CFR (draft), XXXX, XX: 1–29

The Big League Effect

Geoffrey Peter Smith

WP Carey School of Business, Arizona State University; gps@asu.edu

ABSTRACT

This study is critical of implausible stock return effects. Among these are studies linking stock returns to extreme climate variables, outer space activity, sports results, politics, religious observances, weather conditions, and other peculiar phenomena. I argue that these effects are just the predetermined outcomes of endogenous treatment assignments, not true causal effects. To demonstrate, I "discover" a new implausible effect called the big league effect. Win-loss records of NY's two professional baseball teams predict excess returns on popular anomaly strategies. New critical values are calculated by Monte Carlo simulation where treatment assignment follows a Bernoulli distribution with endogenous success probability. These new critical values expose the big league effect as just an artifact of the sample selection mechanism.

Keywords: Stock Return Anomaly, Major League Baseball, Endogenous Treatment Assignment, Monte Carlo Simulation, Randomization

JEL Codes: G11, G12, G14, C10, C12, C18, B40

I thank the anonymous referee for helpful comments and suggestions.

ISSN 2164-5744 ; DOI 10.1561/104.XXXXXXXX

1 Motivation

Splashy stock return effects garner attention in academic journals and popular media outlets. Among these are studies linking stock returns to extreme climate variables (El Niño, geomagnetic storm activity, global warming), outer space activity (lunar phases, planetary conjunction, sunspots), sports results (basketball, cricket, football, ice hockey, rugby, soccer), politics (first 100-days of a new US President, party of the US President), religious observances (Ramadan), weather conditions (clouds, humidity, rain, snow, sunshine, temperature, wind), and other peculiar phenomena (air pollution, daylight savings time, Friday 13th, music sentiment, seasonal depression, summer vacation).[1]

I argue that these effects are nothing more than predetermined outcomes of endogenous treatment assignments (ETA) not true causal effects. To demonstrate, I start by "discovering" a new implausible stock return effect called the big league effect. The big league effect links future stock returns to the win-loss records of New York's two professional baseball teams. Win-loss records for the NY Yankees and NY Mets have significant power to predict excess returns on popular anomaly strategies. Stock traders exploiting the big league effect can earn significant excess returns by strategically holding positions following months in which the Yankees or Mets win (do not win) a majority of their baseball games.

I then expose the big league effect as just the predetermined outcome of an ETA and faulty inference. An assignment mechanism is endogenous when a researcher picks a treatment assignment that tilts the sample selection in favor of finding a significant result. For implausible stock return effects, this involves concocting an entertaining story as a means to se-

[1] See, for example, Kolb and Rodriguez (1987), Kamstra *et al.* (2000) and Kamstra *et al.* (2003), Santa-Clara and Valkanov (2003), Cao and Wei (2005), Dowling and Lucey (2005) and Dowling and Lucey (2008), Keef and Roush (2005) and Keef and Roush (2007), Chang *et al.* (2006), Goetzmann and Zhu (2005), Yuan *et al.* (2006), Edmans *et al.* (2007), Modis (2007), Yoon and Kang (2009), Kang *et al.* (2010), Kaplanski and Levy (2010), Keef and Khaled (2011), Lee and Wang (2011), Białkowski *et al.* (2012), Chang *et al.* (2012), Lu and Chou (2012), Novy-Marx (2014), Apergis *et al.* (2016), Kaustia and Rantapuska (2016), Lepori (2016), Mbanga *et al.* (2019), Chan *et al.* (2020), Fernandez-Perez *et al.* (2020), Dong and Tremblay (2021), Liu *et al.* (2021), and Edmans *et al.* (2022). See also Koppett (1987), Koppett (1988), and Koppett (1989) and Krueger and Kennedy (1990) for discussion on the eponymous Super Bowl indicator discovered in the sports pages of the *New York Times*.

lect extreme returns with greater probability into one sample group than another.

For example, I discover the big league effect by exploiting the preexisting differences in anomaly returns following months in which the Yankees or Mets win (do not win) a majority of their baseball games. Had this treatment assignment not yielded a significant difference in means, then I would have tried the win-loss records of rival teams from Boston and Philadelphia. Had those not worked, then I would have tried the number of rainbows, shark attacks, or UFO sightings. The point is that the stock returns are known before the testing begins. All a motivated researcher needs to do is find a way to tilt the sample selection in favor of yielding a significant difference in means.

The problem with this is that the standard T-test assumes that the treatment and control groups are selected by random sampling. With random sampling, the treatment assignment is made by chance. A monthly return is just as likely to be in the treatment group as it is to be in the control group. Nonrandom sampling inflates the usual T-statistic and renders the critical values from the usual T-distribution invalid. Comparing a T-statistic inflated by an ETA to the critical values for the usual T-distribution makes an implausible stock return effect seem much more improbable than it really is.

I account for this by calculating a new randomization distribution for the T-statistic by Monte Carlo simulation. Specifically, logistic regression of win-loss records on anomaly returns yields fitted values representing propensity scores. I use these propensity scores to populate $100,000$ treatment assignments and calculate difference in means tests. The upper 10%, 5%, 1%, and 0.1% values of the $100,000$ T-statistics calculated by randomization represent new critical values for the T-distribution. These new and larger critical values explain away the big league effect as just the predetermined outcome of an ETA and faulty inference. Future research along these lines may yield new insight on the true significance of other implausible stock return effects.

The remainder of the study proceeds as follows. I discuss the growing literature on statistical bias in financial research in Section 2. I "discover" the big league effect in Section 3. In Section 4, I describe the nature and meaning of an ETA and explain why it renders the usual T-distribution invalid. In Section 5, I calculate the randomization distribution and extract the new critical values. These new and larger critical values expose the big league effect as just a statistical artifact in Section 6. Section 7 summarizes and concludes.

2 Related Literature

A number of studies indicate growing interest in the problem of statistical bias in financial research. For example, Harvey (2017) explores the meaning and limitations of the *p*-value and offers the minimum Bayes factor as an alternative. Robins and Smith (2020) argue that studies on the constancy of stock return anomalies are subject to selection bias and suggest an unbiased alternative. Chen (2021) focuses on the surprisingly large amount of large T-statistics reported in the literature and on the limits to *p*-hacking. Hasler (2023) studies how the small decisions researchers make when they construct anomaly portfolios influences the significance of future portfolio returns.

In this study, I argue that implausible stock return effects are just statistical artifacts not true causal effects. This is important because there is little reason to believe that unusual phenomena have any real causal effect on stock returns even if the test results are statistically significant.

3 The Big League Effect

A study critical of implausible effects needs an implausible effect to criticize. In this section, I "discover" a new stock return effect. The big league effect links future stock returns to the win-loss records of NY's two professional baseball teams, the Yankees and Mets.

3.1 Summary

The game of baseball is such an inveterate part of American culture that it is known as "America's National Pastime." Nicknamed the "big league," Major League Baseball (MLB) stands as North America's premier professional baseball league and one of the most widely-followed sports leagues in the world. At the present time, there are 30 MLB teams throughout the US and Canada, organized into the American (AL) and National Leagues (NL).

