Journal of Economic Behavior and Organization 202 (2022) 703–713

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Female jockeys - what are the odds?

Vanessa Cashmore*, Neil Coster, David Forrest, Ian McHale, Babatunde Buraimo

Centre for Sports Business, University of Liverpool Management School, Liverpool, L69 7ZH, United Kingdom

A R T I C L E I N F O

Article history:
Received 10 January 2022
Revised 3 August 2022
Accepted 10 August 2022

JEL classification:
G14
G41
J16
Z22

Keywords:
Gender bias
Discrimination
Horseracing
Heckman
Female jockey

A B S T R A C T

Under-representation of women persists in many industries and represents an important area of concern for society. We use a revealed preference approach to test for bias against females in an underexplored environment. Whilst much use has been made of the financial industry to examine how market prices reveal implicit views on the relative productivity of men and women, our setting offers advantages through both volume of data and unambiguity of outcome. Over a 20-year period, the effect of jockey gender on fixed price betting odds was examined in National Hunt racing. Employing censored regression to account for non-finishers we find female jockeys to be underestimated by the UK betting market. Results indicate an increasing trend for underestimation in recent years, despite growing representation and rising performance levels of female jockeys. We conclude that mistake-based discrimination and confirmation bias may be impacting efficiency in the betting market. The market might recognise some improvement in female performance but may be failing to adapt at the speed with which female jockeys are professionalising.

1. Introduction

Under-representation of females in the workplace, particularly at managerial level, is well recognised in many sectors. Society's attitudes towards female workers have been studied in settings ranging from orchestras (Goldin and Rouse, 2000) to CEO appointments (Lee and James, 2007), but it is precisely the low participation rates of women that impede such analysis. Horseracing, a sport in which male and female jockeys compete on equal terms, offers an unparalleled opportunity to study gender equality in labour markets owing to the wealth of data available. Each race can be considered a unique hiring event, with a definitive outcome. This offers substantive advantages over professions in which posts are held for extended periods and performance outcomes are often indeterminate. Horseracing betting markets provide insight into public opinion since pricing varies according to betting activity. The accuracy of these opinions can then be measured against race outcome. Any gender bias in the perceptions of horseracing's betting customers would be readily identifiable through mispricing of odds by jockey gender.

Females may be under-bet due to bettors' acceptance of a worse offer to avoid the distaste of backing a woman. Becker (1958) recognises that bias against certain groups may lead to discriminatory behaviour among those who wish to avoid contact with members of these groups. Whilst this 'taste-based' discrimination might account for the hiring bias observed in racing by Binder et al. (2020), it is less likely to drive the behaviour of betting customers, who would not nor-

* Corresponding author.
E-mail address: b.buraimo@liverpool.ac.uk (B. Buraimo).

https://doi.org/10.1016/j.jebo.2022.08.012
0167-2681/

mally have direct contact with the subjects of their bets. More plausible is that bettors misperceive the ability of women relative to men, consistent with Wolfers' (2006) mistake-based theory of discrimination.

Perceptions of the public are difficult to evaluate by stated preference methods since survey responses may be subject to social desirability bias (respondents give the answer that will please the interviewer) or indeed respondents may be unaware of the subconscious beliefs which drive their everyday choices. Wolfers adopted the alternative revealed preference approach and searched for (but did not find) evidence that stock market returns on companies led by female CEOs would, on average, yield higher returns because market participants underestimated female leaders.

More recently, Niessen-Ruenzi and Ruenzi (2019) adopted a revealed preference approach to examine gender bias amongst customers in the financial industry. The setting is comparable to horseracing in that investors would not normally have direct contact with fund managers. Controlling for fund performance and risk profile, the authors document lower inflows into mutual funds where the fund is female-managed, implying customer bias against females. This may also be related to under-representation of women in the financial services sector. Hirers may either take account of customer perceptions or may share these perceptions.

In the case of horseracing, a relationship between customer bias and employer hiring decisions is likely. Underestimation of female jockeys might in part result from the under-representation of females in the jockey population, making it more difficult to assess female performance than that of males. It is equally possible that confirmation bias and cognitive dissonance contribute to bettors' misinterpretation of the available information. Bettors may selectively focus on evidence that supports beliefs held and may discard conflicting information. This propensity to protect pre-existing beliefs has been identified as a factor capable of distorting financial decisions. In a consumer setting, Yin et al. (2016) provide evidence that customers give greater weighting to online reviews that confirm their initial beliefs. Similarly, Antoniou et al. (2013) report delayed investor reaction where information contradicts investor sentiment. This reluctance to adjust existing beliefs in response to new information may well affect bettors' decisions regarding female jockeys.

2. Horseracing and gender bias

Horseracing is remarkably underutilised for the study of customer discrimination in the form of gender bias, and we are aware of only two papers in this area. Brown and Yang (2015) employed UK horseracing data to examine gender bias among bettors and more recently, Binder and Grimes (2021) examined this issue in North American horseracing. Interestingly, whilst the former identified a slight underestimation of the ability of female jump jockeys, the latter reported overestimation of females in flat races. Differences in race type and betting markets may account for the contradictory findings[1] but the field is clearly under-researched and warrants further investigation.

National Hunt (NH) or 'jump' racing offers a particularly interesting setting to examine gender bias. This category of race carries the greatest level of physical risk, since participants are required to negotiate obstacles and race over extended distances. Davies et al. (2021) reported the injury-rate per thousand rides to be five times higher in NH racing than on the flat. Any beliefs regarding female weakness or inability to cope with injury might particularly strongly influence bettors' wagering decisions in the case of NH racing. Similarly, any concerns that women lack strength or stamina would surely be expected to be exaggerated in this most arduous of race genres. Butler and Charles' (2012) ethnographic study supports this assumption, reporting entrenched beliefs amongst trainers and male jockeys, that women could not possess the physical or mental strength required to ride in jump races. This suggests openly discriminatory attitudes towards female jump jockeys at that time. Therefore, unlike Brown and Yang, we focus solely on National Hunt racing, to examine any change in attitude towards female jump jockeys, as reflected by the betting market, and we employ econometric techniques that more appropriately deal with the nuances present in National Hunt racing.

