



האוניברסיטה העברית בירושלים  
THE HEBREW UNIVERSITY OF JERUSALEM

# Do Reemployment Bonus Programs have Externalities? Evidence from a Randomized Experiment

*A thesis submitted in partial fulfillment of the requirements for the M.A. degree at the  
Department of Economics, The Hebrew University of Jerusalem*

*By*

*Gilad Gaibel (302549977)*

*Advisers: Dr. Itay Saporta-Eksten and Dr. Analia Schlosser*

*December 2019*

## **Abstract**

Using data of a large reemployment bonus program in Israel (2015-2018), I exploit random variation in the intensity of the intervention across locally defined labor markets to study its impact on employment of individuals within these markets. Such effect manifests an externality of the program, and a potential violation of the stable unit treatment value assumption (Rubin 1990). I use a two-stage least squares (2SLS) interaction model to estimate the effects. Overall, the results are inconclusive and the confidence intervals are wide, probably due to insufficient statistical power. Future research, using more accurate data and various outcome variables may provide more accurate estimates, and more reliable assessments for either there are externalities for this intervention or not.

# Contents

<b>1. Introduction</b>	<b>3</b>
<b>2. Background</b>	<b>5</b>
2.1. Program Description . . . . .	5
2.2. Reemployment Bonus Programs - Literature Review . . . . .	7
2.3. Externalities in ALMPs - Theory . . . . .	9
2.4. Externalities in ALMPs - Literature Review . . . . .	12
<b>3. Data and Descriptive Statistics</b>	<b>15</b>
<b>4. Empirical Strategy</b>	<b>17</b>
4.1. The Definition of Local Markets . . . . .	18
4.2. Outcome Variable . . . . .	20
4.3. General Structure of the Empirical Model . . . . .	20
4.4. The Intensity of the Intervention . . . . .	21
4.4.1. Introducing $Q$ . . . . .	22
4.4.2. Endogeneity of $Q$ . . . . .	23
4.4.3. The relationship Between $Q$ and $P$ . . . . .	25
4.4.4. Variation Over Time . . . . .	25
4.4.5. Empirical Explorations of the $Q$ Measurement . . . . .	27
4.5. Power . . . . .	29
4.6. Final Empirical Design . . . . .	30
<b>5. Estimation Results</b>	<b>31</b>
5.1. Baseline Results - Locality-Based Markets . . . . .	31
5.2. Locality-Based Markets - Heterogeneity Analysis . . . . .	33
5.3. Proximity-Based Markets . . . . .	35
<b>6. Conclusion</b>	<b>38</b>
<b>Bibliography</b>	<b>39</b>
<b>Appendix</b>	<b>43</b>

# 1. Introduction

In the past few decades ALMPs have been the grounds for a large body of economic research about labor market behaviors of various agents (for a recent survey see Crépon and van der Berg 2016. For meta-analyses see: Kluve 2006; Card *et al.* 2010, 2015). ALMPs are typically aimed at improving the labor market outcomes of unemployed. These programs include subsidized work, training programs of different kinds and job search assistance. In order to obtain precise and reliable measures of effectiveness, a considerable amount of ALMPs were evaluated by randomized controlled trials (RCTs). This was motivated, in part, by an influential paper by LaLonde (1986), which demonstrated the difficulty to reproduce experimental training program estimates using observational data. The typical evaluation uses individual-level data to measure the effect of a program on the participants' labor market outcomes. These evaluations usually compare means of 'treatment' and 'control' groups to calculate treatment-effect (TE) or intent-to-treat (ITT) estimates of the effect of the interventions.

The means-comparison method is based upon two critical assumptions - both are related to independence of the assignment to treatment (Rubin 1990). In the most common case of a binary intervention, the first assumption is that the measured outcomes of a person who was assigned to treatment (control) and did not (did) receive the treatment are the same as if she was initially assigned to control (treatment). This assumption is satisfied by construction in an RCT design where treatment is randomly assigned. The second assumption is that the outcomes of an individual who was assigned to treatment or control were the same if others were assigned to either treatment or control in any way. This assumption is called the Stable Unit Treatment Value Assumption (SUTVA). If, for example, a measured outcome of a person who was assigned to control would have been different if there was no program at all, SUTVA is violated and the means-comparison estimates would be biased.

In the context of ALMPs a violation of SUTVA is likely to occur and it is of major importance for policy-design (Calmfors 1994; Heckman *et al.* 1999; Abbring and Heckman 2007). An evaluation of an ALMP is usually meant to yield reliable recommendations regarding the intervention in a large scale implementation and suggestions for optimal final design. If SUTVA is violated and the evaluation estimates are biased, then the inferred conclusions may not only lead to a waste of public funds, but may also cause harm to the very population it aims at aiding. In ALMPs, it is usually the case that a program in the evaluation stage does not cover the entire target population, but only a fraction of it. However, the proportion of those assigned to treatment out of a local population may reflect the intensity of the intervention in the local level (Crépon *et al.* 2013; Ferracci *et al.* 2014). Hence, while an evaluation may provide reliable results for a certain local intensity level, it may not provide reliable results for the counterfactual situation in which the entire target population actively participates in the program. The main concern here is that the local intensity level may affect the individual-level outcomes which are used to evaluate the program. A simple example for such effect can be a change in equilibrium wage in the local market as a result from the local intensity level, which

corrupts any difference-in-wages estimator. The presence of such externalities manifests a violation of SUTVA. It is important to note that the forms of externalities I consider here are only within labor market framework, since these corrupt program evaluations directly. Externalities of other forms are left out of the scope of this paper, although they also worth a serious discussion (see: Björklund *et al.* 2005; Schochet *et al.* 2006).

The possibility of externalities in ALMPs has been recognized long ago (Woodbury and Spiegelman 1987; Davidson and Woodbury 1993; Heckman *et al.* 1998), however there are still only a handful of empirical studies who actually test for the presence of externalities in ALMPs or try to quantify their magnitude (Davidson and Woodbury 1993; Blundell *et al.* 2003; Lise *et al.* 2004; Albrecht *et al.* 2009; Crépon *et al.* 2013; Ferracci *et al.* 2014; Gautier *et al.* 2018). My research follows this strand in the ALMPs literature and makes three contributions to it. First, I estimate a reduced form model in a rather large program in Israel using exogenous variation in treatment assignment in order to identify externalities. Second, the program that is in the focus of this paper gives a unique opportunity to investigate positive externalities, which were not considered by any past research in the field. Third, I use data of a reemployment bonus program, which a reduced-form identification of externalities in such a program has not been performed so far.

The program that is in the focus of this paper, the Remote Work Grant (RWG), is a reemployment bonus program. The program targets job seekers residing in the poorer peripheral areas in Israel. Eligibility is conditional on finding a job outside the locality of residence within three months of the first notice, and on reporting of at least eleven work days in that job at each month. Eligible participants are granted with a monthly amount of about 600 NIS (~11% of the minimum wage in Israel) for up to five months. The RWG is aimed at removing light-barriers with respect to commuting costs from relatively remote job offers, in order to expand the opportunities horizon of the program's population. The evaluation of the effectiveness of this program is carried out by an RCT.

In this research I investigate whether the RWG has externalities or not, and if there are, of which kind. In order to do so, I construct a measure for the intensity of the intervention in a local level and characterize the necessary requirements for *causal* inference. Specifically, I use an IV interaction model to solve the problem of endogeneity. I then estimate the model using two different approaches, regarding the way local labor markets are being defined. Overall, the results are inconclusive and the confidence intervals are wide. Future research, using more accurate data and various outcome variables may provide more accurate estimates, and more reliable assessments for either there are externalities for this intervention or not.

The remainder of this paper is organized as follows. In Section 2 I provide background about the RWG, about externalities in ALMPs, and review the relevant existing economic literature. In section 3 I describe the data and provide summary statistics. In section 4 I present my empirical strategy, and discuss the main identification challenges and other challenges in the estimation procedure. In section 5 I present and discuss the estimation results. Section 6 concludes.

## 2. Background

### 2.1. Program Description

In the following subsection, I provide relevant details about the RWG setting and context, the design of the program and the experimental design. Then, I present some initial results from the midterm report of the program (Gershoni *et al.* 2018).

Although Israel is quite a small country, there is a great deal of social and political interest in differences between the center and the periphery. The geographic distance, which in terms of other countries may be considered little, is accompanied by differences in important socioeconomic factors of the residing population. In the past, Israeli policy was to refer new immigrants to peripheral localities, mostly in the Negev and the Galilee, either in the name of Zionist ideology or in the name of strategic defensive importance against foreign invasions. This created a situation in which peripheral localities were usually inhabited by lower-means population. In addition, low quality infrastructure and low governmental investment in education, health and transportation in the periphery over the years, deepened the socioeconomic gaps. Since the early 1990's many peripheral localities are defined as "National Priority Areas" and are entitled to a set of benefits, for example, in taxation. The aim of such policies is to reduce the existing socioeconomic gaps.

The RWG is intended to compensate job seekers who suffer from high commuting costs to a relevant occupational center in terms of money and time, and to encourage them to work. Since peripheral localities are usually with low accessibility and bad transportation infrastructure, they were a natural target for the program. The RWG is focused on job seekers residing in the poorer localities in the periphery (localities that are both 4 or below in the Israeli socioeconomic status grade, and 5 or below in the Israeli peripherality index). In addition, the program is focused on young and middle-aged workers (18-55): the older workers are assumed to be less mobile and therefore less responsive to an intervention.

The target population is divided into two different groups: UI claimants and Income Assistance (IA) claimants. UI claimants are usually newly unemployed who are entitled to UI benefits, and they must report each week to their local employment office in order to receive the benefits (they may claim UI benefits for 3-6 months of unemployment since their declaration date). IA claimants are people with very low family income or without family income. In most cases, IA claimants must also report each week to their local employment office in order to receive the benefits. However, if an IA claimant works full time (sometimes even part time) she may be exempted from reporting to her local employment office. Typically, UI claimants and IA claimants are quite different in their socioeconomic and sociodemographic characteristics, such that on average IA population is weaker than UI population.<sup>1</sup>

After a qualification period of one month as a registered job seeker, one who participates in the program

---

<sup>1</sup>In this paper I only focus on IA claimants. I discuss this point later, in subsection 4.2.

and has been allocated to treatment receives a notification about the grant and, if eligible, may claim it. Eligibility is conditional on finding a job outside the locality of residence within three months of the first notice, and on reporting of at least eleven work days in that job at each month. Eligible participants are granted with a monthly amount of 600 NIS (about 11% of the minimum wage) for up to 5 months.

The Israeli Employment Service (IES) initiated the RWG in November 2015. At first, the RWG acted as a pilot within two employment offices. Since November 2016 the program gradually expanded to other employment offices, until it reached a total of 26 employment offices around the country in February 2017.

The RWG is being evaluated by a clustered RCT. Each employment office is responsible for a predefined set of localities, and every job seeker must report to the office which corresponds to her locality of residence. Every month (usually), the employment service forms a list of potential job seekers who answers the eligibility criteria and are divided into three pools: UI claimants, newly registered IA recipients (IA flow), and IA recipients with seniority of more than six months in unemployment (IA stock). A sample out of the listed job seekers is chosen at random out of all qualified job seekers within each local office, such that the amounts of those listed within each pool corresponds to a certain budget constrained quotas. Given these lists, the allocation to treatment is randomly determined within each employment office and within each pool. Therefore, the randomization is clustered at the pool, employment office and month of allocation level (the intersection of these three - month, employment office, pool - will be addressed as “allocation-unit” henceforth). Those who were allocated to treatment receive the relevant information about the program and may claim the RWG grant if eligible. Those who were allocated to control are handled in the offices exactly as they would have been without the program.

Data about the participants in the experiment is gathered in two channels - administrative data and voluntary surveys. The administrative data is from both the IES and the National Insurance Institute (NII).<sup>2</sup> The surveys are conducted in three waves and they only cover a sample of the participants: the baseline survey is conducted before the allocation to treatment, the follow-up survey is conducted about three months after the allocation date, and the second follow-up is conducted about a year after the allocation date. Further information regarding the data, which is in the scope of this paper, is provided in section 3.

Using the data that was available at the time, both from the baseline and the follow-up surveys, combined with the relevant administrative data, the midterm report presented some preliminary findings of the effect of the program (Gershoni *et al.* 2018). It has been found that the program led to a 3 percentage points decrease (significant at the 1% level) in reports to the local offices among the treatment group, and to a decrease of 0.08 months (significant at the 1% level) in cumulative reports to the local offices during the first five months since allocation. Assuming that differences between treatment and control in average amount of absences from the local employment office are a good enough proxy for differences in employment status, these figures

---

<sup>2</sup>At the time of the analysis, NII data was still unavailable.

represent positive results.<sup>3</sup> It was also found that the most responsive pool during the first five months is IA stock, and that their response was the main drive of the results. These results are also consistent qualitatively with the results of the treatment effects estimation during the first eight months since allocation. Preliminary results from the surveys indicated that ten weeks after the allocation date, there was a significant increase in motivation and availability to work outside the locality of residence among the treatment group. These findings were based on comparison of means of the treatment and control groups within allocation-units.

If externalities play a significant role in the RWG, the difference-in-means estimators are biased due to a violation of SUTVA. There is a basis to suspect that externalities do play a role in the program. A thorough discussion regarding this issue appears in subsection 2.3. However, before I address this issue, I first review the existing literature about reemployment bonus programs (subsection 2.2) in order to provide some more context.

## 2.2. Reemployment Bonus Programs - Literature Review

Reemployment bonus programs are thought of as a tool to stimulate job search without reducing the existing unemployment-insurance (UI) level (Chetty 2008). The bonus raises the cost of leisure while keeping the existing ‘safety-net’ of UI unchanged. Evaluations of reemployment bonus programs mostly found that claimants tend to exit unemployment faster, and that the save in UI benefits as a result of the shorter unemployment spells is greater than the total sum of bonuses paid (Crépon and van der Berg 2016).

Previous studies of bonus programs were mostly based on RCTs. Woodbury and Spiegelman (1987) evaluate the first experiment, conducted in US Illinois in 1984. Their main research question was: does monetary incentives affect job search activity? They found that the elasticity of search spell duration relative to bonus transfers was negative and significant; specifically, they estimated an average difference of 1.15 weeks in unemployment spells between treatment and control. An additional finding is that the faster return to employment had no significant effect on post-program wages, i.e. it had no effect on the job quality of claimants. Following the Illinois experiment, four others were conducted in the US and were meant to provide additional information regarding design and targeting. Decker (1994) found that a scheme of a declining bonus over time had little effect on those who were likely to remain unemployed for a rather long period. Decker and O’Leary (1995) found that the reaction of job seekers to the bonus is stronger as the bonus increases, however the overall save in UI benefits was lower than the necessary expenses for the administration and payments of the bonuses. Meyer (1995) argues that a considerable limitation of the US experiments is that they cannot account for the expected effects in a large scale and permanent implementation of a bonus program. He describes mechanisms that could produce negative results. First, such program would incentivize UI eligible workers to file a claim or delay their return to employment, while absent the program

---

<sup>3</sup>The reason behind the idea that differences in absences may serve as a proxy for differences in employment is discussed in detail in subsection 4.2.

they would not have done so. Second, firms with a high probability of lay-off would be considered less risky, thus there would be an increase in their potential share of workers and, in turn, there could be an increase in the unemployment level.

