007), Brown and Huang (2013), and Fang and Yasuda (2014).

55.2.1.2 The conventional bootstrapping and Monte Carlo simulations

Although numerous empirical studies support the predictive power of technical trading rules (see, e.g., Brown, Goetzmann, and Kumar, 1998; Lo, Mamaysky, and Wang, 2000; Savin, Weller, and Zvingelis, 2007; Hsu, Hsu, and Kuan, 2010; among others), such predictive power is likely as a result of data mining, that is, the observed any satisfactory results may simply be obtained by random chance (luck). Sullivan, Timmermann, and White (1999, p. 1648) argue that data mining can "result from a subtle survivorship bias operating on the entire universe of technical trading rules that have been considered historically." White (2000) employs a straightforward procedure to test the null hypothesis that there is no a superior strategy (or model specification) in a "universe" of strategies (or model specifications). However, it is cumbersome to test the hypothesis when the "universe" of strategies is large. White (2000) thus proposes a reality check test to adjust the significance level of the estimated *t*-statistics in the face of possible data mining (see, also, Sullivan et al., 1999). By setting the strategy of zero return at all times as the benchmark, White (2000) suggests the use of the block resampling estimator of Politis and Romano (1994), either the stationary bootstrapping simulation or Monte Carlo simulation, to estimate the *p*-value of a given trading strategy (or model specification).

Brock et al. (1992) conduct a comprehensive study using the daily stock prices over a 90-year sample period, showing that a total of 26 technical trading rules applied to the Dow Jones Industrial Average (DJIA) Index significantly beat the benchmark. Sullivan et al. (1999) examine

the results of Brock et al. (1992) by applying a bootstrapping reality check over the same sample period, confirming that the results are not due to data mining. However, the "universe" of trading strategies in Sullivan et al. (1999) is far from complete, as some well-known strategy classes are not included. Moreover, the distribution of White's (2000) reality check is based on the least favorable configuration, which is referred to as the point least favorable to the alternative (Hsu and Kuan, 2005). Hansen (2005) finds that the power of White's (2000) reality check test can be driven to zero by adding sufficient poor or irrelevant trading rules to the "universe" being tested; also, White's (2000) reality check does not use the standardized *t*-statistics. To overcome these limitations, Hansen (2005) develops a more powerful test for the superior predictive ability (SPA) by using a different way to bootstrap the distribution of the standardized returns. Using Hansen's (2005) SPA test, Hansen, Lunde, and Nason (2005) find significant calendar effect, which is against the results of Sullivan, Timmermann, and White (2001) based on White's (2000) reality check test. Hsu and Kuan (2005) investigate the profitability of technical trading rules using both White's (2000) reality check test and Hansen's (2005) SPA test. They report significantly profitable simple and complex trading strategies in the sample from relatively *young* indices, such as the NASDAQ Composite Index and the Russell 2000 Index, but not in the sample from more *mature* indices, such as the DJIA Index and the Standard & Poor's (S&P) 500 Index. Using White's (2000) reality check test, Qi and Wu (2006) examine 2,127 technical trading rules in the foreign exchange market and find that data mining biases do not change the profitability and statistical significance for the period April 2, 1973 to December 31, 1998.

Romano and Wolf (2005) introduce a reality check based stepwise multiple test to identify the significant models that outperform the benchmark. Hsu et al. (2010) extend Hansen's (2005) SPA test to a stepwise SPA test in attempts to identify the predictive models in the large-scale and multiple testing problems. They find that technical trading rules have significant predictive ability prior to the inception of exchange traded funds. As stated by Hansen (2005), the SPA test could be improved if there is a reliable way to incorporate information about the off-diagonal elements of the covariance matrix. Cai, Jiang, Zhang, and Zhang (2015) propose a new method to test the SPA of a benchmark model against a large group of alternative models, in attempts to reduce the potential data mining bias. Specifically, they model the covariance matrix by factor models and develop a generalized likelihood ratio (GLR) *t*-statistic. In various scenarios, it is shown that the GLR test has asymptotic null distribution independent of nuisance parameters. Also, the GLR test is asymptotically optimal in the sense that it achieves the optimal rate of convergence in the context of semi- and nonparametric settings (see, Jiang, Zhou, Jiang, and Peng, 2007). The Monte Carlo bootstrapping simulation results show that the GLR test is much more powerful and less conservative than Hansen's (2005) SPA test, possibly because the SPA test has

a non-unique null distribution depending on nuisance parameters. The GLR test is also extended to a stepwise GLR (step-GLR) test in the spirit of Romano and Wolf's (2005) step-reality check test and Hsu et al.'s (2010) step-SPA test, allowing to sequentially identify the models that are superior to the benchmark. The modeling of the covariance matrix is semiparametric in nature, as it does not require distributional assumptions. The step-GLR test can identify the most contributed predictive models to the rejection of the null hypothesis.

However, data mining becomes much more serious in the more recent period, as the reality check test examines investment performance by resampling from the distribution of a given sample only, without using the information about *luck* represented by all other return samples. Hence, White's (2000) reality check test discounts the possibility that luck distribution encountered by all other performance distributions also contains highly useful and relevant information (Cuthbertson et al., 2008).

55.2.1.3 The alternative bootstrapping simulations

Kosowski et al. (2006) develop a cross-sectional bootstrapping simulation method on mutual funds research, which resamples the residuals from a large number of time-series individual fund returns independently, but remains the effect of common risk factors *fixed* historically. Fama and French (2010), however, jointly resample both of them, as Kosowski et al. (2006) fail to account for the joint distribution of fund returns, and of fund and explanatory returns, and inevitably bias the inference toward positive performance. Both simulation methods do not impose an assumption that returns follow any specific parametric distribution, nor do they depend on the large sample asymptotic theory; the inference allows for non-normality in the idiosyncratic risk of returns (Cuthbertson et al., 2008; Sørensen 2009). Specifically, examining 1,788 mutual funds over the period January 1975 to December 2002, Kosowski et al. (2006) find that a sizable minority of fund managers have superior ability in selecting stocks even netting of all expenses and transaction costs. Cuthbertson et al. (2008) apply the bootstrapping simulation method of Kosowski et al. (2006) to 935 UK mutual funds from April 1975 to December 2002, and find similar evidence of superior stock picking ability among a small number of the best performing fund manangers. In contrast, using an alternative bootstrapping simulation method, Fama and French (2010) examine 5,238 mutual funds over the period January 1984 to September 2006, showing little evidence of superior skill for fund manangers.

However, the empirical results of Kosowski et al. (2006) and Fama and French (2010) are not directly comparable, given three main differences, i.e., (i) the different fund-inclusion criteria, (ii) the different sample periods, and (iii) the different bootstraping simulation methods—Kosowski

et al. (2006) simulate fund returns and factor returns independently, while Fama and French (2010) simulate these returns jointly. Blake, Caulfield, Ioannidis, and Tonks (2017) apply both bootstrapping simulation methods of Kosowski et al. (2006) and Fama and French (2010) to the same sample of UK mutual funds over the period January 1998 to September 2008. Given that the stock picking skills of fund managers evaluated by Jensen's (1968) alpha could be biased in the presence of market timing skills (see, Treynor and Mazuy, 1966), Blake et al. (2017) incorporate the total performance measure of Grinblatt and Titman (1994)—the sum of Jensen's (1968) alpha and market-timing coefficients—into the bootstrapping simulation methods. Blake et al. (2017) find that the evaluation of investment performance relies heavily on the bootstrapping simulation methods employed. Specifically, the simulation method of Kosowski et al. (2006) is more likely to find superior skill than that of Fama and French (2010). On the contrary, Gallefoss, Hansen, Haukaas, and Molnár (2015) argue that the bootstrapping simulation methods of Kosowski et al. (2006) and Fama and French (2010) generate basically the same results (see, also, Su, Zhang, Bangassa, and Joseph, 2018). In order to enhance the precision of performance predictability, Kosowski, Naik, and Teo (2007) incorporate the seemingly unrelated assets Bayesian approach of Pástor and Stambaugh (2002), which is robust to model misspecification, into the bootstrapping simulation method of Kosowski et al. (2006). More importantly, it takes advantage of information in the seemingly unrelated assets to overcome the short sample period problems and to improve the precision of the performance estimates. They find that the top hedge fund managers possess certain asset selection skill and can take advantage of the simple trading strategies, the results of which challenge the classical point of view that the top hedge funds are just lucky and their performance could not be persistent.

Blake, Caulfield, Ioannidis, and Tonks (2014) argue that the non-parametric bootstrapping simulation methods of Kosowski et al. (2006) and Fama and French (2010) are flawed as the returns are drawn from a uniform distribution.[1] To alleviate the limitations of the non-parametric bootstrapping simulations, Blake et al., (2014) suggest the application of the parametric bootstrapping simulations, in which the time-series returns for each fund are resampled from a stable distribution that shows the distributional properties of the realized returns over the whole sample period. For example, to identify the underlying distribution of alphas, Meyer, Schmoltzi, Stammschulte, Kaesler, Loos, and Hackethal (2012) draw a new alpha that is added to the return series for each portfolio and for each simulation run. The distribution of alphas is drawn from pre-specified distribution, such as normal distributions with negative means and slightly skewed normal distributions with a bit of fat tails. Before adding the drawn alphas to the return series,

[1] Specifically, the non-parametric bootstrapping simulation methods give excessive weights to observations in the tail of the true, but unknown distribution and will underestimate the probability of appearing in the center of the distribution. Moreover, it ignores the skewness and kurtosis in the underlying fund returns data.

Meyer et al. (2012) rescale alphas by the ratio of the residual standard errors of individual portfolio to the average standard error of all portfolios. This rescaling process accounts for the issue that more diversified portfolios are less able to produce extreme alphas than less diversified portfolios. It also captures the fact that portfolios with the same level of diversification have different probabilities to generate large alphas, according to their risk. However, a potential problem of the parametric bootstrapping simulation methods is that any cross-sectional dependence is lost (see, Kapetanios, 2008). Moreover, Cogneau and Zakamouline (2013) suggest that the widely used block bootstrapping simulation methods are generally biased. An improper application of the block bootstrapping simulation method could underestimate the risk of a portfolio with the independent time-varying returns and overestimate the risk of a portfolio with the mean-reverting returns.

Blake et al. (2014) introduce two new simulation methods allowing for appropriate inference in the presence of the non-normal asset returns when evaluating the performance of mutual funds against the benchmark model. Specifically, one simulation method recognizes the panel nature of the dataset and the existence of both fund and time effects, while the other one takes into account the non-normal distribution of individual mutual fund returns. Blake et al. (2014) find little evidence that UK fund managers can generate superior abnormal returns when the non-normality of fund returns using a series of both parametric and non-parametric bootstrapping simulations is allowed.

55.2.1.4 Other simulation methods

Barras, Scaillet, and Wermers (2010) propose an alternative model to distinguish skilled (unskilled) funds from lucky (unlucky) funds. Specifically, they construct three groups of funds, i.e., skilled funds, zero-alpha funds, and unskilled funds, and then use the false discovery rate approach proposed by Storey (2002) to study the distribution of *p*-values on the sampled *t*-statistics. Novy-Marx and Velikov (2016) suggest the use of before- and after-transaction cost performance of financial anomalies to examine data mining. Transaction costs significantly reduce strategy profitability, increasing data mining concerns, in line with Barber et al. (2001) and Timmermann and Granger (2004).

Fama and French (1993) and Carhart (1997) suggest testing market anomalies with the use of the sophisticated asset pricing models. Hou, Xue, and Zhang (2014) argue that nearly half of 80 financial anomalies earn insignificant and average returns for the high-minus-low (HML) deciles constructed with the New York Stock Exchange (NYSE) breakpoints and value-weighted returns, consistent with McLean and Pontiff (2014). Hou, Xue, and Zhang (2017) report that over 100 out

of 161 financial anomalies can be explained by the *q*-factor model of Hou et al. (2014), though they fail to account for transaction costs (Novy-Marx and Velikov, 2016).

Overall, data mining has gained rising attention in the recent literature, especially in the buy-side investment industry, such as mutual funds. However, none of the studies applies the Kosowski et al. (2006) and Fama and French (2010) bootstrapping simulations to a single time-series portfolio returns, to the best of our knowledge. Furthermore, it remains understudied whether the performance of sell-side analysts are lucky or skilful.