Despite the recency of Brown and Yang's analysis (employing data from 2003 to 2013), there have been significant changes in the profile of female riders during the intervening period, with the Google Trends web search tool revealing increased UK interest in female jockeys since 2012.[2] Over the past 20 years female jockeys have increasingly received the opportunity to professionalise, as a result of more regular riding opportunities. The average number of rides per female jockey has almost trebled across the study period,[3] and the associated improvement in female performance is notable. In recent years a series of high-profile female jockeys have enjoyed a level of success previously not seen in British horseracing, with women winning Grade 1 races at both the Aintree and Cheltenham Festival meetings.[4] Prior to 2015 females rarely participated and had never won at this level, but the 2021 Cheltenham Festival saw a single female jockey win five Grade 1 races.[5] Discourse regarding female jockeys in both industry and mainstream media indicates an associated change in attitudes. This quote from *The Telegraph*, following Rachel Blackmore's win in the 2021 Grand National, represents a sentiment increasingly voiced in recent times:

[1] Brown and Yang identified underestimation of female jockeys within UK races over obstacles, using data from fixed-odds betting markets. Binder and Grimes reported overestimation in North American flat races, based on pari-mutuel betting returns.

[2] The percentage of months with no web searches for the term 'female jockeys' prior to 2012 is 71%. After 2012 it is just 26%.

[3] In contrast, there was a 20% increase in the average number of rides for males.

[4] Grade 1 races are the highest classification of NH races, with Aintree and Cheltenham representing the pinnacle of the jump calendar.

[5] 14 Grade 1 races are staged during the meeting.

> "Thanks to Blackmore, whether male or female, from now on riders will surely be referred to simply as jockeys." (White, 2021)

Current media commentary indicates a shift in public consensus, and it might be assumed that changing attitudes towards female jockeys would be reflected in betting markets. In fact, Barrutiabengoa et al. (2021) identified an important link between the volume of media attention and the pricing of matches within the ATP[6] and WTA[7] tennis circuits. The authors discovered a negative association between media attention and over-round.[8] This effect seemingly acts in an opposing direction to our hypothesis that increased coverage of female jockeys should shorten female odds. However, they focus on match pricing in single-sex, paired contests as opposed to individual odds in mixed-sex, multi-competitor events where the overriding media narrative is that of females becoming increasingly competitive with males. In essence though, their findings support our assumption that increased media attention has the capacity to impact pricing and betting market efficiency.

The growing presence of elite female riders operating at the highest level provides strong signalling regarding the quality of female jockeys and has received much media attention. We would therefore expect to see correction of any inefficiencies in the female dimension and, in the contemporary market, would not expect to observe the underestimation of females identified by Brown and Yang. In short, we seek to examine the betting market's reactions to the improved quality of information available, as female jockeys have rapidly professionalised in recent years.

It is important to consider here the relationship between betting markets and the labour market for jockeys. Females are significantly under-represented in the jockey population, restricting the performance information available to the betting market. In addition to limited hiring opportunities, data suggest that female jockeys may also be subject to different selection criteria, regarding the type of rides they receive. Analysis of a range of characteristics highlights distinct differences between male and female rides (Cashmore, 2021). These gender-specific selection criteria may affect the performance differentials of male and female riders and impact public perception of the ability of female jockeys. Furthermore, habitually riding outsiders[9] has the capacity to reinforce the opinion that females are less effective than males. In this way, bias in the labour market, reflected in the betting market, may perpetuate the cycle of gender bias.

In addition to examining betting market responses to recent developments within the labour market for female jockeys, we extend the work of Brown and Yang by refining the methodology. The authors make no mention of how non-finishers are treated and appear to discard these records. This is a risky strategy since failure to finish carries information. We therefore apply censoring to account for non-finishers. This development is particularly important when analysing jump racing as negotiating obstacles and racing over extended distances results in a significant number of runners failing to complete. During our study period 117,426 horses did not complete their race, representing 18% of runners. Inclusion of non-finishers through the application of Heckman's selection model represents an important and novel approach to this previously overlooked issue. The magnitude of non-finishers has the capacity to significantly alter results, particularly as any prejudice in bettors' minds may relate precisely to the ability of a female jockey to complete the course.

Finally, we address the issue of over-round more fully, considering several alternative methods proposed in the literature. Since the total book of a profitable bookmaker will sum to greater than one, the excess (over-round) must be removed to establish the implied probability expectations for each observation. We experiment with several approaches and select the power method, as proposed by Vovk and Zhdanov (2009) and Clarke (2016), to remove over-round whilst addressing favourite-longshot bias.[10] In addition, we introduce a performance measure that accounts for field size, in recognition of the importance of this variable in comparing relative performance across races.

3. Data

Race records were obtained from the British Horseracing Authority (BHA), racing's governing body in Great Britain. The data relate to the full set of Thoroughbred National Hunt (jump) horseraces from 1st January 2001 to 31st December 2020 in England, Scotland and Wales. The data set contains 68,993 races and 644,536 runners. Each record provides race-level information such as date and category of race. The majority of National Hunt races require the horse to jump and are classified as either hurdle races (with small, uniform obstacles) or steeplechase/hunter chase races (with larger, more variable fences and ditches). In addition, there are a limited number of National Hunt Flat (NHF) races which are specifically for young jump horses to gain race experience prior to being asked to negotiate obstacles. It is logical to include NHF races in the analysis as jump jockeys participate in all categories.

