B A C H E L O R A R B E I T

A Vignette Study on Employer Bias and Wage Gaps

vorgelegt von

Christopher William Gardiner

07.01.2019

Fakultät für Wirtschafts- und Sozialwissenschaften

Fachbereich Sozialökonomie

Studiengang: Sozialökonomie

Matrikelnummer: 6772165

Erstprüferin: Prof. Dr. Miriam Beblo

Zweitprüfer: Dr. Tobias Langenberg

Talbachweg 4b
67550 Worms
Mobil 01525 2019282
E-Mail christopher@gardiner.de

Abstract

Extensive labour market research has established that homosexual men have lower average wages than their comparable heterosexual counterparts, while for homosexual women a wage premium compared to similar heterosexual females has been shown. Standardised surveys are useful for determining which extent of these inequalities might stem from employer bias. As part of his master's thesis "Stereotypical Prejudices and Labor Market Behavior - A Case Study for the LGBT Community" (2018), Jens T. Möller has conducted a survey in the form of a vignette study and has identified a significant causal relationship between said employer bias and wage gaps. Since standardised surveys are prone to response errors such as questionnaire context effects, their results might be distorted. By the example of Möller's study, the present thesis examines the relationship between context effects and surveys investigating underlying prejudices against people of different gender and sexual orientation and examines whether context effects need to be adjusted for in the course of such studies. This is done by rerunning the vignette study, having removed potential context effect sources from its questionnaire, analysing the data analogously and comparing the results. The updated questionnaire produces opposing results which suggest that Möller's survey design has facilitated distorting effects on his data. It is shown that context effects play such a significant role that adjusting for them is a necessary measure when conducting a survey in similar studies.

Contents

1. Introduction

Extensive research on wage gaps lays the basis for combating inequity among similarly competent people in the work environment (Antecol, Jong, & Steinberger, 2008, p. 518). For instance, the examination of differences between the earnings of women and men has led to the initiation of anti-discrimination laws and work-family policies (Christofides, Polycarpou, & Vrachimis, 2013, pp. 86-87). Some of these policies have shown to contribute to narrowing down the pay differences between men and women (Christofides, Polycarpou, & Vrachimis, 2013, p. 100; Blau & Kahn, 2017, p. 849). The wage gap between heterosexual people and people of other sexualities got increasing attention from researchers in the course of the last three decades (Klawitter, 2015, p. 4). This has caused the implementation of similar regulations, such as the "Employment Equality (Sexual Orientation) Regulations" in the United Kingdom (The National Archives, 2003). For the "sexual orientation wage gap" (SWG) the research so far has been the most comprehensive concerning homosexual employees (Mize, 2016, p. 1133). While homosexual men were found to earn less than heterosexual men, homosexual women experience a wage premium compared to their heterosexual counterparts (Klawitter, 2015, p. 13; Mize, 2016, p. 1134).

There are various approaches in wage gap research for identifying labour discrimination (Klawitter, 2015, p. 21). The disadvantage of the use of real-life data in the measuring of the wage gap is that it cannot be determined how much of it exactly is caused by discrimination. There are many unobservable factors concerning employees, that figure into their earnings and it is virtually impossible to control for all of them (Christofides, Polycarpou, & Vrachimis, 2013, p. 100). Another method for identifying labour discrimination is to take the affected group out of the equation by examining the decision makers' judgment only, with the help of a survey. In the course of his master's thesis "Stereotypical Prejudices and Labor Market Behavior - A Case Study for the LGBT Community" (2018) Jens T. Möller chooses such an approach. The thesis includes a vignette study which deals with possible stereotypical prejudices of personnel managers and how these biases influence the wages of bisexual and homosexual personnel. For the homosexual groups, his study showed compelling results consistent with the findings of most previous researches. Any employee related factors can be ruled out for having distorted his results. However, the use of a survey has the downside of possible response errors altering its findings (Sudman & Bradburn, 1974, p. 1).

In written questionnaires like Möller's, adjacent items may influence each other's answers; a phenomenon called "context effect" (Converse & Presser, 1988, pp. 39-40). This thesis addresses the relationship between such context effects and surveys investigating underlying prejudices against people of different gender and sexual orientation, by the example of Möller's vignette study. Its aim is to examine whether context effects play such a significant role that they should be adjusted for in the course of such studies. This is done by repeating Möller's vignette study, having removed context effect sources from its questionnaire. Differences in results, shown in the subsequent comparison are assumed to stem from context effects.

My findings suggest that the answers of participants in the vignette study are sensitive to context effects. They further imply that the estimated bias of the participants in Möller's study and the resulting wage gaps are exaggerated to some extent. However, my results also show tendencies that cannot be explained by the removal of context effects. This indicates that there are unobserved factors contributing to the detected differences in results, and that my findings should be interpreted cautiously.

The beginning of my thesis includes a brief overview of existing research on the sexual orientation wage gap and how Jens T. Möller's study is connected to these findings. I then proceed with a review of literature on context effects and in what way they are associated with questionnaire design. The following step involves the application of my findings on context effects and questionnaire design to Möller's questionnaire. After investigating possible causes of context effects in his survey and their potential impact on his results, I establish the hypotheses for my research. Thereafter, the questionnaire design is described in detail and the measures taken to remove sources of context effects are explained. Subsequently, the findings of the data analysis are described. The analysis contains of four parts: an investigation of the evaluation measures' relationship, a comparison of means of evaluated characteristics, an ordinary least squares (OLS) regression analysis for determining wage relevant characteristics and a second OLS regression analysis for an estimation of wage gaps.

2. The sexual orientation wage gap

This section gives an overview of so far conducted research on the "sexual orientation wage gap" (SWG) between homosexual and heterosexual people. By presenting figures from present studies I provide an impression of the estimated extent of these wage gaps. Furthermore, possible causes for the wage differentials are presented. The

impact of discrimination based on stereotypical prejudices, on the wage gaps is assessed, according to the present literature. At the end of this chapter, the examined literature is connected to the results of Jens T. Möller's (2018) vignette study.

The SWG classifies the percental difference between the average gross per hour earnings between employees with differing sexual orientation. For studies on wage gaps in general, it is common practice to control for measurable factors like job choice or education. By doing this, it is attempted to identify the extent of possible labour discrimination (Graf, Brown, & Patten, 2018). After controlling for observable factors, most researchers found that homosexual men earn less than their comparable heterosexual counterparts, and for homosexual women, most studies show a wage premium compared to similar heterosexual females (Klawitter, 2015, p. 21; Mize, 2016, p. 1134). According to a meta-analysis of 31 studies by Klawitter (2015) the average estimated SWG among men is 11% and among women 9%. These findings reveal that there are non-observed factors that play a significant role in forming differences in pay. Researchers suggest that discrimination cannot be ruled out as a crucial factor for the SWG (Mize, 2016, p. 1152; Badgett, 1995, p. 737). Labour discrimination may be based on stereotypical prejudices against the individual groups. "Masculine" attributes are reported to be more associated with competence than "feminine" attributes. The wage related disadvantage of certain groups is suspected to stem from their more feminine reputation (Mize, 2016, p. 1134). However, labour discrimination alone should not be made responsible for the extent of the "adjusted" wage gaps. There are other potential contributors that are not controlled for in studies, due to difficulties of observation or quantification. For example, differing domestic arrangements of homosexuals compared to heterosexuals were found to have a possible impact on the SWG but are not taken into account in every study (Mize, 2016, p. 1132). Another potential, but often neglected contributor is varying work intensity (Klawitter, 2015, p. 4) .

In the course of his study, Möller attempted to identify wage gaps that are exclusively dependent on masculinity related stereotypical prejudices. He established hypotheses on wage gaps and gender stereotypes that are consistent with the findings of the above quoted literature: According to his hypotheses, heterosexual men are perceived as more masculine than heterosexual women, heterosexual men are attributed with a higher masculinity than homosexual men and homosexual women are perceived as more masculine as heterosexual women. Furthermore, perceived masculinity has a positive impact on wage. Hence, homosexual women are given a higher wage than

heterosexual women and heterosexual men are given a higher wage than homosexual men (Möller, 2018, p. 47).[1]

Möller then investigated these hypotheses with the help of a survey: The participants of the survey are presented a text (vignette) on one of six different fictitious job applicants, giving information on work related criteria such as work experience and studies, as well as on gender and sexual orientation. The six people in question only differ in gender and sexual orientation. The participants are then asked to answer two questions regarding wage and suitability as well as a series of evaluation questions regarding different specific characteristics of the concerned job applicant. These characteristics are classified by Möller as stereotypically masculine (ibid., p. 52.). After comparing the mean wage of each vignette and investigating the relationship between wage and perceived masculinity, he finds that each of the above listed hypotheses is confirmed (ibid., p. 62-63). The SWG, identified by the author, lies between 5.44% and 13.54% for homosexual and heterosexual women; for homosexual and heterosexual men the SWG is estimated to lie between 5.26% and 8.15% (ibid., p. 62).

3. Context effects and questionnaire design

As opposed to working with real-life wage data, a survey like Möller's offers the advantage of excluding any unwanted, result distorting characteristics of the examined group and to focus on possible bias in the mind of decision makers. However, surveys are sensitive to various errors, depending on the mode of questioning (Voicu, 2015, p. 993). In the case of standardised written surveys, questionnaire design is a key element for possible response distorting effects (Krosnick & Presser, 2009). Preceding items might induce distorting "context effects" on an answer (Converse & Presser, 1988, pp. 39-40). Therefore, it is investigated if Möller's survey design could have facilitated distorting effects on the results of his study.

At the beginning of this chapter, I introduce two potentially context effect causing phenomena that are likely to occur depending on questionnaire design, namely "priming" and the "preference for consistency" (Voicu, 2015, p. 994; Caldini, 1984, pp. 44-45). I then investigate how these phenomena might manifest themselves in context of a questionnaire, depending on question sequence and question presentation. Finally, the evaluation section of Möller's questionnaire is described in detail. Subsequently,

---

[1] Möller's research also included SWG for bisexual people. For these groups his results were not as evident as for the others. Therefore, these differentials are not further addressed in my study.

Möller's method of question ordering and presentation is examined for possible causes of context effects. This is done in order to recognise in which way Möller's results, based on which he confirms his hypotheses, might have been influenced by context effects in his questionnaire.

3.1. Context effects

In the "Encyclopedia of Survey Research Methods" "priming" is defined as a "psychological process in which exposure to a stimulus activates a concept in memory that is then given increased weight in subsequent judgment tasks." (Parkin, 2008). In the context of a questionnaire, questions and answers represent information that by dealing with, is encoded in the participants' memory ("primed"), and can be accessed later in the questionnaire (Schwarz, Strack, & Mai, 1991, pp. 4-5). The application of primed information is considered a "priming effect". Priming effects are especially likely to occur among questions sharing the same topic and being of varying specificity (Schwarz, Strack, & Mai, 1991, p. 18).

