Are Women (Really) More Lenient?

Gender Differences in Expert Evaluations

Bernd Frick, Clarissa Laura Maria Spiess Bru, Daniel Kaimann

Abstract

This study investigates whether women are more lenient in evaluating the performance of others. We examine the gender-specific behavior of female and male critics in expert evaluations, considering their allocated level of experience by using data from high-prestige wine assessments. We demonstrate that women rate, on average, less generously than men, even in direct comparison. In addition, we show that women with advanced experience levels are less generous than the most experienced same-sex reviewer, whereas this effect is not observed for men. Finally, controlling for self-selection into a particular field (i.e., wine critics), this study confirms previous findings using data, e.g., from professional sports: Unobserved heterogeneity drives results generated in lab experiments.

Keywords: Gender Differences, Information Asymmetry, Competitiveness, Overconfidence, Gender Bias, Reviews and Ratings

JEL classification: J16, C33, L66, J2

This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Centre 901 "On-The-Fly Computing" under the project 160364472-SFB901.

1. Introduction

In the past decades, a large body of literature has highlighted that men and women differ in their (competitive) preferences, (mental) attitudes, and (social) behaviors (e.g., Gneezy et al. 2003), offering evidence for the gender (wage) gap as a labor market phenomenon. The gender (wage) gap is associated with minor outcomes and less access to promotions and higher positions (e.g., top-level management) for women in comparison to their male counterparts (e.g., Kunze 2008, Bertrand and Hallock 2001). The lack of real-world settings with appropriate data is formulated as one research gap (e.g., Nekby et al. 2015) to control for the results and significant effects of gender differences in behavior (under competitive pressure) that have been confirmed in experimental studies. The vast consensus of these studies shows that women are more risk averse (e.g., Borghans et al. 2009, Charness and Gneezy 2012, Sarin and Wieland 2016), less (over-)confident (e.g., Barber and Odean 2001, Kling et al. 1999, Adamecz-Völgyi and Shure 2022), and performance poorer under pressure (e.g., Bucciol and Castagnetti 2020), act more altruistically (e.g., Andreoni and Vesterlund 2001), and tend to avoid competitive settings (e.g., Gneezy et al. 2003, Niederle and Vesterlund 2008, Niederle and Vesterlund 2011, Pekkarinen 2015, Ors et al. 2013, Frick 2011, Jørgensen et al. 2022).

In our study, we link two strands of gender behavior in professional evaluations using the example of a market for experience goods. The ratings given by recognized critics and platforms serve as quality signals for consumers to reduce information asymmetries (Nelson 1970, Akerlof 1970). Consequently, these ratings strongly impact the pricing and demand of wines (e.g., Hilger et al. 2011, Friberg and Grönqvist 2012, Oczkowski and Pawsey 2019). Furthermore, expert opinions have the potential to influence wine prices similar as the respective product's objective qualities (e.g., Oczkowski 2016). At the same time, marginal differences in the ratings can result in different rating outcomes and quality levels for the tested wines, which can then significantly impact future prices (e.g., Kaimann et al. 2023), reputation,

and the firm's overall market and sales performance. As a result, this leads to the following research question: Are women more lenient in evaluating the performance of others in comparison to men?

We aim to shed light on how female and male reviewers evaluate in a competitive environment, such as professional wine critics controlling for self-selection, and whether a difference in rating behavior can be found between the professional judges. The critics are organized in a network of female and male wine experts allocated to specific wine regions and countries. The tastings always proceed with equal product information. Therefore, we conclude that female and male critics encounter the same conditions to deliver their scores. For this purpose, we investigate professional raters and their behavior in a nearly equally distributed group of male and female reviewers, contributing to research studying drivers in performance appraisals (e.g., Spence and Keeping 2011).

Although some studies have already examined the consistency or credibility of professional wine reviewers (e.g., Ashton 2012, Luxen 2018, Bodington 2020), the literature on the gendered behavior of wine reviewers in the industry also considers potential reviewer bias in a somewhat limited manner. Previous studies have examined whether differences in the wine scores of men and women can be found, showing no or only slight differences. Bodington and Malfeito-Ferreira (2018) used 1,726 wine evaluation observations from 23 tastings to analyze whether female and male judges assigned identical scores to the same wines. They found no evidence of differential ratings, indicating that no gender effect can be found in professional wine evaluations. Nevertheless, women are more likely to be attributed to being more generous than men (Aguiar et al. 2009, Bilén et al. 2021) and act in a more altruistic way (e.g., Brañas-Garza et al. 2018).

Using real-world data and controlling for product quality, we obtain a separation of gender effects for female and male judges. We consider past ratings to control for the previous performance of both genders. Furthermore, we introduce experience levels to rule out reviewer-

specific effects of rating performance. Moreover, we are able to compare reviewers in side-by-side roles to generate insights into the gender-specific behavior of female and male reviewers across different experience levels. Thus, we aim to identify how men and women, who have already self-selected into a competitive setting, behave in equivalent roles. Second, in an experience goods market, we describe which gender-specific assessment patterns underlie quality signals for reducing information asymmetries. Our findings have strategic implications for firms, enabling them to decide for or against professional assessments of their goods.

Our study contributes to current research in several ways. First, we analyze gender differences in a competitive environment by examining how male and female critics evaluate, controlling for self-selection and expanding the current (predominant experimental) research by analyzing real-world data. In addition, we utilize generated associations to experience levels to shed light on the individual results of each reviewer within the experience levels and the behavior of the reviewers associated with each level. Each experience level represents a heterogeneous team of reviewers, with a composition of female and male members in varying proportions. Further, the research question addresses whether men and women reveal different patterns of evaluation regarding the performance of others, also with increasing professional experience.

In conclusion, we aim to expand current research in the field of gender performance ratings, which is formulated as one research gap by Levy and Williams (2004). Consequently, we investigate the rating behavior of both genders in professional evaluations (e.g., Furnham and Stringfield 2001). We anticipate that this approach will generate valuable insights that can be utilized to optimize organizational structures and performance.

The remainder of this paper is structured as follows. The next chapter outlines related research and introduces the hypotheses. Chapter 3 displayss the descriptive statistics, the econometric approach, and econometric models before presenting the empirical analysis and

results in chapter 4. Chapter 5 provides a concluding discussion of our results with an outlook on potential future research.

2. Previous Literature and Hypotheses

Competition is prevalent in many industries and educational settings and is linked to may individuals' motivation and performance. Studies have consistently found that men tend to be more competitive than women (e.g., Price 2012). This disparity in competitiveness has been explained by societal norms and prejudices about the roles and skills of men and women. Women are frequently supposed to be cooperative, nurturing, and less competitive, whereas men are frequently expected to be forceful, competitive, and ambitious (Eagly and Karau 2002).

Both beneficial and detrimental effects on performance may result from competition. Competition can boost motivation and improve performance in specific situations (Deci et al. 1999). Competition, however, can also result in anxiety and poor performance in some circumstances, particularly for women (Sekaquaptewa and Thompson 2003).

The purpose of this chapter is to review the existing literature on gender differences in different settings to better understand the underlying mechanisms and potential implications for real-world decision-making. From various sets of (social) behavioral and mental phenomena, we consider the attitudes of altruism (Andreoni and Vesterlund 2001) associated with generosity or leniency, the behavior in competitive settings (Gneezy et al. 2003, Niederle and Vesterlund 2008, Niederle and Vesterlund 2011, Pekkarinen 2015, Ors et al. 2013, Frick 2011), and self-confidence (Barber and Odean 2001, Kling et al. 1999) as relevant for our study.

