The performance of UK stock recommendation revisions: Does brokerage house reputation matter?

Su, C., Zhang, H. & Joseph, N.

Original citation:
Su, C, Zhang, H & Joseph, N 2020, 'The performance of UK stock recommendation revisions: Does brokerage house reputation matter?' Journal of International Finance & Economics, vol. (In-press), DOI: 10.1002/ijfe.2312, pp. (In-press).
https://dx.doi.org/10.1002/ijfe.2312

DOI 10.1002/ijfe.2312
ISSN 1076-9307
ESSN 1099-1158

Publisher: Wiley

Received: 23 September 2019 | Revised: 23 September 2020 | Accepted: 28 September 2020

DOI: 10.1002/ijfe.2312

RESEARCH ARTICLE

WILEY

The performance of UK stock recommendation revisions: Does brokerage house reputation matter?

Chen Su[1] | **Hanxiong Zhang**[2] | **Nathan Lael Joseph**[3]

[1]Newcastle University Business School, Newcastle University, Newcastle upon Tyne, UK

[2]Sheffield University Management School, The University of Sheffield, Sheffield, UK

[3]Centre for Financial and Corporate Integrity, Coventry University, Coventry, UK

Correspondence
Chen Su, Newcastle University Business School, Newcastle University, 5 Barrack Road, Newcastle upon Tyne NE1 4SE, UK.
Email: chen.su@ncl.ac.uk

Funding information
Faculty of Humanities and Social Sciences (HaSS), Newcastle University under the Faculty Research Fund, Grant/Award Number: OSR/0372/FR19/0001

Abstract

Our study examines the impact of brokerage house (BH) reputation on the performance of investment strategies following stock recommendation revisions in the UK stock market. We develop two alternative proxies for BH reputation based either on the past positions on the annual *Institutional Investor* (*II*) All-Europe Research Team or on the past recommendation performance of BHs. We find that BH reputation proxied by the past *II* rankings has no significant impact on the recommendation performance, suggesting that the *II* rankings are largely "popularity contests". However, BH reputation proxied by the past year recommendation performance of BHs has a significantly positive impact on the recommendation performance in the next year, implying that the recommendation performance of BHs in the UK market is persistent. The bootstrap simulations further confirm that the observed performance persistence could be due to BH skill rather than BH luck (i.e., random chance).

KEYWORDS

Bootstrap simulations, brokerage house reputation, financial analysts, performance persistence, stock recommendation revisions

1 | INTRODUCTION

Brokerage houses (BHs) constitute a large segment of the financial services industry and play an important intermediary role to connect buyers and sellers in the capital markets. Analysts working for BHs collect and analyse various publicly available information and/or sensitive information not readily available to the public and then make stock recommendations, which have been widely considered to be valuable to investors when making investment decisions. Whether stock recommendations made by BHs can truly create investment value and promote market efficiency have been of great interest to financial academics and investment professionals, though they are clearly at odds with each other (Barber, Lehavy, McNichols, & Trueman, 2001). Specifically, the semi-strong form of market efficiency posits that investors should not be able to trade profitably using any publicly available information, such as stock recommendations. Although Barber et al. (2001) show the profitability of investment strategies based on stock recommendations, these investment strategies are not easily exploitable in practice as they require a great deal of trading and generate considerable transaction costs (see, also, Jegadeesh, Kim, Krische, & Lee, 2004; Mikhail, Walther, & Willis, 2004). However, BHs, in particular, those bulge bracket houses, invest large amounts of money and resources on security analysis, and investors pay millions of dollars every year to purchase these recommendation data, presumably because they both believe that stock recommendations are able to generate superior returns (Ivkovic & Jegadeesh, 2004).

BHs, as repetitive players in the capital markets, obtain and accumulate their reputation capitals by providing investors with high value-added services to facilitate transactions. In particular, BH reputation could be affected by the accuracy and reliability of their stock recommendations and/or their rankings in financial press (Fang, 2005). These observations provide a compelling empirical motivation for our investigation into the role of BH reputation in improving the recommendation performance and distinguish our study from prior studies that mainly focus on analyst reputation but largely ignore the impact of BH reputation (see, e.g., Stickel, 1992 & 1995; Leone & Wu, 2007; Emery & Li, 2009; Fang & Yasuda, 2009 & 2014; Kucheev, Ruiz, & Sorensson, 2017; among others). Specifically, the objective of this study is to examine the performance of investment strategies following UK stock recommendation revisions, with specific emphasis on the impact of BH reputation. On the one hand, an investigation into this question will not only shed fresh light on the extent to which specific BH reputation measurements are related to the stock recommendation performance, but also provide insights into whether investors have an ex-ante reliable way of enhancing their performance by following stock recommendations issued by certain types of BHs, or ignoring, those of others. On the other hand, despite the existence of extensive analyst research in the US market, there is surprisingly little related research in other developed markets. The existing limited analyst research in the UK stock market shows some different evidence from that in the US market (see, e.g., Dimson & Fraletti, 1986; Ryan & Taffler, 2006; Su, Zhang, Bangassa, & Joseph, 2019; Forbes, Murphy, O'Keeffe, & Su, 2020). Jegadeesh and Kim (2006) point out that an in-depth examination in other developed markets will give us a comprehensive picture of the extent to which the type of stock recommendations is more valuable. The UK stock market, a highly developed and sophisticated market, provides an appropriate setting to conduct such analyst research. In particular, the institutional settings and trading practices of the UK market are partially different from, and independent of, those in the US market;[1] as a result, the existing US evidence may not justify the UK investment practices.

Using a comprehensive sample of 58,647 UK stock recommendation revisions, uniquely created by *Morningstar Company Intelligence*, we examine the impact of BH reputation on the performance of upgrades and downgrades over the period of January 1995 to June 2013. We develop two alternative proxies for BH reputation in the UK stock market and evaluate whether one proxy for BH reputation is systematically superior to the other. First, in each year T, we identify more prestigious BHs as the top five based on their past year ($T - 1$) positions on the annual *Institutional Investor* (*II*) All-Europe Research Team, which has been ignored in prior analyst research. However, the *II* rankings, mainly based on the annual poll of money managers, have been widely criticized by institutional investors and analysts as being "popularity contests" with no substance (see more discussions in Section 2). In practice, it could be reasonable for the investing public to expect more prestigious BHs to issue more valuable stock recommendations, as they maintain closer ties with corporate management and provide more resources to support market research. The second proxy for BH reputation, developed in our study, is based on the past stock recommendation performance of BHs. For example, in each year $T - 1$, we calculate the average return of stock recommendation revisions made by each BH and then identify the most and worst prestigious BHs in each year T as those with the highest and lowest past year ($T - 1$) recommendation performance, respectively. Accordingly, in our study, we test the following two main hypotheses: (a) The (ir) relevant BH reputation hypothesis—BH reputation is (un)related to the recommendation performance; and (b) the BH skill/luck hypothesis—if there exists a significant positive relationship between BH reputation and the recommendation performance, this relationship is due to BH skill, rather than due to BH luck (i.e., random chance).

We find some interesting evidence that not only complements the existing analyst literature but has particular relevance to investors and policymakers in understanding the role of BHs in an important developed market context. First, we find that BH reputation proxied by the *II* rankings has no impact on the recommendation performance in the UK market. That is, it is unlikely for investors to make profits by following upward or downward revisions in the UK market, irrespective of whether they are issued by BHs with high or low past *II* rankings, which seems different from the US evidence (see, e.g., Leone & Wu, 2007; Fang & Yasuda, 2014; Kucheev et al., 2017). We conjecture that the positions of BHs on the annual *II* All-Europe Research Team do not play an important role in determining their reputation in the UK market, compared with the influence of its counterpart in the US market (see Fang & Yasuda, 2014; Hong & Kubik, 2003; Hong, Lim, & Stein, 2000). Our results confirm that it is unlikely for investors to make profits by following upward or downward revisions in the UK market, irrespective of whether they are made by BHs with high or low past *II* rankings. However, when BH reputation is proxied purely by the past stock recommendation performance of BHs, we find a significantly positive relationship between BH reputation and the recommendation performance. That is, BHs that generate superior (inferior) recommendation performance in the past year continue generating superior (inferior) recommendation performance in the next year, indicating that the recommendation performance of BHs in the UK market is persistent, in line with the US evidence (see, Li, 2005).

Furthermore, we test whether the observed performance persistence is simply as a result of BH luck (i.e., random chance) or BH skill. To account for luck, prior studies on portfolio performance evaluation generally use the out-of-sample performance persistence test,[2] which, however, underestimates the likelihood that luck (both good luck and bad luck) can also persist in the short term (see, Neely, Weller, & Ulrich, 2009). To address this problem, we apply the Fama and French (2010) cross-sectional bootstrap simulation method to distinguish BH luck from BH skill (see full details in Section 5). Our simulated results confirm that the reported performance persistence of BHs is not due to BH luck, but due to BH skill.

In summary, to the best of our knowledge, this is the first study that focuses on the relationship between BH reputation and the recommendation performance in the UK. First, we find that there is no relationship between BH reputation, as proxied by the past positions on the annual *II* All-Europe Research Team, and the recommendation performance, in support of Emery and Li (2009) that the *II* rankings are largely "popularity contests." Second, from an investor's perspective, the reported persistence of the recommendation performance of BHs implies that it is likely for investors to make profits by following stock recommendation revisions made by BHs in the UK market, even after controlling for transaction costs, which provide clear evidence of a violation of the semi-strong form of market efficiency. Third, this is the first analyst study that uses the cross-sectional bootstrap simulation method to distinguish BH luck from BH skill, showing that more prestigious BHs have sufficient skills in persistently generating superior recommendation performance, while less prestigious BHs lack such skills and persistently generate inferior recommendation performance.

The remainder of the paper is structured as follows. The next section reviews the relevant analyst literature and develops our hypotheses. Section 3 describes our data and empirical methodology. Section 4 presents our main empirical results, followed by bootstrap simulations in Section 5. The final section concludes.