[6] Association of Tennis Professionals

[7] Women's Tennis Association

[8] Over-round is a measure of by how much the sum of the probabilities, defined by quoted odds, exceeds 1. Bookmakers inflate the probabilities for each competitor winning in order to offer 'unfair' odds, allowing them to make a profit. Over-round may therefore be interpreted as the price of bookmaker services in providing a market for the event.

[9] Across the study period 46% of female rides were taken on horses with odds of 15/1 or longer, in comparison to 39% of male rides. Whilst the trend is for convergence, females still ride proportionally more outsiders than males. In 2001 63% of female rides were at 15/1 or longer compared with 43% for males. In 2020 43% of female rides went off at 15/1 or longer compared with 37% for males.

[10] The tendency for bets placed at higher odds (longshots) to yield a worse rate of return than bets placed at lower odds (favourites).

Runner level variables include horse name, jockey name, jockey gender and starting price (SP). Information on jockey gender was absent for 1,256 individuals.[11] To establish the sex of these riders a combination of approaches was employed. Categorisation by title was used for all jockeys listed as Mr/Mrs/Miss/Ms within the *jockey name* field. Where no title was provided, names were checked manually, and all jockeys known to us were categorised accordingly. Google searches produced titles/images allowing identification for the majority of the remainder. It was not possible to determine the gender of five jockeys, accounting for six rides. These individuals were excluded from the analysis.

Females are under-represented in the jockey labour market. During our study period females comprised 25% of jockey licence holders but accounted for just 3.5% of all NH rides. There is a clear trend for growth in the female percentage share of rides, but progress has been slow, with female jump jockeys taking only 6.5% of rides in 2020. Although women receive a relatively small share of rides this still provides 22,256 female observations.

Starting price (SP) was used to calculate the win-probability for each runner. SP is formed from the average of a sample of fixed-odds bookmaker prices at the close of betting/start of a race.[12] Bookmaker profit margins are incorporated into the odds and consequently probability-odds for each race normally sum to greater than 1. For 128 races the sum of the probability-odds was observed to be less than 1. This occurs when a short-priced horse comes under starter's orders but does not start. The absence of an SP for the non-starter results in the probability-odds for that race totalling less than 1. The recorded odds do not then represent the complete market. These races have been excluded from the analysis as it is reasonable to assume that the non-starters occur randomly[13] and the number of cases is very small (883 runners, 0.14% of all runners). For all remaining races the sum of the win probability-odds was greater than 1, with the excess representing bookmaker margin. The extent of bookmaker over-round varies by race, with a range of 0.003 to 0.917 observed in our dataset. Consequently, probability-odds obtained from SPs over-estimate win-probability and are not comparable across races, requiring us to remove over-round.

To obtain implied probabilities from the win probability-odds, four methods were considered: the additive method, the multiplicative or normalisation method, the Shin method (Shin, 1992) and the power method (Vovk and Zhdanov, 2009; Clarke, 2016). The additive method divides the over-round equally between all outcomes but is not widely used since it can produce negative probabilities for rank outsiders. The multiplicative method removes the over-round proportionally but fails to account for favourite-longshot bias. The Shin method is based on an assumed fraction of knowledgeable bettors and protects against favourite-longshot bias but can adjust outsiders' odds too much. The power method is an extension of the additive and multiplicative methods, raising the probabilities to a constant power.

If p_i represents the implied probability that we wish to obtain and π_i represents the probability-odds for each horse within a race, the power method may be written as:

$$\mathrm{p}_i = {\pi_i}^k$$

where k is selected so that $\mathrm{sum}(\mathrm{p}_i) = 1$

This method effects a greater change to longshots than to favourites and adjusts mid-priced horses the least. Using a data set from Australian horserace betting, Clarke et al. (2017) tested the forecasting performance of sets of odds derived from each adjustment method. The additive method proved superior on each of three measures of forecasting accuracy[14] but the possibility of producing negative probabilities for rank outsiders renders this method impractical. (The additive method produced 124,281 negative probabilities with our dataset.) Of the remaining three methods, the power method was the best on two indicators of forecasting and next-best on the other. Based on this evidence, the power method was selected to remove over-round and obtain implied probabilities for each observation.

The final data set consists of 68,865 races and 643,647 runners, after removing runners with a jockey of unknown gender and excluding all races that do not represent the full betting market.

4. Analysis

4.1. Modelling win-probability

Our initial analysis seeks to examine whether female jockeys win races more frequently than their implied probabilities would predict. A variable equalling 1 for a win and 0 for all other positions is employed as a measure of race success. Using an ordinary least squares (OLS) specification the *win* indicator variable is regressed on the implied win-probability of the runner and a variable equal to 1 if the jockey is female. In this first model we follow Brown and Yang's general approach, employing a linear model for ease of interpretation and using win-probabilities, calculated from starting prices, to test for market efficiency. Bookmaker odds are a direct representation of the betting market's opinion of each runner's probability of winning. These opinions are then measured against race outcome to test the efficiency of the market. Inclusion of a female

[11] These cases are jockeys visiting from abroad for whom the BHA does not hold gender data.

[12] On June 1st, 2020, the sample was altered from a selection of bookmakers at the course to a selection of large off-course operators. Our results are robust to both including and excluding observations after this date, suggesting the change had no material effect on the pricing of races.

[13] Reasons might include the horse losing a shoe, the rider becoming unseated and the horse running loose on course, the horse 'losing its action' (lameness) or any concerns from the trainer or veterinary officer that cause the horse to be deemed unfit to race.