2.1. Gender-Specific Preferences: Are Women More Generous?

Most studies examining altruistic behavior concentrate on different experimental lab settings and the behavioral outcomes they induce (e.g., Heinz et al. 2012, Boschini et al. 2018, Eckel and Grossman 1998). For instance, by including the category price in their experiment, Andreoni and Vesterlund (2001) demonstrated how men and women behave altruistically in

various ways. According to their findings, men are more generous in inexpensive game settings, whereas women are more lenient in inexpensive ones. Aguiar et al. (2009) have presented differences in the expected altruistic behaviors of men and women. They have found men consider women just as generous as men, while women think women are more generous. These findings are partly supported by Brañas-Garza et al. (2018). In one of a few non-experimental studies using comprehensive surveys, they demonstrated that women are more altruistic than men and that both genders also expect women to be such. That women tend to be more altruistic or generous is also supported by Bilén et al. (2021), who have shown that women are more generous than men. Dreber et al. (2014) demonstrate that adolescent girls are more generous and risk-averse than boys. Therefore, we expect women to behave more generously with increasing age in comparison to males. Thus, we formulate our first hypothesis:

H_{1a}: On average, women evaluate more leniently in comparison to their male counterparts.

2.2. Gendered Behavior: Are Women Less Confident than Men?

Overconfidence is a well-researched issue in both psychology and economics. It has been connected to several unfavorable consequences, including making bad decisions and suffering financial losses. Overconfidence, however, has also been linked to a number of issues, including gender and competition (e.g., Reuben et al. 2012, Nekby et al. 2015, Cotton et al. 2015). Overconfidence is the notion that one's skills and assessments are more accurate than they actually are (Moore and Healy 2008). According to research, individuals frequently overestimate their skills and the precision of their forecasts, especially when faced with challenging or ambiguous tasks (Kruger and Dunning 1999). This overconfidence might result in bad choices, such as buying risky stocks or forgoing exploring other options.

Overconfidence and gender behavior have been linked in previous research. According to research, men are more likely than women to be overconfident, especially concerning financial and investment duties (Barber and Odean 2001). The difference between men's and women's

overconfidence in these areas has been connected to societal expectations and preconceptions about men's and women's competence in these fields.

Competition has also been demonstrated to have an impact on overconfidence. Studies have revealed that people often display higher degrees of overconfidence in a competitive context than in a cooperative or non-competitive environment (Mullen and Johnson 1990). This might be because competitiveness fosters a "winner-take-all" mentality, where people are more concerned with outperforming others than honestly evaluating their skills.

In a competitive auction assignment, men considerably overestimate their abilities in comparison to women (Croson and Gneezy, 2009). The gender discrepancy in overconfidence, however, vanished when the same activity was turned non-competitive. This implies that the gender difference in overconfidence may be made worse by competition.

Niederle and Vesterlund (2007) found that males and women perform similarly in a task when the competition is de-emphasized. Still, men often outperform women in the task when the competition is emphasized. This is in line with the theory that competition can make gender differences in overconfidence and performance worse (e.g., Niederle and Vesterlund 2011). This may also affect earnings, as women choosing competition and confidence benefit from higher wages in their early career stages (Kamas and Preston, 2018).

In summary, studies have consistently found that men are more likely to be overconfident than women. Additionally, studies have indicated that competition can worsen the difference between men's and women's overconfidence levels, with males becoming more overconfident and women becoming less so in a competitive setting. Therefore, we formulate our second hypothesis:

H_{2b}: Women evaluate less leniently in comparison to men in equivalent roles.

2.3. Gender in Competitive Settings: Are Women More Lenient in Evaluations?

A large body of literature has focused on gender differences in competitive settings, centering on women and their (lack of) preference for competitive environments and

underperformance under competition (e.g., Niederle and Vesterlund 2008, Croson and Gneezy 2009, Gneezy et al. 2003, Niederle and Vesterlund 2011, Pekkarinen 2015, Jurajda and Münich 2011, Bönte 2015, John 2017). Markowsky and Beblo (2022) have shown in a comprehensive review of 110 experimental studies on gender preferences in the competition that males select competitive situations more often than females. Thus, the majority of studies show that men are more likely to compete or choose to compete (e.g., Vandegrift and Yavas 2009, Sutter and Glätzle-Rützler 2015, Hoyer et al. 2020). This can be ascribed on the one hand to stereotypes or adherence to them and associated gender-typical tasks (e.g., Niederle and Vesterlund 2010, Günther et al. 2010, Iriberri and Rey-Biel 2017, Halladay and Landsman 2022) and, on the other hand, to a higher risk aversion by women (e.g., Sarin and Wieland 2016, Gerdes and Gränsmark 2010), which is reported across different settings investigated, for instance in competitive athletic tournaments (e.g., Frick 2011) or financial decisions (e.g., Charness and Gneezy 2012, Schmidt et al. 2021). However, various studies present contrasting results contradicting a gender gap in competitiveness. For instance, Booth and Nolen (2012a) have demonstrated in their experimental study with students from single-sex and coeducational schools that women do not shy away from competition in same gender environments. Furthermore, they highlight that those young females from single-sex schools act like boys, even when assigned to mixed-gender groups (Booth and Nolen 2012b). In addition, Booth et al. (2014) highlight the composition of gender in the immediate environment to affect the behavior of men and women in terms of risk aversion. They have also shown how females in same-gender groups are more likely to take risks than their female counterparts in mixed-gender groups. Thus, we conclude these findings as plausible to interpret women as being more confident in a same-sex environment or acting more competitively among same-gender colleagues.

Several studies show additional findings disagreeing from the consensus of the literature when considering self-selection in the respective competitive field. First, to start in an athletic

and competitive setting, Nekby et al. (2008) have demonstrated that women who have already self-selected in male-dominated environments are more competitive and confident. Furthermore, Hardies et al. (2013) have confirmed prior research and women's risk aversion while demonstrating that self-selection vanishes gender differences in overconfidence.

In contrast to previous findings, Jetter and Walker (2018) have demonstrated that women compete more aggressively and are more drawn to risk when playing against males, analyzing female behavior in a U.S. game show.

Nonetheless, current research also sheds light on behavior along ages in lifespan as a potential root for gender preferences in competition (e.g., Gneezy and Rustichini 2004, Dreber et al. 2014, Flory et al. 2018). Dreber et al. (2014) analyzed gender differences in adolescence by comparing performance in a competitive environment to that in an uncompetitive environment, observing no gender differences in competitiveness. In addition, Flory et al. (2018) have shown that women seem to become more competitive at around 50 years of age. Following their results, females in this age group show the same preference for competition as males of any age, thus eliminating the gender gap.

Performance appraisals, such as those of employees, are able to illustrate patterns of women's behavior in professional performance evaluations. For instance, Furnham and Stringfield (2001) have pointed out that female supervisors gave the lowest evaluations to male personnel and, furthermore, that male employees were evaluated less favorably than female employees. This finding might suggest a potential reviewer bias based on gender, where women are more critical or less lenient in the performance of others, which is related to assigning lower ratings in professional settings. In addition, Booth et al. (2022) have demonstrated that with increasing experience, gender disparity in terms of skill and performance disappears. In conclusion, we suggest that as women gain experience, they are less lenient in evaluating the performance of others.

Consequently, concerning our study, we do not expect gender differences in entry into competitive fields, e.g., wine critics, and find evidence in our dataset as our sample contains a nearly homogeneous distribution of male and female reviewers. This finding argues for the self-selection of female reviewers in this competitive environment. It provides favorable preconditions to explore gender-specific behavior under comparative circumstances in a real-world setting. Thus, we formulate our third hypothesis:

> *H_3: After controlling for self-selection, women on average evaluate less leniently with increasing experience in in comparison to their male counterparts.*

3. Data and Model

Our panel dataset consists of 81,403 tasting notes from *The Wine Enthusiast Magazine,* one of the most prominent wine magazines providing professional wine ratings for wines produced worldwide. The reviews comprise textual tasting notes and numerical ratings based on the assessment of wine experts, which form a critics network. This network allocates the reviewers to specific wine regions and nations based on their knowledge, expertise, and preferences. Based on the available information from the magazine, the wine assessments follow a specific scheme with several blind tastings, in which the reviewers receive only the vintage and variety information of the specific wine. The descriptive statistics and variables utilized in the econometric model are presented in Table 1.