2 | RELATED ANALYST LITERATURE AND HYPOTHESIS DEVELOPMENT

There has been substantial growth in analyst research regarding the performance of stock recommendations issued by BHs since two influential studies of Stickel (1995) and Womack (1996). One of the most active areas in this stream of research is the impact of analyst reputation on the stock recommendation performance (see, e.g., Jackson, 2005; Leone & Wu, 2007; Bagnoli, Watts, & Zhang, 2008; Emery & Li, 2009; Loh & Stulz, 2011; Fang & Yasuda, 2014; Kucheev et al., 2017; among others). The extant literature provides mixed evidence on the relationship between analyst reputation and the stock recommendation performance. For example, Stickel (1995) examines 16,957 Buy and Sell recommendations issued by 1,510 analysts over the period 1988–1991, showing that analyst reputation, proxied by their positions on the annual *II* All-America Research Team, is positively related to the short-term price reaction, although the influence of analyst reputation appears to be a temporary price pressure effect. Leone and Wu (2007) also document such a positive relationship over the period 1991–2000 and confirm that the recommendation performance of All-Star analysts is persistent. They attribute the performance persistence to All-Star analysts' superior skill, the result of which suggests that the *II* rankings serve a meaningful role in identifying more prestigious analysts. In addition, Kucheev et al. (2017) find that All-Star analysts outperform their non-All Star counterparts for Buy and Strong Buy recommendations, but not for Sell and Strong Sell recommendations. However, the *II* rankings have been widely criticized by institutional investors and analysts as being "popularity contests" with no substance (see, e.g., Emery & Li, 2009), leading the *Wall Street Journal* (*WSJ*) to create its own rankings based on the past stock recommendation performance (see details on the differences between the *II* and *WSJ* rankings in Appendix A). Using a sample of 20,239 stock recommendations issued by 5,941 analysts over the period 1993–2005, Emery and Li (2009) comparatively examine the *II* and *WSJ* rankings, but they find that neither of them has any significant impact on the recommendation performance.

Generally speaking, more prestigious BHs receive more attention in the market and their stock recommendations are more likely to be recognized as a meaningful indicator of a firm's future prospects. If investors realize that they are adversely affected by biased stock recommendations made by a BH on purpose, it will become quite costly and difficult for the BH to convince investors to follow its stock recommendations in the future; the damage to the BH's reputation will be immediate and long-lasting. As reputation is extremely valuable and also fragile, more prestigious BHs invariably attempt to protect their reputation capitals by resisting pressures to make due diligence stock recommendations, which, in turn, drive the market to efficiency (see, Fang & Yasuda, 2009; Mehran & Stulz, 2007).

Accordingly, we extend the commonly employed research design at the analyst level to the BH level and test the following (ir)relevant BH reputation hypothesis that BH reputation is (un)related to the stock recommendation performance:

Hypothesis 1a *BH reputation has no impact on the performance of UK stock recommendation revisions.*

Hypothesis 1b *BH reputation has a significantly positive impact on the performance of UK stock recommendation revisions.*

Although prior analyst research has established a positive relationship between analyst reputation and the stock recommendation performance (see, Kucheev et al., 2017; Stickel, 1995), it is not necessarily true that a BH's reputation is the sum of individual reputation of all analysts employed by the BH, as unskilled analysts are likely to piggyback on the reputation of the BH.[3] More prestigious BHs, compared with their less prestigious counterparts, are supposed to possess greater access to in-house information resources, for example, economists, market strategy experts, and technical analysts, of which in-house economic advisers are the most highly rated in terms of the perceived usefulness (Clement, 1999). As such, even unskilled analysts are able to benefit from these in-house information resources in making valuable stock recommendations. Moreover, Fang and Yasuda (2014, p. 236) argue that some All-Star analysts are not really skilled but achieve their first All-Star status simply due to luck (see, also, Leone & Wu, 2007). Once these unskilled analysts achieve All-Star status, they gain superior access to the management of the firms they cover, which improves the quality of their stock recommendations; in turn, the market is expected to react more strongly to their stock recommendations (see, Stickel, 1995; Hong et al., 2000; Leone & Wu, 2007; Fang & Yasuda, 2009 & 2014; Kucheev et al., 2017). Using a large sample of 392,711 stock recommendations from October 1993 to December 2009, Fang and Yasuda (2014) first divide all stock recommendations into different reputation groups, according to analysts' positions on the annual *II* All-America Research Team, and then construct dynamic portfolios in each group based on these recommendations. To address the question of whether All-Star analysts can really generate superior recommendation performance, they calculate and compare the portfolio alphas in various reputation groups, confirming that skill differences exist among analysts. In addition, they report that superior recommendation performance of All-Star analysts is not significantly eroded after the adoption of Reg-FD in October 2000, suggesting that the superior recommendation performance of All-Star analysts is persistent and the performance persistence is not entirely due to their luck, but due to their better access to company management and/or market influence.

Like Fang and Yasuda (2014), prior studies on portfolio performance evaluation generally account for luck by using the out-of-sample test. For example, Carhart (1997) sorts mutual funds into the winner and loser portfolios based on the lagged one-year returns to examine the short-term performance persistence. Although the out-of-sample performance persistence test is quite popular, it underestimates the likelihood that luck (both good and bad luck) can also persist in the short term, as the allocation of sub-samples (such as the winner and loser portfolios) could be largely based on noises (see, Fama & French, 2010; Kosowski, Timmermann, Wermers, & White, 2006). In addition, the sub-samples may not be directly comparable, as the separation of the whole sample is somewhat arbitrary and thus lacks the expected objectivity (see, Hsu & Kuan, 2005). Therefore, even if BHs as a group do not show performance persistence, we cannot rule out the possibility that there exist relatively fewer BHs with superior recommendation performance and their performance is persistent.

Accordingly, we employ the Fama and French (2010) cross-sectional bootstrap simulation method, which is able to distinguish BH luck from BH skill (see full details in Section 5), to test the following BH skill/luck hypotheses:

Hypothesis 2a *If the stock recommendation performance of BHs in the UK market is persistent, this is due to BH skill, rather than due to BH luck (i.e., random chance).*

Hypothesis 2b *If the stock recommendation performance of BHs in the UK market is persistent, this is not due to BH skill, but due to BH luck (i.e., random chance).*

3 | DATA AND METHODOLOGY

3.1 | Sample selection and descriptive statistics

We obtain the real-time stock recommendations from the Morningstar Extracted Data File: Historic Broker

Recommendations for UK Registered and UK Listed Companies, created by Morningstar Company Intelligence. Each stock recommendation record contains information on the name of the recommended stock, the name of the BH issuing the recommendation, the starting and expiration recommendation dates, and a rating between 1 and 9 (1 = strong buy; 2 = buy; 3 = weak buy; 4 = weak buy/hold; 5 = hold; 6 = hold/sell; 7 = weak sell; 8 = sell; and 9 = strong sell). We exclude stock recommendations that omit the name of BHs, those without releasing the expiration dates, and/or those with data errors. To allow for an intuitive comparison with prior US analyst studies, we reclassify all original stock recommendations into five categories: Strong Buys (1 and 2),

TABLE 1 The distribution of UK stock recommendation revisions

Panel A: The matrix of stock recommendation revisions

	To new rating						
From old rating	**Total**	**%**	**Strong buys (1 and 2)**	**Buys (3 and 4)**	**Holds (5)**	**Sells (6 and 7)**	**Strong sells (8 and 9)**
Strong buys (1 and 2)	18,427	31.42	—	4,927	12,351	329	820
Buys (3 and 4)	9,805	16.72	4,768	—	4,281	577	179
Holds (5)	21,328	36.37	10,758	4,281	—	2,184	4,105
Sells (6 and 7)	3,482	5.94	245	510	2,026	—	701
Strong sells (8 and 9)	5,605	9.56	698	144	4,158	605	—
Overall	**58,647**	—	**16,469**	**9,862**	**22,816**	**3,695**	**5,605**
%	—	100.00	28.08	16.82	38.90	6.30	9.90

Panel B: The distribution of up/downward revisions in the *up/downgrade* portfolio

	The *upgrade* portfolio				The *downgrade* portfolio			
The recommendation year	**No. of covered firms**	**No. of BHs**	**Average rating**	**No. of upward revisions**	**No. of covered firms**	**No. of BHs**	**Average rating**	**No. of downward revisions**
1995	284	23	1.14	514	355	21	3.67	814
1996	513	34	1.21	1,246	501	33	3.52	1,436
1997	552	41	1.23	1,690	547	41	3.55	2,024
1998	558	35	1.28	1,652	559	35	3.58	2,048
1999	523	35	1.26	1,630	523	37	3.52	1,527
2000	416	35	1.26	1,137	421	36	3.45	1,225
2001	447	36	1.26	1,126	518	36	3.64	1,752
2002	478	38	1.24	1,113	486	39	3.74	1,260
2003	446	35	1.22	1,060	488	35	3.64	1,422
2004	490	40	1.24	1,316	474	41	3.63	1,586
2005	510	39	1.27	1,394	530	41	3.61	1,866
2006	490	41	1.26	1,321	509	41	3.54	1,532
2007	477	35	1.24	1,279	451	35	3.52	1,175
2008	404	35	1.17	934	453	34	3.68	1,344
2009	464	41	1.18	1,458	461	45	3.59	1,589
2010	392	35	1.20	933	352	32	3.45	974
2011	339	32	1.17	805	341	31	3.44	865
2012	292	27	1.17	592	342	27	3.48	818

(Continues)

TABLE 1 (Continued)

Panel B: The distribution of up/downward revisions in the *up/downgrade* portfolio

	The *upgrade* portfolio				**The *downgrade* portfolio**			
The recommendation year	**No. of covered firms**	**No. of BHs**	**Average rating**	**No. of upward revisions**	**No. of covered firms**	**No. of BHs**	**Average rating**	**No. of downward revisions**
2013 (January to June)	151	23	1.20	204	175	21	3.55	270
Overall (January 1995 – June 2013)	**1,639**	**95**	**1.23**	**21,404**	**1,760**	**95**	**3.58**	**25,527**

Note: Panel A of this table presents the matrix of 58,647 UK stock recommendation revisions over the period January 1995 to June 2013, while Panel B presents on the distribution of 21,404 upward (25,527 downward) changes in stock recommendations in the *upgrade* (*downgrade*) portfolio over the sample period by the recommendation year, in terms of the number of recommended firms, the number of brokerage houses (BHs), as well as the average rating and number of stock recommendation revisions. We exclude all utilities and financials from the recommended firms and all stock recommendations are obtained from *Morningstar Company Intelligence*. 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, which are reclassified into five categories: Strong Buys (1 and 2), Buys (3 and 4), Holds (5), Sells (6 and 7), and Strong Sells (8 and 9). An *upgrade* portfolio consists of all upward revisions to Strong Buys or Buys from previous Strong Sells, Sells, or Holds, while a *downgrade* portfolio consists of all downgrades to Strong Sells, Sells, or Holds from previous Strong Buys or Buys. The *upgrade* portfolio does not include upward revisions from Strong Sells to Holds, from Strong Sells to Sells, and from Sells to Holds, which can also be interpreted as negative recommendations, while the *downgrade* portfolio does not include downward revisions from Strong Buys to Buys, which can also be interpreted as positive recommendations. We report the average rating for stock recommendation revisions based on the five-point rating scale.

Buys (3 and 4), Holds (5), Sells (6 and 7), and Strong Sells (8 and 9). We exclude all utilities and financials due to their highly regulated nature, according to the two-digit Industry Classification Benchmark (ICB) codes 30, 35, and 65. Like Loh and Stulz (2011), we also exclude those stock recommendations made in the 3 days around quarterly earnings announcements. Furthermore, we require (a) that the gap between the starting and expiration recommendation dates is less than 365 days to ensure that the BH actively follows the recommended stock; and (b) that the relevant financial data of the recommended stocks are available from the London Share Price Database (LSPD).

