LinkedIn and the effects on job finding rates for men and women

07-07-2023

Anne-Jet Doorduin

587188 JD

Supervisor S.V. Kapoor

Second reader A. Baiardi

Erasmus School of Economics

Abstract

What is the effect of having a LinkedIn account on finding jobs of men and women? It focuses on young people with a disadvantaged background in South Africa. LinkedIn gets more important in recruitment processes and job finding trainings, both in schools and in government programs. It is important to know if LinkedIn actually increases the likelihood of finding a job, because it can also be a helpful tool in lowering unemployment if it works. The research is done with the help of data from an experiment from Wheeler et al. (2022). They do an experiment in South Africa with a job readiness program and a LinkedIn training as treatment. The effect of LinkedIn on the job finding rate is significant and increases the end-of-program employment with eight percentage points (approximately 10 percent). Almost all results for females specific are negative, except for the amount of hours worked. Although these results are small, they are an indication of discrimination, which is important to be aware of. The advice is to deeper investigate gender discrimination in South Africa, especially with the eye on technology.

Introduction

LinkedIn is becoming an improving part of the recruitment process, as well as the job finding process. Yet, the research papers around LinkedIn and its actual effects are scarce. Since this research is so scarce, this paper will extend the literature that is there by expanding on the paper of Wheeler et al. (2022). It will also help to understand more about youth unemployment, which is an important problem in many countries. It is especially a prominent issue in South Africa, since it has been higher than in other countries, seeing the rates go up to 51% (World Bank Open Data, 2022). This paper will not only focus on youth as a minority, but also on gender. In South Africa, there is a history of gender discrimination, and although there is evidence that this is diminishing over time, the question lingers whether this is also the case in online territory (Grün, 2004). Therefore, it is important to ask the question: What is the effect of having a LinkedIn account on job finding rates of men and women in the ages between 19 and 36 in South Africa, specifically on those with a low-income background?

This question gets answered with the help of a couple of sub questions. The question that is going to be looked at in the theoretical framework is: Why should we expect differences between men and women in this situation? This can help explain why there are differences in the eventual results and may help understand where this difference is coming from. If there is a difference, it might help explain a gender wage gap. There is found evidence for this in South Africa, by for example the paper by Casale and Posel. (2010).

After this, there will be the empirical part of this research paper, which contains two questions. The first one is: What is the effect of having a LinkedIn account on finding jobs in the ages between 19 and 36? The second one is: What is the difference in the effect of having a LinkedIn account on finding jobs in the ages between 19 and 36 between men and women? This is the main question. The two questions are separate, because it is important to see if the effect is there without adding gender. In addition, the difference will become clear between the results before and after adding gender, which can possibly show if there are variable biases by seeing whether the results change much.

Surrounding LinkedIn, there is some literature, but not so much. As it has an increasingly important presence in this society, there is a need for more insights surrounding this platform. Besides that, a lot of young people, especially in western Europe, are being taught about the importance of having an account on LinkedIn and get training on how to improve their presence on the platform. Because it is being taught in the younger parts of society, it could be a possibility to help resolve youth unemployment and bridge the gap between unexperienced youth that just graduated and their future. Needless to say, there is a lot of time, money and energy invested in improving LinkedIn accounts of young people. If it turns out that this is not useful, this time, energy and money can be spent on better ways to improve youth unemployment.

Gender gaps, however, are widely discussed. As will be further discussed in the theoretical part of this paper, there is not only gaps in wages between men and women, but the discrimination in the labor market can even be experienced in the hiring process (Grün, 2004). It is important to investigate this further, also with the eye on the introduction of technology in the hiring process, because there are also gender gaps when looking at the usage of online media (Ono & Zavodny, 2003). Because of the reasons named above, and because the bridge between the areas of discrimination and technology is still missing, it is important to shed some light regarding these subjects.

In order to investigate the effect of technology on the labor market, the paper of Wheeler et al. (2022) will be used to get the data needed. It is important to name that they use the training as an instrumental variable, which means that they measure the effect LinkedIn has through the training. This can be different from the actual effect. Therefore, there is a first stage and a second stage. Through these stages the direct and indirect effect of the training will be shown.

Wheeler et al. (2022) gathers this data by doing an experiment in South Africa on young men and women between the ages of 19 and 36. They integrate a LinkedIn training of four hours into a job readiness program that is organized by Harambee. This is an institute that helps young people from disadvantaged backgrounds to find a job. They do this through different ways, but their main solution is a job finding training. This is a training where youth is trained to know how to apply for jobs and work in certain sectors.

The trainers of the program could choose which part of the training to drop to accommodate a four-hour long LinkedIn training. This training involved classes in how the platform worked and how to make a complete profile, but also in how to market the skills they have as efficiently as possible. In total, there were 1638 participants in the experiment. Wheeler et al. (2022) collected data three different ways. First, they did their own experiment and got the data through Harambee. Second, they extracted data from LinkedIn itself and matched that up by usernames and email addresses. As third and last, they did surveys to collect information 6 and 12 months after the training. With the last two ways they experienced some difficulties and nonresponses. This explains why there are some differences in the amount of participants with some results.

The participants are divided into thirty cohorts, upon which the randomization is based. This is checked with balance tests and because of the focus of this paper, only for men and women separately. The balance is tested for the whole group in the paper of Wheeler et al. (2022). There are a few variables that could be of concern in the balance surrounding women, five in total. One of this is also significant for men, which is the communication score. Seeing as to how the other scores are balanced and that the control group would have the higher score and therefore possibly an advantage, will only make possible results stronger. Another one of these variables is the completion rate of high school. There is one person in the treatment group of women that has not completed high school, and this is already an exception because this is one of the requirements of Harambee to be let into the program. The other three could be a possible influence, especially the completeness of the LinkedIn profile before the training. Yet, the differences are small, which is also important to take into account.

The experiment takes place in twenty-seven cohorts focused on call centers and three on sales in person. There are requirements in order to apply for the training of Harambee. Participants must come from a disadvantaged background, mostly low income, and have to be young. They have to make tests in order for Harambee to decide whether they have enough potential to join. The last important requirement as named above, is that they have had to finish high school. The training is given to the cohorts randomized based on cohort and location. The location can be in four cities, Cape Town, Pretoria, Durban, and Johannesburg. The training contains classes on how to create accounts, how to market skills, how to make connections and how to apply for jobs. This training is only given to the treatment group.