[Table 1]

Since we are interested in the rating behavior of female and male judges, the variable *rating points* is our dependent variable of interest. We use the variable points on a scale of 80 to 100 as a measure of rating outcome of the judges and proxy of gender-specific behavior and performance. In addition, we include the variables *market price* of the *wine bottle* at the time of evaluation and the product characteristics of the evaluated wine bottle, divided into grape variety, country of origin, and vintage. We exploit this information as controls for wine quality

(e.g., Horowitz and Lockshin 2002, Ling and Lockshin 2003). The price span of the evaluated wines ranges from a minimum of U.S. $4 to a maximum of U.S. $3,300. A kernel density estimation of wine prices revealed a left-skewed distribution. Thus, we utilize the logarithm of prices to obtain a normal distribution of wine prices.

With the present dataset, we cover world-renowned wine-growing regions. The dataset contains wines from eleven countries, with wines from the United States and France forming the group with the most tasting notes (32,372 and 14,379 reviews, respectively). The other wines originated from Italy, Spain, Portugal, Argentina, Chile, Austria, Australia, and Germany. The other wine-producing countries were grouped into one category. Furthermore, the vintages of almost a century are recorded, with the oldest vintage dating back to 1904, while the youngest wine was produced in 2017. In addition, we also include the reviewer of the wines and the gender as variables. A total of 13 reviewers tested the wines. This network of critics consists of seven male and six female judges. Thus, we examine two nearly homogeneous groups of professional critics regarding gender distribution. Since we are also investigating gender-specific review behavior at different experience levels, we assigned the respective experience of the reviewers based on the available number of tasting notes to different levels. This allows us to compare men and women while sharing the same experience levels. The different experience levels have been classified as *junior*, *specialist*, *senior*, and *expert* and allocated to the respective reviewers as dummy variables. The independent variables were tested for correlation at a 5% significance level to avoid multicollinearity and ensure an appropriate selection of variables.

[Table 2]

Table 2 demonstrates a modest correlation for the dependent variables at the 5% significance level. The highest correlation is observed for the variables country of origin and the experience level of senior with a value of 0.589. The correlation coefficient between the dependent variable rating points and wine prices shows a moderate value of 0.404 with a

significance of 5%. This is in line with previous research (e.g., Oczkowski and Doucouliagos 2015). Since critics are assigned to wines on the basis of their experiences, and preferences for wines or regions and test wines from those allocated regions, we control the distribution of wine prices to obtain an unbiased estimation. Therefore, a kernel density estimation was assessed to display the distribution of wine prices for female and male critics.

[Figure 1]

Figure 1 illustrates an approximate normal distribution for both male and female critics. Since the rating points are the measure of performance to compare the critics' rating behaviors, the distribution of the given rating points of each male and female reviewer has also been investigated using a kernel density estimation.

[Figure 2]

Figure 2 shows a normal distribution of the given ratings for male and female judges. Based on these parameters examined, we assume that the reviewers are comparable in their rating behavior. Next, we shed light on the total allocation of reviews to the gender of the reviewers. Table 3 illustrates the distribution of tasting notes in terms of rating points and prices per wine bottle.

[Table 3]

Accordingly, male reviewers contributed 55,311 reviews, while female reviewers authored 26,092 tasting notes. At first glance, women seem to rate higher on average than men, with an average rating of 89.00 points in comparison to 88.44 points. On average, however, women also appear to rate more expensive wines than their male counterparts. The average wine price female critics evaluate is about U.S. $39.40. In contrast, male raters rate wines with an average price per bottle of U.S. $33.36. In the next step, we examine this relationship in more detail, defining different price classes based on the existing price structure to illustrate the distribution of reviews generated by reviewer's gender and price category.

[Table 4]

Table 4 illustrates that women provide the smallest number of reviews in the lowest price class of wine bottles under U.S. $10, holding a 12.51% share of reviews in comparison to their male counterparts in this price class. Female reviewers submit the most reviews in the price category up to U.S. $10-20, with a total number of reviews of about 7,377. In comparison, men also provide the most reviews in this price category (20,482 tasting notes). The percentage of female and male reviews is nearly equal in the wine price category of U.S. $50-100, with a male share of 56.26% and a female share of reviews of about 43.74%. By having men and women critics rate wines in all price classes, a closer look at the average scores in the different price ranges can provide initial indications of whether men and women rate differently.

[Figure 3]

Figure 3 demonstrates the average ratings of male and female reviewers in the different price categories on average. Female reviewers rate higher in the two lowest price categories, contrasted with ratings given by male raters. In the middle price categories of U.S. $20-30, male and female reviewers converge to almost the same average value, indicating no difference in rating outcomes. A different picture emerges in the higher price categories from up to U.S. $30: Women tend to score lower on average in comparison to their male counterparts, with a simultaneous increase in wine price and, thus, presumably increasing product quality. In the highest price category of wines above U.S. $100, the difference between male and female rating averages almost about one rating point. Especially in the context of professional wine rating scales, this difference can be decisive in moving to a different quality level and potentially impact future market prices (e.g., Kaimann et al. 2023).

4. Empirical Analysis and Results

Our empirical analyses shed light on the review patterns of female and male professional critics focusing on two research questions. First, whether female raters evaluate more leniently than men. Second, to what extent do female and male professional critics evaluate differently on an individual level and based on their experience? For the aim of this study, we conduct random effects models with different extensions to test for individual-specific effects of the reviewers. Thus, we have to neglect fixed effects estimations as fixed effects models do not account for time-invariant factors such as gender. Initially, we focus on categories such as market price, vintage, and country of the wines to control for wine quality and reviewers assigned to specific countries of wine origin. The gender of the reviewers is also included in the models. At the same time, we consider the rating scores from the previous evaluation of the reviewers to monitor the consistency of possible evaluation patterns. Equation 1 categorizes rating points as a function, accounting for reviewers, their past performance, and wine quality attributes:

$$\begin{aligned} Rating\ Points_{it} \\ &= \alpha + ß_1\,(log)price_{it} + ß_2\,(Female\ Reviewer)_{it} \\ &+ ß_3\,(Rating\ Points\ Lagged)_{it} + ß_4\,(Variety)_{it} \\ &+ ß_5\,(Country)_{it} + \varepsilon_{it} \end{aligned} \qquad \textbf{(1)}$$

where rating points is the dependent variable and reflects the evaluation of each bottle *j* at time *t* given by the reviewer *g* on a scale of 80-100 points, followed by the log of the market price of the wine bottle at time *t* and the gender of the reviewer *g*, which we utilize as a dummy variable (male = 0, female = 1). The past performance is included in the model by adding the last rating by the respective reviewer. The variable variety describes the specific type of grape for each wine bottle, while the country indicates the origin of the specific product. Both variables serve as dummy control variables and indicators of wine quality.

4.1. Gender-specific Assignments and Past Performance

A sequential random effects model was conducted to empirically test hypotheses H_1 and H_2 and capture the individual effects of the independent variables on rating points (see Table 5). We first estimate the effect of log market price on ratings, including the country and variety dummies as control variables (Model 1). The following model estimates the effect of female reviewers in comparison to male contenders (Model 2), while Model 3 considers past performance but excludes gender. In the last model, the estimation includes all independent variables of interest (Model 4). All estimates show robust and highly significant results at the 99% level. Model 2 shows that female critics score -0.306 points lower on average in comparison to males. When past performance is also considered in Model 4, women assign -0.117 fewer points on average. Model 4 considers 23,113 observations over time, including 14,813 bottles of wine. The explanatory power of the estimated model increases to 58.9% in Model 4. Comprising, the empirical results show that women score lower in professional evaluations and are less lenient or generous than their male counterparts. Consequently, we reject our first hypothesis.