In addition, stock recommendations often remain unchanged for relatively long time periods, and thus become stale and less informative over time (see, Boni & Womack, 2006; Jegadeesh et al., 2004; Jegadeesh & Kim, 2006). Therefore, our study exclusively focuses on stock recommendation revisions—upgrades and downgrades—that tend to convey more valuable information. Panel A of Table 1 presents the matrix of our final sample of 58,647 UK stock recommendation revisions over the period January 1995 to June 2013, that is, 44.88% are Strong Buys and Buys, 38.90% are Holds, and 16.20% are Sells and Strong Sells. As such, our sample is much larger than has been employed in prior UK analyst studies (see, e.g., Dimson & Fraletti, 1986; Ryan & Taffler, 2006).

3.2 | BH reputation measurements and research design

In our study, we develop two alternative proxies for BH reputation in the UK stock market. Specifically, we pair the adjacent 2 years ($T-1$ and T) into a ranking year $T-1$ and an evaluation year T. For example, if 1995 is a ranking year, then 1996 is the evaluation year. First, in each year T, we identify more prestigious BHs as the top five based on their past year ($T-1$) positions on the annual *II* All-Europe Research Team. The second proxy for BH reputation is directly based on the past recommendation performance of BHs. That is, in each year ($T-1$), we calculate the average abnormal return of stock recommendation revisions issued by each BH, using the intercept term (*alphas*) derived from various multi-factor asset pricing models, for example, (a) the Fama and French (1993) three-factor model (3F model, hereafter), (b) the Carhart (1997) four-factor model (4F model, hereafter), and (c) the Fama and French (2015) five-factor model (5F model, hereafter). We then identify the most and worst prestigious BHs in year T as those with the highest and lowest past year ($T-1$) recommendation performance (e.g., top quintile vs. bottom quintile or Best five vs. non-Best five), respectively.

To evaluate the impact of BH reputation on the recommendation performance and to compare whether one BH reputation measurement is systematically superior to the other, we divide the recommended stocks into various BH reputation groups in each evaluation year T over the whole sample period. In each BH reputation group, we construct two portfolios: (a) An *upgrade* portfolio, consisting of all stocks with upward revisions to Strong Buy or Buy recommendations from previous Strong Sell, Sell, or Hold recommendations; and (b) a *downgrade* portfolio, consisting of all stocks with downward revisions to Strong Sell, Sell, or Hold recommendations from previous Strong Buy or Buy recommendations. Panel B of Table 1 presents the distribution of upward and downward revisions included in the *upgrade* and *downgrade* portfolios, respectively, in each recommendation year.[4] The *up/downgrade* portfolio is updated daily; for each revision, the recommended stock enters the *up/downgrade* portfolio at the close of trading on the day the revision is announced. If an up/downward revision is announced on a non-trading day, the recommended stock is added into the *up/downgrade* portfolio at the close of the next trading day, and remains in the portfolio until the stock is either down/upgraded or dropped from coverage by the BH. If a stock is recommended by more than one BH on a given date, then that stock will appear multiple times in the *up/downgrade* portfolio on that date, once for each BH.[5]

Like Barber, Lehavy, and Trueman (2007), we apply an equal monetary investment in each stock recommendation revision, and calculate the daily value-weighted return to the *up/downgrade* portfolio on date t:[6]

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

where $R_{i,t}$ represents the daily return for the recommended stock i on date t;[7] n_t represents the number of up/downward revisions in the *up/downgrade* portfolio p on date t; $x_{i,t}$ represents the compounded daily return for the recommended stock i from the closing of trading on the revision date through date $t-1$.

In each evaluation year, T, we estimate the gross returns to the *upgrade* and *downgrade* portfolios in each BH reputation group using the intercept term of $\alpha_{p,T}$ derived from various multi-factor asset pricing models, for example, the 3F model, 4F model, and 5F model:

$$R_{p,t} - R_{f,t} = \alpha_{p,T} + \beta_p\left(R_{m,t} - R_{f,t}\right) + s_p\mathrm{SMB}_t + h_p\mathrm{HML}_t + \varepsilon_{p,t}, \tag{2}$$

$$\begin{aligned} R_{p,t} - R_{f,t} = {} & \alpha_{p,T} + \beta_p\left(R_{m,t} - R_{f,t}\right) + s_p\mathrm{SMB}_t \\ & + h_p\mathrm{HML}_t + m_p\mathrm{MOM}_t + \varepsilon_{p,t}, \end{aligned} \tag{3}$$

$$\begin{aligned} R_{p,t} - R_{f,t} = {} & \alpha_{p,T} + \beta_p\left(R_{m,t} - R_{f,t}\right) + s_p\mathrm{SMB}_t + h_p\mathrm{HML}_t \\ & + r_p\mathrm{RMW}_t + c_p\mathrm{CAM}_t + \varepsilon_{p,t}, \end{aligned} \tag{4}$$

where $R_{p,t}$ and $R_{m,t}$ are the daily return on the *up/downgrade* portfolio p and on the FTSE All-Share Index, respectively; $R_{f,t}$ represents the daily 3-month UK T-bill rate; SMB_t, HML_t, and MOM_t represent the daily returns on zero-investment factor-mimicking portfolios for size, book-to-market (B/M), and price momentum, respectively; RMW_t and CMA_t represent the daily returns on zero-investment factor-mimicking portfolios for operating profitability and investment, respectively;[8] $\varepsilon_{p,t}$ represents the error term. A significantly positive (negative) $\alpha_{p,T}$ indicates that the *upgrade* (*downgrade*) portfolio in each BH reputation group is profitable after controlling for various market and firm-specific risks.

Barber et al. (2001) argue that investment strategies based on stock recommendations require a great deal of trading and generate considerable transaction costs, so we evaluate the recommendation performance as the average daily *net* abnormal returns to the *up/downgrade* portfolio after accounting for transaction costs, that is, the average daily gross returns net of transaction costs (see details of transaction costs in Appendix B).

4 | EMPIRICAL RESULTS

In this section, we present empirical evidence on the relationship between BH reputation and the performance of UK stock recommendation revisions. In Section 4.1, we find that BH reputation, proxied by the past *II* rankings, has no impact on the recommendation performance, while Section 4.2 shows that BH reputation, proxied by the past year recommendation performance, has a significantly positive impact on the recommendation performance in the next year, suggesting that the stock recommendation performance of BHs in the UK market is persistent. In this section, we mainly focus on discussing empirical results under the 3F model, as our results remain qualitatively similar under the 4F and 5F models.

4.1 | BH reputation proxied by the past *II* rankings

Panel A of Table 2 presents the average daily abnormal returns to the *upgrade* and *downgrade* portfolios within two BH reputation groups over the whole sample period,

TABLE 2 The performance of the *up/downgrade* portfolio within each BH reputation group (Top 5 vs. Non-Top 5), according to the past *II* rankings

	The *upgrade* portfolio							The *downgrade* portfolio						
		3F model		4F model		5F model			3F model		4F model		5F model	
BH reputation group	**No.**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**	**No.**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**
Panel A: BH reputation proxied by the past year ($T-1$) *II* rankings														
Top 5	2,025	1.821	(1.56)	1.798	(1.46)	1.621	(1.34)	2,188	−0.660	(−1.21)	−0.625	(−1.34)	−0.615	(−1.25)
Non-top 5	19,379	1.771	(1.43)	1.723	(1.37)	1.543	(1.15)	23,339	−0.568	(−1.35)	−0.537	(−1.46)	−0.524	(−1.35)
Difference (top 5 – Non-top 5)		0.050		0.075		0.078			−0.092		−0.088		−0.096	
t-stat		(1.21)		(1.29)		(1.42)			(−0.82)		(−0.77)		(−0.72)	
χ^2		(2.09)		(2.21)		(2.14)			(−1.76)		(−1.88)		(−1.82)	
Panel B: BH reputation proxied by the past three-year ($T-3$, $T-2$, $T-1$) *II* rankings														
Top 5	2,115	1.796	(1.23)	1.767	(1.21)	1.595	(1.19)	2,284	−0.642	(−1.18)	−0.603	(−1.06)	−0.622	(−1.05)
Non-top 5	19,289	1.774	(1.16)	1.726	(1.18)	1.545	(1.07)	23,243	−0.570	(−1.44)	−0.539	(−1.28)	−0.524	(−1.16)
Difference (top 5 – Non-top 5)		0.022		0.041		0.050			−0.072		−0.066		−0.077	
t-stat		(1.37)		(1.24)		(1.34)			(−0.77)		(−0.75)		(−0.79)	
χ^2		(2.13)		(2.03)		(2.07)			(−1.68)		(−1.61)		(−1.73)	

Note: This table presents the average daily abnormal returns within each BH reputation, according to their past year positions on the annual *II* All-Europe Research Team, over the whole sample period. The recommended stocks are divided into two BH reputation groups in each evaluation year T. Specifically, in each year T, we identify more (less) prestigious BHs as the Top five (non-Top five) based on their past year positions on the annual *II* All-Europe Research Team. In Panel A, BH reputation is proxied by the past year ($T-1$) *II* rankings, while in Panel B, BH reputation is proxied by the past three-year ($T-3$, $T-2$, $T-1$) *II* rankings. Within BH reputation group, we construct two portfolios: (a) an *upgrade* portfolio, consisting of all stocks with upward revisions to Strong Buy or Buy recommendations from previous Strong Sell, Sell, or Hold recommendations; and (b) a *downgrade* portfolio, consisting of all stocks with downward revisions to Strong Sell, Sell, or Hold recommendations from previous Strong Buy or Buy recommendations. Within each BH reputation group, the daily abnormal returns are estimated using the intercept term (*alpha*) derived from various multi-factor asset pricing models, for example, the 3F model, 4F model, and 5F model, after taking the transaction costs into account (see Appendix B) and displayed as basis points. Both parametric test statistics (*t*-stat) and nonparametric test statistics (Kruskal–Wallis χ^2) are employed to test differences in sub-sample average returns.

based on their past year ($T-1$) *II* rankings. The left side of Panel A shows that the *upgrade* portfolio does not generate significantly positive abnormal returns under various multi-factor asset pricing models, no matter whether these upward revisions are issued by more or less prestigious BHs. For example, upward revisions issued by Top 5 and non-Top 5 BHs generate statistically insignificant average daily abnormal returns of 1.821 basis points (t-stat = 1.56; 4.695% annualized) and 1.771 basis points (t-stat = 1.43; 4.564% annualized),[9] respectively, under the 3F model. In addition, to test the differences in subsample average returns, we employ the parametric statistical test (t-stat) and non-parametric statistical test (Kruskal–Wallis χ^2); in particular, the Kruskal–Wallis non-parametric test does not require equal sample sizes and it is robust to departures from normality. Specifically, both parametric and non-parametric tests show that the difference of the average daily abnormal returns to the *upgrade* portfolio between both BH reputation groups is statistically insignificant (t-stat = 1.21; χ^2 = 2.09), implying that the *II* rankings have no impact on the recommendation performance of the *upgrade* portfolio.