The analysis falls apart into three sections, with the first one being the overall results. This gives the results without taking gender into account yet, which is described as the first question. There are three outcome variables in this stage, which are: the percentage that found work after the training, the amount of accounts during training, and the completeness of the LinkedIn accounts after. Across all three variables, the constant and the treatment effect are both significant. The effect of the treatment on finding work after the training is an eight percentage points increase, the effect of the treatment on having an account during training. is between an 18 and a 19 percentage points increase compared to the mean of the control group. The effect on the completeness of the LinkedIn account after the training is not readable in percentages but has an effect of 0.08 when compared to the control group. This is seen in Table 2 in this paper on page 16.

Moving on to the second question named in the data part: is the effect different for men and women? As named above, this part falls again apart in two different results. There are first stage results, which look at whether the treatment, the LinkedIn treatment, actually influenced the LinkedIn accounts of its participants. Therefore, the outcome variables here contain the LinkedIn profile completeness after the program, the amount of LinkedIn connections after the training and the amount of jobs people applied for using LinkedIn. Only the treatment effect of the first variable is significant, and it is striking that the addition of gender causes a drop in the mean of the control group. The coefficient of gender is also not significant for any of the outcome variables, as is the interaction term between gender and being in the treatment group. The interaction effect is negative in all three. This means that it is not statistically strong to say that being a woman has a negative effect on finding a job, but seeing that all the results are negative, it can indicate that there is a negative effect that could be investigated more.

The second stage results in this paper are things that LinkedIn could have an effect on, which are finding work after the program, the amount of hours worked and placed in targeted sector. The treatment coefficient is significant and positive on all three of these variables. When comparing finding work after the program in Table 4 on page 19 to Table 2 on page 16, it is clear that the results do not change much which indicates that the addition of variables does not influence the results a lot. This often is seen when there is little to no variable bias, which indicates there is not a lot of biases in the results that influence what can be seen in Table 2 (page 16). When looking at the coefficients of the interaction term, none are significant. Two of them, finding work after training and getting a job in the targeted sector, have negative effects, while the amount of hours worked is positive. This indicates that there might be sector biases, which makes it harder for women to find jobs in certain sectors and when they do, they have to compensate in the amount of hours worked.

This paper is divided into sections. Section 1 will discuss the existing literature and talk about hypotheses. Section 2 will be a description of the data set and the experiment that has been done in order to collect the data. Section 3 will discuss results from OLS regression analyses. Section 4 will discuss possible mechanisms. Section 5 will be a summary and conclusion.

Section 1: Theoretical Framework

For this part of this paper, it is important to review some of the literature on the topic at hand. This includes three things, namely the role of LinkedIn in this current world, youth unemployment and gender discrimination on the labor market. It is important to learn about the background of LinkedIn and South Africa, to establish possible mechanisms through which the effect works, and to evaluate possible biases when it comes to the validity of this research.

The amount of LinkedIn users has been growing globally over the last few years, with the number doubling in the time between 2016 and 2022 (*LinkedIn Usage and Revenue Statistics (2023) - Business of Apps*, 2023). The age groups most present on the platform is 18-35 years old, who make up around 80% of the total amount of users and this is a big shift from the roughly 25% in 2011 (Cathy, 2011). It is advocated in schools and higher education to make an account and be active on it, to have a bigger connected network to make it easier for yourself to find a job. The question remains if this is also exactly what the platform does. The paper of Hosain and Liu (2020) shows that it has an effect on passive job seekers who already have a job. They find that passive job seekers are most likely to use LinkedIn out of the options of online job searching and show that this is because of multiple reasons, including information availability and accuracy.

With this major shift in job finding ways, the recruitment process changed as well. The study of Koch, Gerber, and De Klerk (2018) shows that the recruiters in South Africa follow a lot of their international counterparts and search for job candidates via LinkedIn as well. Moreover, they find that it is almost impossible to do the job of a recruiter without the usage of LinkedIn or other social media sites. Since the biggest part of the users on LinkedIn are in the 18-35 range, it is important to see the changes the platform made in the recruitment process. Therefore, the training can be an important part in reducing the youth unemployment in South Africa.

Youth unemployment is one of the biggest challenges South Africa faces, with the rates for the age group 15-24 going up to 51.5% currently (World Bank Open Data, 2022) compared to an estimation of 15.6% worldwide. The theories named as possible reasons for this, include lack of networks, lack of mobility and financial possibilities to relocate, as well as too high of expectations due to the income backgrounds of some families (Yu, 2013). Although all of them seem like plausible explanations, the last one is not relevant in this paper, as all the participants are from low-income backgrounds. Therefore, their expectations are different than those who grew up within a high income family.

The lack of network is an interesting reason to be named in light of this paper. LinkedIn could play an important role in lowering the unemployment rate if the lack of network plays an important role. It could even play a role in the location problems as well, seeing as to how LinkedIn could connect young people with companies. However, like in a lot of societal problems, culture, upbringing, and community also play an important role. An example in South Africa is the fairly recent abandoning of the apartheid. This would introduce discrimination into this equation, making it more difficult to create policies that work. There have been multiple solutions offered, one of them being ALMP, which is short for Active Labor Market Programs. This is part of creating a bridge between secondary education and the labor market, which can also be a part of the problem of youth unemployment.

The paper of Card, Kluver and Weber (2017) found that ALMPs are more likely to have an effect on women and people that have been unemployed longer. This also is something important that should be considered. Since the focus of this paper is on gender differences, it is important to take into account how certain things can influence the results. If these programs are more effective for women, this could tilt the results. To better anticipate the results, it is first important to look at the labor market situation for these different genders, to investigate possible gender differences, which in turn will influence the hypothesis.

There is already a striking difference in the youth unemployment rate between the genders. For females, this rate hangs around the 55.5%, while the males are around the 46.6%. This can already indicate that there are differences present in the South African labor market, because these are only young people who indicate to be looking for a job. This is without considering that there are differences in the participation rates between men and women (Winter, 1999) According to the paper of Grün (2004) there is even a difference in race when women deal with discrimination in the labor market. African women often deal with more discrimination in the hiring process, while white women often experience discrimination in wage differences. Winter (1999) also found evidence of a gender pay gap that was significantly bigger between white women and their male counterparts compared to the African men and women. Wage discrimination can point into the direction of discrimination in the hiring process as well, but there is little research done around this topic. There is, however, evidence on a gender bias in certain sectors.