55.2.2 A brief review of sell-side analyst research

Sell-side analysts typically work for brokerage houses (or investment banks), collecting and analyzing a variety of market, industry, and firm-specific information and then making stock recommendations. These analyst recommendations are disseminated through television appearances or through other electronic and print media, and could be used by investors in their investment decisions. Whether sell-side analyst recommendations can truly create investment value and promote market efficiency has been of great interest to financial academics. The pioneering study of Cowles (1933) concludes that investors are not able to add value to the market when they follow analyst recommendations, confirmed by numerous early studies (see, e.g., Colker, 1963; Groth, Lewellen, Schlarbaum, and Lease, 1979). In contrast, two recent studies of Stickel (1995) and Womack (1996) report that upgrades (downgrades), i.e., favorable (unfavorable) changes in analyst recommendations, are accompanied by significantly positive (negative) returns at the time of their announcements.[2] Barber et al. (2001), Jegadeesh, Kim, Krische, and Lee (2004), Boni and Womack (2006), and Green (2006) also show the existence of profitable investment strategies based on publicly available analyst recommendations, apparently against the semi-strong form of market efficiency. However, Barber et al. (2001) argue that these investment strategies require a great deal of trading and generate considerable transaction costs, suggesting that, although market inefficiencies exist, they are not easily exploitable by investors (see, also, Mikhail et al., 2004; Hall and Tacon, 2010). Altınkılıç and Hansen (2009) and Altınkılıç, Balashov, and Hansen (2013) further call into question the information role played by financial analysts in that their stock

[2] The recent literature on sell-side analyst research shows that the investment value of analyst recommendations is significantly related to the frequency of analyst recommendation revisions (see, Hobbs, Kovacs, and Sharma, 2012), the conflict of interests (see, Mehran and Stulz, 2007; Shen and Chih, 2009; Guan, Lu, and Wong, 2012), the reputation of sell-side analysts (see, Emery and Li, 2009; Fang and Yasuda, 2009 & 2014; Meng, 2015), momentum of analyst recommendations (see, Muslu and Xue, 2013), timing of reaction to analyst recommendations (see, Ivkovic and Jegadeesh, 2004; Green, 2006; Irvine, Lipson, and Puckett, 2007), herding (see, Trueman, 1994; Jegadeesh and Kim, 2010), industry (see, Jegadeesh and Kim, 2006), time stamps reported in the database for analyst recommendations (see, Bradley, Clarke, Lee, and Ornthanalai, 2014), sell-side analysts' industry experience (see, Bradley, Gokkaya, and Liu, 2016), and so on.

recommendation revisions often piggyback on public information (e.g., corporate events and news), and thus provide investors with little incremental information.

The value of sell-side analyst recommendations has been extensively studied in the US, while very few studies have been conducted in other markets. Jegadeesh and Kim (2006) point out that an in-depth examination in other developed markets will provide us with a more comprehensive picture. They report international evidence that stock prices react significantly to analyst recommendation revisions in the Group of Seven (G7) industrialized countries, except for Italy. In addition, Dimson and Fraletti (1986) examine an unpublished sample of 1,649 telephone recommendations made by a leading UK brokerage house in 1983, but they find no significant abnormal returns for the recommended stocks. Ryan and Taffler (2006; p. 372) argue that the study of Dimson and Fraletti (1986) examines analyst recommendations made by "a single UK brokerage house only" and could be biased towards large size stocks. Ryan and Taffler (2006) investigate a sample of 2,506 changes in analyst recommendations made by six London based brokerage houses over the period December 1993 to June 1995, showing that stock prices are significantly influenced by analyst recommendation revisions. Su et al. (2018) argue that the two prior UK studies with the use of the limited observations and short sample periods examined generally suffer from small sample bias. Su et al. (2018) conducts a comprehensive investigation into the investment value of sell-side analyst recommendation revisions in the UK, showing that, on average, upgrades fail to generate any significantly positive abnormal returns in any period of time, even before transaction costs. In addition, although downgrades could generate significantly negative abnormal gross returns over some periods of time, these observed significant returns disappear after accounting for transaction costs. However, an industry-based analysis shows that, within two high-tech industry sectors, i.e., Health Care and Technology sectors, sell-side analysts possess certain skill in making valuable downgrades over some periods of time and, in particular, such skill is sufficient to offset transaction costs.

55.3 Data and Methodology

55.3.1 Data and sample description

We obtain the real-time UK sell-side analyst recommendations from *the Morningstar Extracted Data File: Historic Broker Recommendations for UK Registered and UK Listed Companies*, originally created by *Hemscott Company Guru*, now part of *Morningstar Company Intelligence*. Each analyst recommendation record contains information on the name of the recommended stock, the name of the brokerage house issuing the recommendation, the

recommendation starting and expiration dates, as well as a rating between 1 and 9.[3] A rating of 1 reflects a strong buy; 2, a buy; 3, a weak buy; 4, a weak buy/hold; 5, a hold; 6, a hold/sell; 7, a weak sell; 8, a sell; and 9, a strong sell.

We exclude all stock recommendations that omit the name of brokerage houses and/or contain data errors. Also, we require that: (i) the gap between the starting and expiration dates of each recommendation is less than one year to ensure that the brokerage house actively follows the recommended stock (see, also., Jegadeesh and Kim, 2006; Barber, Lehavy, and Trueman, 2007); and (ii) the relevant financial and accounting data of the recommended stocks are available from the London Share Price Database (LSPD). The final sample is comprised of 294,692 stock recommendations made by 122 brokerage houses on 2,409 distinct firms listed either on the London Stock Exchange (LSE) or on the Alternative Investment Market (AIM) over the period January 1997 to June 2013. Our sample includes 89,014 stock recommendations on 1,346 dead firms to avoid the potential survivorship bias (see more details on descriptive statistics on our final sample in Appendix 55A).

55.3.2 Research design

55.3.2.1 Portfolio construction

We construct two portfolios: (i) a *long* portfolio, consisting of all upgrades to strong buy, buy, weak buy, or weak buy/hold from previous strong sell, sell, weak sell, hold/sell, or hold; and (ii) a *short* portfolio, consisting of all downgrades to strong sell, sell, weak sell, hold/sell, or hold from previous strong buy, buy, weak buy, or weak buy/hold.[4] For each analyst recommendation revision, the recommended stock enters the *long* portfolio at the close of trading on the day the revision is released. If a revision is released on a non-trading day, the recommended stock is added into the *long* portfolio at the close of the next trading day. If a stock is recommended by more than one brokerage house on a given date, then that stock will appear multiple times in the portfolio on that date, once for each brokerage house. The portfolio is updated daily, so the recommended stock

[3] Unlike Su et al. (2018) that reclassify all original analyst recommendations into five categories: Strong Buys, Buys, Holds, Sells, and Strong Sells, this study remains the nine categories, as the purpose of this study is to test the proposed bootstrapping simulation method rather than compare the investment value of analyst recommendations in the UK with studies in other markets.

[4] The *short* portfolio does not include downgrades from strong buy to buy, from strong buy to weak buy, from strong buy to weak buy/hold, from buy to weak buy, from buy to weak buy/hold, or from weak buy to weak buy/hold, as they can also be interpreted as positive recommendations, while the *long* portfolio does not include upgrades from strong sell to sell, from strong sell to weak sell, from strong sell to weak hold/sell, from strong sell to hold, from sell to weak sell, from sell to weak hold/sell, from sell to hold, from weak sell to weak hold/sell, from weak sell to hold, or from weak hold/sell to hold, as they can also be interpreted as negative recommendations (Stickel, 1995).

is removed from the portfolio at the close of the trading on the recommendation expiration date. The *short* portfolio is constructed in an anomalous daily fashion.

Table 55.1 Descriptive statistics on UK sell-side analyst recommendations in the *long* and *short* portfolios

	The *long* portfolio				**The *short* portfolio**			
Year	**No. of recommended firms**	**No. of brokerage houses**	**Average rating**	**No. of recommendations**	**No. of recommended firms**	**No. of brokerage houses**	**Average rating**	**No. of recommendations**
1997	889	45	2.15	12,546	736	43	5.45	10,500
1998	940	47	2.19	12,758	752	42	5.55	10,207
1999	904	45	2.20	11,822	713	39	5.54	8,328
2000	849	50	2.19	10,412	622	44	5.43	6,330
2001	844	45	2.21	7,715	680	41	5.65	7,205
2002	822	43	2.19	7,311	647	41	5.80	6,319
2003	771	47	2.20	8,964	649	41	5.78	7,333
2004	800	49	2.20	10,537	638	47	5.70	8,383
2005	883	49	2.19	10,087	682	48	5.74	9,363
2006	921	54	2.15	10,715	665	47	5.72	7,751
2007	903	49	2.14	8,574	620	41	5.67	5,556
2008	893	50	2.11	9,065	586	47	5.84	5,844
2009	804	51	2.12	9,593	564	51	5.79	6,527
2010	808	40	2.08	8,651	451	37	5.60	4,237
2011	769	39	2.10	8,456	450	39	5.65	4,453
2012	763	38	2.11	7,151	440	32	5.61	3,867
2013 (January-June)	539	31	2.11	2,598	340	26	5.66	1,773
Full sample	**2,341**	**118**	**2.16**	**156,955**	**1,931**	**110**	**5.65**	**113,976**

Notes:

This table presents descriptive statistics on all UK sell-side analyst recommendations in the *long* and *short* portfolios in each year over the period January 1997 to June 2013. All real-time analyst recommendations are obtained from *Morningstar Company Intelligence*, including information on the name of the firm recommended, the name of the brokerage house issuing the recommendation, the recommendation starting date and expiration date, and a rating between 1 and 9. A rating of 1 reflects a strong buy, 2 a buy, 3 a weak buy, 4 a weak buy/hold, 5 a hold, 6 a hold/sell, 7 a weak sell, 8 a sell, and 9 a strong sell. A *long* portfolio includes all upgrades to strong buy, buy, weak buy, or weak buy/hold from previous strong sell, sell, weak sell, hold/sell, or hold, while a *short* portfolio includes all downgrades to strong sell, sell, weak sell, hold/sell, or hold from previous strong buy, buy, weak buy, or weak buy/hold.

Table 55.1 presents descriptive statistics on analyst recommendation revisions in the *long* and *short* portfolios. The *long* portfolio includes 156,955 analyst recommendation revisions for 2,341 stocks, while the *short* portfolio includes 113,976 analyst recommendation revisions for 1,931 stocks. So, the number of analyst recommendation revisions in the *long* portfolio is 37.71% (156,955/113,976 –1) more than that in the *short* portfolio, in line with the argument that sell-side analysts tend to provide more coverage on positive recommendations (see, e.g., Stickel, 1995; Barber et al., 2001; Boni and Womack, 2006).

55.3.2.2 Portfolio performance evaluation

Like Barber et al. (2007), we assume an equal dollar investment in each revision, the return on portfolio p at date t, $R_{p,t}$, is:

$$R_{p,t} = \left(\sum_{i=1}^{n_t} x_{i,t} \times R_{i,t}\right) / \sum_{i=1}^{n_t} x_{i,t}, \tag{55.1}$$

where n_t represents the number of analyst recommendation revisions in the portfolio at the close of trading of the recommendation date through date t –1; $x_{i,t}$ represents the compounded daily return for the recommended stock i from the closing of trading on the recommendation date through date t –1 ($x_{i,t}$ = 1 for a stock recommended on date t –1); $R_{i,t}$ represents the daily return for the recommended stock i on date t.

The abnormal return is estimated using the intercept term, α_p, derived from the Fama and French (1993) three-factor model:[5]

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_p\left(R_{m,t} - R_{f,t}\right) + s_p SMB_t + h_p HML_t + \varepsilon_{p,t}, \tag{55.2}$$

where $R_{p,t}$ and $R_{m,t}$ are the daily return on the portfolio and the FTSE All-Share Index, respectively; $R_{f,t}$ represents the three-month UK T-bill rate; SMB_t and HML_t represent the daily returns on zero-investment factor-mimicking portfolios for size and book-to-market (B/M), respectively; $\varepsilon_{p,t}$ represents the error term. We obtain the daily ($R_{m,t} - R_{f,t}$), SMB_t, and HML_t in the UK stock market from the Xfi Centre for Finance and Investment at the University of Exeter (see, Gregory, Tharyan, and Christidis, 2013).