The right side of Panel A reports the statistically insignificant abnormal returns to the *downgrade* portfolio within both BH reputation groups. For example, downward revisions issued by Top 5 and non-Top 5 BHs generate insignificantly negative average daily abnormal returns of −0.660 basis points (t-stat = −1.21; −1.649% annualized) and − 0.568 basis points (t-stat = −1.35; −1.421% annualized), respectively, under the 3F model. Our parametric and non-parametric tests show no statistically significant difference of the average daily abnormal returns to the *downgrade* portfolio between the two BH reputation groups (t-stat = 0.82; χ^2 = 1.76). In Panel B of Table 2, we replicate all analyses using BH reputation measurement based on the past three-year ($T-3$, $T-2$, $T-1$) moving average of positions on the annual *II* All-Europe Research Team, showing that our results are qualitatively similar.[10]

Furthermore, we report the average daily abnormal returns to the *up/downgrade* portfolio within both BH reputation groups in each evaluation year, to rule out the concern that our evidence shown in Table 2 is due to the extreme results in a specific year. Figure 1 illustrates consistent evidence that the performance of the *up/downgrade* portfolio is not significantly different between Top 5 and non-Top 5 BHs, based on the *II* rankings, in each evaluation year. This confirms that the *II* rankings of BHs do not play an important role in making valuable stock recommendation revisions in the UK market. Therefore, our evidence supports the *irrelevant Hypothesis* 1a that BH reputation, based on the past *II* rankings, has *no* impact on the recommendation performance. That is, it is unlikely for investors to make profits by following upward or downward revisions in the UK market, irrespective of whether they are issued by BHs with high or low past *II* rankings, which seems different from that reported in the US market (see, e.g., Leone & Wu, 2007; Fang & Yasuda, 2014; Kucheev et al., 2017). The discrepancy could be explained by the less influence of the *II*

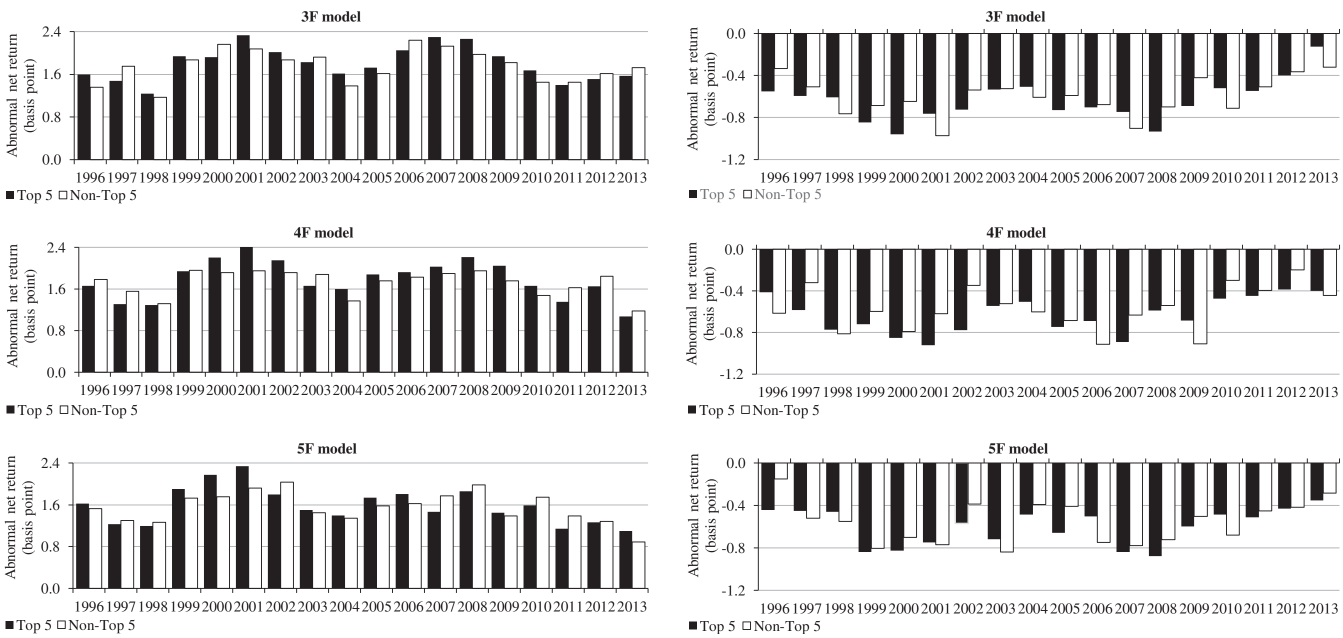

FIGURE 1 The performance of the *up/downgrade* portfolio within two reputation groups (Top 5 vs. Non-Top 5) based on the past year ($T-1$) *II* rankings in each calendar year

TABLE 3 The performance of the *up/downgrade* portfolio within each BH reputation group (*Quintile* 1 vs. *Quintile* 5), according to the past recommendation performance

	The *upgrade* portfolio							The *downgrade* portfolio						
		3F model		4F model		5F model			3F model		4F model		5F model	
BH reputation group	**No.**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**	**No.**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**
Panel A: BH reputation proxied by the past year ($T-1$) recommendation performance														
Quintile 1 (high 20%)	4,281	2.404	(2.71)***	2.343	(2.61)***	2.149	(2.49)**	5,105	−0.684	(−2.74) ***	−0.639	(−2.57)***	−0.617	(−2.48)**
Quintile 2	4,281	2.191	(2.22)**	2.118	(2.12)**	1.938	(2.04)**	5,106	−0.634	(−2.23)**	−0.607	(−2.14)**	−0.597	(−2.10)**
Quintile 3	4,280	1.840	(1.56)	1.806	(1.51)	1.619	(1.48)	5,105	−0.592	(−1.60)	−0.562	(−1.61)	−0.556	(−1.59)
Quintile 4	4,281	1.424	(1.44)	1.395	(1.40)	1.230	(1.36)	5,106	−0.533	(−1.34)	−0.508	(−1.28)	−0.492	(−1.24)
Quintile 5 (low 20%)	4,281	1.020	(1.04)	0.989	(0.99)	0.814	(0.90)	5,105	−0.437	(−0.97)	−0.409	(−0.91)	−0.399	(−0.88)
Difference (high − Low)		1.384		1.354		1.336			−0.247		−0.230		−0.218	
t-stat		(2.93)***		(2.82)***		(2.68)***			(−3.68)***		(−3.38)***		(−3.11)***	
χ^2		(13.08)***		(12.43)***		(11.63)***			(−14.64)***		(−13.65)***		(−11.73)***	
Panel B: BH reputation proxied by the past three-year ($T-3$, $T-2$, $T-1$) recommendation performance														
Quintile 1 (high 20%)	4,281	2.435	(2.74)***	2.372	(2.67)***	2.176	(2.52)**	5,105	−0.676	(−2.70)***	−0.632	(−2.54)**	−0.608	(−2.45)**
Quintile 2	4,281	2.144	(2.17)**	2.072	(2.08)**	1.896	(2.00)**	5,106	−0.642	(−2.26)**	−0.614	(−2.16)**	−0.604	(−2.13)**
Quintile 3	4,280	1.861	(1.59)	1.827	(1.63)	1.637	(1.60)	5,105	−0.599	(−1.72)*	−0.569	(−1.73)*	−0.562	(−1.71)*
Quintile 4	4,281	1.433	(1.35)	1.404	(1.31)	1.237	(1.27)	5,106	−0.522	(−1.32)	−0.497	(−1.25)	−0.481	(−1.21)
Quintile 5 (low 20%)	4,281	1.008	(1.02)	0.977	(0.98)	0.804	(0.89)	5,105	−0.442	(−0.98)	−0.414	(−0.92)	−0.403	(−0.89)
Difference (high − Low)		1.427		1.395		1.372			−0.234		−0.217		−0.205	
t-stat		(2.86)***		(2.77)***		(2.49)**			(−3.15)***		(−2.92)***		(−2.54)**	
χ^2		(9.87)***		(8.47)***		(6.45)**			(−9.32)***		(−8.68)***		(−6.58)**	

Note: This table presents the average daily abnormal returns within each BH reputation over the whole sample period. The recommended stocks are divided into five BH reputation groups in each evaluation year T. Specifically, in each year T, we calculate the average return of stock recommendation revisions issued by each BH, and then we identify the most and worst prestigious BHs in year T as those with the highest and lowest past recommendation performance (i.e., the top and bottom quintiles), respectively. *Quintiles* 1 and 5 represent the most and worst prestigious BH groups generating the highest and lowest recommendation performance in the past year, respectively. In Panel A, BH reputation is proxied by the past year ($T-1$) recommendation performance, while in Panel B, BH reputation is proxied by the past three-year ($T-3$, $T-2$, $T-1$) recommendation performance. Within BH reputation group, we construct two portfolios: (a) an *upgrade* portfolio, consisting of all stocks with upward revisions to Strong Buy or Buy recommendations from previous Strong Sell, Sell, or Hold recommendations; and (b) a *downgrade* portfolio, consisting of all stocks with downward revisions to Strong Sell, Sell, or Hold recommendations from previous Strong Buy or Buy recommendations. Within each BH reputation group, the daily abnormal returns are estimated using the intercept term (*alpha*) derived from various multi-factor asset pricing models, for example, the 3F model, 4F model, and 5F model, after taking transaction costs into account (see Appendix B) and displayed as basis points. Both parametric test statistics (*t*-stat) and nonparametric test statistics (Kruskal–Wallis χ^2) are employed to test differences in sub-sample average returns. ** and *** denote statistical significance at the 5 and 1% levels, respectively.

All-Europe Research Team in the UK market, compared with the influence of its US counterpart (see Fang & Yasuda, 2014; Hong et al., 2000; Hong & Kubik, 2003).

4.2 | BH reputation proxied by the past recommendation performance

4.2.1 | Quintile 1 versus quintile 5

Panel A of Table 3 presents the average daily abnormal returns to the *upgrade* and *downgrade* portfolios within each BH reputation group over the whole sample period, based on their past recommendation performance. Specifically, all BHs are divided into quintiles by their past year ($T-1$) recommendation performance. *Quintiles* 1 and 5 represent the most and worst prestigious BH groups generating the highest and lowest recommendation performance in the past year ($T-1$), respectively. The left side of Panel A shows a significantly positive relationship between BH reputation and the recommendation performance of the *upgrade* portfolio. For example, upward revisions issued by BHs within *Quintile* 1 generate a significantly positive average daily abnormal return of 2.404 basis points (*t*-stat = 2.71; 6.245% annualized), at the 1% level, under the 3F model; an insignificantly positive average daily abnormal return of 1.020 basis points (*t*-stat = 1.04; 2.604% annualized) for *Quintile* 5. Furthermore, our parametric and non-parametric tests show statistically significant difference of the average daily abnormal returns to the *upgrade* portfolio between *Quintiles* 1 and 5 (*t*-stat = 2.93; χ^2 = 13.08), at the 1% level, implying the performance of the *upgrade* portfolio is persistent. That is, the past year ($T-1$) recommendation performance of the *upgrade* portfolio has a significant impact on the recommendation performance in year *T*.