Another RCT has been conducted in Canada in the early 1990s. The Self Sufficiency Project (SSP) was designed to incentivize full-time and long lasting employment. It was targeted at single parents that have been unemployed for at least a year, and provided to those who worked full-time a significant bonus for up to three years. Card and Hyslop (2005) evaluated the effectivity of the program for long-term welfare recipients. They focused on the dynamics of the effect. They found that the program had a strong effect at first, but it gradually faded; at a four-years horizon the program had no significant effect on welfare participation. However, during the first three years of the program, Michalopoulos *et al.* (2005) found a significant increase in the wage of claimants and a decrease in poverty.

Van Der Klaauw and Van Ours (2013) studied a reform in Rotterdam (1997-2002) which introduced a reemployment bonus to those who were unemployed for at least a year and visited regularly at the employment offices. With contrast to previous research they had non-experimental data. They used a difference-in-difference estimation strategy to evaluate the effect of the bonus. Their result was that the bonus had no significant effect whatsoever. They explain the inconsistency of their conclusions with previous results by differences in the characteristics of the populations, specifically, differences in the impatiens of the claimants. Of course, it is also possible that experimental data would have yielded different results.

Another non-experimental study by Ahn (2018) evaluated a nationwide bonus policy in Korea. A reform offered an increased bonus for individuals above the age 55. Ahn used a sharp regression discontinuity design estimation to estimate the local effect of the bonus. They estimate an average reduction of 0.7-1.8 weeks in unemployment among the eligible workers. The study also reports no significant change in job quality (wages), and that the program was cost-effective.

Davidson and Woodbury (1993) were the only ones to study externalities in reemployment bonus programs. They calibrated a partial-equilibrium matching model that would allow them to quantify externalities based on the Illinois experimental data. They were specifically interested in the possible displacement effect in which a job has been filled by a bonus-offered worker on the expense of a non-offered worker. This kind of displacement would manifest a violation of SUTVA and may suggest that the results of the US experiments were biased upwards. Their calibrated model predicted small displacement effects, nonetheless these were responsible for 30%-60% of the gross difference in employment between the UI-eligible and UI-ineligible workers. They concluded that the given program design was a rather unattractive policy alternative. In my research I contribute to this literature by estimating a reduced form model with the purpose of identifying externalities in a large reemployment bonus program.



### 2.3. Externalities in ALMPs - Theory

In this subsection I illustrate and discuss the structure of externalities that will be considered in this paper. For the purpose of clarity, I separate the theoretical discussion from the literature review of externalities in ALMPs (subsection 2.4), but the theoretical discussion will, of course, include some summary of previous studies. In addition, I focus the discussion here on ALMPs, however it can be easily applied to other experimental settings. Some applications from other fields will be reviewed at the end of subsection 2.4.

Let  $T_i$  be an individual binary treatment status indicator, and let  $Y_i$  be an outcome of interest. Following the potential outcomes framework, denote by  $Y_{i1}, Y_{i0}$  the potential outcomes of individual  $i$  when treated and non-treated accordingly (Rubin 1978). Assume that the economy is segmented into  $M$  local labor markets and that the treatment is geographically dispersed. Also assume that each individual  $i$  acts in only one market  $m \in M$ .<sup>4</sup> Let  $P_m$  denote the local saturation levels of the intervention. That is, define:

$$P_m = \frac{\sum_{i \in m} T_i}{|\{i \mid i \in m\}|} = \frac{N_m^T}{N_m^T + N_m^{NT}} = \frac{N_m^T}{N_m}$$

Where  $N_m^T, N_m^{NT}$  are the amounts of those who recieved treatment and of those who did not recieve treatment in a particular market (note that  $N_m^{NT}$  typically includes more than the control group in a real setting). Given  $P_m$ , the average treatment effect of participation in the program is defined by:

$$\Delta(P_m) = E(Y_{i1}|P_m) - E(Y_{i0}|P_m)$$

In an RCT setting, allocation to treatment is random. This implies:

$$\text{Random Assignment} : \{Y_{i1}, Y_{i0}\} \perp T_i$$

A consistent estimate of the ATE also requires the stable unit treatment value condition:

$$\text{SUTVA} : \{Y_{i1}, Y_{i0}\} \perp T_j \quad \forall j \neq i$$

$$\Rightarrow \{Y_{i1}, Y_{i0}\} \perp P_m$$

If both assumptions hold,  $\Delta$  is constant and equals to the simple difference in means:

---

<sup>4</sup>This assumption is called in the literature “regional SUTVA” or “partial interference” (Sopel 2007; Ferracci *et al.* 2014). Manning and Petrongolo (2017) study the relations between local labor-markets in a structural framework. Their results suggest that local markets usually do overlap, and that place-based policies evaluations tend to define too small local markets. I adress this issue later, in subsection 4.2.

$$E(Y_i|T_i = 1) - E(Y_i|T_i = 0)$$

The sample analogue of this expression is simply the difference in average outcomes of the treatment and the control groups.

If, however, SUTVA is violated, then  $\Delta$  becomes a function of  $P_m$ . Therefore, simple difference-in-means estimates will not be of major interest. Estimation which takes variations in  $P_m$  into account may resolve this issue, however, it is harder to perform, especially if the experiment was not initially designed to capture such variation (Angelucci and Di Maro 2016; Huber and Steinmayr 2017; Baird *et al.* 2018). From the policy-maker point of view, the most informative estimate will be of the difference  $\Delta(P_m = p) - \Delta(P_m = 0)$ , which captures the ATE in a situation where the program is implemented such that it is limited to only a share  $p$  of the population. To be more concrete, if the policy agenda is a complete rollout of the program, the relevant estimates will require a comparison of full saturation and zero saturation:  $\Delta(P_m = 1) - \Delta(P_m = 0)$ . Although complete rollout is a common policy objective in ALMPs, to my knowledge, there has not been any experiment that actually estimated this difference using a reduced form approach. In the next subsection I will survey studies which quantified this object using a structural approach. Another important consideration from the policy-maker point of view is the ability to perform social cost-benefit analysis, which critically depends on correct general equilibrium measurements (Calmfors 1994).

Intuitively, variation in  $P_m$  reflects a variation in the intensity of the treatment in the local level. A program is said to have externalities if the level of local intensity affects the local population. This is, of course, an externality since it is not a direct implication of one's treatment status. Consider, for example, a job-search assistance program. The share of individuals allocated to treatment out of the local population may affect the probability of an individual (irrespective of her treatment status) to find a job because the competition on local vacancies becomes fiercer. In a small scale experiment, externalities of this form are usually unlikely to occur, however in a large experiment it is more probable.

In the ALMPs literature, there are two main forms of externalities that have been considered so far. The *displacement effect* describes a situation in which the intensified job-search or the increase in motivation to find a job among the treatment group causes an increase in the probability that those who were allocated to treatment will fill a vacancy, while the probability that others will fill a vacancy decreases (e.g. Calmfors 1994; Blundell *et al.* 2003; Crépon *et al.* 2013). *General equilibrium effects* describe a set of situations in which the equilibrium prices and the supply and demand curves change as a result of an intervention (Angelucci and Di Maro 2016). The most obvious example of such effect is a decrease in equilibrium wage as a result of a shift in the supply curve. Another example is a change in available vacancies as a result of an intervention (Gautier *et al.* 2018).

Other forms of externalities have not received much attention in the context of ALMPs, but are well known in

other experimental contexts. *Social interactions effects* or *peer effects* describe situations in which a treatment affects not only the individual who received it, but also her surrounding social environment (Manski 1993). Such channels were previously investigated in the general context of labor markets (e.g. Bayer *et al.* 2008; Hellerstein *et al.* 2011), and they do seem relevant in ALMPs as well. Consider, for example, an ALMP and a small local market with a high value of  $P_m$ . If those who were allocated to treatment encounter more opportunities, they may share these opportunities with their non-treated neighbors, by referring them to jobs or by giving them access to new information. A broader *norm effect* may arise if local norms with respect to labor market decisions change as a result of the intervention (Eugster *et al.* 2017; for norm effects in the context of medicinal choices see: Avitabile 2012); for example, if local norms of working hours or commuting time changes, or if a local norm regarding the proper reservation wage changes. It is important to note that the forms of externalities I consider here are only within labor market framework, since these corrupt program evaluations directly. Externalities of other forms are left out of the scope of this paper, although they also worth a serious discussion (Björklund *et al.* (2005) study the effect of a Swedish large scale program on the quality of regular education; Schochet *et al.* (2006) study the effect of a large American training program on criminal activities).

It is not *a priori* clear whether displacement effects and general equilibrium effects are expected to be important in the context of the RWG. Landais *et al.* (2013) discuss optimal design of UI insurance schedule, based on a partial job-search model. They show that the generosity of the UI benefits is inversely related to the tightness coefficient in a market where firms have a concave production function (a rather realistic assumption). Within their model there exists an optimal tightness ratio, and so they argue that in a slack market it is arguably better to reduce or cancel the transfers in order to avoid a further decrease in the equilibrium tightness. Crépon *et al.* (2013) analyze a similar model, adapted to a job-search program, and conclude that since the tightness ratio is positively related to the probability of finding a job, in a slack market, displacement effects and general equilibrium effects are expected to be important. Intuitively, a program that operates in a market where employment is mostly constrained by demand is not expected to produce more vacancies, but to change the balance of the competition on existing jobs. Hence, in the RWG, which operates in remote areas with relatively low amount of vacancies, such externalities may be important. However, Since the main incentive at play is encouragement of those allocated to treatment to work outside their usual geographical bounds of labor supply, displacement effects may seem less likely to occur. Incentivizing the treatment group to prefer distant jobs (presumably in areas with high labor demand) over local jobs may not crowd out the control group from either local opportunities nor distant opportunities (assuming the further one is willing to work, the more opportunities she is likely to have). On the contrary, decrease in competition on local vacancies can make it even easier on job seekers to get a local job, at least in the short run. Hence, it is not obvious if these effects are expected to be important, nor it is obvious what their sign is expected to be.

Social interaction effects may seem probable in the RWG. The common participant in the program is from a low-skilled background and belongs to a rather small community. In addition, many of the participants are non-Jews (for descriptive statistics see table 3), i.e. they belong to minorities. Findings from economic research about social networks and neighborhood effects are that, with respect to labor market outcomes, these are more likely to be important in minorities, and in low-skilled cohesive communities (Hellerstein *et al.* 2011; Topa and Zenou 2015). Therefore, it seems probable to expect that such externalities may be at play in the context of the RWG.

There are several empirical challenges in estimating externalities in ALMPs using a reduced form approach. A first necessary assumption is regional SUTVA, which means that there are no interferences between individuals from different markets (Sopel 2007; Huber and Steinmayr 2017; Baird *et al.* 2018). In addition, exogeneity of an intervention is the luxury of an RCT design, however it is not always guaranteed, and measurements of the local intensity of the intervention are often exposed to problems of omitted variable bias, therefore unconfoundedness is a classical obstacle. Assuming unconfoundedness of both variables and identification of the local labor markets, measuring the effect of a program on the non-claimants requires either a ‘pure-control’ group (zero local treatment saturation) or enough variation in treatment intensity which can be used for identification. Furthermore, typically, the effect on the non-claimants is slim, therefore insufficient amount of observations to guarantee the desired power for the statistical tests can be a major obstacle (Crépon and van der Berg 2016). Finally, extrapolation of the results of a small-scale program evaluation may not be convincing, even if externalities are taken into account (Heckman *et al.* 1998). As mentioned above, if the results of an evaluation depend on the local intensity of the treatment, usually, they are not expected to correspond to the effect of a full large-scale implementation.

## 2.4. Externalities in ALMPs - Literature Review

Nowadays, it is widely known in the social sciences and medicinal disciplines that evaluating experimental data without accounting for externalities can be misleading and yield problematic conclusions (Deaton and Cartwright 2018). In ALMPs, the possibility of externalities and their implications with respect to the credibility of the results of evaluations have been recognized long ago (e.g. Woodbury and Spiegelman 1987; Davidson and Woodbury 1993). However, there are still only a few empirical studies who actually test for the presence of externalities or try to quantify their magnitude in this field.

The earlier studies quantified externalities using a structural approach. Davidson and Woodbury (1993) analyze externalities using data from the Illinois 1984 reemployment bonus experiment. They use a partial-equilibrium variant of a Mortensen-Pissarides search-and-matching model with fixed prices to quantify the magnitude of the externalities in a large-scale implementation of the program. They found that 30%-60% of the gross employment effect is due to displacement of UI-ineligible job seekers. Although the absolute

numbers of displaced workers are rather small, they conclude that an adoption of a reemployment bonus program would not yield a Pareto improvement. In addition, they raise the ethical issue that it may be difficult to justify the displacement of UI-ineligible workers by UI-eligible workers.

Blundell *et al.* (2003) use a structural approach to address externalities in the “New Deal for the Young Unemployed,” designed to move young unemployed individuals in the UK. They calibrate a dynamic model using the data of the program, and show that the measured treatment effect can change signs when general equilibrium effects are taken into account.

Lise *et al.* (2004) adopted the model of Davidson and Woodbury (1993) and reevaluated the Canadian SSP. They first calibrated their model using the control group data, then they showed that they were able to predict the outcomes of the treatment group well. Finally, they recalibrated their model using the entire Canadian IA claimant population, and estimated the full large scale impact of the program, taking general equilibrium effects into account. Their main relevant result is that the predicted general equilibrium effects reverse the cost-benefit conclusions of the partial equilibrium evaluation.

Albrecht *et al.* (2009) also used a structural approach. They calibrated a general equilibrium model using the Swedish “Knowledge-Lift” (1997-2002) nationwide skill-upgrade program for adults. Their simulations results indicated that there are quite strong and heterogeneous externalities: those who were low-skilled prior to the program and remained low-skilled suffered negative externalities in terms of their wage and probability of employment, while those who were medium-skilled prior to the program and remained medium-skilled experienced substantial positive externalities.

The later studies quantified externalities using a reduced form approach or a combination of reduced form and structural estimation. Naturally, these studies are more preoccupied with identification problems and data limitations. These studies consider the impact of both receiving treatment and of the local intensity of the treatment on the desired labor market outcomes. Based on the theoretical considerations mentioned above, exogeneity of both treatment assignment and of the local intensity of the treatment is required in order to identify externalities. Another identifying assumption is that the potential outcomes of individuals within a specific locality are independent from treatment assignment of individuals in a different locality (“regional-SUTVA”: Söpel 2007; Huber and Steinmayr 2017; Baird *et al.* 2018).