Two of the most storied MLB teams represent New York City and the NY metropolitan area. Originally known as the Highlanders, the NY Yankees of the AL have played their home games at Yankee stadium in the Bronx, since 1923, while the NY Mets of the NL have played their home games at either the Polo Grounds in Upper Manhattan or at Citi Field (formerly Shea Stadium) in Queens, since 1962.[2]

New York City is also home to the NY Stock Exchange and many professional stock traders live in the NY metropolitan area. Many of these traders follow the Yankees or Mets whose on-field performance elicits a powerful emotional response in their fans. Groundbreaking studies by Saunders (1993) and Hirshleifer and Shumway (2003) have already established a link between stock traders' moods and stock returns, operating through stock traders' emotional response to the local weather conditions.[3] If the on-field performance of the Yankees and Mets also elicits a similar type of emotional response, then there is another plausible link between stock traders' moods and stock returns, operating through stock traders' emotional response to the win-loss records of the Yankees and Mets.

I exploit the link between stock traders' moods and stock returns to discover a new stock return effect called the big league effect. Drawing on largely untapped historical MLB game log data, I discover that the Yankees' and Mets' monthly win-loss records have significant power to predict excess returns on popular anomaly strategies. Stock traders exploiting the big league effect can earn significant excess returns by strategically holding positions following months in which the Yankees or Mets win (do not win) a majority of their baseball games.

Following months in which the Yankees win a majority of their games, I find that long positions in the momentum, residual variance, and variance anomalies earn between significant 1.47% and significant 1.81% more per month than in other months. Short positions in book-to-market, cash-flow price, earnings price, investment, long-term reversal, market beta, short-term reversal, and size earn between significant 0.59% and significant 1.83% more per month than in other months.

Following months in which the Mets win a majority of their games, I

---

[2]Total home game attendance in 2023 was 3.3 and 2.6 million for the Yankees and Mets, respectively. See https://www.baseball-reference.com/leagues/majors/2023-misc.shtml.

[3]See Lucey and Dowling (2005) for a survey of the literature on stock returns and investors' moods.

find that long positions in momentum, net share issues, operating profitability, residual variance, and variance earn between significant 0.94% and significant 3.10% more per month than in other months. Short positions in investment, long-term reversal, market beta, short-term reversal, and size earn between significant 0.85% and significant 2.28% more per month than in other months.

3.2 Data and Method

Data on the Yankees' and Mets' win-loss records comes from Retrosheet, Inc., a non-profit corporation founded in 1989 to computerize historical accounts of MLB games.[4] Retrosheet has game log data on MLB games played since 1871, including the game dates, team names, and final scores.

I use Retrosheet's game log data to calculate a team's monthly winning percentage as the total number of games won divided by the total number of games played.[5] I then set the dummy variable, *win*, equal to one in the months where a team has a monthly winning percentage greater than or equal to 0.500 and zero otherwise.

Data on anomaly returns comes from the internet data library of Ken French.[6] For the 600 months between January 1974 and December 2023, I calculate the monthly percent excess return on 15 popular stock return anomalies by simultaneously buying and selling the extreme equal-weight decile portfolios. These 15 anomalies are: accruals (Sloan, 1996), book-to-market (Fama and French, 1992), cashflow price (Lakonishok *et al.*, 1994), dividend yield (Naranjo *et al.*, 1998), earnings price (Basu, 1977), investment (Cooper *et al.*, 2008), long-term reversal (De Bondt and Thaler, 1985), market beta (Fama and MacBeth, 1973), momentum (Jegadeesh and Titman, 1993), net share issues (Pontiff and Woodgate, 2008), operating profitability (Novy-Marx, 2013), residual variance (Ang *et al.*, 2006), short-term reversal (Jegadeesh, 1990), size (Banz, 1981), and variance (Haugen and Heins, 1975). For book-to-market, cashflow price, dividend yield, earnings price, market beta, momentum, and operating profitability, the high-value portfolio is associated with higher mean return, while

[4] See https://www.retrosheet.org/.

[5] I use the ending dates for games that were started on one date, but were suspended and ended on a later date. For example, the NY Mets and Washington Nationals started a game on August 10, 2021, but that game was suspended due to rain. The game ended on August 11, 2021.

[6] See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

for accruals, investment, long-term reversal, net share issues, residual variance, short-term reversal, size, and variance, the high-value portfolio is associated with lower mean return. I form the portfolios so that the mean return is positive. I then merge the monthly anomaly returns with Retrosheet's game log data such that the anomaly return in month t is matched with Retrosheet's game log data on win-loss records in month $t-1$.

As the first step in the analysis, I depict 15 boxplots of monthly excess returns by win_{t-1}. These boxplots portray the medians, first and third quartiles, and outliers for each series of anomaly returns.

Figure 1 depicts the 15 boxplots of monthly excess returns for the Yankees.

[Figure 1 here]

Figure 1 exposes meaningful differences in the distributions of the returns by win_{t-1} for the Yankees. Extreme returns are generally selected with lesser (greater) frequency following months in which the Yankees win (do not win) at least half of their baseball games.

Figure 2 depicts the 15 boxplots of monthly excess returns for the Mets.

[Figure 2 here]

Figure 2 exposes meaningful differences in the distributions of the returns by win_{t-1} for the Mets. Extreme returns are also generally selected with lesser (greater) frequency following months in which the Mets win (do not win) at least half of their baseball games.

I then calculate the mean returns and test for significant differences in the mean returns by win_{t-1} by estimating

$$r_t = \mu_0 + \varepsilon_t \tag{1}$$

$$r_t = \mu_1 + \mu_2 win_{t-1} + \varepsilon_t, \tag{2}$$

where r_t is the excess return on an anomaly portfolio in month t and win_{t-1} is the win-loss dummy predictor variable. The ε_t is a mean zero error term.[7]

[7] I use R computer software by the R Core Team (2024) for all of the calculations. Major portions of the R code used for this study are given in the Appendix.

The least squares estimate for μ_0 is then the mean excess return over the whole 600-month sample period. The μ_1 is the mean excess return following months in which a team does not win at least half of its games or does not play any games at all. Primary interest lies in the estimate for μ_2, the big league effect. The estimate for μ_2 is the difference in the mean excess return following months in which a team wins at least half of its games and the mean excess return following months in which a team does not win at least half of its games or does not play any games at all. Evidence in favor of a significant big league effect is given by comparing the T-statistic for μ_2 in regression (2) to the 5% critical value for the standard T-distribution.

3.3 Results

Table 1 summarizes the results from regressions (1) and (2) for the Yankees.

[Table 1 here]

The Yankees win a majority of their games ($win_{t-1} = 1$) in 251 months of the 600-month sample period. Long positions in the momentum, residual variance, and variance anomalies earn between significant 1.47% and significant 1.81% more per month in these months than in other months. Short positions in book-to-market, cashflow price, earnings price, investment, long-term reversal, market beta, short-term reversal, and size earn between significant 0.59% and significant 1.83% more per month in these months than in other months.

Table 2 summarizes the results from regressions (1) and (2) for the Mets.

[Table 2 here]

The Mets win a majority of their games ($win_{t-1} = 1$) in 178 months of the 600-month sample period. Long positions in momentum, net share issues, operating profitability, residual variance, and variance earn between significant 0.94% and significant 3.10% more per month in these months than in other months. Short positions in investment, long-term reversal, market beta, short-term reversal, and size earn between significant 0.85% and significant 2.28% more per month in these months than in other months.

3.4 Conclusion

In this section, I "discover" a new implausible stock return effect called the big league effect. Win-loss records for the NY Yankees and NY Mets have significant power to predict excess returns on popular anomaly strategies.