[14] The authors measured the distribution in the adjusted win-probability assigned to the winner, the logloss and the root-mean squared error.

Table 1
Summary statistics for each variable.

	mean	standard deviation	minimum	maximum
win percentage	10.7	–	–	–
percentage of female jockeys	3.45	–	–	–
implied probability	0.1069	0.1168	0.0004	0.9625

Number of runners per year

2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
29,413	30,627	30,122	33,907	34,520	35,364	33,573	36,006	34,009	31,160
2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
34,693	31,504	33,289	31,176	31,557	31,923	31,938	32,333	31,376	25,157*

* The reduction observed in 2020 reflects that racing was shut down for several months due to Covid-19 restrictions.

Table 2
Regression results, dependent variable win indicator.

	Model 1 (linear)	Model 2 (linear)	Model 3 (logistic with spline fitted)
	Coefficient estimates		
implied probability	1.0177***	1.0161***	–
female	0.0068***	–	–
probability*female*year 2001	–	−0.1544	−1.1208
probability*female*year 2002	–	0.1933**	1.0714
probability*female*year 2003	–	−0.0489	−0.2544
probability*female*year 2004	–	0.1347	0.7361
probability*female*year 2005	–	0.0681	0.3586
probability*female*year 2006	–	−0.1204	−0.6612
probability*female*year 2007	–	0.0915	0.5027
probability*female*year 2008	–	−0.0142	−0.0943
probability*female*year 2009	–	0.0210	0.1000
probability*female*year 2010	–	−0.0712	−0.3936
probability*female*year 2011	–	−0.0054	−0.0390
probability*female*year 2012	–	−0.0333	−0.2095
probability*female*year 2013	–	0.0450	0.2326
probability*female*year 2014	–	0.1742***	0.9395*
probability*female*year 2015	–	0.1855***	0.9659**
probability*female*year 2016	–	0.0297	0.1552
probability*female*year 2017	–	0.1327**	0.7152*
probability*female*year 2018	–	0.0661	0.3500
probability*female*year 2019	–	0.0548	0.2871
probability*female*year 2020	–	0.1354***	0.7123*
constant	−0.0020***	0.0018***	−2.7517***
observations	643,647	643,647	643,647
adjusted R^2	0.1477	0.1478	0.1480
AIC	212,858.5	212,859.2	357,766.3

NB For implied probability the significance test relates to the null hypothesis that the coefficient is 1 (which would be consistent with market efficiency).
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

jockey indicator variable allows identification of any gender bias, with a positive coefficient on female indicating female jockeys to be underestimated by the betting market. Unlike Brown and Yang however, we remove over-round from the raw probability-odds to obtain implied probabilities. In transforming the probability-odds, we address the large variability in race pricing,[15] normalising probabilities across races. Additionally, we account for favourite-longshot bias within the pricing, through use of the power method. This is an important consideration since females ride a disproportionate share of long-priced horses.

In Table 2 (Model 1), we see that implied probability is a strong indicator of race outcome, as would be expected. In the complete absence of favourite-longshot bias, and if there were no other inefficiencies in the market, we would expect a coefficient of 1. Whilst very close to 1, the value of our coefficient estimate suggests that the implied probabilities are not entirely purged of favourite-longshot bias, and we will return to this shortly. We estimate a positive female coefficient, revealing that female jockeys have a 0.7% (p<0.001) higher probability of winning a race than their odds imply. Our results indicate an underestimation of females of a similar magnitude to that identified by Brown and Yang in jump racing. Model

[15] A range of 0.914 is observed in over-round in our dataset.

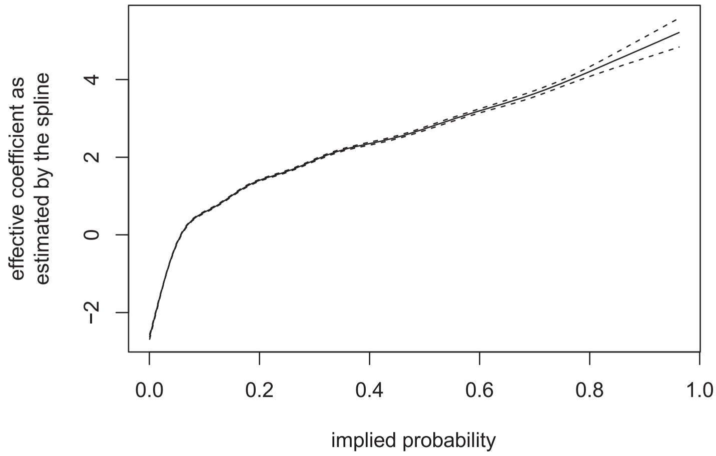

Fig. 1. Plot of implied probability against effective coefficient, as estimated by the spline in Model 3.

1 tests only for underestimation during the full 20-year study period. Further, the specification is such that the impact of female is constrained to be a fixed value addition of 0.0068 to the probability of winning, regardless of bookmaker odds. Since we wish to search for any differences in the level of underestimation across the 20-year time-period, and to allow the impact of a female jockey to vary depending on the odds, in Model 2 we include a three-way interaction term between implied win probability, female and year. Estimating by year allows us to examine the effect of female across time and inclusion of win-probability addresses the differences in the odds spread between male and female riders. Our three-way interaction provides the flexibility to allow the impact of 'female' on the probability of winning to vary between horses at different points in the odds range. For example, a model without this three-way interaction would be constrained to predict the same effect, regardless of whether the horse had a 1% chance of winning or a 50% chance of winning.