[Table 5]

4.2. Individual Reviewer Outcomes

In a subsequent estimation, we focus on comparing the individual performance of raters. For this purpose, we use a sequential random effects model by first including the log wine price and the reviewers at the individual level (Model 1) before regressing the log market price and the individual reviewers as an impact on the dependent variable rating (Model 2). Similarly, Equation 2 deploys rating points as the dependent variable. The independent variables (log) price, past performance, and reviewer are modeled at the individual level:

$$Rating\ Points_{it}$$

$$= \alpha + ß_1\,(log)price_{it} + ß_2\,(Rating\ Points\ Lagged)_{it} + ß_3\,(Reviewer)_{it} + + ß_4\,(Variety)_{it} + ß_5\,(Country)_{it} + \varepsilon_{it} \quad \textbf{(2)}$$

where rating points is the dependent variable, indicating the rating points awarded by rater *g* for each bottle *j* at time *t* on a scale of 80-100 points, followed by the log market price of the wine bottle at time *t*. We include past performance as an additional independent variable in the model by adding the last score awarded by the rater, followed by the individual rater. The male judge with the highest number of tasting notes was selected as the reference category to compare their rating behaviors. Table 6 presents the regression results.

[Table 6]

Model 1 confirms the results of the previous estimations regarding the highly significant positive effect of the market price on the ratings with an effect size of 0.02746 points. This means that the rating increases positively by 0.02746 points when the market price changes by 1%. We have homogeneous groups of female and male raters in each case, with six raters and one male reference category (Rater 6). The r-squared of the first model amounts to 44%, with a total of 81,403 observations and including 58,290 wine bottles, while the second model r-squared increases to 49.7%, counting 23,113 observations and 14,813 wine bottles. Model 1 shows that almost all female raters have highly significant results at the 99% level. While female evaluator 2 is the only one who assigns an higher score than the reference category, it is noticeable that all other female evaluators assign statistically significant fewer rating points compared to the reference category. By comparison, only three male reviewers (2, 4, 6) appear to assign lower scores than the reference critic. Overall, the estimated coefficients' effect sizes are smaller than those of the female reviewers. Based on the results of this model, we expect female raters to rate on average lower than the reference category, confirming our previous findings that females tend to rate less generously than males. Model 2 regresses past ratings and individual raters on the dependent variable rating scores. With one exception, the estimated

coefficients for males and females are highly significant at the 99% level. The estimation results for the female and male raters show a similar relationship: Both have two raters in their group who score below average than the reference category.

4.3. Reviewer Behavior and Experience Levels

Each reviewer's varying number of reviews indicates different experience levels among the critics. The research question now posed is the extent to which evaluation patterns and gender-specific behaviors exist at different levels of reviewer experience. Professional critics are expected to become more precise in their evaluations during their professional life, with increasing experience. This refers on the one hand to the evaluation of taste impressions in the evaluation process and the accumulation of knowledge and, on the other hand, to familiarity with the processes in their specific area of work. Likewise, monitoring the past assessment outcomes of their fellow campaigners may also play a role in helping them fit into the overall rater ecosystem. In the third estimation, we first monitor the rating outcomes of female and male reviewers and the individual experience levels of the reviewers. Based on the existing reviews in the dataset, we utilized quantiles to calculate the limits of each experience level and categorized the levels in ascending order. We consider the level junior as entry-level. Specialists include the group of reviewers with the number of submitted evaluations in the lower midfield up to the 50th percentile. The senior level reflects the midfield from the 50th percentile, while the expert level includes the group of reviewers with the most ratings provided, starting at the 75^{th} percentile. This allows us to compare male and female reviewers with the same experience level and critics with different experience levels. Furthermore, we account for wine quality by adding two dummy control variables, country of origin and grape variety of the wine. We use Equation 3 as a function for five different sequential random effects estimations:

$$
\begin{aligned}
Rating\ Points_{it} \\
&= \alpha + ß_1(log)price_{it} + ß_2\,(Female\ Reviewer)_{it} \\
&+ ß_3\,(Experience\ Level)_{it} + ß_4\,(Gender)_{it} + ß_5\,(Variety)_{it} \\
&+ \varepsilon_{it}
\end{aligned}
\tag{3}
$$

where rating points is the dependent variable and reflects the rating points for each bottle *j* at time *t* given by the reviewer *g* on a scale of 80-100 points, followed by the log market price of the wine bottle at time *t*. Following, we add the experience level of every reviewer and the gender of the reviewer *g* as binary variables. The results of the estimations are displayed in Table 7.

[Table 7]

Model 1 shows the general evaluation behavior of male and female critics, while Model 2 captures rating behavior at different experience levels without considering gender. More, Model 3 provides insight into gender and experience level. Overall, the r-squared increases from 44.4% in Model 1 to 45.4% in Model 3, indicating similar explanatory power of the estimated models. The first results of the three models are highly significant at the 99% level. Overall, the number of observations in the estimations is about 81,403, which includes 58,290 wine bottles. In all three models, the market price effect is similar in size. A 1% change in log price results in a positive change in rating points of 0.02911, 0.02802, and 0.02897 rating points, respectively. Model 1 shows that female raters score, on average, 0.135 points lower than their male counterparts, which confirms our prior findings. Model 2 demonstrates differences between experience levels. The expert level was selected as the reference category. The results show that the entry-level junior awards an average of 1.160 higher points than the most professional level. Specialists and senior categories also award an average of 0.495 and 0.959 points more than the expert reviewers.

Consequently, these findings indicate that male raters below the expert level tend to award more points on average. This also suggests that critics tend to become increasingly discerning

or less generous in assigning rating scores. Model 3 illustrates that women score -0.346 points lower than men when experience levels are included. The previous results remain robust.

4.4. Gender-specific Review Outcomes Across Experience Levels

At this point, we aim to decisively investigate whether our previous findings depict a general phenomenon among professional critics across the various experience levels or if these findings reflect a gender-specific pattern that could assign different rating behavior to men and women. For this purpose, we specifically examine the behaviors of men and women in the different experience levels to compare behaviors. We split the dataset into two parts to compare men and women directly.

Model 5 consistently demonstrates a higher average score by men across the three levels of experience compared to experts. According to the results, men at the junior level assign an average of 1.067 points higher, professionals assess an average of 0.277 rating points higher, and seniors rate 0.939 points higher than experts. Thus, we observe two distinct patterns. First, all reviewers below the expert level evaluate more leniently. Second, we identify a U-shaped rating curve over experience levels. In summary, women evaluate less leniently and tend to assign a lower point rating on average in comparison to their male counterparts.

4.5. Robustness Check

As the last step, we conduct a robustness check to identify the previous findings as gender-specific behavior. Due to the relatively high correlation of the variable experience level senior with country (see Table 2), we neglected the country of origin as a control variable in the final estimations in order to avoid biased estimators. Simultaneously, this approach enables control at the country level. Table 8 demonstrates the allocation of the individual reviewers to the countries of origin and consequently to the ratings provided.

[Table 8]

A group of eleven female and male reviewers submitted most of the evaluations for U.S. wines. We regress the known parameters on rating points, using U.S. wines as a binary variable, to

confirm within the (heterogeneous) group of reviewers the gender effect of women giving fewer points on average than men. The estimation results in Table 8 (Model 6) demonstrate that women assign almost -0.928 points less than their male counterparts. In summary, we are able to identify that women, on average, give fewer points than men.

5. Discussion

This study examined gender-specific behaviors of professional critics. From a theoretical background of current gender research and typical concept attributions of female and male behavior, we observed that females evaluate the performance of others less generously in comparison to their male counterparts. We identified this effect as gender-specific and revealed gender differences in professional rating behavior among men and women. Our findings suggest that once women have succeeded in entering male-dominated, competitive environments, they tend to behave less leniently or generously than their male colleagues, indicating higher level of competitiveness and / or self-confidence. This behavior is also observable among female peers until they reach the highest ranking possible among their hierarchical performance structures, in our case, the expert level. This may support queen bee syndrome behavior (Staines et al. 1974), in which women separate from lower career levels (Derks et al. 2016) but requires further research.