The right side of Panel A reports that the average daily abnormal returns to the *downgrade* portfolios monotonically change from *Quintile* 1 to *Quintile* 5. For example, downward revisions issued by BHs within *Quintile* 1 generate a significantly negative average daily abnormal return of −0.684 basis points (*t*-stat = −2.74; −1.709% annualized), at the 1% level, under the 3F model; an insignificantly negative average daily abnormal returns of −0.437 basis points (*t*-stat = −0.97; −1.095% annualized) for *Quintile* 5. The difference of the average daily abnormal returns to the *downgrade* portfolio between *Quintiles* 1 and 5 is statistically significant (*t*-stat = 3.68; χ^2 = 14.64), at the 1% level, again suggesting that the performance of the *downgrade* portfolio is persistent. Our conclusions hold up well in Panel B of Table 3 when we divide all up/downward revisions into quintiles by BH reputation based on their past three-year ($T-3$, $T-2$, $T-1$) moving average recommendation performance.

Moreover, we report the average daily abnormal returns to the *up/downgrade* portfolio within *Quintiles* 1 and 5 in each evaluation year. Figure 2 illustrates consistent evidence, showing significant difference of the

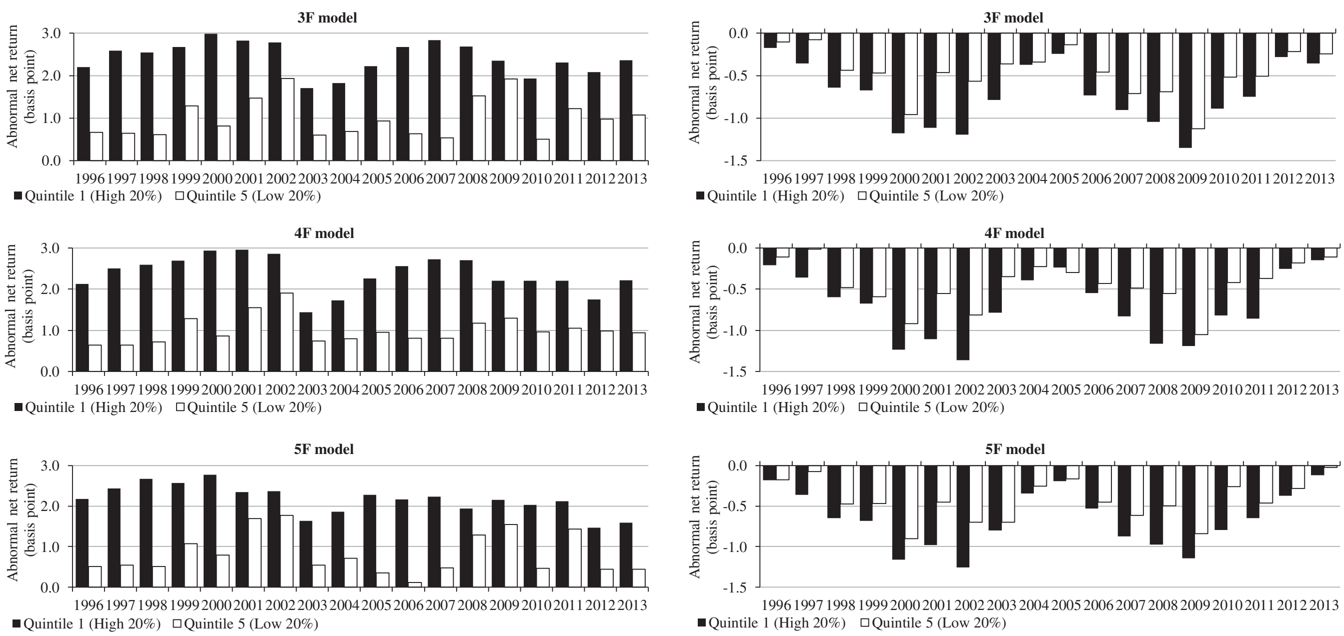

FIGURE 2 The performance of the *up/downgrade* portfolio within two reputation groups (*Quintile* 1 vs. *Quintile* 5) based on the past year ($T-1$) recommendation performance in each calendar year

TABLE 4 The performance of the *up/downgrade* portfolio within each BH reputation group (Best 5 vs. Non-Best 5), according to the past recommendation performance

	The *upgrade* portfolio							The *downgrade* portfolio						
		3F model		4F model		5F model			3F model		4F model		5F model	
BH reputation group	**No.**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**	**No.**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**	**Alpha**	***t*-stat**
Panel A: BH reputation proxied by the past year ($T-1$) recommendation performance														
Best 5	4,182	2.298	(2.44)**	2.191	(2.33)**	2.002	(2.11)**	4,670	−0.667	(−2.51)**	−0.621	(−2.43)**	−0.605	(−2.39)**
Non-best 5	17,222	1.649	(1.26)	1.618	(1.21)	1.440	(1.18)	20,857	−0.556	(−1.46)	−0.528	(−1.40)	−0.516	(−1.36)
Difference (best 5 – Non-best 5)		0.649		0.573		0.562			−0.111		−0.093		−0.089	
t-stat		(2.31)**		(2.23)**		(2.16)**			(−2.89)***		(2.72)***		(−2.65)***	
χ^2		(5.69)**		(5.41)**		(5.14)**			(−12.05)***		(11.77)***		(−11.26)***	
Panel B: BH reputation proxied by the past three-year ($T-3$, $T-2$, $T-1$) recommendation performance														
Best 5	4,487	2.312	(2.48)**	2.221	(2.45)**	2.105	(2.22)**	5,211	−0.674	(−2.55)**	−0.627	(−2.50) **	−0.606	(−2.45)**
Non-best 5	16,917	1.634	(1.25)	1.600	(1.18)	1.403	(1.13)	20,316	−0.551	(−1.41)	−0.524	(−1.34)	−0.514	(−1.30)
Difference (best 5 – Non-best 5)		0.678		0.621		0.702			−0.123		−0.103		−0.092	
t-stat		(2.52)**		(2.35)**		(2.43)**			(−2.96)***		(−2.78)***		(−2.68)***	
χ^2		(5.25)***		(4.94)**		(4.97)**			(−11.78)***		(−10.88)***		(−10.27)***	

Note: This table presents the average daily abnormal returns within each BH reputation over the whole sample period. The recommended stocks are divided into two BH reputation groups in each evaluation year T. Specifically, in each year T, we calculate the average return of stock recommendation revisions issued by each BH, and then we identify Best 5 and non-Best 5 prestigious BHs in year T as those with the Best 5 and non-Best 5 past recommendation performance, respectively. In Panel A, BH reputation is proxied by the past year ($T-1$) recommendation performance, while in Panel B, BH reputation is proxied by the past three-year ($T-3$, $T-2$, $T-1$) recommendation performance. Within BH reputation group, we construct two portfolios: (a) an *upgrade* portfolio, consisting of all stocks with upward revisions to Strong Buy or Buy recommendations from previous Strong Sell, Sell, or Hold recommendations; and (b) a *downgrade* portfolio, consisting of all stocks with downward revisions to Strong Sell, Sell, or Hold recommendations from previous Strong Buy or Buy recommendations. Within each BH reputation group, the daily abnormal returns are estimated using the intercept term (*alpha*) derived from various multi-factor asset pricing models, for example, the 3F model, 4F model, and 5F model, after taking transaction costs into account (see Appendix B) and displayed as basis points. Both parametric test statistics (*t*-stat) and nonparametric test statistics (Kruskal–Wallis χ^2) are employed to test differences in sub-sample average returns. ** and *** denote statistical significance at the 5 and 1% levels, respectively.

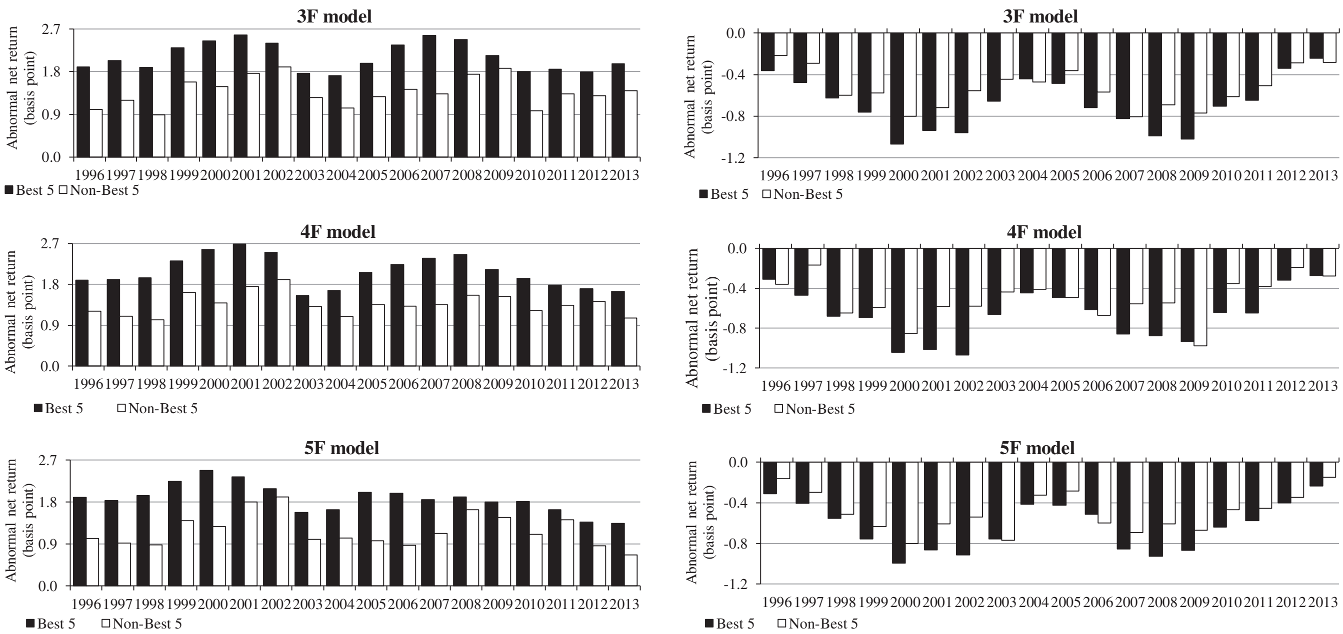

FIGURE 3 The performance of the *up/downgrade* portfolio within two reputation groups (Best 5 vs. Non-Best 5) based on the past year ($T - 1$) recommendation performance in each calendar year

recommendation performance of the *up/downgrade* portfolio between *Quintiles* 1 and 5 in a vast majority of evaluation years. This confirms that the past year recommendation performance of BHs plays an important role in making valuable upward and downward revisions in the next year. Overall, our evidence supports *Hypothesis* 1b that BH reputation, based on the past recommendation performance, has a significantly positive impact on the recommendation performance.