Blundell *et al.* (2004) use a matching difference-in-difference estimation to evaluate the “New Deal” in a reduced-form model. The program has begun as a pilot, then, after a while, has been implemented nation-wide. Hence, they exploit variations in the timing of implementation of the program across areas, and individual-specific eligibility criteria. They do not find a significant effect of the program on ineligible job seekers in areas treated by the program, compared to those in areas untreated by the program.

Crépon *et al.* (2013) evaluated a French job-search assistance program aimed at young university dropouts who had at least 6 months of seniority in unemployment. They used a two-step randomization design: first

they identified relevant, supposedly self-contained, local labor markets and randomly assigned each of them a treatment saturation level of 0%, 25%, 50%, 75% or 100%; next, they randomly assigned individuals to either treatment or control within each market according to its saturation value. This design satisfied all three identifying assumptions, thus theoretically justifying casual inference. The measured effects were of positive treatment effect on the treated and a statistically insignificant negative effect on the non-treated. A statistically significant and negative result has been found for the effect of the program on non-treated males.

It is important to note that the rather slim and insignificant estimates of Crépon *et al.* (2013) did not necessarily capture the ‘true effects’, but may have been the result of two technical factors. First, the high attrition rates may have led to smaller usable sample than necessary. Second, low representation of the participants in the program among the population of eligible workers within each market led to effectively low variance in saturation levels across markets. These two factors may have limited the statistical power of the performed tests, and can explain the insignificant estimates.

With contrast to the ‘clean’ identification in the study of Crépon *et al.* (2013), another study used a quite similar identification approach without the randomized saturation design. Ferracci *et al.* (2014) use a two-step matching method in order to measure externalities in a French training program. In the first step, after identifying the relevant markets as an occupation within region in a given time, they estimate the difference-in-means within each market. In the second step, they estimate the relation between the empirical local treatment saturation and the measured difference-in-means. They find strong evidence for the presence of externalities. However, the identification of each of the two steps depends heavily on a conditional independence assumption, which may seem unconvincing within context, thus weakening the credibility of their results.

Gautier *et al.* (2018) evaluated a Danish job search assistance program. Their study is comprised of two sections: (1) reduced form assessment of the effects of the program, accounting for externalities, and (2) a structural estimation of the effects of the program in a full large scale implementation. For their reduced form estimation, they use a difference-in-difference estimation strategy, comparing treated to non-treated individuals and non-treated individuals in participating regions to those in non-participating regions. They find null effect of the program on post-unemployment wages for both treated and non-treated individuals. They found a slight increase in vacancies in the participating regions, and a statistically significant negative effect of the program on unemployment duration among the non-treated (displacement). The simulated results, using the structural model, indicated a reduction in total welfare due to increased unemployment, increased taxation, and increased search costs for the unemployed.

Treatment externalities have received attention in other fields of research as well. Using a structural model, Heckman *et al.* (1998) demonstrated that subsidized college tuition fee has substantial effects on enrollment rates. Duflo and Saez (2003) found substantial peer-effects in a pension-saving program offered to American university employees. Miguel and Kremer (2004) study the spillover effect of deworming drugs on health and

school attainment of children in Kenya. Avitabile (2012) study spillover effects in cervical cancer screening in rural Mexico.

Finally, the economic literature about externalities in ALMPs is still quite thin. ALMPs are comprised of a rather different programs, and each specific design is very sensitive to the characteristics and environment of its target population. As it is apparent that this specificity is important with respect to the partial-equilibrium evaluations (simple difference in means), it should be recognized that this specificity is also important with respect to the different kinds of externalities. Generally speaking, it has been shown that externalities tend to be rather slim and therefore hard to detect, however they may still be economically significant and must be taken into account, especially when considering a large scale implementation of a program. This paper follows this strand of literature and attempt to test for the presence of externalities within a large Israeli reemployment bonus program.

### 3. Data and Descriptive Statistics

The data that is used for the analysis in this paper is comprised of four datasets. The main dataset is administrative IES individual-level data about the participants of the RWG. Additional locality-level variables are based on IES public municipality-level data, and Central Bureau of Statistics (CBS) public municipality-level and locality-level data.

The individual-level data contains data about participants (both treatment and control) in the RWG that have been allocated since November 2015 until January 2019. The relevant variables for the analysis that follows belong to one of three groups: (1) allocation information - treatment status indicator and allocation-unit information (date of allocation, pool and original employment office); (2) background covariates - gender, ethnic origin, family status, age, locality of residence, index for participation in the “Employment Circles” program etc.<sup>5</sup> (3) reports to the local employment office - monthly indicators for appearance in the local employment office since the month of allocation. Originally, there were 53269 available observations. Some of the individuals that are tracked in the data last reported to their local employment office before their allocation date, and so they could not have known about the program or been affected by their treatment status, therefore I drop 4698 such observations.

My analysis devides into two stages. First, I calculate some measure of the intensity of the intervention at the market level (in the baseine model a markets correspond to localities) using all the available observations. Second, I restrict the sample to IA claimants only with seniority of at least twelve months in the program and perform the estimations on this sample alone. There are two reasons for restricting the data in the second stage: first, my framework is based upon an analysis of the behavior of RWG participants over a time window

---

<sup>5</sup>The IES operates the “Employment Circles” program since 2014. This program is aimed at aiding IA claimants to find jobs by means of intensified meetings in the local employment offices, job-search assistance, personal consulting and more. Participation in the RWG is independent from participation in the Employment Circles program.

of twelve months, therefore observations with a shorter seniority are irrelevant at the second stage; second, the outcome variable that I use is actually a proxy for employment status, and the sense in which this outcome may serve as a proxy is less applicable for UI claimants (see subsection 4.2). Hence, after performing the calculation of the intensity of the intervention for each market, I drop 9482 observations with seniority of less than a year in the program, and drop 17233 observations of UI claimants. I also drop additional 3123 with problematic locality of residence identifier (belong to localities that do not belong to any CBS municipality code - these are almost entirely residents of the Bedouin settlements). Additional 2 observations with missing values in individual-level background characteristics are dropped. 60 observations have missing work-force characteristics values. I replace the missing data in these variables with within-municipality averages when possible, and for the rest, with the global averages. Finally, the sample I use for the estimations contains 18731 observations of IA claimants who participated in the RWG for at least 12 months.

The three additional data sets contain municipality-level and locality-level information. The IES data contain time series of municipality-level amounts of registered job seekers and unemployment rates, which I use to calculate the size of the municipality-level labor force. CBS municipalities data contain various information on municipalities from the years 2015-2017, and I use, for example, the ages distribution among municipalities. CBS localities data contain the locality sizes (population) in 2017. I then map all localities into the list of municipalities in order to transform municipal-level numerical variables into locality level variables, based on the size of the population of each locality relative to the size of the population of its corresponding municipality. Such transformation require a within-municipality uniformity assumption, which in the absence of a more accurate data, I use as a working assumption. Table 1 presents descriptive statistics of the final sample.

The table reports summary statistics of the background variables and the outcome variable that are used for the estimations in the full sample (second stage) of IA beneficiaries in the RWG. The average cumulative absences indicate that on 30% of the first year since allocation the IA population is employed. Also, there is over representation of females (66%) and Arabs (80%) in the sample, very low educational background, and there is a considerable amount of health-limited individuals.

As described above, the RWG was implemented as an RCT. Table 2 reports balance tests for the allocation to treatment and control, and some additional descriptive statistics. The balance tests are performed by estimating a regression of a specified covariate on the treatment status indicator, controlling for allocation unit fixed-effects. The standard errors in this table are clustered at the allocation-unit level. The results indicate that the treatment and control groups are approximately balanced. Although there are some statistically significant differences (at the 5% and 10% level), the magnitude of the estimates is small. The P-value for a joint significance test of all of the covariates against the treatment indicator  $T$ , indicate that the joint association between the covariates and  $T$  is statistically insignificant. This means that the random assignment to treatment did create comparable groups.



Table 1: Summary Statistics

Covariates	Mean	Stdev	Median	Minimum	Maximum
<b>Individual Level Characteristics:</b>					
Cumulative Absences (12 months)	3.579	3.988	2.000	0.000	12.000
Female	0.656	0.475	1.000	0.000	1.000
Arab	0.797	0.403	1.000	0.000	1.000
Ethiopian	0.003	0.057	0.000	0.000	1.000
Married	0.537	0.499	1.000	0.000	1.000
No. of Children	3.340	3.137	3.000	0.000	27.000
Immigrant	0.066	0.248	0.000	0.000	1.000
Health limitation	0.565	0.496	1.000	0.000	1.000
Age	43.051	10.062	45.000	19.000	67.000
Single parent	0.146	0.353	0.000	0.000	1.000
Ultraorthodox	0.030	0.170	0.000	0.000	1.000
Belongs to 'Employment Circles' pop.	0.211	0.408	0.000	0.000	1.000
Matriculation	0.069	0.253	0.000	0.000	1.000
<b>Locality Level Characteristics:</b>					
Number of RWG Participants	198.941	399.033	25.000	1.000	2897.000
Average Unemployment Rate	0.062	0.061	0.039	0.009	0.347
Average Size of Labor Force	9831.661	31443.092	1657.875	36.800	303904.400
Total Pop. (2017)	21039.539	74522.733	3146.000	50.000	901302.000
Pop. in eligibility ages (2017)	10884.489	37724.993	1713.000	27.000	435314.000
Socioeconomic Index (CBS, 2015)	-0.333	0.932	-0.251	-3.110	1.649

Observations: 18731

Localities: 219

*Note:* This table reports raw summary statistics for IA claimants in the RWG and for the local markets (localities) they participate in. Individuals who last reported to thier local employment office before their allocation date are excluded from the sample, as they could not have been influenced by their treatment status. Work force characteristics are computed here as the average over the values which correspond to one's allocation date within each locality.

## 4. Empirical Strategy

In the following section I discuss my estimation framework, my main variables and the identification assumptions in detail. I begin with the definition of local markets. Next, I present the outcome variable and describe a general structure of an externalities-model, which I then adapt for identification of externalities in the RWG, given my data limitations. The main adaptation is of the measure of the intensity of the

Table 2: Balance Tests

Covariates	Treatment	Control	Estimate	P-val.
Female	0.651	0.661	0.007	0.373
Arab	0.781	0.811	0.000	0.923
Ethiopian	0.003	0.003	-0.002	0.062
Married	0.527	0.547	-0.005	0.508
No. of Children	3.234	3.434	-0.010	0.827
Immigrant	0.070	0.062	-0.001	0.702
Health limitation	0.553	0.576	-0.015	0.049
Age	42.560	43.493	0.209	0.134
Single parent	0.148	0.143	0.001	0.802
Ultraorthodox	0.030	0.029	-0.003	0.282
Circles pop.	0.227	0.197	0.009	0.129
Matriculation	0.074	0.064	-0.004	0.261
log(RWG Participants)	6.508	6.528	0.003	0.814
Unemployment Rate	0.097	0.098	0.000	0.864
log(Labor Force)	9.014	9.001	0.010	0.475
log(Total Pop.)	9.941	9.931	0.009	0.470
log(Pop. in eligibility ages)	9.287	9.281	0.008	0.502
Socioeconomic Index	-0.987	-1.006	0.000	0.991

Joint Sig. (P-val.): 0.247

Obs. in Treatment: 8873

Obs. in Control: 9858

*Note:* This table reports balance tests for treatment assignment in the RWG. The two columns on the left contain the raw average of the specified covariate within the treatment and the control groups accordingly. The two columns on the right contain estimates and P-values for the test of a null estimate. The reported statistics are taken from a regression of the specified covariate on the treatment indicator, controlling for allocation-unit fixed effects. Standard errors are clustered at the allocation unit level. The sample contains only IA recipients who have non-missing locality-level information, which are at the focus of this paper. In addition, individuals who last reported to their local employment office before their allocation date are excluded from the sample, as they could not have been influenced by their treatment status. P-value of the F statistic for joint significance of the covariates in a regression of the treatment indicator on the covariates and allocation-unit fixed effects is reported at the bottom of the table.

intervention in the local level, which I elaborate on. For the purpose of clarity, I begin by describing a static version of this measure, and only then I introduce time into the discussion. Throughout this section I discuss both the main identification assumptions and other relevant empirical challenges.

#### 4.1. The Definition of Local Markets

The first element in the model is the definition of local markets. The basic idea behind externalities of the form that I consider in this paper is that an intervention may not only affect those who ‘receive treatment’, but also their surrounding environment. The kind of an environment requires further characterization. Potentially,

one could focus on the effect of the intervention on the ‘labor demand side’, that is, on the rate of vacancies by area (Gautier *et al.* 2018), the productivity of workers, etc. I, however, focus on the ‘labor supply side’ – specifically, on the employment status of job seekers over time. This implies that the externalities I consider here should be discussed in the context of the place of residence and the relations between job seekers in this place, rather than their places of work.

That said, there is still the question of how to define exactly a local labor market. As explained in the theoretical discussion, an assumption regarding the segmentation of local markets is usually necessary for the consistency of the estimators in a model that is meant to measure externalities - regional SUTVA. This assumption is a less restrictive form of SUTVA, that is limited to the market-level. Specifically, regional SUTVA requires that there are no interferences between units from different local markets. If, for example, job seekers compete on vacancies in a regional level, it is possible that treatment status of an individual from one locality may affect the potential outcomes of an individual from a neighboring locality. I think that it is implausible that any given set of locally defined labor markets in Israel would actually characterize separate markets, and strictly satisfy this assumption. Israel is a small country, and commuting to work is common, at least among the wealthier population (Bleikh 2018). Manning and Petreongolo (2017) study the relationship between markets, and show both theoretically and empirically that artificial divisions of local labor markets are most commonly problematic and most likely to be wrong.

Nevertheless, from the point of view of the designers of the RWG, the ultimate goal is to incentivize individuals from the target population to work outside of their locality of residence. Thus in the baseline estimations I rely on the notion that the relevant population is highly engaged with locality-based labor markets. In any case, if there are interferences between units from different local markets, I expect them to manifest in large confidence intervals which would make it harder to detect externalities. Therefore, if some evidence for the importance of externalities is detected, I do not suspect it to be a mechanical product of interferences between units from different local markets.

In a subsequent set of estimations, I use alternative definitions for local labor markets. Using localities’ XY coordinates I define locality-clusters based on Euclidian distances between localities. This is performed by implementing the K-means clustering algorithm on the XY coordinates, setting the amount of clusters to 100, 120, and 140 (there are 219 localities in the sample). I then define each market as the intersection of gender, ethnic origin (Arabs and non-Arabs), and locality-cluster. The idea behind this characterization is that local-markets may be based upon both geographical proximity (in a less restrictive form than localities) and ‘social’ proximity. Henceforth, I will address the baseline characterization of markets as ‘locality-based markets’, and the set of alternative definitions as ‘proximity-based markets’ henceforth.