Social psychologists consider another psychological phenomenon to have a similar impact on the answering behaviour in surveys: it is assumed that people generally prefer to be consistent or even feel pressure to be consistent in their actions and beliefs (Caldini, 1984, pp. 44-45). In questionnaires, respondents are generally tempted to follow a certain answering pattern, reinforcing opinions or evaluations given earlier on (Falk & Zimmermann, 2013, p. 190). Like the priming effect, this "preference for consistency" is more likely to have an effect in surveys with questions regarding the same issue. Unlike the priming effect, a preference for consistency can appear regardless of specificity of the questions (Tourangeau & Rasinski, 1988, p. 311).

3.2. Question order and question presentation

Various studies have analysed question ordering in standardised questionnaires and found that different question sequences have a significant influence on the participants' answers (Converse & Presser, 1988, p. 40; Sudman & Bradburn, 1974, p. 68; Tourangeau & Rasinski, 1988, p. 311). A question represents a reference point for participants, containing information about the subject of the question (e.g. "emotionality") and - if already answered - the participant's answer (e.g. "very emotional"). Therefore, question orders play an important role for priming (Crano & Brewer, 2002, pp. 279-280) The order of questions determines whether, and how recently a piece of

information has been placed in a respondent's memory. The elements directly preceding a question are the most recent source of information. In cases of questions of the same or a similar topic, information provided by already answered questions is more likely to be considered relevant by the participant - and more likely to be applied for following tasks, leading to a priming effect. Particularly the combination of specific and more general questions of the same topic increases the likelihood of priming effects (Schwarz, Strack, & Mai, 1991, p. 18). As they may be interpreted in several different ways, general elements are more susceptible to influence from specific ones than the other way around. Thus, a question that addresses a topic in general is likely to be subject of a priming effect by preceding questions that concern a specific domain of this topic (Krosnick & Presser, 2009, p. 50). Separate studies conducted on this phenomenon drew the same conclusion: If a general question is preceded by a group of specific questions, participants tend compile their preceding evaluations in the general answer. Other important factors, that were not object of previous questions, seem to be rather neglected when assessing the final judgment task (Willits & Saltiel, 1995, p. 663; Tourangeau & Rasinski, 1988, pp. 301-302)[2]. These insights lead to the conclusion that it is favourable in regard to context effects, to place specific questions behind more general questions of the same topic (Krosnick & Presser, 2009, p. 50; Converse & Presser, 1988, p. 41).

While the sequence of questions determines which pieces of information are present in the participant's memory, question presentation affects the extent of their presence. The number of questions presented on the same page influences how accessible pertinent information from earlier questions is. The more questions are presented on the same page, the more information is visible and more directly accessible. This makes it more probable that information from these visible questions and answers is incorporated in subsequent evaluations. Hence, the more questions are presented on one page, the higher the probability of a priming effect should be. Moreover, a higher number of questions in view likely facilitates the consistent answering throughout a greater part of the questionnaire.

A study by Tourangeau, Couper and Conrad (2004) on the effect of question presentation on one vs. multiple screens has shown results that could be connected to both, a preference for consistency, and priming: Answers to multiple questions that were

---

[2] On the contrary, a single specific question that is followed up by a more general question, might cause the participant to actively disregard the specific domain in the general evaluation (Tourangeau & Rasinski, 1988, pp. 302-303) (Schwarz, Strack, & Mai, 1991, p. 18).

presented on a single screen showed a higher intercorrelation than answers to the same questions being presented on separate screens. The authors suggest that participants tend to see a stronger relationship between items that are visually grouped than there actually is (Tourangeau, Couper, & Conrad, 2004, p. 391). Another possible explanation is that participants spend more time on questions that are presented separately. This might produce a more thought through answer and a reduction of context effects (Bishop, Hippler, & Schwarz, 1987, p. 321).

3.3. Context effects and Möller's (2018) vignette study

The main part of Jens Möller's questionnaire contains a series of judgment tasks concerning the characteristics of the fictitious job applicant introduced in the vignette. Every evaluation is to be made on a 5-point Likert scale. The opening question is "How suitable is (the candidate) for the job of a project manager in your opinion?"[3]. The Likert scale starts with "Not suitable" (value: 1) and ends with "Very suitable" (value: 5). The following five questions concern stereotypical "masculine" traits, higher values on the Likert scale indicating a higher masculinity. The first one of these asks for an evaluation of the candidate's risk appetite, ranking from "Not at all risk loving" (1) to "Very risk loving" (5). It is followed up by a question concerning the candidate's behaviour in negotiations. The question contains three levels the participant is asked to evaluate using the respective scale: A judgment of the candidate's behaviour is to be made on a scale from "Submissive" to "Dominant", from "Not aggressive" to "Very aggressive" and from "No confident appearance" to "Very confident appearance". Subsequently, the participant is asked about the behaviour of the participant in an "extremely stressful" situation. The question has two specifications: the first one contains a scale with the lowest value being "Very emotional" and the highest value "Not emotional" and in the second one the respondent is given a scale from "Does not hide emotions" to "Hides emotions". The question is followed up by an evaluation task concerning competitiveness of the job candidate, with a scale ranging from "Does not like competition" to "Likes competition". The last masculinity related evaluation in Möller's questionnaire regards assertiveness of the applicant in their role as a project manager, ranging from "Not assertive" to "Very assertive". The evaluation part of Möller's survey is concluded by an estimate of the monthly pay the person deserves to receive, on a range from 3000€ to 6000€ (Möller, 2018, pp. 52-53). Apart from education and work experience, the author

---

[3] All questions and answer options were translated from German to English.

consciously provided only little information in the vignette, based on which an objective evaluation of each of these characteristics could be made (Möller, 2018, pp. 51-52). The evaluation part of the questionnaire is presented on two pages. The first page includes a question regarding overall suitability followed by an evaluation of risk appetite. The second page includes all the remaining evaluation questions with the proposed wage question at the end (Möller, 2018, appendix A 5).

All the evaluation questions regard the same overall topic, which is a contributing factor for context effects. Apart from the questions regarding overall suitability and wage, all the questions concern specific characteristics of the person, which is why question order is important. The fact that the wage question is placed at the end, and therefore is preceded by a group of more specific questions could contribute to priming effects: participants could give increased weight to previously evaluated masculinity related factors in their wage proposal. Furthermore, the masculinity related evaluation questions vary in specificity among each other. Some more specific questions precede more general questions. This may cause priming effects on masculinity related questions that are more general. Figure 1 presents the question sequence and visualises the specificity of each question:

Figure 1: Evaluation question sequence in Möller's (2018) questionnaire

The colour of the respective question indicates the rank in specificity: yellow signifies that the question addresses the topic in general, higher specificity is indicated by an increasingly dark colour (orange –> red). The questions have been ranked and coloured based on following considerations: Deserved wage and suitability (both yellow) are supposedly the most general traits for a job applicant. Assertiveness, risk taking, and competitiveness (orange) are more specific traits figuring into overall suitability and deserved pay. These three characteristics are assumed to be on the same level of specificity. The final two questions address the job applicant's behaviour in specific situations: in a stressful situation and in a negotiation. A "stressful situation" (light red) may be interpreted in more different ways than a "negotiation" (dark red). Furthermore, the question regarding the stressful situation asks for emotions in general, while the question regarding the negotiation includes a specific emotion, namely "aggressiveness". Therefore, the question concerning behaviour in a stressful situation is assumed to be more general than the question concerning behaviour in a negotiation. The image is to be interpreted as follows: every question that is preceded by one or multiple questions with a darker colour is assumed to be especially prone to priming effects from these questions.

It stands to reason that the specific traits and deserved pay are difficult to judge having read only the vignette. This makes it more likely for the participant to resort on information from previous questions, which in the case of most of the evaluation part are grouped together on one page and especially easily accessible. The grouping of the questions in Möller's questionnaire can also promote a consistency effect. Moreover, the scaled answers create a visible pattern, making it easier to follow through with consistent replies. Although requiring a written answer, the wage related question is accompanied by a scale, visualising the range of the possible pay from 3000€ to 6000€. Because this form of presentation emphasizes a relationship of the items it could increase the probability of the consistency effect also being carried over to the final wage question (Tourangeau & Rasinski, 1988, p. 301).

4. Hypotheses

The insights from the preceding chapter lead to the conclusion that the evaluation questions in Möller's questionnaire may be object of context effects distorting his results. Masculinity traits may have been given a disproportionate amount of importance for the wage evaluation, resulting in an overestimation of the relationship between perceived masculinity and proposed pay. Consequently, the detected wage gaps may be caused or exaggerated by context effects. Moreover, the relationship of all the evaluation items is possibly overestimated through context effects. I predict that with context effects removed, Möller's results are not replicated and should deviate as follows:

Hypothesis 1: *All the evaluation measures show a weaker relationship among each other.*

Hypothesis 2: *The relationship between the group of masculinity traits and proposed wage is weaker.*

Hypothesis 3: *Wage gaps are smaller or not identifiable.*

5. Empirical strategy

In order to check my hypotheses, a second vignette study on stereotypical prejudices and wage gaps is conducted. As of content, the vignette study is virtually identical to Möller's. However, changes are made to question order and question presentation in the evaluation part, aiming at a minimisation of context effects. This lays the basis for checking my hypotheses by comparing the results to Möller's findings.

At the beginning of this chapter I elaborate the general questionnaire design of my vignette study. I then point out the specific adjustments made compared to Möller's survey, concerning context effects, and explain how they should help to check my hypotheses. The following chapter deals with the process of the data collection. Thereafter, the composition of the sample is described.

5.1. Questionnaire design

An online questionnaire was programmed in order to collect the data, with the help of the survey provider *SosciSurvey* (Leiner, 2018)[4]. Four versions of the questionnaire were prepared, a female heterosexual (henceforth referred to as FHE), a female homosexual (FHO), a male heterosexual (MHE) and a male homosexual version (MHO). The language of the questionnaire is German. As of content and overall structure, the survey does not differ from Möller's survey: The main part of the survey includes a vignette describing one of four fictitious job applicants at the beginning, followed by evaluation questions about the person, and demographic questions at the end. The vignette contains the same information, all the evaluation questions were kept, and no evaluation question was added. The mode of answering, using a text box for the wage evaluation and 5-point Likert scale for every other evaluation, was replicated. A few general amendments were made: Firstly, the introduction was worded in a more gender-neutral way. Some questions and instructions were rephrased for reasons of comprehensibility. Furthermore, instructions were adjusted to the change from a physical to an online survey. For instance, participants are asked to click "continue" to proceed to the next page. A manipulation check including the same questions as in Möller's survey was programmed to ensure that the vignette has been read and understood, before moving on to the evaluation part. The questions ask for the year of the applicant's master's degree and the number of years of their work experience. In case of an incorrect answer to the comprehension questions the programme would show the respondent a new page with the instruction to read the vignette again thoroughly and then direct them back to the vignette page. The final part of the survey includes demographic questions, addressed at the respondent personally. The first two questions concern gender ("female", "male") and age (in years) were adopted from Möller's 2018 questionnaire. Two demographic questions were added to Möller's questions. The first additional question addresses the highest achieved level of education and has the possible answers "No Graduation", "Basic Degree of Secondary Education", "Intermediate Degree of Secondary Education", "General Qualification for University Entrance", "Bachelor's Degree", "Master's Degree", "Doctorate" and "Other; please specify: …". It is followed up by a question concerning the participant's marital status which gives the options "single", "married", "divorced" and "widowed". Both questions were considered appropriate for controlling for in the process of data analysis and providing further

---

[4] The questionnaire is presented in Appedix A.

helpful information for the interpretation of results. Furthermore, the intermediate option "Suburbs" was added to the options of the question addressing the respondent's place of residence ("City" and "Country"). This was done to provide an appropriate answering option for the respondents living in areas that are neither clearly urban nor rural.