4.2.2 | Best 5 versus non-best 5

To make a point-by-point comparison with BH reputation proxied by the *II* rankings, we also divide all BHs into two groups (i.e., Best 5 vs. non-Best 5), based on their past recommendation performance. Specifically, Panel A of Table 4 presents the average daily abnormal returns to the *upgrade* and *downgrade* portfolios within the two reputation groups over the whole sample period, based on their past year ($T - 1$) recommendation performance. The left side of Panel A shows that upward revisions issued by Best 5 BHs generate a significantly positive average daily abnormal return of 2.298 basis points (t-stat = 2.44; 5.961% annualized), at the 5% level, under the 3F model; an insignificantly positive average daily abnormal return of 1.649 basis points (t-stat = 1.26; 4.243% annualized) for non-Best 5 BHs. Moreover, our parametric and non-parametric tests show statistically significant difference of the average daily abnormal returns to the *upgrade* portfolio between the two groups (t-stat = 2.31; $\chi^2 = 5.69$), at the 5% level.

The right side of Panel A shows that downward revisions issued by Best 5 BHs generate a significantly negative average daily abnormal return of −0.667 basis points (t-stat = −2.51; −1.667% annualized), at the 5% level, under the 3F model; an insignificantly negative average daily abnormal return of −0.556 basis points (t-stat = −1.46; −1.391% annualized) for non-Best 5 group. The difference of the average daily abnormal returns to the *downgrade* portfolio between the two groups is statistically significant (t-stat = −2.89; $\chi^2 = -12.05$), at the 1% level. Our results are qualitatively the same when we divide all BHs into Best5 and non-Best5 groups by BH reputation based on their past three-year (T −3, T −2, T − 1) moving average recommendation performance (see Panel B of Table 4).

In addition, the significant difference of the performance of the *up/downgrade* portfolio between the Best 5 and non-Best 5 groups is shown in all evaluation years (see Figure 3). Overall, our results are consistent with those shown in Section 4.2.1—the recommendation performance of BHs in the UK market is persistent—confirming that the past recommendation performance of BHs plays an important role in making valuable upward and downward revisions in the UK.[11]

5 | BOOTSTRAP SIMULATIONS

Thus far, we find empirical evidence that the recommendation performance of BHs is persistent in the UK, in line with Leone and Wu (2007) and Fang and Yasuda (2014).

In this section, we further test whether the reported performance persistence is due to BH luck (i.e., random chance) or due to BH skill by using the Fama and French (2010) cross-sectional bootstrap simulation method. To the best of our knowledge, we are the first to apply the cross-sectional bootstrap simulation method to distinguish BH luck for BH skill. The bootstrap simulation method, initially proposed by Kosowski et al. (2006) in mutual funds research, resamples the residuals from individual fund returns independently but keeps the effect of common risk factors unchanged 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," and thus jointly resample both of them. In our study, we explain the Fama and French (2010) cross-sectional bootstrap simulation method with the 3F model, but the same bootstrap procedure can be extended to the 4F and 5F models.[12]

First, we estimate the 3F model to generate the estimated abnormal returns, factor loadings, and residuals using the time-series of daily excess returns for the *up/downgrade* portfolio including up/downward revisions made by each BH j, ($R_{j,p,t} - R_{f,t}$; $j = 1, \ldots, N$):[13]

$$R_{j,p,t} - R_{f,t} = \hat{\alpha}_{j,p} + \hat{\beta}_{j,p}\left(R_{m,t} - R_{f,t}\right) + \hat{s}_{j,p}\mathrm{SMB}_t + \hat{h}_{j,p}\mathrm{HML}_t + \hat{\varepsilon}_{j,p,t}. \quad (5)$$

Second, we save the coefficient estimates, $\left\{\hat{\alpha}_{j,p}, \hat{\beta}_{j,p}, \hat{s}_{j,p}, \hat{h}_{j,p}\right\}$, the time-series of estimated residuals, $\left\{\hat{\varepsilon}_{j,p,t}; t = T_{j,0}, \ldots, T_{j,1}\right\}$, and the actual t-statistic of abnormal return, $\hat{t}_{\hat{\alpha}_{j,p}}$, where $T_{j,0}$ and $T_{j,1}$ are the dates of the first and last daily returns available for BH j, respectively.

Third, we generate a pseudo-time-series of resampled residuals, $\left\{\hat{\varepsilon}^b_{j,p,t_b}; t_b = T^b_{j,0}, \ldots, T^b_{j,1}\right\}$, by randomly drawing residuals from the saved residual vector, $\left\{\hat{\varepsilon}_{j,p,t}; t = T_{j,0}, \ldots, T_{j,1}\right\}$, with replacements, where b is the bootstrap simulation index. In the same way, we generate a pseudo-time-series of risk factors, $\left\{\left(R_{m,t_b} - R_{f,t_b}\right)^b, \mathrm{SMB}^b_{t_b}, \mathrm{HML}^b_{t_b}\right\}$, by randomly drawing risk factors from the original risk factor vector, $\{(R_{m,t} - R_{f,t}), \mathrm{SMB}_t, \mathrm{HML}_t\}$, with replacements.

Fourth, we generate a time-series of pseudo-daily excess returns, $\left(R_{j,p,t_b} - R_{f,t_b}\right)^b$, imposing the null hypothesis of zero true recommendation performance ($\hat{\alpha}_{j,p} = 0$):

$$\left\{\left(R_{j,p,t_b} - R_{f,t_b}\right)^b = 0 + \hat{\beta}_{j,p}\left(R_{m,t_b} - R_{f,t_b}\right)^b + \hat{s}_{j,p}\mathrm{SMB}^b_{t_b} + \hat{h}_{j,p}\mathrm{HML}^b_{t_b} + \hat{\varepsilon}^b_{j,p,t_b}\right\}. \quad (6)$$

Fifth, we regress the pseudo-daily excess returns, $\left(R_{j,p,t_b} - R_{f,t_b}\right)^b$, on the three factors:

$$\left(R_{j,p,t_b} - R_{f,t_b}\right)^b = \hat{\alpha}^b_{j,p} + \hat{\beta}_{j,p}\left(R_{m,t_b} - R_{f,t_b}\right) + \hat{s}_{j,p}\mathrm{SMB}_{t_b} + \hat{h}_{j,p}\mathrm{HML}_{t_b} + \hat{\varepsilon}_{j,p,t_b}. \quad (7)$$

Repeating the above steps across all BHs, $j = 1, \ldots, N$ ($N = 95$ for the *upgrade* or *downgrade* portfolio in our study), we obtain a draw from the cross-section of simulated returns, $\left\{\hat{\alpha}^b_{j,p}\right\}$, and their corresponding t-statistics, $\left\{\hat{t}^b_{\hat{\alpha}_{j,p}}\right\}$. The simulated $\hat{\alpha}^b_{j,p}$ and $\hat{t}^b_{\hat{\alpha}_{j,p}}$ represent the sampling variation around a zero true recommendation performance, entirely due to BH luck. We then order all simulated $\hat{t}^b_{\hat{\alpha}_{j,p}}$ into a separate cross-sectional distribution from the best-performing BH to the worst-performing BH. We repeat the above bootstrap simulation 10,000 times, say, $b = 10{,}000$.

Like Fama and French (2010), our study focuses on presenting the distribution of the t-statistics of the actual return, $\hat{t}_{\hat{\alpha}_{j,p}}$, which represents *information ratio*, because the t-statistic scales the return by its standard errors and thus has superior statistical properties (Cuthbertson, Nitzsche, & O'Sullivan, 2008; Gallefoss, Hansen, Haukaas, & Molnár, 2015). We compare the actual t-statistic of each BH with its 10,000 simulated t-statistic. For the *upgrade* (*downgrade*) portfolio, if the simulated t-statistics are greater (less) than the actual t-statistics in less than 5% of the 10,000 simulations, we reject the hypothesis that the statistically significant portfolio performance is due to BH luck, and vice versa.

Table 5 presents the actual t-statistics, $\hat{t}_{\hat{\alpha}_{j,p}}$, against the average simulated t-statistics, $\hat{t}^b_{\alpha_{j,p}}$, for each BH at the orders from the best-performing BH to the worst-performing BH, along with the fraction of the 10,000 simulations that generate higher (lower) simulated t-statistics than the corresponding actual t-statistics for the *upgrade* (*downgrade*) portfolio. Specifically, Panel A of Table 5 shows that, for the *upgrade* portfolios, the actual t-statistics are always above their corresponding average simulated t-statistics, and, in particular, less than 5% of the 10,000 simulated t-statistics are higher than their corresponding actual t-statistics for the best-performing BHs (e.g., the Best 5 BHs). For example, for the #01 (#02) BH, only 1.44% (2.49%) of the simulated t-statistics are higher than the actual t-statistics, clearly suggesting that

TABLE 5 Bootstrap simulation results

Panel A: The *upgrade* portfolio

	3F model			4F model			5F model		
BH rank	$\hat{t}_{\hat{\alpha}j,p}$	$\hat{t}^{b}_{\hat{\alpha}j,p}$	$\%\left(\hat{t}^{b}_{\hat{\alpha}j,p} > \hat{t}_{\hat{\alpha}j,p}\right)$	$\hat{t}_{\hat{\alpha}j,p}$	$\hat{t}^{b}_{\hat{\alpha}j,p}$	$\%\left(\hat{t}^{b}_{\hat{\alpha}j,p} > \hat{t}_{\hat{\alpha}j,p}\right)$	$\hat{t}_{\hat{\alpha}j,p}$	$\hat{t}^{b}_{\hat{\alpha}j,p}$	$\%\left(\hat{t}^{b}_{\hat{\alpha}j,p} > \hat{t}_{\hat{\alpha}j,p}\right)$
#01 (best)	3.86	3.28	1.44	3.75	3.12	1.57	3.57	3.38	1.45
#02	3.35	2.90	2.49	3.31	2.81	2.75	3.21	2.93	2.81
#03	2.98	2.59	3.89	2.96	2.52	3.30	2.88	2.61	3.45
#04	2.50	2.17	4.01	2.48	2.11	4.26	2.41	2.19	4.44
#05	2.22	1.89	4.86	2.16	1.83	5.79	2.09	1.94	4.68
#06–10	1.56	1.36	10.83	1.56	1.32	11.94	1.51	1.36	10.76
#11–20	1.44	1.21	14.12	1.39	1.03	14.83	1.18	1.24	14.67
#21–30	1.24	0.95	19.00	1.08	0.69	19.69	0.79	1.08	19.57
#31–40	0.80	0.64	22.06	0.73	0.51	23.57	0.58	0.73	25.65
#41–50	0.56	0.45	27.25	0.52	0.39	28.30	0.45	0.49	26.87
#51–60	0.33	0.27	33.02	0.31	0.23	31.44	0.26	0.29	33.87
#61–70	0.25	0.20	37.63	0.23	0.18	38.58	0.21	0.21	39.12
#71–80	−0.28	−0.28	43.42	−0.28	−0.34	42.85	−0.35	−0.35	43.66
#81–90	−0.66	−0.69	48.77	−0.67	−0.77	48.68	−0.75	−0.83	46.75
#91	−1.17	−1.26	52.64	−1.23	−1.29	51.58	−1.26	−1.47	52.34
#92	−1.51	−1.63	57.76	−1.59	−1.66	58.03	−1.62	−1.89	58.42
#93	−2.07	−2.23	61.55	−2.18	−2.28	62.44	−2.22	−2.59	61.90
#94	−2.54	−2.72	66.69	−2.65	−2.82	66.06	−2.75	−3.18	65.41
#95 (worst)	−3.29	−3.31	68.91	−3.43	−3.34	69.13	−3.59	−3.40	69.72