According to the paper of Gradín (2021), women are often overrepresented in low-paying jobs, which could be a cause for the gender wage differences previous studies show. This is getting better over time, as more women are getting better education, which allows them to go for higher paying jobs. There is still trouble with getting into higher positions within firms, which is seen in the almost completely male dominated management positions also shown in the paper of Gradín (2021). In these male dominated positions and the occupations that women are now getting more into, the gender pay gap stays persistent and is significantly bigger than in the female dominated sectors (Adeleken & Bussin, 2022).

In the Global Gender Gap Report (2017), there is discussed how the gender gaps changed over time, discussing 144 different countries, and taking four different types of gender gaps into account. This includes but is not limited to the economic participation and opportunity and educational attainment. In this report, they found that, although there is a decline in this gap, in Africa it stays around the 32% gap. This entails that the remaining gap between men and women still has an effect of 32%. Although this is relatively big compared to for example Europe, with its 25%, it also means that 68% percent of this gap has been successfully taken care of.

Although evidence above suggests there still are gender differences in the labor market and in day-to-day life, there is also evidence that suggests that there is also a big role for the inherit differences between men and women. The paper of Nyakudya et al. (2017) found that the reason there are less self-employed women is because of a lack of internal skills, which make women choose other jobs instead of entrepreneurship.

Besides this, there is also an interesting paper from Grant and Behrman (2010) that explains that there might also be a gap between men and women when it comes to education. The surprising results from their paper, however, show that among children who have ever attended school, the girls under the age of sixteen have an equal or even greater schooling progress than boys their age. This indicates that they might thrive better within schooling environments and therefore might get into higher education easier. This could eventually erase a part of the differences in the workplace in the future, since women will become the better educated group.

For the last part of this section, the literature found and described above will be used to make predictions for the results later in this paper. The results in the paper of Wheeler et al. (2022) will also be used for the first part of the results, the effect of the training will be positive for the job finding rate. Wheeler et al. (2022) uses an ALMP which is already there, focused on youth from a disadvantaged background. When looking at the second question for the results section, the prediction gets harder to make because of the gender component. Still, since there seem to be at least some gender differences in the labor market, it seems plausible that the effect of LinkedIn will be less for women compared to men. This is because of the idea that there could be a dislike towards women, which makes women less likely to be hired. If, in that case, women expand their search area, they will still be less likely to be hired than men, because less companies will be interested. It does intrigue what the results will precisely be, because of the difference just the ALMP already have as described above. This might influence the effect that is shown, because the overall effect of the training is bigger for women than for men. Thus, expected is a low, but negative number for the job finding rate when looking at the gender effect.

Section 2: Data Description

For this part, the data and experiment explanation are drawn from the paper of Wheeler et al. (2022). They use a job finding training that is already there in order to have a place to educate the treatment group on using LinkedIn. They call it the job readiness program. The program was already given by an organization called Harambee. It is a full-time program lasting six to eight weeks. They cover anything from workplace simulations to teambuilding and developing noncognitive skills. They also help with applying for jobs. Harambee has partner firms where they have long-term relations with. This can help with placing the participants from the training. Harambee helps through this training, with the application process for jobs and sometimes sets up an interview.

All candidates are from disadvantaged backgrounds and had to do three types of tests to confirm their potential before being let into the program. This comes down to the excluding of people from middle and upper-income families. They all applied to get into the program, which implies that they are active work seekers. The three tests are numeric, communication and cognitive. Since they have to score well on these tests to be let into the program, they have a positive selection bias in comparison to the rest of the population in terms of employment prospects. There is one more important note to be made. The race of the participants is not known in this dataset. Therefore, an analysis on the differences that might be left behind from the apartheid cannot be performed.

The training program included classes on how to construct a profile, what information should be in the profile, how to add in the job readiness program on the profile, how to identify and join groups, how to make connections, how to view profiles and how to ask for recommendations. The training was four hours in total and it was done in two different trainings, namely in May 2016 and January 2018. It is also important to name that there are in total five treated cohorts that did not fully finish the LinkedIn curriculum.

The sample size eventually contained 1638 observations, although there are some missing values. They are divided into thirty cohorts, of which twenty-seven are focused on call center jobs and the remaining three focus on the sales jobs. The call center cohorts train for work both in the business process outsourcing firms and in banks or insurers with call centers in-house. The training is both for customer service roles as it is for sales jobs. The trainers only got to know if they were in the control or treatment group when it was too late to change or transfer groups. The trainers were allowed to make a choice on what part of the training to drop to be replaced by the LinkedIn training. No one observes these choices.

The assignment of the cohorts was at random and based on the number of the cohort within a city. They therefore are sequentially paired in the randomization. For example, numbers one, three, five are in treatment and the groups two, four, six are in control group within a certain region. So, if Cape Town had ten cohorts, cohorts one, three, five, seven and nine would be in the treatment group and the rest in the control group. Treated cohorts received an in-detail training on LinkedIn as described above. There is data on whether they had LinkedIn before starting the training, the activeness of this account, the use of the account during the training and the amount of connections before, during and after the training. There is even more data surrounding LinkedIn, but these variables will not be used in this paper and therefore will not be discussed.

Because the training is mostly focused on disadvantaged youth, there are very few people with a higher degree, around 6% which can be seen in Table 1 in Appendix A on page 24. They focused the training in four big cities, being Cape Town, Durban, Johannesburg, and Pretoria, which is the region in Appendix A. They combined four rounds of survey data with the administrative data of Harambee and data from LinkedIn itself. The final two surveys had some nonresponses, as they were follow-ups done after 6 and 12 months after the training. They extracted data from LinkedIn itself, using the names and e-mail addresses of the participants to keep track of them after the training. This was often difficult, hence the drop in observations in Appendix A when this data is needed. The age in the training is relatively young, with 36.8 being the highest, but the mean being 23.7. This means there are a lot of younger people in this dataset, which would make sense. The training is mostly focused on people who have little work experience, which young people are more likely to have. There are around 61% women in the groups. When looking at the work after, about 74% of the participants find a job after the training of Harambee, which is a lot. Especially when comparing that with the high unemployment rate explained above.

Added to that, the amount that is placed in the targeted sector from the total number of participants is 63%. When compared to each other, there is only 11% which started working in another sector than they were training for, which is not much. The scores seen from these tests in Table 1 in Appendix A (page 24) are standardized to have a mean of zero, these are also seen in Table 2 and 3 in Appendix A (page 25 and 26), as well as the differences between control and treatment group in Table 1. As mentioned before, not everyone completed the program. The percentage of participants who did, is 87%. Lastly, something that is also striking is the hours they work. The average there is 28.25. When comparing this to the normal fulltime workweek in South-Africa of 45, this is not a lot and therefore it is something interesting to investigate as well (Andre, 2023).