Moreover, women with advanced experience tend to judge less leniently than the most experienced female judge. In contrast, men below the expert level score higher on average than the most experienced same-sex evaluators. In particular, women appear to be harsher in groups of the same gender, while men tend to be more generous. One possible explanation could stem from the composition of the reviewer genders in the proximate environment (e.g., Booth et al. 2012). Our findings align with studies that examine highly competitive athletic settings. Using data from large running races, Nekby et al. (2008) observed that women tend to be more competitive or self-confident once they have self-selected in male-dominated environments. In

addition, using data from Alpine and Nordic skiing, Frick and Moser (2021) found that differences between men and women in competitiveness vanish when controlling for self-selection. Finally, controlling for self-selection into a particular field (i.e., wine critics), this study confirms previous findings using data from professional sports, for example: Unobserved heterogeneity drives results generated in lab experiments.

Further, our findings have implications for performance assessments of employees, where male managers judge subordinates more leniently than female managers. As a result, male employees with male supervisors are more likely to be promoted to managerial positions due to higher-rated performance appraisals, thus maintaining the gender gap. That women are worse off than men in terms of promotion opportunities is highlighted by Roth et al. (2012). Their findings suggest that although women receive better performance reviews, they fare worse than men regarding promotion opportunities.

Since we also address rating behavior in different price categories in our study, we contribute to research that addresses gender-specific behavior, considering prices and different price categories. For instance, in their experimental study Andreoni and Vesterlund (2001) have shown a gender difference in altruistic behavior when the price is accounted for. Thus, men are more generous in low-priced settings, while women are more lenient in higher-priced ones. The opposite is found in our study (see Figure 3). Figure 3 reveals that men and women behave equally in their rating behavior at about the point range of U.S. $20 to 30, while we notice that women assess fewer points on average for higher-priced wine categories.

In contrast, men are less lenient in low to standard-price categories. Production companies participating in professional assessments benefit from this strategic positioning. This can enhance reputation, sales, and prices which, in the long run, pays off in terms of company performance. These companies are well advised to evaluate this strategic decision against the background of the respective pricing of the evaluated product to avoid any drawbacks.

Future research could also examine the editorial boards of journals in more detail. For example, Bransch and Kvasnicka (2022) have examined the composition of editorial boards of scientific journals and found that women are more likely to publish their research when there is a higher proportion of women on the editorial board, at once, the quality of the publications increases. Particularly for German wine guides, Frick (2020) has demonstrated that changes in editorial boards or the composition of professional wine critic teams have a statistically significant, yet minor impact on wine ratings. Thus, the (exogenous) gender-mixed composition of professional reviewer teams based on experience and their outcomes could be investigated in the future, especially considering the findings by Fenoll and Zaccagni (2022). They have shown that female-dominated teams perform worse than other types of teams when teams are formed exogenously. However, these disparities disappear when teams are formed endogenously. Against this background, the fictitious (exogenous) assignment of reviewers to experience levels may provide insights into the gender-specific behavior of reviewers. For this purpose, the respective gender composition of professional reviewer teams needs to be examined in more detail (e.g., Apesteguia et al. 2012).

6. References

Adamecz-Völgyi, A., & Shure, D. (2022). The gender gap in top jobs - the role of overconfidence. *Labour Economics*, 79, 102283.

Aguiar, F., Brañas-Garza, P., Cobo-Reyes, R., Jimenez, N., & Miller, L. M. (2009). Are women expected to be more generous?. *Experimental Economics*, 12(1), 93-98.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488-500.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics*, 116(1), 293-312.

Apesteguia, J., Azmat, G., & Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58(1), 78-93.

Ashton, R. H. (2012). Reliability and consensus of experienced wine judges: Expertise within and between?. *Journal of Wine Economics*, 7(1), 70-87.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1), 261-292.

Bertrand, M., & Hallock, K. F. (2001). The gender gap in top corporate jobs. *ILR Review*, 55(1), 3-21.

Bilén, D., Dreber, A., & Johannesson, M. (2021). Are women more generous than men? A meta-analysis. *Journal of the Economic Science Association*, 7(1), 1-18.

Bodington, J. (2020). Rate the raters: A note on wine judge consistency. *Journal of Wine Economics*, 15(4), 363-369.

Bodington, J., & Malfeito-Ferreira, M. (2018). Do female and male judges assign the same ratings to the same wines? Large sample results. *Journal of Wine Economics*, 13(4), 403-408.

Bönte, W. (2015). Gender differences in competitive preferences: New cross-country empirical evidence. *Applied Economics Letters*, 22(1), 71-75.

Booth, A. L., & Nolen, P. (2012a). Gender differences in risk behaviour: Does nurture matter?. *The Economic Journal*, 122(558), F56-F78.

Booth, A., & Nolen, P. (2012b). Choosing to compete: How different are girls and boys?. *Journal of Economic Behavior & Organization*, 81(2), 542-555.

Booth, A., Cardona-Sosa, L., & Nolen, P. (2014). Gender differences in risk aversion: Do single-sex environments affect their development?. *Journal of Economic Behavior & Organization*, 99, 126-154.

Booth, A., Hayashi, R., & Yamamura, E. (2022). Gender differences in tournament-performance over time in single-sex and mixed-sex environments. *Labour Economics*, 76, 102173.

Borghans, L., Heckman, J. J., Golsteyn, B. H., & Meijers, H. (2009). Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association*, 7(2-3), 649-658.

Boschini, A., Dreber, A., von Essen, E., Muren, A., & Ranehill, E. (2018). Gender and altruism in a random sample. *Journal of Behavioral and Experimental Economics*, 77, 72-77.

Brañas-Garza, P., Capraro, V., & Rascon-Ramirez, E. (2018). Gender differences in altruism on Mechanical Turk: Expectations and actual behaviour. *Economics Letters*, 170, 19-23.

Bransch, F., & Kvasnicka, M. (2022). Male Gatekeepers: Gender bias in the publishing process?. *Journal of Economic Behavior & Organization*, 202, 714-732.

Bucciol, A., & Castagnetti, A. (2020). Choking under pressure in archery. *Journal of Behavioral and Experimental Economics*, 89, 101581.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1), 50-58.

Cotton, C. S., Li, C., McIntyre, F., & Price, J. P. (2015). Which explanations for gender differences in competition are consistent with a simple theoretical model?. *Journal of Behavioral and Experimental Economics*, 59, 56-67.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448-474.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627-668.

Derks, B., Van Laar, C., & Ellemers, N. (2016). The queen bee phenomenon: Why women leaders distance themselves from junior women. *The Leadership Quarterly*, 27(3), 456-469.

Dreber, A., Von Essen, E., & Ranehill, E. (2014). Gender and competition in adolescence: Task matters. *Experimental Economics*, 17(1), 154-172.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573-598.

Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal*, 108(448), 726-735.

Fenoll, A. A., & Zaccagni, S. (2022). Gender mix and team performance: Differences between exogenously and endogenously formed teams. *Labour Economics*, 79, 102269.

Flory, J. A., Gneezy, U., Leonard, K. L., & List, J. A. (2018). Gender, age, and competition: A disappearing gap?. *Journal of Economic Behavior & Organization*, 150, 256-276.

Friberg, R., & Grönqvist, E. (2012). Do expert reviews affect the demand for wine?. *American Economic Journal: Applied Economics*, 4(1), 193-211.

Frick, B. (2011). Gender differences in competitive orientations: Empirical evidence from ultramarathon running. *Journal of Sports Economics*, 12(3), 317-340.

Frick, B. (2020). The legacy of gurus: The impact of Armin Diel and Joel Payne on winery ratings in Germany. *Journal of Wine Economics*, 15(4), 370-377.

Frick, B., & Moser, K. (2021). Are women really less competitive than men? Career duration in Nordic and Alpine skiing. *Frontiers in Sociology*, 5, 539766.

Furnham, A., & Stringfield, P. (2001). Gender differences in rating reports: Female managers are harsher raters, particularly of males. *Journal of Managerial Psychology*, 16(4), 281-288.

Gerdes, C., & Gränsmark, P. (2010). Strategic behavior across gender: A comparison of female and male expert chess players. *Labour Economics*, 17(5), 766-775.

Gneezy, U., Niederle, M., Rustichini, A. (2003): Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics* 118(3), 1049–1074.

Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, 94(2), 377-381.

Günther, C., Ekinci, N. A., Schwieren, C., & Strobel, M. (2010). Women can't jump?—An experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization*, 75(3), 395-401.