## 4.2. Outcome Variable

The available administrative data contains information about monthly-report statuses to the local employment office of each participant in the RWG since her month of allocation. In the available administrative data there is no information regarding actual employment status, wages, or other labor market outcomes. Therefore, the outcome variable I use is the total amount of monthly absences of an individual from the local employment office over the twelve months since allocation.<sup>6</sup> The importance of reporting to the employment office for each individual is that it is a condition for receiving UI or IA benefits. That said, absence is assumed to be the result of either expired eligibility, a job-seeker found a job, or she became discouraged. I assume that the rates of expiration of eligibility and of becoming discouraged are similar across treatment arms. Hence a difference between treatment and control with respect to this sum proxies a difference in the length of time in employment during the first year since allocation. The reason for using aggregated reports is based on the notion that the externalities that are on the focus of this research may take some time to take effect. However, aggregation over too many months may mask any existing effect. Therefore, the outcome variable must capture a long enough period, but not too long; I chose to use a time-window of one year.

Unfortunately, using this outcome variable can actually be a sensible proxy for the length of employment in the first year since allocation only for IA claimants. Because eligibility for UI benefits in Israel is limited to 3-6 months since the declaration date, absences from the employment office after the expiration month do not seem like a sensible proxy for employment among UI claimants. In future research, with NII supplemental data, the use of direct measures could provide richer and more accurate conclusions, and would allow the use of UI claimants data in the estimations.

## 4.3. General Structure of the Empirical Model

Following the notation above, I assume a model in which the economy is segmented into the defined local labor markets, and that each individual acts in only one market. My objective is to investigate the effect of  $P_m$  on some relevant outcome  $Y_{i,m}$ . Initially, I would have wanted to estimate a continuous interaction model as the following:

$$(1) \quad Y_{i,m} = c_0 + \alpha T_i + \beta T_i P_m + \delta P_m + \gamma X_{i,m} + \varepsilon_{i,m}$$

Where  $Y$  is the outcome of interest,  $T$  is the individual treatment status indicator,  $P$  is the local saturation level of the intervention,  $X$  is a vector of controls and  $\varepsilon$  is the error term.  $i, m$  are indices for individuals and markets respectively. The standard errors must be clustered at the market level <sup>7</sup>

---

<sup>6</sup>For a discussion about the influence of the choice of the outcome variable see Card *et al.* 2007 and Card *et al.* 2010

<sup>7</sup>As will be mentioned below, in my actual model I cluster the standard errors both at the market level and at the allocation unit level.

For simplicity, assume now that both  $T$  and  $P$  are exogenous. Then  $c_0$  captures the mean outcome of a non-treated individual in a market with zero saturation.  $\delta$  captures the average marginal effect of an increase in the local saturation level on the non-treated;  $\beta$  captures the average effect of being assigned to treatment rather than control in a market with a given saturation level. Alternatively,  $\beta$  could be interpreted as the effect of being assigned to treatment in a fully saturated market, relative to being assigned to treatment in market with zero saturation. If SUTVA holds,  $\alpha$  is the difference in means estimator (in practice it will be a scaled version). If SUTVA is violated, then the ITT effect varies with  $P$ , and it must be that  $\delta \neq 0$  or  $\beta \neq 0$  (or both).

An empirical test for externalities using this model can then be performed in two ways: (1) test for the null hypothesis that  $\delta = 0$ ; (2) test for the null hypothesis that  $\beta = 0$ . If the null is rejected in one of these tests, it can be interpreted as a sign for the presence of externalities, and the magnitude of the estimated coefficients can be interpreted in a *causal* fashion. More specifically, in the case where  $Y$  signifies the employment status of individuals, if negative externalities, such as displacement or general equilibrium effects, are at play for all of the population, one should expect for both  $\beta$  and  $\delta$  to be negative. If positive externalities, such as peer effect or network effects, are at play, the sign of the coefficients should be positive.

However, there are some empirical challenges for the reliability of this model. First, as previously discussed, a known challenge in the empirical literature is statistical power, which limits the ability to detect externalities even if they do exist. Namely, there should be enough variation in  $P_m$  and enough observations to capture an existing effect. This is a challenge in the case of the RWG and I adress it below in detail.

Second, one should convince that  $T$  and  $P$  are indeed exogenous. Exogeneity of  $T$  in the RWG is guaranteed by design as long as the model includes a proper controlling term for allocation-units (fixed-effects), since the allocation to treatment is randomly assigned only within allocation-units. With contrast, in the RWG,  $P$  is not only endogenous, it is not actually observed. The available data contains only data regarding the participants of the RWG, and not of the entire population of each local market (these populations are not well-defined by themselves, and can only be approximated). I therefore use a proxy for  $P$  that is based on the participants of the RWG alone. This proxy is endogenous by itself, and I use an IV estimation approach to solve this issue.

#### 4.4. The Intensity of the Intervention

The general mechanism which has been considered by previous studies as the cause for externalities in ALMPs is the intensity of the intervention within a local labor market. Specifically, the object of interest is the amount of individuals who recieved treatment relative to the size of the local population (or some measure of the size of the local substitutable work force with respect to the target population of the intervention). I also construct a measure of local intensity of the RWG intervention. Defining a usable and interpretable

measure is a key feature of the empirical design in this paper. In this subsection I present my measure, and discuss both its' limitations and application in the empirical model.

#### 4.4.1. Introducing Q

An exogenous and direct measure of  $P_m$  (as defined in the theoretical section) is infeasible using the currently available data. With contrast to the randomized saturation design of Crépon *et al.* (2013), the design of the RWG does not include pre-defined (and random) local saturation levels. Therefore any measure of local intensity of the intervention will not be randomly assigned *per se*. I then follow the lines of Ferracci *et al.* (2014) and use observed variation in local intensities. However, instead of using a matching procedure and a CIA assumption, I exploit a random variation that is generated by the original randomization procedure of the RWG to instrument for these observed local intensities in a two-stage IV model.

To explain my procedure, it is useful to first consider an experiment with a simple (non-clustered) randomization design. Also assume that: (1) the randomization occurs in a single point in time, and (2) that the randomization procedure is based upon a prior proportion  $\bar{Q}$  of those allocated to treatment out of the population, and carried out without considering the segmentation to markets. Define:

$$Q_m = \frac{\sum_{i \in m} T_i}{||i \in m, i \in experiment||} = \frac{N_m^T}{N_m^T + N_m^C}$$

Where  $N_m^T$ ,  $N_m^C$  are the amounts of those allocated to treatment and control in a given market respectively. Also define:

$$\nu_m = \bar{Q} - Q_m$$

$$\Rightarrow Q_m = \bar{Q} + \nu_m$$

That is,  $\nu_m$  is the market-level sample-deviation from the prior proportion  $\bar{Q}$ , and is equal to demeaned  $Q_m$ . Obviously,  $\nu_m$  is a random output of the randomization procedure in which the different local markets are not taken into consideration. Thus, I argue that it is safe to assume that  $\nu_m$  is orthogonal to any background characteristic of either individual or market (as long as the randomization procedure ignores them), i.e.  $Q_m$  in place of  $P_m$  would be exogenous in a regression model such as model (1). In addition, note that the variability of  $\nu_m$  depends on the size of the market, where in large markets it is expected to be very close to zero, while in small markets it could receive values that are farthure from zero with greater probability.

In a clustered randomization design such as in the RWG the situation is different.  $Q_m$ , as it was defined, would

not be orthogonal to other covariates since the randomization procedure takes some observed characteristics into consideration. An analogous measure, however, would be useful. Define:

$$Q_{m,a} = \frac{\sum_{i \in m, i \in a} T_i}{||i \in m, i \in a, i \in experiment||} = \frac{N_{m,a}^T}{N_{m,a}^T + N_{m,a}^C}$$

Where  $a$  is an index for allocation-unit (or cluster). Observe that  $Q_{m,a}$  is similar to  $Q_m$  in the sense that it is defined over the randomization-unit of the experiment (in the simple design it is all of the participants, while in the clustered version it is the cluster) and measured within markets. Orthogonality of  $Q_{m,a}$  is guaranteed given the allocation-unit  $a$ . That is,  $Q_{m,a}$  would be exogenous in a regression model such as model (1) when controlling for allocation-unit fixed effects in the regression. Empirical tests for this assertion are presented at subsection 4.4.5.

#### 4.4.2. Endogeneity of Q

Observe that the economic meaning of  $P_m$  or  $Q_m$  does not apply to  $Q_{m,a}$ . The main segmentation of interest is only for markets, but not within allocation-units. However, using  $Q_m$  in a regression model that is based on a clustered experiment, even when controlling for allocation-unit fixed-effects, raises a problem of confoundedness. To see this, note the following:

$$Q_m = \frac{\sum_{i \in m} T_i}{N_m^T + N_m^C} = \frac{\sum_a \sum_{i \in m, i \in a} T_{i,a}}{N_m^T + N_m^C}$$

$$Q_m = \frac{\sum_a N_{m,a}^T}{N_m^T + N_m^C}$$

Define  $N_{m,a} = N_{m,a}^T + N_{m,a}^C$ . Thus:

$$Q_m = \frac{\sum_a N_{m,a} * Q_{m,a}}{N_m^T + N_m^C}$$

$$\Rightarrow Q_m = \sum_a \frac{N_{m,a}}{N_m^T + N_m^C} * Q_{m,a}$$

This formulation describes the relationship between  $Q_m$  and the  $Q_{m,a}$ s in a clustered experimental design. We have that  $Q_m$  is a weighted sum of the different  $Q_{m,a}$ s, where the weights are a function of the amount of the participants in the experiment within a market and the size of an allocation-unit within a market. Obviously, the weights may depend on an outcome of interest (say, larger markets tend to be closer to optional vacancies thus allowing for the residing population to be more responsive to an intervention). Thus,

plugging  $Q_m$  to model (1) in place of  $P_m$  where the randomization is clustered, even when controlling for allocation-unit fixed-effects would be exposed to problems of omitted variable bias. My strategy, in face of this endogeneity problem, is to use the  $Q_{m,a}$ s as instruments in an IV model.

It is relatively straight forward to justify the validity of the four identifying assumptions of such an IV model.

**Independence** of  $Q_{m,a}$  from the potential outcomes follows from the fact that the variance within  $Q_{m,a}$  is a random product of the allocation to treatment algorithm, which is indifferent of the potential outcomes.

**Relevance** would be satisfied if the  $Q_{m,a}$ s would be sufficiently good predictors of  $Q_m$ . In order to assess the validity of this assumption I calculate the Sanderson-Windmeijer (2014) conditional F statistic in my estimations.<sup>8</sup> **Monotonicity** requires that the potential treatment of every individual changes monotonically as a result of a change in the instrument. Since the formula of  $Q$  (and  $P$ ) is known, this assumption can be

verified directly. As the last equation shows, the following holds:  $\frac{\partial Q_m}{\partial Q_{m,a}} = \frac{N_{m,a}}{N_m^T + N_m^C} \geq 0$ , which satisfies weak monotonicity. Strict monotonicity (compliance) is only guaranteed if using an instrument which corresponds to the allocation-unit of each observation; in that case, it cannot be that  $N_{m,a} = 0$  thus  $\frac{N_{m,a}}{N_m^T + N_m^C} > 0$ . Denote by  $Q_{m,self}$  the variable which contains the proportion of those allocated to treatment in market  $m$  and belong to the allocation-unit of each observation. In order to maximize the observations in which I am making inference on (compliers), and in order to maintain a high degree of interpretability, I will use  $Q_{m,self}$  alone in the 2SLS procedure. Note that using this instrument alone may jeopardise the validity of the relevance assumption, because  $Q_{m,self}$  is responsible for only a small part of  $Q$  (or  $P$ ). **Exclusion** states that the entire effect of  $Q_{m,a}$  on the total absences from the employment office during the first year since allocation is through  $Q_m$ , that is, through the local intensity of the intervention. This seems quite reasonable, as  $Q_{m,a}$  captures a measure of this intensity in a sub-group within a market by itself.

Given this set of four assumptions, the estimated model captures *the local average casual response*, that is, a weighted average of the marginal effect of  $Q$  on  $Y$  among the compliers at each value of the continuous treatment (Angrist and Pischke, 2009). In this context, the compliers are those who would take higher treatment value  $Q_m$  if and only if  $Q_{m,self}$  strictly increases. In other words, the compliers are those who an increase in  $Q_{m,self}$  causes an increase in the local intensity of the intervention in their environment. Observing the formulation of  $Q_m$ , the compliers are those who have a strictly positive weight multiplied by  $Q_{m,self}$ , which is the entire population. The local average casual response is then the average of the marginal casual effect of the local intensity level on the total absences from the local employment offices, within the available range of the local intensities. The weighting function is defined by the relative size of the group of compliers at each value of the continuous treatment.<sup>9</sup>

<sup>8</sup>I estimate a two-stage interaction model, which means that I use in fact, two endogenous variables. This F statistic is adjusted for multiple endogenous variables and multiple instruments.

<sup>9</sup>An additional layer of complication regarding the meaning of the IV estimator is due to the interaction framework (Angrist and Pischke, 2009). Since the second stage (similar to the form of eq. 1) will contain the endogenous variable both interacted and uninteracted, there are in fact two endogenous variables in the model. Analogically, there would be two instruments -  $Q_{m,self}$  and  $TQ_{m,self}$  - and two first stage equations for each of the endogenous variables. Each of these two instruments defines a different population of compliers: while an increase in  $Q_{m,self}$  causes an increase in  $Q_m$  among the entire population (all observations would be compliers), it is not obvious that an increase in  $TQ_{m,self}$  also causes an increase



#### 4.4.3. The relationship Between Q and P

An important difference between  $Q_m$  and  $P_m$  is in the definition of the denominator. In  $P_m$ , the denominator -  $N_m$  - includes the entire population (or substitutable work force, etc.) of a given market, while in  $Q_m$  it only includes those who participate in the experiment and were allocated to either treatment or control -  $N_m^T + N_m^C$ . By definition,  $N_m^T + N_m^C \leq N_m$ . Therefore, there exists some  $N_m^U \geq 0$ , which represents the size of the population within each market *that is unobserved in the data*, such that:

$$N_m^T + N_m^C + N_m^U = N_m$$

I can now rewrite  $P_m$  as:

$$P_m = \frac{Q_m * (N_m^T + N_m^C)}{N_m^T + N_m^C + N_m^U}$$

Therefore:

$$(*) \quad Q_m = P_m \left(1 + \frac{N_m^U}{N_m^T + N_m^C}\right)$$

The equation (\*) describes the relationship between  $Q_m$  and  $P_m$ , and it serves as a basis for the final empirical design. Without reliable information regarding the actual  $N_m$  (the existing data is very crude), I will assume that  $\frac{N_m^U}{N_m^T + N_m^C}$  is approximately constant over markets and use  $Q_m$  as a scaled version of  $P_m$  in the estimated models (Baird *et al.* 2018). This assumption serves only as a working assumption, as this ratio may vary considerably across markets, although it may have some validity if participation in the RWG experiment is randomly determined. Assuming so, the scaling of  $P_m$  will not affect the results of the hypotheses tests (OLS is equi-variant), but will harm the economic interpretation of the magnitude of the estimators, as long as  $N_m^U$  remains unknown.