Since there is randomization in the treatment effects, it is important to make sure that the randomization went correctly. Therefore, there is a balance table constructed in Table 1 on page 14. These balance tests are split up based on gender, because of the main focus of this paper. Since the results will mostly focus on the differences between these two, it is important to make sure that there are no alternate differences between the treatment and control group based on this. For the extra balance tests for the entire group, it is advised to look at Wheeler et al. (2022). Besides this, the complete balance tables can be found as Table 2 and 3 in Appendix A on page 25 and 26.

As expected, there are some variables that are significant in differences. When looking at the males, there is only one, which is the communication score on one of the tests. Since the other tests are not significant at all, it is doubtful on whether this is important to keep in mind. For the women, there are a couple more variables that are significantly different from one another. One of them again being the communication score. This variable is a little bit odd on why this is significantly different, because the other scores are balanced. That is why this will not be thoroughly looked at.

The next one is age, which is about half a year difference between the control group and the treatment group. When looking at males and comparing the two, we see that the males also have a difference, although not significant, and the same way (treatment group being a little younger) which makes it believable that there is no problem there. The next one is high school education. It is not particularly important to keep this one in mind, as it is mandatory to have the secondary education completed.

Within the university degree, there is also a significant and substantial difference. Especially since the mean of the overall sample is 6%, a 4% difference between the two groups is a lot. The positive thing, however, is that the percentage is higher in the control group. If college degree would impact the usage of LinkedIn positively, or it would impact the placement in a job after positively, it would only lessen the effect of LinkedIn.

The last one that is significant, is the LinkedIn profile completeness before the training. This could become a problem, as this is something the treatment will have an effect on and will probably make this difference even bigger. This is a variable difference that should be kept in mind when interpreting the results. The biggest issue possibly at play here is the sample size. The bigger the sample size, the smaller chances for biases. This will also be seen later. A bigger sample size would have helped to clear up whether some of the results will be significant. In addition, it would also help to clear up whether the differences are there because of an error in the randomization or not. The complete and separate tables for men and women can be found in Appendix A. The eventual analysis will have as main method OLS regressions, with robust standard errors.

Table 1: Balance Table for men and women

Variable	*Women*	*Men*	*Observations*
Age	-0.56**	-0.38	1636
	(0.18)	(0.24)	
Region	0.04	0.09	1638
	(0.05)	(0.06)	
Numeracy Score	-0.05	-0.10	1547
	(0.06)	(0.08)	
Communication Score	0.12**	0.17**	1610
	(0.06)	(0.08)	
Cognitive score	0.04	0.09	1617
	(0.06)	(0.08)	
High School Education	0.01*	0.00	1500
	(0.00)	(0.01)	
University Education	-0.04**	0.00	1500
	(0.02)	(0.02)	
LinkedIn Profile Before	-0.01	0.05	1520
	(0.02)	(0.03)	
LinkedIn Profile Completeness Before	0.19***	0.11	863
	(0.06)	(0.08)	

Note: This table shows the difference between the treatment and control group for women and men for the data from an experiment with a LinkedIn training from Wheeler et al. (2022). It shows the means for both groups with the standard deviation between brackets. The significance levels are denoted by stars. *p<0.1 **p<0.05 ***p<0.01. The last column shows the observations for each variable. The full tables separately can be found in Appendix A on page 25 and 26.

Section 3: Results

Table 2 on page 16 shows the results without taking gender into account. These are the results that are also discussed in the paper of Wheeler et al. (2022). In Table 2, there are a couple of things worth pointing out. First, the three variables looked at in this table are: found work after, had an account during training and the completeness of a LinkedIn account after. These will also come back when looking at the gender effects in later results with the help of an interaction term of gender and treatment. With first stage results, it is meant to look at the direct effect of the training itself. This means effects on the amount of accounts during training, and the completeness of these accounts. The second stage is more about what the effect is from these better accounts. An example of this stage is the found work variable. The biggest effect of the treatment is on an account during training, which is about thirty-nine percentage points. This comes down to an average of 48% mean of the treatment group compared to a 30.5% mean in the control group. Comparing this to the LinkedIn completeness variable is not really a great comparison, as this variable is measured in values between 1 and 4, which means that the coefficients are not translatable to percentages. The coefficient of the treatment is still significant and comes down to 0.208. The one that Wheeler et al. (2022) was also most interested in, the found work after variable, the constant and treatment variable are both strongly significant, and come down to a constant of 70.1% and a treatment effect of eight percentage points. When comparing this to the mean of the control group, there is still a 4% difference, which indicates that the usage of LinkedIn has an effect on the job finding rate. Something that is also important to note is the vast differences in observations between the first two columns and the last column. This is almost half compared to the first two columns. Still, the constant, treatment coefficient and the mean of the control group are all significant with the first two significant on a 1% level and the last one significant on a 5% level. Together with the differences between the treatment and the control group, this indicates that the treatment increases the Completeness of a LinkedIn account and therefore indicates the active participation in the treatment. The same goes for the second column, where it is clear that the treatment strongly influences the amount of accounts there are. It shows that the treatment has a direct effect on the usage of LinkedIn, through both the Completeness of the accounts, but also through the amount of accounts. This is needed in order to establish if there is a relationship between the usage of LinkedIn and the job finding rate. Looking at this rate in column one (finding placement after) and seeing the difference between the control group and the treatment group, there is a significant difference, as named before. This tells us that there has to be an effect of the usage of LinkedIn on the job finding rate, because the treatment assignment was random, which was secured by the balance table.

Table 2: Results

Variable	***Work After***	***Account During Training***	***LinkedIn Completeness After***
Constant	0.701***	0.094***	1.779***
	(0.017)	(0.011)	(0.037)
Treatment	0.080***	0.393***	0.208***
	(0.021)	(0.020)	(0.444)
Mean Control Group	0.744	0.305	1.911**
	(0.437)	(0.461)	(0.619)
Number Of Respondents	1626	1566	863
Number Of Cohorts	30	30	30

Note: This table shows the regression results using the data from Wheeler et al. (2022). It shows the treatment effect of having LinkedIn on finding a workplace after, whether it had an effect on having an account during treatment and whether it had an effect on the LinkedIn profile completeness after the training. The p-values are noted with stars, *p = 0.10 **p = 0.05 ***p = 0.01.