To minimise possible context effects the questionnaire design was adjusted. This includes a re-ordering of the questions in the evaluation part, because the original question-sequence does not take specificity of the questions into account. In the questionnaire of the follow up study questions are ordered by ascending specificity, illustrated in figure 2:

Figure 2: Evaluation question sequence in the present study's questionnaire
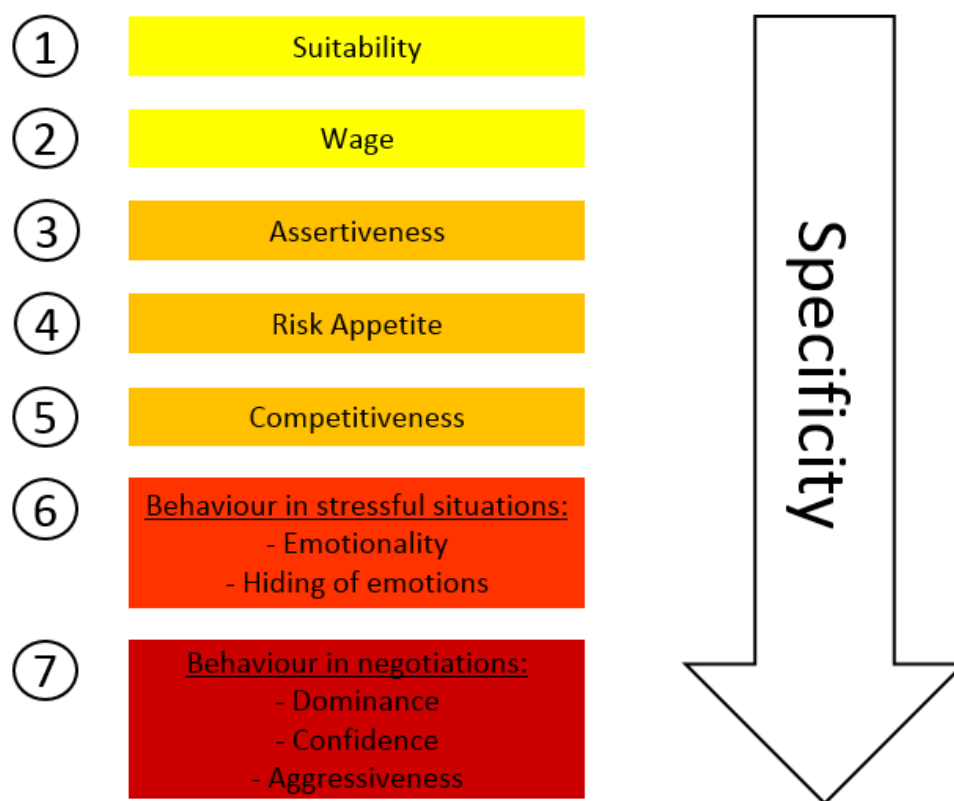
The two questions that are assumed to be the most general, namely the questions regarding suitability and wage, are put in front of the questionnaire. By this, any point of reference that is related to specific, supposedly masculine characteristics is removed when evaluating the wage. They are followed up by the more specific questions regarding assertiveness, risk appetite and competitiveness. As they are assumed to be on the same level of specificity, these three questions are in no particular order. The next question addresses the job applicant's behaviour in a "stressful situation".

The supposedly most specific question, regarding the person's behaviour in a negotiation, is located at the end. Based on these considerations, the new question order is less susceptible for priming effects because participants are not able to summarise specific evaluations into a more general one.

In the present questionnaire, each question is given a separate page and participants are only able to see the following page after having given an answer and clicking "continue". "Relevant" information in the form of previous questions and answers is now less accessible, not being in the field of sight. Specific information, placed further back, is not accessible to the respondent at all. This further reduces the probability of priming effects. Moreover, the respondents should spend more time and be more focused on each question independently, paying less attention to previously given answers, which decreases the likelihood of a consistency effect.

Based on the insights from chapter 2, sources of context effects caused by unfavourable aspects of the question order and question presentation in Möller's survey are removed whilst keeping the essential nature of the survey. Any deviation of results may be linked to context effects in Möller's questionnaire.

5.2. Sampling process

The online questionnaire was put online on the first of November 2018. Thereafter it could be retrieved on www.soscisurvey.com. When following the hyperlink, the participant was randomly assigned one of the four questionnaires. On the 30th of November 2018 a sample size of 144 had been reached and the data was downloaded from the survey-provider's server. 5 completed questionnaires were removed from the data set due to a spent time period of below 20 seconds on the survey page featuring the vignette. Such a short reading time implies that the text is "skimmed" rather than read thoroughly and thus, is likely not being read for the first time. Due to the automatic random assignment of the different questionnaires it cannot be ruled out that a respondent read the vignette, for some reason quit the survey and later reopened the link, then being assigned a different vignette. The participant could have noticed the change of the concerned fictitious person in the vignette and realise the nature of the study, possibly leading to biased answers. In the case of them not having noticed the change of the concerned fictitious person they would have proceeded to answer the questions, with the wrong vignette in mind. In both cases, the participant's answers would have distorted the overall results.

5.3. Sample

A total of 139 participants were included in the sample. The average age of the participants is 35.83 years, with women being on average 33.94 years old and men being on average 37.79 years old. The sample contains of 69.1% of participants living in the city, 18.0% of participants living in the suburbs and 12.9% of participants residing in the countryside. Measured by the demographic parameters that my study and Möller's study have in common, both are relatively similar which should facilitate a comparison of results.[5]

1.4 % of participants have reached a basic degree of secondary education (Haupt-/ Volksschulabschluss) as their highest degree of education. An intermediate degree of secondary education (mittlere Reife) is the highest degree for 11.5% of the participants while 23% have lastly completed secondary school with a general qualification for university entrance ((Fach-)Abitur). 60.4% of participants have reached a university degree (Bachelor's/ Master's/ Diploma/ Staatsexamen) and 3.6% have been awarded a doctorate. As the sample contains a high percentage of university graduates, and a low percentage of people with a basic degree it is not representative of the distribution of education levels in the German population (Statistisches Bundesamt, 2013, p. 76).[6]

65.5 % of people who have completed the survey are single, 28.8% are married and 5.8% are divorced. None of the participants stated to be widowed. In the category "marital status" the sample does not reflect the distribution in the German population either, as the single population in Germany is smaller and the married population in Germany is greater (Statistisches Bundesamt, 2013, p. 33).

Concomitant with the composition of the overall sample, it is equally important to know the distribution of participants across the individual vignettes. 33 participants completed the FHE vignette's questionnaire. 39 people participated in the MHE vignette study. Questions for the FHO vignette were answered by 35 people. The MHO vignette study was completed by 32 participants. Tables 1 – 4 depict the distribution of respondents amongst each of the four questionnaires regarding the five demographic parameters gender, mean age, place of residence, education and marital status. The differences in mean age are considered neglectable.

---

[5] It is assumed that the percentages for his sample excluding the bisexual vignettes is roughly the same as the ones for his whole sample.

[6] However, it is debatable whether a high percentage of highly educated participants is unsuitable for the purpose of this research. Möller's assumptions account for decisions of human resource managers (Möller, 2018, p. 47). Typically, the position of a human resources manager requires a higher education, in the form of a university degree (Human Resources EDU, 2018).

The FHO vignette shows a slight overbalance of female respondents (ca. 57%) as the MHO vignette does for male respondents (ca.56%) (table 1).

Table 1: Sample distribution among the individual vignettes in terms of mean age and gender

Vignette		Gender		
	Mean age	Female	Male	Total
FHE	38,12 years	51,52%	48,48%	100,00%
FHO	35,63 years	57,14%	42,86%	100,00%
MHE	34,77 years	51,28%	48,72%	100,00%
MHO	34,97 years	43,75%	56,25%	100,00%

The MHO vignette contains of a relatively low percentage of urban residents (less than 60 %). The MHE vignette has an especially high proportion of people living in the city (almost 80%) and a low proportion of people living in the countryside: just under 8% of participants of this vignette are rural residents. This percentage is considerably lower than for the other vignettes (table 2).

Table 2: Sample distribution among the individual vignettes in terms of residence

Vignette		Residence		
	Countryside	Suburbs	City	Total
FHE	15,15%	15,15%	69,70%	100,00%
FHO	14,29%	20,00%	65,71%	100,00%
MHE	7,69%	12,82%	79,49%	100,00%
MHO	15,63%	25,00%	59,38%	100,00%

The sample of the MHO vignette is a bit more educated on average, as the share of university graduates is about ten percentage points higher than in the other vignettes (table 3).

Table 3: Sample distribution among the individual vignettes in terms of education

Vignette	Education					
	Basic	Intermediate	Uni quali	Uni degree	Doctorate	Total
FHE	3,03%	18,18%	18,18%	57,58%	3,03%	100,00%
FHO	2,86%	11,43%	25,71%	57,14%	2,86%	100,00%
MHE	0,00%	10,26%	25,64%	58,97%	5,13%	100,00%
MHO	0,00%	6,25%	21,88%	68,75%	3,13%	100,00%

In the MHE vignette there is an especially high proportion of divorced participants compared to the other vignettes, with just short of 13% versus values around 3% (table 4).

Table 4: Sample distribution among the individual vignettes in terms of marital status

Vignette	Marital status			
	Single	Married	Divorced	Total
FHE	66,67%	30,30%	3,03%	100,00%
FHO	65,71%	31,43%	2,86%	100,00%
MHE	61,54%	25,64%	12,82%	100,00%
MHO	68,75%	28,13%	3,13%	100,00%

The present sample is assumed to be relatively well-balanced overall. However, it does show slight inequalities, that are to be considered when interpreting the results. Inequalities in education- and marital status distribution in the sample, as well as imbalances among the different vignettes' samples give reason for a cautious interpretation of the comparison of means in the following chapter.