In the following sections, I will expose the big league effect as just the predetermined outcome of an ETA not a true causal effect. Endogenous assignment inflates the usual T-statistic by tilting the treatment assignment in favor of finding a significant difference in means. Comparing a T-statistic inflated by ETA to the 5% critical value from the usual T-distribution makes the big league effect seem much more improbable than it really is.

4 Endogenous Treatment Assignment

The difference in means test in regression (2) requires a series of stock returns and a series of dummy variables. Let $R_i = (r_t, \ldots, r_T)$ represent the series of stock returns. Let $D_i = (d_t, \ldots, d_T)$ represent one *possible* series of dummy variables. For example, D_i could be an ordered series with a clear breakpoint like $D_i = (0, 0, 0, 1, 1, 1)$ or an unordered series like $D_i = (0, 0, 1, 1, 0, 1)$.

The series D_i is known as the treatment assignment. The probability of a particular treatment assignment is known as the assignment mechanism. With a sample size of 600 stock returns and two sample groups (treatment and control), there are 2^{600} possible treatment assignments.

The dummy variables in D_i are what determine how the stock returns in R_i are selected into treatment or control groups where

$$d_t = \begin{cases} 0 & \text{then } r_t \text{ is selected into the control group} \\ 1 & \text{then } r_t \text{ is selected into the treatment group.} \end{cases} \tag{3}$$

A motivated researcher selects among which of the 2^{600} possible treatment assignments to study.[8] For example, I select the treatment assignments $D_{yankees}$ and D_{mets} using the win-loss records of the Yankees and Mets. Had $D_{yankees}$ and D_{mets} not yielded a significant difference in means test, then I would have tried the treatment assignments $D_{red\ sox}$ and $D_{phillies}$

[8] See Hasler (2023) for evidence on how a motivated researcher's selection among which of the possible R_i to study affects statistical significance.

using the win-loss records of rival MLB teams from Boston and Philadelphia. Had these not yielded a significant difference in means, then I would have tried treatment assignments $D_{rainbows}$, D_{sharks}, or D_{ufos} by concocting a story about the causal effect of rainbows, shark attacks, or UFO sightings on stock returns.

The problem with this is that the test statistic for μ_2 in regression (2), $T(D_i, R_i)$, only follows the usual T-distribution if the conditional probability of sample group selection is random. This means that the conditional probability of treatment group selection must be $\Pr(d_t = 1 \mid r_t) = 0.50$ and thus the conditional probability of control group selection must also be $\Pr(d_t = 0 \mid r_t) = 0.50$. Standard methods assume that the d_t are exogenous in the sense that the anomaly returns were equally likely to have been in the treatment group as they were to have been in the control group before the testing begins.

Selection of the treatment assignment by endogenous methods violates this exogeneity requirement by tilting the probability of sample group selection in favor of finding a significant difference in means. This inflates the usual T-statistic, making an implausible stock return effect seem much more improbable than it really is.[9]

To account for this, I calculate a new randomization distribution for the standard difference in means test. Specifically, logistic regression of win-loss records on anomaly returns yields fitted values representing propensity scores. I use these propensity scores to populate 100,000 treatment assignments by Monte Carlo simulation. Each treatment assignment represents a series of random draws from independent Bernoulli trials where the propensity scores serve as probabilities of success. The upper 10%, 5%, 1%, and 0.1% values of the 100,000 T-statistics calculated by randomization represent new critical values for the T-distribution. Using these new and larger critical values to evaluate the big league effect exposes the big league effect as just the predetermined outcome of an ETA not a true causal effect.

[9]The difference in means test is essentially a structural break test when the treatment assignment is an ordered series with a clear breakpoint like $D_i = (0,0,0,1,1,1)$. Robins and Smith (2020) show how a researcher's endogenous selection of this breakpoint inflates the usual T-statistic and suggest an unbiased alternative. This study extends that analysis by showing how a researcher's endogenous selection of an unordered series like $D_i = (0,1,1,1,0,1)$ also inflates the usual T-statistic and renders the usual T-distribution invalid.

5 Randomization-based Inference

In this section, I describe how I calculate the randomization distribution by Monte Carlo simulation. This involves populating 100,000 treatment assignments, $D_i = (d_t, \ldots, d_T)$, where each d_t is the result of an independent Bernoulli trial with success probability of p_t.

The p_t is a propensity score calculated as a fitted value from the logistic regression

$$\ln\left(\frac{\Pr(win_{t-1} = 1 \mid r_t)}{1 - \Pr(win_{t-1} = 1 \mid r_t)}\right) = \beta_0 + \beta_1 r_t, \tag{4}$$

where $\Pr(win_{t-1} = 1 \mid r_t)$ is the conditional probability of treatment group selection.[10] The β_0 is a constant and the β_1 represents the change in the log odds of treatment group selection given a change in the monthly excess return r_t.

The fitted value p_t calculated as

$$p_t = \frac{1}{1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 r_t)}} \tag{5}$$

becomes a Bernoulli success probability used to populate 100,000 treatment assignments $D_i = (d_t, \ldots, d_T)$ where $d_t \sim Bin(1, p_t)$.

Table 3 reports the results of the logistic regressions for the Yankees. Also reported are the average marginal effects (AME) as $^{\Delta \Pr(win_{t-1}=1|r_t)}/_{\Delta r_t}$.

[Table 3 here]

Higher returns are selected into the treatment group with greater probability for the momentum, residual variance, and variance anomalies. The β_1 are significant 0.040 for momentum, significant 0.039 for residual variance, and significant 0.034 for variance. The AME for momentum is significant 0.010, meaning that a 1% increase in the monthly excess return on momentum increases the probability of treatment group selection by 1%. The AME for residual variance is significant 0.009 and the AME for variance is significant 0.008.

[10] The idea to correct for sample selection using the output from a binary response model originates in Heckman (1979). See Kennedy (1995) for discussion on the benefits of randomization tests in econometrics. See Branson and Bind (2018) for evidence on the validity of randomization-based inference for Bernoulli trial experiments in observational studies.

Lower returns are selected into the treatment group with greater probability for book-to-market, cashflow price, earnings price, investment, long-term reversal, market beta, short-term reversal, and size. The β_1 ranges from from significant −0.033 for market beta to significant −0.099 for investment. The AME ranges from significant −0.008 for market beta and short-term reversal to significant −0.024 for investment.

Table 4 reports the results of the logistic regressions for the Mets.

[Table 4 here]

Higher returns are selected into the treatment group with greater probability for momentum, net share issues, operating profitability, residual variance, and variance. The β_1 ranges from significant 0.048 for momentum to significant 0.083 for operating profitability. The AME ranges from significant 0.010 for momentum to significant 0.017 for operating profitability.

Lower returns are selected into the treatment group with greater probability for investment, long-term reversal, market beta, short-term reversal, and size. The β_1 ranges from from significant −0.038 for short-term reversal to significant −0.103 for long-term reversal. The AME ranges from significant −0.008 for short-term reversal to significant −0.021 for long-term reversal.

