

## *How does a rating system affect private providers' market shares in active labor market policy?*

Christer Gerdes, Swedish Employment Service; Swedish Institute for Social Research;  
[christer.gerdes@arbetsformedlingen.se](mailto:christer.gerdes@arbetsformedlingen.se)  
[christer.gerdes@sofi.su.se](mailto:christer.gerdes@sofi.su.se)

### ***Abstract***

One service in Swedish active labor market policy is job search assistance. For several years unemployed using this service could choose among several private providers of job coaching. In 2013 a performance indicator of job coaching providers was made available at an online website. The indicator, called "rating", was based on previous records on how successful job coaches had been in helping participants to find work. This study examines to what extent this information affected job coaches' possibility to attract new participants. The results indicate that rating established a (causal) link between "quality" of services and number of new participants.

***Paper prepared for the 5th Joint IZA/IFAU Conference on Labor Market Policy Evaluation in Uppsala, October 09, 2014 - October 10, 2014***

## *1 Introduction*

This study presents a first evaluation of the introduction of a quality indicator for providers of job coaching in Sweden in April 2013. There are several reasons to offer an assessment (subsequently called rating) of provider performance in a system based on free choice by consumers. In the particular case of job coaching the main reason for introducing a rating has been to offer job seekers appropriate information to increase their chances of finding employment with the help of a job coach. Ideally this would lead to that resources through job seekers' choices would be allocated to those private providers of job coaching who best met their objectives, ultimately leading to more efficient services within job search assistance.

Since the change of government in Sweden in autumn 2006, there has been an opening of broader parts of publicly provided active labor market services, among others giving the job seeker the possibility to choose among several private providers within certain services; the regulation underlying this way of contracting private providers is called “the Act on System of Choice” (“Lagen om valfrihetssystem”, LOV). For a system of choice to work properly requires that job seekers make their choice based on reliable information on what differentiates the various private providers, especially how well they manage to help job seekers to find employment. Without information about how such companies are performing, it becomes difficult to make rational choices. The need for information has been highlighted in the literature on privatization of services that have the character of being public goods; see e.g. Shleifer (1998). Lundin (2011) describes in his study of private actors in labor market policies readily available and reliable information as the key to enable participants to make rational choices.

Job seekers registered with the Public Employment Service (PES) during the period January 2012 - October 2013 had the option to choose a provider of job coaching according to LOV regulation. A supplier of job coaching was expected to help the job seeker to find work, for example by getting the job seeker to connect with different employers, helping to prepare for a job interview, giving advice how to write résumés and more. There were no clear instructions how a provider of job coaching was expected to work with its clients more in detail. The main reason for this was to open up for new and

innovative ways to provide job search services. Instead of detailed regulations on how the service should be performed the PES focused on creating financial incentives for the providers of job coaching to help job seekers to find work as soon as possible. This was intended to be achieved through a payment model that rewarded the suppliers of job coaching in case they were successful.<sup>1</sup> Case workers at the Swedish PES were not allowed to recommend providers of job coaching, even if they would have had some knowledge that some providers were likely to be better than others. This restriction was imposed to ensure impartiality from the side of case workers at the PES. To resolve this kind of information shortages, in April 2013 a rating system for job coaching was published on the PES website.

The rating system is constructed such that providers were awarded a three graded rating based on how well they succeeded in getting participants into work. The best providers were awarded three stars, those with an average record two stars, while those providers who were less successful than others received only one star. About half of all providers offering job coaching services did not receive a rating value. The PES decided in advance that only those providers with at least ten participants within the period underlying the calculation of the rating would have their rating published. The main reason for putting up such requirement was to guarantee reliable estimations of the rating values. See Appendix, A.1, for a detailed explanation of how the rating values were calculated. In the following analysis only providers of job coaching are included that received a rating in April 2013.<sup>2</sup>

---

<sup>1</sup> Provider of job coaching gets paid according to a payment model that is divided into three steps. Firstly, there is an initial award of 5000 SEK for each participant who starts the program and an additional amount of 5000 SEK if the participant remains with the supplier during the entire program period. In case the participant begins to work during this period or within a further period of 30 days after the end of the program, the provider will be paid an additional amount of 5000 SEK. Overall, a provider of job coaching can get 15000 SEK for one participant.

<sup>2</sup> We will address the question of the importance of not having received a rating in a separate chapter at the end of the study.

The question addressed in this study is as follows: Does information about the coaches' previous performance play a role in the ability of these operators to attract new participants?

Information regarding private actors' previous performance in active labor market policy (ALMP) is generally scarce. Most studies on the use and usefulness of performance measures in labor market programs have been focusing on public management and accounting. Partly, this follows from the fact that a system of allowing free choice of provider in a voucher based system is still rather rare. To the best of my knowledge, no study has so far addressed the specific issue of use and usefulness of performance information in ALMP for choosing a provider in a voucher based system.<sup>3</sup> During the last decade a number of countries (such as Australia, Denmark, England, Germany or the Netherlands) have experimented with voucher based systems in ALMP, e.g. for job training, where the voucher is used to pay the training provider. The results of studies evaluating these systems have been rather mixed. The evaluations conducted for Sweden that looked at the effect of privatization of parts of ALMP have been focusing on the question if private actors are better at getting job seekers to work than the PES's own case workers; see e.g. Gartell (2011) or Liljeberg et.al. (2012). Both studies do not find major differences between private and public job search assistance after controlling for observable

---

<sup>3</sup> In other areas, such as (high) school choice, the use of performance indicators has been studied on a more regularly basis. Most of these studies look at ranking of schools, e.g. by the use of school league tables, and universities, e.g. high school rankings. Sometimes also the issue of stratification (sorting/segregation) caused by high school rankings or school league tables has been addressed. Most studies point to that there is an effect on stratification; see Clarke (2007) for a literature survey. One opposing view is provided by Burgess et al (2013), claiming that there is no evidence of social stratification. Even within health care there are a number of studies examining the impact of quality indicators, see for example Jin and Sorensen (2006) and Varkevisser et al. (2012). The first study finds a "meaningful influence on individuals' choices", while the other finds that patients have a high propensity to choose hospitals with a good reputation. So in both studies there is a positive connection between a positive rating and the actual choice. One interesting study on the influence of quality rating is Hilger et al. (2011), discussing the possible impact of higher visibility that rating might lead to, i.e. that it is not the information of the rating as such that affect consumer choice.

background factors among participants.<sup>4</sup> The present study does not aim to compare private and public job search assistance. Instead it focuses on whether there is empirical evidence that job seekers make use of information about the performance-based assessments of private providers of job coaching. This means that here only private providers of job coaching are studied.<sup>5</sup>

The prospects of conducting an evaluation of the introduction of the rating are somewhat limited due to the fact that the rating values were launched simultaneously all over the country. Consequently, there is no opportunity to see how the choice of job coaches would have looked like without the publication of the rating values (the counterfactual outcome). However, to some extent it has been possible to conduct estimations based on non-continuous changes brought about by regulatory thresholds, which will allow us to measure causal effects in different quasi-experimental settings.

The following sections examine the relationship between a job coach's rating value and the possibility to contract new participants in a number of different settings. To start with, section 2 present some upfront statistical estimations focusing on the association between the numbers of new participants and rating values, while in section 3 we look at the "market share" before and after

---

<sup>4</sup> One problem with this type of evaluation is that it is based on the participants' own decision whether to use a private provider of job coaching or if they prefer using a job coach connected to the PES, which raises concerns of selection bias. A recent study by Benmarker et al. (2013) looks at the outcome of a trial scheme where participants randomly were assigned to private placement agencies. They did not find any major differences in their chances of finding a job compared to the control group that were participating in the PES ordinary activities. For the Netherlands, a country more prominent when it comes to privatization of its employment services, it has been shown that privatization has not led to an improvement in the outflow to work, see e.g. Koning and Heinrich (2010). As a consequence, outsourcing of welfare services was partly reversed in 2007. In their review of the Australian and the Dutch system Struyven and Steurs (2005) find that privatization of parts of the employment service has meant some cost savings (in Australia), but it has not been accompanied by an improvement in the efficiency of services.

<sup>5</sup> Those choosing among private providers are probably positively selected from the group of job seekers registered with the PES. Gartell (2011) reports that individuals who are less attached to the labor market prefer internal job coaching by PES case workers over external job coaching.

publication of the rating. Section 4 shows the outcome of statistical calculations measuring causal relations between job coaches’ rating value and the ability to attract more new participants based on regression discontinuity estimations. In section 5 the impact of not having been awarded a rating value will be analyzed. Section 6 summarizes and discusses the results.

***2 Do better rated job coaches get more participants?***

One principal aim of introducing a performance-based assessment was to facilitate participants of job coaching to choose providers that proved to be more successful in helping participants to find a job, which should give those providers better opportunities to attract new participants and thereby raise their market share.