As before, implied win-probability is a strong predictor of race outcome. We see some noise in the female effect when examined by year but continue to observe underestimation of female riders, particularly in more recent years. Although only five of the years appear to show significance this may be a result of the reduced statistical power, given the lower number of observations at year level. The 20 estimated coefficients for the interaction term were tested as a group, using the Wald test, and found to be jointly significant (p<0.001). Results suggest that the tendency for females to outperform their odds has become more pronounced in recent years.[16]

Because the conversion of quoted odds to implied probabilities does not entirely purge the data of the effects of longshot bias, there is a risk that the results on the female rider indicator variable will misrepresent the existence of gender bias. For example, females disproportionately ride longshots and market odds may systematically overestimate the chance of a longshot winning. To allow full flexibility and account for any favourite-longshot bias within the estimated probabilities, a spline was used on the probability-odds (Model 3), within a logistic regression. This approach allows our estimates to represent the data at the extremes more accurately (i.e., extreme favourites and longshots).

Fig. 1 shows that higher implied probabilities from the bookmaker's odds have a larger effective coefficient in the logistic regression than smaller implied probabilities. This suggests that, as the implied probability of winning increases, the true probability increases at a faster rate. This is the classic favourite-longshot bias. Even in the presence of the spline which should purge the model of the favourite longshot bias, we continue to observe underestimation of female jockeys.

These first three models clearly indicate the presence of inefficiencies within the British betting market, in the form of underestimation of female jockeys. However, using a *win* indicator as the measure of success fails to capture the relative performance of runners that do not win. This is an important detail since females ride proportionally more longshots and therefore have less opportunity to win.[17] In the next section we make more complete use of the information available, specifically that contained in finish position.

4.2. Modelling finish position

To further examine the underestimation of female jockeys we redefine our performance measure to consider all finishing positions in the analysis. Similar to Brown and Yang, we first calculate each runner's predicted finish position (R) by ranking the win-probabilities for each race.[18] We then construct a variable to represent the horse's relative finish position (relative

[16] Findings are robust to different methods for removing over-round.

[17] It is also possible that SPs may to some degree reflect place-probabilities and normalised odds may not be a pure estimate of win probability, although this is of lesser concern, given the closeness of the coefficient estimate to 1.

[18] Horses with the same odds in a race were ranked according to sporting convention, with both horses tying for the higher placing. For example, joint favourites in a race would both be allocated a ranking of 1, with the next most fancied horse allocated a ranking of 3.

Table 3
Summary statistics for additional variables.

	mean	standard deviation	minimum	maximum
relative performance	0.0348	0.2791	−0.9500	0.9500
relative rank	0.5346	0.2829	0.0300	1.0000

Type of race			
national hunt flat	hurdle	steeplechase	hunter chase
57,016	367,470	200,560	18,601

performance) by comparing actual finish position (A) to predicted finish position. We also take into account the number of runners (n).

$$relative\ performance = \frac{R - A}{n}$$

This variable provides a more practical measure of performance relative to betting market opinions. It also allows better use of the available data since it is possible to produce and analyse this measure for all runners that complete the race. Our relative performance measure differs from that of Brown and Yang through the inclusion of field size. The number of participants in a race is highly relevant when considering the relationship between predicted and actual finish position. Furthermore, it varies greatly, with a range of 2 to 40 runners in our data set.[19] Using Brown and Yang's measure, the spread of relative performance values is excessively wide and lacks comparability between races. Our refined measure of relative performance produces values in the range −1 to 1 that are normalised across races, whilst still placing additional weighting on runners that fail to meet/exceed expectations by many places. Not only does the inclusion of number of runners improve inter-race comparability but it also addresses the non-uniform distribution of female riders across field size. The mean relative performance of 0.0348 (Table 3) demonstrates a tendency for both genders to improve on predicted finish position. This is a consequence of some horses failing to complete the course. Interestingly, the mean relative performance for females is 0.082, giving an indication that females may be outperforming their odds ranking, relative to males.

Using an OLS specification, our redefined measure of relative performance is initially regressed on an indicator variable equal to 1 if the jockey is female (Table 4, Model 4).[20] A positive coefficient for female indicates that relative to males, female jockeys are undervalued and improve their finishing position by more than male jockeys. We do indeed see an underestimation of females, confirming the market inefficiencies identified through win-probability analysis. The magnitude of the effect is economically meaningful. Across the data period, our model estimates that in a field of 20, females finish one place better than their odds would imply ($p < 0.001$).

As before, we focus on any differences in the level of underestimation across the time-period, by introducing an interaction term between female and year in Model 5. The superior AIC for Model 5 suggests improved fit when the effect of 'female' is allowed to vary across years. The positive coefficient estimated for each of the interaction terms indicates an underestimation of females in each year, but contrary to our win-probability analysis, the magnitude of underestimation shows a downward trend. In 2001 we estimate that in a race with 10 runners, a female would finish 1 place better than their odds imply. In 2020 we estimate the effect to be reduced to ⅓ of a place. These results appear to support our hypothesis; the downward trend seemingly mirrors recent changes to public opinion regarding female jockeys, as evidenced in the media. However, our approach in Models 4 and 5 discards data which may significantly affect the results.

At the level of the individual race, we have a problem of missing information. Suppose we have a field of fourteen runners and ten finish. The finishers are recorded in positions 1 to 10. All others are recorded as not finishing. Any distinction between these non-finishers is suppressed in the data because horse racing has no formula for assigning different levels of performance to horses that do not complete the race. In employing relative finish as our dependent variable, we exclude all non-finishers from the analysis since they are not assigned a finish position. Rejection of these runners is risky, due to the large proportion of horses involved (18% of all runners) and the possibility that females are allotted horses with an inherently lower probability of finishing.