We estimate Eq. (2) on a rolling window basis, i.e., a one-year window length rolling one trading day forward, to track the performance of the underlying variables over time. For example, the first rolling window is from January 2, 1997 to December 31, 1997, which covers 253 trading

[5] We explicitly exclude the returns on the first trading day following the recommendation as many investors, particularly small investors, tend to react to information contained in analyst recommendation revisions with a delay (Barber et al., 2001).

days. The chosen length of a rolling window is one year in this study as average investors generally evaluate their portfolios on an annual basis (see, Benartzi and Thaler, 1995).Then the return for the first trading day is dropped and the return for a new trading day is added, meaning that the return for January 2, 1997 is dropped and the return for January 2, 1998 is added. This process continues so that we estimate a total of 3,912 rolling windows over the whole sample period. In each rolling window, the three-factor regression yields parameter estimates of α_p, β_p, s_p, and h_p; a statistically positive (negative) α_p indicates that the *long* (*short*) portfolio is profitable after controlling for the risk factors of market, size, and value. This calculation, therefore, generates a time-series of daily abnormal returns for the *long* or *short* portfolio.

55.3.3 Time-series bootstrapping simulation methods

To study whether the time-varying abnormal returns to the *long* and *short* portfolios are due to random chance (i.e., sell-side analysts' luck), we propose a time-series bootstrapping simulation method in the spirit of Kosowski et al. (2006) and Fama and French (2010). In particular, we employ two bootstrapping procedures, i.e., (i) Procedure I: Resampling the residuals independently with fixed common risk factors; and (ii) Procedure II: Jointly resampling both the residuals and common risk factors.

55.3.3.1 Procedure I: Resampling the residuals independently with fixed common risk factors

Like Kosowski et al. (2006), our first rolling window based bootstrapping simulation resamples the residuals from a time-series of returns independently, keeps the order of common risk factors *fixed*, and rolls the procedure forwards throughout a large number of windows. In this section, we demonstrate the time-series bootstrapping simulation method with the Fama and French (1993) three-factor model, but the application of the bootstrapping procedure to other asset pricing models, such as the CAPM, the Carhart (1997) four-factor model, and the Fama and French (2015) five-factor model, is very similar, with the only modification of the steps being the substitution of appropriate models.

Specifically, the time-series bootstrapping simulation method includes five major steps as follows:

(i) We estimate the Fama and French (1993) three-factor model to calculate the estimated alphas, factor loadings, and residuals using the time-series of daily excess returns for the portfolio $\{(R_t - R_{f,t}), t = T_1, \dots, T_{253}\}$, where T_1 and T_{253} are the first and last trading dates, respectively, in each rolling window:

$$R_t - R_{f,t} = \hat{\alpha} + \hat{\beta}\left(R_{m,t} - R_{f,t}\right) + \hat{s}SMB_t + \hat{h}HML_t + \hat{\varepsilon}_t. \quad (55.3)$$

(ii) We save the coefficient estimates, $\{\hat{\alpha}_p, \hat{\beta}_p, \hat{s}_p, \hat{h}_p\}$, the time-series of estimated residuals, $\{\hat{\varepsilon}_t = T_1, \dots, T_{253}\}$, and the *t*-statistic of alpha, $\hat{t}_{\hat{\alpha}}$.

(iii) We generate a time-series of pseudo residuals $\{\hat{\varepsilon}_{t_b}^b = T_1^b, \dots, T_{253}^b\}$ by randomly drawing the residuals from the saved residual vector $\{\hat{\varepsilon}_t\}$ with replacements, where b is the bootstrapping simulation index. It is important to reiterate that only the residuals are reordered, while the common risk factors are not resampled, in line with the Kosowski et al. (2006).

(iv) We produce a time-series of pseudo daily excess returns $\left(R_t - R_{f,t}\right)^b$ in the rolling window, imposing the null hypothesis of zero true performance ($\alpha = 0$):

$$\left\{\left(R_t - R_{f,t}\right)^b = 0 + \hat{\beta}\left(R_{m,t_b} - R_{f,t_b}\right)^b + \hat{s}SMB_{t_b}^b + \hat{h}HML_{t_b}^b + \hat{\varepsilon}_{t_b}^b\right\}, \quad (55.4)$$

where $t = T_1, \dots, T_{253};\ t_b = T_1^b, \dots, T_{253}^b$

(v) For the bootstrapping simulation index $b = 1$, we regress the pseudo daily excess returns $\left(R_t - R_{f,t}\right)^b$ on the three factors:

$$\left(R_t - R_{f,t}\right)^b = \hat{\alpha}^b + \hat{\beta}\left(R_{m,t} - R_{f,t}\right) + \hat{s}SMB_t + \hat{h}HML_t + \hat{\varepsilon}_t. \quad (55.5)$$

The simulated $\hat{\alpha}^b$ represents the sampling variation around zero true performance, purely due to random chance (luck). Repeating the above procedures forward by one observation each time, we yield a time-series of simulated alphas, $\{\hat{\alpha}_{w,}^b\ w = 1, \cdots, 3{,}912\}$, and their corresponding *t*-statistics, $\{\hat{t}_{\alpha_w}^b, w = 1, \dots, 3{,}912\}$, where w is the number of rolling windows throughout the full sample period January 1998 to June 2013. We then sort all simulated $\hat{\alpha}_w^b$ into a cumulative distribution function (CDF) of simulated $\hat{\alpha}_w^b$, $f(\hat{\alpha}_w^b)$, a separate time-series of *luck* distribution from the worst performing rolling window to the best performing rolling window. We repeat the above bootstrapping simulation a large number of times, say, b = 1, …, 10,000, generating a similar time-series distribution of bootstrapped *t*-statistics $\{\hat{t}_{\alpha_w}^b, w = 1, \dots, 3{,}912; b = 1, \dots, 10{,}000\}$, which can be compared with the distribution of the actual distribution $\{\hat{t}_{\alpha_w}, w = 1, \dots, 3{,}912\}$, once both sets of *t*-statistics have been resorted from the lowest value ($\hat{t}_{\alpha_{min}}$) to the highest value ($\hat{t}_{\alpha_{max}}$).

We compare the *t*-statistics rather than the alphas simply because the use of the *t*-statistics controls for differences in risk-taking across subsamples. The rationale behind the bootstrapping

simulations is to investigate the number of subsamples that one might expect to achieve a given level of abnormal returns by random chance alone and then compare this with the number of subsamples that actually achieve this level of abnormal returns in the 'real world'. For the outperforming (underperforming) subsamples measured by *t*-statistics of alpha, if the simulated $\hat{t}^{b}_{\alpha_{max}}$ is greater than the actual $\hat{t}_{\alpha_{max}}$ in less than 5% of the 10,000 simulations, at any given performance order, we reject the null hypothesis that the outperforming (underperforming) subsample is due to good luck (poor stock picking skill) at the 95% confidence level and infer that the strategy is genuine (bad luck) (see, Cuthbertson et al., 2008; Blake et al., 2017). The same rule can be applied for any other point in the CDF.

Like Meyer et al. (2012), we do not account for autocorrelation for two reasons. First, the majority of rolling window regressions do not report autocorrelation at the conventional significance level using the Breusch-Godfrey test. Second, it has the advantage of enhancing comparability between simulated and actual t-statistics through a uniform test specification. This is because the bootstrap simulations consist of random drawings of individual daily returns with replacements, which means the time series drawn cannot contain any true underlying autocorrelation by design. We thus employ the robust standard errors that control for the autoregressive conditional heteroscedastic (ARCH) effect (see Appendix 55B).

55.3.3.2 Procedure II: Jointly resampling both the residuals and common risk factors

Motivated by Fama and French (2010), we repeat Procedure I by jointly resampling both the residuals and common risk factors in Step (iii), *ceteris paribus*. Specifically, in Step (iii), we generate two time-series of pseudo residuals $\left\{\hat{\varepsilon}^{b}_{t} = T^{b}_{1}, \dots, T^{b}_{252}\right\}$ and risk factors $\left\{\left(R_{m,t} - R_{f,\mathrm{t}}\right)^{b}, SMB^{b}_{t}, HML^{b}_{t}\right\}$ jointly by randomly drawing residuals and risk factors from the original residual vector $\{\hat{\varepsilon}_t\}$ and risk factor vector $\left\{\left(R_{m,t} - R_{f,t}\right), SMB_t, HML_t\right\}$ with replacements, respectively. The major difference between the two bootstrapping procedures is that Procedure II considers the distribution of the residuals conditional on the realization of the systematic risk factors (see, Fama and French, 2010), while Procedure II employs the unconditional distribution of the residuals and assumes the influence of the common risk factors is not fixed at their historical realizations (see, Kosowski et al., 2006).

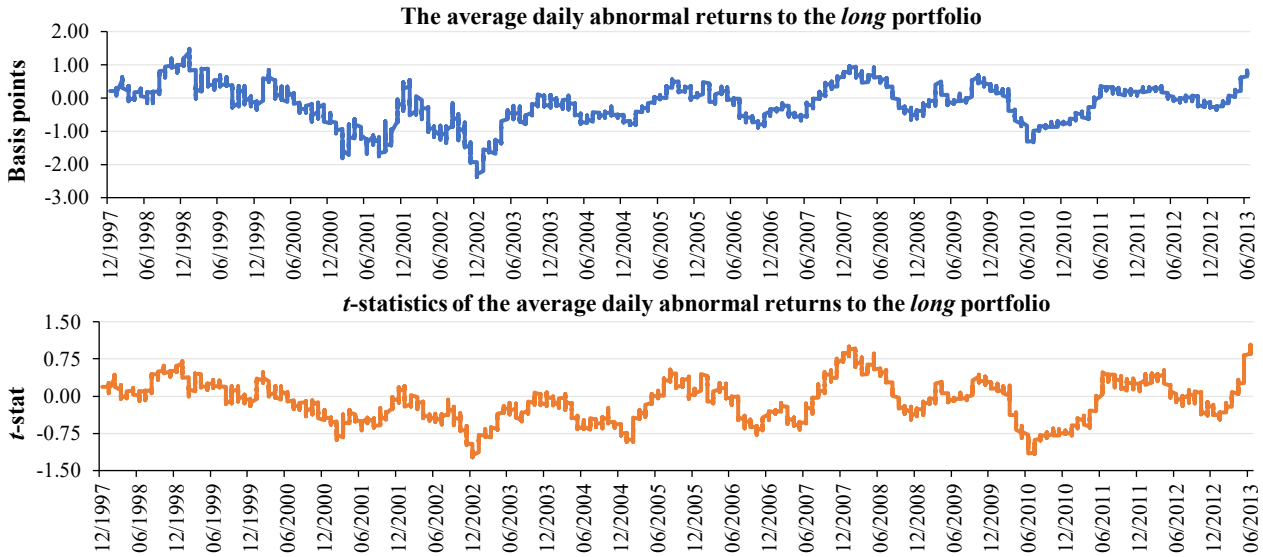

Figure 55.1a. The time-varying average daily abnormal returns to the *long* portfolio and the corresponding *t*-statistics under the Fama and French (1993) three-factor model, based on the whole sample of UK sell-side analyst recommendations