#### 4.4.4. Variation Over Time

So far I present  $P_m$ ,  $Q_m$  and  $Q_{m,a}$  as time-invariant. Obviously, since the RWG was carried out over time and since allocation-units also differ in terms of time, time should be part of the analysis. Incorporating time into the measure of the local intensity of the intervention adds some degree of complication, however the core arguments that were presented above will also apply here.

---

in  $Q_m$  among the entire population (it will surely include all of the treatment group). The final IV estimators would represent a weighted average of the estimators that would have been produced using one instrument at a time. The weights are determined by the strength of the first stage of each instrument separately. However, since the identification of the parameters in the model is dependent on both instruments, actual analysis of the weights becomes more complicated, and will be left for future analysis.

An important distinction to begin with is between objective time and relative time. The short analysis in this subsection focuses on time that is relative to the participants' allocation dates. Another time line, that will remain mostly in the background, is the actual year and month of allocation. For simplicity, I will present my measure for the local intensity level of the intervention while referring to a single cohort in the RWG. This will allow me to focus on the relative time line. Notice, however, that eventually the measures I will present vary over the objective time line; that is why at the end of this subsection I add the index  $a$  to the measures I present, which signifies that it changes over the different cohorts.

In my model I examine the aggregated behavior of the participants of the RWG over a time window of twelve months since their allocation date. I use a single measure of the local intensity of the intervention. To do that, I generate the final version of the local intensity of the intervention in two stages: first, I calculate the intensity in each period (relative to ones' allocation date); second, I take a simple average of the intensity levels over the relevant periods (I use a window of twelve months). Note that the choice of taking a simple average is actually a choice of a functional form. The justification for a simple average is that absent any prior assessment about the importance of the intensity level of specific periods, I assign all the periods equal importance weights and account for the entire time trajectory using an average transformation.

Still, an accurate definition of the local intensity of the intervention in a given period is required. For the purpose of clarity, I add to the variables that I used so far an index  $t$  ( $P_{t,m}$ ,  $Q_{t,m}$  and  $Q_{t,m,a}$ ) which represents *the months relative to the allocation date* of each participant. That is,  $P_{0,m}$  signifies the real local intensity of the intervention at the month of allocation,  $P_{-1,m}$  signifies the real local intensity of the intervention one month before the month of allocation and so on. Recall that allocation-units may actually be different in terms of objective time, but I refer now to a single cohort and the use of the index  $t$  allows me to address relative time periods in a clear way.

In the measure I define the denominator remains as before - the entire participants within a specific market over all time periods; that is used as a proxy for the size of the market as discussed in point 4.4.3, and it remains time-invariant (the denominator varies only across markets, thus it is only indexed by  $m$ ). The numerator of my measure takes into account all of those allocated to treatment in previous periods with respect to period  $t$  within a market  $m$ , i.e. the full history of allocations relative to time  $t$ . Formally:

$$Q_{t,m}^{FH} = \frac{\sum_{\tau \leq t} N_{\tau,m,a}^T}{N_m^T + N_m^C}$$

$$\iff Q_{t,m}^{FH} = \sum_{\tau \leq t} \frac{N_{\tau,m,a}}{N_m^T + N_m^C} * Q_{\tau,m,a}$$

Taking a simple average over a twelve-months period of  $Q_{t,m}^{FH}$  results with the following:<sup>10</sup>

---

<sup>10</sup>Notice again that  $\bar{Q}_m^{FH}$  is only indexed with  $m$ . The reason is that I am only referring to a single cohort in the RWG, thus this measure changes for this cohort only across markets.

$$\bar{Q}_m^{FH} = \frac{1}{12} \sum_{c=0}^{11} Q_{c,m}^{Full} = \frac{1}{12} \sum_{c=0}^{11} \sum_{\tau \leq c} \frac{N_{\tau,m,a}}{N_m^T + N_m^C} * Q_{\tau,m,a}$$

Rearranging the last expression:

$$\bar{Q}_m^{FH} = \sum_{\tau \leq 0} \frac{N_{\tau,m,a}}{N_m^T + N_m^C} * Q_{\tau,m,a} + \sum_{0 < \tau \leq 11} \frac{12 - \tau}{12} \frac{N_{\tau,m,a}}{N_m^T + N_m^C} * Q_{\tau,m,a}$$

Intuitively,  $\bar{Q}_m^{FH}$  assigns importance weights to every element  $\frac{N_{\tau,m,a}}{N_m^T + N_m^C} * Q_{\tau,m,a}$  for  $t \leq 11$ . Earlier allocations are taken into account in all of the later intensity measures, thus all allocations prior to ones' allocation month are considered in full (weighted 1), while later allocation months have linearly diminishing importance.

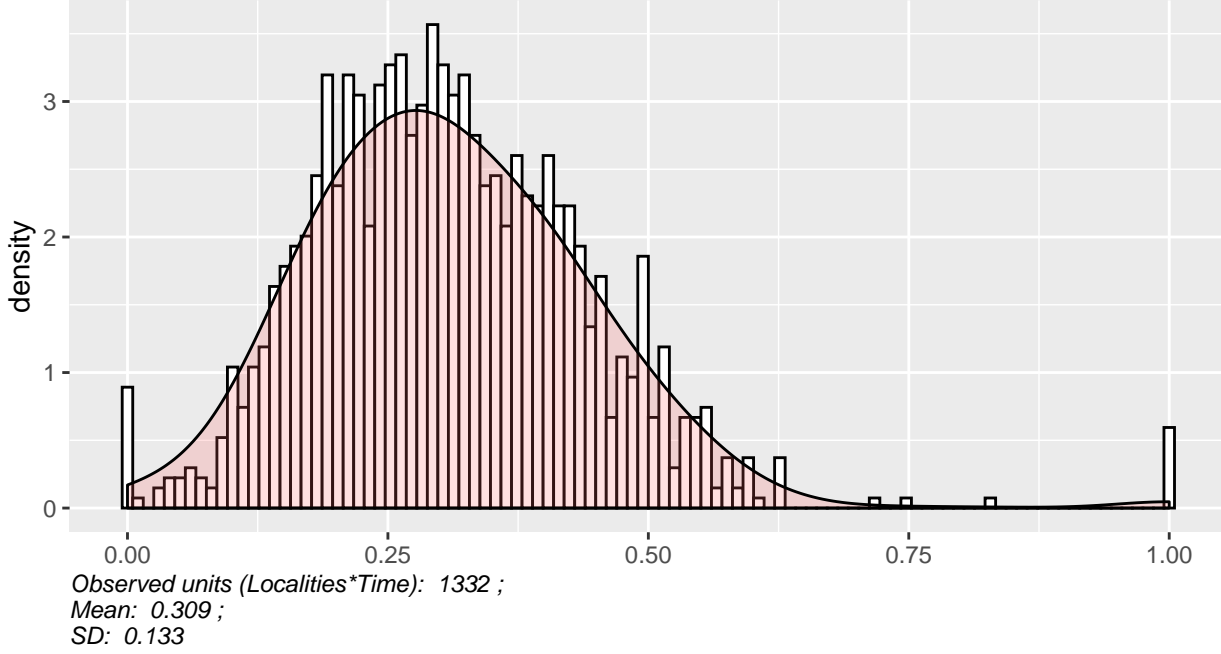
It is easy to see that the arguments above regarding the problem of endogeneity also apply to both  $\bar{Q}_m^{FH}$ . In addition,  $Q_{m,self}$  is a valid instrument for these variables by the same arguments above, and can be used in an IV regression.

Recall that the notation I have just used was referring to a single cohort in the RWG. In reality, the measure that I have presented varies over the different cohorts in the experiment. That is why an additional index  $a$  is required -  $\bar{Q}_{m,a}^{FH}$ . For a cleaner representation, I will address this variable by  $Q^{FH}$  henceforth.

#### 4.4.5. Empirical Explorations of the $Q$ Measurement

After defining the main  $Q$  variables -  $Q^{FH}$ ,  $Q_{m,self}$  - In this subsection I will present the distribution of  $Q^{FH}$  and orthogonality tests for  $Q_{m,self}$ . Figure 1 presents the raw distribution of  $Q^{FH}$ , which is the main independent variables in this paper. It is apparent that  $Q^{FH}$  has an approximately bell-shaped symmetric distribution, and that there is substantial variation within it. There are peaks at the extreme values 0 and 1, and a little positive skew. Since  $Q^{FH}$  is endogenous, of course, not all of the observed variation may be utilized in a causal model.

Figure 1: Raw Distribution of Q-FH



Note: This figure present the raw distribution of the main independent variable – Q-FH. The unit of observation here is locality interacted with allocation-month. Mean and standard deviation are reported at the bottom of the graph. Individuals who last reported to thier local employment office before their allocation date are excluded from the sample, as they could not have been influenced by their treatment status.

In order to solve the endogeneity of  $Q^{FH}$ , I use  $Q_{m,self}$  as an instrument in a 2SLS procedure. In order to demonstrate the exogeneity of  $Q_{m,self}$  (thus support the validity of the independence assumption), table 3 presents orthogonality tests for this variable in the locality-based version of markets. The presented orthogonality tests are analogous to the simple balance tests that are presented in table 2. For each covariate among the set of controls  $X$  I run a regression of the covariate on  $Q_{m,self}$ , controlling for allocation-unit fixed effects, and clustering the standard errors at both the allocation-unit and at the market level. The estimated coefficient of  $Q_{m,self}$  is then tested for the null hypothesis of a null estimate. If  $Q_{m,self}$  is indeed independent from all cavariates, the null should not be rejected. In addition, I perform a joint significance test of all of the covariates against  $Q_{m,self}$ . I repeat this procedure four times, each time for a different subsample that will be used in the estimation (see subsection 4.5). Orthogonality tests for  $Q_{m,self}$  in the proximity-based version of markets is presented in appendix table A.1.

The estimates in table 3 are all insignificant at the 95% level except for one, and most of them are far from the rejection area. Since I perform here many tests, it is sensible that some of them will be rejected (to be specific, about 5%). In any case, the coefficients are mostly very small (they correspond to a unit difference in  $Q_{m,self}$ , which is the entire range of  $Q_{m,self}$ ). In addition, the P-values of an F test for joint significance of the covariates in explaining  $Q_{m,self}$  are all far from the rejection area. It does seem that there is some selection in health limitation (it also exists in  $T$  - see table 2). However, the coefficients are sufficiently small. Therefore, the estimates that are reported in this table support the claim that  $Q_{m,self}$  is as good as randomly

Table 3: Orthogonality Tests of  $Q_{m,self}$  - Baseline Model (Locality-based Markets)

Covariates	All		High Unemp.		High N Ratio		Small Markets	
	Est.	P-val	Est.	P-val	Est.	P-val	Est.	P-val
Female	-0.031	0.264	-0.053	0.164	-0.043	0.196	-0.018	0.601
Arab	0.026	0.485	0.042	0.436	0.017	0.709	-0.020	0.553
Ethiopian	-0.001	0.563	0.000	0.847	0.001	0.359	-0.002	0.423
Married	0.014	0.661	-0.014	0.757	0.016	0.694	-0.006	0.874
No. of Children	-0.060	0.728	0.043	0.888	0.074	0.737	-0.155	0.439
Immigrant	-0.018	0.363	-0.024	0.250	-0.006	0.707	0.004	0.811
Health limitation	-0.045	0.061	-0.011	0.763	-0.049	0.072	-0.068	0.014
Age	0.058	0.913	0.842	0.298	0.254	0.680	-0.282	0.674
Single parent	-0.007	0.725	0.026	0.401	0.012	0.579	0.010	0.676
Ultraorthodox	-0.017	0.060	-0.004	0.277	0.005	0.484	-0.011	0.266
Circles pop.	0.024	0.268	-0.010	0.712	0.019	0.447	0.021	0.459
Matriculation	-0.020	0.211	-0.036	0.065	-0.020	0.222	-0.012	0.525
log(RWG Participants)	0.032	0.821	0.231	0.278	0.146	0.378	0.012	0.929
Unemployment Rate	0.001	0.786	0.001	0.897	0.001	0.944	-0.003	0.389
log(Labor Force)	0.103	0.427	0.290	0.184	0.161	0.340	0.151	0.268
log(Total Pop.)	0.095	0.451	0.253	0.235	0.153	0.351	0.123	0.323
log(Pop. in eligibility ages)	0.089	0.477	0.240	0.260	0.138	0.397	0.113	0.358
Socioeconomic Index	0.004	0.943	-0.082	0.414	-0.057	0.429	0.084	0.082
Joint Sig. (P-val.):	0.342		0.613		0.241		0.227	
Observations:	18731		12671		16226		11100	
Amount of Markets:	219		65		109		188	

*Note:* This table present orthogonality tests for  $Q_{m,self}$  based on the available individual-level covariates, where markets are defined as localities. Each row contains a set of estimates for the association between  $Q_{m,self}$  and the specified covariate and P-value for the null hypothesis of null estimates. The estimates are taken from a regression model of the specified covariate on  $Q_{m,self}$ , controlling for allocation-unit fixed effects. Standard errors are clustered both at the allocation unit level and at the market level. The four columns-pairs correspond to four different samples: all available observations; markets with high unemployment rate (above the 0.75 percentile); High N Ratio - markets with high amount of participants in the RWG relative to the local labor force (above the median); and markets that had less than 15,000 residents in the ages 20-59 during 2015. All samples contain only IA recipients who have non-missing locality-level information, which are at the focus of this paper. In addition, individuals who last reported to their local employment office before their allocation date are excluded from all samples, as they could not have been influenced by their treatment status. P-values of the F statistics for joint significance of the covariates in a regression of  $Q_{m,self}$  on the covariates and allocation-unit fixed effects are reported at the bottom of the table.

assigned conditional on the allocation-unit fixed effects.

## 4.5. Power

Baird *et al.* (2018) showed analytically that there is trade-off in statistical power between the ability to detect treatment effect, and the ability to detect externalities that are due to local intensity levels of an intervention. They investigate the optimal randomized saturation experimental design that maximizes the power of both

tests. As previously discussed, even in the randomized saturation design of Crépon *et al.* (2013), insufficient power to detect externalities is a main issue. In the RWG, the experimental design is meant to maximize the ability to detect treatment effects, and therefore insufficient power to detect externalities must be of major concern. Furthermore, I use proxies of both employment (absences from the local offices) and the real local intensity level of the intervention ( $Q$ ). This means that I measure my latent variables with additional noise. Finally, I measure  $Q^{FH}$  using an instrumental variable, which means that it is measured with even more noise. For these reasons, lack of power is a major concern for my model. Nevertheless, since the prior probabilities for treatment assignment in the RWG were subjected to changes over different allocation units (mainly due to budget constraints), and since  $Q^{FH}$  also varies over time, there could still be enough variance in the local saturation levels to identify externalities.