To take account of this issue and to allow us to include all observations, rather than include only finishers, we undertook censored regression analysis where the performance of all non-finishers has been allocated a common value. An obvious model choice would be Tobit (Tobin, 1958). The Tobit model assumes that, had they finished, these horses would have finished in positions 11 to 14 and attempts to deal with the censoring problem i.e., in reality, one should be 11th, one 12th, one 13th and one 14th. Tobit imposes a strong assumption that the same variables affect the probability of being non-censored and the expected value of the outcome, and with the same degrees of importance relative to each other. If it can be assumed that the same variables affect both the probability of finishing and the predicted finishing position, Tobit may be considered an appropriate model. However, any selection bias within the censored data may induce bias in

[19] There were six examples of a male vs female two-horse race. Females won three of these races.
[20] For now, the sample is restricted to horses which completed the course.

Table 4
Regression results, dependent variable relative performance.

	Model 4 (linear) coefficient estimates	Model 5 (linear)
female	0.0489***	–
female*year 2001	–	0.1021***
female*year 2002	–	0.0856 ***
female*year 2003	–	0.0764 ***
female*year 2004	–	0.0905 ***
female*year 2005	–	0.0503 ***
female*year 2006	–	0.0702 ***
female*year 2007	–	0.0742 ***
female*year 2008	–	0.0638 ***
female*year 2009	–	0.0604 ***
female*year 2010	–	0.0595 ***
female*year 2011	–	0.0398 ***
female*year 2012	–	0.0522 ***
female*year 2013	–	0.0470 ***
female*year 2014	–	0.0424 ***
female*year 2015	–	0.0517 ***
female*year 2016	–	0.0404 ***
female*year 2017	–	0.0404 ***
female*year 2018	–	0.0229 ***
female*year 2019	–	0.0227 ***
female*year 2020	–	0.0333 ***
constant	0.0332 ***	0.0332 ***
observations	526,221	526,221
adjusted R^2	0.0010	0.0011
AIC	149,561.5	149,505.6

** $p < 0.05$, * $p < 0.1$.
*** $p < 0.01$.

the estimates of our outcome of interest. If there are non-observed variables that affect both the probability of selection (finishing the race) and the outcome (finishing position if the horse finished), selection bias would violate the assumptions of the Tobit. In our case, non-observed variables may reside within bookmaker odds. A horse's win-probability will represent a combination of the reliability and the speed of the horse, impacting both probability of finishing and expected finish position. A horse's starting price, however, does not differentiate between these components. The unobserved ratio between speed and reliability will clearly have a direct effect on probability of finishing and expected finish position. The resultant selection bias means that if only the finishers were modelled, coefficient estimates would be biased if the regressors were correlated with the non-observed variables.

The Heckman two-step selection model offers an alternative solution for estimating regression models which suffer from selection bias. The technique was formulated by Heckman (1979) to address selection bias in estimating a wage equation for women. Step 1 modelled the probability of a woman being employed and Step 2 estimated salary, given that the woman was employed. Heckman recognised that employed women may possess attributes that non-employed women lacked and his technique used the inverse Mills ratio to adjust the estimation of prospective wages for women not currently employed. Similarly here, horses which finish might have non-observed characteristics which are correlated with covariates in the ordinary least squares model, such that coefficient estimates on those covariates are biased (because some of the effects of the non-observed characteristics are attributed to those covariates). The Heckman is able to estimate the probability of each runner completing the race, together with the conditional expectation of relative performance. The two-step process both tests and corrects for any selection bias in runners that fail to finish. The model has been demonstrated to produce unbiased estimates even when the proportion of missing observations is substantial (Koné et al., 2019). This is an important point since almost $^1/_5$ of horses in our data set fail to complete their race.

The Heckman model consists of two separate equations. The first focuses on selection, using a probit to estimate the probability that a runner will complete the course. The second equation employs linear regression to model the relative performance of each runner, conditional on finishing the race. In order to identify the Heckman model, it is usually necessary to specify a variable relevant only to selection and not to the outcome of interest. Type of race serves this purpose. When National Hunt racing is examined by subcategory the proportion of non-finishers increases progressively, in line with the jumping demands of the race. For our data, 2.2% of runners fail to finish in NHF races, which do not include obstacles. 15.9% of runners do not finish in hurdle races, where 'small' obstacles are negotiated. In steeplechase and hunter chase races, which include the most difficult and varied obstacles, 25.6% and 35.0% of runners fail to complete the course. Type of race is clearly an important variable regarding the probability of finishing but bears no significance on finish position among those who complete in any given race. It therefore provides a valid exclusion restriction for the selection equation.

For step 1 the dependent variable is an indicator variable, equal to 1 if the horse completes the race and 0 for non-finishers. This variable represents selection. In step 2 we use *relative performance* as the dependent variable ($\frac{R-A}{n}$). As before,

Table 5
Step 2 of Heckman selection model, outcome equation (relative performance).

	coefficient estimate
female*year 2001	−0.0129
female*year 2002	0.0077
female*year 2003	−0.0041
female*year 2004	−0.0386**
female*year 2005	−0.0266
female*year 2006	−0.0108
female*year 2007	0.0342*
female*year 2008	0.0198
female*year 2009	0.0227
female*year 2010	0.0507***
female*year 2011	0.0294*
female*year 2012	0.0465***
female*year 2013	0.0283*
female*year 2014	0.0228
female*year 2015	0.0213
female*year 2016	0.0171
female*year 2017	0.0534***
female*year 2018	0.0465***
female*year 2019	0.0492***
female*year 2020	0.0370***
constant	−0.1996***
observations	526,221
adjusted R^2	0.168
inverse mills ratio	0.7769***
sigma	0.5317
rho	1.4610

*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

the predicted rankings used to generate this performance measure are calculated using all runners, since the model itself accounts for non-finishers. Predictor variables include a female indicator variable and an interaction term between female and year, to examine any changes over time. We also include a variable representing *relative rank* in step 1, to serve as a measure of expected horse quality. Relative rank is calculated by ranking the starting price for each horse within a race and dividing by the number of runners in that race, to ensure comparability. Lastly, we include type of race as our instrumental variable in step 1. Coefficient estimates on the race type indicator variables (where NHF is the reference category) are of the expected size and relative magnitudes. The probability of finishing is notably reduced with more demanding jumping requirements.