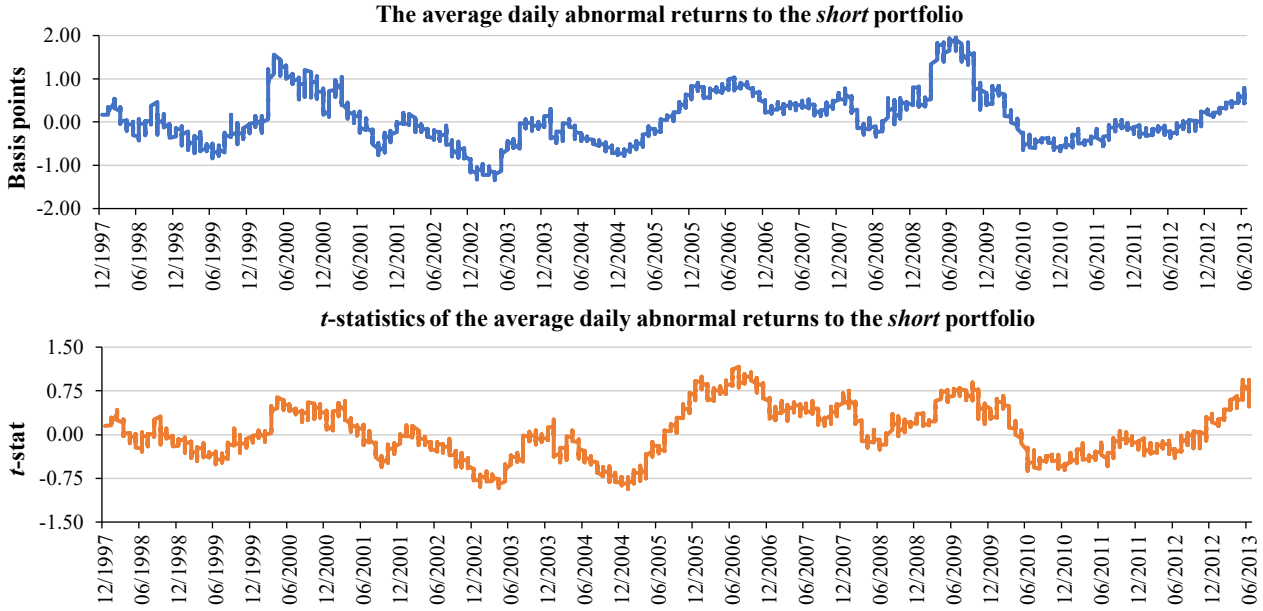

Figure 55.1b. The time-varying average daily abnormal returns to the *short* portfolio and the corresponding *t*-statistics under the Fama and French (1993) three-factor model, based on the whole sample of UK sell-side analyst recommendations

Notes:

Figure 55.1a (b) illustrates the time-varying average daily abnormal returns to the *long* (*short*) portfolio using the whole sample of UK sell-side analyst recommendations over the period January 1997 to June 2013. The *long* portfolio includes all upgrades to Strong Buy or Buy from previous Strong Sell, Sell, or Hold, while the *short* portfolio includes all downgrades to Strong Sell, Sell, or Hold from previous Strong Buy or Buy. For each analyst recommendation made by a brokerage house, the recommended stock enters the *long* or *short* portfolio at the close of trading on the day the recommendation is released, and remains in the portfolio until the close of trading on the day the recommendation expires. The abnormal return is estimated as the intercept term derived from the Fama and French (1993) three-factor model on a one-year window length rolling one trading day forward. The reported abnormal returns in this section are the abnormal net returns adjusted by the round-trip transaction costs in the UK, i.e., 1.5% for purchasing stocks and 3.0% for short selling stocks.

55.4 RESULTS

In this section, we first report the results for the average daily abnormal returns to the *long* and *short* portfolios, followed by the simulated results. The abnormal returns to the *long* and *short* portfolios are calculated as the gross abnormal returns less the estimated transaction costs multiplied by the corresponding daily portfolio turnover in each rolling window. We employ a relatively cautious estimate of the average round-trip transaction costs in the UK for purchasing stocks at 1.5% and for short selling stocks at 3.0% (see, e.g., Hudson et al., 1996; Ellis and Thomas, 2004; Li, Brooks, and Miffre, 2010).[6] Throughout this section, we present the results under the Fama and French (1993) three-factor model, while our results remain qualitatively the same under the CAPM and the Carhart (1997) four-factor model, ruling out the concern on a poor model of asset pricing raised by Barber et al. (2001). Our empirical and simulated results under the the CAPM and the Carhart (1997) four-factor model are not presented for the sake of brevity, but are available on request from the authors.

55.4.1 Time-varying performance of the long and short portfolios

Figure 55.1a (1b) shows the time-varying average daily abnormal returns to the *long* (*short*) portfolio and the corresponding *t*-statistics. Overall, we find that the abnormal returns to the *long* and *short* portfolios fluctuate over time during the whole sample period, while these positive or negative abnormal returns are not statistically significant at the conventional level. Specifically, the abnormal returns to the *long* portfolio range from –2.38 basis points (*t*-stat = –1.23) to 1.48 basis points (*t*-stat = 0.70), while the abnormal returns to the *short* portfolio range from –1.34 basis points (*t*-stat = –0.91) to 1.95 basis points (*t*-stat = 0.80). Our rolling window analysis results suggest that it is unlikely for investors to make profits either by purchasing stocks with positive revisions or by short selling stocks with negative revisions during any period of time.

[6] Keim and Madhavan (1998) categorize transaction costs into explicit costs (e.g., brokerage commissions and taxes) and implicit costs (e.g., bid-ask spread and market impact of trading). Hudson et al. (1996) show that the total round-trip transaction costs in the UK stock market for the most favored of investors is upward of 1.0%, including government stamp duty of 0.5%, negotiated brokerage commission of 0.1% (soft commissions could be zero if alternative services are offered in lieu of cash), and bid-ask spread of 0.5%.

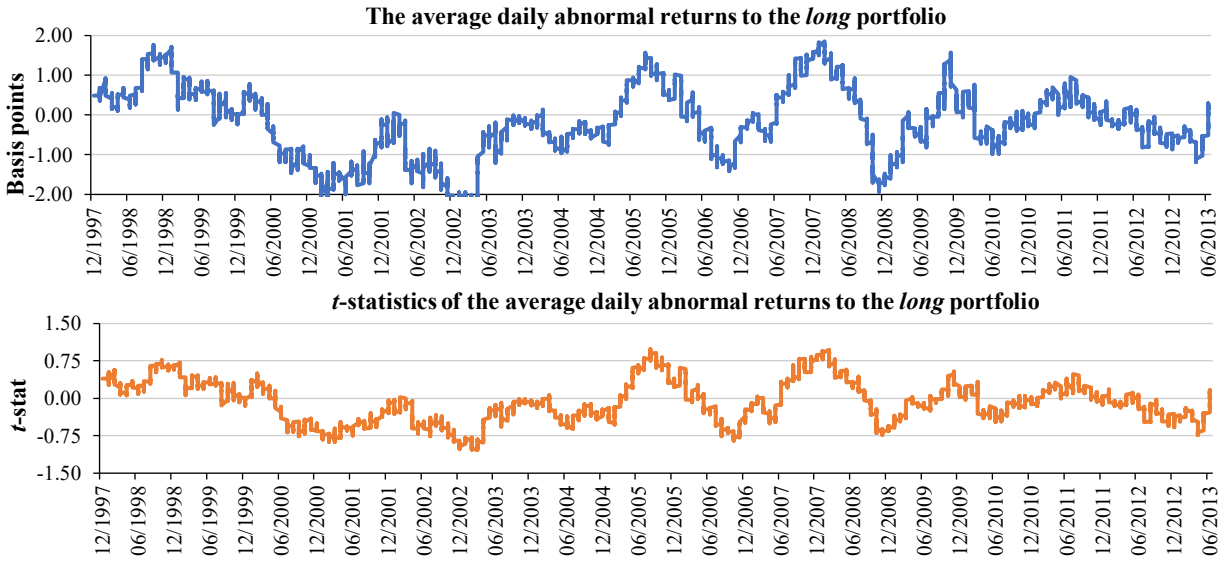

Figure 55.2a. The time-varying average daily abnormal returns to the *long* portfolio and the corresponding *t*-statistics under the Fama and French (1993) three-factor model, based on the subsample of UK sell-side analyst recommendations made exclusively by Top 5 brokerage houses

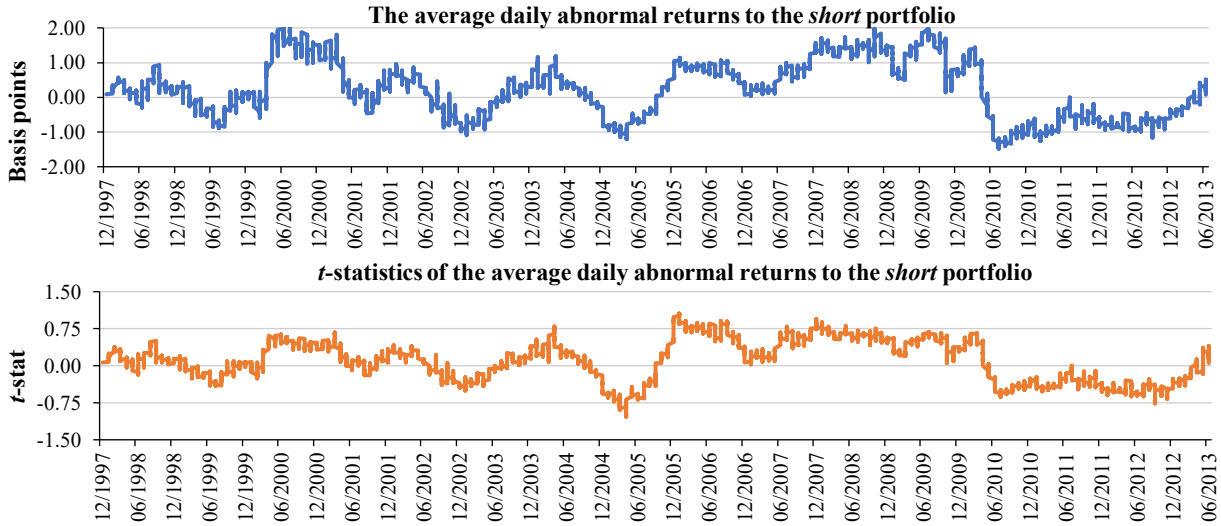

Figure 55.2b. The time-varying average daily abnormal returns to the *short* portfolio and the corresponding *t*-statistics under the Fama and French (1993) three-factor model, based on the subsample of UK sell-side analyst recommendations made exclusively by Top 5 brokerage houses

Notes:

Figure 55.2a (b) illustrates the time-varying average daily abnormal returns to the *long* (*short*) portfolio using the subsample of UK sell-side analyst recommendations made exclusively by Top 5 largest brokerage houses over the period January 1997 to June 2013. The Top 5 brokerage houses are identified by their three-year moving average (t –3, t –2, t –1) of positions on the annual All-European Research Team published by *Institutional Investor*. The *long* portfolio includes all upgrades to Strong Buy or Buy from previous Strong Sell, Sell, or Hold, while the *short* portfolio includes all downgrades to Strong Sell, Sell, or Hold from previous Strong Buy or Buy. For each analyst recommendation made by a brokerage house, the recommended stock enters the *long* or *short* portfolio at the close of trading on the day the recommendation is released, and remains in the portfolio until the close of trading on the day the recommendation expires. The abnormal return is estimated as the intercept term derived from the Fama and French (1993) three-factor model on a one-year window length rolling one trading day forward. The reported abnormal returns in this section are the abnormal net returns adjusted by the round-trip transaction costs in the UK, i.e., 1.5% for purchasing stocks and 3.0% for short selling stocks.

Although a number of investment strategies based on analyst recommendation revisions fail to generate significant abnormal returns, Barber et al. (2001; p. 537) point out that "it remains an open question whether other types of trading strategies could be profitable". A growing body of evidence shows that more prestigious brokerage houses have superior access to company managers and information; thus, their stock recommendations are supposed to be more readily available to investors (see, Stickel, 1995; Womack, 1996; Barber et al., 2001; Hong and Kubrik, 2003; Green, 2006; Jegadeesh and Kim, 2006; Fang and Yasuda 2014). For the robustness purposes, we construct two alternative *long* and *short* portfolios, based on a subsample of 55,092 analyst recommendation revisions made exclusively by Top 5 brokerage houses, identified by their three-year moving average (t –3, t –2, t –1) of positions on the annual All-Europe Research Team published by *Institutional Investor*. For example, the ranking of each brokerage house in the calendar year of 1998 is determined based on its average position in the years of 1997, 1996, and 1995.[7] Replicating all analyses in Figures 55.1a and 1b, we find that our results remain qualitatively the same (see Figures 55.2a & 2b). That is, investment strategies following UK sell-side analyst recommendation revisions could not generate significant abnormal returns, irrespectively of whether these stock recommendations are made by sell-side analysts working for more or less prestigious brokerage houses.