Nonetheless, power considerations led me to estimate externalities within segments of the data where the externality should be more prominent, that is, segments in which externalities are expected to vary significantly or be measured with less noise. First, in markets where the ratio  $\frac{N_m^T + N_m^C}{N_m}$  is large (above the median), i.e. where the amount of the participants in the experiment is large relative to the size of the local population, it is sensible that variation in  $Q_m$  will reflect better the real variation in the ‘real’ levels of local intensity. This, however, depends on data about  $N_m$ , which is available but with a poor accuracy. Nevertheless I use the available data about the local average size of the work force to construct this ratio. In addition, one of the theoretical insights from Crépon *et al.* (2013) and Landais *et al.* (2013) is that in a slack market, displacement effects and general equilibrium effects are expected to be important. Therefore, another relevant market-level segmentation is of the level of local labor demand. In the absence of data about vacancies, and since the demand is relatively low in all peripheral localities, I use the local unemployment rate as a proxy for the local demand level as in Crépon *et al.* (2013) and estimate my model in high unemployment localities (above the 0.75 percentile). Another segmentation I use is to simply limit the sample to relatively small markets (below 15,000 people in the eligibility ages), where the values of  $Q^{FH}$  may be less restricted by the law of large numbers.

## 4.6. Final Empirical Design

Finally, I estimate the following two-stage model:

$$(FS1) \quad Q_{m,a}^{FH} = c_1 + a_1 T_{i,a} + b_1 T_{i,a} Q_{m,self} + d_1 Q_{m,self} + g_1 X_{i,m} + \psi_a + \nu_{i,m,a}$$

$$(FS2) \quad T_{i,a} Q_{m,a}^{FH} = c_2 + a_2 T_{i,a} + b_2 T_{i,a} Q_{m,self} + d_2 Q_{m,self} + g_2 X_{i,m} + \psi_a + e_{i,m,a}$$

$$(SS) \quad Y_{i,m,a} = \theta + \alpha T_{i,a} + \beta T_{i,a} Q_{m,a}^{FH} + \delta Q_{m,a}^{FH} + \gamma X_{i,m} + \psi_a + \varepsilon_{i,m,a}$$

Where  $(FS1)$ ,  $(FS2)$  and  $(SS)$  are the two first-stage and the second-stage equations.  $\psi_a$  is the allocation-unit fixed-effects term. The vector  $X$  contains both the individual-level and locality-level characteristics that were presented above. I cluster the standard errors both at the allocation-unit and at the market level. In each estimation I report the Sanderson-Windmeijer (2014) conditional F statistic to account better for having multiple endogenous variables and multiple instruments. In addition, in the baseline estimation, I use the locality-based market definition, and estimate the model four times: on the full sample, on the sample of localities with  $\frac{RWG \text{ Participants}}{WorkForce}$  ("N ratio") above the median, on the sample of localities with high unemployment rate (above the 0.75 percentile), and on localities with less than 15,000 residents at the eligibility ages. In subsequent set of estimations I use the proximity-based definition for markets, and estimate the model using different geographic-proximity parameters (on the full sample only).

## 5. Estimation Results

In the following section I present and discuss the results of the estimations compared to an estimation of a simple difference-in-means model.<sup>11</sup> This section is divided into three subsections. In subsection 5.1 I report the results of the main baseline estimation, using all available observations of IA claimants with locality-based markets. In subsection 5.2 I report the results of the same estimation procedure while restricting the data to specific groups, in order to investigate heterogeneity in the measured effects. In subsection 5.3 I report the results of the estimations of the proximity-based model.

### 5.1. Baseline Results - Locality-Based Markets

The detailed estimation results of the models using locality-based markets are reported in table 4. I begin the analysis with the validity of the IV model. The results support the validity of both the relevance assumption and of the monotonicity. Observing the reported F statistics at the bottom of panels C and D, the instruments are jointly significant. The general rule of thumb for a sufficiently strong instrument is for the F statistic to be greater than 10. This rule does not necessarily qualify here since it is intended for IV models with only a single endogenous variable and a single instrument (Sanderson and Windmeijer, 2016). Furthermore, even if the value of the F statistic is lower than 10, it does not necessarily mean that it is a weak instrument. Eventually, the relevance assumption is supported here by the high and consistent values of the F statistic across the samples. Observing the first stage coefficients (panels C and D), each endogenous

<sup>11</sup>The DIM model is the following:  $(DIM) \quad Y_{i,m,a} = c_0 + \alpha T_{i,a} + \gamma X_{i,m} + \psi_a + \nu_{i,m,a}$ . In this model, since the assumption is that local markets are unimportant, I cluster the standard errors only at the allocation-unit level.

variable's corresponding instrument is highly statistically significant, and all are positive, as expected (an increase in  $Q_{m,self}$  causes an increase in  $Q^{FH}$ ). We have that the estimation results support the validity of the IV model.

Panel A reports the difference-in-means (DIM) estimators. The coefficients are precise and stable across the four samples. They indicate an average 0.23 of a month difference on the entire sample of IA claimants, which is equivalent to an increase of about a week (given there are 4.2 weeks in a month) or a 7.2% among the treatment group, relative to the control. Assuming that SUTVA is valid, and assuming that differences in absences are a good proxy for differences in employment, it means that the RWG causes an increase of that magnitude in employment during the first year since allocation. The magnitude of this effect is consistent with previous studies. Woodbury and Spiegelman (1987) estimated an average difference of 1.15 weeks in unemployment spells between treatment and control in the 1984 Illinois experiment. Card and Hyslop reported a 14% increase over an 18 month time window in the Canadian SSP, however this program offered much more generous bonuses. In addition, there is some heterogeneity in the magnitude of the DIM estimator over localities, and in small localities the measured effect reaches about 10.5% increase.

Panel B reports the reduced form estimates of the casual effect of  $Q_{m,self}$  on  $Y$ . The coefficients of both  $Q_{m,self}$  and  $TQ_{m,self}$  are statistically insignificant across all subsamples. It is apparent, however, that their signs are consistent. The point estimates suggest that while the control group experiences a negative (and insignificant) effect of  $Q_{m,self}$ , the treatment group experiences a subtler effect or even opposite in sign. Specifically, within the full sample, a 10% increase at the average value of  $Q_{m,self}$  (4.7 percentage points) causes a 0.11 weeks decrease in total absences among the control group, and of a 0.8 weeks decrease among the treatment group. Again, these are statistically insignificant results, and there are no strong conclusions to draw from these coefficients.

The coefficient of the treatment status in Panel B is less accurate than in the DIM estimation. The magnitudes of the  $T$  coefficients in the reduced form and in the DIM model are not directly comparable, because of the interaction term in the reduced form equation. However, it is clear that since the coefficients of the interaction term are positive across all subsamples, and the coefficients of  $T$  are all greater in the reduced form estimation, the total effect of  $T|Q_{m,self}$  (point estimates) is greater than the one estimated at the DIM model. For example, in the sample of localities with high unemployment, at the average  $Q_{m,self}$ , the total difference in absences between treatment and control equals  $0.47 \times 0.58 + 0.38 \approx 0.65$ , which is 2.25 times the point estimate of 0.29 in the DIM model. In the other samples, the difference between treatment and control at the average  $Q_{m,self}$  equals  $\sim 1.8$  times the DIM coefficient. However, the standard errors in the reduced-form model are too large to determine if this difference is statistically meaningful.

Panel E reports the second stage estimation results. Overall, the coefficients are very imprecise. As discussed above, this estimates represent the local average causal response of  $Y$  for a change in  $Q^{FH}$ , that is the average marginal effect over the existing range of  $Q^{FH}$ . The point estimates of the interaction term are all



negative across the different samples, while the sign of the coefficient of  $Q^{FH}$  is inconsistent both in sign and in magnitude. The coefficient of the interaction term is consistently statistically insignificant, so the fact that it is stable and negative does not guarantee that there are negative externalities. In the full sample, at the average  $Q^{FH}$ , the effect of a 10% increase in  $Q^{FH}$  (3 percentage points) causes about a 0.2 decrease ( $0.03 \times -0.54 \approx -0.2$ ) in monthly absences among the control group. Such a change in  $Q^{FH}$ , causes about a 0.09 decrease ( $0.03 \times -2.84 \approx -0.09$ ) in monthly absences among the treatment group. However, these numbers could just be the result of noise in the data.

An interesting point is that although the coefficient of  $T$  in panel E is statistically insignificant across all subsamples, evaluation of the average difference between treatment and control at the average  $Q^{FH}$  provides very close point estimates to those that are reported at the DIM panel (A). In the full sample we have that this point estimate is  $0.03 \times -2.84 + 1.1 \approx 0.24$  which is very close to the 0.22 estimate in the DIM model, and within the 95% level confidence interval; in the sample of localities with high unemployment rate, the difference (evaluated at the average  $Q^{FH}$ ) is 0.28, in the sample of high N ratio it is 0.27 and in the sample of small localities it is 0.35. This suggests that the IV model may have only added noise to the DIM estimators.

## 5.2. Locality-Based Markets - Heterogeneity Analysis

In this subsection I report and discuss the results of the baseline estimation, carried out on restricted versions of the estimation samples. Specifically, I repeated the same estimation procedure, where I restricted the data to one of two groups at the time: females and Arabs.<sup>12</sup> It is important to clarify that the measures of  $Q_{m,self}$  and  $Q^{FH}$  are still based on the entire sample; only the final estimations are performed on the restricted samples. The tables with the detailed estimation results can be found in the appendix (tables A.2 - A.3).

Panels C and D in both tables report the first stage estimation results. The reported F statistics are mostly above 10, even though the sample sizes are now smaller. In the first stage of  $Q^{FH}$  (uninteracted) they are much larger, as in the unrestricted samples (table 4). By the values of the F statistics it seems that the first stage is stronger among females than among arabs. The coefficients of the main instruments corresponding to each first stage equation are all positive and highly statistically significant. These results, again, provide some support to the monotonicity and for the relevance assumption.

---

<sup>12</sup>The choice of these groups is mainly led by sample sizes considerations. Both females and Arabs provide sufficiently large samples to perform estimations on.

Table 4: Baseline - Regression Results for Locality-Based Markets

	All	High Unemp.	High N Ratio	Small Markets
Panel A: Difference in Means				
$T$	0.229*** (0.060) p = 0.0002	0.289*** (0.072) p = 0.0001	0.256*** (0.064) p = 0.0001	0.326*** (0.083) p = 0.0001
<i>Average <math>Y_{Control}</math></i>	3.21	2.9	3.09	3.1
Panel B: Reduced Form				
$T$	0.353** (0.158) p = 0.026	0.381* (0.204) p = 0.061	0.416** (0.178) p = 0.020	0.531*** (0.182) p = 0.004
$TQ_{m,self}$	0.090 (0.257) p = 0.728	0.577 (0.367) p = 0.116	0.117 (0.298) p = 0.697	0.156 (0.320) p = 0.627
$Q_{m,self}$	-0.260 (0.285) p = 0.361	-0.254 (0.349) p = 0.468	-0.335 (0.316) p = 0.290	-0.432 (0.321) p = 0.179
<i>Average <math>Q_{m,self}</math></i>	0.474	0.469	0.47	0.465
Panel C: First Stage - $Q^{FH}$				
$Q_{m,self}$	0.047*** (0.007) p = 0.000	0.021*** (0.005) p = 0.0002	0.027*** (0.005) p = 0.00000	0.047*** (0.008) p = 0.000
$TQ_{m,self}$	-0.005 (0.004) p = 0.210	0.001 (0.003) p = 0.580	0.001 (0.003) p = 0.813	-0.004 (0.005) p = 0.455
$F_{stat}$	49.29	15.03	28.86	41.33
Panel D: First Stage - $TQ^{FH}$				
$Q_{m,self}$	-0.040** (0.016) p = 0.011	-0.063*** (0.022) p = 0.004	-0.055*** (0.017) p = 0.002	-0.041** (0.017) p = 0.018
$TQ_{m,self}$	0.092*** (0.025) p = 0.0003	0.098*** (0.033) p = 0.003	0.095*** (0.028) p = 0.001	0.090*** (0.021) p = 0.00003
$F_{stat}$	14.49	10.18	12.81	20
Panel E: Second Stage				
$T$	1.101 (1.030) p = 0.286	1.145 (1.288) p = 0.374	1.324 (1.112) p = 0.234	1.812 (1.211) p = 0.135
$TQ^{FH}$	-2.843 (3.292) p = 0.388	-2.871 (4.115) p = 0.486	-3.518 (3.567) p = 0.324	-4.838 (3.959) p = 0.222
$Q^{FH}$	-0.535 (5.172) p = 0.918	18.825 (18.163) p = 0.300	-2.848 (10.158) p = 0.780	-0.945 (6.407) p = 0.883
<i>Average <math>Q^{FH}</math></i>	0.303	0.303	0.299	0.303
Observations	18,731	12,671	16,226	11,100

*Notes:* This table reports the estimation results for locality-based markets, using the  $Q^{FH}$  intensity measure, of the following regressions: difference-in-means, reduced-form, two first-stage equations and second-stage. All of the regressions include allocation-unit fixed-effects, as well as the set of background characteristics. Standard errors are in parentheses. In panel A, the main variable is the treatment value indicator. In this panel, the standard errors are clustered at the allocation-unit level only. In the rest of the estimations, the standard errors are clustered at both the allocation-unit and at the market level. Panels B report the estimates of the reduced-form interaction model of  $Y$  on  $Q_{m,self}$ . Sanderson-Windmeijer (2014) conditional F statistic are reported at the bottom of panels C and D. Panel E reports the second stage IV estimates. The average  $Q$  is reported at the bottom of panels B and E. The estimations are performed on four sub-samples: all available observations; markets with high unemployment rate (above the 75th percentile); markets with high amount of participants in the RWG relative to the local labor force (above the median); and markets with less than 15,000 residents in the ages 20-59 (2015). All samples contain only IA recipients only. In addition, individuals who last reported to their local employment office before their allocation date are excluded from all samples, as they could not have been influenced by their treatment status.

The difference in means estimators (panel A) show that the average differences between treatment and control are relatively high among Arabs. The coefficient in the sample of all localities is almost 30% higher than among the corresponding coefficient in the full sample of IA claimants. The same coefficient among females points to a lower difference between treatment and control. The sample of small localities is different with respect to the measured difference between treatment and control, where both famels and Arabs present greater differences (point estimates) than those of the full sample of small markets. However, these differences between the coefficients are statistically insignificant.