From the raw data it appears that females are less likely to finish than males[21] but we estimate no such effect within step 1 (see online supplementary information, Table 1). When type of race and relative rank are held constant the probability of completing the course is the same regardless of gender. The probability of completion has increased over time for both males and females. This is probably a result of the BHA's introduction of various safety measures and may also be related to decreasing field sizes.

Relative rank is a significant predictor of the likelihood of completion, as would be expected, with a negative relationship observed. The probability of completion decreases as relative rank increases. For example, in a field of 10 runners, a horse ranked last (relative rank 1) is less likely to finish than the favourite (relative rank 0.1).

The results confirm the need to allow for selection bias in modelling. The highly significant inverse Mills ratio suggests the presence of unobserved variables which raise the probability of completion and the performance level for those who do complete. The application of censoring, using a method to account for selection bias is therefore key to producing unbiased estimations in this analysis.

In our outcome equation (Table 5, Step 2) we uncover some interesting findings. Using uncensored regression in Model 5 we identified a downward trend in the underestimation of female jockeys. Once we account for non-finishers though, the trend is reversed.[22] We now estimate an increasing propensity to under-bet female jockeys. The effect shows stronger significance in recent years but applying the Wald test to all estimated coefficients for the interaction term in the output equation, we found the group to be jointly significant (p<0.00016). These results are qualitatively more similar to the win-probability analysis, where it was not necessary to discard non-finishers.

[21] In our data 20.3% of female ridden horses fail to finish, compared to 18.2% of male ridden horses.
[22] These results are qualitatively robust to both Tobit and the Heckman selection model.

It appears that the betting market tended towards correct assessment of female performance in finishing order at the beginning of the period but more recently the market has been inclined to underestimate female jockeys. This could be considered unexpected given the rising profile of female jockeys and increasing opportunities afforded them in recent years. Statistical discrimination (Phelps, 1972) would predict the contrary; less accurate assessment of female ability in earlier years, since a lack of information would cause individuals to be judged according to the group they belong to. The market now has access to more information on female riders and should be approximating towards efficiency. As females are increasingly able to signal their true ability, bettors should rely less on group information and are better able to assess individuals. However, this assumes that the quality of female riders remains constant, relative to males. If female jockeys have capitalised on increased opportunities, the standard of females may be rising more rapidly than males. We propose that the surprising upward trend in the underestimation of female jockeys may be a result of the market failing to appreciate the rate at which females have been professionalising. Whether due to improved skills, increased opportunity or perhaps both, the mean win rate[23] for females increased more than that of males during the study period. Diminishing field sizes have driven an increased win rate in both male and female jockeys, but while male win rate has increased by 1.8 percentage points, that of females has risen by 4.7 percentage points. It is feasible that bettors may have failed to recognise this growing professionalisation of females.

To eliminate the possibility that certain high-profile female jockeys might be the driving force for this underestimation, the model was re-run without races in which Bryony Frost and/or Rachael Blackmore appeared. This resulted in increased female coefficients in the final three years of the study.[24] Greater underestimation of the general female jockey population, compared to the highest-profile females suggests that bettors are able to compartmentalise their bias. The betting market might acknowledge that there are a small number of female superstars who are exceptional, however, in order to reduce cognitive dissonance, the perception of females in general may remain unchanged, causing underestimation in their odds.

5. Conclusion

Gender discrimination presents an important and significant area of concern for society, that is inherently difficult to quantify. We take advantage of an underutilised arena in which to examine gender bias and our findings suggest that the problem of gender discrimination is far from over. Employing data from 68,865 races across a 20-year period we estimate the effect of jockey gender on fixed price betting odds, to identify statistically significant inefficiencies within the market. Comparison of expected and observed performance, in terms of both wins and finishing positions, indicates an underestimation of female jump jockeys. While previous work has neglected to account for non-finishers, we apply censored regression in the form of Heckman's selection model to reveal an increasing trend for the UK betting market to underestimate females. These findings are highly surprising given the progress of female riders, both in terms of their heightened media profile and raised performance levels, in recent years. Bettors are increasingly forgoing a degree of profitability, despite the improved information available to inform their wagers.

If non-standard preferences and mistake-based discrimination exist amongst the betting public, despite the associated reduction in profitability, it is not inconceivable that gender bias may be present within the industry itself. In fact, the notable under-representation of females among the jump jockey population may be an indicator of entry barriers and hiring bias within the market for jockeys. Our results highlight the importance of recognising the challenges faced by female jockeys. Contrary to the current media representation that females are now considered equal to males, this analysis suggests the persistence of gender bias in British horseracing.

The results are highly relevant to the horseracing industry and serve to refocus attention towards the elimination of gender bias, an area that is at risk of being neglected as a result of the misguided belief that it is no longer an issue for female jockeys. More broadly, the work also indicates the possibility for gender discrimination in other areas of the industry and highlights the need for examination of gender bias at all levels. These findings are also pertinent to the wider labour market, drawing attention to the potential for mistake-based discrimination in occupations where females are under-represented. Furthermore, even in the event of growing representation and increasing performance levels of females, reduction of gender bias does not necessarily ensue. This work reveals a requirement for greater focus on gender equality in sectors where information available on female performance is limited, as misperceptions around female ability may persist despite signalling to suggest otherwise.