55.4.2 Bootstrapping simulations

As our results suggest that sell-side analysts cannot provide valuable stock recommendation revisions, we further examine whether the observed insignificant abnormal returns are simply as a result of these analysts' bad luck or their inferior skill (Fama, 1998; Barber et al., 2001). We compare any actual t-statistic of alpha with its appropriate luck distribution to distinguish whether sell-side analysts have sufficient skill in making valuable stock recommendations. Like, Kosowski et al. (2006), Barras et al. (2010), and Fama and French (2010), we focus on the distribution of the t-statistic of alpha rather than the actual alpha, as the t-statistic has superior statistical properties (see, Brown, Goetzmann, Ibbotson, and Ross, 1992). Specifically, the t-statistic accounts for differences in the precision of alpha. Consider, for example, an investor who takes high risks or who only has a short time series of portfolio returns. That investor can more easily exhibit extreme estimated alphas, which, however, are likely to be spurious outliers. The t-statistic corrects for that by scaling the alphas by its standard errors. Through this rescaling, the t-statistic accounts even more generally for the differences in risk-taking and number of observations in the whole rolling

[7] Su et al. (2018) provide a list of Top 5 brokerage houses based on their three-year moving average of positions on the annual All-Europe Research Team published by *Institutional Investor* over the period 1996 to 2013. For each brokerage house in each year of t, its positions in previous three years (t –3, t –2, t –1) are shown in bracket.

windows. The time-series distribution of the *t*-statistic of alpha, therefore, has the statistically preferable attribute of being closer to a normal distributed than the time-series distribution of alpha. Therefore, we compare the *t*-statistics of alphas at selected percentiles of the CDF of the actual *t*-statistics of alphas with the average of the 10,000 simulated *t*-statistics at the same percentiles. For the outperforming subsamples, if the simulated *t*-statistic is higher than the actual *t*-statistic in *less* than 5% of the 10,000 simulated *t*-statistics,[8] we reject the null hypothesis that the abnormal return is due to luck at the 95% confidence level and infer that the outperforming subsamples show superior skill and *vice versa* (see, also, Berk and Green, 2004; Gallefoss et al., 2015). In this subsection, we focus on presenting the simulated results based on *Procedure II* (see Table 55.2), as the simulated results based on *Procedure I* (see Table 55.3) are quite similar, in line with the results shown in Gallefoss et al. (2015).

Panel A of Table 55.2 shows the CDFs of the actual *t*-statistics of the abnormal returns and the average of the simulated CDFs for the *long* and *short* portfolios, based on the whole sample. The left side of Panel A shows that for the *long* portfolio, the actual *t*-statistics of the abnormal returns are always smaller than the corresponding average values from the simulations for all percentiles. For example, the left tail 5th and 10th percentiles of the actual *t*-statistics are −1.09 and −0.87, respectively, smaller than the corresponding average simulated values of −0.55 and −0.35. Similarly, the right tail 90th and 95th percentiles of the actual *t*-statistics are 0.77 and 0.98, respectively, smaller than the average simulated values of 0.92 and 1.08. For the outperforming subsamples, there are more than 30% of simulated *t*-statistics higher than the actual *t*-statistics. The simulated results clearly suggest sell-side analysts' lack of skill in making valuable upward revisions for stocks that can generate significantly positive abnormal returns in any period of time.

The right side of Panel A of Table 55.2 shows that, for the *short* portfolio, the left tail percentiles of the actual *t*-statistics of the abnormal returns are also below the corresponding average values from the simulations for all percentiles. For example, the 5th and 10th percentiles of the actual *t*-statistics are −1.06 and −0.75, respectively, much lower than the average simulated values of −0.60 and −0.30. Similarly, the right tail percentiles of the actual *t*-statistics confirm that sell-side analysts do not have certain skill in making valuable downward revisions in any period of time. For example, the 90th and 95th percentiles of the actual *t*-statistics are 1.01 and 1.18, respectively, smaller than the average simulated values of 1.28 and 1.45. Figure 55.3a (3b) illustrates the CDFs of the actual *t*-statistics of the abnormal returns and the corresponding average simulated CDFs for the *long* (*short*) portfolio, using the whole sample.

[8] In this subsection, we present the performance of the *short* portfolio is measured as the signed abnormal returns multiplied by −1 to make direct comparison with the performance of the *long* portfolio.

Table 55.2 Percentiles of *t*-statistics for actual and simulated abnormal returns (Procedure I)

	The *long* portfolio			**The *short* portfolio**		
%	**Simulated *t*-stat**	**Actual *t*-stat**	**% (Simulated > Actual)**	**Simulated *t*-stat**	**Actual *t*-stat**	**% (Simulated > Actual)**
Panel A: The whole sample						
1	–0.92	–1.48	0.20	–1.15	–1.47	0.23
2	–0.75	–1.30	0.21	–0.92	–1.32	0.30
3	–0.67	–1.21	0.23	–0.80	–1.22	0.49
4	–0.61	–1.15	0.24	–0.70	–1.13	0.49
5	–0.55	–1.09	0.24	–0.60	–1.06	0.81
10	–0.35	–0.87	0.26	–0.30	–0.75	0.94
20	–0.11	–0.56	0.27	0.08	–0.37	1.19
30	0.04	–0.35	0.32	0.31	–0.13	1.85
40	0.18	–0.19	0.53	0.48	0.04	2.71
50	0.30	–0.04	1.02	0.62	0.18	4.34
60	0.42	0.10	3.64	0.75	0.33	5.84
70	0.55	0.27	4.20	0.89	0.51	7.23
80	0.71	0.47	7.11	1.05	0.72	13.53
90	0.92	0.77	12.02	1.28	1.01	26.95
95	1.08	0.98	15.71	1.45	1.18	32.66
96	1.12	1.03	30.84	1.50	1.22	36.13
97	1.17	1.09	33.73	1.58	1.28	44.66
98	1.25	1.18	38.89	1.70	1.35	46.56
99	1.45	1.37	44.01	1.94	1.45	48.19
Panel B: The subsample of UK analyst recommendations made by Top 5 brokerage houses						
1	–0.89	–1.45	0.20	–1.12	–1.44	0.23
2	–0.72	–1.28	0.21	–0.89	–1.30	0.30
3	–0.64	–1.19	0.23	–0.77	–1.20	0.49
4	–0.59	–1.13	0.24	–0.67	–1.11	0.49
5	–0.53	–1.07	0.24	–0.57	–1.04	0.80
10	–0.33	–0.86	0.26	–0.27	–0.73	0.93
20	–0.09	–0.55	0.27	0.10	–0.36	1.18
30	0.05	–0.34	0.32	0.33	–0.12	1.83
40	0.19	–0.19	0.53	0.49	0.05	2.68
50	0.30	–0.05	1.01	0.63	0.18	4.30
60	0.42	0.10	3.61	0.76	0.33	5.78
70	0.55	0.27	4.16	0.89	0.51	7.16
80	0.71	0.46	7.05	1.05	0.72	13.40
90	0.91	0.76	11.92	1.28	1.00	26.68
95	1.07	0.97	15.59	1.44	1.17	32.34
96	1.11	1.01	30.63	1.49	1.21	35.77
97	1.15	1.07	33.46	1.57	1.26	44.22
98	1.23	1.16	38.58	1.68	1.33	46.10
99	1.43	1.35	43.66	1.92	1.43	47.71

Notes:

This table presents the values of *t*-statistics at selected percentiles (%) of the distribution of *t*-statistics of the actual and simulated abnormal returns, as well as the percent of the 10,000 simulation runs that produce higher values of *t*-statistics at the selected percentiles than those actual abnormal returns (% Simulated > Actual), over the period January 1997 to June 2013. The abnormal return is estimated as the intercept term derived from the Fama and French (1993) three-factor model, either using the whole sample of UK sell-side analyst recommendations (in Panel A), or using the subsample of UK sell-side analyst recommendations made by Top 5 brokerage houses (in Panel B). The Top 5 brokerage houses are identified by their three-year moving average (t –3, t –2, t –1) of positions on the annual All-Europe Research Team published by *Institutional Investor*.

We also report the simulated results using the subsample of analyst recommendation revisions made by Top 5 brokerage houses. Panel B of Table 55.2 shows that our subsample simulated results are qualitatively similar to those based on the whole sample, confirming that the reputation of brokerage houses does not matter. That is, on average, even sell-side analysts working for more prestigious brokerage houses in the UK have no sufficient skill to make valuable upward or downward revisions for stocks in any period of time.

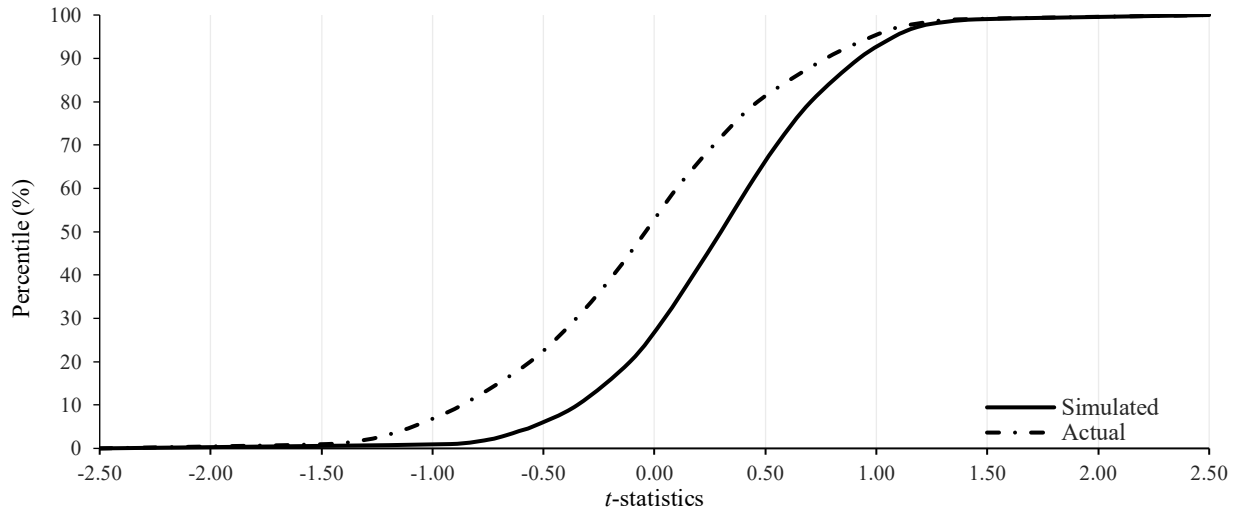

Figure 55.3a: Simulated and actual cumulative density function (CDF) of *t*-statistics for the abnormal returns to the *long* portfolio, under the Fama and French (1993) three-factor model, using the whole sample

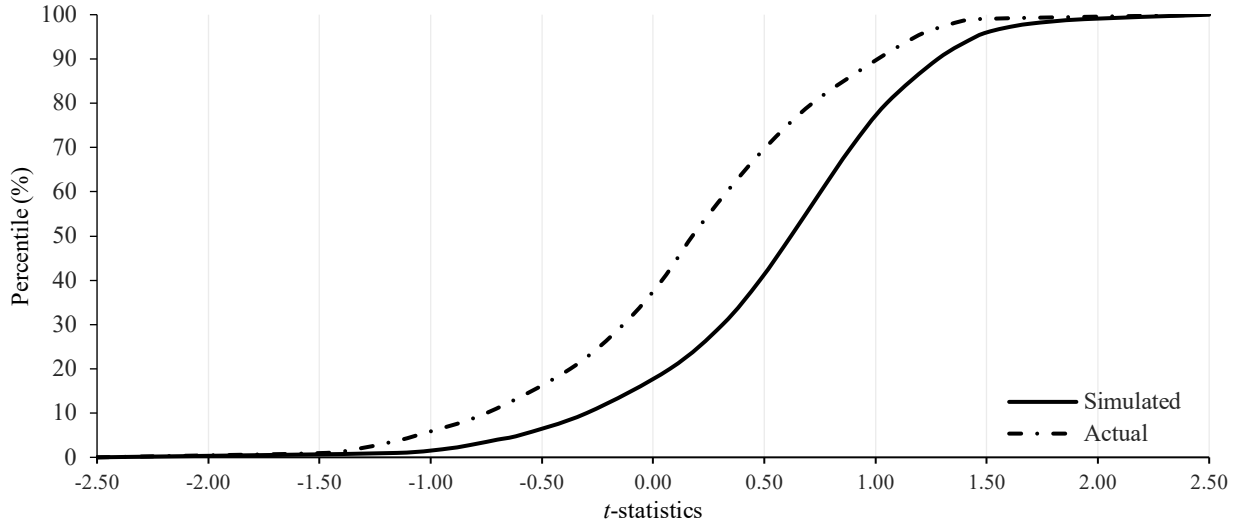

Figure 55.3b: Simulated and actual cumulative density function (CDF) of *t*-statistics for the abnormal returns to the *short* portfolio, under the Fama and French (1993) three-factor model, using the whole sample