6. Data analysis

The analysis is performed with *SPSS Statistics* (IBM Corp, 2017). The data is analysed in four steps; in the last three, Möller's analysis is replicated (see Möller, 2018, p. 55). Each analysis is concluded with an evaluation whether my respective assumption was met.

The first step involves the examination of the relationship of all evaluation measures with the help of a correlation test and a test of internal consistency. This is done to

assess whether my attempt to reduce context effects was effective. This should lay the foundation for a comparison of the results. It is predicted that the evaluation measures show a relatively low intercorrelation and internal consistency. Secondly, a comparison of means is conducted to give a first impression of the data. Using a t-test, it is investigated whether there are significant differences in perceived masculinity between the concerned groups. In the course of this analysis, Möller confirmed all assumptions regarding differences in perceived masculinity, meaning that the first requirement for his main hypothesis was met. To be in line with my hypotheses, these differences are expected to be smaller in the present data due to a weaker relationship among the evaluation measures. The third step involves an OLS regression investigating the relationship between perceived masculinity and wage. Möller found a significant positive relationship between perceived masculinity and suggested pay, and met the second requirement of his main hypothesis, ultimately identifying masculinity related prejudices as a cause of wage gaps. My hypotheses suggest that perceived masculinity will not be identified as such a strong predictor for suggested pay in the present data, because context effects on the final wage questions were minimised in my questionnaire. The fourth and final step of the analysis contains of a linear regression of both genders separately, determining the extent of these prejudice related wage gaps between heterosexual and homosexual vignettes. In the course of this analysis, Möller confirmed his main assumption of there being significant wage gaps in all investigated comparisons, that are mainly caused by masculinity related prejudices. As a consequence of the previous two assumptions, I expect there to be no significant wage gaps, or wage gaps of a smaller extent for my data.

6.1. Investigation of the evaluation measures' relationship

Now it is investigated how strongly the evaluation measures are connected to each other in the present data. In other words, it is estimated how independently from each other the questions were assessed by the participant. This is done in order to give an impression of whether the reordering and separation of questions was effective in terms of context effect removal. Initially, the Pearson correlation coefficient (PCC) for every combination of measures is looked at. If context effects were successfully reduced, the PCC values should be relatively low for most relations.

Table 5 depicts the results of the correlation test.

Table 5: Correlation matrix for all evaluation measures in the questionnaire

Variable	Wage	Suit.	Assert.	Risk	Compet.	Emot.	EmotHid	Dom.	Aggress.	Conf.
Wage	1,000									
Suit.	0,320***	1,000								
Assert.	0,262***	0,312***	1,000							
Risk	0,129**	0,028	0,229***	1,000						
Compet.	0,075	0,044	0,121*	0,386***	1,000					
Emot.	0,024	0,031	0,217***	-0,060	0,076	1,000				
EmotHid	-0,086	0,022	0,061	-0,064	-0,025	0,304***	1,000			
Dom.	-0,031	0,066	0,237***	-0,017	0,132*	0,292***	0,194**	1,000		
Aggress.	0,065	-0,129*	-0,008	-0,037	0,170**	0,037	-0,020	0,132*	1,000	
Conf.	0,146	0,223***	0,337***	0,205***	0,166**	0,361***	0,143**	0,429***	-0,020	1,000

*. The correlation is significant at the 10% level
**. The correlation is significant at the 10% level
***. The correlation is significant at the 10% level

It is shown that 19 out of 41 relationships are significantly correlated, all significant relationships being positive. Six correlations show a value of above 0.3, implying that there are six "moderate" linear positive relationships among the evaluation measures (Ratner, 2009, p. 139). Additionally, a test of internal reliability, measured by Cronbach's alpha (α) gives an impression of the consistency within the group of measures, and is related to the PCC (Tavakol & Dennik, 2011, p. 54). With little influence of context effects on the present data, it is also assumed to be relatively low. It has a value of 0.556, which would typically be classified as "inacceptable" in terms of reliability (ibid.)[7]. In the context of this research, the low values for the PCC and α support the assumption that the reordering and separating of questions reduced context effects in the questionnaire.[8]

Considering that overall suitability should be the most decisive factor for the wage decision it makes sense that the PCC is relatively high for this relationship. It is noticeable that the among the highest values are measures of characteristics that belong to one question and are thus presented on the same screen – confidence and dominance (PCC = 0.429) as well as emotionality and emotion-hiding (PCC = 0.304). It is furthermore striking that emotion-hiding otherwise only shows correlation values that are far from 0.3. Perhaps context effects were a contributing factor here; even if the weak

---

[7] The results of the statistical test are presented in Appendix B.
[8] Even though I have no reference values in terms of correlation and internal reliability for Möller's results it seems to confirm the assumption.

relationship between aggressiveness and dominance (PCC = 0.132) and the nearly non-existent relationship between aggressiveness and confidence (PCC = -0.020) seem to contradict this notion. Although the PCC and α are not given in Möller's study, this seems to provide first evidence for the plausibility of my first hypothesis.

6.2. Comparison of means

In the following it is investigated which average values were assigned each of the concerned groups by the participants of the survey, being measured on the five-point Likert-scale for the character traits and on a range from 3000€ to 6000€ for the wage proposal. This is done by comparing the means of the values assigned to the respective character traits.

Möller found that the participants of his study suggested the highest pay for the female homosexual person, followed by the male heterosexual job applicant, the male homosexual person, and lastly the female heterosexual candidate. For the all the remaining characteristics, the respective rank orders are very similar. In general, participants of his study have collectively ranked the four groups in only two different orders: either the female homosexual vignette or the male heterosexual vignette are in first place, always followed by the male homosexual vignette and then the female heterosexual vignette (Möller, 2018, p. 57). This pattern indicates that there is a relatively strong positive relationship between the variables in Möller's data. For the data of the present study the assumption is that the variables show a weaker relationship, meaning that they were evaluated more independently from another (Hypothesis 1). Therefore, the values should not follow such a strict pattern. Generally, higher and lower values should be more evenly distributed. Which of the vignettes have higher or lower specific values than others, may not be explained by my hypothesis.

Table 6 shows the means for every evaluation made for each vignette.

Table 6:Table of means for all evaluation measures and the newly created measure "masculinity"

Variable	Vignette			
	FHE	FHO	MHE	MHO
Suitability	4,33	4,54	4,38	4,34
Assertiveness	3,91	3,89	3,74	3,63
Risk taking	3,12	2,89	2,69	2,88
Competitiveness	3,61	3,43	3,05	3,31
Dominance	3,52	3,43	3,41	3,31
Aggressiveness	2,73	2,63	2,82	2,47
Confidence	3,91	3,80	3,74	3,72
Emotionality	3,18	3,34	3,51	3,41
Emotion hiding	3,42	2,97	3,31	3,00
Masculinity	3,42	3,30	3,29	3,21
Wage	4.583,33 €	4.534,29 €	4.547,10 €	4.415,63 €
N	33	35	39	32

Green numbers indicate the highest values among the vignettes and red numbers indicate the lowest values among the vignettes. The FHE vignette is attributed the highest pay, followed by the MHE vignette and the FHO vignette, by only a slim margin respectively. Male homosexuals are assigned the lowest wage on average (170€ less than heterosexual females). With regards to overall suitability, heterosexual women and men and homosexual men are evaluated nearly identically on average (less than 0.05-point difference). The FHO vignette leads this category by a 0.2-point margin on the Likert scale. The MHO vignette ranks last in three out of eight masculine categories. Taking the mean of the masculinity related evaluation measures has the MHO vignette come out at the bottom as well (these measures are displayed in the means table next to "Masculinity"). Based on these criteria, one might draw the conclusion that homosexual males are perceived as the least masculine of the four groups by the participants. These tendencies contradict Möller's results. The FHE vignette leads in the most as masculine classified characteristics (five out of eight) and shows the lowest mean across masculine categories. This may suggest that the female heterosexual person in the vignette is considered the most masculine by the participants. The general mean

values don't follow such a strict pattern as in Möller (2018): there are eight different rank orders for the ten categories. This indicates a weaker relationship between the individual categories, as predicted. It is curious why the FHE vignette in particular was evaluated so differently in both studies, as this cannot be explained by the removal of context effect sources.

In the following, a test of statistical significance determines differences that may be interpreted further: For a pairwise comparison of the means for every category of evaluation, an independent samples t-test with unequal variances is conducted for every combination of vignettes. In Möller's data, every comparison of means of the vignettes follow his main assumptions: heterosexual women are being perceived as less masculine than heterosexual men and homosexual women as well as homosexual men being perceived as less masculine than heterosexual men. Of these comparisons he found the great majority to be statistically significant and all differences pointing in the expected direction (e.g. homosexual men were given lower values for every trait than heterosexual men). Thus, he was able to confirm his hypotheses (Möller, 2018, p. 56). It follows from the previous test that a lower perceived relationship between the evaluation measures should lead to a more diverse distribution of high and low values by the participants of the present study. Therefore, fewer significant values for above mentioned comparisons should point in the same direction (Hypothesis 1). Table 7 shows the results of the t-tests for my data:

Table 7: T-test for selected group differences

Variable	Group Differences				
	MHO - MHE	FHO - FHE	MHE - FHE	MHO - FHE	FHO - MHE
Wage	-131,48 €	-49,05 €	-36,23 €	-167,71 €	-12,82 €
Suitability	-0,041	0,210	0,051	0,010	0,158
Assertiveness	-0,119	-0,023	-0,166	-0,284	0,142
Risk taking	0,183	-0,235	-0,429**	-0,246	0,193
Competitiveness	0,261**	-0,177	-0,555***	-0,294	0,377*
Emotionality	-0,107	0,161	0,331**	0,224	-0,170
Emotion hiding	-0,308	-0,453**	-0,117	-0,424**	-0,336*
Dominance	-0,098	-0,087	-0,105	-0,203	0,018
Aggressiiveness	-0,352**	-0,099	0,093	-0,259	-0,192
Confidence	-0,025	-0,109	-0,166	-0,190	0,056

*. The difference is significant at the 10% level
**. The difference is significant at the 10% level
***. The difference is significant at the 10% level

For most of the characteristics, the FHE vignette has got higher means than the FHO vignette. One relationship is significant: The FHE vignette got a significantly (5% level) higher mean score for the evaluation of the ability to conceal emotions in stressful situations than the FHO vignette. The comparisons between both male vignettes favour heterosexual men in most cases. Two of the comparisons reach statistical significance, one in favour of the MHE vignette and one in favour of the MHO vignette: The MHE vignette was attributed higher aggressiveness in negotiations than the MHO vignette (significant at the 5% level), while the MHO vignette got a significantly (5% level) higher mean for competitiveness. The comparison between the heterosexual vignettes reveals that in most of the characteristics the FHE vignette was given higher average values and two of these differences reach significance: The FHE vignette was perceived as more risk taking than the MHE vignette. According to the t-test this relation is significant at the 5% level. Heterosexual women are also perceived as more competitive than heterosexual males. This relation is classified as significant at the 1% level. The emotionality domain is the only one in which the MHE vignette was given a significantly (5% level) higher mean value than the FHE vignette, which is to be interpreted as heterosexual men being regarded as less emotional than heterosexual women. Three more relationships reach significance according the t-test: Homosexual women are perceived as being more competitive than heterosexual men. This relation shows a significance on the 10% level. Furthermore, heterosexual men are perceived as more emotion-concealing than homosexual women which is significant on the 10% level. Also, heterosexual females are assumed to be significantly (5% level) more emotion-hiding than homosexual males. No significant differences are shown for the wage category.