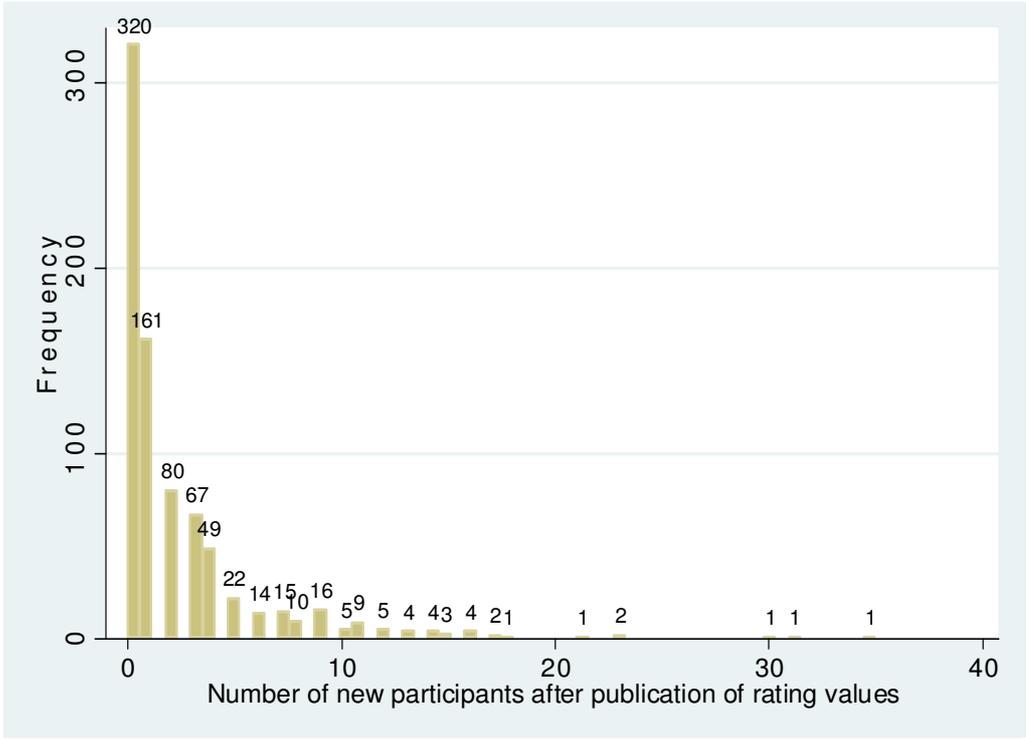
Table 1 presents some descriptive results regarding the providers of job coaching that were awarded a rating in April 2013. It shows that about 19 percent of job coaching providers (of those who actually received a rating) were assigned one star, about 63 percent two stars, and about 17 percent three stars. Also shown are the participants’ average number of participants according to the time period when the rating values were calculated. The average number of participants is somewhat larger for those providers with two stars. To some extent this follows from that larger providers were by design given more weight when calculating the star rating values; see Appendix A1 for more details.

**Table 1.** Distribution of providers of job coaching according to their rating.

	Number of providers	Number of participants during the time rating values were calculated	Std. dev.
Rating values			
One star	155	23.79	20.65
Two stars	506	36.31	32.09
Three stars	136	29.43	34.18
Total	797	32.70	30.97

In figure 1 the number of new participants after the period 30 April to 30 August 2013, i.e. from the date the rating was published to the date the data underlying the estimations of the current study was extracted, is shown. A few providers received more than two new participants, while 320 providers did not receive any new participants.

**Figure 1**



In this chapter the relation between the scope of getting new participants and a job coach's rating value is scrutinized. Due to the fact that the job coaches' results (none, one, two, three, etc., new participants) are an enumeration of individual events, also called "count data", one has to use an adequate estimation model. Here, a negative binomial regression model has been used.<sup>6</sup> Because larger providers of job coaching can utilize economies of scale (or because they are more known companies in the market for personnel services), there are different preconditions to tie new participants across different providers, an aspect that has to be controlled for in order to find out on the importance of the

<sup>6</sup> For a detailed discussion of the various possibilities to analyze this kind of data, see O'Hara and Kotze (2010). The basic idea of the model is the assumption that the data are negative binomial distributed. We use the standard estimation procedure offered in Stata, see Pérez Trugliay (2009) for details.

rating itself. Furthermore, it is important to take into account differences across regional labor markets, as the situation on the local labor market may imply different opportunities for providers of job coaching to attract new participants. Therefore the models estimated include controls for different procurement areas where providers of job coaching had been operating and the "size" of the job coaches as measured by the number of participants in the period January to November 14, 2012, i.e. the period underlying the calculations of rating values. The first set of outcomes is shown in Table 2. The variable named "rating" measures the importance of having a rating value of one, two or three stars for the provider of job coaching to get new participants (note that only those providers with a rating are included in the estimations). The model also includes control variables for each of the delivery areas (as 0/1-indicator variables) and also an account of how many delivery areas a provider of job coaching has been working in.

The rating variable, which can take on the values 1, 2 or 3 stars, is positive and statistically significant, saying that the more stars a provider of job coaching has been awarded the more participants did sign up after publication of the rating. Next we look in more detail at the role of a supplier of job coaching with one, two, or three stars in the rating that was published in April 2013, see column (2). The model has the same structure as in column (1), with the difference that now there are two separate indicators included: one indicator that marks a rating equal to one star, the other a rating equal to three stars. By including rating values in this way allows us to look for non-symmetric relationship between rating and outcome. The respective coefficient estimates indicate the difference in having one and three stars compared to having two stars, respectively. According to the estimations shown in Table 2, column (2), the provider of job coaching with three stars receive about 37% more new participants (during the period studied) compared to a provider of job coaching with two stars controlling for the delivery area and initial size.<sup>7</sup> Suppliers of job coaching with just one star receive

---

<sup>7</sup> Interpretation "% more participants" is not entirely correct. More accurate would be to write that there was a "37% higher incidence in log differences" between providers of job coaching with two compared with providers of job coaching with three stars. Since such an interpretation is somewhat difficult to handle we use the

about 14% fewer new participants; however, the latter result is not statistically significantly different from 0, which means that one cannot establish any statistically significant differences between job coaches with one and two stars. Thus, it looks like that there is a larger positive impact of having three stars compared to having two stars, compared to what a provider of job coaching loses by having one star instead of two stars.

**Table 2.** The significance of the publication of the rating values. Actual number of participants since the publication of the rating values. Negative binomial estimations.

	(1)	(2)
<b>Rating values (1, 2, 3 stars)</b>	<b>0.255 ***</b>	
	<b>(0.0893)</b>	
<b>One star</b>		<b>-0.137</b>
		<b>(0.156)</b>
<b>Two stars</b>		<b>reference</b>
<b>Three stars</b>		<b>0.367 ***</b>
		<b>(0.116)</b>
Observations	797	797
Ln(alpha)	-0.103	-0.100
	(0.103)	(0.105)
Robust standard errors in parentheses *** p <0.01, ** p <0.05, * p <0.1		

Note: in addition to those above reported control variables also controls for delivery areas, the number of participants at the time of the determination of the rating value, and the total number of areas a provider of job coaching has been represented, have been included.

An alternative to focus at the importance of having one, two or three stars for *how many participants* a job coach can attract is to investigate whether the job coach may receive any new participants at all, i.e. *one or more new* participants or not. Such an approach can be considered less

simpler formulation. See [http://www.ats.ucla.edu/stat/stata/output/stata\\_nbreg\\_output.htm](http://www.ats.ucla.edu/stat/stata/output/stata_nbreg_output.htm) for an example on the correct interpretation of the coefficient estimates in negative binomial estimations.

“partial” in the sense that we are reducing the significance of *the mere number of* new participants, which in general will benefit larger providers. The outcome in this situation thus becomes a binomial variable that can take the values 0 and 1, where "0" indicates no new participants and "1" at least one new participant. In the estimations shown below OLS is used.<sup>8</sup>

Column (1) of Table 3 shows that for each additional grade level (star) in the rating increases the probability of getting one or more new participants by more than 8%. Results with individual indicators of one star or three stars are reported in column (2). Suppliers of job coaching with three stars are about 8% more likely to have at least one new participant compared to the suppliers of job coaching which have been awarded two stars. Similarly, providers of job coaching with only one star are about 9% less likely to get one or more new participants. Accordingly, the impact of rating for the probability of attracting one more participant is about as large in both negative as positive direction, i.e. participants are opting out inferior providers of job coaching as much as they opt for suppliers of job coaching with a better rating value compared to providers of job coaching that had two stars.

**Table 3.** The significance of the publication of the rating values. Binomial outcome none vs. any participant. OLS estimations.