Halladay, B., & Landsman, R. (2022). Perception matters: The role of task gender stereotype on confidence and tournament selection. *Journal of Economic Behavior & Organization*, 199, 35-43.

Hardies, K., Breesch, D., & Branson, J. (2013). Gender differences in overconfidence and risk taking: Do self-selection and socialization matter?. *Economics Letters*, 118(3), 442-444.

Heinz, M., Juranek, S., & Rau, H. A. (2012). Do women behave more reciprocally than men? Gender differences in real effort dictator games. *Journal of Economic Behavior & Organization*, 83(1), 105-110.

Hilger, J., Rafert, G., & Villas-Boas, S. (2011). Expert opinion and the demand for experience goods: An experimental approach in the retail wine market. *Review of Economics and Statistics*, 93(4), 1289-1296.

Horowitz, I., & Lockshin, L. (2002). What price quality? An investigation into the prediction of wine-quality ratings. *Journal of Wine Research*, 13(1), 7-22.

Hoyer, B., van Huizen, T., Keijzer, L., Rezaei, S., Rosenkranz, S., & Westbrock, B. (2020). Gender, competitiveness, and task difficulty: Evidence from the field. *Labour Economics*, 64, 101815.

Iriberri, N., & Rey-Biel, P. (2017). Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. *Journal of Economic Behavior & Organization*, 135, 99-111.

Jetter, M., & Walker, J. K. (2018). The gender of opponents: Explaining gender differences in performance and risk-taking?. *European Economic Review*, 109, 238-256.

John, J. P. (2017). Gender differences and the effect of facing harder competition. *Journal of Economic Behavior & Organization*, 143, 201-222.

Jørgensen, L. K., Piovesan, M., & Willadsen, H. (2022). Gender differences in competitiveness: Friends matter. *Journal of Behavioral and Experimental Economics*, 101, 101955.

Jurajda, Š., & Münich, D. (2011). Gender gap in performance under competitive pressure: Admissions to Czech universities. *American Economic Review*, 101(3), 514-518.

Kaimann, D., Spiess Bru, C. L. M., & Frick, B. (2023): Ratings meet prices: The dynamic relationship of quality signals. *Journal of Wine Economics*, 18(3), 226-244.

Kamas, L., & Preston, A. (2018). Competing with confidence: The ticket to labor market success for college-educated women. *Journal of Economic Behavior & Organization*, 155, 231-252.

Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin*, 125(4), 470-500.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134.

Kunze, A. (2008). Gender wage gap studies: Consistency and decomposition. *Empirical Economics*, 35(1), 63-76.

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30(6), 881-905.

Ling, B. H., & Lockshin, L. (2003). Components of wine prices for Australian wine: How winery reputation, wine quality, region, vintage, and winery size contribute to the price of varietal wines. *Australasian Marketing Journal (AMJ)*, 11(3), 19-32.

Luxen, M. F. (2018). Consensus between ratings of red Bordeaux wines by prominent critics and correlations with prices 2004–2010 and 2011–2016: Ashton revisited and expanded. *Journal of Wine Economics*, 13(1), 83-91.

Markowsky, E., & Beblo, M. (2022). When do we observe a gender gap in competition entry? A meta-analysis of the experimental literature. *Journal of Economic Behavior & Organization*, 198, 139-163.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502-517.

Mullen, B., & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology*, 29(1), 11-28.

Nekby, L., Thoursie, P. S., & Vahtrik, L. (2008). Gender and self-selection into a competitive environment: Are women more overconfident than men?. *Economics Letters*, 100(3), 405-407.

Nekby, L., Skogman Thoursie, P., & Vahtrik, L. (2015). Gender differences in examination behavior. *Economic Inquiry*, 53(1), 352-364.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2), 311-329.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. *The Quarterly Journal of Economics*, 122(3), 1067-1101.

Niederle, M., & Vesterlund, L. (2008). Gender differences in competition. *Negotiation Journal*, *24*(4), 447-463.

Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, 24(2), 129-144.

Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, 3(1), 601-630.

Oczkowski, E. (2016). Identifying the effects of objective and subjective quality on wine prices. *Journal of Wine Economics*, 11(2), 249-260.

Oczkowski, E., & Doucouliagos, H. (2015). Wine prices and quality ratings: A meta-regression analysis. *American Journal of Agricultural Economics*, 97(1), 103-121.

Oczkowski, E., & Pawsey, N. (2019). Community and expert wine ratings and prices. *Economic Papers: A Journal of Applied Economics and Policy*, 38(1), 27-40.

Ors, E., Palomino, F., & Peyrache, E. (2013). Performance gender gap: Does competition matter?. *Journal of Labor Economics*, 31(3), 443-499.

Pekkarinen, T. (2015). Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations. *Journal of Economic Behavior & Organization*, 115, 94-110.

Price, C. R. (2012). Gender, competition, and managerial decisions. *Management Science*, 58(1), 114-122.

Reuben, E., Rey-Biel, P., Sapienza, P., & Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization*, 83(1), 111-117.

Roth, P. L., Purvis, K. L., & Bobko, P. (2012). A meta-analysis of gender group differences for measures of job performance in field studies. *Journal of Management*, 38(2), 719-739.

Sarin, R., & Wieland, A. (2016). Risk aversion for decisions under uncertainty: Are there gender differences?. *Journal of Behavioral and Experimental Economics*, 60, 1-8.

Schmidt, U., Friedl, A., Eichenseer, M., & de Miranda, K. L. (2021). Social comparison and gender differences in financial risk taking. *Journal of Economic Behavior & Organization*, 192, 58-72.

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, 39(1), 68-74.

Spence, J. R., & Keeping, L. (2011). Conscious rating distortion in performance appraisal: A review, commentary, and proposed framework for research. *Human Resource Management Review*, 21(2), 85-95.

Staines, G., Tavris, C., & Jayaratne, T. E. (1974). The queen bee syndrome. *Psychology today*, 7(8), 55-60.

Sutter, M., & Glätzle-Rützler, D. (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science*, 61(10), 2339-2354.

Vandegrift, D., & Yavas, A. (2009). Men, women, and competition: An experimental test of behavior. *Journal of Economic Behavior & Organization*, 72(1), 554-570.

7. Tables and Figures

Table 1: Descriptive Statistics

Variables	Definition	Observations	Mean	Std. Dev.	Minimum	Maximum
Points	The rating points on a scale of 80-100 points for each bottle *j* at time *t*.	81,403	88.62	3.01	80	100
Price	The market price of the wine bottle *j* at time *t* in U.S. $.	81,403	35.29	44.15	4	3300
Reviewer	The reviewer *i* for corresponding wine bottle *j*.	81,403	7.85	3.29	1	13
Gender	Dummy variable indicating reviewer's gender with male = 0, female =1.	81,403	.32	.47	0	1
Variety	The grape variety of every bottle *j* at time *t*.	81,403	351.96	188.95	1	690
Country	The country of origin of every wine bottle *j* at time *t*.	81,403	7.94	3.14	1	11
Wine Bottle	The total number of wine bottles.	81,403	29132.32	16759.71	1	58290
Harvest Year	The vintage of every wine bottle *j* given in years.	81,403	2011.53	3.31	1904	2017
Junior Level	The dummy variable indicates the reviewer's experience level based on their tasting notes.	81,403	.11	.31	0	1
Specialist Level	The dummy variable indicates the reviewer's experience level based on their tasting notes.	81,403	.16	.36	0	1
Senior Level	The dummy variable indicates the reviewer's experience level based on their tasting notes.	81,403	.27	.45	0	1
Expert Level	The dummy variable indicates the reviewer's experience level based on their tasting notes.	81,403	.47	.5	0	1