The fitted values p_t from equation (5) are used to populate $100,000$ random treatment assignments D_i. With each treatment assignment I then calculate $100,000$ difference in means tests by estimating

$$r_t = \mu_1 + \mu_2 d_t + \varepsilon_t, \tag{6}$$

where r_t is the excess return on an anomaly portfolio in month t and d_t is a dummy variable such that $d_t \sim Bin(1, p_t)$. I keep the absolute value of the $100,000$ T-statistics calculated for μ_2 as the exact randomization distribution for the difference in means test. These T-statistics are corrected for heteroskedasticity and autocorrelation using the methods of Newey and West (1987) and Andrews (1991).

The upper 10%, 5%, 1%, and 0.1% values of the $100,000$ T-statistics calculated by Monte Carlo simulation represent new critical values for the T-distribution.

Table 5 reports the new critical values for the Yankees.

[Table 5 here]

Table 6 reports the new critical values for the Mets.

[Table 6 here]

As expected, the new critical values are larger than the usual 10%, 5%, 1%, and 0.1% asymptotic critical values of 1.645, 1.960, 2.576, and 3.291 for the standard T-distribution. With 598 degrees of freedom, the 10%, 5%, 1%, and 0.1% critical values for the standard T-distribution are 1.647, 1.964, 2.584, and 3.307. In the following section, I use the new critical values from the randomization distribution to explain away the big league effect as just the predetermined outcome of an ETA and faulty inference.

6 Exposing the Big League Effect

For the Yankees, the maximum (absolute value) T-statistic for μ_2 in Table 1 is 4.175 for long-term reversal, which is highly significant when compared to 0.1% critical value of 3.307 for the usual T-distribution. When compared to the 5% critical value of 5.784 from the randomization distribution in Table 5, the big league effect for long-term reversal is no longer significant. In fact, comparing all of the T-statistics for μ_2 in Table 1 to the 5% critical values in Table 5 explains away all of the big league effects for the Yankees.

The same holds for the Mets. The maximum (absolute value) T-statistic for μ_2 in Table 2 is 6.007 for long-term reversal, which is highly significant when compared to 0.1% critical value for the usual T-distribution. This effect disappears though when the T-statistic is compared to the 5% critical value of 6.761 from the randomization distribution in Table 6. Comparing all of the T-statistics for μ_2 in Table 2 to the 5% critical values in Table 6 explains away all of the big league effects for the Mets.

7 Summary and Conclusion

How probable is it to find a significant relation between the win-loss records of the Yankees and Mets and future stock returns? What about extreme climate variables, outer space activity, sports results, politics, religious observances, and weather conditions? Answering this type of question involves knowledge of the underlying distribution of the stock returns and a complete understanding of the sample selection mechanism.

In this study, I argue that implausible stock return effects are nothing more than predetermined outcomes of endogenous treatment assignments, not true causal effects. I demonstrate this by "discovering" a new implausible effect called the big league effect. Win-loss records of NY's two professional baseball teams predict excess returns on popular anomaly strategies. Stock traders exploiting the big league effect can earn significant excess returns by strategically holding positions following months in which the Yankees or Mets win (do not win) a majority of their baseball games.

I then expose the big league effect as just a statistical artifact of ETA and faulty inference. Selecting returns into treatment and control groups using the win-loss records of the Yankees and Mets is just a means to tilt the treatment assignment in favor of finding a significant result. Had the win-loss records of the Yankees and Mets not yielded a significant difference in means, then I would have tried the win-loss records of rival MLB teams from Boston and Philadelphia. Had those not worked, then I would have tried the number of rainbows, shark attacks, or UFO sightings.

The problem with this is that nonrandom sampling inflates the usual T-statistic and renders the critical values for the standard T-distribution invalid. Comparing a T-statistic inflated by ETA to the usual critical values from the standard T-distribution makes an implausible stock return effect seem much more improbable than it really is.

To account for this, I calculate a new randomization distribution for the T-statistic by Monte Carlo simulation. This involves populating $100,000$ treatment assignments where extreme returns are selected with greater probability for one sample group than another. The upper 10%, 5%, 1%, and 0.1% values of the $100,000$ T-statistics calculated by randomization represent new critical values for the T-distribution. These new critical values are larger that the critical values for the standard T-distribution because these new values account for inflation by ETA.

These new and larger critical values explain away the big league effect as nothing more than a statistical artifact. Future research along these lines may yield new insight into whether other implausible stock return effects succumb to the same critique.

References

Andrews, D. W. K. 1991. "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". *Econometrica*. 59(3): 817–858.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. "The cross-section of volatility and expected returns". *Journal of Finance*. 61(1): 259–299.

Apergis, N., A. Gabrielsen, and L. A. Smales. 2016. "(Unusual) weather and stock returns—I am not in the mood for mood: further evidence from international markets". *Financial Markets and Portfolio Management*. 30: 63–94.

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

Basu, S. 1977. "Investment performance of common stocks in relation to their price-earnings ratios: a test of the efficient market hypothesis". *Journal of Finance*. 32(3): 663–682.

Białkowski, J., A. Etebari, and T. P. Wisniewski. 2012. "Fast profits: investor sentiment and stock returns during Ramadan". *Journal of Banking & Finance*. 36: 835–845.

Branson, Z. and M.-A. Bind. 2018. "Randomization-based inference for Bernoulli trial experiments and implications for observational studies". *Statistical Methods in Medical Research*. 28(5): 1378–1398.

Cao, M. and J. Wei. 2005. "Stock market returns: a note on temperature anomaly". *Journal of Banking & Finance*. 29: 1559–1573.

Chan, K. F., P. Gray, S. Gray, and A. Zhong. 2020. "Political uncertainty, market anomalies and presidential honeymoons". *Journal of Banking and Finance*. 113: 1–11.

Chang, S.-C., S.-S. Chen, R. K. Chou, and Y.-H. Lin. 2012. "Local sports sentiment and returns of locally headquartered stocks: a firm-level analysis". *Journal of Empirical Finance*. 19: 309–318.

Chang, T., C.-C. Nieh, M. J. Yang, and T.-Y. Yang. 2006. "Are stock market returns related to the weather effects? Empirical evidence from Taiwan". *Physica A*. 364: 343–354.

Chen, A. Y. 2021. "The limits of *p*-hacking: some thought experiments". *Journal of Finance*. 76(5): 2447–2480.

Cooper, M. J., H. Gulen, and M. J. Schill. 2008. "Asset growth and the cross-section of stock returns". *Journal of Finance*. 63(4): 1609–1651.

De Bondt, W. F. M. and R. Thaler. 1985. "Does the stock market overreact?" *Journal of Finance*. 40(3): 793–805.

Dong, M. and A. Tremblay. 2021. “Does the weather influence global stock returns?” *Critical Finance Review*. 10(2): 207–249.

Dowling, M. and B. M. Lucey. 2005. “Weather, biorhythms, beliefs and stock returns—some preliminary Irish evidence”. *International Review of Financial Analysis*. 14: 337–355.

Dowling, M. and B. M. Lucey. 2008. “Robust global mood influences in equity pricing”. *Journal of Multinational Financial Management*. 18: 145–164.

Edmans, A., A. Fernandez-Perez, A. Garel, and I. Indriawan. 2022. “Music sentiment and stock returns around the world”. *Journal of Financial Economics*. 145: 234–254.

Edmans, A., D. García, and Ø. Norli. 2007. “Sports sentiment and stock returns”. *Journal of Finance*. 62(4): 1967–1998.

Fama, E. F. and K. R. French. 1992. “The cross-section of expected stock returns”. *Journal of Finance*. 47(2): 427–465.