The findings also have implications for the study of betting market efficiency. Sauer's survey of the literature (Sauer, 1998) showed no persistent anomalies in race betting markets beyond the favourite-longshot bias, and there has been little convincing evidence since. However, this paper demonstrates that there remains the potential for anomalies to arise in betting markets.

Our results also reinforce the importance of addressing selection bias in gender-related studies. Application of a two-step Heckman selection model brought about qualitatively different results to preceding models which did not account for selection bias. Failing to attend to sample selection bias in this case would have given rise to misleading results.

[23] Wins per ride.

[24] This exercise was repeated for each of the models in this study. In every case the removal of Blackmore and Frost resulted in greater underestimation of female riders in recent years. The full results are available in the online supplementary information, Tables 2–4.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2022.08.012.

References

Antoniou, C., Doukas, J., Subrahmanyam, A., 2013. Cognitive dissonance, sentiment, and momentum. J. Financ. Quant. Anal. 48 (1), 245–275. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.642.3252&rep=rep1&type=pdf.

Barrutiabengoa, J.M., Corredor, P., Muga, L., 2021. Does the betting industry price gender? Evidence from professional tennis. J. Sports Econom. doi:10.1177/15270025211071026.

Becker, G.S., 1958. Economics of Discrimination. University of Chicago Press.

Binder, A.D., Grimes, P.W., 2021. Defying the odds: horseplayers and female jockeys. Appl. Econ. Lett. doi:10.1080/13504851.2021.1947469.

Binder, A.D., Grimes, P.W., Winterbotham, R.G., 2020. In the money: gender and jockey success on the thoroughbred racetrack. J. Sports Econom. 2 (3), 295–328. doi:10.1177/1527002520975839.

Brown, A., Yang, F., 2015. Does society underestimate women? Evidence from the performance of female jockeys in horse racing. J. Econ. Behav. Organ. 111, 106–118. doi:10.1016/j.jebo.2014.12.031.

Butler, D., Charles, N., 2012. Exaggerated femininity and tortured masculinity: embodying gender in the horseracing industry. Sociol. Rev. 60 (4), 676–695. doi:10.1111/j.1467-954X.2012.02129.x.

Cashmore, V., 2021. All jockeys are equal, but some jockeys are more equal than others. In: Butler, R. (Ed.), Advances in Sports Economics. Agenda Publishing.

Clarke, S.R., 2016. Adjusting true odds to allow for vigorish. In: Stefani, R., Shembri, A. (Eds.), Proceedings of the 13th Australasian Conference on Mathematics and Computers in Sport. MathSport, Melbourne, pp. 111–115.

Clarke, S.R., Kovalchik, S., Ingram, M., 2017. Adjusting bookmakers odds to allow for overround. Am. J. Sports Sci. 5 (6), 45–49. doi:10.11648/j.ajss.20170506.12.

Davies, M., Jackson, K.A., Mackinnon, A.L., Turner, A., Kuznik, K., Hill, J., Newton, J.L., Sanchez Santos, M., 2021. Epidemiology of race day injury in young professional jockeys in Great Britain from 2007 to 2018: a retrospective cohort study. BMJ Open 11 (8). doi:10.1136/bmjopen-2020-044075.

Goldin, C., Rouse, C., 2000. Orchestrating impartiality: the impact of 'blind' auditions on female musicians. Am. Econ. Rev. 90 (4), 715–741. https://www.jstor.org/stable/117305.

Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161. doi:10.2307/1912352.

Koné, S., Bonfoh, B., Daouda, D., Koné, I., Fink, G., 2019. Heckman-type selection models to obtain unbiased estimates with missing measures outcome: theoretical considerations and an application to missing birth weight data. BMC Med. Res. Methodol. 19 (231), 1–13. doi:10.1186/s12874-019-0840-7.

Lee, P.M., James, E.H., 2007. She'-e-os: gender effects and investor reactions to the announcements of top executive appointments. Strateg. Manag. J. 28 (3), 227–241. doi:10.1002/smj.575.

Niessen-Ruenzi, A., Ruenzi, S., 2019. Gender bias in the mutual fund industry. Manag. Sci. 65 (7), 3001–3025. doi:10.1287/mnsc.2017.2939.

Phelps, E.S., 1972. The statistical theory of racism and sexism. Am. Econ. Rev. 62 (4), 659–661. https://www.jstor.org/stable/1806107.

Sauer, R.D., 1998. The economics of wagering markets. J. Econ. Lit. 36 (4), 2021–2064. https://www.jstor.org/stable/2565046.

Shin, H.S., 1992. Prices of state contingent claims with insider traders, and the favorite-longshot bias. Econ. J. 102 (411), 426–435. doi:10.2307/2234526.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26 (1), 24–36. doi:10.2307/1907382.

Vovk, V., Zhdanov, F., 2009. Prediction with expert advice for the brier game. J. Mach. Learn. Res. 10, 2445–2471. https://www.jmlr.org/papers/volume10/vovk09a/vovk09a.pdf.

White, J., 2021. Rachael Blackmore's victory ends any debate about relevance of a jockey's gender. retrieved from https://mobile.twitter.com/telegraphsport/status/1380956827639234561

Wolfers, J., 2006. Diagnosing discrimination: stock returns and CEO gender. J. Eur. Econ. Assoc. 4 (2–3), 531–541. https://www.jstor.org/stable/40005119.

Yin, D., Mitra, S., Zhang, H., 2016. When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. Inf. Syst. Res. 27 (1), 131–144. SSRN https://ssrn.com/abstract=2686218 .