	(1)	(2)
<b>Rating values (1, 2, 3 stars)</b>	<b>0.0848 ***</b>	
	<b>(0.0268)</b>	
<b>One star</b>		<b>-0.0915 **</b>
		<b>(0.0441)</b>
<b>Two stars</b>		<b>reference</b>
<b>Three stars</b>		<b>0.0775 *</b>
		<b>(0.0442)</b>
Observations	797	797
R2	0.262	0.262

<sup>8</sup> Alternatively, one could use probit or logit estimates, however, OLS has the advantage that the coefficient estimates are easier to interpret.

Robust standard errors in parentheses \*\*\* p <0.01, \*\* p <0.05, \* p <0.1

---

Note: in addition to those above reported control variables also controls for delivery areas, the number of participants at the time of the determination of the rating value, and the total number of areas a provider of job coaching has been represented, have been included.

It is important to keep in mind that these figures are showing associations and should not to be interpreted as causal effects. One cannot know if it was the publication of the rating values that led to that the suppliers of job coaching received more participants, or if this was due to other factors. For example, it might have been the case that some participants received signals that some suppliers were better / more successful by using their personal contacts or networks both before and after the rating was published.<sup>9</sup> We will address this issue more explicitly in the following sections.

### *3 Estimation of the “net impact” of the publication of the rating values*

As stated earlier it is difficult to measure the real impact of rating values for job seekers' choice of provider. Other factors may have been important for the outcome, e.g. job seekers who were about to choose a provider of job coaching may have used some other kind of information. Such information could be shared through a participant's social networks or through information from employment agencies (although the latter was not allowed according to the PES instructions).<sup>10</sup> It is not possible to clarify all processes underlying each individual's choice of job coach. Yet, it might be possible to get a reliable appreciation of the overall impact of offering a rating to increase better providers' share of the market. This can be achieved by studying the choice of providers of job coaching before and after publication of the rating values. We thus exploit the fact that during a period prior to the publication of the rating values there were participants who chose a provider of job coaching without having access

---

<sup>9</sup> One problem in estimating the influence of quality indicators on demand is that products receiving a positive rating tend to be of better quality, and it is difficult to determine whether the rating or the quality is responsible for higher demand, see Reinstein and Snyder (2005).

<sup>10</sup> Previous research has shown that access and dissemination of information usually go via social networks, where information dissemination favors some groups more than others; see e.g. Hipp and Warner (2007).

to public information about the job coach’s past performance. See Figure 2 for the time frames that apply to those persons included in the calculation of the rating values, and the persons who define the comparison group in the subsequent estimations.

**Figure 2**



By looking at the difference between the job seekers who chose a provider of job coaching during the period November 14, 2012 to April 30, 2013, with job seekers who chose a provider of job coaching after publication of the rating values, we can produce a more reliable estimate of the publication’s actual importance. In practical terms this is done by including the two time periods enclosing the publication of rating values the 30th April 2013 in one and the same estimation model. By including interaction variables indicating the period after the publication of the rating we aim to achieve a more reliable estimate of the "net-impact" between the job coaches’ *published* rating values regarding their importance to attract new participants.

However, there is one caveat to this approach. The number of participants decreased profoundly around the time of the publication of rating as a result of the PES’s decision to withdraw job coaching as a service in ALMP to the end of 2013. Because of that the actual number of participants in job coaching fell considerably in the months after the rating was published. In order to make outcomes comparable the following analysis is conducted by looking at the *market share* instead of the mere number of participants for each provider in the two periods. This is done by dividing the number of participants a provider of job coaching had under the period before (the “pre-publication” period) and after (the “post-publication” period) the 30th of April with the total number of people who chose a supplier within the corresponding period. Thus, instead of the number of new participants as before,

here it is the market share a provider of job coaching had before and after the publication of the rating values that will be studied.

Since the outcome variable under study has changed and now considers shares instead of numbers, a different type of estimation model is applied. Here so called Tobit-estimations are used, which take into account that the outcome is not overall normally distributed.<sup>11</sup> The estimated model looks as follows:

$$\text{Share of new participants} = \beta_1[\text{rating}] + \beta_2[\text{rating} \times \text{after publication}] + \beta_3[\text{after publication}] + \mathbf{XB} + \varepsilon$$

The variable "after publication" is an indicator that takes the value 1 or 0 according to the time of publication of rating values, and where the variable "rating  $\times$  after publication" is defined as the product of the variable "rating" and "after publication". Variable vector  $\mathbf{X}$  includes indicator-variables for delivery areas and the number of local delivery areas a provider of job coaching has been registered. The latter is done for the purpose of accounting for differences across providers in terms of their (initial) size.

The main interest is on the coefficient  $\beta_2$  indicating the marginal role of having a particular rating value *after* the rating has been made public. Figure 3 shows the distribution of market shares before and after the publication of the rating values. It indicates that many providers of job coaching did not receive any participant after the publication of the rating values the 30th April 2013, i.e. the frequency at point 0 is much higher in the preceding period. To some extent, this reflects that there were fewer people registered at the employment office who got the opportunity to choose a job coach. For this reason, we choose to include a control variable that captures the systematic decline in the number of participants. This is done by including the control variable "after publication" that is coded as 1 after and 0 before the publication of the rating values.

---

<sup>11</sup> See the discussion in McDonald and Moffit (1980) for clarifications how to use Tobit estimations. Using OLS outright would lead to systematic underestimation of the coefficient estimates one wants to measure. As the outcome does not involve count data it is not appropriate to use negative-binomial estimations as in the previous section.

**Figure 3**

Distribution of market shares before and after publication of rating values

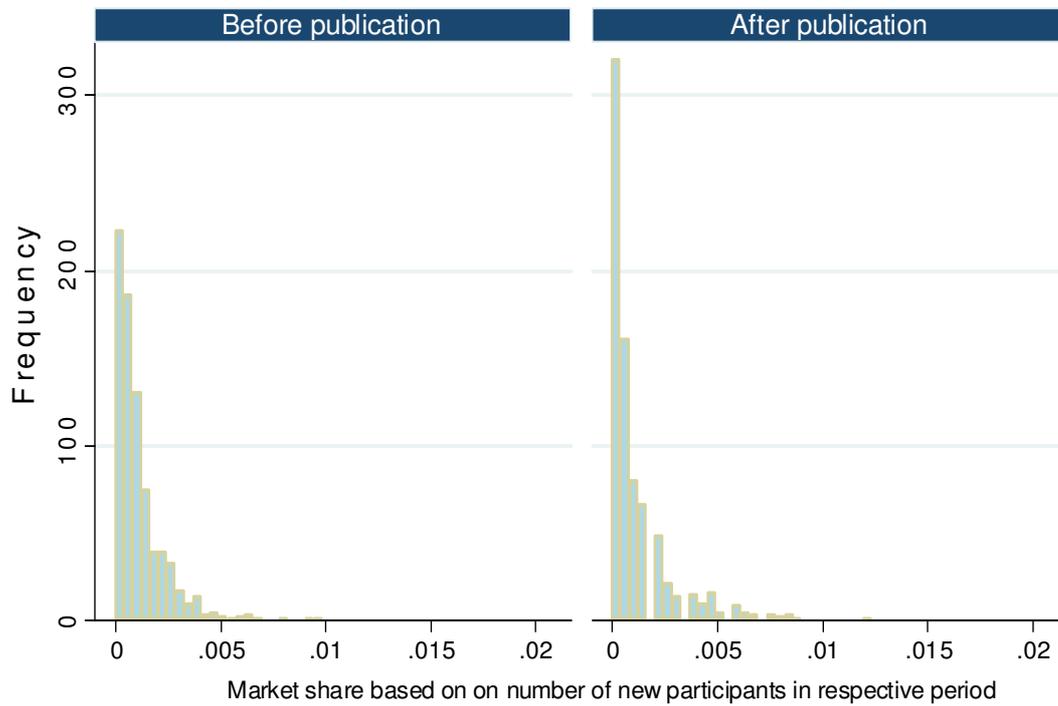


Table 4, column (1) shows the estimation results when using a single variable that takes values one, two, three stars. Our first variable of interest regards "Rating values (1, 2, 3 stars) after publication". As can be seen, the coefficient is positive and statistically significant (at the 10 percent level). The fact that the coefficient estimate is as small as 0.00027 reflects that there were many providers in the market, which resulted in an average market share (across both periods) of about 0.00125; this implies that the relative importance of having a better star rating is rather substantial, i.e. approximately 22% ( $0.00027/0.00125$ ).

**Table 4.** Effects of the publication of the rating values. Actual changes in market share.

Tobit estimations.