Generally, the reduced form estimation results among Arabs and females are consistent with those of the full sample in signs, and in statistical significance. In the sample of all localities, the average effect of an increase of 10% (4.6 percentage points) at the average  $Q_{m,self}$  on females who were allocated to treatment is a positive 0.05 increase in weekly total absences from the local emplyment office; among Arabs, this effect is of a 0.03 weeks increase. Similarly, the effect of a 10% increase at the average  $Q_{m,self}$  on females who were allocated to control is of a 0.06 weeks decrease in total absences, while among arabs it is of a 0.07 decrease in weekly absences. Of course, these effects are statistically insignificant and are very slim.

Evaluation of the differences bewtween treamtent and control at the average  $Q_{m,self}$  present some heterogeneity relative to the full sample of IA claimants. In the smaple of all localities, the difference between treatment and control, evaluated at the average  $Q_{m,self}$  is  $0.47 \times 0.16 + 0.46 \approx 0.53$  which is 1.8 times the DIM estimator. For females, the difference is of 0.48, which is 1.6 times the corresponding DIM estimator. In the sample of localities with high unemployment, the difference (evaluated at the average  $Q_{m,self}$ ) for females is 2.9 times the DIM estimator, and for Arabs it is 2.7. These figures, although they are not statisitically significant, may suggest that  $Q_{m,self}$  do add some level of information to the regressions, and adds more than just noise over the DIM model.

As in the full sample of IA claimants, assesing the difference between treatment and control at the average  $Q^{FH}$  in the second stage provides almost identicle figures to the corresponding estimates of the DIM model. This may suggest that the 2SLS mostly capture noise.

### 5.3. Proximity-Based Markets

As described earlier, in the discussion about the definition of markets (subsection 4.1), in addition to the baseline model, where the local markets are defined as localities, I report and discuss in this subsection the estimation results based on alternative definitions for local markets. The baseline model corresponds to the initial motivation behind the RWG. However, there is no reason to assume that localities capture local labor markets perfectly, and therefore I also estimate my model using alternative definitions for local markets. More specifically, I model local markets using a data-driven geographic parameter, and assume that markets are segmented not only on a geographical basis, but also on cultural and political basis. I assume that local

markets are defined by some degree of geographical proximity, on gender and on ethnic origin (Arab or non-Arab). Using CBS locality XY coordinates I construct locality-clusters with the K-means algorithm, and then construct the data as before. Orthogonality tests for  $Q_{m,self}$  with respect to individual-level covariates are presented in appendix table A.1.

There is a reason to assume that in the geographically-flexible definitions for local labor markets, if externalities do exist in the RWG, then they are most likely negative. In locality-based markets, the geographical proximity may also capture within-locality social interactions. However, if the definition of a market allows for a greater geographical distance between individuals in the same market, it is less reasonable that two such individuals socially interact with equal or higher probability. In order to make such a distinction visible, I first estimate the model using localities as the geographical segmentation (K=219, where there are 219 localities in the sample), and then I gradually force the geographic-cluster parameter, setting K to be 140, 120, and 100.

The estimation results using the proximity-based markets are presented in table 5. The first stage F statistics are all highly significant, and the signs and significance of the coefficients matches the attributes of the IV model. The reduced form estimates are not stable as they were in the model with the locality-based markets. The point estimates of  $Q_{m,self}$  and  $TQ_{m,self}$  show that a 10% increase at the average  $Q_{m,self}$  causes a change of -0.02-0.03 in weekly absences from the local employment office during the first year since the allocation date for the treatment group. For the control group, the effect is negative for K = 219, while for K=140,120,100 the effect is positive. The signs of the coefficients do not support my hypothesis that negative externalities would be more pronounced for K<219. Evaluation of the average difference between treatment and control at the average  $Q_{m,self}$  for the different values of K shows that these differences are generally lower for K<219, relative to the DIM estimators. However, the coefficients are statistically insignificant, and so there are no substantial conclusions to draw from them. As before, the second stage average differences between treatment and control, evaluated at the average  $Q^{FH}$ , are almost identical to the point estimates of the DIM model.

Table 5: Regression Results for Proximity-Based Markets

	$K = 219$	$K = 140$	$K = 120$	$K = 100$
Panel A: Difference in Means				
$T$	0.229*** (0.061) p = 0.0002	0.229*** (0.061) p = 0.0002	0.230*** (0.061) p = 0.0002	0.229*** (0.061) p = 0.0002
<i>Average <math>Y_{Control}</math></i>	3.21	3.21	3.21	3.21
Panel B: Reduced Form				
$T$	0.310** (0.136) p = 0.023	0.197 (0.134) p = 0.142	0.154 (0.144) p = 0.287	0.191 (0.140) p = 0.174
$TQ_{m,self}$	0.091 (0.219) p = 0.678	0.016 (0.223) p = 0.943	-0.092 (0.254) p = 0.719	0.146 (0.234) p = 0.532
$Q_{m,self}$	-0.176 (0.229) p = 0.442	0.058 (0.237) p = 0.806	0.165 (0.256) p = 0.520	0.048 (0.252) p = 0.851
<i>Average <math>Q_{m,self}</math></i>	0.474	0.473	0.474	0.474
Panel C: First Stage - $Q^{FH}$				
$Q_{m,self}$	0.068*** (0.007) p = 0.000	0.061*** (0.008) p = 0.000	0.059*** (0.008) p = 0.000	0.055*** (0.007) p = 0.000
$TQ_{m,self}$	-0.012** (0.005) p = 0.029	-0.010* (0.005) p = 0.057	-0.014** (0.006) p = 0.020	-0.008 (0.005) p = 0.102
$F_{stat}$	107.93	54.52	63.02	63.77
Panel D: First Stage - $TQ^{FH}$				
$Q_{m,self}$	-0.019 (0.012) p = 0.125	-0.037** (0.015) p = 0.013	-0.028** (0.013) p = 0.041	-0.039** (0.015) p = 0.011
$TQ_{m,self}$	0.079*** (0.019) p = 0.00005	0.075*** (0.020) p = 0.0003	0.074*** (0.021) p = 0.0004	0.078*** (0.021) p = 0.0002
$F_{stat}$	18.82	14.02	13.84	14.45
Panel E: Second Stage				
$T$	0.884 (0.887) p = 0.319	-0.048 (1.008) p = 0.962	-0.419 (1.086) p = 0.700	-0.083 (1.027) p = 0.936
$TQ^{FH}$	-2.126 (2.831) p = 0.453	0.888 (3.247) p = 0.785	2.122 (3.522) p = 0.547	0.966 (3.321) p = 0.772
$Q^{FH}$	0.755 (3.004) p = 0.802	0.809 (3.600) p = 0.823	-0.564 (4.274) p = 0.896	3.351 (4.196) p = 0.425
<i>Average <math>Q^{FH}</math></i>	0.306	0.304	0.304	0.304
Observations	18,731	18,731	18,731	18,731

*Notes:* This table reports the estimation results for proximity-based markets (geographical proximity interacted with gender and ethnic background), using the  $Q^{FH}$  intensity measure, of the following regressions: difference-in-means, reduced-form, two first-stage equations and second-stage. All of the regressions include allocation-unit fixed-effects, as well as the set of background characteristics. Standard errors are in parentheses. In panel A, the main variable is the treatment value indicator. In this panel, the standard errors are clustered at the allocation-unit level only. In the rest of the estimations, the standard errors are clustered at both the allocation-unit and at the market level. Panels B report the estimates of the reduced-form interaction model of  $Y$  on  $Q_{m,self}$ . Sanderson-Windmeijer (2014) conditional F statistic are reported at the bottom of panels C and D. Panel E reports the second stage IV estimates. The average  $Q$  is reported at the bottom of panels B and E. The estimations are performed with respect to four geographical segmentations which define the markets: segmentation by locality ( $K = 219$ ); segmentation by the results of K-mean clustering of the XY coordinates of the localities in the sample, for  $K < 219$ . All samples contain only IA recipients only. In addition, individuals who last reported to their local employment office before their allocation date are excluded from all samples, as they could not have been influenced by their treatment status.

## 6. Conclusion

In this paper, I investigated the existence of externalities of the Israeli Remote Work Grant program, which is a large reemployment bonus program that has operated since November 2015 in peripheral Israel. Using the administrative IES data about the RWG participants and some supplementary public data, I estimated the effect of the local intensity of the intervention on individuals who were allocated to treatment and on those who were not. Given the unique locality-based design of the RWG, the program poses an opportunity to investigate not only negative externalities, as done in previous research, but also positive externalities. Data limitations restricted my analysis with respect to outcome variables. Eventually I used a single outcome variable, which is the individual-level total amount of monthly absences from the local employment office during the first year since the date of allocation. Differences in this variable between treatment and control served as a proxy for differences in the amount of months spent in employment during this time period. My baseline model relies on the definition of local labor markets as the population of a certain locality. I also investigated alternative definitions for local markets that allow for relationships between neighboring localities. In order to solve the endogeneity problem that is generated by the size of the localities, I used a 2SLS interaction model to estimate the effect of the local intensity level of the program on the total monthly absences from the local employment office. I use an exogenous variation that is generated by the initial randomization procedure of the allocation to treatment as an instrument for the intensity measurement. As in previous studies, the main challenge in my empirical strategy was insufficient power to detect an existing effect.

The estimation results are overall inconclusive, and lack statistically significant effects of the local intensity levels of the RWG intervention. The second stage estimates show that evaluation of the difference between treatment and control at the average intensity level, is practically identical to the difference-in-means estimator. Overall it seems that the externalities model only captured noise. The reduced form estimation results, however were slightly more informative, and suggested that there could possibly be more extensive differences between treatment and control than those captured by the difference-in-means estimators.

Finally, these results do not supply any evidence for externalities of the RWG. However, it is unclear whether the lack of results stems from insufficient power or from the possibility that there are truly no externalities. In any case, the importance of this research for quality cost-benefit analysis and for the understanding of both local markets, and the behavior of the target population justify further research on the topic. Fortunately, future NII and IES data would allow to investigate externalities more accurately by using some more refined measurements and various outcome variables such as wages and actual employment status.

## Bibliography

- Abbring J.H., & Heckman, J.J. (2007). “Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation.” *Handbook of Econometrics*, Vol. 6, Part B.
- Ahn T. (2018). “Assessing the Effects of Reemployment Bonuses on Job Search: A Regression Discontinuity Approach.” *Journal of Public Economics*, Vol. 165.
- Albrecht J., Van Den Berg G.J. & Vroman S. (2009). “The Aggregate Labor Market Effects of the Swedish Knowledge Lift Program.” *Review of Economic Dynamics*, Vol. 12.
- Angelucci M., & Di Maro V. (2016). “Programme Evaluation and Spillover Effects.” *Journal of Development Effectiveness*, Vol. 8, No. 1.
- Angrist J.D., & Pischke J.S. (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- Avitabile C. (2012). “Spillover Effects in Healthcare Programs: Evidence on Social Norms and Information Sharing.”
- Baird S., Aislinn B.J., McIntosh C., & Özler B. (2018). “Optimal Design of Experiments in the Presence of Interference.” *Review of Economics and Statistics*, Vol. 100, No 5.
- Bayer P., Ross S.L., & Topa G. (2008). “Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes.” *Journal of Political Economy*, Vol. 116, No. 6.
- Björklund A., Clark M., Edin P., Fredriksson P., & Krueger A. (2005). *The Market Comes to Education in Sweden: An Evaluation of Sweden’s Surprising School Reforms*. New York: Russell Sage Foundation.
- Bleikh H. (2018). “Back and Forth: Commuting for Work in Israel.” Taub Center for Social Policy Studies in Israel, No. 05.2018.
- Blundell R., Costa Dias M., & Meghir C. (2003). “The Impact of Wage Subsidies: A General Equilibrium Approach.” Unpublished manuscript.
- Blundell R., Costa Dias M., Meghir C., & Van Reenen J. (2004). “Evaluating the Employment Impact of a Mandatory Job Search Program.” *Journal of the European Economic Association*, Vol. 2, No. 4.
- Calmfors, L. (1994), “Active Labor Market Policy and Unemployment - a Framework for the Analysis of Crucial Design Features.” *OECD Economic Studies*, No. 22.
- Card D., & Hyslop R. (2005). “Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers.” *Econometrica*, Vol. 73, No. 6.

- Card D., Chetty R., & Weber A. (2007). “The Spike at Benefit Exhaustion: Leaving the Unemployment System or Starting a NewJob?” *The American Economic Review*, Vol. 97, No. 2.
- Card D., Kluve J., & Weber A. (2010). “Active Labour Market Policy Evaluations: A Meta-Analysis.” *The Economic Journal*, Vol. 120.
- Card D., Kluve J., & Weber A. (2015). “What Works? A Meta-Analysis of Recent Active Labor Market Program Evaluations.” Discussion Paper No. 9236, IZA.
- Chetty R. (2008). “Moral Hazard Versus Liquidity and Optimal Unemployment Insurance.” *Journal of Political Economy*, Vol.116, No. 2.
- Crépon B., Duflo E., Gurgand M., Rathelot R., & Zamora P. (2013). “Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment.” *The Quarterly Journal of Economics*, Vol. 128, No. 2.
- Crépon B., & Van Den Berg G.J. (2016). “Active Labor Market Policies.” *Annual Review of Economics*, Vol. 8.
- Davidson C., & Woodbury S.A. (1993). “The Displacement Effect of Reemployment Bonus Programs.” *Journal of Labor Economics*, Vol. 11, No. 4.
- Deaton A., & Cartwright N. (2018). “Understanding and Misunderstanding Randomized Controlled Trials.” *Social Science & Medicine*, Vol. 210.
- Decker P.T. (1994). “The Impact of Reemployment Bonuses on Insured Unemployment in the New Jersey and Illinois Reemployment Bonus Experiments.” *The Journal of Human Resources*, Vol. 29, No. 3.
- Decker P.T., & O’Leary C.J. (1995). “Evaluating Pooled Evidence from the Reemployment Bonus Experiments.” *The Journal of Human Resources*, Vol. 30, No. 3
- Duflo E., & Saez E. (2003). “The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment.” *The Quarterly Journal of Economics*, Vol. 118, No. 3.
- Eugster B., Lalive R., Steinhauer A., & Zweimuller J. (2017). “Culture, Work Attitudes, and Job Search: Evidence from the Swiss Language Border.” *Journal of the European Economic Association*, Vol. 15, No. 5.
- Ferracci M., Jolivet G., & Van Den Berg G.J. (2014). “Evidence of Treatment Spillovers Within Markets.” *Review of Economics and Statistics*, Vol. 96, No. 5.
- Gautier P., Muller P. & Van Der Klaauw B., Rosholm M., & Svarer M. (2018). “Estimating Equilibrium Effects of Job Search Assistance.” *Journal of Labor Economics*, Vol. 36, No. 4.