Fama, E. F. and J. D. MacBeth. 1973. “Risk, return, and equilibrium: empirical tests”. *Journal of Political Economy*. 81(3): 607–636.

Fernandez-Perez, A., A. Garel, and I. Indriawan. 2020. “Music sentiment and stock returns”. *Economics Letters*. 192: 1–6.

Goetzmann, W. N. and N. Zhu. 2005. “Rain or shine: where is the weather effect?” *European Financial Management*. 11(5): 559–578.

Harvey, C. R. 2017. “Presidential address: the scientific outlook in financial economics”. *Journal of Finance*. 72(4): 1399–1440.

Hasler, M. 2023. “Looking under the hood of data-mining”. *Unpublished Working Paper*.

Haugen, R. A. and A. J. Heins. 1975. “Risk and the rate of return on financial assets: some old wine in new bottles”. *Journal of Financial and Quantitative Analysis*. 10(5): 775–784.

Heckman, J. J. 1979. “Sample selection bias as a specification error”. *Econometrica*. 47(1): 153–161.

Hirshleifer, D. and T. Shumway. 2003. “Good day sunshine: stock returns and the weather”. *Journal of Finance*. 58(3): 1009–1032.

Jegadeesh, N. 1990. “Evidence of predictable behavior of security returns”. *Journal of Finance*. 45(3): 881–898.

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

Kamstra, M. J., L. A. Kramer, and M. D. Levi. 2000. “Losing sleep at the market: the daylight saving anomaly”. *American Economic Review*. 90(4): 1005–1011.

Kamstra, M. J., L. A. Kramer, and M. D. Levi. 2003. “Winter blues: a SAD stock market cycle”. *American Economic Review*. 93(1): 324–343.

Kang, S. H., Z. Jiang, Y. Lee, and S.-M. Yoon. 2010. “Weather effects on the returns and volatility of the Shanghai stock market”. *Physica A*. 389: 91–99.

Kaplanski, G. and H. Levy. 2010. “Exploitable predictable irrationality: the FIFA World Cup effect on the U.S. stock market”. *Journal of Financial and Quantitative Analysis*. 45(2): 535–553.

Kaustia, M. and E. Rantapuska. 2016. “Does mood affect trading behavior?” *Journal of Financial Markets*. 29: 1–26.

Keef, S. P. and M. S. Khaled. 2011. “Are investors moonstruck? Further international evidence on lunar phases and stock returns”. *Journal of Empirical Finance*. 18: 56–63.

Keef, S. P. and M. L. Roush. 2005. “Influence of weather on New Zealand financial securities”. *Accounting and Finance*. 45: 415–437.

Keef, S. P. and M. L. Roush. 2007. “Daily weather effects on the returns of Australian stock indices”. *Applied Financial Economics*. 17: 173–184.

Kennedy, P. E. 1995. “Randomization tests in econometrics”. *Journal of Business & Economic Statistics*. 13(1): 85–94.

Kolb, R. W. and R. J. Rodriguez. 1987. “Friday the thirteenth: 'Part VII'—a note”. *Journal of Finance*. 42(5): 1385–1387.

Koppett, L. 1987. “Quarterbacking the market”. *New York Times*. Nov.: A31.

Koppett, L. 1988. “A free insider tip”. *New York Times*. Jan.: 31.

Koppett, L. 1989. “The Super Bowl’s perfect 22”. *New York Times*. Jan.: 23.

Krueger, T. M. and W. F. Kennedy. 1990. “An examination of the Super Bowl stock market predictor”. *Journal of Finance*. 45(2): 691–697.

Lakonishok, J., A. Shleifer, and R. W. Vishny. 1994. “Contrarian investment, extrapolation, and risk”. *Journal of Finance*. 49(5): 1541–1578.

Lee, Y.-M. and K.-M. Wang. 2011. “The effectiveness of the sunshine effect in Taiwan’s stock market before and after the 1997 financial crisis”. *Economic Modelling*. 28: 710–727.

Lepori, G. M. 2016. “Air pollution and stock returns: evidence from a natural experiment”. *Journal of Empirical Finance*. 35: 25–42.

Liu, H., W. Zhang, X. Zhang, and J. Liu. 2021. “Temperature and trading behaviours”. *International Review of Financial Analysis*. 78: 1–19.

Lu, J. and R. K. Chou. 2012. "Does the weather have impacts on returns and trading activities in order-driven stock markets? Evidence from China". *Journal of Empirical Finance*. 19: 79–93.

Lucey, B. M. and M. Dowling. 2005. "The role of feelings in investor decision-making". *Journal of Economic Surveys*. 19(2): 211–237.

Mbanga, C., J. S. Jones, and S. A. Hoelscher. 2019. "The impact of politics and economic policy uncertainty on anomaly-based investment strategies". *Managerial Finance*. 45(5): 654–670.

Modis, T. 2007. "Sunspots, GDP and the stock market". *Technological Forecasting & Social Change*. 74: 1508–1514.

Naranjo, A., M. Nimalendran, and M. Ryngaert. 1998. "Stock returns, dividend yields, and taxes". *Journal of Finance*. 53(6): 2029–2057.

Newey, W. K. and K. D. West. 1987. "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix". *Econometrica*. 55(3): 703–708.

Novy-Marx, R. 2013. "The other side of value: the gross profitability premium". *Journal of Financial Economics*. 108: 1–28.

Novy-Marx, R. 2014. "Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars". *Journal of Financial Economics*. 112: 137–146.

Pontiff, J. and A. Woodgate. 2008. "Share issuance and cross-sectional returns". *Journal of Finance*. 63(2): 921–945.

R Core Team. 2024. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Robins, R. P. and G. P. Smith. 2020. "Selection bias and pseudo discoveries on the constancy of stock return anomalies". *Review of Quantitative Finance and Accounting*. 55: 1407–1426.

Santa-Clara, P. and R. Valkanov. 2003. "The presidential puzzle: political cycles and the stock market". *Journal of Finance*. 58(5): 1841–1872.

Saunders Edward M., J. 1993. "Stock prices and Wall Street weather". *American Economic Review*. 83(5): 1337–1345.

Sloan, R. G. 1996. "Do stock prices fully reflect information in accruals and cash flows about future earnings?" *Accounting Review*. 71(3): 289–315.

Yoon, S.-M. and S. H. Kang. 2009. "Weather effects on returns: evidence from the Korean stock market". *Physica A*. 388: 682–690.

Yuan, K., L. Zheng, and Q. Zhu. 2006. "Are investors moonstruck? Lunar phases and stock returns". *Journal of Empirical Finance*. 13: 1–23.