	(1)	(2)
Rating values (1, 2, 3 stars)	0.000122*	
	(6.54e-05)	
<b>Rating values (1, 2, 3 stars) after publication</b>	<b>0.000268*</b>	
	<b>(0.000155)</b>	
One star		-1.97e-05
		(9.63e-05)
Two stars		reference
Three stars		0.000232**
		(0.000115)
<b>One star after publication</b>		<b>-0.000328</b>
		<b>(0.000260)</b>
<b>Three stars after publication</b>		<b>0.000201</b>
		<b>(0.000236)</b>
After publication	-0.00103***	-0.000470***
	(0.000321)	(0.000111)
Observations	1594	1594
sigma	0.00170***	0.00170***
	(0.000106)	(0.000106)

Robust standard in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note: in addition to those above reported control variables also controls for delivery areas, the number of participants at the time of the determination of the rating value, and the total number of areas a provider of job coaching has been represented, have been included.

Column (2) presents results that apply to the situation when rating values are divided into two separate indicators, i.e. "one star" and "three stars" dummy variables. According to the estimations providers of job coaching with a grade of one star are losing market shares, compared with providers of job coaching that have two stars in the rating. The importance of having a rating value of three stars

compared to having two stars is positive, but somewhat smaller in size. However, both coefficients are not statistically significant.

Table 5 below shows the results of the impact of the rating values to have or not have received *one or more new* participants. Here we change the outcome variable from shares to an outcome that is either zero or one. Note that the interpretation of the estimates differs due to this change of outcome and that the results are to be read as probabilities.

**Table 5.** Effects of the publication of the rating values. Actual changes in market share > 0 or 0. OLS estimations.

	(1)	(2)
Rating values (1, 2, 3 stars)	0.00578 (0.0131)	
<b>Rating values (1, 2, 3 stars) after publication</b>	<b>0.0870 ***</b> <b>(0.0285)</b>	
One star		0.0176 (0.0225)
Two stars		reference
Three stars		0.0312 (0.0200)
<b>One star after publication</b>		<b>-0.129 ***</b> <b>(0.0469)</b>
<b>Three stars after publication</b>		<b>0.0419</b> <b>(0.0446)</b>
After publication	-0.526 *** (0.0603)	-0.336 *** (0.0219)
Observations	1594	1594
R2	0.334	0.335
Robust standard in parentheses *** p<0.01, ** p<0.05, * p<0.1		

Note: in addition to those above reported control variables also controls for delivery areas, the number of

---

participants at the time of the determination of the rating value, and the total number of areas a provider of job coaching has been represented, have been included.

The results indicate statistically significant positive association between a better rating value and the probability that a provider of job coaching would be able to attract one or more participants. Each star higher rating value increases the probability of getting new participants by more than eight percent, see column (1). This result is reminiscent to the situation when only the post-publication period was included, shown in table 3, column (1). If the rating values are divided into separate "star indicators", it turns out that it's more important to have a weak rating value than having a better rating value, see column (2). A supplier with a rating value of "one star" is 13 percent less likely to get one or more new participants as compared to one with two stars. For providers of job coaching with three stars, there is a marginally higher probability of getting one or more new participants compared to a providers with two stars, but the coefficient is not statistically significant. Overall, the results in Table 5 are consistent with those shown in Table 3: Worse rating decreases the chances of attracting new participants, and vice versa.<sup>12</sup>

#### *4 Estimating effects of rating in a quasi-experimental setting*

As has been mentioned earlier, the introduction of rating values does not allow estimating more general causal effects, because there is no possibility to quantify the counterfactual outcome, i.e. what would have happened if the PES would not have published rating values. However, it is possible to use a quasi-experimental approach to distinguish causal effects of the introduction of rating values regarding a subgroup of the studied job coaches. This can be done for the suppliers of job coaching that were just on the verge of getting one vs. two stars, and three vs. two stars, respectively. As shown in Figure A1 in the Appendix, there are two thresholds that determine the number of stars a job coach has been awarded. The underlying idea that will allow us to comment on the causal effects of having

---

<sup>12</sup> To some extent some job coaches might have chosen to end their activity as job coaches as they anticipated that a bad rating would minimize their chances of attracting new clients. The extent to which this has occurred is not possible to sort out in the context of this investigation.

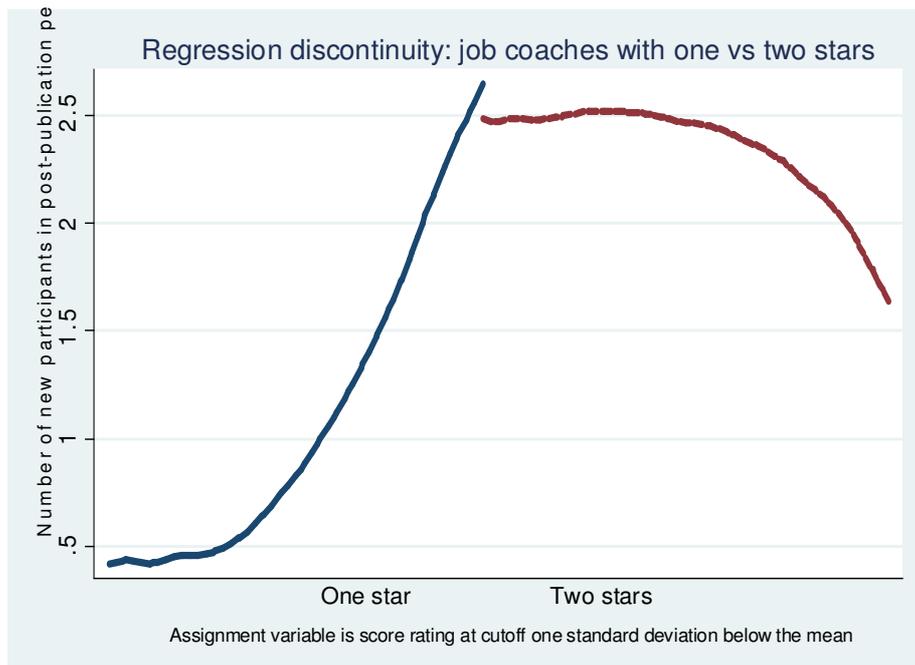
one, two or three stars in the rating is that it has been pure coincidence that determined whether a provider of job coaching close to the threshold ended up just to the left or right of it. It is reasonable to believe that the crucial assumption of randomness close to the cut off for this method to provide accurate results is fulfilled. The observations underlying the calculation of the rating values relate to a period long before it was announced that the PES would introduce a performance indicator. Even more, no one knew beforehand where the thresholds would be set, which made it impossible to “game the system” and end up with a particular rating value.

We conduct the analysis by studying the closer environment surrounding the threshold value in the framework of a regression discontinuity approach.<sup>13</sup> The following figures display the outcome as measured by the number of new participants since rating values were published the 30th April 2013 as the outcome (depicted on the y-axis) and the score values as the independent variable; see Appendix A1 for a detailed explanation of how the score values have been calculated. Figure 4 shows results at the threshold that determines whether a provider of job coaching received one or two stars. The figure suggests that there are signs of discontinuity at the threshold point, but the difference between having one vs. two stars is rather small; the precise estimates are shown in Appendix A2.

---

<sup>13</sup> See e.g. Imbens and Lemieux (2008) for a detailed discussion. In the calculations reported here we used the "rd" command in STATA, see Nichols (2011) for details.

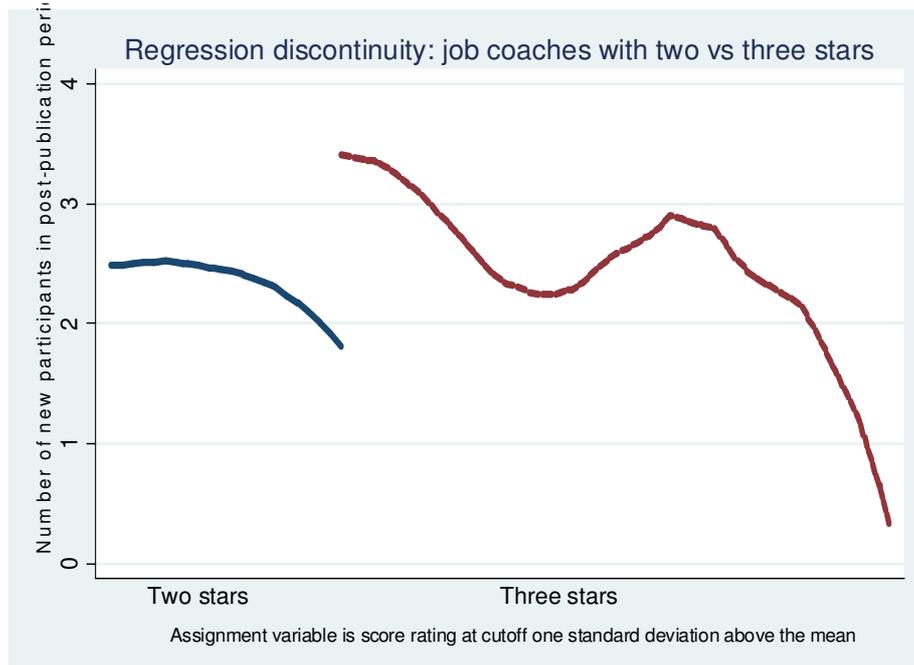
**Figure 4**



The situation looks quite different for those providers of job coaching just to the left and right of the threshold that determined if a job coach would get two or three stars, respectively. The figure shows a remarkable difference, which can be measured to about 1.5 additional participants, i.e. a difference between 3.5 new entrants for a provider with three stars, compared with just 2 new entrants for a provider with two stars. This should be read as follows: providers of job coaching that had a calculated score value that were slightly larger than the threshold 0.068 and thus received three stars had approximately 1.5 more participants than providers of job coaching that lay just to the left of the threshold (and thus received two stars).<sup>14</sup>

<sup>14</sup> The according Wald estimations include controls for the 52 delivery areas, the number of participants at the time of the determination of the rating value, and score-values squared. The results are statistically significant even without including such controls.