All in all, the t-test show that only eight comparisons reach statistical significance. Out of these eight significant relations only two are in line with Möller's masculinity related assumptions. Three comparisons contradict them, leading to the fact that some of the significant relationships contradict each other in terms of perceived masculinity by pointing in opposing directions (e.g. the FHE vignette was rated higher in terms of risk appetite, but lower in terms of emotionality than the MHE vignette). The conclusions that Möller drew in his study cannot be drawn from these results. At the same time, these findings indicate that the categories were viewed more independently from each other by the participants. This supports my first hypothesis. However, it is to be noted

that the results only account for this specific sample and the results may differ for sampling related reasons (Assael & Keon, 1982, p. 114). Naturally, the composition of the sample is not identical to Möller's, shown by the three demographic indicators the samples share. Also, the sample distribution among each vignette is not equal in this study.[9] As the t-test does not offer any mechanism to control for demographic factors, sampling errors cannot be ruled out as a contributing factor for the differences in results.

6.3. Determination of wage relevant factors

Possible sampling errors should be less problematic in the following procedure as the demographic indicators can now be controlled for. A linear regression is used to investigate the relationship between the masculinity factors and proposed earnings. At first, a linear multiple regression model is chosen for the whole sample of 139. Proposed wage is predicted for, the demographic variables and overall suitability are controlled for and the masculinity related characteristics act as predictor variables. As in Möller's study, the wage scale is logarithmised (Möller, 2018, p. 57). Since nominal variables with more than two categories raise problems for calculation and interpretation of results in a linear regression, three demographic variables are adapted before being included in the model (Lunt, 2013, p. 1142). The education variable(seven-categories) is dummy coded, as it is common practise for performing regression analyses (ibid.). All observations of an education level with no university degree are compiled into the value 0 being "Without a university degree" and all observation of university degrees were given the value 1 ("With a university degree"). The nominal variable regarding residence (three levels) is dummy-coded as well. The two resulting dummy variables are "Suburbs" - with the value 1 indicating that "Suburbs" has been chosen for an answer and the value 0 referring to any of the other two possible answers - and "Country" (1= Country, 0= any other). Following the usual approach, the most frequently chosen answering option, namely "City", is left out of the regression model, functioning as the reference category (ibid.). The variable concerning marital status is treated in the same way. The option "Single", being the reference category, is not contained in the model, whereas the resulting dummy variables "Married" and "Divorced" are included. The model is run six times, resulting in six different model specifications. The first, basic

---

[9] For lack of information on the sample distribution among Möller's vignettes, I assume that his sample is evenly distributed.

model specification includes the variables that are controlled for, the demographic variables and *Suitability*. The predictor variables are then successively in the same order as in Möller (2018, p 58). The sixth and final model specification looks as follows:

$$\ln(wage) = \beta_0 + \beta_1\,Gender + \beta_2\,Age + \beta_3\,Suburbs + \beta_4\,Country + \beta_5\,Education + \beta_6\,Married + \beta_7\,Divorced + \beta_8\,Suitability + \beta_9\,Risk + \beta_{10}\,Dominance + \beta_{11}\,Aggress + \beta_{12}\,Confidence + \beta_{13}\,Emotionality + \beta_{14}\,EmotHide + \beta_{15}\,Competitiveness + \beta_{16}\,Assertiveness + \varepsilon$$

Initially, the regression model is checked for the assumptions of multiple linear regression, in order to determine whether the regression model chosen can provide valid results (Williams, Gomez Grajales, & Kurkiewicz, 2013, p. 11)[10]. The independent variables should have a linear relationship with the dependent variable. The linearity assumption is checked with the help of a normal p-p plot and turns out to be met (Casson & Farmer, 2014, p. 593). The assumed absence of multicollinearity can be confirmed, as the PCC shows values well below 0.7 for every combination of the explaining variable (Williams, Gomez Grajales, & Kurkiewicz, 2013, pp. 10-11). The data is consistent with the requirement of the disturbance values not being autocorrelated, shown by a Durbin Watson value of close to 2 (Cliff & Ord, 1972, p. 267). The data should show a homoscedastic distribution of errors, and by observing the normal p-p plot this assumption is found to be met (Williams, Gomez Grajales, & Kurkiewicz, 2013, pp. 9-10). The assumption of normally distributed disturbance values is met, which can be seen when observing the normal q-q plot (ibid., p.10). Furthermore, the presence of potentially problematic outlying values can be ruled out by checking Cook's distance (ibid., p. 11). For Möller's model, the coefficient of determination R^2 showed a value of 0.3279, indicating that 32.79% of variances are predicted by the independent variables. Furthermore, Möller found four of eight relationships of masculinity related factors with the logarithmic wage to be positive and significant, three of which being significant on all six model specifications. Based on these results, Möller confirmed the hypothesis of a positive relationship between perceived stereotypical masculine factors and wage. Based on my hypotheses, the present data is expected to show fewer significant relationships between perceived masculinity and logarithmic wage, and therefore a lower predictive ability of the independent variables, indicated by the coefficient of

---

[10] The statistical tests regarding the assumption checks are presented in Appendix B.

determination (Moore, Notz, & Flingner, 2016, p. 138). The results of the regression analysis are depicted in table 8. For the final model specification, the R² value is 0.2475 indicating that 24.75% of the variances in the dependent variable are explained by the model. However, the adjusted R², which takes the number of predictors in the model into consideration, is only 0.1488 (ibid., p.140). It is shown that *Suitability* has a significant impact on the logarithmic wage on all six stages. According to the β-value of 0.0431, for every point added on the 5-point Likert Scale there is a 4.31% increase in proposed wage, ceteris paribus. *Risk* is the first predictor variable added to the basic model specification and turns out to be a significant predictor on the second (5% level) and third (10% level) stage. On the remaining three stages, *Risk* is not a significant predictor. Two of three negotiation-behaviour related variables, being added to the third stage, *Aggression* and *Confidence* show no significance; the exception being *Dominance*, which is significant in specification six (10% level). The variable reaches a β value of -0.0408. This implies that dominant behaviour in negotiations is penalised with a decrease of 4.08% in wage, per point added on the Likert-scale, all other factors considered constant. *Emotion* and *EmotHide* are not significant in any of the specifications, as well as *Compet*.

Table 8: Results of the regression analysis for the whole sample

Model / Variable	1 β	2 β	3 β	4 β	5 β	6 β
(Constant)	8,118*** (0,110)	8,065*** (0,112)	8,035*** (0,138)	8,063*** (0,151)	8,061*** (0,153)	8,014*** (0,151)
Suitability	0,061*** (0,018)	0,057*** (0,018)	0,056*** (0,018)	0,056*** (0,018)	0,056*** (0,018)	0,043** (0,019)
Gender	0,026 (0,027)	0,022 (0,026)	0,018 (0,026)	0,016 (0,027)	0,016 (0,027)	0,033 (0,028)
Age	-0,002 (0,001)	-0,003** (0,001)	-0,003** (0,001)	-0,003** (0,001)	-0,003** (0,001)	-0,004** (0,001)
Country	0,029 (0,041)	0,020 (0,040)	0,011 (0,041)	0,010 (0,041)	0,010 (0,042)	0,007 (0,041)
Suburbs	-0,039 (0,034)	-0,039 (0,034)	-0,043 (0,034)	-0,046 (0,034)	-0,046 (0,035)	-0,049 (0,034)
Education	0,029 (0,028)	0,035 (0,028)	0,036 (0,028)	0,034 (0,028)	0,035 (0,028)	0,039 (0,028)
Married	0,034 (0,037)	0,048 (0,037)	0,052 (0,037)	0,051 (0,037)	0,052 (0,037)	0,052 (0,037)
Divorced	0,081 (0,066)	0,095 (0,065)	0,089 (0,065)	0,091 (0,065)	0,092 (0,066)	0,092 (0,065)
Risk		0,032** (0,016)	0,028* (0,017)	0,028 (0,017)	0,027 (0,018)	0,020 (0,018)
Dominance			-0,034 (0,022)	-0,032 (0,022)	-0,033 (0,023)	-0,041* (0,022)
Aggress			0,024 (0,017)	0,022 (0,017)	0,022 (0,018)	0,020 (0,017)
Confidence			0,028 (0,020)	0,026 (0,021)	0,026 (0,021)	0,018 (0,021)
Emotionality				0,011 (0,020)	0,011 (0,020)	0,006 (0,019)
EmotHide				-0,018 (0,016)	-0,018 (0,016)	-0,016 (0,016)
Competitiveness					0,002 (0,017)	0,002 (0,017)
Assertiveness						0,048** (0,020)
N	139	139	139	139	139	139
R²	0,151	0,177	0,204	0,212	0,213	0,248
Adjusted R²	0,099	0,119	0,128	0,124	0,117	0,149

*. The relationship is significant at the 10% level
**. The relationship is significant at the 10% level
***. The relationship is significant at the 10% level
Standard errors in parentheses

The adjusted R² value decreases with the introduction of the emotion related variables on the fourth stage variable indicating perceived competitiveness on the fifth stage. This might indicate that not every variable has got a predictive ability for the logarithmic

wage individually (Hoffmann & Shafer, 2015, p. 92). *Assert* was added in the final stage of the model and reaches significance at the 5% level, with a β of 0.0479. Therefore, according to the model, every point added on the assertiveness-Likert-scale accounts for a 4.79% increase in suggested wage, ceteris paribus. Except for the basic specification in which there is no significance shown, the participants' age has a significant influence on the evaluation of wage in all model specifications, at the 5% level. With all other variables included in the model, every additional year of age is responsible for an estimated 0.36% decrease in proposed wage (β =0.0036). The demographic factors *Education* and *Gender* show no significant influence. The dummy variables *Married* and *Divorced* show no significant difference to the reference category of single participants. The β value is positive for both variables, indicating that there also likely is no significant difference between the two of them. Therefore, the marital status of the respondents is assumed to have no influence in general. The dummy variables *Suburbs* and *Country* show no significant influence compared to the reference category representing participants in the city.[11]

All in all, significant relationships are determined for five variables, three of which are masculinity related (*Assertiveness*, *Risk and Dominance)*. The variable *Risk* shows no significance in the final model, so its influence is questionable. The impact on proposed wage that is calculated for perceived dominance is negative, contradicting Möller's hypotheses. Only perceived assertiveness follows the assumption to have a significant positive impact on the suggested wage. Moreover, my model shows a considerably lower R² value (0.248) than Möller's does (0.3279) for the final specification, which indicates a lower explanatory ability of my predictor variables. R² values of below 0.3 are classified as indicating a non-existent none or a very weak influence (Moore, Notz, & Flingner, 2016, p. 138). Consequently, the results of the regression do not give reason to assume that the traits classified as stereotypical masculine have a positive impact on wage. They therefore support my second hypothesis that with context effect sources removed from the questionnaire, the identified relationship is weaker, if not gone.