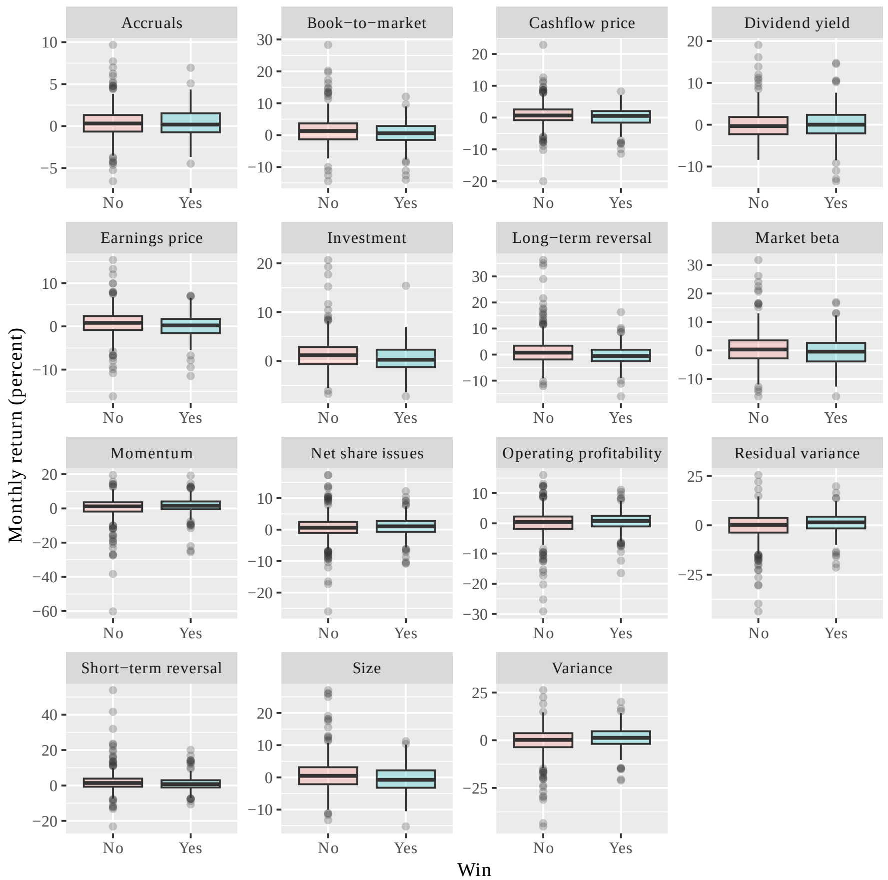

Figure 1: Boxplots of monthly excess returns by sample group for the NY Yankees

Description: Depicted are boxplots for the monthly excess returns on 15 popular anomaly portfolios. Win is "Yes" when the NY Yankees win a majority of their baseball games in the prior month and "No" otherwise. The sample period represents the 600 months from January 1974 to December 2023.

Interpretation: Boxplots depict the median, first and third quartiles, and outliers. These are nonparametric visualizations that make it easy see the preexisting differences in the distributions of excess returns between the treatment (Yes) and the control (No) groups.

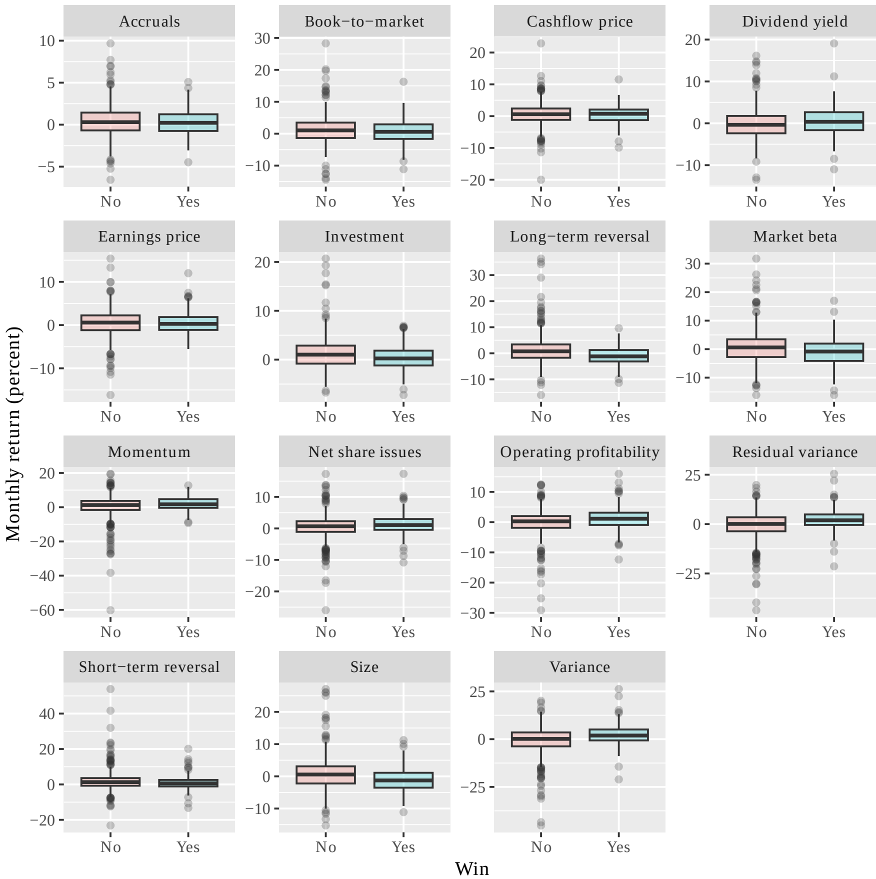

Figure 2: Boxplots of monthly excess returns by sample group for the NY Mets

Description: Depicted are boxplots for the monthly excess returns on 15 popular anomaly portfolios. Win is “Yes” when the NY Mets win a majority of their baseball games in the prior month and “No” otherwise. The sample period represents the 600 months from January 1974 to December 2023.

Interpretation: Boxplots depict the median, first and third quartiles, and outliers. These are nonparametric visualizations that make it easy see the preexisting differences in the distributions of excess returns between the treatment (Yes) and the control (No) groups.

Table 1: Excess returns and NY Yankees' win-loss record

Anomaly	μ_0	μ_1	μ_2	$Adj.\,R^2$
Panel A: Strategies performing significantly better				
Momentum	0.759**	0.145	1.468**	0.011
	[2.816]	[0.425]	[3.238]	
Residual variance	0.133	−0.624	1.808***	0.015
	[0.386]	[−1.378]	[3.397]	
Variance	0.124	−0.623	1.786**	0.012
	[0.346]	[−1.304]	[3.162]	
Panel B: Strategies performing significantly worse				
Book-to-market	1.082***	1.499***	−0.997**	0.011
	[5.207]	[5.474]	[−2.763]	
Cashflow price	0.655***	0.937***	−0.674*	0.008
	[4.427]	[4.911]	[−2.428]	
Earnings price	0.571***	0.818***	−0.590*	0.007
	[4.114]	[4.558]	[−2.235]	
Investment	0.984***	1.362***	−0.904**	0.018
	[6.233]	[6.530]	[−3.121]	
Long-term reversal	0.680**	1.445***	−1.829***	0.027
	[2.642]	[4.080]	[−4.175]	
Market beta	0.220	0.652*	−1.032*	0.006
	[0.848]	[2.002]	[−2.272]	
Short-term reversal	1.676***	2.088***	−0.984*	0.006
	[7.404]	[6.404]	[−2.552]	
Size	0.341	0.981**	−1.530***	0.021
	[1.404]	[2.982]	[−3.735]	

Description: Reported are the results from regressions

$$r_t = \mu_0 + \varepsilon_t \quad (1)$$

$$r_t = \mu_1 + \mu_2 win_{t-1} + \varepsilon_t, \quad (2)$$

where the dependent variable, r_t, is the percent excess return on an anomaly portfolio in month t and the independent variable, win_{t-1}, is a dummy variable set to one in months where the NY Yankees win at least half of their baseball games and zero otherwise. Reported in brackets below each estimate are T-statistics, corrected for heteroskedasticity and autocorrelation using the methods of Newey and West (1987) and Andrews (1991). ***, **, and * represent statistical significance at the 0.001, 0.01, and 0.05 levels, respectively. The sample period represents the 600 months from January 1974 to December 2023.