**Figure 5**



When interpreting the results of this section, it is important to keep in mind that these are effects for providers of job coaching near the thresholds. Therefore the results should not be interpreted as average effects in the sense that they reflect the overall differences *for all* providers of job coaching with three stars compared to *all providers of job coaching* with two stars.<sup>15</sup> However, the results shown in figure 4 and figure 5 are consistent with the outcome of the estimations shown in table 2, column (2), indicating that the relative effect of having three stars compared to two stars is more important than to have a rating below average, i.e. a rating similar to one star compared to two stars. This indicates that the found results are likely to be of more general relevance than just applying close to the cut-of that determine which provider received two or three stars.

We test the reliability of this estimation procedure by performing similar calculations for the period preceding the publication of the rating values. These are calculated for a period where participants chose a provider of job coaching without having knowledge of the job coaches' past

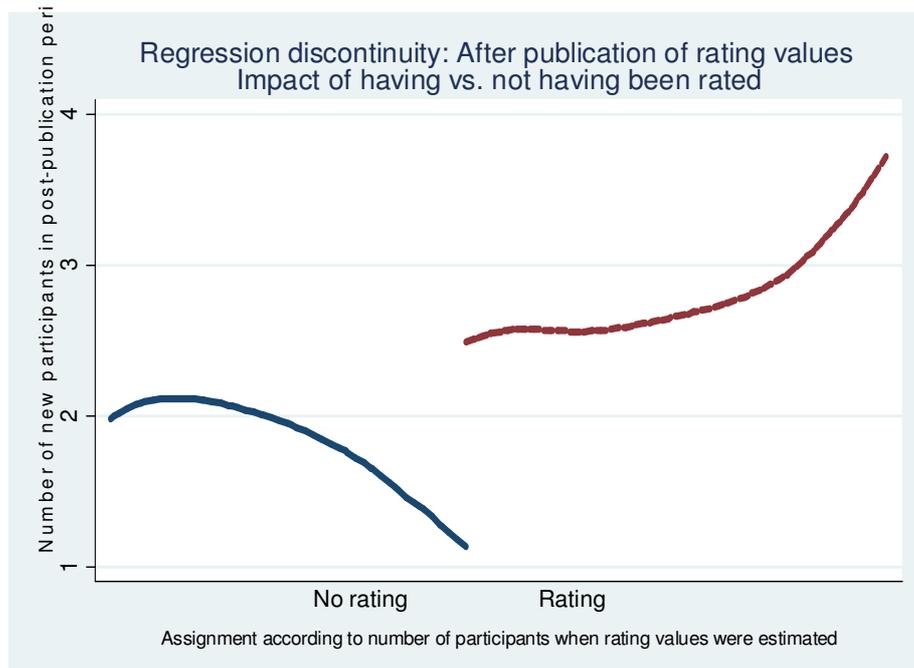
<sup>15</sup> The method is not appropriate to analyze non-continuous outcome variable see Imbens and Lemieux (2008), p 618 ("Assumption 2.2"). For this reason, only estimates of the number of participants as outcomes are estimated, not 0/1 outcome of having or not having received one or more new participants.

performance as measured by the rating values (i.e. the “pre-publication” period shown in figure 2 above). Since there was no published rating during that time, there should be no "discontinuity" at the thresholds of the score values that generate the respective rating value. The figures in Appendix A3 show that this is indeed the case. It supports the conclusion that the found discontinuities around the thresholds determining a star rating of two vs. three stars *after* publication of the rating values were caused by the publication and are not spuriously caused by some other factors.

### *5 The importance of having been rated vs. not having been rated*

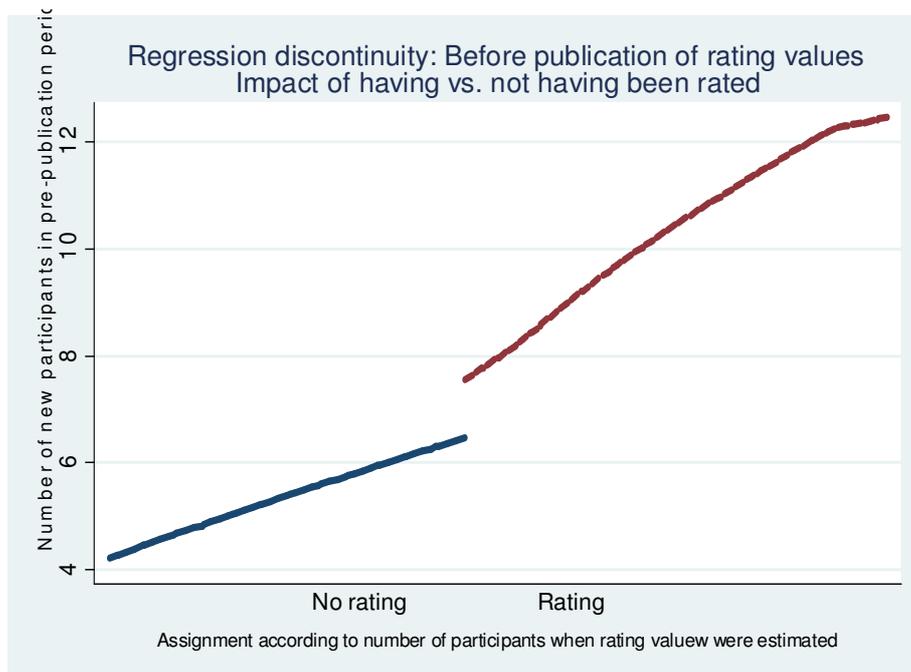
One issue that has not been addressed so far is whether it makes a difference to a provider of job coaching to have received a rating value, compared to not having received a rating value. This section examines this aspect. This is done by exploiting the fact that a provider of job coaching would receive a rating value in April 2013 only if that provider had at least ten participants under the period January to November 2012. This suggests that one can compare how those with nine or fewer participants were able to attract new participants compared with providers who had ten or more participants. The statistical method used here is similar to that in the previous section, making use of a regression discontinuity approach. The results in figure 6 show clear differences. A supplier who had at least ten participants in the period January-November 2012 received about 1.4 additional participants compared with a supplier who had nine participants or less in the same period. The difference is statistically significant at the 1% significance level. It is not possible to say what mechanisms that drive these results. To some extent it could be that a (non-) published rating leads to higher visibility of that provider, which might transmit to a larger propensity to attract new clients (see Hilger et al., 2011 for a discussion). Alternatively it might be the case that participants implied that no rating was seen a kind of “zero”-rating, which seen as an ordinal scale variable took a lower position than a rating of one, two or three stars.

**Figure 6**



As a sensitivity test a similar approach covering the period preceding the publication of the rating values are shown in figure 7. The relative difference is much smaller around the threshold. In contrast to the outcome shown in figure 6 it is not statistically significant (with a p-value of .340). This confirms that there indeed is a positive effect of receiving a rating, no matter what value, compared to not being awarded a rating for those providers close to the threshold.

**Figure 7**



## 6 Final Discussion

This study shows that there are clear indications that the publication of the job coaches' past performance in terms of rating values has been a relevant piece of information to the job seekers. The publication of the rating values clearly strengthened the link between "quality" (in terms of the registered ability to get participants to jobs) and the number of new participants that providers of job coaching were able to attract. A number of different estimation procedures have been used to examine that relation, all leading to consistent results, which strengthens the reliability of the found results. In particular the results based on regression discontinuity estimations provide convincing evidence. They show that it is of particular importance to have a better rating (three stars) for the purpose of attracting new participants. The relative size is notable and can be estimated to more than 80 percent increase: 3.5 participants for the job coach who got rating of three stars, compared with barely 2 participants for an approximately equal good provider who received two stars. On the other hand, if a provider of job coaching was awarded one instead of two stars does not seem to play as an important role. Furthermore, there is reliable evidence that job coaching providers with published rating (no matter whether they got one, two or three stars) attracted more participants than the job coaches who did not receive a rating.