From Table 2 (page 16), we move on to Table 3 (page 17). It shows the first stage results with the gender component and the interaction term included. The variables used to describe these first stage results are the following three: the completeness of a LinkedIn account after training, the amount of LinkedIn connections after and the amount of jobs applied through LinkedIn. The first important thing to point out is that the results of LinkedIn account completeness do not change much. This is an indication that the addition of variables does not correct for potential left-over variable biases and therefore the results are strong. The second thing that is clear, is that none of the interaction term coefficients are significant, which is interesting. It is a negative sign with all three variables, which can be an indication that the treatment effect is less, and that the sample size just was not big enough. This is, however, speculation. The amount of jobs applied via LinkedIn have a small treatment effect as well. One last thing that is interesting to point out is about the amount of LinkedIn connections column in Table 3. The standard errors and the coefficients were indicative of a possible outlier, maybe multiple, which is seen in Table 4 in Appendix A (page 27). This is checked in the graph shown in Figure 1 in

Appendix B (page 29). As shown in this figure, there is one exceeding the 10000 mark, which is a lot of LinkedIn connections and definitely an outlier. Because of this outlier and multiple others around the 4000 mark, the decision was made to make this a logarithmic regression. Through this, the treatment coefficient immediately became significant, which is an indication that the method worked. These results are more in line with the results found in the other variables and therefore are more believable.

Table 3: First stage results

Variable	*LinkedIn Completeness After*	*Log (LinkedIn Connections After)*	*Jobs Applied Using LinkedIn*
Constant	1.79*** (0.06)	2.624 (14.15)	0.23** (0.10)
Treatment	0.25*** (0.07)	0.525** (0.01)	0.00 (0.14)
Female	-0.02 (0.08)	-0.364 (0.12)	0.10 (0.15)
Female * Treatment	-0.07 (0.09)	-0.197 (0.46)	-0.09 (0.19)
Control Group Mean	1.78*** (0.65)	2.391 (1.77)	0.30 (1.37)
Number Of Respondents	862	1005	862
Number Of Cohorts	30	30	30

Note: This table shows the regression results using the data from Wheeler et al. (2022). It shows the direct treatment effects of the LinkedIn training on the usage of LinkedIn. The interaction term is the interaction effect of the treatment and the gender. Gender is denoted by female and is equal to one if the participant is female. There are robust standard errors used. The LinkedIn Connections after is in logarithmic values. The p-values are noted with stars, *p = 0.10 **p = 0.05 ***p = 0.01.

Moving on to the second part of the results, namely Table 4 (page 19). Above, there is a discussion about the first stage results, these are the second stage results. These include the following three variables: found work after, amount of hours worked and placed in targeted sector. The first thing that is important to name is that for all the variables, the constant and the treatment variable is significant on a 1% significance level. This, just like with the results of LinkedIn Completeness, can indicate that the results are robust against variable biases. The results of the interaction term and of gender are both not significant. The interaction coefficient does show a substantial difference between the variables, as it is -0.02 and -0.06 for the placement after and the targeted sector, respectively. As can be seen in Table 5 in Appendix A (page 28), the indication was given with a standard regression that there might be a lot of observations with a zero, which was holding the results down. This was checked by doing a scatterplot, which is seen in Figure 2 in Appendix B (page 30). In this Figure, there is a clear mean around the 40-hour mark, and not around the 25-hour mark. Therefore, the regression seen in Table 4 (page 19) has excluded the zero observations but included everything above that zero mark. The constant in the regression doubled as a result. This is more logical, as the average of hours worked within a job in South Africa is 45 hours (Andre, 2023). It is striking that women still seem to work more on average than men. Especially the amount that is seen in Table 4, with 1.74 hours. This is almost two hours a week, which is quite a lot. It can indicate that women must work harder to prove themselves. Since this coefficient is not significant, a bigger sample size will be necessary to investigate this effect further.

The decrease in likelihood of getting placed in the targeted sector seen in Table 4 on page 19 could be an indication of sector discrimination. This would mean that some sectors could be more female friendly than others, since the decrease in targeted sector is bigger than the decrease of getting placed overall. It is also important to note that here the amount of respondents is a lot higher than in Table 3 on page 17. This can play an important role in whether the coefficients found are significant. Although it is higher, it is likely not high enough to get significant results around men and women. Because of adding this division between these groups, the groups must be even bigger than just a comparison between the control and treatment group.

Table 4: Second stage results

Variable	*Work After*	*Hours worked*	*Targeted Sector*
Constant	0.70***	41.17***	0.51***
	(0.03)	(0.77)	(0.03)
Treatment	0.09***	0.33	0.20***
	(0.03)	(0.89)	(0.04)
Female	0.00	-0.91	0.05
	(0.03)	(0.91)	(0.04)
Female * Treatment	-0.02	1.74	-0.06
	(0.04)	(1.13)	(0.05)
Control Group Mean	0.70	40.63***	0.54
	(0.46)	(7.52)	(0.50)
Number Of Respondents	1625	755	1625
Number Of Cohorts	30	30	30

Note: This table shows the regression results using the data from Wheeler et al. (2022). It shows the indirect treatment effects of the LinkedIn training on the work after, hours worked excluding the zero observations and the targeted sector. This last one denotes the amount of participants that got into the field the training prepared them for. The interaction term is the interaction effect of the treatment and the gender. Gender is denoted by female and is equal to one if the participant is female. The p-values are noted with stars, *p = 0.10 **p = 0.05 ***p = 0.01.

Section 4: Mechanisms

There are a couple of mechanisms that are suggested by Wheeler et al. (2022) for the way the results work in their paper. There are a total of five suggested, which are: *Self-beliefs* (Things taken into account like trust and reservation wage), *Program Engagement* (The treatment would change the level of engagement with the program), *Search Costs* (Lowers the costs to look for jobs), *Referrals* (LinkedIn increases believability through visible accounts and connections), *Demand-Side Information* (LinkedIn increases the information available to the employer) and *Supply-Side Information* (LinkedIn increases the information available to the job-seeker). The evidence found in Wheeler et al. (2022) supported this last mechanism the most.

However, this does not explain the mechanism behind the found results surrounding gender. An explanation could be discrimination, but this is quite broad. Too broad to help with making decisions surrounding the fixing of this gap. Let us first focus on the first stage results found. In this, the results for females are already less. This means the completeness of the profile is less, but also the amount of connections and the amount of jobs applied to through LinkedIn are less. It is unlikely that these results are less because of discrimination. The likeliness that people would not accept connections because it is a woman who is asking, is unlikely. LinkedIn completeness rate is even more intriguing, because this implies that there might be less *Program Engagement* for women than for men. They are putting less time in updating their LinkedIn page and on finding connections.