Interpretation: The NY Yankees' win-loss record has significant power to predict excess returns on popular anomaly strategies assuming random sample selection.

Table 2: Excess returns and NY Mets' win-loss record

Anomaly	μ_0	μ_1	μ_2	$Adj.\ R^2$
Panel A: Strategies performing significantly better				
Momentum	0.759**	0.272	1.642***	0.012
	[2.816]	[0.797]	[3.851]	
Net share issues	0.768***	0.489*	0.941**	0.010
	[4.066]	[2.351]	[2.859]	
Operating profitability	0.204	−0.213	1.405***	0.020
	[0.996]	[−0.919]	[4.000]	
Residual variance	0.133	−0.769*	3.040***	0.038
	[0.386]	[−2.034]	[5.588]	
Variance	0.124	−0.795*	3.098***	0.035
	[0.346]	[−1.979]	[5.477]	
Panel B: Strategies performing significantly worse				
Investment	0.984***	1.236***	−0.850***	0.013
	[6.233]	[6.543]	[−3.322]	
Long-term reversal	0.680**	1.356***	−2.281***	0.036
	[2.642]	[4.433]	[−6.007]	
Market beta	0.220	0.808**	−1.979***	0.024
	[0.848]	[2.814]	[−4.483]	
Short-term reversal	1.676***	1.976***	−1.013*	0.005
	[7.404]	[7.106]	[−2.365]	
Size	0.341	0.968***	−2.115***	0.035
	[1.404]	[3.446]	[−5.590]	

Description: Reported are the results from regressions

$$r_t = \mu_0 + \varepsilon_t \quad (1)$$
$$r_t = \mu_1 + \mu_2 win_{t-1} + \varepsilon_t, \quad (2)$$

where the dependent variable, r_t, is the percent excess return on an anomaly portfolio in month t and the independent variable, win_{t-1}, is a dummy variable set to one in months where the NY Mets win at least half of their baseball games and zero otherwise. Reported in brackets below each estimate are T-statistics, corrected for heteroskedasticity and autocorrelation using the methods of Newey and West (1987) and Andrews (1991). ***, **, and * represent statistical significance at the 0.001, 0.01, and 0.05 levels, respectively. The sample period represents the 600 months from January 1974 to December 2023.

Interpretation: The NY Mets' win-loss record has significant power to predict excess returns on popular anomaly strategies assuming random sample selection.

Table 3: Logistic regression results for the NY Yankees

Anomaly	β_0	β_1	**AME**
Panel A: Strategies performing significantly better			
Momentum	-0.367^{***}	0.040^{**}	0.010^{**}
	(0.085)	(0.015)	(0.004)
Residual variance	-0.342^{***}	0.039^{**}	0.009^{**}
	(0.084)	(0.013)	(0.003)
Variance	-0.340^{***}	0.034^{**}	0.008^{**}
	(0.084)	(0.012)	(0.003)
Panel B: Strategies performing significantly worse			
Book-to-market	-0.278^{**}	-0.053^{**}	-0.013^{**}
	(0.085)	(0.020)	(0.005)
Cashflow price	-0.295^{***}	-0.057^{*}	-0.014^{*}
	(0.084)	(0.025)	(0.006)
Earnings price	-0.299^{***}	-0.059^{*}	-0.014^{*}
	(0.084)	(0.027)	(0.006)
Investment	-0.242^{**}	-0.099^{***}	-0.024^{***}
	(0.087)	(0.029)	(0.007)
Long-term reversal	-0.297^{***}	-0.074^{***}	-0.017^{***}
	(0.084)	(0.018)	(0.004)
Market beta	-0.326^{***}	-0.033^{*}	-0.008^{*}
	(0.083)	(0.015)	(0.004)
Short-term reversal	-0.275^{**}	-0.035^{*}	-0.008^{*}
	(0.086)	(0.017)	(0.004)
Size	-0.318^{***}	-0.065^{***}	-0.016^{***}
	(0.084)	(0.018)	(0.004)

Description: Reported are the results from logistic regressions

$$\ln\left(\frac{\Pr(win_{t-1} = 1 \mid r_t)}{1 - \Pr(win_{t-1} = 1 \mid r_t)}\right) = \beta_0 + \beta_1 r_t, \tag{4}$$

where the dependent variable is the log odds ratio for the probability of treatment group selection and the dependent variable is the percent excess return on an anomaly portfolio in month t. The average marginal effect (AME) represents the change in the probability of treatment group selection given a change in the monthly excess return, ${}^{\Delta \Pr(win_{t-1}=1|r_t)}/_{\Delta r_t}$. Standard errors are reported in parentheses below each estimate. ***, **, and * represent statistical significance at the 0.001, 0.01, and 0.05 levels, respectively. The sample period represents the 600 months from January 1974 to December 2023.

Interpretation: There is a significant relation between the sample selection probability and the monthly excess return. Fitted values are propensity scores.

Table 4: Logistic regression results for the NY Mets

Anomaly	β_0	β_1	**AME**
Panel A: Strategies performing significantly better			
Momentum	−0.920***	0.048**	0.010**
	(0.094)	(0.017)	(0.003)
Net share issues	−0.924***	0.063**	0.013**
	(0.094)	(0.024)	(0.005)
Operating profitability	−0.907***	0.083***	0.017***
	(0.092)	(0.024)	(0.005)
Residual variance	−0.927***	0.075***	0.015***
	(0.094)	(0.016)	(0.003)
Variance	−0.921***	0.067***	0.014***
	(0.094)	(0.015)	(0.003)
Panel B: Strategies performing significantly worse			
Investment	−0.788***	−0.095**	−0.020**
	(0.092)	(0.032)	(0.006)
Long-term reversal	−0.852***	−0.103***	−0.021***
	(0.092)	(0.022)	(0.004)
Market beta	−0.878***	−0.068***	−0.014***
	(0.091)	(0.018)	(0.004)
Short-term reversal	−0.809***	−0.038*	−0.008*
	(0.092)	(0.019)	(0.004)
Size	−0.878***	−0.098***	−0.020***
	(0.092)	(0.021)	(0.004)

Description: Reported are the results from logistic regressions

$$\ln\left(\frac{\Pr(win_{t-1}=1 \mid r_t)}{1-\Pr(win_{t-1}=1 \mid r_t)}\right)=\beta_0+\beta_1 r_t, \qquad (4)$$

where the dependent variable is the log odds ratio for the probability of treatment group selection and the dependent variable is the percent excess return on an anomaly portfolio in month t. The average marginal effect (AME) represents the change in the probability of treatment group selection given a change in the monthly excess return, $\Delta\Pr(win_{t-1}=1|r_t)/\Delta r_t$. Standard errors are reported in parentheses below each estimate. ***, **, and * represent statistical significance at the 0.001, 0.01, and 0.05 levels, respectively. The sample period represents the 600 months from January 1974 to December 2023.

Interpretation: There is a significant relation between the sample selection probability and the monthly excess return. Fitted values are propensity scores.