From a theoretical perspective, the results imply that an essential condition for potential economic efficiency gains are met, namely that the suppliers that perform worse are losing customers, while those who perform better increase their market share. One aspect that has not been addressed in this study, and which is of vital importance for the issue of efficiency gains of providing a rating, regards the issue how the choice of provider of job coaching translates to improved opportunities to find work for those participating. This regards both those who chose the best rated providers, but even those who chose one of the lower ranked providers of job coaching. Another question left for future research regards the issue of social stratification, i.e. the question if the provided information is used equally across social groups. Research within other areas, especially in school choice, suggest that certain social groups are more likely to make use of performance indicators, such as school league tables, meaning that such kind of information is not used equally across groups, which can imply that access to good quality services will not be equally distributed in a system based on free choice of provider.

## *References*

Benmarker, Helge, Erik Grönqvist, Björn Öckert (2013), "Effects of outsourcing employment services: evidence from a randomized experiment", *Journal of Public Economics*, Vol. 98, 68–84.

Burgess, Simon, Deborah Wilson, Jack Worth (2013), "A natural experiment in school accountability: The impact of school performance information on pupil progress", *Journal of Public Economics*, Vol. 106, 57–67.

Clarke, Marguerite (2007). "The impact of higher education rankings on student access, choice, and opportunity. College and university ranking systems". *Global perspectives and American challenges*. Washington, DC, Institute for Higher Education Policy.

Gartell, Marie (2011), "Interna och externa coachningstjänster- en utvärdering av resultatet", Arbetsförmedlingen URA 2011:2.

Hilger, James, Greg Rafert and Sofia Villas-Boas (2011) “Expert Opinion and the Demand for Experience Goods: An Experimental Approach in the Retail Wine Market”, *The Review of Economics and Statistics*, Vol. 93 (4), 1289–1296.

Hipp, Lena and Mildred E. Warner (2007), “Market Forces for the Unemployed? Training vouchers in Germany and the U.S.”, *Social Policy & Administration*, Vol. 42, 77-101.

Imbens, Guido W. and Thomas Lemieux (2008), “Regression discontinuity designs: A guide to practice”, *Journal of Econometrics*, Elsevier, Vol. 142 (2), 615–635.

Institute for Digital Research and education (idre),  
[http://www.ats.ucla.edu/stat/stata/output/stata\\_nbreg\\_output.htm](http://www.ats.ucla.edu/stat/stata/output/stata_nbreg_output.htm) .

Jin, Ginger and Alan Sorensen (2006) “Information and Consumer Choice: The Value of Publicized Health Plan Ratings,” *Journal of Health Economics*, Vol. 26 (2), 248–275.

Koning, Pierre and Caroline J. Heinrich (2010), “Cream-Skimming, Parking and Other Intended and Unintended Effects of Performance-Based Contracting in Social Welfare Services”, IZA Discussion Paper No. 4801

Liljeberg, Linus, Sara Martinson och Jonas Thelander (2012), ”Vad innebär det att bli coachad? En utvärdering av jobbcoachningen vid Arbetsförmedlingen”, IFAU-Rapport 2012:24.

Lundin, Martin (2011), ”Arbetsmarknadspolitik – om privata komplement till Arbetsförmedlingen”, in (red) Laura Hartman *Konkurrensens konsekvenser. Vad händer med svensk välfärd?*, SNS Förlag

McDonald, John F. and Robert A. Moffitt (1980), “The Uses of Tobit Analysis”, *Review of Economics and Statistics*, Vol. 62, 318–321.

Nichols, Austin. 2011th rd 2.0: Revised Stata module for regression discontinuity estimation.  
<http://ideas.repec.org/c/boc/bocode/s456888.html>

O'Hara, B. and D. Robert Johan Kotze (2010), "Do not log-transform count data", *Methods in Ecology and Evolution*, Vol. 1, 118–122.

Pérez Trugliay, Ricardo Nicolás (2009), "Applied Econometrics using Stata", Department of Economics Harvard University, Early draft

Reinstein, David A. and Christopher M. Snyder (2005), "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics", *The Journal of Industrial Economics*, Vol. 53 (1), 27–51.

Shleifer, Andrei (1998), "State versus Private Ownership", *Journal of Economic Perspectives*, Vol. 12, 133–150.

Struyven, Ludo and Geert Steurs (2005), "Design and redesign of a quasi-market for the reintegration of job seekers: empirical evidence from Australia and the Netherlands", *Journal of European Social Policy*, Vol. 15, 211–229.

Varkevisser, Marco, Stephanie A. van der Geest, and Frederik T. Schut (2012) "Do Patients Choose Hospitals with High Quality Ratings? Empirical Evidence from the Market for Angioplasty in the Netherlands," *Journal of Health Economics*, Vol. 31 (2), 371–378.

## *A.1 Calculation of rating values*

Rating values regarding the job coach's past performance were obtained by calculations based on the information of participants who were registered with that job coach in the period January to November 14, 2012. A supplier could operate in one or more of the 52 different supply areas where companies (i.e. providers) could sign up by contacting the local PES. When calculating the rating value, each provider of job coaching within an area has been treated as an isolated provider, which means that if a supplier had at least ten participants in that area this would result in a unique rating for that provider for that area. This implies that larger providers that had been actively engaged in several supply areas did receive as many, possibly different, ratings. See the map (Figure A2) showing how supply areas were distributed across Sweden.

More in detail the rating is done as follows. For every job seeker in job coaching that was registered between January and November 2012 a value is calculated that indicates the likelihood that he/she will get a job within a certain period of time. The calculation is done based on a statistical model that take into account the individual's age and sex, country of birth, unemployment record, job skills, education, etc. By adding the different expected values for all job seeker that were registered with a supplier of job coaching, a measure of the expected outcome (i.e., the number of participants expected to get a job) is estimated for that job coach. This number is then compared with the actual outcome, i.e. how many participants that were registered with that supplier that did actually find a job during the period measured.

Example: Suppose that a supplier had three participants, which, based on estimations were estimated to have a probability of getting a job that is equal to 0.1, 0.2, and 0.4, respectively. Putting these three values together this adds to 0.7 ( $= 0.1+0.2+0.4$ ). If it turned out that one of these three people got a job, the actual value becomes equal to 1 ( $= 1+0+0$ ). According to this example the difference between the actual and expected value becomes 0.3 ( $= 1-0.7$ ). Dividing the result by the number of participants returns 0.1 ( $= 0.3/3$ ). For ease of presentation, let call this number "score value". This score value is the basis for determining how well a supplier has done in helping participants to find a job in comparison to other suppliers of job coaching. In the estimations a minimum of ten participants for a

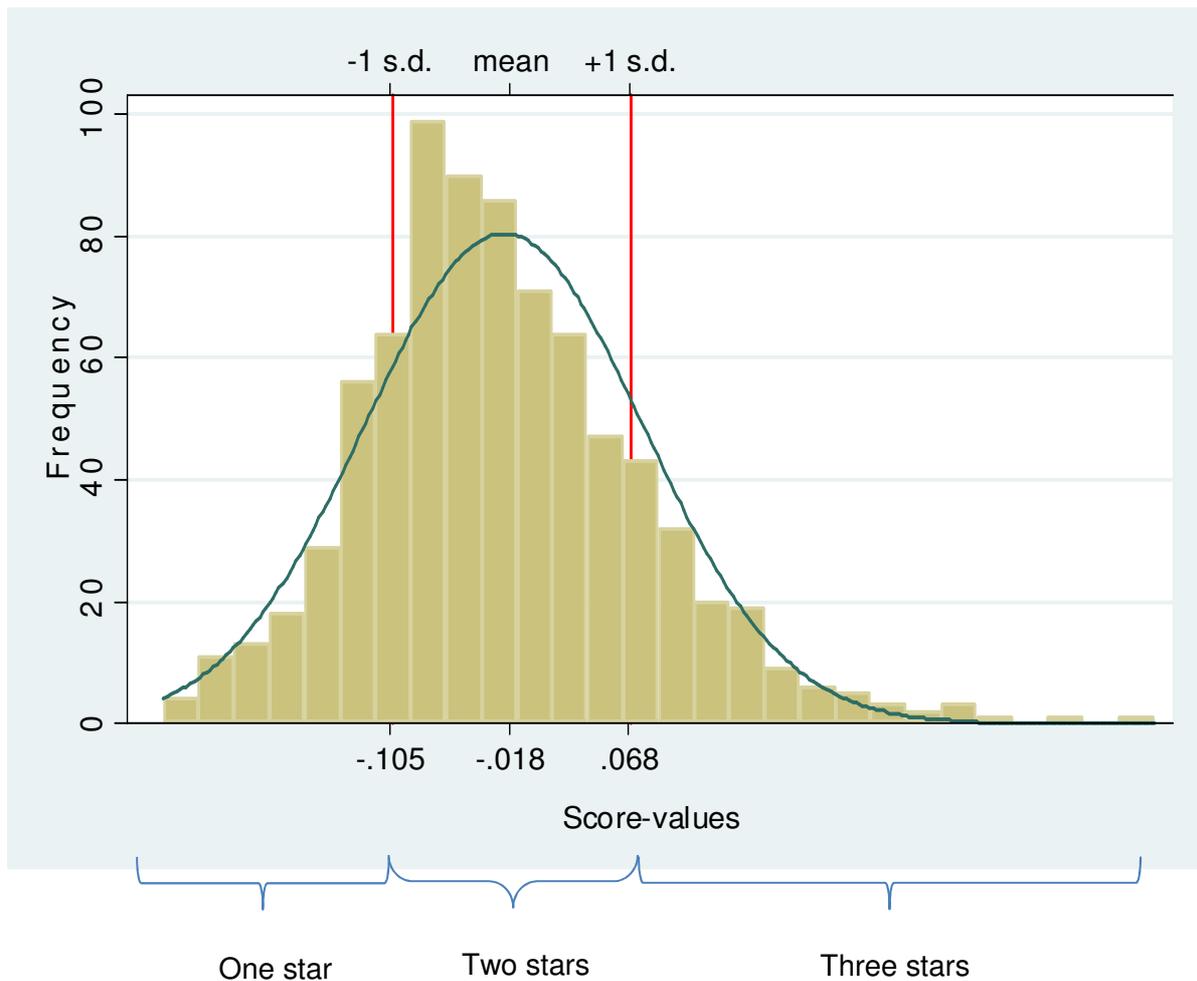
job coach in each respective supply area under the observation period January to November 2012 were required to trigger publication of a rating value, which means that those who did not have at least ten participants did not receive a rating value. The reason to apply a lower bound of ten participants was to guarantee that the estimated ratings were actually the result of a provider's work rather than the outcome of random events.