---

[11] There might be a significant difference between the wage decision of participants from the suburbs and participants from the country but this relationship plays no role for my hypotheses and is therefore not further investigated.

6.4. Estimation of wage gaps

The following step has the objective to determine the extent of possible wage gaps among both genders depending on sexual orientation. For this, the sample is split into male and female vignettes and on both samples two linear regression analyses are performed. Based on a comparison of the results of both of these regression models, Möller determined a range in which the real wage gaps should lie (Möller, 2018, p. 62). The first one used on the present data is a linear regression with a newly created dummy variable indicating the vignette's sexual orientation (*Homosexual*), as the only explanatory variable for the logarithmic wage. This variable is assigned the value 0 for heterosexual vignettes and 1 for homosexual vignettes. The second regression analysis includes *Homosexual* while accounting for relevant masculinity related factors. Only masculinity related coefficients that show significant impact on the wage in the final specification of the previous model are considered relevant, namely *Dominance* and *Assert.* Additionally, the model includes interaction terms that investigate an influence of *Gender* in combination with each of the predictor variables. *Suitability* and the demographic variables are being controlled for in the second model. The model looks as follows:

$$\ln(wage) = \beta_1\ Homosexual + \beta_2\ Gender + \beta_3\ Age + \beta_4\ Suburbs + \beta_5\ Country + \beta_6\ Education + \beta_7\ Married + \beta_8\ Divorced + \beta_9\ Suitability + \beta_{10}\ Dominance + \beta_{11}\ Assert + \beta_{12}GenSuit + \beta_{13}\ GenDom + \beta_{14}\ GenAssert + \varepsilon$$

As for the previous regression model, both samples are checked for the assumptions for multiple regressions. For both samples the model is consistent with all assumptions. For his analysis, Möller included the masculinity related variables regarding competitiveness, assertiveness and risk appetite, because they had shown a significant impact on the logarithmic wage in the earlier stage of the analysis (Möller, 2018, p. 61). He found a significant wage gap between homosexual and heterosexual men, indicating that homosexual men are attributed a 5.26% lower pay than their heterosexual counterparts, masculinity related factors concerned. The raw wage gap is estimated to be considerably higher (8.15%). After controlling for masculinity related factors, it was shown that homosexual women were attributed an estimated 5.44% higher wage than heterosexual women, also a significant difference on the 10% level. The raw wage gap lied at 13.54% (ibid., p. 62). According to my third hypothesis, wage gaps shown for the present data should be smaller and less significant. The results are depicted in table 9.

Table 9: Results for the regression analyses for the samples of female and male vignettes

Female Vignettes

Model / Variable	1 β	2 β
(Constant)	8,424*** (0,068)	7,913*** (0,208)
Homosexual	-0,010 (0,043)	-0,023 (0,039)
Gender		-0,005 (0,050)
Age		-0,002 (0,002)
Country		0,089 (0,065)
Suburbs		-0,019 (0,059)
Education		0,037 (0,043)
Married		0,017 (0,058)
Divorced		-0,030 (0,137)
Suitability		0,061* (0,032)
Dominance		0,001 (0,032)
Assertiveness		0,075** (0,032)
GenSuit		0,014 (0,062)
GenDom		0,056 (0,065)
GenAssert		-0,047 (0,064)
N	68	68
R²	0,001	0,384
Adjusted R²	-0,014	0,222

Male Vignettes

Model / Variabe	1 β	2 β
(Constant)	8,442*** (0,052)	8,575*** (0,201)
Homosexual	-0,030 (0,034)	-0,045 (0,036)
Gender		0,084** (0,037)
Age		-0,003 (0,002)
Country		0,001 (0,061)
Suburbs		0,003 (0,046)
Education		0,043 (0,039)
Married		0,047 (0,051)
Divorced		0,026 (0,083)
Suitability		0,001 (0,025)
Dominance		-0,068** (0,030)
Assertiveness		0,016 (0,030)
GenSuit		0,005 (0,048)
GenDom		0,078 (0,058)
GenAssert		0,021 (0,056)
N	71	71
R²	0,011	0,281
Adjusted R²	-0,003	0,102

*. The relationship is significant at the 10% level
**. The relationship is significant at the 10% level
***. The relationship is significant at the 10% level
Standard errors in parentheses

For the female vignette sample, the R² has a value of 0.384 while the adjusted R² has a value of 0.222. *Sexuality* has a negative β value (ranging from -0.097 to -0.228), meaning that homosexual females would earn less than heterosexual females, which opposes Möller's results. However, for both models this relationship not statistically significant. Suitability (10%) and Assertiveness (5%) show a significant positive

influence on the wage evaluation. For the male vignettes sample's model, the models shown an insignificant F-value which questions the overall predictive ability of the model (Hoffmann 2010 p. 79)[12]. The R² has a value of 0.281 and adjusted R² lies at 0.102. The negative impact of *Sexuality* (homosexual men earn between 2.98% and 4.52 % less than heterosexual men) is classified as insignificant. Within the model, *Gender* is a significant predictor for the logarithmic wage, at the 5% level: men suggest an estimated 8.35% higher wage for the male vignette than women do. *Dominance* is a significant predictor (5% level) with a β value of **-**0.068 meaning that this character trait is supposedly penalised in the wage decision, with a 6.8% lower wage per added point on the Likert scale, ceteris paribus.

This analysis does not confirm Möller's results, as the wage gaps determined by the model are insignificant. It is to be noted that the gap between amongst women of different sexualities points in the other direction, while the model for the male SWG do not reject the null-hypothesis and the predictors therefore cannot be attributed any explanatory value with certainty. These findings are also consistent with hypothesis 3.

7. Conclusion

In the course of this thesis I evaluated the accuracy of the results of Jens T. Möller's study on employer prejudices as a cause for the SWG. I derived from existing research on psychology and survey design that his questionnaire design increased the likelihood of context effects on the participants' answers. The question order in Möller's survey was suspected to cause priming effects: because some specific questions precede more general questions, the participants could have been influenced by information of the specific questions when answering the general ones. Especially the wage question, placed at the very back, was suspected to be influenced by the specific masculinity measures in the way that respondents overvalued the perceived masculinity for judging the appropriate wage of the candidate. Furthermore, the fact that most of the questions are presented on the same page increased the possibility of priming effects by further augmenting the accessibility of specific information. This circumstance has also increased the probability of consistent answers throughout the questionnaire promoted by the participants' general preference for consistency.

---

[12] The test is displayed in Appendix B.

The findings from a comparison of Möller's and my results appear to be consistent with my assumptions: A comparison of means revealed that the distribution of high and low values does not follow such a strong pattern as in the original survey. Therefore, participants seemed to have evaluated the characteristics more independently from each other. Möller's findings of a higher perceived masculinity of men compared to women, of homosexual women compared to heterosexual women and of heterosexual men compared to homosexual men was not confirmed. The following regression analysis revealed that the relationship between masculinity related evaluation measures and the proposed pay is considerably weaker for my data. As opposed to Möller, I didn't identify a significant impact of perceived masculinity on wage. The final regression analysis, accounting for significant predictors of the wage evaluation, showed no significant SWG for both genders. This also contradicts Möller's results.

I conclude from these findings that Möller's estimations of employer bias towards people depending on their gender and sexual orientation, and the resulting wage gaps are likely exaggerated. Especially the evidence for a distortion by context effects of the relationship between perceived masculinity and wage seems compelling. However, I cannot infer that employer bias does not cause wage differences between people of differing sexual orientation. The absence of statistical significance in the results of my regression analyses only determines that the given variables could not be identified as predictors with certainty. Furthermore, my results show features that cannot be explained by context effect removal alone. For instance, the FHE vignette seemed to be attributed with a slightly higher masculinity and wage compared to the other vignettes. Möller had completely opposing results. This circumstance suggests that there are other sources of deviation, in addition to context effects. One possible source could be sampling errors. According to the existing information on Möller's sample composition, our samples are relatively similar. However, the three demographic questions (gender, age and residence) shared by both surveys provide only limited information on the comparability of the samples. The small sample size is sensitive to statistical variations, possibly leading to two very "different" samples, in terms of certain unobserved characteristics. For example, the different survey methods (online survey vs street survey) could have contributed to such variations. Moreover, my sample shows some imbalances among the vignettes that could have affected my results. Particularly the results of the means comparison should be interpreted with caution, as it is not controlled for demographic factors in this analysis.

Regardless of the exact values that were shown by the statistical tests, my research demonstrated that the results of the vignette study are highly sensitive to external influences. It was shown that adjusting for context effects is a measure that should be taken when conducting a survey in similar studies.

Appendix

Appendix A: Questionnaire

Questionnaire: Introduction text

Umfrage

Im Rahmen meiner Bachelorarbeit führe ich eine Umfrage durch, die die Beurteilung von Bewerbungen durch die Personalabteilung thematisiert. Für die Beantwortung der Fragen werden Sie nun gebeten, sich in die Personalleitung hinein zu versetzen. Die gesamte Umfrage dauert 5-10 Minuten und erfolgt anonym.

Bitte verwenden Sie **nicht** die Pfeiltasten Ihres Browsers/ Smartphones, um zwischen den Seiten zu navigieren, da dies zum Abbruch der Umfrage führt.

Questionnaire: FHO vignette text with instructions

Stellen Sie sich bitte vor Sie sind Personalmanager(in) eines großen Unternehmens mit Sitz in Hamburg. Für den Aufbau und die Entwicklung einer vollkommen neuen Produktpalette benötigen Sie einen Projektmanager (m/w), der alle Prozesse steuern und auch Verhandlungen führen soll. Auf die Stellenausschreibung haben sich mehrere geeignete Personen beworben. Nachfolgend finden Sie eine Kurzbeschreibung einer dieser Personen. Bitte lesen Sie sich diese aufmerksam durch und beurteilen Sie anschließend die Eignung der Person für die Stelle des Projektmanagers.