Table 2: Pairwise Correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Points	1.00											
(2) Price	0.404*	1.00										
(3) Reviewer	-0.026*	0.057*	1.00									
(4) Gender	0.086*	0.064*	-0.178*	1.00								
(5) Variety	0.00	-0.049*	-0.157*	0.063*	1.00							
(6) Country	0.091*	0.016*	0.216*	0.078*	0.119*	1.00						
(7) Bottle	-0.00	-0.013*	-0.051*	0.057*	0.055*	0.056*	1.00					
(8) Harvest Year	0.054*	-0.097*	-0.035*	0.111*	0.01	0.079*	-0.007*	1.00				
(9) Junior Level	0.033*	-0.060*	-0.390*	0.201*	0.031*	-0.014*	0.029*	0.086*	1.00			
(10) Specialist Level	-0.00	-0.018*	-0.308*	-0.011*	0.042*	0.031*	0.051*	-0.022*	-0.149*	1.00		
(11) Senior Level	0.161*	0.067*	0.406*	0.082*	-0.01	0.589*	0.066*	0.041*	-0.212*	-0.262*	1.00	
(12) Expert Level	-0.162*	-0.010*	0.104*	-0.189*	-0.045*	-0.539*	-0.113*	-0.075*	-0.324*	-0.400*	-0.570*	1.00

**** p<0.01, ** p<0.05, * p<0.1*

Table 3: Number of Tasting Notes by Price and Reviewer's Gender

Points					
Gender	**Tasting Notes**	**Mean**	**Std. Dev.**	**Minimum**	**Maximum**
Female	26,092	89.00	2.83	80	100
Male	55,311	88.44	3.07	80	100

Price per Bottle					
Gender	**Tasting Notes**	**Mean**	**Std. Dev.**	**Minimum**	**Maximum**
Female	26,092	39.40	35.62	4	800
Male	55,311	33.36	47.53	4	3,300

Table 4: Number of Tasting Notes by Price Category and Reviewer's Gender

Price per Wine Bottle	**Tasting Notes**		
	Men	**Women**	**Total**
> U.S. $100	1,417	936	2,353
	60.22	39.78	100.00
U.S. $50-100	6,044	4,699	10,743
	56.26	43.74	100.00
U.S. $30-50	11,408	6,497	17,905
	63.71	36.29	100.00
U.S. $20-30	12,114	6,033	18,147
	66.75	33.25	100.00
U.S. $10-20	20,482	7,377	27,859
	73.52	26.48	100.00
<U.S. $10	3,846	550	4,396
	87.49	12.51	100.00
Total	55,311	26,092	81,403
	67.95	32.05	100.00

Note: The first row has *frequencies*, and the second row has *row percentages*

Table 5: Regression Results I

	(1)	(2)	(3)	(4)
VARIABLES			**Dependent Variable: Rating Points**	
Log (Price)	2.800***	2.813***	2.089***	2.097***
	(0.016)	(0.016)	(0.031)	(0.031)
Female Reviewer		-0.306***		-0.117***
		(0.024)		(0.044)
Rating Points (Lagged N-1)			0.293***	0.292***
			(0.006)	(0.006)
Variety Dummies	YES	YES	YES	YES
Country Dummies	YES	YES	YES	YES
Constant	80.372***	80.305***	55.162***	55.228***
	(1.594)	(1.592)	(0.607)	(0.607)
Observations	81,403	81,403	23,113	23,113
Number of Wine Bottles	58,290	58,290	14,813	14,813
R-squared Overall	0.452	0.454	0.589	0.589

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Regression Results II

	(1)	(2)
VARIABLES	**Dependent Variable: Rating Points**	
Log (Price)	2.746***	
	(0.014)	
Rating Points (Lagged N-1)		0.515***
		(0.006)
Reference Category: Male Reviewer #6		
Female Reviewer		
#1	-0.021	0.097
	(0.045)	(0.081)
#2	1.839***	1.161***
	(0.049)	(0.081)
#3	-0.428***	0.301***
	(0.033)	(0.065)
#4	-0.468***	-0.360**
	(0.065)	(0.148)
#5	-1.287***	-0.725***
	(0.076)	(0.224)
#6	-0.665***	0.372***
	(0.033)	(0.067)
Male Reviewer		
#1	0.317***	0.371***
	(0.044)	(0.082)
#2	-0.158***	0.385***
	(0.042)	(0.089)
#3	0.623***	0.615***
	(0.039)	(0.078)
#4	-1.071***	-0.886***
	(0.030)	(0.058)
#5	0.037	0.293***
	(0.034)	(0.063)
#7	-0.407***	-0.156**
	(0.041)	(0.069)
Constant	79.677***	43.269***
	(0.050)	(0.490)
Observations	81,403	23,113
Number of Wine Bottles	58,290	14,813
R-squared Overall	0.440	0.497

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Regression Results III

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	**Dependent Variable: Rating Points**					
Log (Price)	2.911***	2.802***	2.897***	2.640***	3.024***	3.038***
	(0.015)	(0.016)	(0.015)	(0.029)	(0.018)	(0.031)
Female	-0.135***		-0.346***			-0.928***
	(0.022)		(0.022)			(0.033)
Reference Category: Expert Level						
Junior Level		1.160***	1.035***	0.049	1.067***	1.692***
		(0.048)	(0.033)	(0.077)	(0.047)	(0.291)
Specialists Level		0.495***	0.219***	-0.705***	0.277***	1.234***
		(0.054)	(0.030)	(0.084)	(0.034)	(0.291)
Senior Level		0.959***	0.659***	-0.167**	0.939***	1.685***
		(0.055)	(0.026)	(0.072)	(0.031)	(0.289)
Constant	81.117***	80.535***	81.155***	80.462***	80.811***	78.697***
	(1.604)	(1.587)	(1.591)	(2.233)	(1.575)	(2.351)
Observations	81,403	81,403	81,403	26,092	55,311	32,372
Number of Bottles	58,290	58,290	58,290	19,585	39,587	23,026
Variety Dummies	YES	YES	YES	YES	YES	YES
R-squared Overall	0.444	0.457	0.454	0.393	0.489	0.356

Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Reviewer, Country, Mean and Standard Deviation

Reviewer	Country											
	Argentina	**Australia**	**Austria**	**Chile**	**France**	**Germany**	**Italy**	**Others**	**Portugal**	**Spain**	**USA**	**Total**
Female Reviewer #1						90.023		86.715			87.447	88.445
						2.393		*2.436*			*2.12*	*2.599*
Female Reviewer #2			90.75		90.712			91.894				90.755
			2.264		*2.68*			*1.821*				*2.429*
Male Reviewer #1											88.645	88.645
											2.736	*2.736*
Male Reviewer #2		88.784		85.05	88.755	89.23	85.923	88.337	85.933		85.518	88.558
		2.98		*2.762*	*2.727*	*2.868*	*2.362*	*2.484*	*2.344*		*2.873*	*2.859*
Female Reviewer #3							88.976					88.976
							2.554					*2.554*
Female Reviewer #4					86.901			87.842	84	83	85.308	87.49
					2.269			*2.432*			*.855*	*2.42*
Male Reviewer #3				89							90.104	90.104
											2.586	*2.586*
Male Reviewer #4	86.612			86.449	82.778		89.123	85.553	91.5	87.215	86.373	86.8
	3.187			*2.691*	*3.993*		*3.094*	*2.804*	*.707*	*3.109*	*3.434*	*3.04*
Male Reviewer #5					89.692			89.548			89.127	89.136
					2.016			*2.322*			*2.853*	*2.843*
Male Reviewer #6			89.129		88.673		88.807	89.118	88.324		89.5	88.608
			2.605		*3.108*		*2.361*	*2.315*	*3.059*		*2.121*	*3.076*
Male Reviewer #7								90.31			88.769	88.779
								1.734			*2.509*	*2.507*
Female Reviewer #5								87.205			84.541	86.631
								2.199			*1.713*	*2.371*
Female Reviewer #6											89.22	89.22
											3.042	*3.042*
Total	86.612	88.784	90.241	86.442	88.761	89.884	88.948	87.828	88.316	87.214	89.063	88.619
	3.187	*2.98*	*2.492*	*2.693*	*3.101*	*2.5*	*2.571*	*2.595*	*3.06*	*3.109*	*2.858*	*3.008*