Table 55.3 Percentiles of *t*-statistics for actual and simulated abnormal returns (Procedure II)

	The *long* portfolio			**The *short* portfolio**		
%	**Simulated *t*-stat**	**Actual *t*-stat**	**% (Simulated > Actual)**	**Simulated *t*-stat**	**Actual *t*-stat**	**% (Simulated > Actual)**
Panel A: The whole sample						
1	–0.88	–1.45	0.20	–1.10	–1.44	0.23
2	–0.72	–1.28	0.21	–0.88	–1.29	0.30
3	–0.64	–1.19	0.23	–0.76	–1.20	0.48
4	–0.58	–1.13	0.24	–0.66	–1.11	0.48
5	–0.52	–1.07	0.24	–0.57	–1.04	0.80
10	–0.33	–0.86	0.26	–0.27	–0.73	0.93
20	–0.10	–0.55	0.27	0.10	–0.36	1.17
30	0.05	–0.35	0.32	0.33	–0.12	1.83
40	0.19	–0.19	0.52	0.49	0.04	2.67
50	0.29	–0.05	1.01	0.63	0.18	4.28
60	0.42	0.10	3.59	0.76	0.33	5.76
70	0.55	0.26	4.14	0.89	0.51	7.13
80	0.70	0.46	7.01	1.05	0.71	13.35
90	0.91	0.75	11.86	1.27	1.00	26.59
95	1.06	0.96	15.50	1.43	1.16	32.23
96	1.10	1.01	30.42	1.48	1.20	35.65
97	1.15	1.07	33.27	1.56	1.26	44.07
98	1.22	1.15	38.36	1.68	1.33	45.94
99	1.42	1.34	43.41	1.92	1.43	47.55
Panel B: The subsample of UK analyst recommendations made by Top 5 brokerage houses						
1	–0.84	–1.41	0.20	–1.07	–1.41	0.23
2	–0.68	–1.25	0.21	–0.84	–1.27	0.30
3	–0.60	–1.16	0.23	–0.73	–1.17	0.48
4	–0.56	–1.11	0.24	–0.63	–1.09	0.48
5	–0.50	–1.05	0.24	–0.53	–1.02	0.79
10	–0.30	–0.84	0.26	–0.24	–0.71	0.92
20	–0.07	–0.53	0.27	0.13	–0.34	1.16
30	0.07	–0.33	0.32	0.35	–0.11	1.81
40	0.20	–0.18	0.52	0.51	0.06	2.64
50	0.31	–0.05	1.00	0.64	0.18	4.24
60	0.42	0.10	3.57	0.77	0.33	5.70
70	0.55	0.27	4.11	0.89	0.51	7.06
80	0.71	0.45	6.97	1.05	0.71	13.22
90	0.90	0.75	11.78	1.28	0.99	26.32
95	1.06	0.95	15.42	1.43	1.16	31.90
96	1.09	0.99	30.34	1.47	1.19	35.28
97	1.13	1.05	33.09	1.55	1.24	43.62
98	1.21	1.14	38.15	1.66	1.31	45.47
99	1.40	1.32	43.18	1.90	1.41	47.06

Notes:

This table presents the values of *t*-statistics at selected percentiles (%) of the distribution of *t*-statistics of the actual and simulated abnormal returns, as well as the percent of the 10,000 simulation runs that produce higher values of *t*-statistics at the selected percentiles than those actual abnormal returns (% Simulated > Actual), over the period January 1997 to June 2013. The abnormal return is estimated as the intercept term derived from the Fama and French (1993) three-factor model, either using the whole sample of UK sell-side analyst recommendations (in Panel A), or using the subsample of UK sell-side analyst recommendations made by Top 5 brokerage houses (in Panel B). The Top 5 brokerage houses are identified by their three-year moving average (t –3, t –2, t –1) of positions on the annual All-Europe Research Team published by *Institutional Investor*.

55.5 CONCLUSIONS

We examine the time-varying performance of investment strategies—the *long* and *short* portfolios—constructed following UK sell-side analyst recommendation revisions over the period January 1997 to June 2013. We find that the abnormal returns to the *long* and *short* portfolios are not statistically significant at the conventional level in any period of time. Importantly, we develop a time-series bootstrapping simulation method confirming that the observed statistically insignificant abnormal returns to the *long* and *short* portfolios are due to sell-side analysts' lack of skill in making valuable stock recommendation revisions, rather than their bad luck. Our conclusions hold up fairly well with the use of an alternative subsample of analyst recommendation revisions exclusively made by sell-side analysts working for more prestigious brokerage houses under various single- and multi-factor asset pricing models. Our rolling window based bootstrapping simulation method can be applied in other time-series investment performance to distinguish luck from skill.

BIBLOGRAPHY

Agarwal, V., & Naik, N.Y. (2001). Multi-period performance persistence analysis of hedge funds. *Journal of Financial and Quantitative Analysis*, 35, 327–342.

Altınkılıç, O., Balashov, V.S., & Hansen, R. S. (2013). Are analysts' forecasts informative to the general public? *Management Science*, 59, 2550–2565.

Altınkılıç, O., & Hansen, R.S. (2009). On the information role of stock recommendation revisions. *Journal of Accounting and Economics*, 48, 17–36.

Baquero, G., Horst, J.T., & Verbeek, M. (2005). Survival, look-ahead bias, and persistence in hedge fund performance. *Journal of Financial and Quantitative Analysis*, 40, 493–517.

Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *Journal of Finance*, 56, 531–563.

Barber, B., Lehavy, R., & Trueman, B. (2007). Comparing the stock recommendation performance of investment banks and independent research firms. *Journal of Financial Economics*, 85, 490–517.

Barras, L., Scaillet, O., & Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *Journal of Finance*, 65, 179–216.

Benartzi, S., & Thaler, R.H. (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, 110, 73–92.

Berk, B.J., & Green, C.R. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112, 1269–1295.

Blake, D., Caulfield, T., Ioannidis, C., & Tonks, I. (2014). Improved inference in the evaluation of mutual fund performance using panel bootstrap methods. *Journal of Econometrics*, 183, 202–210.

Blake, D., Caulfield, T., Ioannidis, C., & Tonks, I. (2017). New evidence on mutual fund performance: A comparison of alternative bootstrap methods. *Journal of Financial and Quantitative Analysis*, 52, 1279–1299.

Boni, L., & Womack, K.L. (2006). Analysts, industries, and price momentum. *Journal of Financial and Quantitative Analysis*, 41, 85–109.

Bradley, D., Clarke, J., Lee, S., & Ornthanalai, C. (2014). Are analysts' recommendations informative? Intraday evidence on the impact of time stamp delays. *Journal of Finance*, 69, 645–673.

Bradley, D., Gokkaya, S., & Liu, X.I. (2016). Before an analyst becomes an analyst: Does industry experience matter? *Journal of Finance*, 72, 751–792.

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47, 1731–1764.

Brown, L., & Huang, K. (2013). Recommendation-forecast consistency and earnings forecast quality. *Accounting Horizons*, 27, 451–467.

Brown, S.J., Goetzmann, W.N., Ibbotson, R.G., & Ross, S. A. (1992). Survivorship bias in performance studies. *Review of Financial Studies*, 5, 553–580.

Brown, S.J., Goetzmann, W.N., & Ibbotson, R.G. (1999). Offshore hedge funds: Survival and performance, 1989–95. *Journal of Business*, 72, 91–117.

Brown, S.J., Goetzmann, W.N., & Kumar, A. (1998). The Dow Theory: William Peter Hamilton's track record reconsidered. *Journal of Finance*, 53, 1311–1333.

Cai, Z., Jiang, J., Zhang, J., & Zhang, X. (2015). A new semiparametric test for superior predictive ability. *Empirical Economics*, 48, 389–405.

Carhart, M.M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52, 57–82.

Cogneau, P., & Zakamouline, V. (2013). Block bootstrap methods and the choice of stocks for the long run. *Quantitative Finance*, 13, 1443–1457.

Colker, S. (1963). An analysis of security recommendations by brokerage houses. *Quarterly Review of Economics and Business*, 3, 19–28.

Cowles, A. (1933). Can stock market forecasters forecast? *Econometrica*, 1, 309–324.

Cuthbertson, K., Nitzsche, D., & O'Sullivan, N. (2008). UK mutual fund performance: Skill or luck? *Journal of Empirical Finance*, 15, 613–634.

Dimson, E., & Fraletti, P. (1986). Brokers' recommendations: The value of a telephone tip. *Economic Journal*, 96, 139–159.

Edwards, F.R., & Caglayan, M.O. (2001). Hedge fund performance and manager skill. *Journal of Futures Markets*, 21, 1003–1028.

Ellis, M., & Thomas, D.C. (2004). Momentum and the FTSE 350. *Journal of Asset Management*, 5, 25–36.

Elton, E.J., Gruber, M.J., & Blake, C.R. (1996). The persistence of risk-adjusted mutual fund performance. *Journal of Business*, 69, 133–157.

Elton, E.J., Gruber, M.J., & Blake, C.R. (2012). Does mutual fund size matter? The relationship between size and performance. *Review of Asset Pricing Studies*, 2, 31–55.

Emery, D.R., & Li, X. (2009). Are the Wall Street analyst rankings popularity contests? *Journal of Financial and Quantitative Analysis*, 44, 411–438.

Ertimur, Y., Sunder, J., & Sunder, S.V. (2007). Measure for measure: The relation between forecast accuracy and recommendation profitability of analysts. *Journal of Accounting Research*, 45, 567–606.

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

Fama, E.F., & French, K.R. (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance*, 65, 1915–1947.

Fama, E.F., & French, K.R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116, 1–22.

Fang, L., & Yasuda, A. (2009). The effectiveness of reputation as a disciplinary mechanism in sell-side research. *Review of Financial Studies*, 22, 3735–3777.

Fang, L., & Yasuda, A. (2014). Are stars' opinions worth more? The relation between analyst reputation and recommendation values. *Journal of Financial Services Research*, 46, 235–269.

Gallefoss, K., Hansen, H.H., Haukaas, E.S., & Molnár, P. (2015). What daily data can tell us about mutual funds: Evidence from Norway. *Journal of Banking and Finance*, 55, 117–129.

Goetzmann, W.N., & Ibbotson, R.G. (1994). Do winners repeat? *Journal of Portfolio Management*, 20, 9–18.

Green, T.C. (2006). The value of client access to analyst recommendations. *Journal of Financial and Quantitative Analysis*, 41, 1–24.

Gregory, A., Tharyan, R., & Christidis, A. (2013). Constructing and testing alternative versions of the Fama–French and Carhart models in the UK. *Journal of Business Finance and Accounting*, 40, 172–214.

Grinblatt, M., & Titman, S. (1994). A study of monthly mutual fund returns and performance evaluation techniques. *Journal of Financial and Quantitative Analysis*, 29, 419–444.

Groth, J.C., Lewellen, W.G., Schlarbaum, G.G., & Lease, R.C. (1979). An analysis of brokerage house securities recommendations. *Financial Analysts Journal*, 35, 32–40.

Guan, Y., Lu, H., & Wong, M.H.F. (2012). Conflict-of-Interest reforms and investment bank analyst's research biases. *Journal of Accounting, Auditing and Finance*, 27, 443–470.

Hall, J.L., & Tacon, P.B. (2010). Forecast accuracy and stock recommendations. *Journal of Contemporary Accounting & Economics*, 6, 18–33.

Hansen, P.R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23, 365–380.

Hansen, P.R., Lunde, A., & Nason, J.M. (2005). Testing the significance of calendar effects. *Working Paper*, Federal Reserve Bank of Atlanta.

Harri, A., & Brorsen, B.W. (2004). Performance persistence and the source of returns for hedge funds. *Applied Financial Economics*, 14, 131–141.

Hendricks, D., Patel, J., & Zeckhauser, R. (1993). Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *Journal of Finance*, 48, 93–130.

Hobbs, J., Kovacs, T., & Sharma, V. (2012). The investment value of the frequency of analyst recommendation changes for the ordinary investor. *Journal of Empirical Finance*, 19, 94–108.

Hong, H., & Kubik, J.D. (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *Journal of Finance*, 58, 313–351.

Hou, K., Xue, C., & Zhang, L. (2014). Digesting anomalies: An investment approach. *Review of Financial Studies*, 28, 650–705.