Kurzbeschreibung:
Anna Müller ist 34 Jahre alt und lebt mit ihrer Lebensgefährtin Jana in Hamburg. Anna hat das Abitur auf einem Wirtschaftsgymnasium absolviert und darauffolgend Betriebswirtschaftslehre studiert. Sie hat 2008 ihren Bachelorabschluss und 2010 ihren Masterabschluss erlangt. Beide Abschlüsse wurden innerhalb der Regelstudienzeit und mit guten Noten absolviert. Während ihres Studiums hat sie als Werkstudentin bei einem namhaften Konkurrenten gearbeitet und konnte so erste Berufserfahrung sammeln. Kurz nach absolviertem Studium hat Anna dann eine erste feste Anstellung als Projektmanagerin gefunden und ununterbrochen in diesem Berufsfeld gearbeitet. Somit kommt sie auf 8 Jahre Berufserfahrung. Arbeitszeugnisse ehemaliger Arbeitgeber bescheinigen Annas Befähigung zur Durchführung der geplanten Produkterweiterung.

Questionnaire: FHE vignette text

Kurzbeschreibung:

Anna Müller ist 34 Jahre alt und lebt mit ihrem Lebensgefährten Jan in Hamburg. Anna hat das Abitur auf einem Wirtschaftsgymnasium absolviert und darauffolgend Betriebswirtschaftslehre studiert. Sie hat 2008 ihren Bachelorabschluss und 2010 ihren Masterabschluss erlangt. Beide Abschlüsse wurden innerhalb der Regelstudienzeit und mit guten Noten absolviert. Während ihres Studiums hat sie als Werkstudentin bei einem namhaften Konkurrenten gearbeitet und konnte so erste Berufserfahrung sammeln. Kurz nach absolviertem Studium hat Anna dann eine erste feste Anstellung als Projektmanagerin gefunden und ununterbrochen in diesem Berufsfeld gearbeitet. Somit kommt sie auf 8 Jahre Berufserfahrung. Arbeitszeugnisse ehemaliger Arbeitgeber bescheinigen Annas Befähigung zur Durchführung der geplanten Produkterweiterung.

Questionnaire: MHO vignette text

Kurzbeschreibung:

Sebastian Müller ist 34 Jahre alt und lebt mit seinem Lebensgefährten Jan in Hamburg. Sebastian hat das Abitur auf einem Wirtschaftsgymnasium absolviert und darauffolgend Betriebswirtschaftslehre studiert. Er hat 2008 seinen Bachelorabschluss und 2010 seinen Masterabschluss erlangt. Beide Abschlüsse wurden innerhalb der Regelstudienzeit und mit guten Noten absolviert. Während seines Studiums hat er als Werkstudent bei einem namhaften Konkurrenten gearbeitet und konnte so erste Berufserfahrung sammeln. Kurz nach absolviertem Studium hat Sebastian dann eine erste feste Anstellung als Projektmanager gefunden und ununterbrochen in diesem Berufsfeld gearbeitet. Somit kommt er auf 8 Jahre Berufserfahrung. Arbeitszeugnisse ehemaliger Arbeitgeber bescheinigen Sebastians Befähigung zur Durchführung der geplanten Produkterweiterung.

Questionnaire: MHE vignette text

Kurzbeschreibung:

Sebastian Müller ist 34 Jahre alt und lebt mit seiner Lebensgefährtin Jana in Hamburg. Sebastian hat das Abitur auf einem Wirtschaftsgymnasium absolviert und darauffolgend Betriebswirtschaftslehre studiert. Er hat 2008 seinen Bachelorabschluss und 2010 seinen Masterabschluss erlangt. Beide Abschlüsse wurden innerhalb der Regelstudienzeit und mit guten Noten absolviert. Während seines Studiums hat er als Werkstudent bei einem namhaften Konkurrenten gearbeitet und konnte so erste Berufserfahrung sammeln. Kurz nach absolviertem Studium hat Sebastian dann eine erste feste Anstellung als Projektmanager gefunden und ununterbrochen in diesem Berufsfeld gearbeitet. Somit kommt er auf 8 Jahre Berufserfahrung. Arbeitszeugnisse ehemaliger Arbeitgeber bescheinigen Sebastians Befähigung zur Durchführung der geplanten Produkterweiterung.

Questionnaire: Comprehension questions (by the example of the female version)

1. Zuerst zwei allgemeine Verständnisfragen

Wann erlangte Anna ihren Bachelor?

- ○ 2006
- ○ 2008

Wie viel Berufserfahrung hat Anna?

- ○ 8 Jahre
- ○ 6 Jahre

Questionnaire: Instruction page shown when incorrectly answering the comprehensive questions

Sie haben mindestens eine der Verständnisfragen falsch beantwortet. Bitte lesen Sie sich nochmals aufmerksam die Kurzbeschreibung durch.

Über "Weiter" gelangen Sie direkt zur Kurzbeschreibung.

Questionnaire: Evaluation part (by the example of the female version)

2. Kommen wir nun zu Ihrer persönlichen Beurteilung und Bewertung von Anna Müller

Bitte geben Sie Ihre persönliche Einschätzung zu den jeweiligen Bereichen ab. Dabei müssen Sie zu unterschiedlichen Bereichen auf einer Skala von 1 bis 5 anklicken, was Ihrer Meinung nach auf Anna zutrifft. Mit den Werten zwischen beiden Extremen können Sie Ihre Einschätzung abstufen.

2.1. Für wie geeignet halten Sie Anna für die Stelle der Projektmanagerin?

1	2	3	4	5
○	○	○	○	○
Nicht geeignet				Sehr geeignet

2.2. Im Falle einer Einstellung, welches monatliche Bruttogehalt im Bereich von 3000€ - 6000€ würden Sie Anna zahlen?

3000€ ist das niedrigst- und 6000€ das höchstmögliche Gehalt. Bitte notieren Sie Ihre Antwort im freien Feld.

[   ] €

2.3. Für wie durchsetzungsfähig halten Sie Anna in ihrer Rolle als Projektleitung?

1	2	3	4	5
○	○	○	○	○
Nicht durchsetzungsfähig				Sehr durchsetzungsfähig

2.4. Wie schätzen Sie Annas Risikobereitschaft ein?

1	2	3	4	5
○	○	○	○	○
Gar nicht risikobereit				Sehr risikobereit

2.5. Befindet sich Anna gerne im Wettbewerb mit anderen?

1	2	3	4	5
○	○	○	○	○
Mag keinen Wettbewerb				Mag Wettbewerb

Questionnaire: Evaluation part (continued)

2.6. Wie denken Sie, würde sich Anna in einer extrem stressigen Situation verhalten?

Bitte geben Sie für beide Bereiche Ihre Einschätzung ab.

1	2	3	4	5
○	○	○	○	○
Sehr emotional				Nicht emotional

1	2	3	4	5
○	○	○	○	○
Versteckt Gefühle nicht				Versteckt Gefühle

2.7. Wie schätzen Sie Annas Verhalten in Verhandlungen ein?

Bitte geben Sie für alle drei Bereiche Ihre Einschätzung ab.

1	2	3	4	5
○	○	○	○	○
Unterwürfig				Dominant

1	2	3	4	5
○	○	○	○	○
Nicht aggressiv				Sehr aggressiv

1	2	3	4	5
○	○	○	○	○
Kein selbstsicheres Auftreten				Sehr selbstsicheres Auftreten

3. Zum Abschluss noch einige Fragen zu Ihrer Person

3.1. Welches Geschlecht haben Sie?

- ○ weiblich
- ○ männlich

3.2. Wie alt sind Sie?

Notieren Sie Ihr Alter bitte im freien Feld.

[ ] Jahre

3.3. Wo leben Sie?

- ○ Auf dem Land
- ○ In der Vorstadt
- ○ In der Stadt

3.4. Welchen höchsten Bildungsabschluss haben Sie?

- ○ ohne Abschluss
- ○ Volks-/ Hauptschule
- ○ Realschule (mittlere Reife)
- ○ (Fach-)Abitur
- ○ Bachelor
- ○ Master
- ○ Promotion
- ○ Sonstiges, und zwar:
  [   ]

3.5. Welchen Familienstand haben Sie?

- ○ ledig
- ○ verheiratet
- ○ geschieden
- ○ verwitwet

Appendix B: Statistical tests

Statistical test: Cronbach's alpha for all evaluation measures (chapter 6.1.)

Reliability Statistics

Cronbach's Alpha	N of Items
,556	9

Statistical test: Normal P-P Plot for whole sample (chapter 6.3.)

Normal P-P Plot of Regression Standardized Residual

Dependent Variable: logwage
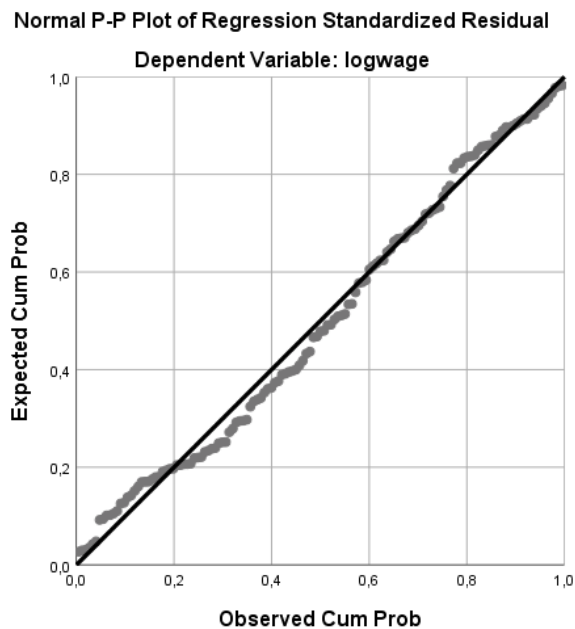
Statistical test: Durbin-Watson (chapter 6.3.)

Model Summary[g]

Model	Durbin-Watson
6	1,852

g. Dependent Variable: logwage

Statistical test: Normal Q-Q Plot for whole sample (chapter 6.3.)

Normal Q-Q Plot of logwage
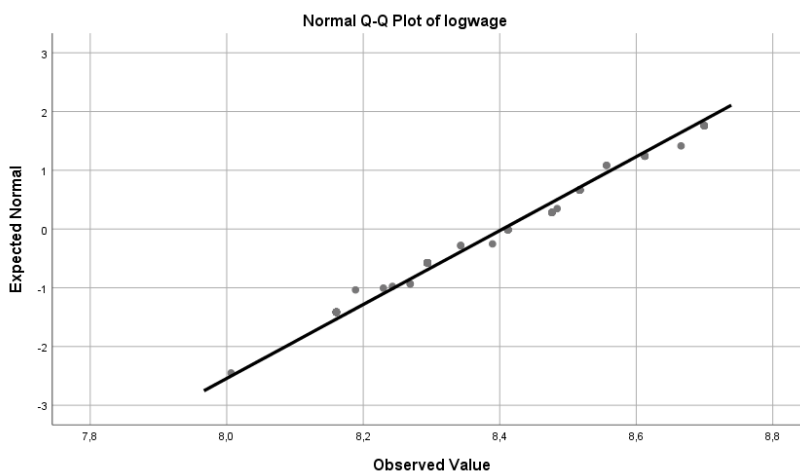
Statistical test: Normal P-P Plot for female vignettes (chapter 6.4.)