- Gershoni N., Saporta-Eksten I., & Schlosser A. (2018). “Remote Work Grant: Midterm Report.” Working Paper No. 19, The Foerder Institute for Economic Research.
- Heckman J.J., Lochner L., & Taber C. (1998). “General-Equilibrium Treatment Effects: A Study of Tuition Policy.” *AEA Papers and Proceedings*, Vol. 88, No. 2.
- Heckman J.J., LaLonde R.J., & Smith J.A. (1999). “The Economics and Econometrics of Active Labor Market Programs.” *Handbook of Labor Economics*, Vol. 3.
- Hellerstein J.K., McInerney M., & Neumark D. (2011). “Neighbors and Coworkers: The Importance of Residential Labor Market Networks.” *Journal of Labor Economics*, Vol. 29, No. 4.
- Huber M., & Steinmayr A. (2017). “A Framework for Separating Individual Treatment Effects from Spillover, Interaction, and General Equilibrium Effects.” Discussion Paper No. 10648, IZA.
- Kluve J. (2006). “The Effectiveness of European Active Labor Market Policy.” Discussion Paper No. 37, RWI.
- LaLonde R.J. (1986). “Evaluating the Econometric Evaluations of Training Programs with Experimental Data.” *The American Economic Review*, Vol. 76, No. 4.
- Landais C., Michaillat P., & Saez E. (2013). “Optimal Unemployment Insurance over the Business Cycle.” Discussion Paper No. 1303, Centre for Macroeconomics.
- Lise J., Seitz S., & Smith J. (2004). “Equilibrium Policy Experiments and the Evaluation of Social Programs.” Working Paper No. 10283, NBER.
- Manning A., & Petrongolo B. (2017). “How Local Are Labor Markets? Evidence from a Spatial Job Search Model.” *The American Economic Review*, Vol. 107, No. 10.
- Manski C.F. (1993). “Identification of Endogenous Social Effects: The Reflection Problem.” *The Review of Economic Studies*, Vol. 60, No. 3.
- Meyer B.D. (1995). “Lessons from the U.S. Unemployment Insurance Experiments.” *Journal of Economic Literature*, Vol. 33, No. 1.
- Michalopoulos C., Robins P.K., & Card D. (2005). “When Financial Work Incentives Pay for Themselves: Evidence from a Randomized Social Experiment for Welfare Recipients.” *Journal of Public Economics*, Vol. 89.
- Miguel E., & Kremer M. (2004). “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica*, Vol. 72, No. 1.
- Rubin D.B. (1990). “Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies.” *Statistical Science*, Vol. 5, No. 4.

- Sanderson E., & Windmeijer F. (2016). “A Weak Instrument Test in Linear IV Models with Multiple Endogenous Variables”. *Journal of Econometrics*, Vol. 190.
- Schochet P.Z., Burghardt J., & McConnell S. (2006). *National Job Corps Study and Longer-term Follow-up Study: Impact and Benefit-Cost Findings using Survey and Summary Earnings Records Data*. Princeton, NJ: Mathematica Policy Research.
- Sopel M.E. (2006). “What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Interference.” *Journal of the American Statistical Association*, Vol. 101, No. 476.
- Topa G., & Zenou Y. (2015). “Neighborhood and Network Effects.” *Handbook of Regional and Urban Economics*, Vol. 5A.
- Van Der Klaauw B., & Van Ours J.C. (2013). “Carrot and Stick: How Reemployment Bonuses and Benefit Sanctions Affect Exit Rates from Welfare.” *Journal of Applied Econometrics*, Vol. 28, No. 2.
- Woodbury S.A., & Spiegelman R.G. (1987). “Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois.” *The American Economic Review*, Vol. 77, No. 4.

## Appendix

Table A.1: Orthogonality Tests of  $Q_{m,self}$  - Proximity-Based Markets)

Covariates	K = 219		K = 140		K = 120		K = 100	
	Est.	P-val	Est.	P-val	Est.	P-val	Est.	P-val
Female	0.043	0.370	0.070	0.315	0.056	0.502	0.084	0.310
Arab	0.004	0.906	-0.017	0.701	-0.040	0.446	0.001	0.991
Ethiopian	-0.002	0.218	-0.002	0.349	-0.002	0.284	-0.002	0.377
Married	-0.011	0.657	-0.026	0.442	-0.025	0.492	-0.040	0.262
No. of Children	-0.086	0.542	-0.164	0.389	-0.227	0.265	-0.168	0.431
Immigrant	-0.003	0.826	0.008	0.704	0.012	0.593	0.006	0.802
Health limitation	-0.052	0.018	-0.058	0.043	-0.056	0.063	-0.052	0.111
Age	-0.092	0.820	0.001	0.999	-0.373	0.489	-0.300	0.591
Single parent	0.004	0.789	0.029	0.166	0.022	0.316	0.027	0.205
Ultraorthodox	-0.012	0.064	-0.017	0.058	-0.013	0.131	-0.015	0.111
Circles pop.	0.027	0.099	0.013	0.503	0.019	0.331	0.031	0.102
Matriculation	-0.006	0.607	-0.003	0.845	-0.004	0.818	0.008	0.612
log(RWG Participants)	-0.029	0.745	-0.124	0.213	-0.122	0.301	-0.195	0.076
Unemployment Rate	0.001	0.823	0.002	0.721	0.000	0.976	0.003	0.576
log(Labor Force)	0.057	0.473	0.105	0.302	0.108	0.276	0.054	0.659
log(Total Pop.)	0.053	0.480	0.102	0.287	0.101	0.280	0.071	0.528
log(Pop. in eligibility ages)	0.049	0.510	0.099	0.302	0.095	0.308	0.066	0.561
Socioeconomic Index	0.001	0.971	0.023	0.686	0.024	0.694	-0.012	0.868
Joint Sig. (P-val.):	0.395		0.168		0.237		0.156	
Observations:	18731		18731		18731		18731	
Amount of Markets:	390		260		233		218	

*Note:* This table present orthogonality tests for  $Q_{m,self}$  based on the available individual-level covariates, where markets are defined by gender, ethnic origin and geographical proximity (defined by the number of means ( $K$ ) that is fed to the K-means clustering algorithm using the XY coordinates of each locality). Each row contains a set of estimates for the association between  $Q_{m,self}$  and the specified covariate and P-value for the null hypothesis of null estimates. The estimates are taken from a regression model of the specified covariate on  $Q_{m,self}$ , controlling for allocation-unit fixed effects. Standard errors are clustered both at the allocation unit level and at the market level. The four columns-pairs correspond to four different values of  $K$ , which governs the geographical segmentation. If  $K = 219$  the geographical segmentation matches localities; if  $K < 219$  some localities are grouped together. All samples contain only IA recipients who have non-missing locality-level information, which are at the focus of this paper. In addition, individuals who last reported to their local employment office before their allocation date are excluded from all samples, as they could not have been influenced by their treatment status. P-values of the F statistics for joint significance of the covariates in a regression of  $Q_{m,self}$  on the covariates and allocation-unit fixed effects are reported at the bottom of the table.

Table A.2: Regression Results for Locality-Based Markets - Arabs Only

	All	High Unemp.	High N Ratio	Small Markets
Panel A: Difference in Means				
$T$	0.295*** (0.069) p = 0.00002	0.297*** (0.078) p = 0.0002	0.296*** (0.073) p = 0.00005	0.397*** (0.087) p = 0.00001
$Average Y_{Control}$	3	2.83	2.93	2.96
Panel B: Reduced Form				
$T$	0.460*** (0.170) p = 0.007	0.472** (0.200) p = 0.019	0.510*** (0.185) p = 0.006	0.598*** (0.183) p = 0.002
$TQ_{m,self}$	0.155 (0.279) p = 0.579	0.726* (0.381) p = 0.057	0.279 (0.322) p = 0.388	0.195 (0.341) p = 0.568
$Q_{m,self}$	-0.355 (0.296) p = 0.231	-0.443 (0.330) p = 0.180	-0.472 (0.320) p = 0.141	-0.439 (0.328) p = 0.180
$Average Q_{m,self}$	0.465	0.464	0.465	0.456
Panel C: First Stage - $Q^{FH}$				
$Q_{m,self}$	0.029*** (0.006) p = 0.00001	0.019*** (0.006) p = 0.002	0.024*** (0.006) p = 0.00002	0.032*** (0.008) p = 0.00003
$TQ_{m,self}$	-0.001 (0.004) p = 0.853	0.004 (0.003) p = 0.164	0.002 (0.003) p = 0.473	-0.002 (0.005) p = 0.720
$F_{stat}$	25.48	10.83	20.3	27.39
Panel D: First Stage - $TQ^{FH}$				
$Q_{m,self}$	-0.040** (0.017) p = 0.022	-0.055** (0.023) p = 0.019	-0.049*** (0.019) p = 0.009	-0.044** (0.019) p = 0.019
$TQ_{m,self}$	0.077*** (0.025) p = 0.003	0.081*** (0.030) p = 0.007	0.084*** (0.028) p = 0.003	0.087*** (0.023) p = 0.0002
$F_{stat}$	9.8	7.75	9.92	16.7
Panel E: Second Stage				
$T$	1.729 (1.285) p = 0.179	2.196 (1.496) p = 0.143	2.033 (1.276) p = 0.112	1.950 (1.305) p = 0.136
$TQ^{FH}$	-4.610 (4.055) p = 0.256	-6.320 (4.871) p = 0.195	-5.609 (4.014) p = 0.163	-5.065 (4.276) p = 0.237
$Q^{FH}$	-0.996 (9.586) p = 0.918	20.194 (21.000) p = 0.337	0.182 (12.653) p = 0.989	-0.796 (10.173) p = 0.938
$Average Q^{FH}$	0.307	0.301	0.306	0.303
Observations	14,922	11,211	13,405	9,828

*Notes:* This table reports the estimation results for Arabs in locality-based markets, using the  $Q^{FH}$  intensity measure, of the following regressions: difference-in-means, reduced-form, two first-stage equations and second-stage. All of the regressions include allocation-unit fixed-effects, as well as the set of background characteristics. Standard errors are in parentheses. In panel A, the main variable is the treatment value indicator. In this panel, the standard errors are clustered at the allocation-unit level only. In the rest of the estimations, the standard errors are clustered at both the allocation-unit and at the market level. Panels B report the estimates of the reduced-form interaction model of  $Y$  on  $Q_{m,self}$ . Sanderson-Windmeijer (2014) conditional F statistic are reported at the bottom of panels C and D. Panel E reports the second stage IV estimates. The average  $Q$  is reported at the bottom of panels B and E. The estimations are performed on four sub-samples: all available observations; markets with high unemployment rate (above the 75th percentile); markets with high amount of participants in the RWG relative to the local labor force (above the median); and markets with less than 15,000 residents in the ages 20-59 (2015). All samples contain only IA recipients only. In addition, individuals who last reported to their local employment office before their allocation date are excluded from all samples, as they could not have been influenced by their treatment status.

Table A.3: Regression Results for Locality-Based Markets - Females Only

	All	High Unemp.	High N Ratio	Small Markets
Panel A: Difference in Means				
$T$	0.223*** (0.075) p = 0.004	0.223** (0.093) p = 0.017	0.241*** (0.082) p = 0.004	0.381*** (0.096) p = 0.0001
$Average Y_{Control}$	3.04	2.79	2.95	2.93
Panel B: Reduced Form				
$T$	0.367** (0.182) p = 0.045	0.307 (0.234) p = 0.191	0.401** (0.199) p = 0.044	0.478** (0.209) p = 0.022
$TQ_{m,self}$	0.250 (0.325) p = 0.443	0.725 (0.449) p = 0.107	0.466 (0.383) p = 0.224	0.165 (0.386) p = 0.669
$Q_{m,self}$	-0.329 (0.318) p = 0.300	-0.259 (0.381) p = 0.498	-0.391 (0.344) p = 0.256	-0.226 (0.386) p = 0.558
$Average Q_{m,self}$	0.467	0.462	0.464	0.456
Panel C: First Stage - $Q^{FH}$				
$Q_{m,self}$	0.044*** (0.008) p = 0.00000	0.019*** (0.005) p = 0.0003	0.025*** (0.006) p = 0.00002	0.043*** (0.010) p = 0.00002
$TQ_{m,self}$	-0.005 (0.005) p = 0.285	0.004 (0.003) p = 0.171	0.002 (0.004) p = 0.676	-0.004 (0.006) p = 0.532
$F_{stat}$	34.73	15.19	20.71	27.09
Panel D: First Stage - $TQ^{FH}$				
$Q_{m,self}$	-0.041*** (0.016) p = 0.010	-0.059*** (0.022) p = 0.008	-0.056*** (0.017) p = 0.002	-0.040** (0.018) p = 0.024
$TQ_{m,self}$	0.091*** (0.025) p = 0.0004	0.094*** (0.033) p = 0.005	0.095*** (0.028) p = 0.001	0.088*** (0.023) p = 0.0002
$F_{stat}$	13.73	9.97	13.28	17.03
Panel E: Second Stage				
$T$	1.288 (1.168) p = 0.270	1.374 (1.479) p = 0.354	1.524 (1.255) p = 0.225	1.143 (1.380) p = 0.408
$TQ^{FH}$	-3.473 (3.689) p = 0.347	-3.871 (4.750) p = 0.416	-4.250 (3.993) p = 0.288	-2.518 (4.508) p = 0.577
$Q^{FH}$	2.396 (6.671) p = 0.720	26.214 (21.351) p = 0.220	9.034 (12.895) p = 0.484	1.493 (7.880) p = 0.850
$Average Q^{FH}$	0.306	0.304	0.302	0.302
Observations	12,293	8,370	10,693	7,487

*Notes:* This table reports the estimation results for females in locality-based markets, using the  $Q^{FH}$  intensity measure, of the following regressions: difference-in-means, reduced-form, two first-stage equations and second-stage. All of the regressions include allocation-unit fixed-effects, as well as the set of background characteristics. Standard errors are in parentheses. In panel A, the main variable is the treatment value indicator. In this panel, the standard errors are clustered at the allocation-unit level only. In the rest of the estimations, the standard errors are clustered at both the allocation-unit and at the market level. Panels B report the estimates of the reduced-form interaction model of  $Y$  on  $Q_{m,self}$ . Sanderson-Windmeijer (2014) conditional F statistic are reported at the bottom of panels C and D. Panel E reports the second stage IV estimates. The average  $Q$  is reported at the bottom of panels B and E. The estimations are performed on four sub-samples: all available observations; markets with high unemployment rate (above the 75th percentile); markets with high amount of participants in the RWG relative to the local labor force (above the median); and markets with less than 15,000 residents in the ages 20-59 (2015). All samples contain only IA recipients only. In addition, individuals who last reported to their local employment office before their allocation date are excluded from all samples, as they could not have been influenced by their treatment status.