Table 5: Critical values from the randomization distribution for the NY Yankees

Anomaly	**10%**	**5%**	**1%**	**0.1%**
Panel A: Strategies performing significantly better				
Momentum	3.929	4.241	4.823	5.476
Residual variance	4.372	4.703	5.351	6.082
Variance	4.165	4.506	5.158	5.888
Panel B: Strategies performing significantly worse				
Book-to-market	3.975	4.328	4.974	5.732
Cashflow price	3.513	3.842	4.484	5.231
Earnings price	3.454	3.799	4.456	5.144
Investment	4.703	5.040	5.663	6.385
Long-term reversal	5.461	5.784	6.379	7.083
Market beta	3.507	3.865	4.536	5.288
Short-term reversal	3.475	3.817	4.449	5.141
Size	4.936	5.275	5.912	6.694

Description: Reported are the critical values for the T-distribution calculated by means of randomization and Monte Carlo simulation. The Monte Carlo simulation calculates 100,000 treatment assignments $D_i = (d_t, \ldots, d_T)$, where each d_t follows a Bernoulli distribution with endogenous success probability such that $d_t \sim Bin(1, p_t)$. The p_t is a propensity score calculated as a fitted value from logistic regression (4). The randomization distribution is calculated by estimating the regression

$$r_t = \mu_1 + \mu_2 d_t + \varepsilon_t \qquad (6)$$

100,000 times where the dependent variable, r_t, is the percent excess return on an anomaly portfolio in month t. These critical values represent the upper 90%, 95%, 99%, and 99.9% of the absolute value of 100,000 T-statistics calculated for μ_2. These T-statistics are corrected for heteroskedasticity and autocorrelation using the methods of Newey and West (1987) and Andrews (1991). The sample period represents the 600 months from January 1974 to December 2023.

Interpretation: An endogenous treatment assignment renders the usual T-distribution invalid. With 598 degrees of freedom, the 10%, 5%, 1%, and 0.1% critical values for the standard T-distribution of 1.647, 1.964, 2.584, and 3.307 are too small.

Table 6: Critical values from the randomization distribution for the NY Mets

Anomaly	**10%**	**5%**	**1%**	**0.1%**
Panel A: Strategies performing significantly better				
Momentum	4.267	4.592	5.184	5.839
Net share issues	3.798	4.119	4.749	5.461
Operating profitability	4.780	5.103	5.711	6.451
Residual variance	6.227	6.557	7.184	7.899
Variance	6.051	6.388	7.010	7.774
Panel B: Strategies performing significantly worse				
Investment	4.440	4.793	5.473	6.190
Long-term reversal	6.431	6.761	7.387	8.110
Market beta	5.326	5.682	6.349	7.096
Short-term reversal	3.542	3.897	4.554	5.272
Size	6.161	6.500	7.147	7.921

Description: Reported are the critical values for the T-distribution calculated by means of randomization and Monte Carlo simulation. The Monte Carlo simulation calculates 100,000 treatment assignments $D_i = (d_t, \ldots, d_T)$, where each d_t follows a Bernoulli distribution with endogenous success probability such that $d_t \sim Bin(1, p_t)$. The p_t is a propensity score calculated as a fitted value from logistic regression (4). The randomization distribution is calculated by estimating the regression

$$r_t = \mu_1 + \mu_2 d_t + \varepsilon_t \qquad (6)$$

100,000 times where the dependent variable, r_t, is the percent excess return on an anomaly portfolio in month t. These critical values represent the upper 90%, 95%, 99%, and 99.9% of the absolute value of 100,000 T-statistics calculated for μ_2. These T-statistics are corrected for heteroskedasticity and autocorrelation using the methods of Newey and West (1987) and Andrews (1991). The sample period represents the 600 months from January 1974 to December 2023.

Interpretation: An endogenous treatment assignment renders the usual T-distribution invalid. With 598 degrees of freedom, the 10%, 5%, 1%, and 0.1% critical values for the standard T-distribution of 1.647, 1.964, 2.584, and 3.307 are too small.

Appendix

R Computer Code

Below are major segments of the R computer code used in the study. The R code uses functions from the packages: `data.table`, `lmtest`, `margins`, `sandwich`, and `zoo`.

```
#Load anomaly returns and calculate excess returns

anomalys <- list(
 `Accruals` = fread(
  cmd = "unzip -p Portfolios_Formed_on_AC_CSV.zip",
  skip = 751, nrows = 729,
  select = list(character = 1, numeric = c(7, 16))
 ),
 ...
)

lapply(anomalys, setnames, c("month", "lo", "hi"))

anomalys <- rbindlist(anomalys, idcol = "anomaly")

anomalys[TRUE, ret := lo - hi]

anomalys[anomaly %chin% negative_returns, ret := hi - lo]

#Load Retrosheet game log data

gamelogs <- lapply(seasons, fread, header = FALSE,
 select = list(character = c(1, 4, 7, 14),
 integer = c(10, 11)), col.names = c("date", "awayteam",
 "hometeam", "completed", "awayscore", "homescore"))

gamelogs <- rbindlist(gamelogs)

gamelogs[completed != "", date := substr(completed, 1, 8)]

#Select data on Yankees (NYA) and Mets (NYN)

gamelogs <- list(
 NYA = gamelogs[hometeam == "NYA" | awayteam == "NYA"],
 NYN = gamelogs[hometeam == "NYN" | awayteam == "NYN"])
```

```
gamelogs <- rbindlist(gamelogs, idcol = "team")

#Calculate total wins and total games by team and month

gamelogs[TRUE, winteam := fifelse(homescore > awayscore,
 hometeam, awayteam)]

gamelogs[TRUE, wingame := fifelse(team == winteam, 1, 0)]

gamelogs[TRUE, playgame := 1]

gamelogs <- gamelogs[TRUE, .(nwins = sum(wingame),
 ngames = sum(playgame)), by = .(team, month)]

#Calculate lagged dummy variable and merge with anomalys

gamelogs[TRUE, win := fifelse(nwins / ngames >= 0.50,
 "Yes", "No", "No")]

setorder(gamelogs, team, month)

gamelogs[TRUE, win := shift(win, type = "lag"), by = team]

anomalys <- gamelogs[anomalys, on = "month",
 allow.cartesian = TRUE]

anomalys <- split(anomalys, by = c("team", "anomaly"))

#Estimate logistic regressions

logistic_regression <- lapply(anomalys, function(x)
 glm(win ~ ret, family = binomial(link = "logit"),
 data = x))

#Calculate propensity scores and merge with anomalys

binps <- lapply(logistic_regression, fitted.values)

anomalys <- mapply(cbind, anomalys, pr = binps,
 SIMPLIFY = FALSE)
```

```
#Estimate regressions

regression_1 <- lapply(anomalys, function(x)
 lm(formula = ret ~ 1, data = x))

lapply(regression_1, coeftest, vcov = kernHAC)

regression_2 <- lapply(anomalys, function(x)
 lm(formula = ret ~ win, data = x))

lapply(regression_2, coeftest, vcov = kernHAC)

#Monte Carlo simulations

simulation_function <- function(anomaly) {
 model <- lm(ret ~ rbinom(600, 1, pr), data = anomaly)
 tstat <- coeftest(model, vcov = kernHAC)[2, "t value"]
 abs(tstat)
}

simulation <- lapply(anomalys, function(anomaly)
 replicate(100000, simulation_function(anomaly)))

lapply(simulation, quantile,
 probs = c(0.90, 0.95, 0.99, 0.999))
```