Each score-value is related to all other providers' score-values. If a supplier's score-value deviates positively from the average of all providers' score-values, the provider will receive a rating value of three stars. Those close to the average get two stars, while those who are below average receive just one star. To increase the reliability of the calculation of the mean value, values are weighted by the number of participants who were registered with a provider. This means that suppliers with a larger number of participants gain more weight than providers with fewer participants in the calculation of the mean and the measured standard deviation. This implies that larger providers will most likely receive two stars as they more likely are located closer to the mean in the distribution, which by design generates two stars.

### ***Defining threshold values for the rating***

An important issue is how much a rating value needs to differ from the mean to result in a rating better or worse than average. The PES decided that a threshold corresponding to one standard deviation larger or smaller than average should be applied. The distribution of the score values follows approximately a normal distribution, see the figure shown below. Figure A1 shows the spread of score values. The borders, defined by one standard deviation larger or smaller than the mean value, determine which rating a supplier is attributed. When the rating was published in late April 2013 those providers with score values lower than  $-.105$  received one star, those with score values between  $-.105$  and  $+.068$  received two stars, and those who were above  $+.068$  got three stars.

**Figure A1**



In defining transition to work there is a window of time that begins with the date of the recorded start for a participant using the services of a job coach and ends 120 days later (the time within the program plus an additional period of 30 days). If the registered participant within this interval switches to a registration code that marks that she has obtained work, it counts as 1 ("Successful match"), if not as 0 ("No match"). One restriction here is that the registered work must have lasted for at least 30 days. To determine whether the efforts of a job coach resulted in a successful match to job or not it is therefore necessary to apply a follow-up period of 150 days (120+30). This also explains why no individuals that were registered to a job coach after the 14th of November 2013 were included in the calculation of the rating values (the data underlying the calculations of score values were extracted the 14th of April 2013): The follow-up timeframe would not be sufficient (i.e. at least five months). See figure A2 for a clarification on what time periods that form the basis for calculating the rating values.

**Figure A2**

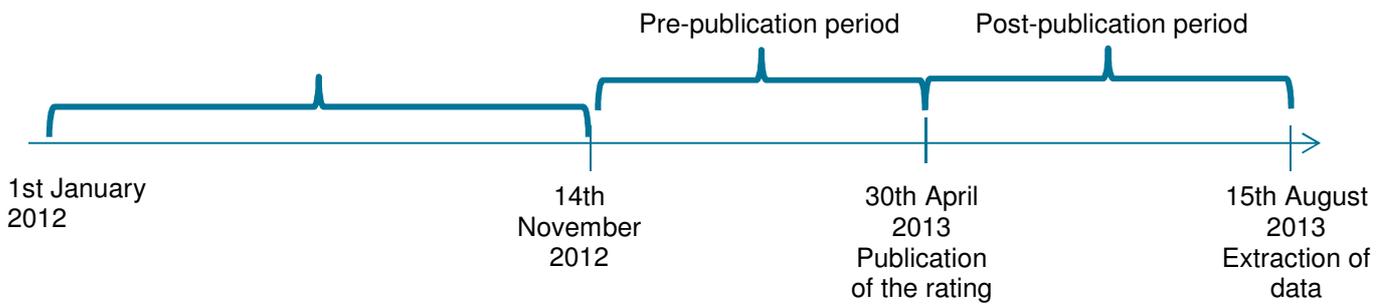
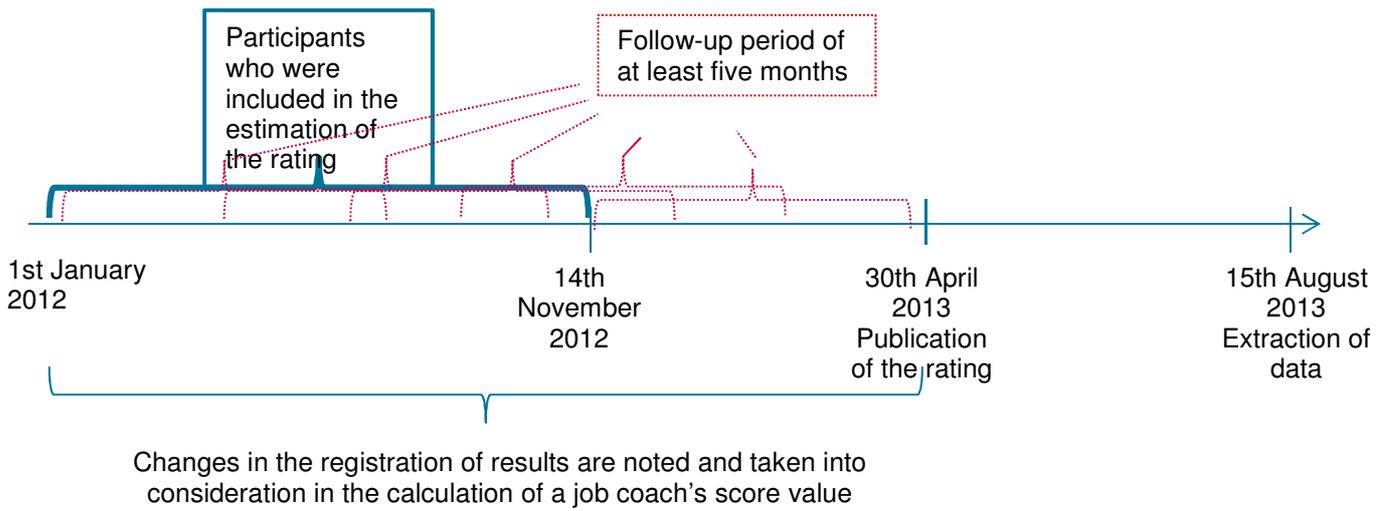
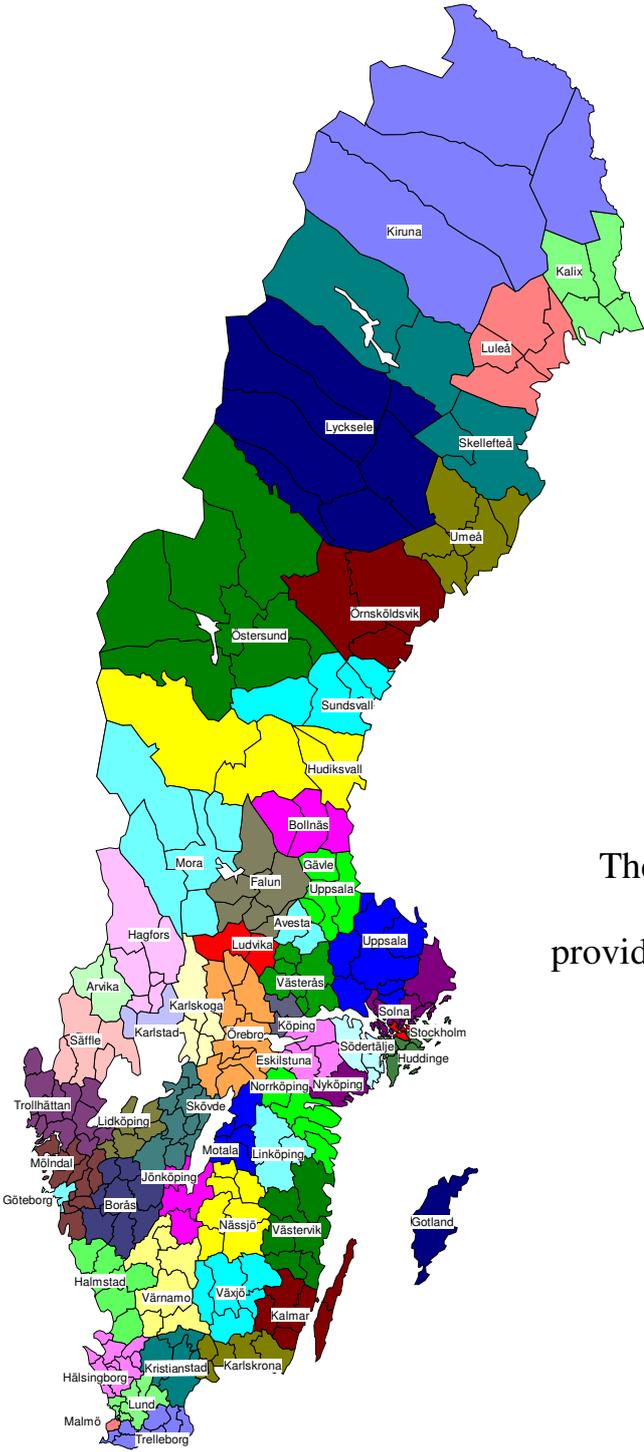