Normal P-P Plot of Regression Standardized Residual

Dependent Variable: logwage
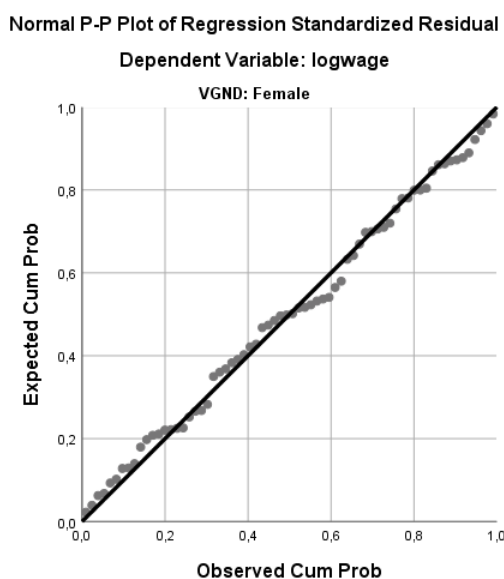
VGND: Female

Statistical test: Normal P-P Plot for male vignettes (chapter 6.4.)

Normal P-P Plot of Regression Standardized Residual

Dependent Variable: logwage
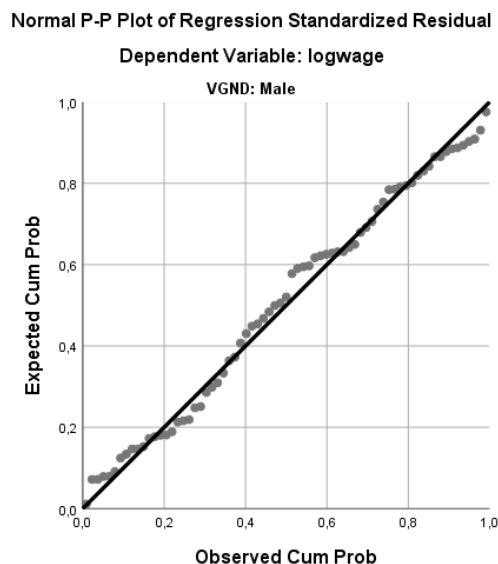
VGND: Male

Statistical test: Durbin-Watson for female vignettes (chapter 6.4.)

Model Summary[a,c]

Model	Durbin-Watson
1	2,043

a. VGND = Female

Statistical test: Durbin-Watson for male vignettes (chapter 6.4.)

Model Summary[a,c]

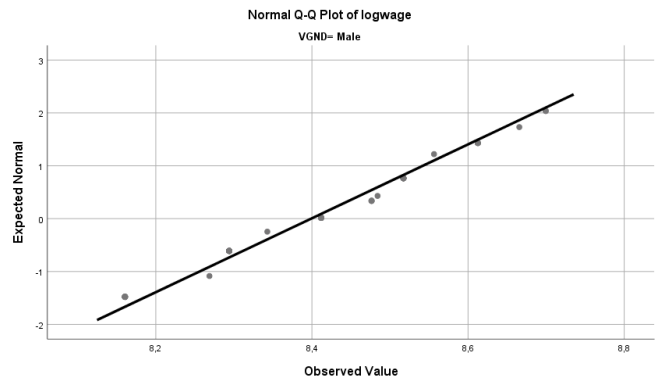| Model	Durbin-Watson
1	1,766

a. VGND = Male

Statistical test: Normal Q-Q Plot for female vignettes (chapter 6.4.)
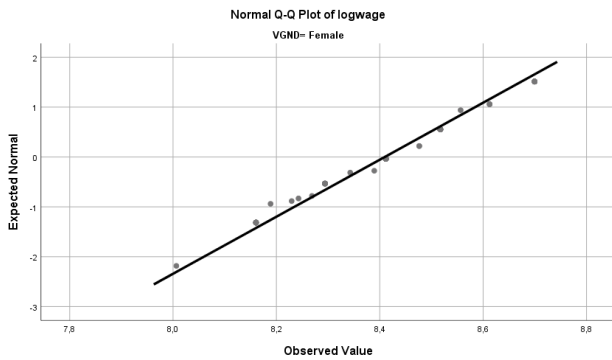
Statistical test: Normal Q-Q Plot for male vignettes (chapter 6.4.)

Statistical test: F- test for male vignettes (chapter 6.4.)

ANOVA[a,b]

Model		F	Sig.
1	Regression	1,565	,119[c]

a. VGND = Male

b. Dependent Variable: logwage

Statistical test: Cook's distance for regression with whole sample (chapter 6.3.)

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Cook's Distance	,000	,085	,008	,013	139

a. Dependent Variable: logwage

Statistical test: Cook's distance for regression with female sample (chapter 6.4.)

Residuals Statistics[a,b]

	Minimum	Maximum	Mean	Std. Deviation	N
Cook's Distance	,000	,182	,018	,029	68

a. VGND = Female

b. Dependent Variable: logwage

Statistical test: Cook's distance for regression with male sample (chapter 6.4.)

Residuals Statistics[a,b]

	Minimum	Maximum	Mean	Std. Deviation	N
Cook's Distance	,000	,299	,025	,046	71

a. VGND = Male

b. Dependent Variable: logwage

References

Antecol, H., Jong, A., & Steinberger, M. D. (2008). The Sexual Orientation Wage Gap: The Role of Occupational Sorting and Human Capital. *Industrial Labor and Relations Review, Vol 61 (4)*, 518-543.

Assael, H., & Keon, J. (1982). Nonsampling vs. Sampling Errors in Survey Research. *Journal of Marketing, Vol. 42 (2)*, 114-123.

Badgett, M. V. (1995). The Wage Effects of Sexual Orientation Discrimination. *ILR Review, Vol. 48 (4)*, 726-739.

Bishop, G. E., Hippler, H.-J., & Schwarz, N. (1987). *A Comparison of Response Effects in Self-administered and Telephone Surveys.* Mannheim: Zentrum für Umfragen, Methoden und Analysen - ZUMA.

Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature, Vol. 55 (3)*, 789-865.

Caldini, R. B. (1984). *Influence, the Psychology of Persuasion.* New York: Harper Collins.

Casson, R. J., & Farmer, L. D. (2014). Understanding and checking the assumptions of linear regression: a primer for medical researchers. *Clinical and Experimental Ophthalmology, Vol. 42*, 590–596.

Christofides, L. N., Polycarpou, A., & Vrachimis, K. (2013). Gender wage gaps, 'sticky floors' and 'glass ceilings' in Europe. *Labour Economics, Vol. 21*, 86-102.

Cliff, A., & Ord, K. (1972). Testing for Spatial Autocorrelation Among Regression Residuals. *Geographocal Analysis, Vol. 4 (3)*, 267-284.

Converse, J. M., & Presser, S. (1988). *Survey questions : handcrafting the standardized questionnaire.* Beverly Hills, California: Sage University Papers.

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research, 2nd ed.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Falk, A., & Zimmermann, F. (2013). A Taste for Consistency and Survey Response Behavior. *CESifo Economic Studies, Vol. 59*, 181-193.

Graf, N., Brown, A., & Patten, E. (2018, April 9). *The narrowing, but persistent, gender gap in pay*. Retrieved from Pew Research Center: http://www.pewresearch.org/fact-tank/2018/04/09/gender-pay-gap-facts/

Hoffmann, J. P., & Shafer, K. (2015). *Linear Regression Analysis Assumptions and Applications.* Washington DC: NASW Press.

Human Resources EDU. (2018). *How to Become a Human Resource Manager*. Retrieved from Human Resources EDU: https://www.humanresourcesedu.org/human-resource-manager/

IBM Corp. (2017). IBM SPSS Statistics for Windows. Armonk, New York.

Klawitter, M. (2015, January). Meta-Analysis of the Effects of Sexual Orientation on Earnings. *INDUSTRIAL RELATIONS Vol. 54*, pp. 4-32.

Krosnick, J. A., & Presser, S. (2009). Question and Questionnaire Design. *Handbook of Survey Research (2nd Edition)*.

Leiner, D. J. (2018). SoSci Survey (Version 3.1.06-i) [Computer software]. Available at https://www.soscisurvey.de.

Lunt, M. (2013). Introduction to statistical modelling 2: categorical variables and interactions in linear regression. *Rheumatology, Vol. 54*, 1141-1144.

Mize, T. D. (2016). Sexual Orientation in the Labor Market. *American Sociological Review, Vol. 81 (6)*, 1132-1160.

Möller, J. T. (2018). *Stereotypical Prejudices and Labor Market Behavior - A Case Study for the LGBT Community.* Hamburg: Universität Hamburg.

Moore, D., Notz, W. I., & Flingner, M. A. (2016). *The Basic Practise of Statistics, 6th ed.* New York: W. H. Freeman and Company.

Parkin, M. (2008). Priming. In P. J. Lavrakas, *Encyclopedia of Survey Research Methods* (p. 612). Thousand Oaks, CA: SAGE Publications, Inc.

Ratner, B. (2009). The correlation coefficent: Its values range between+1/-1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing*, 139-142.

Schwarz, N., Strack, F., & Mai, H.-P. (1991). Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis. *Public Opinion Quarterly Vol. 55*, pp. 3-23.

Statistisches Bundesamt. (2013). *Statistisches Jahrbuch Deutschland und Internationales.* Wiesbaden: Statistisches Bundesamt.

Sudman, S., & Bradburn, N. M. (1974). *Response Effects in Surveys - A Review and Synthesis.* Chicago: ALDINE Publishing Company.

Tavakol, M., & Dennik, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 53-55.

The National Archives. (2003). *The Employment Equality (Sexual Orientation) Regulations 2003*. Retrieved from legistlation.gov.uk: http://www.legislation.gov.uk/uksi/2003/1661/regulation/3/made

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin Vol. 103*, pp. 299-314.

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, Position, and Order - Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly, Vol. 68*, 368-393.

Voicu, B. (2015). Priming Effects in Measuring Life Satisfaction. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, Vol. 124 (3)*, 993-1013.

Williams, M. N., Gomez Grajales, C. A., & Kurkiewicz, D. (2013). Assumptions of Multiple Regression: Correcting Two Misconceptions. *Practical Assessment, Research & Evaluation, Vol. 18 (11)*, 1-14.

Willits, F. K., & Saltiel, J. (1995). Question Order Effects on Subjective Measures of Quality of Life. *Rural Sociology, Vol. 60 (4)*, 654-665.

List of Figures

List of Tables