Panel B: The *downgrade* portfolio

	3F model			4F model			5F model		
BH rank	$\hat{t}_{\hat{\alpha}j,p}$	$\hat{t}^{b}_{\hat{\alpha}j,p}$	$\%\left(\hat{t}^{b}_{\hat{\alpha}j,p} < \hat{t}_{\hat{\alpha}j,p}\right)$	$\hat{t}_{\hat{\alpha}j,p}$	$\hat{t}^{b}_{\hat{\alpha}j,p}$	$\%\left(\hat{t}^{b}_{\hat{\alpha}j,p} < \hat{t}_{\hat{\alpha}j,p}\right)$	$\hat{t}_{\hat{\alpha}j,p}$	$\hat{t}^{b}_{\hat{\alpha}j,p}$	$\%\left(\hat{t}^{b}_{\hat{\alpha}j,p} < \hat{t}_{\hat{\alpha}j,p}\right)$
#01 (best)	−3.52	−3.12	1.96	−3.50	−3.13	1.20	−3.45	−3.16	1.44
#02	−3.12	−2.61	2.59	−3.02	−2.67	1.78	−2.82	−2.79	2.36
#03	−2.25	−1.91	3.46	−2.24	−1.91	3.30	−2.02	−2.04	3.49
#04	−2.57	−2.18	3.94	−2.48	−2.23	4.40	−2.39	−2.29	4.97
#05	−1.93	−1.66	5.53	−1.91	−1.67	4.90	−1.79	−1.74	5.20
#06–10	−1.70	−1.44	10.72	−1.64	−1.47	12.82	−1.57	−1.52	12.96
#11–20	−1.38	−1.17	14.37	−1.33	−1.20	13.52	−1.28	−1.23	14.15
#21–30	−1.10	−0.93	16.66	−1.05	−0.97	17.26	−1.04	−0.97	16.71
#31–40	−0.90	−0.75	22.55	−0.85	−0.78	24.76	−0.83	−0.79	22.18
#41–50	−0.78	−0.62	27.35	−0.71	−0.66	28.01	−0.68	−0.68	27.29
#51–60	−0.39	−0.29	30.49	−0.35	−0.32	29.92	−0.31	−0.31	30.25
#61–70	0.03	0.03	35.99	0.02	0.02	32.18	0.01	0.02	32.42
#71–80	0.14	0.15	38.48	0.13	0.14	35.50	0.12	0.12	35.95
#81–90	0.76	0.84	40.60	0.70	0.79	40.09	0.67	0.69	41.22
#91	1.21	1.64	43.26	0.88	1.20	43.17	0.58	1.05	44.56
#92	1.66	1.73	47.01	1.46	1.65	47.33	1.34	1.43	49.09
#93	2.03	2.14	50.50	1.80	2.03	52.77	1.66	1.75	54.94
#94	2.38	2.49	58.53	2.09	2.36	57.53	1.93	2.04	60.12

(Continues)

TABLE 5 (Continued)

Panel B: The *downgrade* portfolio									
	3F model			4F model			5F model		
BH rank	$\hat{t}_{\hat{\alpha}_{j,p}}$	$\hat{t}^b_{\hat{\alpha}_{j,p}}$	$\% \left(\hat{t}^b_{\hat{\alpha}_{j,p}} < \hat{t}_{\hat{\alpha}_{j,p}}\right)$	$\hat{t}_{\hat{\alpha}_{j,p}}$	$\hat{t}^b_{\hat{\alpha}_{j,p}}$	$\% \left(\hat{t}^b_{\hat{\alpha}_{j,p}} < \hat{t}_{\hat{\alpha}_{j,p}}\right)$	$\hat{t}_{\hat{\alpha}_{j,p}}$	$\hat{t}^b_{\hat{\alpha}_{j,p}}$	$\% \left(\hat{t}^b_{\hat{\alpha}_{j,p}} < \hat{t}_{\hat{\alpha}_{j,p}}\right)$
#95 (worst)	2.41	2.58	62.41	2.28	2.43	65.34	2.02	2.19	67.50

Note: This table presents the values of *t*-statistics at selected percentiles (%) of the distribution of *t*-statistics of the actual $\left(\hat{t}_{\hat{\alpha}_{j,p}}\right)$ and simulated $\left(\hat{t}^b_{\hat{\alpha}_{j,p}}\right)$ abnormal returns, as well as the percentage (%) of the 10,000 simulation runs that produce lower values of *t*-statistics at the selected percentiles than those actual abnormal returns (% Simulated $\hat{t}^b_{\hat{\alpha}_{j,p}}$ > Actual $\hat{t}_{\hat{\alpha}_{j,p}}$) for the *upgrade* portfolio and (% Simulated $\hat{t}^b_{\hat{\alpha}_{j,p}}$ < Actual $\hat{t}_{\hat{\alpha}_{j,p}}$) the *downgrade* portfolio. #01 (#95) represents the BH with the best (worst) recommendation performance among the total of 95 BHs involved in making up/downward revisions included in the *up/downgrade* portfolio. To save space, the average values of the actual and simulated *t*-statistics as well as the selected percentiles are reported for #06–10, #11–20, #21–30, #31–40, #41–50, #51–60, #61–70, #71–80, and #81–90.

the superior performance of the *upgrade* portfolio is due to BH skill, rather than due to BH luck. Moreover, for the relatively poor-performing BHs (e.g., #06–95 BHs), more than 10% of the simulated *t*-statistics are higher than the corresponding actual *t*-statistics. Similar evidence is found for the *downgrade* portfolio as shown in Panel B of Table 5.

Overall, our simulated results are in support of *Hypothesis* 2a that the reported performance persistence of BHs could be due to BH skill, rather than due to BH luck (i.e., random chance).

6 | CONCLUSIONS

Our study examines the impact of BH reputation on the performance of investment strategies following stock recommendation revisions in the UK. We develop two alternative proxies for BH reputation in the UK stock market either based on the past positions on the annual *II* All-Europe Research Team or based on the past recommendation performance in each calendar year. Using a unique dataset of 58,647 UK stock recommendation revisions over the period of January 1995 to June 2013, we find some interesting evidence that BH reputation proxied by the past positions on the annual *II* All-Europe Research Team has no impact on the recommendation performance, supporting Emery and Li (2009) that the *II* rankings are largely "popularity contests." However, BH reputation proxied by the past year recommendation performance has a significantly positive impact on the recommendation performance in the next year, implying that the recommendation performance of BHs is persistent. The reported persistence of the recommendation performance of BHs provides clear evidence of a violation of the semi-strong form of market efficiency, and from an investor's perspective, it is likely for investors to make profits by following stock recommendation revisions made by BHs in the UK market, even after controlling for transaction costs. Finally, our cross-sectional bootstrap simulations confirm that the observed performance persistence of BHs could be due to BH skill, rather than due to BH luck (i.e., random chance). That is, more prestigious BHs have sufficient skills in persistently generating superior recommendation performance, while less prestigious BHs lack such skills and persistently generate inferior recommendation performance.

ACKNOWLEDGEMENTS

The authors wish to thank for the helpful comments and suggestions from four anonymous reviewers, the Editor (Mark P. Taylor), Carol Alexander, and Brian M. Lucey. Su also thanks financial supports from the Faculty of Humanities and Social Sciences (HaSS), Newcastle University under the Faculty Research Fund – Autumn 2019 (OSR/0372/FR19/0001). All remaining errors and omissions are our own.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Chen Su https://orcid.org/0000-0003-4115-6207

ENDNOTES

[1] For example, the US regulatory oversight appears to be more fragmented, thereby giving rise to significant gaps in the monitoring and enforcement roles. That is, the US financial market is more exposed to regulatory risk that might have an influence on the recommendation performance. See more discussions on the differences of market structure in the US and UK markets in the speech of *"Comparing UK and US Macroprudential Systems: Lessons for China" given by* Donald Kohn at the Global Financial Forum, Tsinghua University, Beijing, on May 11, 2014 (available at: https://goo.gl/SHWkdX).

[2] The rationality behind the out-of-sample test is if BHs truly possess skill or information advantages over the market, they are likely to continue generating abnormal returns in the out-of-

sample period as well; otherwise, their performance is due to luck or random chance, it is likely to disappear in the out-of-sample period. For example, Fang and Yasuda (2014) use Regulation Fair Disclosure (Reg-FD) by the US Securities and Exchange Commission (SEC) in October 2000 as a natural experiment. They find that the superior recommendation performance of All-Star analysts is not significantly eroded after the adoption of Reg-FD, suggesting that the performance persistence of All-Star analysts is not entirely due to their luck, in line with Leone and Wu (2007).

[3] Altinkilic and Hansen (2009) also call into question the information role played by analysts in that their stock recommendation revisions often piggyback on public information (e.g., corporate events and news), thus providing investors with little incremental information (see, also, Altinkilic, Balashov, & Hansen, 2013).

[4] The total number of stock recommendation revisions included in the *upgrade* and *downgrade* portfolios is 46,931 (= 21,404 + 25,527), apparently less than the number of 58,647, as shown in Panel A of Table 1, which is not surprising, however. The upward revisions from Strong Sells to Sells, from Strong Sells to Holds, and from Sells to Holds are not included in the *upgrade* portfolio, as they can also be interpreted as negative recommendations. Similarly, the downward revisions from Strong Buys to Buys are not included in the *downgrade* portfolio, as they can also be interpreted as positive recommendations (see, also, Stickel, 1995).

[5] Specifically, in our sample, a very small proportion (4.585%) of upgrades and downgrades is made by more than one BH on a given date, for example, 3.401% (728 out of 21,404) of upgrades and 5.578% (1,424 out of 25,527) of downgrades.

[6] The value-weighted returns enable us to better capture the economic significance of our results, while the equal-weighted returns are, on average, biased upward due to the bid-ask bounce, that is, the returns of large size firms will be more heavily represented in the aggregate returns than those of small size firms (see, Barber et al., 2001).