When looking at the second stage results, the effects are around the same size as the ones from the first stage, without considering the amount of hours worked. This makes it more likely that the *Program Engagement* could seriously play a role in these results. It could also be that there is a certain form of sector bias, which causes women to not work in the sectors that they are trained for in the program. This could also influence the amount of hours worked, as this sometimes depends on sectors. The amount of hours worked, is still sticking out. It could be interesting to further investigate this, together with the sectors, to figure out if there is indeed a connection there.

Section 5: Summary and Conclusion

To summarize, the effect of LinkedIn on finding jobs is positive and significant, around the 10% mark (7 percentage points). Treatment increases the likelihood to work in the targeted sector of the training. Treatment also increases the amount of accounts during training and the completeness of an account afterwards, which indicates the treatment worked. Most likely the way that LinkedIn influences the job finding rate is through supplying information to the participants/job seekers about the employers. This way, job seekers know where they can apply. The most likely mechanism through which the LinkedIn accounts have effect is called the *Supply-Side Information*. This works through informing users on what options are there in terms of employers. It familiarizes the users with the names of different companies and therefore it expands their field of possibilities.

The effect of gender in combination with treatment is negative on all fronts except for the amount of hours worked. The negative numbers in the first stage could indicate there might be a decrease in the program engagement, which could lead to negative results for women in the second stage. The amount of hours worked for women is around two hours higher, which is striking if looked at the means, because both the constant and the control group mean are around the 40–41-hour mark. The significance is not found in any of the results, but this is likely due to the amount of observations.

For future research, it is highly recommended that there is more investigating into the effects of LinkedIn in the job market industry. With these results and possibly more evidence, more classes and trainings can be organized to reduce unemployment. Since youth unemployment is a problem worldwide, it would be interesting to see if the same results surface when applying it to different continents. It could be helpful to connect more people via LinkedIn, because of the big shortages on the labor market right now.

Another interesting direction to take this research in is discrimination, and then with a focus on race as well. Because of the lack of variables in this data set, this is not possible with this data. This is especially interesting in South Africa, there are multiple studies showing the remainders of the apartheid, which can add an interesting dimension onto the results. It could also be interesting to look at gender more thoroughly, because of the lack of significance in this case and do it with a bigger experiment.

References

Adeleken, A., & Bussin, M. H. R. (2022). Occupational segregation and gender pay gap dynamics in the formal sector of South Africa. *Journal of Human Resource Management*, *20*(1). https://journals.co.za/doi/abs/10.4102/sajhrm.v20i0.1660

Andre. (2023). Working Hours and Overtime. *Labour Guide South Africa*. https://labourguide.co.za/employment-condition/working-hours-and-overtime/

Card, D., Kluve, J., & Weber, A. (2017). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association*, *16*(3), 894–931. https://doi.org/10.1093/jeea/jvx028

Casale, D., & Posel, D. (2010). Unions and the Gender Wage Gap in South Africa. *Journal of African Economies*, *20*(1), 27–59. https://doi.org/10.1093/jae/ejq029

Cathey, G. (2011). LinkedIn User Demographics and Visitor Statistics 2011. *Boolean Black Belt-Sourcing/Recruiting*. https://booleanblackbelt.com/2011/09/linkedin-user-demographics-and-visitor-statistics2011/#:~:text=26%25%20of%20LinkedIn%20users%20are,at%20the%20graduate%20school%20level.

Gradín, C. (2021). Occupational Gender Segregation in Post-Apartheid South Africa. *Feminist Economics*, *27*(3), 102–133. https://doi.org/10.1080/13545701.2021.1906439

Grant, M. J., & Behrman, J. R. (2010). Gender Gaps in Educational Attainment in Less Developed Countries. *Population and Development Review*, *36*(1), 71–89. https://doi.org/10.1111/j.1728-4457.2010.00318.x

Grün, C. (2004). Direct and indirect gender discrimination in the South African labour market. *International Journal of Manpower*, *25*(3/4), 321–342. https://doi.org/10.1108/01437720410541425

Hosain, S., & Liu, P. (2020). LinkedIn for Searching Better Job Opportunity: Passive Jobseekers' Perceived Experience. *The qualitative report*. https://doi.org/10.46743/2160-3715/2020.4449

Koch, T., Gerber, C., & De Klerk, J. J. (2018). The impact of social media on recruitment: Are you LinkedIn? *SA journal of human resource management*, *16*. https://doi.org/10.4102/sajhrm.v16i0.861

Nyakudya, F. W., Simba, A., & Herrington, M. R. (2017). Entrepreneurship, gender gap and developing economies: the case of post-apartheid South Africa. *Journal of small business and entrepreneurship*, *30*(4), 293–324. https://doi.org/10.1080/08276331.2017.1412612

Ono, H., & Zavodny, M. (2003). Gender and the Internet*. *Social Science Quarterly*, *84*(1), 111–121. https://doi.org/10.1111/1540-6237.t01-1-8401007

LinkedIn Usage and Revenue Statistics (2023) - Business of Apps. (2023, 26 mei). Business of Apps. https://www.businessofapps.com/data/linkedin-statistics/

Schwab, K., Samans, R., Zahidi, S., Leopold, T. A., Ratcheva, V., Hausmann, R., & Tyson, L. D. (2017b). The global gender gap report 2017. *World Economic Forum*. https://apo.org.au/node/208501

Wheeler, L., Garlick, R. L., Johnson, E., Shaw, P. J., & Gargano, M. (2022). LinkedIn(to) Job Opportunities: Experimental Evidence from Job Readiness Training. *American Economic Journal: Applied Economics*, *14*(2), 101–125. https://doi.org/10.1257/app.20200025

Winter, C. (1999). Women workers in South Africa: participation, pay and prejudice in the formal labor market. *Non-Journal*, 1. http://files.eric.ed.gov/fulltext/ED453132.pdf

World Bank Open Data. (2023). World Bank Open Data. https://data.worldbank.org/indicator/SL.UEM.1524.ZS?locations=ZA

Yu, D. (2013). Youth unemployment in South Africa revisited. *Development Southern Africa*, *30*(4–05), 545–563. https://doi.org/10.1080/0376835x.2013.830964