Figure 1: Log (Price) per Wine Bottle by Reviewer's Gender

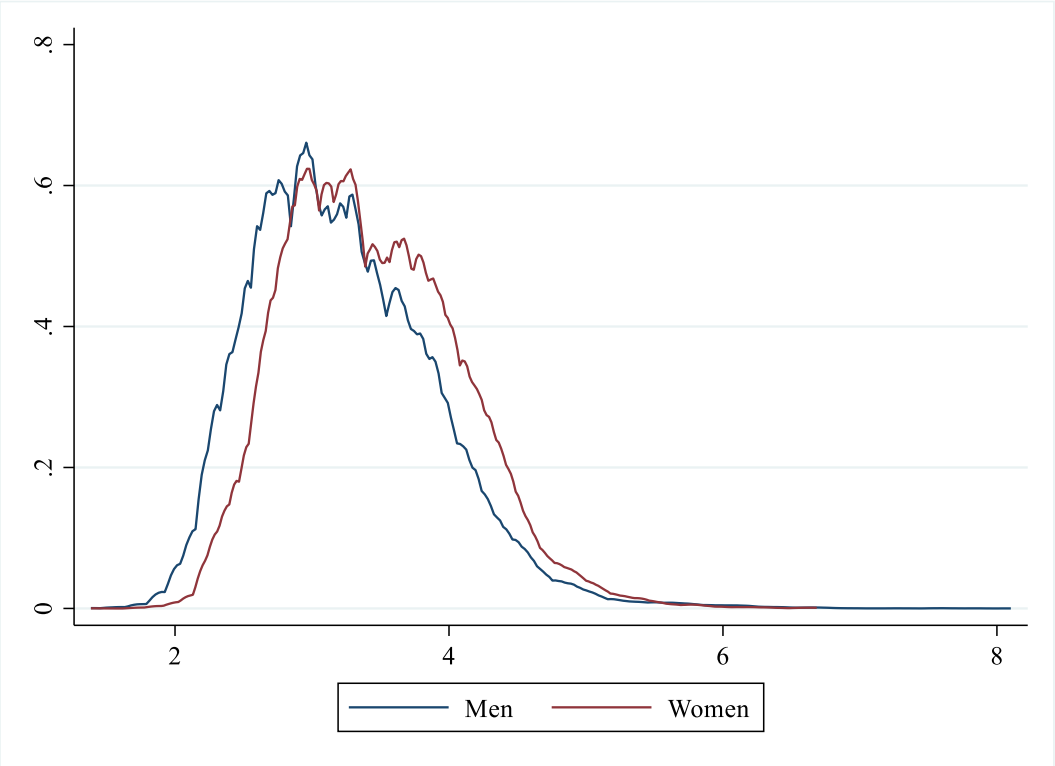


Figure 2: Rating Points by Reviewer's Gender

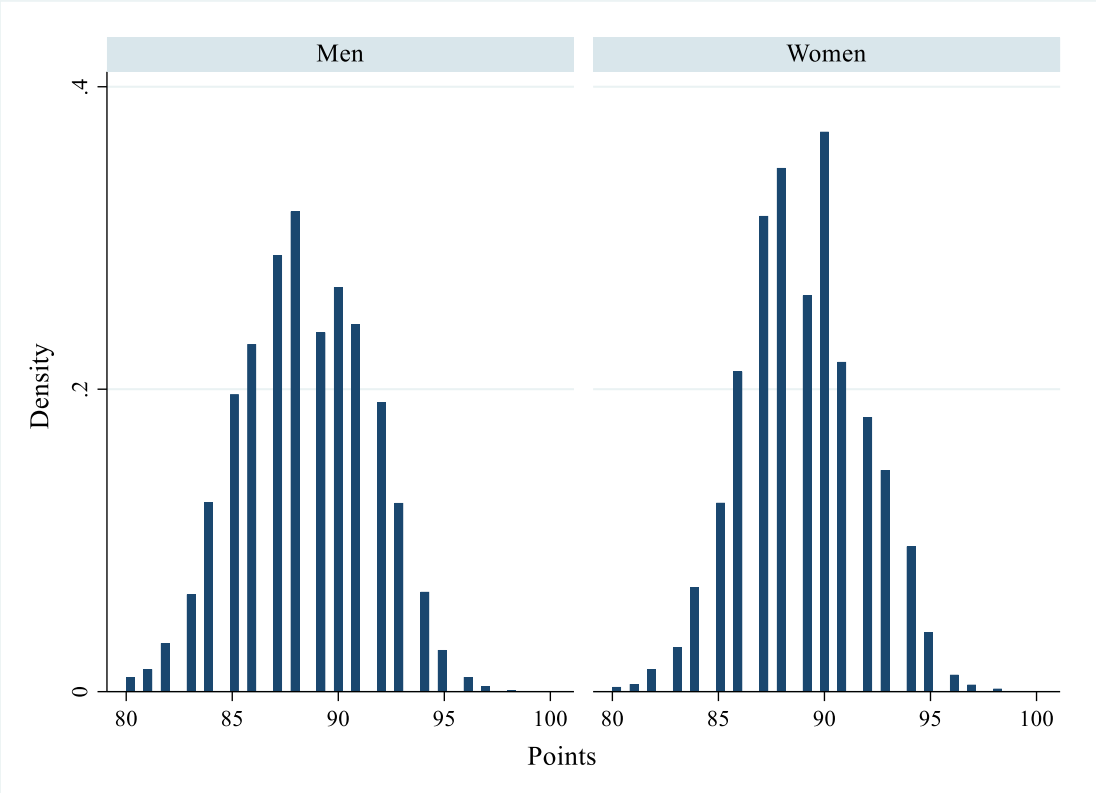
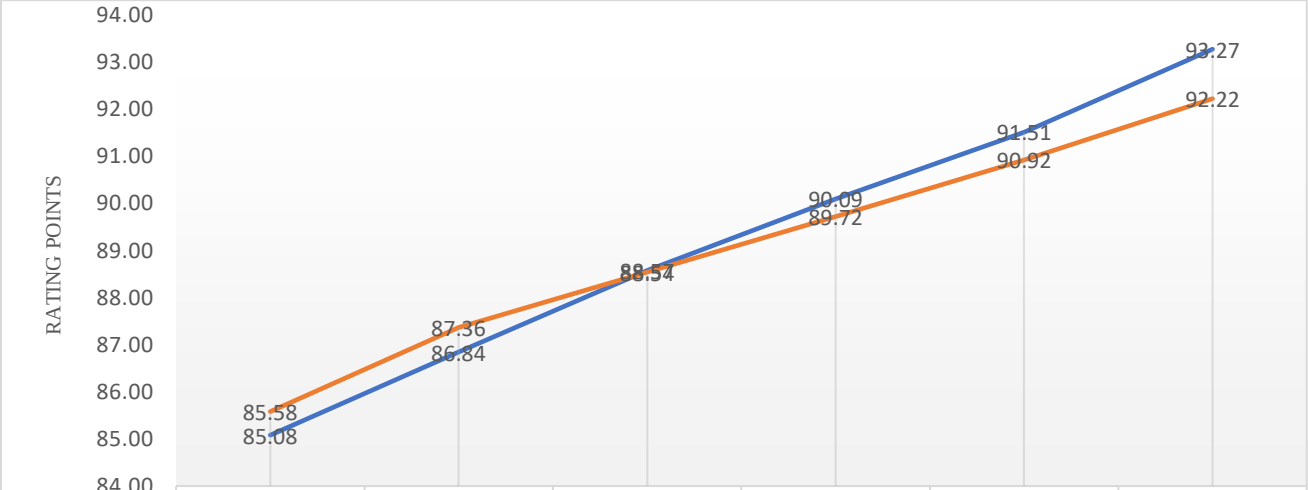


Figure 3: Average Number of Points by Price Category and Reviewer's Gender



	<10 US $	10-20 US $	20-30 US $	30-50 US $	50-100 US $	>100 US $
Men	85.08	86.84	88.57	90.09	91.51	93.27
Female	85.58	87.36	88.54	89.72	90.92	92.22