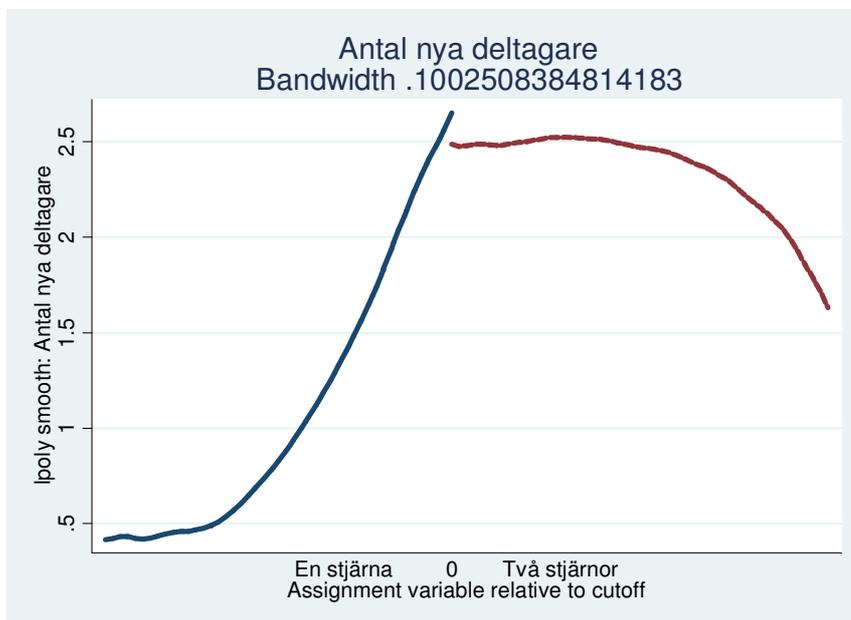
Figure A3



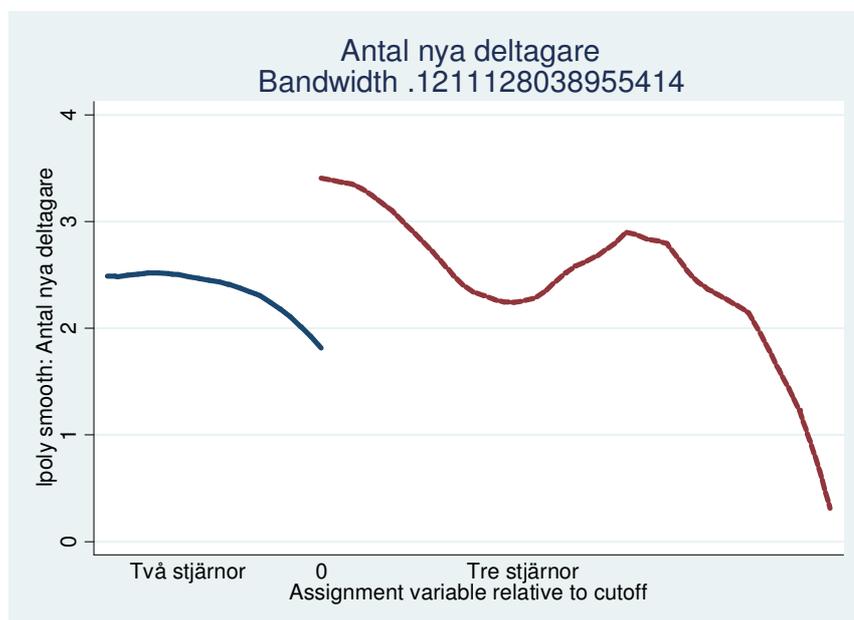
The supply areas where providers of job coaching

## A.2 REGRESSION DISCONTINUITY

One star vs. two stars. Outcome: Number of new participants						
Estimations on the data <i>after</i> publication of the rating values						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	.4655113	.8912675	0.52	<b>0.601</b>	-1.281341	2.212363
lwald50	.739329	1.108242	0.67	<b>0.505</b>	-1.432786	2.911444
lwald200	.532121	.6867992	0.77	<b>0.438</b>	-.8139807	1.878223

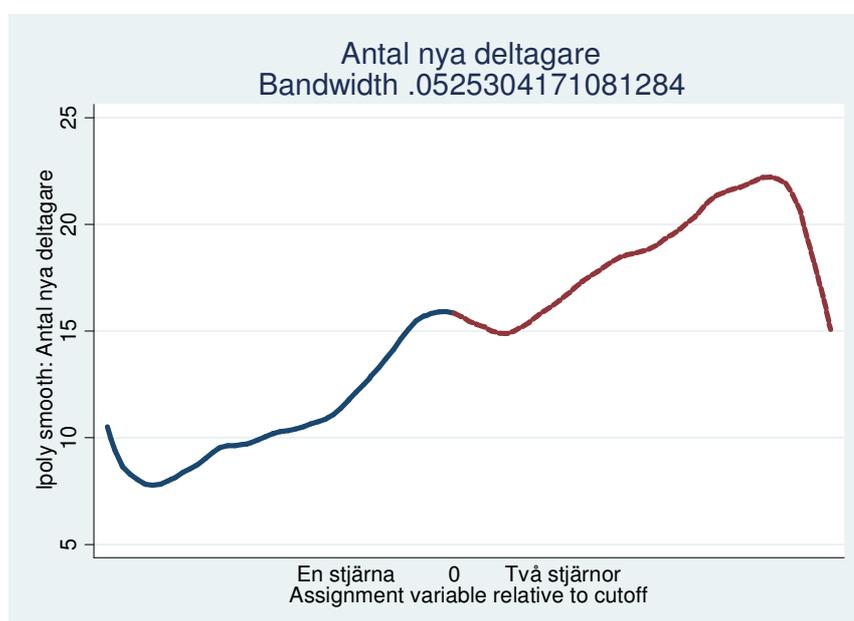


Three stars vs. two stars. Outcome: Number of new participants						
Estimations on the data <i>after</i> publication of the rating values						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	1.830836	.6013351	3.04	0.002	.6522406	3.009431
lwald50	1.591916	.8526606	1.87	0.062	-.0792684	3.2631
lwald200	1.534553	.5095715	3.01	0.003	.5358116	2.533295



### A.3 REGRESSION DISCONTINUITY: SENSITIVITY ANALYSIS

One star vs. two stars. Outcome: Number of new participants						
Estimations on the data <i>before</i> publication of the rating values						
	Coef.	Std. Err.	z	P> z	[95%Conf. interval]	
lwald	1.157894	3.35875	0.34	0.730	-5.425136	7.740923
lwald50	8.149067	6.608037	1.23	0.217	-4.802448	21.10058
lwald200	-1.278655	2.216747	-0.58	0.564	-5.6234	3.066089



Three stars vs. two stars. Outcome: Number of new participants						
Estimations on the data <i>before</i> publication of the rating values						
	Coef.	Std. Err.	z	P> z	[95%Conf. nterval]	
lwald	-5.991643	4.358306	-1.37	0.169	-14.53377	2.55048
lwald50	-3.825027	5.513708	-0.69	0.488	-14.6317	6.981642
lwald200	-4.124864	3.287986	-1.25	0.210	-10.5692	2.31947