Appendix A: Tables

Table 1: Summarized Characteristics

Variable	*Observations*	*Mean*	*SD*	*Min*	*Max*
Age	1636	23.65	2.95	17.81	36.81
Female	1633	0.61	0.49	0	1
Region	1638	2.03	0.84	1	4
Numeracy	1547	-0.034	0.997	-2.178	4.118
Communication	1610	0.076	0.955	-3.239	2.649
Cognitive	1617	0.037	0.979	-4.294	2.401
Treatment	1638	0.54	0.50	0	1
High School	1500	0.99	0.08	0	1
University	1500	0.06	0.24	0	1
Work after	1626	0.74	0.44	0	1
LinkedIn Account Before	1520	0.17	0.37	0	1
Completed Program	1612	0.87	0.33	0	1
LinkedIn Completeness Before	691	1.53	0.59	1	3
LinkedIn Completeness After	863	1.91	0.62	1	3
LinkedIn Connections Before	1366	5.53	10.97	0	40
LinkedIn Connections After	1580	15.64	27.17	0	92
Jobs Applied LinkedIn Before	668	0.14	1.05	0	16
Jobs Applied LinkedIn After	863	0.26	1.36	0	17
Placed In Targeted Sector	1626	0.63	0.48	0	1
Hours Worked	1107	28.25	20.24	0	96

Note: This table shows summary statistics for the data set of Wheeler et al. (2022). The test scores are standardized to have a mean of zero. The cognitive test has similarities to a Raven's test, which is a multiple-choice intelligence test. The completed program mean is about the amount of people that finished the program, not the amount that finished the LinkedIn training. The amount placed in the targeted sector is a percentage of the total amount that followed the training.

Table 2: Balance Table For Men

Variable	*Control Group*	*Treatment Group*	*Difference*	*Observations*
Age	24.19 (3.09)	23.78 (3.02)	-0.38 (0.24)	631
Region	1.97 (0.74)	2.07 (0.86)	0.09 (0.06)	632
Numeracy Score	0.35 (0.97)	0.25 (1.02)	-0.10 (0.08)	600
Communication Score	0.05 (1.06)	0.21 (0.92)	0.17** (0.08)	622
Cognitive score	0.21 (1.10)	0.30 (1.00)	0.09 (0.08)	624
High School Education	0.99 (0.11)	0.99 (0.10)	0.00 (0.01)	582
University Education	0.06 (0.24)	0.06 (0.24)	0.00 (0.02)	582
LinkedIn Profile Before	0.15 (0.37)	0.20 (0.40)	0.05 (0.03)	590
LinkedIn Profile Completeness Before	1.46 (0.60)	1.57 (0.62)	0.11 (0.08)	261

Notes: This table shows the balance between the treatment and control group for men for the data from an experiment with a LinkedIn training from Wheeler et al. (2022). It shows the means for both groups with the standard deviation between brackets. The third column shows the differences with the significance levels denoted by stars. *p<0.1 **p<0.05 ***p<0.01. The last column shows the observations for each variable conditional on the gender.

Table 3: Balance Table For Women

Variable	*Control Group*	*Treatment Group*	*Difference*	*Observations*
Age	23.76 (2.90)	23.19 (2.83)	-0.56** (0.18)	1000
Region	2.00 (0.84)	2.04 (0.86)	0.04 (0.05)	1001
Numeracy Score	-0.21 (0.96)	-0.26 (0.92)	-0.05 (0.06)	946
Communication Score	-0.03 (0.96)	0.09 (0.91)	0.12** (0.06)	987
Cognitive score	-0.13 (0.91)	-0.08 (0.90)	0.04 (0.06)	992
High School Education	0.99 (0.08)	1 (0)	0.01* (0.00)	918
University Education	0.08 (0.28)	0.04 (0.20)	-0.04** (0.02)	918
LinkedIn Profile Before	0.16 (0.37)	0.15 (0.35)	-0.01 (0.02)	930
LinkedIn Profile Completeness Before	1.39 (0.60)	1.58 (0.55)	0.19*** (0.06)	430

Notes: This table shows the balance between the treatment and control group for men for the data from an experiment with a LinkedIn training from Wheeler et al. (2022). It shows the means for both groups with the standard deviation between brackets. The third column shows the differences with the significance levels denoted by stars. *p<0.1 **p<0.05 ***p<0.01. The last column shows the observations for each variable conditional on the gender.

Table 4: LinkedIn account connections after training normal regression

	LinkedIn Connections After
Constant	71.30
	(40.05)
Treatment	-27.10
	(40.23)
Female	27.79
	(67.79)
Female * Treatment	-44.27
	(67.94)
Mean of Control Group	89.23
	(702.29)
Observations	1005
Cohorts	30

Note: This table shows the normal regression results including the outliers and offers a first look at the results of the amount of connections after training. Because of the big constant and mean of control group and added to that the negative treatment coefficient, the results in the text are done with a logarithmic regression. There are no significant coefficients.