Hou, K., Xue, C., & Zhang, L. (2017). Replicating anomalies. *Working Paper*, *Fisher College of Business*.

Hsu, P.-H., Hsu, Y.-C., & Kuan, C.-M. (2010). Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. *Journal of Empirical Finance*, 17, 471–484.

Hsu, P.-H., & Kuan, C.-M. (2005). Reexamining the profitability of technical analysis with data snooping checks. *Journal of Financial Econometrics*, 3, 606–628.

Hudson, R., Dempsey, M., & Keasey, K. (1996). A note on the weak form efficiency of capital markets: The application of simple technical trading rules to UK stock prices – 1935 to 1994. *Journal of Banking and Finance*, 20, 1121–1132.

Irvine, P., Lipson, M., & Puckett, A. (2007). Tipping. *Review of Financial Studies*, 20, 741–768.

Ivković, Z., & Jegadeesh, N. (2004). The timing and value of forecast and recommendation revisions. *Journal of Financial Economics*, 73, 433–463.

Jegadeesh, N., Kim, J., Krische, S.D., & Lee, C.M.C. (2004). Analyzing the analysts: When do recommendations add value? *Journal of Finance*, 59, 1083–1124.

Jegadeesh, N., & Kim, W. (2006). Value of analyst recommendations: International evidence. *Journal of Financial Markets*, 9, 274–309.

Jegadeesh, N., & Kim, W. (2010). Do analysts herd? An analysis of recommendations and market reactions. *Review of Financial Studies*, 23, 901–937.

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

Jensen, M.C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance*, 23, 389–416.

Jiang, J., Zhou, H., Jiang, X., & Peng, J. (2007). Generalized likelihood ratio tests for the structure of semiparametric additive models. *Canadian Journal of Statistics*, 35, 381–398.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *Econometrics Journal*, 11, 377–395.

Kat, H.M., & Menexe, F. (2003). Persistence in hedge fund performance: The true value of a track record. *Journal of Alternative Investments*, 5, 66–72.

Keim, D.B., & Madhavan, A. (1998). The cost of institutional equity trades. *Financial Analysts Journal*, 54, 50–69.

Kosowski, R., Naik, N.Y., & Teo, M. (2007). Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *Journal of Financial Economics*, 84, 229–264.

Kosowski, R., Timmermann, A., Wermers, R., & White, H.A.L. (2006). Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance*, 61, 2551–2595.

Kouwenberg, R. (2003). Do hedge funds add value to a passive portfolio? Correcting for non-normal returns and disappearing funds. *Journal of Asset Management*, 3, 361–382.

Lakonishok, J., Shleifer, A., & Vishny, R.W. (1992). The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32, 23–43.

Li, X. (2005). The persistence of relative performance in stock recommendations of sell-side financial analysts. *Journal of Accounting and Economics*, 40, 129–152.

Li, X., Brooks, C., & Miffre, J. (2010). Transaction costs, trading volume and momentum strategies. *Journal of Trading*, 5, 66–81.

Lo, A.W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, 55, 1705–1765.

Loh, R.K., & Mian, G.M. (2006). Do accurate earnings forecasts facilitate superior investment recommendations? *Journal of Financial Economics*, 80, 455–483.

McLean, D.R., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, 71, 5–32.

Mehran, H., & Stulz, R.M. (2007). The economics of conflicts of interest in financial institutions. *Journal of Financial Economics*, 85, 267–296.

Meng, X. (2015). Analyst reputation, communication, and information acquisition. *Journal of Accounting Research*, 53, 119–173.

Meyer, S., Schmoltzi, D., Stammschulte, C., Kaesler, S., Loos, B., & Hackethal, A. (2012). Just unlucky? – A bootstrapping simulation to measure skill in individual investors' investment performance. *Working Paper*, Goethe University Frankfurt.

Mikhail, M.B., Walther, B.R., & Willis, R.H. (2004). Do security analysts exhibit persistent differences in stock picking ability? *Journal of Financial Economics*, 74, 67–91.

Muslu, V., & Xue, Y. (2013). Analysts' momentum recommendations. *Journal of Business Finance and Accounting*, 40, 438–469.

Neely, C.J., Weller, P.A., & Ulrich, J.M. (2009). The Adaptive Markets Hypothesis: Evidence from the foreign exchange market. *Journal of Financial and Quantitative Analysis*, 44, 467–488.

Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*, 29, 104–147.

Okunev, J., & White, D. (2003). Do momentum-based strategies still work in foreign currency markets? *Journal of Financial and Quantitative Analysis*, 38, 425–447.

Olson, D. (2004). Have trading rule profits in the currency markets declined over time? *Journal of Banking and Finance*, 28, 85–105.

Pástor, L., & Stambaugh, R.F. (2002). Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, 63, 315–349.

Politis, D.N., & Romano, J.P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89, 1303–1313.

Qi, M., & Wu, Y. (2006). Technical trading-rule profitability, data snooping, and Reality Check: Evidence from the foreign exchange market. *Journal of Money, Credit and Banking*, 38, 2135–2158.

Romano, J.P., & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73, 1237–1282.

Ryan, P., & Taffler, R.J. (2006). Do brokerage houses add value? The market impact of UK sell-side analyst recommendation changes. *British Accounting Review*, 38, 371–386.

Savin, G., Weller, P., and Zvingelis, J. (2007). The predictive power of "head-and-shoulders" price patterns in the U.S. stock market. *Journal of Financial Econometrics*, 5, 243–265.

Shen, C.-H., & Chih, H.-L. (2009). Conflicts of interest in the stock recommendations of investment banks and their determinants. *Journal of Financial and Quantitative Analysis*, 44, 1149–1171.

Sørensen, L.Q. (2009). Mutual fund performance at the Oslo stock exchange. *Working Paper*, Norwegian School of Economics.

Stickel, S.E. (1995). The anatomy of the performance of buy and sell recommendations. *Financial Analysts Journal*, 51, 25–39.

Storey, J.D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64, 479–498.

Su, C., Zhang, H., Joseph, N.L, and Bangassa, K. (2018). On the investment value of sell-side analyst recommendation revisions in the UK. *Review of Quantitative Finance and Accounting*, Forthcoming.

Sullivan, R., Timmermann, A., and White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, 54, 1647–1691.

Sullivan, R., Timmermann, A., and White, H. (2001). Dangers of data-driven inference: The case of calendar effects in stock returns. *Journal of Econometrics*, 105, 249–286.

Timmermann, A., & Granger, C.W.J. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, 20, 15–27.

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

Trueman, B. (1994). Analyst forecasts and herding behaviour. *Review of Financial Studies*, 7, 97–124.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68, 1097–1126.

Womack, K.L. (1996). Do brokerage analysts' recommendations have investment value? *Journal of Finance*, 51, 137–167.

APPENDIX 55A DESCRIPTIVE STATISTICS ON UK SELL-SIDE ANALYST RECOMMENDATIONS

This appendix presents descriptive statistics on 294,692 UK sell-side analyst recommendations in each year over the period January 1997 to June 2013. All real-time analyst recommendations are obtained from *Morningstar Company Intelligence*, including information on the name of the firm recommended, the name of the brokerage house issuing the recommendation, the recommendation starting date and expiration date, and a rating between 1 and 9. A rating of 1 reflects a strong buy, 2 a buy, 3 a weak buy, 4 a weak buy/hold, 5 a hold, 6 a hold/sell, 7 a weak sell, 8 a sell, and 9 a strong sell.

Year	No. of recommended firms	No. of brokerage houses	No. of recommendations	Average rating	Recommendation frequency									Positive recommendations		Negative recommendations	
					1	2	3	4	5	6	7	8	9	No.	%	No.	%
1997	975	45	24,770	3.75	459	10,556	1,638	487	9,295	247	364	1,515	209	13,140	53.05	11,630	46.95
1998	1,021	47	25,074	3.77	732	9,853	2,704	426	8,580	147	886	1,448	298	13,715	54.70	11,359	45.30
1999	985	45	21,842	3.65	672	8,898	2,786	254	7,121	105	739	1,037	230	12,610	57.73	9,232	42.27
2000	931	50	18,189	3.43	752	7,892	2,551	185	5,509	85	426	656	133	11,380	62.57	6,809	37.43
2001	939	45	16,687	3.94	369	6,153	1,832	228	5,750	77	768	1,436	74	8,582	51.43	8,105	48.57
2002	911	44	15,052	3.93	343	5,942	1,420	235	4,713	137	778	1,461	23	7,940	52.75	7,112	47.25
2003	867	48	17,666	3.87	211	7,420	1,630	331	5,498	41	868	1,631	36	9,592	54.30	8,074	45.70
2004	885	50	20,609	3.82	208	8,722	2,198	213	6,521	22	957	1,735	33	11,341	55.03	9,268	44.97
2005	956	52	21,511	3.98	213	8,326	2,317	96	7,264	140	1,018	2,083	54	10,952	50.91	10,559	49.09
2006	1,000	54	20,186	3.72	255	9,125	2,152	6	6,019	20	974	1,620	15	11,538	57.16	8,648	42.84
2007	977	51	15,560	3.59	92	7,632	1,648	4	4,463	6	587	1,126	2	9,376	60.26	6,184	39.74
2008	963	53	16,117	3.66	119	8,268	1,047	142	4,370	2	517	1,652	0	9,576	59.42	6,541	40.58
2009	856	53	17,686	3.72	147	8,625	1,231	227	5,148	0	495	1,791	22	10,230	57.84	7,456	42.16
2010	834	40	13,821	3.32	199	7,834	997	83	3,584	0	221	902	1	9,113	65.94	4,708	34.06
2011	804	40	13,636	3.38	195	7,522	1,002	117	3,604	0	217	978	1	8,836	64.80	4,800	35.20
2012	805	40	11,694	3.41	104	6,442	815	135	3,196	0	140	862	0	7,496	64.10	4,198	35.90
2013 (Jan.–Jun.)	613	31	4,562	3.62	0	2,361	309	20	1,432	0	42	428	0	2,690	58.58	1,902	41.42
Full sample	**2,409**	**122**	**294,692**	**3.70**	**5,070**	**131,571**	**28,277**	**3,189**	**92,067**	**1,029**	**9,997**	**22,361**	**1,131**	**168,107**	**57.04**	**126,585**	**42.96**

APPENDIX 55B: STATA CODES FOR THE TIME-SERIES BOOTSTRAPPING SIMULATIONS

Our bootstrapping simulations can be conducted by the main body of the Stata codes (*bs_resid.do* and *bs_resid.ado*) with a loop function.[9] Specifically, the *.do* file is a text file containing commands to execute the commands stored in the *.ado* file, which could be downloaded and saved in a folder under the *personal* directory, along with the corresponding help file (*.hlp*) in the same folder. In this appendix, we detail the bootstrapping simulation method under the Fama and French (1993) three-factor model, but the application of the bootstrapping procedure to other asset pricing models, e.g., the CAPM, the Carhart (1997) four-factor model, the Fama and French (2015) five-factor model, and so on, is very similar, with the only modification of the steps being the substitution of appropriate models.