[7] We explicitly exclude the return on the first trading day as many investors, particularly small investors, tend to react to information with a delay. Barber et al. (2001, p. 534) argue that "it is impractical for them to engage in the daily portfolio rebalancing that is needed to respond to the changes."

[8] The daily returns on size, value, and momentum in the UK stock market are collected from the Xfi Centre for Finance and Investment at University of Exeter, while we construct the profitability and investment factors in the UK stock market, strictly following Fama and French (2015).

[9] Like Li (2005), we also report the annualized abnormal return as $(1 + \text{daily abnormal net return to the } \textit{up/downgrade} \text{ portfolio})^{252}-1$.

[10] We also replicate our analysis by identifying more (less) prestigious BHs as the top 10 (non-top 10) on the *II* All-Europe Research Team, showing qualitatively similar results. That is, BH reputation proxied by the past *II* rankings has no impact on the recommendation performance, the results of which are not reported for the sake of brevity but available on request.

[11] Similar to Emery and Li (2009), we also calculate *information ratio* as the alternative recommendation performance, which is the *t*-statistic of the average daily abnormal return to the *up/downgrade* portfolio. We then divide all BHs into quintiles (or Best 5 and Non-Best 5 groups) by their past *information ratio* and replicate all analyses in Tables 2–4, obtaining consistent conclusions. These results are not reported to save space, but available on request.

[12] In contrast, Su et al. (2019) develop a rolling window-based time-series bootstrap simulation method, producing simulated results for the *up/downgrade* portfolio including all stocks recommended by Top 5 BHs in a total of 4,420 one-year rolling windows over the whole sample period January 1995 to June 2013 (see, also., Su & Zhang, 2020). The objective of the time-series bootstrap simulations is to test whether Top 5 BHs are able to generate superior recommendation performance in certain time periods.

[13] We conduct bootstrap simulations for the *upgrade* and *downgrade* portfolios separately. *N* represents a total of 95 BHs involved in making up/downward revisions included in the *up/downgrade* portfolio (see Table 1).

[14] For example, in 2001, the *II* sent ballots to more than 780 institutions and received the opinions of more than 3,200 money managers from about 400 institutions.

[15] On October 29, 1991, the "Heard on the Street" column of the *WSJ* reported that, at most BHs, the three most important factors determining analyst pay are an evaluation of the analyst by (a) the brokerage sales force, (b) the standing in the *II* poll, and (c) job offers from competitors.

[16] It is quite usual that analysts attempt to influence the *II* poll by visiting money managers about the time they vote. This practice is confirmed by a research director quoted in the *WSJ* as saying, "most of the guys know that they will be visiting for the *II* in the spring. I am a lonely guy in March and April shortly before the balloting" (Stickel, 1992).

[17] Like the *WSJ*, the Thomson Reuters StarMine Analyst Awards also recognize the world's top individual analysts and sell-side BHs based on the objective measurement of their estimate accuracy and recommendation performance (see details in Kucheev et al., 2017), which has been rigorously tested and proven in the marketplace since 1998. Given that our sample period starts before the year of 1998, we do not discuss the details of StarMine Analyst Awards in Appendix B.

[18] Barber et al. (2001) estimate the average round-trip transaction costs of 1.31% in the US. Despite the lack of readily available data regarding short selling costs in the UK, we assume a short selling cost of 1.50%, according to Su et al. (2019) and Su, Zhang, and Hudson (2020).

REFERENCES

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

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

Bagnoli, M., Watts, S. G., & Zhang, Y. (2008). Reg-FD and the competitiveness of all-star analysts. *Journal of Accounting and Public Policy*, *27*, 295–316.

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

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

Bradley, D., Gokkaya, S., & Liu, X. (2017). Before an analyst becomes an analyst: Does industry experience matter? *Journal of Finance*, *72*, 751–792.

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

Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics*, *27*, 285–303.

Cuthbertson, K., Nitzsche, D., & O'Sullivan, N. (2008). UKmutual 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.

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

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. H. (2005). Investment bank reputation and the price and quality of underwriting services. *Journal of Finance*, *60*, 2729–2761.

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

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

Forbes, P. W., Murphy, A., O'Keeffe, C., & Su, C. (2020). Are financial analysts eager postmen of bubble psychology? Evidence in the United Kingdom. *International Journal of Finance and Economics*, *25*, 120–137.

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.

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

Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance*, *55*, 265–295.

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 UKstock prices - 1935 to 1994. *Journal of Banking and Finance*, *20*, 1121–1132.

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

Jackson, A. R. (2005). Trade generation, reputation, and sell-side analysts. *Journal of Finance*, *60*, 673–717.

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.

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

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.

Kucheev, Y. O., Ruiz, F., & Sorensson, T. (2017). Do stars shine? Comparing the performance persistence of star sell-side analysts listed by institutional investor, the wall street journal, and StarMine. *Journal of Financial Services Research*, *52*, 277–305.

Leone, A. & Wu, J. (2007). What does it take to become a superstar? Evidence from institutional investor rankings of financial analysts. Working Paper, University of Rochester.

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

Loh, R. K., & Stulz, R. M. (2011). When are analyst recommendation changes influential? *Review of Financial Studies*, *24*, 593–627.

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

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.

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.

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

Stickel, S. E. (1992). Reputation and performance among security analysts. *Journal of Finance*, *47*, 1811–1836.

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

Su, C., & Zhang, H. (2020). A time-series bootstrapping simulation method to distinguish sell-side analysts' skill from luck. In C. F. Lee & J. C. Lee (Eds.), *Handbook of financial econometrics, mathematics, statistics, and machine learning* (Vol. 2, pp. 2011–2052). Singapore: World Scientific Publishing.

Su, C., Zhang, H., Bangassa, K., & Joseph, N. L. (2019). On the investment value of sell-side analyst recommendation revisions in the UK. *Review of Quantitative Finance and Accounting*, *53*, 257–293.

Su, C., Zhang, H., & Hudson, R. S. (2020). The time-varying performance of UKanalyst recommendation revisions: Do market conditions matter? *Financial Markets, Institutions and Instruments*, *29*, 65–89.

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

How to cite this article: Su C, Zhang H, Joseph NL. The performance of UK stock recommendation revisions: Does brokerage house reputation matter? *Int J Fin Econ*. 2020;1–20. https://doi.org/10.1002/ijfe.2312

APPENDIX

THE *INSTITUTIONAL INVESTOR* (*II*) AND *WALL STREET JOURNAL* (*WSJ*) RANKINGS

Each year, in March, April, or May, the *II* sends surveys to a variety of money managers in the US, European, and Asian investment funds to evaluate analysts on the basis of a series of criteria, such as accessibility/responsiveness, industry knowledge, special services, stock selection, earnings estimates, written reports, and so on (see, Bradley, Gokkaya, & Liu, 2017).[14] In its October issue each year, the *II* publishes its first, second, third, and runner-up teams determined by the weighted average of the returned scores. Typically, only one analyst per industry is listed in each of the first and second research teams, but multiple analysts from the same industry are common in the third and runner-up research teams.

Given that the *II* rankings are weighted by the size of respondent's institution, they are likely to favour analysts at large BHs. In fact, the *II* survey forms ask respondents to name and rank four best analysts in each industry without identifying any analysts, so respondents have to recall or look up analyst names to be able to vote. Compared with analysts at large BHs, analysts at small BHs are less likely to be known, so it is hard for them to achieve enough recognition to score well on the ballot. Also, there is a potential for conflicts of interest for large BHs, as they often have analysts in one division and money managers who are surveyed in another division and these managers may be biased in favour of analysts in their own BHs.

Although no claim is made that the *II* rankings are indicative of future recommendation performance, it is likely that many investors interpret them in this manner. The positions on the annual *II* All-America Research Team can be viewed as a proxy for analyst reputation and as one of the most important criteria for determining analyst pay at most BHs.[15,16] The directors of research at major BHs confirm that All-Star analysts are generally paid higher salaries. In particular, BHs tend to display their past recommendation performance or the number of All-Star analysts in advertisements designed to attract new clients.

In addition, the *II* rankings are widely criticized due to their biased election criteria, which add more weights on "accessibility/responsiveness" than on "earnings estimates" and "stock selection." For example, "earnings estimates" and "stock selection" are typically listed near the bottom of all election criteria, while "accessibility/responsiveness" is ranked highly. This suggests that money managers tend to value information that is passed along in private communications rather than analysts' research reports.

To rule out the concern that the *II* rankings are not mainly based on the past stock recommendation performance, the *WSJ* created its own ranking in 1993.[17] Specifically, the *WSJ* publishes a quarterly listing of the largest BHs, ranked by their recommendation performance during the past calendar year. Although the *WSJ* rankings are determined solely by the recommendation performance, it explicitly imposes eligibility requirements on analysts, which could bias the rankings (Emery & Li, 2009). For example, each year, the *WSJ* ranks the top five analysts in each specific industry, based on their recommendation performance. To be eligible for the *WSJ* rankings, an analyst must cover five or more qualified stocks in the industry and at least two of them must be among the 10 largest stocks. The *WSJ* requirements put analysts at small BHs at a disadvantage, as small BHs typically focus on the coverage of small size stocks, which are less likely to be included in the 10 largest stocks in each specific industry. For example, in 2001, only 1,370 of over 4,000 analysts were eligible for the *WSJ* rankings; the similar proportions are shown in other years.

TRANSACTION COSTS

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). According to Hudson, Dempsey, and Keasey (1996), the total round-trip transaction costs in the UK stock market for the most favoured 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%. Based on 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%,[18] we measure transaction costs multiplied by the corresponding average daily portfolio turnover.

Specifically, the daily turnover for the portfolio on the trading date t is defined as the percentage of stocks in the portfolio as of the close of trading on date $t-1$ that has changed by the close of trading on date t. That is, like Barber et al. (2001), we measure the daily turnover as the percentage of the portfolio that has been moved into some other set of stocks on date t. For each stock i in portfolio p as of the close of trading on date $t-1$, we calculate its fraction of the portfolio, $G_{i,t}$, at the end of trading on date t without accounting for portfolio rebalancing:

$$G_{i,t} = \omega_{i,t-1} \times (1+R_{i,t}) / \sum_{i=1}^{n_{p,t-1}} \omega_{i,t-1} \times (1+R_{i,t}). \tag{A1}$$

Then, $G_{i,t}$ is compared to the actual fraction $F_{i,t}$ that stock i makes up of portfolio p as of the close of trading on date t, after accounting for any portfolio rebalancing. Finally, the change in the percentage holding of each stock on date $t-1$ is summed, generating the portfolio turnover on date t:

$$TURNOVER_{p,t} = \sum_{i=1}^{n_{p,t}} |G_{i,t} - F_{i,t}|. \tag{A2}$$

We calculate the net abnormal return as the gross return less the estimated transaction costs multiplied by the corresponding daily portfolio turnover.