Table 5: First results total hours worked including all observations

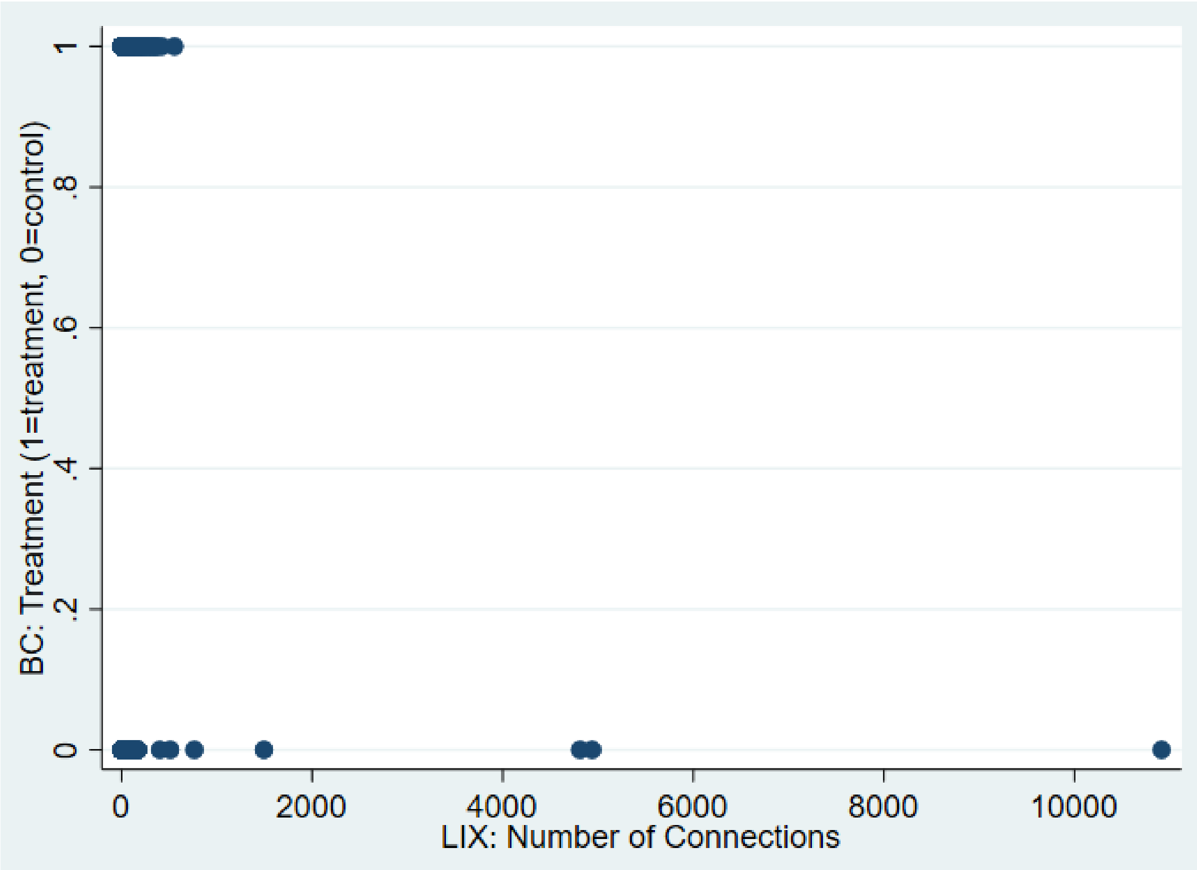
	Hours worked
Constant	27.60***
	(1.51)
Treatment	3.96**
	(1.92)
Female	-2.90
	(1.89)
Female * Treatment	0.77
	(2.48)
Mean of Control Group	25.79
	(20.48)
Observations	1107
Cohorts	30

Note: This table shows the total amount of hours worked including all observations, whereas the table in the text only includes observations above zero. Because the mean was so different than the mean found in a lot of papers and information, it was interesting to look further into why it was so different. The standard errors are robust and are between the brackets. The significance level is denoted by stars. *p<0.1 **p<0.05 ***p<0.01

Appendix B: Figures

Figure 1

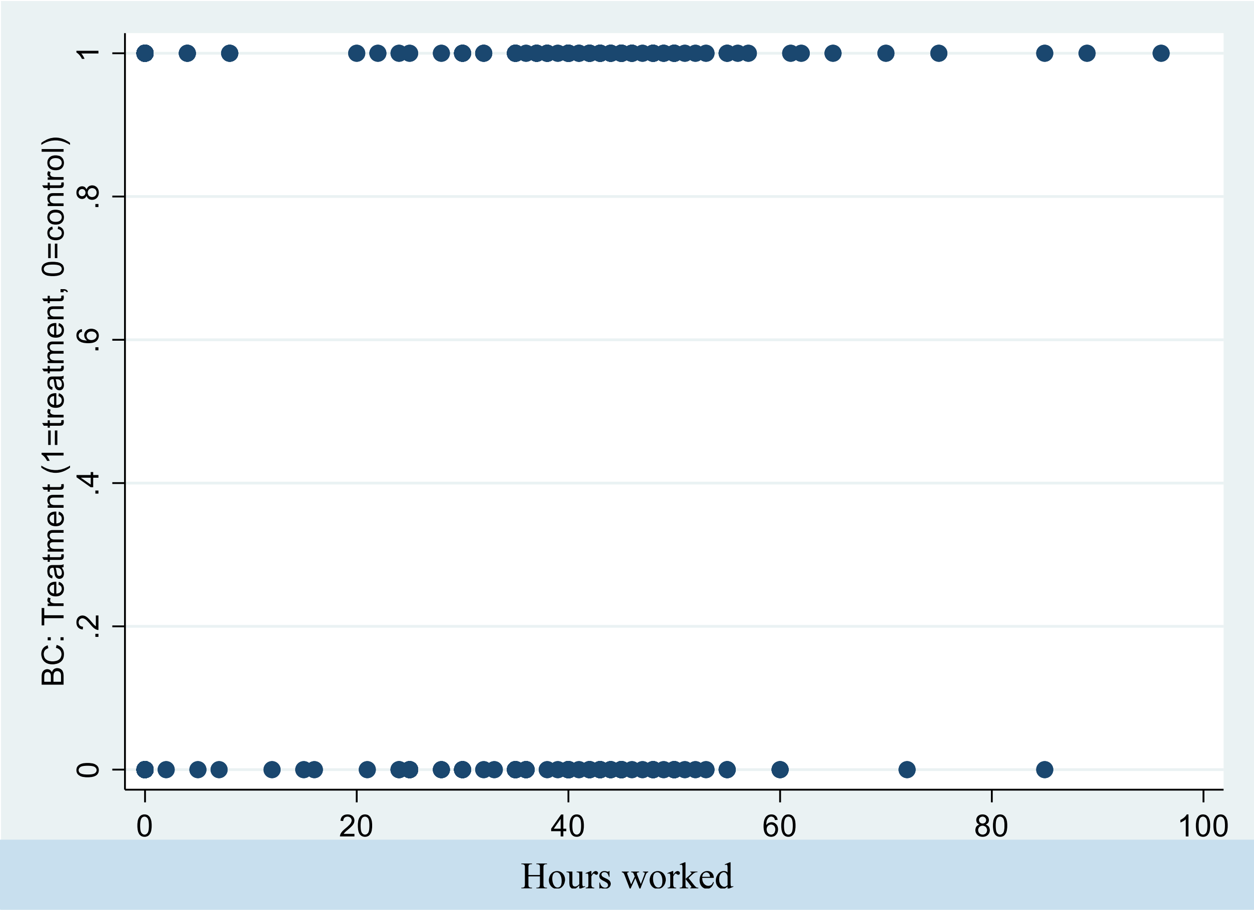
Treatment and Number Of Connections



Note: This figure is shown because of the possibilities of outliers. There are clearly a couple of outliers interesting to investigate with the help of a logarithmic regression, which is used in the Table 3 in the text.

Figure 2

Hours worked scatterplot



Note: This figure shows the observations around hours worked conditional on treatment.