The Stata .ado file, bs_resid.ado, for Procedure I: Resampling the residuals independently with fixed common risk factors—Kosowiski et al. (2006) bootstrapping simulations

```
program bs_resid

version 13.1

syntax, RESidual(varname numeric) MATrix(name)
```

Step (i): The time-series daily excess portfolio returns ($R_t - R_{f,t}$) of the portfolio are regressed on the Fama and French (1993) three factors ($R_t - R_{f,t}$, SMB_t, and HML_t) in each rolling window, to calculate the estimated alphas ($\hat{\alpha}$), factor loadings ($\hat{\beta}, \hat{s}, \hat{h}$), and residuals ($\hat{\varepsilon}_t$). Specifically, R_t and $R_{m,t}$ are the daily return of the portfolio and the FTSE All-Share Index, respectively; $R_{f,t}$ represents the three-month UK T-bill rate; SMB_t and HML_t represent the daily returns on zero-investment factor-mimicking portfolios for size and B/M, respectively. The daily ($R_{m,t} - R_{f,t}$), SMB_t, and HML_t in the UK stock market are collected from the Xfi Centre for Finance and Investment at the University of Exeter (see, Gregory et al., 2013). The excess return of portfolio and risk factors for the Fama and French three-factor model will be specified in the *.do* file, *bs_resid.do* (see descriptive statistics and correlations of various factor returns in Appendix 55C).

```
local xvars : colna `matrix'

local CONS _cons
```

[9] The main body of our Stata codes are developed by Jeff Pitblado at StataCorp LP. This is a sample for demonstration purpose only.

```
local xvars : list xvars - CONS
```

Step (ii): For each loop (i.e., each rolling window), we save the coefficient estimates $\{\hat{\alpha}_p, \hat{\beta}_p, \hat{s}_p, \hat{h}_p\}$, the time-series of estimated residuals ($\hat{\varepsilon}_t$), and the *t*-statistic of the intercept ($\hat{t}_{\hat{\alpha}}$), generated from Step (i).

```
tempvar xb idx y

matrix score double `xb' = `matrix'
```

Step (iii): We randomly select the residuals from the saved residual vector $\{\hat{\varepsilon}_t\}$ in Step (ii), with replacements, in attempts to prepare for a time-series of the pseudo residuals $\{\hat{\varepsilon}_{t_b}^b\}$ in the next step.

```
gen long `idx' = ceil(_N*runiform())
```

Step (iv): Using the pseudo time-series of residuals $\{\hat{\varepsilon}_{t_b}^b\}$ and fixed time-series of risk factors $\{\hat{\beta}_p, \hat{s}_p, \hat{h}_p\}$, we generate a time-series of pseudo daily excess returns $\left(R_t - R_{f,t}\right)^b$ in each rolling window; the time-series of pseudo excess returns impose the null hypothesis of zero true performance ($\alpha = 0$). In this step, we only record the residuals rather than resampling the common risk factors, according to Kosowski et al. (2006).

```
gen double `y' = `xb' + `residual'[`idx']

regress `y' `xvars', vce(robust)
```

Step (v): We regress the pseudo daily excess returns $\left(R_t - R_{f,t}\right)^b$ generated from Step (iv) against the Fama and French (1993) three factors used in Step (i), generating the simulated $\hat{\alpha}^b$, which represents the sampling variation around zero true performance, purely due to random chance (i.e., sell-side analysts' luck). We repeat the above procedures in each rolling window and generate a time-series of simulated alphas, $\{\hat{\alpha}_w^b\}$, and their corresponding *t*-statistics, $\{\hat{t}_{\alpha_w}^b\}$, where w is the number of rolling windows, throughout the whole sample period. The standard errors are corrected for heteroscedasticity through the Newey–West procedure with 0 lag.

```
End
```

All simulated $\hat{\alpha}_w^b$ are then sorted into a CDF of the simulated $\hat{\alpha}_w^b$, $f(\hat{\alpha}_w^b)$, a separate time-series of *luck* distribution from the worst performing rolling window to the best performing

rolling window. We repeat the above bootstrapping procedure a large number of times, say, b = 1, …, 10,000, thus generating a similar time-series distribution of bootstrapped *t*-statistics $\{\hat{t}^b_{\alpha_w}\}$, which can be compared with the distribution of the actual distribution $\{\hat{t}_{\alpha_w}\}$, once both sets of *t*-statistics have been resorted from the lowest value ($\hat{t}_{\alpha_{min}}$) to the highest value ($\hat{t}_{\alpha_{max}}$). For the outperforming (underperforming) subsamples measured by *t*-statistics of alpha, if the simulated $\hat{t}^b_{\alpha_{max}}$ is greater than the actual $\hat{t}_{\alpha_{max}}$ in less than 5% of the 10,000 simulations, at any given performance order, we reject the null hypothesis that the outperforming (underperforming) subsample is due to good luck (poor stock picking skill) at the 95% confidence level and infer that the strategy is genuine (bad luck).

The Stata .ado file, bs_resid.ado, for Procedure II: Jointly resampling both the residuals and common risk factors—Fama and French (2010) bootstrapping simulations

Precodure II repeats *Procedure I* by jointly resampling both the residuals and the common risk factors generated in Step (iii), *ceteris paribus*. The major difference between the two bootstrapping procedures is that *Procedure II* considers the distribution of the residuals conditional on the realization of the systematic risk factors (see, e.g., Fama and French, 2010). Also, *Procedure II* employs the unconditional distribution of the residuals and assumes that the influence of the common risk factors is not fixed at their historical realizations (see, e.g., Kosowski et al., 2006).

```
program bs_resid
version 13.1
syntax, RESidual(varname numeric) MATrix(name)
```

Step (i): The time-series daily excess portfolio returns ($R_t - R_{f,t}$) of the portfolio are regressed on the Fama and French (1993) three factors ($R_t - R_{f,t}$, SMB_t, and HML_t) in each rolling window, to calculate the estimated alphas ($\hat{\alpha}$), factor loadings ($\hat{\beta}, \hat{s}, \hat{h}$), and residuals ($\hat{\varepsilon}_t$). Specifically, R_t and $R_{m,t}$ are the daily return of the portfolio and the FTSE All-Share Index, respectively; $R_{f,t}$ represents the three-month UK T-bill rate; SMB_t and HML_t represent the daily returns on zero-investment factor-mimicking portfolios for size and B/M, respectively. The daily ($R_{m,t} - R_{f,t}$), SMB_t, and HML_t in the UK stock market are collected from the Xfi Centre for Finance and Investment at the University of Exeter (see, Gregory et al., 2013). The excess return of portfolio and risk factors for the Fama and French three-factor model will be specified in the *.do* file, *bs_resid.do*.

```
local xvars : colna `matrix'

local CONS _cons

local xvars : list xvars - CONS
```

Step (ii): For each loop (i.e., each rolling window), we save the coefficient estimates $\{\hat{\alpha}_p, \hat{\beta}_p, \hat{s}_p, \hat{h}_p\}$, the time-series of estimated residuals ($\hat{\varepsilon}_t$), and the *t*-statistic of the intercept ($\hat{t}_{\hat{\alpha}}$), generated from the Step (i).

```
tempvar xb idx y

matrix score double `xb' = `matrix'
```

Step (iii): We randomly select the observations from the saved residual vector $\{\hat{\varepsilon}_t\}$ and the risk factor vectors $\{\hat{\beta}_p, \hat{s}_p, \hat{h}_p\}$in Step (ii), with replacements, in attempts to prepare for a time-series of the pseudo residuals $\{\hat{\varepsilon}_{t_b}^b\}$ and a time-series of the pseudo risk factors $\left\{\left(R_{m,t_b} - R_{f,t_b}\right)^b, SMB_{t_b}^b, HML_{t_b}^b\right\}$ in the next step.

```
gen long `idx' = ceil(_N*runiform())
```

Step (iv): Using the pseudo time-series of residuals and fixed time-series of risk factors, we generate a time-series of pseudo daily excess returns $\left(R_t - R_{f,t}\right)^b$ in each rolling window; the time-series of pseudo excess returns impose the null hypothesis of zero true performance ($\alpha = 0$). We resample both the residuals and the common risk factors, according to Fama and French (2010). Specifically, we randomly select the residuals from the saved residual vector $\{\hat{\varepsilon}_t\}$ in Step (ii), with replacements. Similarly, we generate the pseudo time-series of risk factors by randomly collecting values with replacement from the original risk factor vectors $\{(R_{m,t} - R_{f,t}), SMB_t, HML_t\}$.

```
gen double `y' = `xb'[`idx'] + `residual'[`idx']

regress `y' `xvars', vce(robust)
```

Step (v): We regress the pseudo daily excess returns $\left(R_t - R_{f,t}\right)^b$ generated from Step (iv) against the Fama and French (1993) three factors used in Step (i), generating the simulated $\hat{\alpha}^b$, which represents the sampling variation around zero true performance, purely due to random chance (i.e., sell-side analysts' luck). We repeat the above procedures in each rolling

window and generate a time-series of simulated alphas, $\{\hat{\alpha}_w^b\}$, and their corresponding *t*-statistics, $\{\hat{t}_{\alpha_w}^b\}$, where w is the number of rolling windows, throughout the whole sample period. The standard errors are corrected for heteroscedasticity through the Newey–West procedure with 0 lag.

```
end
```

The Stata do file, "bs_resid.do", for bootstrapping simulation

```
set seed 12345
```

```
regress Rpt_Rft Rm_Rf SMBt HMLt, vce(robust)
```

The daily excess portfolio returns (i.e., the difference between the daily return of the portfolio and the three-month UK T-bill rate) are regressed on the Fama and French (1993) three factors (i.e., the difference between the FTSE All-Share Index and the three-month UK T-bill rate, as well as the daily returns on zero-investment factor-mimicking portfolios for size and B/M. The standard errors are corrected for heteroscedasticity through the Newey–West procedure with 0 lag.

```
matrix b = e(b)
```

The *b* matrix contains the coefficients from the original regression. This matrix is used to generate the list of independent variables stored in the macro *xvars*, which is a place where we can hold a piece of text including numeric as well as alphabetic characters, and then to produce/simulate a new dependent variable with the resampled residuals.

```
local icons = colnumb(b, "_cons")
```

```
matrix b[1,`icons'] = 0
```

The time-series of pseudo excess returns impose the null hypothesis of zero true performance ($\alpha = 0$).

```
predict double resid, residuals
```

```
histogram resid
```

```
simulate _b _se, reps(10000) : bs_resid, res(resid) mat(b)
```

We use the *bs_resid.do* file to execute the *bs_resid.ado* file for the corresponding bootstrapping simulation methods. Specifically, *_b* represents the coefficient on risk factor; *_se* represents the standard error; *reps()* specifies the number of replications to be performed, say, 10,000 in this study.

APPENDIX 55C: SUMMARY STATISTICS AND CORRELATIONS FOR THE DAILY RISK FACTOR RETURNS

This appendix presents descriptive statistics (in Panel A) and correlation (in Panel B) of the daily risk factor returns in the UK stock market over the period January 1997 to June 2015. *Mean* and *Std. Dev.* are the mean and standard deviation of the daily returns; *Min* and *Max* are the minimum and maximum of the daily returns; *t*-statistic is the ratio of the mean return over its standard error. *** stands for the statistical significance at the 1% level. We obtain the daily ($R_{m,t} - R_{f,t}$), SMB_t, and HML_t in the UK stock market from the Xfi Centre for Finance and Investment at the University of Exeter, available at https://goo.gl/oDGL47 (see, Gregory et al., 2013). In the construction of the factors and test portfolios, Gregory et al., (2013) only include stocks in the Main Market of the London Stock Exchange (LSE) and exclude financials, foreign companies, and stocks listed in the Alternative Investment Market (AIM). $R_{m,t} - R_{f,t}$ represents the market factor (market risk premium); $R_{m,t}$ represents the total return on the FTSE All-Share Index, and $R_{f,t}$ represents the risk free rate, the return on three month Treasury Bills.

The size (*SMB*) and vaule (*HML*) factors are constructed from six portfolios formed on size (market capitalization) and value/growth (B/M). Gregory et al. (2013) independently sort the sample firms on market capitalization and B/M at the beginning of October in each year. Sorting on market capitalization first, Gregory et al., (2013) form two size groups: small (*S*) and big (*B*) using the median market capitalization of the largest 350 companies in each year as the break point. Then, sorting on B/M, Gregory et al., (2013) form the three B/M groups: High (*H*), medium (*M*), and Low (*L*), using the 30th and 70th of the 350 firms as break points. Using two size and three BTM portfolios, Gregory et al., (2013) form six intersecting portfolios: *SH* (small size and high B/M portfolio), *SM* (small size and medium B/M portfolio), *SL* (small size and low B/M portfolio), *BH* (big size and high B/M portfolio), *BM* (big size and medium B/M portfolio), and *BL* (big size and low B/M portfolio). These portfolios are then used to form the *SMB* and *HML* factors. Specifically, SMB = (*SL* + *SM* + *SH*)/3 – (*BL* + *BM* + *BH*)/3 and *HML* = (*SH* + *BH*)/2 – (*SL* + *BL*)/2.

	$R_{m,t} - R_{f,t}$ (%)	SMB_t (%)	HML_t (%)
Panel A: Descriptive statistics			
Mean	0.02	–0.00	0.01
Std. Dev.	1.18	0.80	0.70
Min.	–8.36	–6.30	–4.19
Max.	9.20	3.56	5.78
t-statistic	0.93	–0.06	0.94
Panel B: Correlations			
	$R_{m,t} - R_{f,t}$	SMB_t	HML_t
$R_{m,t} - R_{f,t}$	1.00		
SMB_t	–0.54***	1.00	
HML_t	0.09***	–0.01	1.00